NBS SPECIAL PUBLICATION 617

U.S. DEPARTMENT OF COMMERCE / National Bureau of Standards

Precision Measurement and Fundamental Constants II

Precision Measurement and Fundamental Constants II

Proceedings of the Second International Conference
held at the National Bureau of Standards
Gaithersburg, MD, June 8-12, 1981

Edited by:

B.N. Taylor and W.D. Phillips

Center for Basic Standards
National Measurement Laboratory
National Bureau of Standards
Gaithersburg, MD 20899

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, Secretary
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director

Issued August 1984

Library of Congress Catalog Card Number: 84-601083

National Bureau of Standards Special Publication 617
Natl. Bur. Stand. (U.S.), Spec. Publ. 617, 646 pages (Aug. 1984)
CODEN: XNBSAV

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1984

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402

Abstract

This volume presents the Proceedings of the Second International Conference on Precision Measurement and Fundamental Constants, held at the National Bureau of Standards in Gaithersburg, MD, from June 8–12, 1981. Like its 1970 predecessor, the Conference provided an international forum for theoretical, experimental, and applied scientists actively engaged in research on precision measurements relating to the fundamental physical constants, and on the testing of related fundamental theory. More specifically, the purpose of the Conference was to assess the present state of the precision measurement–fundamental constants field, to examine basic limitations, and to explore the prospects for future significant advances. The principal subjects covered were: frequency, wavelength, spectroscopy, quantum electrodynamics, the gas constant, x-ray interferometry, nuclidic masses, uncertainty assignment, gravitational acceleration, mass, electrical quantities, gravity, and relativity. These proceedings contain the vast majority of both the invited review papers and the contributed current research papers presented at the Conference. The new results reported at the Conference were considered for inclusion in the 1983 least-squares adjustment of the constants carried out under the auspices of the Task Group on Fundamental Constants of the Committee on Data for Science and Technology (CODATA).

Key words: data analysis; experimental tests of theory; fundamental constants; least-squares adjustments; precision measurements.

Preface

Like its 1970 predecessor, the Second International Conference on Precision Measurement and Fundamental Constants (PMFC-II), held at the National Bureau of Standards, Gaithersburg, MD, June 8–12, 1981, provided an international forum for scientists actively engaged in experimental and theoretical research on precision measurements relating to the fundamental physical constants, and on the testing of related fundamental theory. More specifically, the Conference assessed the present state of the precision measurement–fundamental constants (PMFC) field, examined its current limitations, and explored the prospects for future significant advances.

The Conference was unique in that it brought together to an even greater extent than did PMFC-I scientists from fields as apparently diverse as precision mass metrology and quantum field theory. But this of course was the main reason for holding the Conference; that is, its principal purpose was to foster the interchange of information and ideas among researchers who at first glance have little in common but who upon closer scrutiny have much to gain from each other because of their mutual interest in and need for precision measurements and accurate values of the fundamental constants. The reasons for this interest and need were well described in the Preface to the proceedings of PMFC-I by D. N. Langenberg and B. N. Taylor:

> Precision measurement plays an essential and crucial role in the development of all science and technology. Scientific theories cannot be tested, instruments and machines cannot be designed and built, and even the daily routine of the market place cannot proceed without precise quantitative measurement. The demand for precision and accuracy is greatest at the frontiers of science and technology. Here our increasing understanding of the fundamental nature of the universe and our increasing ability to put our knowledge to practical use depends on continuing refinement of our theories and of the technology used in testing these theories against reality.
>
> In all of our fundamental theories there appear a certain few parameters which characterize the fundamental particles and interactions we find in nature. These are the fundamental physical constants. The precision determination of the numerical values of these constants has long been and remains one of the principal objectives of experimental science. This is not because there is any intrinsic virtue in accumulating lists of ever longer numbers, but because the fundamental constants are the quantitative links between our most basic theories and the physical reality we wish them to describe. Our theories must stand or fall according to their ability to make quantitative predictions which agree with experimental observations to the maximum accuracy possible. The progress of our understanding of the physical world is therefore very much intertwined with the advance of the art of precision measurement and its application to the determination of the fundamental physical constants.

To this we must add the idea that because of the interrelationships which exist among the fundamental constants and related quantities, for example, the proton gyromagnetic ratio, the Faraday, Avogadro, and fine-structure constants, $2e/h$ via the Josephson effect, the absolute ohm and ampere, the quantized Hall resistance, the electron and muon anomalous magnetic moments, the proton to electron mass ratio, etc., it is appropriate, indeed even necessary, for the solid state theorist to be aware of the work of the electrical metrologist; the atomic mass spectroscopist of the work of the precision electrochemist; and the quantum electrodynamics theorist of the work of the experimental semiconductor physicist. The fundamental constants of nature and closely related precision measurements are truly the common meeting ground for many of the disciplines of science and technology.

The technical program of the Conference was designed to give participants both a broad and in-depth view of the PMFC field as well as many opportunities to interact with one another. The Keynote Address by Ian W. Roxburgh served to stimulate thinking about the fundamental constants and their relationship to the laws of nature; the invited review talks provided a broad overview of the PMFC field including its current limitations and future possibilities; the orally presented contributed papers gave a more detailed sampling of the current research underway in the field; the five poster sessions (one each day with about 20 papers per session) allowed everyone to hold in-depth discussions of their current research work with interested colleagues; and the informal, off the record, evening session enabled those who were sufficiently brave to present their most highly speculative or "far out" ideas.

Each day of the Conference was devoted to a small number of broad topical areas and each of the papers presented that day—invited review, oral contributed, and poster—more or less dealt with the selected topics. This arrangement has been followed in this volume except for slight modifications in order to improve overall coherence. Thus, included in these Proceedings are the written versions of the Keynote Address, the 21 invited review papers (one in abstract form only), the 16 oral contributed papers (one in abstract form only); and 93 of the 100 poster papers (one in abstract form only). Nothing has been included from the evening session because it was truly "off the record."

A quick perusal of the contents of these Proceedings would convince even the most casual observer that the precision measurement–fundamental constants field is full of vitality. Enormous advances have been made over the last decade or so since PMFC-I. The uncertainty in c, the speed of light, is now limited by the present ^{86}Kr definition of the meter; a definition of length based on c is destined to replace the krypton definition within the next several years. The Rydberg constant is now known to one part in 10^9, the Avogadro constant to one part in 10^6, the fine structure constant to one part in 10^7, the Faraday constant to better than two parts in 10^6, the proton-to-electron mass ratio and proton gyromagnetic ratio to two parts in 10^7, the anomalous moment of the electron to 3 parts in 10^8, and the Newtonian gravitational constant to 6 parts in 10^5. On the theoretical side, significant advances have been made in the theory of the anomalous moment of the electron and muon, Lamb-shift, and muonium hyperfine splitting. Moreover, completely new techniques and phenomena which promise to have a significant impact on the future of the PMFC field have been developed and identified. Especially noteworthy in this regard is the methodology for trapping a single, charged particle pioneered by H. G. Dehmelt and colleagues, and the discovery of the quantized Hall resistance by K. v. Klitzing. This latter phenomenon, which may eventually yield the most accurate value of the fine-structure constant, has already attracted considerable attention since it was first reported in June 1980 as evidenced by the eight papers on the subject presented at the Conference. It could have the same impact on the PMFC field in the 1980's as did the Josephson effect in the 1960's.

The one to two orders of magnitude reduction in the uncertainties of many of the constants which has taken place since PMFC-I continues the trend of the last 20 to 30 years and again raises the question as to whether it can continue indefinitely. Will improvements in and the application of existing technologies such as high-speed computers, lasers, and cryogenics, and the discovery of new techniques and phenomena, allow the proceedings editors of a hypothetical PMFC-III to make the same comment a decade or so hence? We are inclined to believe that those editors will be able to do so, especially if the enthusiasm of the Conference participants was any indication of what lies ahead. In this regard we were especially pleased to see so many in attendance from outside of the United States (103 out of 257 or 40%) and to see the large number of graduate students who are being trained to become the next generation of "precision measurers."

We would like to thank personally the many organizations and individuals who helped to make PMFC-II the highly worthwhile Conference it turned out to be. These include its sponsors, contributors, and the members of the organizing committees as listed on the following pages, as well as many members of the staff of the National Bureau of Standards, especially Joanne Lorden, Greta Pignone, and Kathy Stang.

B. N. Taylor
W. D. Phillips

Conference Organizing Committees

Conference Committee

K. M. Baird (National Research Council, Canada)
E. R. Cohen (Science Center, Rockwell International, U.S.A.)
R. D. Deslattes (National Bureau of Standards, U.S.A.)
J. L. Hall (Joint Institute for Laboratory Astrophysics, U.S.A.)
K. G. Kessler (National Bureau of Standards, U.S.A.)
D. Kleppner (Massachusetts Institute of Technology, U.S.A.)
D. R. Lide (National Bureau of Standards, U.S.A.)
R. C. Ritter (University of Virginia, U.S.A.)
W. G. Schweitzer (National Bureau of Standards, U.S.A.)
B. N. Taylor (Chairman; National Bureau of Standards, U.S.A.)

International Advisory Committee

S. J. Brodsky (Stanford Linear Accelerator Center, U.S.A.)
H. G. Dehmelt (University of Washington, U.S.A.)
V. I. Goldansky (Institute for Chemical Physics, U.S.S.R.)
T. W. Hänsch (Stanford University, U.S.A.)
H. W. Hellwig (Frequency and Time Systems, Inc., U.S.A.)
T. Kinoshita (Cornell University, U.S.A.)
B. Kramer (Physikalisch-Technische Bundesanstalt, F.R.G.)
M. Morimura (National Research Laboratory of Metrology, Japan)
B. W. Petley (National Physical Laboratory, U.K.)
T. G. Quinn (Bureau International des Poids et Mesures, France)
N. F. Ramsey (Harvard University, U.S.A.)
A. Rich (University of Michigan, U.S.A.)

Honorary Committee

E. Ambler (National Bureau of Standards, U.S.A.)
W. R. Blevin (CSIRO Division of Applied Physics, Australia)
L. M. Branscomb (International Business Machines Corporation, U.S.A.)
P. Dean (National Physical Laboratory, U.K.)
A. Ferro-Milone (Institute Elettrotechnico Nazionale "Galileo Ferraris," Italy)
P. Giacomo (Bureau International des Poids et Mesures, France)
Kanjie Ju (National Institute of Metrology, P.R.C.)
M. Kawata (National Institute of Metrology, Japan)
D. Kind (Physikalisch-Technische Bundesanstalt, F.R.G.)
H. Preston-Thomas (National Research Council, Canada)
Y. V. Tarbeyev (Mendelelyev Research Institute of Metrology, U.S.S.R.)
J. Terrien (Bureau International des Poids et Mesures, France)

Sponsors

International Union of
Pure and Applied Physics

Committee on Data for Science and
Technology of the International
Council of Scientific Unions

U.S. National Academy of Sciences—
National Research Council, Committee
on Fundamental Constants

Bureau International des
Poids et Mesures

National Bureau of Standards
U.S. Department of Commerce

Contributors

The Second International Conference on Precision Measurement and Fundamental Constants expresses its deep appreciation to the following organizations for their generous financial support which contributed greatly to making this meeting possible.

Airco, Inc.
Avco Corporation
Ball Technical Products Group
Beckman Instruments, Inc.
Bendix Automation & Measurement Division
COMSAT Laboratories
Corning Glass Works
E.I. duPont de Nemours & Company, Inc.
EG&G Princeton Applied Research Corporation
Electro Scientific Industries, Inc.
Exxon Research and Engineering Company
Fisher Scientific Company
John Fluke Mfg. Co., Inc.
The Foxboro Company
Frequency and Time Systems, Inc.
General Atomic Company
General Electric Company
Guildline Instruments Inc.
Gulf Research & Development Company
Hewlett-Packard Company
IEEE Instrumentation and Measurement Society
International Business Machines Corporation
International Union of Pure and Applied Physics
Leeds & Northrup Company
Matheson Gas Products
Mettler Instrument Corporation
The 3M Company
Mobil Research and Development Corporation
Monsanto Company
Motorola Inc.
National Bureau of Standards
National Conference of Standards Laboratories
National Science Foundation
Rockwell International
Sandia Laboratories
Shell Development Company
Tektronix, Inc.
Texas Instruments, Inc.
U.S. Air Force Geophysics Laboratory
U.S. Army Research Office
U.S. Office of Naval Research
Westinghouse Electric Corporation
Xerox Corporation

Some of the participants in the Second International Conference on Precision Measurement and Fundamental Constants, held June 8–12, 1981, at the National Bureau of Standards, Gaithersburg, MD.

Contents

Page

Preface v

Conference Organizing Committees vii

Conference Sponsors and Contributors viii

Frontispiece ix

Keynote Address: The Laws and Constants of Nature 1
I. W. Roxburgh

Frequency, Wavelength, and Stored Ions

Frequency Standards Based on Magnetic Hyperfine Structure Resonances (Review) 11
H. Hellwig

Performance of the Three NRC 1-Meter CsVI Primary Clocks 21
A. G. Mungall, H. Daams, and J.-S. Boulanger

Prospects for Cesium Primary Standards at the National Bureau of Standards 25
L. L. Lewis, F. L. Walls, and D. A. Howe

A Cesium Beam Atomic Clock with Laser Optical Pumping, as a Potential Frequency Standard 29
M. Arditi

Frequency Measurement of Optical Radiation (Review) 35
K. M. Baird

Optical Frequency Standards: Progress and Applications (Review; abstract only) 43
J. L. Hall

Measurement of Frequency Differences of Up to 170 GHz Between Visible Laser Lines Using Metal-Insulator-Metal Point Contact Diodes 45
H.-U. Daniel, M. Steiner, and H. Walther

Precision Frequency Metrology for Lasers in the Visible and Application to Atomic Hydrogen 49
B. Burghardt, H. Hoeffgen, G. Meisel, W. Reinert, and B. Vowinkel

System for Light Velocity Measurement at NRLM 53
K. Tanaka, T. Sakurai, N. Ito, T. Kurosawa, A. Morinaga, and S. Iwasaki

Laser Wavelength Measurements and Standards for the Determination of Length (Review) ... 57
W. R. C. Rowley

Double-Mode Method of Sub-Doppler Spectroscopy and Its Application in Laser Frequency Stabilization 65
N. G. Basov, M. A. Gubin, V. V. Nikitin, A. V. Nikulchin, D. A. Tyruikov, V. N. Petrovskiy, and E. D. Protscenko

He-Ne ($^{127}I_2$) Lasers at 0.633 μm (and at 0.604 μm) 69
F. Bertinetto, B. I. Rebaglia, P. Cordiale, S. Fontana, and G. B. Picotto

Recent Work on 612 nm He-Ne Stabilized Lasers 73
A. Brillet, P. Cérez, and C. N. Man-Pichot

Iodine and Methane Stabilized He-Ne Lasers as Wavelength Standards 77
Shen Nai-cheng, Wu Yao-xiang, Sun Yi-min, Li Cheng-yang, Zhang Xue-bin, and Wang Chu

Page

Spatial Coherence and Optical Wavelength Metrology 81
P. Bouchareine

Spectroscopy of Stored Ions (Review) 83
D. J. Wineland

Progress Toward a Stored Ion Frequency Standard at the National Bureau of Standards 93
W. M. Itano, D. J. Wineland, J. C. Bergquist, and F. L. Walls

High Resolution Microwave Spectroscopy on Trapped Ba^+ Ions 99
W. Becker, R. Blatt, and G. Werth

Observation of High Order Side Bands in the Spectrum of Stored $^3He^+$ Ions 103
H. A. Schuessler and H. S. Lakkaraju

Quantum Limits in the Measurements of e.m. Fields and Frequency (Abstract only) 109
V. B. Braginsky

Spectroscopy, Quantum Electrodynamics, and Elementary Particles

Precision Laser Spectroscopy (Review) 111
T. W. Hänsch

Atomic Beam, Linear, Single-Photon Measurement of the Rydberg Constant 117
S. R. Amin, C. D. Caldwell, and W. Lichten

Current Work on Two Photon Excitation in a Hydrogen Beam for the Measurement of the Rydberg Constant and m_e/m_p 123
D. Shiner and C. Wieman

Measurement of the $2^2P_{3/2}$ - $2^2S_{1/2}$ Fine-Structure Interval in Atomic Hydrogen 127
K. A. Safinya, K. K. Chan, S. R. Lundeen, and F. M. Pipkin

Measurement of the Lamb Shift in Hydrogen, $n = 2$ 131
S. R. Lundeen and F. M. Pipkin

Atomic Interferometer Method Measurement of the Lamb Shift in Hydrogen ($n = 2$) 135
Y. L. Sokolov

Measurement of the $4^2S_{1/2}$ - $4^2P_{1/2}$ Lamb Shift in He^+ 141
J. J. Bollinger, S. R. Lundeen, and F. M. Pipkin

Lamb Shift in the Hydrogenic Ion Cl^{16+} 145
E. T. Nelson, O. R. Wood II, C. K. N. Patel, M. Leventhal, D. E. Murnick, H. W. Kugel, and Y. Niv

Helium Fine Structure and the Fine Structure Constant 149
W. Frieze, E. A. Hinds, A. Kponou, V. W. Hughes, and F. M. J. Pichanick

Preliminary Measurement of the $J = 0$ to $J = 2$ Fine Structure Interval in the $3\,^3P$ State of Helium 153
M. Feldman, T. Breeden, L. DiMauro, T. Dong, and H. Metcalf

Laser Microwave Precision Measurements of $2\,^3S_1$ and $2\,^3P$ Term Splittings in Helium-Like Li^+ 159
U. Kötz, J. Kowalski, R. Neumann, S. Noehte, H. Suhr, K. Winkler, and G. zu Putlitz

An Optically Pumped Metastable Hydrogen Beam 163
K. C. Harvey

Precise Determination of the S and P Quantum Defects in Sodium and Cesium by Millimeter and Submillimeter Spectroscopy Between Rydberg States 169
P. Goy, J. M. Raimond, G. Vitrant, C. Fabre, S. Haroche, and M. Gross

A New Method of Measuring the Fine-Structure Constant Using Stark Spectroscopy 173
M. G. Littman and W. D. Phillips

Time Resolved Sub-Natural Width Spectroscopy 177
W. D. Phillips and H. J. Metcalf

Page

The Spectroscopy of Atoms and Molecules in Gases: Corrections to the Doppler-Recoil Shift 181
M. P. Haugan and F. V. Kowalski

The Implications of QED Theory for the Fundamental Constants (Review) 185
G. P. Lepage and D. R. Yennie

Uncertainties in QED Fine Structure Calculations 195
G. W. Erickson

Sixth Order Contributions to g-2 of the Electron 201
M. J. Levine and R. Z. Roskies

Calculation of the Eighth Order Anomalous Magnetic Moment of the Electron 203
T. Kinoshita and W. B. Lindquist

Experimental Determinations of the Anomalous Magnetic Moments of the Free Leptons (Review) 207
R. Conti, D. Newman, A. Rich, and E. Sweetman

Preliminary Comparison of the Positron and Electron Spin Anomalies 215
P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt

Geonium Without a Magnetic Bottle—A New Generation 219
G. Gabrielse and H. Dehmelt

The Magnetic Moment of Positive Muons in Units of the Proton Magnetic Moment 223
E. Klempt, R. Schulze, H. Wolf, M. Camani, F. N. Gygax, W. Rüegg, A. Schenck, and H. Schilling

Measured G_J-Factor Ratio of $^4He^+(1\ ^2S_{1/2})$ and $^4He(2\ ^3S_1)$ 229
H. G. Robinson and C. E. Johnson

Determination of the Neutron Magnetic Moment 233
G. L. Greene, N. F. Ramsey, W. Mampe, J. M. Pendlebury, K. Smith, W. B. Dress, P. D. Miller, and P. Perrin

Precision Exotic Atom Spectroscopy (Review) 237
V. W. Hughes

Fundamental Tests and Measures of the Structure of Matter at Short Distances (Review) 249
S. J. Brodsky

Experimental Limit for the Charge of the Free Neutron 257
R. Gähler, J. Kalus, and W. Mampe

Gas Constant, X-Ray Interferometry, Nuclidic Masses, Other Constants, and Uncertainty Assignment

Methods for the Determination of the Gas Constant (Review) 263
A. R. Colclough

An Ultrasonic Determination of the Gas Constant 277
W. C. Sauder

Spherical Acoustic Resonators: Promising Tools for Thermometry and Measurement of the Gas Constant 281
M. R. Moldover and J. B. Mehl

Proposed Method for the Determination of the Molar Gas Constant, R 287
L. A. Guildner and M. L. Reilly

A Radiometric Determination of the Stefan-Boltzmann Constant 291
T. J. Quinn and J. E. Martin

On the Radiometric Measurement of the Stefan-Boltzmann Constant at NRLM 299
A. Ono

Applications of X-ray Interferometry (Review) 303
R. D. Deslattes

Work Related to the Determination of the Avogadro Constant in the PTB 313
P. Seyfried

Page

Absolute Determination of the (220)-Lattice Spacing in Silicon 317
P. Becker and H. Siegert

Systematic Uncertainties in the Determination of the Lattice Spacing $d(220)$ in Silicon 321
H. Siegert and P. Becker

High Precision Studies of Pionic X Rays: Some Past Results and Future Prospects 325
G. Dugan, L. Delker, C. S. Wu, and D. C. Lu

Isotope Shifts of K X-Rays of Lead 331
G. L. Borchert, O. W. B. Schult, J. Speth, P. G. Hansen, B. Jonson, H. Ravn, and J. B. McGrory

The Measurement of Atomic Masses by Mass Spectroscopic Methods and a Role for Atomic Masses in the Determination of the Fundamental Constants (Review) 335
W. H. Johnson

Absolute Determination of the Threshold Energies of ^{7}Li(p,n), ^{10}B(p,n), and ^{14}N(p,n) 345
P. H. Barker, M. J. Lovelock, H. Naylor, R. M. Smythe, and R. E. White

Preliminary Proton/Electron Mass Ratio Using a Precision Mass-Ratio Spectrometer 349
R. S. Van Dyck, Jr. and P. B. Schwinberg

A Direct Determination of the Proton-Electron Mass Ratio 353
G. Gräff, H. Kalinowsky, and J. Traut

A New Determination of the Atomic Weight of Silver and an Improved Value for the Faraday 357
L. J. Powell, T. J. Murphy, and J. W. Gramlich

High Resolution Magnetic Measurements on Rotating Superconductors to Determine h/m_e 359
B. Cabrera, S. B. Felch, and J. T. Anderson

High Precision Measurement of the Electron Compton Wavelength (h/m_e) Using Cryogenic Metrological Techniques 365
J. C. Gallop, B. W. Petley, and W. J. Radcliffe

Preliminary Determination of h/m_n 369
E. Krüger, W. Nistler, and W. Weirauch

The Assignment of Uncertainties to the Results of Experimental Measurements (Review) 375
J. W. Müller

On the Statement of a Total Confidence Interval Based on the Concept of Randomization of Systematic Errors: Large and Small Sample Sizes 383
W. Wöger

Measurement Assurance 385
A. F. Dunn

An Extended-Least-Squares Treatment of Discrepant Data 391
E. R. Cohen

Gravitational Acceleration, Mass, and Electrical Quantities

Present Status of the Absolute Measurement of Gravitational Acceleration (Review) 397
A. Sakuma

A New, Portable, Absolute Gravimeter 405
M. A. Zumberge, J. E. Faller, and R. L. Rinker

"Super Spring"—A Long Period Vibration Isolator 411
R. L. Rinker and J. E. Faller

Transportable Gravimeter for the Absolute Determination of Gravity 419
Guo You-guang, Huang Da-lun, Li De-xi, Zhang Guang-yuan, Gao Jing-long, Fang Yong-yuan, and Huang Cheng-qing

New Techniques for Absolute Gravity Measurement 423
J. A. Hammond, R. L. Hiff, and R. W. Sands

Page

The Mass Unit "Kilogram," Precision Measurement of Mass, Attainable Uncertainties, and Possibilities of a New Definition (Review) 427
M. Kochsiek

Measurement of Air Density for High Accuracy Mass Determination 437
D. B. Prowse

On a More Precise Correction for Buoyancy and Gas Adsorption in Mass Measurement 441
Y. Kobayashi

Density Standards—The Density and Thermal Dilatation of Water 445
G. A. Bell and J. B. Patterson

Precision Measurements on Solid Artifacts for a Redetermination of the Density of Water 449
A. Peuto, A. Sacconi, R. Panciera, W. Pasin, and M. Rasetti

The Influence of Dissolved Air on the Density of Water 453
G. Girard and M.-J. Coarasa

The Helium Melting Curve and the Linkage of Fundamental Constants, Pressure, Density, and Mass 457
C. T. Van Degrift

Realization of the Electrical SI Units (Review) 461
B. P. Kibble

An Absolute Determination of the Volt at LCIE 465
N. Elnékavé and A. Fau

The CSIRO Absolute Volt Project 469
G. J. Sloggett, W. K. Clothier, D. J. Benjamin, M. F. Currey, and H. Bairnsfather

Status of the Measurement of the NBS Ampere in SI Units 475
P. T. Olsen, W. D. Phillips, and E. R. Williams

A Feasibility Study of an Absolute Determination of the Magnetic Flux Quantum 479
K. Hara, F. Shiota, and T. Kubota

The Work Done at the Mendeleyev Research Institute of Metrology (VNIIM) To Improve the Values of the Fundamental Constants 483
Y. V. Tarbeyev

Realization of a Josephson Potentiometer 489
M. Koyanagi, T. Endo, and A. Nakamura

A Transportable Josephson Voltage Standard 493
K. Lahdenperä, H. Seppä, and P. Wallin

The Proton Gyromagnetic Ratio in H_2O—A Problem in Dimensional Metrology (Review) 497
E. R. Williams, P. T. Olsen, and W. D. Phillips

The Development of Precision Measurement and Fundamental Constants in China 505
Wang Zhu-xi

The γ_p-Experiment at PTB 509
K. Weyand

A New Method for the Determination of the Proton Gyromagnetic Ratio 515
G. L. Greene

The Quantized Hall Resistance in Two-Dimensional Systems (Review) 519
K. von Klitzing, H. Obloh, G. Ebert, J. Knecht, and K. Ploog

Hall Effect in Silicon MOS Inversion Layers for h/e^2 Determination 529
C. Yamanouchi, K. Yoshihiro, J. Kinoshita, K. Inagaki, J. Moriyama, S. Baba, S. Kawaji, K. Murakami, T. Igarashi, T. Endo, M. Koyanagi, and A. Nakamura

Cryogenic Method for the Determination of the Fine-Structure Constant by the Quantized Hall Resistance 535
E. Braun, P. Gutmann, G. Hein, F. Melchert, P. Warnecke, S. Q. Zue, and K. v. Klitzing

Page

Status of the NBS-NRL Determination of the Fine-Structure Constant Using the Quantized Hall Resistance Effect 539
M. E. Cage, R. F. Dziuba, B. F. Field, C. F. Lavine, and R. J. Wagner

Use of a Cryogenic Current Comparator to Determine the Quantized Hall Resistance in a Silicon MOSFET 543
A. Hartland

A Resistance Standard Using the Quantized Hall Resistance of GaAs-$Al_xGa_{1-x}As$ Heterostructures 549
A. C. Gossard and D. C. Tsui

A Quantitative Theory for the Determination of h/e^2 from the Hall Effect in Two-Dimensional Conductors 553
L. Bliek and G. Hein

Quantum Hall Effect: Role of Inversion Layer Geometry and Random Impurity Potential 557
R. W. Rendell and S. M. Girvin

Gravity and Relativity

Experiments Relating to the Newtonian Gravitational Constant (Review) 561
H. de Boer

Redetermination of the Newtonian Gravitational Constant "G" 573
G. G. Luther and W. R. Towler

The Design of a Beam Balance for a Determination of G 577
C. C. Speake and A. J. F. Metherell

Optimizing the Shape of the Attracting Mass in Precision Measurements of G 581
A. J. F. Metherell, C. C. Speake, Y. T. Chen, and J. E. Faller

Vacuum Polarization and Recent Measurements of the Gravitational Constant as a Function of Mass Separation 587
D. R. Long

Tests of the Gravitational Inverse Square Law Using Torsion Balances 591
J. K. Hoskins, R. Newman, J. Schultz, and R. Spero

Measurement of Gravitational Forces at Separations Around 10 Meters 595
H. Yu, W. Ni, C. Hu, F. Liu, C. Yang, and W. Liu

Non-Newtonian Gravity: Geophysical Evidence 597
F. D. Stacey and G. J. Tuck

Experimental Test of a Spatial Variation of the Newtonian Gravitational Constant at Large Distances 601
H. A. Chan and H. J. Paik

The Measurement of G for Small Inter-Mass Spacings 607
W. C. Oelfke

Tests of Gravitation and Relativity (Review) 611
R. F. C. Vessot

Is the Gravitational Constant Changing? 625
T. C. Van Flandern

Experiments on Variation of the Gravitational Constant Using Precision Rotations 629
G. T. Gillies and R. C. Ritter

Interpreting Dirac's Large Numbers Hypothesis 635
W. Davidson

New Laboratory Test of the Equivalence Principle 639
P. T. Keyser, J. E. Faller, and K. H. McLagan

What Test Masses Are Best for an Eötvös Experiment? 643
D. F. Bartlett, J. Shepard, and C. D. Zafiratos

Page

Equivalence Principles and Precision Measurements 647
W. Ni

An Inertial Clock to Test the Non-Metricity of Gravity 653
W. S. Cheung and R. C. Ritter

Assessment of the Prospects for a Measurement of Relativistic Frame Dragging by 1990 659
R. A. Van Patten

The Status of the Velocity of Light in Special Relativity 667
E. Breitenberger

Relativistic Time Dilation: A Latter-Day Ives-Stillwell Experiment 671
P. Nachman, M. D. Rayman, and J. L. Hall

An Experiment to Measure Relative Variations in the One-Way Velocity of Light 675
D. G. Torr and P. Kolen

High-Energy Gamma Rays Might Be Faster than Visible Light 681
K. Fujiwara

Fiber Optic Ring as a Gravitational Wave Detector 685
C. L. Mehta, D. Ranganathan, and G. Bose

A Possible Laser Gravitational Wave Experiment in Space (Abstract only) 689
J. E. Faller and P. L. Bender

Author Index 691

Conference Registrants 695

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

KEYNOTE ADDRESS

The Laws and Constants of Nature

Ian W. Roxburgh

Department of Applied Mathematics, Queen Mary College, University of London

The paper concentrates on a few problems; geometry, mechanics, gravitation, and the large numbers (10^{40}) that relate microphysics to the large scale structure of the universe. My purpose is not so much to describe what is known, but to question how well anything is known and to provoke the reader into asking questions and proposing experiments that probe the foundations of our understanding. Is geometry locally Euclidean? Do different clocks keep the same time? Does our existence depend on the exact form of the laws of nature and the exact values of the constants of nature? Do the constants of nature vary in time? Why are there laws at all? Readers can, and I hope will, add questions of their own.

Key words: anthropic principle; constants of nature; fundamental natural laws; gravitation; large number hypothesis; time variation of constants.

1. Introduction

Science as we know it probably began with man's attempts to understand the regularity of nature as seen in the rising and setting of the sun, the phases of the moon, the annual recurrence of the seasons, and the motion of the planets across the sky. Indeed, the Shu Ching, the ancient Chinese Book of History, opens in the Cannon of Yao with the sovereign giving instructions [1]:

> "He commanded the brothers Hsi and the brothers Ho, in reverent accord with their observations of the wide heavens, to calculate and delineate the movements of the sun, the moon, the stars and the zodiacal spaces and so to deliver respectfully the seasons to be observed by the people."

Science in those days was a dangerous activity; we later learn that the astronomers Hsi and Ho were executed for failing to predict a solar eclipse. If the supporting details are correct (the moon was in Sieou Fang in the center of Tsing Lung on the first day of autumn) this was in the year 2137 B.C.

Whilst we do not know how the astronomers could have predicted eclipses at that time, we learn from the 'Chou pi' of the 'Kai-T'ien' model of the motion of the sun [2]. The 'Chou pi' probably dates from about 1100 B.C., but it is possible that the cosmological model was a later addition either indigenous to China or imported from India where a very similar model is described in the Suryaprajnapti (circa 500 B.C.) [3]. The model has the sun describing a set of seven circles about the celestial pole at a height of 80,000 li above a flat earth; the innermost circle has a diameter of 238,000 li, the outermost circle a diameter of 476,000 li. The sun moves on the inner circle at summer solstice and moves from circle to circle during the course of six months until at winter solstice it is on the outermost circle. The sun's rays only extend a distance of 167,000 li so explaining night and day.

This model is quite good. Indeed, one can compute that at a latitude of 36° the model correctly predicts the elevation angles of the sun at noon on both midsummer's and midwinter's day. But is it Science? Indeed, what is Science?

The ancient Chinese were clearly aware of the regularity of the motion of the sun across the sky and of the seasons. They constructed a model based on these observations, the model explained past observations (at least to some level of accuracy) and successfully predicted future observations. What more do we require before giving such a model the accolade of being scientific? Is not this model as scientific as Kepler's Laws?

The ancient Chinese scientists had made that first step in science of believing that the future could be predicted on the basis of past experience. This is the act of faith we make in science although as pointed out by Hume, it is an act of faith and we cannot on the basis of past experience prove that our predictions will be true in the future [4]. The Chinese model was quantitative and not just qualitative. It was narrow, applying only to one set of phenomena, so it lacked the generality we associate with scientific reasoning. But I nevertheless think we should describe these Chinese atronomers as early scientists.

2. Geometry as an Example of Science

A better example of early science is geometry. The origins of geometrical knowledge also go very far back in the history of mankind; some special laws such as Pythagoras' Theorem in the form $6^2 + 8^2 = 10^2$ were known to the Chinese. The fact that the circumference of a circle is a constant multiple of its diameter was also

known to several ancient civilizations, as was the existence of similar triangles. However, the main body of geometrical knowledge was accumulated by the Greeks and it is to this culture that I turn for my discussion on geometry as a science.

Let us take a few simple geometrical results and ask what they tell us about the world. Three such examples are:

(a) The sum of the angles of any triangle equals two right angles.

(b) In a right angled triangle the sum of the squares of the lengths of the two smallest sides equals the square of the length of the hypotenuse.

(c) The circumference of any circle is a constant multiple of its diameter.

These results can properly be considered as laws of nature. They are found experimentally by drawing triangles and circles and measuring lengths and angles. They tell us quite a lot about the world we live in: The results are the same no matter where we carry out the experiments; the laws are true for all orientations of triangles; they were true in 400 B.C. and are true today; in order to conduct such experiments we need to define standards and measuring procedures ..... These laws are general, that is when stating them we are not only saying that these laws were satisfied for all triangles and circles that have been measured, but that they are true for *all* triangles and circles. We have made that inductive leap from the particular to the general that is characteristic of science and are prepared to use these laws to predict the outcome of future experiments. In doing so the ancient Greeks were accepting that nature satisfies 'laws of nature' that can be uncovered by experiment.

But the contribution of the Greeks does not stop at the discovery of such laws. Their main contribution to geometry was to recognize that all these empirical laws could be deduced from a small set of fundamental laws: the axioms of Euclidean geometry together with definitions of the objects that enter those laws and the rules of logic that we use in deductive reasoning. Examples of these latter two categories are [5]:

> A straight line is a line which lies evenly with the points on itself.
>
> Things which are equal to the same thing are equal to each other.

The axioms or fundamental laws as defined by Euclid are [5]:

(1) It is possible to draw a straight line from any point to any point.

(2) It is possible to extend a finite straight line continuously in a straight line.

(3) It is possible to describe a circle with any center and any radius.

(4) All right angles are equal to one another.

(5) If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are less than two right angles.

These fundamental laws clearly represent an extrapolation from experience. Our experience is finite. Within this finite domain we find these laws to be satisfied (to some level of accuracy!) so we postulate that they can be generalized to the above. Given these fundamental laws all the other laws of geometry can be deduced by logical (i.e., mathematical) reasoning. This is just what we try to do in modern physics. We find laws applicable to different sets of phenomena and then seek more fundamental laws from which this variety of subsidiary laws can be deduced.

3. The First Constant of Nature?

Let me return to one of the early geometrical laws:

"The circumference of any circle is a constant multiple of the diameter."

What is this constant multiple, and how can it be determined? The multiple is given the label π and we determine it by measurement. If I take a circle and some measuring device, for example my wife's tape measure (borrowed from her sewing box!) I can first of all test that it is a circle by measuring the diameter at different orientations. I always find the same value, say 8 inches. I then measure the circumference and find that it is say 25 inches. The ratio is $\pi = 25/8 = 3.125$. I repeat the experiment in London and in Washington, and I find more or less the same answer. I do the same experiment with different size circles and I find more or less the same answer. I have therefore determined the first known constant of nature as $\pi = 3.125$.

Such a simple experiment raises several interesting questions: First one might ask about the measuring apparatus. A plastic tape measure is not perhaps the most accurate of measuring rods and I doubt if it would compare very favorable with the standard definition of length whether by reference to the old standard meter or to the new standard using the velocity of light. Yet it has some advantages. I can fold it back on itself and compare the one inch intervals on one part of the tape with those on another part; they seem to agree. Thus the subdivisions may not be equal to a standard inch but they are equal to each other. But since the value of π is given by the ratio of the circumference to the diameter it does not matter whether the scale on my measure agrees with the standard or not; the ratio of lengths and therefore the value I find for π is only dependent on an accurate division of my measuring tape. This independence of the definition of a measuring standard is not confined to the determination of π; the fine structure constant, or the ratio of proton to electron masses are pure numbers and therefore independent of the standard units of mass, length, charge, or time. Units are only an intermediary that enable us to attach labels of mass, length, etc., to other quantities, but the constants of nature are really pure numbers, independent of units. This is a point I will return to later on in this article.

Again this simple experiment can teach us something about experimental procedures. If I repeat the experiment at different times on circles of different dimensions and at different places I do not get exactly the same result, only 'more or less' the same result. Thus a sequence of five measurements may give the values

3.125 3.163 3.142 3.120 3.157 .....

I now recognize that one measurement does not give an exact result so I take an average and obtain $\pi = 3.1414$, and believe that this is a more accurate estimate of π. Indeed, if I follow the standard practice of estimating the uncertainty to be attached to this value I would then calculate the standard deviation to be 0.01695 and give the experimental determination of π as

$$\pi = 3.141 \pm 0.017 .$$

It is important to realize that in following the above procedure I have made an act of faith, namely that there

is a unique value of π that can be determined to some accuracy by taking many measurements. I have ruled out the possibility that π is really a stochastic quantity and that my measurements reflect such a fact about nature. On the basis of my experimental determination of π I cannot justify such an assumption. The same problem necessarily arises with all measurements, for example, why should not the fine structure constant be a stochastic quantity? It is true that past experience shows that as we make more accurate experiments the latest determinations do not show the earlier fluctuations so we deduce that these earlier fluctuations were experimental error. But if we then claim that this experience proves the present set of values do not reflect the stochastic nature of the world, we are again up against Hume's argument that we cannot prove that our laws will hold true in the future just because they held good in the past.

Of course if we now wish to evaluate π we would not do it by an experiment of the kind I described; we would calculate it mathematically. For example we might use the power series for $\tan^{-1}x$ obtained from the integral obtained by Newton, Leibniz, and Gregory [6]:

$$\tan^{-1}x = \int_0^x \frac{dy}{1+y^2} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \ldots\ldots$$

Taking $x = 1$ gives the Leibniz expression [6]

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \ldots\ldots\ldots$$

This result, though elegant in relating π to simple fractions, is of no practical use. A much better formula is that obtained by Machin [7] which is

$$\frac{\pi}{4} = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239}$$

$$= 4\left[\frac{1}{5} - \frac{1}{3}\frac{1}{5^3} + \frac{1}{5}\frac{1}{5^5} - \frac{1}{7}\frac{1}{5^7} + \frac{1}{9}\frac{1}{5^9} - \ldots\ldots\right]$$

$$- \left[\frac{1}{239} - \frac{1}{3}\frac{1}{239^3} + \ldots\ldots\right]$$

Thus π can be determined to any required accuracy; it is no longer a constant of physics but a constant of mathematics! What does this mean? If a physical constant like the ratio of the circumference to the diameter of a circle can be determined by mathematics, why not other physical constants like the fine structure constant, the mass ratios of elementary particles . . . . Is physics reducible to mathematics?

4. The Relationship Between Mathematics and Physics

The mathematical reasoning that led to the power series for $\tan^{-1}x$ and hence for π is part of Euclidean geometry; that is, provided the fundamental laws, or Euclid's axioms, are true, then the ratio of the circumference of a circle to its diameter is calculable from the above series. But what does it mean to ask if the axioms of a branch of mathematics are true? All that we ask in mathematics is that the axioms be consistent, that is they do not lead to contradictory conclusions. Mathematics is not true or false, it is simply a procedure for deducing the consequences of a particular set of axioms. What we are doing is to ask whether the axiom set of Euclidean geometry is an appropriate mathematical model for the physical world, so that the theorems of Euclidean geometry can be taken over into statements about physical geometry. This clearly requires us to define the objects we call lines and angles in a physical way. Whether these definitions then satisfy the axioms of Euclidean geometry is an empirical question that can only be settled by experiment.

There are several definitions of a straight line that conform to everyday concepts. One is the path of a light ray. Indeed, it is possible that this is what was behind Euclid's statement that "a straight line is a line which lies evenly with the points on itself" but it is not the only definition. Another is the geodesic definition that a straight line is the shortest distance between two points; that is, of all the possible paths joining two points the straight line is that which requires the minimum number of measuring rods laid end to end to cover the distance from one point to another. A right angle can be defined as one quarter of a revolution such that four such angles which coincide when laid one on top of another constitute a complete revolution when laid side by side. An alternative definition is that it is the angle between two straight lines such that the distance from a point on one line to a point on the other is a minimum. Whether these alternative definitions are equivalent is again an empirical issue, and having chosen one definition it is again an empirical issue as to whether these physical definitions have the same properties as the lines and angles in Euclidean geometry. Of course we can only verify that they do to a certain level of accuracy, so whilst our mathematical deductions from the axioms are correct we cannot be sure that they correspond to consequences for the physical behavior of objects.

There is a philosophical point here of some substance as was emphasized by Poincaré [8]. Suppose we found that our physical geometry did not agree with Euclidean geometry. How are we to interpret such a finding? One possible route is to stick to our definition of a straight line as, say, the path of a light ray, in which case the geometry of light rays is non-Euclidean. On the other hand we could say that light does not follow straight lines since they do not satisfy Euclidean geometry. The physicist would then seek to explain this departure from Euclidean geometry by saying that light rays are bent by the presence of other objects in the universe. Within present day science the situation is somewhat confused. On the one hand we talk about geodesics and curved space—on the other hand we talk about the bending of light, that is bending compared to a Euclidean straight line. This illustrates Poincaré's conventionality thesis, subsequently elaborated by Reichenbach and other philosophers of science, that there are many ways of interpreting a given physical situation depending on the definitions used. Contrary to oft repeated claims, there is no information conveyed in the statement "space is non Euclidean." Our knowledge is of the behavior of objects, not space. It may be convenient to talk of light as defining straight lines and that the light geometry is non Euclidean, but it is no more (or less) correct than to claim the straight lines are those that satisfy Euclidean geometry and that light rays are bent.

Yet another lesson to be learnt from geometry is that there is no clear division between what we consider as axioms or fundamental laws, and what we consider as theorems or deductions from those laws. For example, the axioms of Euclid can be used to prove Pythagoras' theorem, or alternatively, Pythagoras' theorem can be taken as the fifth axiom and then the statement that straight lines meet if produced indefinitely becomes a theorem. Alternatively, the fifth axiom could be taken to be the statement that similar triangles exist, or the sum of the angles of a triangle is half a revolution . . . . . One

man's axiom is another man's theorem! So it is in the other branches of science; we choose to take particular statements as fundamental laws and others as deductions from those laws but laws and consequences can be interchanged and we have no way of establishing one set of laws as 'the laws of nature'.

5. The Truth of Euclidean Geometry

Even in Greek times there was some unease about the fifth axiom and attempts were made either to replace it by some more self-evidently true axiom or to deduce it from the other four. The best known attempt was by Ptolomy who tried to deduce the parallel axiom from the first four of Euclid's axioms together with the first twenty eight theorems derived by Euclid which are independent of the fifth axiom, but in this as in other attempts some other assertion entered the argument and effectively replaced the fifth axiom.

However, there does not seem to have been much doubt that Euclidean geometry was indeed the physical geometry of the world. Indeed, it was so firmly held to be true that attempts to demonstrate the validity of other branches of mathematics used the model of Euclidean geometry. Almost all philosophers, mathematicians, and scientists held Euclidean geometry to be necessarily true. Isaac Barrow (Newton's predecessor at Cambridge) held that this truth was derived from innate reason. Newton, Leibniz, Lock, and Hobbes held that Euclidean geometry was inherent in the design of the universe. One notable exception was David Hume, to whose ideas I have already drawn attention. Hume argued both against the necessity of laws and contended that no amount of past experience could demonstrate that the future will obey such laws; knowledge is empirical and the laws of Euclidean geometry, like other laws, are not necessary truths.

On the other hand Immanuel Kant maintained that Euclidean geometry is an *a priori* synthetic truth, that is (roughly interpreted) that our minds compel us to organize our experiences around certain intuitions of space and time and that Euclidean geometry was one such *a priori* synthetic truth. So pervasive was Kantian philosophy that even when Gauss had convinced himself that non-Euclidean geometry was possible he states in a letter to Bessel that he would not publish his findings because he feared ridicule.

As with many of the discoveries of science and mathematics there is no one person that can be called the discoverer of non-Euclidean geometry. Lambert reached the conclusion that there could be many geometries derived from different axioms but it was Gauss who recognized that the non-Euclidean geometry could apply to the physical world; in a letter to Olbers written in 1817 he concluded [9] " . . . we must place geometry not in the same class as arithmetic which is purely *a priori* but with mechanics . . ."

This empirical approach to geometry led to Gauss surveying a triangle formed by the mountain peaks Brocken, Hohenhagen, and Inselsberg to determine the sum of the angles of a triangle [10]. In fact the result obtained was 180° 14.85'' but Gauss realized that the error exceeded 14.85'' so that the correct result could have been 180° or less, and that it was any way unlikely that we could find a positive result with such a small triangle (the sides were 69, 85, and 197 km). However, at least the distinction had been made between physical and mathematical knowledge.[1]

6. Newtonian Gravitation and Mechanics

I now wish to examine the laws of mechanics and gravitation by which Newton and his successors were able to account for the motion of planets, the rate of pendula clocks, the motion of projectiles ... According to Newton's laws the motion of a body of mass m_1 under the gravitational attraction of a second body of mass m_2 is given by [11]

$$\frac{d^2\mathbf{r}_{12}}{dt^2} = \frac{-G(m_1 + m_2)}{r_{12}^3}\,\mathbf{r}_{12}$$

where $\mathbf{r}_{12}$ is the (vector) distance between the two masses, G is the Newtonian constant of gravitation, a universal constant of nature to be determined by experiment [11].

This constant G is of a different kind to π. It has dimensions and therefore its value depends on the units in which it is measured. The quantity π on the other hand is a pure number; its value is unit independent. This should make us ask what is meant by calling G a constant of nature.

Let us examine carefully the content of Newton's law. It contains the symbols r for length, t for time, m for mass. What do these symbols mean? How are mass, length, and time to be measured?

Firstly, the laws only apply in an 'inertial frame', that is in a system of reference in which freely moving particles continue with constant velocity. Yet to know whether particles have a constant velocity presupposes a measure of time and distance! Suppose we use a rigid rod to measure distance. We can now define a measure of inertial time such that a freely moving particle covers equal distances in what we define as equal times. This particle, of course, necessarily covers equal distances in equal times. It then becomes a matter of empirical investigation as to whether other particles also cover equal distances in equal times as measured by that inertial clock; Newtonian mechanics is based on the hypothesis that such frames of reference do exist in nature. In practice time was not measured using freely moving particles, but either by the rotation of the earth around its axis, or some gravitational clock such as a pendulum. Do these clocks keep Newtonian inertial time?

It is an elementary deduction from Newton's laws that the total angular momentum of an isolated system of interacting particles does not change in inertial time. Thus the angular velocity of a rigid body is constant in inertial time so the rotation of the earth provides a tolerably good clock. However, it is not self evident that Newton's laws are true. We could imagine a world in which the angular momentum of an 'isolated' body changed (slowly) as measured in inertial time. Whether or not this is the case is an empirical question.

A pendulum clock is a gravitational rather than an inertial clock, as is the motion of the moon around the earth, the earth around the sun, etc.; these depend on the law of gravitation as well as the laws of motion. Newton's law of gravity is not self evidently true and

[1] As far as I am aware, no one has proposed an improved version of Gauss' experiment; for 3 satellites in orbit around the earth the sum of the angles according to general relativity would be 180° plus an excess of about 0.005''.

needs to be tested: is the symbol t that enters the law of gravity the same as inertial time? Is the constant of gravity, G, really a constant in time and/or space? These are empirical questions that can only be decided by experiment. Such conceptually simple experiments would compare the period of a pendulum, or the length of the year with the length of a day. Suppose Newton's law of gravity is true for some time scale t_g. The period of a pendulum P_c, and the length of the year P_E would then be

$$P_c = 2\pi \left[\frac{lR_E^2}{M_E}\right]^{1/2} \frac{1}{G^{1/2}}, \quad P_E = \frac{2\pi L_E^{3/2}}{M_\odot^{1/2}} \frac{1}{G^{1/2}},$$

where l is the length of the pendulum, M_E the mass and R_E the radius of the earth, L_E the radius of the earth's orbit around the sun, and $M_\odot$ the solar mass. If P_c and P_E varied in time (as measured by an inertial clock) we could interpret this as the constant of gravity $G = G(t)$. However, this would require P_E/P_c to be independent of time since they are both gravitational clocks. If P_E/P_c varied in time then we would require a spatial as well as temporal variation in G.

So far I have paid little attention to the meaning of the symbols r for length and m for mass. Within Newtonian mechanics we measure length by two means, rigid rulers, and triangulation. It is an empirical fact about the world that objects that are the same length at one place remain equal in length when transported to another place. We can therefore usefully base our measurement of length on such a basis. We cannot however test that bodies do not change in length when transported, so it is purely a convention to assume that they do not. The triangulation method of length measurement is based on the assertion that Euclidean geometry is valid both for rigid bodies and for the path of light rays. As I pointed out earlier, whether or not these assertions are true are empirical questions.

The mass that enters Newton's laws is the inertial mass defined in terms of momentum conservation. Having set up a measure of length and time we can choose one standard mass m_0. By measuring initial and final velocity in a collision experiment we determine the inertial mass M of another body through the relation

$$m_0 u_i + MU_i = m_0 u_f + MU_f,$$

where (u_i, U_i) and (u_f, U_f) are the initial and final velocities of m_0 and M. Any one such experiment gives a value for M. The empirical content in the law of momentum conservation is that any initial conditions (u_i, U_i) give the same value of M for a given object. Newton's law of gravity states that the source of the gravitational field is proportional to the inertial mass, or that the constant of gravity, G, is actually a constant independent of the nature of the bodies.

7. The Newtonian Constant of Gravity

Having laid down a procedure for measuring mass, length, and time we can now ask what meaning can be given to the constant of gravity within mechanics and gravitation. The basic units are arbitrary; let us call them m_0, l_0, t_0. We can now imagine an experiment to determine G by measuring the period of a pendulum of length $L = N_1 l_0$, on the surface of a planet of mass $M = N_2 m_0$, whose radius is $R = N_3 l_0$; if the period $P = N_4 t_0$ we have

$$G = 4\pi^2 \frac{l_0^3}{m_0 t_0^2} \cdot \frac{N_1 N_3^2}{N_2 N_4^2}.$$

Clearly N_1, N_2, N_3, N_4 are measurable (by counting) so G can be determined in units of m_0, l_0, and t_0. Moreover we could imagine a world in which this experiment would give a different value of N_4 and hence of G depending on position and time. It is therefore meaningful to talk about a variation of the constant of gravity within the framework of Newtonian mechanics.

If we also introduce the velocity of light, c, we can compare a gravitational clock with a light clock. For example we could imagine an experiment in which two bodies were in orbit about each other and reflected a light signal back and forth. The number of such round trip signals in one orbit (relative to an inertial frame) is a measurable; it could be constant in time or vary in time. Similarly, given our choice of standards l_0, t_0, we can measure the velocity of light c. This could be constant or varying as measured in inertial time. Indeed, we could imagine a world in which both c and G varied in inertial time and in which the number of round trip light signals in a gravitational orbit also varied.

While we can imagine a world in which $c = c(t)$, $G = G(t)$, what would it mean to say that G were constant but different in value? How would this manifest itself in measurable ways? To say that $G = G(t)$ we mean that having chosen our units m_0, l_0, t_0, we find a different result of a given experiment when done at different times but referred to the same units. But we cannot transport our arbitrary units m_0, l_0, t_0 from this world to some other I might imagine. The problem is that the world of Newtonian mechanics is continuous. There are no intrinsic standards of mass, length, and time defined within the system so that measurement cannot be entirely reduced to counting and I cannot compare my two worlds. Within Newtonian mechanics the value of G is meaningless; it just reflects the arbitrariness of our choice of units. For illustration let us imagine that with the standard pre-atomic definitions of kilogram, meter, and second, we found that

$$G = 6.67 \times 10^{-7}\ \mathrm{m^3 \cdot s^{-2} \cdot kg^{-1}}.$$

The relative behavior of the moon and planets would be unchanged. The only change would be in the number of days in a year, and the number of swings of a 1 meter pendulum in one day; that is, comparisons of gravitational phenomena with our arbitrarily defined inertial clock. The observed phenomena could be equally explained by stating that the earth was spinning ten times faster so that there were only 8640 seconds in a day and G had its usual value.

This should be contrasted with a world model with discrete intrinsic structure. For example, if all bodies were made up of a basic unit m_0 of radius l_0, by using the propagation of light l_0 can also be a unit of time t_0, and all measurements are now reducible to counting. In terms of these basic units G is a number; if it had a different value physics would be different. For example the period of rotation of the earth is some pure number times the intrinsic unit t_0 so I can meaningfully state that the rotation rate is different and so separate a change in the earth's rotation from a change in G. Since our present models of the world do have a discrete quantized structure, an absolute determination of G becomes meaningful.

8. Quantum Clocks

The previous discussion is somewhat removed from actual practice. Time and distance standards are normally based on quantum phenomena and the properties of light. Electric charge apparently comes in identical 'lumps', in-

teger multiples of the charge on an electron e (or possibly multiples of $e/3$?). Matter also comes in lumps, m_e, m_p, etc., and angular momentum comes in multiples of $\hbar$. This allows us to define an atomic unit of time, for example the period of revolution of an electron in the ground state of a hydrogen atom as

$$P_0 = \frac{2\pi\hbar^3}{e^4 m_e},$$

and a unit of length, r_0, as the radius of this orbit:

$$r_0 = \frac{\hbar^2}{m_e e^2}.$$

We can now measure time in units of P_0, length in units of r_0, and mass in units of m_e. If Newton's laws are true the period P_E of the earth around the sun is given by

$$\frac{P_E}{P_0} = \left[\frac{e^2}{Gm_e^2}\right]^{1/2} \left[\frac{m_e}{M_0}\right]^{1/2} \left[\frac{L}{r_0}\right]^{1/2}.$$

Each term in parentheses is a pure number and therefore represents at least in principle an empirical measurement. Thus we can meaningfully ask if this relation is satisfied (i.e., is Newton's law true in atomic units) and determine the value of G by determining the dimensionless number e^2/Gm_e^2. We can also meaningfully ask if this dimensionless number ($\simeq 10^{42}$) is constant in time, that is do we get the same value when we conduct an experiment at different times? We should, however, remember that should we find a time variation we have a choice between saying e and m_e are constant in time and G varies, or taking G to be constant and letting e and/or m_e vary in time.

9. The Fine-Structure Constant and Other Numbers

So far we have not considered the velocity of light. Since we have a quantum definition of length and time it is now an empirical question as to whether the velocity of light is constant as measured in these quantum units. For example, the time required for a light signal to travel a distance $2\pi r_0$ measured in units of P_0 is

$$\alpha = \frac{2\pi r_0}{cP_0} = \frac{e^2}{\hbar c},$$

which is the fine-structure constant. Whether or not α is constant in time is therefore an empirical result about the world, and this can only be settled by experiment.

Similarly, we may ask if the mass ratio of elementary particles, i.e., m_p/m_e, varies in time. Indeed, any pure number is independent of our choice of units and can be considered as a possible candidate for time variation. In the end it is experiment that has to decide; all philosophy can do is to try and separate the meaningless from the meaningful questions.

One might reasonably ask why we should expect any time variation of numerical constants. To answer this we must consider the large scale structure of the universe.

10. Cosmology

There are very few measured properties of the large scale structure of the universe, and even fewer that have any claim to accuracy. The best known is probably the Hubble parameter H_0. This is often thought of as the constant of proportionality in the velocity-distance relation for the expanding universe, $V = H_0 r$. However, distances and velocities are inferred quantities; the actual measurables are the energy per unit time per unit area F, and the red shift Z, of light from distant galaxies; H_0 is then calculated from the relation

$$F = \frac{L\,H_0^2}{4\pi Z^2}\,[1 + (q_0 - 1)Z + \ldots],$$

where L is the intrinsic luminosity of such galaxies. Estimates of L and hence H_0 vary by factors of one to four (and maybe more). Consequently H_0 is only known to a factor of about 2 (or more!); to order of magnitude $H_0^{-1} \simeq 10^{10}$ years.

The parameter q_0 in the above expression is called the deceleration parameter and is related to the rate of change of the Hubble constant H_0. One of the challenging problems of observational cosmology is to determine the actual value of q_0 since (with the aid of some theory!) it will tell us whether the universe will continue to expand or contract. If the mean density of matter in the universe, ρ_0, can also be determined, these parameters will tell us whether our simple cosmological models are valid.

This mean density, ρ_0, is difficult to estimate. If we just estimate the masses of galaxies then $\rho_0 \simeq 0.01\rho_c$ where ρ_c is called the critical density and is such that

$$\frac{4}{3}\,\frac{\pi G}{H_0^2}\,\rho_c = \frac{1}{2}.$$

But of course there may be mass that we have not yet detected; halos of galaxies, matter in intergalactic space (e.g., bricks!), dead stars, black holes, massive neutrinos, gravitational waves ..... As I said earlier there are few reliable measurements in cosmology!

11. Dimensionless Constants of Nature

The two cosmological parameters, H_0 and ρ_0, can be combined with other constants of nature, G, c, e, and m_e, to give the following pure numbers:

$$\frac{m_e c^3}{e^2 H_0} \simeq 6 \times 10^{40}, \quad \frac{m_e^4\, c^6}{e^6 \rho_0} \simeq 4 \times 10^{40};$$

$$\frac{e^2}{Gm_e^2} \simeq 4 \times 10^{42}.$$

The first parameter is often referred to as the age of the universe in atomic units; the second is the ratio of atomic to cosmological density; and the third is the ratio of electrical to gravitational forces. Other pure numbers we have considered are the fine-structure constant and the mass ratio of the electron to proton:

$$\alpha = \frac{e^2}{\hbar c} \simeq \frac{1}{137}; \quad \frac{m_e}{m_p} \simeq \frac{1}{1836}.$$

Why do these numbers have these values? Why is the ratio of the circumference to radius of a circle equal to 3.1415926...? Can these numbers be deduced from a single theory such that no other values are possible? Are these numbers just random initial conditions for the universe? What would the universe be like if these numbers were different?

I think most scientists would conjecture that one day we will be able to explain why the particle masses are as observed, and why the fine-structure constant has its particular value. At the moment we do not have a deep enough understanding of particle physics to provide the explanation. A similar view can be taken about the cosmological numbers of the order of 10^{40}. Such fundamental theories linking cosmology and microphysics have

been pursued by several people, but without apparent success [12, 13]. The point of view of Dirac [14] is that whatever the theory is, it will entail that where such large numbers as 10^{40} occur, then parameters of this magnitude are proportional to each other, so that

$$\frac{e^2}{Gm_e^2} = k_1 \frac{m_e c^3}{e^2 H_0} = \frac{k_2 m_e^4 c^6}{e^6 \rho_0},$$

where k_1 and k_2 are of order unity. This statement is known as *The Large Number Hypothesis*.

I should point out that the standard, general relativistic cosmological model does not satisfy this hypothesis; the first parameter is constant in time, the second parameter varies linearly with time, and the third parameter varies with the square of the time. (These statements are only strictly true for the Einstein-de Sitter model.) It is therefore pure accident that these three numbers happen to have approximately the same value at the present time.

The large number hypothesis leads to the result that either all three parameters are constant in time, or they vary in unison. The first proposal leads directly to the steady state cosmology in which G, H_0, and ρ_0 are constant. This in turn implies the creation of matter [15]. The steady state theory has few advocates these days as its very specific predictions are apparently contradicted by observation, for example, the radio source counts. I must admit that I am not wholly convinced that some form of the steady state theory cannot stand up to this data, for example, a statistically steady state with local fluctuations.

The second proposal advanced by Dirac is that as H_0^{-1} is a time, related to the age of the universe, then we should find

$$G \propto \frac{1}{t}; \quad H_0 \propto \frac{1}{t}; \quad \rho_0 \propto \frac{1}{t}.$$

This is one origin of the idea that the Newtonian constant of gravity varies in time, although this proposal was advanced somewhat earlier by E. A. Milne [16]. Such a variation is incompatible with general relativity and some new theory needs to be advanced that successfully explains this time variation, and whose predictions can be evaluated and subjected to empirical investigation. The scalar-tensor theory of Brans and Dicke [17] was such an attempt, but it cannot successfully explain the time variation of G and ρ_0 [18]. I have considered many other possible theories but I have not, as yet, been able to find one that explains all the large number relations and agrees with solar system tests on the orbits of planets [19].

Recently Dirac introduced a two metric framework. In the gravitational metric, ds_g, Einstein's equations are valid, and this is conformally related to an atomic metric, $ds_A = ds_g/\beta$, where β is to be determined [20]. Without any theory for the determination of β all that can be done is to ask general questions such as what classes of theory satisfy the large number hypothesis [21] and what consequences would such theories have in other branches of physics [22]. My own view is that little progress can be made without some more precise theory.

However, experimentally there is clearly a testable prediction, namely that the constant of gravity G decreases in atomic time, $\dot{G}/G \simeq -10^{10}$ years. This can be tested either by radar ranging experiments or by laboratory tests. A second prediction is that there could be creation of matter proportional to matter already present, and this might be detectable in the laboratory [23]. At the present time no experiments are sensitive enough to detect these effects.

12. The Anthropic Principle

I have already pointed out that in standard theory (i.e. general relativity and quantum mechanics) the large numbers vary in time, and it is just a chance event that these three numbers happen to be about 10^{40}. These numbers could be different, in which case the properties of the universe would be different, and in particular we, man the observer, would not be here to ask why the universe is as it is. This idea has been elevated to the status of *The Anthropic Principle* [24, 25, 26].

Since man contains heavy elements there must have been nucleosynthesis, therefore there must have been stars that have evolved to supernovae and ejected heavy elements into the interstellar material. This material then formed other stars with planets on which life as we know it evolved.

A star is an object in which pressure balances gravity and in which gas pressure dominates over radiation pressure; this is sufficient to show that stars have masses up to a maximum M_0 where

$$M_0 \sim \frac{m_p}{\alpha_G^{3/2}}; \quad \text{and } \alpha_G = \frac{Gm_p^2}{\hbar c}.$$

On the other hand the central temperature must be high enough for nuclear reactions to take place before the gas is degenerate. This gives a minimum mass of about $0.01\ M_0$. The luminosity of a star is essentially determined by radiative energy transport. The principal opacity source is scattering by electrons, and this permits the evaluation of the luminosity L_0. The evolution time for a star is essentially the time to convert a significant fraction of hydrogen to helium and so is proportional to M_0/L_0. The actual result gives

$$t_{\text{evol}} \sim \frac{1}{\alpha_G} \frac{\hbar}{m_p c^2}.$$

The anthropic principle is now used to argue that if the age of the universe H_0^{-1} is less than t_{evol} then no stars would have evolved. On the other hand, if the universe is very much older than t_{evol} there would be no stars left. Conditions are therefore ripe for life to develop when the universe is just a little older than t_{evol} or

$$\frac{1}{H_0} \sim \frac{1}{\alpha_G} \frac{\hbar}{m_p c^2} \text{ or } \frac{m_p c^3}{e^2 H_0} \sim \left[\frac{e^2}{Gm_p^2}\right] \frac{1}{\alpha^2},$$

where α is the fine-structure constant. This 'explains' why this large number coincidence is observed.

The second large number coincidence is 'explained' by appealing to cosmological models. In such general relativistic models the deceleration parameter q_0 is

$$q_0 = \frac{4\pi}{3} \frac{G\rho_0}{H_0^2}.$$

In the Einstein de Sitter model $q_0 = 1/2$. If q_0 were very much less than 1/2 the universe would not be unstable to the formation of galaxies, and hence stars. On the other hand, if q_0 were much greater than 1/2, the age of the universe, which for such models is approximately $1/H_0 q_0^{1/2}$, would be so short that there would not have been time for stars to have been formed and to have evolved. Hence we expect to find $q_0 \sim 1/2$.

An alternative explanation is that general relativity is incomplete since it does not predict a unique cosmological model. When we have found how to correct this deficiency we will find that the only possible cosmological model is the Einstein-de Sitter model with $q_0 = 1/2$.

Either way the combination of $q_0 = 1/2$ with the other large number relation gives the results:

$$\frac{m_e^2 c^6}{e^6 \rho_0} \sim \frac{e^2}{G m_e^2} \sim \frac{m_e c^3}{e^2 H_0}.$$

This line of argument has been pushed further by Carter [25] who argues that we need conditions which give some stars with convective outer envelopes since these stars have planets, but we cannot have completely mixed convective stars since supernovae require unmixed stars. This argument is quantified by requiring that the effective temperature of stars be approximately the ionization temperature of hydrogen ($0.1\alpha^2 m_e c^2/k$). This condition determines the mass of such stars as

$$M_* = \alpha_G^{-1/2} \alpha^{10} M_0.$$

The argument is now used that $\alpha_G^{-1/2}\alpha^{10} \simeq 1$ since if it were much greater all stars would be convective and mixed, whereas if it were much less then there would be no stars with outer convective zones, and so no planets. With $\alpha_G \simeq \alpha^{20}$ we have $\alpha_G \sim 10^{-42}$ and the actual magnitude of the large numbers is 'explained'. Similar arguments to those above can be used to explain other numerical constants of nature [26].

13. Critique of the Anthropic Program

My first objection to the anthropic program is that the technical arguments have not been carried through; the behavior of models with substantial different parameters awaits evaluation. For example, could we not have a universe with $q_0 = 10^{40}$? In such a model stars would have evolved provided

$$\frac{1}{H_0 q_0^{1/2}} \sim \frac{1}{\alpha_G} \frac{\hbar}{m_p c^2}.$$

The arguments on galaxy and star formation are a little weak since we are still struggling to understand how gravitational condensations occur. The argument advanced by Carter to explain why $\alpha_G \simeq 10^{-42}$ is very weak. Firstly supernovae could occur in homogeneous evolution after evolution through to carbon stars, and secondly it is by no means obvious that convective zones are a prerequisite for planetary formation. This criticism however can and should be quantified by calculations on the properties of a universe with different values of the large numbers.

But if the program is successful, what does it tell us? There are laws of nature (L_i) containing arbitrary constants (α_j) and a set of observations (O_k); and a subset of these observations is the existence of ourselves. If the theories are 'correct' then with the observed values of the constants they must predict our existence. If they did not predict our existence the theories would be wrong—that is contradicted by observation. If we had the same theories with different constants the predictions would necessarily be different, and there is no reason why those predictions should include the emergence of mankind. The anthropic program is therefore bound to be successful. We exist in the universe and if we had a full knowledge of the laws of physics, we would predict our own existence.

The more fundamental problem is why there are any constants of nature at all, or indeed why there are any laws of nature. Is there a Creator who set up the universe with just those initial conditions, laws, and constants of nature that would lead to the emergence of mankind? Is it just chance? Is there an infinity of 'parallel' universes with different laws and different constants, each with its own properties? Our universe which contains us then necessarily has those laws that give rise to mankind. Does the universe go through cycles in which the laws and constants of nature change? Could it be that there are really no laws or constants of nature, but that we just live in a statistical fluctuation of order in a universe of chaos? Or are there unknown laws, constraints, consequences yet to be discovered that will ultimately lead us to the conclusion that there can only be one universe, one set of laws, one set of constants, one geometrical structure? Is there only one possible universe with the properties we observe including our own existence?

14. Conclusions

I have tried to emphasize throughout this article that experiments are the final arbiter; in Poincaré's words "Experiment is the sole source of truth" [27]. Yet mathematical physics has a vital role to play in developing ideas through to predictions that can then be compared to experimental data; and philosophy too has a role to play in challenging our firmly held views, in separating the meaningful from the meaningless questions.

I would therefore like to finish by encouraging the reader to let his hair down occasionally and think of the outlandish possibilities, which can in turn be developed and tested. Is 'space' as determined by light signals locally Euclidean? Is gravitational time the same as inertial time, spin time, electric time ..... Are the laws of physics 'here and now' the same as they were 'there and then'? Is matter being created or destroyed? At what level of accuracy will we find general relativity to be incorrect? Does special relativity hold good as $v \to c$? Do the fine-structure constant and the particle mass ratios vary in time at some level? The reader can add his own ideas to such a list!

References

[1] Shu Ching, *The Book of History* (circa 2200 B.C.). For translations, see J. Legge, *The Shoo King* (The Author's Hongkong, London, 1865); C. Waltham, *Shu Ching* (Allen and Unwin, 1971). See also C. P. S. Menon, *Early Astronomy and Cosmology* (London, 1932).

[2] Chou Pi (circa 1100 B.C.). For a discussion of this model, see Thaibut, J. Asiatic Soc. of Bengal (1880); and C. P. S. Menon, *Early Astronomy and Cosmology* (London, 1932).

[3] Suryaprajnapti, Book 1, chapter 8 (circa 500 B.C.); see also W. Brennand, *Hindu Astronomy* (London, 1896).

[4] D. Hume, *An Inquiry Concerning Human Understanding* (1748) [see also Open Court Edition, Chicago (1927)].

[5] Euclid, *Elements* (circa 300 B.C.); see also T. L. Heath, *The Thirteen Books of Euclid's Elements* (Dover, New York, 1956).

[6] G. Leibniz (1674), published in *Math. Schrifters*, vol. 5, Ed. by C. I. Gerhardt (Ascher Schmidt, 1849), p. 88. See also M. Kline, *Mathematical Thought from Ancient to Modern Times* (Oxford University Press, 1972).

[7] J. Machin, in *W. Jones* (Synopsis, London, 1706); see also K. Kopp, *Infinite Series* (Hafner, New York, 1971), p. 253.

[8] H. Poincaré, *La Science et l'Hypothèse* (Villiers, Paris, 1902). [English transl.: G. B. Halsted (Dover, New York, 1952.]

[9] K. F. Gauss, *Werke* vol. 8 (B. G. Teubner, Leipzig, 1900), p. 177.

[10] K. F. Gauss, *Werke*, vol. 4 (B. G. Teubner, Leipzig, 1900), p. 258.

[11] I. Newton, *Philosphial Naturalis Principia Mathematica*, London (1687).

[12] A. S. Eddington, *Fundamental Theory* (Cambridge University Press, 1946); E. A. Milne, *Kinematic Relativity* (Oxford University Press, 1948).
[13] I. W. Roxburgh, in *The Encyclopedia of Ignorance*, Ed. by R. Duncan and M. Weston-Smith (Pergamon Press, Oxford, 1977), p. 38.
[14] P. A. M. Dirac, Proc. R. Soc. London, Ser. A: 165, 199 (1938).
[15] H. Bondi and T. Gold, Mon. Not. R. Astron. Soc. 108, 372 (1948).
[16] E. A. Milne, Nature (London) 130, 508 (1932); and *Relativity, Gravitation and World Structure* (Oxford University Press, 1935).
[17] C. H. Brans and R. H. Dicke, Phys. Rev. 124, 925 (1961).
[18] S. Weinberg, *Cosmology* (Wiley, New York, 1972), p. 619.
[19] I. W. Roxburgh, Bicentennial Lectures on Cosmology, University of Virginia, 1976 (unpublished).
[20] P. A. M. Dirac, Proc. R. Soc. London, Ser. A: 338, 439 (1974).
[21] I. W. Roxburgh, Nature (London) 268, 504 (1977).
[22] V. M. Cannuto, P. J. Adams, S. H. Hsieh, and E. Tsiang, Astrophys J. Suppl. Ser. 41, 243 (1979).
[23] R. Ritter, these proceedings.
[24] R. H. Dicke, Nature (London) 192, 440 (1961).
[25] B. Carter, in *Confrontation of Cosmological Theories and Observation Data: Proceedings*, Ed. by M. S. Longair (Kluwer Boston, 1974).
[26] B. J. Carr and M. J. Rees, Nature (London) 278, 605 (1979).
[27] H. Poincaré, *La Science et l'Hypothèse* (Villiers, Paris, 1902). [English transl.: G. B. Halsted (Dover, New York, 1952), p. 140.]

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

FREQUENCY, WAVELENGTH, AND STORED IONS

Frequency Standards Based on Magnetic Hyperfine Structure Resonances

H. Hellwig

Frequency and Time Systems, Inc., 34 Tozer Road, Beverly, MA 01915

Magnetic hyperfine resonances in alkali atoms were the physical basis of practical atomic frequency standards at the beginning of the age of the realization of the unit of time via atomic phenomena. This dates back more than 25 years; it appears to be no accident that today's time and frequency technology still rests on the same physical basis: practical frequency standards and clocks use magnetic hyperfine transitions in cesium, rubidium, and hydrogen, and the unit of time is defined—as well as practically realized—via the cesium resonance.

This paper explores the basis for this phenomenon which is a result of a combination of mature electronics and physics technologies with proven principles of experimental physics such as beam spectroscopy and optical pumping. This paper will also address the limitations of these "traditional" or "microwave" frequency standards as well as opportunities still open for further improvements. These limitations and opportunities center around the desirability to achieve a spectrally narrow line; i.e., a high line-Q within the microwave region. Options to realize improved Q-values will be discussed.

Key words: atomic clocks; cesium beam; frequency standards; hydrogen maser; rubidium gas cell.

1. Introduction

The resonances of cesium, rubidium, and hydrogen in the microwave region have dominated the atomic time and frequency standard field since the 1960's. In fact, rubidium and cesium standards now have been commercially available for two decades [1, 2]. The 1960's saw refinements in atomic clock design and their use in many scientific applications and in metrology including the generation of time on a world-wide basis. The 1970's can be characterized by the beginning of large-scale system-use of atomic time and frequency standards reaching far beyond scientific studies, making possible today's communication and navigation systems.

An atomic time and frequency standard consists of three elements: the external interfaces, the electronic systems, and the atomic resonator. The atomic resonator is the heart of the atomic standard and serves the same function as the pendulum in a mechanical clock; however, with one most fundamental difference: The atomic resonance frequency itself is given by nature, it will not drift, nor will it age according to all of our present experimental and theoretical knowledge. Furthermore, all atoms are the same. A standard built independently from another standard will show the same characteristics and the same frequency as its predecessors or as standards in different locations. Atomic clocks feature an *a priori* accuracy; therefore, atomic clocks and frequency standards are often called primary standards.

The frequency of the emitted or absorbed radiation is related to the energy difference between the two atomic energy levels. In the three alkali atoms, hydrogen, rubidium, and cesium, which are most prominently in use today, as they were for several decades, the magnetic hyperfine transition in the ground state ($^2S_{1/2}$) is being used. Table 1 shows a summary of these three standards and their characteristic frequencies. The transitions are between energy levels which are created by the spin-spin interaction between the atomic nucleus and the outer electron in the ground state of the atom. As such, the separation of the two energy levels, i.e., the transition frequency, is a function of the magnetic field. This dependency is described by the Breit-Rabi equation [3, 4] and depicted in Fig. 1.

In order to define the precise resonance frequency, the external magnetic field has to be well defined and kept

TABLE 1. *Summary of most widely used atoms.*

Atom	Atomic Mass	Nuclear Spin	ν_R(Hz)
Hydrogen (H)	1	1/2	1 420 405 752
Rubidium (Rb)	87	3/2	6 834 682 608
Cesium (Cs)	133	7/2	9 192 631 770

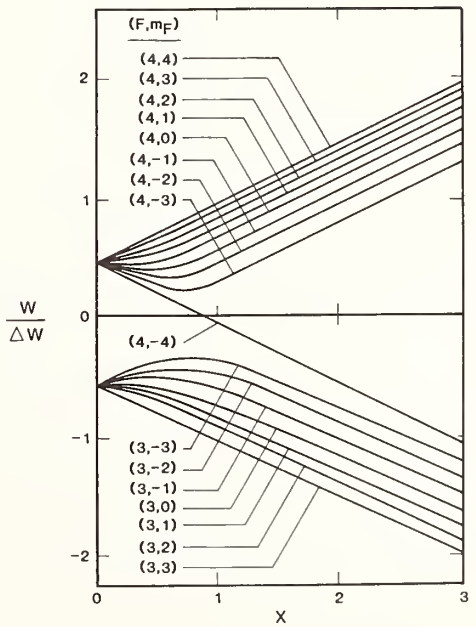


FIGURE 1. *Energy level diagram of ^{133}Cs in the $^2S_{1/2}$ ground state as a function of the applied magnetic field.*

constant. This necessitates magnetic shielding which is a characteristic design feature of all presently used atomic frequency standards. The shielding can be quite elaborate and reduces the external magnetic fields, foremost the earth's magnetic field, to one percent or less of its normal value. It is obvious that only the $\Delta m_F = 0$, the so-called clock transition, can provide highly stable reference frequencies. The transition probability for this transition must be maximized and the corresponding probabilities for all other magnetic-field-dependent transitions must be minimized. This is done by applying a constant magnetic field, the so-called C-field. In order to force the $\Delta m_F = 0$ transition, the magnitude of the C-field is typically at least an order of magnitude above all anticipated residual field components, and its orientation is such that its vector is parallel to the vector of the oscillating field which is coupled to the resonance of the atom. The resulting frequency shift can range up to parts in 10^9 as compared to the frequency of the atom in a zero magnetic field.

The magnitude of the field can be measured quite precisely, actually to an accuracy which is much better than required in view of all other frequency stability limitations. Such a measurement is done by using the atom itself as the magnetometer. In Fig. 2, which illustrates cesium for regions of relatively low magnetic field [4], the clock transition is indicated by the solid line. The magnetic field H can then be measured precisely by measuring the magnetic field dependent or Zeeman resonances at higher or lower frequencies as indicated by the dotted lines between the energy levels. We note that we can measure a magnetic field H by measuring at microwave frequencies or, alternately, by injecting a low frequency.

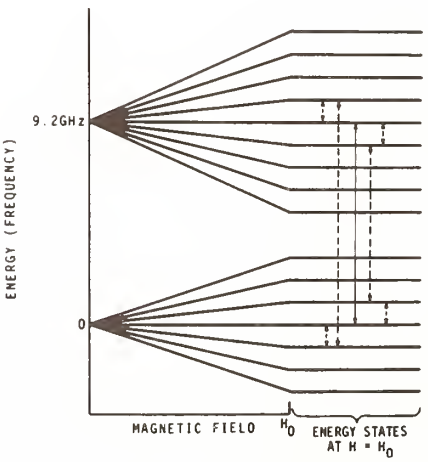


FIGURE 2. *Magnetic field dependence of the hyperfine energy levels in the groundstate of the cesium atom.*

2. Design Principles of an Atomic Resonator

Atomic frequency standards, fundamentally, are devices which allow the measurement of a microwave frequency with very high resolution and accuracy. The Doppler effect represents perhaps the most important problem limiting the accuracy and resolution of any frequency standard. In the usual way we can say that if an absorber of radiation moves relative to the source, the observed resonance frequency is shifted to the value [5]

$$\omega_{obs} \simeq \omega_0 + \mathbf{k} \cdot \mathbf{v} - \frac{\omega}{2}\left[\frac{v}{c}\right]^2 + \frac{\hbar\omega_0^2}{2Mc^2}$$

where the velocity $\mathbf{v}$ and wave-vector $\mathbf{k}$ are measured relative to the source. The first order Doppler shift ($\mathbf{k} \cdot \mathbf{v}$), the second order Doppler shift $(v/c)^2/2$, and the recoil shift (the last term) can be understood in terms of conservation of energy and momentum in the emission or absorption process. Specifically, the so-called second order Doppler shift is merely the relativistic time dilation factor resulting from the movement of the atom relative to the apparatus. Its effect is small but ultimately important.

The term "natural linewidth" normally refers to the emission of visible light where the natural linewidth describes the spectral linewidth associated with an optical transition. Lifetimes of optical transitions can be calculated from quantum mechanics using Planck's equation for the black-body radiator. The lifetime in a given state is proportional to the cube of the frequency; thus, at microwave frequencies we are faced with natural lifetimes of the states involved of hundreds of seconds. In other words, the natural linewidth is unmeasurably small in the microwave region as compared to other limitations. As a consequence, one may interpret for microwave frequency standards (Fig. 3), that the lifetime is given by the observation time allowed by the frequency standard apparatus or atomic resonator to coherently observe the atomic radiation. In general, this lifetime is limited by either of two general effects: (1) the atom enters and leaves the apparatus after a time Δt, or (2) the atom stops oscillating due to a collision with other atoms or the walls of the container after a time Δt. Thus atomic frequency standards are characterized by maximizing Δt and, with it, the line-Q.

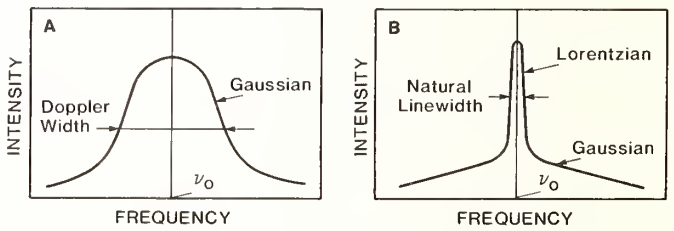


FIGURE 3. *Spectrum of an atomic transition. On the left, part A shows the situation when the atoms are unbound and the resonance feature has the full Doppler width $\Delta\nu \sim (v/c)\,\nu_0$. When the atom is confined to dimensions less than the wavelength, the Doppler profile is suppressed and the central feature has the natural width $\Delta\nu$ as shown in part B. (From Ref. [5].)*

Fundamentally related to this, again via Planck's equation, is the fact that the population difference between the two levels separated by microwave frequencies is very small. The population difference is a function of ν. In the low gigahertz region where $h\nu \ll kT$, the population difference is only a very small fraction of one percent. Therefore, in order to observe effectively the atomic resonance, it is a common design principle in microwave frequency standards to modify the population in the upper and lower state. Either one has to be significantly depopulated. The result is an atomic ensemble which either has a significant net emission or net absorption of energy. The way in which this is accomplished is a significant determinant in the design of an atomic resonator.

In summary, the design of an atomic resonator is characterized by the particular method of maximizing the coherent observation time, minimizing first order Doppler effects, and achieving a significant net population difference between the two microwave energy levels.

3. Performance Principles

Time and frequency standards can be characterized in numerous ways. However, most frequently, a time domain characterization is most lucid and useful. The quantity measured [6] is the two sample or Allan variance $\sigma_y^2(\tau)$ defined as

$$\sigma_y^2(\tau) = \left\langle \frac{(\bar{y}_{k+1} - \bar{y}_k)^2}{2} \right\rangle ,$$

where y is the fractional frequency of adjacent measurements each with a sample time τ, and the brackets denote an infinite time average.

An idealized presentation [1] for all precise time and frequency standards is shown in Fig. 4. The first part, I, of the stability plot is determined by the fundamental noise processes present in the standard and can be characterized by the equation $\sigma_y(\tau) = k_1\tau^{\alpha_1}$. The second part, II, is called the flicker or frequency floor, which describes an independence of $\sigma_y(\tau)$ from the averaging time. We shall call this value σ_{yF}. Its level depends on the particular frequency standard and is not fully understood in its physical basis, but relates to fluctuations in the value of critical, frequency-determining parameters of the standard. The last section, III, can again be characterized by $\sigma_y(\tau) = k_2\tau^{\alpha_2}$. The coefficients k_2 and α_2 cannot usually be determined very accurately because of the long measurement times needed in order to obtain statistical confidence. These coefficients are subject largely to environmental effects and are thus not stable. If linear frequency drift is present, $\alpha_2 = +1$. Of course, drift is always caused by some systematic change in a frequency determining parameter with time. In contrast, Regions I and II have more quantifiable physical relationships. The approximate relationship between frequency stability in Region I and the signal to noise available in the device can be given by $\sigma_y(\tau) \simeq (Q \cdot \mathrm{S/N})^{-1}$. S/N is the signal to noise ratio which, of course, is a function of the sampling time τ. The quality factor of the resonance line enters as a key determining factor: the larger the Q, the better the stability. However, Q and S/N can be traded against one another.

A device with a relatively low Q and a large S/N gives the same short-term stability performance as a device with a high Q and a low S/N. S/N can be expressed in terms of physical parameters of the atomic resonators which typically are shot-noise-limited: If n is the number of signal events per unit time, $\mathrm{S/N} = \sqrt{n\tau}$.

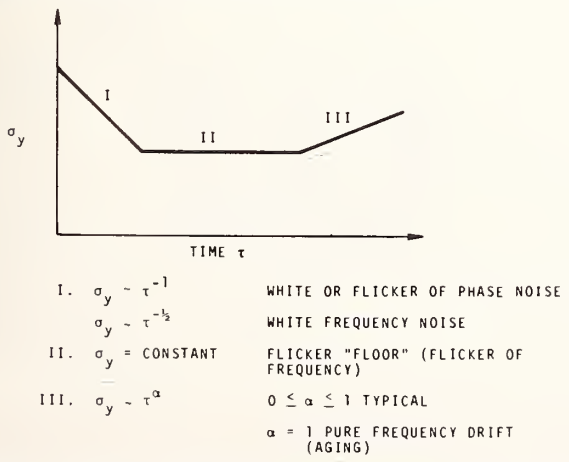


FIGURE 4. *Typical frequency stability behavior of a frequency standard. The two-sample variance is* $\sigma_y^2(\tau)$.

The flicker floor, Region II in Fig. 4 (σ_{yF}), appears to be related to the Q-value of the resonance. Qualitatively, this is not surprising. Atomic resonators with Q-values of 10^8 or higher may be expected to have accuracies of one part in 10^8 or better because the linewidth is acting like an ultimate limit on the variability of the frequency of the resonance [7]. We can summarize: A high resonator Q is essential to achieve high accuracy and ultimate stability. Thus, the time keeping potential of an atomic clock is governed by the Q-value and not by the signal-to-noise. Signal-to-noise is only important to assure sufficient short-term frequency stability so that measurement periods of acceptably short duration are needed to realize the available accuracy.

4. Electronic Systems

Figure 5 depicts the general concept of the overall design of an atomic standard where the atomic resonator serves as the frequency reference for a slave oscillator (usually a quartz crystal) in a servo loop configuration.

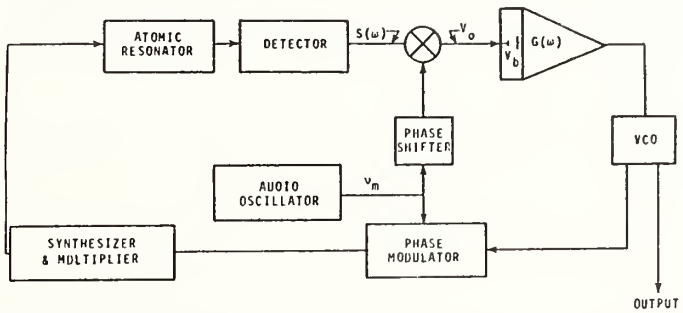


FIGURE 5. *Block diagram of an atomic clock with a frequency lock servo.*

Frequency-lock servos are used in passive atomic clocks (e.g., cesium beam and rubidium gas cell) to steer the frequency of the reference oscillator so that it coincides with that of the atomic transition. The atomic resonator behaves similarly to a simple band-pass filter followed by a square-law detector. Therefore, in the following, we will assume that the "line shape" of the detected signal $S(\omega)$ versus applied rf frequency is Lorentzian, i.e.,

$$S(\omega) = S_0\,(1 + (\nu - \nu_0)^2/\Delta\nu_l^2)^{-1},$$

where ν_0 is the atomic frequency, $\Delta\nu_l$ is the line width, and ν is the frequency of the source (multiplied reference oscillator). $S(\omega)$ is plotted in Fig. 6. Although some devices do not exhibit such line shapes, the qualitative results discussed below still apply. We can use the $S(\omega)$ curve as a frequency discriminator; one common way is to use sinewave frequency modulation (FM) on the source. In this case $\nu = \nu_s + \Delta\nu_m \sin\omega_m t$ where $\Delta\nu_m$ is the FM swing and $\omega_m = 2\pi\nu_m$ the modulation frequency.

If we now mix this signal with a reference signal proportional to $\sin\omega_m t$, as shown in Fig. 5, then the output of the mixer has output near dc (i.e., neglecting $2\omega_m$ terms) of

$$V_0 = -F\,\frac{\Delta\nu_m(\nu_s - \nu_0)}{(\Delta\nu_l)^2\left[1 + \left[\dfrac{\nu_s - \nu_0}{\Delta\nu_l}\right]^2\right]^2} + D\ ,$$

where F depends on the input levels to the mixer and mixer efficiency and D is a possible mixer output offset.

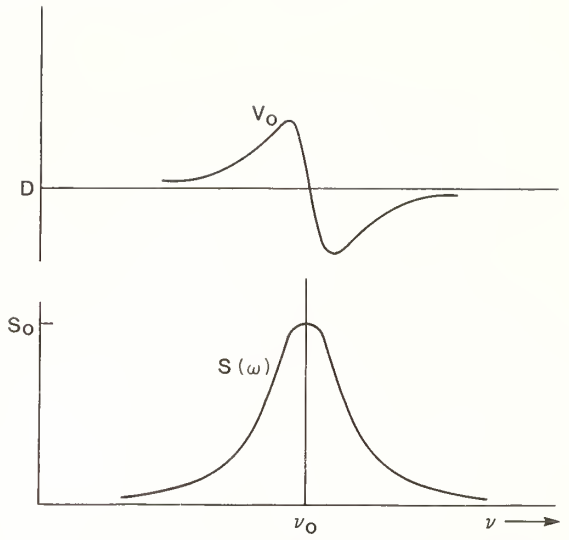


FIGURE 6. *Plot of atomic resonance features assuming a Lorentzian line shape. $S(\omega)$ is detected signal versus frequency. $V_0(\omega)$ is the detected signal component (at frequency ν_m) when the amplitude of the frequency modulation is small ($\Delta\nu_m \ll \Delta\nu_l$).*

This output voltage as a function of frequency is plotted in Fig. 6 and can be used to servo the frequency of the oscillator (VCO) as shown in Fig. 5. For good long-term stability and accuracy, we want the gain $G(\omega = 0) \rightarrow \infty$; therefore this part of the servo is usually an integrator. Depending on the particular design of the servo electronics, frequency offsets, frequency instabilities, and frequency drifts can be induced which usually relate to phase and gain changes. However, proper design and alignment can nearly arbitrarily reduce such phenomena.

It is possible to operate an atomic standard as an active oscillator or maser oscillator. It is necessary, nevertheless, to include frequency synthesis and electronic servo principles in such a device because the output signal is not a standard frequency and, typically, has a very low power level (order of picowatts).

The electronic system of a maser oscillator frequency standard is depicted in Fig. 7. The output of the maser is amplified in a wideband amplifier, and then fed into a double heterodyne receiver which ultimately translates the frequency down to dc. This dc voltage is used to control the frequency of the crystal oscillator which drives the double heterodyne system. In other words, the crystal oscillator is phase-locked to the maser frequency. Such frequency standards are called "active" devices (in contrast to the "passive" principles discussed above).

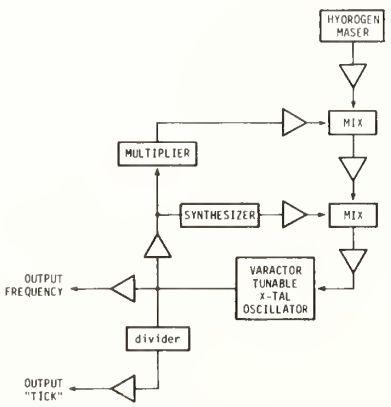


FIGURE 7. *Schematic of a maser oscillator frequency standard.*

In both "passive" as well as "active" systems, the slave oscillator is of special importance because its performance determines, in part, the frequency stability of the whole standard. For averaging times of less than those corresponding to the unity gain of the servo loop, the frequency stability is essentially that of the slave oscillator; for times larger than that, the atomic resonator dominates the stability performance. It must be noted here that an inferior slave oscillator, i.e., an oscillator with a frequency stability worse than that of the atomic resonator at the unity gain point, will correspondingly reduce the frequency stability of the system for some range of averaging times. Thus, in the design of a complete frequency standard, the choice of both the stability characteristics of the slave oscillator as well as the atomic resonator in combination with the selection of the time constant of the servo loop or bandwidth which determines the unity gain point, is critical to achieve optimum overall system performance.

5. The Cesium Beam Standard

The cesium standard was the first and still is the most important of all atomic standards. Developed in the early 1950's, its first operational use as a laboratory clock came about in 1955 at the National Physical Laboratory in the United Kingdom, and its first commercial realization took place in 1958 at the National Company in Massachusetts, U.S.A. Today, a variety of laboratory-type and commercial standards exist [1], and the world's internationally coordinated time system is based on a large number of cesium standards of worldwide distribution. The laboratory devices, or primary standards, are used to define the unit of time, the second [8]. This is done by constructing the apparatus in such a way that most known perturbing parameters can be independently and separately measured. Then, corrections can be applied to the output frequency in order to realize, as best as possible, the unperturbed resonance frequency of the cesium atom at rest in free space [9, 10, 11].

Figure 8 depicts the schematic design of a cesium beam standard omitting the electronic systems. The oven is constructed in such a way that an ampule containing cesium can be inserted and opened inside of the oven via a remote mechanism after the beam device has been evacuated to very low pressures. Pressures of better than 10^{-7} Torr are needed to allow sufficient mean free path from source to detector. Oven temperatures used vary between 60 and 120 °C, providing vapor pressures

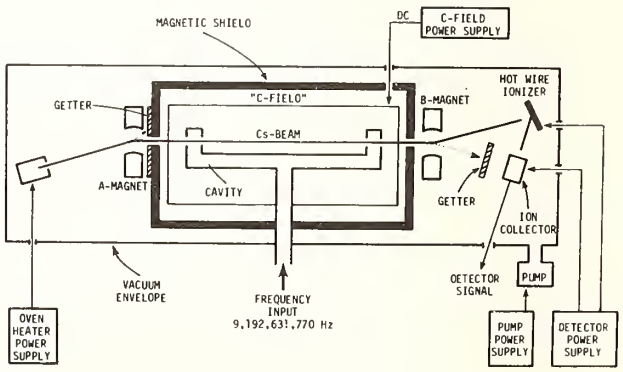


FIGURE 8. *Schematic of a cesium atomic-beam resonator. The frequency input is derived from a quartz-crystal oscillator (typically at 5 MHz) with a frequency multiplier and synthesizer to generate the atomic resonance frequency. A feedback servo from the detector output then controls this oscillator.*

in the range from 10^{-4} to 10^{-2} Torr inside of the oven. The state selectors for a cesium beam are typically dipole magnets, but multiple dipole, double-dipoles, and hexapole magnets have been used. Dual-beam devices have been built which utilize simultaneously the upper and lower state by generating two different beams deflected in opposing directions [10].

Vacuum is typically maintained by ion pumps; these feature rather sizeable pump capacities in the case of laboratory-type standards, but are as small as 0.1 l/s internal pumps in the case of commercial cesium standards. These small pumps serve only to scavenge residual gases, most importantly noble gases. Since prior to permanent seal-off a commercial tube is baked at high temperatures, there is very little evolution of gases inside of the tube; furthermore the expended cesium acts as an effective getter for many gases. The waste-cesium itself is gettered by strategically placed graphite in solid, as well as in surface deposition form.

The detector in cesium standards is typically a hot wire. The low ionization potential of the cesium atom makes it possible to surface-ionize cesium atoms on metals with a sufficiently large work function. Metals which are being used include tungsten, niobium, and tantalum, as well as platinum-iridium alloys. Since most of these metals contain significant amounts of contaminants, in particular, potassium, the ions formed on the surface of the hot wire are extracted through a mass spectrometer. The mass spectrometer is typically a magnetic field, bending the ions by 90 degrees out of the original path of the incoming neutral atoms. The ions are collected on an electrode. In some high intensity beam laboratory devices the current is directly detected by an electronic amplifier [9, 10, 11], typically with a field-effect transistor front end. In all commercial devices, however, a tube-internal electron multiplier is used. Thus, the collector electrode becomes the first dynode of the multiplier. The electron multiplier provides sufficient gain to raise the signal level by orders of magnitude above the noise level of electronic amplifiers. The surface ionization efficiency, together with the low noise of an electron multiplier, allows an essentially noise-free detection of single atoms in cesium beam devices.

The speed of the atoms and the length of the interaction region determine the Q of the cesium resonator. Atomic velocities at the most probable thermal velocity are 200 to 250 m/s. In many practical cesium beam tubes, a lower velocity is selected by the beam optics, typically close to 100 m/s. Thus we find Q values ranging from a few 10^7 to a few 10^8 over the spectrum of commercial and laboratory devices. Typical beam intensities at the detector are of the order of 10^7 to 10^8 atoms per second corresponding to electronic currents of the order of picoamperes. The resulting signal-to-noise for averaging times of one second is several thousand and allows fractional frequency stabilities at one second in the range from 10^{-12} to 10^{-10}. Cesium is preferred over other atoms because of its convenient frequency, permitting the use of readily available electronic techniques while featuring a sufficiently high resonance frequency for achieving high Q values; cesium has sufficient vapor pressure not far above room temperature, and the detectability of cesium by surface ionization is nearly 100% efficient.

The most significant limitation to frequency stability and accuracy of cesium beam devices appears to be the distributed cavity phase shift [12, 13]. It arises from the fact that the cavity phase is distributed along the axis as well as across both coordinates of the plane perpendicular to the beam. As a result, different trajectories sample different cavity phases in each of the two interrogation regions; and the cavity phase difference which affects the shift of the Ramsey pattern becomes trajectory dependent. Since with spatial state selection the trajectories are velocity dependent, we have a two-parameter dependence of the effective cavity phase shift, and thus of the apparent resonance frequency. Since the distributed cavity phase shift can be assumed to be of the order of 10^{-4} to 10^{-3} radian, cesium tubes suffer from this effect, depending on their resonance line Q, anywhere between 10^{-13} to 10^{-11}. Since optimum power is a function of the velocity, we have a change of the average trajectory location with microwave power, and thus frequency shifts as a function of microwave power. Other related shifts [13] include temperature gradient related frequency shifts, acceleration or gravitationally induced frequency shifts, vibration induced perturbations, and voltage related frequency shifts, e.g., those caused by change in the collector voltage potential near the hot wire [14] (causing a sampling of different regions of the hot wire corresponding to different atomic trajectories). Thus, state selection and detection techniques which are uniform across the beam and retain spatial isotropy of the velocity distribution would essentially remove all effects related to the distributed cavity phase shift. Optical pumping appears to be an elegant means to this end, but at this point in time insufficient experimental data are available to support its practical feasibility. Optical detection would be a combined function of the second state selector which, if optically executed, would interrogate the population levels and act as a photon transformer for the detection of microwave-induced atomic transitions.

6. The Rubidium Gas Cell Standard

The traditional design of a rubidium standard features a lamp filled with rubidium 87 isotope, a filter with rubidium 85 isotope, and the cell featuring again rubidium 87 isotope [15]. As illustrated in Fig. 9, it is a happy coincidence that the optical absorption lines of rubidium 85 largely overlap with one of the hyperfine splitting components of the rubidium 87 pumping light; thus the insertion of the rubidium 85 filter effectively suppresses the light which otherwise would couple to the upper microwave level of rubidium 87. This results in efficient optical pumping, i.e., in population of the upper

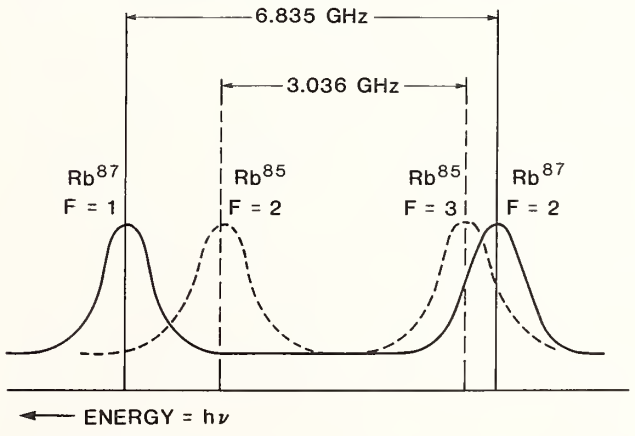


FIGURE 9. *The solid curves represent optical emission corresponding to the desired (F = 1) and the undesired (F = 2) states of ^{87}Rb. The dashed curves represent absorptions corresponding to the filter action of ^{85}Rb.*

microwave level, while the lower level essentially is fully depopulated.

Figure 10 shows the schematic of a rubidium gas cell frequency standard. In order to achieve full efficiency of the system, the vapor pressure in the gas cell has to be adequately high; partial pressures of 10^{-6} to 10^{-5} Torr are a reasonable compromise between achieving good signals while avoiding significant numbers of spin exchange collisions between rubidium atoms which would lead to frequency shifts [16] and also shorten the relaxation time and thus the line Q. In order to obtain the needed light intensity, the rubidium pressure in the lamp has to be at a higher level, and, typically, temperatures between 120 °C and 180 °C in the lamp are used. The filter temperature is adjusted to a temperature between those of the lamp and the gas cell. The gas cell features a buffer gas mixture typically containing two components, one with a positive pressure shift on the rubidium frequency, the other with a negative shift, compensating each other's temperature coefficients [17]. In practice the residual temperature coefficient is not much less than 1 part in 10^{10}/°C.

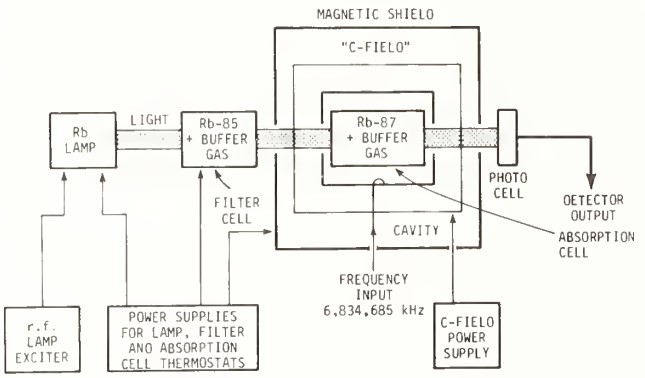


FIGURE 10. *Schematic of a rubidum gas cell frequency standard.*

Usually, the rubidium is in a liquid/gas equilibrium, provided via a relatively cooler appendix in the rubidium gas cell. The appendix is cooled via thermoelectric cooling or by providing a thermal gradient oven with higher temperatures for most of the gas cell with the exception of the appendix. Alkali-resistant glass must be used to minimize reaction as well as absorption and diffusion of rubidium into the glass. Some diffusion of rubidium into the glass always exists; thus, overfilling with rubidium is typically the procedure in manufacturing the lamp, filter, as well as the gas cell. Rubidium depletion into the glass is responsible for the fact that the rubidium vapor pressure is not in true equilibrium with the liquid phase as would be calculated from the temperature of the appendix containing the liquid phase. The lamp is typically excited by an rf-discharge in the hundred megahertz range. The rubidium gas cell is housed in a cavity with low Q. The microwave signal can be generated externally or by harmonic generation internal to the cavity using a varactor diode. A separate filter cell is not absolutely necessary. If the rubidium gas cell is filled with an isotopic mixture of rubidium 85 and rubidium 87, the light of the lamp penetrating the first layers of this isotopic mixture will shift in its spectral characteristics in such a way as to effectively pump the remainder of the gas cell. In other words, this "integrated" gas cell [18] acts like a filter cell at the side facing the lamp and, on the opposite side, like the gas cell resonator. The detector is typically a photocell, i.e., a photo-voltaic device, which has a surface area commensurate with the effectively used light beam (of the order of 1 centimeter diameter). Unlike the case of the cesium standard, there is no optimum power. The reason for this is the fact that we have no transit-time phenomena as in a cesium beam, but rather the relaxation spectrum of the rubidium atoms due to collisions or radiation phenomena. Best conditions are obtained by selecting the microwave power in such a way that the product of signal-to-noise and line Q is optimized.

As discussed above, the atomic ensemble does not perform spatial averaging but rather acts as a superposition of the individual resonance frequencies of the spatially fixed atoms leading to an inhomogeneous resonance line. As a result, dependencies occur against all parameters which are spatially selective [19]. This includes light intensity, magnetic field gradients across the cell, and microwave power due to the mode structure of the microwave cavity. In summary, the rubidium gas cell standard is a less "primary" device than the cesium beam tube. Basic frequency shifts due to buffer gas and light are of the order of parts in 10^9 to parts in 10^8. The rubidium standard thus is sensitive to the stability in time of those parameters which cause these shifts [20]. Therefore, it is no surprise that rubidium gas cell standards show systematic frequency changes on the order of 10^{-13} to 10^{-12}/day, corresponding to parameter changes as a function of time on the order of 10^{-4} to 10^{-5}/day.

One of the primary limitations of the rubidium gas cell standard is the fact that it features an inhomogeneous line. In order to transfer the rubidium gas cell into a standard with a homogeneous line, the buffer gas has to be essentially removed in order to realize effective spatial averaging of the individual atoms. As a result, wall collisions dominate as was shown experimentally [21] using high-polymer coatings such as paraffin and polyetheylene as wall materials. This approach allows the removal of the buffer gas while retaining the optical pumping principle with separate filter or isotope mixture in the gas cell. Unresolved at this time is the stability in time of the coating materials when exposed to rubidium.

Another limitation of rubidium standards appears to be the residual light shift. One portion of the light shift is due to the inhomogeneous line and would be cured with the wall coating approach and spatial averaging. The other remaining light shift is due to the insufficient overlap of the rubidium 85 absorption and the rubidium 87 emission (see Fig. 9). This leads to an asymmetric spectral filtering of the rubidium 87 line which pumps the lower microwave level. This light shift could be removed if a light source is employed which is centered and symmetrically coupled to the upper microwave level. Obviously, this can only be a laser source. The limitation of a practical standard would then be the stability of the laser frequency with temperature.

7. The Hydrogen Maser

An atomic hydrogen maser is depicted schematically in Fig. 11. It is shown in operation as an active maser oscillator. The beam source is a glass container in which atomic hydrogen is created in a radio-frequency discharge from molecular hydrogen. Molecular hydrogen is supplied from a hydrogen source, such as a bottle, or in some more recent experiments, from metal-hydrides at elevated temperatures. The hydrogen beam is formed by a collimator which can be a single channel or a multichannel design. The hydrogen beam intensity is determined by the power of the dissociating discharge, as well as the

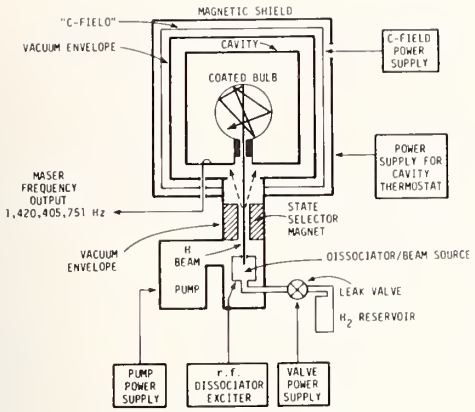


FIGURE 11. *Schematic of an atomic hydrogen maser frequency standard.*

flux of molecular hydrogen. In order to achieve stable beam intensities, it is mandatory to regulate both the power of the dissociating discharge as well as the supply of molecular hydrogen. Palladium metal has the property that, at elevated temperature, it has a very high diffusion coefficient for molecular hydrogen, while essentially remaining impermeable to all other gases. Thus via temperature changes it can operate as a regulated leak while at the same time purifying and rejecting impurities in the hydrogen gas. For effective dissociation, the typical pressure in the source which operates at around 100 MHz is 10^{-3} Torr.

The beam, which is formed at the output of the source, must be state selected [22]. The state selector almost always is a multipole, axially symmetric magnet, which achieves high field strengths while providing an axially symmetric state selected beam as well as some limited spatial focusing. The most probable velocity is focused on the entrance aperture of the storage bulb. The storage bulb is placed inside of a TE011 cavity. In order to retain a high cavity-Q, the storage bulb is typically made out of quartz which shows very low dielectric losses. The inside of the storage bulb is coated with a high-polymer substance. Fluorocarbons such as Teflon have provided the best surfaces, allowing highly elastic collisions. Typically, many thousands of collisions can take place before sufficient phase error is accumulated to de-correlate the radiation process. If the cavity-Q is high enough, i.e., cavity losses are low and the supply of state selected hydrogen is sufficiently high, the energy transferred into the cavity by the state selected hydrogen atoms exceeds the energy losses of the cavity and its associated electronics: the maser becomes an oscillator, providing a frequency close to the atomic resonance frequency of atomic hydrogen at 1,420,405,752 Hz. It is desirable to achieve oscillations at the lowest possible beam intensities to make the storage time in the storage bulb as long as possible. Since storage times in excess of one second are possible, the escape relaxation time of the bulb, i.e., the flow of gas in or out of the bulb, is arranged in such a way as to be just below the wall collision relaxation time of the hydrogen atoms for optimum performance. Storage times of about one second lead to line-Q values of above 10^9. These are the highest Q values in all presently used atomic standards, and are the basic reason for the exceptional performance of the atomic hydrogen maser. The output of the device is a signal at the picowatt level.

For an active maser oscillator, the crystal oscillator is still needed in order to provide a standard frequency output. A signal which is derived from the crystal oscillator is generated at the atomic hydrogen maser frequency. This signal is compared to the output of the atomic hydrogen maser using a superheterodyne, multistage receiver, as was discussed in section 4. Ultimately, the output of a phase detector is used to servo the crystal oscillator to be phase-coherent with the atomic hydrogen maser output signal. For this type of servo, which is phase sensitive, the frequency stability improves as the inverse of the averaging time τ, as compared to a $\tau^{-1/2}$ dependency for frequency-lock servos such as those in the "passive" cesium beam and rubidium gas cell standards. In an active device, cavity pulling is more pronounced [23]. Cavity pulling can be approximated by the following equation:

$$\nu - \nu_R = \frac{Q_c}{Q_l}(\nu_c - \nu_R) .$$

We note that the pulling factor translating the cavity offset into the frequency offset of the standard is the simple ratio of the cavity-Q and the line-Q as contrasted to the square of this ratio for passively operating devices (far below oscillating threshold). The pulling factor in typical hydrogen masers is of the order of 10^{-5} for the translation of cavity frequency offsets into output frequency offsets. Thus, active hydrogen masers are characterized by designs to stabilize the microwave cavity. This stabilization can be done by using low thermal expansion materials such as pure quartz as the cavity material within an oven enclosure. Such ovens may yield temperature stabilities of better than one millidegree.

There is an alternate way to stabilize the microwave cavity's resonant frequency [22]. By increasing the amount of hydrogen supplied to the storage bulb, an increasing number of spin-exchange collisions between the hydrogen atoms take place, shortening the radiation lifetime of the hydrogen atoms. The result is a broadening of the resonance line, i.e., a decrease in the line-Q. If the cavity is not tuned perfectly, the pulling equation leads to a frequency change as a result of the change in Q due to spin-exchange. Thus, a modulation of the hydrogen beam intensity results in a modulation of Q which can be used to determine the point where $(\nu - \nu_R)$ becomes zero and the cavity is properly tuned.

For hydrogen beam intensities insufficient to provide self-sustained oscillations, microwave radiation, i.e., its amplification, can be detected by using a microwave receiver, and a crystal oscillator can be servoed to the atomic resonance signal [24]. In this process, frequency or phase modulation in conjunction with either the absorption feature or the dispersion feature of the atomic resonance can be used to generate the reference signal for locking the crystal oscillator. This is the principle of the passive atomic hydrogen maser [25]. Since the detection of the atomic resonance still relies on the (phase-sensitive) measurement of the atomic radiation, the pulling equation for the active maser still applies. Therefore, cavity stabilization is as important for a passive maser as for the active one with the only difference that lower cavity-Q values may be used. An elegant method for cavity stabilization is the absolute probing of the cavity resonance by injecting sidebands into the wings of the cavity resonance and detecting any mistuning of the cavity. This is shown in Fig. 12 where two servos are employed [25], one at a high frequency such as 10 kHz to probe the cavity resonance and a much slower modulation, e.g., at 1 Hz, locking the crystal oscillator to the atomic reference frequency.

It must be noted that one of the exquisite features of the atomic hydrogen maser in the active mode is that its

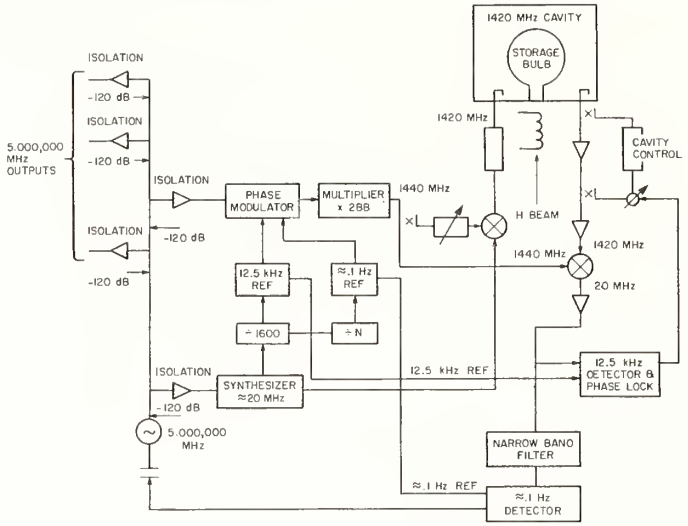


FIGURE 12. *Block diagram of a passive hydrogen maser frequency standard.*

stability improves as τ^{-1}. Stabilities of 10^{-15} at 1000 seconds have been achieved and stabilities in the 10^{-16} range have been repeatedly reported. We are reminded that the main reason for this superb performance of the hydrogen maser is the line-Q which exceeds that of the longest cesium beam tube by almost an order of magnitude. We remember that the concept of optimum power, so convenient in the case of the cesium atomic resonator, is absent in the case of the hydrogen storage principle, as it was in the case of the rubidium gas cell. Again, we have an exponential relaxation phenomenon in the storage bulb, which is due to the escape of atomic hydrogen out of the storage bulb. Thus, increasing microwave power leads to interrogation of atoms with lesser relaxation times and corresponding line broadening, much as in the case of the rubidium gas cell. Again, in the passive hydrogen maser, the microwave power is set in such a way as to maximize the product of line-Q and signal-to-noise.

It is difficult to imagine a storage-type device with a performance exceeding that of the hydrogen maser. This is based on the fact that the hydrogen atom is the smallest of all atoms and permits a maximum number of collisions with minimum relaxation.

As we discussed many times, the very high line-Q of the hydrogen storage principle leads to a very low flicker or frequency floor, as well as to excellent long term time-keeping performance. The key to frequency fluctuations in medium and long term is cavity pulling. The elimination of cavity pulling appears most elegant with the passive maser principle. The passive principle also allows less costly, more practical and smaller devices [26, 27, 28, 29]; however, the reduced cavity volume (e.g., of a dielectric cavity) leads to less available volume for the storage container, and thus increased wall-relaxation and consequential reduction in the line-Q. Unfortunately, for the purpose of this discussion, it appears that the reduced sensitivity to cavity pulling due to the lower cavity-Q and the passive control principle is compensated by the lower line-Q which leads to increased susceptibility to frequency perturbing parameter changes. Thus, it may be concluded that for excellence in clock performance (i.e., a very high line-Q), a so-called full-size passive hydrogen maser appears to be the most promising device realization.

8. Fundamental Advances

As compared to today's microwave frequency standards, fundamental improvements in both Q and signal-to-noise appear to be possible only by the use of higher frequencies, including infrared and optical frequencies. If it is assumed that these higher frequencies do not pose additional technical problems, the advantages are obvious: For the same length of the device, the Q improves with the resonance frequency. Thus a device with a resonance at 100 GHz and otherwise similar dimensions, has a 10 times improved Q as compared to cesium, and in the near infrared and visible regions, an enhancement by a factor of 10,000 is available. With such enhancement factors, Q could be traded against signal-to-noise, by building smaller devices making more efficient use of the spatial acceptance limitations, or by multiple interrogation of the atoms or molecules, cycling each atom several times through its two states. Another technically elegant solution is optical pumping for state selection and possibly also detection. Optical pumping was tried on the rubidium beam and, more recently, on the cesium beam, originally with rf-excited lamps, but more recently with lasers, including diode lasers and dye-lasers. The principle advantage of this technique is that it removes the distributed cavity phase shift limitation.

Apparently, ion storage is the only known technique offering line-Q's well in excess of 10^{10} in the microwave region. Although ions with resonance frequencies in the infrared or visible radiation region are typically considered, very attractive resonances can be found in the GHz region. Most importantly this includes mercury with hyperfine resonances at about 26 and 40 GHz [30, 31, 32, 33]. The achievement of very high line-Q's is a result of stable storage in electromagnetic fields [34]. The most attractive and elegant solution is the radio-frequency ion trap. An analysis of the frequency shifts due to this confinement method has shown their influence to be less than 10^{-14}. The main drawback of this method is the difficulty in determining the speed of the ions which is needed to calculate the second order Doppler effect.

Both the first and second order Doppler shifts could be reduced in a fundamental way if the atoms could be slowed down or cooled. Ion traps offer the possibility of using radiation pressure to cool. Recently tunable, narrow-band optical sources have become available which make detection of the small number of ions much easier than in the past and effective cooling possible. Typically, not many more than 10^4 ions can be stored and cooling of electromagnetically confined atoms or ions has been demonstrated [35, 36]. Optical pumping appears to be necessary to effectively interrogate any stored ion-ensemble without perturbation. Other auxiliary techniques in need of refinement for practical uses include the generation of ions and the interrogation by a microwave signal.

References

[1] H. Hellwig, Proc. IEEE **63**, 212 (1975).
[2] N. F. Ramsey, IEEE Trans. Instrum. Meas. **IM-21**, 90 (1972).
[3] N. F. Ramsey, *Molecular Beams* (Oxford University Press, London, 1956).
[4] R. C. Mockler, Adv. Electron Electron Phys. **15**, 1 (1961).
[5] H. Hellwig, K. M. Evenson, and D. J. Wineland, Phys. Today, December, p. 23 (1978).

[6] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McGunigal, J. A. Mullen, Jr., W. L. Smith, R. L. Sydnor, R. F. C. Vessot, and G. M. R. Winkler, IEEE Trans. Instrum. Meas. **IM-20**, 105 (1971).
[7] H. Hellwig, Radio Sci. 14, 561 (1979).
[8] M. Granveaud and J. Azoubib, IEEE Trans. Instrum. Meas. **IM-25**, 469 (1976).
[9] G. Becker, Metrologia 13, 99 (1977).
[10] D. J. Wineland, D. W. Allan, D. J. Glaze, H. W. Hellwig, and S. Jarvis, Jr., IEEE Trans. Instrum. Meas. **IM-25**, 453 (1976).
[11] A. G. Mungall, H. Daams, D. Morris, and C. C. Costain, Metrologia **12**, 129 (1976).
[12] L. L. Lewis, D. A. Howe, S. R. Stein, C. Manney, M. Mohler, J. C. Berquist, D. J. Wineland, and F. L. Walls, *CPEM Digest 1980* (IEEE Catalog No. 80CH1497-7 IM, 1980), p. 129.
[13] D. J. Wineland, Metrologia 13, 121 (1977).
[14] D. W. Allan, H. Hellwig, S. Jarvis, Jr., D. A. Howe, and R. M. Garvey, Proc. 31st Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C., 1977), p. 555.
[15] P. Davidovits and R. Novick, Proc. IEEE 54, 155 (1966).
[16] M. Tetu, R. Fortin, and J. Y. Savard, IEEE Trans. Instrum. Meas. **IM-25**, 477 (1976).
[17] G. Missout and J. Vanier, IEEE Trans. Instrum. Meas. **IM-24**, 180 (1975).
[18] E. Jechart, Proc. 27th Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C. 1973), p. 387.
[19] A. S. Risley and G. Busca, Proc. 32nd Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C., 1978), p. 506.
[20] P. Cerez and F. Hartmann, IEEE J. Quantum Electron. **QE-13**, 344 (1977).
[21] G. Busca, R. Brousseau, and J. Vanier, IEEE Trans. Instrum. Meas. **IM-24**, 291 (1975).
[22] D. Kleppner, H. C. Berg, S. B. Crampton, N. F. Ramsey, R. F. C. Vessot, H. E. Peters, and J. Vanier, Phys. Rev. 138, A972 (1965).
[23] J. Viennet, C. Audoin, and M. Desaintfucien, IEEE Trans. Instrum. Meas. **IM-21**, 204 (1972).
[24] H. Hellwig and H. E. Bell, Metrologia 8, 96 (1972).
[25] F. L. Walls and H. Hellwig, Proc. 30th Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C., 1976), p. 473.
[26] H. E. Peters, Proc. 32nd Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C., 1978), p. 469.
[27] H. T. M. Wang, Proc. 34th Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C., 1980), p. 364.
[28] E. M. Mattison, E.L. Blomberg, G.U. Nystrom, and R. F. C. Vessot, Proc. 33rd Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C., 1979), p. 549.
[29] D. A. Howe, F. L. Walls, H. E. Bell, and H. Hellwig, Proc. 33rd Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C., 1979), p. 554.
[30] M. Jardino, M. Desaintfuscien, R. Barillet, J. Viennet, P. Petit, and C. Audoin, Proc. 34th Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C., 1980), p. 353.
[31] H. A. Schuessler, Metrologia 7, 3 (1971).
[32] M. D. McGuire, Proc. 31st Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C., 1977), p. 612.
[33] D. J. Wineland, W. M. Itano, J. C. Berquist, and F. L. Walls, Proc. 35th Annual Symp. on Freq. Control (Electronics Ind. Assoc., Wash., D.C.), to be published.
[34] H. G. Dehmelt, Adv. At. Mol. Phys. 3, 53 (1967).
[35] D. J. Wineland, R. E. Drullinger, and F. L. Walls, Phys. Rev. Lett. 40, 1639 (1978).
[36] W. Neuhauser, M. Hohenstatt, P. Toschek, and H. Dehmelt, Phys. Rev. Lett. 41, 233 (1978).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Performance of the Three NRC 1-Meter CsVI Primary Clocks

A. G. Mungall, H. Daams, and J.-S. Boulanger

Division of Physics, National Research Council, Ottawa, Ontario, Canada, K1A OR6

The performance of the three 1-m interaction length NRC primary cesium clocks, CsVIA, CsVIB, and CsVIC, is outlined for their initial year and a half of operation as primary clocks, which commenced in December, 1979. Measurements of their systematic corrections made then and again early in 1981 indicate only small changes, within the accuracy limit of $\pm$ 1.5 $\times$ 10^{-13}. The frequencies of the three clocks compared with that of CsV remained within $\pm$ 2 $\times$ 10^{-13} despite certain problems concerning the electronics systems which became apparent during this test period and which limited their long-term frequency stability. These problems, which caused frequency shifts arising from spectral impurities in the microwave exciting signal, have been partially corrected, with consequent improvement of the agreement between all four NRC primary clocks to better than 5 $\times$ 10^{-14}. The potential long-term frequency stability appears to be a few parts in 10^{15}.

Key words: frequency stability; performance; primary cesium clocks; spectral impurities; systematic corrections.

1. Introduction

When the CsVI clocks were designed early in 1976, the intent was to achieve a stability as good as that of the NRC long beam primary cesium clock, CsV, and an accuracy only slightly less, in a unit much smaller in physical dimensions at a significantly lower cost [1–4]. It was also hoped that clocks embodying such a design might prove attractive from the point of view of later commercial exploitation, since their potential accuracy would be about 50 times better than that of currently available commercial clocks. Three cesium clocks were built of this design, as shown in Fig. 1, during 1977 and 1978, and began operation as secondary clocks late in 1978. As mentioned in an earlier publication [5], which outlined their salient design characteristics, certain difficulties arising from thermo-electric currents flowing in the magnetic shields and the aluminum interior support cylinder, and from mechanical distortions were encountered. The former were overcome by improved clock and room temperature control, and the latter by spring-mounting the ion pumps and the microwave excitation system.

Following an initial test, evaluation, and adjustment period, a complete accuracy evaluation was made of each

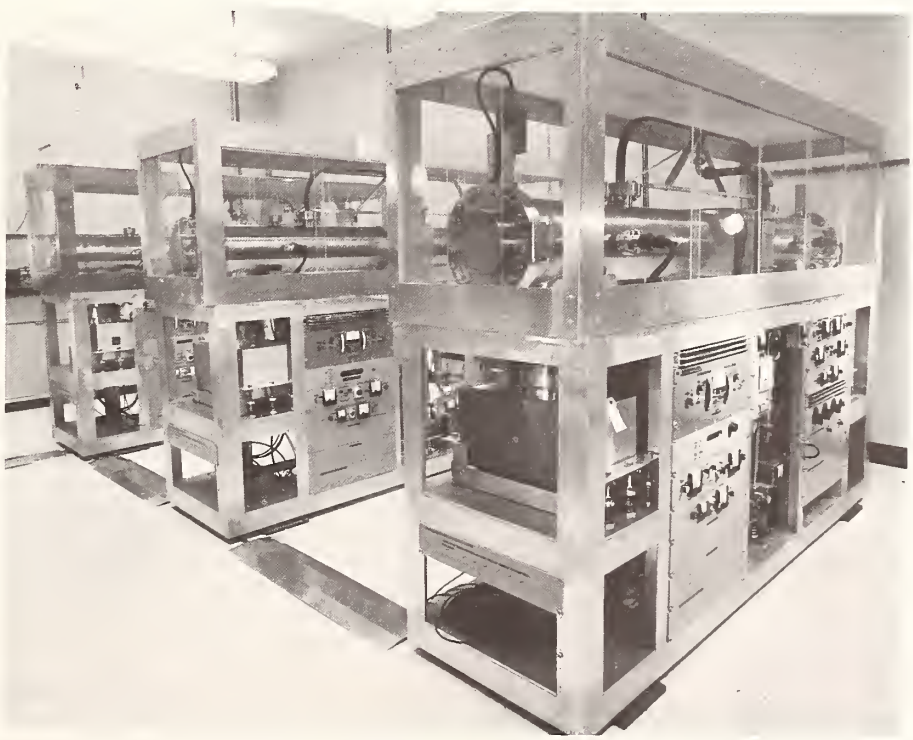


FIGURE 1. *The three CsVI clocks.*

clock late in 1979. This was repeated early in 1981, and only minor alterations in the systematic corrections were observed.

Since December, 1979, all three clocks have operated completely independently as primary standards, producing the scales of proper time PT(NRC CsVIA), PT(NRC CsVIB), and PT(NRC CsVIC). The relations between these three scales and UTC(NRC), which is based solely on CsV, have been reported at 10-day intervals to the Bureau International de l'Heure.

This paper will describe their performance during the first year and a half of their operation as primary clocks, based not only on their accuracy evaluations, but also on comparisons between their scales of proper time and UTC(NRC), and between UTC(NRC) and TA(NRC), and UTC(USNO), UTC(BIH), TA(NBS), and TA(PTB).

2. Results

2.1 Accuracy Evaluations

It is essential in primary clock operation that the various systematic corrections such as the cavity phase difference, the second order Doppler shift, and the frequency offset dependent on the direction of the C field remain constant. They should also be as small as possible, with magnitudes similar to that of the total clock uncertainty. The two complete evaluations so far performed on the CsVI clocks show that these two requirements have been met. None of the corrections exceeds 2 or 3 $\times$ 10^{-13}, and each remains constant within several parts in 10^{14} except when known changes have been made or have occurred to some component of the clock.

The latter proviso is particularly applicable in the case of the C-field reversal effect, which appears sensitive primarily to changes in the form of the C field in the vicinity of the two microwave interaction regions. Such a sensitivity is to be expected on the basis of an analysis presented in a previous publication [6]. Measurements of the C-field uniformity, using the set of six axial coils which excite the (4, -4) $-$ (4, -3) transition, are performed regularly to check the constancy of this frequency shift.

Measurements during the first year of operation, during 1980, disclosed unexpected instabilities in the microwave excitation systems which originated in all the component units, the 5–90 MHz frequency multiplier, the 12.6 MHz frequency synthesizer, the step recovery diode matching network, and the Gunn diode phase-locked frequency source. These all affected the spectral purity of the beam excitation signal and led to time-dependent frequency shifts. Modifications made to all these units have improved their performance, but further work is necessary to attain consistently the long-term frequency stability of which the clocks are capable.

2.2 Short-Term Frequency Stability

Measurements of the short-term frequency stability of each of the CsVI clocks intercompared, or compared with CsV, indicate that for values of τ between 2 and 10^5 s, the value of $\sigma(2, \tau)$ for any combination of the clocks can be described by the expression $4.6 \times 10^{-12}\tau^{-0.52}$. The value of $\sigma(2, 10^5\ \text{s})$ is about 1.2×10^{-14}, or 8×10^{-15} per clock. The dependence of σ on τ is shown in Fig. 2.

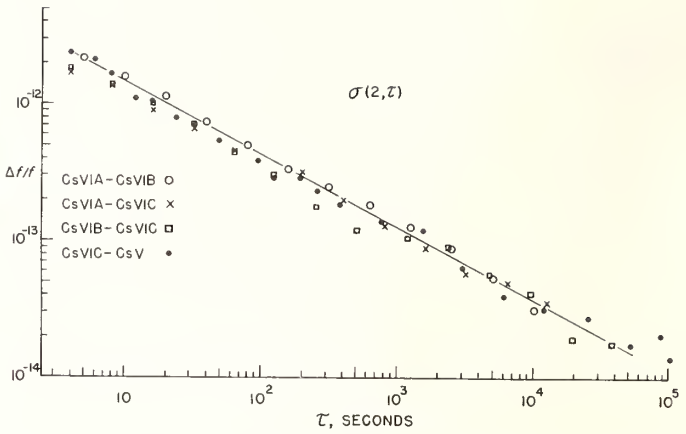


FIGURE 2. *The short-term frequency stability of the CsVI clocks both intercompared and compared with CsV.*

2.3 Long-Term Frequency Stability

The long-term frequency stability of the CsVI clocks is indicated in Fig. 3, which shows comparisons of their scales of proper time and UTC(NRC), based on CsV. Also shown in the figure are comparisons of UTC(NRC) and TA(NRC) with UTC(BIH), UTC(USNO), TA(PTB) and TA(NBS). The complete accuracy evaluations for both CsV and the CsVI clocks are indicated in Fig. 3 as well as the dates for a number of partial evaluations of the latter. Also shown are the times at which a variety of clock adjustments or equipment modifications were made. In a number of cases, slope changes, indicating frequency shifts, are evident at these times. The figure shows that throughout the measurement period the relative frequencies of CsV and the CsVI clocks were generally within $\pm 2 \times 10^{-13}$ despite recognized difficulties with the electronics systems. Subsequent to the 1981 CsVI evaluations, and after partial correction of these difficulties, this value decreased to about 5×10^{-14}.

During periods for which no adjustments or modifications were made the long-term relative frequency stability, $\sigma(2, \tau)$, drops below 1×10^{-14}, as indicated in Table 1. This table also shows that if the C-field frequency steering corrections required in primary clock operation are taken into account then the value of $\sigma(2, \tau)$ decreases to a few parts in 10^{15}.

Examination of the lower portion of Fig. 3 showing comparisons between the NRC time scales based on CsV and those of the BIH and other national laboratories indicates that the frequency variations between these scales are of the order of $\pm 2 \times 10^{-13}$ with the sole exception of TA(NRC) $-$ TA(PTB), which exhibits maximum changes of about $\pm 1 \times 10^{-13}$. These comparisons, based principally on LORAN C measurements, are subject to daily and annual signal propagation delay fluctuations, but the latter appear to be appreciably smaller than the variations shown in the figure. All these variations are similar in magnitude to those measured for the CsVI clocks with respect to CsV during a period when known causes existed to explain the CsVI frequency fluctuations. It should therefore be possible to reduce the limits of the fluctuations between the NRC primary clocks to values significantly lower than those at present occurring between the scales of the various different laboratories.

3. Conclusions

The three 1-m NRC CsVI clocks, now in operation for about two and a half years, provide first order accuracy

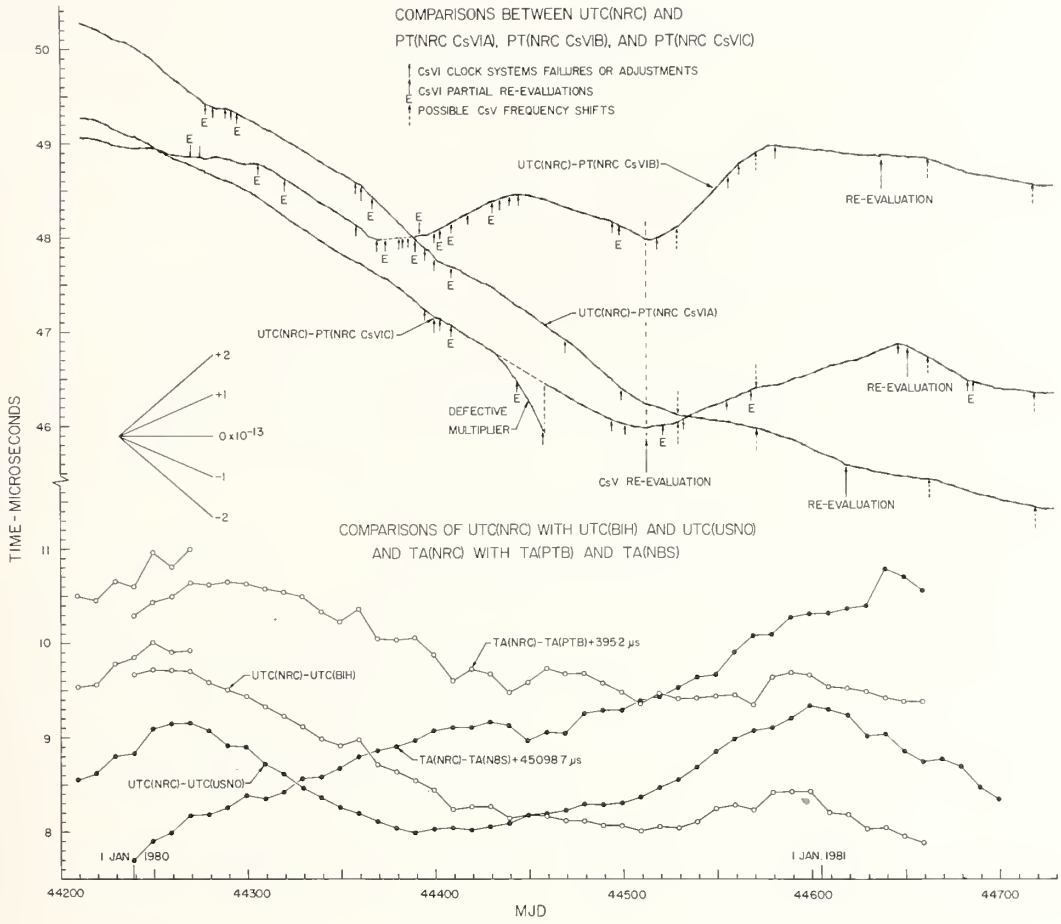


FIGURE 3. *Time scale comparisons between UTC(NRC) and PT(NRC CsVIA), PT(NRC CsVIB), and PT(NRC CsVIC) shown in the upper portion, and UTC(NRC) and TA(NRC) with UTC(BIH), UTC(USNO), TA(NBS), and TA(PTB) in the lower portion.*

TABLE 1. *Relative frequency stability for the NRC scales of proper time, with and without corrections for C field steering of the clock frequencies*

Time scales compared	Meas. Period MJD to MJD		Total time, days	τ days	$\sigma(2, \tau) \times 10^{-15}$ Uncorrected	Corrected
CsVIA-CsV	44303	44354	51	17	6.2	2.2
CsVIB-CsV	44583	44658	75	25	1.2	1.7
CsVIC-CsV	44301	44430	129	43	6.3	3.6

and frequency stability in a clock appreciably smaller than CsV or other comparable primary standards. It is likely that further improvements in the electronics systems will lead to the realization of a long-term stability of a few parts in 10^{15}.

References

[1] A. G. Mungall, R. Bailey, H. Daams, D. Morris, and C. C. Costain, Metrologia **9**, 113 (1973).

[2] A. G. Mungall, H. Daams, D. Morris, and C. C. Costain, Metrologia **12**, 129 (1976).

[3] A. G. Mungall and C. C. Costain, Metrologia 13, 105 (1977).

[4] A. G. Mungall, IEEE Trans. Instrum. Meas. **IM-27**, 330 (1978).

[5] A. G. Mungall, H. Daams, and J.-S. Boulanger, IEEE Trans. Instrum. Meas. **IM-29**, 291 (1980).

[6] A. G. Mungall, Metrologia **12**, 151 (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Prospects for Cesium Primary Standards at the National Bureau of Standards

L. L. Lewis, F. L. Walls, and D. A. Howe

Time and Frequency Division, National Bureau of Standards, Boulder, CO 80303

An application of optical pumping, in conjunction with a number of design improvements, may permit the development of a cesium primary standard with an accuracy an order of magnitude better than that of our present primary frequency standards, NBS-4 and NBS-6. Limitations to short-term stability, as well as possible errors in accuracy, are discussed.

Key words: atomic clock; atomic frequency standards; cesium frequency standard; light shift; optical pumping.

1. Introduction

The United States' primary frequency standard, NBS-6, was placed into service in 1975. This cesium standard has an accuracy of about 8×10^{-14}, long-term stability better than 1×10^{-14}, and short-term stability of about $5 \times 10^{-13}\ \tau^{-1/2}$. Research in progress at NBS is directed towards the design of a new cesium standard with an order of magnitude improvement in accuracy and long-term stability and a factor of two or three improvement in short-term stability. Much of this improvement may be obtained through the use of optical pumping state preparation and fluorescence detection, and the rest of the advancement will come from a number of technological innovations in the clock system.

The accuracy of NBS-6 is limited in large part by microwave phase shifts associated with losses in the Ramsey cavity [1]. The results of the 1980 evaluation of NBS-6 are given in Table 1, clearly illustrating this fact. The size of the apparent shift of the microwave resonance frequency is approximately

$$\Delta \nu_\phi = \frac{\Delta \phi}{\pi}\, \Delta \nu_\mu, \tag{1}$$

where $\Delta\phi$ includes both end-to-end and distributed cavity phase shift, and $\Delta \nu_\mu$ is the microwave linewidth of the central Ramsey feature. Since $\Delta\nu_\mu \sim 30$ Hz, one concludes from Table 1 that $\Delta\phi \sim 3 \times 10^{-4}$ radian. This is a reasonable value for a cavity structure as large as the one used in NBS-6 [2]. Assuming a linear dependence of the distributed cavity phase shift upon position across the microwave cavity window of $\sim 1 \times 10^{-14}$ rad/mm [1, 2], it would be necessary to obtain retrace of the cesium beam to within $\sim 100\ \mu$m in order to ensure a frequency error associated with the cavity phase shift of less than 10^{-14}. We believe that the use of optical pumping state preparation and fluorescence detection techniques in a cesium atomic beam frequency standard should make

TABLE 1. *NBS-6 uncertainties, 1980 evaluation.*

		Source of Uncertainty	Bias ($\sigma_y \times 10^{13}$)	Uncertainty ($\times 10^{13}$)
1.	(a)	Cavity Phase Shift (for one direction) (residual first-order Doppler Shift)	3.3 (typical)	0.80
	(b)	Second-order Doppler Shift	−2.8 (typical)	0.10
2.		Pulling by neighboring transitions	+0.3	0.20
3.		Magnetic Field Effects		
		(a) Offset due to finite field	+1767 (typical)	0.02
		(b) Magnetic field inhomogeneity	+0.02	0.02
		(c) Majorana Transitions	—	?
4.		Servo System Offset		
		(a) Amplifier offsets	0	0.1
		(b) Second harmonic distortion	0	0.2
5.		RF Spectrum	0	0.1
6.		Cavity Pulling	0	0.01
		RMS error due to systematic frequency biases		0.87
		Random Uncertainty		0.15

such a precision retrace possible. In addition, different microwave cavity structures than that used in NBS-6 may reduce the cavity phase shift dependence upon position, thereby relaxing the retrace requirement by as much as an order of magnitude.

The effect of Majorana transitions on the accuracy of Cs standards is not clear. Various authors [3] have suggested that they may produce uncertainties as large as one part in 10^{12}. If optical pumping is used for state preparation and detection, such Majorana transitions should not occur, since a uniform C-field strength can be maintained throughout the clock. In addition, by using laser techniques, it should be possible to measure Majorana transitions which might occur.

Optical pumping may reduce other sources of uncertainty as well. If only a single magnetic sublevel is prepared, there will be no $\Delta m_f = 0$ neighboring transitions which would shift the central resonance. This would permit a much lower value of C-field, which would relax restrictions on measurement of the finite field, as well as reduce the effect of magnetic field inhomogeneities and magnetic field changes with time.

The other contributions to uncertainty listed in Table 1 are not considered to be serious limitations to accuracy at the 10^{-14} level, assuming some reasonable improvements in electronics and measurement techniques are made. However, as discussed below, the introduction of optical pumping will create new sources of error which must be considered.

2. Optical Pumping and Fluorescence Detection

The development of high performance, cw, single-mode diodes for the communications industry has fortunately provided an optical pumping source appropriate for use in atomic frequency standards. The linewidth, intensity noise, frequency noise, and wavelength tuning characteristics of these devices make them very attractive for this purpose [4]. At the present time, it is reasonable to expect laser diodes to operate tuned to an atomic transition for months at a time.

The first use of laser diodes to optically pump a cesium beam atomic clock was made by Arditi and Picqué [5]. They used a single GaAlAs laser tuned to the D_2 line of Cs (852 nm) in order to pump atoms into one of the $F=3, 4$ ground-state hyperfine levels (Fig. 1). Detection of a microwave transition within the Ramsey cavity was accomplished by fluorescence detection using the same laser.

The work at the National Bureau of Standards (NBS) has moved in a somewhat different direction, requiring more than one laser. It is possible to pump nearly every Cs atom in an atomic beam into a single magnetic sublevel [6]. If one laser is tuned to the $6\ ^2S_{1/2}\ F=3 \rightarrow 6\ ^2P_{3/2}\ F'=4$ transition, and a second laser is tuned to the $6\ ^2S_{1/2}\ F=4 \rightarrow 6\ ^2P_{3/2}\ F'=4$ transition with plane electric polarization parallel to a weak magnetic field (π-polarization), only atoms in the $F=4$, $m_F=0$ sublevel of the ground state will remain unaffected. This selection rule is clear when one notes that the Clebsch-Gordon coefficient $\langle Fjm_Fm_j \mid F'm_{F'}\rangle = \langle F\,100 \mid F\,0\rangle = 0$, where j is the angular momentum of the photon. Eventually, most of the atoms will be pumped into this magnetic sublevel. Alternatively, a similar arrangement can pump atoms into the $F=3$, $m_F=0$ state.

Fluorescence detection can be performed with a third laser which is tuned to either the $F=4 \rightarrow F'=5$ or the

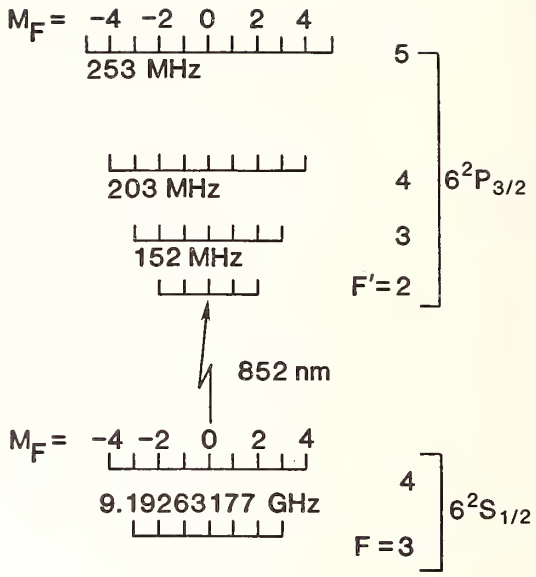


FIGURE 1. *Term diagram for ^{133}Cs.*

$F=3 \rightarrow F'=2$ transition. In these two cases (with appropriate polarization of the $F=3 \rightarrow F'=2$ laser light to avoid pumping into the $m_F=\pm 3$ sublevels), atoms in the excited state return to the original ground-state hyperfine level. Thus, a large number of fluorescence photons (limited by excited-state energy separations) may be obtained from each atom in the atomic beam. This technique should permit unity quantum detection efficiency, even if the light collection efficiency is considerably less than one. However, as discussed below, there may be systematic effects which would make this detection method undesirable. In this case, it is still possible to use a laser tuned to a pumping transition, and take care to obtain a high collection efficiency at the detector. An additional advantage a pumping transition provides in detection is that every atom is weighted equally regardless of velocity, which may provide immunity from certain systematics. On the other hand, detection by fluorescence on a cycling transition produces a relatively larger signal for slower atoms which remain in the laser beam longer, which would increase the effective microwave Q. Still another advantage of optical pumping is that it should be possible to operate atomic beams in opposite directions simultaneously without interference. This would permit direct measurement of retrace, as well as very rapid modulation of beam direction (or even continuous operation of two beams) for purposes of cancellation of cavity phase shift. A final consideration in the use of fluorescence detection is that the laser beam can be made to intersect the atomic beam at a slight angle, thereby selecting low velocity atoms. This would cause an increase in the microwave transition Q, with some accompanying loss of signal. This result is a consequence of the Doppler shift, which changes the effective velocity distribution of the cesium beam. As further experiments are performed, the relative merits of these various detection methods will become clearer.

Using values of atomic beam current $I_B = 1.0$ nA (6×10^9 atoms/s) and microwave transition $Q \sim 10^8$, it should be possible to obtain stabilities of $\sim 1 \times 10^{-13}\ \tau^{-1/2}$, even if a pumping transition is used for detection purposes. The high beam current suggested here would come partially from the increased number of useful cesium atoms (a factor of 8 more than that of NBS-6,

through optical pumping of all magnetic sublevels), and partially from new oven design. A reflux-type oven is presently under consideration.

3. Additional Uncertainties and Biases

Perhaps the most serious source of frequency uncertainty introduced by optical pumping techniques is that of near-resonant light shifts [7, 8]. Although it should be possible to prevent stray laser light from entering the microwave region, fluorescence light from the atoms will pass through the cavity, and interact with atoms in the "C" region, changing the ground-state hyperfine splitting slightly. The form of this light shift is dispersive, but averaging over excited-state hyperfine levels and including Doppler shifts associated with the Maxwellian velocity distribution in the atomic beam, gives a nonzero value for the shift. Preliminary calculations have been made at NBS [9] for the frequency shift from fluorescence light for pumping by two lasers of π-polarization, with the C-field transverse to the atomic beam direction, and including the effect of the tensor light shift [8]. The computations, assuming lasers of equal power driving the $F = 4 \rightarrow F' = 4$ and $F = 3 \rightarrow F' = 4$ transitions in the pumping region, give a light shift of approximately 4 Hz for a power flux of 1 $\mu W/cm^2$ at the site of the cesium atom in the C-field region. A simple computer calculation gives ten as the average number of photons emitted per atom in this pumping scheme. Referring to Fig. 2, reasonable values of the clock dimensions would be $l_1 = l_2 = l_3 = 50$ cm, and $L = 200$ cm. Assuming a beam flux of 6×10^9 atoms/s at the detection region, and a free aperture of 2 mm everywhere in the cesium beam, the total fluorescent power arriving at the first window of the Ramsey cavity from the pumping region would be about 2×10^{-5} $\mu W/cm^2$. This gives a fractional frequency shift of $\sim 9 \times 10^{-15}$, which is comparable to the desired accuracy. Additional collimation of the atomic beam before the pumping region would reduce this predicted light shift by a factor of ten or more. Nevertheless, a more careful calculation of the effect will be made, including the effective velocity distribution associated with cycling fluorescence detection, and including the shift caused by light originating in the detection region. In addition, versions of optically pumped frequency standards being considered contain provision for measurement of the light shift.

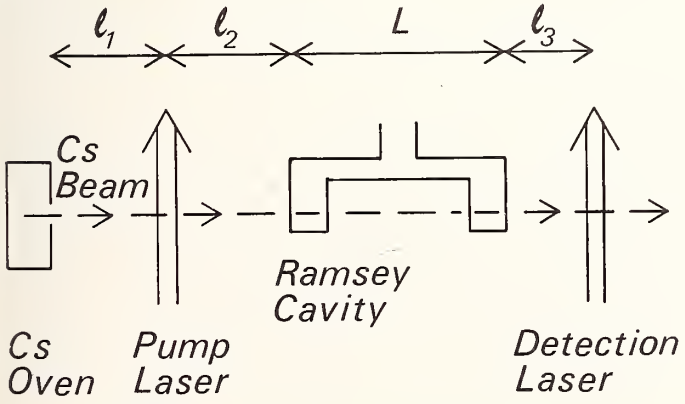


FIGURE 2. *Schematic of Cs atomic beam frequency standard.*

A second source of trouble associated with the laser optical pumping is deflection of the atomic beam through photon recoil. The deflection is about 1.6×10^{-5} rad/photon for an average atomic velocity of ~ 200 m/s, and proportionately greater for slower velocities. If the Cs beam is irradiated symmetrically on two sides, an expansion of the atomic beam should occur, with magnitude $\sim 5 \times 10^{-5}$ rad. With the same assumptions used above to estimate the limitations imposed by beam retrace, this should give an uncertainty in frequency considerably less than 10^{-15}.

Still another correction to the Cs microwave frequency, not unique to optical pumping, but which has not been fully considered in the past, is the light shift caused by blackbody radiation. The ac Zeeman shift of the ground-state hfs of Cs, due to the magnetic field of the blackbody radiation is only $\sim 10^{-16}$ [11]. However, the light shift associated with the rms electric field of the blackbody radiation is considerably larger [10]. The approximate magnitude of the shift correction at $T = 300$ K is about

$$\frac{\delta\nu_{ss}}{\nu_{hfs}} = -1.8 \times 10^{-14} . \qquad (2)$$

The sensitivity of an anticipated new cesium primary standard would be adequate to measure this effect in a suitably designed apparatus.

4. Conclusion

Optical pumping techniques may improve both the accuracy and short-term stability of Cs primary frequency standards. The greatest anticipated improvement comes from better retrace upon beam reversal. New problems associated with laser diode optical pumping will require serious consideration, but are not seen as major obstacles to the design of an improved standard.

The authors wish to thank their colleagues in the Time and Frequency Division for many fruitful discussions. The contributions of D. Wineland, E. Smith, and D. Glaze were especially helpful.

References

[1] D. J. Wineland, D. W. Allan, D. J. Glaze, H. W. Hellwig, and S. Jarvis, Jr., IEEE Trans. Instrum. Meas. **IM-25**, 453 (1976).

[2] R. F. Lacey, Proc. 22nd Annual Symp. on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, NJ, 545 (1968).

[3] G. Becker, IEEE Trans, Instrum, Meas. **IM-27**, 319 (1978); D. W. Allan, H. Hellwig, S. Jarvis, D. A. Howe, and R. M. Garvey, Proc. 31st Annual Symp. on Frequency Control, Ft. Monmouth, NJ, 555 (1977); S. Urabe, K. Nakafiri, Y. Ohta, M. Kabayashi, and Y. Saburi, IEEE Trans. Instrum. Meas. **IM-29**, 304 (1980).

[4] L. L. Lewis and M. Feldman, Proc. 35th Annual Symp. on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, NJ (1981) 612.

[5] M. Arditi and J.-L. Picqué, J. Phys.-Lett. (Paris) 41, L-379 (1980); see also M. Arditi, these proceedings.

[6] This method was made known to us by L. Cutler. See also H. J. Gerritsen and G. Nienhuis, Appl. Phys. Lett. **26**, 347 (1975).

[7] We are grateful to A. Brillet for emphasizing the seriousness of this systematic to us.

[8] B. S. Mathur, H. Tang, and W. Happer, Phys. Rev. 171, 11 (1968); W. Happer and B. S. Mathur, Phys. Rev. 163, 12 (1967).

[9] E. Smith, private communication.

[10] W. M. Itano, L. L. Lewis, and D. J. Wineland, Phys. Rev. A 25, 1233 (1982).

[11] T. F. Gallagher and W. E. Cooke, Phys. Rev. Lett. **42**, 835 (1979).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A Cesium Beam Atomic Clock with Laser Optical Pumping, as a Potential Frequency Standard

M. Arditi

Institut d'Electronique Fondamentale*, Université Paris XI, Bât. 220, 91405 Orsay, France

A passive microwave cesium beam resonator using optical pumping and optical detection, with a cw tunable GaAs diode laser, has been realized. The "0-0 clock transition" is detected through a change in the intensity of the fluorescence of the cesium beam. Experimental recordings of the Ramsey pattern agree with a Maxwellian distribution of atomic velocities. Results of preliminary tests, to an accuracy of a few parts in 10^{11}, show good potential for a frequency standard of higher accuracy.

Key words: atomic clock; cesium beam; frequency standard; laser diode; optical pumping.

1. Introduction

With a cesium beam frequency standard of the Rabi-Ramsey type, an accuracy better than 10^{-13} in the realization of the second appears to be a realistic probability [1]. This accuracy is limited mainly by second-order Doppler-shift and cavity phase-shift whose estimation depends on a precise measurement of the velocity distribution of the atoms in the beam [2].

The purpose of this paper is to describe a cesium beam atomic frequency standard where optical pumping could, perhaps, further extend the limits of accuracy due to these effects.

As early as 1950, A. Kastler [3] had proposed to replace the Stern-Gerlach magnets in a Rabi classical atomic beam magnetic resonance apparatus by "optical pumping," to create and detect differences in the Zeeman sublevel populations of the ground state of the alkali atoms [4]. A successful application of these principles was obtained in a ^{87}Rb atomic beam to measure accurately the hyperfine structure separation of the ground state [5]. In that experiment, optical pumping was achieved using the isotopic shift of the ^{85}Rb and ^{87}Rb resonance lines emitted by spectral lamps. However, limited by a weak signal to noise ratio of the detection, the device could not be used conveniently as a practical frequency standard.

The advent of stable, monochromatic and tunable laser sources (diode lasers or dye lasers) in the near infra-red, had raised the hope of increasing the signal to noise ratio in optically pumped atomic beam clocks. After the successful experiments of J.-L. Picqué on hyperfine pumping of a cesium beam using a GaAs laser diode [6], an attempt was made to set up an optically pumped cesium beam clock with optical detection. The preliminary results were described in a brief report [7]. In this paper more details are given on the experimental procedure and on the potentialities for a fully developed frequency standard.

2. Optical Pumping of a Cesium Beam

The experimental set-up is shown schematically in Fig. 1. In region A, optical pumping increases the population difference between the $F = 4$ and $F = 3$ levels of the ground state of cesium atoms. In region C, microwave interaction takes place, and in region B the microwave resonance is detected with a photocell by a change in the intensity of the beam's fluorescence.

In this device, the atomic beam is axially symmetrical, and the velocity distribution of the cesium atoms is the same across the beam since there is no spatial selection as with the A magnet, thus reducing considerably the problem of the spatially distributed cavity phase-shift [8]. Moreover, the device can be made completely reversible by placing a cesium oven at each end, thus simplifying the retrace measurement of the residual cavity phase-shift. Also, as will be seen later, because of the high spectral density of the laser light, the cesium atoms will all be pumped with the same efficiency, regardless of their velocity, so that the velocity distribution will be broad, very close to a Maxwell-Boltzmann distribution and thus the second-order Doppler-shift can be more easily evaluated.

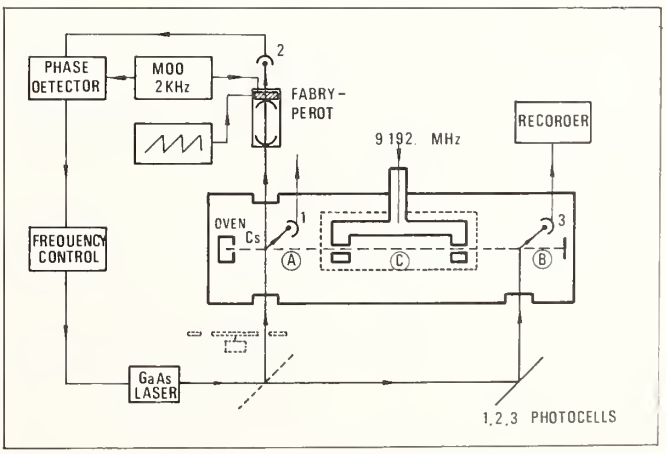


FIGURE 1. *Cesium beam resonator with optical pumping and optical detection.*

*Laboratoire associé au C.N.R.S.

The time evolution for the population of the ground state hyperfine Zeeman levels of cesium atoms has been previously analyzed for the case where a cesium beam is optically pumped under different conditions of polarization, either with a weak monochromatic laser light or with the resonant light from a spectral lamp filtered by a cesium cell [9].

With cesium spectral lamps and cesium filter-cells, no good experimental results could be obtained in population changes in a cesium atomic beam, although very good signal to noise ratio resulted in the optical detection of the 0-0 transition in a gas cell [10]. This failure could be attributed perhaps to a lack of pumping light due to too much light absorption by the filter-cell.

More successful results were achieved with a semiconductor laser diode (type LD-33, from Laser Diode Lab. Inc.) originally intended for pulse-operation and adapted by J.-L. Picqué for cw, single mode operation in a temperature-stabilized cryostat at about 25 K.

For an injection current of 200 mA this laser diode delivers a few milliwatts of power, and the output wavelength is coarsely adjusted in the vicinity of the cesium D_2 resonance line (852.1 nm) by changing the temperature (about 2.4 Å per °C) and finely tuned to one of the hyperfine components by varying the injection current (about 200 MHz per mA). The phase-jitter of the laser light can be reduced considerably by locking the frequency of the laser to an external Fabry-Perot resonator. More details on the operation of the laser diode and cryostat can be found in Ref. [11]. By sweeping the piezo-crystal of the Fabry-Perot with a dc voltage ramp, the laser frequency can be tuned over a wide range. A photocell, under the cesium beam, is used to detect the fluorescence of the beam.

The cesium atomic beam is produced in a stainless steel vacuum chamber about 1.5 m long, fitted with several glass windows to admit the light of an external laser diode. The collimator of the cesium oven is an array of about 185 nickel tubes (diam. = 0.25 mm, length = 10 mm) inside a 6 mm diam. cylinder. The directivity of the beam being rather poor at the oven exit, the cesium beam is further collimated by several apertures (diam. = 8 mm) distributed along the axis, and resulting in a beam divergence from the axis of less than 1 degree. The aperture supports are coated with lamp-black to act as an efficient getter for diverging cesium atoms, and a movable stop is provided in front of the oven to check that the detected signals are due to the atomic beam and not to some residual cesium vapor. At the end, an ion detector can be placed to give a measurement of the cesium atomic density in the detection region, which typically for an oven temperature of 96 °C is of the order of 2×10^8 atoms per cm^3. Right under the cesium beam, in the regions of illumination, silicon photocells are placed in shallow cups (to protect them from cesium deposits) to detect the fluorescence of the beam.

Figure 2(b) shows a recording of the derivatives of the absorption lines originating on the $F = 3$ level and detected by the fluorescence of the beam as the laser diode was swept in frequency. The line-width of each hyperfine component is about 50 MHz. Similar results were obtained for transitions originating on the $F = 4$ level.

Due to selection rules which allow only $\Delta F = 1$, $\Delta m_F = 0, \pm 1$ transitions to take place, atoms in the ground state undergoing $3 \rightarrow 2$ or $4 \rightarrow 5$ transitions can only return to their original level and the total population of these ground levels should not change due to these

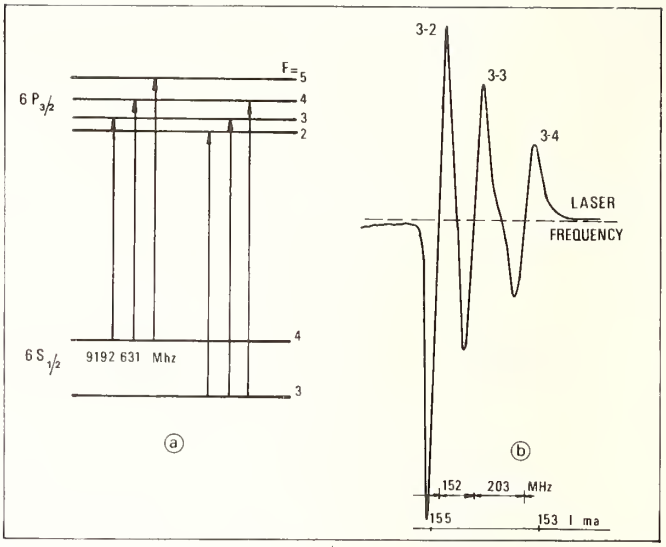


FIGURE 2. *a) Energy levels for cesium D_2 transitions; b) Derivatives of D_2 absorption lines originating on $F = 3$ level, recorded by fluorescence of a cesium beam.*

two transitions. However, as was shown in Ref. [9] when detailed balancing of the populations of all the Zeeman sublevels is carried out, a population difference can appear between the $F = 3$, $m_F = 0$ and $F = 4$, $m_F = 0$ levels, even though the overall population of the ground level should not change.

To study in detail these effects, the set-up of Fig. 1 was completed by placing a rotating light-chopper in front of the pumping window. Without microwave excitation, by alternatively pumping the cesium beam on and off, the populations are redistributed in the various Zeeman sublevels, and an unmodulated probe beam from the laser can be used to detect the total population changes by monitoring the modulation of the beam fluorescence.

Figure 3(a) shows such a recording obtained by sweeping the laser frequency. It can be seen that the largest signal is obtained with the laser diode tuned to the $3 \rightarrow 3$ transition frequency. It is interesting to note also that a distorted signal is detected at the $3 \rightarrow 2$ transition although, according to the selection rules, the total population of the $F = 3$ level should not change. Similar recordings were obtained with the laser tuned to the frequencies of the transitions originating on the $F = 4$ level. Changing the polarization of the laser light from linearly to circularly polarized, in all possible combinations for pumping or detecting, did not change significantly the signal shapes or magnitudes and it is finally simpler to use the light as it comes directly from the laser diode.

Since there is no saturation with the laser power involved in these experiments, it is easy to compute the relative change of fluorescence in the detection region B produced by optical pumping of the cesium beam in region A: here, it is about 7 percent, for the value of $I_o t$ corresponding to experimental conditions (I_o = pumping laser intensity: 6×10^{15} photons/sec, and t = pumping time illumination: about 50 microseconds). A dual frequency optical pumping scheme, with two lasers, has also

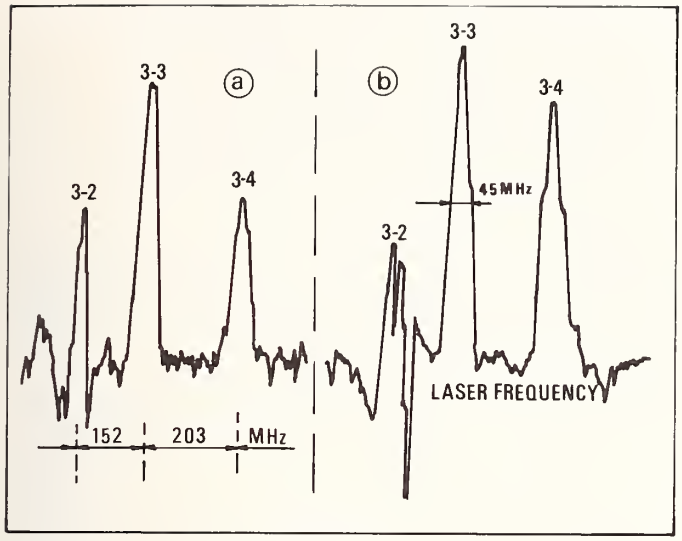


FIGURE 3. *a) Optical detection of population change in $F = 3$ level produced by optical pumping of cesium beam, b) Optical detection of 0-0 microwave transition.*

been suggested, by which all of the atoms could be pumped into the $F = 3$, $m_F = 0$ (or $F = 4$, $m_F = 0$) level [12].

With the same experimental set-up, if in addition the microwave cavity is excited at the 0-0 frequency, i.e., 9192 , . . . MHz, it is found that the signals previously recorded in Fig. 3(a) are reduced by about 1/7 of their value, indicating that the change of fluorescence produced by the microwave interaction is about 1 percent of the total fluorescence of the cesium beam in B. This is also in agreement with the value given by a simple linear theory.

To increase the signal to noise ratio of the optical detection of the 0-0 transition, of interest in atomic clocks, the experimental set-up was modified as follows:

The light-chopper is removed and a low frequency modulation is applied to the frequency of the microwave oscillation. A maximum signal is obtained in photocell B when the microwave frequency is set to the side of maximum slope of the 0-0 resonance curve. With the microwave frequency fixed at this value, the laser diode is swept in frequency and the signal out of photocell B is recorded with a synchronous detector. Figure 3(b) shows such a recording (similar results were obtained for laser frequencies corresponding to transitions originating on the $F = 4$ level). Here again there is a predominance of signal at the $3 \rightarrow 3$ transition frequency and also an unstable and peculiar shape signal from the $3 \rightarrow 2$ transition frequency. In this 0-0 detection, the width of the laser frequency range corresponds to the width of the absorption lines in the cesium beam, about 50 MHz.

In conclusion: from these experiments it appears that the laser frequency giving the largest signal in the optical detection of the 0-0 transition is the frequency of the $3 \rightarrow 3$ transition, although, as we shall see later, the $3 \rightarrow 4$ transition frequency gives a better signal to noise ratio of the detection (in most of these experiments, the photon flux collected by fluorescence by the photocell in B was of the order of 5×10^{10} photons/sec).

For applications to atomic frequency standards, it is much better to lock the laser frequency directly to one of the absorption lines of the cesium beam itself [11]. As shown in Fig. 2(b), the derivative of these absorption lines are narrow-width S curves with a well defined cross-over which can be used for locking the laser frequency, and this was done in the following experiments.

3. Optical Detection of the 0-0 "Clock Transition"

The experimental set-up of Fig. 1 was modified as follows: the laser frequency was set to correspond to one of the absorption lines of the cesium beam and was slightly frequency modulated (at a 3 kHz rate, for example). The modulated fluorescence of the cesium beam was detected by a photocell in A feeding a phase detector. The output of this phase detector controls the current of the laser diode, thus locking the laser to the absorption frequency of the cesium beam (see also Fig. 5).

The microwave cavity was of the Ramsey-type with oscillating fields separated by a distance L. Circular holes (8 mm diam.) were provided at cavity ends for the passage of the atomic beam. This cavity was made of a rectangular waveguide operating in a TE_{01N} mode and is resonant for a length equal to $N\lambda_g/2$. Experimentally two different cavities have been used: one with $L = 21.5$ cm ($N = 14$) and the other with $L = 36.5$ cm ($N = 20$), in order to compare the experimental resonant curves with the theory. The cavity was finely tuned to resonance with a $E - H$ waveguide tuner in the cavity feed.

The uniform magnetic field in region C was produced by a U shaped electromagnet of Conetic metal. The whole assembly, cavity and electromagnet, was placed inside a cylindrical magnetic shield of μ-metal which was demagnetized by means of a 50 Hz ac current. The distribution of the residual magnetic field inside the cavity was measured with a Hall effect magnetometer probe. The mean value of the field could be obtained also, in operation, by inducing Zeeman low frequency transitions with two coils placed at the cavity ends, which at resonance perturbed the 0-0 detected signal.

A photocell in B detected the change of fluorescence of the cesium beam when the 0-0 transition was induced by microwave resonance. To study the shape of the 0-0 resonance curve, the microwave frequency was swept slowly around the resonance frequency. By modulating, at low frequency, the frequency of the microwave excitation, the output of a phase detector fed by photocell B gave a dc signal, the polarity of which depended on the relative position of the microwave excitation and the 0-0 resonance frequencies. By sweeping the exciting frequency this signal gave the derivative of the resonance curve. An experimental curve is shown in Fig. 4 for a cavity with $\ell = 1$ cm and $L = 21.5$ cm (ℓ = length of the oscillating field region, and L = distance between the two separated oscillating fields), and an oven temperature of about 85 °C.

According to Ramsey [13] the separated oscillating fields transition probability for a cesium atom in the beam, with velocity v, is:

$$P = \sin^2(2b\tau)\cos^2[(\omega_0 - \omega)L/2v], \qquad (1)$$

where b the Rabi nutation frequency, is proportional to the square-root of the microwave power in the cavity, ω_0 the Bohr resonant frequency, ω the excitation frequency,

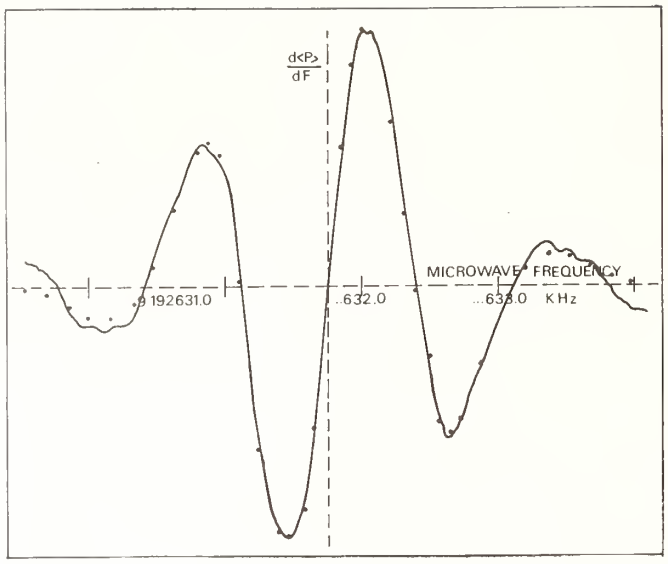


FIGURE 4. *Experimental derivative of Ramsey pattern of cesium beam resonator with optical pumping by laser (solid curve) and theoretical values (dots) for Maxwellian distribution of atomic velocities (sweep speed: 16 Hz/s; recorder time constant: 3 s).*

and τ the time the atom spends in each end of the cavity.

Assuming the usual Maxwellian velocity distribution for effusive beams:

$$f(v) = 2(v^3/\alpha^4)\exp(-v^2/\alpha^2), \tag{2}$$

where α is the most probable velocity in the atomic beam, $\alpha = (2kT/m)^{1/2}$ (m, mass of the cesium atom; T, temperature of the cesium oven), the transition probability averaged over this velocity distribution is given by:

$$<P> = 2\int_0^\infty e^{-y^2} y^3 \sin^2\left(\frac{2b\ell}{\alpha y}\right) \cos^2\left[\frac{(\omega - \omega_0 L}{2\alpha y}\right] dy \tag{3}$$

with

$$y = v/\alpha \tag{4}$$

Near resonance, the derivative of $<P>$ versus frequency reduces to:

$$\frac{d<P>}{df} = (1/4)K'(\chi + \frac{4b\ell}{\alpha}) + (1/4)K'(\chi - \frac{4b\ell}{\alpha}) - (1/2)K'(\chi), \tag{5}$$

with $\chi = (\omega - \omega_0)L/\alpha$ and $K'(\chi)$

$$= \int_0^\infty e^{-y^2} y^2 \sin(\chi/y)\, dy. \tag{6}$$

On Fig. 4 such a theoretical curve has been plotted for the following values of the parameters: $\ell = 1$ cm, $L = 21.5$ cm, $\alpha = 213$ m/sec and $2b\ell/\alpha = 2.0$ (corresponding to a microwave signal power slightly over that for optimum transition probability) and with the ordinate of the curve normalized to coincide with the maximum of the experimental curve. The agreement between the experimental curve and the computed points is quite good, and gives evidence of a broad Maxwellian distribution of velocities in the beam. The frequency width between the central peaks of the derivative curve is about 500 Hz and 320 Hz for the microwave cavities with lengths $L =$ 21.5 cm and 36.5 cm, respectively.

The shape of the resonance curve in this experiment is different from the one obtained in the optical pumping of a rubidium beam with spectral lamps [5] where there was a line-narrowing due to an enhanced contribution of the slower atoms to the overall signal [4]. This is probably due to the fact that the spectral density of the monochromatic laser radiation is about 10^3 times higher than for a spectral lamp, so that the efficiency of optical pumping is about the same for all the atoms, regardless of their transit time across the light beam. However, the lack of this resonance narrowing with laser pumping is more than compensated by the large increase in the signal to noise ratio (S/N) of the detection in this case and this makes the device attractive for atomic clock applications.

In the present experiments, the S/N is limited by the noise due to residual phase-jitter of the laser diode. In this respect, it was observed that the S/N was very bad when locking the laser frequency to the $3 \rightarrow 2$ absorption line of the cesium beam, and better with the $3 \rightarrow 4$ than with the $3 \rightarrow 3$ transition. This effect may have been instrumental, the slope of the S curve (Fig. 2(b)) for the $3 \rightarrow 4$ transition being smaller than for the $3 \rightarrow 2$ or $3 \rightarrow 3$ transitions. Improvement in the S/N could be obtained by increasing the laser intensity at the pumping region, and by increasing the bandwidth of the amplifier and the loop gain of the laser frequency servo. Also electro-optical devices could be used to reduce amplitude noise of the diode laser. Further work is contemplated in that direction.

4. Operation as an Atomic Clock

4.1 Experimental Set-Up

Referring to Fig. 5, the microwave excitation signal is synthesized from a stable 5 MHz voltage-controlled quartz oscillator (H.P. 107 BR). The error signal from the phase detector acts on a voltage-variable capacitor to lock the frequency of this oscillator to the atomic transition. The output of the 5MHz oscillator is compared with the frequency of a commercial cesium-beam frequency standard (H.P. 5061 A) either by frequency multiplication and counting, or in a phase comparator. In this way the precision resetability of the clock and its accuracy could be checked versus various parameters such as: magnetic field, rf power, cavity tuning, rf modulation, laser light intensity or polarization, oven temperature, servo-gain, etc.

However, in these preliminary measurements, the electronics and the mechanical assembly were not sufficiently developed for measurements of 2nd-order Doppler shift or cavity phase-shift of the order of 10^{-12} to 10^{-14}. Since the commercial standard for frequency comparison has a limited accuracy of $\pm 2 \times 10^{-11}$, and with the inclusion of various biases in the electronics, it is believed that the absolute accuracy of the measurements in these first tests was perhaps no better than $\pm 3 \times 10^{-11}$. Also, since the S/N could probably be improved considerably with further work on the laser servo, the emphasis of the measurements was mostly on the accuracy of the clock rather than on short-term stability. Finally it should be

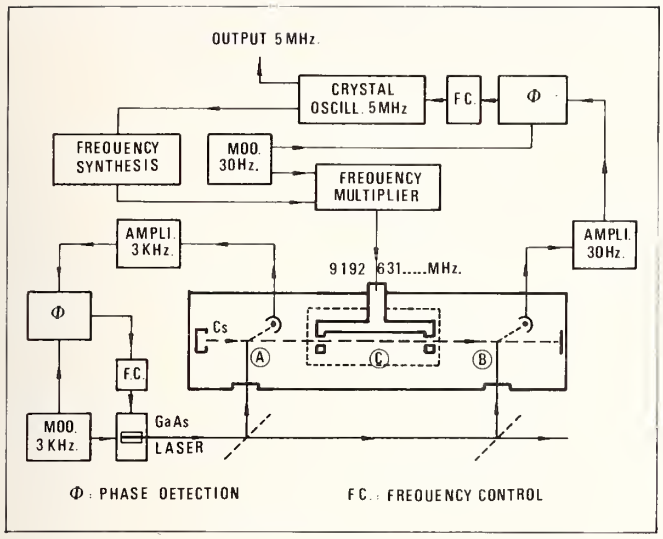


FIGURE 5. *Overall experimental set-up for optically pumped cesium beam frequency standard.*

noted that the laser diode used could lase on the same frequency for slightly different conditions of current and temperature, corresponding to different modes of oscillation, and since some modes were less noisy than others, the values of stability or precision quoted here may vary slightly from one experiment to the other.

4.2 Experimental Results

Measurements of the locked oscillator frequency every 10 seconds, for a period of one hour, show that the distribution of the data is Gaussian ($\chi^2 = 3.6$) with a standard deviation of about $S = 5 \times 10^{-11}$ and an Allan variance $\sigma_y(10) = 3 \times 10^{-11}$.

As said previously, the short-term stability is better with the laser frequency locked to the $3 \rightarrow 4$ than to the $3 \rightarrow 3$ transition. For example, in one experiment, with the $3 \rightarrow 4$ transition: $S = 7.8 \times 10^{-11}$ and $\sigma_y(10) = 3 \times 10^{-11}$ whereas with the $3 \rightarrow 3$ transition: $S = 1.4 \times 10^{-10}$ and $\sigma_y(10) = 4.6 \times 10^{-11}$.

The precision resetability from an offset resonance frequency was on the average $\pm 3 \times 10^{-11}$.

The effect of rf power variations on the precision of the clock depended very much on cavity tuning. For example, for a slight frequency detuning of the microwave cavity the frequency shift could be as large as 1.3×10^{-9} for a 3 dB variation in rf power from optimum power. However, with fine adjustment of the cavity tuning, this shift could be made smaller than 2×10^{-11} and probably much smaller with improvements in the measuring equipment.

Variation of the laser light intensity was also studied for possible frequency shifts due to the fluorescence of the cesium beam. Within the accuracy of the measuring equipment ($\pm 2 \times 10^{-11}$) no light-shift of the microwave resonance could be detected when the laser light was reduced by as much as 50 percent of its maximum intensity.

The absolute accuracy of the clock was observed by changing the static magnetic field H in the microwave cavity and measuring the corresponding frequency shifts. With a large number of observations, a least-squares fit of the results gave the following relation, very close to the theoretical value:

$$f_{\mathrm{Hz}} = 9\ 192\ 631\ 770.05 + 426.7\overline{H}^{\,2} \qquad (7)$$

($\overline{H}$ in gauss = 10^{-4} tesla),

with a regression coefficient $r^2 = 0.99978$.

5. Conclusions

Although these preliminary measurements are too coarse to establish claims of high accuracy regarding 2nd-order Doppler shift or cavity phase-shift, the results obtained are encouraging and the experiments will be continued with improved electronics, especially concerning the stabilization of the laser frequency (which incidentally, could be used also, in the same apparatus, as a secondary length standard).

The potentiality of this frequency standard for accuracy of the order of 10^{-14} will also depend on the absence of an eventual "light-shift" produced by stray laser light or fluorescence of the beam and precautions should be taken against this effect. Also, the need for liquid helium for the laser cryostat is a drawback for operation of the clock over a long period of time. However, laser diodes have already been developed for operation at room temperature, with lifetimes in excess of several thousands of hours of operation, and this should allow the realization of more practical devices than the laboratory apparatus just described.

The author wishes to thank Professor A. Kastler, who initiated the experiment, for his continuous encouragement. The work was done in collaboration with Dr. J.-L. Picqué of Laboratoire Aimé Cotton, Orsay, and with the helpful advice of S. Roizen. Appreciation is also expressed to Dr. T. Yabuzaki[1] and I. Hirano[2] for temporary help in the early phase of the work, and to G. Faucheron, D. Guitard, and P. Pagès, for contributing technical help.

References

[1] D. J. Glaze, H. Hellwig, D. W. Allan, and S. Jarvis, Jr., Metrologia 13, 17 (1977).
[2] H. Hellwig, S. Jarvis, D. Halford, and H. E. Bell, Metrologia 9, 107 (1975).
[3] A. Kastler, J. Phys. Radium 11, 155 (1950); J. Opt. Soc. Am. 47, 460 (1957).
[4] J. Brossel, B. Cagnac, and A. Kastler, J. Phys. Radium 15, 6 (1954).
[5] P. Cerez, M. Arditi, and A. Kastler, C. R. Acad. Sci. Ser. B: 267, 282 (1968).
M. Arditi and P. Cerez, IEEE Trans. Instrum. Meas. IM-21, 391 (1972).
P. Cerez and F. Hartmann, IEEE Trans. Quantum Electron. QE-13, 344 (1977).
[6] J.-L. Picqué, Metrologia 13, 115 (1977).
[7] M. Arditi and J.-L. Picqué, C. R. Acad. Sci. Ser. B: 290, 461 (1980), also in English: J. Phys. (Paris) Lett. 41, L379 (1980).
[8] S. Jarvis, Jr., Metrologia 10, 87 (1974).
[9] M. Arditi, I. Hirano, and P. Tougne, J. Phys. D. 11, 2465 (1978).
[10] E. Bernabeu, P. Tougne, and M. Arditi, C. R. Acad. Sci. Ser. B: 268, 321 (1969).
M. Arditi and P. Tougne, C. R. Acad. Sci. Ser. B: 280, 405 (1975);
Rev. Phys. Appl. 11, 665 (1976).
[11] J.-L. Picqué, S. Roizen, H. H. Stroke, and O. Testard, Appl. Phys. 6, 373 (1976).

---

[1]Ionosphere Research Laboratory, Kyoto, Japan
[2]National Research Laboratory of Metrology, Ibaraki, Japan

J.-L. Picqué, and S. Roizen, Appl. Phys. Lett. **27**, 340 (1975).
[12] L. S. Cutler and L. Lewis, National Bureau of Standards, Boulder, CO (private communication).
[13] N. F. Ramsey, *Molecular Beams*, (Oxford University Press, London, 1963), p. 129.
[14] P. Cerez, C. R. Acad. Sci. Ser. B: **272**, 897 (1971).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Frequency Measurement of Optical Radiation

K. M. Baird

National Research Council of Canada, Ottawa, Canada K1A 0R6

The feasibility of directly relating the frequency of visible radiation to microwave standards has been demonstrated and a number of frequency comparison systems linking infrared frequencies to the cesium primary standard have already been operated. These have yielded sufficient accuracy that together with wavelength measurement based on the ^{86}Kr line used to define the Meter, the standard of length can now be based without fear of a significant discontinuity, on a conventional value for the speed of light and the Cs standard for time. This paper reviews present and proposed frequency comparison chains and discusses their possibilities. Limitations for the general use of frequency comparison methods in the optical region are described.

Key words: frequency chains; laser frequency phase locking; optical frequency measurement; standard of length.

1. Introduction

Among the remarkable facts of modern technology, I think one of the most striking is that the possibility is at hand of making an exact count of events that occur at a rate of over 500 million in a microsecond. As you will hear in this and a later paper of this conference, the recent work leading to a redefinition of the Meter has involved the measurement of optical frequencies. Although the actual counting of optical frequencies or phase locking them to the cesium frequency standard has not yet been achieved, the feasibility of doing this has been demonstrated and a number of laboratories are now engaged in setting up equipment for its realization. Before discussing the techniques that are used to accomplish the remarkable feat mentioned above, let us look briefly at the basic principles used to measure a high frequency in terms of a lower one.

Figure 1 illustrates how a device that has a non-linear response will convert a sinusoidal signal into an output that is distorted, i.e., it contains harmonics of the original signal. Depending on the response characteristic, the output may contain harmonics of significant amplitude up to very high orders (a hundred or more). Similarly, as shown in Fig. 2, such a device will convert a mixture of

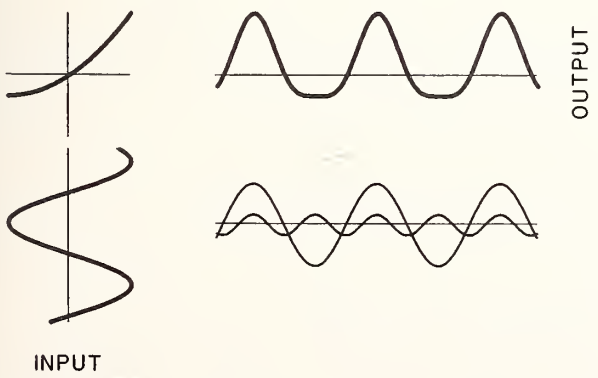


FIGURE 1. *A purely sinusoidal input is converted by a device whose response is non-linear, as shown at the upper left, into a distorted sinusoidal output, i.e., one that contains one or more harmonics as shown in the lower right.*

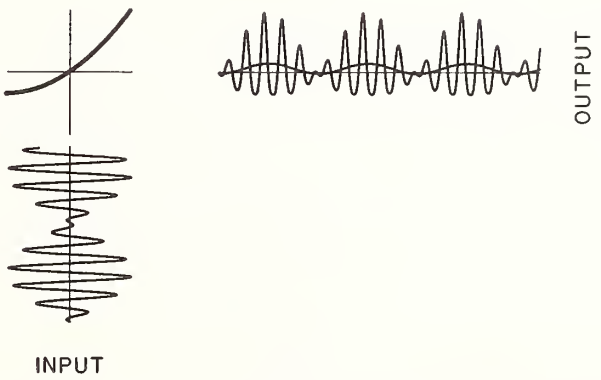


FIGURE 2. *A signal consisting of the superposition of two purely sinusoidal waves of slightly different frequencies is converted by a device whose response is non-linear into an output that contains a component whose frequency is equal to the difference or beat between the original signals.*

two sinusoidal signals of slightly different frequencies, into a distorted output that contains a signal corresponding to the difference frequency or beat between the two original signals. A count of the beats over a period of time yields *exactly* the difference in the number of cycles of the original signals. It can be seen that if two signals, one of which is nearly equal to a harmonic of the other, are impressed upon such a device having a suitable non-linear response, the rather complicated output will contain a low frequency signal of frequency equal to the difference between the one and the harmonic of the other. Thus a measurement of the frequency of the beat, f_b, will yield the value of the higher frequency f_h in terms of the lower f_1, $f_h = Nf_1 + f_b$. The frequency of the beat can be low enough to handle in convenient electric circuits and be measured or counted. The only requirement for high speed is in the device for harmonic generation and mixing. Of course the oscillations must be sufficiently coherent (i.e., they must suffer frequency changes sufficiently slowly) to allow observation of enough beat oscillations to make a significant measurement. When suitably applied, the method makes it possible to measure the number of oscillations of the higher frequency signal during a given number of periods of the lower, *without missing one.*

2. Techniques for Measurement of Optical Frequencies

Two major developments have made it possible to apply the above methods to extend frequency measurement to the optical region of the spectrum: the first was the invention of the laser, which satisfied the coherence requirement, and with the associated development of Doppler-free spectroscopy and techniques for stabilization, provided very precise reference standards; the second was the development of very high speed non-linear devices of which the most important for the present discussion is the point contact metal-oxide-metal (MOM) diode. It can be used for harmonic generation and mixing to produce an electrical output from signals in the range from dc to the infrared. Of nearly equal importance was the production of optical non-linear crystals which can be used to generate second harmonics and for the addition or subtraction of infrared and optical frequencies. They require a detector for conversion into an electrical signal.

The development of the MOM diode was an extension of the well known point contact technique (such as used in old crystal radio receivers). This has been adapted to achieve very high response speed by the use of small, low resistance, low capacitance junctions that result when the very fine tip (Fig. 3) of an etched tungsten wire is brought into contact with an oxidized nickel post. The contact area can be of the order of 10^{-11} cm^2 so that the very low capacitance and the high speed of the electrical response, thought to be due to electron tunneling through the 8–10 Å nickel oxide layer, allows its use as an electrical device at frequencies up to about 200 THz (λ = 1.5 μm).

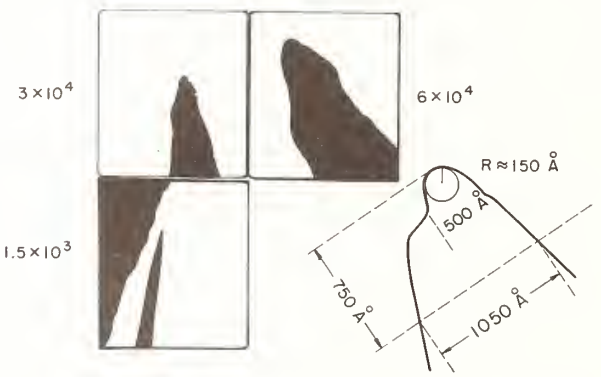


FIGURE 3. *An electron photomicrograph of the etched tip of a fine tungsten wire showing the very small radius that can be used to create junction diodes of extremely small area.*

Overlapping this region, from about 30 THz (10 μm) up to frequencies of radiation in the visible and ultraviolet, non-linear optical crystals are used for harmonic generation and signal mixing, giving an output that with suitable sensitive detectors yields beat frequencies in the form of usable electrical signals. There are, however, limitations to the use of crystals: firstly, the crystals have limited ranges of transparency, and secondly, there is the problem of phase matching. This problem arises because the harmonic signal generated at a given place in the crystal is extremely small. In order to get usable signals, the contributions along a considerable length of path must add up in phase, i.e., the harmonic generated "downstream" must be in phase with the harmonic that has come from "upstream." Because of dispersion the phase velocity of the generating beam and the harmonic will normally differ and this adding up will not occur, but it can often be effected by compensating for dispersion by making use of double refraction, a trick that requires precise adjustment of the indexes of refraction at the two frequencies, either by temperature or by the direction of propagation with respect to the crystal axes.

Although the above developments have opened up the possibility of direct frequency comparison of optical with microwave frequencies, its realization is far from simple and a great deal remains to be done before the frequency of visible radiation can be correlated directly with microwave standards. Problems arise from the fact that the frequencies involved are so very high, so that even very small relative differences result in beat frequencies that are high compared to those that can be handled in convenient state-of-the-art circuitry. This makes trouble in two ways: small frequency instabilities in the lasers cause very large excursions in the beat . . . e.g., a jitter of $10^{-6}f$ at 100 THz (λ = 3 μ m) is 100 MHz, and if the sought for beat signal is of this order, measurement becomes very difficult to say the least. Secondly, matching the harmonic of one source to the frequency of another source to produce low frequency beats is not easy because of the limited choice of reference lasers and of their limited tuning range. This situation is illustrated in Fig. 4; the top half shows on a log scale the five decades in frequency from the cesium standard at 9 GHz to the visible. The ranges of klystrons and the range of operation of the W-Ni MOM diode are indicated, as well as important bench mark frequencies of Cs, HCN, H_2O, CO_2, etc., up to the I_2 lines on which HeNe and argon lasers can be stabilized. Since our frequency comparison method depends to a considerable extent on the simple addition of frequencies, particularly at the higher ranges and since it yields output signals that correspond to differences, the difficulties are more appropriately illustrated by the use of a linear scale, as done in the bottom half of the figure. On this scale the total range of frequency comparison and measurement from sub-kHz through the GHz microwave bands that have been made possible with commercial equipment development and by the pioneering experiments in the infrared done in the sixties at MIT covers only a very small part at the left of the scale. The enormous range yet to be covered to reach the visible as well as the large size of the gaps between available bench mark lasers is evident. This can be emphasized further by considering an application of the rich 10 μ m bands of CO_2 laser lines indicated on the figure.

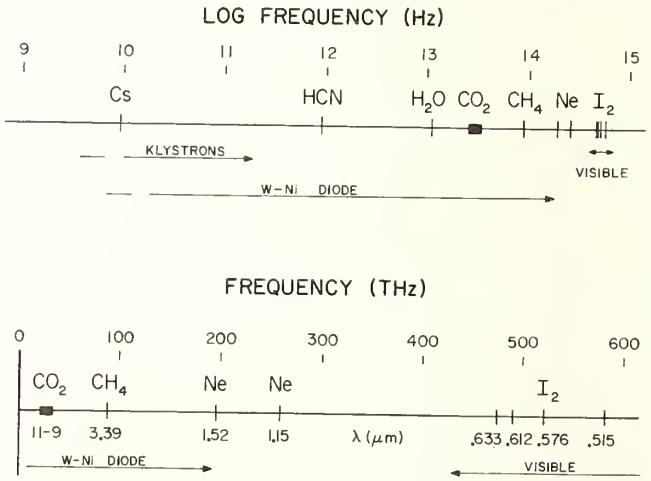


FIGURE 4. *The spectrum covered by optical frequency measurement on a log scale (top) and a linear scale (bottom) of frequency showing the important "bench mark" lines.*

One of these bands is shown in enlarged detail in Fig. 5. These very conveniently produced lines are spaced about 40 GHz apart covering a range of about two terahertz and a number of such bands can be produced using various isotopes. However the gap between the lines is usually more than 100 times the gain width (i.e., tuning range) of the lines; in other words the band is more than 99% empty space. Thus, for example it might be thought straightforward to measure a CO laser line (at about 5 μm) by comparison with the second harmonic of one of the CO_2 lines; in fact the chance of finding a coincidence closer than several GHz is very slight. In such a case the gap usually has to be covered by the addition of a klystron generated frequency as will be seen in several examples below.

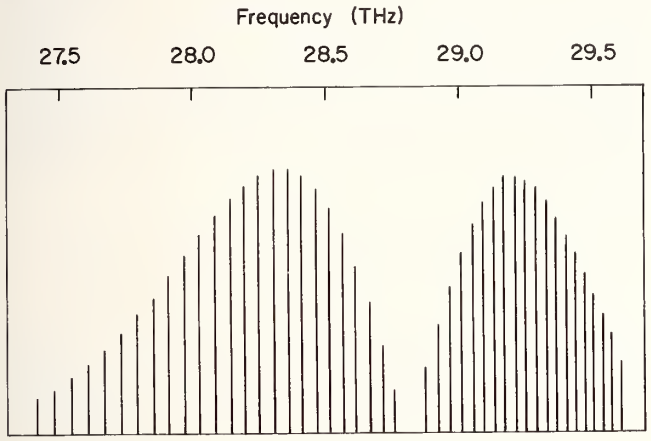


FIGURE 5. *The 10 μm band of the CO_2 laser.*

The CO_2 band shown and similar overlapping bands produced by the use of CO_2 sequence lines and CO_2 of different isotopic composition provide a very important set having frequencies very accurately known relative to one another and covering the range from about 26 to 33 THz (9 to 11 μm). These radiations and their harmonics as well as radiations generated by the sums and differences of pairs, and harmonics of the differences, can be added to and subtracted from other laser radiations by the use of W-Ni diodes up to about 1.5 μm; above this point the diodes appear not to respond electrically. The CO_2 laser is thus of great importance in frequency comparison chains. In certain cases CO_2 laser frequencies can be added to radiations above this limit, thus making it possible in effect to transfer the band of frequencies to another region of the spectrum, (e.g., at 0.633 μm) by the use of non-linear crystals. However the requirement that the crystal be transparent to all radiations involved, the requirement of phase matching, and the very small nonlinear coefficients very much limit the possibilities, as will be seen.

Referring again to Fig. 4 the scarcity of bench mark laser lines can be somewhat alleviated by the use of tunable dye lasers in the visible, and, recently, of color center lasers in the range from ~2.5 μm to 1 μm as will be discussed later.

From the foregoing discussion it is evident that, despite the simplicity of the basic principle, extension to the optical region of frequency comparison with respect to microwave standards is far from easy. Nevertheless the first measurement of a visible frequency has already been demonstrated [1], and several measurements of infrared frequencies have been made with very high accuracy [2–4]. Taken in conjunction with wavelength measurement with respect to the ^{86}Kr length standard, they make possible a new definition of the Meter based on the standard of time and the adoption of a conventional value for the speed of light; this will be discussed in another paper at this conference.

3. Chains to Compare Infrared and Microwave Frequencies

The frequency chains to the infrared will be described with reference to Fig. 6 where they are shown, greatly simplified, in roughly the chronological order in which they were successfully operated (or are expected to operate). The first measurement of the frequency of a CH_4 stabilized He-Ne laser was made at NBS (Boulder) by Evenson and his colleagues [5]. Following up the pioneering work at MIT on W-Ni diodes in the far infrared, they gradually extended the chain, and succeeded in measuring the CH_4 line at 88 THz (3.39 μm), in 1972. Almost simultaneously the group at NPL in London set up a very similar chain [6] and obtained results in very good agreement with NBS. In these chains as shown, the output of a klystron of frequency 74 GHz, was compared with an HCN laser by generation of the 12th harmonic and mixing in a conventional tungsten silicon microwave diode. The HCN laser output was in turn compared to the H_2O laser line at about 10 THz by generation of its 12th harmonic and mixing in a W-Ni diode. The H_2O line was multiplied by three and compared with CO_2 laser radiation at 30 THz, again in a W-Ni diode, and finally, the 30 THz radiation was multiplied by three and compared with a CH_4 stabilized HeNe laser at 88 THz. As mentioned, this description is very much simplified: for example, at nearly all stages above the first klystron stage, additional klystron frequencies had to be added or subtracted to reduce the output beat signal frequency to the MHz range so as to be convenient for measuring; also the

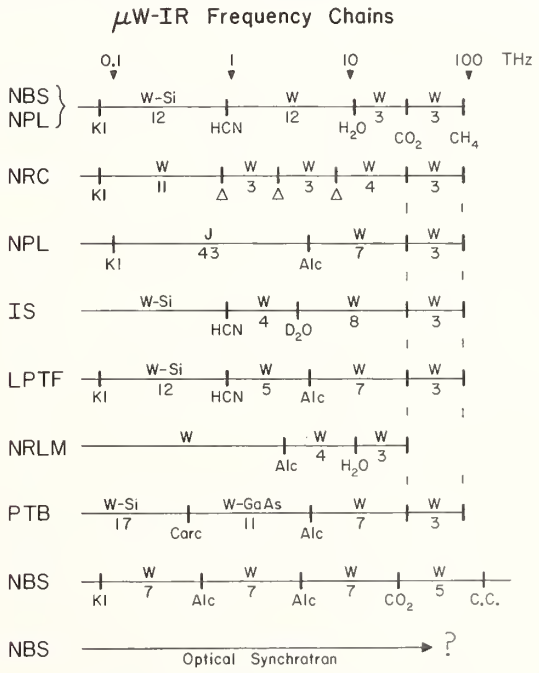


FIGURE 6. *Simplified schematic diagram of frequency comparison chains at the National Laboratories indicated on the left, covering the range from the microwave region to the 3 μm wavelength region. The numbers indicate the harmonics used, Kl = klystron; W = tungsten-nickel point contact diode; W-Si = tungsten-silicon diode, etc.; Δ = difference frequency generated by two CO_2 lasers.*

CO_2 line used for comparison to the H_2O line was not the same line that was used in the comparison with CH_4 and the difference had to be measured against a klystron frequency. At NPL a different pair of CO_2 lines was used in this stage. These early experiments suffered loss of accuracy in the transfer from one stage to the next but in more refined repetitions of the experiments produced frequency values for the CH_4 and CO_2 lines of accuracy approaching better than 10^{-9}. It was these values that, taken together with wavelength measurements at NBS, NPL, and several other national laboratories, led to the acceptance of 299792458 m/s as the best value for the speed of light [7].

More recently, at NRC (Ottawa) [8] and at IS (Moscow) [2] the frequency chains shown were put into operation. At NRC the HCN and H_2O laser stages were replaced by the use of difference frequencies generated in W-Ni diodes by simultaneous input from two CO_2 lasers operating on appropriately chosen transitions. A considerable advantage in the simplicity of the lasers is partially offset by the low signal strengths of the difference frequencies. This required the addition of an extra stage to reduce the harmonic numbers to 3, 3, and 4. The accuracy of the first measurement at NRC was limited by the fact that each stage was measured separately, depending on saturated fluorescence stabilization of the CO_2 lasers for the transfer accuracy. This defect is being corrected by the use of phase-locking in a new system now being set up; phase-locking of the beat frequency of two lasers, operating on lines of different CO_2 isotopes, to a rubidium standard has already been performed successfully. At IS, a D_2O laser and harmonic factors of 8 and 3 were used instead of the H_2O laser and factors of 12 and 3 used at NBS; an OsO_4 stabilized CO_2 laser was used as the 30 THz transfer.

In a recent revised version of the NPL chain [3] the number of stages was reduced as shown by the use of 43rd harmonic generation in a Josephson junction to go from the klystron stage to a CO_2 pumped alcohol laser at 4.25 THz; from there two stages of 7x and 3x were used to go via a CO_2 laser to the CH_4 stabilized HeNe laser at 3.39 μm. In this measurement the stages were operated simultaneously and phase locking or continuous beat frequency counting was used at all points to give a much higher accuracy (3×10^{-11}) than in the previous experiment.

The last chain that, to date, has been reported as having operated is that at LPTF (Paris) [4]. It makes use of HCN and an alcohol laser in the lower stages. The NRLM (Tokyo) chain, which is to be described at this conference, makes use of alcohol and H_2O lasers to reach the CO_2 lines at 30 THz. The PTB (Braunschweig) chain is in the stage that all the parts are operational but final stabilization and locking to make a measurement of the CH_4 line has yet to be done.

The newest NBS chain, shown near the bottom of Fig. 5, is part of a continuous system designed to go to the visible. It will use harmonic numbers of 7, 7, 7, and 5 to go from the klystron region, via alcohol and CO_2 lasers, to a color center laser at 2.52 μm (130 THz); more will be said later about the upper stages to the visible.

The last system shown in the figure, also being set up at NBS (Boulder), is quite different from all the others and is intended to go directly from the rf region to at least the infrared by using the principle of the synchrotron. An electron orbiting in synchronism with an rf field will pass through a focussed laser beam in such a manner that, if the orbital period, i.e., the rf, is an exact submultiple of the laser frequency, the electron will get a "kick" from the electric field of the laser on each revolution and a resonance will be observed. The method has an advantage, in addition to that of having only one stage, in that it is a frequency division, rather than frequency multiplication method, and as such will not suffer from the problem of phase jitter amplification.

Before proceeding to a description of techniques to extend frequency measurements into the visible, it is appropriate to consider the accuracy demonstrated in measurements of the CH_4 line frequency reported up to the present. Figure 7 shows the latest values reported by each laboratory with standard deviation error bars. It is seen that these agree to within one in 10^9, i.e., well within the precision of the present meter definition. There is only one case (IS) where the error bars do not overlap.[1] The uncertainties are still far from the limit imposed in principle by proper frequency comparison with the Cs standard (10^{-13}) but the NPL value (3×10^{-11}) is within the uncertainty known to be attributable to lack of reproducibility of the CH_4 lasers as used. Evidence for this uncertainty comes from two international direct comparisons of CH_4 lasers, one amongst NPL, PTB, and BIPM [9] and the other between BIPM and VNIIFTRI (Moscow) [10].

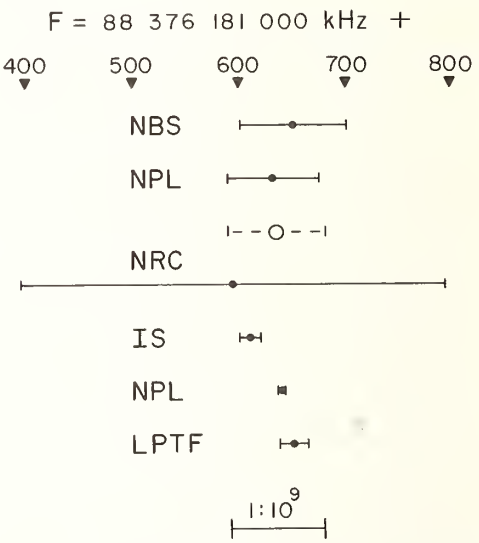


FIGURE 7. *Results obtained at various laboratories for the frequency of the He-Ne laser line at 3.39 μm stabilized to the P(7) absorption line in methane.*

4. Extension of Frequency Measurement to the Visible

Systems designed to extend direct frequency comparison to the visible part of the spectrum are shown in very simplified schematic form in Fig. 8. Only the first of these, the NBS-NRC chain [11, 1], has been demonstrated to completion, yielding a frequency for an I_2 absorption line in the visible. The ISP (Novosibirsk) system has been demonstrated to function but the component frequencies have not been measured [12]. The other three systems are under active development and many parts of them are already operational.

In the NBS-NRC experiments the frequency of a Xe laser at 150 THz (2.02 μm) was first measured by comparison in a W-Ni diode with the sum frequency of a He-Ne laser at 3.39 μm (88 THz) and two 10 μm CO_2 lines

[1] Recent measurements have revised the IS value upwards by 17.4 kHz [20].

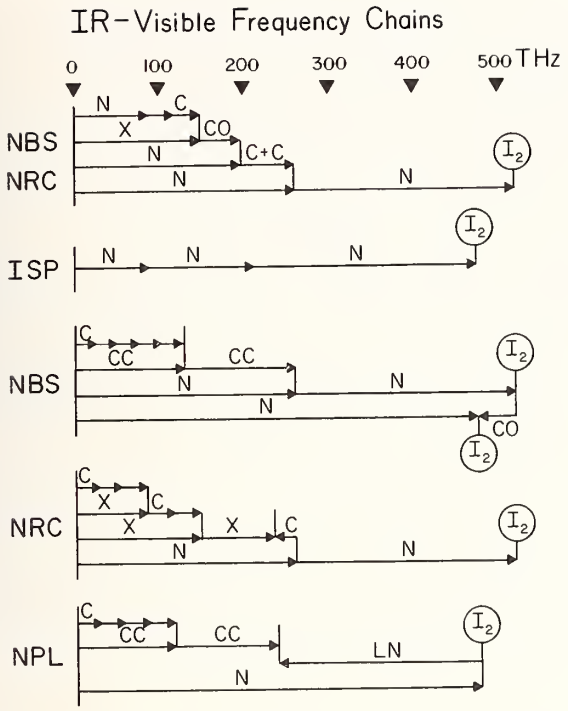


FIGURE 8. *Simplified schematics of frequency comparison chains at the National Laboratories indicated on the left covering the range from 10 μm wavelength (30 THz) to the visible (600 THz). C = CO_2 laser; CC = color center laser; X = Xe laser; N = He-Ne or pure Ne laser; LN = frequency generated in a $LiNbO_3$ crystal.*

(~30 THz). The sum of the Xe line and a CO line at 50 THz was used to measure the 1.5 μm (196 THz) He-Ne laser line, again in a W-Ni diode. In the final part of the experiment done at NBS, the He-Ne 1.5 μm line was added in a proustite crystal to the sum of two CO_2 lines (produced by addition in a CdGeAs crystal) resulting in a sum frequency very near 1.15 μm. A pure Ne 1.15 μm (260 THz) laser output was compared by the use of a Schottky diode. The Ne laser, stabilized on its Lamb dip, was taken to NRC and compared to a 1.15 μm He-Ne laser whose radiation, after doubling, was locked to an I_2 line by saturated absorption. This experiment, done in 1979, thus demonstrated for the first time the direct measurement of a frequency in the visible. It ought to be mentioned however that the CO frequency was not directly measured but was inferred from molecular constants based on frequency measurements. In fact, to date, the best knowledge of the 520 THz frequency is derived from its measured wavelength. The weak CO link could be corrected, but present efforts are directed to what appear to be more attractive systems as described below.

In the ISP system three Ne transitions at 3.39, 2.39, and 1.15 μm were excited simultaneously in a He-Ne plasma; non-linear interaction in the plasma itself produced a sum frequency at 0.633 μm (474 THz). Although very elegant in principle the scheme has not been widely adopted because of apparent difficulties in measurement and control of the component frequencies.

The third chain shown, under development at NBS (Boulder), starts with a CO_2 line (shown at the end of their chain in Fig. 6) whose fifth harmonic generated in a W-Ni diode is used to control the frequency of a color center laser at 130 THz. The frequency of the color center laser is doubled to 260 THz in a $LiNbO_3$ crystal, and used to control a He-Ne 1.15 μm laser. The frequency of the latter is doubled in a second $LiNbO_3$ crystal for reference to the I_2 line at 520 THz as in the NBS-NRC experiment. An additional stage has been proposed that would use the difference frequency of the 520 THz line and a carbon monoxide laser line to measure the widely used He-Ne, I_2 stabilized 0.633 μm line. However a suitable crystal was not identified.

The fourth system shown in the figure, being developed at NRC, makes use of two Xe laser lines excited simultaneously in the same laser and measured by reference to CO_2 laser lines in W-Ni diodes. The Xe lines are added in $LiNbO_3$ and compared with the difference frequency generated by mixing a 1.15 μm He-Ne laser and a CO_2 line in a proustite crystal, thus giving the frequency of the He-Ne line. The latter can be compared with the I_2 line at 520 THz as described above. Some difficulties in this system are associated with the proustite difference frequency generation and low S/N but the problem could be very much alleviated by the use of a $AgGaS_2$ crystal, expected to be available in the near future.

The last system shown, nearing completion at NPL, is similar to the NBS system in the use of a color center laser whose frequency is doubled in a $LiNbO_3$ crystal but will make use of parametric conversion in $LiNbO_3$ to detect the difference between an I_2 stabilized He-Ne laser (0.633 μm) and four times the frequency of the color center laser. Also it uses only the fourth harmonic of a different starting CO_2 line and a lower frequency color center laser, than in the NBS case, to reach the lower frequency (474 THz) visible line.

In concluding this description of the schemes for reaching visible frequencies it is emphasized that, as in the case of the infrared frequency measurement, the experiments are much more complex than one might judge. This can be illustrated by describing in more detail the apparently simple stage of stabilizing doubled Ne 1.15 μm radiation with reference to an I_2 line as used in the NRC-NBS experiment.

This elegant experiment, performed by Hanes at NRC [13] is shown schematically in Fig. 9. It employs a double resonant cavity seen at ABCBE and EBFG. The 1.15 μm radiation produced by the He-Ne plasma tube is focussed into the $LiNbO_3$ crystal which is accurately temperature controlled for phase matching so as to produce doubled frequency at 520 THz (0.576 μm). An additional phase matching requirement is satisfied by a special dispersive reflector at E to ensure that the reflected second harmonic is in phase with the second harmonic generated from the reflected fundamental radiation (1.15 μm). An

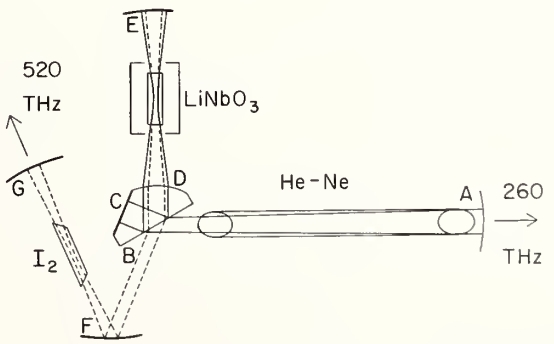


FIGURE 9. *Schematic diagram of apparatus for stabilization of doubled He-Ne 1.15 μm laser on an I_2 line at 576 nm (520 THz).*

I_2 cell in the second cavity, resonant at the doubled frequency, produces saturated absorption features that can be used for stabilization. Scanning and servo control of the cavity arms AC, CE, and EG are suitably coordinated so as to be resonant simultaneously at the required frequencies. About 100 μW of 1.15 μm radiation is emitted at A and about 20 μW of 0.576 μm radiation at G. The fortunate I_2 hyperfine spectrum at 520 THz is shown in Fig. 10. Note the strong component at the left where the gain has been reduced by a factor of 10.

At present it looks as if the optical frequency measurement systems described above stand an excellent chance of working and could, at least in principle, result in phase correlation of visible laser radiation with the standard Cs frequency. This might be done by a completely phase locked system or by a simultaneous count of beats at some of the stages. In any case the frequency of good reference lines in the visible will be measured, but they may amount to only a few bench marks at 0.576 μm, 0.633 μm, probably at 0.612 μm, and possibly an I_2 line near the H_α line at 0.656 μm. Let us next consider the problems and some methods for establishing and of making use of such bench marks for measurement of other lines in the visible and infrared.

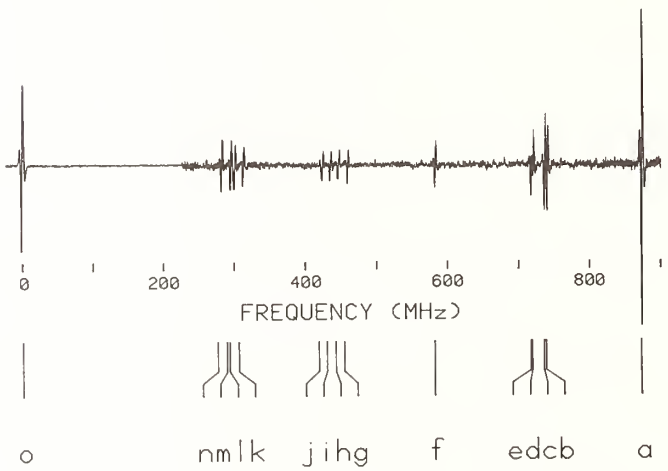


FIGURE 10. *I_2 Hyperfine spectrum at 520 THz obtained with the apparatus of Fig. 9.*

5. The Use of Optical Frequency Standards

Once a bench mark is established in a given part of the spectrum, either in the form of a precisely reproducible absorption line or a laser locked to a frequency chain, other lines or bench marks can be measured by the use of methods like those described; this may require one or a number of steps. The process has become relatively straightforward in the region covered by the point contact MOM diodes, i.e., up to about 1.5 μm (200 THz). Differences of up to a few tens of GHz can be measured directly in the diode output; for greater separations, up to about 100 GHz, klystron radiation can be mixed to measure the differences. For yet greater separations, up to about seven terahertz, two CO_2 lasers having the appropriate difference in frequency can be used; and finally, CO_2 or other appropriate laser emission can be used to measure separations upwards of 25 THz.

One example of such a process is the work of Clairon *et al.* [14], who have established a grid of very precise standards, separated by about 50 GHz, covering a range of 265 GHz in the 10 μm region. This was done by stepping off from a well measured OsO_4 standard [2] by the use of a klystron, to OsO_4 and SF_6 absorption lines that coincide with CO_2 laser lines. Saturated absorption in the very narrow OsO_4 and SF_6 lines made possible such precise setting (±1 kHz and ±3 kHz respectively) that they contributed practically no significant error, even allowing for accumulation in a number of steps. Another example of technique is the method used by Siemsen at NRC [15] to measure the frequency of a laser that is well outside a grid of reference lines. As shown in Fig. 11, when two appropriate known frequencies f_1 and f_2 are mixed with the unknown frequency f_3, the latter can be deduced from the beat frequency $f_b = (f_1 - f_2) - (f_2 - f_3)$.

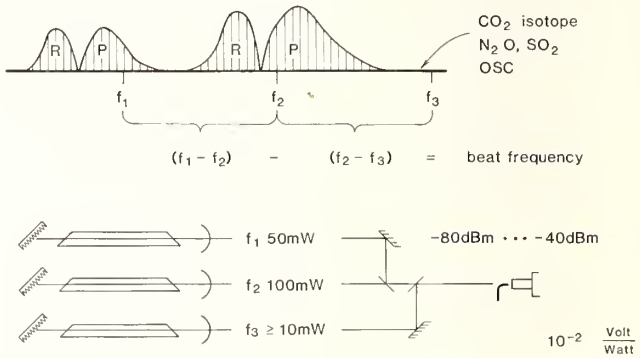


FIGURE 11. *The use of three wave mixing to measure the frequency of a laser line that lies outside a grid of known lines.*

The measurement of large frequency differences gets considerably more difficult above the electrical response limit of the MOM diodes (~1.5 μm). Photo electric detectors can be used as square law devices to measure directly differences up to several GHz in the region from 3 μm through the visible. Schottky diodes have also been used to measure differences up to 122 GHz [16] by mixing with klystron generated frequencies. In order to measure greater differences or to generate harmonics, non-linear crystals must be used.

The use of non-linear crystals for frequency comparison involves far more restrictions than does the application of MOM diodes because of their limited range of transparency, the problems of phase matching, and the small non-linear coefficients; they cannot be used to mix klystron frequencies. We have seen examples of their application in the frequency chains described but these tend to be special cases and general applicability cannot be assumed. For example $AgGaS_2$ is a rather remarkable crystal that is transparent from 13 μm to well into the visible and has a good conversion efficiency: one would expect it to be ideal for the last stage in the NBS chain going from 520 THz to the 0.633 laser line. Unfortunately however, according to published data [17], it is not possible to realize phase matching for this case. The number of "special cases," with the right transparency, phase matching, and suitable laser lines may well become quite large, however, with the further development of crystals and tunable dye and color center lasers. There are cases already where one can, in effect, transfer the 7 THz-wide grid of CO_2 laser frequencies into part of the visible spectrum by mixing in proustite, and perhaps it is not unreasonable to hope that more possibilities like this will turn up.

Special tricks can be used to increase the measurement separation limit beyond that imposed by the photo electric detector. One such is that suggested by Hänsch and Wong [18] whereby a widely tunable laser is frequency modulated so as to match the side band separation to the

intermode spacing; a comb of frequencies covering about 0.5 THz in the visible may be generated and locked to an absorption reference by two-photon absorption. Another example is the stepping procedure proposed by Meisel and his colleagues [19]. Using two stabilized dye lasers, they plan to make two hundred steps of 80 GHz each in order to measure the H_α line with respect to the 0.633 I_2 laser line.

6. Conclusion

It is clear from the foregoing review that, although the direct measurement of frequencies in the optical region is now possible, it still often involves the use of individual ingenious and difficult experiments, particularly in the visible spectrum. For some time to come the best means for interpolation or extrapolation with respect to the few very accurate bench marks will be by the use of wavelength interferometry; certainly it is the most convenient for accuracies of 10^{-8} or less. On the other hand perhaps it is not an unreasonable hope that development of new devices, such as possibly a broad band non-linear reflector, will make possible "day-to-day" use in the optical region of the great accuracy and convenience inherent in the methods of frequency comparison.

The author acknowledges with thanks the help of his colleagues and especially the major contribution of G. R. Hanes to the preparation of this paper.

References

[1] K. M. Baird, K. M. Evenson, G. R. Hanes, D. A. Jennings, and F. R. Petersen, Opt. Lett. 4, 263 (1979).
[2] Y. S. Domnin, N. B. Koshelyaevskii, V. M. Tatarenkov, and S. Shumyatskii, Pis'ma Zh. Eksp. Teor. Fiz. 30, 273 (1979) [JETP Lett. 30, 253 (1979)].
[3] D. J. E. Knight, G. J. Edwards, P. R. Pearce, and N. R. Cross, Nature **285**, 388 (1980).
[4] A. Clairon, B. Dahmani, and J. Rutman, IEEE Trans. Instrum. Meas. **IM-29**, 268 (1980).
[5] K. M. Evenson, J. S. Wells, F. R. Petersen, B. L. Danielson, and G. W. Day, Appl. Phys. Lett. **22**, 192 (1972).
[6] C. C. Bradley, G. Edwards, and D. J. E. Knight, Radio Electron. Eng. **42**, 321 (1972).
[7] Comptes Rendus des Seances de la Conf. Gen. des Poids et Mesures, 15th, 103 (1975).
[8] B. G. Whitford, Opt. Commun. 31, 363 (1979).
[9] B. W. Jolliffe, G. Kramer, and J.-M. Chartier, IEEE Trans. Instrum. Meas. **IM-25**, 447 (1976).
[10] N. B. Koshelyaevskii, A. Obukhov, V. M. Tatarenkov, A. N. Titov, J.-M. Chartier, and R. Felder, Metrologia 17, 3 (1981).
[11] D. A. Jennings, F. R. Petersen, and K. M. Evensen, Appl. Phys. Lett. **26**, 510 (1975); Opt. Lett. **4**, 129 (1979).
[12] V. P. Chebotayev, V. M. Klementyev, and Y. A. Matyugin, Appl. Phys. **11**, 163 (1976).
[13] G. R. Hanes, Appl. Opt. 18, 3970 (1979).
[14] A. Clairon, A. VanLerberghe, C. Salomon, M. Ouhayoun, and C. J. Borde, Opt. Commun. **35**, 368 (1980).
[15] K. J. Siemsen, Opt. Lett. **6**, 114 (1981).
[16] H. -U. Daniel, M. Steiner, and H. Walther, Appl. Phys. **25**, 7 (1981).
[17] G. D. Boyd, H. Kasper, and J. H. McFee, IEEE J. Quantum Electron. **QE-7**, 563 (1971).
[18] T. W. Hänsch and N. C. Wong, Metrologia 16, 101 (1980).
[19] B. Bukhard, H. J. Hoeffgen, G. Meisel, W. Reinert, and B. Vowinkel, these proceedings.
[20] U.S.S.R. State Committee of Standards Bulletin 16, Domnin, *et al.*, (1981) [ISSN 0135–2415].

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Optical Frequency Standards: Progress and Applications

J. L. Hall

Joint Institute for Laboratory Astrophysics, National Bureau of Standards and University of Colorado, Boulder, CO 80309

The decade since the First International Conference on Precision Measurement and Fundamental Constants has witnessed dramatic progress in stable laser technology. For example, frequency stability [1], linewidth [2], and reproducibility [3] of $\approx 3 \times 10^{-14}$ have been reported for the methane-stabilized HeNe laser at 3.39 μm. Impressive performance has also been obtained with CO_2 lasers stabilized with CO_2 [4] and other molecular resonances [5]. In the visible, argon ion lasers with an I_2 molecular reference have given 10^{-12} reproducibility [6]. Recently, a frequency reproducibility of 6×10^{-13} has been reported for the orange HeNe laser (612 nm) stabilized to an external I_2 cell [7]. HeNe lasers operating on the usual red line can give reproducibility in the 10^{-11} domain using intracavity absorption in I_2 [8].

However, to reach spectral transitions of particular physical interest (e.g., H [9]) or of special promise as standards (e.g., Ca at 657 nm [10]) in general will require use of a broadly tuneable laser, typically using color center crystals or a flowing dye solution as the active medium. Comparable stabilization results with such tuneable lasers—especially dye lasers—is vastly more difficult than with gas lasers, although kilohertz linewidth dye lasers have just been reported [11]. A technique suggested by Drever [12] recently allowed achievement of sub-100 Hz dye laser linewidth [13].

High resolution optical interaction techniques based on Ramsey's method of "separated oscillating fields" have been studied theoretically [14] and demonstrated experimentally using two-photon [15] and/or multizone saturated absorption techniques [16] in atomic beams. As in the high resolution methane cell work [17], uncertainty in the second order Doppler shift provides the principal limitation to the accuracy of optical frequency standards [18], even using atomic beam/Ramsey resonance techniques [19]. Measurement of atomic beam velocity distributions may be accomplished with gated excitation of the Ramsey zones, analogous to the techniques employed with cesium microwave standards [20]. Alternatively, the interacting molecular beam velocity may be precisely defined with dual frequency, longitudinal saturation spectroscopy [21]. Useful signal/noise ratios may be feasible using sensitive cryogenic bolometeric detection of the excited beam [22] or using the recently-introduced optical heterodyne technique [23, 13] which appears to approach closely the fundamental quantum detection limit [24].

The ultimate solution to the Doppler shift problem is surely to slow [25] and/or deflect [26] the atomic beam or to cool [27] electromagnetically-trapped ions using radiative processes. The only meaningful limitation to the latter technique may be inferred from the following: we provide an expanded list of interesting candidate transitions for atomic beam frequency standard investigations; however there are only three entries in a comparable list of suitable ions, filtered for compatibility with existing potentially-stable laser sources.

Although the ultrastable laser technology is still immature and of imperfect reliability—especially for cw dye lasers—there is a certain temptation to begin applying these optical frequency standard techniques to physical measurement problems of outstanding fundamental interest. One such measurement, a HeNe laser version of the Michelson-Morley isotropy of space experiment, has already appeared [28]. Other precision experiments underway include: determination of the ground state Lamb-shift using two-photon spectroscopy [9] or construction of a fundamental frequency standard using the same 1S-2S transition [29]; remeasurement of the Rydberg constant using precision atomic beam spectroscopy [30]; precise measurement of metastable and Rydberg energy levels in helium to test the quantum-defect formalism; remeasurement of the relativistic time dilation with vastly higher accuracy; and measurements to set a limit on any possible vector anisotropy in the speed of light. These and related stable laser techniques also appear relevant to interferometric gravity wave detectors [31] and some methods of detecting the influence of parity-nonconservation effects [32].

References

[1] J. L. Hall, in *Atomic Physics 3*, Ed. by S. J. Smith and G. K. Walters (Plenum, New York, 1973), p. 615.

[2] 7×10^{14}. S. N. Bagev, L. S. Vasilenko, V. G. Gol'dort, A. K. Dmitriev and A. S. Dychkov, Sov. J. Quantum Electron. 7, 665 (1977).

[3] S. N. Bagev and V. P. Chebotayev, Appl. Phys. 7, 71 (1975).

[4] C. Freed and R. G. O'Donnell, Metrologia 13, 151 (1977).

[5] C. J. Bordé, M. Ouhayoun, A. vanLerberghe, C. Salamon, S. Avrillier, C. D. Cantrell, and J. Bordé, in *Laser Spectroscopy IV*, Ed. by H. Walther and K. W. Rothe (Springer-Verlag, Heidelberg, 1979), p. 142.

[6] L. A. Hackel, R. P. Hackel, and S. Ezekiel, Metrologia 13, 171 (1977).

[7] P. Cerez, A. Brillet, C. N. Man-Pichot, and R. Felder, IEEE Trans. Instrum. Meas. **IM-29**, 352 (1980).

[8] G. R. Hanes and K. M. Baird, Metrologia 5, 32 (1969). W. G. Schwitzer, E. G. Kessler, R. D. Deslattes, H. P. Layer, and J. R. Whetstone, Appl. Opt. 12, 2927 (1973). J. M. Chartier, J. Helmcke, and J. A. Wallard, IEEE Trans. Instrum. Meas. **IM-25**, 450 (1976).

[9] See for example A. I. Ferguson, J. E. M. Goldsmith, T. W. Hänsch, and E. W. Weber, in *Laser Spectroscopy IV*, *op. cit.*, p. 31.

[10] R. L. Barger, T. C. English, and J. B. West, Opt. Commun. **18**, 58 (1976).

[11] J. Helmcke, S. A. Lee, and J. L. Hall, Appl. Opt. **21**, 1686 (1982).
[12] R. W. P. Drever, private communication.
[13] R. W. P. Drever, J. L. Hall, F. V. Kowalski, J. Hough, G. M. Ford, and A. J. Munley, in preparation.
[14] Y. V. Baklanov, B. Y. Dubetsky, and V. P. Chebotayev, Appl. Phys. 9, 171 (1976).
[15] S. A. Lee, J. Helmcke, and J. L. Hall, in *Laser Spectroscopy IV*, *op. cit.*, p. 130.
[16] J. C. Berquist, S. A. Lee, and J. L. Hall, Phys. Rev. Lett. 38, 159 (1977).
[17] J. L. Hall, C. J. Bordé, and K. Uehara, Phys. Rev. Lett. 37, 1339 (1976).
[18] J. L. Hall, C. Bordé, and C. V. Kunasz, Bull. Am. Phys. Soc. 19, 448 (1974).
[19] R. L. Barger, Opt. Lett. 6, 145 (1980).
[20] D. Halford, H. Hellwig, and D. Glaze, private communication.
[21] J. L. Hall, Opt. Commun. 18, 62 (1976).
[22] T. E. Gough, R. E. Miller, and G. Scoles, Appl. Phys. Lett. 30, 338 (1977).
[23] G. C. Bjorklund, Opt. Lett. 5, 15 (1979).
[24] L. Hollberg, T. Baer, H. Robinson, and J. L. Hall, Appl. Phys. Lett. 39, 680 (1981).
[25] T. W. Hänsch and A. Schawlow, Opt. Commun. 13, 68 (1975).
[26] J. E. Bjorkholm, R. R. Freeman, and D. B. Pearson, Phys. Rev. 23, 491 (1981).
[27] D. Wineland and H. Dehmelt, Bull. Am. Phys. Soc. 20, 637 (1975).
[28] A. Brillet and J. L. Hall. Phys. Rev. Lett. 42, 549 (1979).
[29] E. V. Baklanov and V. P. Chebotalyev, Opt. Commun. 12, 312 (1974).
[30] C. Weiman, private communication; W. L. Lichten, private communication. (See also these proceedings.)
[31] R. W. P. Drever and K. S. Thorne, private communication: P. L. Bender and J. E. Faller, private communication.
[32] I. B. Khriplovich, JETP Lett. 20 315 (1974).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Measurement of Frequency Differences of Up to 170 GHz Between Visible Laser Lines Using Metal-Insulator-Metal Point Contact Diodes*

H.-U. Daniel†, M. Steiner††, and H. Walther†, ††

Frequency differences of up to 170 GHz between the lines of a cw dye laser and a krypton laser at 568 nm were measured by mixing laser and microwave radiation in a metal-insulator-metal point contact diode. The beat signals exhibit good signal-to-noise ratio and no frequency "roll-off" is observed when increasing the laser frequency difference from a few hundred MHz to 170 GHz. It follows that the point contact diode could be used at still much higher difference frequencies. Furthermore, these investigations show a diode response which is different at microwave and visible laser frequencies. Video detection experiments performed in the visible show the influence of thermal phenomena in the diode junction having a roll-off frequency of a few megahertz.

Key words: frequency measurements; heterodyne spectroscopy; metal-insulator-metal diodes.

1. Introduction

Metal-insulator-metal point contact diodes (MIM diodes) have already been well known for about 15 years as effective nonlinear mixers and harmonic generators for infrared and microwave radiation [1]. On these grounds they have been extensively used for absolute measurements of infrared laser frequencies [2] and for precision infrared heterodyne spectroscopy (see, for example, Ref. [3]). However, until recently all efforts to use MIM diodes in the visible spectrum for these purposes have failed [4], thus causing a rather extensive discussion on the diode action in general. Free electron tunnelling [5], thermally enhanced field emission [6], and photoexcitation of tunnel electrons [7] are but a few of the possible mechanisms proposed.

This contribution reports on recent efforts made by the authors [8, 9] to extend the range of application of MIM diodes into the visible spectrum. By mixing the radiation of a cw ring dye laser and a krypton laser at 568 nm with an appropriate microwave frequency it could be demonstrated that MIM diodes are valuable tools for difference frequency measurements even in the visible: The 170 GHz we report is the largest frequency difference yet measured between two visible laser lines. Furthermore, our results strongly indicate different mechanisms to be effective in the microwave and visible frequency ranges. These different mechanisms are found to cause serious signal losses when mixing microwaves and visible laser radiation and, therefore, make it difficult to use microwave harmonics simultaneously generated in the diode.

2. Experimental

Video detection experiments could be done simply by focusing the multi-mode radiation of an argon ion laser (514.6 nm) onto the point contact diode and measuring the laser-induced diode currents at the kilohertz chopping frequencies and/or at the intermode beat frequencies. Mixing experiments, however, were partly performed with an optically pumped sodium dimer ring laser emitting two or more lines simultaneously and in the single-mode regime [8]. As this set-up showed certain drawbacks, the main part of the mixing experiments was done with the apparatus schematically shown in Fig. 1

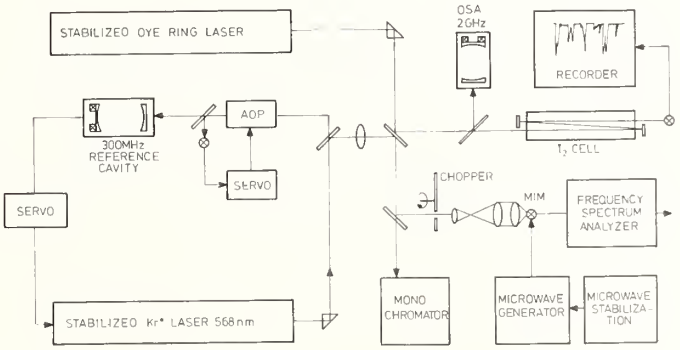


FIGURE 1. *Block diagram of the experimental set-up.*

[9]. The dye laser used was an actively stabilized ring dye laser (Coherent 699) providing a single-mode output of up to 200 mW at 568 nm (Rhodamine 6G). A single-mode krypton ion laser (Spectra Physics 171 with intracavity etalon) which was locked to a temperature controlled confocal Invar Fabry-Perot cavity, served as a second laser source. Additionally its intensity was stabilized using an external ADP crystal device. The detuning of the dye laser emission relative to the fixed krypton laser frequency could be monitored by measuring an I_2 absorption spectrum (see Fig. 1) and comparing it with the iodine line atlas of Gerstenkorn and Luc [10].

Both laser beams were adjusted for optimal collinearity, expanded in a telescope and finally focused onto the diode by means of a microscope objective. The diode itself consisted of a 25 μm thick tungsten wire with an electrolytically etched tip (radius of curvature 50–80 nm) in mechanically adjustable contact with a polished cobalt

*Work supported in part by the Deutsche Forschungsgemeinschaft.

†Max-Planck-Institut f. Quantenoptik, D-8046 Garching, Fed. Rep. Germany.
††Sektion Physik, Universität München, D-8046 Garching, Fed. Rep. Germany.

platelet. This electrode material seemed to offer a somewhat superior stability compared with previously used metals.

Microwave frequencies tunable from 60 to 90 GHz were produced by a wobble generator (Marconi Model 6600/1 with a Siemens backward wave oscillator BWO 6655, peak output power 50 mW) and could be controlled by a three-stage frequency locking stabilization scheme which has been described elsewhere [8]. In the G-band a reflex klystron (Varian VRT 2122 A) provided an approximate power of 10 mW at 170 ± 1 GHz; however, the spectral density was rather low since, owing to a power supply failure, the line width of the klystron emission was 10 MHz. The microwave frequencies were coupled into the MIM diode while it stood in a small groove in the front end of a suitably twisted piece of waveguide. In the E-band a small gold reflector at an appropriate distance from the diode and the waveguide helped to form a favorable cavity-like field distribution.

Beat signals detected by the MIM diode were amplified (Avantek AMT 2006 M, 0.1–2 GHz, 49.5 dB gain, 4 dB noise figure) and then measured in a microwave spectrum analyzer. Owing to mechanical chopping of the laser beams the video output of each spectrum analyzer scan could be detected by lock-in techniques. When the laser difference frequency and a microwave harmonic were mixed, a digital signal averager following the lock-in amplifier had to be applied to recover the beat signals from noise.

3. Frequency Difference Measurements

For reasons yet to be explained we always found the optimal detection of visible laser light in low ohmic point contact diodes. Evidently, this caused signal losses when mixing laser radiation and microwaves because the latter are best detected by highly nonlinear, i.e., high-resistance, diode junctions. Consequently, mixing experiments always had to begin with the observation of a low frequency beat note (about 40 dB above noise) between the two laser lines with a low resistance (10–20 Ω) MIM diode. Increasing the diode resistance subsequently led to improved microwave detection and finally to reproducible observation of laser-microwave beat signals. These beats were obtained within a wide range of diode resistances, but for stability reasons the best results fell into the 30–50 Ω range. Furthermore, most effective mixing was achieved when the detected dc signals of laser and microwave intensity showed positive polarity.

Figure 2 shows mixing signals between the two different laser lines and the 170 GHz klystron emission obtained with a 30 ohm diode. The best signal-to-noise ratio observed was 10 dB. We found similar mixing signals at a few hundred MHz as well as in the whole band from 60 to 90 GHz with amplitudes up to 14 dB above noise. This, however, certainly does not indicate a frequency roll-off, but has to be attributed to the mentioned fact that the klystron emission line width at 170 GHz was nearly a hundred times as broad as the BWO emission line width.

All beat signals between laser lines and fundamental microwave frequencies were observed by real-time spectrum analysis and without the need of signal averaging. Averaging was found necessary, however, in the case of mixing the laser light with a second harmonic of a microwave frequency generated simultaneously on the diode. The beat signals achieved between two green Na_2 dimer laser lines (frequency difference 122 GHz) and the second harmonic of a 61 GHz microwave frequency were heavily buried in noise (about −20 dB). This conversion loss of more than 30 dB is considerably higher than the loss usually observed when generating microwave harmonics in a MIM diode. As mentioned above it has to be ascribed to the low-resistance diode characteristic necessary for the light mixing experiment. Consequently, the latter still turns out to be a major obstacle in mixing the laser light with microwave harmonics; solving this problem would open an even wider range of application to MIM point contact structures.

4. Detection Mechanisms

While elastic electron tunnelling has finally been established as the diode mechanism effective in the infrared and millimeter ranges [11], the response of the diode to visible laser light is still under discussion. Here thermal,

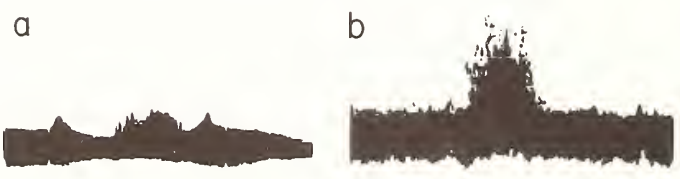


FIGURE 2. *Beat signal obtained by mixing the light of the dye laser and the krypton laser with a microwave frequency of 170 GHz. a: photograph of the beat signal (central peak between two strong noise bursts) taken from spectrum analyzer (vertical scale 10 dB/div, intermediate frequency 409 MHz, IF bandwidth 1 MHz, scan width 5 MHz/div, scan time 5 ms/MHz). b: the same signal after phase-sensitive detection.*

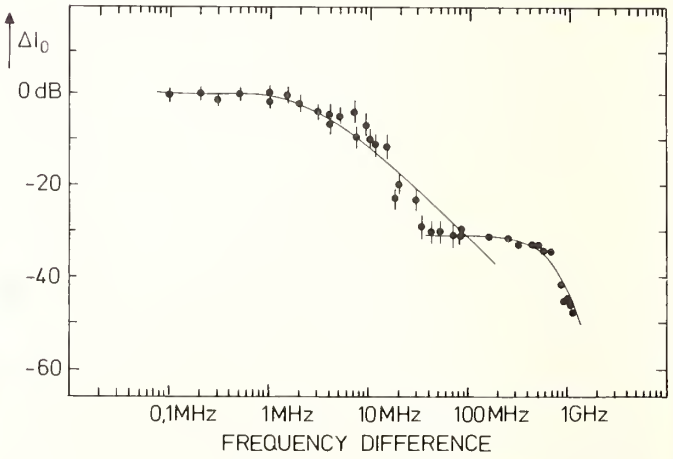


FIGURE 3. *Laser induced diode current increase versus frequency of the laser intensity modulation (diode resistance 50 ohm).*

photoemissive, and geometrical phenomena are involved, thus leading to rather complex diode behaviour.

Two especially interesting findings which have been discussed in detail elsewhere [8] should still be mentioned here: In video detection experiments there is frequency roll-off observed in the detection characteristic of MIM diodes (Fig. 3), and there can be laser driven currents of different polarity depending on their modulation frequency [8]. Both indicate that there is a substantial change in the diode action when going from near dc detection, where most of the investigations were performed (see, for example, Ref. [5]), to true mixing conditions.

The results given in Fig. 3 were obtained by detecting the diode current induced by multimode argon ion laser radiation (514.6 nm) in a 50 ohm junction. Between 50 kHz and 80 MHz the beam was modulated in an acousto-optical device; higher intermediate frequencies were produced by intermode beat signals. While the high frequency roll-off is due to diode mismatch, the low frequency decrease of 30 dB is caused by the time characteristic of the thermal diode response. The solid line drawn in this frequency range (Fig. 3) displays the theoretical tunnelling current driven by thermal modulations of a whisker tip with a 1.7 MHz thermal cut-off frequency [6, 8]. In addition, noise temperature measurements give tip temperatures between 500 and 1000 K. It is therefore concluded that thermal heating of the junction provides the main contribution to the diode response below 100 MHz. At higher frequency differences it probably assists the field emission of electrons at both sides of the contact [8, 12], thus causing the polarity changes mentioned above. So the main difference between the observed diode response to infrared and microwave frequencies and to visible laser light seems to consist of thermal influences in the latter case.

5. Summary

It has been shown that MIM point contact diodes can be used as effective nonlinear mixer elements for visible laser light and microwave frequencies. The signal-to-noise ratio can be further improved by using more powerful and frequency stabilized microwave sources and a better mechanical adjustment for the diode. There was no frequency roll-off observed in the mixing experiments; this suggests that for the measurement of larger frequency differences the microwaves can be replaced by infrared or far-infrared laser frequencies, which are known to couple even better to the diode whisker. Such a link between the visible and infrared regions is of considerable importance for metrology. In the visible spectrum the range of frequency differences now obtainable with MIM diode mixers is far wider than before, which will make it possible to exploit the obvious advantages of frequency measurements as compared to wavelength measurements. The simplicity of the MIM diode technology and already widespread knowledge of it should be of particular advantage compared with, for example, the use of integrated Schottky mixers.

References

[1] L. O. Hocker and A. Javan, Phys. Lett. 26A, 255 (1968).

[2] K. M. Evenson, D. A. Jennings, F. R. Petersen, and J. S. Wells, in *Laser Spectroscopy III*, Ed. by J. Hall and J. L. Carlsten (Springer, Berlin, New York, 1977) p. 56.

[3] J. S. Wells, G. E. Streit, and F. R. Petersen, Opt. Commun. 19, 248 (1976).

[4] D. A. Jennings, F. R. Petersen, and K. M. Evenson, in *Laser Spectroscopy IV*, Ed. by H. Walther and K. W. Rothe (Springer, Berlin, 1979) p. 39.

[5] S. M. Faris, T. K. Gustafson, and J. C. Wiesner, IEEE J. Quantum Electron. **QE-9**, 737 (1973).

[6] A. A. Lucas and P. H. Cutler, Solid State Commun. 13, 361 (1973).

[7] G. M. Elchinger, A. Sanchez, C. F. Davis, Jr., and A. Javan, J. Appl. Phys. 47, 591 (1976).

[8] H.–U. Daniel, M. Steiner, and H. Walther, Appl. Phys. 25, 7 (1981).

[9] H.–U. Daniel, M. Steiner, and H. Walther, Appl. Phys. 26, 19 (1981).

[10] S. Gerstenkorn and P. Luc, *Atlas du spectre d'absorption de la molecule de l'iode, 14800–20000 cm^{-1}*, (Editions du CNRS, Paris 1978).

[11] A. Sanchez, C. F. Davis, Jr., K. C. Liu, and A. Javan, J. Appl. Phys. 49, 5270 (1978).

[12] M. J. G. Lee, R. Reifenberger, E. S. Robbins, and H. G. Lindenmayr, J. Appl. Phys. 51, 4996 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Precision Frequency Metrology for Lasers in the Visible and Application to Atomic Hydrogen

B. Burghardt, H. Hoeffgen, G. Meisel, W. Reinert, and B. Vowinkel

Institut für Angewandte Physik and Radioastronomisches Institut, Universität Bonn, D-5300 Bonn, F.R.G.

A multi-step method is discussed that permits the determination of frequency differences between lasers in the visible in cases where the beat frequency is too large for direct detection. A step width of 80 GHz is used; the beat signal is picked up with millimeter-wave GaAs photodiodes. The resulting beat signals can be measured without further smoothing using a frequency counter. We report on experiments with atomic hydrogen, applying the method to measure transition frequencies aiming to determine the Rydberg frequency and the electron/proton mass ratio with increased precision.

Key words: atomic hydrogen transitions; electron-proton mass ratio; frequency measurement for visible laser radiation; Rydberg frequency.

1. Introduction

Heterodyne techniques have been proven to be extremely powerful in all fields of frequency metrology. For optical laser frequencies this principle is applied by superimposing the parallel beams of the lasers so that the beat frequency can be detected via a photodiode. The result is the difference between the frequencies of the two lasers. This method is widely used to determine the spacings, e.g., between hyperfine components and other details in optical spectra [1]. An upper limit for the size of the measurable spacings is set by the reaction speed of the diode. Commercial fast photodiodes have cut-off frequencies ranging from 5 to about 10 GHz with a fast fall-off if they are used at higher frequencies.

An alternative task is to determine the absolute frequency of laser radiation. The laser can be one that has been tuned to some atomic or molecular transition of interest or it may be a new reference laser with a stable but not yet accurately known frequency. There are three ways to determine the frequency of such a laser: First, the laser can be compared with the frequency of one or several other lasers through a frequency chain that reaches into the visible [2]. Nonlinear crystals and MIM diodes can be used for this purpose though the method is difficult to apply routinely since it requires that phase matching conditions for laser radiation over a very wide range be met. The nonlinear efficiencies of crystals that meet such conditions are small, so that in many cases the signals may be too small. Even if this method can be applied to selected cases only, its great value lies in the fact that it can be used to establish one or several reference laser frequencies in the visible with high accuracy.

The second method is to compare the wavelength of the laser with unknown frequency with that of a laser of known frequency [3]. The main problems arise from the fact that in an interferometric wavelength comparison geometric properties of the two laser fields such as their parallelism and wavefront shape influence the result systematically. Some progress has been achieved in this field by carefully controlling the laser beam parameters [4].

2. Optical Laser Frequency Measurement

A third method by which to determine the absolute frequency of visible laser radiation is discussed in this paper. It makes use of a reference laser with well-known frequency and a beat frequency determination. Thus it avoids the dangers of interferometric methods and gains from the fact that beat frequencies are not shifted if the two superimposed beams are out of parallelism. The obvious principle of the method is to determine the frequency difference between the reference laser and the laser whose frequency is to be determined. The difference is added to or subtracted from the reference frequency for the final result. It is the scope of this paper to discuss how this method can be applied in practice.

Clearly, the reference frequency should be close to the unknown frequency so that the beat frequency is low and thus easily measurable. The octave of visible light, however, spans about 375 THz. If a set of, e.g., 10 reference lasers with almost equally spaced frequencies is established within this band, the difference with respect to an arbitrarily chosen unknown frequency can be as large as 20 THz approximately. At present the only well-known reference lasers are stabilized to iodine transitions at 633 nm (HeNe, [5]) and 514 nm (Ar, [6]). There are no photodiodes that are fast enough to follow a 10 THz beat oscillation. A solution to the problem of measuring such a high difference frequency is to cut it into many smaller differences that can be measured separately. This is achieved by using two cw dye lasers as interpolating oscillators. The principle is schematically presented in Fig. 1: Laser 1 has the unknown (or only approximately known) frequency ν_x. Dye laser 2 is tuned to a frequency ν_2 so that $(\nu_2 - \nu_x)$ is measurable. Dye laser 3 in turn is tuned to ν_4, etc., until ν_{ref} is reached. The sum of all partial differences is the final large difference.

It is clear that the practical application of this scheme requires that the step width be as high as possible in order to avoid excessively large step numbers and cumulative errors. We found a step width of 80 GHz (approximately 1 Å at 6000Å) a reasonable compromise between costs and efficiency of the method [7]. A special GaAs

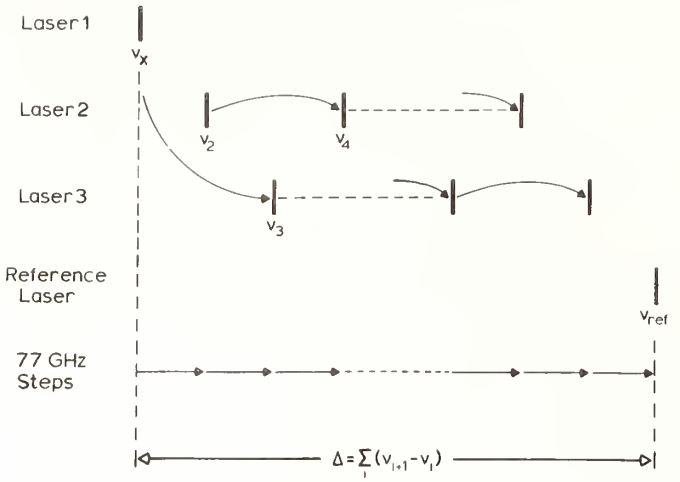


FIGURE 1. *Schematic diagram to illustrate the procedure using two interpolating oscillators (lasers 2 and 3) to close the gap between laser 1 of unknown frequency (ν_x) and a reference laser (ν_{ref}).*

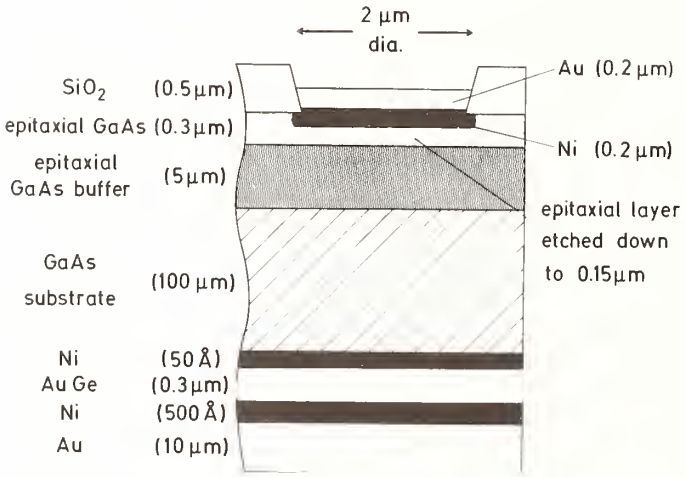


FIGURE 2. *Cross section of the GaAs Schottky photodiode used to detect beat signals of 80 GHz and over.*

Schottky diode was used to pick up the beat oscillation [8]. The active area of the diode has been enlarged to increase the beat signal. The resulting signal/background level was high, namely 40 db for two 2 mW laser beams. Figure 2 is a cutaway view of a single diode. Several thousand are manufactured on a diode chip which is mounted inside an E-band wave guide light detector (Fig. 3). The diodes were used up to a 70 mW cw light level focused to a spot size of 10 μm without damage to the diode.

3. Application to Hydrogen Transitions

An experiment is being prepared to measure the frequency of the hydrogen H_α transition and of other transitions in this way. For H_α, the HeNe reference laser will be used, which requires about 210 steps of 80 GHz each. The expected uncertainty is 0.1 to 1 kHz per step resulting in an overall error of 20 to 200 kHz for the total difference with respect to the HeNe laser. The other main contribution to the error arises from the uncertainty to which laser 1 (which in this case is identical to laser 3) can be tuned to the center of the 30 MHz wide H_α transition. A conservative estimate of this uncertainty is 1% of the line width or 300 kHz [9]; with careful control of the atomic lineshape and the laser light distribution the uncertainty might be as low as 10^{-3} of the linewidth or 30 kHz. Altogether it is planned to determine the H_α frequency to within 50 to 500 kHz or with a relative uncertainty of 10^{-9} to 10^{-10}. The result will be an improved value of the Rydberg frequency, $R_\infty c$ with an accuracy approaching the 10^{-10} level.

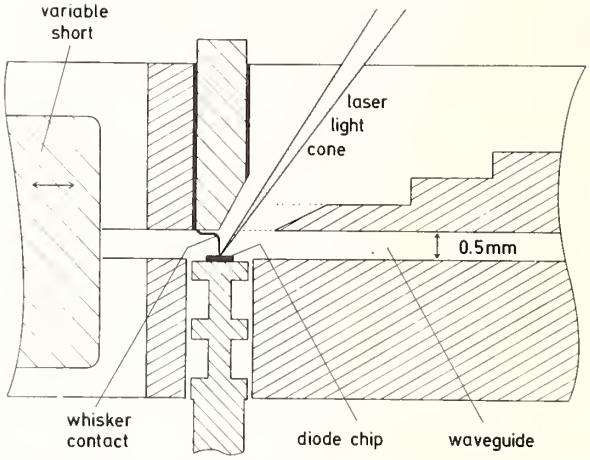


FIGURE 3. *Schematic diagram of the E-band beat frequency detector.*

The experiment is performed with the "free" atoms of an atomic beam in order to reduce any perturbations to the lowest possible level. The light interaction region is designed to avoid Doppler effects, making use of two spatially separated light fields from two exactly counter-running laser beams [10]. This method ensures that the resulting Lamb dip cannot be shifted by non-perfect alignment of the atomic beam with respect to the laser.

Other atomic hydrogen transitions that are within the range of available cw lasers are, e.g., H_β (4860 Å) and H_γ (4341 Å). Since the 4p and 5p states involved have considerably longer natural lifetimes than the 3p state associated with the H_α transition, the accuracy might even be higher. If one accurately measured hydrogen frequency is divided by another, the Rydberg frequency as a leading factor is eliminated, which allows interesting tests for the remaining calculated factors [11].

In another experiment we are trying to observe the two-photon transition from 2s to 7s, 8s or 9s of hydrogen which requires readily available laser radiation between 750 and 800 nm. Besides an absolute frequency determination, the experiment aims to measure the isotope shift in order to improve the accuracy of the proton/electron mass ratio.

References

[1] G. Nowicki, H. Bekk, S. Göring, A. Hauser, H. Rebel, and G. Schatz, Phys. Rev. C 18, 2369 (1978).

[2] K. M. Baird, K. M. Evenson, G. R. Hanes, D. A. Jennings, and F. R. Petersen, Opt. Lett. 4, 263 (1979).

[3] H. P. Layer, R. D. Deslattes, and W. G. Schweitzer, Jr., Appl. Opt. **15**, 734 (1976).

[4] J. P. Monchalin, M. J. Kelly, J. E. Thomas, N. A. Kurnit, A. Szöke, F. Zernike, P. H. Lee, and A. Javan, Appl. Opt. **20**, 736 (1981), and references therein.

[5] F. Bayer-Helms, Ed., PTB-Bericht ME-17 (Physikalisch-Technische Bundesanstalt, 1977).

[6] F. Spieweck, IEEE Trans. Instrum. Meas. **IM-29**, 361 (1980).

[7] B. Burghardt, H. Hoeffgen, G. Meisel, W. Reinert, and B. Vowinkel, to be published

[8] B. Burghardt, H. Hoeffgen, G. Meisel, W. Reinert, and B. Vowinkel, Appl. Phys. Lett. 35, 498 (1979).

[9] B. Burghardt, M. Dubke, W. Jitschin, and G. Meisel, Phys. Lett. **69A**, 93 (1978).

[10] B. Burghardt, H. Hoeffgen, H. Kritz, and G. Meisel, PTB-Bericht E-18 (Physikalisch-Technische Bundesanstalt, 1981).

[11] G. W. Erickson, J. Phys. Chem. Ref. Data 6, 831 (1977).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

System for Light Velocity Measurement at NRLM

K. Tanaka, T. Sakurai, N. Ito, T. Kurosawa, A. Morinaga, and S. Iwasaki

National Research Laboratory of Metrology 1-4, 1-chome, Umezono, Sakura-Mura, Niihari-Gun, Ibaraki 305, Japan

A system for making an absolute measurement of the wavelength and frequency of a stabilized carbon-dioxide laser is under construction. The wavelength has been measured by an up-conversion technique using Proustite with reference to an iodine stabilized laser. The determined value is 9.31724631 μm with a standard error of the mean of 1.4×10^{-8} of the wavelength and the systematic uncertainty is roughly estimated to be 3×10^{-9} of the wavelength. For the frequency measurement, a water vapor laser and an optically pumped alcohol laser have been constructed. Tungsten-nickel and tungsten-cobalt point contact diodes with precision mounts as harmonic generators and mixers have been developed and used for evaluating the stability of the carbon-dioxide laser by beat frequency counting.

Key words: CO_2 laser; light velocity; optical frequency difference; wavelength.

1. Introduction

Since a precision value of the velocity of light was reported by Evenson *et al.* [1] at the fifth meeting of the CCDM in 1973, the value was confirmed by Blaney *et al.* [2] and Baird *et al.* [3]. The present authors, considering the likelihood that a definition of the meter based on the velocity of light will be adopted in the very near future, have been developing a system, which is shown in Fig. 1, for measuring the absolute wavelength and frequency of a carbon-dioxide laser to reconfirm the light velocity and to provide for the establishment of a future wavelength and optical frequency standard.

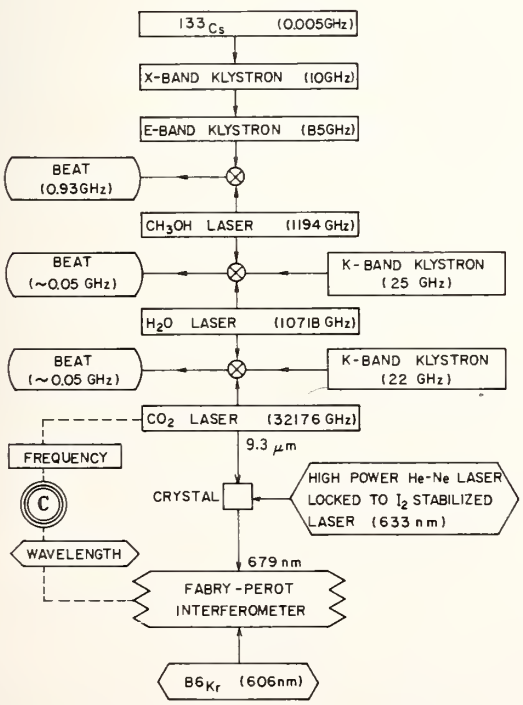


FIGURE 1. *System for light velocity measurement.*

2. Wavelength Measurement

In order to measure the absolute wavelength of a 9.3 μm stabilized CO_2 laser, 9.3 μm radiation is up-converted to deep red 0.679 μm radiation generated by difference frequency mixing with the radiation from a 0.633 μm stabilized He-Ne laser in a nonlinear crystal, Proustite (Ag_3AsS_3) [4, 5].

As the speed of light is independent of wavelength in vacuum, the relation $1/\lambda_{9.3} = 1/\lambda_{0.633} - 1/\lambda_{0.679}$ holds. The wavelength of 9.3 μm can be calculated from the two visible wavelengths, if they are known. The wavelength of 0.679 μm is measured with a pressure scanned Fabry-Perot interferometer with reference to a 0.633 μm offset lock He-Ne laser whose wavelength is determined from the $^{127}I_2$ stabilized He-Ne laser.

The experimental system is schematically shown in Fig. 2. The 1.6 m long CO_2 laser having an intracavity CO_2 cell is stabilized to the R (12) line of the 9.4 μm band by the Javan-Freed method, locking to the zero crossing point of the first derivative signal of 4.3 μm saturated fluoresence. TEM_{00} output power of this laser is typically 0.3 W (after a chopper), and its frequency stability is 3×10^{-10}.

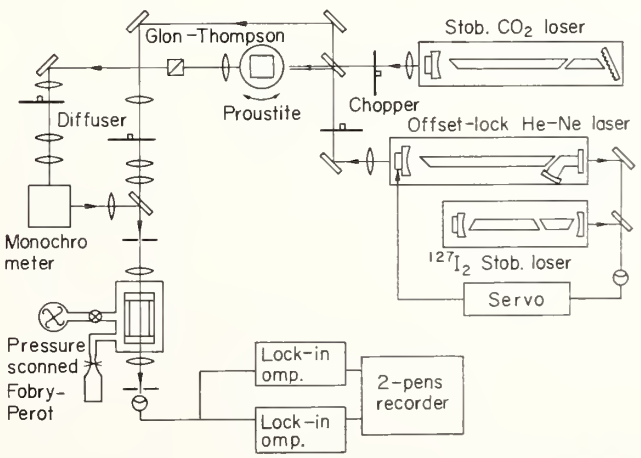


FIGURE 2. *Experimental system for wavelength measurement of carbon-dioxide laser.*

The 2 m long 0.633 μm He-Ne laser having a Fox-Smith type mode selector is operated at a single frequency [6]. Its frequency is +4.8 MHz offset locked, with a relative stability of 1×10^{-11}, to the *i*-component of the R(127) line of a $^{127}I_2$ stabilized He-Ne laser, which has a reproducibility of 6×10^{-11} [7]. The absolute wavelength of this laser locked to the *i*-component of $^{127}I_2$ has been evaluated to be 0.632 991 400 0 μm with a standard error of the mean of 3×10^{-9} of the wavelength [8]. As we have confirmed that the measured value agrees with the recommended value of the CCDM in 1973 within the uncertainty of the measurement, the absolute vacuum wavelength of the offset locked laser is determined as 0.632 991 392 6 μm based on the recommended value of the iodine stabilized laser. The usual output power of the TEM_{00} mode of this laser is 5 mW whereas that of the iodine stabilized laser is 30 μW.

The beams of both the CO_2 laser and the He-Ne laser are weakly focused into the center of a 6 mm long Proustite crystal so that the same confocal parameter of 0.2 m can be obtained. The Proustite is cut and polished so that Type II phase matching is realized by angle tuning, and it is positioned in a liquid nitrogen cryostat cooled to 77 K.

The 0.679 μm radiation of about 10^{-9} W is introduced to the Fabry-Perot interferometer along with that from the offset locked He-Ne laser. A Glan-Thompson prism and a grating monochrometer are used for suppressing the intensity of background 0.633 μm radiation. A plastic diffuser and a set of lenses are also used for achieving uniform illumination onto the Fabry-Perot interferometer.

Scanning of the optical path length of the interferometer is made in a nearly linear fashion by introducing dry nitrogen gas as a supersonic flow into the Fabry-Perot chamber through a needle valve. Simultaneous scanning for both radiations of 0.679 μm and 0.633 μm eliminates the systematic error caused by thermal drift and misalignment of the interferometer, or by the difficulty in realizing the same reference pressure, 4 Pa, when the scanning is started. The combined light of the two lasers is detected by a photomultiplier, and its output signal is supplied to two lock-in amplifiers, which are synchronized with two light choppers at two different chopping frequencies of 400 Hz and 1 kHz, respectively. The outputs from the two lock-in amplifiers are fed into a 2-pen recorder, and a pair of interferograms are obtained on the same chart.

Dispersion of phase change on reflection is eliminated by using a pair of data of fractional orders which are obtained by using two spacers of the Fabry-Perot interferometers with the lengths of 50 mm and 170 mm, respectively.

Figure 3 shows typical data of Fabry-Perot fringes of 0.679 μm and 0.633 μm. The centers of the interference fringes are determined from abscissa readings on both sides of each peak at six intensity levels between 30 and 80% of the peak height. Fractional order number is determined with a standard deviation of 0.001 fringe by extrapolating from four or five peaks by means of a least-squares method.

Thirty-three independent measurements were made for each of the short and long interferometers respectively. The absolute wavelength of the 0.679 μm difference frequency radiation was calculated using the method of exact fractions with respect to the "virtual spacer," whose length is the difference between the lengths of the long and short interferometers. On the basis of the wavelength of the 0.633 μm offset locked laser, the vacuum wavelength of

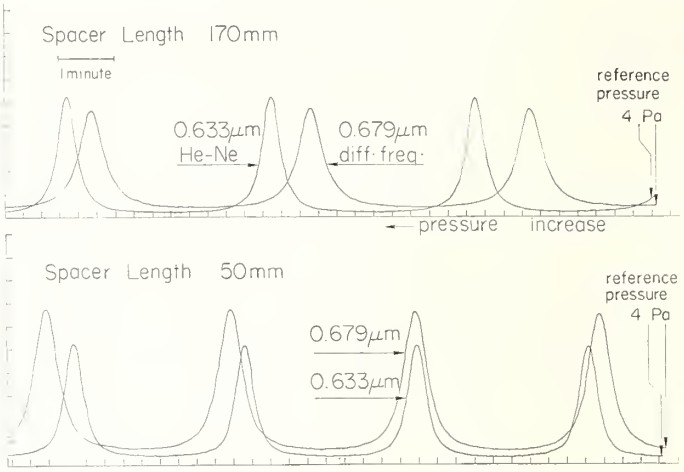


FIGURE 3. *Interferograms of 0.679 μm up-converted light and 0.633 μm helium-neon laser offset-locked to iodine stabilized laser.*

the 0.679 μm radiation was determined to be 0.679 129 847 4 μm with a standard error of the mean of 1×10^{-9} of the wavelength. Thus, the infrared wavelength is calculated to be

$$\lambda_{9.3} = 9.317\ 246\ 31\ \mu m .$$

It should be noted that the uncertainty in $\lambda_{9.3}$ is increased by a factor of 14. Therefore, the statistical uncertainty is estimated to be 1.4×10^{-8} of the wavelength as a standard error of the mean. Although the systematic uncertainties are not fully investigated, the uncertainty due to interferometric measurement is roughly estimated to be 3×10^{-9} of the wavelength.

3. Optically Pumped FIR Laser

An optically pumped far-infrared laser shown in Fig. 4 has been studied. The 2 m long carbon-dioxide laser for the pumping is stabilized by an opto-acoustic method. The laser has a 1.5 m long, 8 mm bore discharge tube. A 25% CO_2, 20% N_2 and 55% He gas mixture flows at a pressure of 2.0 kPa at the entrance port. A maximum output power of 31 W has been obtained under fundamental mode operation. The FIR laser has a 2.1 m long, 38.5 mm bore Pyrex waveguide. The resonator consists of a gold coated flat mirror of fused silica with a 2 mm entrance hole, and a flat aluminium mirror with a 5 mm

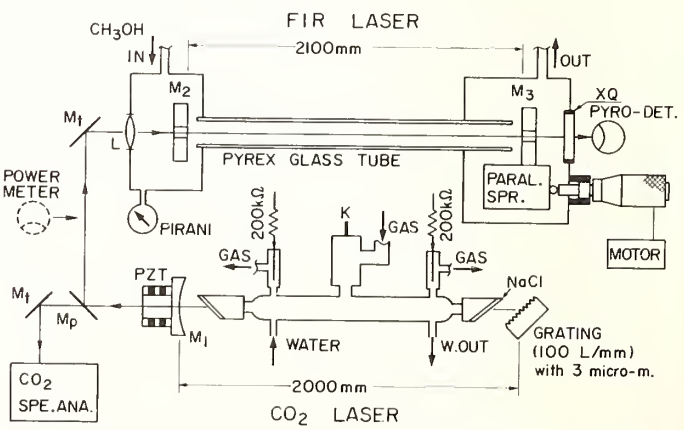


FIGURE 4. *Schematic diagram of optically pumped FIR laser.*

FIR coupling hole. Frequency tuning of the resonator is made by a parallel spring mechanism and a micrometer driven by a small motor. We have made studies concerned with increasing its output power and also surveyed the most suitable oscillation line for frequency synthesis with a water vapor laser and an E-band klystron.

4. Water Vapor Laser

Two internal-mirror-type water vapor lasers have been constructed to examine the optimum discharge conditions, the characteristics of the output beam and the stabilities of output power and frequency. One has a 2.4 m long resonator and a 25 mm bore discharge tube, and the other has a 5.4 m long resonator and a 40 mm bore discharge tube. Both lasers have similar designs. The laser tube has a water cooling jacket, a copper anode and a water cooled copper hollow cathode in the side arms. The laser resonator is composed of two gold coated Pyrex mirrors which are connected rigidly by four invar rods. Frequency tuning of the resonator can be done by the same mechanism as that of the FIR laser.

The water vapor is generated from distilled water in a reservoir and gas additives are supplied to the laser tube from the anode side; they are pumped out from the cathode side by using a continuous gas flow system. The mass flow rate of each gas is regulated by stainless steel needle valves and the gas pressure in the laser tube is adjusted by an exhaust valve.

The short resonator laser was used to study the operating characteristics and discharge conditions of the 28 μm output. It was found that the peak output power was obtained at a discharge current which dissociated the H_2O rapidly [9]. An output power of about 30 mW is obtained under conditions of a H_2O pressure of 80 Pa and a H_2 pressure of 120 Pa. The long resonator laser oscillated nearly at the condition which was estimated using a scaling law and the lasing condition in the short resonator laser. The output power was more than 150 mW multimode, but it was found that the optimum output power was obtained at a different condition due to different systems of gas flow in the two lasers.

For extracting the 28 μm output from the laser resonator, there are three methods where a coupling hole, an intracavity beam divider and a Michelson-type coupler are used. In the coupling hole method, higher order transverse modes oscillated easily and the polarization azimuth of the output beam varied with cavity-scanning [10]. On the other hand, the alignment of the Michelson-type coupler was not easy. Therefore, the output beam was coupled out using a 45° polyethylene intracavity beam divider. The output beam was linearly polarized. The TEM_{00} mode was obtained easily by inserting an aperture in the cavity.

At present, we are aiming to stabilize the frequency of the laser to the Lamb dip with an FWHM of about 5 MHz. The frequency of the laser is stabilized by using a piezoelectric translator. To obtain a good reproducibility, the shape of the Lamb dip is being investigated. By analyzing the beat frequency between two water vapor lasers, origins of frequency fluctuations will be clarified.

5. M-I-M Point Contact Diode

For making precise frequency measurements, we have made tungsten-nickel and tungsten-cobalt point contact diodes [12, 13] of the metal-insulator-metal type [11]. A short tungsten wire 25 μm in diameter was spot welded to the top of a 2 mm thick and 15 mm long brass rod. The wire was bent and the top of the bent wire was etched by conventional electrolytic polishing techniques. The flat, top surfaces of nickel or cobalt posts 30 mm in length and 3 mm in diameter were polished so that they had a flatness of better than one interference fringe with the 589 nm line of a sodium lamp. To obtain resettability and stability of the point contact diode, a precise diode mount which has fine mechanical adjustability and stability of contact pressure, was made. The tungsten-cobalt point contact diode had a S/N ratio 5–10 dB better than that of the tungsten-nickel diode for detecting the rf beat note between harmonics of an X-band klystron and two CO_2 lasers locked to 9R(24) and 9R(26) lines, respectively, as shown in Fig. 5. This diode was used for measuring the stability of our CO_2 laser, which was 3×10^{-10}.

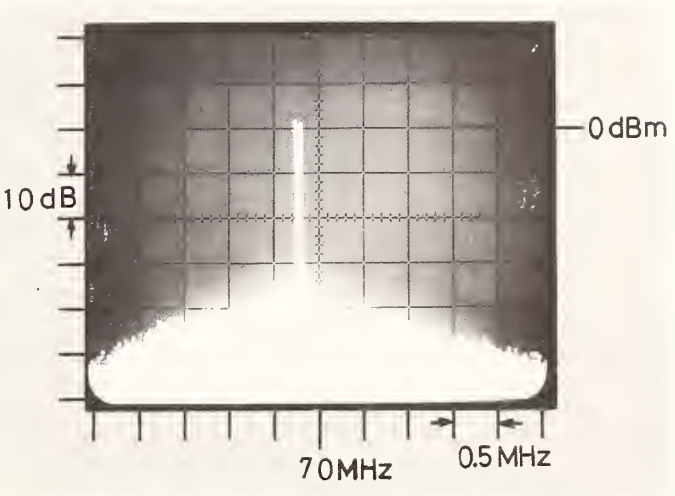


FIGURE 5. *Spectrum of beat signal detected using W-Co point contact diode. Beat is obtained by mixing outputs of X-band klystron and two CO_2 lasers locked to 9R(24) and 9R(26) lines.*

6. Conclusion

The absolute wavelength of a frequency stabilized carbon-dioxide laser has been evaluated to be 9.31724631 μm with a statistical uncertainty of 1.4×10^{-8}. A water vapor laser with a 28 μm output of 150 mW and a tungsten-cobalt point contact diode with a S/N ratio improved by 5–10 dB have been developed. We have been studying ways to increase the output of the FIR laser for linking the CO_2 laser to the cesium frequency standard.

References

[1] K. M. Evenson, J. S. Wells, F. R. Petersen, B. L. Danielson, G. W. Day, R. L. Barger, and J. L. Hall, Phys. Rev. Lett. 29, 1346 (1972).

[2] T. G. Blaney, C. C. Bradley, G. J. Edwards, B. W. Jolliffe, D. J. E. Knight, W. R. C. Rowley, K. C. Shotton, and P. T. Woods, Nature 251, 46 (1974); Proc. Soc. London Ser. A 355, 89 (1977).

[3] K. M. Baird, D. S. Smith, and B. G. Whitford, Opt. Commun. 31, 367 (1979).

[4] K. M. Baird, H. D. Riccius, and K. J. Siemsen, Opt. Commun. 6, 91 (1972).

[5] B. W. Jolliffe, W. R. C. Rowley, K. C. Shotton, A. J. Wallard, and P. T. Woods, Nature 251, 46 (1974).

[6] A. Morinaga and K. Tanaka, Jpn. J. Appl. Phys. 17, 881 (1978).

[7] K. Tanaka, T. Sakurai, and T. Kurosawa, Jpn. J. Appl. Phys. 16, 2071 (1977).

[8] N. Ito and K. Tanaka, Metrologia **14**, 47 (1978).
[9] A. Morinaga and K. Tanaka, IEEE J. Quantum Electron. **QE-16**, 406 (1980).
[10] A. Morinaga, Appl. Opt. **20**, 2395 (1981).
[11] L. D. Hocker, D. R. Sokoloff, V. Daneu, A. Szöke, and A. Javan, Appl. Phys. Lett. **12**, 401 (1968).
[12] T. Kurosawa, T. Sakurai, and K. Tanaka, Appl. Phys. Lett. **36**, 751 (1980).
[13] T. Kurosawa, A. Kuriyagawa, and K. Tanaka, Bull. Natl. Res. Lab. Metrology (Japan) **30**, 8 (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Laser Wavelength Measurements and Standards for the Determination of Length

W. R. C. Rowley

Division of Mechanical and Optical Metrology
National Physical Laboratory, Teddington, Middlesex TW11 OLW, U.K.

The light emitted by portable stabilized lasers used as wavelength standards for length and spectroscopic measurements is reproducible to at least three parts in 10^{11}, and different wavelengths can be intercompared to this level of uncertainty by interferometry. Their absolute wavelength accuracy, limited at present to four parts in 10^9 by the ^{86}Kr standard of the meter, will be improved at least tenfold by a redefinition of the meter, based on the fixed value 299 792 458 m/s for the speed of light. Length measurements, however, are seldom more accurate than one part in 10^7, except in lunar and interplanetary ranging; although changes in length can be measured to better than one part in 10^{14}.

Key words: meter definition; precision length measurement; stabilized lasers.

1. Introduction

Developments in precision length and wavelength measurement over the last decade have been dominated to a large extent by laser techniques. For example, the primary unit of length, the meter, was defined in 1960 by the wavelength of a krypton-86 transition. By the mid-1960's, stabilized visible helium-neon lasers were becoming widely used for the practical measurement of length by interferometry. The lasers available at that time were stabilized to the Lamp dip in the center of the power tuning curve. The stability and reproducibility, at a few parts in 10^8, was significantly worse than that afforded by the ^{86}Kr lamp. This accuracy was, and to a large extent still is, adequate for the majority of length measurement tasks involving material objects, such as length bars and standard gauges. Gradually laser measurement systems have become so widespread that, for example, the majority of precision length measurement tasks at the National Physical Laboratory are now carried out with laser interferometer systems.

Lasers for interferometric measurement are important because of three main characteristics:

(a) Temporal coherence—Their narrow monochromatic bandwidth removes the former path-length restriction of less than 1 m.

(b) Spatial coherence—The laser emits in a narrow beam of almost plane wavefront, so that the light can be easily and efficiently used in optical systems.

(c) Intensity—The amount of light and its efficiency of use makes photoelectric detection easy, leading to the application of electronic methods and automated measurements.

The length measurement technique that is normally used with lasers is fringe-counting interferometry. Commercial systems are available for general measurements of modest accuracy up to lengths of a few tens of meters. For the highest accuracy, however, purpose-built mechanical systems are mandatory, with particular care taken to avoid alignment, temperature, and refractive index errors. The accuracy then attainable when using a well calibrated laser standard is as good as can be achieved by using the ^{86}Kr standard lamp directly. For a 1-m length bar this accuracy is about $\pm 1 \times 10^{-7}$, or up to a factor of two better, at the 99% confidence level [1].

2. Saturated Absorption Stabilization

A great advance in the performance of lasers as reference standards for length measurement was the introduction of the saturated absorption technique of stabilization [2, 3]. It improved the reproducibility of laser standards from being a factor of 10 worse than the ^{86}Kr lamp to a factor of 100 or more better than ^{86}Kr. For several years, two such stabilized laser systems were widely studied and developed. One of them, the 633-nm visible helium-neon laser, stabilized with iodine, is of particular importance because of its direct application to length measurement and its use as a standard for the calibration of other lasers by beat frequency or interferometric wavelength comparison. The other laser system is the methane-stabilized He-Ne laser at 3.39 μm. Although not so suitable for length measurement, this laser afforded the promise of better stability and reproducibility. It has proved to be a particularly important link in the chain of frequency measurements from the cesium frequency standard towards visible frequencies. The 3.39 μm wavelength is also close enough to the visible for infrared/visible wavelength measurements to be made. This has enabled the ^{86}Kr and ^{133}Cs standards of length and time to be interrelated, giving a value of the speed of light to the accuracy limitation imposed by the ^{86}Kr standard [4].

It is usual and convenient for these stabilized lasers to have the saturable absorber within the laser cavity itself. This gives a neat and portable laser system. It is, however, not ideal. The interaction region is too small in diameter so that thermal movement takes the absorbing molecules out of the beam, shortening the effective lifetime; the length of the cell is restricted, so that the pressure of molecules must be raised to give sufficient signal;

and the high beam power density oversaturates the transitions, accentuating any asymmetry. Nevertheless, the performance of these lasers is impressive enough. The stability of a typical internal-cell 633-nm iodine-stabilized He-Ne laser is 3×10^{-12} for an averaging time of 10 s, and the corresponding figure is 1×10^{-13} for the 3.39 μm methane-stabilized laser [5].

A more important measure of performance, however, is the reproducibility of the stabilized frequency (or wavelength). As a result of many comparisons between lasers of different origin, at various times and in various laboratories, it seems that a reproducibility of $\pm 2 \times 10^{-11}$ can normally be obtained with internal-cell 633-nm He-Ne lasers, provided that the operating conditions are the same. The stabilized frequency is affected, for example, by alignment errors, the iodine pressure, the amplitude of frequency modulation, the iodine cell wall temperature, the internal power, etc. The iodine cell must also be free from impurities, such as may result from outgassing during or after sealing-off. From a joint study made recently between NPL and BIPM, it is clear that such contamination is not rare, and accounts for at least some of the discrepancies of inter-laboratory reproducibility that are occasionally observed. It is clear that the quality of iodine cells needs strict monitoring [6, 7].

With internal-cell 3.39 μm methane-stabilized lasers, the methane hyperfine structure is not resolved, and an inter-laboratory reproducibility of only $\pm 3 \times 10^{-11}$ has been reported in one study, although the reproducibility figure of $\pm 5 \times 10^{-12}$ has been obtained for lasers constructed and compared within one laboratory [5]. By putting the saturable absorber outside the laser cavity, however, the conditions may be optimized, and much better reproducibility achieved. The penalty is a non-portable system. Thus although a reproducibility of 1×10^{-14} has been reported [5], it has not been possible to confirm, by direct beat-frequency comparison, that this can be achieved internationally with independently constructed laser systems.

External-cell systems are also advantageous for iodine-stabilized lasers in the visible, and are probably 10 or 100 times better than the internal-cell systems. A number of radiations of the ionized argon and krypton lasers have been thus stabilized for use as wavelength standards, particularly the argon 514-nm line [8]. The orange 612-nm radiation and other transitions of the helium-neon system also show great promise for excellent stability and reproducibility [9]. Indeed, the number of laser wavelength standards that could be realized, particularly by using dye lasers, is almost unlimited. The limit will be set in practice by requirements and resources.

3. The Role of Wavelength Measurements and Standards

Wavelength standards are the basis for interferometric length measurement, but they serve a wider purpose. They are, for example, used in spectroscopy, where it is customary to specify absorption and emission lines by their wavelength or wavenumber. The reason for this is historical and a matter of experimental convenience, as spectrographs allow such measurements to be made by interpolation from a reference set of wavelength standards. It would be more fundamental, however, to specify such radiations by the energy difference between the two states involved in the transition. Alternatively, and more conveniently, the frequency of the radiation, which is directly related to energy by the well known relation $E = h\nu$, is more fundamental than wavelength. I hope and expect that there will be a gradual changeover from the specification of wavelength or wavenumber to the identification of spectral transitions by their frequency. Thus one should perhaps talk about laser frequency standards rather than laser wavelength standards.

In comparing similar stabilized lasers to measure their reproducibility, beat frequency techniques are used. These are rapid, convenient and precise. Unfortunately it is not yet possible to use beat frequencies in the general case to compare two lasers operating on completely different spectral transitions. The frequency differences are too great to be detected so easily. It is possible to work out schemes to overcome this problem using non-linear crystals as mixers and harmonic generators, synthesizing the difference frequency approximately from infrared and microwave oscillators. The final mixer/detector need then only generate a low-frequency beat. At present, however, it is not clear whether such schemes will be viable for more than a few visible radiations of special interest. The alternative well-established technique is interferometric wavelength comparison. Although widely applicable, it is unfortunately subject to a number of practical limitations that introduce errors and thus limit the accuracy. Nevertheless, accuracies of a few parts in 10^{11} have been achieved [10]. Wavelength comparison is thus still a vital means for establishing the wavelength ratios, and thus also the frequency ratios, of visible and near infrared stabilized lasers. As such, it complements the frequency-chain measurements that are now extending to visible frequencies. It will be interesting to see, however, to what extent frequency techniques may replace the present role of wavelength measurement.

4. Interferometer Design

Wavelength measurements are carried out by interferometric comparison, using the optical length of the interferometer as a temporary or virtual intermediate reference. The interferometers that are used may be divided into two main classes:

(a) Low-finesse, usually with two-beam interference, such as the Michelson interferometer and its variants.

(b) High-finesse, with multiple-beam interference, of which the Fabry-Perot etalon is the normal form.

In their basic forms, both these systems form a Haidinger interference ring pattern in the image plane of an extended source, the transmission maxima for a radiation of wavelength λ being given by:

$$m\lambda = 2t\cos\theta,$$

where m is the integer order of interference, t is the real or effective separation of the reflecting surfaces, and θ is the angle of incidence.

As illustrated in Fig. 1, the main difference between the two types of interferometer is the distribution of light in the pattern between the maxima. The two-beam system has a sinusoidal variation superposed on a constant background, whereas the multiple-beam system gives rise to a sharply peaked intensity distribution. It would seem that the multiple-beam system defines the intensity maxima more precisely and should thus be advantageous. In practice, however, this advantage is offset by other factors, such as transmission loss and asymmetry of the peaks due to maladjustment or optical defects. Thus, both types of interferometer are roughly equally utilized in precision measurement.

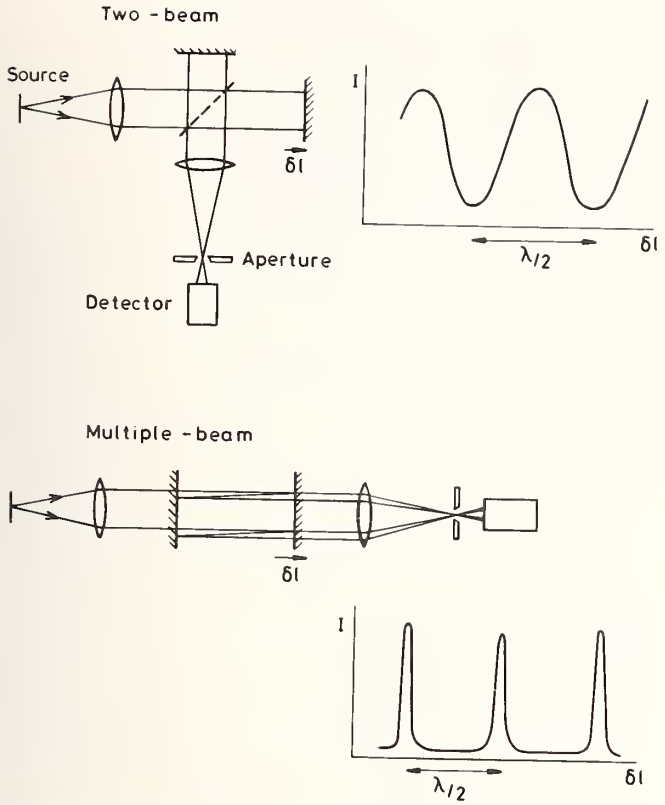


FIGURE 1. *Interferometers used for wavelength measurement. The two-beam (Michelson) form gives a sinusoidal intensity variation, the multiple-beam (Fabry-Perot) gives sharp intensity maxima.*

Interferometer designs may also be divided into various sub-classes. One such sub-class is according to the reflector shape:

(a) Flat reflectors
(b) Curved (spherical) reflectors.

Flat reflectors are the classical form which, as just described, give rise to a Haidinger ring pattern. Measurements can be made by measuring the diameters of these rings, so that the fractional order at the centre may be determined by extrapolation. This used to be the common measurement technique, but it is now rarely employed. More usually, as illustrated in Fig. 1, the central portion of the interference pattern is isolated by transmission through a small hole, and measurements are made by varying the optical length in some linear or predictable manner. The angular size of the pinhole that may be used depends inversely on the finesse and the reflector separation t. Thus, for example, with the 1-m Fabry-Perot instrument at NPL, the optimum size of the pinhole is only 0.1 mm radius for a focal length of 1 m. This makes the instrument sensitive to focus imperfections and aberrations of the imaging system.

Curved mirrors in the form of adjustable cat's-eye reflectors are used in the two-beam interferometer at the BNM in France [11]. This field-widened interferometer enables a much larger detector pinhole to be used, which is particularly advantageous when using relatively weak incoherent light sources such as the ^{86}Kr lamp. With laser sources, the field-widening technique enables greater reflector separations to be contemplated than with flat mirrors, and it is planned to extend the BNM instrument to $t = 2$ m.

Curved mirrors are also used with Fabry-Perot etalons. With laser sources these take the form of mode-matched cavities, similar to laser cavities. Such cavities can be made long without requiring tiny pinholes, and they have a high finesse and transmission, giving excellent sensitivity.

A further subdivision of interferometer designs is according to the method of illumination:

(a) Spatially incoherent, or diffused laser light
(b) Coherent laser beam.

The forms of interferometer shown in Fig. 1 are designed for an incoherent source that has a uniform intensity over the area corresponding to the exit pinhole, and that also radiates uniformly over a sufficient solid angle to fill the entrance lens. Laser beams do not provide such illumination [12]. Partially coherent light, however, may be formed with a diffuser. This gives rise to a speckle field, and quasi-uniform illumination is often achieved by moving or rotating this diffuser. Care must be taken, as illustrated in Fig. 2, to avoid a Doppler shift due to deviation of the light at the diffuser. Reversing the direction of rotation reverses the shift, so that it may be measured or cancelled.

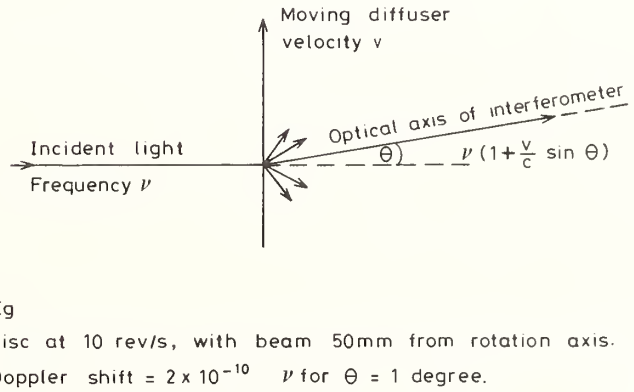


FIGURE 2. *Doppler shift from a moving (or rotating) diffuser.*

Coherent laser beams, however, are essential for the mode-matched form of interferometer. As shown in Fig. 3, the equation specifying the exact length at which the transmission maxima occur is slightly different, and incorporates a phase shift term [13]. This term arises from the diffraction limited propagation of a laser beam [14] and may have a magnitude of a significant fraction of a wavelength. Undiffused laser beams may also be used in two-beam interferometers. As shown in Fig. 4, the two wavefronts that interfere have different radii of curvature and different diameters. The combined effect causes a slight reduction of visibility [15] together with a phase shift, which is additional, and of comparable magnitude, to the propagation phase shift [16, 17, 18, 19].

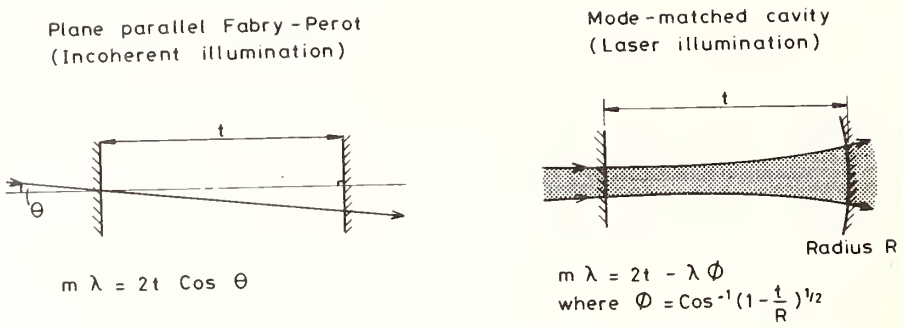
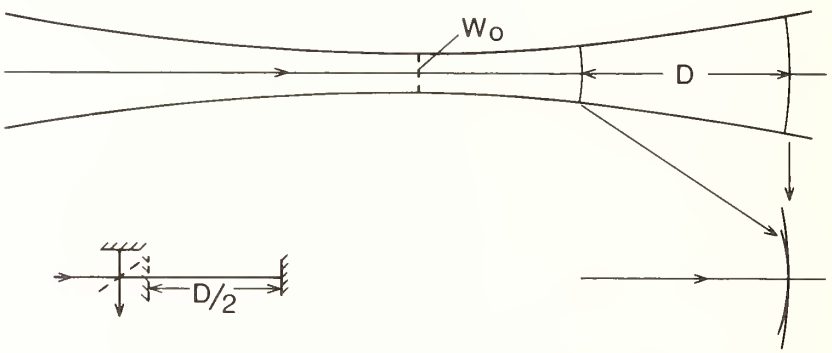


FIGURE 3. *Resonance conditions (giving transmission maxima) for the plane-parallel Fabry-Perot etalon, and mode-matched cavity interferometers.*



In addition to the propagation phase shift, there is a phase shift due to the unequal wavefront curvatures

Combined effect : $D' = D - \frac{\lambda}{2\pi} \arctan \frac{D\lambda}{2\pi W_o^2}$

FIGURE 4. *Michelson interferometer with coherent laser illumination.*

5. The Evolution of Wavelength Measurement Technique

During the last decade, the form of interferometer most used for precision wavelength measurement has been the pressure-scanned plane-mirror Fabry-Perot. With diffused laser light, it is very suitable for measurements directly with the ^{86}Kr lamp. The use of the primary standard radiation, however, restricts the precision because its bandwidth is wide and the etalon length must normally be less than 20 cm. At the NPL [20, 21], the etalon was used at near atmospheric pressure and scanned with a motorized piston. By scanning alternately with pressure increasing and decreasing, some effects of unequal electronic time delays could be eliminated, but scanning at high pressure causes unfortunate temperature changes and non-linearities. At the NBS (Gaithersburg) [22], the NRC (Canada) [23], and NRML (Tokyo) [24] the etalon was placed in a vacuum chamber and scanned by leaking dry nitrogen in through a supersonic nozzle. The interference patterns are usually detected photoelectrically using chopped light, and digitized for computer analysis. At the NRC, however, a moving photographic plate is used, and this avoids the problems of unequal electronic delays. At the IMGC (Italy) [25] the interferometer is kept evacuated and the Haidinger ring image is scanned across an analysing slit in front of the photodector. The Fabry-Perot of the NSL (Australia) [26, 27] is not usually scanned, but is adjusted onto the interference maxima with a magnetic coil system that has a linear translation characteristic and good reproducibility.

Michelson interferometer measurements of stabilized lasers relative to ^{86}Kr have also been carried out in several laboratories. In the interferometers at BIPM [28] and PTB (West Germany) [17] the compensator plate may be tilted slightly so that the phase of the interference pattern can be calculated by fitting a sinusoid to the intensities at four equispaced values of path difference. The instrument at NPRL (South Africa) [29] has one mirror on a parallel-spring translation member, driven by a

piezoelectric tube, with its position monitored by a capacitance transducer. The linearity and calibration of this transducer are checked every time measurements are made. The observations are made by servolocking the mirror to the positions corresponding to interference maxima or minima and determining fringe fractions from the capacitance transducer readings. With the field-widened instrument of the BNM (France) [11], one of the cat's-eye reflectors may be moved on a parallel-spring system by a differential screw together with a piezoelectric element. It may be scanned continuously over a few fringes, or servolocked to successive fringes of the reference laser. The measurement uncertainties in all these measurements of stabilized lasers relative to ^{86}Kr is generally about 1×10^{-9} due to the practical limitations imposed by the ^{86}Kr source, apart from the systematic error in realizing the defined wavelength.

Significantly better accuracies are possible in laser/laser comparisons. The path lengths may be longer, and the source intensities are greater, so that more powerful observational methods may be used. In my opinion the most important development in this field has been the introduction of servolock methods, together with tunable "slave" lasers. An early foretaste of this was the work of Bay, Luther, and White [30] shown in Fig. 5. Sidebands separated by $\pm\omega$ from the laser carrier ν pass through an optical isolator and the mode-matched cavity, and one servo-system locks the length of the cavity to the mean of the two fringe patterns, while the other modifies the frequency ω so that the two sidebands resonate at the same cavity length. As well as pointing the way to servolock methods, the modulation technique used in this experiment is a useful way of establishing the approximate length, and hence the integer order number, of a long interferometer [13, 31].

The use of a slave laser as the wavelength tunable element was introduced by Barger and Hall at the NBS, Boulder [32]. The slave laser is servolocked so as to give a maximum transmission of the interferometer, and hence an integer number of wavelengths, and its detuning from the stabilized laser is measured by the beat frequency. Their particular measurement was relative to ^{86}Kr so that the accuracy of the result was modest. Their technique was, however followed at the NBS, Gaithersburg [13] in a measurement of the 3.39 μm methane-stabilized He-Ne laser relative to the 633-nm iodine-stabilized laser. The reproducibility of these measurements was a few parts in 10^{12}, although the accuracy was limited to two parts in 10^{10} by systematic effects. The high precision of such measurements is due to the advantageous use that is made of the observation time, by comparison with scanning methods, and by the accuracy of beat frequency counting instead of fringe interpolation as the measurement procedure. Another advantage of the method is that because of the high precision, systematic effects, due for example to adjustment errors, may be investigated quickly and without involving an undue amount of data gathering. The method is also used at the PTB [34] and NPL [21, 33], and enables an accuracy of a few parts in 10^{11} to be achieved in visible laser comparisons [10].

Another wavelength measurement technique that has been developed in the last decade is that of fringe-counting, in which a moving reflecting element is traversed over a significant fraction of a meter, while counting the interference fringes due to the unknown and a reference laser. Its advantage is that no prior approximate knowledge is needed of the unknown wavelength, or of a path difference. It is thus particularly suited to the swift determination of the wavelength output of tunable dye lasers to an accuracy of 1×10^{-6} or 1×10^{-7}. As a precision measurement technique it is generally limited by alignment errors due to the use of cube-corner reflectors; but when designed with accuracy in mind, the technique may be as good as other two-beam interferometers. Thus with the MIT instrument [19] an accuracy of a few parts in 10^{11} should be possible.

The main limitations to accuracy in wavelength meas-

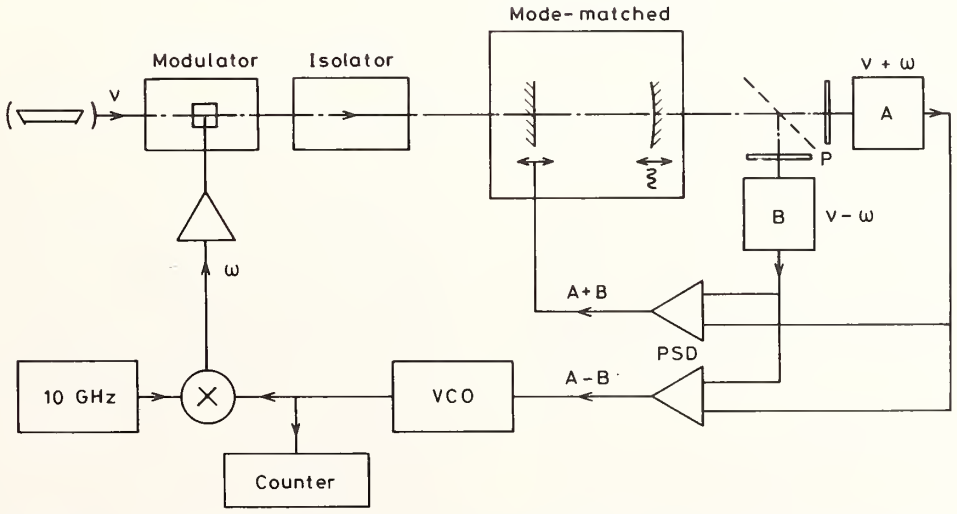


FIGURE 5. *The apparatus of Bay, Luther, and White (1972), using a servo-locked mode-matched cavity interferometer, with feedback tuning of the source so that both sidebands* $(\nu + \omega)$ *and* $(\nu - \omega)$ *resonated simultaneously.*

urement are:

(a) Flatness errors (or non-uniformity of curvature) of surfaces,
(b) Non-uniformity of illumination (or mode impurity),
(c) Defects of imaging (or mode matching),
(d) Non-linearity of interpolation,
(e) Diffraction effects,
(f) Prismatic dispersion of windows,
(g) Parallelism and alignment errors,
(h) Electronic imperfections.

The flatness or errors of figure are the most serious problem. A 1-m interferometer is roughly 3×10^6 half-wavelengths in length, so that to achieve an accuracy of 3×10^{-11} each interference fringe must be subdivided to 10^{-5} of the fringe separation. The flatness of optical surfaces is seldom better than $\lambda/100$. The extra accuracy is achieved by averaging over the irregularities. The problem is to get the same average with both the radiations being compared, and this is a particularly serious problem with interferometers that use spatially coherent light, as the illuminated area then depends upon the wavelength.

6. Wavelength Results

Many precision wavelength measurements have been carried out on the methane- and iodine-stabilized He-Ne laser radiations at 3.39 μm and 633 nm, a few on the CO_2 laser system, and a few also on other iodine-stabilized visible radiations. Measurements of the 633-nm laser, relative to ^{86}Kr have been reported by 10 laboratories, using a variety of interferometer designs. As mentioned above, the measurement uncertainties are generally in the region of 1×10^{-9}, and their agreement well within 4×10^{-9}. Measurements of the 3.39 μm radiation directly with respect to ^{86}Kr, reported by four laboratories, have a slightly worse measurement uncertainty of about 4×10^{-9}.

The ratio of the two wavelengths 3.39 μm/0.633 μm has also been determined interferometrically. Direct measurements suffer the problem of the large diffraction effect for the infrared radiation. In the NBS (Gaithersburg) measurement [13] this limited the accuracy to 2×10^{-10}. A measurement at the PTB [34], however, was limited to an uncertainty of 2×10^{-9}, as also was a measurement at the NPL.

A measurement by the NRC [34] avoided this difficulty by mixing the infrared radiation with that from a 633-nm laser in a non-linear crystal, so as to give sum-frequency radiation in the green. From a wavelength measurement of this green light relative to a 633-nm iodine standard, the required 3.39/0.633 ratio was determined. By this method the severe diffraction problem is avoided; but the penalty is that the uncertainty of measurement is multiplied by the wavelength ratio in calculating the infrared wavelength. Thus the uncertainty of the infrared/visible ratio was 7×10^{-9}.

This non-linear optical mixing method has also been used to measure the wavelength ratio of a 9.7 μm CO_2 laser to the 633-nm laser, in this case by generating the difference frequency in the deep red end of the visible spectrum. A measurement at the NPL by a pressure scanning method [21] had an uncertainty of 5×10^{-9}. Subsequently with a servolock flat-plate Fabry-Perot system an uncertainty of 4×10^{-10} was achieved for the infrared/visible ratio [10]. A direct measurement of the CO_2 wavelength is subject to serious limitation by diffraction. Thus for the MIT measurement [19], the diffraction correction was -2×10^{-8} with a ±20% uncertainty, corresponding to an uncertainty in the wavelength ratio of 4×10^{-9} and making the total wavelength ratio uncertainty 6×10^{-9}.

Only a few visible/visible laser wavelength ratios have been measured precisely. The ratio of the iodine-stabilized Ar^+ laser radiation to the 633-nm laser has been measured with an uncertainty of 4×10^{-10} at the BIPM [35] and by a servolock two-beam method at the PTB [34], with an uncertainty of 5×10^{-10}. The BNM have also made a measurement [36], with an uncertainty 3×10^{-10}. The 612-nm iodine-stabilized He-Ne laser has been measured with respect to the 633-nm laser by the BIPM with an uncertainty of 3×10^{-10} [35].

These results, however, do not represent the limit of what may be achieved by existing techniques. With the MIT system it is suggested that with visible radiations, where diffraction is less important, the measurement uncertainty would be reduced to a few parts in 10^{11}. The BNM instrument is also believed to be capable of similar accuracy [37], and a redesigned PTB instrument is being considered that would also give performance to this level. The CO_2/visible measurement at the NPL [10] involved a visible/visible wavelength ratio determination with an uncertainty of 3×10^{-11}. It is thus realistic to quote 3×10^{-11} as the level of performance that may be achieved for visible laser radiations by current methods of interferometric wavelength measurement.

The present wavelength standards shown in Table 1 have their uncertainties set by the 4×10^{-9} figure associated with the ^{86}Kr standard. If it were not for this limit,

TABLE 1. *Reference standards for length measurements.*[a]

^{86}Kr, transition $2p_{10}$ - $5d_5$,		wavelength	605 780 211 fm
c, the speed of light			299 792 458 m/s
Laser	Absorber	Transition	Component	Wavelength, fm
He-Ne	CH_4	ν_3 ; $P(7)$	; $F_2^{(2)}$	3 392 231 400
He-Ne	$^{127}I_2$	11-5 ; $R(127)$	; i	632 991 399
He-Ne	$^{127}I_2$	9-2 ; $R(47)$	; o	611 970 771
Ar^+	$^{127}I_2$	43-0 ; $P(13)$	; a_3	513 673 467

[a] All have identical uncertainties of 4 parts in 10^9.

the uncertainties of the laser standards would now be a factor of 10 better, with prospects for a further factor of 10 improvement. There is thus pressure to change the definition of the meter so as to avoid this limitation to accuracy. The discussion on this has been active during the past decade, and action is now imminent. The proposed redefinition is such that the speed of light will become a fixed constant with the value given in 1973 (299 792 458 m/s). This value was determined from frequency and wavelength measurements of the 3.39 μm methane-stabilized laser ($c = f \lambda$), with an uncertainty that reflects essentially only the ^{86}Kr uncertainty of realization. A number of different formulations have been suggested for the wording of the new meter definition [38, 39, 40]. After a joint CCDM/CCU discussion in April 1981, however, only two alternatives remain, with the bulk of opin-

ion being in favor of:

(a) "The meter is the length equal to the distance travelled by plane electromagnetic waves in free space in a time interval of 1/299 792 458 of a second."

The alternative is:

(b) "The meter is the length equal to f/299 792 458 wavelengths in free space of plane electromagnetic waves of which the frequency, expressed in hertz, has the numerical value f."

The majority of people prefer wording (a) as it is easier to understand, being simpler and more elegant in concept. A minority of metrologists prefer wording (b) because it is closer to the means of realization of 1-m lengths by means of laser wavelength standards. It is most likely, however, that wording (a) will prevail.

A redefinition of the meter in this form may be recommended in 1982, for formal adoption in 1983 by the General Conference of Weights and Measures. When this takes place, the uncertainties of the stabilized laser wavelengths will have to be revised, and so will their numerical values. Present evidence suggests that the 3.39 μm wavelength will be reduced by 3 fm, and the three visible laser values will each be reduced by about 1 fm, with an extra digit being added after the decimal point, as the uncertainties improve by a factor of 10.

7. Length Measurement

In interferometric wavelength comparison, the mechanical length of the interferometer is a kind of intermediate standard. It is effectively "measured" by the standard radiation, although its length is not actually calculated. An analogous case is the Michelson-Morley type of experiment carried out by Brillet and Hall [41] in which the length of a 30-cm mode-matched cavity was monitored by reference to a 3.39 μm methane-stabilized laser. The cavity length drifted at a rate of about 5×10^{-13}/s and distorted by $\pm 1 \times 10^{-12}$ by gravitational stretching as it was rotated. By averaging over many days, however, it was possible to show that its optical length was independent of its orientation in inertial space to an uncertainty of 2.5×10^{-15}. This is metrology of extreme precision.

Even better precision is sought in interferometer systems designed to detect gravitational waves. The first such system was built and operated at the Hughes Research Laboratories in Malibu, California, in the early 1970's [42]. This instrument had folded arms of effective length 4.25 m and had a strain resolution of 10^{-16} $(\text{Hz})^{-1/2}$, so that it would have been capable of detecting gravitational waves that gave a total strain level of 10^{-14} over the audio band (1 to 20 kHz). Now several groups around the world are developing interferometers of greatly improved sensitivity. At the California Institute of Technology, for example, the aim is to build an interferometer with a baseline of 40 m so as to detect strains of about 3×10^{-19} on millisecond timescales. A long term goal is to construct large scale systems, with baselines of about 1 km, in order to achieve a strain sensitivity of 10^{-21} for frequencies from about 30 Hz to 10 kHz. Such sensitivity would match the estimated strength of gravitational waves that should occur reasonably often. Apart from the mechanical problems, such measurements must be limited by (a) the detection statistics (photon-counting error), which means that high-power lasers should be used, and (b) the disturbance to the momentum, and hence the subsequent position of the reflecting mirrors, by radiation pressure fluctuations. An interesting discussion has been taking place on this latter aspect [43], regarding details of its applicability, and whether or not some arrangement may be devised to reduce its effect.

If interferometers are rigidly connected to the ground, then they measure earth strain. A number of such instruments have been developed around the world for geophysical studies. The motion of the earth's surface behaves like a stochastic process, with the notable exception of earthquakes and earth tides. This strain noise spectrum [44] seems to be similar at different sites, and forms the limitation to observations. Against this background, the normal modes of oscillation of the earth may readily be measured after an earthquake, and earth tides studied in detail and correlated with the local geology and topography [45].

Distances of kilometer dimensions measured in the earth's surface by optical means are limited in accuracy by the refractivity of the earth's atmosphere. This is of magnitude 3×10^{-4}, so that to get an accuracy of 10^{-7} an accurate correction must be applied. The use of two optical wavelengths allows partial cancellation, but the water vapor still presents some problem. The addition of a radio-frequency measurement should be a great improvement. A three-wavelength instrument of this kind being developed in NBS Boulder [46] is aimed at a measurement accuracy of 5×10^{-8} over a 50 km range. This is, incidentally, similar to the relative accuracy achieved for 1-m length bars measured by interferometry.

Over the larger distances in space beyond the earth's atmosphere, refractive index is no longer a problem. After a decade of observations at the McDonald Observatory, lunar ranging has provided a lot of information about the orbital dynamics of the earth-moon system [47, 48]. The fitting of this data to the various models of the orbits, rotations, etc. is a complex task. As a length measurement project, the problems are formidable. Nevertheless, a range precision of 4×10^{-10} is achieved; a remarkable achievement on a moving target at which 10^{18} photons are transmitted, but only one comes back 2.5 seconds later. It should be possible to use lunar ranging results made at two different earth stations to measure the secular variation of the distance between them. Data gathered over 10 years could determine this parameter to an accuracy of 1 cm/year.

Even higher relative accuracy has been achieved in the radio ranging experiments made for the Viking relativity experiment [49, 50]. Two Viking landers implanted on the surface of Mars, and orbiting spacecraft, echoed modulated microwave signals back to earth with a round trip of 2500 seconds. The total time uncertainty of 10 ns, in favorable circumstances, corresponds to a fractional precision of 4×10^{-12} in the distance to Mars.

The proposed redefinition of the meter will allow such precise measurements to be expressed absolutely in meters instead of only in time delay, and it will thus benefit astronomers, as well as all who are concerned with precise laboratory measurements and stabilized lasers.

References

[1] Report of the Advisory Committee for the Definition of the Meter, 1979, Appendix M 3, pp. M46–M55. (Bureau International des Poids et Mesures, F92310 Sèvres, France, 1979).

[2] R. L. Barger and J. L. Hall, Phys. Rev. Lett. 22, 4 (1969).

[3] G. R. Hanes and K. M. Baird, Metrologia 5, 32 (1969).

[4] Report of the Advisory Committee for the Definition of the Meter, 1973. Recommendation M2. (Bureau International des Poids et Mesures, F92310 Sèvres, France, 1979).
[5] Report of the Advisory Committee for the Definition of the Meter, 1979. Appendix M2-C, pp. M31–M45. (Bureau International des Poids et Mesures, F92310 Sèvres, France, 1979).
[6] B. R. Marx and W. R. C. Rowley, NPL Report MOM 51, April 1981.
[7] W. R. C. Rowley and B. R. Marx, Metrologia 17, 65 (1981).
[8] F. Spieweck, IEEE Trans. Instrum. Meas. **IM-29**, 361 (1980).
[9] P. Cerez, A. Brillet, C. N. Man-Pichet, and R. Felder, IEEE Trans. Instrum. Meas. **IM-29**, 352 (1980).
[10] P. T. Woods, K. C. Shotton, and W. R. C. Rowley, Appl. Opt. 17, 1048 (1978).
[11] P. Bouchareine, Bulletin BNM, No. 24, 3, (1976).
[12] D. A. Solomakha, Meas. Tech. 16, 1167, (1973).
[13] H. P. Layer, R. D. Deslattes, and W. G. Schweitzer, Jr., Appl. Opt. 15, 734 (1976).
[14] H. Kogelnik and T. Li, Appl. Opt. 5, 1550 (1966).
[15] W. R. C. Rowley, Opt. Acta 16, 159 (1969).
[16] F. Bayer-Helms, PTB-Bericht Me-16 (Physikalisch-Technische Bundesanstalt, January 1977).
[17] G. Bönsch, PTB-Bericht Me-17 (Physikalisch-Technische Bundesanstalt, May 1977).
[18] K. Dorenwendt and G. Bönsch, Metrologia 12, 57 (1976).
[19] J. -P. Monchalin, M. J. Kelly, J. E. Thomas, N. A. Kurnit, A. Szoke, F. Zernike, P. H. Lee, and A. Javan, Appl. Opt. 20, 736 (1981).
[20] W. R. C. Rowley and A. J. Wallard, J. Phys. E. 6, 647 (1973).
[21] T. G. Blaney, C. C. Bradley, G. J. Edwards, D. J. E. Knight, W. R. C. Rowley, K. C. Shotton, and P. T. Woods, Proc. R. Soc. London Ser. A: 355, 89 (1977).
[22] W. G. Schweitzer, Jr., E. G. Kessler, Jr., R. D. Deslattes, H. P. Layer, and J. R. Whetstone, Appl. Opt. 12, 2927 (1973).
[23] G. R. Hanes, K. M. Baird, and J. DeRemigis, Appl. Opt. 12, 1600 (1973).
[24] N. Ito and K. Tanaka, Metrologia 14, 47 (1978).
[25] F. Bertinetto and A. Sacconi, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, New York, 1976), p. 357.
[26] J. B. Cole and C. F. Bruce, Appl. Opt. 14, 1303 (1975).
[27] C. F. Bruce and R. M. Duffy, Rev. Sci. Instrum. 46, 379 (1975).
[28] A. J. Wallard, J. M. Chartier, and J. Hamon, Metrologia 11, 89 (1975).
[29] F. H. Müller and R. Turner, J. Phys. E: Sci. Instrum. 13, 1024 (1980).
[30] Z. Bay, G. G. Luther, and J. A. White, Phys. Rev. Lett. 29, 189 (1972).
[31] F. Bien, M. Camac, H. J. Caulfield, and S. Ezekiel, Appl. Opt. 20, 400 (1981).
[32] R. L. Barger and J. L. Hall, Appl. Phys. Lett. 22, 196 (1973).
[33] W. R. C. Rowley, K. C. Shotton, and P. T. Woods, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, New York, 1976), p. 410.
[34] G. Bönsch, Document CCDM/79–20, submitted to the Advisory Committee for the Definition of the Meter, 1979. (Bureau International des Poids et Mesures, F92310 Sèvres, France, 1979).
[35] Document CCDM/79–12, submitted to the Advisory Committee for the Definition of the Meter, 1979. (Bureau International des Poids et Mesures, F92310 Sèvres, France, 1979).
[36] P. Bouchareine and B. Rougie, Document CCDM/79–18, submitted to the Advisory Committee for the Definition of the Meter, 1979. (Bureau International des Poids et Mesures, F92310 Sèvres, France, 1979).
[37] P. Bouchareine, Bulletin BNM, No. 43, 9, (1981).
[38] Document GT-M.U./N° 2. submitted to the CCDM/CCU working group on the Definition of the Meter, April 1981. (Bureau International des Poids et Mesures, F92310 Sèvres, France, 1981).
[39] D. T. Goldman, J. Opt. Soc. Am. 70, 1640 (1980).
[40] P. Bouchareine, Bulletin BNM, No. 43, 12, (1981).
[41] A. Brillet and J. L. Hall, Phys. Rev. Lett. 42, 549 (1979).
[42] R. L. Forward, Phys. Rev. D 17, 379 (1978).
[43] B. R. Marx, Nature (London) 287, 276 (1980).
[44] J. Berger and J. Levine, J. Geophys. Res. 79, 1210 (1974).
[45] J. Levine and J. C. Harrison, J. Geophys. Res. 81, 2543 (1976).
[46] S. E. Moody and J. Levine, Tectonophys. 52, 77 (1979).
[47] Yu. L. Kokurin, Sov. J. Quantum Electron. 6, 645 (1976).
[48] J. D. Mulholland, Rev. Geophys. Space Phys. 18, 549 (1980).
[49] I. I. Shapiro, R. D. Reasenberg, P. E. MacNeil, R. B. Goldstein, J. P. Brenkle, D. L. Cain, T. Komarek, A. I. Zygielbaum, W. F. Cuddihy, and W. H. Michael, Jr., J. Geophys. Res. 82, 4329 (1977).
[50] R. D. Reasenberg, I. I. Shapiro, P. E. MacNeil, R. B. Goldstein, J. C. Briedenthal, J. P. Brenkle, D. L. Cain, T. M. Kaufman, T. A. Komarek, and A. I. Zygielbaum, Astrophys. J. 234, L219 (1979).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Double-Mode Method of Sub-Doppler Spectroscopy and Its Application in Laser Frequency Stabilization

N. G. Basov, M. A. Gubin, V. V. Nikitin, A. V. Nikulchin, and D. A. Tyruikov

P. N. Lebedev Physical Institute, U.S.S.R., Moscow, 117924

and

V. N. Petrovskiy and E. D. Protscenko

Institute of Physical Engineering, U.S.S.R., Moscow, 115409

We present some results of the investigation of the proposed high sensitivity method of sub-Doppler spectroscopy and laser frequency stabilization which is based on the parameters of a double-mode (DM) gas laser containing an internal absorption cell. A short DM He-Ne/CH_4 laser was constructed which has relative frequency stability better than 10^{-14} and radiation spectral width $\lesssim 10$ Hz. When a telescopic beam expander was used inside the cavity of the short DM He-Ne/CH_4 laser, super-narrow reference spectral lines of about 3 kHz in width were obtained, and the magnetic hyperfine structure (hfs) of the $F_2^{(2)}$ methane line was resolved.

Key words: double-mode method; He-Ne laser; methane cell; sub-Doppler spectroscopy.

1. Introduction

The recent development of optical frequency standards is based on obtaining super-narrow spectral lines with nonlinear sub-Doppler laser spectroscopy methods. Extremely small values of collision, saturation, and transit flight time broadening are required for obtaining spectral lines of 10^2-10^3 Hz in width. In particular, if one uses the method of inverted Lamb-dip with an internal or external absorption cell, very low gas pressures and weak optical fields are needed. An essential difficulty in experimentally obtaining such narrow resonances is the rapid decrease of the signal-to-noise ratio when low absorption gas pressures of 0.1 to 1 mPa are used.

The present paper reports the results of studying the double-mode method of sub-Doppler spectroscopy, which has much higher sensitivity for obtaining narrow resonances in a laser with an internal absorption cell as compared to single mode laser operation. The investigations have been carried out with a double-mode He-Ne/CH_4 laser (at wavelength $\lambda = 3.39$ μm, $F_2^{(2)}$ line of methane), which attracts attention as a possible high accuracy optical frequency standard [1].

Stable double-mode operation of the laser is realized with the help of a Fabry-Perot cavity, containing two quarter wave plates [2]. The laser provides stable operation near the gain curve center on the two adjacent axial modes possessing orthogonal linear polarizations.

In contrast to a single-mode laser (SML) two types of narrow resonances, "amplitude" and "frequency" (which possess a number of advantages) can be observed in the double-mode laser (DML) radiation (Fig. 1).

2. Amplitude Resonances of the DML

In the observation of the output power of a separate mode there appear "amplitude" resonances (AR) (Fig. 2) which are similar to the inverted Lamb-dip, but may be

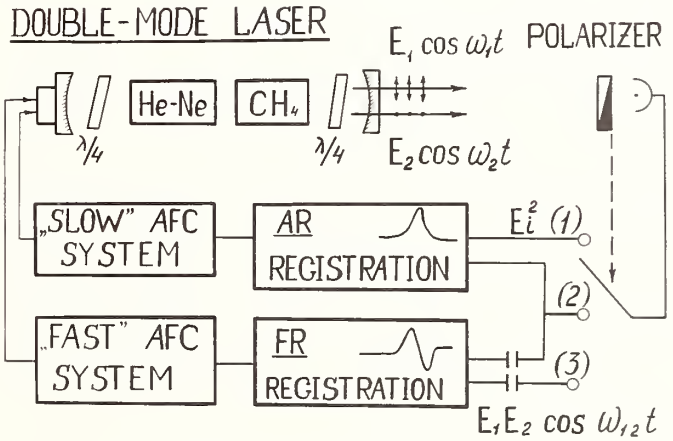


FIGURE 1. *The scheme of the AR and FR observation in the DML. With respect to polarizer position, AR (1), FR (3), or both resonances simultaneously (2) may be observed. In position (2) the resonances are used as discriminators for the two-loop AFC systems.*

about two orders of magnitude higher than the SML resonances. Such a sharp increase in the sensitivity is reached by creating a strong interaction regime between orthogonally polarized modes. The ratio between resonance amplitude (a_-) in SML and DML is as follows:

$$a_-(\mathrm{DML}) = \frac{a_-(\mathrm{SML})}{f_{\mathrm{int}}} \qquad (1)$$

where $f_{\mathrm{int}} = 10^{-1} - 10^{-2}$ is the mode interaction parameter [2].

The physical mechanism of the sharp increase in the AR in the DML consists in the following: when the frequencies of the two modes (ω_1, ω_2) are scanned their intensities can be made approximately equal. If, for example, the first mode frequency ω_1 coincides with the

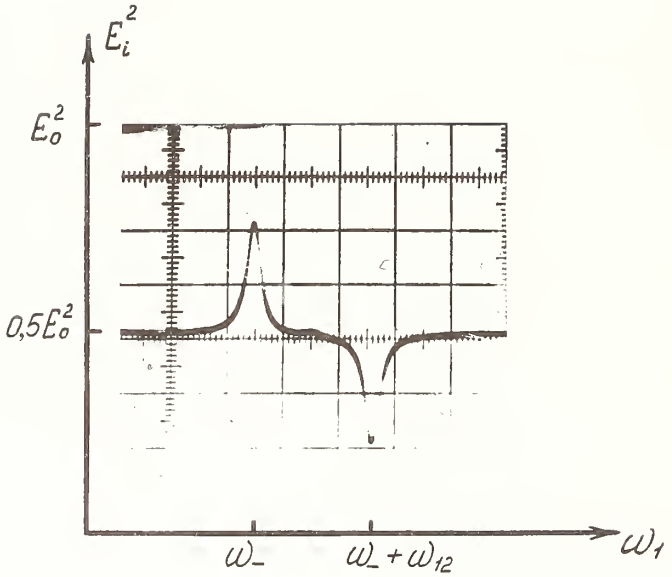


FIGURE 2. *The amplitude of resonances of the DML. E_i^2 is the intensity of one of the two modes; E_0^2 the total DML intensity; $p_- = 400$ mPa, $\gamma_- = 150$ kHz.*

center of an inhomogeneously broadened absorption line ($\omega_1 \approx \omega_-$), the mode losses change slightly due to the formation of the Lamb-dip. For weakly absorbing or poorly saturated transitions in molecular gases at low pressures, the change in losses is usually about 0.1% and may be an order of magnitude smaller. Approximately the same contrast of narrow inverted Lamb-dip resonances is observed in SML operation.

But in DML operation the strong coupling between the modes leads to a sharp intensity re-distribution among the modes. It is possible to realize the situation when approximately all the active atoms radiate in the first mode.

AR contrast may constitute 30–40% of the total laser power even in the case of a short laser. When the length of the He-Ne cavity ℓ_0 is 60–80 cm, the transverse size of the mode distribution, d_-, in the cell about 2 mm, and the methane pressure p_- within the range of (0.2–0.7) Pa, the AR width γ_- is equal to 100–150 kHz. Under these conditions a large value of AR and high S/N ratio allowed us to obtain a laser frequency stability of better than 10^{-14} (Allan variance for $\tau = 10$ sec averaging time).

Increase of the AR by changing the mode interaction (decreasing f_{int}) is limited by two factors: the increase in natural laser intensity noise and instability of the double-mode regime.

3. Frequency Resonances of the DML

The frequency resonances (FR) of the DML are observed as changes in the beat frequency between modes and are caused by resonant changes in the Doppler-broadened dispersion line of the absorption medium under the influence of standing wave field saturation (frequency pulling to the line center [3]). In the DML with an internal absorption cell the frequency spacing between modes $\omega_{12} = \omega_1 - \omega_2$ changes in the following way (with $|\omega_1 - \omega_-| \sim \gamma_-$; $\omega_{12} \gg \gamma_-$):

$$\omega_{12} = \omega_{12}^0 + \Delta\omega_{12} \sim \omega_{12}^0 + \frac{\alpha_- c \ell_-}{\ell_0} \mathcal{I} \frac{\gamma_-(\omega_1 - \omega_-)}{\gamma_-^2 + (\omega_1 - \omega_-)^2} \quad (2)$$

where α_- is the nonsaturated absorption coefficient per unit length; c is the speed of light; and $\mathcal{I}_1$ is the normalized intensity of the mode which coincides with ω_-.

In our experiments with the He-Ne/CH_4 DML containing two $\lambda/4$ plates the beat frequency ω_{12}^0 was determined for the range (2–5) MHz, and its change ($\Delta\omega_{12}$) was $\sim 10^5$ Hz, when $p_- = (0.2 - 0.4)$ Pa and $\ell_- \sim 0.5$ m.

Reference [4] describes the first observation of FR in the DML with an absorption cell. FR in the SML has also been studied [5, 6] with the use of an additional heterodyne laser. In the DML the heterodyne function is performed by the second mode.

3.1 Increasing DML Short-Term Frequency Stability

One of the requirements for obtaining super-narrow spectral lines with $\gamma_- \sim 10^3$ Hz is high short-term frequency stability of the laser. The laser spectral width ($\Delta\omega_S$) must be considerably less than the resonance width γ_-. It is well known that the frequency spectrum narrowing in any type of laser used in precision spectroscopy is accomplished by means of frequency stabilization with a broad-band automatic frequency control (AFC) system. The best results in He-Ne/CH_4 laser spectrum narrowing have been achieved by Chebotayev *et al.* [7]. They obtained a spectral width $\Delta\omega_S \sim 1$ Hz when using intense ($a_- \sim 1$ mW) and narrow ($\gamma_- \sim 50$ kHz) resonances in a 5-meter length laser.

In the present paper we suggest another method of laser spectrum narrowing. The main idea is that the laser frequency be stabilized to the reference line by a two-loop AFC system, using different molecular frequency discriminators (Fig. 1). The first AFC reduces "slow" frequency fluctuations in the range 0–10 Hz and uses as a reference point the top of the AR. This is the commonly used AFC where laser frequency changes are detected as intensity changes. The second AFC reduces "fast" fluctuations with a frequency of more than 10 Hz. The FR of the DML is used as a frequency discriminator for this AFC.

Note the main advantages of such a combined system: the "fast" feedback loop does not need laser cavity modulation and, consequently, the AFC band (Δf_{AFC}) is limited only by γ_-, not by the modulation frequency, and may be, therefore, 5–10 times wider than in ordinary AFC. As the amplitude detection is replaced by frequency detection, so the demands on AR intensity are considerably less. At the same time the value of FR($\Delta\omega_{12}$) is proportional only to a linear absorption and hence, weakly depends on AR contrast. Therefore one can obtain for short lasers a FR of $\sim 10^5$ Hz in magnitude.

Figure 3b shows gain-frequency dependences of the constructed ARC, which has frequency band width $\Delta f_{\text{AFC}} = 20$ kHz.

Figure 3a shows the mode beat signal, observed on a spectrum analyzer screen with the DML frequency stabilized by the double AFC. The full width at half maximum (FWHM) of the observed signal corresponds to the laser frequency spectrum, multiplied by the derivative $\partial\omega_1/\partial\omega_{12}$ in the center of the absorption lines. For our experimental conditions the value of $\partial\omega_1/\partial\omega_{12}$ varied within the range 5–1.2. Taking into account the value of $\partial\omega_1/\partial\omega_{12}$ the measurements gave the stabilized laser frequency spectrum $\Delta\omega_S^{\text{act}} \lesssim 10$ Hz. The passive laser frequency spectrum was about 8 kHz. The main contribution to the $\Delta\omega_S^{\text{act}}$ was made by the fluctuations of a radiogenerator used in the AFC frequency-to-voltage converter. Other parameters of the DML are as follows: $\ell_0 = 1.5$ m, $\ell_- = 0.8$ m, $d = 2.10^{-3}$ m, $p_- = 0.3$ Pa. We

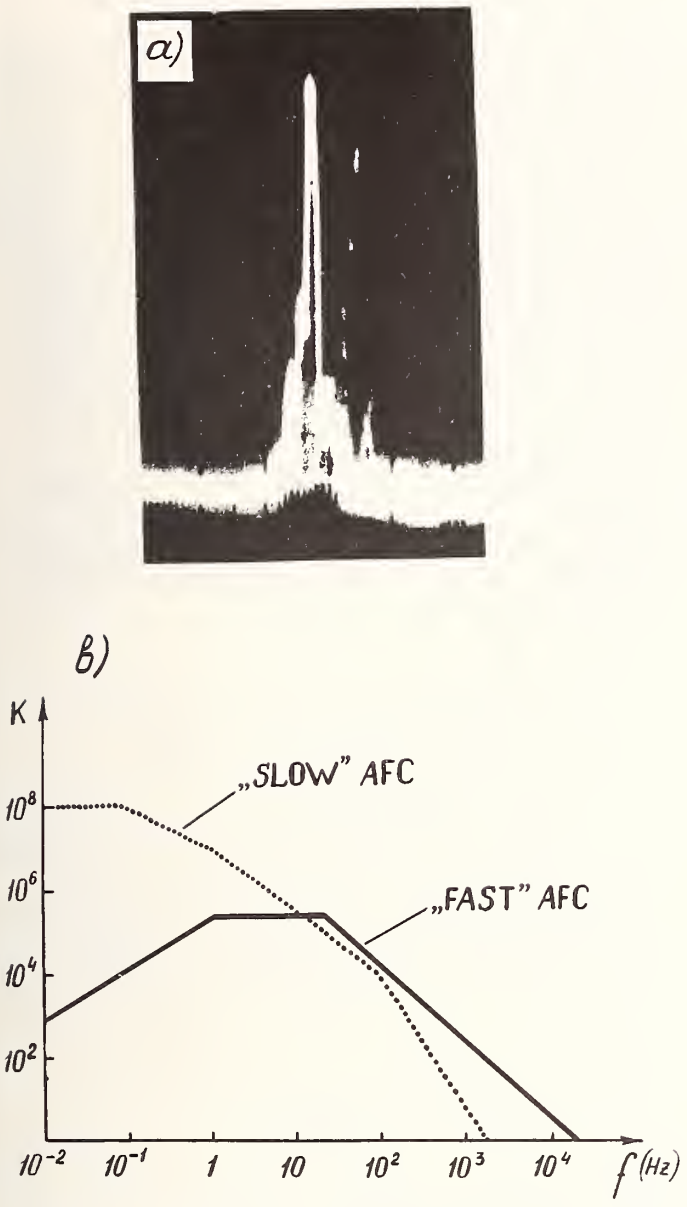


FIGURE 3. *a) The mode beat frequency spectrum of the DML stabilized with the two-loop AFC system. b) The gain-frequency dependence of the "slow" and "fast" AFC parts. The unity gain is at 20 kHz.*

hope that further development of the method would allow one to obtain the natural laser spectral width.

3.2 Sensitivity of FR Registration in DML

The investigation of the FR shows that frequency detection of resonances in the DML may be more sensitive in comparison with AR and FR registration in the SML. This feature of the DML FR is especially important for super-narrow reference line detection ($\gamma_- \lesssim 1$ kHz) with low pressures of the absorption gas ($p_- = 1-10$ mPa).

Three noise sources must be taken into account for the FR detection: additive noise, F/V transformation noise, and natural frequency noise of the radiation. Using well known relations [8] we may show that for additive noise the S/N ratio (by voltage) for the frequency detection is essentially more than for the amplitude detection:

$$\left.\frac{(S/N)_{FR}}{(S/N)_{AR}}\right|_{add} \approx \left[\frac{\Delta\omega_{12}}{f_m}\right] \cdot \left[\frac{E_1 E_2}{a_-}\right] \qquad (3)$$

where E_1, E_2 are electric fields of the modes; f_m is the modulation frequency (synchronous detection is supposed); a_- is the value of AR in SML or in DML; and $\Delta\omega_{12}$ is defined by Eq. (2).

The a_-/E_1E_2 ratio is equal approximately to the AR contrast. Equation (3) shows that when $f_m \sim 10^2$ Hz, $a_-/E_1E_2 \sim 10^{-1}$, $\Delta\omega_{12} \sim 10^5$ Hz the gain in the S/N ratio may be a few orders of magnitude. In practice the frequency characteristics of the photodetector must be taken into account in addition to (3).

As to the F/V transformation noise the main contribution to its spectral density is made in the optical heterodyne process. In the DML the second mode is used as the internal local oscillator and its frequency technical fluctuations are strongly correlated with the first mode's fluctuations if the following condition is realized: $\gamma_- \ll \Delta\omega_{12} \ll \gamma_+$, $\Delta\omega_D^{\pm}$ (γ_+ is the gain line homogeneous width, $\Delta\omega_D^{\pm}$ is the Doppler width of the gain and absorption lines).

In the SML an external heterodyne laser is needed for the FR registration. If the heterodyne laser frequency is highly stabilized with the help of a frequency-feedback system to the reference laser, the main contribution to the F/V transformation noise is made by the short-term instability of the laser under investigation.

The spectral density of the technical frequency noise of gas lasers in the low frequency range (less than 10 kHz) is about $10^4 - 10^8$ Hz^2/Hz. When synchronous detection is used the S/N ratio is determined by the spectral density at the modulation frequency. The problem can't be solved by means of active laser stabilization since the AFC system will decrease the frequency fluctuations simultaneously with the signal ($\Delta\omega_{12}$) and hence the S/N ratio will not be improved. It is possible in Refs. [5, 6] that poor passive short-term stability gave no opportunity to obtain any gain in the S/N ratio in case of the FR registration in contrast to AR.

So, DML allows one to detect FR with the noise level equal to the natural frequency noises. On the other hand the FR signal, given by Eq. (2), under conditions where $\mathcal{K}_1 \sim 1$, is determined only by the linear absorption coefficient, i.e., linearly depends on the pressure p_-. (As is known, for the AR $a_- \sim p_-^3$.)

We now estimate the limit of sensitivity of weak line detection by the proposed method. As was mentioned above for the $F_2^{(2)}$ methane line, which has an absorption coefficient per unit length of 0.135 m^{-1} Pa^{-1}, the FR is about 10^5 Hz when $p_- = 100$–200 mPa, $\ell_- \sim 0.5\,\ell_0$. Assuming that the spectral density of the natural frequency noises is $\sim 10^{-3}$ Hz^2/Hz one may find that the minimum detectable methane pressure is $(10^{-4}-10^{-5})$ mPa. Therefore FR registration with S/N $\sim 10^3$ is possible if $p_- \sim 10^{-1}$ mPa, $\tau = 1$ s and short absorption cells are used.

3.3 Resolution of the $F_2^{(2)}$ Methane Line hfs in DML

For magnetic hfs resolution of the $F_2^{(2)}$ line [9] a He-Ne/CH_4 DML with an internal telescope expander (TE) was constructed. The light beam diameter in the absorption cell was about 3×10^{-2} m, while the spot size in the active medium was 3×10^{-3} m; $\ell'_- = 0.8$ m, $\ell'_0 = 1.8$ m. The whole experimental set-up also consisted of the

reference DML and heterodyne laser. The present scheme is acceptable when precision spectroscopic investigations are made. The reference He-Ne/CH_4 DML (without TE) has a narrow spectrum ($\Delta\omega_S^{act} \sim 10$ Hz) and was constructed according to the scheme described in section 3.1. DML/TE was studied in the pressure range $p_- \leqslant 300$ mPa. Since there exists a time shift of the beat frequency ω_{12}^0, e.g., due to a slow change of the active medium parameters, synchronous detection of the FR with $f_m = 400$ Hz was used.

The first and second derivative signals of the FR are shown in Figs. 4 and 5 for different p_-, $\mathcal{K}_1$ values. Under high methane pressures (Fig. 4a, $p_- = 270$ mPa) when magnetic hfs isn't resolved the DML/TE frequency stabilization was made by means of the second derivative of the FR. The measurement of the relative frequency stability of the DML/TE and DML gave an Allan variance $\sigma = 0.5 \times 10^{-14}$ ($\tau = 10$ s) which is better by two orders of magnitude than the one reached in Refs. [5, 6]. (In Refs. [5, 6] a SML stabilized by the FR was used.) The asymmetry of the FR and the "crossing" lines [9] are observed clearly at the $p_- = 40$ mPa (Fig. 4b).

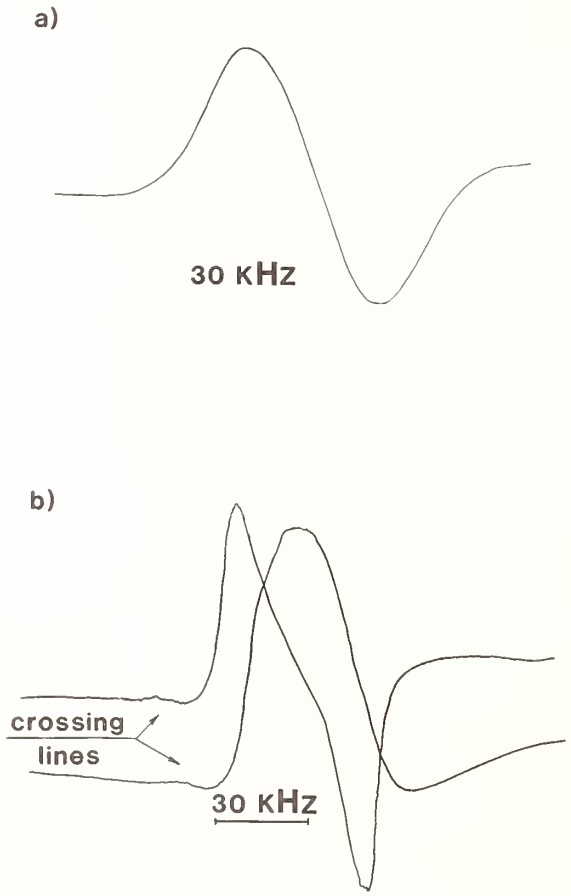


FIGURE 4. *The first and second derivative FR signals. a) $p_- = 270$ mPa; the time constant $\tau = 0.3$ s; b) $p_- = 40$ mPa, $\tau = 1$ s; the "crossing" lines are observed at the red side of the main components.*

The spectrograms in Fig. 5a, b were obtained at equal pressure ($p_- = 3$–4 mPa) but under two different intensities of the first mode; $\mathcal{K}_1^a = 0.2$ and $\mathcal{K}_1^b = 1.0$. Under conditions corresponding to Fig. 5b, the half width at half maximum (HWHM) of the first derivative FR signal is $\gamma'_- = (2.7 \pm 0.3)$ kHz. The transit flight time width of the FR is about 5 kHz. The last spectrograms show that the spectral resolution and S/N ratio obtained in the present work are equal to those of Ref. [9] where the hfs of the $F_2^{(2)}$ line had been resolved for the first time. However the parameters of the experimental set-ups differ in essential ways. Comparison of the absorption particle number in the mode volume and the signal registration time show that the detection sensitivity of the supernarrow resonances reached in the DML is about two orders higher than the one in Ref. [9].

In conclusion it may be noted that the DML with internal absorption cell represents in itself an optical spectroscope with frequency detection of the signal which may significantly increase the sensitivity and resolution of sub-Doppler spectroscopy and distinguish high quality reference lines as well.

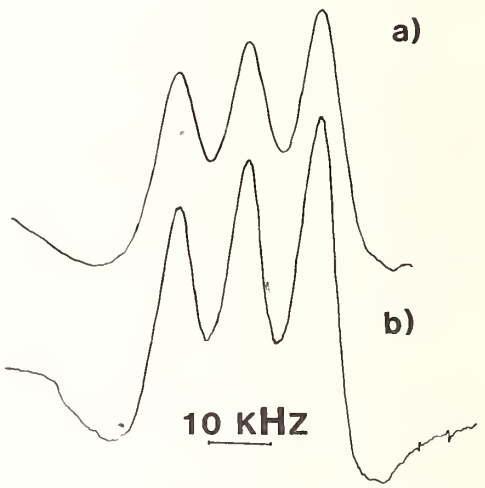


FIGURE 5. *Magnetic hfs of the $F_2^{(2)}$ methane line. $p_- = 3$–4 mPa, $\tau = 15$ s, light beam diameter = 30 mm, length of the cell = 0.8 m; a) $\mathcal{K}_1 = 0.2$; b) $\mathcal{K}_1 = 1.0$.*

References

[1] J. L. Hall, Science **202**, 147 (1978).
[2] M. A. Gubin *et al.*, Kvant. Elektron. (Moscow) **6**, 63 (1979) [Sov. J. Quantum Electron. **9**, 34 (1979)].
[3] V. S. Letokhov, Pis'ma Zh. Eksp. Teor. Fiz. **6**, 597 (1967) [JETP Lett. **6**, 101 (1967)].
[4] Yu. A. Vdovin *et al.*, Kvant. Elekton. (Moscow) **2**, 105 (1973). [Sov. J. Quantum Electron. **2**, 565 (1973)].
[5] G. Kramer *et al.*, Z. Naturforsch. **30a**, 1128 (1975).
[6] S. N. Bagaev *et al.*, Appl. Phys. **10**, 231 (1976).
[7] V. P. Chebotayev, Report at XIX General Assembly of URSI, Helsinki, Finland (1978).
[8] J. Klapper and J. Frankle, *Phase-locked and Frequency-Feedback Systems*, (Academic Press, New York, 1972).
[9] J. L. Hall and C. J. Borde, Phys. Rev. Lett. **30**, 1101 (1973).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

He-Ne ($^{127}I_2$) Lasers at 0.633 μm (and at 0.604 μm)

Fabrizio Bertinetto, Bruno I. Rebaglia, Paolo Cordiale, Sergio Fontana, and Gian Bartolo Picotto

Istituto di Metrologia G. Colonnetti, Torino, Italy

Although iodine stabilized, 0.633 μm lasers are used as practical wavelength standards, no common agreement exists as to the operating conditions. It is shown that on the basis of such an agreement, reproducibility of ±20 kHz or $\pm 4 \times 10^{-11}\nu$ can be attained. This study proposes such conditions and shows that for reproducibility to exceed $10^{-10}\nu$, certain cavity configurations must be discarded. Preliminary observations of strong absorption lines of iodine at the emission wavelength of 0.604 μm are also reported.

Key words: frequency stabilized lasers; frequency standards; wavelength standards.

1. He-Ne ($^{127}I_2$) Lasers at 0.633 μm

1.1 Stability and Reproducibility

Figure 1 is a plot of the Allan variance of two independently stabilized He-Ne ($^{127}I_2$) lasers with the temperature of the iodine cell walls as a parameter. The operating conditions were: iodine pressure $P = 17.4$ Pa and modulation depth $\Delta\nu_{p-p} = 6$ MHz. The improvement in the short term stability due to enhanced signal to noise ratio [1] is evident and can be appreciated in the practical utilization of the laser.

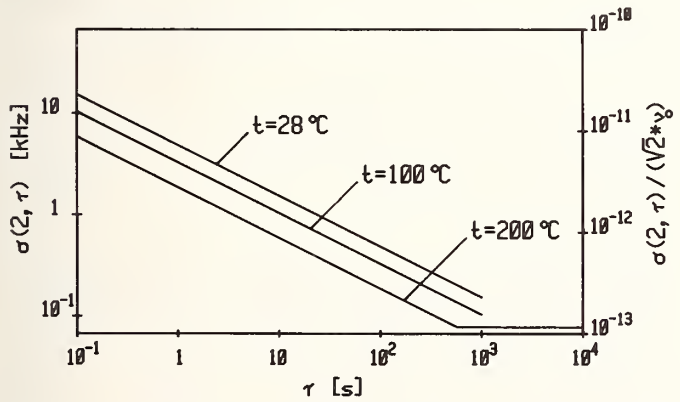


FIGURE 1. *Allan variance plot for two 0.633 μm He-Ne ($^{127}I_2$) lasers at different temperatures of the iodine cell.*

In order to check the frequency reproducibility, three lasers with identical cavity structures and servo-systems were used. Interchanging these servo-systems amongst the lasers, all combinations of lasers and servos were obtained. For each of these combinations separately, frequency differences between the lasers were measured in such a way that in every two sets of measurements, i.e., two different combinations of lasers and servos, one laser with its servo-system was kept unaltered so as to be used as a reference laser. It was further assumed that the frequency of this reference laser remained unchanged between two measurements. As a result the reproducibility was approximately ±5 kHz or $10^{-11}\nu$ as shown in Fig. 2.

Similar results can be obtained in any metrological laboratory and it has already been shown that lasers constructed in different laboratories, but under the same operating conditions, show a repeatibility of ±20 kHz or $\pm 4 \times 10^{-11}\nu$ [2]. However, for arbitrary operating conditions and for certain cavity geometries, the repeatibility of the frequency output[1] of He-Ne ($^{127}I_2$) lasers can be no better than $\pm 5 \times 10^{-10}\nu$, as will be shown below.

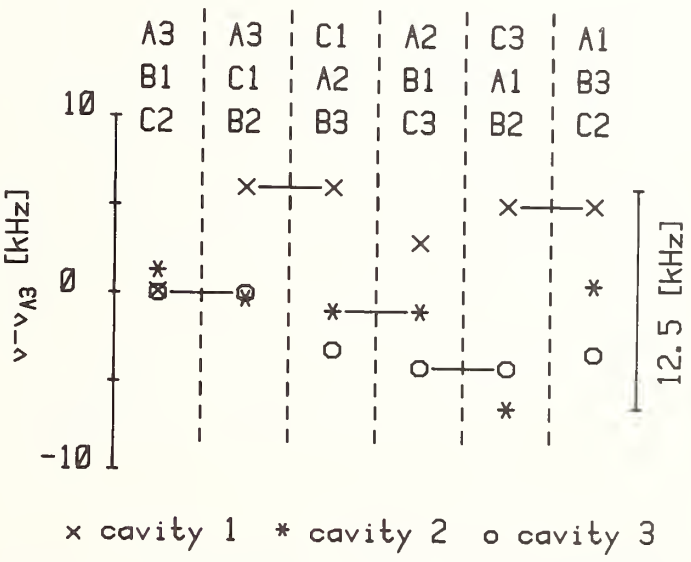


FIGURE 2. *Reproducibility of He-Ne ($^{127}I_2$) lasers constructed at IMGC. The three servo-systems employed are designated A, B, and C and the three cavities 1, 2, and 3.*

[1]This study deals with the frequency measurements of the hfs components *d*, *e*, *f*, *g* of iodine ($^{127}I_2$) against a fixed reference. Only the behavior of the center of gravity of these peaks is reported here. The frequency, therefore, refers to that of the center of gravity. It is worth mentioning here that the group of hfs components *h*, *i* and *j* behaves similarly to that of *d*, *e*, *f* and *g*.

1.2 Iodine Pressure

Figure 3 is a typical plot of frequency vs. pressure. Our result differs from that of P. Cérez *et al.* [3]. The straight line of −9.8 kHz/Pa slope was drawn only through the experimental points for a cold finger temperature $t_c \geq 9$ °C. It can be seen that the extrapolation below this temperature does not fit the data. At low pressures as the contrast of the peaks becomes poor, the offsets introduced by the electronics of the servo-system become important. Nevertheless the behavior shown in Fig. 3 has been obtained consistently.

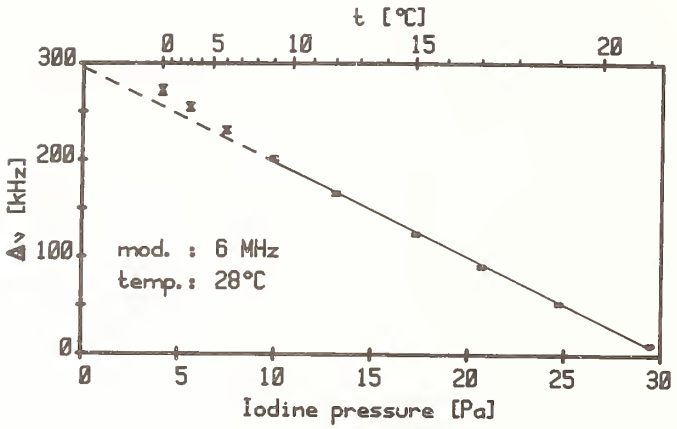


FIGURE 3. *Frequency output vs. iodine pressure. Extrapolation below $t_c = 9$ °C seems incorrect for lasers at 0.633 μm.*

1.3 Misalignment and Cavity Geometry

The effects of misalignment and cavity geometry are shown in Fig. 4. These curves were obtained by tilting the mirror on the cell side (M1). Tilting the other mirror (M2) did not change the curves significantly. At maximum output power (best alignment), the frequency output was not affected by the geometry of the laser cavity, except for curve d, in which case there was an offset of approximately 100 kHz. Repeating these measurements with different cavity lengths and laser powers, similar results were obtained which were independent of the output power at best alignment. Shortening the cavity to 300 mm using a shorter tube (otherwise similar to the previous ones), the frequency offset observed with the same mirrors as of curve d (Fig. 4) was 35 kHz. For this cavity length and for mirrors as of curve a and a′ output was insensitive to misalignment within ±10 kHz.

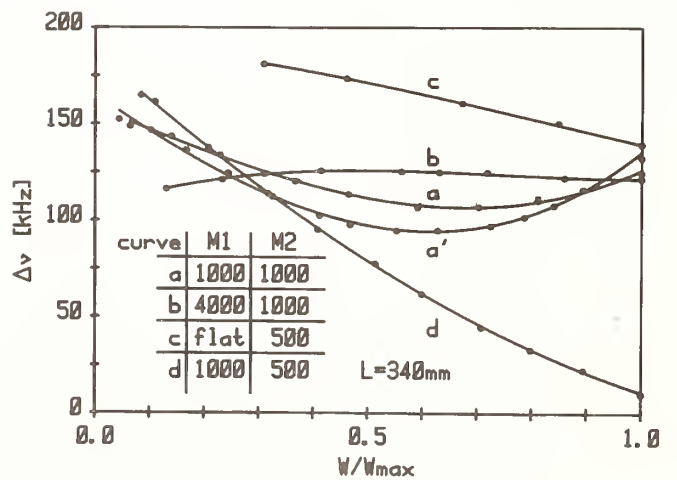


FIGURE 4. *Effect of misalignment and cavity geometry on frequency output. Radiis of curvature of M1 and M2 are in millimeters. Mirror reflectivity for curve a 99% for both mirrors, and for a′ 99.0% and 99.4%. The effect of reflectivity and hence irradiance is small compared to that of cavity configuration.*

Misalignment also had an effect on the slope of the modulation dependent frequency shift. For a cavity similar to that of curve d (Fig. 4) and at iodine pressure of 17.4 Pa, the slope at $\Delta\nu_{p-p} = 6$ MHz was −10.5 kHz/MHz at best alignment and −4.7 kHz/MHz when the power was reduced to 35% by tilting one of the mirrors.

1.4 Irradiance

There have been attempts to interpret these results only as an effect of irradiance of the beam traversing the iodine cell [4]. To this end the frequency shift expected from irradiance has been recently measured using two different methods independent of cavity misalignment. The data shown in Fig. 5 were obtained using a laser cavity similar to that of curve a (Fig. 4), but with mirrors of different reflectivities. The error bars are twice the standard deviation. Optical surfaces were cleaned between two measurements using the same mirrors. The dispersion of the frequency output is well within the reproducibility of He-Ne ($^{127}I_2$) lasers.

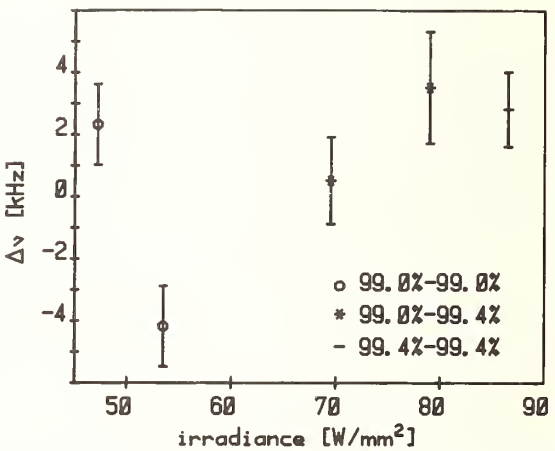


FIGURE 5. *Frequency output vs. irradiance for a He-Ne ($^{127}I_2$) laser at 0.633 μm.*

In another set of measurements the power was varied from 44 μW to 91 μW by varying the discharge current from 2.5 mA to 3.5 mA. The variations in the frequency output were within 1.5 kHz, or $3 \times 10^{-12}\nu$. It can then be concluded that irradiance had a very small effect on the frequency output of a He-Ne ($^{127}I_2$) laser at 0.633 μm.

However the proximity of the neighboring lines causes some spread in frequency through power broadening. Nevertheless, the modulation-dependent shift in the frequency of the center of gravity of a group of adjacent peaks is not affected by irradiance (see Table 1). That the pressure-dependent shift is only slightly affected by irradiance can also be seen from Table 1.

1.5 Temperature of the Cell

Variations in the frequency output of a He-Ne ($^{127}I_2$) laser with temperature of the iodine cell are shown in Fig. 6. At a pressure of 17.4 Pa and $\Delta\nu_{p-p} = 6$ MHz the frequency shift associated with the cell temperature was found to be approximately 0.16 kHz/K. The effect

TABLE 1. *Pressure-dependent and modulation-dependent frequency shift at different values of irradiance. The measurements were taken in conjunction with those of Fig. 5.*

H [W/mm²]	$\Delta\nu/P$ [kHz/Pa]	$\Delta\nu/\Delta\nu_{p\text{-}p}$ [kHz/MHz]
47	-9.0	-10.2
54	—	-11.2
70	-9.8	-10.4
79	—	-10.7
87	-10.2	-10.6

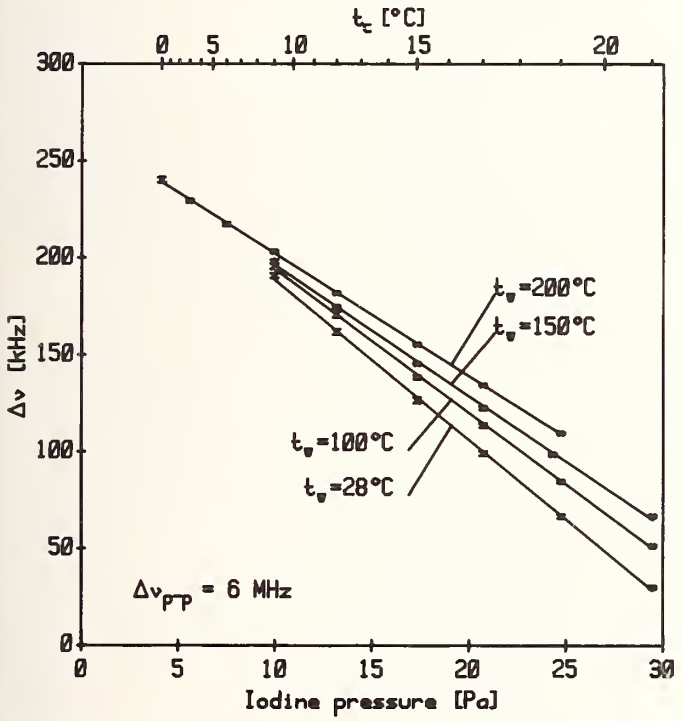


FIGURE 6. *Frequency output vs. iodine pressure at different temperatures of the cell. For t_w < 200 °C, data below t_c = 9 °C are not shown.*

shown in Fig. 3 disappears at 200 °C and a solid line has then been drawn for this temperature starting from a pressure of 4.1 Pa.

Heating of the cell seemed to improve the line shape as a small decrease in the modulation-dependent frequency shift, from −9.9 kHz/MHz at 28 °C to −8.4 kHz/MHz at 200 °C at a constant iodine pressure of 17.4 Pa, was observed.

2. He-Ne ($^{127}I_2$) Laser at 0.604 μm

It is known [5, 6] that the 0.612 μm orange line of He-Ne is coincident with iodine transitions which are strong and easily saturable. Further investigations are in progress at IMGC.

A laser using a commercial plasma tube is under construction and with the proper choice of mirrors is expected to give 25 mW output at 0.633 μm. Using two mirrors each of 1200 mm radius of curvature and 99.8% reflectivity, the laser could be operated at 0.604 μm [7], with a power output of only 25 μW. The effect of an intracavity cell was the same as that at 0.612 μm and the laser had single mode operation over several hundreds of megahertz. Figure 7 shows the emission profile and the third derivative signal at an iodine pressure of 0.8 Pa and a temperature of 160 °C. The lines seen are as yet unidentified. A large modulation (nearly 15 mHz peak-to-peak) had to be applied because of a frequency jitter of approximately 10 MHz due mainly to the poor mechanical mount.

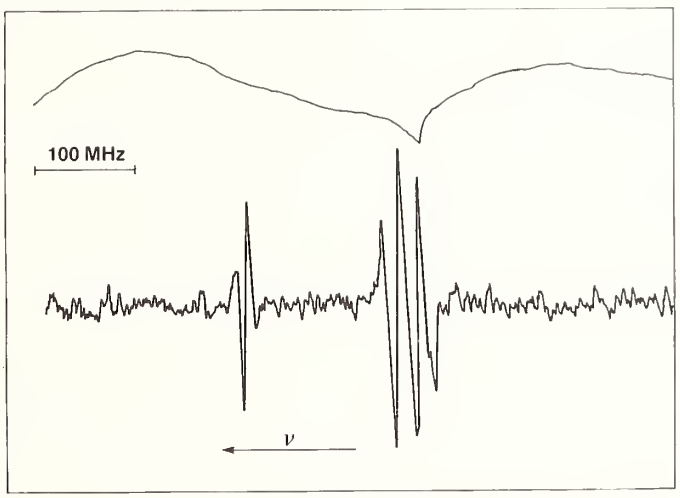


FIGURE 7. *Output power and third harmonic signal vs. laser frequency for a He-Ne ($^{127}I_2$) laser at 0.604 μm.*

3. Conclusions

In spite of the fact that the stability of the He-Ne ($^{127}I_2$) lasers at 0.633 μm approaches $10^{-13}\,\nu$ and the reproducibility obtainable in a single laboratory is better than ±10 kHz or $\pm 2 \times 10^{-11}\,\nu$, the repeatability between lasers constructed and operated in different laboratories may fall short of $10^{-10}\,\nu$ in the absence of an agreement about the operating conditions and specifications of cavity geometry. Towards using these lasers as practical secondary frequency standards in the optical region, the following operating conditions are suggested:

$$P = (17.4 \pm 1)\ \text{Pa}$$

$$\Delta\nu_{p-p} = (6 \pm 0.2)\ \text{MHz}$$

$$t_w = (200 \pm 5)\ ^\circ\text{C}$$

Several cavity configurations give similar performances; however, cavities as of curve d in Fig. 4 should be discarded. It is probable that for He-Ne ($^{127}I_2$) lasers at 0.612 μm and at 0.604 μm a similar problem of choosing the cavity may occur in the future, if the lasers are operated with an intracavity cell.

References

[1] P. Cérez and S. J. Bennet, IEEE Trans. Instrum. Meas. IM-27, 396 (1978).

[2] J. M. Chartier *et al.*, IEEE Trans. Instrum. Meas. **IM-25**, 450.
[3] P. Cérez *et al.*, PTB-Bericht **Me-17**, Ed. by F. Bayer-Helms (Physikalisch-Technische Bundesanstalt, Braunschweig, May 1977) p. 71.
[4] F. Bertinetto *et al.*, CPEM Digest, Conference on Precision Electromagnetic Measurements, IEEE Cat. No. 78CH1320–1 IM, p. 79.
[5] P. Cérez *et al.*, IEEE Trans. Instrum. Meas. **IM-29**, 352 (1980).
[6] K. Dschao *et al.*, IEEE Trans. Instrum. Meas. **IM-29**, 354 (1980).
[7] A. D. White and J. D. Rigden, Appl. Phys. Lett. **2**, 211 (1963).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Recent Work on 612 nm He-Ne Stabilized Lasers

A. Brillet, P. Cérez, and C. N. Man-Pichot

Laboratoire de l'Horloge Atomique, Equipe de Recherche du CNRS, associée à l'Université Paris-Sud, Bât. 221 - Université Paris-Sud, 91405 - Orsay - France

We report on the metrological properties of 612 nm He-Ne lasers frequency stabilized on $^{127}I_2$ saturated absorption lines. We describe the new results of spectroscopic and metrological interest, obtained both with conventional internal cell devices and with the new technique using an external cell inside a Fabry-Perot resonator. A reproducibility of 6×10^{-13} is obtained with this last technique.

Key words: He-Ne laser; hyperfine predissociation; optical frequency standards; saturated absorption.

1. Introduction

He-Ne lasers stablized by saturated absorption in $^{127}I_2$ at 633 nm [1–3] or 612 nm [4] are now widely used in precision interferometry and high resolution spectroscopy at the $10^{-10} - 10^{-11}$ precision level. We show in this paper how both the high saturation level obtained with internal cell devices and the very high resolution capability of the external cell technique [5] allowed us to obtain new spectroscopic results on the 612 nm R(47)9-2 iodine line, which in turn, helped in improving the metrological properties of these optical frequency standards.

2. Experimental Techniques

2.1 Internal Cell Devices

The cavity of our He-Ne laser is made of three silica rods and contains a commercial He-Ne tube, designed to produce 5 to 7 mW of 633 radiation, a 10 cm long iodine cell with Brewster windows, and a Littrow prism for wavelength selection. Its free spectral range is about 200 MHz. Because of the low gain at 612 nm, both mirrors are high reflectivity, and the power emitted through each mirror at low iodine pressure is 300 μW at most, strongly depending on the cleanliness of the optical surfaces. This corresponds to an intracavity power of nearly 1 W! Even at the lowest iodine pressure available (0.2 Pa), the laser is single frequency and continuously tunable over 1–200 MHz, due to a differential saturation effect [4].

These lasers were used for internal cell saturated absorption studies, with iodine pressures ranging between 2 and 5 Pa. They were also used as sources for the external cell studies, with a much lower iodine pressure (0.1 to 1 Pa). Their emission linewidth, with good passive isolation, has now been reduced to 30 kHz (6×10^{-11}).

2.2 External Cell Technique

2.2.1 Optical Set-up

This new saturated absorption technique was first proposed by Cole in 1975 [6], and first demonstrated by C. Pichot-Man *et al.* [7]. The high sensitivity of this technique is easy to understand with simple arguments. Let us consider a resonator with an input mirror M_1, an absorbing cell, and an output mirror M_2, having respectively the transmissivities t_1, t, and t_2 for the field amplitude. If the mirrors are perfect, their reflectivities are $r_1 = |1 - t_1^2|^{1/2}$ and $r_2 = |1 - t_2^2|^{1/2}$. At resonance, the intensity transmission of the cavity is then:

$$T = \frac{t_1^2 t_2^2}{(1 - t\ r_1 r_2)^2}, \tag{1}$$

and its variation with t is:

$$\frac{\partial T}{\partial t} = \frac{-2t_1^2 t_2^2 t}{(1 - t\ r_1 r_2)^3}. \tag{2}$$

For $r_1^2 = r_2^2 = 0.98$, $t^2 = 0.98$, we get $T = 0.45$, and $\partial T/\partial t = 40$, which represents the gain in sensitivity compared to the case of a simple saturated absorption experiment. Under the same conditions, the energy storage factor inside the resonator is:

$$S = \frac{t_1^2}{(1 - t\ r_1 r_2)^2} = 22.7. \tag{3}$$

Thus this technique allows one to perform saturated absorption spectroscopy even with very weak laser sources.

The numerical values given above correspond approximately to our experimental conditions. Let us note that Eqs. (1) to (3), which do not include the intensity dependence of t, are valid only for low saturation parameters. They constitute, however, a valid approximation, since we always tried to avoid power broadening. If the saturation effects become important, the sensitivity $\partial T/\partial t$ further increases, up to the limit of bistability.

A block diagram of the experimental set-up is shown in Fig. 1. The resonator is made of two invar rods, 50 cm long. It contains a 20 cm or 35 cm iodine cell, and an anti-reflection coated lens acting as a telescope. The beam waist on mirror M_2 can be varied between $w_0 = 0.2$ and $w_0 = 3$ mm, by translating the lens.

Feedback to the laser is prevented by a Faraday or an acousto-optic isolator.

2.2.2 Electronic Set-up

To operate this spectrometer, one wants to keep the resonator frequency locked onto the laser frequency. This is realized by a conventional servo-loop using frequency modulation of the resonator at 20 kHz (kc in the figure) and synchronous detection. It is important to obtain a very tight lock in order to avoid the excess amplitude noise which would result from a relative frequency jitter between the laser and the resonator.

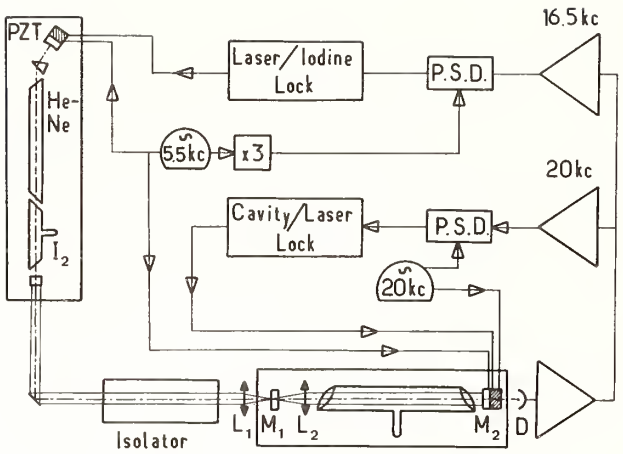


FIGURE 1. *Block diagram of an external cell system.*

The iodine saturated absorption features are detected by a second frequency modulation, at 5.5 kHz, which is applied simultaneously to the laser and the resonator with the same phase and amplitude so that the signal detected at 5.5 kHz or its 3rd harmonic is due only to the iodine features and not to a residual frequency mismatch between the laser and the resonator. This signal is synchronously detected and recorded, or used as an error signal to lock the laser frequency.

3. Some Relevant Spectroscopic Properties of $^{127}I_2$ at 612 nm

3.1 The Saturated Absorption Spectrum

Figure 2 shows the whole spectrum which can be registered with an internal cell device, completed by higher resolution studies with an external cell. One can easily assign the 21 hyperfine components of R(47)9-2, the 15 from R(48)15-5, and the 10 highest frequency components from P(48)11-3. Furthermore, the spectrum contains over 40 additional lines, shown pointing downwards on Fig. 2. They appear clearly only at high intensity, and we could assign most of them to "forbidden," $\Delta F = 0$ transitions and to Doppler generated cross-over resonances between these transitions and allowed ones. Moreover, this assignment allows one to predict the frequency of some other cross-over resonances which could not be resolved because they were too close to strong allowed transitions, like m, *q*, r (Fig. 3).

The relative intensities of the allowed transitions compared to R(47)9-2 are about 0.8 for P(48)11-3 and 0.07 for R(48)15-5 at room temperature. It is possible to enhance by a factor 10 the 15-5 line intensity, using a heated cell.

At low power, the computed intensity of the saturated absorption signal is about 10^{-4} for $\Delta F = 0$ lines [10] and 10^{-2} for the cross-overs. With increasing power, however, the allowed lines broaden and their intensity saturates, while the $\Delta F = 0$ lines keep increasing linearly, so that the intensity ratio is only about 10 in the usual operating conditions of an internal cell device.

3.2 Linewidths and Hyperfine Predissociation

3.2.1 Instrumental Broadening

In internal cell devices, the problem has been treated earlier [8] for 633 nm lasers. At 612 nm, the linewidth of all the iodine components is about 6 MHz (FWHM), mainly due to power broadening. Changing the pressure does not affect the linewidth very much, because any decrease (increase) of the pressure broadening is nearly compensated by an increase (decrease) of the power broadening.

With the external cell technique, it is possible to obtain low saturation parameters. The main source of instrumental broadening is then the pressure broadening which is about 250 kHz/Pa around 1 Pa. Smaller additional ef-

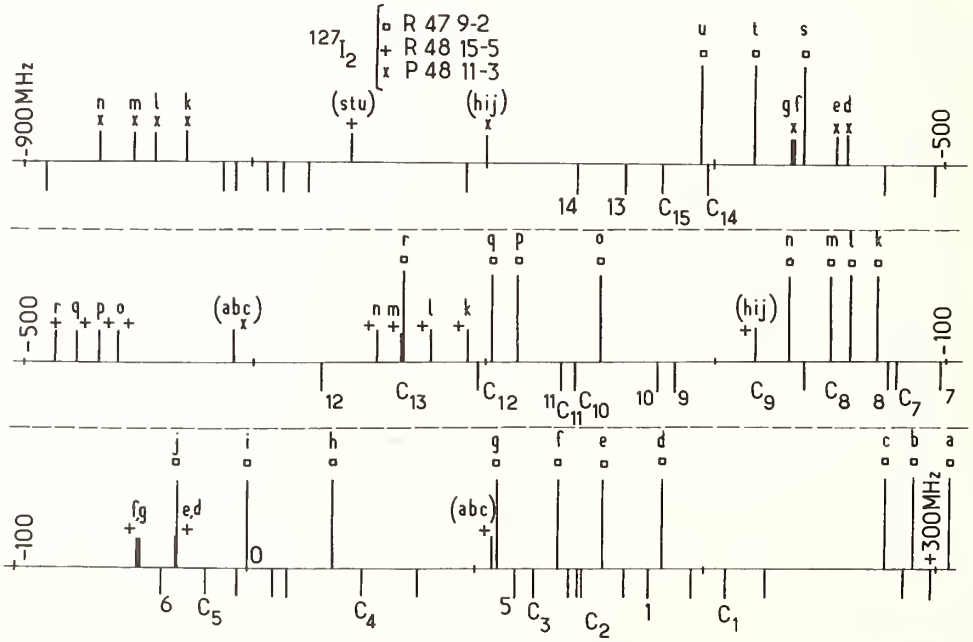


FIGURE 2. *$^{127}I_2$ saturated absorption spectrum at 612 nm. The lines shown pointing downwards are forbidden transitions and Doppler generated level crossings. The $\Delta f = 0$ lines, and cross-over resonances assignable to R(47)9–2 are labelled 1 . . . 14 and C_1 . . . C_{15}, respectively.*

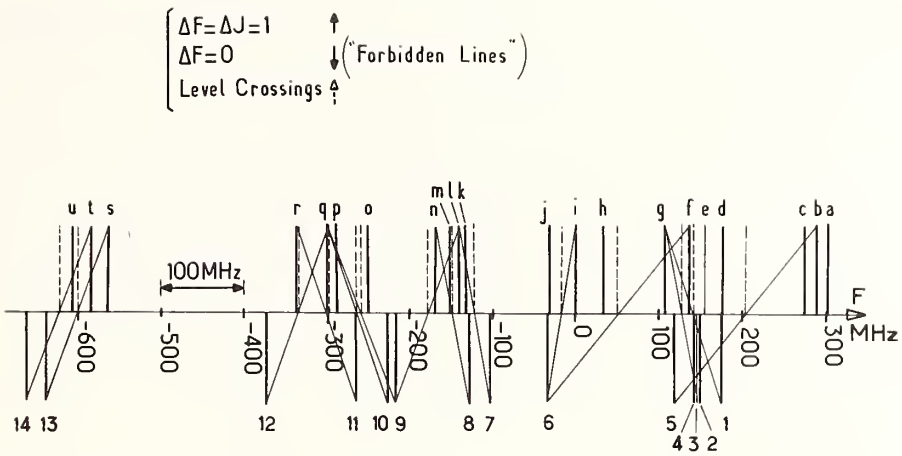


FIGURE 3. *R(47)9−2 hyperfine structure. The intersection of the diagonal lines (which connect some forbidden and allowed resonances) and the frequency axis give the cross-over positions.*

fects come from the laser frequency jitter (30 kHz) and from the finite size of the laser beam waist (10 kHz·mm).

3.2.2 Radiative Lifetimes and Predissociation

The natural linewidth of a given iodine hyperfine transition can be expressed by the formula

$$W = \frac{1}{2\pi}\left[\Gamma_I + \Gamma_V + C_V^2 J(J+1) + \Gamma_{HP}\right], \qquad (4)$$

where Γ_I is the instrumental broadening, Γ_V is the radiative lifetime of the level, $C_V^2 J(J+1)$ is the gyroscopic predissociation, and Γ_{HP}, the hyperfine predissociation rate, is given by [9]:

$$\Gamma_{HP}(V,J,,F) = \sum_I |\alpha(I,F,J)|^2 \times \left[\frac{a_V^2}{3}\left[\mathbf{I}^2 + \frac{3(\mathbf{I}\cdot\mathbf{J})^2 + \frac{3}{2}\mathbf{I}\cdot\mathbf{J} - \mathbf{I}^2\mathbf{J}^2}{(2J-1)(2J+3)}\right] - a_V C_V \sqrt{2}\,\mathbf{I}\cdot\mathbf{J}\right], \qquad (5)$$

where a_V and C_V are the predissociation constants. The projection coefficients $\alpha(I,F,J)$ can be determined by fitting the well-known hyperfine hamiltonian with the measured hyperfine splittings [10].

We specially studied the hyperfine predissociation of the R(47)9-2 line by carefully measuring the linewidth of all its components, and comparing with Eqs. (4, 5), after a zero pressure extrapolation. For a pressure of 1 Pa, for instance, the observed linewidths vary between 495 kHz for the component p and 1070 kHz for s. Choosing $a_g^2 = 135\ 10^3\ s^{-1}$, $C_g^2 = 118\ s^{-1}$, and $\Gamma_9 + \Gamma_I = 2.344\ 10^6\ s^{-1}$, the rms difference between experimental and calculated linewidths is only 12 kHz. However, the values of a_V and C_V are probably slightly underestimated because of the possible presence of a residual background gas in the iodine cells: since the collision broadening rates are larger for narrow lines than for wider lines due to velocity changing collisions [11], the zero pressure extrapolations would be biased. We estimate our values of a_V and C_V to be too small by about 10%, leading to $a_g^2 = (150 \pm 10) \times 10^3\ s^{-1}$ and $C_g^2 = (130 \pm 10)\ s^{-1}$. Our results also confirm the negative sign of $a_V C_V$.

4. Metrological Results

4.1 Internal Cell Devices

In these devices, the most important criterion for choosing the best frequency reference lines is the absence of neighboring lines or cross-over resonances. In this respect, only h, o, and t are suitable in the R(47)9−2 line. The stability of the lasers locked to o or t was comparable with the results obtained at 633 nm with heated cells [12]. Their reproducibility has been checked by comparison with a BIPM laser: both lasers produced the same frequency within 5 kHz (10^{-11}). The causes which limit this reproducibility are not understood. They are very difficult to isolate experimentally because all the parameters are strongly coupled in these compact systems. Anyway, it seems reasonable to use these simple devices for interferometry at the 10^{-10} level and to prefer external cell systems for higher precision measurements.

4.2 External Cell Systems

In this case, most lines are well isolated, because they are narrow and because the spectrum is simple, the intensity of forbidden lines being negligible at low power. Since the intensity of a saturated absorption feature is inversely proportional to the square of its width (see Fig. 4), the figure of merit,

$$M = \frac{(S/N)}{W}, \qquad (6)$$

is proportional to W^{-3}. It is then strongly recommended to use the narrowest lines as frequency references: m, n, p, and t are the best choices, j being perturbed by e R(48)15-5.

Using m and n components, we obtained a short term stability (Allan variance) which varies as $5 \times 10^{-12} \times \tau^{-1/2}$, with a flicker floor at 2×10^{-13}. We expect in the next experiments a tenfold improvement since the excess noise due to the frequency mismatch between the laser and the resonator has been greatly reduced.

The most important step forward is the improvement of the reproducibility which is 300 Hz (6×10^{-13}) with these systems. The iodine pressure shift is $-4 \times 10^{-12}\ Pa^{-1}$, and we can control the iodine pressure with a precision better than 0.05 Pa. A variation by a factor of 2 in

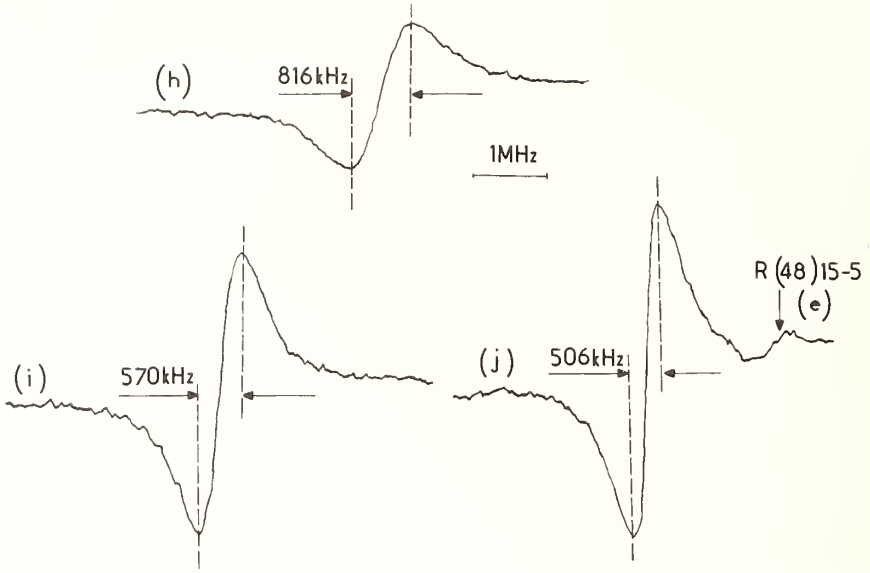


FIGURE 4. *First derivative lineshape of R(47)9−2, h, i, and j components, showing hyperfine predissociation.*

the power from any of the lasers induced no measurable frequency shift. Geometrical effects due to the mismatch of the counterpropagating waves [13] are completely negligible with this technique. The only limitation we could observe was due to the offsets of the servo-loops, so we think we might be able to improve this reproducibility which is already the best one ever reported for a visible optical frequency standard.

5. Conclusions

A detailed spectroscopic study of the $^{127}I_2$ properties at 612 nm allows us to determine the best reference lines to be used in internal and external cell devices, and helps in understanding the limitations of these systems. The new saturated absorption technique using an external cell inside a Fabry-Perot resonator proves to be highly efficient for high resolution spectroscopy and for use in optical frequency standards.

References

[1] G. R. Hanes and C. E. Dahlstrom, Appl. Phys. Lett. 14, 362 (1969).
[2] A. J. Wallard, J. Phys. E 5, 926 (1972).
[3] A. Brillet, P. Cérez, and H. Clergeot, IEEE J. Quantum Electron. **QE**-10, 526 (1974).
[4] P. Cérez and S. J. Bennett, Appl. Opt. 18,1079 (1979).
[5] P. Cérez, A. Brillet, C. N. Pichot-Man, and R. Felder, IEEE Trans. Instrum. Meas. **IM-29**, 352 (1980).
[6] J. B. Cole, J. Phys. D 8, 1392 (1975).
[7] C. N. Pichot-Man, P. Cérez, and A. Brillet, in Summaries of Contributions to the 11th Annual EGAS Conference (Laboratoire de Spectroscopie Hertzienne de l'ENS, Paris, 1979), p. 133.
[8] A. Brillet and P. Cérez, Metrologia 13, 137 (1977).
[9] M. Broyer, J. Viqué, and J. C. Lehmann, J. Chem. Phys. 64, 4793 (1976).
[10] J. P. Picque, private communication.
[11] J. L. Le Gouët, J. Phys. B 11, 3001 (1978).
[12] P. Cérez, S. J. Bennet, and C. Audoin, C. R. Acad. Sci. Ser. B 286, 53 (1978).
[13] J. L. Hall and C. J. Bordé, Appl. Phys. Lett. 23, 788 (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Iodine and Methane Stabilized He-Ne Lasers as Wavelength Standards

Shen Nai-cheng, Wu Yao-xiang, Sun Yi-min, Li Cheng-yang, and Zhang Xue-bin

National Institute of Metrology, Beijing, P.R.C.

and

Wang Chu

Peking University, Beijing, P.R.C.

The iodine and methane stabilized lasers designed by NIM and Peking University can be used as wavelength standards. We have compared the frequency differences of the lasers between NIM and BIPM in Paris in April 1980. The relative frequency differences are 2.9×10^{-11} for iodine and 6.3×10^{-12} for methane, respectively. When the laser power is given a fixed value, the frequency variation of the iodine stabilized laser can be very small. The power shift and standard power value are discussed in this paper.

Key words: frequency reproducibility; iodine and methane stabilized lasers; power shift; power standard value; wavelength standard.

1. Introduction

Iodine and methane stabilized He-Ne lasers have been used as wavelength standards since 1973 [1]. In particular, the 633 nm wavelength of $^{127}I_2$ is very useful and convenient. International intercomparisons of lasers have been done many times. We performed the first comparison between NIM and BIPM in April 1980 [2]. The results of our comparison were satisfactory. Although the construction of our lasers is different from that of BIPM, the frequency difference of the iodine stabilized lasers is less than 3×10^{-11}. This result gives strong proof of the good reproducibility among different 633 nm iodine stabilized lasers.

2. The Results of the Comparison between NIM and BIPM in 1980

2.1 The $^{127}I_2$ Stabilized He-Ne Laser

Table 1 shows the design characteristics of the $^{127}I_2$ stabilized lasers. The measurements were done at BIPM, April 2–17, 1980. The frequency difference and standard deviation of the two lasers were each determined from the small matrix of the measured values from the *d* to *g* components by a least-squares calculation. The mean frequency difference of thirteen measurements was calculated to be

$$f_{\mathrm{NIM2}} - f_{\mathrm{BIPM2}} = \quad 13.8 \text{ kHz}, \quad \sigma = 4.5 \text{ kHz}$$
$$f_{\mathrm{NIM1}} - f_{\mathrm{BIPM2}} = -20.3 \text{ kHz}, \quad \sigma = 3.8 \text{ kHz}$$

In the comparison, all lasers were adjusted to an iodine pressure of 17.3 Pa (15 °C) and a modulation amplitude of 6 MHz. These are standard operation conditions. The laser powers are slightly different between NIM and BIPM. The power regions used in the measurements were

$$17\ \mu\mathrm{W} < P_{\mathrm{NIM2}} < 27\ \mu\mathrm{W},$$
$$20\ \mu\mathrm{W} < P_{\mathrm{NIM1}} < 28\ \mu\mathrm{W},$$
$$25\ \mu\mathrm{W} < P_{\mathrm{BIPM2}} < 43\ \mu\mathrm{W}.$$

We have also measured the influence of the iodine pressure, modulation amplitude, and the laser power on the laser frequency. For example, the pressure shift of the *f* component is −9.3 and −9.2 kHz/Pa for NIM2 and NIM1, respectively, and the modulation shift of the *f* component is −7.9 kHz/MHz_{pp} for both lasers. For the other components, the shifts are slightly different. The

TABLE 1. *Design characteristics of $^{127}I_2$-stabilized laser.*

	Laser tube	I_2-cell	Reflective mirror		Output power
BIPM 2	CW301	10 cm	99.7%	99.5%	40 μW
	25 cm, ^{3}He-^{20}Ne		50 cm	∞	
NIM (1,2)	made in NIM	10 cm	99.7%	99.7%	25 μW
	11 cm, ^{3}He-^{nat}Ne		120 cm	120 cm	

power shifts are −0.05, −0.06, and −0.18 kHz/μW for NIM2, NIM1, and BIPM2, respectively.

2.2 The Methane Stabilized He-Ne Laser

The mean frequency difference was calculated to be

$$f_{\text{NIM-CH1}} - f_{\text{BICH4-6}} = 0.56 \text{ kHz}, \quad \sigma = 1.29 \text{ kHz}.$$

The coefficient of frequency variation for NIM-CH1 on the discharge current in the laser tube is −1.39 kHz/mA. The modulation shift of NIM-CH1 is +657 Hz/100 kHz_{pp}. For the comparison, the modulation amplitude of NIM-CH1 was adjusted to 1.2 MHz.

3. The Results of the Comparison Between NIM2 and NIM1 in 1981

The frequency difference between I_2-stabilized lasers NIM2 and NIM1 was determined to be 34 kHz at BIPM in 1980. We have measured the difference once again in January 1981. The mean value of thirteen measurements at various powers was calculated to be

$$f_{\text{NIM2}} - f_{\text{NIM1}} = 22.9 \text{ kHz}, \quad \sigma = 9.8 \text{ kHz}.$$

The laser tubes of NIM2 and NIM1 were replaced with new ones which were filled with the isotope ^{20}Ne in April 1981. The other characteristics of the lasers were the same as in 1980. The power of one laser was fixed, but that of the other was changed from its minimum to its maximum during the measurement, then the power condition of the two lasers was interchanged. When the region of the power variation was within 9 μW to 42 μW, the mean frequency difference varied from 7 kHz to 20 kHz and the standard deviations were about 6−11 kHz. So the power shift is too great for good frequency reproducibility. But as the powers of the two lasers were adjusted to about 30−32 μW, both the mean frequency difference of the small matrix and the standard deviations were very small. The value of the difference between NIM1 and NIM2 was only 6 kHz, with σ = 2 kHz. This result is better than the intercomparison for NIM at BIPM in 1980. We shall discuss this in the next section.

4. The Selection and Measurement of the Standard Value of the Laser Power for Good Reproducibility

In order to get good frequency stability and reproducibility, the absorption peak must have a high contrast and a narrow width. On the basis of Greenstein's theory [3], when the other operating conditions are fixed, we find the contrast of the absorption signal to be roughly proportional to $f(X)$, which is given by:

$$f(X) = (1 + I/I_A)^{-1/2} - (1 + 2I/I_A)^{-1/2} \qquad (1)$$

where $X = I/I_A$, I is the laser output power, and I_A is the saturation power of the iodine line. The linewidth half width at half maximum is well represented by the relation

$$\gamma = \gamma_0(1 + I/I_A)^{1/2} \qquad (2)$$

where γ_0 and γ are the linewidth without and with power broadening, respectively.

If we take $I/I_A = 0.6$, $f(X) = 0.117 = 0.89\, f(X)_{\max}$ (see Fig. 1), the broadening factor $(1 + I/I_A)^{1/2} = 1.26$. In this case, the contrast and width of the peak is better for a stabilized laser. The reason for selecting a standard value of I/I_A instead of I is that the contrast and width

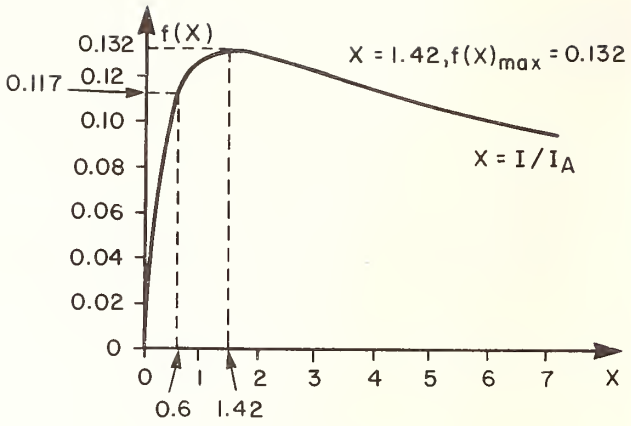


FIGURE 1. *f(X) versus X.*

depend on I/I_A. At the same time, the power broadening increases the asymmetry of the line; this is possibly a main source of the power shift. Thus, the power shift depends on I/I_A.

We have measured γ and I_A of NIM2 and NIM1 with a third harmonic locking technique. The third harmonic signal line shape $A_3(x)$ is given [4] by:

$$A_3(x) = -16x/m^3 + (\sqrt{2}/m^3 d) \times \{x(3m^2 - 4x^2 + 12)(d + 1 + m^2 - x^2)^{1/2} + (12x^2 - 3m^2 - 4)[d - (1 + m^2 - x^2)]^{1/2}\} \quad (3)$$

where

$$d = [(1 + m^2 - x^2)^2 + 4x^2]^{1/2} \qquad (4)$$

and x and m are the detuning and modulation amplitude divided by γ. The shape of the third harmonic signal is shown in Fig. 2.

We can see from Fig. 2 that the third harmonic signal crosses the axis at three points. Point 0 is the reference point of stabilization. The points on the two sides can be also locked, but their frequency values and γ depend on m. Typical relationships between these quantities are illustrated in Fig. 3. We have measured γ and I_A by using these relationships and the broadening formula of Eq. (2). The results are that γ_0 is 2.1 MHz and 1.9 MHz, I_A is 54 μW and 50 μW, and $I = (0.6)\, I_A$ is 32 μW and 30 μW for NIM1 and NIM2, respectively.

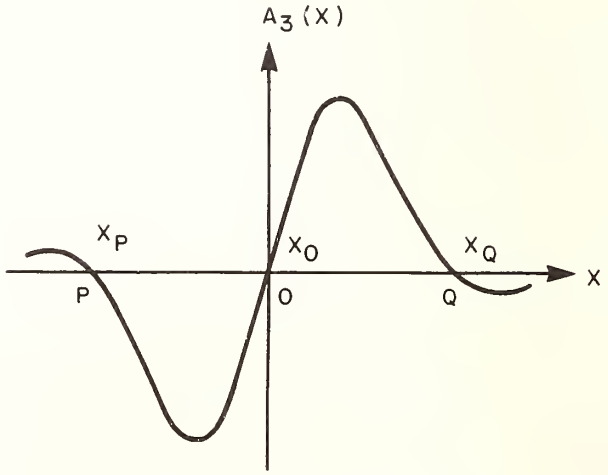


FIGURE 2. *Profile diagram of $A_3(x)$.*

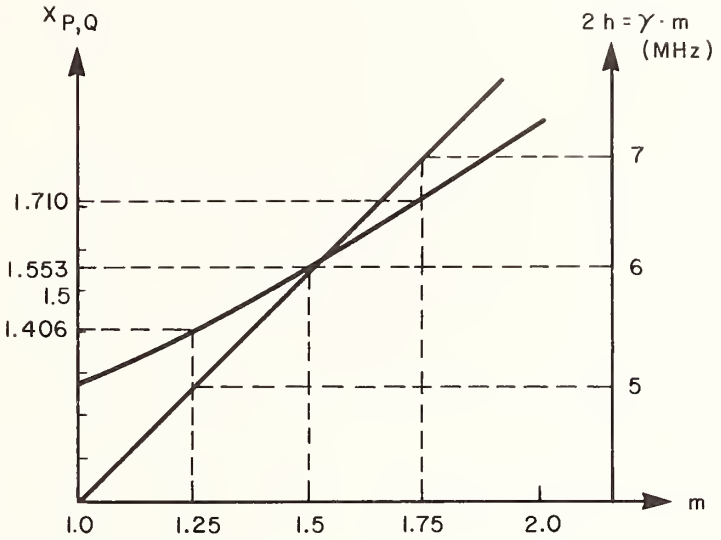


FIGURE 3. *Curve showing the dependence of $x_{P,Q}$ and γ on m.*

We have also measured the line width at various pressures by this method. The pressure broadening is about 36.5 kHz/Pa [6], so that the total broadening is about 630 kHz at 17.3 Pa, and the natural width is about 1.25 MHz.

The asymmetry of the line was observed in locking measurements. The amount of asymmetry can be expressed as the difference between the absolute values of x_P and x_Q. This asymmetry increased as the power was gradually increased. This may have been caused by a frequency-dependent gas lens effect [5] in iodine. We have not found any increase in asymmetry when the pressure or modulation amplitude was increased. It is possible that the above asymmetry is the main reason for the power shift. Thus we select a small value of I/I_A as the standard value of laser power. Our selected value of I/I_A is 0.6; in the case of our lasers, the power is about 30 to 32 μW.

5. Conclusion

The power shift is the main effect limiting the reproducibility of laser frequency. We hope to determine an appropriate standard value of the laser power in a manner similar to that for the pressure and modulation amplitude. We shall check the method by further experiments [6].

References

[1] CCDM Recommendation M1 (1973); J. Terrien, Metrologia 10, 75 (1974).
[2] N. C. Shen, C. Y. Li, Y. M. Sun, C. Wang, J.-M. Chartier, and R. Felder, CCDM/82–27 (document submitted to the 1982 meeting of the Consultative Committee for the Definition of the Meter).
[3] H. Greenstein, J. Appl. Phys. 43, 1732 (1972).
[4] R. Lenaour *et al.*, Opt. Quantum Electron. 10, 119 (1978).
[5] G. Kramer *et al.*, Z. Naturforsch. 30a, 1128 (1975).
[6] Shen Nai-cheng, CCDM/82–12.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Spatial Coherence and Optical Wavelength Metrology

P. Bouchareine

Institut National de Métrologie, Conservatoire National des Arts et Métiers, 292 rue Saint Martin F 75141 Paris cedex 03

We discuss the influence on precise measurements of optical wavelengths of spatial coherence of laser beams, or of classical sources when the "étendue" of the interferometer is reduced by isolating the central fringe at high path differences. The interferometric observation of the phase angle at a fixed path difference is currently made with an uncertainty less than 10^{-3} of a fringe. With spatially coherent illumination, phase shifts of this order can be given by scratches or dust particles, or by mirror aberrations, and cannot be compensated by subtracting two phase angles at two path differences. If the spatial coherence of a laser beam is destroyed, these errors vanish, but only if the étendue of the interferometer is large enough. At very high path differences allowed by the good temporal coherence of lasers, compensated field interferometers may be useful to eliminate these spurious shifts and the uncertainty in diffraction corrections.

Key words: diffraction correction; optical wavelength metrology.

1. Introduction

The unit of length is defined in terms of a wavelength. The definition of the meter will probably be soon changed, but for a long time, practical measurements of length up to several meters will be best achieved by optical interferometry.

The wavelength is the spatial periodicity of the phase and of the amplitude of a monochromatic infinite plane wave. If the wave is limited, optical rays are bent by diffraction, and a correction has to be made to the apparent wavelength. The correction can be calculated only in some simple cases, as for instance with a Gaussian beam.

2. Spatial Coherence of Laser Beams

Lasers deliver a beam of high temporal coherence, which is essential for observing interference fringes over path differences of several hundred meters. Lasers also have quite perfect spatial coherence which might be a good condition if one wants to have a plane wave and to look at the true wavelength. Nevertheless, optical wavelength metrology needs interpolation within a fringe at high path differences. Interpolation within 10^{-3} of a fringe is often obtained. With a spatially coherent wave, local phase shifts and errors are produced by any parasitic wave created from the main wave by reflection, diffraction, or diffusion. For example any punctual defect on a mirror or a lens, like a scratch, or a dust particle, gives rise to a spherical wave which interferes with the main wave in almost all space (Fig. 1). At the point where the wave is observed, or where it interferes with the reference wave of the interferometer, this spurious wave brings a phase shift of the main wave, depending on the ratio of the amplitudes A_o and A_p of the two waves, and on the phase difference between them. The maximum phase error is observed when the phase difference is $\pm \pi/2$ (Fig. 2) and the statistical mean of this error is zero. To be sure that this shift is less than 10^{-3} of a fringe, the energy of the parasitic beam must be less than 10^{-6} of the energy of the main beam. Such an error occurs with any spatially coherent wave, plane, spherical, or Gaussian, and to a lesser extent with a partially coherent wave.

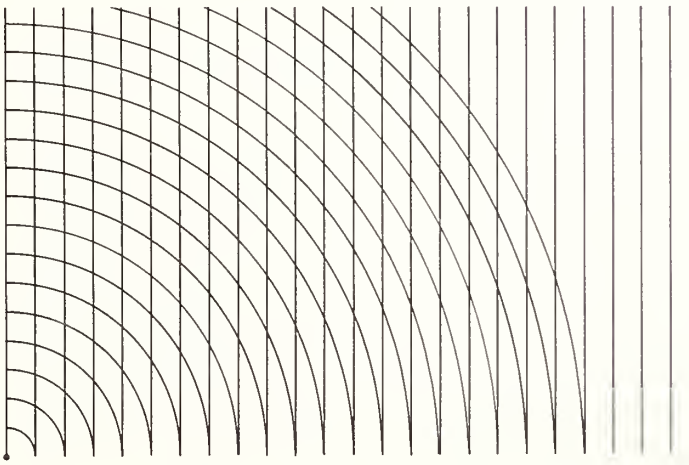


FIGURE 1. *The influence of a local defect of a mirror on a plane wave. The spherical diffracted wave interferes with the main wave in all space, giving phase shifts.*

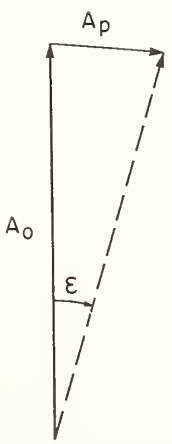


FIGURE 2. *Interference of the main beam of amplitude A_o with the parasitic beam A_p gives a phase shift of the main beam. The statistical mean of this shift is zero, and the maximum value is observed at the points where the waves are in quadrature.*

3. Compensation of Mirror Shape Defects

If the optical surface of the mirror of an interferometer is not perfect (plane, spherical, or otherwise) the path difference between rays reflected by an area $dx\ dy$ around the point with coordinates x and y is different from Δ_0 by an amount $\delta(x, y)$. Δ_0 can be the path difference at the point $x = 0$, $y = 0$, or the mean path difference over the whole area of the mirror, or any arbitrary path difference, so that $\delta(x, y)$ is defined with an arbitrary origin $\delta = 0$.

In interferometric wavelength metrology, wavelengths and path differences are measured with an uncertainty much less than the amplitude of $\delta(x, y)$, because a spatially incoherent illumination of the mirror is assumed, and the intensities of the interference signals of different areas $dx\ dy$ are added. The phase error resulting from $\delta(x, y)$ in the interference signal is compensated by taking the phase difference between two path differences, assuming that the energy distribution on the mirrors is kept constant. Thus, fractional phase excesses can be measured in two-wave interferometry with an uncertainty less than 10^{-3} of a fringe with mirrors whose shape is manufactured only within a tenth of a fringe.

If the mirrors of the interferometer are illuminated by spatially coherent beams, we have to add the amplitudes of different areas $dx\ dy$, and the consequence of the shape error is no longer the same for a given mirror at two different positions, even if the amplitude distribution on the mirror remains constant.

4. Compensated Field Interferometers and Spatially Incoherent Illumination

It is possible to destroy the spatial coherence of a laser beam with one or two moving diffusers, and a sufficiently long time response of the receiver. If such a diffused beam is travelling through a classical interferometer (Michelson or Fabry-Perot), it must be limited at infinity by the hole isolating the central fringe. The solid angle of this hole is $2\pi/p$, where p is the interference order. For high path differences, this hole may give a partial spatial coherence to the illumination of the mirror area, and the interferometer is more or less sensitive to the previously described phase errors.

To avoid such spatial coherence, with a diffused laser beam as well as with a spatially incoherent source, it is necessary to work with a large étendue interferometer, with compensated field, like a spherical Fabry-Perot or a cat's eye Michelson [1]. The gain in étendue is particularly high for path differences of 10^6 fringes and more, and the phase errors due to diffraction can become negligible.

However, a fundamental question remains with these interferometers, in which at least one mirror has to be changed when the path difference is changed. With incoherent illumination, the weighted mean of the path difference is the same for two radiations of different wavelengths only if the error function weighted by the incoherent illumination $I(x, y)$,

$$F(\delta) = \int_{\delta=\text{const}} I(x,y)\, dx\ dy \ ,$$

is an even function of δ, or can be made even by a change in origin of δ.

5. Conclusion

With stabilized lasers, the main uncertainty in wavelength comparisons will probably lie in the diffraction correction for imperfect mirrors and phase errors due to parasitic waves. The large étendue of field compensated interferometers used with spatially incoherent illumination reduces the influence of diffraction. The parasitic waves are also present with incoherent illumination, but the statistical mean of the phase error is zero.

When we built our cat's eye interferometer with a compensated field, it was because of the limited luminosity of the krypton lamp. With lasers this is not a factor, but we hope that our apparatus will be useful in studying those aspects of optical wavelength metrology that were not in sight at the beginning.

Reference

[1] P. Bouchareine, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum, New York, 1976), p. 417.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Spectroscopy of Stored Ions

D. J. Wineland

**Frequency and Time Standards Group, Time and Frequency Division
National Bureau of Standards, Boulder, CO 80303**

The benign environment and long confinement times obtained with ion storage techniques have led to some unique experiments in the area of precision measurements and fundamental constants. This is perhaps epitomized by the single electron g factor measurements at the University of Washington in which a precision of 4 parts in 10^{11} has been attained. Now, use of lasers to cool stored ions has allowed the experimentalist to approach the goal of unperturbed atomic ions nearly at rest; most recently, spectroscopy has been performed on single "cold" trapped ions. Stored ion experiments in the area of precision measurements and fundamental constants will be briefly reviewed. These include experiments on e^-/e^+ g factors, mass spectroscopy, lifetimes, and atomic spectroscopy. The intent is to emphasize the unique environment provided by ion storage techniques for these measurements.

Key words: atomic ion spectroscopy; electron g factor; fundamental constants; ion storage; mass spectroscopy; precision measurements.

1. Introduction

The purpose of this paper is twofold. The first purpose is to outline "trapped ion" experiments of the last decade which have contributed to the area of precision measurements and fundamental constants. The second purpose is to emphasize why the stored ion "environment" is particularly well suited to the area of precision measurements.

The main advantage of the stored ion technique is that the ideal of an unperturbed species at rest in space is approached to a high degree. Specifically, charged particles such as electrons and atomic ions can be stored for long periods of time (essentially indefinitely) without the usual perturbations associated with confinement (for example the perturbations due to collisions with walls or buffer gasses in a traditional optical pumping experiment). These unique properties were exploited most notably by Dehmelt and co-workers [1] before 1970; since then,that group and others have continued to extend these techniques with dramatic results.

Unfortunately, there is a price to be paid for this property of long storage times with small perturbations—the number of particles that can be stored is typically small ($\lesssim 10^6$ for a "trap" with centimeter dimensions); the resulting low densities are governed by the competition between space charge repulsion and the confining electromagnetic forces obtained under normal laboratory conditions. These low numbers, of course, require very sensitive detection techniques and preclude many types of experiments—for example spectroscopic experiments on complex molecular ions where only a small fraction of the ions is in a given state. Although some effort has been made to increase the number of ions by space charge neutralization [2, 3] and trap "arrays" [4], these methods introduce other experimental complications. In spite of the low numbers obtained, sensitive techniques have been developed to detect simple species such as electrons and atomic ions so that *single* electrons [5, 6] and ions [7, 8] have been observed.

Since the intent is to discuss experiments relating to precision measurements and fundamental constants, many interesting experiments using stored ion techniques are not discussed in this paper. The reader is referred elsewhere to experiments on, for example, photodetachment [9, 10], chemical reactions [11], electron-ion recombination [12], charge transfer [13, 14], and nonneutral plasma studies [15]. Also, more general reviews are available [1, 16, 17].

In Sec. 2, methods of trapping are briefly reviewed. In Secs. 3–6, experiments on electrons/positrons, mass spectroscopy, atomic spectroscopy, and lifetime measurements are described with emphasis on the ion's environment. In the last section, speculations about other future experiments are made.

2. Ion Trapping Methods

Four types of traps have been useful: the rf (or Paul) trap, the Penning trap, and electrostatic and magnetostatic traps. These traps and their relevant properties are briefly described here. More detailed descriptions are given elsewhere [1, 16].

2.1 The rf Quadrupole Trap

The ideal rf or Paul [18] trap uses hyperbolic electrodes in a vacuum apparatus as shown in Fig. 1. These electrodes are symmetric about the z axis, so we can describe the potential in cylindrical coordinates. If an alternating voltage of frequency Ω is applied between the endcaps and ring electrode, then the instantaneous potential inside the trap is given by (following the notation of Ref. [1]):

$$\phi(r,z) = A(r^2 - 2z^2), \qquad A = A_0 \cos \Omega t. \qquad (1)$$

An ion experiences an rf electric field such that its motion (the "micromotion") is 180° out of phase with respect to the electric force. Because the electric field is inhomogeneous, the force averaged over one period ($T = 2\pi/\Omega$) of the micromotion is in a direction of weaker field amplitude (independent of the sign of the charge), i.e.,

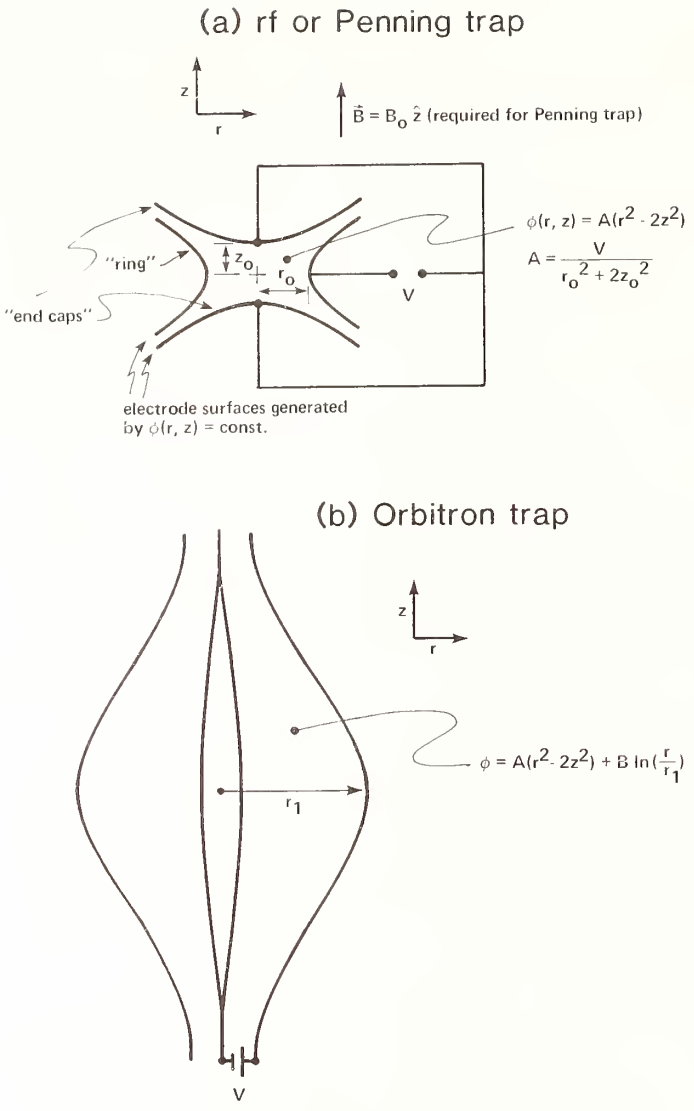


FIGURE 1. *Electrode configuration for rf, Penning, and orbitron ion traps. All electrode surfaces are figures of revolution about the axial (z) direction. In a practical orbitron trap, it is desirable to make the center electrode as thin as possible to reduce ion loss from collisions.*

towards the center of the trap. For Ω sufficiently high, this restoring force gives rise to a pseudopotential

$$\Psi(r,z) = \frac{eA_0^2}{M\Omega^2}\,[(\bar{r})^2 + 4(\bar{z})^2]$$

where $\bar{r}$ and $\bar{z}$ are the positions of the ion averaged over T, e is the ion charge, and M is the ion mass. The resulting "secular" motion is

$$\begin{aligned} \bar{z} &= \bar{z}_0 \cos(\bar{\omega}_z t + \phi_z) \\ \bar{x} &= \bar{x}_0 \cos(\bar{\omega}_r t + \phi_x) \qquad \bar{\omega}_z = 2\bar{\omega}_r = \frac{2\sqrt{2e}A_0}{M\Omega}. \\ \bar{y} &= \bar{y}_0 \cos(\bar{\omega}_r t + \phi_y) \end{aligned}$$

In more detail, we have for the z motion:

$$z \simeq \bar{z}_0 \left[1 + \frac{\sqrt{2}\,\bar{\omega}_z}{\Omega} \cos \Omega t\right] \cos \bar{\omega}_z t,$$

where we require $\Omega/\bar{\omega}_z \gg 1$.

For ease of comparison with the Penning trap, it will be useful to consider the case of a spherical pseudopotential well in the rf trap. This is accomplished by simultaneously applying a static potential between the ring and endcaps so that

$$\phi = (U_0 + A_0 \cos \Omega t)(r^2 - 2z^2).$$

If we choose

$$U_0 = \frac{e}{M\Omega^2} A_0^2, \quad \text{then}$$

$$\Psi(r,\ z) = \frac{2eA_0^2}{M\Omega^2} \left[(\bar{r})^2 + (\bar{z})^2\right]. \tag{2}$$

In this case:

$$z = \bar{z}_0 \left[1 + \frac{2\bar{\omega}_z}{\Omega} \cos \Omega t\right] \cos \bar{\omega}_z t$$

$$\begin{Bmatrix} \bar{x} \\ \bar{y} \end{Bmatrix} = \begin{Bmatrix} \bar{x}_0 \\ \bar{y}_0 \end{Bmatrix} \left[1 + \frac{\bar{\omega}_z}{\Omega} \cos \Omega t\right] \cos \bar{\omega}_r t\ , \tag{3}$$

where $$\bar{\omega}_z = \bar{\omega}_r = \frac{2eA_0}{M\Omega}.$$

Approximate experimental values for stored atomic ions are such that $r_0 \simeq 1$ cm, A_0 is a few hundred volts/cm^2, $\Omega/2\pi$ is a few MHz, and $\Omega/\bar{\omega}_z \simeq 10$.

If the secular motion is "cold" then we expect the maximum density of ions to be given when the force from the space charge potential ϕ_i cancels that due to the trap. In this case, for the pseudo-potential given by Eq. (2), the ion cloud is a uniformly charged sphere of charge density ρ with internal potential

$$\phi_i = (2/3)\ \pi\rho\ (r^2 + z^2).$$

Therefore for a well with $2eA_0^2/M\Omega^2 = 10$ V/cm^2, the condition $\phi_i = -\Psi(r,\ z)$ gives a density $n = 3.3 \times 10^7$/cm^3. In practice, attainable densities are approximately 1 to 2 orders of magnitude lower due to "rf heating"—a process which couples kinetic energy from the micromotion into the secular motion. One way this happens is through collisions with background gas. However, at high vacuum other coupling mechanisms can occur partly due to the presence of impurity ions or imperfections in the trap electrodes. This process can give kinetic energies of ions of as much as electron volts and shortened storage times of less than 1 s. If one uses a light buffer gas such as He to provide viscous damping [1] then storage times of many days can be achieved [19, 20]. However, this increases perturbations due to collisions; therefore, for experiments on large numbers of ions in an rf trap, there appears to be a practical tradeoff between the temperature of the ions and possible perturbations due to collisions with a cooling buffer gas. However, this is not a fundamental limitation and one hopes it will be solved in the future.

We note that in principle, densities could be increased by increasing the rf field strengths; in practice, higher densities (close to 10^9/cm^3 [21]) have been achieved by decreasing the electrode dimensions. This, of course, may not increase the ion number and it also may increase the rf heating because the field imperfections may be relatively larger.

It is interesting to note that the methods proposed for trapping of neutral particles in laser beams [22] are very analogous to the rf trapping of charged particles. In the laser beam case, the outer atomic electron experiences net forces in the (inhomogeneous) laser field. But since the electron is bound to the nucleus, dispersive effects

occur. If the laser is tuned above the frequency of an electronic transition, similar to an ion in a rf trap, the atom is forced to regions of low field intensity. If the laser is tuned below resonance, the atom is forced to regions of high field intensity.

2.2 The Penning Trap

The Penning trap [23] uses the same electrode configuration as the rf trap (Fig. 1) but now A in Eq. (1) is a constant U_0 such that the charged species see a static potential well along the z axis. We have $\phi_T = U_0(r^2 - 2z^2)$. This causes a repulsive potential in the $x-y$ plane which can be overcome by superimposing a static magnetic field along z ($\boldsymbol{B} = B_0\hat{z}$). For a single ion in the trap, the equations of motion are [1, 24]

$$\ddot{z} + \omega_z^2 z = 0,$$

$$\omega_z^2 = \frac{4eU_0}{M} = \frac{4eV_0}{M(r_0^2 + 2z_0^2)},$$

$$\ddot{\mathbf{r}} = \frac{1}{2}\omega_z^2 \mathbf{r} - i\omega_c \dot{\mathbf{r}},$$

$$\omega_c = \frac{eB_0}{Mc},$$

where

$$\mathbf{r} = x + iy.$$

Therefore

$$z = z_0 \cos \omega_z t \text{ and } \mathbf{r} = \mathbf{r}_c e^{-i\omega_c' t} + \mathbf{r}_m e^{-i\omega_m t} \tag{4}$$

where

$$\begin{Bmatrix} \omega_c' \\ \omega_m \end{Bmatrix} = \frac{1}{2}\omega_c \begin{Bmatrix} + \\ - \end{Bmatrix} [(\omega_c/2)^2 - \omega_z^2/2]^{1/2}. \tag{5}$$

Some useful expressions are:

$$\omega^2 + \omega_z^2/2 = \omega_c \omega \qquad (\omega = \omega_m \text{ or } \omega_c'), \tag{6a}$$

$$\omega_c' + \omega_m = \omega_c, \tag{6b}$$

$$\omega_c' \omega_m = \omega_z^2/2. \tag{6c}$$

Typical operating conditions in a Penning trap are such that for electrode spacings about 1 cm, V_0 is a few volts and B_0 is larger than a few tenths of a tesla. The required large magnetic field can either be an advantage or disadvantage, depending on the experiment. It should be noted that the magnetron motion (the $\mathbf{r}_m e^{-i\omega_m t}$ term) is in an unstable equilibrium in the trap. For example, if collisions with background neutrals occur, the ions will diffuse out of the trap. (When the magnetron velocity is much less than the velocity of the neutrals, $\mathbf{r}_m$ random walks in the x–y plane with step size $\sim r_c$.) This is a potential disadvantage when compared to the rf trap for which all three degrees of freedom (x, y, z) are in stable equilibrium. In practice however, this is not a limitation because ions can be stored for days in a room temperature apparatus [25] and electrons for weeks [5] in an 80 K apparatus. Moreover, the technique of sideband or radiation pressure cooling [6, 21, 25, 26] can reverse this diffusion process.

If the voltage V_0 applied to the electrodes becomes too high, then the radial electric field is high enough to overcome the $e\mathbf{v} \times \mathbf{B}/c$ magnetic force and the ions strike the ring electrode in exponentially increasing orbits (argument of square root in Eq. (5) becomes negative). For singly ionized atoms, the voltage where this occurs is given by $V_c \simeq 1200\ B_0^2(r_0^2 + 2z_0^2)/M$ where V_c is in volts, B_0 in tesla, M in u (atomic mass units) and dimensions in cm. This same mechanism limits the densities achievable in the Penning trap since space charge also gives radial electric fields. If the axial and cyclotron temperatures are low enough, then we can approximate the equilibrium distribution of the ion cloud [27] by a uniformly charged ellipsoid which has an (internal) potential of:

$$\phi_i(r) = (2/3)\ \pi\rho\ (ar^2 + bz^2)$$

where a and b are constants. The $\mathbf{r}$ motion of an individual ion is now given by Eq. (4) with

$$\begin{Bmatrix} \omega_c' \\ \omega_m \end{Bmatrix} = \frac{\omega_c}{2} \begin{Bmatrix} + \\ - \end{Bmatrix} \left[\left[\frac{\omega_c}{2} \right]^2 - \frac{\omega_z^2}{2} - \frac{4\pi e\rho a}{3M} \right]^{1/2} \tag{7}$$

where ω_z is the axial frequency for a single ion in the trap. For a spherical distribution, $a = b = 1$; and the maximum density allowable (argument of square root term kept positive) is given by (using $\phi_i(z) = -\phi_T(z)$):

$$n = \frac{\rho}{e} < \frac{B_0^2}{8\pi c^2 M}$$

or

$$n < 2.7 \times 10^9 B_0^2/M$$

where B_0 is in tesla and M in u. For $B_0 = 1$ T, $M = 100$ u, $n < 2.7 \times 10^7/\text{cm}^3$ which is similar to the rf trap case. In experiments at the National Bureau of Standards (NBS), densities within a factor of about 3 of the theoretical maximum have been observed as indicated by the space charge shifted magnetron rotation frequency.

2.3 Electrostatic Traps

The orbitron trap [28] has the advantage of simplicity and although lifetimes are observed to be relatively short, precision lifetime and spectroscopic measurements have been made by Prior and his colleagues at Berkeley [29, 30].

The orbitron trap is purely electrostatic and has axial symmetry as shown in Fig. 1b. A potential which is useful is given by the general form [31]:

$$\phi = A(r^2 - 2z^2) + B \ln(r/r_1),$$

because a harmonic well is provided along the z axis. The equations of motion for the $x-y$ plane are not solvable analytically, but the motion is basically composed of precessing orbits about the z axis [32]. Storage times in such a trap should be shorter because a single collision with a background neutral is sufficient to cause an ion to collide with the center electrode and be lost. Also the cooling discussed in Sec. 6 would be more difficult to realize. Nevertheless, with a trap whose electrodes approximate equipotentials of this function, trapping times of about one second were achieved by Knight [31] at a pressure of 10^{-6} Pa ($\simeq 10^{-8}$ Torr). (The center electrode was a 100 μm dia. wire.)

An axial magnetic field superimposed along the z axis has been investigated theoretically [33]; in this case, storage times should increase dramatically. Ions will eventually be lost by diffusing in towards the center electrode; therefore, storage times should be comparable to that of the Penning trap. However, the frequency of the drift motion about the center of the trap will depend on the distance from the z axis; for the Penning trap this drift (magnetron) frequency is independent of radial posi-

tion (neglecting space charge) and may be part of the reason for the slow evolution of the ion cloud under the influence of the ion-ion collisions. A possible advantage over the Penning trap is that experiments may be possible at much lower magnetic fields [33].

2.4 Magnetostatic Traps

Magnetic bottles are of course used extensively for plasma confinement and also in ion spectroscopy. The disadvantage in terms of spectroscopy appears to be that rather large magnetic field inhomogeneities are required for trapping, thus causing inhomogeneities and broadening in paramagnetic ion line shapes. The main application in terms of precision measurements has been the Michigan e^+/e^- experiments [34]; these are reviewed by A. Rich at this conference.

3. e^-/e^+ $g-2$ Measurements in a Penning Trap

This experiment has been pursued primarily by the groups of Dehmelt at the University of Washington, and Gräff at Mainz [35–37]. Only the Washington single electron experiment [6] will be discussed here, since its accuracy significantly exceeds that of the experiments on clouds of electrons [35–38]. Experiments on clouds of electrons are susceptible to rather large and difficult to measure space charge shifts. For example, the fractional shift of the cyclotron frequency of one electron due to another electron a distance r away (in a direction perpendicular to the magnetic field) is approximately given by [39]

$$\frac{\Delta\omega}{\omega_0} = \frac{1}{2}\,\frac{e}{m\,\omega_c^2}\,\left[\frac{1}{r}\,\frac{\partial}{\partial r}\,(rE_r)\right]$$

where m is the electron mass, ω_c is the cyclotron frequency, and E_r is the electric field at one electron from the other. For a magnetic field of 1 T, and $r = 100$ μm, $\Delta\omega/\omega_0 = 4.1 \times 10^{-9}$ which gives a shift of about 4 ppm in $g-2$. In 1971, the Michigan group had measured the electron $g-2$ to 3 ppm [40]; this result and the problems associated with the space charge shifts led the Washington group to develop an experiment based on single electrons. A complete account of this experiment is given in Refs. [6, 41]. A simpler account is given in Ref. [42].

Briefly, a single electron can be detected [5, 42, 43] by monitoring the currents induced in the electrodes by the axial (z) oscillation at frequency ω_z. The harmonically bound electron is equivalent to a series L-C circuit which shunts the electrodes [5]; this L-C oscillator can be phase locked to an external oscillator by feeding back on the trap voltage V_0 [6]. It is useful to think of the single electron bound to the Penning trap as a one electron pseudo atom whose "nucleus" is the earth (to which the Penning trap is fixed), thus the name "geonium" [6]. The quantum mechanical solution for the energy levels of this "atom" are given by:

$$E/h = m\nu_s + (n + 1/2)\nu_c' + (k + 1/2)\nu_z - (q + 1/2)\nu_m$$

where m is the spin quantum number ($m = \pm 1/2$) and n, k, and q are the cyclotron, axial and magnetron quantum numbers (n, k, $q = 0, 1, 2 \ldots$). (The magnetron energy is mostly potential energy and is negative.)

In the $g-2$ and e^+/e^- mass ratio experiments it is necessary to detect the occurrence of spin flip and cyclotron transitions. This can be accomplished by superimposing a weak magnetic bottle centered on the trap such that the field is [6]:

$$B = (B_0 + \beta(z^2 - r^2/2)\hat{z} - \beta zy\hat{y} - \beta zx\hat{x}\,.$$

The spin, cyclotron, and magnetron magnetic moments interact with this magnetic bottle to shift the axial frequency by an amount

$$\delta\nu_z = [m + n + 1/2 + (\nu_m/\nu_c)q\,]\delta$$
$$\delta \simeq \mu_B\,\beta/(2\pi^2 m\,\nu_z)$$

For $\nu_z = 60$ MHz and $\beta = 0.012$ T/cm^2, $\delta \simeq 1$ Hz, therefore for example, spin flip transitions would show up as 1 Hz changes in ν_z or as an accompanying step increase in feedback voltage to the locked axial resonance [6].

For reproducibility, it was desirable to have the electron at the center of the trap. This could be accomplished by sideband excitation [6, 26, 44, 45] as follows: with the (damped) axial motion driven by an inhomogeneous rf field of frequency $\nu_z + \nu_m$, the electron sees a sideband at ν_z because of the magnetron motion through the rf field. Therefore the electron absorbs quanta of energy $h(\nu_z + \nu_m)$ and reradiates (by the damping) quanta of energy $h\nu_z$. The deficit, $h\nu_m$, goes into the magnetron energy which shrinks the magnetron orbit. In principle, this technique allows for infinite storage times.

The electron $g - 2$ has now been measured with an accuracy of 0.04 ppm [46] and the positron $g - 2$ to 0.05 ppm [47]. The details of these remarkable experiments are described in Ref. [6]; rather than give a complete description of the experiments, only the experimental environment is described here. From Ref. [6], the cyclotron degree of freedom was radiatively thermalized to 4 K. The axial temperature was somewhat higher. Pressure was estimated to be as low as 10^{-12} Pa so that collisions with background gas could be neglected. At temperatures near 4 K, axial excursions of 0.1 mm and cyclotron orbit sizes of ~60 nm would be obtained. During cyclotron excitation to $n \simeq 15$, this gave a corresponding relativistic shift of only -6×10^{-9}. Using sideband excitation, $r_m \simeq 14$ μm. From Eq. (6b), there is, of course a large systematic shift (~6×10^{-4} in $g - 2$) between the observed and free space cyclotron frequencies. Due to imperfections in the trap, a direct measurement of ν_m (by magnetron excitation) and the calculated value from Eq. (6c) can disagree slightly. However, it can be shown that the error in ν_m is much less [6, 48] than the above difference which was measured to be $10^{-4}\,\nu_m$. Thus the error in $g - 2$ was negligible. Errors in the present experiments are estimated from field jitter and drift and dependence of $g - 2$ on applied spin flip power [47]. Ultimately, the experiment will be limited by systematic shifts associated with the superimposed magnetic bottle. To eliminate this problem, new ways have been proposed to detect the spin flips [6, 49]. For example, it might be possible to use a switchable magnetic bottle so that while the cyclotron and $g - 2$ resonances are driven, the bottle would be off and transitions would be detected by turning the bottle on [50]. Even without a magnetic bottle, the axial motion is still linked to the cyclotron degees of freedom via the relativistic mass shift [6, 51]. This coupling is about an order of magnitude smaller than for the magnetic bottle; the required increase in axial well depth stability might be provided with a reference electron in an adjacent trap whose axial frequency is locked to a particular value [52].

4. Mass Spectroscopy

With the long storage times and high vacuums in the traps, one could expect extremely narrow motional resonances. In the Washington electron experiments cyclotron excitation resonances reproducible to a few parts in 10^9 have been observed [53], and in principle the resolution should only be limited by effects due to inhomogeneous or time varying fields. Using the methods described in Sec. 3, the electron/positron mass ratio has recently been measured to about 0.1 ppm [63]. This very important experiment represents an improvement by more than two orders of magnitude over previous experiments. In addition, the cyclotron resonances of both atomic ions *and* electrons can be measured in the same magnetic field and in the same apparatus. This rather unique feature has led to experiments designed to make a direct measurement of the electron/proton mass ratio in a Penning trap.

The three current experiments using this method [53–55] all basically measure the (shifted) cyclotron frequencies of electrons and protons in the same magnetic field. In the experiment of Gärtner and Klempt [54], resonances were detected by electron/proton loss; the largest uncertainty was due to a space charge shift in the proton cyclotron resonance. In the Van Dyck and Schwinberg experiment, resonances were detected in several (non-destructive) ways [53]; the limit here was the uncertainty in the relative positions of the electrons and protons in the trap and also possibly ion number (space charge) dependent shifts. In the experiment of Gräff, Kalinowsky, and Traut, resonances were detected by an energy dependent time of flight technique [55]. The experiment of Van Dyck and Schwinberg has the smallest uncertainty (0.14 ppm) and substantial improvements can be expected.

With regard to space charge shifts, it is worth noting that in an ideal trap described by a quadratic electric potential and uniform magnetic field, if only one species is trapped and if the cyclotron motion is excited by a uniform electric field, then *no* space charge shifts can occur because only the center of mass mode is excited [26, 56]. Unfortunately, deviations from the ideal trap fields can occur, impurity ions can be present, and the excitation usually occurs from non-uniform fields. All of these things can lead to (partial) excitation of the internal modes which then results in space charge shifts. These systematic shifts increase with the mass of the ion and therefore are much more important for protons and atomic ions than for electrons.

After the initial single electron work, it became apparent that it should be possible to observe single ion cyclotron resonances by observing the induced currents in the ring electrode if it is divided in half [57] or into quadrants [53]; the key advantage here is that space charge shifts are totally absent. Since the cyclotron frequency of protons in a field of 4 T is about 60 MHz, electronic detection should be similar to the detection of the axial resonance of single electrons.

Other, related mass spectroscopy experiments should be possible. First, using techniques as those referred to above, mass ratios of various atomic (molecular) ions could be measured. It should also be possible to observe the cyclotron resonance of single atomic ions by laser techniques. For example, if a single cold Be^+ ion is localized at the center of the trap [8] with a laser beam whose diameter is approximately 3 μm, then it should be possible to observe cyclotron excitation which increases the orbit radius to about 3 μm by the decrease in fluorescence scattering. This corresponds to a temperature ($k_B T = (1/2) M (\omega_c')^2 r_c^2$) of 14 K in a field of 5 T and a relativistic shift of 1.4×10^{-13}. The axial and magnetron frequencies could be measured in a similar fashion. If the electron cyclotron frequency were measured in the same magnetic field, the accuracy in M_e/M_p from this indirect experiment would be limited by the accuracy of M_{Be}/M_p which is about 0.04 ppm [58].

In preliminary experiments at NBS, the axial resonance ($\nu_z \simeq 200$ kHz) of a small cloud (<10) of $^{25}Mg^+$ ions has been measured in this way; however significant broadening (a few parts in 10^4) is observed—presumably due to impurity ions. For an initial experiment on Be^+, the electron cyclotron frequency can be measured indirectly by measuring the spin flip frequency on the same ion; since g_J (Be^+) may be calculable to as good as a few parts in 10^8 [59], the error in M_e/M_p could still be limited by the error in M_{Be}/M_p. Ultimately when M_e/M_p is known better by other means, such measurements should lead to better experimental determinations of g_J (Be^+) and M_{Be}/M_p.

We conclude by remarking that the extreme confinement possible with the stored ion techniques considerably relaxes the requirements on magnetic field homogeneity in these experiments.

5. Radiative Lifetime Measurements

Although perhaps not "high precision" in an absolute sense, the radiative lifetime measurements of Prior and colleagues at Berkeley should be included in this review because these measurements, made on simple atomic ions, can be compared with various theories. Lifetimes of 2^1S_0 Li^+ ($\tau \simeq 503$ μs) in a Penning trap [60], 2s He^+ ($\tau \simeq 1.92$ ms) in an orbitron [29], and 2^3S_1 Li^+ ($\tau \simeq 58.6$ s) in an rf trap [61] have been measured. One should also include the $5^2D_{3/2}$ Ba^+ ($\tau \simeq 17.5$ s) measurement in an rf trap by Schneider and Werth [62] and the $5^2D_{5/2}$ Ba^+ ($\tau \simeq 47$ s) measurement in an rf trap by Plumelle *et al.* [19]. The key advantage of the traps in these measurements is the very long storage times and the benign environment which minimizes the effects of other relaxations. For the very long lifetimes, perhaps no other method is available.

6. Atomic Spectroscopy

As discussed in Sec. 1, the small number of ions involved makes it desirable to study simple atomic ions, but this does not preclude the study of simple molecular ions [1, 64]. As we will see, the potential accuracy and precision for atomic spectroscopy using ion storage techniques is quite remarkable. This was apparent from the early experiments on the $^3He^+$ ground state hyperfine structure [1, 65] where linewidths of about 10 Hz at 8.7 GHz were obtained. Since that time the 2s $^3He^+$ hyperfine structure (hfs) has been measured in an orbitron trap by Prior and Wang [66]. The importance of this measurement is that a comparison between the 1s and 2s hfs suppresses the effects of nuclear structure corrections and allows an accurate check of the QED corrections; such experimental/theoretical comparisons are most meaningful for hydrogenic ions.

A significant interest in stored atomic ion spectroscopy stems from the desire to make a frequency standard. Stimulated by the He^+ experiments, various proposals [65, 67–78] have been made for microwave and optical stored ion frequency standards. For a microwave stand-

ard, considerable attention has been given to the $^{199}Hg^+$ ion because (1) its ground state hyperfine frequency (~40 GHz) may be the largest of any conveniently usable ion, (2) its large mass gives a relatively small second order Doppler shift at a given temperature (~2×10^{-13} at 300 K), and (3) a $^{202}Hg^+$ lamp source (194 nm) is available for optical pumping. Starting with the work of Major and Werth [69], groups at Mainz [74], Orsay [76], and at least one commercial company [74] have sought to develop a frequency standard based on $^{199}Hg^+$ ions stored in an rf trap. This work has been developed to a fairly high level; the group at Orsay [76] has made a working standard whose stability compares favorably with that of a commercial cesium beam frequency standard.

However, the full potential of the stored ion technique has yet to be realized. Historically, it appears that this has been due to two problems. (1) Because it has been difficult to cool the ions, second order Doppler or time dilation shifts can be sizeable (e.g.: ~10^{-11} in Ref. [74], ~4×10^{-10} in Ref. [65]). (2) Signal-to-noise ratio has been poor due to the small number of stored ions.

In the past two or three years, both of these problems have been addressed. In 1978, groups at NBS and Heidelberg demonstrated [21, 25] that radiation pressure from lasers [44, 70, 79] could be used to cool ions to temperatures <0.1 K, thereby reducing the second-order Doppler shift by more than three orders of magnitude below the room temperature case. As discussed below, the reduction of kinetic energy is most favorable for very small numbers of ions (down to one ion), so that there is a trade-off between the maximum number of ions we can use and the minimum second-order Doppler shift that can be achieved.

With regard to signal-to-noise ratio, in certain optical-pumping, double-resonance experiments, it is possible to scatter many optical photons from each ion for each microwave or optical "clock" photon absorbed [70, 73, 79–81]. This can allow one to make up for losses in detection efficiency due to small solid angle, small quantum efficiency in the photon detector, etc., so that the transition probability for each ion can be measured with *unity* detection efficiency. This means that the signal-to-noise ratio need be limited only by the statistical fluctuations in the number of ions that have made the transition [78, 81]. Therefore the signal-to-noise ratio can be maximized for the given number of ions. The ability to sensitively detect the fluorescence from ions is emphasized by experiments on single ions [7, 8] where it is even possible to make photographs as shown in Fig. 2.

More recently, the narrow linewidths anticipated for the stored ion technique have been observed; a resonance linewidth of about 0.012 Hz at 292 MHz has been observed [82] for the $(m_I, m_J) = (-3/2, 1/2) \longleftrightarrow (-1/2, 1/2)$ hyperfine transition of $^{25}Mg^+$ at a magnetic field of about 1.24 T where the first derivative of the transition frequency with respect to magnetic field is zero. (The Ramsey interference method was implemented by applying two rf pulses of 1 s duration separated by 41 s.) These narrow linewidths should be preserved with hyperfine transitions of higher frequency, such as in Hg^+, but, of course, more attention must be paid to field homogeneity and stability. In any case, the above results allow one to contemplate microwave frequency standards with $Q > 10^{12}$ and inaccuracies $<10^{-15}$ [78]. Of course, in addition to the interest in frequency standards, very high precision hyperfine and optical spectroscopy on Hg^+ [74], Ba^+ [20, 82], Mg^+ [25, 82, 83], and other alkali-like ions [75] is possible.

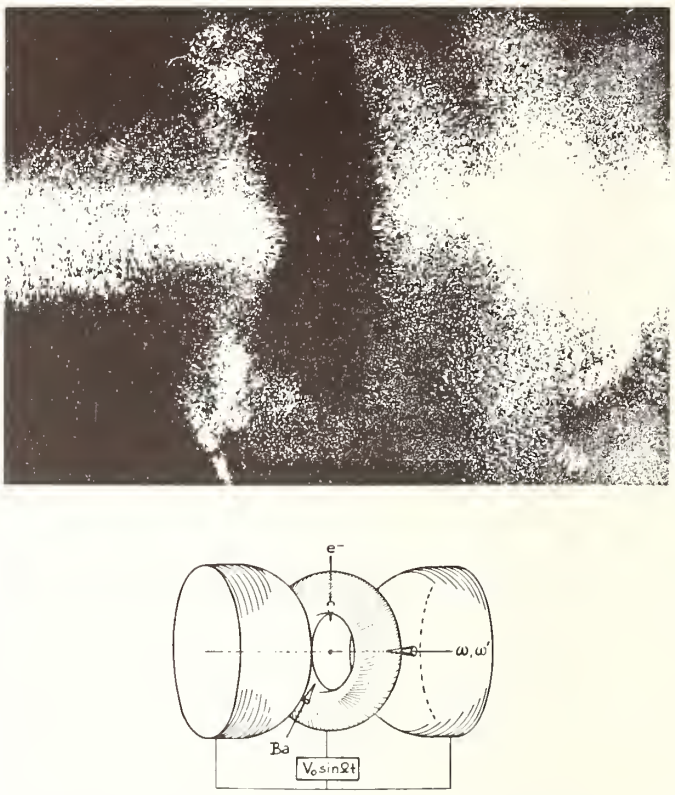


FIGURE 2. *Photograph of fluorescence from single Ba^+ ion in an rf trap (T ≈ 10-36 mK; from Ref. [86]). The schematic diagram of the rf trap from same view. Inner diameter of ring is 0.7 mm.*

The possibilities for an optical frequency standard are even more dramatic [70, 71, 77, 78, 84, 85]. For a given interaction time, the Q of a transition will scale with the frequency. Therefore, in principle, an optical frequency standard would have clear advantages over a microwave frequency standard. The increased Q would allow one to work with smaller numbers, even down to one ion [70, 77, 78], and therefore obtain the "ultimate" atomic frequency standard. Example optical frequency standard candidates are the $^3P_0 \leftarrow {}^1S_0$ transition in Tl^+, Al^+ or Ga^+ [70], the $^3P \leftarrow {}^1S$ transition in B^+ [85], the $D \leftarrow P \leftarrow S$ Raman transitions in Ba^+ [77], the S → D two photon transition in Hg^+ [71, 78], and the $^2P_{3/2} \leftarrow {}^2P_{1/2}$ transition in Pb^+ [72]. All have $Q \geq 10^{14}$. Unfortunately, before the full potential of an ion-storage optical frequency standard can be realized, tunable lasers with suitable spectral purity must become available. This problem may be nearing solution [84]. Also, if an optical frequency standard is to provide time, the phase of the radiation must be measured. This appears to present a much more formidable problem [84, 85]. Therefore it is expedient to continue to pursue a microwave frequency standard where both of these problems have been solved.

Rather than concentrate on the details of the various experiments it is useful to examine the environment of the stored ions in order to identify possible systematic effects. We will be primarily concerned with Stark and second order Doppler shifts. Usually only second order Stark shifts will be important; therefore we will be interested only in $\langle E^2 \rangle$. As an example, for the strongly polarizable cesium (neutral) atom, the ground state hyperfine shift [88] is equal to: $\delta\nu/\nu = -2.5 \times 10^{-16} E^2$ (E in volts/cm); therefore electric fields of volts/cm become

important in very high accuracy experiments [90]. Magnetic field shifts due to instabilities and inhomogeneities are of course important in the Penning trap but can be made negligible in many experiments. (See for example Refs. [78, 82].) Other effects such as electric quadrupole frequency shifts could be important in certain experiments [70, 78].

For single ions, laser cooling has already achieved temperatures between 10 mK and 100 mK [7, 8]. Theoretically, when the motional oscillation frequencies Ω_i ($\bar{\omega}_z$ and $\bar{\omega}_r$ for the rf trap and ω_z, ω_c', and ω_m for the Penning trap) are less than the natural linewidth ($\Delta\nu = \gamma/2\pi$) of the optical cooling transition, then the limiting "temperature" (governed by photon recoil) in each degree of freedom is given by $k_B T \simeq \hbar\gamma/2$ [21, 25, 44, 87]. (For a single ion, the precise minimum temperature depends on the angle of incidence of the laser beam and on the spatial distribution of recoil photons [87].) For strongly allowed transitions as in Ba^+ or Mg^+, this limiting temperature is about 1 mK. For more weakly allowed transitions the temperature is correspondingly less but other limits such as thermodynamic limits can come into play [44]. When the condition $\Omega_i \gg \gamma$ is fulfilled then the limiting energy [44, 86] is given by $E_i = \hbar\Omega_i(\langle n_i\rangle + 1/2)$ where $\langle n_i\rangle \simeq 5\gamma^2/(16\Omega_i^2)$. Therefore the limiting kinetic energy is given by $E_{ki} \simeq \hbar\Omega_i/4$ but the uncertainty in energy (and therefore second order Doppler effect) is much less [44]. For simplicity we will assume only the case $\Omega_i \ll \gamma$ below, however even better results are potentially obtained for the opposite condition. For a single ion in an rf trap, when $U_0 = 0$ the nonthermal micromotion has an average kinetic energy equal to that of the secular motion [1]; this is also true in the spherical trap. In the Penning trap the kinetic energy in the nonthermal magnetron motion can be much less than in the cyclotron or axial modes. Therefore, the minimum second order Doppler shifts are given approximately by:

$$\frac{\Delta\nu_D}{\nu_0} = \frac{E_k}{Mc^2} = \begin{cases} \frac{3}{2}\,\hbar\gamma/Mc^2 & \text{rf trap} \\ \frac{3}{4}\,\hbar\gamma/Mc^2 & \text{Penning trap} \end{cases}$$

For a single ion in an rf trap (assuming the spherical trap of Eqs. (2) and (3)), $\langle E^2\rangle$ is primarily due to the oscillating rf fields and is largest for the z motion. A simple calculation gives $\langle E^2\rangle_z = M\Omega^2\hbar\gamma/e^2$ for maximum laser cooling *or* $\langle E^2\rangle_z = 2M\Omega^2 k_B T/e^2$ for a given temperature in the z secular motion. For a single ion in a Penning trap, it is usually possible to make r_m, $r_c \ll z$ [8, 87], therefore Stark shifts from the static fields are primarily due to the z motion. We find $\langle E^2\rangle_z = \hbar\gamma M\omega_z^2/(2e^2)$ for maximum laser cooling or $\langle E^2\rangle_z = k_B T M\omega_z^2/e^2$ at temperature T. In the Penning trap, a larger effect can be caused by the motional electric field $\mathbf{E} = \mathbf{v}\times\mathbf{B}/c$. We have $\langle E^2\rangle_M = \hbar\gamma B_0^2/(Mc^2)$ (maximum laser cooling) and $\langle E^2\rangle_M = 2k_B T B_0^2/(Mc^2)$. In Table 1 are shown examples of the second order Doppler shift and $\langle E^2\rangle$ for single ions in rf and Penning traps.

For clouds of identical ions, we first consider the electric fields due to collisions between ions. For the rf trap, we neglect the energy in the micromotion since the ions are driven in phase, therefore ion collisional effects in the rf and Penning traps are treated the same. $\langle E^2\rangle$ will of course depend on the cloud density and temperature but some idea of the magnitude can be given by calculating the electric field for one ion on another at the distance of closest approach. Assuming the maximum energy available for closest approach is given by $3k_B T$, we have $E_{\max} = 6.7\times10^{-8}$ V/cm ($\gamma/2\pi$ = 10 MHz and max-

TABLE 1. *Fractional second order Doppler shifts, Stark fields, and classical rms axial amplitudes (z_{rms}) for single ions in rf and Penning traps. When $\gamma/2\pi$ is given, we assume maximum theoretical laser cooling ($\Omega_i \ll \gamma$). For both traps we assume M = 100 u. For the rf trap $\Omega/2\pi$ (rf drive frequency) = 1 MHz, A_0 = 300 V/cm². For the Penning trap, $\omega_z/2\pi$ = 20 kHz, B = 1 T. T is the temperature of the secular motion for the rf trap and the temperature of the cyclotron and axial motion for the Penning trap. $\langle E^2\rangle_z$ is the mean square electric field for motion along the z axis, $\langle E^2\rangle_M$ is the mean square "motional" electric field for the* $\mathbf{v}\times\mathbf{B}/c$ *force. Note that z_{rms} for the Penning trap can be reduced at the expense of increasing $\langle E^2\rangle_z$.*

	$\gamma/2\pi$	——	——	10 MHz	1 kHz
rf	T(K)	300	4	2.4×10^{-4}	2.4×10^{-8}
	$\Delta\nu_D/\nu_\sigma$	8.3×10^{-13}	1.1×10^{-14}	6.6×10^{-19}	6.6×10^{-23}
	$\langle E^2\rangle_z$ (V²/cm²)	210	2.8	1.7×10^{-4}	1.7×10^{-8}
	$\bar{z}_{\rm rms}$(μm)	170	20	0.15	1.5×10^{-3}
Pen.	$\Delta\nu_D/\nu_0$	4.2×10^{-13}	5.5×10^{-15}	3.3×10^{-19}	3.3×10^{-23}
	$\langle E^2\rangle_z$ (V²/cm²)	4.2×10^{-2}	5.7×10^{-4}	3.4×10^{-8}	3.4×10^{-12}
	$\langle E^2\rangle_M$ (V²/cm²)	4.99	0.067	4.0×10^{-6}	4.0×10^{-10}
	$\bar{z}_{\rm rms}$(μm)	1300	145	1.1	0.011

imum laser cooling) and $E_{max} = 7.4$ V/cm at $T = 4$ K. Therefore at modest temperatures, ion-ion collision induced Stark shifts can be quite small.

For clouds of ions, other effects can contribute to Stark and second order Doppler shifts. We will consider only theoretical limits and therefore neglect effects such as rf heating which may be the real limitation in a practical experiment. We will assume that the secular motion in an rf trap and the axial and cyclotron modes in a Penning trap have been cooled to negligible values. For both traps we will assume that it is desirable to maximize the number of ions N.

Then in an rf trap we must consider the effects of the micromotion and corresponding electric fields for ions on the edge of the cloud. We impose the constraint that the maximum fractional second order Doppler shift not exceed a certain value (ϵ). Therefore for the rf spherical well we set $\Psi(r,z) = -\phi_i$, and with the condition on ϵ (the maximum second order Doppler effect is due to ions at $z = r_i$, $r = 0$) we find

$$N_{max} = 6.48 \times 10^{15}\, r_i M \epsilon$$

when M is in u, and r_i is the cloud radius.

For the Penning trap, the maximum second order Doppler effect is due to the magnetron motion of ions on the edge of the cloud ($r_m = r_i$, $z = 0$). We set $\phi_i(z) = -\phi_T(z)$ and for a spherical cloud we find from Eqs. (5–7),

$$N_{max} = 1.96 \times 10^{13}\, B \sqrt{\epsilon} \left[r_i^2 - \frac{440\, M \sqrt{\epsilon}}{B} r_i \right]$$

where B is in tesla, M in u. Negative solutions are unphysical because they correspond to parameters where the magnetron second order Doppler shift cannot be made as large as ϵ (for a spherical cloud).

For very small values of ϵ, very low density ion clouds are required which implies very small applied potentials. From the condition $\phi_i = -\Psi_T$ for the rf trap we have (for the spherical well):

$$\frac{2eA_0^2}{M\Omega^2} = 7.2 \times 10^{-8} \frac{N_{max}}{r_i^3} \left[\frac{V}{cm^2}\right]$$

and similarly, for the Penning trap we have:

$$A = 3.6 \times 10^{-8} \frac{N_{max}}{r_i^3} \left[\frac{V}{cm^2}\right].$$

We can also calculate the corresponding electric fields. As before, for the rf trap, the maximum fields occur on ions for $z = r_i$ and $r = 0$ and we have

$$\langle E^2 \rangle_{z(max)} = 8\, r_i^2\, A_0^2 = \frac{2M\Omega^2 N_{max}}{r_i}.$$

In the Penning trap, the electric fields cancel along the z axis. Along the radial direction,

$$\langle E^2 \rangle_{r(max)} = (2Mc^2\epsilon/(er_i))^2.$$

In Table 2 are shown some representative values of maximum numbers of stored ions, trap potentials, and Stark shifts for various values of ϵ and r_i on clouds of ions. In certain configurations, second order Doppler and Stark shifts could still be a problem; however, with small enough numbers of ions these can be overcome. We note that in many cases, the electric fields from black-body radiation ($\langle E^2 \rangle_{bb} \simeq (10\ V/cm)^2$) can be much larger than those due to trapping conditions [78, 89, 90]. Therefore, operation at reduced environmental temperatures may ultimately be required.

TABLE 2. *Maximum numbers, pseudopotential well depths for rf traps, applied potentials for Penning traps, and electric fields for "cold" spherical ion clouds in rf and Penning traps. A maximum fractional second order Doppler shift ϵ is assumed. The secular motion for the rf trap and the axial and cyclotron motion for the Penning trap are assumed to be frozen out (i.e., cooled to negligible values). r_i = ion cloud radius: $M = 100$ u, $\Omega/2\pi = 1$ MHz, $B = 1$ T.*

	ϵ	10^{-12}	10^{-12}	10^{-15}	10^{-15}
rf	r_i (cm)	0.5	0.01	0.5	0.01
	N_{max}	3.2×10^5	6500	320	~6
	$\frac{2eA_0^2}{M\Omega^2}\left[\frac{V}{cm^2}\right]$	0.19	470	1.9×10^{-4}	0.47
	$\langle E^2 \rangle_z \left[\frac{V^2}{cm^2}\right]$	760	760	0.76	0.76
Pen.	N_{max}	4.5×10^6	—	1.5×10^5	53
	$A_0\left[\frac{V}{cm^2}\right]$	1.3	—	0.043	1.9
	$\langle E^2 \rangle_{r_i} \left[\frac{V^2}{cm^2}\right]$	0.14	—	1.4×10^{-7}	3.5×10^{-4}

7. Other Possibilities

With the special features of the ion traps in mind, it is worthwhile speculating on possible future experiments. A very interesting class of experiments has already been initiated on multiply charged ions. Electron capture rates of Ne^{q+} $(1 \leq q \leq 10)$ in an orbitron trap have recently been measured [14] and Penning traps are being developed for high Z ion storage [91]. These experiments may pave the way for interesting spectroscopy on for example high Z hydrogenic ions where precise theoretical calculations are available. Various energy separations become higher; for example the $^{235}U^{91+}$ ground state hfs transition may be in the visible [92]. In the future, more exotic species may be available for study such as antiprotons, e^-/e^- pairs in a Penning trap or e^+/e^-, e^+/p^-, e^-/p^+ or p^+/p^- pairs in an rf trap [92].

The Penning trap may be an ideal place to measure g_J factors in certain atomic ions. First, because very small samples (<100 μm) can be used, the effects of inhomogeneous fields are greatly reduced [81]. Second, atomic g_J factors could be measured in terms of the Bohr magnetron by measuring the cyclotron frequency of free electrons in the same magnetic field. This technique has already been used to measure the g_J factor in S^- [93]. More generally, these features plus the very high resolutions possible [82] should allow precise measurements of other magnetic-field-dependent structure. For example, deviations in the Breit-Rabi formula for hfs could be measured and it may be possible to measure the effects of nuclear diamagnetism [94].

From Tables 1 and 2, it is evident that Stark shifts can be very small on cooled ions; this suggests that the spectroscopy of Rydberg ions with minimal Stark shifts may be possible. With this in mind, perhaps a measurement of the Rydberg along the lines of Kleppner's proposal [95] is possible in an ion.

Finally, ion storage techniques might also be used for measurement devices, for example, to measure magnetic fields [96] or frequencies of lasers [45, 97].

Quite generally, it is now feasible to think of experiments on stored ions where the perturbations are extremely small. Hopefully, this can lead to many new precision experiments other than the ones discussed here.

The preparation of this paper was supported in part by the Air Force Office of Scientific Research and the Office of Naval Research. The author thanks W. M. Itano, G. Dunn, J. C. Bergquist, R. E. Drullinger and F. L. Walls for a critical reading of the manuscript and R. Ray and B. Barrett for manuscript typing.

References

[1] H. G. Dehmelt, Adv. At. Mol. Phys. 3, 53 (1967); 5, 109 (1969).
[2] F. G. Major and J. P. Schermann, Bull. Am. Phys. Soc. 16, 838 (1971).
[3] C. Audoin, private communication (1979).
[4] F. G. Major, J. de Phys. (Paris) 38, L-221 (1977).
[5] D. Wineland, P. Ekstrom, and H. Dehmelt, Phys. Rev. Lett. 31, 1279 (1973).
[6] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, in *New Frontiers in High-Energy Physics*, Ed. by B. M. Kursunoglu, A. Perlmutter, and L. F. Scott (Plenum, New York, 1978), p. 159; H. G. Dehmelt, in *Atomic Physics 7*, Ed. by D. Kleppner and P. M. Pipkin (Plenum, New York, 1981), p. 337.
[7] W. Neuhauser, M. Hohenstatt, P. E. Toschek, and H. Dehmelt, Phys. Rev. A 22, 1137 (1980).
[8] D. J. Wineland and W. M. Itano, Phys. Lett. 82A, 75 (1981).
[9] W. A. M. Blumberg, R. M. Jopson, and D. J. Larson, Phys. Rev. Lett. 40, 1320 (1978).
[10] W. A. M. Blumberg, W. M. Itano, and D. J. Larson, Phys. Rev. A 19, 139 (1979).
[11] R. T. McIver, Jr. in *Lecture Notes in Chemistry*, Vol. 7, Ed. by H. Hartmann and K. P. Wanczek (Springer-Verlag, New York, 1978), p. 97.
[12] F. L. Walls and G. H. Dunn, Phys. Today 27(8), 30 (August, 1974).
[13] H. A. Schuessler, Metrologia 13, 109 (1977); D. A. Church and H. M. Holzscheiter, Chem. Phys. Lett. 76, 109 (1980).
[14] C. R. Vane, M. H. Prior, and Richard Marrus, Phys. Rev. Lett. 46, 107 (1981).
[15] J. H. Malmberg and C. F. Driscoll, Phys. Rev. Lett. 44, 654 (1980); T. M. O'Neil, Proc. Inter. Conf. on Plasma Physics, II (April 1980).
[16] H. A. Schuessler, in *Progress in Atomic Spectroscopy Part B*, Ed. by W. Hanle and H. Kleinpoppen (Plenum, New York, 1979), p. 999.
[17] G. Werth, Acta Phys. Polon. A61, 213 (1982).
[18] W. Paul, O. Osberghaus, and E. Fisher, Forchungsber. Wirtsch. Verkehrsministeriums Nordhein-Westfalen No. 415 (1958).
[19] F. Plumelle, M. Desaintfuscien, J. L. Duchene, and C. Audoin, Optics Commun. 34, 71 (1980); F. Plumelle, Thesis, Universite de Paris-Sud, Orsay (1979).
[20] R. Blatt and G. Werth, Z. Phys. 299, 94 (1981). Also see these proceedings.
[21] W. Neuhauser, M. Hohenstatt, P. Toschek, and H. G. Dehmelt, Phys. Rev. Lett. 41, 233 (1978); W. Neuhauser, M. Hohenstatt, P. Toschek, and H. G. Dehmelt, Appl. Phys. 17, 123 (1978).
[22] For a review see A. Ashkin, Science 210, 1081 (1980).
[23] F. M. Penning, Physica 3, 873 (1936).
[24] J. Byrne and P. S. Farago, Proc. Phys. Soc. (London) 86, 801 (1965).
[25] D. J. Wineland, R. E. Drullinger, and F. L. Walls, Phys. Rev. Lett. 40, 1639 (1978); R. E. Drullinger, D. J. Wineland, and J. C. Bergquist, Appl. Phys. 22, 365 (1980).
[26] D. J. Wineland and H. G. Dehmelt, Int. J. Mass. Spectr. Ion Phys. 16, 338 (1975); and erratum, 19, 251 (1976); H. G. Dehmelt, Nature 262, 777 (1976).
[27] See for example: S. A. Prasad and T. M. O'Neil, Phys. Fluids 22, 278 (1979).
[28] K. H. Kingdon, Phys. Rev. 21, 408 (1923).
[29] M. H. Prior, Phys. Rev. Lett. 29, 611 (1972).
[30] M. H. Prior and E. C. Wang, Phys. Rev. Lett. 35, 29 (1975).
[31] R. D. Knight, Appl. Phys. Lett. 38, 221 (1981).
[32] The case of a purely logarithmic potential has been investigated by R. H. Hooverman, J. Appl. Phys. 34, 3505 (1963). See also R. R. Lewis, J. Appl. Phys. 53, 3975 (1982).
[33] C. E. Johnson, private communication.
[34] A. Rich and J. C. Wesley, Rev. Mod. Phys. 44, 250 (1972). Also see the review by A. Rich, these proceedings.
[35] G. Gräff, E. Klempt, and G. Werth, Z. Phys. 222, 201 (1969).
[36] G. Gräff and M. Holzscheiter, Phys. Lett. 79A, 380 (1980).
[37] D. A. Church and B. Mokri, Z. Phys. 244, 6 (1971).
[38] F. L. Walls and T. Stein, Phys. Rev. Lett 31, 975 (1973).
[39] S. Liebes, Jr., and P. A. Franken, Phys. Rev. 116, 633 (1959).
[40] J. C. Wesley and A. Rich, Phys. Rev. A 4, 1341 (1971).
[41] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Phys. Rev. Lett. 38, 310 (1977).
[42] P. Ekstrom and D. Wineland, Scientific American 243, August (1980).
[43] R. S. Van Dyck, Jr., D. J. Wineland, P. A. Ekstrom, and H. G. Dehmelt, Appl. Phys. Lett. 28, 446 (1976).
[44] D. J. Wineland and W. M. Itano, Phys. Rev. A 20, 1521 (1979).
[45] D. J. Wineland, J. Appl. Phys. 50, 2528 (1979).
[46] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Bull. Am. Phys. Soc. 24, 758 (1979).

[47] P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt, Phys. Rev. Lett. 24, 1679 (1981). Also see these proceedings.
[48] G. Gabrielse and R. S. Van Dyck, Jr., Bull. Am. Phys. Soc. 26, 598 (1981).
[49] H. Dehmelt, R. Van Dyck, P. Schwinberg, and G. Gabrielse, Bull. Am. Phys. Soc. 24, 675 (1979); and H. G. Dehmelt and G. Gabrielse, Bull. Am. Phys. Soc. 24, 758 (1979).
[50] P. B. Schwinberg and R. S. Van Dyck, Jr., Bull. Am. Phys. Soc. 24, 1202 (1979); Bull. Am. Phys. Soc. 26, 598 (1981).
[51] G. Gabrielse and H. G. Dehmelt, Bull. Am. Phys. Soc. 25, 1149 (1980). Also see these proceedings.
[52] P. B. Schwinberg and R. S. Van Dyck, Jr., Bull Am. Phys. Soc. 23, 40 (1978).
[53] R. S. Van Dyck, Jr., P. B. Schwinberg, and S. H. Bailey, in *Atomic Masses* and *Fundamental Constants 6*, Ed. by J. A. Nolen, Jr., and W. Benenson (Plenum, New York, 1980), p. 173; R. S. Van Dyck, Jr., and P. B. Schwinberg, these proceedings.
[54] G. Gärtner and E. Klempt, Z. Phys. A 287, 1 (1978).
[55] G. Gräff, H. Kalinowsky, and J. Traut, Z. Phys. 296, 35 (1980). Also see these proceedings.
[56] D. J. Wineland and H. G. Dehmelt, J. Appl. Phys. 46, 919 (1975).
[57] D. J. Wineland, Bull. Am. Phys. Soc. 18, 1573 (1973).
[58] A. M. Wapstra and K. Bos, Atomic Data and Nuc. Data Tables 19, 177 (1977).
[59] R. Hegstrom, private communication.
[60] M. H. Prior and H. A. Shugart, Phys. Rev. Lett. 27, 902 (1971).
[61] R. D. Knight and M. H. Prior, Phys. Rev. 21, 179 (1980).
[62] R. Schneider and G. Werth, Z. Phys. 293, 103 (1980).
[63] P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt, Phys. Lett. 81A, 119 (1981).
[64] S. C. Menasian and H. G. Dehmelt, Bull. Am. Phys. Soc. 18, 408 (1973); F. J. Grieman, B. H. Mahan, and A. O'Keefe, J. Chem. Phys. 72, 4246 (1980).
[65] H. A. Schuessler, E. N. Fortson, and H. G. Dehmelt, Phys. Rev. 187, 5 (1969).
[66] M. H. Prior and E. C. Wang, Phys. Rev. Lett. 35, 29 (1975); and Phys. Rev. A 16, 6 (1977).
[67] F. G. Major, NASA Report X-521.69.167, Goddard Space Flight Center (1969).
[68] H. A. Schuessler, Metrologia 7, 103 (1971).
[69] F. G. Major and G. Werth, Phys. Rev. Lett. 30, 1155 (1973); F. G. Major and G. Werth, Appl. Phys. 15, 201 (1978).
[70] H. Dehmelt, Bull. Am. Phys. Soc. 18, 1521 (1973); H. Dehmelt, Bull. Am. Phys. Soc. 20, 60 (1975); D. J. Wineland and H. Dehmelt, Bull. Am. Phys. Soc. 20, 637 (1975); H. Dehmelt, IEEE Trans. Instrum. Meas. IM-31, 83 (1982).
[71] P. L. Bender, J. L. Hall, R. H. Garstang, F. M. J. Pichanick, W. W. Smith, R. L. Barger, and J. B. West, Bull. Am. Phys. Soc. 21, 599 (1976).
[72] F. Strumia, Proc. 32nd Ann. Symp. Freq. Control, June 1978, p. 444. (Copies available from Electronic Industries Assn., 2001 Eye St., NW., Washington, DC 20006.)
[73] F. L. Walls, D. J. Wineland, and R. E. Drullinger, Proc. 32nd Ann. Symp. Freq. Control, June 1978, p. 453. (Copies available from Electronic Industries Assn., 2001 Eye St., NW., Washington, DC 20006.)
[74] M. D. McGuire, R. Petsch, and G. Werth, Phys. Rev. A 17, 1999 (1978). M. D. McGuire, Bull. Am. Phys. Soc. 26, 615 (1981).
[75] W. M. Itano and D. J. Wineland, Bull. Am. Phys. Soc. 24, 1185 (1979).
[76] M. Jardino, M. Desaintfuscien, R. Barillet, J. Viennet, P. Petit, and C. Audoin, Appl. Phys. 24, 1 (1981).
[77] P. E. Toschek and W. Neuhauser, in *Atomic Physics 7*, Ed. by D. Kleppner and F. M. Pipkin (Plenum, New York, 1981), p. 529.
[78] D. J. Wineland, W. M. Itano, J. C. Bergquist, and F. L. Walls, Proc. 35th Ann. Symp. Freq. Control, June, 1981. (Copies available from Electronic Industries Assn., 2001 Eye St., NW., Washington, DC 20006.) Also see W. M. Itano, D. J. Wineland, J. C. Bergquist and F. L. Walls, these proceedings.
[79] T. W. Hänsch and A. L. Schawlow, Opt. Commun. 13, 68 (1975).
[80] M. H. Prior and R. D. Knight, Opt. Comm. 35, 54 (1980).
[81] D. J. Wineland, J. C. Bergquist, W. M. Itano, and R. E. Drullinger, Opt. Lett. 5, 245 (1980).
[82] W. M. Itano and D. J. Wineland, Phys. Rev. A 24, 1364 (1981).
[83] W. Nagourney and H. Dehmelt, Bull. Am. Phys. Soc. 26 (1981).
[84] J. L. Hall, Science 202, 147 (1978); in *Atomic Physics 7*, Ed. by D. Kleppner and F. M. Pipkin (Plenum, New York, 1981), p. 267.
[85] D. J. Wineland, Proc. 11th Ann. Precise Time and Time Interval Meeting (NASA Publ. 2129, Nov. 1979), p. 81.
[86] W. Neuhauser, M. Hohenstatt, P. Toschek, and H. Dehmelt, Proc. 5th Int. Conf. on Spectral Lineshapes, West Berlin, July 1980.
[87] W. M. Itano and D. J. Wineland, Phys. Rev. A 25, 35 (1982).
[88] R. D. Haun, Jr., and J. R. Zacharias, Phys. Rev. 107, 107 (1957).
[89] T. F. Gallagher and W. E. Cooke, Phys. Rev. Lett. 42, 835 (1979).
[90] W. M. Itano, L. L. Lewis, and D. J. Wineland, Phys. Rev. A 25, 1233 (1982).
[91] D. A. Church, in *Physics of Electron and Atomic Collisions*, Ed. by S. Datz (North Holland, 1982), p. 533.
[92] H. G. Dehmelt, R. S. Van Dyck, Jr., P. B. Schwinberg, and G. Gabrielse, Bull. Am. Phys. Soc. 24, 757 (1979).
[93] R. M. Jopson and D. J. Larson, Phys. Rev. Lett. 47, 789 (1981).
[94] N. P. Economou, S. J. Lipson, and D. J. Larson, Phys. Rev. Lett. 38, 1394 (1977).
[95] D. Kleppner, Bull. Am. Phys. Soc. 20, 1458 (1975).
[96] P. B. Schwinberg and R. S. Van Dyck, Bull. Am. Phys. Soc. 24, 35 (1979).
[97] J. C. Bergquist and D. J. Wineland, Proc. 33rd Ann. Symp. Freq. Control, June 1979. (Copies available from Electronic Industries Assn., 2001 Eye St., NW., Washington, DC 20006.)

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Progress Toward a Stored Ion Frequency Standard at the National Bureau of Standards*

Wayne M. Itano, D. J. Wineland, J. C. Bergquist, and F. L. Walls

Frequency and Time Standards Group, Time and Frequency Division, National Bureau of Standards, Boulder, CO 80303

Two fundamental problems with the development of a primary frequency standard based on stored ions have long been apparent—the second-order Doppler shift and the low signal-to-noise ratio. Both problems have been addressed in experiments at the National Bureau of Standards (NBS)—the first by the development of the laser cooling technique and the second by the development of laser-optical-pumping techniques with high detection efficiency. Also, a hyperfine transition in $^{25}Mg^+$ has been observed by rf-optical double resonance with a linewidth of 0.012 Hz and a Q of 2.4×10^{10}. A possible microwave frequency and time standard based on a ground-state hyperfine transition in $^{201}Hg^+$ and a possible optical frequency standard based on the two-photon $5d^9\ 6s^2\ ^2D_{5/2} \longleftrightarrow 5d^{10}\ 6s\ ^2S_{1/2}$ transition in $^{199}Hg^+$ or $^{201}Hg^+$ are described.

Key words: atomic frequency standards; atomic hyperfine structure; double resonance; Hg^+; laser cooling; Mg^+; optical pumping; stored ions.

1. Introduction

It has long been realized that microwave or optical transitions of ions stored in electromagnetic traps might form the basis for new types of frequency standards. The basic advantage of such devices is that very long interrogation times (up to many seconds) and, therefore, high transition line Q's can be achieved, while the perturbations which usually accompany the confinement of atoms, such as collisions with buffer gas molecules or the cell walls, can be substantially avoided. Ground-state hyperfine transitions of $^3He^+$ [1], $^{199}Hg^+$ [2–4], and $^{137}Ba^+$ [5] ions stored in rf quadrupole traps have been observed with high resolution by direct or indirect optical pumping techniques. In fact, Q's greater than 10^{10} have been observed on the 40.5 GHz hyperfine transition of $^{199}Hg^+$. The accuracies of the measurements were limited largely by the uncertainty of the second-order Doppler (time dilation) shift. The signal-to-noise ratios were relatively low, because the maximum number of ions stored was about $10^5 - 10^6$ and the transitions were detected with an efficiency much less than one. Techniques developed at NBS in the last three years address both of these problems. Resonant light pressure has been used to cool stored ions to temperatures below 1 K, thus reducing the second-order Doppler shift. Laser-optical-pumping techniques have been developed which are, in principle, capable of detecting transitions with nearly unit efficiency, so that the signal-to-noise ratio is limited only by the statistical fluctuations in the number of ions that make the transition.

2. Laser Cooling and Double Resonance of Mg^+

Laser cooling of Mg^+ ions stored in a Penning trap has been investigated in a series of experiments at NBS [6–10]. Ground-state hyperfine transitions have been detected with extremely high resolution by rf-optical double resonance techniques [10]. Laser cooling of Ba^+ ions stored in a quadrupole rf trap has been demonstrated in a related series of experiments at Heidelberg [11].

2.1 Apparatus

In a Penning trap, ions are confined by a combination of a uniform magnetic field and a quadrupolar electrostatic potential [12, 13]. The trap used had dimensions $r_0 = 1.64\ z_0 = 0.63$ cm in the notation of Ref. [12]. Typical operating parameters were $V_0 \simeq 7$ V and $B_0 \simeq 1$ T, where V_0 is the potential across the electrodes and B_0 is the magnetic field. The storage time for Mg^+ ions in the ultrahigh vacuum environment ($P \lesssim 10^{-7}$ Pa) was about a day.

Narrowband, tunable radiation resonant with the $3s\ ^2S_{1/2} \rightarrow 3p\ ^2P_{3/2}$ transition at 279.6 nm was required for the laser cooling and the double resonance experiments. This was produced by frequency doubling the output of a single-mode cw Rhodamine 110 dye laser in a 90° phase-matched AD*P crystal. Between 5 and 30 μW of UV power was generated in a bandwidth of about 1 MHz. The dye laser could be long-term stabilized at discrete frequencies to less than 1 MHz by locking it to saturated absorption features in I_2. Fine tuning could be accomplished by varying the trap magnetic field to Zeeman shift the Mg^+ levels.

The UV radiation was focused at the center of the trap, passing through holes in the ring electrode. Resonance fluorescence photons emitted by the ions in the backward direction were detected by a photomultiplier tube and counted. The net detection efficiency for a photon emitted from an ion, taking into account the collection solid angle, mirror reflectivity, transmission of the lenses and filters, and photomultiplier quantum efficiency, was about 2×10^{-5}.

*Supported in part by the Air Force Office of Scientific Research and the Office of Naval Research.

2.2 Laser Cooling

The basic principle of laser cooling of free or harmonically bound atoms [14, 15] can be explained as follows: Consider a monochromatic, directed beam of light whose frequency is slightly below that of a strong resonance line of the atoms. Atoms whose velocities are directed toward the light source see a frequency Doppler shifted closer to resonance; if their velocities are directed away from the source, the frequency shift is away from resonance. Therefore, the atoms tend to absorb photons when their velocity is directed toward the source. This slows them down, since the momentum of the absorbed photons reduces the atomic momentum. (The photons are reemitted in random directions.) All velocity components could be reduced with six laser beams directed along the $\pm x$, $\pm y$, and $\pm z$ directions. A single beam suffices to cool all oscillational modes of an atom bound in a three-dimensional harmonic potential, if the modes are nondegenerate and if the beam is not directed along one of the principal axes [11, 15, 16].

For an ion in a Penning trap, cooling of the axial and cyclotron modes takes place in the same way as for harmonically bound atoms, if the laser is tuned below resonance [16]. Cooling of the magnetron mode, which is an $\mathbf{E} \times \mathbf{B}$ drift of the center of the cyclotron orbit around the trap axis, can be accomplished by focusing the laser beam so that it is more intense on the side of the trap axis on which the magnetron motion recedes from the laser [6–9, 16].

It is easiest to achieve very low temperatures with small numbers of ions. This is because the radial electric field due to space charge increases the magnetron velocity. This problem does not exist for a single, isolated ion, and the lowest temperatures were observed for this case [9]. A recording of the fluorescence intensity from a small number of $^{24}Mg^+$ ions is shown in Fig. 1. After the ions were cooled and localized at the trap center, an oven containing ^{25}Mg was heated in order to induce the resonant charge exchange reaction ($^{24}Mg^+ + {}^{25}Mg \rightarrow {}^{24}Mg + {}^{25}Mg^+$). The resulting $^{25}Mg^+$ ions were ejected from the trap by resonant cyclotron-magnetron rf excitation. We attribute the step decreases in the fluorescence to individual charge exchange events and the last plateau above background to the fluorescence from a single ion.

The "temperature" of a single ion was determined from the Doppler width by optical-optical double resonance (see Fig. 2). (We define the "temperature" of a single ion in terms of its average kinetic energy.) When the light is polarized perpendicular to the magnetic field and is close to resonance with the ($^2S_{1/2}$, $M_J = -1/2$) → ($^2P_{3/2}$, $M_J = -3/2$) Zeeman component, most of the $^2S_{1/2}$ population is pumped into the $M_J = -1/2$ sublevel by the following mechanism [8]: The electric dipole selection rules allow M_J to change by ± 1 in the $^2S_{1/2} \rightarrow {}^2P_{3/2}$ transition and by 0 or ± 1 in the subsequent $^2P_{3/2} \rightarrow {}^2S_{1/2}$ decay. Thus, the $-1/2 \rightarrow -3/2$ and $+1/2 \rightarrow +3/2$ transitions (we denote the $^2S_{1/2}$ and $^2P_{3/2}$ M_J-values by the first and second numbers respectively) do not optically pump the ground state, since the ion must decay back to the same sublevel. The other allowed transitions, $+1/2 \rightarrow -1/2$ and $-1/2 \rightarrow +1/2$, do cause optical pumping, because the ion can decay to either sublevel. These off-resonance transitions are driven in their Lorentzian wings (the Zeeman splitting is much greater than the Doppler broadening), at a rate $\gtrsim 1\ s^{-1}$, which is much faster than any other relaxation rate between ground-state sublevels. The detuning of the light from the $-1/2 \rightarrow +1/2$ transition is four times its detuning from the $+1/2 \rightarrow -1/2$ transition. Hence, the $M_J = +1/2$ ground-state sublevel is depopulated 16 times faster than the $M_J = -1/2$ sublevel. In the steady state, $16/17 \simeq 94\%$ of the population is in the $M_J = -1/2$ sublevel. Since the photon scattering rate is proportional to the number of ions in the $M_J = -1/2$ sublevel, any process that changes this number is detectable as a change in the fluorescence intensity. This is the basis of the double resonance technique. To obtain the data shown in Fig. 2, one laser was tuned slightly below the $-1/2 \rightarrow -3/2$ transition, in order to provide cooling and fluorescence detection, while a low power laser was swept across the $-1/2 \rightarrow -1/2$ transition in order to deplete the $M_J = -1/2$ sublevel. The resulting lineshape reflects both the natural and Doppler broadenings. Also shown in Fig. 2 are simulated curves for temperatures $T = 0$ K and 100 mK. We estimate from this data that $T = 50 \pm 30$ mK. Since the light beam was incident at an angle of 82° with respect to the magnetic field, this is primarily a measurement of the cyclotron-magnetron (x-y) temperature. The axial (z) motion was not cooled as efficiently as the cyclotron-magnetron motion because of the beam angle. The axial temperature was estimated to be about 600 mK by lowering the well depth and noting the decrease in the fluorescence due to the decreased fraction of time the ion spent in the laser beam. According to a theoretical model, it should be possible to obtain a cyclotron-magnetron temperature of about 1 mK and an

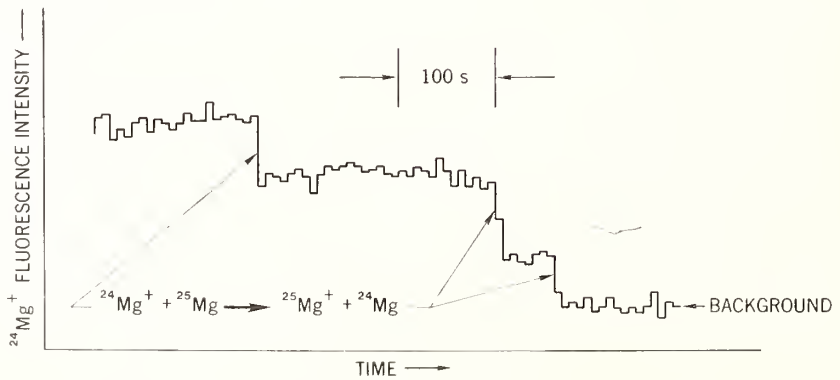


FIGURE 1. *$^{24}Mg^+$ fluorescence as a function of time. Each point represents a 10 s integration. The three large steps are due to single charge exchange events, which remove the $^{24}Mg^+$ ions, one at a time. The last plateau above background is due to a single ion. (From Ref. [9].)*

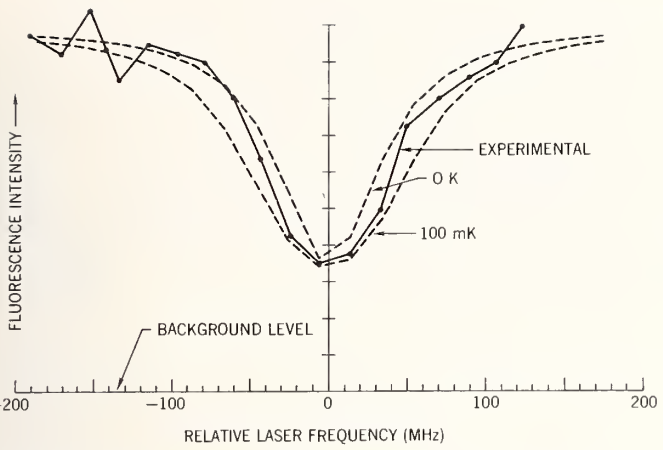


FIGURE 2. *Optical-optical double resonance of a single $^{24}Mg^+$ ion. The fluorescence due to a strong laser tuned to the ($^2S_{1/2}$, $M_J = -1/2$) → ($^2P_{3/2}$, $M_J = -3/2$) transition is plotted as a function of the frequency of a weak laser, which is tuned continuously across the ($^2S_{1/2}$, $M_J = -1/2$) → ($^2P_{3/2}$, $M_J = -1/2$) transition. Each dot represents a 10 s integration. The dashed curves are simulated data for ion "temperatures" 0 K and 100 mK. (From Ref. [9].)*

axial temperature of about 11 mK for these conditions [16]. At present, the discrepancy is not understood, but may be due to the presence of impurity ions in the trap.

2.3 RF-Optical Double Resonance

The microwave Zeeman transition in the ground state ($^2S_{1/2}$, $M_J = -1/2 \rightarrow +1/2$) also can be detected by a decrease in the fluorescence intensity [8]. This "flop-out" detection method can be very efficient. In effect, a transition due to a single microwave "photon" interrupts the stream of scattered optical photons until the ion is pumped back to the $M_J = -1/2$ sublevel by weak, off-resonance scattering. Since the decrease in the number of scattered photons can be very large, it is possible to make up for poor light collection and detector quantum efficiencies, so that the transition can be detected with nearly unit efficiency. "Flop-in" detection where a large increase in the number of photons scattered is caused by absorption of a single microwave photon, is also possible [17]. Similar detection methods have been proposed previously [18].

Various ground-state rf and microwave transitions in $^{25}Mg^+$ were detected by double resonance. $^{25}Mg^+$ has nuclear spin $I = 5/2$; in a high magnetic field, the ground-state sublevels can be denoted by (M_I, M_J). If the laser is tuned to the ($^2S_{1/2}$, $M_I = -5/2$, $M_J = -1/2$) → ($^2P_{3/2}$, $M_I = -5/2$, $M_J = -3/2$) transition, pumping into the $M_J = -1/2$ ground-state manifold takes place in the same way as for $^{24}Mg^+$, which has $I = 0$. The pumping of nearly all of the population into the $M_I = -5/2$ sublevels occurs because of hyperfine coupling in the excited state [7, 8]. The excited-state sublevels which are labeled (M_I, $M_J = -3/2$) all contain small admixtures of states with lower M_I, except for ($M_I = -5/2$, $M_J = -3/2$), which is pure. Hence, if an ion in the (M_I, $-1/2$) ground-state sublevel is excited to the (M_I, $-3/2$) excited-state sublevel, it has a small probability of decaying to a ground-state sublevel with lower M_I, unless $M_I = -5/2$. Eventually, the ions are trapped in the $M_I = -5/2$ sublevels. After this pumping takes place, the ($-5/2$, $-1/2$) → ($-5/2$, $+1/2$) electronic spin flip transition and the ($-5/2$, $-1/2$) → ($-3/2$, $-1/2$) nuclear spin flip can be detected by a decrease in the fluorescence intensity when resonant rf is applied. When the laser is tuned to the ($^2S_{1/2}$, $M_I = +5/2$, $M_J = +1/2$) → ($^2P_{3/2}$, $M_I = +5/2$, $M_J = +3/2$) transition, most of the population is pumped into the ($+5/2$, $+1/2$) sublevel, and the ($+5/2$, $+1/2$) → ($+5/2$, $-1/2$) and ($+5/2$, $+1/2$) → ($+3/2$, $+1/2$) transitions can be observed.

The observed resonances were broadened by magnetic field fluctuations. However, the magnetic field derivative of the ($-3/2$, $+1/2$) ⟷ ($-1/2$, $+1/2$) transition frequency goes to zero at $B_0 \simeq 1.2398$ T. The transition was observed near this field with linewidths as narrow as 12 mHz [10]. The following method was used to detect the transition: The ions were optically pumped into the ($-5/2$, $-1/2$) sublevel. Some population was transferred to the ($-3/2$, $1/2$) sublevel by saturating the ($-5/2$, $-1/2$) ⟷ ($-5/2$, $+1/2$) and ($-5/2$, $+1/2$) ⟷ ($-3/2$, $+1/2$) transitions with rf. The ($-3/2$, $+1/2$) → ($-1/2$, $+1/2$) transition was driven with the light blocked and the saturating rf off. The laser and saturating rf were turned on again, and the number of transitions made to the ($-1/2$, $+1/2$) sublevel was indicated by a decrease in the fluorescence level. The Ramsey interference method [19] was implemented by pulsing the rf, resulting in an oscillatory lineshape (see Figs. 3a, 3b). Several transition frequencies were measured in order to determine the hyperfine constant $A = -596.254\ 376(54)$ MHz and the nuclear-to-electronic g factor ratio $g_I/g_J = 9.299\ 484(75) \times 10^{-5}$ [10]. (Estimated standard deviations in parentheses.) The uncertainties of A and g_I/g_J are dominated by the uncertainties of the field-sensitive transitions, which are due to fluctuations of about 1 ppm in B_0. Further experimental efforts are directed toward developing a trap with greatly improved light collection efficiency in order to continue studies of Mg^+ and to initiate similar studies of Be^+.

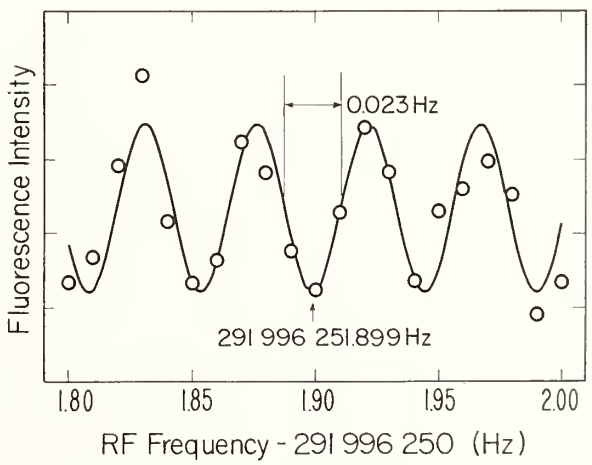


FIGURE 3a. *(M_I, M_J) = ($-3/2$, $+1/2$) ⟷ ($-1/2$, $+1/2$) $^{25}Mg^+$ ground-state hyperfine resonance. Each circle represents the average of four measurements (total fluorescence detection integration time of 8 s). The oscillatory lineshape results from the use of the Ramsey method. Two coherent rf pulses of duration $\tau = 1.02$ s separated by $T = 20.72$ s were applied. The solid curve is a theoretical fit. The vertical arrow marks the central minimum, which corresponds to the resonance frequency. The magnetic field B_0 was set so that the ($-5/2$, $-1/2$) ⟷ ($-5/2$, $+1/2$) electronic spin flip resonance was in the range of 36 248.374 ± 0.750 MHz, which corresponds to $B_0 \simeq 1.2398$ T.*

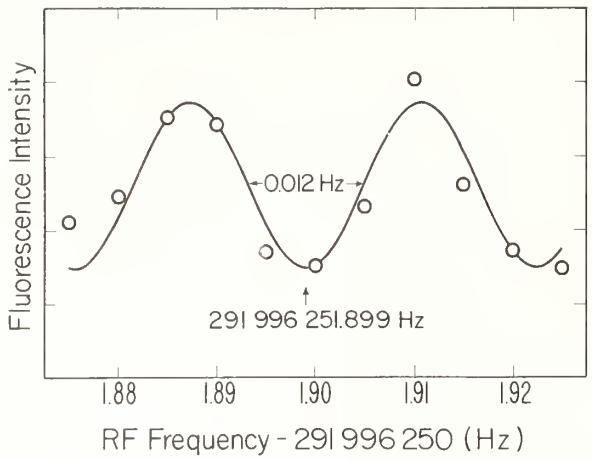


FIGURE 3b. *(−3/2, +1/2) ⟷ (−1/2, +1/2) $^{25}Mg^+$ hyperfine resonance. The total fluorescence detection integration time was 16 s per point and T was 41.40 s. The experimental conditions were otherwise the same as for Fig. 3a. The vertical arrow marks the central minimum. (From Ref. [10].)*

3. Proposals for Hg^+ Frequency Standards

Here we outline proposals for a microwave frequency and time standard based on a hyperfine transition in $^{201}Hg^+$ and for an optical frequency standard based on a two-photon transition in $^{199}Hg^+$ or $^{201}Hg^+$. These proposals incorporate the laser cooling and laser-optical-pumping techniques demonstrated with Mg^+. They use the Penning trap instead of the rf trap because cooling (of more than one ion) appears to be easier. Preliminary experimental work on these devices has begun at NBS. Further details are published elsewhere [17].

3.1 Microwave Frequency Standard

The basic advantages of Hg^+ for a microwave frequency standard are the large hyperfine structure (hence high Q for a given linewidth) and high mass (hence low second-order Doppler shift for a given temperature). Other groups [2–4, 20] have worked with the $^{199}Hg^+$ $(F, M_F) = (1,0) \longleftrightarrow (0,0)$ transition using a $^{202}Hg^+$ lamp for optical pumping and an rf trap near zero magnetic field. For a Penning trap, the best transition appears to be the $^{201}Hg^+$ $(F, M_F) = (1,1) \longleftrightarrow (2,1)$ transition, which is field-independent to first order at $B_0 \simeq 0.534$ T, with frequency $\simeq$25.9 GHz. If B_0 can be controlled to slightly better than 0.1 ppm over the volume of the ion cloud, the fractional second-order field shift can be kept below 10^{-15}. A laser tuned just below a particular hyperfine-Zeeman component of the $6s\ ^2S_{1/2} \rightarrow 6p\ ^2P_{1/2}$ 194.2 nm line could be used to cool the ions. All of the ground-state sublevels would have to be rf-mixed, in order to permit cyclic interaction with the laser. Essentially all of the ions could then be pumped into the $(F, M_F) = (1,1)$ sublevel by leaving it unmixed by the rf. The laser and mixing rf could then be shut off, and the $(1,1) \rightarrow (2,1)$ transition driven. The number of ions that made the transition could be determined by turning the laser on, rf-mixing all but the (1,1) sublevel, and counting the scattered photons. In order to keep the second-order Doppler shift below 10^{-15}, the "temperature" must be kept below 1.45 K. This should be easy for the cyclotron and axial modes, but the space-charge-induced magnetron velocity is a problem for a large cloud of ions. If the shift is to be kept below 10^{-15}, and the ions are to be contained within a 1 cm diameter spherical volume, the maximum number of ions is about 8.2×10^4. All other systematic shifts, such as those due to collisions, the trap fields, or thermal radiation, appear to be much less than 10^{-15}. For an interrogation time of 50 s, a Q of 2.6×10^{12} would be obtained using the Ramsey method. With 8.2×10^4 ions and a signal-to-noise ratio limited only by statistical fluctuations in the number of ions that make the transition, the fractional frequency uncertainty for measurement time, τ, is calculated to be $\sigma_y(\tau) \simeq 2 \times 10^{-15}\ \tau^{-1/2}$ for $\tau > 100$ s. Actual short-term stability may be limited by available local oscillators.

Various possibilities exist for the required 194.2 nm source. One is a narrowband, high repetition rate ArF* excimer laser. Another, which is being pursued at NBS, is sum-frequency mixing in a KB5 crystal of the output of a single-mode 790 nm cw ring dye laser and the second harmonic, generated in an ADP crystal, of the output of a 514 nm stabilized, single-mode cw Ar^+ laser. The basic method has been demonstrated previously with pulsed lasers [21].

3.2 Optical Frequency Standard

It has previously been pointed out that the two-photon-allowed $5d^{10}\ 6s\ ^2S_{1/2} \rightarrow 5d^9\ 6s^2\ ^2D_{5/2}$ Hg^+ transition, which has a Q of 7.4×10^{14}, could be used in an optical frequency standard [22]. The first-order Doppler effect can be eliminated by driving the transition with counter-propagating 563.2 nm laser beams. Hyperfine-Zeeman components whose magnetic field derivatives vanish at particular values of B_0 exist in $^{199}Hg^+$ and $^{201}Hg^+$. The fractional frequency shifts are much smaller than for the microwave frequency standard for the same fractional change in B_0. The second-order Doppler shift can be reduced to the same degree as for the microwave standard by cooling with a 194.2 nm laser. The two-photon transition can be detected with high efficiency by a change in the 194.2 nm fluorescence intensity. Taking full advantage of the high Q transition would require a laser with a linewidth of less than 1 Hz, which does not exist at present. However, linewidths $\simeq$100 Hz appear feasible and could be used for initial experiments. The largest systematic shift appears to be the ac Stark shift due to the 563.2 nm laser beams. If the laser linewidth is less than the natural linewidth, and the transition is near saturation, this shift is about 2×10^{-15}. Assuming ideal lasers are available, the stability of this standard is calculated to be $\sigma_y(\tau) \simeq 2 \times 10^{-18}\ \tau^{-1/2}$ for 8.2×10^4 ions and $\sigma_y(\tau) \simeq 6 \times 10^{-16}\ \tau^{-1/2}$ $(\tau \gtrsim 1$ s) for a single ion.

References

[1] H. A. Schuessler, E. N. Fortson, and H. G. Dehmelt, Phys. Rev. 187, 5 (1969).
[2] F. G. Major and G. Werth, Phys. Rev. Lett. 30, 1155 (1973); Appl. Phys. 15, 201 (1978).
[3] M. D. McGuire, R. Petsch, and G. Werth, Phys. Rev. A 17, 1999 (1978).
[4] M. Jardino and M. Desaintfuscien, IEEE Trans. Instrum. Meas. IM-29, 163 (1980); M. Jardino, M. Desaintfuscien, R. Barillet, J. Viennet, P. Petit, and C. Audoin, Appl. Phys. 24, 107 (1981).
[5] R. Blatt and G. Werth, Z. Phys. 299, 24 (1981); these Proceedings.
[6] D. J. Wineland, R. E. Drullinger, and F. L. Walls, Phys. Rev. Lett. 40, 1639 (1978).
[7] R. E. Drullinger, D. J. Wineland, and J. C. Bergquist, Appl. Phys. 22, 365 (1980).
[8] D. J. Wineland, J. C. Bergquist, W. M. Itano, and R. E. Drullinger, Opt. Lett. 5, 245 (1980).
[9] D. J. Wineland and W. M. Itano, Phys. Lett. 82A, 75 (1981).

[10] W. M. Itano and D. J. Wineland, Phys. Rev. A, 24, 1364 (1981).
[11] W. Neuhauser, M. Hohenstatt, P. E. Toschek, and H. Dehmelt, Phys. Rev. Lett. 41, 233 (1978); Appl. Phys. 17, 123 (1978); Phys. Rev. A 22, 1137 (1980).
[12] H. Dehmelt, in *Advances in Atomic and Molecular Physics*, Ed. by D. R. Bates and I. Estermann (Academic Press, New York, 1967), Vol. 3, p. 53; Vol. 5, p. 109 (1969).
[13] D. J. Wineland, these Proceedings.
[14] T. W. Hänsch and A. L. Schawlow, Opt. Commun. 13, 68 (1975).
[15] D. J. Wineland and W. M. Itano, Phys. Rev. A 20, 1521 (1979).
[16] W. M. Itano and D. J. Wineland, Phys. Rev. A 25, 35 (1982).
[17] D. J. Wineland, W. M. Itano, J. C. Bergquist, and F. L. Walls, Proc. 35th Ann. Symp. on Freq. Control (U.S. Army Electronics Command, Fort Monmouth, NJ, 1981).
[18] H. Dehmelt, Bull. Am. Phys. Soc. 20, 60 (1975).
[19] N. F. Ramsey, *Molecular Beams*, (Oxford University Press, London, 1956).
[20] M. D. McGuire, Bull. Am. Phys. Soc. 26, 615 (1981).
[21] R. E. Stickel and F. B. Dunning, Appl. Opt. 17, 981 (1978).
[22] P. L. Bender, J. L. Hall, R. H. Garstang, F. M. J. Pichanick, W. W. Smith, R. L. Barger, and J. B. West, Bull. Am. Phys. Soc. 21, 599 (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

High Resolution Microwave Spectroscopy on Trapped Ba^+ Ions*

W. Becker, R. Blatt, and G. Werth

Institut für Physik, Universität Mainz, 6500 Mainz, F.R.G.

We performed an optical double resonance experiment on the ground state of $^{137}Ba^+$ and $^{135}Ba^+$ ions. About 10^5 particles were confined in an rf quadrupole trap for many hours. Hfs state selection by pulsed laser optical pumping was followed by microwave transitions, which were observed via change in the ionic fluorescence intensity. Linewidths of the order of the laser repetition frequency (1–20 Hz) and a complicated line structure were observed in the "field independent" $F = 1$, $m = 0 \rightarrow F = 2$, $m = 0$ transitions. The statistical uncertainty of the line center was below 0.1 Hz. The results for the hyperfine separations, including corrections to zero magnetic and electric field, are

$$\Delta\nu(^{137}Ba^+) = 8\ 037\ 741\ 667.69\ (0.37)\ \text{Hz, and}$$
$$\Delta\nu(^{135}Ba^+) = 7\ 183\ 340\ 234.35\ (0.57)\ \text{Hz}.$$

Key words: frequency standard; hyperfine structure; ion storage.

Microwave resonance spectroscopy on electrodynamically confined ions can achieve very high resolution and accuracy; the relaxation times between ground state sublevels can be of the order of many seconds and uncontrolled perturbations can be made very small ($<10^{-14}$). Thus this method may have applications to frequency and time standards. Best suited for this purpose are heavy ions with large hyperfine separation to reduce the limitation imposed by the second order Doppler effect.

We performed an optical double resonance experiment on both stable odd Ba isotopes with masses 135 and 137. Preliminary results on the latter have been published recently [1]. The ions were created by surface ionization of a sample of isotope separated Ba on a hot Pt filament placed close to an endcap of an rf quadrupole ion trap, whose operation is described in detail in the literature [2, 3]. At 10^{-6} mbar of He background pressure about 2% of the total ions emitted from the filament are trapped and remain inside the trap for many hours. At typical operating conditions ($V_{DC} = 8$ V, $V_{ac} = 750$ V, $\Omega/2\pi = 500$ kHz, $r_0 = 20$ mm) the trap formed a pseudopotential valley of about 30 eV depth. The ions were maintained at a temperature which corresponds to 10% of the potential well depth [4]. Figure 1 shows a block diagram of the experimental setup. Ground state population inversion was achieved by optical pumping with a linearly polarized pulsed dye laser, tuned to one of the hyperfine components of the $6S_{1/2} - 6P_{1/2}$ transition at 493.4 nm (Fig. 2). The laser spectral width (1 GHz) and the optical Doppler width (3 GHz) were small compared to the hyperfine splitting (8 and 7.1 GHz, respectively) for the two isotopes. Taking into account that the excited $P_{1/2}$ state partially decays into the metastable $5D_{3/2}$ state, whose radiative decay time is as long as 17.5 s [5], we achieved almost complete population inversion after 10–15 laser pulses, which were well above saturation intensity.

To induce the hyperfine transition, microwaves were coupled into the trap by a hairpin antenna, placed between the ring and one of the endcap electrodes. The microwaves were created by a temperature stabilized klystron and phase locked to the 17th harmonic of a quartz oscillator, referenced to a Rb atomic clock.

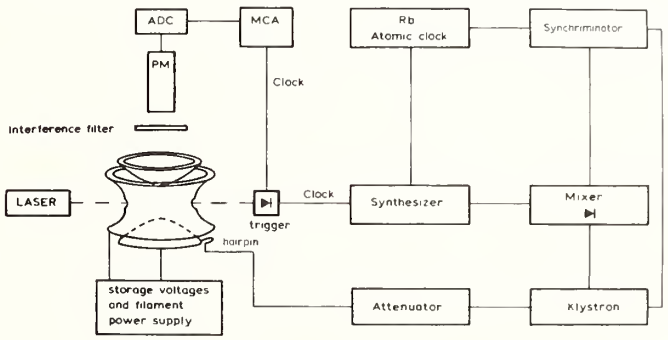


FIGURE 1. *Setup of the experiment.*

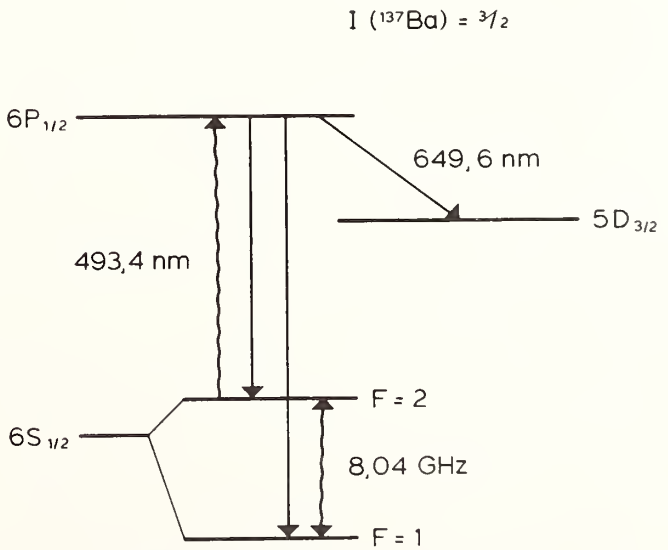


FIGURE 2. *Relevant energy levels of* Ba^+.

*Work supported by the "Deutsche Forschungsgemeinschaft."

Resonance on the possible $\Delta F = 1$, $\Delta m = 0, \pm 1$ transitions was monitored via the change in fluorescence intensity perpendicular to the primary light either at 493.4 nm or 649.6 nm. In spite of the low quantum efficiency of our photomultiplier in the red ($\lesssim 1\%$) we decided to monitor the $P \rightarrow D$ transition to be independent of any laser stray light. Among the possible transitions the $F = 2$, $m = 0 \rightarrow F = 1$, $m = 0$ line depends only in second order on the magnetic field. At field strengths of about 100 mG inhomogeneities do not represent a significant contribution to line broadening effects. The remaining field dependent $\Delta F = 1$, $\Delta m = 0, \pm 1$ transitions were used to determine the magnetic field strength. The lineshape of the "field independent" transition is dominated by the pulsed laser excitation scheme. If T is the time between two consecutive laser pulses, the well known theory for a two level system predicts for the transition probability

$$P(\omega) = \frac{\gamma^2}{(\omega - \omega_0)^2 + \gamma^2} \sin^2\{T/2((\omega - \omega_0)^2 + \gamma^2)^{1/2}\}$$

where ω is the microwave frequency, ω_0 the transition frequency, and γ the transition rate, depending on the microwave power. In our case, however, the diameter of the ion cloud, which assumes a Gaussian density distribution [4], is about 10 mm—larger than the laser beam diameter (4 mm). Thus there is a chance for an ion to escape one or even more laser pulses and extend its free interaction time with the microwave field to multiples of T. Consequently the total transition probability is given by

$$P(\omega) = \frac{\gamma^2}{(\omega - \omega_0)^2 + \gamma^2} \times \sum_{k=1}^{n} a_k \sin^2\{k\, T/2((\omega - \omega_0)^2 + \gamma^2)^{1/2}\}.$$

The constants a_k become rapidly smaller with increasing k. Throughout our analysis of the observed transitions we found that $n = 5$ was sufficient to fit all the data. Normalizing to $\Sigma a_k = 1$ we found $a_1 = 0.781$, $a_2 = 0.156$, $a_3 = 0.039$, $a_4 = 0.016$, $a_5 = 0.008$. These values were held constant throughout the whole analysis. In addition we had to add a small term, $b \exp\{-(\omega - \omega_0)^2/\gamma^2\}$ to the transition probability to account for spatial variation of the microwave power over the trap volume. Finally ion loss was assumed to be linear, since the storage times of many hours far exceeded the averaging time of a few minutes for one line. With this lineshape formula we found excellent agreement with all the observed transitions, within our limits of error. Figure 3 shows an example of a $^{137}Ba^+$ line which was taken with high spectral resolution (0.6 Hz per point) at a laser repetition rate of 16 Hz. The width of the central maximum depends linearly on the laser repetition rate. Even at repetition rates of 1 Hz we found a FWHM of 0.87 Hz in the "field independent" transition (Fig. 4) and no indication of relaxation broadening.

The limitations on accuracy in our experiment were mainly given by the applied fields and the available equipment. We varied the magnetic field between 0.2 and 2.5 G. To obtain a decent rate of information we operated the laser typically at 10–20 Hz, thus limiting the spectral resolution of the transition. The extrapolation to zero magnetic field according to the Breit-Rabi-Formula (Fig. 5) shows an uncertainty of 0.23 Hz for $^{137}Ba^+$ and 0.51 Hz for $^{135}Ba^+$. Variation of the electric field changes the potential depth of the trap and accordingly the ion temperature. This leads to a shift from the second order Doppler effect combined with a possible Stark shift. The contribution of the latter however is assumed to be very small. Experimentally we found $\delta\nu/\nu = D(-2.5 \pm 1.1) \times 10^{-12}$, if D is the depth of the trapping potential. Extrapolation to zero electric field results in an error of 0.27 Hz. We operated the trap at He buffer gas densities of about 10^{-6} mbar in order to trap a sufficient number of ions. Once the ions are trapped, the gas could have been removed. However, since a possible pressure shift of the hyperfine transition

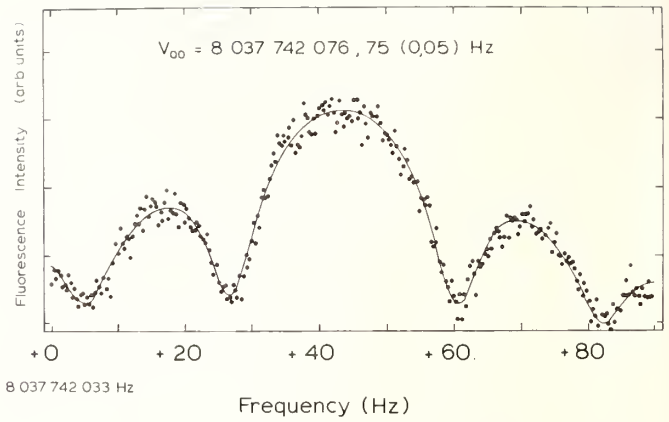


FIGURE 3. *$F = 1$, $m = 0 \rightarrow F = 2$, $m = 0$ hyperfine transition in $^{137}Ba^+$. Laser repetition rate was 16.6 s^{-1}, averaging time 15 s per point, microwave frequency steps 0.6 Hz. The full line represents the theoretical lineshape according to the formula given in the text. Statistical uncertainty of the line center (one standard deviation) is 0.07 Hz.*

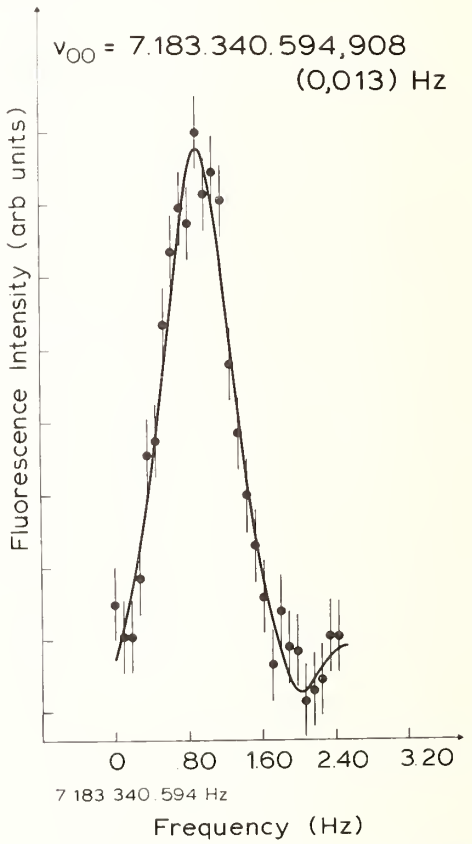


FIGURE 4. *"Field independent" transition in $^{135}Ba^+$. Laser repetition rate 1 s^{-1}, averaging time 30 s per channel, microwave steps 0.1 Hz.*

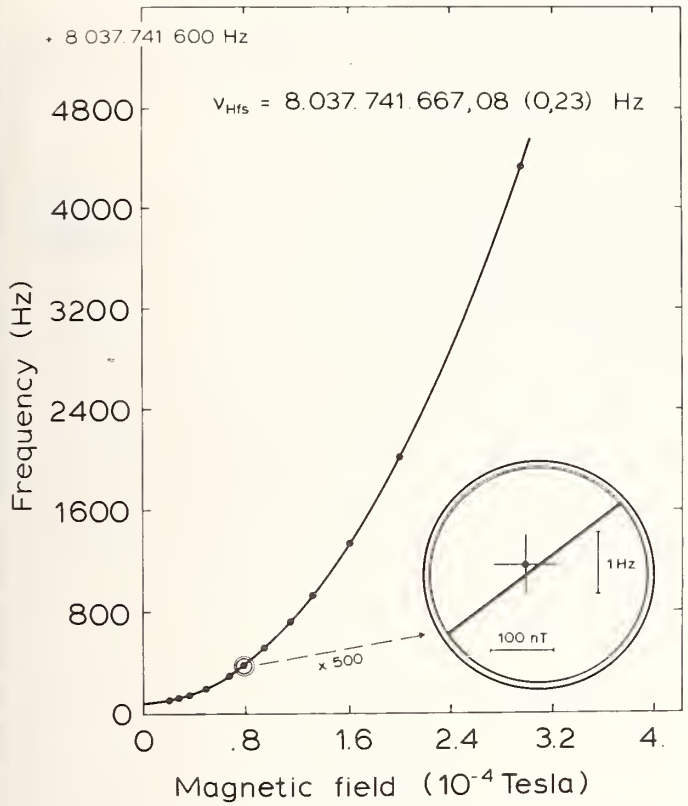


FIGURE 5. *Magnetic field dependence of the $F = 1$, $m = 0 \rightarrow F = 2$, $m = 0$ transition in $^{137}Ba^+$. The magnetic field was determined by several $\Delta F = 1$, $\Delta m = 0$, ± 1 transitions. Full line: Breit-Rabi formula fitted to the experimental points.*

according to the known data [6] is $\delta\nu/\nu \simeq 6 \times 10^{-13}$ and collisional relaxation rates between the hfs levels obviously are smaller than 1 sec^{-1} (corresponding to $\sigma < 2 \times 10^{-16}$ cm^2), we maintained this pressure during our measurements, since it increased the signal to noise ratio by collisional deexcitation of the long living metastable D-state. The final results for both Ba^+ isotopes are listed in Table 1. Further improvement in accuracy is easily possible by more careful determination of the electric and magnetic field dependence of the "field independent" line. Limitations in line-Q by collisional relaxation may be estimated from our measurements: Since at 10^{-6} mbar we obtained about $\nu/\Delta\nu = 10^{10}$, one may easily gain 3 or 4 orders of magnitude by operating at UHV pressures. We did not extend our measurements into that range, since the stability of our available crystal oscillators was too poor to obtain the required resolution.

The short term frequency stability (Allan variance $\delta f/f$) of this system operated in feedback mode as a frequency standard can be estimated. The optimum feedback signal is obtained if the frequency is chopped between the opposite points of maximum signal change of the resonance curve. For this case we have $\delta f/f = (3/4Q)(1/SNR)$ where Q is the quality factor $\omega_0/\delta\omega$ of the line and SNR the signal to noise ratio. Taking Fig. 4 as an example, where we have $SNR = 100$ at 30 s averaging time, we obtain $\delta f/f = 5 \times 10^{-12} t^{-1/2}$. This is in the same region as the short-term stability of commercially available Cs standards. Technical improvements on the signal to noise ratio are easily possible if one considers the poor solid angle of our optics of 2% and the quantum efficiency of our photomultiplier at 650 nm of about 1%. In addition the successful demonstration of optical sideband cooling [7] indicates that the main source of error, the second order Doppler effect, may be reduced by several orders of magnitude.

References

[1] R. Blatt and G. Werth, Z. Phys. A 299, 93 (1981).
[2] H. G. Dehmelt, Adv. At. Mol. Phys. 3, 53 (1976).
[3] J. F. J. Todd, G. Lawson, and R. F. Bonner, in *Quadrupole Mass Spectrometra*, Ed. by P. H. Dawson (Elsevier, Amsterdam, 1976).
[4] H. Schaaf, U. Schmeling, and G. Werth, Appl. Phys. 25, 249 (1981).
[5] R. Schneider and G. Werth, Z. Phys. A 293, 183 (1979).
[6] H. Ackermann *et al.*, Phys. Lett. 44A, 515 (1973).
[7] W. Neuhauser, M. Hohenstatt, P. Toschek, and H. Dehmelt, Appl. Phys. 17, 123 (1978).

TABLE 1. *Ground state hfs separation.*

Isotope	2A [Hz]	statistical uncert. of line center	error of magn. field extrapol.	error of pot. depth extrapol.	total error (1σ)
$^{137}Ba^+$	8 037 741 667 .69	<0.1	0.23	0.27	0.37
$^{135}Ba^+$	7 183 340 234 .35	<0.1	0.51	0.25	0.57

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Observation of High Order Side Bands in the Spectrum of Stored $^3He^+$ Ions*

H. A. Schuessler and H. S. Lakkaraju

Department of Physics, Texas A&M University, College Station, TX 77843

The magnetic resonance spectrum of the $1\ ^2S_{1/2}$ ground state of $^3He^+$ ions was remeasured using a radio-frequency quadrupole ion trap and the ion storage exchange collision technique. In particular the line shape of the $\Delta F = \pm 1$, $\Delta m_F = \pm 1$ hyperfine structure transition was studied. The spectrum was observed to consist not only of the component at the main transition frequency (ν_0 = 8655.662 MHz, H_0 = 7.13 G) but also of components at discrete frequencies of up to 400 kHz above and below the central transition frequency ν_0. A calculation of the spectrum based on the correlation function formalism reproduced the main features of the experimental spectrum although differences are discernible. The dependence of the side-band spectrum on the energy and energy distribution of stored ions in both standing and traveling wave fields was also investigated numerically. The present investigation has a strong bearing on side-band cooling techniques and on improving the precision of spectroscopic measurements.

Key words: Doppler effect of bound radiators; precision spectroscopic measurements; side bands; side-band cooling; stored ions.

1. Introduction

It is well-known that the absorption and emission of radiation from a moving atom is shifted by the Doppler effect. In general the change in photon frequency is proportional to the velocity component of the radiating atom along the direction of the photon wave vector. In the special case of a bound atomic system moving periodically about a guiding center, the Doppler effect generates side bands in the spectrum of the radiating atom. The first study [1] of these side bands was made for the spectrum of ions stored in a radio-frequency (rf) quadrupole ion trap, where an ensemble of ions is bound practically harmonically in three dimensions. It was shown that *low* order side bands occur in the hyperfine structure (hfs) spectrum of the $^3He^+$ ion at separations determined by the various characteristic ion oscillation frequencies and that the spacings are determined by the ion trap operation voltages. In this paper we report on the observation of *high* order side bands in the same system and present a theoretical interpretation valid within the limits of first order perturbation theory.

The occurrence of the side bands in the spectra of harmonically bound systems has significant impact on the precision measurements of atomic and molecular spectra, as demonstrated by the side-band cooling experiments of stored ions in the recent past [2], making it possible to improve the precision of ultrahigh resolution spectroscopy experiments by reducing the Doppler effects.

2. Experiments

We have used the ion storage exchange collision technique [3] to observe the $\Delta F = \pm 1$ and $\Delta m_F = \pm 1$ hfs transition of the $1\ ^2S_{1/2}$ ground state of $^3He^+$. The apparatus used in our work is illustrated in Fig. 1. $^3He^+$ ions are produced inside the rf quadrupole ion trap by electron bombardment of the background helium gas and

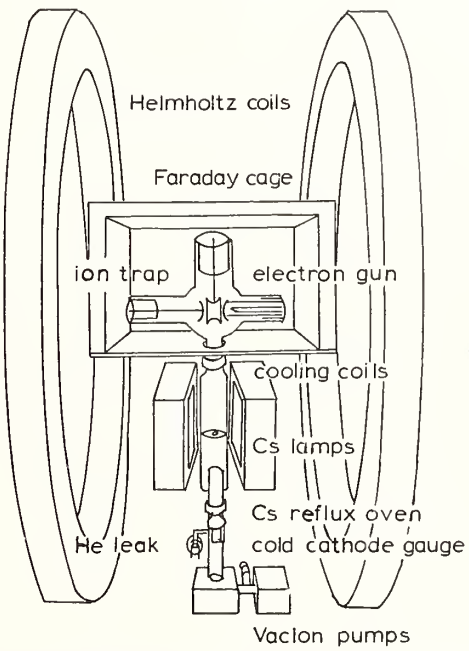


FIGURE 1. *Experimental arrangement of the ion trap apparatus for studying motional side bands in the microwave spectrum of $^3He^+$ ions.*

are subjected to an atomic beam of cesium polarized by optical pumping. In the ensuing collisions, the ions quickly assume the polarization of the cesium atoms by spin exchange and are subsequently slowly lost by charge exchange. The rate at which the $^3He^+$ ions are lost is determined by the ion polarization. In this way the spin dependent charge exchange is used to monitor the changes in the polarization of the $^3He^+$ ions when magnetic dipole transitions are induced between the hyperfine structure levels of the $^3He^+$ ground state. Magnetic dipole transitions are then detected by counting the number of ions left in the trap after a fixed interaction

*Supported by the National Science Foundation under Grant No. Phy 81–11943 and by the Center for Energy and Mineral Resources at Texas A&M University.

time with the cesium atomic beam. The ion trap structure is not only used to store the ions but also as a microwave cavity. The microwave field required for the magnetic resonance is coupled into the trap by a $\lambda/2$ slot antenna cut into one of the end cap electrodes. The microwave field inside the trap cavity is similar to that of a TE_{013} mode of an equivalent cylindrical cavity whose effective dimensions are slightly larger than the trap dimensions. In the weak external magnetic field of 7.13 G produced by a pair of Helmholtz coils, the transition frequency is measured at 8655.662 MHz.

In Fig. 2 a typical experimental spectrum is shown. The spectrum is a point-by-point plot of the data obtained by varying the frequency in discrete steps. As is evident from the data, the spectrum consists of a central peak at the main transition frequency (ν_0 = 8655.662 MHz) and of a series of side peaks spaced at separations of the various characteristic ion oscillations frequencies. Only the central peak and the components at frequencies higher than ν_0 are shown in the figure. The additional components in the spectrum are side bands and are the manifestation of the first order Doppler effect in harmonically bound physical systems [4].

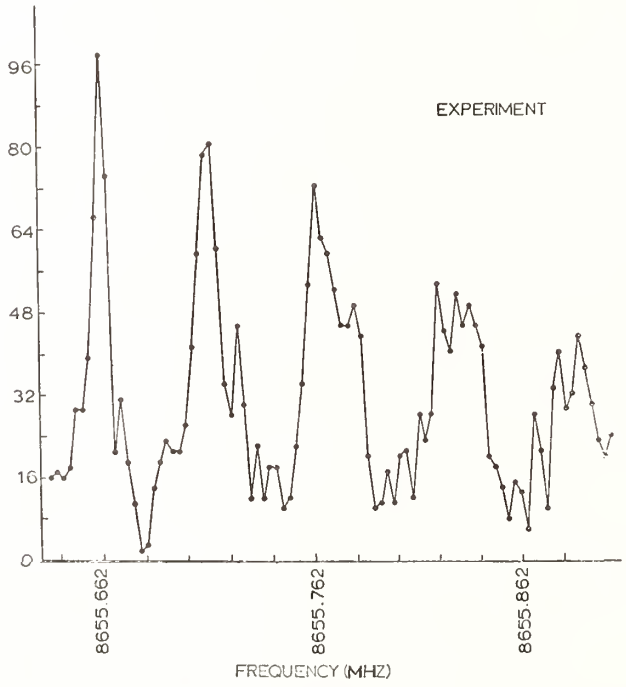


FIGURE 2. *Experimental microwave spectrum of the stored $^3He^+$ ions ($\Delta F = \pm 1$, $\Delta m_F = \pm 1$ transition of $1\,^2S_{1/2}$ ground state).*

3. Theory

In order to quantitatively account for the side-band spectrum of the oscillating ensemble of ions, a calculation using the correlation function formalism [5] was made. In this calculation the power spectrum, or alternately the transition rate, associated with the frequency $\omega_0 = (E_m - E_k)/h$ is proportional to the Fourier transform of the correlation function

$$P(\omega, \omega_0) \propto \int_{-\infty}^{\infty} G_{mk}(\tau)\, e^{-i\omega_0 \tau}\, d\tau \qquad (1)$$

where E_m and E_k are the energies of the levels involved and $G_{mk}(\tau)$ is the correlation function defined as the ensemble average of the matrix elements of the perturbation Hamiltonian in the stationary perturbation approximation given by

$$G_{mk}(\tau) = \langle \mathcal{H}_{mk}(t)\, \mathcal{H}_{km}(t') \rangle \qquad (2)$$

for the case of no or weak saturation. The matrix elements of the perturbation Hamiltonian $\mathcal{H}$ involve the microwave field inducing the transition. The microwave field in turn implicitly is a function of the motion of the ions.

The motion of the ions in a rf quadrupole trap is determined by the operating parameters of the trap and is described by a Mathieu type differential equation in both the axial and radial directions. The motion of the ions consists of the large amplitude macromotion at frequencies ω_r and ω_z with the small amplitude micromotion at the driving frequency Ω superimposed on the macromotion. In the harmonic approximation used in the following, only the dominant mode of oscillation at the macromotion is considered. The motion of the ions is then described by

$$z = \xi \sin(\omega_z t + \phi_z) \text{ and } r = \rho \sin(\omega_r t + \phi_r). \qquad (3)$$

The amplitudes ξ and ρ together with the phase angles ϕ_z and ϕ_r determine the exact trajectory that a particular ion assumes in the trap. With the motion of the ions in the trap given by Eq. (3) and the microwave field appropriate for the TE_{013} mode of a cylindrical cavity, the calculation of the intensities of the side bands in the spectrum of stored ions proceeds in two steps. First an expression for $G_{mk}(\tau)$ is derived and next the Fourier transform of $G_{mk}(\tau)$ is formed which yields the relative intensities of the side bands in the spectrum. The power spectrum thus calculated is obtained as

$$\begin{aligned} P(\omega, \omega_0) = {} & P(\omega_0) H_1^2 \sum_{m,n=0}^{\infty} \int_0^{r_{\max}} \int_{-z_{\max}}^{z_{\max}} \\ & \rho\, d\rho\, d\xi\, f(\rho, \xi) J_m^4(k_1 \rho/2) J_{2n}^2(k_3 \xi) \\ & \times [\delta(\omega - (\omega_0 + |(\pm 2m\omega_r \pm 2n\omega_z)|)) \\ & + \delta(\omega - (\omega_0 + |\mp 2m\omega_r \pm 2n\omega_z)|))] \\ & + H_1'^2 \sum_{k,l=0}^{\infty} \int_0^{r_{\max}} \int_{-z_{\max}}^{z_{\max}} \rho\, d\rho\, d\xi \\ & \times J_k^2(k_1 \rho/2) J_{k+1}^2(k_1 \rho/2) J_{2l+1}^2(k_3 \xi) f(\rho, \xi) \\ & \times [\delta(\omega - (\omega_0 + |(\mp(2k+1)\omega_r \pm (2l+1)\omega_z)|)) \\ & + \delta(\omega - (\omega_0 + |(\mp(2k+1)\omega_r \pm (2l+1)\omega_z)|))] \end{aligned} \qquad (4)$$

where the constant $P(\omega_0)$ contains all other constants, and $f(\rho, \xi)$ is the ion oscillation amplitude distribution function. In Eq. (4) k_1 and k_3 are determined by the boundary conditions of the standing wave field in the equivalent cavity with $k_1 r_c = x_{01}$ the first zero of $J_1(x)$, r_c being the radial dimension of the cavity, and $k_3 z_c = 3\pi/2$, where z_c is the axial dimension of the equivalent cavity. Also $H_1' = 1.73 H_1$ is the relationship between the field amplitudes of the radial and axial microwave fields. The above expression gives the spectrum on the high frequency side of the central line. The actual spectrum is symmetric about the central line.

The power spectrum given by Eq. (4) not only contains

the central line at $\omega = \omega_0$ but also contributions at the various multiples of ω_r and ω_z. For a given side band the contribution from ω_r and ω_z is such that either both even or both odd order side bands contribute. There are no side bands for which an odd harmonic of ω_r appears with an even harmonic ω_z or vice versa. This feature is a consequence of the spatial symmetry of the microwave field in the trap. Furthermore, it is seen that the pair of integers m and n, with specified values, gives rise to more than one spectral component, with the same intensity. For instance, two side bands are generated from the combination of the frequencies $2\omega_r$ and $2\omega_z$, one at $(\omega_0 + 2\omega_r + 2\omega_z)$ and the other at $(\omega_0 + 2\omega_z - 2\omega_r)$.

No relaxation process is present in the calculation and this leads to infinitely sharp lines represented by the delta functions. However, the observed line shapes have a finite width. For a comparison between the calculated spectrum and the observed spectrum, the delta functions are later replaced by Lorentzian functions with an empirically determined value of the linewidth parameter. Fig. 3 shows calculated spectra for a uniform distribution and a Boltzmann distribution with an equilibrium temperature $kT_e = 0.43$ eV.

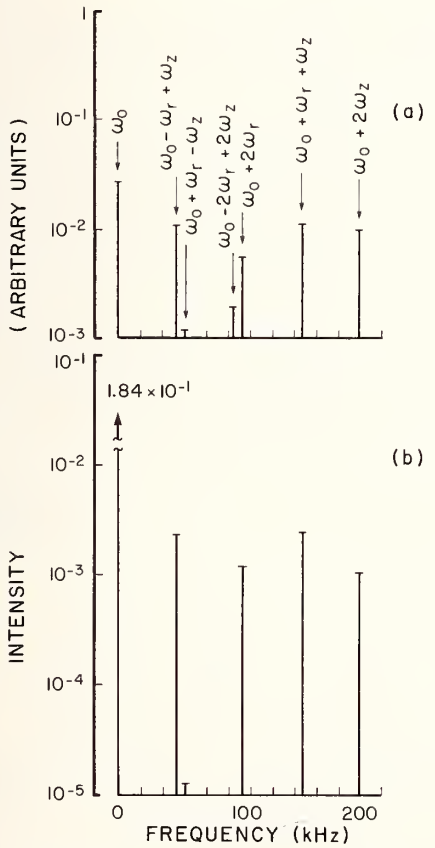


FIGURE 3. *Side-band spectrum calculated for a pure standing wave field in the ion trap cavity and no relaxation processes. (a) Uniform ion energy distribution. (b) Boltzmann distribution of ion energies for an equivalent temperature* $kT_e = 0.43$ *eV.*

So far the case of a standing microwave field in the ion trap has been discussed for the particular example of the TE_{013} mode of interest in the present experiment. Other field geometries can be handled in a similar way by replacing the microwave field by the corresponding ones for the other geometries. The calculation of the spectrum for the traveling wave field was also carried out and the power spectrum for this case is

$$P_T(\omega, \omega_0) = P_T(\omega_0) H_{1T}^2 \sum_{m=0}^{\infty} \int_0^{r_{max}} \int_{-z_{max}}^{z_{max}} \rho \, d\rho \, d\xi \, f(\rho, \xi) \times [J_m^2(k\xi) \delta(\omega - (\omega_0 + m\omega_z))] \quad (5)$$

where m is an integer, H_{1T} is the amplitude and $k = 2\pi/\lambda$ the wave number of the traveling microwave field. In order to match the conditions of the experiment it is assumed in this spectrum that the wave travels in the z direction with a plane wavefront. In this case the radial motion does not contribute and the side bands generated are due to the z motion only. Figure 4 displays the spectra calculated from Eq. (5) for a uniform distribution and a Boltzmann distribution with an equilibrium temperature of $kT_e = 0.43$ eV. The operating parameters of the experiment are shown in Table 1, and Table 2 summarizes the side-band intensities computed for a uniform energy distribution and a Boltzmann distribution for different equilibrium temperatures.

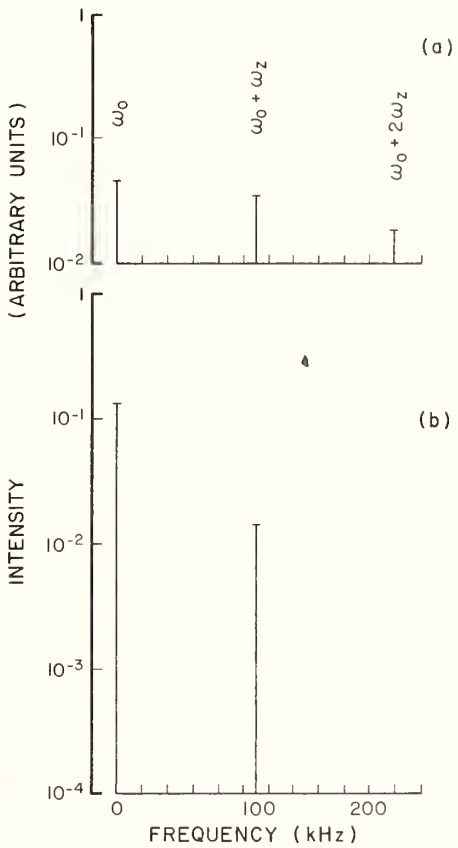


FIGURE 4. *Side-band spectrum calculated for a pure traveling wave field in the ion trap cavity (propagation along the z direction) and no relaxation processes. (a) Uniform ion energy distribution. (b) Boltzmann distribution of ion energies for an equivalent temperature* $kT_e = 0.43$ *eV.*

4. Results

A comparison of Fig. 3 and Fig. 4 clearly indicates that the spectrum due to a standing wave excitation field is richer in side bands than that due to the traveling wave.

TABLE 1. *Operating parameters.*

Trap axial dimension	z_0	=	2.55 cm
Trap radial dimension	r_0	=	3.60 cm
RF trapping frequency	Ω	=	$2\pi \times 10^6$ rad/sec
Axial oscillation frequency	ω_z	=	$2\pi \times 110 \times 10^3$ rad/sec
Radial oscillation frequency	ω_r	=	$2\pi \times 57 \times 10^3$ rad/sec
RF trapping voltage	V_0	=	127 V (Peak)
DC trapping voltage	U_0	=	0.24 V
Axial well depth	eD_z	=	5.05 eV
Radial well depth	eD_r	=	2.35 eV
Axial dimension of the equivalent microwave cavity	z_c	=	2.95 cm
Radial dimension of the equivalent microwave cavity	r_c	=	4.16 cm
Microwave power coupled into trap	P	=	300 mW
Number of trapped ions	n	=	10^6

TABLE 2. *Relative intensities of the sidebands.*

Side band: I_{mn} $m \rightarrow \omega_r$ $n \rightarrow \omega_z$	Relative Intensity				
	Uniform Distribution	Boltzmann Distribution			
		$T_e = 10^4$ K	$T_e = 5 \times 10^3$ K	$T_e = 3 \times 10^2$ K	$T_e = 1$ K
Standing Wave Field:					
I_{00}	2.73×10^{-2}	1.43×10^{-1}	1.84×10^{-1}	2.29×10^{-1}	2.70×10^{-1}
I_{11}	1.14×10^{-2}	6.00×10^{-3}	2.42×10^{-3}	1.53×10^{-5}	1.75×10^{-10}
I_{20}	5.32×10^{-3}	3.12×10^{-3}	1.18×10^{-3}	6.70×10^{-6}	7.74×10^{-11}
I_{02}	9.79×10^{-3}	2.66×10^{-3}	9.87×10^{-4}	5.00×10^{-6}	6.58×10^{-11}
Traveling Wave Field in the z Direction:					
I_0	4.69×10^{-2}	1.19×10^{-1}	1.31×10^{-1}	1.57×10^{-1}	1.59×10^{-1}
I_1	2.26×10^{-2}	2.14×10^{-2}	1.35×10^{-2}	1.09×10^{-3}	3.70×10^{-6}
I_2	1.73×10^{-2}	4.13×10^{-4}	1.12×10^{-4}	5.67×10^{-6}	3.99×10^{-11}
I_3	1.08×10^{-2}	3.37×10^{-4}	6.84×10^{-4}	2.23×10^{-8}	8.32×10^{-16}

The origin of this is the fact that in the rest frame of the oscillating ions a traveling wave appears to be frequency modulated while the standing wave appears to be frequency as well as amplitude modulated. A comparison between the experimental (solid line) spectrum and a calculated spectrum (dashed line) is made in Fig. 5. The dashed curve represents the sum of the calculated spectra, with 95% contribution from the standing wave field and 5% contribution from the traveling wave field, for the case of a uniform ion energy distribution. The calculated spectrum reproduces the observed spectrum reasonably well as far as the gross structure is concerned, although several differences are noticeable. It is evident that the widths of the experimental lines become larger for the higher order side bands, and the experimental spectrum appears to have more structure. This can be understood by noting that the number of side bands which are present depends on the spatial symmetry of the exciting radiation. In our calculation it was assumed that the field geometry is that of a TE_{013} mode, which yielded a particular combination of either even or odd harmonics. In view of the off axis coupling of the microwave field into the trap and the low Q of the trap employed, departures from the axial symmetry of the TE_{013} mode are expected. This departure will result in additional combinations of the harmonics. For example, odd

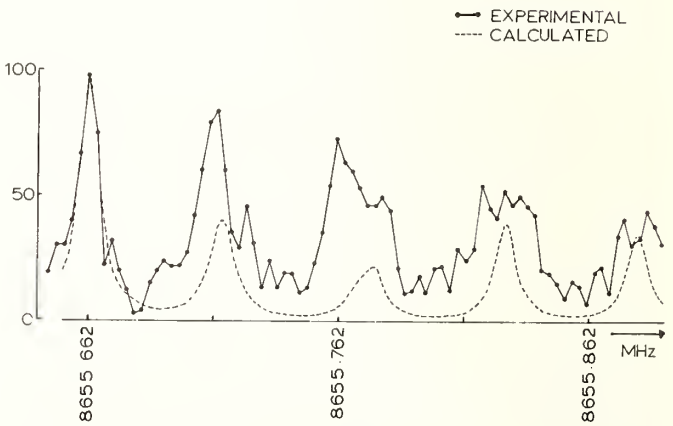


FIGURE 5. *Comparison of experimental (solid curve) and calculated (dashed curve) spectra.*

harmonics in the radial direction may weakly couple to the even harmonics in the axial direction and vice versa. In this case the observed spectrum is richer in components. In addition, side bands [6] generated by the cyclotron motion of the ions about the static magnetic field occur. As a consequence, each side band contains three

components, namely, an unshifted one and two components on either side at frequencies of $\pm\omega_c$, where ω_c is the cyclotron motion frequency. In the present experiment the magnetic field dependent side bands have a separation $\omega_c = 3.6$ kHz and are not resolved. Therefore the magnetic field dependent side bands overlap with the ion motional side bands and contribute to the width of each line. Assuming that the purely power broadened part of the line width is 4 kHz and that the height of the magnetic side bands is about 40% of the central line, a composite line width of about 10 kHz results, in agreement with our measurements. Any residual anharmonicity in the trap will also lead to a broadening of the side bands, particularly of the ones of high order, which are produced when hot ions sample the whole trap.

5. Conclusion

The line shape of transitions on bound radiating systems exhibits Doppler-effect-generated side bands which were observed in high order in the microwave spectrum of stored $^3He^+$ ions. A theoretical interpretation was developed using the correlation function formalism and is valid in the case of weak and intermediate fields for which the transition of interest is not saturated.

References

[1] H. A. Schuessler, Appl. Phys. Lett. 18, 117 (1971).

[2] D. J. Wineland, R. E. Drullinger, and F. L. Walls, Phys. Rev. Lett. 40, 1639 (1978): W. Newhauser, M. Hohenstatt, P. Toschek, and H. G. Dehmelt, Phys. Rev. Lett. 41, 233 (1978).

[3] H. A. Schuessler, E. N. Fortson, and H. G. Dehmelt, Phys. Rev. 187, 5 (1969).

[4] R. H. Dicke, Phys. Rev. 89, 472 (1952).

[5] C. P. Slichter, *Principles of Magnetic Resonance* (Harper and Row, New York, 1963) p. 127; J. H. Van Vleck and D. L. Huber, Rev. Mod. Phys. 49, 939 (1977).

[6] H. A. Schuessler, Bull. Am. Phys. Soc. 18, 532 (1971).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Quantum Limits in the Measurements of e.m. Fields and Frequency

V. B. Braginsky

Department of Physics, Moscow State University, Moscow 117234 U.S.S.R.

The quantum limit on the frequency stability of an electromagnetic (e.m.) autooscillator may be written in the following form (see Refs. [1, 2] for details):

$$\frac{\Delta\omega_e}{\omega_e} \simeq \sqrt{\frac{\hbar}{YV\tau}}\ , \tag{1}$$

where $\Delta\omega_e/\omega_e$ is the random relative deviation of the frequency during an interval τ, $\hbar$ is Planck's constant divided by 2π, Y is the Young's modulus of the material from which the e.m. resonator is made, and V is the volume of the resonator. Equation (1) is valid (i) if the power W of the e.m. autooscillator is optimal:

$$W_{\text{optimal}} \simeq \frac{YV\omega_e}{Q_e^2}\ , \tag{2}$$

where Q_e is the quality factor of the e.m. resonator in the autooscillator; (ii) if $kT < \hbar\omega_e$; and (iii) if the thermal relaxation time of the e.m. resonator is less than Q_e/ω_e. A special case of Eqs. (1) and (2) is the well-known Schawlow-Townes formula for the quantum limit on frequency stability. Substituting in Eq. (1) $Y = 4 \times 10^{12}$ dyn/cm^2 (this is the Young's modulus of sapphire), $V = 10$ cm^3, $\tau = 10$ sec, we obtain $\Delta\omega_e/\omega_e \simeq 1 \times 10^{-21}$. This value is five orders smaller than the smallest $\Delta\omega_e/\omega_e$ currently obtained by J. Turneaure [3] at Stanford University.

From a technical point of view one of the most serious obstacles to reaching the level of instability $\Delta\omega_e/\omega_e \simeq 10^{-20}$ to 10^{-21} is the factor $(1/\omega_{\text{res}})\partial\omega_{\text{res}}/\partial T$—i.e., the dependence of eigen frequency ω_{res} of an e.m. resonator on the temperature. For example a niobium microwave resonator at $T = 2$ K has $(1/\omega_{\text{res}})\partial\omega_{\text{res}}/\partial T \simeq 4 \times 10^{-9}$/K due to a strong dependence of the superconductive impedance on temperature [4]; to reach the level $\Delta\omega_e/\omega_e \simeq 10^{-16}$ it is necessary to have $\Delta T < 1 \times 10^{-6}$ K. On the other hand, if it were possible to use only a dielectrical material to make a resonator, then the factor $(1/\omega_{\text{res}})\partial\omega_{\text{res}}/\partial T$ would be approximately equal to the thermal expansion coefficient α_T. The measurements of α_T for sapphire [5] have shown that $\alpha_T \simeq 5 \times 10^{-12}$ at $T = 2$ K; in other words this factor is three orders less than for the niobium resonator. It is possible to make a microwave resonator with high Q_e using rings or disks of sapphire exciting modes with complete internal reflection. Simple calculations [6] show that the conditions on the shape of the resonator and its impurities are not so severe as to require great efforts to achieve $Q_e \simeq 10^9$. The losses in this dielectric are small ($\tan\delta < 1.5 \times 10^{-9}$; see Ref. [4]). Preliminary measurements of Q_e in a sapphire disk with a radius 5 cm and thickness 2 cm at a frequency 9×10^9 Hz have shown this approach to be promising: at $T = 300$ K we obtained $Q_e = 2 \times 10^5$, at $T = 77$ K, $Q_e \simeq 10^6$, and at $T = 4$ K the value $Q_e \gtrsim 3.5 \times 10^6$ (see Ref. [7] for details).

If an experimentalist wants to measure the values of electric field E or magnetic field B in an e.m. resonator during a short time τ, then quantum limitations become important. For example if one makes simultaneous measurements of E or B in the bandwidth $\Delta\omega_e \simeq \omega_e$ during $\tau \simeq 1/\omega_e$ then the smallest error for E and B is

$$\Delta E = \Delta B \simeq \sqrt{\frac{2\pi\hbar\omega_e}{V_e}}\ ,$$

where V_e is the volume of the resonator. The perturbations of E and B due to this measurement are described by the same Eq. (3) (see review articles, Refs. [8] and [9]). In principle it is possible to realize such procedures that beat the limit given in Eq. (3). These procedures are described in Refs. [8, 9, 10, 11]. As an example of such measurements (which now are called quantum-non-demolition measurements) we discuss here a procedure described in detail in Ref. [11]. Suppose that an optical resonator (which has the form of a closed ring of optical fiber) is located in the capacity part of a Klystron-type microwave resonator. Suppose also that the volume of the optical resonator and the volume of the capacitor are approximately equal. If the dielectric constant $\epsilon = \epsilon_0(1 + \alpha E^2)$, then if there are some optical quanta in the optical resonator, one can register a shift of the eigen frequency of the microwave resonator. If the Q is high enough and the level of losses in the fiber and in the microwave cavity are low enough, it is possible with this method to count optical quanta without absorption. In this measurement the number and phase uncertainties must satisfy $\Delta n\, \Delta\phi \geq 1/2$, but the value Δn may be much less than unity and the phase will be seriously perturbed.

References

[1] V. B. Braginsky and S. P. Vyatchanin, Sov. Phys. JETP 74, 828 (1978).

[2] V. B. Braginsky, S. P. Vyatchanin, and V. Z. Panov, Sov. Phys. Dokl. **247**, 583 (1979).
[3] J. Turneaure, private communication.
[4] V. B. Braginsky and V. Z. Panov, IEEE Trans. Magn. **MAG-17**, 955 (1981).
[5] V. B. Braginsky, S. I. Vasiliev, and V. I. Panov, JETP Lett. **6**, 665 (1980).
[6] V. B. Braginsky and S. P. Vyatchanin, Sov. Phys. Dokl. **252**, 584 (1980).
[7] V. B. Braginsky, H. S. Bagdasarov, V. I. Panov, and A. V. Timaschov, JETP Lett. **7**, 10 (1981).
[8] C. M. Caves, K. S. Thorne, R. W. P. Drever, V. D. Sandberg, and M. Zimmermann, Rev. Mod. Phys. **52**, 341 (1980).
[9] V. B. Braginsky, Y. I. Vorontsov, and K. S. Thorne, Science **209**, 547 (1980).
[10] V B. Braginsky and F. Y. Khalili, Sov. Phys. JETP **78**, 1712 (1980).
[11] V. B. Braginsky and S. P. Vyatchanin, "On the Quantum-Non-Demolition Measurement of the Energy of Optical Quanta", Preprint No. 9/1981, Physical Department, Moscow State University, 1981.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

SPECTROSCOPY, QUANTUM ELECTRODYNAMICS, AND ELEMENTARY PARTICLES

Precision Laser Spectroscopy*

T. W. Hänsch

Department of Physics, Stanford University, Stanford, CA 94305

Precision laser spectroscopy of atomic hydrogen and deuterium will be reviewed. The Balmer-α line has been studied by Doppler-free saturated absorption spectroscopy, polarization spectroscopy, optical-radiofrequency double quantum spectroscopy, and by laser-quenching of a beam of metastable atoms. These experiments have led to an eightyfold improvement in the accuracy of the Rydberg constant. Two-photon spectroscopy of the 1S-2S transition has made possible an accurate measurement of the ground state Lamb shift, and further advances in resolution promise new stringent tests of quantum electrodynamic theory.

Key words: fundamental constants; high resolution spectroscopy; hydrogen; lasers; precision measurements; quantum electrodynamics.

1. Introduction

Precision laser spectroscopy of simple quantum mechanical systems can be a powerful tool to determine better values of fundamental constants and to probe the limits of quantum electrodynamic theory. At present, only the hydrogen atom and its isotopes, the heavier hydrogen-like ions, and certain hydrogen-like exotic atoms permit sufficiently accurate model calculations for such an approach. Future theoretical advances should make it possible to add three body systems such as the neutral helium atom or the hydrogen molecular ion to the list of promising candidates.

Of these candidates, only hydrogen and deuterium have so far been studied extensively by precision laser spectroscopy [1], and we will limit our discussions to the work on these two simplest of the stable atoms. We will review the experiments reported to date, and we will investigate some of the possible directions for future progress.

2. Precision Laser Spectroscopy of the Hydrogen Balmer-α Line and the Rydberg Constant

The simple and regular Balmer spectrum of atomic hydrogen can be readily observed in the light emitted by a glow discharge. The exploration of this spectrum has played a crucial role in the development of atomic physics and quantum mechanics [1]. More than once, seemingly minute discrepancies between experiment and theory have led to major revolutions in our understanding of quantum physics. However, no classical spectroscopic observation has ever succeeded in fully resolving the intricate fine structure of the Balmer lines. The spectra always remained blurred by Doppler broadening due to the rapid thermal motion of the light hydrogen atoms.

2.1 Saturation Spectroscopy

Dramatic progress in spectral resolution became possible only with the advent of highly monochromatic tunable dye lasers together with techniques of Doppler-free laser spectroscopy. The prominent red Balmer line was the first hydrogen line to be studied by Doppler-free saturated absorption spectroscopy [2]. In this technique, two monochromatic laser beams are sent in opposite directions through the absorbing gas sample. When the laser is tuned to the center of a Doppler broadened line both beams can interact with the same atoms, those with zero axial velocity. The signal is observed as a bleaching of the absorption of the probe beam, caused by the saturating beam.

Even though only a relatively simple pulsed dye laser was available in our initial experiments at Stanford, we were thrilled by the spectra that could be recorded in a simple Wood type gas discharge. As illustrated in Fig. 1, we were able to resolve single fine structure components of the Balmer-α line, and the $n = 2$ Lamb shift could be observed directly in the optical spectrum. In 1974, Munir Nayfeh at Stanford [3] completed an absolute wavelength measurement of the strong $2P_{3/2} - 3D_{5/2}$ component of hydrogen and deuterium which yielded an eightfold improved value of the Rydberg constant, compared to the value recommended in the 1973 adjustment of the fundamental constants [4].

Since then, frequency stabilized cw dye lasers, whose resolution is not limited by pulse length, have become

*Work supported by the National Science Foundation under Grant PHY80-10689 and by the U.S. Office of Naval Research, contract ONR N00014-78-C-0403.

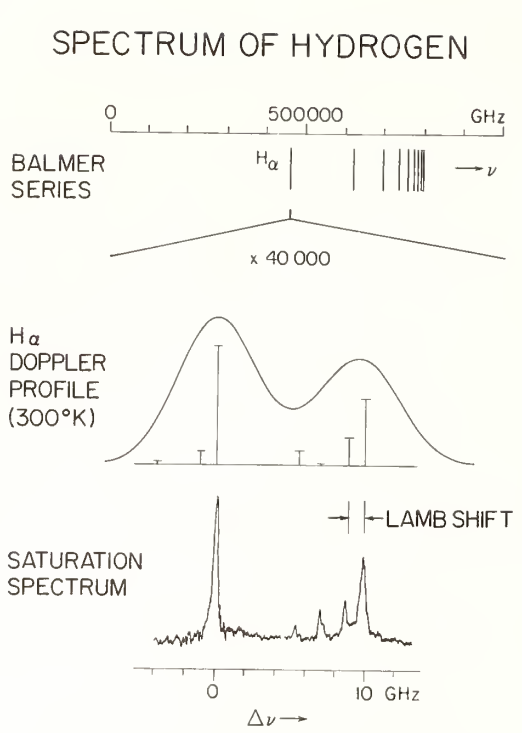


FIGURE 1. *Top: Balmer spectrum of atomic hydrogen. Center: Doppler profile of the red Balmer-α line at room temperature and theoretical fine structure components. Bottom: Doppler-free spectrum of Balmer-α, recorded by saturated absorption spectroscopy with a pulsed dye laser [2, 3].*

readily available. Moreover, new techniques of saturation spectroscopy have been developed, which can reach shot-noise limited sensitivity despite laser intensity fluctuations [5, 6]. In polarization spectroscopy [5], the probe beam monitors the dicroism and birefringence of the sample induced by a circularly or linearly polarized saturating beam. The technique takes advantage of the fact that small changes in light polarization can be detected with higher sensitivity than changes in intensity. Because of its higher sensitivity, the method permits measurements at lower atom densities and lower laser intensities, so that pressure broadening, power broadening, and related problems can be much reduced.

In 1978, J. E. M. Goldsmith *et al.* at Stanford [7] completed a new measurement of the Rydberg constant, observing the hydrogen Balmer-α line in a mild helium-hydrogen discharge by polarization spectroscopy with a cw dye laser. As shown in Fig. 2, the resolution of the weak but narrow $2S_{1/2} - 2P_{1/2}$ fine structure component was at least 5 times better than in the earlier pulsed experiments. The absolute wavelength of this component was measured relative to the i^{th} hyperfine component of the $^{127}I_2$ B − X R(127) 11 − 5 transition at 632.8 nm, using a near-coincident I_2 line (the i^{th} hyperfine component of the $^{127}I_2$ B − X R(73) 5 − 5 transition) as an intermediate reference [8].

Systematic line shifts due to the discharge plasma were studied very carefully in a series of measurements which yielded several results interesting in their own right [9, 10]. For instance, anomalous pressure shifts have been observed which could be explained in terms of collisional decoupling of the 3P hyperfine structure [9]. The final evaluation of all measurements gave a threefold improvement in the accuracy of the Rydberg value. Another twofold improvement could be obtained immediately with a more accurate absolute wavelength measurement of the intermediate iodine reference line.

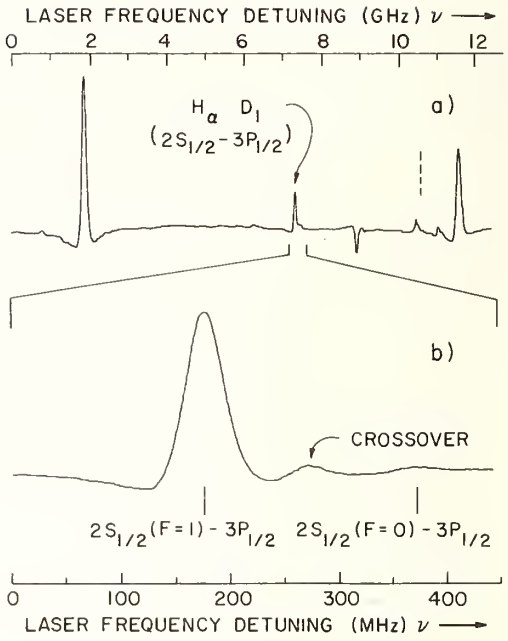


FIGURE 2. *Doppler-free polarization spectrum of the Balmer-α line, recorded with a cw dye laser in a mild He-H_2 discharge [7].*

The two Stanford measurements, as originally published, used values for the iodine reference wavelength (^{20}Ne: $^{127}I_2$, i) that differed slightly from the rounded value of 632.991 399 nm recommended by the Committee for the Definition of the Meter [11]. The results in Table 1 and Fig. 3 have been adjusted to be consistent with this rounded value. Specifically, the 1974 Rydberg value has been adjusted downwards by 0.0024 ppm, and the 1978 value has been adjusted upwards by 0.0013 ppm. The 1974 result is a composite of two somewhat differing Rydberg values for deuterium and hydrogen. The adjusted 1974 deuterium value, R_∞ = 109 737.314 8 ± 0.001 0 cm^{-1} is in excellent agreement with the 1978 measurement. The adjusted 1974 hydrogen value, R_∞ = 109 737.312 8 + 0.001 0 cm^{-1}, is almost 0.02 ppm smaller. However, the latter includes substantial systematic corrections for the larger $2P_{3/2}$ hyperfine splitting of the light isotope, and the uncertainty of these corrections in the presence of collisions may have been underestimated in 1974.

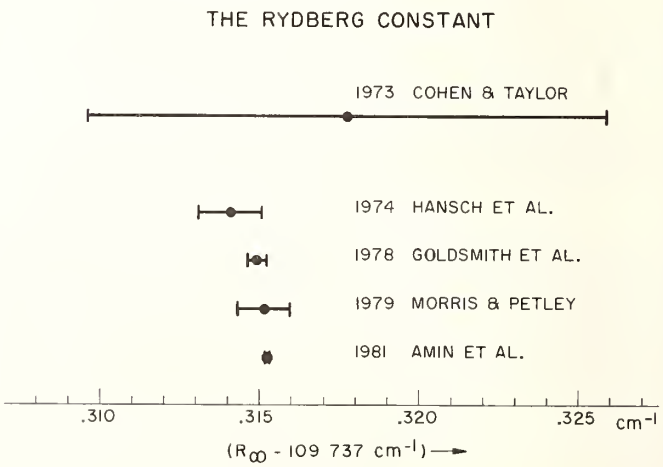


FIGURE 3. *Recent measurements of the Rydberg constant.*

Table 1 and Fig. 3 also give the results of an independent measurement of the Rydberg constant, reported in 1980 by B. W. Petley *et al.* [12]. The Balmer-α line of hydrogen was observed in a Wood type gas discharge by saturated absorption spectroscopy with a cw dye laser. The result is in good agreement with the earlier Stanford values.

TABLE 1. *Measurements of the Rydberg Constant*

Year	Authors	R_∞ [cm^{-1}]
73	COHEN, TAYLOR [6]	109 737.317 70 ± 0.008 30
74	HANSCH NAYFEH, LEE, CURRY, SHAHIN [5]	109 737.314 10 ± 0.001 00
78	GOLDSMITH, WEBER, HANSCH [10]	109 737.314 90 ± 0.000 32
79	PETLEY, MORRIS, SHAWYER [15]	109 737.315 13 ± 0.000 85
81	AMIN, CALDWELL, LICHTEN [16]	109 737.315 21 ± 0.000 11

2.2 Laser Spectroscopy of an Atomic Beam

The accuracy of the best measurements in hydrogen discharges [7] does not appear to be limited by pressure shifts or Stark effect in the discharge plasma. Nonetheless, it has long been obvious that a collisionless beam of metastable hydrogen atoms would be a more ideal sample for precision laser spectroscopy [3].

S. R. Amin *et al.* of Yale University are the first who have succeeded with such an atomic beam experiment, and they report on a new Rydberg measurement elsewhere in these proceedings [13]. The Balmer-α line of hydrogen and deuterium is observed by exciting the metastable 2S (F = 1) atoms with cw dye laser beams which cross the atomic beam at a right angle. Most of the excited 3P atoms quickly decay into the 1S ground state, and the resulting quenching of the 2S state can be observed with a detector for metastable atoms, just as in the classical Lamb-Retherford experiment [14]. Such linear atomic beam spectroscopy requires fewer systematic corrections than nonlinear saturation spectroscopy, and a wavelength measurement by direct comparison with an iodine stabilized He-Ne laser has yielded a new Rydberg value accurate to one part in 10^9, as shown in Table 1 and Fig. 3.

2.3 Two-Photon Spectroscopy

Although the accuracy of linear laser spectroscopy of a metastable hydrogen beam appears amenable to further improvements, the resolution of the narrowest Balmer-α components will always be limited by the short lifetime of the upper 3P state to no better than 29 MHz.

D. E. Roberts and E. N. Fortson [15] were the first to point out that narrower lines can be obtained if an additional radiofrequency field is applied so that radiofrequency optical double quantum transitions are induced from the $2S_{1/2}$ level to the longer living $3S_{1/2}$ and $3D_{1/2}$ level. E. W. Weber and J. E. M. Goldsmith [16] have observed lines as narrow as 20 MHz by applying this technique to hydrogen atoms in a gas discharge, and they have been able to measure the small $3P_{3/2} - 3D_{3/2}$ Lamb shift directly by comparing single- and double-quantum signals. C. E. Wieman and collaborators [17] have recently begun to apply the same technique to a beam of metastable hydrogen atoms, and they expect to reach a resolution better than 1 MHz, corresponding to the natural width of the 3S level.

The same narrow lines could also be observed by excitation with two laser photons of equal frequency. If the two photons come from opposite directions, first order Doppler broadening is automatically eliminated without any need to select slow atoms, because from a moving atom the two photons have equal but opposite Doppler shifts, so that their sum-frequency is constant [18]. Doppler-free two-photon spectroscopy of the Balmer-α transition has so far been stifled by the lack of suitable highly monochromatic tunable lasers in the near infrared. But visible dye lasers should make it possible to study transitions from 2S to high Rydberg levels by this technique.

By comparing the wavelengths of such transitions with that of the Balmer-α line, one might detect, for instance, some small deviations from Coulomb's law, which may exist within atomic dimensions, but which may have escaped detection in the past.

3. Two-Photon Spectroscopy of Hydrogen 1S–2S

There is another, even more intriguing transition in hydrogen which can be studied by Doppler-free two-photon spectroscopy: the transition from the 1S ground state to the metastable 2S state. The 1/7 sec lifetime of the upper level implies an ultimate natural linewidth as narrow as 1 Hz.

There is no intermediate near resonant level which would enhance the two-photon transition rate. However, even small numbers of excited 2S hydrogen atoms can be detected with high sensitivity, by monitoring the vacuum ultraviolet Lyman-α radiation emitted after conversion to the 2P state by collisions or external fields, or by photoionizing the 2S atoms and observing charged particles.

Unfortunately, however, two-photon excitation of 1S–2S requires monochromatic ultraviolet radiation near 243 nm, where there are still no good tunable laser sources available. Intense coherent radiation at this wavelength can be generated by frequency doubling of a pulsed dye laser in a nonlinear optical crystal, and 1S–2S two-photon spectra have been observed at Stanford with such sources [19–21]. But the resolution remained limited by the bandwidth of the pulsed lasers.

3.1 Spectroscopy with Pulsed Lasers

The best 1S–2S spectra so far have been recorded by C. E. Wieman, [21] who reached a resolution of 120 MHz (FWHM at 243 nm) with the help of a blue single-mode cw dye laser oscillator with nitrogen-pumped

pulsed dye laser amplifier chain and lithium niobate frequency doubler. The hydrogen atoms were generated in a Wood type discharge tube and carried by gas flow and diffusion into the observation chamber, where they were excited by two counterpropagating beams from the laser system. The emitted vacuum ultraviolet Lyman-α photons were observed through a magnesium fluoride side window by a solar blind photomultiplier. The spectral resolution was sufficient to resolve the hyperfine doublets (Fig. 4), although another 100 million-fold improvement should ultimately be possible. But even these crude spectra permitted a measurement of the 611 GHz H-D isotope shift to within 6.3 MHz and provided a first qualitative confirmation of the predicted small 11.9 MHz relativistic correction due to nuclear recoil [21].

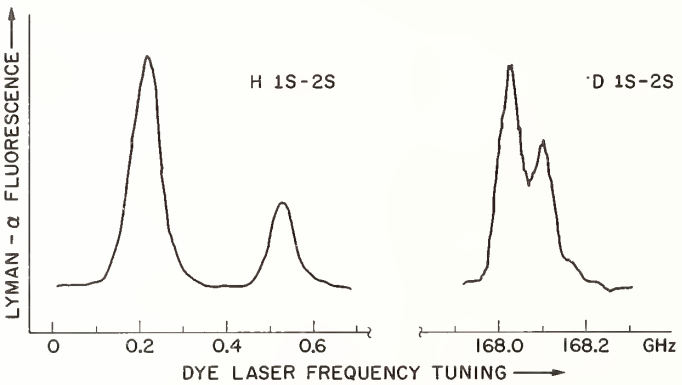


FIGURE 4. *Doppler-free two-photon spectrum of the 1S–2S transition in hydrogen and deuterium with resolved hyperfine splittings [21].*

In the same series of experiments [19–21], the 1S–2S energy interval was compared with the $n = 2 - 4$ interval, by simultaneously observing the Balmer-β line with the visible dye laser output. If the simple Bohr theory were correct, the $n = 1 - 2$ would be exactly four times the $n = 2 - 4$ interval, and both transitions would be observed at exactly the same laser frequency. In reality, this degeneracy is lifted by relativistic and quantum electrodynamic corrections, and we expect line splittings and displacements as illustrated in Fig. 5. By measuring the separation of the 1S–2S resonance from one of the Balmer components, one can determine an experimental value of the Lamb shift of the 1S ground state, which cannot be measured by radiofrequency techniques, because there is no 1P state which could serve as a reference.

The predicted 8149.43 ± 0.08 MHz Lamb shift of the hydrogen 1S state has been confirmed within 0.4%, by comparing the 1S–2S spectrum with a polarization spectrum of the Balmer-β line, observed with the cw dye laser output in a Wood type gas discharge [21]. The uncertainty in this experiment was dominated by laser frequency shifts due to rapid refractive index changes in the pulsed dye amplifiers. Such chirping introduces unknown phase parameters into the calculation of the line-shape of the two-photon signals. Pressure shifts of the Balmer-β spectrum were the next largest source of error.

Both problems can be overcome with mere technical improvements, and the intrinsic narrow natural linewidths of the observed transitions make it appear likely that measurements of the 1S Lamb shift will eventually reach a higher accuracy than radiofrequency measurements of the 2S Lamb shift [22]. The latter provide one of the most stringent current tests of quantum electrodynamics, but they are plagued by a 100 MHz natural linewidth due to the short lifetime of the 2P state. Persisting discrepancies between experiment and the predictions of different computational approaches [23, 24] make further accurate Lamb shift measurements highly desirable.

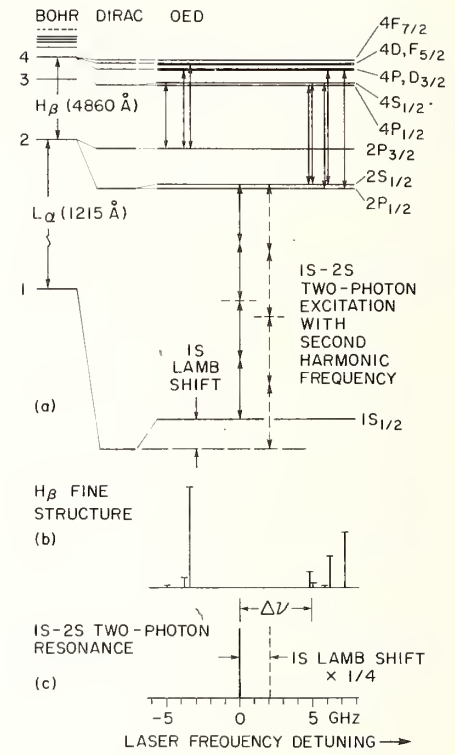


FIGURE 5. *Top: Simplified diagram of hydrogen energy levels and transitions. The Dirac fine structure and QED corrections are shown on an enlarged scale; hyperfine structure has been ignored. Bottom: Fine structure spectrum of the Balmer-β line and relative position of the 1S–2S two-photon resonance, as recorded with the second harmonic frequency. The dashed line gives the hypothetical position of the 1S–2S resonance if there were no 1S lamb shift.*

3.2 Current Efforts Towards Higher Resolution

In order to avoid the limitations of a pulsed laser source, A. I. Ferguson, J. E. M. Goldsmith, B. Couillaud, A. Siegel, J. E. Lawler and other collaborators at Stanford have invested considerable effort into an experiment designed to observe the hydrogen 1S–2S two-photon transition with low power cw ultraviolet radiation. While known nonlinear optical crystals do not permit efficient 90 degrees phase matched second harmonic generation down to 243 nm, cw ultraviolet radiation at this wavelength can be produced as the sum frequency of a violet krypton ion laser and a yellow rhodamine 6G dye laser in a crystal of ammonium dihydrogen phosphate (ADP), cooled close to liquid nitrogen temperature [25]. With frequency stabilized single frequency lasers of 0.6 and 2.5 watts power, respectively, focused to a waist diameter of 100 μm inside a 5 cm long ADP crystal, about 700 μW of tunable cw ultraviolet power have been produced. A frequency locked external passive enhancement cavity increases this power to several mW at the sample. Such a power should be sufficient for a resonant signal of several hundred Lyman-α photons per second under the chosen experimental conditions.

Unfortunately, however, the ADP crystal is damaged within less than a minute under these conditions. We speculate that the ultraviolet light produces color centers, perhaps associated with heavy ion impurities, and that these centers then absorb enough of the intense primary beams to damage the crystal. We are presently investigating whether mixing of more widely separated primary wavelengths in ADP near room temperature can provide a cure for this problem, and we are actively pursuing alternative approaches. Once the crystal damage problem is solved, it should be possible to reduce the bandwidth of such a cw ultraviolet source with the help of fast servo controls to a few kHz.

However, other causes of line broadening, in particular pressure broadening, transit broadening, and transverse Doppler broadening have to be overcome before such a resolution can be approached in the 1S–2S two-photon spectrum [25]. A beam of ground state hydrogen atoms, cooled close to liquid helium temperature [26], and interacting with nearly collinear counterpropagating laser beams would minimize such problems, and it appears technically quite feasible to observe the 1S–2S two-photon transition with a line width of a few tens of kHz, or a resolution approaching one part in 10^{11}. The line center could then be determined to within 1 part in 10^{13} or better, once accurate frequency standards become available in the visible and ultraviolet region.

Such a precise measurement of the 1S–2S frequency could, of course, be used to determine a still better value of the Rydberg constant. However, the current uncertainty of the electron/proton mass ratio [27] (about 0.14 ppm) limits the accuracy of such a Rydberg value to about 1 part in 10^{10}. Considerable improvements of direct measurements of the electron/proton mass ratio have been predicted [27]. Alternatively, a better mass ratio could be determined from a precision measurement of the 611 GHz H-D isotope shift of the 1S–2S frequency. However, uncertainties of the fine structure constant and of the mean square radii of the nuclear charge distributions would still impose error limits of about 4 parts in 10^{11} for the Rydberg.

If the electron/proton mass ratio and the fine structure constant can be measured independently with improved accuracy, then a precise measurement of the 1S–2S H-D isotope shift could provide an accurate probe for nuclear structure and recoil shifts.

To determine a precise Rydberg value that is not limited by nuclear structure corrections, one could combine the 1S–2S measurement with a precise measurement of a two-photon transition from 2S to one of the higher nS levels. Taking advantage of the fact that the lowest order nuclear structure corrections scale with the inverse cube of the principal quantum number n, one can easily construct differences of transition frequencies which are no longer sensitive to the exact nuclear sizes.

Neither a measurement of the 1S–2S frequency nor of the isotope shift by itself can provide a very stringent test of quantum electrodynamics, because we are free to adjust the values of fundamental constants until the calculations agree with the observations. However, if we form the ratio of the 1S–2S frequency to the frequency of a different hydrogenic transition, such as a Balmer transition, or a transition to or between Rydberg states, we arrive at a dimensionless quantity, which, to lowest order, no longer depends on the Rydberg constant, and which can be calculated with very high precision. An accurate measurement of such a frequency ratio would permit a very interesting test of theory, and if the past is any guide, it may well lead to some surprising discovery.

References

[1] T. W. Hänsch, G. W. Series, and A. L. Schawlow, Sci. Am. 2403, 94 (March 1979).
[2] T. W. Hänsch, I. S. Shahin, and A. L. Schawlow, Nature 235, 63 (1972).
[3] T. W. Hänsch, M. H. Nayfeh, S. A. Lee, S. M. Curry, and I. S. Shahin, Phys. Rev. Lett. 32, 1336 (1974).
[4] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[5] C. E. Wieman and T. W. Hänsch, Phys. Rev. Lett. 36, 1170 (1976).
[6] R. K. Raj, P. Bloch, J. J. Snyder, G. Camy, and M. Ducloy, Phys. Rev. Lett. 44, 1251 (1980).
[7] J. E. M. Goldsmith, E. W. Weber, and T. W. Hänsch, Phys. Rev. Lett. 41, 940 (1978).
[8] J. E. M. Goldsmith, E. W. Weber, F. V. Kowalski, and A. L. Schawlow, Appl. Opt. 18, 1983 (1979).
[9] E. W. Weber and J. E. M. Goldsmith, Phys. Lett. 70A, 95 (1979).
[10] E. W. Weber, Phys. Rev. A 20, 2278 (1979).
[11] Comité Consultatif pour la Définition du Mètre, 5e Session—1973, Bureau International des Poids et Mesures, p. M23.
[12] B. W. Petley, K. Morris, and R. E. Shawyer, J. Phys. B: Atom. Molec. Phys. 13, 3099 (1980).
[13] S. R. Amin, C. D. Caldwell, and W. Lichten, these proceedings; Phys. Rev. Lett. 47, 1234 (1981).
[14] W. E. Lamb and R. C. Retherford, Phys. Rev. Lett. 79, 549 (1950).
[15] D. E. Roberts and E. N. Fortson, Phys. Rev. Lett. 31, 1539 (1973).
[16] E. W. Weber and J. E. M. Goldsmith, Phys. Rev. Lett. 41, 940 (1978).
[17] C. E. Wieman, private communication; D. Shiner and C. Wieman, these proceedings.
[18] N. Bloembergen and M. D. Levenson, in *High Resolution Laser Spectroscopy* (Topics in Applied Physics, Vol. 13), Ed. by K. Shimoda (Springer-Verlag, Berlin, 1976), p. 315.
[19] T. W. Hänsch, S. A. Lee, R. Wallenstein, and C. E. Wieman, Phys. Rev. Lett. 34, 307 (1975).
[20] S. A. Lee, R. Wallenstein, and T. W. Hänsch, Phys. Rev. Lett. 35, 1262 (1975).
[21] C. E. Wieman and T. W. Hänsch, Phys. Rev. A 22, 192 (1980).
[22] S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. 46, 232 (1981).
[23] G. W. Erickson, Phys. Rev. Lett. 27, 780 (1971).
[24] P. J. Mohr, Phys. Rev. Lett. 34, 1050 (1975).
[25] A. I. Ferguson, J. E. M. Goldsmith, T. W. Hänsch, and E. W. Weber, in *Laser Spectroscopy IV*, Ed. by H. Walther and K. W. Rothe (Springer Series in Optical Sciences, Vol. 21, Springer-Verlag, Berlin, 1979) p. 31.
[26] S. B. Crampton, T. J. Greytak, D. Kleppner, W. D. Phillips, D. A. Smith, and A. Weinrib, Phys. Rev. Lett. 42 (1979).
[27] R. S. Van Dyck, Jr. and P. B. Schwinberg, Phys. Rev. Lett. 47, 395 (1981); these proceedings.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Atomic Beam, Linear, Single-Photon Measurement of the Rydberg Constant*

S. R. Amin, C. D. Caldwell, and W. Lichten

Department of Physics, Yale University, Box 6666, New Haven, CT 06511

The Rydberg constant has been measured to a standard error of one part in 10^9 for the first time by atomic beam, laser spectroscopy. The results are as follows: $R_\infty \lambda$ (^{3}He-^{22}Ne: $^{129}I_2$, B) = 6.946 263 169 (± 1 part in 10^9). With the CCDM definition of λ (^{3}He-^{20}Ne: $^{127}I_2$, i) rounded exactly to 632 991.399 pm, R_∞ = 109 737.315 21 (± 1 part in 10^9) cm^{-1}. The results are consistent with recent, less precise measurements.

Key words: atomic beam; fundamental constants; hydrogen; laser spectroscopy; Rydberg constant.

1. Introduction

High precision, optical spectroscopy of atomic hydrogen is of considerable current interest [1–10]. It is agreed [1–7] that an atomic beam should be used to eliminate pressure and current corrections of the order of $1 - 2 \times 10^{-8}$, which occur in gas cell measurements. One attempt has been reported [7]. Others are underway at Michigan [8], Bonn, Heidelberg [9], and NBS Washington [10]; we present a Yale effort in this paper. Efforts to use the two-photon 1S ⟷ 2S transition [11–14] by improving the waveforms [15] by means of cw lasers are current in Stanford [16] and Novosibirsk [17]. Another proposal is microwave transitions between Rydberg states of atoms to make a direct frequency determination of the Rydberg constant [18, 19].

We report an atomic beam, single photon determination of the Rydberg constant, by means of the Balmer-α line. The four transitions observed are H(2S, $F = 1$) → H(3P, $J = 1/2$, $F = 1, 0$), → H(3P, $J = 3/2$, $F = 2, 1$); D(2S, $F = 3/2$) → D(3P, $J = 1/2$, $F = 1/2, 3/2$), → D(3P, $J = 3/2$, $F = 1/2, 3/2, 5/2$). (See Fig. 1.)

1.1 Plan of Experiment

The experiment [6] is an optical analogue of the classic Lamb-Retherford paper [20]. A beam of atomic hydrogen comes from a W oven at 2850 K. A small fraction of the atoms ($\sim 10^{-6}$) is excited by electron bombardment to the metastable 2S state and is detected by secondary electron emission. Typical currents are $\sim 3 \times 10^{-14}$ A, with a signal-to-noise ratio of 300/1. A chopped, low intensity (~ 0.002 W/cm^2), tunable, cw dye laser beam crosses the atomic beam at right angles and quenches the metastable atoms with Balmer-α (2S − 3P) light. A retroflector eliminates first order Doppler shifts. A lock-in amplifier, referred to the chopper, detects the quenched part of the atomic beam.

Figure 2 shows the lasers and optics. We use standard H and D lamps, a Hartridge reversion spectroscope and quartz etalons (P.P.E.3), 0.1 mm or 1 mm thick, to allow tuning to within one GHz of the H or D Balmer-α line.

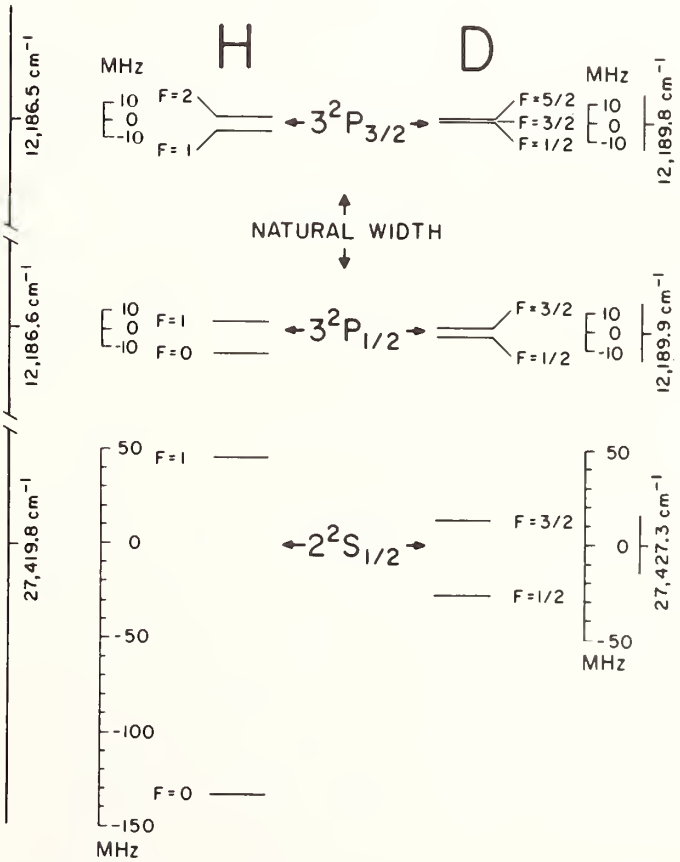


FIGURE 1. *Energy levels of H and D atoms involved in the present experiment. The term separations between fine structure levels and between levels of different principal quantum numbers are suppressed.*

*This research was aided by Grants PHY 78-26138 and PHY 80-20465 of the National Science Foundation.

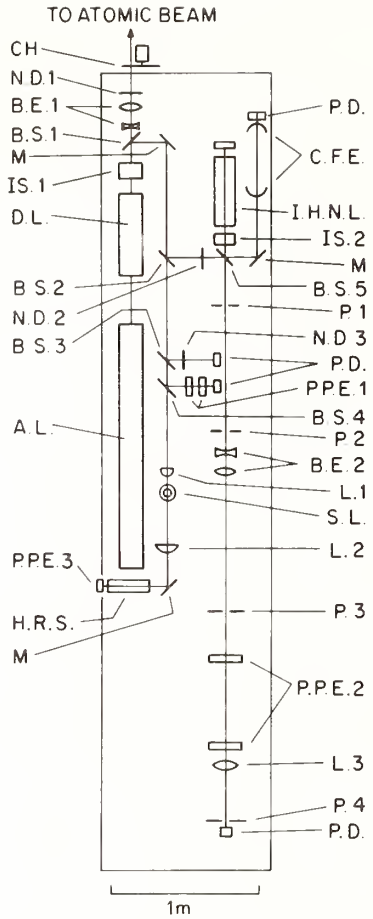


FIGURE 2. *Plan of the lasers and optics of the present experiment. CH, chopper; N. D., neutral density filter; B.E., beam expander; M, Mirror; IS, isolator; D.L., dye laser; P.D., photodiode; C.F.E., confocal scanning Fabry-Perot etalon; I.H.N.L., iodine stabilized, ^4He-^{22}Ne laser (primary wavelength standard; plasma tube Jodon CE-16–1.5 CS); A.L., argon ion laser; P.1, P.2, pinholes, 1 mm diameter; P.P.E., plane parallel Fabry-Perot etalon, 1 and 2 pressure scanned; L, Lens; P.3, pinhole, 2–5 mm diameter; P.4, pinhole, 50 μm diameter; S.L., atomic wave length standard lamp; H.R.S., Hartridge reversion spectroscope.*

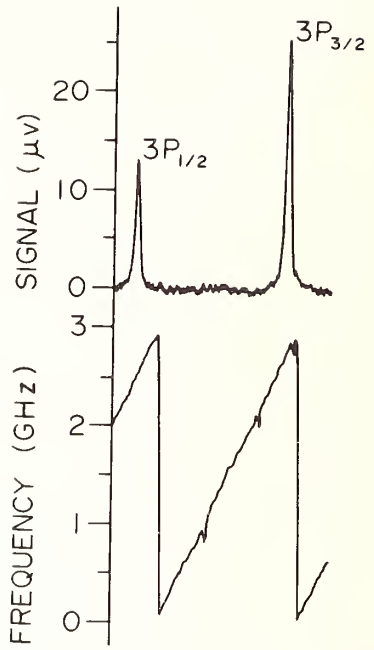


FIGURE 3. *Laser scan of the atomic hydrogen line. The plot is from a two pen recorder, with the time axis horizontal. Lower plot: scanning Fabry-Perot interferometer (free spectral range = 3 GHz). Upper plot: lock-in detector. Since the frequency scan is linear with time (lower figure), the upper figure is equivalent to a plot of signal vs. frequency. The dye laser intensity is reduced by a factor of four to ensure linearity of response with line strength. The theoretical intensity ratio of 2/1 is in excellent agreement with experiment.*

2. Methods

2.1 The Lines: Structure, Width and Shape

Figure 3 shows a dye laser scan of the two fs components of the hydrogen line. There are no crossovers or other artifacts characteristic of non-linear techniques [2–4, 11–15]. There is no hfs because the excitation process at ~ 1300 gauss leads to population of only the level $F = I + 1/2$ of the 2S state [20]; also, the hfs of the 3P state is unresolvable. (see Fig. 1.)

Figure 4 compares the D transitions to the $^2P_{3/2}$ levels with a theoretical line shape. The extrapolated line width at zero power is measured to be 42 ± 3 MHz. Early runs are with an unexpanded beam and in a stray field of 3.3 gauss. Later runs, with an expanded laser beam and magnetic shielding (stray field 0.3 gauss) produce slightly better line widths, typically power broadened to ~ 50 MHz. The line at zero optical power has a Voight profile with a natural, Lorentzian width equal to 30 MHz [21]. The Gaussian component is the root-mean-square sum of laser jitter of 5 MHz, hyperfine and Zeeman broadening of 13 MHz, and Doppler broadening of 12.5 ± 1 MHz, which arises from spread of the atomic beam (1.25 ± 0.15 mrad). The total Gaussian width of 19 ± 1 MHz, by the tables of Davies and Vaughan [22], gives a final width of 40 ± 1 MHz, which agrees with the observed value of 42 ± 3 MHz [21].

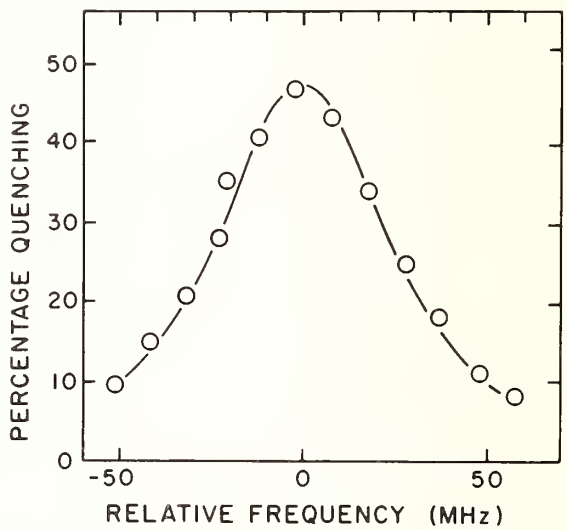


FIGURE 4. *Comparison of theoretical and experimental line shapes for D, $2^2S_{1/2} \rightarrow 3^2P_{3/2}$. Circles are experimental points; size of the circle shows error in frequency. The solid line is a theoretical curve; the only parameter is percentage quenching.*

From the measured width at zero power and transition probability at line center, the line shape (Fig. 4) is calculable with no further adjustable parameters. Within the experimental errors, the theoretical and experimental line shapes agree.

Splitting a line to 1% of its width presupposes an understanding of its shape, especially as to any possible asymmetry present [20]. We consider our understanding to be satisfactory.

2.2 Wavelength Measurements

In earlier runs with the etalon (P.P.E.2) spacing at 4.3 cm, we use a simple, manual laser tuning. For later runs, with 12.25 and 31.4 cm spacings, we frequency modulate the laser with a 15 Hz square wave of peak-to-peak shift of 50 MHz, approximately equal to the line width. With the modulating voltage as a reference for lock-in detection, the line shape becomes a derivative trace. We then hold the laser on the line center by keeping the lock-in output at zero.

We compare the wavelength of the laser line in the etalon (Fig. 2, P.P.E.2) to a ^{3}He-^{22}Ne laser, stabilized to the $^{129}I_2$ B 6329 Å peak by a third derivative lock [23]. The etalon is pressure scanned, has Cervit spacers, and has multi-layered dielectric coated mirrors with 99.5% reflectivity. Minimum finesse is $F = 40$. The standard is compared once with a portable ^{3}He-^{20}Ne: $^{127}I_2$ laser tuned to the "i" and "g" peaks, set up by Howard Layer of the U.S. National Bureau of Standards, Washington, D.C. His wavelength standard has been compared with standards in other countries [23, 24]. The $^{129}I_2$ "B" laser is $(5 \pm 5) \times 10^{-10}$ parts high in frequency.

Two methods of wavelength measurement are used. By use of two or more spacings, the method of virtual mirrors [25] eliminates the effects of phase shifts in the mirror coatings. Alternatively, we use the reflection curves furnished by the vendors to calculate the phase shift [26]. The second method uses only one spacing (31.4 cm).

The integral order numbers of the etalons are made unambiguous by measurement of the fractional order with several atomic lines, all obtained in atomic beams: H 6563, Ne 6533, Ne 6334, Ne 6266 and the standard 6329 Å. Wavelengths for the neon lines will be published elsewhere [27].

Figure 5 shows part of a pressure scan of the etalon. A typical scan covers 4 orders. Data for the deuterium 3 $^{2}P_{3/2}$ peak is taken at all three etalon spacings. Data for the other three lines is taken at the 4.3 cm and 31.4 cm spacings. A total of 272 lines is measured.

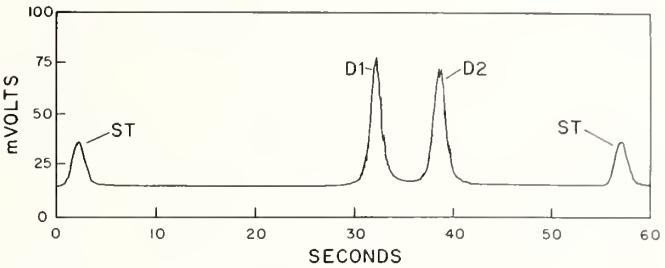


FIGURE 5. *Pressure scan of Fabry-Perot etalon with plane-parallel faces 31.4 cm apart (FSR = 478 MHz). ST, standard laser line; D_1, D_2 dye laser lines; each order is doubled by the frequency modulation.*

3. Corrections and Results

3.1 Miscellaneous Corrections

To correct for asymmetric excitation of Zeeman levels at 3.3 gauss, we average the strengths and frequencies of all the Zeeman transition components by use of the Breit-Rabi equation [28]. We also take into account the effect of the polarization of the incident light. In 0.3 gauss, where the possibility of Majorana transitions makes the magnetic sub-level populations unknown, we use the theoretical line strengths [29] to calculate shifts. Light shifts are negligible at the power levels used in this experiment. We find no observable systematic effects in the recording system and no asymmetry in the line shapes. Tables 1 and 2 list the results and corrections. Table 3 gives estimated errors.

TABLE 1. *Corrections to R_x (parts in 10^{10}).*

Transition / Correction[b]	H, $3P_{1/2}$	H, $3P_{3/2}$	D, $3P_{1/2}$	D, $3P_{3/2}$
Phase shift[c]	466	466	463	463
Refractive index[a,c]	34	34	34	34
Non-linear scan	6	9	26	22
Magnetic field[d]	−8	−5	−4	3
2S hfs	972	972	299	299
3P hfs	32	−32	10	−10
Optical pumping	0	0	−1	−2
Wavelength standard	5	5	5	5
Diffraction	−3	−3	−3	−3
2nd order Doppler	5	5	2	2
Photon recoil	−10	−10	−5	−5
Total	999[d] 1507[c]	941[d] 1446[c]	329[d] 830[c]	311[d] 805[c]

[a]See Ref. [36].
[b]For both sets of data, unless noted.
[c]Phase shift data only.
[d]Virtual mirrors data only.

TABLE 2. *Calculation of the Rydberg Constant* R_∞[a].

Upper State	H, $3P_{1/2}$	H, $3P_{3/2}$	D, $3P_{1/2}$	D, $3P_{3/2}$
Wave number[b] $1/\lambda - 15\ 230\ cm^{-1}$				
Phase shift	3.256 773	3.365 148	7.401 652	7.510 138
Virtual mirrors	3.256 749	3.365 154	7.401 639	7.510 102
Theory[c]	3.257 094	3.365 505	7.401 998	7.510 438
Rydberg constant[d] $R_\infty - 109\ 737\ cm^{-1}$				
Phase shift	0.315 39	0.315 13	0.315 21	0.315 55
Virtual mirrors	0.315 22	0.315 17	0.315 11	0.315 27
Combined by atoms:	H		D	
Phase shift	0.315 26		0.315 38	
Virtual mirrors	0.315 19		0.315 19	
Grand average:				
Phase shift	$0.315\ 32 \pm .000\ 09_4$			
Virtual mirrors	$0.315\ 19 \pm .000\ 03_4$			
Weighted average	$R_\infty = 109\ 737.315\ 21 \pm .000\ 03\ cm^{-1}$			

[a]Corrections given in Table 1.
[b]Based on CCDM $\lambda(^3He\text{-}^{20}Ne: {}^{127}I_2, i) = 632\ 991.399$ pm [30].
[c]$R_{TH} = 109\ 737.3177\ cm^{-1}$.
[d]$R_\infty = R_{TH}\lambda_{TH}/\lambda$ [37].

TABLE 3. *Error estimates (parts in* 10^{10}).

Statistical	3
Magnetic field	2
Recording and pressure scanning	6
3P hfs, optical pumping, and light shifts	5
Wavelength standard	5
Non-uniform mirror coatings	3
Diffraction	1
2nd order Doppler, photon recoil, 2S hfs, index of refraction, and minor effects	0
Root mean square error	10

3.2 Wavelengths

The measured quantities are the four wavelength ratios for the Balmer-α lines: λ (^{3}He-^{22}Ne: $^{129}I_2$, B)/λ(H or D, $2^2S_{1/2} \rightarrow 3^2P_{3/2}$ or $3^2P_{1/2}$). These, with the wavelength definitions of the CCDM [30] and the quantum electrodynamical, reduced mass, and nuclear structure corrections of Erickson [31], give four independent results for the virtual mirrors and four for the phase shift method. To average out systematic effects, we take each of the four results with equal weight. Both sets of data are combined into a weighted average for the Rydberg constant.

The primary datum of the experiment is the product of the Rydberg constant with a stabilized laser wavelength:

$$R_\infty \lambda(^3He\text{-}^{22}Ne: {}^{129}I_2, B) = 6.946\ 263\ 169 \quad (\pm 1 \text{ part in } 10^9)\ . \qquad (1)$$

With the CCDM meter [30], we have

$$\text{for } \lambda(^3He\text{-}^{20}Ne: {}^{127}I_2, i) = 632\ 991.399\ pm\ , \qquad (2)$$

$$R_\infty = 109\ 737.315\ 21\ cm^{-1}\ (\pm 1 \times 10^{-9})\ . \qquad (3)$$

To obtain R_∞ in frequency units, we use the following chain:

$$f(\text{He-Ne: } CH_4) = 88\ 376\ 181.616\ MHz \quad (\pm 0.03 \times 10^{-9}) \qquad (4)$$

as measured by Knight *et al.* [32], and

$$\frac{\lambda(\text{He-Ne: } CH_4)}{\lambda(^3He\text{-}^{20}Ne: {}^{129}I_2, k)} = 5.359\ 049\ 260\ 6 \quad (\pm 0.2 \times 10^{-9})\ , \qquad (5)$$

as measured by Layer *et al.* at NBS [33]. This gives

$$f(^3He\text{-}^{20}Ne: {}^{129}I_2, k) = 473\ 612\ 310.7\ MHz \quad (\pm 0.2 \times 10^{-9}) \qquad (6)$$

which, when combined with the known offset [34]

$$f(^3\text{He-}^{22}\text{Ne: } ^{129}\text{I}_2,\ \text{B}) - f(^3\text{He-}^{20}\text{Ne: } ^{129}\text{I}_2,\ \text{k}) = 892.563 \text{ MHz}\ , \quad (7)$$

gives

$$f(^3\text{He-}^{22}\text{Ne: } ^{129}\text{I}_2,\ \text{B})_{\text{NBS}} = 473\ 613\ 203.2 \text{ MHz} \ (\pm 0.2 \times 10^{-9})\ . \quad (8)$$

The chain of Woods *et al.* [35] gives the frequency of the (^{3}He-^{20}Ne: ^{127}I$_2$, d) laser to be 473 612 380.5 MHz ($\pm 0.6 \times 10^{-9}$) which, when combined with the offset [34] $f(^{129}\text{I}_2,\ \text{B}) - f(^{127}\text{I}_2,\ \text{d}) = 823.21$ MHz, gives

$$f(^3\text{He-}^{22}\text{Ne, } ^{129}\text{I}_2,\ \text{B})_{\text{NPL}} = 473\ 613\ 203.7 \text{ MHz} \ (\pm 0.6 \times 10^{-9})\ . \quad (8')$$

A weighted average of the two results is:

$$f(^3\text{He-}^{22}\text{Ne, } ^{129}\text{I}_2,\ \text{B})_{\text{AV}} = 473\ 613\ 203.4 \text{ MHz} \ (\pm 0.2 \times 10^{-9})\ . \quad (8'')$$

By combining Eqs. (1) and (8''), we have

$$c\,R_\infty = 3\ 289\ 841\ 951 \text{ MHz } (\pm 1 \times 10^{-9}) \quad (9)$$

We obtain a different value for R_∞ if we assume the tentative definition of the meter [34] in terms of the velocity of light: $c \equiv 299\ 792\ 458$ m/sec; $R_\infty = 109\ 737.315\ 37$ cm^{-1} ($\pm 1 \times 10^{-9}$).

4. Discussion

The results for the four lines agree very well for the virtual mirrors; the phase shift calculations agree less well. The phase shift data carries 11% of the weight of the final result and shifts it upward by two parts in 10^{10}. Table 4 and Fig. 6 compare our results with other authors [2–4, 36].

TABLE 4. *Comparison with earlier measurements.*

Reference	R − 109 737 cm^{-1}
Taylor *et al.*[b]	0.312 ± .005
Hänsch *et al.*[a,c]	0.314 ± .001
Petley *et al.*[d]	0.315 13 ± .000 85
Goldsmith *et al.*[a,e]	0.314 90 ± .000 32
Present result	0.315 21 ± .000 11

[a] Corrected to: (^{3}He-^{20}Ne: ^{127}I$_2$, i) = 632 991.399 pm
[b] Ref. [36]
[c] Ref. [2]
[d] Ref. [3]
[e] Ref. [4]

The results of Hänsch *et al.* [2] are in slight disagreement with the present paper and also are internally inconsistent. The results of Goldsmith *et al.* [4] are in fair agreement; the results of Petley *et al.* [3] and Taylor *et al.* [36] are in excellent agreement with the present results.

The results of the phase shift calculations are encouraging. Such calculations should be useful for all but the most precise measurements and help to answer an unsolved problem in metrology; that is, how to use light waves to measure dimensions of objects to very high accuracy. Recent attempts to measure objects accurately have led to inconsistent results [38]. The present phase shift data deviate by approximately one part in 10^9 from our final value. This corresponds to a phase shift on reflection of 1/6° or an error of 1.5 Å (0.00015 μm) in the location of the reflecting surface. Further investigation in this area is desirable [37].

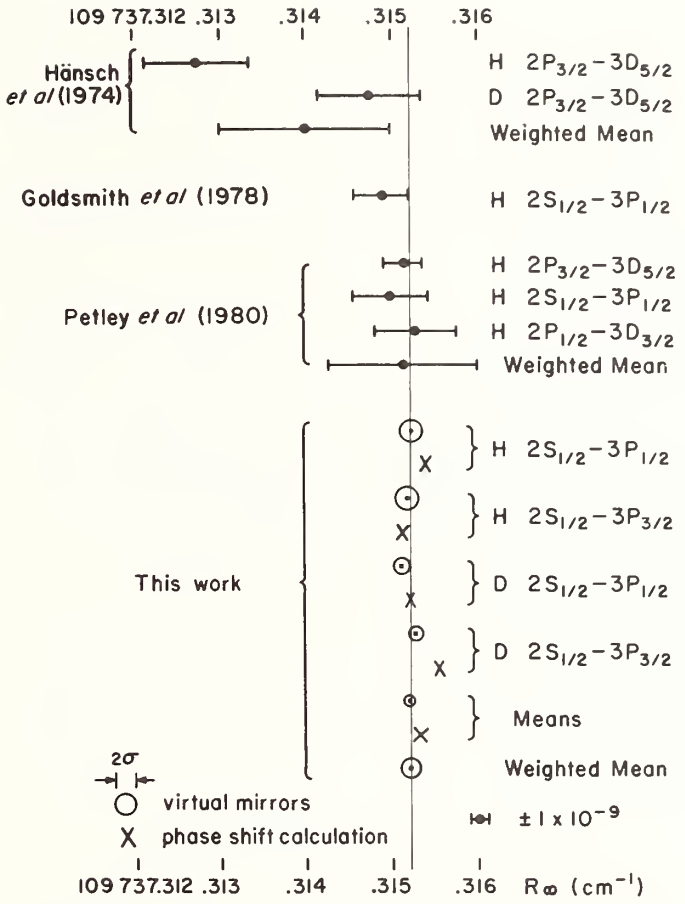


FIGURE 6. *Comparison of recent measurements of the Rydberg constant. Errors for individual measurements are statistical. Errors for weighted means include systematic errors. See Table 4.*

Minor improvements in techniques would improve Rydberg measurements to a precision of $1 - 2 \times 10^{-10}$, comparable to that of the most reliable, current definition of the meter by means of optical wavelengths. Further improvements would allow accurate measurement of the electron-proton mass ratio [6, 8, 39].

We thank Howard Layer for advice on and calibration of our wavelength standard, and for the loan of an ^{129}I$_2$ absorption cell; Barry Taylor for helpful conversations; Munir Nayfeh, Salim Jabr, and Kim Woodle for aid in building the wavelength standard; Zoe Goodman, David Mariani, and Alain Rostain for aid in the experiment; K. M. Baird, R. L. Barger, J. L. Hall, T. W Hänsch, W. R. C. Rowley, and J. R. Whetstone for helpful discussions; and W. R. Bennett, Jr. for use of his laboratory.

References

[1] For an early review of the subject, see G. W. Series in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.) Spec. Publ. 343 (Aug. 1971), p. 73.

[2] T. W. Hänsch, M. H. Nayfeh, S. A. Lee, S. M. Curry, and I. S. Shahin, Phys. Rev. Lett. **32**, 1336 (1974); M. H. Nayfeh, Ph.D. Thesis, Stanford University, 1974 (M. L. Report 2318), pp. 20 and 85.

[3] B. W. Petley, K. Morris, and R. E. Shawyer, J. Phys. B. **13**, 3099 (1980).

[4] J. E. M. Goldsmith, E. W. Weber, and T. W. Hänsch, Phys. Rev. Lett. **41**, 1525 (1978); J. E. M. Goldsmith, E. W. Weber, F. V. Kowalski, and A. L. Schawlow, Appl. Opt. **18**, 1983 (1979). Their definition of the meter differs from Refs. [2, 3] and the recommendations of the CCDM. Our Table 4 and Fig. 6 give values for R_∞ which are corrected to a single, consistent standard for the meter. After our paper was delivered, Mr. van Dierendonck informed us of his recalculation of the results of Goldsmith *et al.*, based on new measurements listed in Ref. [34] and in F. Spieweck, IEEE Trans. Instrum. Meas. **IM-29**, 361 (1980). He gives R_∞ = 109 737.314 88 cm^{-1}.

[5] J. L. Hall, Science **202**, 147 (1978).

[6] T. Wik, W. R. Bennett, Jr., and W. Lichten, Phys. Rev. Lett. **40**, 1080 (1978). See Fig. 1 for the basic plan of the present experiment. See also S. R. Amin, C. D. Caldwell, and W. Lichten, Phys. Rev. Lett. **47**, 1234 (1981).

[7] J. B. West, R. L. Barger, and T. C. English, IEEE Trans. Quant. Electron **QE-11**, 18 (1975); R. L. Barger, T. C. English, and J. B. West, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, London, 1976), p. 565.

[8] D. Shiner and C. Wieman, these proceedings.

[9] E. W. Weber and co-workers, University of Heidelberg; B. Burghardt *et al.*, University of Bonn, these proceedings.

[10] K. C. Harvey, these proceedings.

[11] D. E. Roberts and E. N. Fortson, Phys. Rev. Lett. **31**, 1539 (1973).

[12] T. W. Hänsch, S. A. Lee, R. Wallenstein, and C. Wieman, Phys. Rev. Lett. **34**, 307 (1975).

[13] S. A. Lee, R. Wallenstein, and T. W. Hänsch, Phys. Rev. Lett. **35**, 1262 (1975).

[14] C. Wieman and T. W. Hänsch, Phys. Rev. A **22**, 192 (1980).

[15] J. N. Eckstein, A. I. Ferguson, and T. W. Hänsch, Phys. Rev. Lett. **40**, 847 (1978).

[16] T. W. Hänsch, these proceedings. For positronium, see S. Chu and A. P. Mills, Jr., Phys. Rev. Lett. **48**, 1333 (1982).

[17] V. Chebotayev, private communication.

[18] D. Kleppner, Bull. Am. Phys. Soc. **20**, 1458 (1975).

[19] See C. Fabre, P. Goy, and S. Haroche, J. Phys. B **10**, L183 (1977). See also Ref. [10].

[20] W. E. Lamb and R. C. Retherford, Phys. Rev. **79**, 549 (1950). See par. 5 for production of a polarized beam of atoms which, at low magnetic fields, becomes a beam in one hf level.

[21] If the non-Lorentzian parts are not Gaussian, we add the variances and set the non-Lorentzian width equal to 2.35 σ, as in the Gaussian distribution.

[22] J. T. Davies and J. M. Vaughan, Astrophys. J. **137**, 1302 (1963).

[23] W. G. Schweitzer, E. G. Kessler, R. D. Deslattes, H. P. Layer, and J. R. Whetstone, Appl. Opt. **12**, 2927 (1973).

[24] H. P. Layer, W. R. C. Rowley, and B. R. Marx, Opt. Lett. **6**, 188 (1981).

[25] H. Buisson and Ch. Fabry, J. Phys. Theor. Appl. **7**, 169 (1908); Ch. Fabry and H. Buisson, *ibid.*, p. 417; W. F. Meggers, Bull. Natl. Bur. Stand. (U.S.) **12**, 198 (1915–16); K. W. Meissner, J. Opt. Soc. Am. **31**, 405 (1941); D. H. Rank, E. R. Shull, J. M. Bennett, and T. A. Wiggins, *ibid.* **43**, 952 (1953); D. H. Rank and H. E. Bennett, *ibid.* **45**, 69 (1955); R. W. Stanley and K. L. Andrew, *ibid.* **54**, 625 (1964); R. L. Barger and J. L. Hall, Appl. Phys. Lett. **22**, 196 (1973).

[26] J. M. Bennett, J. Opt. Soc. Am. **54**, 612 (1964); C. J. Koester, J. Res. Natl. Bur. Stand. Sect. A: **64**, 191 (1960); D. H. Rank and H. E. Bennett, *op. cit.*; P. W. Baumeister and F. A. Jenkins, J. Opt. Soc. Am. **47**, 57 (1957); J. Bauer, Ann. Phys. (Leipzig) **20**, 181 (1934); C. F. Bruce and P. E. Ciddor, J. Opt. Soc. Am. **50**, 295 (1960); M. Born and E. Wolf, *Principles of Optics*, 2nd ed. (Pergamon Press, New York, 1975), pp. 52–59.

[27] S. R. Amin, Ph.D. Thesis, Yale University, Dec. 1982; submitted to J. Opt. Soc. Am.

[28] G. Breit and I. I. Rabi, Phys. Rev. **38**, 2082 (1931); N. F. Ramsey, *Molecular Beams* (Oxford University Press, London, 1956), p. 84.

[29] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, London, 1935), p. 243.

[30] Comité Consultatif pour la Définition du Mètre (CCDM), 5^e Session, 1973, Bureau International des Poids et Mesures (BIPM), p. M23.

[31] G. W. Erickson, J. Phys. Chem. Ref. Data **6**, 831 (1977).

[32] D. J. E. Knight, G. J. Edwards, P. R. Pearce, and N. R. Cross, IEEE Trans. Instrum. Meas. **IM-29**, 257 (1980).

[33] H. P. Layer, R. D. Deslattes, and W. G. Schweitzer, Jr., Appl. Opt. **15**, 734 (1976).

[34] For a discussion of comparisons of He-Ne lasers, see CCDM, 6^e Session, 1979, BIPM, especially CCDM/78–13.

[35] P. T. Woods, K. C. Shotton, and W. R. C. Rowley, Appl. Opt. **17**, 1048 (1978).

[36] B. N. Taylor, W. H. Parker, and D. N. Langenberg, Rev. Mod. Phys. **41**, 375 (1969).

[37] W. Lichten, to be submitted for publication.

[38] E. Englehard, Natl. Bur. Stand. (U.S.) Cir. 581, p. 13 (1957); H. A. Bowman, R. M. Schoonover, and C. L. Carroll, J. Res. Natl. Bur. Stand. Sect. A: **78**, 13 (1974); NBS Internal Report 75–768.

[39] R. S. Van Dyck and P. B. Schwinberg, these proceedings; Phys. Rev. Lett. **47**, 395 (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Current Work on Two Photon Excitation in a Hydrogen Beam for the Measurement of the Rydberg Constant and m_e/m_p*

D. Shiner and C. Wieman

Department of Physics, University of Michigan, Ann Arbor, MI 48109

We have observed the two quantum 2S to 3S transition in a beam of hydrogen in the metastable 2S state. The transition was excited by 6563 Å laser light plus a 315 MHz radio frequency field and is much narrower than the 2S to 2P single photon transition. The transition is detected by observing the increase in the number of metastable atoms which survive passage through the region containing the two fields. Work is underway to precisely measure the transition energy and the shift between hydrogen and deuterium. These results will provide more precise values for the Rydberg constant and the electron to proton mass ratio.

Key words: fundamental constants; hydrogen; laser spectroscopy.

1. Introduction

We report on the current status of an experiment to measure the 2S-3S energy separations in atomic hydrogen and deuterium to very high precision. From these measurements we expect to obtain values for the Rydberg constant and for the electron to proton mass ratio more precise than the current values, and to provide an improved test of the relativistic nuclear recoil correction to the energies of the hydrogenic states. To measure this energy separation we excite the 2S → 3S two quantum transition in a beam of 2S (metastable) hydrogen using a dye laser tuned to the 2S → 3P transition at 6563 Å and a radiofrequency field of ~300 MHz. This two quantum transition requires comparable laser power but has a natural linewidth 30 times narrower than the 2S → 3P one photon transition. Weber and Goldsmith [1] have detected this two photon transition using polarization spectroscopy in a Wood's discharge but the discharge environment resulted in linewidths only slightly less than the 3P natural width. We are presently observing the 2S → 3S transition and studying the line shape.

2. Background

The theory of double quantum transitions has been discussed by many authors, and Roberts and Fortson [2] consider our particular case of laser plus radiofrequency exciting the 2S → 3S transition in hydrogen. Although their treatment makes some simplifying assumptions (primarily low laser intensity) in solving the coupled 2S - 3P - 3S problem, these assumptions, and hence their results, match our experimental conditions with only slight changes.

From Eq. 1 (or pictorially in Fig. 2) of their paper one can see the key features of this resonance. First, when the rf is tuned to the 3S - 3P difference and the laser frequency varied, the transition rate into the 3S state has a sharp resonance which for low rf power has a width limited only by the 1 MHz 3S state natural linewidth. We have detected this process in our apparatus by observing the Lyman-α photons which are emitted when the 3S state decays via the 2P state. However, we have found there is a second way to observe the 2S → 3S resonance which, for technical reasons which will be discussed below, is far superior. This second approach uses the fact that along with the narrow rise in the 3S transition rate there is a corresponding decrease in the transition rate out of the 2S state. Although this may not be obvious intuitively it is quite evident in the results of Roberts and Fortson. This reduced transition rate means the 2S atoms will have a greater probability of passing through the interaction region without undergoing an excitation and subsequent decay to the ground state. One can therefore observe the two photon transition by seeing an increase in the number of metastable atoms detected downstream of the interaction region. When power broadening is included the exact line shape depends on the rf and laser intensities, and the length of the interaction region. To give a specific example however, to quench the beam off resonance to e^{-2} and on resonance to e^{-1} of its initial value, the powers required broaden the line to ~2 MHz.

3. Experimental Setup

The present experimental apparatus is shown schematically in Fig. 1. The metastable hydrogen source is similar in design and performance to that used by Lamb [3]. Molecular hydrogen is disassociated in a hot tungsten oven producing a beam of atomic hydrogen. This beam passes through an electron bombardment region where a small fraction is excited to the 2S state. This provides ~2×10^{10} H(2S)/sr·sec from a source slit 1 cm high and 0.04 cm wide. When the beam is collimated to give a residual Doppler width in the perpendicular direction of 1 MHz, about 10^4 H(2S)/sec remain. We have investigated the use of an rf disassociator for the molecular hydrogen, but have found that, at the present stage of the experiment, the increased Lyman-α background this technique introduces outweighs the advantages of a lower beam.

The H(2S) beam passes through an interaction region where there is an rf electric field on the order of 1V/cm and a dye laser beam with ~1 mW/cm^2 intensity. The rf field is produced in a parallel plate 50 Ω transmission

*This work supported by National Bureau of Standards Precision Measurements Grant #NB79NADA0018.

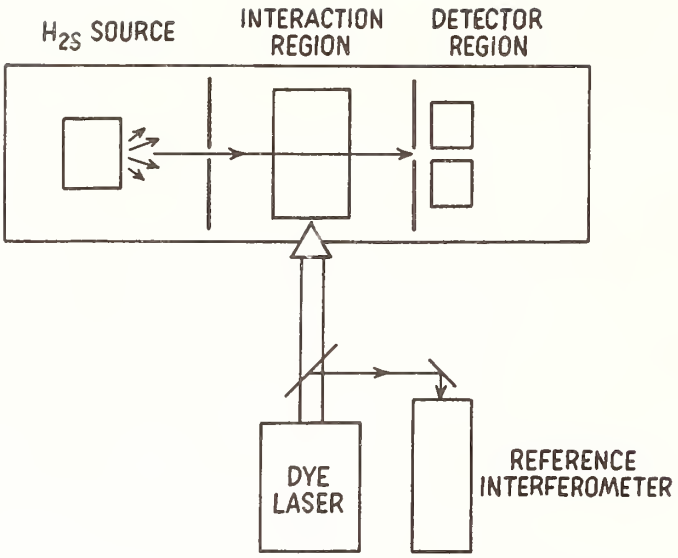


FIGURE 1. *Schematic of the apparatus.*

line. The laser beam is ~2 cm in diameter and passes through the central region of the rf field perpendicular to the atomic beam. The laser can be retroreflected to check for misalignments.

The metastable atoms remaining after the interaction region go through a narrow collimation slit and are then detected by quenching them with a dc electric field between two closely spaced Lyman-α counters. These subtend a solid angle of ~πsr each and see a negligible amount of background Lyman-α photons. These two factors make this method superior to the direct observation of the 3S decay. In that method the detected Lyman-α photons are emitted from a region of appreciable size and with the additional problem that the requirements on rf field homogeneity force the detectors to be a substantial distance from it. This situation results in small detection solid angle and substantial problems with Lyman-α background suppression.

The Lyman-α photon counters are photoionization detectors of our own design. They have 3 Torr of benzene vapor contained in a metal tube with a magnesium window on one side. We measure the current pulses on a wire down the center of the tube to which a bias voltage is applied. The gain is sufficiently high (~5×10^3) that the single photon pulses can be timed and counted with standard "Nim bin" electronics. At the present time the detectors have a quantum efficiency of about 15%, which is 30–40% lower than we expect. The difference is probably caused by contamination on the window and in the benzene.

The dye laser is a Spectra-Physics model 380A which is frequency locked to a stable Fabry-Perot interferometer. This provides a short term linewidth of about 1 MHz.

4. Present Results

In Fig. 2 we show typical spectra obtained when the laser frequency was not locked to the Fabry-Perot interferometer but was scanned over the $2S_{1/2} \rightarrow 3P_{1/2}$ single photon transition when no rf field was present. Note that the quantity being observed is the number of surviving 2S atoms and thus the resonance is a decrease in the count rate.

In Fig. 3, on an expanded frequency scale, we show a similar scan of the $2S_{1/2}(F=1) \rightarrow 2P_{1/2}(F=1,0)$ component with a 324 MHz rf field present. Although the two photon transition is clearly narrower than the 3P width it is substantially broader than the 4 MHz residual Doppler width set by the beam collimation. The additional broadening is due to the 10+ MHz frequency jitter of the laser as it is scanned. We have now locked the laser frequency to the interferometer and in future work will scan only the rf frequency to observe the transition. We anticipate this will provide significantly narrower lines.

5. Future Work

Our immediate effort is to make a detailed study of the two photon line shape and the various mechanisms which cause broadening and shifts. The most difficult shifts to deal with are those which depend on the atomic velocity. We believe these can be accurately determined however by appropriate chopping of the laser and atomic beams and timing the arrivals of the metastable atoms. This can be done quite well since the time response of the electron gun, the electroptic laser amplitude modulator, and the detector are all very short compared to the flight time of the atoms, and the time interval between the arrival of individual metastable atoms.

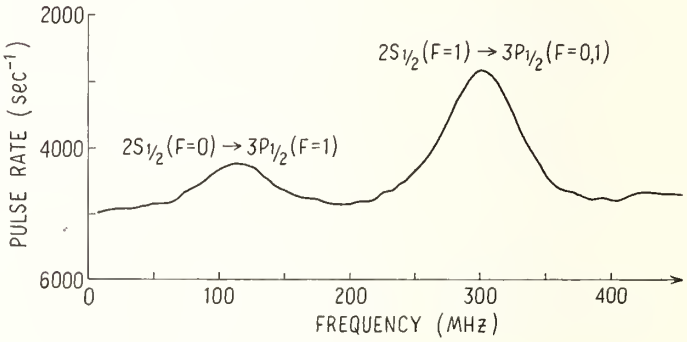


FIGURE 2. *Single photon transition spectrum.*

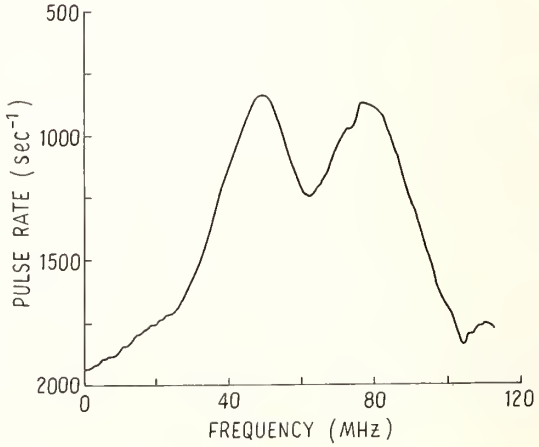


FIGURE 3. *Two photon $2S_{1/2}$ $(F=1) \rightarrow 3P_{1/2}$ $(F=0,1) \rightarrow 3S_{1/2}$ $(F=1)$ transition.*

Once these studies are completed we will measure a series of beat frequencies between two tuneable dye lasers to compare the $2S \rightarrow 3S$ energy with I_2 lines which are absolute wavelength references to obtain an improved value of the Rydberg constant. Also we will measure the energy difference between the hydrogen and deuterium resonances. These can be observed simultaneously in one apparatus. This will provide a greatly improved test of the relativistic nuclear recoil correction to the hydrogenic energy levels [4], and a measurement of the electron to proton mass ratio.

References

[1] E. Weber and J. Goldsmith, Phys. Rev. Lett. 41, 940 (1978).
[2] D. E. Roberts and E. N. Fortson, Phys. Rev. Lett. 31, 1539 (1973).
[3] W. Lamb and R. Retherford, Phys. Rev. 79, 549 (1950).
[4] C. Wieman and T. Hänsch, Phys. Rev. A 22, 192 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Measurement of the $2^2P_{3/2} - 2^2S_{1/2}$ Fine-Structure Interval in Atomic Hydrogen*

K. A. Safinya, K. K. Chan, S. R. Lundeen, and F. M. Pipkin**

Lyman Laboratory, Harvard University, Cambridge, MA 02138

A separated-microwave-oscillatory-field technique has been used in conjunction with a fast atomic-hydrogen beam to measure the $2^2P_{3/2} - 2^2S_{1/2}$ fine-structure interval in atomic hydrogen. The value obtained for the $2^2P_{3/2} - 2^2S_{1/2}$ fine-structure interval is 9911.117(41) MHz. This value is compared with other measurements of this fine structure interval and the potential precision obtainable in a more definitive measurement is discussed.

Key words: fast atomic beam; fine structure constant; fine structure of hydrogen; Lamb shift; separated-oscillatory-fields.

At present the accepted value for the fine structure constant is derived indirectly using the equation [1]

$$\alpha^{-2} = \frac{c}{4R_\infty}\left[\frac{\Omega_{NBS}}{\Omega}\right]^{-1}\left[\frac{\mu_p'}{\mu_B}\right]\frac{(2e/h)_{NBS}}{\gamma_p'(\text{low})_{NBS}}.$$

Here c is the speed of light in vacuum; R_∞, the Rydberg constant for infinite mass; Ω_{NBS}/Ω, the ratio of the National Bureau of Standards (NBS) as-maintained ohm to the absolute or SI (International System) ohm; μ_p'/μ_B, the magnetic moment of the proton in units of the Bohr magneton; $(2e/h)_{NBS}$, the ratio of twice the elementary charge to the Planck constant measured in terms of the NBS as-maintained volt using the ac Josephson effect; and $\gamma_p'(\text{low})_{NBS}$, the gyromagnetic ratio of the proton measured in terms of the NBS as-maintained ampere by the low-field method. The most recent value is [2]

$$\alpha^{-1} = 137.035963(15) \quad (0.11 \text{ ppm}).$$

An alternative value for the fine structure constant of comparable precision can be determined from the measurement of $g - 2$ for the electron with the assumption that the quantum electrodynamic calculation of $g - 2$ is correct. The value obtained in this manner is [2, 3][1]

$$\alpha^{-1} = 137.035993(10) \quad (0.073 \text{ ppm}).$$

These two values for α disagree by over 2 standard deviations. As an aid for resolving this discrepancy it would be valuable to have a third equally precise value for the fine structure constant which was free from theoretical ambiguities.

A theoretically unambiguous value for the fine-structure constant can be determined from the $2^2P_{3/2} - 2^2P_{1/2}$ fine-structure interval in atomic hydrogen [1]. This fine-structure interval can be calculated from independent measurements of the $2^2S_{1/2} - 2^2P_{1/2}$ (Lamb shift) and the $2^2P_{3/2} - 2^2S_{1/2}$ fine-structure intervals. At present, the uncertainty in the $P^2P_{3/2} - 2^2P_{1/2}$ fine-structure interval is due primarily to the 4 ppm error in the most precise measurement of the $2^2P_{3/2} - 2^2S_{1/2}$ interval and the disagreement between the several measurements. This limits the precision of the value of the fine-structure constant calculated from this interval to 2 ppm [1]. This paper reports the preliminary results of a new measurement of the $2^2P_{3/2} - 2^2S_{1/2}$ fine-structure interval with the use of a separated-microwave-oscillatory-field technique [4, 5] which is as precise as previous measurements and which gives promise of an order-of-magnitude increase in the precision of the determination of this fine-structure interval.

Figure 1(a) shows a schematic drawing of the apparatus. Fast hydrogen atoms in the $2^2S_{1/2}$ metastable state were formed through charge-exchange collisions in nitrogen gas from a 50- or 100-keV proton beam. The atoms passed through a state-selecting radio-frequency field (1100 MHz) set to quench atoms in the $2^2S_{1/2}$ ($F = 1$) state and entered the spectroscopy region. The number of atoms in the $2^2S_{1/2}$ ($F = 0$) state traversing the spectroscopy region was observed by quenching the atoms with a radio-frequency field (910 MHz) set at the central frequency for the $2^2S_{1/2}$ ($F = 0$) $\rightarrow$ $2^2P_{1/2}$ ($F = 1$) transition and observing the number of Lyman-α photons emitted.

Figure 1(b) shows a schematic drawing of the microwave apparatus used in the spectroscopy region to drive the transition. The microwave source was a klystron phase locked to the 25th harmonic of a frequency synthesizer. A traveling-wave-tube amplifier was employed to obtain enough power (700 mW) to drive the transition. A custom-made magic tee provided two equal components of the microwave field which were either in phase or 180° out of phase. A high-power three-pole switch altered the power between the 0°, off, and 180° configurations. The beam passed through 0.634-cm-diameter holes in the narrow side of the spectroscopy waveguides, which were separated by 2.721(3) cm. Tubes with a length of 0.953 cm soldered to each side of the guide acted as waveguides beyond cutoff and eliminated the leakage of microwave power. Two Schottky diodes in matched crystal holders were used to monitor the microwave power in each waveguide.

Figure 2 shows typical resonance curves for a 49 keV proton beam. The quenching signal, $\overline{Q}$, is the average percentage change in the surviving number of $2^2S_{1/2}$ ($F = 0$) atoms when the microwave field is turned on

*Research supported in part by National Science Foundation Grant No. 78-09657.

**Present address: SRI International, Palo Alto, CA 95205.

[1]Editors' note: See also the paper by T. Kinoshita and W. B. Lindquist in these proceedings.

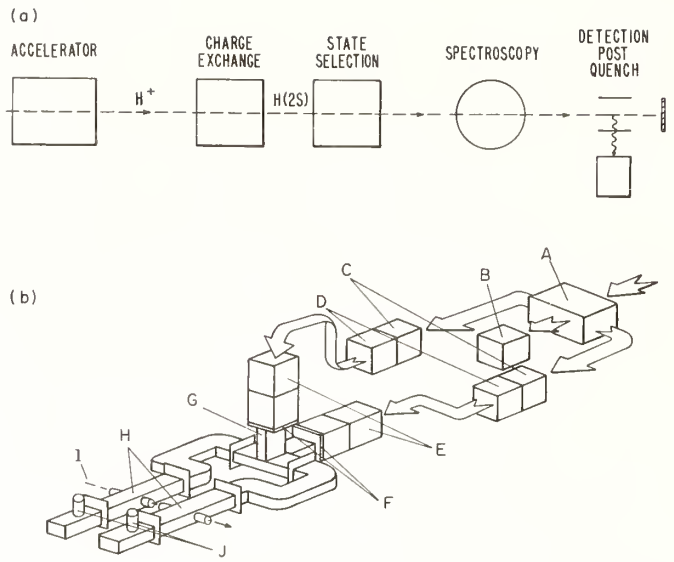


FIGURE 1. *(a) A schematic drawing of the apparatus. (b) A schematic drawing of the microwave plumbing in the spectroscopy region. A, three-pole switch; B, matched load; C, low-pass filters; D, precision variable attenuators; E, 50-dB circulators; F, vacuum windows; G, magic tee; H, interaction waveguides; I, beam; J, detector diodes.*

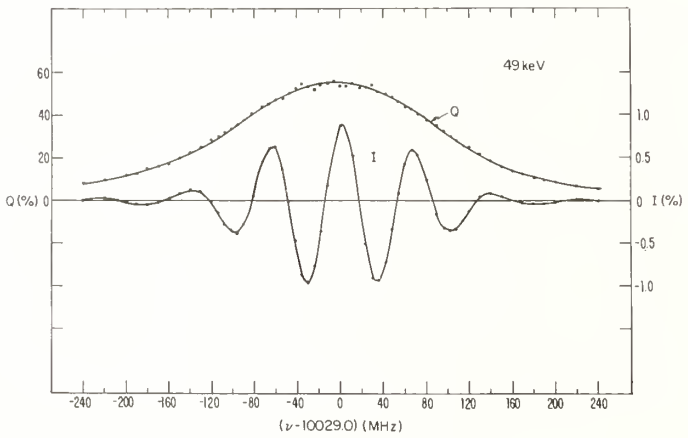


FIGURE 2. *Typical quenching ($\overline{Q}$) and interference (I) signals for a 40-keV beam. None of the corrections for asymmetries in the line shape have been applied to these data.*

with a relative phase of 0° and 180°; the interference signal, I, is the difference in the percentage change in the surviving number of $2^2S_{1/2}$ atoms when the microwave field is turned on with relative phases of 0° and 180°. The width of the interference signal is less than the 100 MHz natural linewidth due to the selection of long-lived P-state atoms by the separated oscillatory fields.

In order to cancel errors due to the Doppler shift and to a possible error in the phase between the two microwave fields at the location of the beam, data were taken in four configurations for each major run. The Doppler shift was cancelled by first making measurements with the microwaves propagating in one direction in the spectroscopy region and then reversing the feed so that the microwaves propagated in the opposite direction. The phase error was cancelled by interchanging the two waveguides and their driving components in the interaction region so that the time order in which the beam passed through the two interaction regions was inverted.

The method of symmetric points was used to determine the centers of the resonance curves. Table 1 summarizes the results of the measurements, the average corrections required to determine the $2^2P_{3/2} - 2^2S_{1/2}$ fine-structure interval, and the error in the corrections. The data at each point were corrected for the overlap of the residual neighboring transitions, the variation of the microwave power over the line, the incomplete $\overline{Q}$ subtraction, and the shift due to the motional electric field arising from the microwave magnetic field in the waveguide.

Prior to the experiment the two power measuring diodes and their mounts were calibrated with the use of a 10-dB attenuator and a Hewlett-Packard power meter. During the spectroscopy measurements the average power in the two diodes was set to be the same in both the 0° and 180° configurations. The voltages across each diode in both the 0° and 180° configurations were recorded for all data points. The sum of the two diode voltages was monitored with the computer and the data were corrected on line for the small power variation during each data point. Subsequent to taking data the diodes were recalibrated and the reflection coefficients of the components which could affect the microwave electric field at the beam were carefully measured. The new measurements showed that the microwave electric field decreased by 1.9(3)%/(100 MHz) over the region where measurements of the resonance line were made. The data were corrected for this variation in the microwave electric field. From computer simulations we estimated that

TABLE 1. *A summary of the measured line centers and the corrections required to determine the $2^2P_{3/2} - 2^2S_{1/2}$) fine-structure interval.*

Data Set	1	2	3
Beam energy (keV)	107.5(2.3)	48.7(1.0)	48.7(1.0)
Signal size (% peak to peak)	4.5	1.8	1.8
Raw center (MHz)	10028.628(6)	10029.054(14)	10028.966(16)
Corrections			
$F = 1$ overlap	−0.091(5)	+0.052(9)	+0.052(9)
IQS	−0.018(3)	−0.005(1)	−0.005(1)
v × B	−0.098(20)	−0.070(14)	−0.070(14)
Bloch-Siegert and rf Stark shift	−0.025(5)	−0.025(5)	−0.025(5)
Time dilation	+1.148(24)	+0.520(11)	+0.520(11)
rf-power slope	0.000(6)	0.000(6)	0.000(6)
$\nu(2^2P_{3/2}\ (F = 1) - 2^2S_{1/2}\ (F = 0))$ (MHz)	10029.544(33)	10029.526(26)	10029.438(27
Hyperfine structure (MHz)	−118.386	−118.386	−118.386
$\nu(2^2P_{3/2} - 2^2S_{1/2})$ (MHz)	9911.158(33)	9911.140(26)	9911.052(27)

because of the oscillatory shape of the interference curve a 1%/(100 MHz) error in the relative electric field would shift the center of the interference curve by less than 10 kHz.

Variations in the relative magnitudes of the microwave fields in the two waveguides between the 0° and 180° configurations can result in a spurious interference signal which persists even when the waveguides are very far apart. This signal, which is termed the incomplete $\overline{Q}$ subtraction (IQS), was determined by measurements made with a wide separation for both 50- and 100-keV beams. They were fitted to straight lines and used to correct the observed interference signals. The spurious signal had a slope of −0.22(4)%/(100 MHz) for the 100-keV data and a slope of −0.17(5)%/(100 MHz) for the 50-keV data.

The motional electric field due to the magnetic field in the waveguide produces a shift ($\mathbf{v} \times \mathbf{B}$) which has a weak dependence on the position of the resonance line. A program which simulated the resonance was used to calculate this shift. A simulation program was also used to calculate the Bloch-Siegert and rf Stark shifts for the $\overline{Q}$ and I signals. The second-order Doppler shift was calculated with use of the result of a separate experiment in which the first-order Doppler shift was employed to determine the velocity of the beam [6]. The primary source of the 2.1% error in the second-order Doppler shift was the error in the measurement of the magnetic field used to set the energy of the accelerator.

Since the three independent measurements of the $2^2P_{3/2} - 2^2S_{1/2}$ fine-structure summarized in Table 1 differ by more than the internally derived errors indicate they should, we have taken the final statistical error to be the standard deviation of the mean for the three measurements. The final error was obtained by combining in quadrature the statistical error (0.033 MHz) and the average systematic error (0.025 MHz). This procedure gave

$$\nu(2^2P_{3/2} - 2^2S_{1/2}) = 9911.117(41)\ \text{MHz}.$$

Figure 3 shows a comparison between this measurement, the results of previous measurements [7–9], and the theoretical value for this interval. The theoretical value was calculated using $\alpha^{-1} = 137.035963(15)$ from Ref. [2] and Mohr's calculation [10] of the Lamb shift with the radius of the proton [11] taken equal to $0.862(12) \times 10^{-13}$ cm. These calculations gave

$$\nu(2^2P_{3/2} - 2^2P_{1/2}) = 10969.044(2)\ \text{MHz},$$

$$\nu(2^2S_{1/2} - 2^2P_{1/2}) = 1057.884(13)\ \text{MHz},$$

and

$$\nu(2^2P_{3/2} - 2^2S_{1/2}) = 9911.160(13)\ \text{MHz}.$$

The use of Erickson's calculation [12] of the Lamb shift would give

$$\nu(2^2P_{3/2} - 2^2S_{1/3}) = 9911.114(10)\ \text{MHz}.$$

Mohr's value has been used in Fig. 3 since it is most favored by the recent measurements of the Lamb shift [5]. The new measurement reported here agrees with the theoretical value for this interval and is clearly competitive with the other measurements.

Table 1 shows that the primary limitation in the measurement of the $2^2P_{3/2} - 2^2S_{1/2}$ fine-structure interval is the systematic error and not the statistical error. Of the systematic errors the most troublesome is the variation in the relative phase of the microwave fields in the two waveguides when the apparatus is taken apart and reassembled. It is our conclusion from a recently reported measurement of the Lamb shift [5] that these systematic errors can be reduced through improvements in the apparatus, the procedure used to take data, and further theoretical studies of the resonance line shape. The replacement of the present detector, which consists of a single solar blind photomultiplier, by the recently developed large solid angle proportional counter for Lyman-α radiation [13] should increase the detection efficiency by a factor of 100 and enable us to work with resonance lines with a width of 15 MHz and the same signal to noise ratio as the resonance lines used in the preliminary measurement. This combination of these improvements should conservatively yield an order-of-magnitude increase in the precision of the determination of the fine-structure interval to the level of 0.4 ppm. This should provide a value of the fine-structure constant with a precision of 0.2 ppm or, alternatively, with the use of α determined from other sources, a measurement of the Lamb shift to 4 ppm.

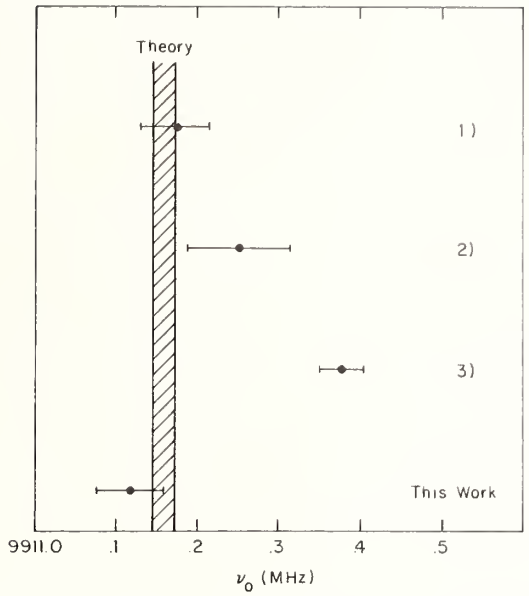


FIGURE 3. *Comparison of the theoretical and measured values for the $2^2P_{3/2} - 2^2S_{1/2}$ fine-structure interval in hydrogen. The points marked 1, 2, and 3 are the results from Refs. [7], [8], and [9], respectively. The origin of the theoretical value is described in the text.*

References

[1] For a general discussion of the determination of the fine structure constant, see B. N. Taylor, W. H. Parker, and B. N. Langenberg, Rev. Mod. Phys. 41, 375 (1969).

[2] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. 42, 1575 (1979).

[3] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Phys. Rev. Lett. 38, 310 (1977); and Bull. Am. Phys. Soc. 24, 758 (1979).

[4] For a report on the use of the separated-oscillatory-field technique to measure the Lamb shift in H(n = 3), see C. W. Fabjan and F. M. Pipkin, Phys. Rev. A 6, 556 (1972).

[5] For reports on the use of the separated-oscillatory-field technique to measure the Lamb shift in H(n = 2), see S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. 34, 1368 (1975); Phys. Rev. Lett. 46, 232 (1981); these proceedings.

[6] K. A. Safinya, Ph.D. thesis, Harvard University, 1979 (unpublished).

[7] B. L. Cosens and T. V. Vorburger, Phys. Rev. A 2, 16 (1970). [$\nu(2^2P_{3/2} - 2^2S_{1/2}) = 9911.173(42)$ MHz.]

[8] T. W. Shyn, T. Rebane, R. T. Robiscoe, and W. L. Williams, Phys. Rev. A 3, 116 (1971). [$\nu(2^2P_{3/2} - 2^2S_{1/2})$ = 9911.250(63) MHz.]
[9] S. L. Kaufman, W. E. Lamb, Jr., K. R. Lea, and M. Leventhal, Phys. Rev. A 4, 2128 (1971). [$\nu(2^2P_{3/2} - 2^2S_{1/2})$ = 9911.377(26) MHz.]
[10] P. J. Mohr, Phys. Rev. Lett. **34**, 1050 (1975).
[11] G. G. Simon, F. Borkowski, Ch. Schmitt, and V. W. Walther, Z. Naturforsh. **35A**, 1 (1980).
[12] C. W. Erickson, Phys. Rev. Lett. **27**, 780 (1971).
[13] J. J. Bollinger and F. M. Pipkin, Rev. Sci. Instrum. **52**, 938 (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Measurement of the Lamb Shift in Hydrogen, $n = 2$*

S. R. Lundeen and F. M. Pipkin

Lyman Laboratory of Physics, Harvard University, Cambridge, MA 02138

A measurement based on the fast atomic beam separated oscillatory field method of sub-natural linewidth spectroscopy gives for the Lamb shift in hydrogen: $S\,(n = 2) = 1057.845(9)$ MHz. The result is not in good agreement with theory.

Key words: fast atomic beam; hydrogen fine structure; Lamb shift; quantum electrodynamics; separated oscillatory fields.

Measurements of the Lamb shift interval ($2^2S_{1/2} - 2^2P_{1/2}$) in hydrogen and hydrogenic ions are among the most sensitive tests of QED at low energies, and in particular of QED's description of the electromagnetic vacuum state. The Dirac Equation, which describes the interaction of nucleus and electron relativistically, but in the absence of vacuum fluctuations, predicts incorrectly that this interval is zero. More sophisticated theoretical descriptions of these fundamental bound states reproduce the measured interval to at least 100 ppm. Improvements in the measurement precision beyond this level are difficult because the resonance linewidth is normally limited by the 100 MHz natural width of the $2P_{1/2}$ state which is about 10% of the Lamb shift interval. In the measurement described here [1], the method of separated oscillatory fields [2] is used to observe resonances that are narrowed considerably below this natural width. The increased resolution thus obtained, combined with the reduced sensitivity to systematic effects and the convenient availability of systematic checks lead to significant improvement in the experimental precision.

The fast atomic beam apparatus used for this measurement is shown schematically in Fig. 1. Hydrogen atoms in the $2^2S_{1/2}$ ($F = 0$) state are prepared by charge exchange from a fast proton beam, using radiofrequency hyperfine-state selection techniques. The beam passes through two separated coherent oscillatory fields and the surviving number of metastable atoms is monitored by observing the Lyman-α photons emitted in Stark induced decay to the ground state. The fractional quenching of the metastable beam by the separated oscillatory fields is measured for relative phases of 0° and 180° between the two fields. When averaged these two signals give a large broad signal ($\overline{Q}$) whose width is determined by the natural linewidth and the transit time through the rf plates. The difference of the two signals ($I = Q(0°) - Q(180°)$) gives the interference signal whose width is determined by the transit time between the two regions. Measurements are made in zero magnetic field and the resonance observed by changing the frequency of the applied rf field while keeping its amplitude constant.

Figure 2 shows a typical line profile for the $2^2S_{1/2}$ ($F = 0$) $- 2^2P_{1/2}$ ($F = 1$) transition. For each configuration, line profiles are taken with (a) the direction of the propagation of the rf field in the interaction region reversed, (b) the time order in which the beam passed through the interaction region reversed, and both (a) and (b). The average center for four such lines is free from shifts due to the Doppler effect and rf phase errors. The average of the four line scans is found to be very symmetric; the raw data shown no asymmetry larger than 0.1% of the peak to peak interference signal, all of which is accounted for by systematic effects. Thus the line centers can be determined by the method of symmetric points. As a check on systematic effects, measurements are made in eight different experiment configurations which differ in beam velocity, plate separation, and plate geometry. Table 1 lists the eight configurations together with the sizes and widths of the interference signals.

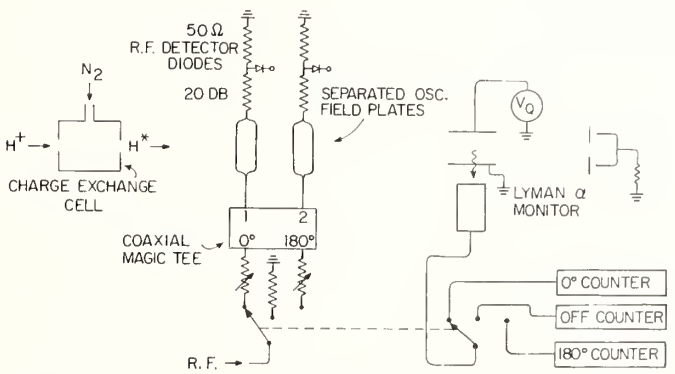


FIGURE 1. *Fast-beam, separated oscillatory field apparatus. The coaxial magic tee divides the rf power equally between the two separated rf regions with relative phase of 0° or 180°.*

The raw experimental line centers obtained in each configuration are shown in Table 1 with the associated statistical error (1 standard deviation). These numbers must be corrected for several small systematic effects, the total of which is shown in Table 1. The origin of the corrections, a more detailed tabulation of which is given in Ref. [1], is given below as well as an indication of the maximum systematic uncertainty (one standard deviation) which each produces in any of the eight configurations of Table 1.

*Research supported in part by National Science Foundation Grant No. 78-09657

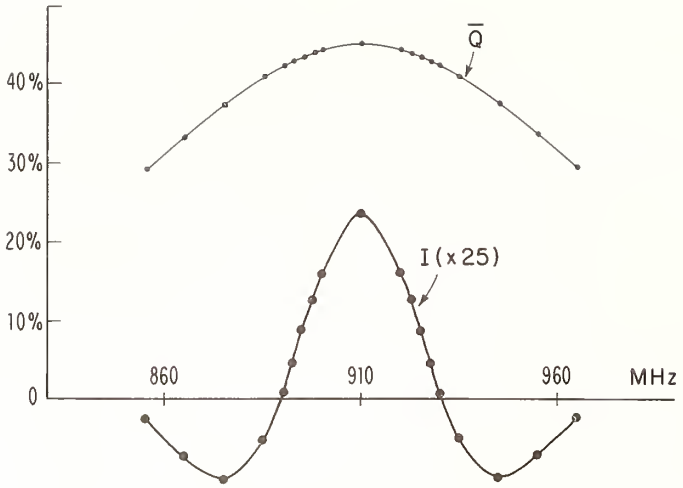


FIGURE 2. *Measured interference (I) and average quenching ($\bar{Q}$) signals for configuration 4. The radius of the plotted points is 10σ.*

1) Time dilation ($\leq \pm 0.004$ MHz): The resonance frequency should be measured in the moving atom's rest frame.
2) Bloch-Siegert and rf Stark shifts ($\leq \pm 0.004$ MHz): The shifts produced by the anti-resonant $2^2S_{1/2} - 2^2P_{1/2}$ and non-resonant $2^2S_{1/2} - 2^2P_{3/2}$ coupling. The net shift depends on plate spacing through a calculated geometric factor.
3) Plate coupling ($\leq \pm 0.002$ MHz): The impedance of the rf plates depends slightly on whether they are driven with 0° or 180° relative phase.
4) $F=1$ ($\leq \pm 0.001$ MHz): There are residual traces of the $2^2S_{1/2}$ ($F=1$) state.
5) Incomplete $\bar{Q}$ subtraction ($\leq \pm 0.013$ MHz): Systematic errors in the control of relative rf power between 0° and 180° states may produce a small spurious signal.
6) Off-axis distortion ($\leq \pm 0.004$ MHz): The non-uniform rf-field polarity seen by atoms traveling through the rf plates off axis may lead to small distortions.
7) rf-field slope ($\leq \pm 0.003$ MHz): Laboratory calibration of the rf electric field strength within the plates indicates that under experimental conditions it varies by $\Delta E/E = -0.08(7)\%/100$ MHz.
8) Additional field slope ($\leq \pm 0.010$ MHz): The need for this ad-hoc correction, equivalent to a variation in rf electric field strength by 0.5(2)%/100 MHz, is indicated by three independent lines of internal evidence. This may indicate that an error was made in calibrating the rf fields. On the other hand, it may indicate that some other systematic effect is present whose physical origin is unknown but which, in effect, mimics an error in the rf field calibration. Work is in progress to clarify the origin of this problem, but in the meantime the proper treatment of the data seems clear. A more detailed discussion of this correction is given in Ref. [1].

The sensitivity of the line center to individual systematic errors varies widely between the eight configurations. For instance, although correction eight results in a systematic uncertainty of ±0.010 MHz in configuration 1, it gives an uncertainty of only ±0.002 MHz in configuration 6. Several other potential sources of systematic error have been studied and found to be negligible. They are: overlapping resonances in high n states detected through cascades; stray electric and magnetic fields; and coherent population of S and P states.

The corrected line centers shown in Table 1, constitute eight independent determinations of the $2^2S_{1/2}$ ($F=0$) – $2^2P_{1/2}$ ($F=1$) interval, spanning a range of 32 ppm of the Lamb shift interval. To compute the final result, the results from 55 keV and 107 keV configurations are first combined in separate weighted averages. We assume that each systematic error is completely correlated between different configurations, but that different systematic effects are mutually uncorrelated. The results for the two beam energies are in excellent agreement and when averaged give:

$$\nu[2^2S_{1/2}\ (F=0) - 2^2P_{1/2}\ (F=1)] = 909.887(9)\ \text{MHz}$$

with a random error of ±0.004 MHz and a systematic error of ±0.008 MHz. Correcting for the known hyperfine structure of the $2^2S_{1/2}$ and $2^2P_{1/2}$ states, the Lamb shift interval is found to be

$$S\ (\text{H},\ n=2) = 909.887(9) + 147.958 = 1057.845(9)\ \text{MHz}.$$

TABLE 1. *A summary of the configurations in which data were taken. One independent result requires four line profiles.*

Config.	Beam Energy (keV)	Plate Spacing (cm)	Peak to Peak I Signal (%)	FWHM (MHz)	Number of Independent Results	Raw Center (MHz)	Total Systematic Correction (MHz)	Corrected Center (MHz)	Weight
1	55.1(1.1)	2.34	6.08	45	5	909.930(8)	−0.030(12)	909.900(15)	0.230
2	55.1(1.1)	3.30*	3.65	41	8	909.887(10)	−0.017(9)	909.870(13)	0.230
3	55.1(1.1)	3.30	2.95	38	4	909.893(6)	+0.008(6)	909.901(9)	0.344
4	55.1(1.1)	4.30	1.35	32	4	909.847(11)	+0.020(8)	909.867(14)	0.142
5	55.1(1.1)	5.08	0.66	29	4	909.844(18)	+0.027(14)	909.871(23)	0.054
6	106.8(2.1)	5.08	1.82	35	8	909.818(9)	+0.069(7)	909.887(11)	0.590
7	106.8(2.1)	6.62	0.64	28	6	909.807(11)	0.079(9)	909.886(14)	0.365
8	106.8(2.1)	7.60	0.37	25	2	909.788(38)	0.081(14)	909.869(40)	0.045

*For configuration 2, the inner diameter of the cylinder pipe surrounding the rf field region was increased from 3.81 cm to 5.08 cm.

The result, shown in Fig. 3, is in good agreement with previous measurements of Lamb, *et al.* [3, 4] and Robiscoe, *et al.* [4, 5] both of which were obtained using thermal atomic beams and Zeeman tuning for the Lamb shift resonance, and also with the more recent measurement of Newton, Andrews, and Unsworth [6], obtained with a fast atomic beam in zero magnetic field without the use of separated oscillatory fields.[1] The predictions of QED for the Lamb shift interval have been evaluated most precisely by Erickson [7] and Mohr [8]. When modified to reflect recent redeterminations of the proton rms charge radius (r_p = 0.862(12) fm) [9], their results are (in MHz):

Erickson: $S\,(n = 2) = 1057.930(10)$, $T\text{-}E = 0.085(13)$

Mohr : $S\,(n = 2) = 1057.884(13)$, $T\text{-}E = 0.039(16)$.

As illustrated in Fig. 3, neither result is in good agreement with experiment.[2]

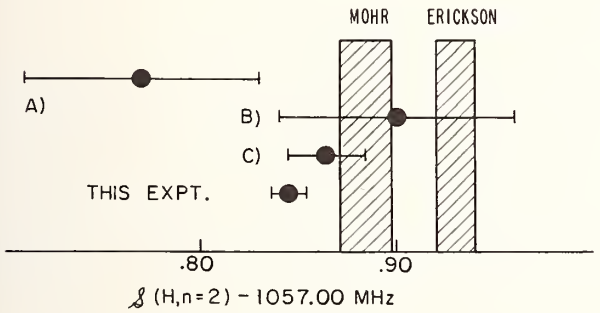


FIGURE 3. *Comparison of the result of this experiment with previous measurements A (Refs. [3, 4]), B (Refs. [4, 5]), C (Ref. [6]) and with theoretical calculations of Erickson [7] and Mohr [8]. The experimental error bars represent ±1 standard deviation, as do the widths of the vertical bars showing the theoretical results.*

Based on the results obtained to date, the prospects for further improvements in this measurement appear promising. The statistical signal to noise ratio can be improved by a factor 100 by more efficient detection of the metastable flux. This will allow precise studies of even narrower lines (FWHM $\lesssim$ 5 MHz) and better checks of systematic effects. At this time, except for the need to clarify the origin of correction 8, no fundamental systematic problems are yet evident. It appears reasonable, therefore, to hope for a measurement with a precision of ±0.001 MHz (1 ppm) using this technique. Full utilization of such a result would require a slight improvement in knowledge of the proton size to 0.4 percent and a considerable improvement in the precision of the QED calculations.

References

[1] S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. 46, 232 (1981).

[2] N. F. Ramsey, Molecular Beams (Oxford Univ. Press, London, 1956), p. 124.

[3] S. Triebwasser, E. S. Dayhoff, and W. E. Lamb, Jr., Phys. Rev. 89, 98 (1953).

[4] B. N. Taylor, W. H. Parker, and D. N. Langenberg, Rev. Mod. Phys. 41, 375 (1969).

[5] R. T. Robiscoe and T. W. Shyn, Phys. Rev. Lett. 24, 559 (1970).

[6] G. Newton, D. A. Andrews, and P. J. Unsworth, Philos. Trans. R. Soc. London, Ser. A:290, 373 (1979).

[7] G. W. Erickson, Phys. Rev. Lett. 27, 780 (1971).

[8] P. J. Mohr, Phys. Rev. Lett. 34, 1050 (1975).

[9] G. G. Simon, F. Borkowski, Ch. Schmitt, and V. W. Walther, Z. Naturforsch. Teil A:35, 1 (1980).

[1]Editors' note: See also the paper by Yu. L. Sokolov in these proceedings.

[2]Editors' note: Additional nuclear size corrections proposed by E. Borie (Phys. Rev. Lett. 47, 568 (1981)) would reduce the theoretical values of S by 40 ppm, bringing them into much better agreement with the experimental value reported here, but see also the comment by G. P. Lepage, D. R. Yennie, and G. W. Erickson, Phys. Rev. Lett. 47, 1640 (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Atomic Interferometer Method Measurement of the Lamb Shift in Hydrogen ($n = 2$)

Yu. L. Sokolov

Kurchatov Atomic Energy Institute, Moscow, U.S.S.R.

The frequency of the $(2S_{1/2}, F = 0) - (2P_{1/2}, F = 1)$ transition in the hydrogen atom has been measured by means of an atomic interferometer. The Lamb shift has been found to be $S = 1057.8583 \pm 0.0022$ MHz, where the uncertainty is the statistical standard deviation of a single observation.

Key words: atomic interferometer; atom velocity measurement; hydrogen atom; Lamb shift.

1. Introduction

The measurement of the Lamb shift S in the hydrogen atom is known to be one of the most important verifications of QED. However, in spite of the fact that more than a quarter of a century has been devoted to the measurement of S,[1] the increase in accuracy has proved to be rather modest. The spread in the experimental values of the Lamb shift measured by current methods is obviously too great to make a definitive comparison with the theory. As a matter of fact, the theoretical value of S cannot be calculated to an accuracy much better than 0.01 MHz. Thus, if a discrepancy is found to exist between S_{exp} and S_{theo}, it is unclear to what fact it should be ascribed—most likely to errors in the measurements and calculations, rather than to a failure in QED.

We performed a measurement of S by means of the atomic interferometer described in Refs. [2] and [3]. The results obtained indicate that we may expect a further increase in accuracy.

2. Experimental Setup

Figure 1 shows a schematic diagram of the "Pamir" unit which was used for the Lamb shift measurements. Protons with energies of about 20 keV emerging from the ion source (1) were passed through the velocity analyser consisting of a magnet (2) and a 0.02 cm wide slit diaphragm (3). The magnet dispersion measured in the plane of this diaphragm is 163 eV/cm; thus the energy spread of protons passing through the diaphragm does not exceed 3.25 eV, about 1.6×10^{-4} of their energy.

[1]The classic work by S. Triebwasser, E. S. Dayhoff, and W. E. Lamb [1] was performed in 1953.

Neutral hydrogen atoms were produced in the charge exchange chamber (4). The mixed beam was passed through a weak magnetic field (5), deflecting the proton component by an angle of about 0.2°; the proton current was measured by the Faraday cup (6). A parallel plate capacitor (7) was used for quenching $2S_{1/2}$ atoms, which was necessary for the background determination. The atom velocity was measured by means of system (8) which was essentially an ordinary interferometer with a longitudinal field [2] and a fixed inter-electrode distance (i.e., with a fixed time-of-flight). The $2P_{1/2}$ atoms produced under the action of the nonadiabatic field were recorded by a fixed detector (8a) and a detector (8b) which could be moved along the beam trajectory over a range of 1.5 cm. When not operating, the electrodes of system (8) were removed from the beam trajectory.

For monitoring the $2S_{1/2}$ atom flux a portion of the beam was transmitted through the pickup (9) consisting of a parallel plate capacitor where $2S_{1/2}$ atoms were quenched and a L_α detector. To eliminate the $2S_{1/2}$ state component with total angular momentum $F = 1$, the beam, as usual, was transmitted through rf fields with frequencies 1147 and 1087 MHz. The intensity of the $F = 0$ component was monitored by means of a pickup (11) analogous to pickup (9). The remaining part of the beam passed through the "double" interferometer (12) with the detector (13). Pickup (14) had a quenching field and was intended for observing the interference of the $2S_{1/2}$ state components formed by the nonadiabatic fields. The total neutral atom current was measured by the end detector (15).

In the experiments described, the neutral H atom beam length measured from charge exchange chamber (4) to the inlet slit of the interferometer (12) was 300 cm. Such a length was chosen so that the states with $n = 3$ –

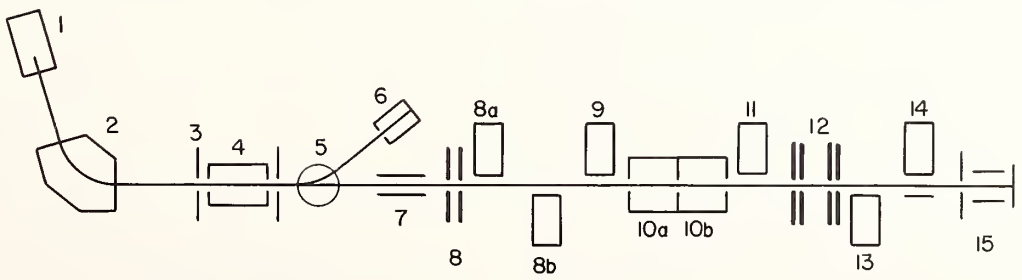


FIGURE 1. *Diagram of "Pamir" unit (notations in the text).*

7 which are populated rather densely and at the same time are comparatively long-lived had a sufficient time for de-excitation since their cascade transitions to the 1S ground state produce the L_α quantum background.

3. Ion Source

The main requirement imposed on the ion source was to obtain a thin (3–4 mm cross section) proton beam which was only weakly divergent and stable in time.

We developed a source which produced a two-component beam whose total current was 2.5–3.0 mA. This beam contained a comparatively weak central component with an angular divergence of about 10^{-3} and a current of 70–100 μA. The diverging part of the beam was stopped by the diaphragm and the "jet" component, without any focusing, was directed onto the inlet slit (3) of the charge exchange chamber.

Molecular hydrogen was chosen as the gas where proton charge exchange took place. Such a choice may seem not quite reasonable since the process of proton interaction with H_2-molecules can proceed in several channels, which will result in production of $2S_{1/2}$ atoms with different velocities. However as the collimation conditions are very stringent, a beam fraction with a divergence not exceeding 10^{-4} came into the interferometer. Such a selection by trajectories ensured a high uniformity in neutral atom velocities.

4. Background Estimation

Let us introduce the following notations:

m_1 — "natural" detector background;

m_2 — background produced by deexcitation of highly-excited atoms;

m_3 — background due to atom collisions with the residual gas;

m_4 — background due to interaction of the beam atoms with the edges of the interferometer slits;

m_5 — background arising because of penetration of L_α quanta to the detector through the outlet interferometer slit.

The main contribution to the total background is the m_2 component. As noted above, its intensity may be reduced if the observation point is put at an appropriate distance from the charge exchange chamber.

We calculated contributions to the m_2 component from the $(n\,S - 2P)$ and $(n\,D - 2P)$ transitions for $3 \leqslant n \leqslant 25$. It follows from the results obtained that the main part of the L_α background is due to the $(4S - 2P)$ and $(5S - 2P)$ transitions with lifetimes of 3.88×10^{-7} sec and 7.76×10^{-7} sec, respectively. The flux of L_α quanta produced by a beam section with length $\Delta x = 1$ cm (in the region 300 cm from the charge exchange chamber) reaches 5×10^{-6} of the neutral atom flux.

The control experiments showed that the measurements performed when the quenching field (7) is switched on give the same results with the non-adiabatic field switched either on or off, and, hence, such a field does not have a noticeable effect on highly excited atoms.

Estimation of the L_α quantum flux arising due to collisions of H_{1S} atoms with the residual gas (which is mainly hydrogen flowing from the charge exchange chamber) shows that this background component proves to be negligibly small (of the order 5 quanta/sec at the gas pressure $\sim 5 \times 10^{-8}$ Torr). Equally as small are the contributions from excited atoms and from $2S_{1/2}$ atoms because their flux is about 1.5% of the $1S_{1/2}$ and the cross section of $2S_{1/2}$ atom quenching in collisions is about 100 times smaller than that of the $H_{1S} + H_2^0 \rightarrow H_{2P} + R$ process.

The m_4 and m_5 background components are eliminated in adjustment of the device.

The total background value was determined in the experiment when no fields were on. In this case comparison of readings of the end meter (15) and detectors (9) and (13) allowed us to conclude that the background recorded is proportional to the neutral atom current within the limits of the attainable accuracy of L_α quantum flux measurements.

In the experiments on atomic state interference the value

$$I = \frac{N_1 - N_{1b}}{N_2 - N_{2b}}$$

was determined, where N_1 and N_2 are the readings of detector (13) and monitor (11) with quenching field (7) switched off and N_{1b} and N_{2b} are the background values measured, i.e. the readings of the same detectors with the quenching field switched on. For the sake of reducing the measurement time, in some experiments reading of the end meter (15), monitor (11), and detector (13) were recorded with the quenching field switched off, i.e. the values N_1 and N_2 were determined. The corresponding values N_{1b} and N_{2b} were found from readings of the end meter (15) calibrated in advance.

5. Interferometer

The analysis performed by V. P. Yakolev shows that the Lamb shift may be determined using a two-region interferometer consisting of electric fields I and II (see Fig. 2) changing nonadiabatically on the boundaries [3].

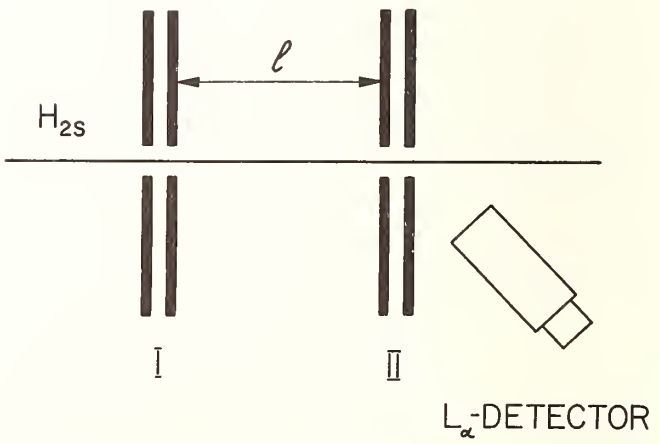


FIGURE 2. *Diagram of the "dual" atomic interferometer.*

When using such a device the interference pattern of the two phase-shifted 2P state components induced by fields I and II is described by the following formula:

$$F(\Delta l) = \left\{A\cos\left(\frac{\omega}{v}\sqrt{1-\frac{v^2}{c^2}}\,\Delta l + \alpha\right) + B\cos\left(\frac{\omega_1}{v}\sqrt{1-\frac{v^2}{c^2}}\,\Delta l + \beta\right)\right\} \times \exp\left[-\frac{\Delta l}{2l_o}\sqrt{1-\frac{v^2}{c^2}}\right], \quad (1)$$

where $l_o = v\tau$, v is the atom velocity, τ is the lifetime of the 2P atom; $\omega = 2\pi\nu$ is the $(2S_{1/2}, F = 0) - (2P_{1/2}, F = 1)$ transition frequency; $\omega_1 = 2\pi\nu_1$ is the fine struction splitting frequency; Δl is the change in the distance between the inlet I and outlet II systems of the interferometer. The A, B amplitudes and phases α, β depend on the condition in systems I and II and enter as parameters which are not calculated.

Equation (1) can be rewritten as:

$$F'(\Delta l) = A'\left[\cos\left(\frac{\omega}{v}\sqrt{1-\frac{v^2}{c^2}}\,\Delta l\right) + \frac{B'}{A'}\cos\frac{\omega_1}{v}\sqrt{1-\frac{v^2}{c^2}}\,\Delta l\right] \times \exp\left[-\frac{\Delta l}{2l_o}\sqrt{1-\frac{v^2}{c^2}}\right]. \quad (2)$$

Here A' is the scale factor which is of no interest; thus in fitting the theoretical curve to the experimental points the values ω/v and B'/A' are determined.

If the field strengths in systems I and II are about 250 V/cm then B'/A' will be equal to about 0.1. Thus curve $F'(\Delta l)$ will represent a superposition of two cosinusoids modulated by the attenuating exponent; the second cosinusoid has a frequency a factor of 10 higher and an amplitude a factor of 100 lower than the first cosinusoid, and thus produces small, fast oscillations about the latter.

In the experiments the L_α quantum yield recorded by detector (13) was measured as a function of the time-of-flight T between systems I and II. The strengths of the nonadiabatic fields E_1 and E_2 were held constant during the measurement.

A change in the time-of-flight T was achieved by changing the distance l between systems I and II. For this purpose the inlet system I was displaced relative to the fixed system II over a range of 0–2 cm by means of a precision mechanism. The displacement Δl was read (or monitored) by a precise scale rigidly connected to system I. Readings were made by means of a microscope provided with an eyepiece micrometer having a double hair line. The reading accuracy was 1×10^{-5} cm.

In determining the $I_{2P}(T)_{E_1, E_2}$ dependence, the following procedure was adopted:

1. At given field strengths in systems I and II and a given distance l the number of L_α quanta from the 2P state decay is determined by detector (13) after atoms have passed through the interferometer, i.e., the value $I_{2P}(T)_{E_1, E_2} = I(\Delta l)$ is found.

2. The direction of the field in system II is reversed, with other conditions remaining constant. The measurement gives the value $I_{2P}(T)_{E_1, E_2} = \bar{I}(\Delta l)$.

3. Difference $I - \bar{I}$ is found.

4. The field E_2 is returned to the initial value. Then the distance l is changed by Δl and two more measurements are made with different directions of the field E_2.

5. The procedure is repeated for a chosen series of increments Δl. If a certain L_o (its value is of no importance) is taken as the initial distance between the systems, the increments Δl may be both positive and negative. The dependence $F(\Delta l) = I(\Delta l) - \bar{I}(\Delta l)$ obtained in this way is described by the theoretical formula Eq. (1).

Prior to the initiation of the measurements, with the distance $l = 0$ and the outlet electrode of system II grounded, an appropriate strength of the field E_1 in system I was determined, corresponding to the interference maximum. Then, with system II on, the field E_2 was similarly chosen. The strengths of both fields were of the order of 300 V/cm. Under this condition the number of L_α quanta recorded by detector (13) in the interference maximum for $l = 0$, was about 5×10^5 per second. The measurements were carried out for $0 \leq l \leq 2$ cm at 41 points, i.e., for $\Delta l = 0.05$ cm. Determination of the ω/v and B'/A' values in Eq. (2) was done by a least squares fitting procedure.

To check the state of the setup, its resolution relative to the Lamb shift was periodically determined using the method described in Ref. [2].

6. Measurement of $2S_{1/2}$-Atom Velocity

It follows from Eq. (2) that for determination of the Lamb frequency S the $2S_{1/2}$ atoms velocity must be measured by an independent method. Stabilization and measurement of the velocity are the most complicated parts of this experiment and the source of the main limitations of the method.

The $2S_{1/2}$ atoms' velocity can be determined, in principle, by transmitting them through the quenching electric field and observing the resulting L_α radiation at a small angle with respect to the beam trajectory. Then the atom velocity can be found by the Doppler shift of the L_α line. This method was tested on the "Pamir" setup with a one meter spherical diffraction grating having 1200 lines/mm. However, the results obtained revealed that such a method does not permit one to determine the velocity with the required accuracy (not better than 10^{-5}), mainly because of the low aperture ratio of the device.

In all the Lamb-shift-experiments the atom velocity was measured by observation of de-excitation of $2P_{1/2}$ atoms produced under the action of the non-adiabatic field.

When the detector is moved along the beam trajectory, the L_α radiation intensity will vary as $I = I_o \exp(-x/l_o)$, where x is the distance measured from an arbitrary reference point, $l_o = v\tau$ where v is the atom velocity, and τ is the 2P atom lifetime. Thus, the velocity (more accurately the value $1/l_o$) can be determined by the slope of the straight line $\ln(I) = \text{const} - (x/l_o)$. The corresponding measurement system must include two detectors one of which (monitor) is fixed and the other moves along the beam trajectory.

As has been said, in the experiments on the Lamb shift the determination of the velocity was performed by means of system (8) which has a detector (8b) with a displacement mechanism similar to that of the interferometer detector (13). For determination of the atom velocity the movable system I was also used in combination with detector (13). In this case the system II electrodes were grounded.

At first glance it might appear that determination of

the value l_o using the method under consideration can be made by measuring the intensity at two points of the beam trajectory, divided by a distance x_o, which is a typical procedure in the case of determination of an absorption coefficient. However, such a procedure is only valid when it is known beforehand that the intensity of radiation recorded decreases exponentially with x and that the experimental parameters do not vary with time. As a consequence the "two point" method proves to be invalid, in principle, for a measurement of the beam atoms' velocity since a supposition of the exponential nature of the $I_{2P}(x)$ dependence is accepted *a priori* and, generally speaking, may be invalid either because of experimental defects (e.g., velocity variations) or because of the influence of some unknown factors. Thus the analysis of the experimental data must include verification of the linearity of the function studied.

In addition, if the intensity is recorded only at two points of the trajectory, an error in the measurement of the distance x_o appears as a systematic error. When, on the other hand, the measurement is performed at many points, errors of this type become random, which permits one to make a statistical treatment of the data using the least squares method.

The following procedure was adopted in accordance with the considered formulation of the problem: The L_α radiation intensity was measured at several points of the beam trajectory, i.e., for a number of discrete values $x_i (i = 1), 2, 3, \ldots . k)$, n separate measurements being performed within equal time intervals at each i-th point. Then for the set of values $y_{i,m} = \ln(I_{i,m})$, $(m = 1, 2 \ldots n)$ corresponding to the theoretical dependence $Y = \alpha + \beta x$ where $Y = \ln I$, $\alpha = \ln I_o$ and $\beta = 1/l_o$, the empirical dependence (regression line) $y = \alpha + \beta x$ was found. Calculation of its parameters allows one to find l_o and determine its standard error.

Measurements of the I_{2P} intensity were made at 26 points for $\Delta x = 0.04$ cm at the 1 cm long trajectory section; 20 measurements were performed at each point. The number of counts at the first point ($i = 1$) being at a distance of 0.03 cm from the outlet electrode of system (8) was about 6×10^4 pulses/sec. In the control experiments, in addition to the statistical error Δl_o, the resolution of the setup (considered as an instrument for atom velocity measuring) was determined by the method described in Ref. [2]. In an experiment of this type the difference in the velocities of H_{1S} and H_{2S} atoms produced in the charge exchange chamber was determined. The measured difference in l_o values corresponds to a difference in energies of about 10 eV, which is indicative, first, of a sufficiently high resolution of the method and, second, of the circumstance that in the beam one monoenergetic group of $2S_{1/2}$ atoms is present.

In the last version of the "Pamir" device the atom velocity is stabilized to a sufficient degree after a 3 hour warm up of the entire system. If no failures occur in the operation of the apparatus, the velocity stays constant for about 2.5 hours and then begins to decrease slowly. The atom velocity changes somewhat after the apparatus is turned on again.

7. Results

For two years about 200 Lamb shift measurements were carried out. However, it followed from the data obtained that only in 34 cases the atom velocity remained constant within permissible limits when the $I_{2P}(T)_{E_1, E_2}$ dependence was determined (the velocity was measured in the beginning and in the end of the experiment).

Figure 3 shows the histogram for the ν values. The blocks shaded represent the ν values obtained in treating the measurements according to the criterion of selection by velocity constancy. They form a compact group with a mean value of

$$\nu = 909.9003 \pm 0.0022 \text{ MHz}$$

The error is equal to the standard deviation of a single observation and represents statistical uncertainty only. The Lamb shift

$$S = 1057.8583 \pm 0.0022 \text{ MHz}$$

corresponds to this ν.

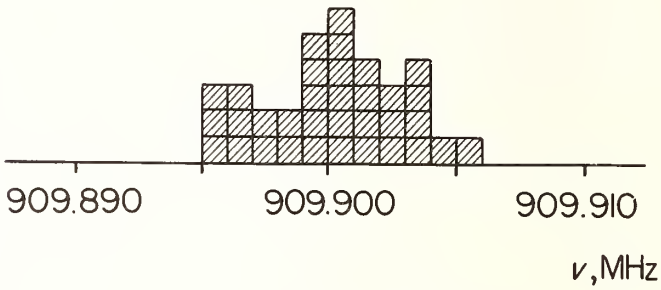


FIGURE 3. *The histogram of ν values.*

The value of S given here is somewhat different from that reported earlier [4]. This is due to the fact that by the time the present paper had been written, V. P. Yakovlev improved the value of the 2P atom lifetime ($\tau = 1.596185 \times 10^{-9}$ sec was found).

Thus, after an appropriate correction was introduced, S was found to be in a satisfactory agreement with the experimental values 1057.862(20) and 1057.845(9) MHz obtained in Refs. [5] and [6], respectively. However, the value found differs from the theoretical data available [7, 8], which was also pointed out in Ref. [6].[2] But it seems one need not lend a great importance to this discrepancy since the differences between the theoretical and experimental S values do not definitely go beyond the error limits. Therefore one cannot speak with confidence about the existence of any difference between S_{exp} and $S_{theo.}$. It appears that further improvement of the S measurement and calculation accuracy is needed.

The general impression the author has after many years of work with the atomic interferometer is that such an increase in the accuracy is quite feasible. The method must be improved first of all by replacement of the l-measuring arrangements consisting of scales and microscopes with precision transducers which transform the linear displacement of movable elements into electrical signals expressed in numerical code.

The author is deeply grateful to V. P. Yakovlev for his constant help and numerous discussions.

References

[1] S. Triebwasser, E. S. Dayhoff, and W. E. Lamb, Jr., Phys. Rev. 89, 98 (1953).

[2] Editor's note: see also footnote number 2 to the paper by Lundeen and Pipkin in these proceedings.

[2] Yu. L. Sokolov, Zh. Eksp. Teor. Fiz. 63, 461 (1972) [Sov. Phys. JETP 36, 243 (1973).]
[3] Yu. L. Sokolov, in *Atomic Physics 6*, Ed. by R. Damburg (Plenum Press, New York, 1978), p. 207.
[4] Yu. L. Sokolov, Program and Abstracts of this conference, p. 63.
[5] G. Newton, D. A. Andrews, and P. J. Unsworth, Philos. Trans. R. Soc. London, Ser. A: **290**, 373 (1979).
[6] S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. 46, 232 (1981); these proceedings.
[7] G. W. Erickson, Phys. Rev. Lett. 27, 780 (1971).
[8] P. J. Mohr, Phys. Rev. Lett. 34, 1050 (1975).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Measurement of the $4^2S_{1/2} - 4^2P_{1/2}$ Lamb Shift in He^+*

J. J. Bollinger, S. R. Lundeen, and F. M. Pipkin

Lyman Laboratory, Harvard University, Cambridge, MA 02138

The separated oscillatory field technique is being used with a fast 120 keV He^+ beam and a miniature microwave spectroscopy region to make a zero magnetic field measurement of the $n = 4$ Lamb shift interval in He^+. The $4^2S_{1/2}$ state population is monitored by observing its decay to the 2P state with a large solid angle VUV photoionization detector. With 10 minutes of counting time, the signal to noise ratio is sufficient to make a 100 ppm measurement of the line center. It is believed that with this signal to noise ratio the systematics can be well enough understood to determine the line center better than previous He^+ Lamb shift measurements and test the Erickson and Mohr calculations of the Lamb shift.

Key words: cascades; He^+ Lamb shift; large solid angle VUV detector; overlapping transitions; separated oscillatory fields.

1. Introduction

Measurements of the Lamb shift of hydrogenic ions provide sensitive low energy tests of quantum electrodynamics. The sensitivity of these measurements is determined both by the accuracy of the experimental measurements and by the accuracy of the theoretical calculations. In hydrogen, a major source of uncertainty is the proton rms charge radius. Because the uncertainty in the α particle rms charge radius [1] is about a factor of five less than the uncertainty in the proton rms charge radius [2] ($\Delta r/r = 0.03\%$ and 1.4% respectively), measurements of the Lamb shift intervals in $^4He^+$ provide, potentially, some of the best tests of quantum electrodynamics in hydrogenic ions. In addition, the absence of hyperfine structure in $^4He^+$ provides an experimental simplification over Lamb shift measurements in H. These factors have motivated us to make a Lamb shift measurement in $^4He^+$ using the fast beam, separated oscillatory field technique (SOF) developed in this laboratory [3]. The $n = 4$ $^2S_{1/2} - {}^2P_{1/2}$ Lamb shift has been chosen as an experimentally convenient Lamb shift to measure with this technique and our present 120 keV fast beam.

Figure 1 shows all direct sub 250 ppm Lamb shift measurements in $^4He^+$ together with two recent calculations. Comparisons between measurements of the Lamb shift intervals $n^2S_{1/2} - n^2P_{1/2}$ for different n are made by normalizing all measurements to the $n = 2$ frequency scale. The overall agreement between theory and experiment is not good. The precision of these measurements (approximately 100 ppm) is more than a factor of nine worse than the current best Lamb shift measurements in hydrogen [3, 10]. Significant improvements are needed in the precision of He^+ Lamb shift measurements to reduce the experimental uncertainty to the level of uncertainty in the theoretical calculations and take full advantage of the precise knowledge of the $^4He^+$ nuclear charge radius.

Although the Mader, Leventhal, and Lamb measurement [6] in Fig. 1 is the only sub 250 ppm Lamb shift measurement involving a non-metastable ($n \geq 3$) $^2S_{1/2}$ state in He^+, many others have been made much less precisely. A major systematic problem common to most of the $n \geq 3$ experiments has been contributions to the Lamb shift resonance signal due to overlapping transitions in higher n states. For this or any experiment involving a Lamb shift measurement with $n \geq 3$ to be successful, contributions due to overlapping transitions in higher n states must be identified and taken into account.

2. Description of the Experiment

Figure 2 gives a sketch of the basic experimental layout. Excited He^+ atoms are created by collision by pass-

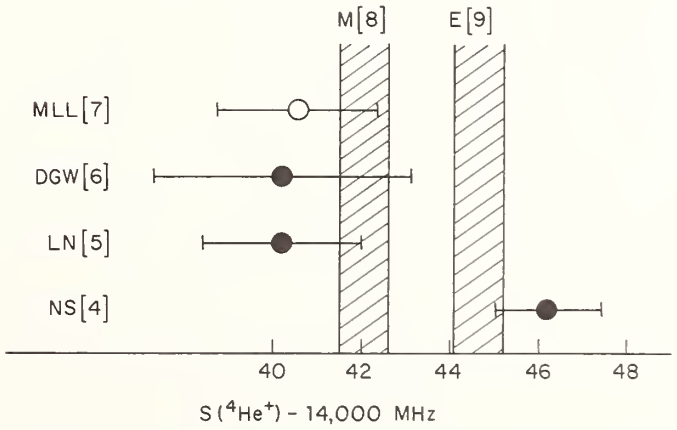


FIGURE 1. *Previous measurements of the Lamb shifts in $^4He^+$ [4–7] with two calculations by Mohr [8] and Erickson [9]. The dark circles are $n = 2$ measurements. The open circle is an $n = 3$ measurement normalized to the $n = 2$ frequency scale by multiplying by the ratio of Erickson's values for the $n = 2$ and $n = 3$ $^4He^+$ Lamb shifts.*

*Research supported in part by National Science Foundation Grant No. 78-09657.

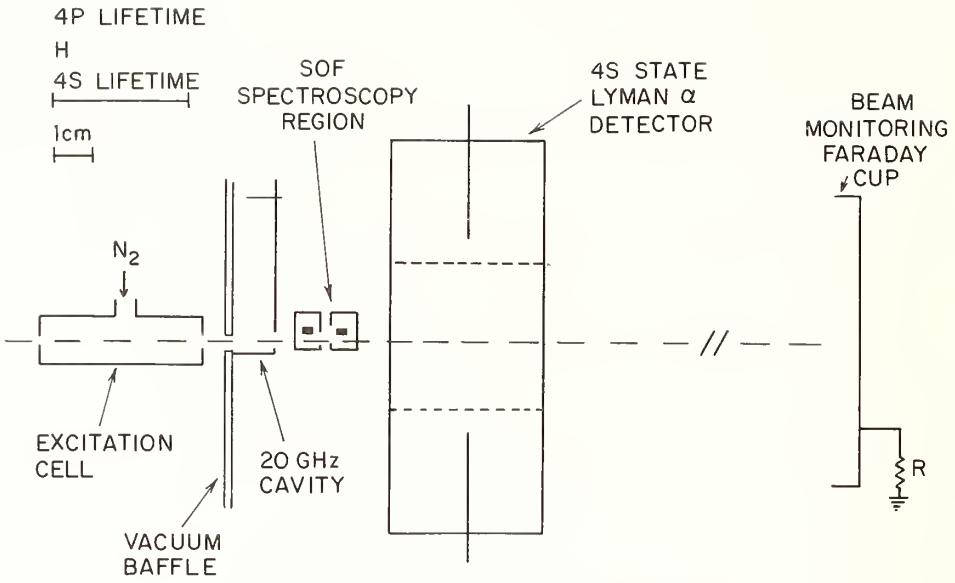


FIGURE 2. *Sketch of the experimental arrangement.*

ing a fast (120 keV, 2.4 × 10^8 cm/sec) He^+ beam through a short, 4 cm cell at 10–40 mTorr pressure. Due to their lifetime, 4^2P He^+ atoms will decay almost immediately, lasting on the average less than 2 mm down the beam line. After passing through a vacuum baffle isolating the excitation cell chamber from the rest of the vacuum system, the He^+ atoms enter several microwave spectroscopy regions. The second region is the SOF region consisting of identical, 0.7 cm wide, rectangular 50Ω coaxial transmission lines separated by a small distance (typically 0.00, 0.32 or 0.64 cm). After the microwave spectroscopy regions, the $4^2S_{1/2}$ state population is monitored by observing its decay (121.5 nm) to the 2P state with a large solid angle vacuum ultraviolet photoionization detector.

Applying microwave power near the $n = 4^2S_{1/2} - {}^2P_{1/2}$ resonant frequency to the SOF region transfers population from the $4^2S_{1/2}$ state to the short lived $4^2P_{1/2}$ state, resulting in a decrease of the $4^2S_{1/2}$ state population and the output of the photodetector. For a fixed phase difference of the two SOF regions the quenching signal of the $4^2S_{1/2}$ atoms is defined by the percent charge in the output of the photodetector when microwave power is applied. The width of the quenching signal is determined by the natural linewidth and the transit time through one of the rectangular transmission lines. The difference in the quenching signals when the two SOF regions are driven in phase and 180° out of phase is the interference signal. The width of the interference signal is determined by the transit time between the two SOF regions.

In addition to the SOF spectroscopy region described above, a fixed 20.18 GHz spectroscopy region consisting of a K-band waveguide cavity is located just before the SOF region. Applying resonant microwave power to this cavity quenches the $4^2S_{1/2}$ state via the $4^2S_{1/2} - 4^2P_{3/2}$ transition (see Fig. 3) and permits the observation of any contributions due to overlapping transitions in higher n states. These contributions are subtracted from the unquenched Lamb shift signal.

Data are taken in zero magnetic field and as a function of the frequency of the microwave power applied to the SOF region across a 300 MHz bandwidth centered on the 1769 MHz Lamb shift transition. In addition, to cancel first order Doppler shifts and possible phase errors between the two SOF transmission lines, data are taken with the direction of propagation of the microwave field in the SOF region reversed, and with the time order in which the beam passes through the SOF transmission lines reversed. The average of these four interchanges will be free of first order Doppler shifts and phase errors between the two SOF transmission lines. After any asymmetries in the lineshape are understood and corrected for, the method of symmetric points is used to calculate the line center.

The large solid angle Lyman α (121.5 nm) detector consists of six NO photoionization chambers with rectangular MgF_2 windows which totally surround the beam (see Fig. 4). The detector housing is made out of aluminum and consists of an inner and outer cylinder, two end flanges, six perforated chamber separations, and a filling tube, all aluminum dip brazed together into one piece. The electrons from the NO photoionization are collected on 0.76 mm diameter probes extending from the rear of each chamber and separated from the chamber housing by ceramic insulators and a metal guard ring to reduce leakage currents. The Lyman α detector is operated in the proportional mode with the aluminum housing at negative voltage. The average fractional solid angle subtended by the photodetector over its 3.5 cm length is greater than 0.55. The combined detection efficiency of the 2 mm MgF_2 window transmission and the NO photoionization quantum efficiency is estimated to be 35%. Consequently the overall detection efficiency for Lyman α photons is approximately 20% over the 3.5 cm length of the detector.

3. Discussion

Figure 5 shows some data taken with the plates separated by 0.32 cm. The contribution to the unquenched Lamb shift signal due to the overlapping transitions in higher n states is approximately 10% of the Lamb shift signal height. The dominant overlapping tran-

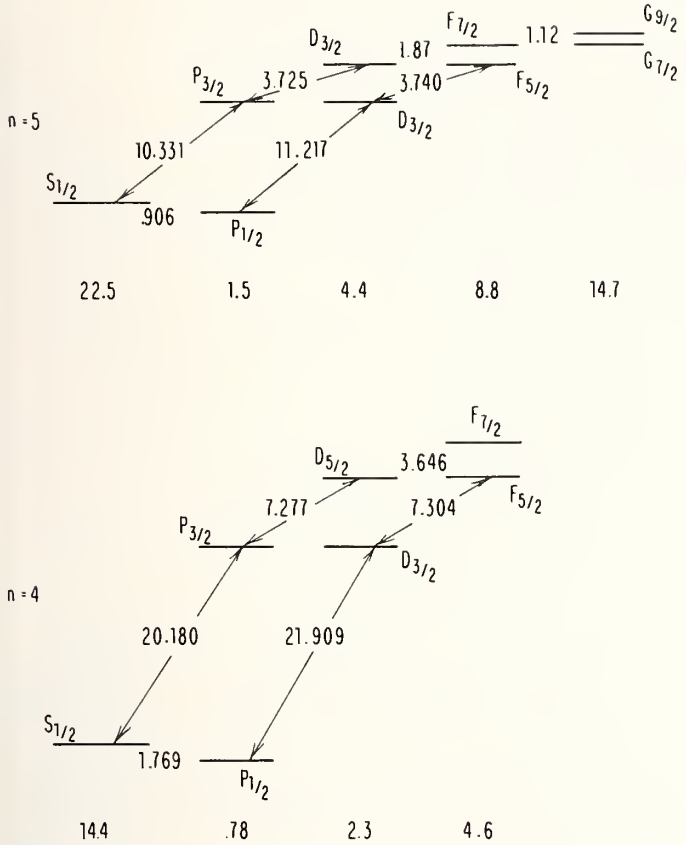


FIGURE 3. *Fine structure intervals (GHz) and lifetimes (nsec) in the $n = 4$ and $n = 5$ manifolds of $^4He^+$. The $n = 5$ $(J = 5/2) - (J = 7/2)$ transitions at 1.87 GHz overlap the $n = 4$ 1.769 GHz Lamb shift.*

sitions which appear in the Lamb shift resonance signal are the $n = 5$ $(J = 5/2) - (J = 7/2)$ electric dipole transitions at 1.87 GHz (see Fig. 3). Subtracting this contribution from the unquenched Lamb shift signal, shifts the resonance line center by approximately 2 MHz down in frequency. If a 100 kHz (50 ppm) measurement is to be made, the subtraction technique to get rid of contributions due to overlapping transitions must work very well. To test how well this technique is working, data have been taken at different spacings of the SOF regions and at different levels of microwave power applied to the SOF region. Changing either one of these parameters changes the relative sizes of the signals due to Lamb shift and overlapping transitions and will therefore help in determining how well the subtraction technique is working.

In addition to contributions due to overlapping transitions, the calibration of the microwave power applied to the atoms in the SOF region as a function of frequency will be a major systematic problem. From past work done in this laboratory, it is believed that this calibration

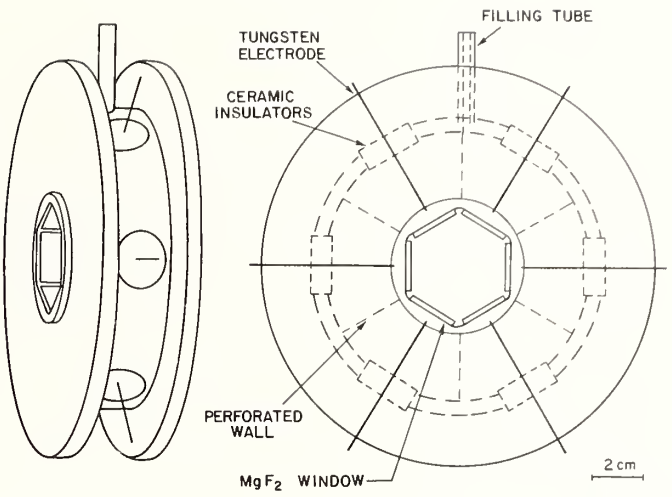


FIGURE 4. *Isometric and cross sectional sketches of the large solid angle Lyman-α detector.*

can be done well enough to permit at least a 50 ppm measurement of the line center.

Over 200 hours of data-taking have been completed. With the large solid angle detector and short geometry of this experiment, good signal to noise has been obtained. With 10 minutes of counting time, the signal to noise ratio is sufficient to make a 100 ppm measurement of the corrected (with the overlapping transition contributions subtracted out) line center. With this signal to noise ratio, we believe the systematics can be well enough understood to determine the line center better than previous He^+ Lamb shift measurements and experimentally distinguish the Erickson [9] and Mohr [8] calculations of the Lamb shift.

References

[1] A. Bertin *et al.*, Phys. Lett. 55B, 411 (1975); G. Carboni *et al.*, Nuovo Cimento A 34, 493 (1976).

[2] G. G. Simon, F. Borkowski, Ch. Schmitt, and V. W. Walther, Z. Naturforsch. Teil A: 35, 1 (1980).

[3] S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. 34, 1368 (1975); S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. 46, 232 (1981), and these proceedings.

[4] M. A. Narasimham and R. L. Strombotne, Phys. Rev. A 4, 14 (1971) [$S(n = 2) = 14\,046.2 \pm 1.2$ MHz].

[5] E. Lipworth and R. Novick, Phys. Rev. 108, 1434 (1957) [$S(n = 2) = 14\,040.2 \pm 1.8$ MHz].

[6] G. W. F. Drake, S. P. Goldman, and A. Van Wijngaarden, Phys. Rev. A 20, 1299 (1979) [$S(n = 2) = 14\,040.2 \pm 2.9$ MHz].

[7] D. L. Mader, M. Leventhal, and W. E. Lamb, Jr., Phys. Rev. A 3, 1832 (1971) [$S(n = 3) = 4183.17 \pm 0.54$ MHz].

[8] P. J. Mohr, Phys. Rev. Lett. 34, 1050 (1975), and in *Beam-Foil Spectroscopy*, Ed. by I. A. Sellin and D. J. Pegg (Plenum Press, New York, 1976), p. 89 [$S(n = 2) = 14\,042.05 \pm 0.55$ MHz].

[9] G. W. Erickson, J. Phys. Chem. Ref. Data 6, 831 (1977) [$S(n = 2) = 14\,044.64 \pm 0.54$ MHz].

[10] G. Newton, D. A. Andrews, and P. J. Unsworth, Philos. Trans. R. Soc. London, Ser. A: 290, 373 (1979).

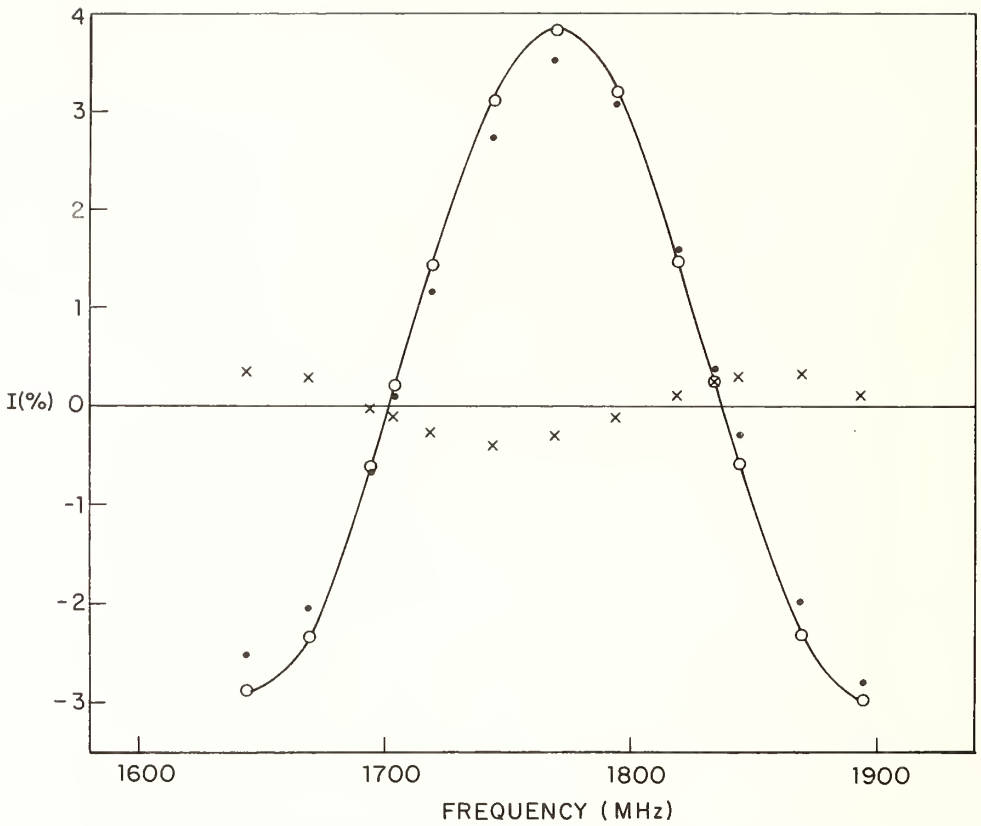


FIGURE 5. *Typical raw interference (I) signal for a 120 keV beam and a SOF separation of 0.32 cm. The solid dots are taken with no 20 GHz quenching of the 4S states. The x's are taken with 20 GHz quenching of the 4S states, and the hollow dots are the differences of the two.*

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Lamb Shift in the Hydrogenic Ion Cl^{16+}

E. T. Nelson, O. R. Wood II, C. K. N. Patel, M. Leventhal, and D. E. Murnick

Bell Laboratories, Murray Hill, NJ 07974

and

H. W. Kugel* and Y. Niv**

Rutgers University, Piscataway, NJ 08854

The Cl^{16+} $n = 2$ Lamb shift has been measured by resonant laser quenching of a fast ($v/c \sim 0.1$) metastable beam. The result for the $2S_{1/2} - 2P_{1/2}$ splitting is 31.19(22)THz in agreement with the calculation of Mohr and with the series expansion in powers of $(Z\alpha)$.

Key words: high Z hydrogenic ions; Lamb shift; laser resonance spectroscopy.

1. Introduction

Studies of the $n = 2$ Lamb shift, S, in high Z hydrogenic systems are important tests of the high field limits of QED [1]. In addition, these measurements offer a sensitive test of modern calculational procedures. Since these calculations are based on expansions in α and $Z\alpha$, the effects of higher order terms play an increasingly important role in the high Z hydrogenics. By scaling any potential discrepancy between experiment and theory with Z, it is in principle possible to identify which terms may be in error. The best detailed calculations have been done independently by Erickson [2] and by Mohr [3] and differ by 2% at $Z = 17$.

The experimental method is a modification of the classic Lamb-Retherford experiment [4], in which an optical frequency electromagnetic field probes the $2S_{1/2}$–$2P_{1/2}$ energy level splitting, and has been described in a recent review [5].

2. Experiment

A schematic diagram of the experimental apparatus is shown in Fig. 1. A 190 MeV chlorine beam from the Brookhaven National Laboratory double tandem van de Graaff accelerator facility was fully stripped prior to entering a switching magnet to the Lamb shift resonance

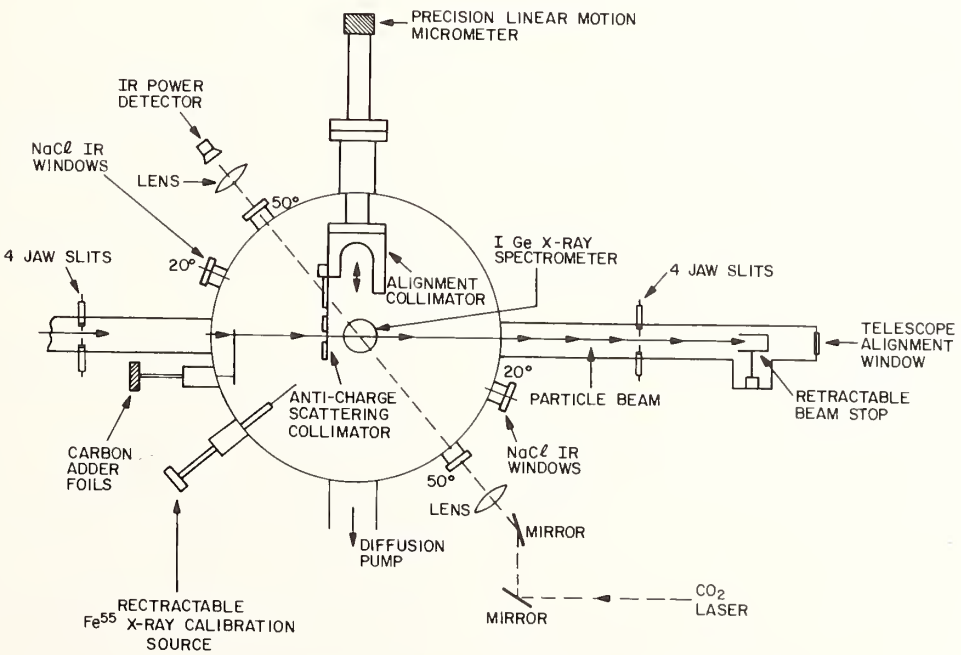


FIGURE 1. *Schematic diagram of Cl^{16+} experimental system.*

*Present address: Princeton Plasma Physics Laboratory; Princeton, NJ.
**Dr. Chaim Weizmann Fellow; Resident Visitor at Bell Laboratories.

beam line. A thin adder foil was then used to preferentially populate $2S_{1/2}$ Cl^{16+} ions by electron pickup. Of order 0.1% of the total beam was promoted to this state resulting in an estimated metastable beam current of 400 pA. As the metastable state pickup was probably a surface interaction, and lower states of ionization and scattered ions were the major source of background radiation, ultrathin carbon foils (0.5 to 5.0 $\mu g/cm^2$) were used. (The x-ray spectrum observed with 300 mm^2 × 7 mm thick planar intrinsic Ge detectors consisted of two photon E1 and M1 radiation from the 2S state, Lyman α radiation from the 2P levels, and radiation from n = 2 Cl^{15+} as well as analogous transitions from other ionization states formed in the foil and by the scattered beam.) A 152 μm Be window absorbed 85% of the dominant two photon decay from the Cl^{16+} $2S_{1/2}$ state. Typical total count rates were 70 kHz per detector.

The CO_2 laser consisted of a grating tuned oscillator with pulsed electrical excitation and cavity Q-switching that produced 25 mJ, 150 μsec pulses (10.6 μm, P(20)) at a repetition rate of 480 Hz. The oscillator output was mode-matched into a 13 m folded amplifier with a saturated gain of 14. The amplified laser beam had good spatial coherence, and was focused with $f/3.8$ KCl optics to a 150 μm spot (FWHM) coincident with the metastable beam. Frequency tuning was accomplished by the Doppler effect and line tuning of the laser across the 9 μm and 10 μm bands between 28.3 THz and 32.33 THz.

A retractable alignment collimator with position reproducibility of ±25 μm was used to precisely define the interaction volume and align the system so that both particle and laser beams were focused to the same point in space with a well defined angle, between the x-ray detectors. The collimator system had to be removed during measurement as any particle interactions caused unacceptably high x-ray backgrounds, and even the edges of the focused laser beam had sufficient power to remove material from the metal collimator. In separate experiments the particle beam waist was determined to be 0.4 mm and the laser beam waist to be 0.15 mm.

The laser induced fluorescence was monitored by detection of enhanced Lyman α radiation at 2.95 keV from the interaction volume within a 200 μsec interval coincident with the laser pulse. Typical laser induced count rates of 20 Hz were observed at the resonance peak. The accurate recovery of this small excess count rate within the 140 kHz background was achieved by using 20 MHz scalars and a 2 MHz clock to subtract the counts accumulated during the time the laser was off (scaled by the laser duty cycle) from the counts accumulated during the time the laser was on. No excess counts were observed when the laser beam was blocked or displaced so as to not intersect the Cl^{16+} particle beam. To enhance the signal to noise ratio the laser parameters were varied to maximize an expression proportional to the laser induced count rate divided by the fluctuation in the background counts accumulated during the laser pulse:

$$\frac{P}{\sqrt{f\,\Delta t}}\,, \tag{1}$$

where P is the average laser power, f is the pulse repetition rate, and Δt is the laser pulse duration. The detectors were also collimated to view only the interaction volume so as to reduce the background rate. The discrete tuning range of the laser was limited so that about one half-width of the resonance curve could be probed at fixed intersection angle and beam energy. An angle of 130° was chosen for obtaining most data in order to have highest sensitivity to line shape and center.

3. Line Shape Theory and Curve Fitting

For the case of $n = 2$, $Z = 17$, and $I = 3/2$, one finds that the hyperfine splittings involved are of the order of two percent of the natural linewidths of the transition. Since it is known on general grounds that this effect does not shift the center of gravity of the $2S_{1/2} - 2P_{1/2}$ transition the hyperfine interaction does not play a significant role in the lineshape. In the laboratory the particle beam makes an angle θ with respect to the laser beam. Atoms in the $2S_{1/2}$ state are driven to the $2P_{1/2}$ state by the laser and the subsequent Lyman α decay x-ray is detected. The basic induced transition probability per unit line time μ is given by the Wigner-Weisskopff theory. In the rest frame of the atom:

$$\mu = \frac{e^2\,\Gamma\,S_m}{2\,\pi\,c\,\hbar^2}\,\frac{|r|^2}{[(\nu_m - \nu_0)^2 + (\frac{\Gamma}{4\,\pi})^2]}\,, \tag{2}$$

where $|r|^2$ is the relevant matrix element squared, S_m is the power density in the moving frame, (i.e., the atom's rest frame), ν_m is the laser frequency in the moving frame, Γ is the linewidth, and $\nu_0 \equiv S$.

The quantity μ is a function of θ because both S_m and ν_m are functions of θ.

From the Doppler shift formalism,

$$\nu_m = \frac{\nu_L(1 - v/c\,\cos\theta)}{\sqrt{1 - v^2/c^2}}\,, \tag{3}$$

where ν_L is the laboratory laser frequency. An expression relating the laboratory laser power density S_L with S_m can be obtained from the Lorentizan transformation of the field components,

$$S_m = S_L\,\gamma^2(1 - \beta\cos\theta)^2\,,$$
$$\beta \equiv v/c\,. \tag{4}$$

Data is obtained at each laser frequency normalized to integrated particle current and average laser power. Frequencies are changed in random order approximately every minute to average out any long term shifts or variations in the system. The signal is proportional to the product of laser power density, interaction time, and fraction of particle beam intercepted by the laser. As the focused laser beam is smaller than the particle beam this product is independent of small variations in laser spot size.

Since data is obtained at fixed angle θ, very far from saturation of the transition, the signal expression becomes to a good approximation a simple Lorentizan with the natural linewidth. The best fit parameters are the Lamb shift and an amplitude parameter.

A typical set of data resulting from 8 h of measurement, the fitted resonance curve, the final experimental result for the Lamb shift from all the data, and the calculated values are shown in Fig. 2. About 50 h of data were obtained, most of it at a beam energy of 190 MeV and θ = 130°. In order to check for systematic errors some data were obtained at a different energy (150 MeV) and different interaction angles (θ = 50° and 160°). Also the foil to interaction volume distance was changed. The results of these measurements were consistent with the results obtained using the standard parameter values.

The final result for the Lamb shift was obtained by making an overall fit to all the data, with a single center frequency ν_0 as a free parameter, together with free am-

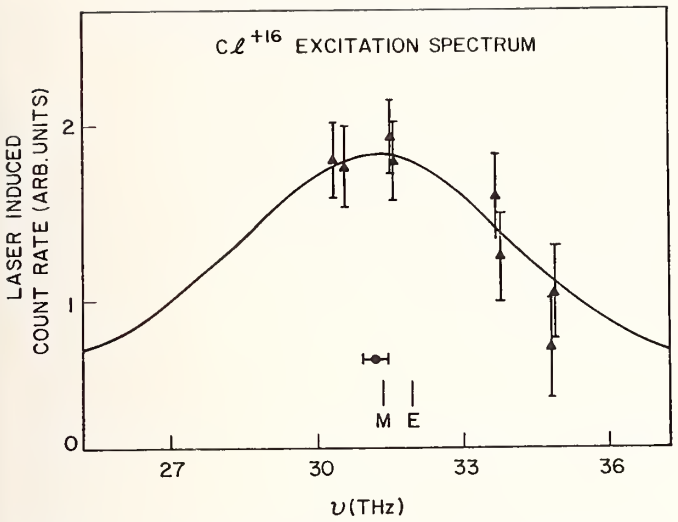


FIGURE 2. *Cl^{16+} excitation spectrum. The circular point with error bar indicates the final experimental value for the Lamb shift. M and E indicate values calculated by Mohr [3] and Erickson [2].*

plitude parameters for each individual resonance curve. This fit yielded a value for the Lamb shift of ν_0^{exp} = 31.19(22) THz with χ^2 = 29 for 41 degrees of freedom. The error stated is the statistical one standard deviation. Employing a separate fit for each curve alone and averaging the results thus obtained yields a similar frequency ν_0^{exp}, while estimating the error in the mean from the scatter of the different results around the mean gives the same 0.7% relative error.

A number of identifiable systematic effects have been considered in detail. These include neglect of hyperfine structure as noted above along with neglect of effects due to the $2P_{3/2}$ level. Also possible errors in ion beam velocity, intersection angle, and particle beam-laser beam overlap have been considered. We estimate that the resultant error from all these effects is less than 0.03 THz. Hence the experimental error is determined predominantly by statistics.

4. Conclusion

The experimental result for the Lamb shift for the Z = 17 hydrogenic system Cl^{16+} is in agreement with the calculation of Mohr and with the series expansion in powers of $(Z\alpha)$; and is three standard deviations below the calculation of Erickson. Scaling the discrepancy recently reported for the Z = 1 Lamb shift [6] with an order higher than Z^5 is in disagreement with the experimental result and indicates that theoretical corrections of the order of $(\alpha Z)^{14}$ must be considered possibly due to proton size, structure, or relativistic recoil effects.

References

[1] S. J. Brodsky and P. J. Mohr, in *Structure and Collisions of Ions and Atoms*, Ed. by I. A. Sellin (Springer, Berlin, 1978), p. 3.
[2] G. W. Erickson, Phys. Rev. Lett. **27**, 780 (1971).
[3] P. J. Mohr, Phys. Rev. Lett. **34**, 1050 (1975).
[4] W. E. Lamb and R. C. Retherford, Phys. Rev. **79**, 549 (1950).
[5] H. W. Kugel and D. E. Murnick, Rep. Prog. Phys. **40**, 297 (1977).
[6] S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. **46**, 232 (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Helium Fine Structure and the Fine Structure Constant

W. Frieze, E. A. Hinds, A. Kponou, and V. W. Hughes

Yale University, New Haven, CT 06520

and

F. M. J. Pichanick

University of Massachusetts, Amherst, MA 01003

A series of precision measurements have been made of the fine structure interval $\nu_{JJ'}$ in the 2^3P state of helium. The results are ν_{01} = 29 616.864(36) MHz (1.2 ppm), ν_{12} = 2 291.196(5) MHz (2.2 ppm), and ν_{02} = 31 908.040(20) MHz (0.6 ppm). These results are self-consistent and in good agreement with theory. Experiment and theory can be combined to produce an independent value for the fine structure constant α^{-1} = 137.036 13(11) (0.8 ppm), in good agreement with the more accurate value currently accepted α^{-1} = 137.035 963(15) (0.11 ppm).

Key words: fine structure constant; helium fine structure; optical-microwave spectroscopy; quantum electrodynamics.

1. Introduction

One of the most important tests of quantum electrodynamics is in the study of simple atomic systems, and in this respect experimental and theoretical work on the 2^3P fine structure of helium has been particularly fruitful. In particular the helium 2^3P fine structure is more favorable for precise measurement than its counterpart in the 2^2P state of atomic hydrogen. The helium 2^3P state has a relatively long lifetime, 1.0×10^{-7} s, compared to 1.6×10^{-9} s for the 2^2P state of hydrogen. This gives for the $2^3P_0 - 2^3P_1$ and $2^3P_1 - 2^3P_2$ intervals respectively, a ratio of fine-structure interval to natural linewidth of 85 and 7 times the ratio for the $2^2P_{1/2} - 2^2P_{3/2}$ interval for hydrogen. The helium fine structure measurements currently provide an independent measurement of the fine structure constant with a precision of 0.8 ppm, where the accuracy is limited by the theory. Although this is the most accurate value available from atomic spectroscopy it is about a factor of eight less accurate than values obtained from the ac Josephson Effect [1] and the anomalous g-factor of the electron [2].

2. Experimental Technique

The energy level diagram of helium is given in Fig. 1 with an inset showing the 2^3P fine structure to an expanded scale. The 2^3P state is inverted, the lowest value of J lying highest, because the spin-other-orbit interaction makes a larger contribution than does the spin-self-orbit. In addition the Landé interval rule is not even obeyed approximately because the spin-spin interaction is comparable to the spin-orbit.

Early measurements were made using optical spectroscopy [3] with a precision of about 1%. The first precise experiment (150 ppm) utilizing magnetic dipole transitions between 2^3P fine structure levels was performed [4] by exciting the levels by electron bombardment and observing changes in intensity and polarization of the $2^3P - 2^3S$ optical radiation. The precision was improved to about 10 ppm by level-crossing measurements [5], and in

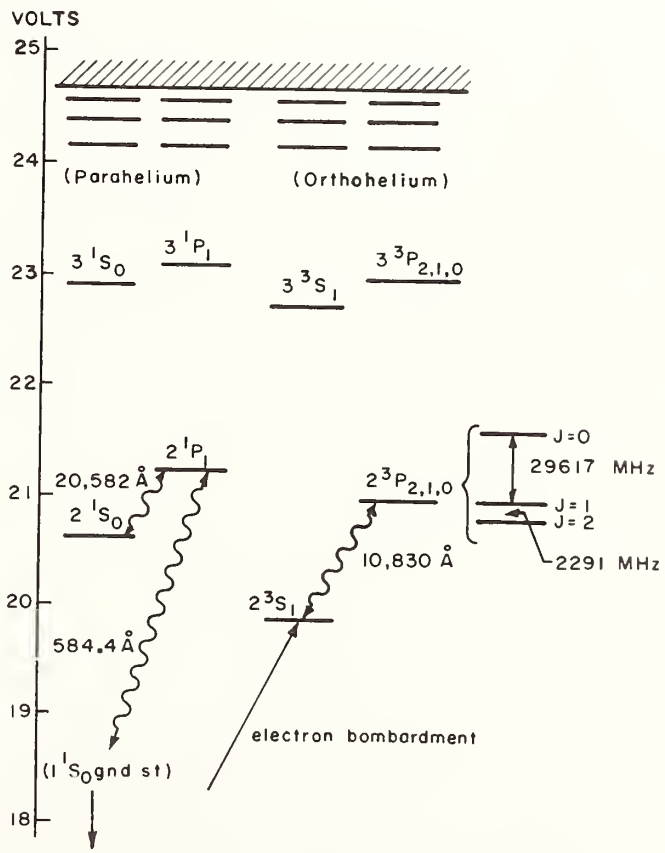


FIGURE 1. *Energy level diagram of helium, with the 1^1S ground state suppressed. The inset shows the 2^3P fine structure on an expanded scale.*

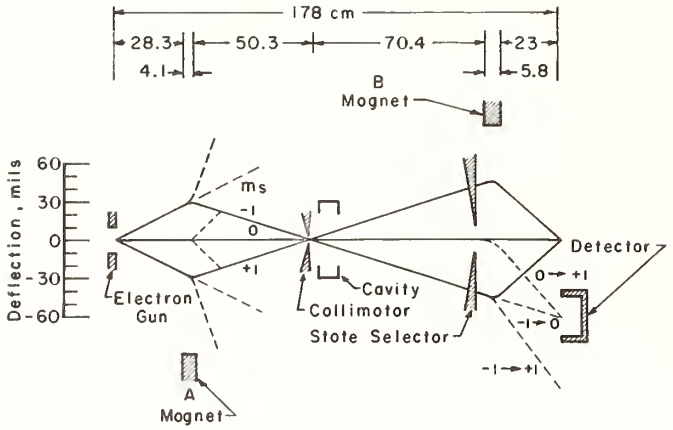


FIGURE 2. *Trajectory diagram illustrating the experimental technique.*

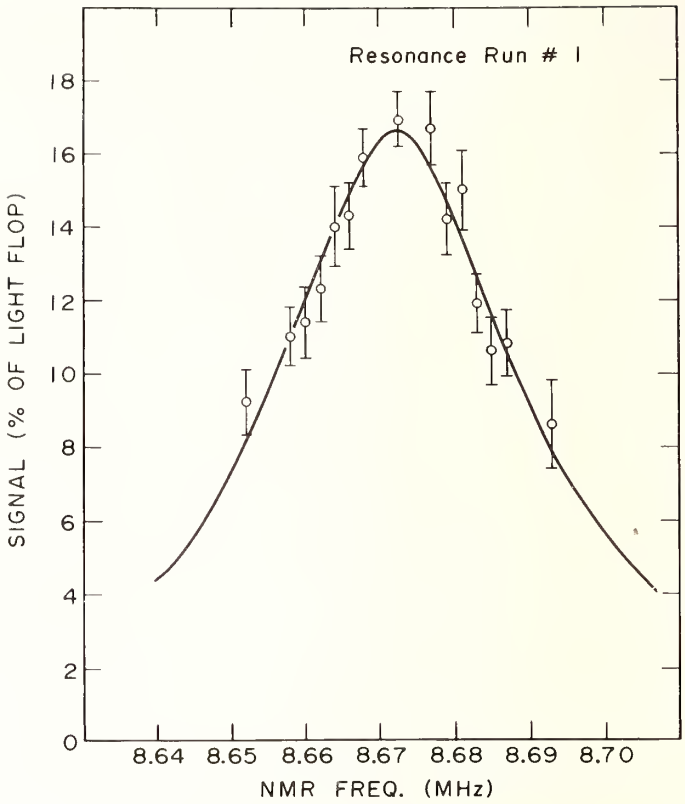


FIGURE 3. *Resonance line shape for a single run on the transition* $2^3P\ (J, m_J)$*:* $(0,0) \rightarrow (2,0)$.

fact the optical level-crossing technique was discovered in the 2^3P state. Level-crossing experiments of higher precision have been conducted recently [6], but this technique is severely limited by its inability to make measurements of fine structure intervals independent of Zeeman parameters.

The method we have employed, the optical-microwave atomic beam magnetic resonance technique does not have this limitation. Furthermore the experiments are performed on a beam of free atoms, and level shifts due to collisions are not a problem. The basic experimental setup is illustrated in Fig. 2. A well-collimated beam of atoms in the metastable 2^3S state is obtained by electron bombardment of the 2^1S ground state. Two inhomogeneous deflecting magnets A, B are used, in conjunction with state-selector stops, to define specific spatial trajectories for atoms which leave the electron gun in a particular m_s sublevel and arrive at the detector in another. Between the A and B magnets is a region of uniform magnetic field where the atoms are optically excited to the 2^3P state while they also pass through a microwave cavity. During the process of optical excitation and subsequent decay back to the 2^3S state a fraction of the atoms will transfer from one m_s sublevel to another, and as illustrated in Fig. 2 our detector was located to measure the fraction making the transfer $m_s = 0 \rightarrow m_s = +1$. If during the excitation and decay process a microwave magnetic-dipole transition is induced between a pair of 2^3P (J, m_J) sublevels, the fractional 2^3S (m_s) transfer will be modified. In this way the microwave transition in the 2^3P state is rendered observable in the detected 2^3S beam. The fractional microwave signal depends on relative optical excitation and decay matrix elements for the pair of 2^3P (J, m_J) sublevels in question.

In the fine-structure measurements three transitions were measured involving the 2^3P $(m_J = 0)$ sublevels for all possible pairs of $J = 0, 1, 2$. The $m_J = 0$ sublevels were chosen to minimize the magnetic field dependence of the transition frequencies. In the presence of a magnetic field sublevels of different J and same m_J are admixed resulting in a slight field-dependence of the $m_J = 0$ sublevels. This admixture, however, had the virtue of enabling us to observe the otherwise forbidden $J = 0 \rightarrow J = 2$ transition. Thus the $J = 0 \rightarrow J = 1$ and $J = 1 \rightarrow J = 2$ intervals were measured as well as their sum $J = 0 \rightarrow J = 2$ to provide a check on the experimental consistency. Other transitions were measured to obtain independent values of the Zeeman parameters for the 2^3P state.

In frequency units the transition natural linewidth was 3 MHz, and optimal signal-to-noise ratio was obtained when this was power-broadened to about 5 MHz. This necessitated microwave magnetic fields of between 2 G and 3 G. The lines were plotted by sweeping the uniform magnetic field utilizing the slight dependence thereon of the transition frequency. Figure 3 shows a typical resonance line for the $J = 0 \rightarrow J = 2$ interval measurement. The precision of that run was 0.05 MHz or about 1% of the linewidth.

Further details of the experiments can be found in the literature [7].

3. Results and Discussion

The results for the fine structure intervals $\nu_{JJ'}$ are summarized below:

$$\nu_{01} = 29\,616.864 \pm 0.036 \text{ MHz (1.2 ppm)}$$
$$\nu_{12} = 2\,291.196 \pm 0.005 \text{ MHz (2.2 ppm)}$$

$$\nu_{01} + \nu_{12} = 31\,908.060 \pm 0.036 \text{ MHz (1.1 ppm)}$$
$$\nu_{02} = 31\,908.040 \pm 0.020 \text{ MHz (0.6 ppm)}$$

TABLE 1. *Theoretical contributions to the fine structure of 2^4P helium in MHz.*

Interval	α^2Ry	α^3Ry	(m/M)α^2Ry	Second Order	α^4Ry	ν_{theory}
ν_{01}	29564.587	54.708	−10.707	11.657	−3.331	29616.914
	±0.006		±0.00044	±0.042	±0.0039	±0.043
	(0.21 ppm)		(0.015 ppm)	(1.42 ppm)	(0.13 ppm)	(1.44 ppm)
ν_{12}	2317.204	−22.548	1.952	−6.866	1.542	2291.284
	±0.0018		±0.00088	±0.081	±0.0068	±0.081
	(0.76 ppm)		(0.39 ppm)	(35 ppm)	(3.0 ppm)	(35 ppm)

It can be seen that the measured ν_{02} is in excellent agreement with the sum of the measured ν_{01}, ν_{12}, establishing experimental consistency and suggesting that there were no unaccounted systematic errors. The most precise theoretical number is for ν_{01}, and the most accurate experimental comparison for this is the difference $(\nu_{02} - \nu_{12})$:

$$\nu_{01} = 29\ 616.844(21)\ \text{MHz}\ (0.7\ \text{ppm}) \quad \text{expt.}$$

$$\nu_{01} = 29\ 616.914(43)\ \text{MHz}\ (1.5\ \text{ppm}) \quad \text{theory.}$$

The theory is summarized in Table 1. The current theory includes quantum electrodynamic terms [8] to order α^4Ry as well as corrections to the spin-dependent operators [9] to the same order.

An alternative approach is to use the theory to obtain an independent value for the fine structure constant with the result

$$\alpha^{-1} = 137.036\ 13(11) \quad (0.8\ \text{ppm}).$$

This is to be compared with the more accurate value currently accepted [1]

$$\alpha^{-1} = 137.035\ 963(15) \quad (0.11\ \text{ppm}).$$

We believe that refinements in the experimental technique could improve the accuracy of the measured interval ν_{01} to 0.2 ppm, making possible a 0.1 ppm determination of the fine structure constant. This would require the theory to be done to order α^5Ry.

The research was supported in part by the National Science Foundation under grant No. PHY 78–25655.

References

[1] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. **42**, 1575 (1979).
[2] R. S. Van Dyck, Jr., R. P. Schwinberg, and H. A. Dehmelt, Phys. Rev. Lett. **38**, 310 (1977).
[3] J. P. Barrat, J. Phys. Radium **20**, 541 (1959); **20**, 633 (1959); **20**, 657 (1959).
[4] I. Wieder and W. E. Lamb, Jr., Phys. Rev. **107**, 125 (1957).
[5] F. D. Colegrove, P. A. Franken, R. R. Lewis, and R. M. Sands, Phys. Rev. Lett. **3**, 420 (1959).
[6] C. Lhillier, J. P. Faroux, and N. Billy, J. Phys. (Paris), **37**, 335 (1976).
[7] F. M. J. Pichanick, R. D. Swift, C. E. Johnson, and V. W. Hughes, Phys. Rev. **169**, 55 (1968); S. A. Lewis, F. M. J. Pichanick, and V. W. Hughes, Phys. Rev. A **2**, 86 (1970); A. Kponou, V. W. Hughes, C. E. Johnson, S. A. Lewis, and F. M. J. Pichanick, Phys. Rev. A **24**, 264 (1981); W. Frieze, E. A. Hinds, V. W. Hughes, and F. M. J. Pichanick, Phys. Rev. A **24**, 279 (1981).
[8] M. Douglas and N. M. Kroll, Ann. Phys. (N.Y.) **82**, 89 (1974).
[9] M. L. Lewis and P. H. Serafino, Phys. Rev. A **18**, 867 (1978).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Preliminary Measurement of the $J = 0$ to $J = 2$ Fine Structure Interval in the 3^3P State of Helium*

Mark Feldman†, Tom Breeden, Louis DiMauro, Taiqian Dong, and Harold Metcalf

Department of Physics, State University of New York at Stony Brook, Stony Brook, NY 11794

We have made time resolved measurements of the level crossing signal from the $J = M = 2$, and the $J = M = 0$ sublevels of the 3^3P_0 state of He near 2277 gauss. A thermal beam of He 2^3S metastables is excited by a pulse of 3889 Å dye laser light and a temporal slice of the fluoresence at 90° to the field is recorded by fast electronics as a function of field. This experiment has been carefully designed to minimize the effects of very many possible systematic errors. We have a S/N of about 15 in each of about 100 data sets (≃ 1 hour run time). The result has a statistical standard error of 1.4 ppm and a systematic uncertainty of about 1.4 ppm, resulting in a net uncertainty of 2 ppm. The distribution of the results is approximately a Gaussian of width 14 ppm and the average, corrected for field inhomogeneity, is 9695.013 kHz (NMR, H_20).

Key words: fine structure; helium; level crossing; line narrowing; time resolution.

1. Introduction

Measurements of the fine structure in the 2^3P state of helium can now be compared with theory at the ppm level [1], and show promise of providing an independent measure of the fine structure constant α. Although a comparably accurate theory for the 3^3P fine structure does not presently exist, the possibility of such a calculation makes the measurement of this interval very attractive. Furthermore, line narrowing techniques employing time delayed spectroscopy [2] offer the possibility of sub-natural width signals that enhance both accuracy and precision of such measurements.

We have made time resolved measurements of the level crossing signal from the $J = 2$, $M = \mp 2$ and the $J = M = 0$ sublevels of the 3^3P state of He near 2277 gauss (see Fig. 1). A thermal beam of He 2^3S metastables is excited by a pulse of 3889 Å dye laser light and a temporal slice of the fluoresence at 90° to the field is recorded by fast electronics as a function of field. A frequency-locked NMR system is used to sweep the stabilized magnetic field through several linewidths in a series of 100 steps, repeatable to about 1 ppm. [3]

The level crossing field depends on both the zero field splitting and the Zeeman effect. The precise measurement reported here can be combined with an accurate evaluation of the Zeeman effect to determine the fine structure splitting. This can be compared with the relativistic theory of the helium atom, including QED contributions, for a test of that theory or as a measure of the fine structure constant α.

2. Apparatus

A schematic diagram of the apparatus is shown in Fig. 2. The source of metastable triplet He atoms consists of

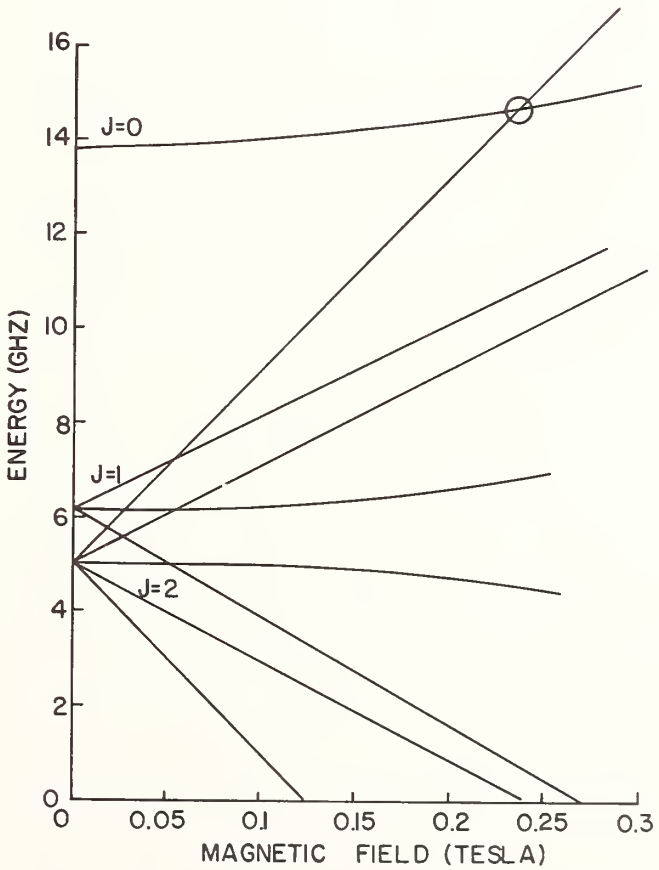


FIGURE 1. *The Zeeman effect in 3^3P helium in the field range of interest. The zero of energy is arbitrary.*

*Supported by the National Science Foundation

†Present address: Time and Frequency Division, N.B.S., Boulder, CO 80303

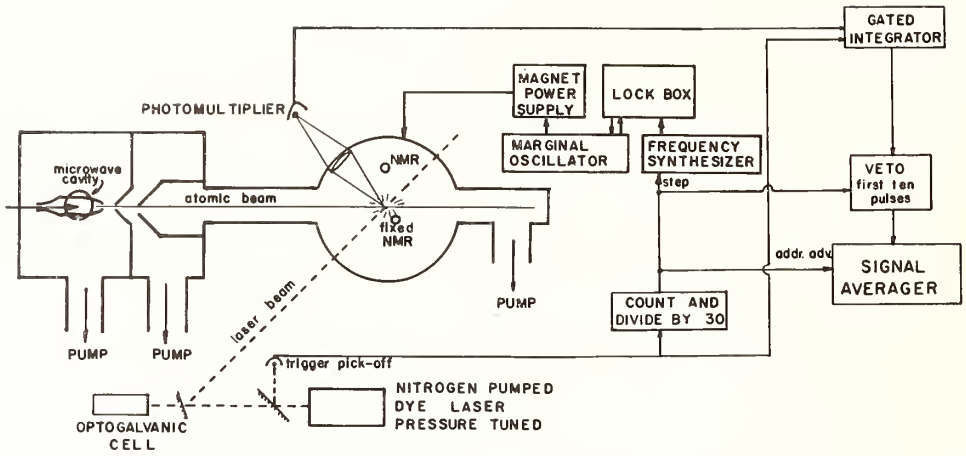


FIGURE 2. *Partial schematic diagram of apparatus.*

two differentially pumped chambers with an exit to a 1.5 m beam line that is maintained at 5×10^{-6} Torr.

We use a large jet booster pump on this specially designed vacuum system enabling a helium flow of several Torr·liters/sec from a microwave discharge source, and a skimmer aperture that passes only a small fraction of the total helium flow into an intermediate vacuum region. A second skimmer allows only the helium beam to enter into the high-vacuum ($\simeq 10^{-6}$ Torr) beam pipe. The resulting high pressure (a few Torr), high flow source produces a thermal beam of more than 10^{18} atoms/sec and a density of more than 10^5 metastable He atoms/cm^3 in our interaction region 1 m from the source. This is sufficient for us to detect 3889 Å fluorescence of several photons/laser pulse.

Our dye laser has a bandwidth close to the design limit of about 3 GHz, peak output power of about 6 kW, and is pressure tuned. An airspaced, intracavity etalon enables us to resolve the various fine structure transitions, thus reducing the background (and therefore the noise) from many possible transitions that contribute to the fluoresence but not to the signal. Figure 3 shows that many of these contributions can be reduced or eliminated by appropriate tuning of the laser. Figure 4 shows the fluorescence obtained by sweeping the laser through ~30 GHz. The width of the peaks is caused by both the 3 GHz laser width and the 3 GHz Doppler width (45° intersection of beams).

Our field measurement and sweep is accomplished by a series of single steps, each of which is reliable and repeatable. At the heart of the system (see central part of Fig. 2) is a Hewlett Packard programmable frequency synthesizer that is stepped by 0.3 kHz by a pulse from a trigger pick-off photodetector every 30 laser pulses (laser is free running). An NMR marginal oscillator is locked to the synthesizer with a mixer-amplifier-varactor system so that it also steps frequency. The NMR signal from the marginal oscillator is amplified by a lock-in detector which provides a dispersion-shaped error signal that is fed to the programmable sweeping magnet power supply so that the magnet is always brought to the center of the resonance. The field is thus locked to the NMR at each of 100 steps corresponding to the 100 channels of the signal averager. Flyback for multiple sweeps is arranged with appropriate electronics. With this system the accuracy of our individual field measurements is about 1 ppm, and the overall field calibration has internal precision of about 0.3 ppm.

Fluorescence is detected by a photomultiplier (PMT) that views about 1 cm of the laser beam path through a telescope of approximately unit magnification. The PMT

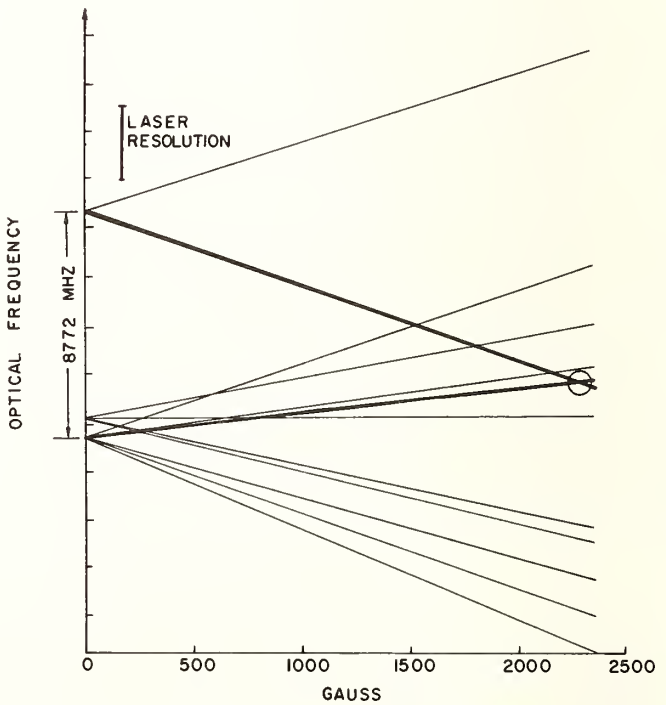


FIGURE 3. *Frequency of allowed optical transitions ($\Delta M = \pm 1$) in magnetic fields of interest. The zero of energy is arbitrary.*

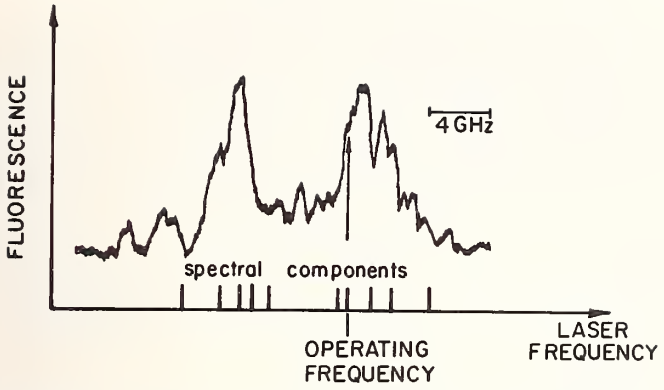


FIGURE 4. *Trace of fluorescence obtained by sweeping the pressure of the laser tuning chamber. Compare with Fig. 3.*

is well shielded from fringing fields and is located well outside of the magnet. The atomic beam is much larger than either the PMT field of view or the 5 mm laser beam.

The time-delayed detection of fluorescent light is accomplished by a LeCroy 227 gated integrator that receives the output of the PMT. The trigger pick-off pulse that signals the field and frequency stepping is also used to start the delay interval of the integrator (see Fig. 2). After this delay has elapsed (typically 80–100 nsec) the integration begins for a measurement period (typically 80–100 nsec). Each of these time intervals is adjustable. The output of the integrator is averaged for 20 laser pulses in one channel of our analyzer, and the channel is incremented with the field and NMR frequency.

The effect of time resolution on the signal shape deserves further comment. The precision of experiments on excited states has always been limited by the natural width of the transition. It is possible to modify the experiment to reduce the effect of the finite lifetime in limiting the precision of measurements of excited state parameters [2].

The modification is accomplished by arranging the experimental apparatus to interact with the atoms under study for two relatively short and well separated intervals. In our level crossing experiments we excite the species under study with a pulse of light of duration short compared with the excited state lifetime, observe the decay fluorescence for a well-defined interval, and separate these intervals by a selectable time span. For delay times considerably longer than the lifetime, the central part of the oscillatory signal [4] is considerably narrower than that from a continuous experiment. What is more important though, is that because these signals have high frequency information far into the wings, they will be much more useful for defining the position of the center of the level crossing. [2]

3. Data Acquisition and Processing

Each data set consists of the averaged fluorescence from 8–10 sweeps over 100 channels (corresponding to 100 field points). We veto the first 10 laser pulses of each step in order to minimize the effect of magnetic field settling time. Each sweep step has 30 laser pulses, resulting in 16,000–20,000 laser pulses averaged during the 45–60 minute data collection period. We detect 2–4 photons/laser pulse, about 500 in each channel, producing an expected scatter of 4–5% of the total detected fluorescence. Since the signal is about 40% of the total fluorescence, the S/N we expect is about 8–10, and the typical signal shown in Fig. 5 corroborates this estimate.

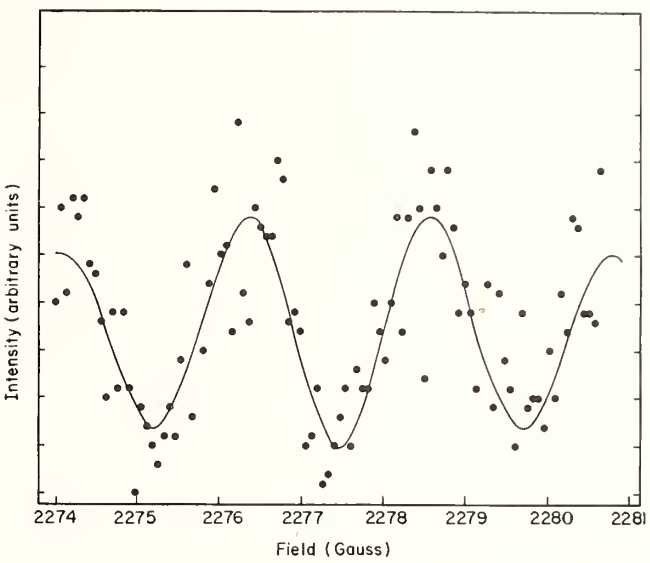


FIGURE 5. *Typical signal (unsmoothed) and fitted lineshape. This data set was taken in 50 minutes.*

Using a standard non-linear least squares routine, we fit a theoretical lineshape to the data by varying the center frequency and 2–4 other experimental parameters. From the results of the fit we can determine the NMR frequency of the position of the central maximum. The result of averaging over 100 data sets is 9695.013 kHz (corrected to H_2O frequency). The distribution of our results, shown in Fig. 6, has a standard deviation of about 14 ppm resulting in a statistical standard error of 1.4 pm.

4. Systematic Errors

This experiment has been carefully designed to minimize the effects of possible systematic errors. There is no rf to broaden or shift the signal. It is done in a beam to eliminate effects of both collisions and stray charged particles. All the data is taken long after the laser pulse ends so that there are no optical power effects. The data averaging system is digital so that the signal-to-noise ratio is not influenced by extraneous signals. The principle systematic effects under investigation derive from alignment uncertainties and field calibration.

In order to test for effects associated with magnetic field settling between steps, we have swept the field in both directions. In order to test for a variety of effects (alignment, field sensitivity of PMT, etc.), we have reversed the direction of the magnetic field. These experiments divide our data into four roughly equal populations providing the following results. Sweep up results aver-

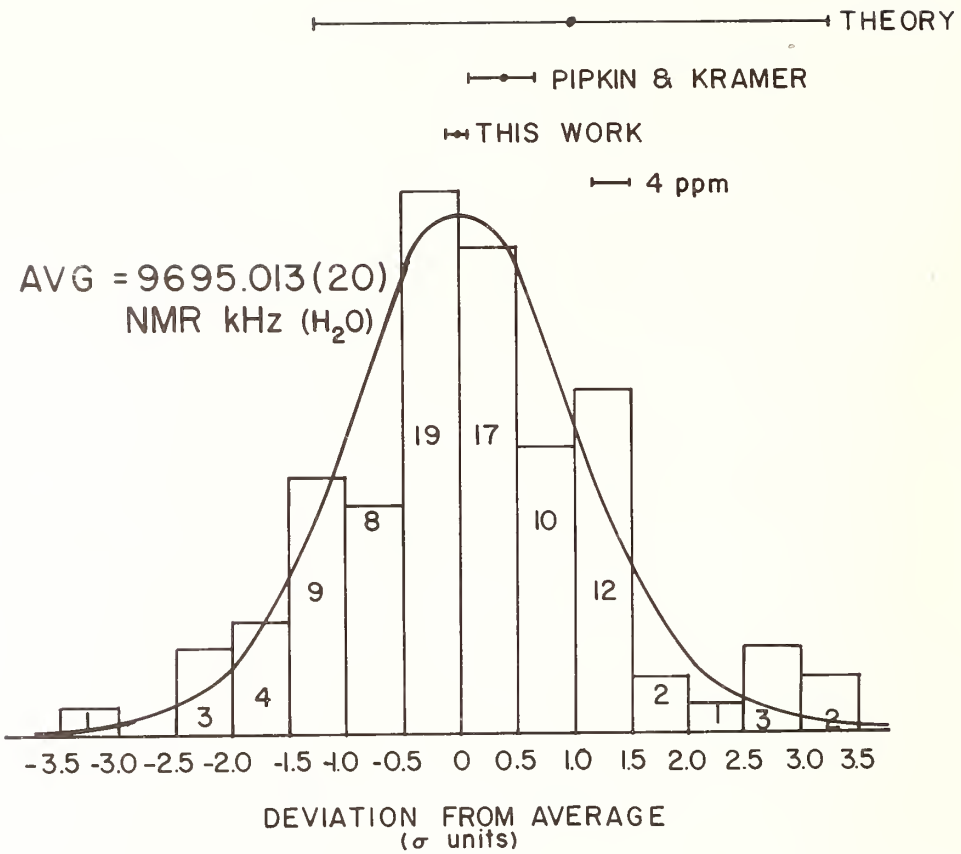


FIGURE 6. *Histogram of results of center frequency and a Gaussian of appropriate area. The results of the measurement reported here, shown above the histogram, includes both statistical and systematic uncertainty which combine to 2.0 ppm.*

age 5 ± 2 ppm lower than sweep down; field "reversed" data average 11 ± 2 ppm lower than field "normal". We attribute these differences to shifts associated with time lags (sweep direction) and misalignment (field direction) and average over the results. Because these shifts of 2.5 ± 1 ppm and 5.5 ± 1 ppm may not be exactly compensated by averaging, we allow a 1 ppm systematic uncertainty associated with these effects.

The magnetic field is measured by a probe located a few cm from the interaction region, and the measured field value is corrected by the measured difference between this and a movable probe which is centered in the interaction region before and after each day's running time (5 to 15 data sets). We have found this correction to be constant over a period of a few days, as long as the magnet is left on, but to change by as much as 25 ppm for different magnet cycles. Our uncertainty in measuring this correction on a given day is usually less than 1 ppm, and the errors are expected to be random. We therefore assign a residual uncertainty of 1 ppm/$\sqrt{\text{No of days}} \simeq 0.4$ ppm to this effect.

We have mapped the field over considerable distances a number of times in order to determine the effect of the field inhomogeneity over the interaction region. This can cause a systematic shift because the probe does not have the same shape (spherical) as the region of intersection of the laser beam, atomic beam, and field of view of the PMT. Although the inhomogeneity over the interaction region can be a few ppm, a conservative estimate of the uncertainty arising from a linear variation of the field is about 0.7 ppm. Any non-linear field variation does not significantly change this estimate.

We have considered the contribution from a number of other sources to possible systematic errors in this measurement but have found them to be negligible. These include laser frequency setting and drift, overlapping resonances, diamagnetic shielding, time base of synthesizer (measured to be 0.3 ppm), and PMT field dependence. The total systematic uncertainty arises from the field sweep and direction results (1.0 ppm), the probe displacement correction (0.7 ppm), and the field inhomogeneity (0.7 ppm). The quadrature sum of these contributions is 1.4 ppm.

5. Results[1]

We combine the systematic and statistical uncertainty (1.4 ppm each) in quadrature to obtain 2.0 ppm for the

[1]A new possibility of light shifts in quantum beats has been discovered while this paper was in press. It will be described in the June 1982, issue of Phys. Rev. A. It may affect these results by as much as 5 ppm.

total. Our present result, pending further work, is quoted as the NMR frequency of protons in water at the center of the time-resolved level crossing signal: f = 9695.013(19) kHz. This is in good agreement with the work of Kramer and Pipkin [5] that produced 9695.072(33) kHz and the result of Lhuiller et al. [6], 2277.125(15) gauss or 9695.099(64) kHz.

References

[1] W. Frieze *et al.*, these proceedings.
[2] H. Metcalf and W. Phillips, Opt. Lett. 5, 540 (1980); these proceedings.
[3] For more details of all aspects of this experiment, see the thesis of Mark Feldman, State University of N.Y., 1980.
[4] P. Schenck, R. Hilborn, and H. Metcalf, Phys. Rev. Lett. 31, 189 (1973).
[5] P. Kramer and F. Pipkin, Phys. Rev. A18, 212 (1978).
[6] C. Lhuillier, P. Riviere, and J. Faroux, C.R. Acad. Sci. Ser. B 276, 607 (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Laser Microwave Precision Measurements of 2 $^3S_1$ and 2 ^3P Term Splittings in Helium-Like Li^+

U. Kötz, J. Kowalski, R. Neumann, S. Noehte, H. Suhr, K. Winkler, and G. zu Putlitz*

Physikalisches Institut der Universität Heidelberg, F.R.G.

The hyperfine structure (hfs) splittings of the metastable $1s2s\ ^3S_1$ state of helium-like $^{6,7}Li^+$ have been measured with combined laser optical pumping and microwave resonance. A low energy Li^+ ion beam, optically excited by an intersecting laser beam, passed through a waveguide where radio frequency transitions were induced. The resulting population transfer between the hfs levels of the $^3S_1$ was detected via the change of fluorescence light intensity from a second crossing region of laser light and ion beam located behind the waveguide. The magnetic hfs constants $A(^6Li^+, 2\ ^3S_1)$ and $A(^7Li^+, 2\ ^3S_1)$ were extracted from the measurements and compared with theory. The hfs anomaly is in good agreement with the value obtained from the magnetic hfs constants of the atomic $2\ ^2S_{1/2}$ ground states. In addition rf measurements of the $2\ ^3P$ hfs have been performed.

Key words: fine and hyperfine structure; hyperfine structure anomaly; laser-microwave spectroscopy; singly ionized lithium; two-electron system.

Singly ionized lithium belongs, together with helium, to the class of simple two-electron systems which may be regarded as fundamental in the sense that the energy states of these systems can be calculated with high precision. The exact knowledge of such energies can serve either as an experimental test of the underlying theory or, if this theory is accepted, as a means for the determination of fundamental constants. The precise determination of the He fine structure splitting of the $1s\,2p\ ^3P$ term [1] has become the source for an exact value of the fine structure constant α. In this paper results of radio frequency (rf) measurements of the hyperfine structure (hfs) splittings of the metastable $1s\,2s\ ^3S_1$ term ($\tau = 50\ s$) and of the short-lived $1s\,2p\ ^3P_2$ term ($\tau = 43$ ns) of Li^+ are reported. The relevant part of the energy level diagram is shown in Fig. 1.

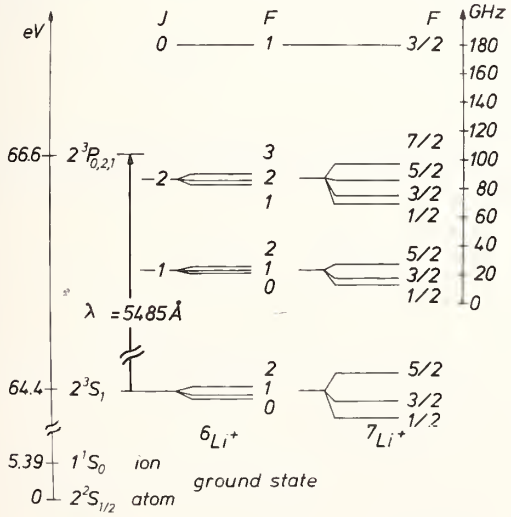


FIGURE 1. *Energy level scheme of the 1s2s $^3S_1$ and 1s2p ^3P terms of $^{6,7}Li^+$, the isotope shift is omitted.*

The hfs of the $2\ ^3S_1$ term of $^6Li^+$ and $^7Li^+$ has been measured by means of the optical pumping method [2–4]. Li atoms were ionized and excited to the metastable $2\ ^3S_1$ term by electron impact when leaving the aperture of an oven, and were focused to a well collimated beam of 200–400 eV (Fig. 2). Selective optical excitation of, e.g., the $^3S_1$, $F = 5/2 - {}^3P_2$, $F = 5/2$ transition with a single mode dye laser leads to a depopulation of the $^3S_1$, $F = 5/2$ term, which can be monitored by the same laser in a second interaction region 15 cm downstream in the ion beam. Between the two laser ion-beam intersection points the ions pass a waveguide, where rf transitions $^3S_1$, $F = 3/2 - 5/2$ are induced. This leads to a repopulation of the $F = 5/2$ term, which is observed as an increase of the fluorescent light in the monitor region. The linewidth (0.6–1.3 MHz) of the rf signal is dominated by the interaction time of the ions with the rf field. In order to eliminate the Doppler shift in the rf transition of several MHz the waveguide is terminated to produce a counterpropagating wave. Two signals shifted by $\pm\nu_0(v/v_{\text{phase}})$ with respect to the center frequency of the

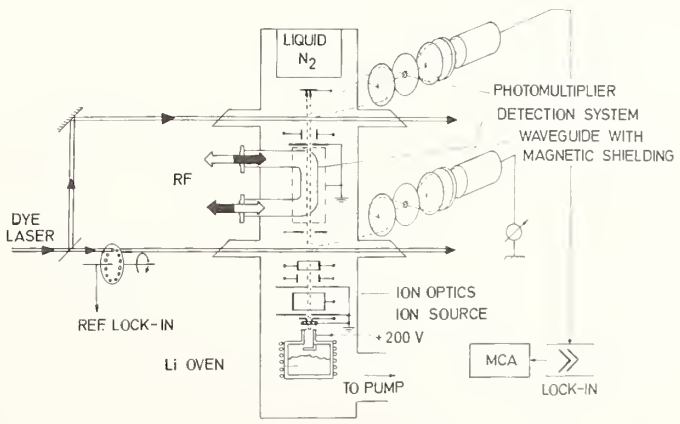


FIGURE 2. *Experimental arrangement of the laser-microwave spectrometer and ion beam chamber.*

*Also: Gesellschaft für Schwerionenforschung, Darmstadt, Federal Republic of Germany.

hfs transition ν_0 are obtained. Both signals are recorded simultaneously and stored in a multichannel analyzer. A more detailed description of the experimental set-up is given in Ref. [4].

Figures 3 and 4 show examples of microwave resonance signals for the $2\,^3S_1$ transitions in $^6Li^+$ and $^7Li^+$. The results of the $2\,^3S_1$ measurements are compiled in Table 1a and b. From the hfs splittings in $^{6,7}Li^+$ the magnetic hfs interaction constants $A(^6Li^+)$ and $A(^7Li^+)$ and small energy depressions $\delta(^6Li^+)$ and $\delta(^7Li^+)$ of the $F = I$ hfs sublevels were extracted. Within the experimental errors the resulting hfs anomaly Δ, defined by the equation $\Delta = (^6A/^7A)\,(^7g_I/^6g_I) - 1$ [5], where $^{6,7}g_I$ are the nuclear g-factors, is in good agreement with the value [6, 7] obtained from the magnetic hfs constants of the atomic $2\,^2S_{1/2}$ ground states (Table 2).

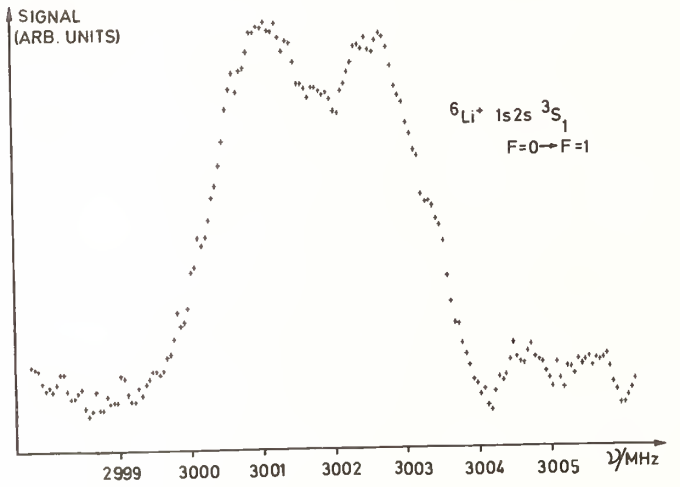


FIGURE 3a. *Microwave transition $F = 0 \rightarrow F = 1$ in the $2\,^3S_1$ state.*

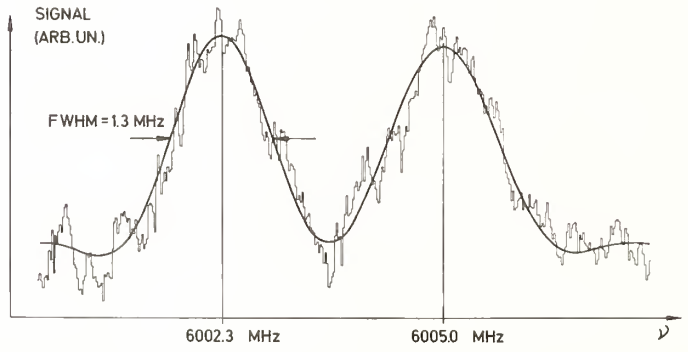


FIGURE 3b. *Microwave transition $F = 1 \rightarrow F = 2$ in the $2\,^3S_1$ state.*

The equation for the energy splitting between two hfs sublevels F and F' can be found in Ref. [8], and is written in the following form:

$$\Delta E = \frac{8}{3}\alpha^2 \frac{m_e}{m_p} g_I Z^3 [(1 + \epsilon) - 3\frac{m_e}{M} + \frac{\alpha}{2\pi}] R_\infty (Y_F - Y_{F'}).$$

Thereby ϵ is the correction for the second electron [9], and

$$Y_F = (1/4)[F(F + 1) - I(I + 1) - S(S + 1)].$$

The equation was used for a preliminary calculation of the A factors which corrects only for the second electron, the anomalous magnetic moment in lowest order ($\sim\alpha/2\pi$), and for the reduced mass. The error for $A_{\text{theor.}}$ in Table 3 is dominated by the uncertainty of ϵ. The difference from the experimental value is partly due to higher order QED and relativistic corrections. As is well known from ^3He [10] there should exist also a significant nuclear structure correction.

As can be seen from the experimental splitting values in Table 1, the ratio of the splittings within one isotope deviates from the Landé interval rule. For example, in $^7Li^+$ this ratio should be 5/3. However hyperfine mixing of the $2\,^3S_1$ term with the $2\,^1S_0$, with the latter

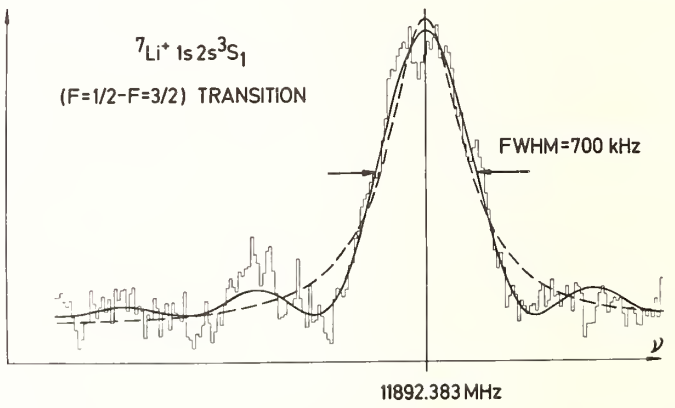


FIGURE 4a. *(a) $2\,^3S_1$, $F = 1/2 - F = 3/2$ microwave transition in $^7Li^+$ with fit curves (solid line: "Ramsey-type" function as discussed in [21], dashed line: Lorentzian).*

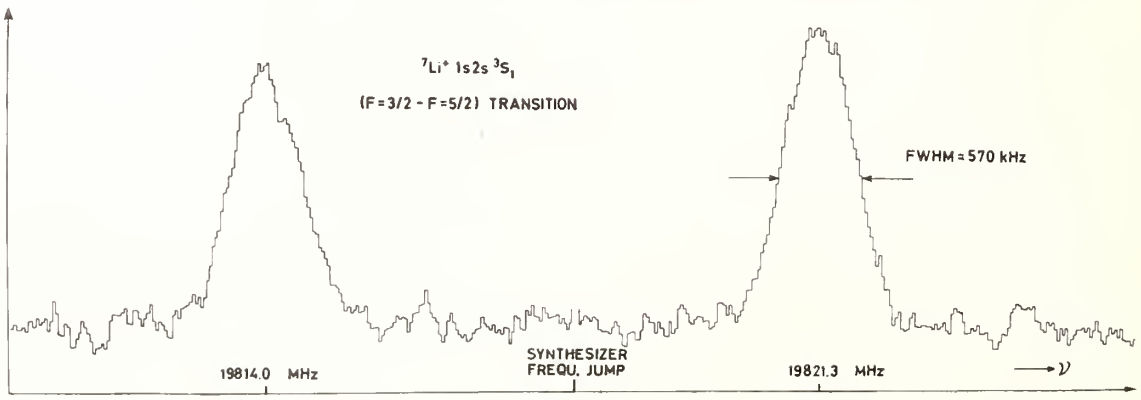


FIGURE 4b. *$2\,^3S_1$, $F = 3/2 - F = 5/2$ transition in $^7Li^+$ with two Doppler shifted signal curves.*

TABLE 1. *Experimental values of the 2 $^3S_1$ hfs splittings in $^6Li^+$ (a) and $^7Li^+$ (b).*

a.	$F = 0 - F = 1$	3001.780(50) MHz
	$F = 1 - F = 2$	6003.600(50) MHz
	Magnetic hfs constant A ($^6Li^+$, 2 $^3S_1$)	3001.793(17) MHz
	$F = 1$ depression δ($^6Li^+$, 2 $^3S_1$)	13(37) kHz
b.	$F = 1/2 - F = 3/2$	11890.018(40) MHz
	$F = 3/2 - F = 5/2$	19817.673(40) MHz
	Magnetic hfs constant A ($^7Li^+$, 2 $^3S_1$)	7926.923(14) MHz
	$F = 3/2$ depression δ($^7Li^+$, 2 $^3S_1$)	366(29) kHz

TABLE 2. *Hyperfine structure anomaly* Δ.

Li^+ ion (2 $^3S_1$)	Li atom (2 $^2S_{1/2}$)
$6.76(60) \times 10^{-5}$	$6.806(63) \times 10^{-5}$ [6, 7]

TABLE 3. *Theoretical estimates of A and* δ.

	$^6Li^+$	$^7Li^+$
$A_{theor.}$ (in MHz)	3000.81(10)	7925.15(23)
$\delta_{theor.}$ (in kHz)	39	511

represented by the nuclear spin $I = 3/2$, increases the distance between the 2 $^3S_1$ ($F = 3/2$) and 2 $^1S_0$ ($F = 3/2$) sublevels [11, 12] and causes a depression of the 2 $^3S_1$, $F = 3/2$ hfs state (Fig. 5). This effect can be calculated. Results based on hfs theory of two-electron systems, given in [13–15] are listed in Table 3. More refined calculations using better wavefunctions are in progress [16].

The method of laser-microwave spectroscopy has been extended to the excited short-lived 2 ^3P multiplet of $^7Li^+$, in order to measure the 2 ^3P hfs with rf methods. The experimental scheme is the following: The low-velocity Li^+ ion beam passes a microwave cavity. A dye laser beam crosses the ion beam inside the cavity and excites ions from a certain 2 $^3S_1$ hfs sublevel to one of the 2 ^3P hfs states, e.g., with the transition 2 $^3S_1$, $F = 3/2 - 2\ ^3P_2$, $F = 5/2$ (Fig. 6). Microwave transitions ($\Delta F = 1$) in the 2 $^3P_2$ to one of the neighboring hfs sublevels with $F = 3/2$ or 7/2 change the decay rate back to 2 $^3S_1$, $F = 3/2$. The same laser monitors the resulting population change of the 2 $^3S_1$, $F = 3/2$ in a second crossing region far outside the cavity. An example of a microwave resonance signal is shown in Fig. 7. The method can be compared with the magnetic deflection method used by V. W. Hughes and co-workers for the very precise 2 ^3P fine structure measurement in ^4He [1]. In the Li^+ experiment the monitoring laser beam replaces the deflecting magnetic field. The technique is very successful for ion-beam precision spectrosopy.

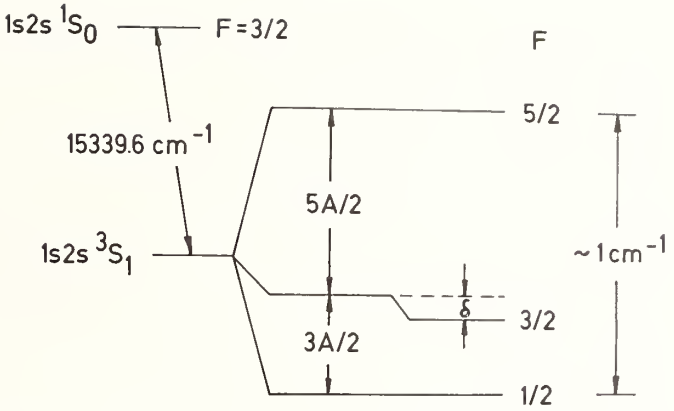


FIGURE 5. *Diagram of the 2 $^3S_1$ hyperfine multiplet of $^7Li^+$ including the $F = 3/2$ depression (enlarged).*

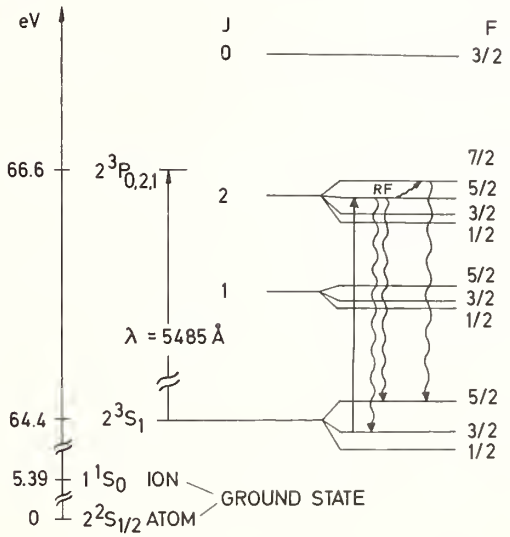


FIGURE 6. *Energy diagram of the 2 $^3S_1$ and 2 ^3P multiplets of $^7Li^+$ including the optical pumping cycle described in the text.*

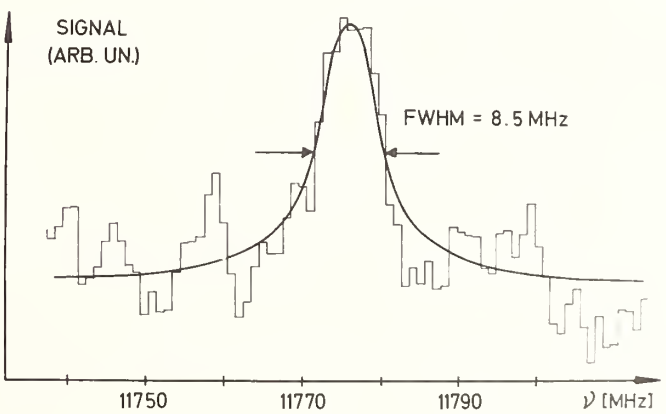


FIGURE 7. *Signal curve of the 2 $^3P_2$ ($F = 5/2 - F = 7/2$) microwave transition in $^7Li^+$.*

For the determination of the larger frequency splittings between different fine structure (fs) levels of the ^3P term a recently tested [17] radio frequency controlled two dye laser spectrometer will be used. It allows one to measure laser difference frequencies continuously up to 18 GHz. Larger energy differences between Li^+ ^3P fs levels will be measured stepwise by using, for example,

I_2 lines as references between the corresponding Li^+ transitions to which one laser can be stabilized. These measurements should be capable of determining the fs with ppm precision.

For the first time experimental data can now be obtained which allow one to test the very precise hfs [18] and fine structure [19, 20] theories existing for Li^+.

References

[1] W. E. Frieze, E. A. Hinds, V. W. Hughes, and F. M. Pichanick, Phys. Lett. 78A, 322 (1980).

[2] S. D. Rosner, R. H. Holt, and T. D. Gaily, Phys. Rev. Lett. 35, 785 (1975).

[3] W. Ertmer and B. Hofer, Z. Phys. A 276, 9 (1976).

[4] U. Kötz, J. Kowalski, R. Newmann, S. Noehte, H. Suhr, K. Winkler, and G. zu Putlitz, Z. Phys. A 300, 25 (1981).

[5] H. M. Foley, in *Atomic Physics*, Ed. by B. Bederson, V. W. Cohen, and F. M. J. Pichanick (Plenum Press, New York, 1969) p. 509.

[6] A. Beckmann, K. D. Böklen, and D. Elke, Z. Phys. 270, 173 (1974) (Earlier $2\ ^2S_{1/2}$ hfs precision measurements in Li are cited in Ref. [7].

[7] E. Arimondo, M. Inguscio, and P. Violino, Rev. Mod. Phys. 49, 31 (1977).

[8] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Springer, Berlin, 1957).

[9] P. J. Luke, R. E. Meyerott, and W. W. Clendenin, Phys. Rev. 85, 401 (1952).

[10] S. D. Rosner and F. M. Pipkin, Phys. Rev. A 1, 571 (1970).

[11] I. I. Sobel'man, *Introduction to the Theory of Atomic Spectra* (Pergamon Press, New York, 1972).

[12] M. M. Sternheim, Phys. Rev. Lett. 15, 545 (1965).

[13] G. Breit and L. A. Wills, Phys. Rev. 44, 470 (1933).

[14] C. Schwartz, Phys. Rev. 97, 380 (1955).

[15] A. Lurio, M. Mandel, and R. Novick, Phys. Rev. 126, 1758 (1962).

[16] R. Herman, J. Kowalski, R. Neumann, and G. zu Putlitz, to be published.

[17] J. Kowalski, F. Mayer, R. Neumann, S. Noehte, R. Schwarzwald, H. Suhr, K. Winkler, and G. zu Putlitz, Abstract Book, p. 1068, European Conference on Atomic Physics (ECAP), held in Heidelberg, 6–10 April 1981.

[18] N. A. Jette, T. Lee, and T. P. Das, Phys. Rev. A 9, 2337 (1974).

[19] B. Schiff, Y. Accad, and C. L. Pekeris, Phys. Rev. A 1, 1837 (1970).

[20] Y. Accad, C. L. Pekeris, and B. Schiff, Phys. Rev. A 4, 516 (1971).

[21] N. F. Ramsey, *Molecular Beams* (Oxford University Press, London, 1956).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

An Optically Pumped Metastable Hydrogen Beam

K. C. Harvey

Center for Absolute Physical Quantities, National Bureau of Standards, Washington, DC 20234

A beam source of atomic hydrogen is described which produces metastable atoms in the $2S_{1/2}$ state by optical pumping. An effusive beam of atomic hydrogen is generated in the ground state. The atoms pass in front of a VUV lamp producing Lyman-β (1026 Å) radiation. The atoms are excited to the 3p level and then cascade to the $2S_{1/2}$ state. The metastable atoms are measured by quenching them with an electric field and detecting the emitted Lyman-α radiation. Beams with a flux of 10^6 atoms/sec were obtained. A metastable-beam effective temperature of 100 K was measured. In addition preliminary measurements of the transition from $n = 9$ to $n = 20$ in hydrogen using a Woods discharge are presented and directions for improvement are indicated.

Key words: atomic beam; hydrogen; metastable states; optical pumping.

The metastable $2S_{1/2}$ state of atomic hydrogen has long been a useful system for measurements in the area of fundamental constants as well as basic physics. With the need for increasingly precise and accurate measurements, a suitable source of metastable atoms has not always been available. For such measurements, the atom must be isolated from uncontrolled external influences which means an atomic beam must be used. In such an arrangement the number density is small. In addition the velocity distribution of the beam should be well characterized so that experimental systematics may be properly taken into account. An optically pumped hydrogen beam is a source which offers the possibility of high number densities and at the same time a well defined, slow velocity distribution.

There are a number of existing and proposed methods for producing metastable hydrogen beams. Electron bombardment was the original technique [1] and is still frequently employed. Typical fluxes are about 10^6 atoms/sec with temperatures of about 2000 K, but their usefulness is limited by the momentum transfer produced in the excitation process by the electrons. This leads to a velocity distribution which strongly depends on the electron energy and collimation parameters [2]. Charge exchange between a proton beam and cesium vapor [3] produces typical fluxes of 10^{12} atoms/sec. These beams have, however, a large energy of 500 eV or more and limited state selectivity since long-lived Rydberg states are simultaneously produced with the $2S_{1/2}$ state in the charge exchange process. Proposals for neutral atomic traps [4], though promising, have not yet been realized. Schemes employing laser-induced nonlinear processes [5] are now limited by the low duty cycle of pulsed lasers and the photoionization which is always present.

The technique of optical pumping offers a way to overcome these limitations. Although the advantages of this approach were early appreciated by Lamb and Retherford [1], its development was inhibited by the absence of an adequate source of radiation [6]. Recently an appropriate VUV lamp has been developed [7]. In the arrangement presented here, a dense, collimated beam of neutral hydrogen in the ground state is produced and passes in front of the VUV lamp. The lamp irradiates the atoms with Lyman-β (1026 Å) radiation. The atoms are transferred to the 3P state from which they cascade to the $2S_{1/2}$ state with a branching ratio of about 12% (see Fig. 1). From the lamp region the metastable atoms travel to an experimental region or an area where the metastable flux is measured. Since Lyman-β light may be readily filtered, it is possible to produce highly state selected beams. The number of metastables produced is limited by the flux of ground state atoms and the flux of Lyman-β photons which suggest that high densities are possible. In principle, the entire neutral beam can be converted to metastable atoms.

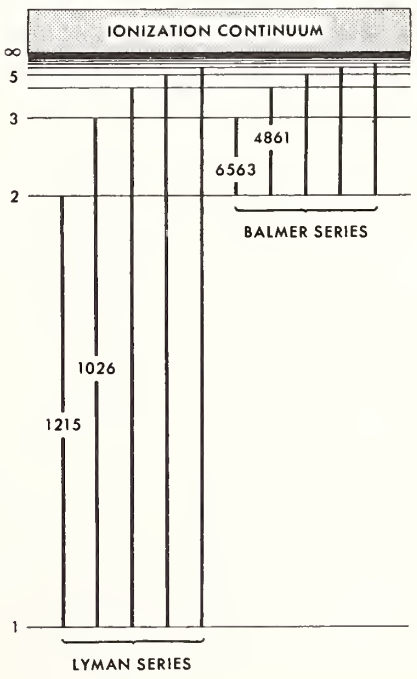


FIGURE 1. *Energy level diagram of hydrogen.*

A schematic of the atomic beam system is shown in Fig. 2. The vacuum chamber is a stainless steel box in

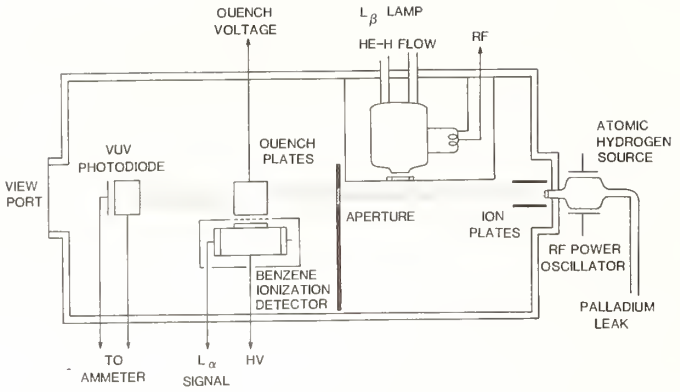


FIGURE 2. *Schematic of the atomic beam apparatus showing the hydrogen source, lamp region, and metastable detector.*

which various items are mounted on plates supported kinematically on stainless steel rods. Most of the components in vacuum are gold plated to improve conductivity and reduce charge accumulation. The chamber has a number of feedthrough ports along its length as well as view ports. All dielectric surfaces on windows are shielded with metallic mesh to reduce stray fields from accumulated charges. The apparatus is pumped by a 1000 ℓ/sec diffusion pump which uses a low vapor pressure silicone oil to reduce backstreaming. A high conductivity liquid nitrogen trap is located between the chamber and pumps. The end vacuum of the system is about 10^{-8} Torr.

The neutral atomic hydrogen beam is produced by dissociation of molecular hydrogen in an rf discharge and then effusion through a multichannel array. The molecular hydrogen is first purified with a palladium leak. The hydrogen is introduced into a pyrex bulb 4.0 cm in diameter where it is dissociated by an rf oscillator delivering 25 W at 30 MHz. The degree of dissociation is monitored by observing in the discharge the relative intensity of the atomic Balmer series and the yellow molecular hydrogen lines. The glass multichannel array is attached to the discharge bulb to collimate the beam. The array has capillaries 50 μm in diameter and 2 mm in length, arranged so that the array has transparency of 50%. The total area of the array is that of a circle 3 mm in diameter. Supports on the discharge bulb allow the direction of the beam to be adjusted. The hydrogen beam is estimated to be about 50% dissociated when it enters the vacuum chamber. Using the theory of multichannel gas collimating arrays [8], the divergence of the beam is found to be 2.4° in the low pressure limit when the pressure in the discharge is about 10 μm. The beam flux is measured using the pumping speed and the vacuum chamber pressure. A pressure of 10^{-6} Torr in the chamber indicates a beam intensity of 3×10^{16} atoms/sec. The above method for producing a beam was chosen since it generates an intense beam without differential pumping and it can be readily modified to produce a beam at cryogenic temperatures [9].

The rf discharge produces ions such as H^+, H_2^+ and H_3^+ as well as neutral H and H_2. Since these ions might quench the metastables in the beam and create fast neutrals by charge exchange, they are swept away by an electric field produced by the ion plates. This is an electrode structure [10] which creates a uniform field between the electrodes but very weak fields a short distance from them.

The Lyman-β lamp has been described elsewhere [7] and only details relevant to the operation of the beam will be repeated here. This lamp is a relatively high intensity source of hydrogen resonance radiation with a narrow spectral width and low self-reversal to better match the beam's absorption profile. The lamp's window is a LiF crystal 12 mm in diameter and 1 mm thick which when cooled will transmit radiation at 1026 Å [11]. Electrodes near the window excite an rf discharge in a gas composed of 1% hydrogen and 99% helium. The lamp is surrounded by a Dewar filled with liquid nitrogen which cools the window and the gas. With an rf power of 10 W at 130 MHz the lamp produces $10^{12} - 10^{13}$ photons $sec.^{-1}$ $steradian^{-1}$ at 1026 Å. A gold-plated copper box surrounds the lamp to reduce rf interference and to isolate the hydrogen beam from the lamp's rf fields. The VUV radiation passes out of the box through a 2.0 cm by 2.0 cm opening covered with a gold mesh having 40 lines/cm and 82% transparency. The lamp's window is located 2 cm from the center of the beam and 25.0 cm down the beam from the multichannel array. At this point, ground state atoms are optically pumped to the metastable state. The diameter of the atomic beam is about 1.0 cm. In order to reduce scattered Lyman-α light from the lamp, an aperture is located about 31 cm from the multichannel array. This is a large gold-plated copper sheet 3 mm thick with a 1 cm hole for the beam to pass through.

The flux of metastable atoms is measured by applying an electric field to mix the $2S_{1/2}$ and $2P_{1/2}$ states and then detecting the emitted Lyman-α radiation when the atom decays to the ground state. This quenching region is located approximately 48 cm from the multichannel array where the atomic beam is 2 cm in diameter. The quenching electrodes [12] are arranged so that a well defined field occurs in a restricted portion of the atomic beam. This allows the metastable beam to be unperturbed except immediately in front of the Lyman-α detector where the fluorescence may be efficiently detected. The electrodes are a pair of bars 2.0 cm long, 0.5 cm wide and 0.15 cm thick. They are located symmetrically about the beam 2.0 cm apart so that the long axis of the bar is perpendicular to the beam. Each electrode is mounted on a grounded plane 6.5 cm square with a 0.1 cm insulating stand-off. The potentials applied to each electrode are equal in magnitude but of opposite polarity. The field at the center of the electrode structure is 0.2 V/cm per volt of applied potential difference. The entire electrode structure is gold plated and highly polished to reduce stray fields. It is estimated that 95% of the quenching occurs in 2.0 cm along the beam axis between the electrodes.

The Lyman-α fluorescence is detected with a benzene-filled, ionization detector. The detector has a cylindrical, gold-plated, copper body 3.5 cm in diameter and 7.5 cm long with a 0.05 cm diameter tungsten wire electrode down its center. On one side is a MgF_2 window 4.0 cm by 2.0 cm and 0.2 cm thick and on the other side is a high-vacuum valve which allows the detector to be removed from the chamber for refilling. The fill gas was 2.0 Torr benzene, 2.9 Torr argon and 9.8 Torr carbon dioxide. Under these conditions, the gain was unity to about 100 V and approximately exponential thereafter. The quantum efficiency of the benzene mixture is 40% and the transmission of the MgF_2 window is 50% for Lyman-α radiation. The detector is located 3.0 cm from the

quench region in a gold-plated copper box with a gold mesh window. This allows approximately 5% of the fluorescence to be collected.

The final element of the beam line is a VUV photodiode which facilitates the final beam alignment. It consists of a stainless steel photocathode and a collector electrode consisting of a stainless steel tube. After all of the elements in the chamber are properly arranged the hydrogen beam can be unambiguously aligned by maximizing the diode's photoelectric current. The multichannel array collimates the VUV radiation from the discharge bulb as well as the neutral atom flux so that one travels the same path as the other.

The flux of metastable $2S_{1/2}$ hydrogen was measured using the arrangement just described. The maximum metastable flux was obtained with the lamp containing a 1% hydrogen, 99% helium mixture, possibly due to low self-reversal of the Lyman-β line. For a beam intensity of 1.5×10^{16} atoms/sec, a quench voltage of ±25 V and a detector bias of 800 V, a metastable signal of 2.0 pA and a Lyman-α background of 120 pA was measured. When the beam, lamp and quench field were independently turned off and on the metastable signal appeared or disappeared as expected. The Lyman-α background comes from the lamp and could be eliminated by improved shielding. Using the previously given parameters for the transmission of the detector window, the detector's quantum efficiency, the collection efficiency for the fluorescence and the measured gain of the detector at 800 V, the current of 2.0 pA corresponds to a metastable flux of 2×10^6 atoms/sec.

The metastable signal measured as a function of the potential applied to the quenching electrodes is shown in Fig. 3. In order to eliminate the large Lyman-α background from the lamp, synchronous detection was used. A signal generator supplied a square wave voltage of 25 V and 1 kHz to a wire located between the aperture and quench electrodes. This caused the metastable flux to be quenched at this frequency while the lamp background was unaffected. The detector current went to a lock-in amplifier which was referenced by the signal generator. A least-squares fit to the data was made using the function,

$$Y = C_1[1 - \exp(-C_2 X)] \tag{1}$$

where Y is the detected signal and X is the quench voltage squared. The choice of this function is motivated by the Bethe-Lamb theory of the lifetime of the 2S state in external fields [1] [12]. The fit yields $C_1 = 5.61(14)$ and $C_2 = 0.0019(2)$. C_1 is the asymptotic limit for the saturation curve while C_2 is directly related to the beam's mean velocity.

An approximate calculation of the number of metastable atoms produced by the lamp may be done by considering the transverse velocity components of the beam. Since the hydrogen beam is collimated, the overlap of the Doppler broadened absorption line with the emission line of the lamp is reduced. An expression for the fraction of photons from the lamp absorbed by the beam is given by [13],

$$A_\alpha = \frac{k_0 \ell}{(1 + \alpha^2)^{1/2}} - \frac{(k_0 \ell)^2}{2!(1 + 2\alpha^2)^{1/2}} + \ldots \tag{2}$$

where k_0 is the atomic beam's Doppler absorption coefficient, ℓ is the absorption depth, and α is the ratio of the lamp Doppler width to the atomic-beam absorption width which is reduced by collimation. If the beam source is assumed to operate at 300 K and the lamp at 77 K, $\alpha = 29$, $k_0 = 0.19$ cm^{-1} for a beam flux of 1.5×10^{16} atoms sec^{-1} and $A_\alpha = 6.0 \times 10^{-3}$. The number of metastables produced is then given by the product of the fractional absorption A_α, the number of Lyman-β photons from the lamp, 10^{11} to 10^{12} sec^{-1}, and the optical pumping efficiency, 0.12. The result is 1.4×10^8 metastables per second. This is clearly more than was measured and reasons for this will be discussed later.

The number of metastable atoms F which decay in the quenching electrode region can be written as [1]

$$F = F_0[1 - \exp(-\ell/v\,\tau_S)] \tag{3}$$

where F_0 is the total metastable flux, ℓ is the length of the interaction region, v is the atomic velocity and τ_S is the lifetime of the 2S state induced by the electric field. The lifetime is given by $\tau_S = (E/475)^{-2} t_P$ for weak mixing [15] where E is the electric field in volts/cm and t_P is the lifetime of the 2P state. The average velocity may be used for v to a good approximation [14]. The velocity distribution of the metastable atoms produced by optical pumping can be obtained by first considering rate equations for the 1S, 3P and 2S three level system. Assuming the weak optical pumping limit where most atoms are in the ground state and $\dot{\eta}_{1S} = -\dot{\eta}_{2S}$ one obtains

$$\eta_{2S}(v) = W_L B_{31} \left(\frac{A_{32}}{A_{32} + A_{31}}\right) \eta_B f_B(v) \left(\frac{d}{v}\right) \tag{4}$$

where A and B are the Einstein coefficients, W_L is the energy density of the lamp photons, η_B the number density of the beam, d the interaction length and f_B is the Maxwell-Boltzman velocity distribution for the beam [16]. This indicates that the metastable atoms will have a velocity distribution $f_{2S}(v)$,

$$f_{2S}(v) = \frac{2}{\alpha^2} v \exp\left(-\frac{v^2}{\alpha^2}\right) \tag{5}$$

where $\alpha = \sqrt{2kT/m}$. The mean velocity for the distribution f_{2S} is $\bar{v} = \sqrt{\pi}\,\alpha/2$. This may be used to relate

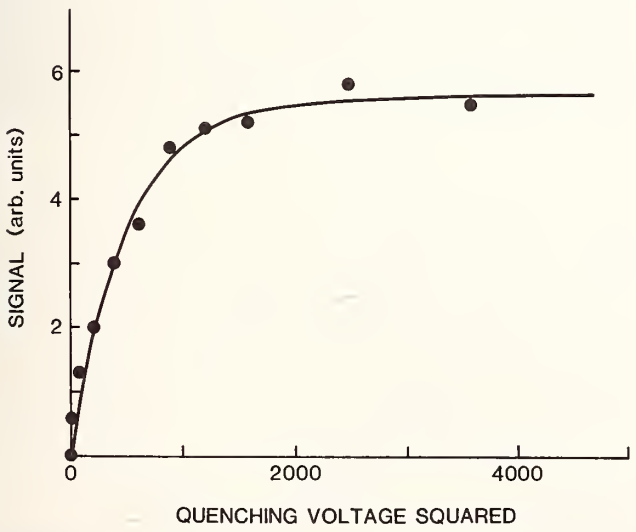


FIGURE 3. *Quenching curve for the metastable beam. The points are the measured values and the curve is from a least squares fit to the Bethe-Lamb theory with* $Y = C_1[1 - \exp(-C_2X)]$ *and* $C_1 = 5.61$ *and* $C_2 = 0.0019$.

Eq. (3) with Eq. (1) and the experimentally derived parameter C_2. For $C_2 = 1.9 \times 10^{-3}$, a metastable beam temperature T_B is found to be 100 K. This is much below the neutral beam temperature of 300 K. Although Eq. (5) indicates that the optical pumping process leads to a slower velocity distribution, its magnitude is not great enough to explain the metastable beam temperature.

The measured metastable flux at 10^6 atoms/sec is considerably less than the flux estimated by using Eq. (2). If the assumptions involved in this estimation are valid, a number of effects may be present to reduce the experimental yield, such as quenching due to stray fields, motional electric fields, blackbody radiation, and collisions with the background gas. The stray fields arise from contact potentials and accumulated charges on insulating films mostly arising from electrons produced by the photoelectric effect and the VUV radiation from the lamp. The gold plating reduces the accumulated charge, and from Eq. (3) it is clear that contact potentials on the order of 0.1 V are insufficient to explain the discrepancy. The motional Stark quenching may be readily calculated using standard techniques [1]. The magnetic field in the metastable beam was measured to be less than 30 milli Gauss, which leads to $\tau_S = 10^5$ sec, too large to explain the discrepancy. The $2S_{1/2}$ state may also be quenched by blackbody radiation which is present in the vacuum chamber at $T = 300$ K. Previous calculations of this effect [17] but ignoring the fine structure and Lamb shift find $\tau_S = 7.0 \times 10^4$ sec, while similar considerations for quenching through $P_{1/2}$ yield $\tau_S = 3.2 \times 10^5$ sec and through $P_{3/2}$, $\tau_S = 3.3 \times 10^3$ sec, which is again much too large. Finally, the quenching due to collisions with the background gas is especially effective at low velocities. Using a cross section of $\sigma = 2 \times 10^{-14}$ cm^2 [18] one finds a loss in beam intensity of only 1%. Thus it seems most likely that the discrepancy should be explained by an overestimation of the flux from the lamp or, more likely, that of the neutral hydrogen beam.

The mean velocity of the metastable beam is anomalously low at $T = 100$ K. Quenching due to the motional Stark effect is proportional to the velocity [14]. This would lead to a reduction in the temperature of the beam. However, as seen above, for the magnetic fields occurring in the vacuum chamber, this is a negligible effect. A more likely explanation involves the spectral overlap of the lamp at liquid nitrogen temperature with the absorption profile of the beam. Previous calculations have assumed that the radiation traveled perpendicular to the beam. This is not so since the window is only 2 cm from the beam. Atoms traveling towards the lamp have their absorption profile shifted up in frequency by the Doppler effect. However only the slower atoms can absorb the lamp radiation since it is emitted at a lower temperature. Atoms traveling away from the lamp have their absorption lines Doppler shifted down in frequency, but once again only the low velocity atoms can absorb the lamp's radiation. Therefore, the low velocity atoms are preferentially optically pumped and it is not surprising that the temperature of the beam roughly corresponds to the temperature of the lamp.

The source of metastable atomic hydrogen described here has a flux which compares favorably with other

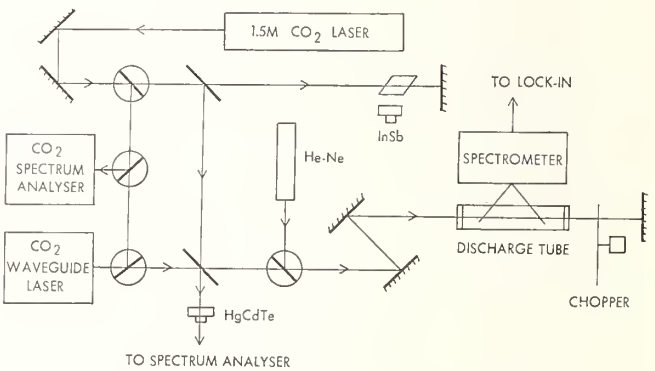


FIGURE 5. *Schematic of experiment.*

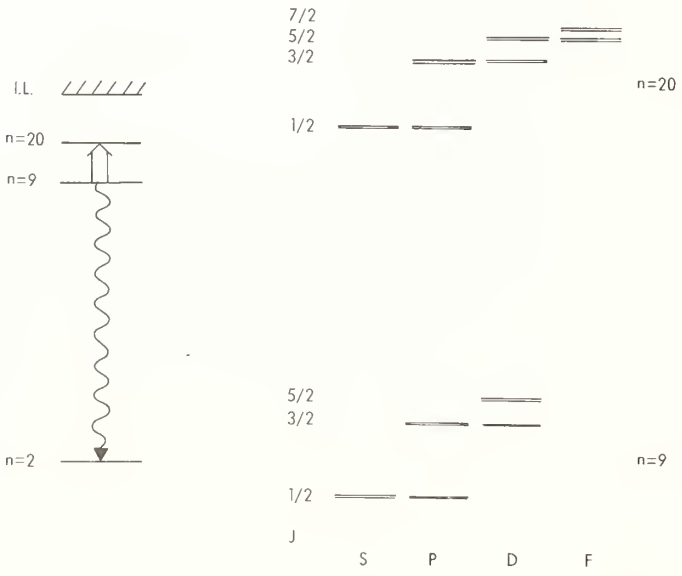


FIGURE 4. *Energy level diagram for saturated fluorescence in the Rydberg states of hydrogen.*

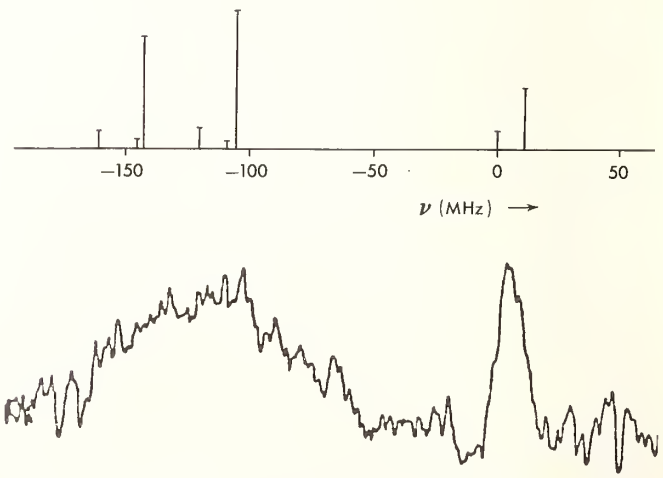


FIGURE 6. *Measured spectra for $n = 9$ to $n = 20$ transition.*

sources which use electron bombardment. Moreover, the beam spread due to recoil from absorption of the Lyman-β photon is only 0.1 degree and the second-order Doppler effect is only two parts in 10^{11} since a slow beam with a well defined velocity distribution is used. This suggests it may find use in precision experiments in optical spectroscopy or level crossing spectroscopy. In addition, there are several avenues for improvement. The Lyman-α background may be reduced by further shielding. The lamp may be scaled to larger lamps or arrays of lamps with higher flux. Laser sources may become available to greatly increase the radiation at 1026 Å [19]. Finally, intense cooled sources of atomic hydrogen beams may be used at liquid nitrogen [9] or liquid helium temperatures [20]. Such improvements could lead to a source which is ideally suited to experiments requiring high state selectivity and large intensity [21].

A precision measurement which would require the absence of perturbing electric and magnetic fields afforded by the atomic beam is indicated by Fig. 4. This relies on a coincidence of the 9 μm $R(20)$ CO_2 laser line and the transition from $n = 9$ to $n = 20$ in hydrogen. A preliminary measurement using a Woods discharge has been done. The technique of saturated fluorescence is used where the transition from $n = 9$ to $n = 2$ is monitored. Since the selection rules for dipole-allowed transition only connect the S, P and D levels of $n = 9$ to $n = 2$, the levels observed by the technique are shown in Fig. 4. The experimental arrangement is shown in Fig. 5. A CO_2 waveguide laser is used to induce the transitions. It is directed through a hydrogen discharge, chopped and reflected back on itself. The laser has a power of 500 mW, 1 GHz tunability, and less than 100 kHz passive stability. The $n = 9$ to $n = 2$ fluorescence is filtered by a 1/4-meter spectrometer and detected by a photomultiplier and lock-in detector. The frequency of the hydrogen transition is measured by heterodyning with a 1.5 m positive-column CO_2 laser in a HgCdTe diode. The beat signal from the diode is measured with a rf spectrum analyzer. The 1.5 m CO_2 laser can be locked with a CO_2 cell so that measurement of the absolute frequencies is possible [22]. The transitions to be expected are shown at the top of Fig. 6 [23]. It is noteworthy that the lifetimes of these states are relatively long. The lifetime of 9S is 1.3×10^{-6} sec while 20P is 1.4×10^{-6} sec. This gives a transition with a Q thirty times greater than the Balmer-α line. At the bottom of Fig. 6, a trace is shown of the spectrum actually observed. It is clear that the plasma environment of the discharge has considerably broadened the lines. The transition for 9S to 20P which occurs on the right side of the trace is the least broadened. This is possibly due to these levels having the smallest Stark polarizabilities. Undoubtedly, future measurement which can reduce external perturbations will yield improved results.

The author would like to thank R. D. Deslattes for valuable support at the initiation of this work, W. L. Williams for numerous suggestions and details of the quenching of metastable hydrogen, and L. Lèvy for information on the ionization detector which he designed and was used in this work.

References

[1] W. E. Lamb, Jr., and R. C. Retherford, Phys. Rev. **79**, 549 (1950).
[2] R. T. Robiscoe and T. W. Shyn, Phys. Rev. Lett. **24**, 559 (1970).
[3] B. L. Donnally, T. Clapp, W. Sawyer, and M. Schultz, Phys. Rev. Lett. **12**, 502 (1964).
[4] W. H. Wing, Phys. Rev. Lett. **45**, 631 (1980).
[5] H. Zacharias, H. Rottke, J. Danon, and K. H. Welge, Opt. Commun. **37**, 15 (1981).
[6] J. W. Heberle, H. A. Reich, and P. Kusch, Phys. Rev. **101**, 623 (1956).
[7] K. C. Harvey, Appl. Opt. **20**, 2883 (1981).
[8] C. B. Lucas, Vacuum **23**, 395 (1973).
[9] H. Wilsch, J. Chem. Phys. **56**, 1412 (1972).
[10] S. Czuchlewski, S. R. Ryan, and W. H. Wing, Rev. Sci. Instrum. **47**, 1026 (1976).
[11] W. R. Hunter and S. A. Malo, J. Phys. Chem. Solids **30**, 2739 (1969).
[12] R. T. Robiscoe, Phys. Rev. **138**, A22 (1965).
[13] M. W. Zemansky, Phys. Rev. **36**, 219 (1930).
[14] W. E. Lamb, Phys. Rev. **85**, 259 (1952).
[15] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Plenum, New York, 1977).
[16] G. Comsa, Vacuum **19**, 277 (1969).
[17] J. W. Farley and W. H. Wing, Phys. Rev. A **23**, 2397 (1981).
[18] S. J. Czuchlewski, Ph.D. thesis, Yale University, 1973.
[19] G. Pellegrini, Nucl. Instrum. Meth. **177**, 227 (1980).
[20] R. W. Cline, D. A. Smith, T. J. Greytak, and D. Kleppner, Phys. Rev. Lett. **45**, 2117 (1980).
[21] R. R. Lewis and W. L. Williams, Phys. Lett. **59B**, 70 (1975).
[22] K. M. Evenson, J. S. Wells, F. R. Petersen, B. L. Danielson, and G. W. Day, Appl. Phys. Lett. **22**, 192 (1973).
[23] G. W. Erickson, J. Phys. Chem. Ref. Data **6**, 831 (1977).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Precise Determination of the S and P Quantum Defects in Sodium and Cesium by Millimeter and Submillimeter Spectroscopy Between Rydberg States

P. Goy, J. M. Raimond, G. Vitrant, C. Fabre, S. Haroche, and M. Gross

Laboratoire de Physique de l'Ecole Normale Supérieure 24, rue Lhomond, 75231 Paris Cedex 05, France

Well-stabilized millimeter and submillimeter sources in the frequency range 50–500 GHz permit one to induce narrow transitions between Rydberg states of alkalis with the principal quantum numbers n in the range $23 \leq n \leq 41$. The levels are prepared by laser excitation. Detection of the atomic levels is performed through the selective field ionization technique. Precise experimental values for quantum defects and fine structure intervals are reported for the S and P states in sodium and cesium. The hyperfine structure of Rydberg states $n\,S_{1/2}$ and $n\,P_{1/2}$ has been observed for the first time in cesium. The extension of these experiments with increased accuracy to the spectroscopy of hydrogen would provide a new way to determine the Rydberg constant in frequency units.

Key words: alkali atoms; fine structure, hyperfine structure; millimeter and submillimeter waves; quantum defects; Rydberg constant; Rydberg states.

1. Introduction

Since they have a single valence electron, alkali atoms are like hydrogen in many ways. The simple energy spectrum $E = -R_H/n^2$ in hydrogen becomes $E = -R_A/(n - \epsilon)^2$ in alkalis, where the energy E is measured from the ionization limit, n is the principal quantum number, R is the Rydberg constant and ϵ is the quantum defect which reflects the perturbing effect of the core on the valence electron. The quantum defect ϵ depends on the alkali and on the angular momentum ℓ of the level. For a given species, it decreases with increasing ℓ values. For a given ℓ, it increases with the atomic number of the alkali. In first approximation, ϵ is independent of the principal quantum number n. Only very precise experiments make evident the systematic variation of ϵ with the binding energy. High resolution experiments performed in the optical range can determine the slight variation of ϵ with n. See for instance Refs. [1, 2] for the case of the quantum defects in cesium, $\epsilon_\ell^{Cs}(n)$.

In this paper, we present the most precise determination of the S and P quantum defects close to ionization. It is obtained by measuring the energy difference between successive excited states. The first state is excited from the ground state in an atomic beam by laser excitation. The transition to the second state is induced by a microwave source, the frequency of which is carefully measured. Due to the long lifetime of the Rydberg states (about 50 μs for n around 30), some very narrow (10 kHz range) resonances can be observed, yielding resonance Q-factors in the 10^7 range. The best atomic frequency determination is in the 10^{-8} range and the ultimate uncertainty in the determination of the quantum defect can be as low as a few times 10^{-7}.

2. Experimental Set-Up

2.1 Rydberg Atoms Excitation and Detection [3]

The experiment is performed on an alkali atomic beam (Na or Cs) propagating in a vacuum chamber with a background pressure of $10^{-6} - 10^{-7}$ Torr. From the ground state (3S or 6S), the atoms are excited to an nS state by a stepwise process, via the intermediate $3P_{3/2}$ (Na) or $6P_{3/2}$ (Cs) level, by two pulsed dye lasers. These two dye lasers (5896 Å and ~4110 Å for Na, 8521 Å and ~5100 Å for Cs) are themselves excited by the same pulsed pump laser of the nitrogen or Yag type.

After the pulsed laser excitation, the Rydberg states interact with the cw microwave radiation. The microwave power necessary to saturate transitions between close levels is very small: 10^{-6} W/cm^2 or less for two-photon transitions and 10^{-10} W/cm^2 or less for single photon transitions. The interaction lasts a variable time interval between 2 and 100 μs. Then the atoms enter between the plates of a capacitor. An electric field ramp is applied to the atoms. This ramp ionizes each Rydberg level at a characteristic field value. Thus each level is ionized at a specific time. The electrons produced are collected and amplified in an electron multiplier. This well-known field ionization technique [3] is selective and very sensitive: a single Rydberg atom gives a macroscopic current. The final state population is averaged over a few tens of laser pulses by a PAR 162 boxcar integrator, and recorded as a function of the microwave frequency.

2.2 The Microwave Set-Up

The microwave set-up is shown in Fig. 1. Among the microwave sources used were nine carcinotrons (Thomson-C.S.F. backwave oscillators) working respectively in the frequency ranges (in GHz): 51–61, 75–77, 77–82, 88–97, 91–100, 107–125, 226–274, 277–314, 417–485. The frequency F_c of a carcinotron is simply determined by the voltage applied to it. In order to count F_c and to stabilize it, we used as a reference an X-band klystron (X-13) locked to a quartz oscillator (MOS-5 microwave system). The frequency of the klystron F_k is directly measured on a counter, the clock reference of which comes from a long wave receiver for standard frequency with an accuracy better than 10^{-9} (Adret 4801 A). A mixer [4] using a packaged small Schottky diode placed across a waveguide generates the p^{th} harmonic of the klystron ($5 \leq p \leq 38$) and mixes it with the carcinotron

radiation. The beat note signal at the intermediate frequency $F_{if} = |F_c - pF_k|$ is amplified in the interval 0–500 MHz, counted and monitored on a spectrum analyzer (HP 141T, 8552B, 8553B). The beat note is also used to lock the carcinotron. Phase-locking carcinotrons is easy up to 200 GHz by comparing the phase of the beat signal with that of the signal given by a synthesizer. The linewidth of the free running carcinotron (~0.5 MHz) is then reduced below the resolution limit of the spectrum analyzer (a few Hz). Above 200 GHz, the phase noise of the multiplied frequency of the klystron becomes noticeable. Then the frequency-locking arrangement is more convenient. The linewidth of a 460 GHz carcinotron can thereby be reduced from 2 MHz to 1 kHz.

Above the cut-off of the waveguide the same harmonic mixer is used for all frequencies. For instance, $F \geq 80$ GHz is measured with an F-band (90–140 GHz) waveguide. Naturally the conversion loss increases with increasing harmonic mixing number p, going from 25 dB with $p = 7$ at $F_c = 80$ GHz, up to 85 dB with $p = 38$ at $F_c = 460$ GHz. Nevertheless the signal-to-noise is still reasonable at the upper frequency (≥ 20 dB in 100 kHz bandwidth).

The carcinotron can be very powerful (4 W at 80 GHz). Microwave leaks sometimes make it difficult to desaturate the most sensitive transitions for which powers in the nW range are sufficient. Instead of the carcinotron, we have also used the millimeter wave coming out of the harmonic mixer working as a frequency multiplier (top of Fig. 1). The harmonics $F = pF_k$ of the klystron frequency F_k are available at the millimeter waveguide output of the frequency multiplier as soon as their frequency is larger than the cut-off frequency of the waveguide. Starting with 100 mW from the klystron near 11.5 GHz, the following optimized microwave powers have been obtained: 0.1 mW at 80 GHz; 10 μW at 115 GHz; 1 μW at 230 GHz and so on [4]. Reducing this microwave power is easily achieved by reducing the klystron power. Practically all single-photon transitions with $F \leq 340$ GHz (P ≥ 1 nW) and two-photon transitions with F ≤ 110 GHz (P ≥ 10 μW) can be obtained with this very simple arrangement.

For any transition $n\ell \rightarrow n'\ell'$ the resonant frequency F is given by:

$$F/R = [n - \epsilon_\ell(n)]^{-2} - [n' - \epsilon_{\ell'}(n')]^{-2} \qquad (1)$$

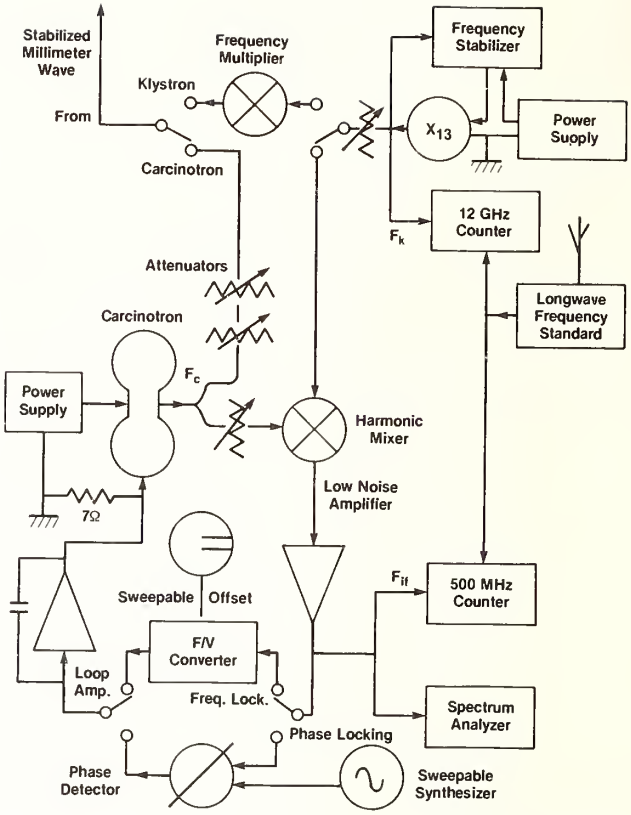


FIGURE 1. *Microwave arrangement. On the top right is the centimeter klystron, the frequency of which (F_k) is quartz-stabilized. This klystron can be used either of two ways: 1st—the centimeter power is sent down to a harmonic mixer, the carcinotron frequency F_c beats with the p^{th} harmonic of F_k and a feedback loop is used to stabilize the intermediate frequency F_{if}; 2nd—the centimeter microwave is sent up to a frequency multiplier, the millimeter frequency F is directly obtained at the multiplier output $F = pF_k$ and the carcinotron is not used.*

3. Experimental Results

3.1 Two-Photon Doppler Free nS → (n + 1)S Transitions in Sodium [5]

To cancel the Doppler broadening, it is possible to observe the two-photon absorption of two counter-propagating photons in the standing wave of a microwave cavity. This cavity is of the Fabry-Perot semi-confocal type and the atomic beam crosses it in a direction perpendicular to its optical axis. It is tuned by adjusting the distance between mirrors while monitoring the transmitted microwave power.

The following nS → (n + 1)S transitions have been observed: $n = 32$, 33, 34, 36, 39, and 40. With interaction times up to 100 μs the resonance linewidths are respectively: $\Delta F_c = 36$, 37, 28, 25, 11, and 13 kHz. The narrowest transitions (Fig. 2) are transit-time limited and the resonance Q is of the order of $Q = F_c/\Delta F_c = 59\ 260\ 221.5/11 \sim 5 \times 10^6$. The uncertainty in frequency position ±2 kHz includes an "instrumental" uncertainty of a tenth of a linewidth and a ±1 kHz uncertainty due to Stark shift induced by stray electric fields (a few tens of mV/cm).

Using Eq. (1) and the sodium Rydberg constant ($R_{Na} = 3.289\ 763\ 409(20)\ 10^{15}$ Hz), we can deduce the following formula for the S-quantum defect in sodium, valid for n ≥ 17:

$$\epsilon_S^{Na}(n) = 1.347\ 969\ 2(4) + 0.061\ 37(10)\, n^{*-2} \qquad (2)$$

where $n^* \simeq n - 1.348$.

Any nS → n'S two-photon transition in Na can thus be predicted with good accuracy for $n \geq 17$. Since the $n = 17$ level is 6 THz below the ionization limit, a large number of transitions in the 0–6 THz range could be absolutely calibrated, and could serve as frequency markers in the centimeter, millimeter, submillimeter and far-infrared range.

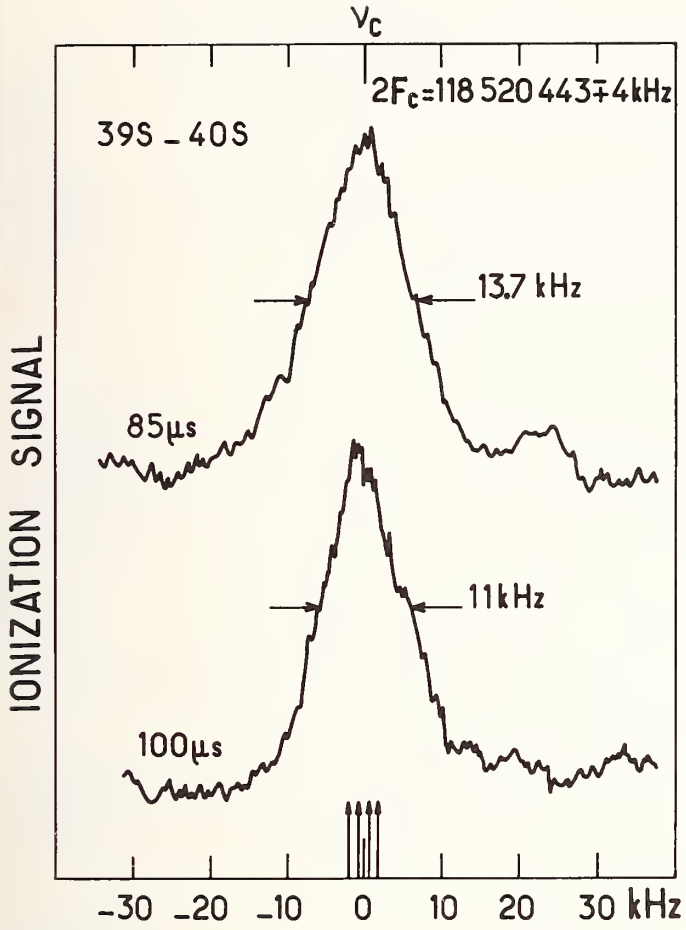


FIGURE 2. *Two-photon 39S → 40S transition in Na. The two traces are obtained for a Doppler free two-photon transition in sodium with a carcinotron stabilized frequency F_c around 59 GHz. Increasing the interaction time from 85 μs to 100 μs decreases the resonance linewidth in the same ratio: this suggests that this linewidth is transit-time limited. The arrows show the unresolved hyperfine structure difference of the $nS_{1/2}$ levels.*

3.2 One-Photon nS → n'P Transitions in Sodium [6]

Transitions $n\mathrm{S} \rightarrow n\,\mathrm{P}_{1/2}$ and $n\,\mathrm{S} \rightarrow n\,\mathrm{P}_{3/2}$ have been observed with n = 23, 24, 25, 32, 33, 34, and 36. We have also observed $n\mathrm{S} \rightarrow (n+1)\mathrm{P}$ transitions with n = 34 and 35. The uncertainty on the frequency measurements, of the order of 0.5 to 1 MHz, reflects the finite width of resonance (Doppler effect) and residual Stark and Zeeman effects.

According to Eq. (2), the S quantum defect is well-known for each transition. The ϵ_P quantum defects can then be obtained. One can deduce the following formula (which predicts the observed value within ±1 MHz):

$$\epsilon_{P_{1/2}}^{Na} = 0.855\ 424(6) + 0.122\ 2(2)\ n^{*-2} \qquad (3)$$

where $n^* \simeq n - 0.855\ 4$; and

$$\epsilon_{P_{3/2}}^{Na} = 0.854\ 608(3) + 0.122\ 0(2)\ n^{*-2} \qquad (4)$$

where $n^* \simeq n - 0.854\ 6$

The observed fine structure separation $\Delta F(n) = n\,\mathrm{P}_{1/2} \rightarrow n\,\mathrm{P}_{3/2}$ has an n^{*-3} dependence:

$$\Delta F^{Na}(n) = 5.367(5)\ 10^{12}(n - 0.855)^{-3}\,\mathrm{Hz} \qquad (5)$$

where 0.855 is the averaged $\mathrm{P}_{1/2}$, $\mathrm{P}_{3/2}$ quantum defect.

3.3 Two-Photon nS → (n + 1)S Transitions in Cs [7]

Due to their higher mass and lower operating temperature of the oven, the thermal velocity of Cs atoms is lower than that of Na. Thus the Doppler effect is less important, although it is not quite negligible (~100 kHz). The experiments described here are not performed in the cavity because a large fraction of the slow Cs atoms decays during the transit of the 50 mm diameter cavity.

The following $n\,\mathrm{S} \rightarrow (n+1)\mathrm{S}$ transitions have been observed: n = 27, 32, 33, 34, 35, 36, 37 and 38 with the following respective linewidths ΔF = 390, 210, 200, X, 250, 186, X and 430 kHz respectively (X when unmeasured). The interaction time is in the range 2 to 10 μs. The frequency determination accuracy is ±100 kHz and is essentially due to the Doppler effect and Stark effects from stray electric fields.

Using Eq. (1) and the value of the Rydberg constant in Cs (R_{Cs} = 3.289 828 299(20) 10^{15} Hz), one obtains the following expression for the S-quantum defects:

$$\epsilon_S^{Cs}(n) = 4.049\ 325(15) + 0.246\ 2(20)\ n^{*-2} \qquad (6)$$

where $n^* \simeq n - 4.0493$.

3.4 One-Photon nS → nP Transitions in Cs [7]

3.4.1 Quantum Defects of the nP Levels

The following transitions $n\mathrm{S} \rightarrow n\,\mathrm{P}_{1/2}$ have been observed: n = 24, 25, 26, 27, 28, 31, 32 and 38. Also the $n\,\mathrm{S} \rightarrow n\,\mathrm{P}_{3/2}$ transitions: n = 25, 26, 27, 28, 30, 32, 34, 35 and 38. For all these transitions a large splitting of the resonance signal is evident. It is due to the hyperfine structure of the $n\,\mathrm{S}_{1/2}$ initial state (see in Fig. 3 the $23\mathrm{S} \rightarrow 23\mathrm{P}_{3/2}$ transition split into two peaks "5" and "6"). Moreover, one can observe also a second splitting in the $n\mathrm{S} \rightarrow n\,\mathrm{P}_{1/2}$ transitions due to the hyperfine components of the $n\,\mathrm{P}_{1/2}$ levels (see in Fig. 3 the four peaks "1", "2", "3", and "4"). Taking into account this structure and measuring the frequency at the line center, we obtain from Eqs. (1) and (6) the following expressions:

$$\epsilon_{P_{1/2}}^{Cs}(n) = 3.591\ 56(3) + 0.3714(36)\,n^{*-2} \qquad (7)$$

where $n^* \simeq n - 3.5916$; and,

$$\epsilon_{P_{3/2}}^{Cs} = 3.559\ 06(3) + 0.374(4)\,n^{*-2} \qquad (8)$$

where $n^* \simeq n - 3.5591$

The observed $n\mathrm{S} \rightarrow n\mathrm{P}$ transition frequencies agree within ±3 MHz with the values calculated from Eqs. (1), (6), (7) and (8).

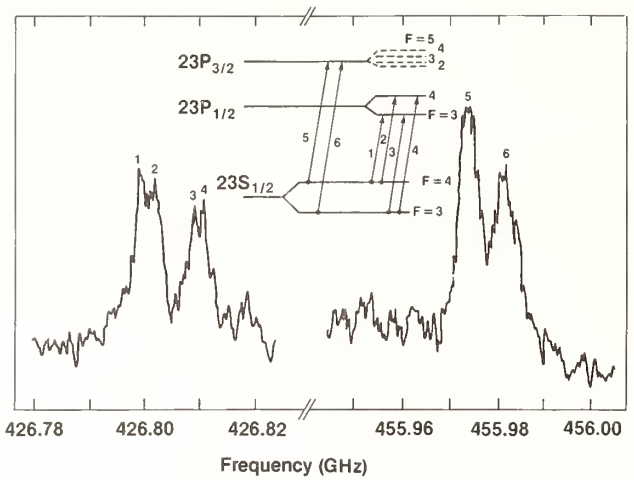


FIGURE 3. *Single photon 23S → 23P transition in Cs. The large frequency difference (29172 MHz) between the two groups "1–2–3–4" and "5–6" is due to the fine structure $23P_{1/2} - 23P_{3/2}$ of the 23P level. Also observed are: the hyperfine structure of the $23S_{1/2}$ level (9.2 MHz) in the separation "5" to "6" and "1–2" to "3–4", and the hyperfine structure of the $23P_{1/2}$ level on the splitting (2.2 MHz) "1" to "2" and "3" to "4". The hyperfine structure of $23P_{3/2}$ is unresolved here.*

3.4.2 Fine Structure of the *n*P Levels

Observed on the levels 23, 25, 26, 27, 28, 32 and 34, the fine structure interval is determined to a good relative accuracy due to its large absolute value:

$$\Delta F^{Cs}(n) = 2.138\ 3(2)\ 10^{14}(n - 3.5757)^{-3}\ \text{Hz}, \quad (9)$$

where 3.5757 is the average $P_{1/2}$, $P_{3/2}$ quantum defect.

3.4.3 Hyperfine Structure of the $nS_{1/2}$ Levels

The large splitting of the resonance traces of the $n\,S_{1/2} \rightarrow n\,P_{1/2}$ or $n\,S_{1/2} \rightarrow n\,P_{3/2}$ transitions gives the following experimental values of the hyperfine structure ($F = 3 \rightarrow F = 4$) of the $n\,S_{1/2}$ levels with the absolute accuracy ± 0.8 MHz: 9.2, 5.8, 5.7 and 4.7 MHz for n = 23, 25, 26 and 28 respectively (see for instance in Fig. 3 with n = 23 the distance "5" to "6" or "1–2" to "3–4").

3.4.4 Hyperfine Structure of the P Levels

The $n\,P_{3/2}$ levels are split in three sublevels giving rise to a pattern more complicated (unresolved on "5" and "6" in Fig. 3) than for the $n\,S_{1/2} \rightarrow n\,P_{1/2}$ transition. Observing only the double splitting of the $n\,S_{1/2} \rightarrow n\,P_{1/2}$ transitions (distance "1" to "2" and "3" to "4" in Fig. 3), we obtain the experimental values for the $P_{1/2}$ hyperfine splitting ($F = 3 \rightarrow F = 4$) with the absolute uncertainty ± 0.2 MHz: 2.23, 1.6, 1.25 and 1.1 for n = 23, 25, 26 and 28, respectively.

4. Conclusion

An improvement by a factor of 10 to 100 of the resolution obtained here is certainly possible. Cooling the atomic beam apparatus to liquid He temperature should suppress blackbody effects and cancel stray electric fields [8]. One can imagine for instance that for the two photon Doppler free resonance, the natural linewidth ~1 kHz could be observed. In this way a relative accuracy in the range 10^{-10} on the frequency could be obtained on $n\,S \rightarrow (n + 1)S$, $n\,S \rightarrow (n + 3)S$, $n\,S \rightarrow (n + 5)S$ . . . transitions. The alkali Rydberg levels would provide a secondary frequency standard for the calibration of electromagnetic radiation from the microwave to the far infrared domain.

The extension of these experiments to the spectroscopy of hydrogen Rydberg levels is worth considering [9]. In this case, the quantum defects vanish and the Rydberg constant R_H can be directly measured in frequency units.

Thanks are due to Dr. R. Adde (Electronique Fondamentale, Orsay) and to Dr. A. Clairon (LPTF, Observatoire de Paris) for the lending of equipment. One of the authors (P. G.) would like to express his gratitude to A. Clairon for many fruitful discussions.

References

[1] C. J. Lorenzen, K. H. Weber, and K. Niemax, Opt. Comm. 33, 271 (1980).
[2] C. J. Lorenzen and K. Niemax, J. Quant. Spectros. Radiat. Transfer 22, 247 (1979).
[3] C. Fabre, S. Haroche, and P. Goy, Phys. Rev. A 18, 229 (1978).
[4] P. Goy, Int. J. Infrared Millimeter Waves 3, 221 (1982).
[5] P. Goy, C. Fabre, M. Gross, and S. Haroche, J. Phys. B 13, L 83 (1980).
[6] C. Fabre, S. Haroche, and P. Goy, Phys. Rev. A 22, 778 (1980).
[7] P. Goy, J. M. Raimond, G. Vitrant, and S. Haroche, Phys. Rev. A 26, 2733 (1982).
[8] F. Witteborn and W. Fairbank, Rev. Sci. Instrum. 48, 1 (1977).
[9] D. Kleppner, Bull. Am. Phys. Soc. 20, 1458 (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A New Method for Measuring the Fine Structure Constant Using Stark Spectroscopy

Michael G. Littman

**Department of Mechanical and Aerospace Engineering, Princeton University
Princeton, NJ 08540**

and

William D. Phillips

Electrical Measurements and Standards Division, National Bureau of Standards, Washington, DC 20234

An experiment to determine a value for α, the fine structure constant, is proposed. The determination is to be based on a measurement of the Stark effect of hydrogen Rydberg states. Hydrogen atoms in a uniform field of known strength will be excited to Rydberg levels using intense tunable lasers. The presence of excited atoms will be detected using the sensitive technique of field ionization. A precise measurement of the linear energy shift of an extreme Stark state is to be made, and from this measurement, in conjunction with reported values of R_∞, the Rydberg constant, and $2e/h$, the Josephson frequency-voltage ratio, a value of α will be determined. The estimated accuracy of the determination is expected to be competitive with or better than the 0.11 ppm accuracy of the current best non-QED determination of α.

Key words: fine structure constant; hydrogen atom; Rydberg states; Stark effect.

1. Introduction

The need for an improved value for α, now known to an uncertainty of 0.11 ppm [1], is almost self-evident. The fine structure constant is the primary expansion parameter used in quantum electrodynamics (QED). An accurate value of α is needed for virtually any comparison of QED theory with experiment. Currently, there are several experiments whose comparison with theory is limited by the uncertainty in α. Of these, one of the most notable is the recent "g-2" determination by researchers at the University of Washington [2]. Here, the uncertainty in the measurement of the electron anomalous moment is less than the uncertainty imposed by α on the theoretical prediction. A 2.5 standard deviation discrepancy[1] with QED theory is obtained [1]. The value of α itself is also subject to controversy. While the best QED-independent value of α is based on measurements of the gyromagnetic ratio of the proton, γ'_p, a discrepancy of 6 ppm (12 standard deviations) exists between recent precision measurements of γ'_p [1, 3]. This represents a 3 ppm discrepancy in α. Such discrepancies point to the need for redundant, independent determinations of α.

The motivation for the demonstration of the usefulness of precision Stark spectroscopy is also clear. In addition to the obvious contributions to the area of fundamental constants, precision Stark spectroscopy can be of considerable importance in the area of precision electrical measurements. Of particular interest are the possible applications to transfer voltage standards and the measurement of large potential differences.

2. Theory

In the following discussion we will summarize the theoretical basis for the measurement and some of the more important experimental considerations.

In an electrical field of magnitude F, the non-relativistic energy eigenvalues for hydrogen are given [4] by the perturbation expansion (atomic units[2])

$$W = -\frac{1}{2n^2} + \frac{3}{2}nkF - \frac{1}{16}n^4F^2 \times (17n^2 - 3k^2 + 9m^2 + 19) + \ldots \quad (1)$$

Here n, m, and k are, respectively, the principal, azimuthal, and electric quantum numbers [m can assume integral values between $(n-1)$ and $-(n-1)$, and k can assume bi-integral values between $(n - |m| - 1)$ and $-(n - |m| - 1)$]. A map of levels having $n = 12 - 14$ and $|m| = 0, 1$ according to Eq. (1) is given in Fig. 1. It is clear from this Stark map that the linear effect is dominant over the range of fields presented.

To derive an expression for α, consider the linear term of Eq. (1),

$$W_1 = 3nkF/2. \quad (2)$$

[1] Recent calculations by Levine and Roskies and by Kinoshita and Lindquist (elsewhere in these proceedings) have reduced this discrepancy.

[2] The field is expressed in units of that field produced by one electron charge at a distance equal to the Bohr radius, and the energy in units of the electrical potential energy of a pair of charges separated by a Bohr radius.

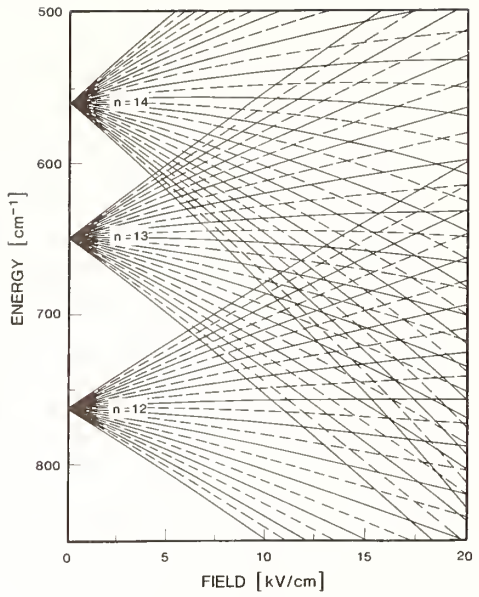


FIGURE 1. *Hydrogen Stark structure according to perturbation theory. The $m = 0$ levels are indicated by solid lines; the $|m| = 1$ levels are indicated by dashed lines.*

Converting to SI units and expressing the energy shift in terms of a frequency, ν_1, we have:

$$h\nu_1 = 3nkEea_0/2$$

where a_0 is the Bohr radius, E is the field in V/m, and e is the electron charge. Using the relation $R_\infty = \alpha/4\pi a_0$, we can write

$$\nu_1 = \frac{3}{16} nkE \left(\frac{2e}{h}\right) \alpha R_\infty^{-1} \quad . \tag{4}$$

Although Eq. (4) is expressed in SI units, it is easy to see that the expression is still valid if the electric field, E, is expressed as E_{NBS} (i.e., is measured in units of V_{NBS}/meter where V_{NBS} is the NBS unit of voltage) and the Josephson frequency-voltage ratio, $2e/h$, is expressed in units of $\mathrm{Hz}/V_{\mathrm{NBS}}$. Since the NBS volt is realized in terms of $2e/h$ to 0.03 ppm [5] and R_∞ is known to 0.001 ppm [6], the determination of α at the 0.1 ppm level will not be limited by the accuracy of the other constants in Eq. (4).

It can be shown that higher (N^{th}) order terms in the Stark shift may be written (in SI units) as:

$$\nu_N = C\, E^N \left(\frac{2e}{h}\right)^N \alpha^N c^{N-1} R_\infty^{2N-1} \quad , \tag{5}$$

where C is a numerical constant which depends on the quantum numbers. Only the speed of light, c, appears as an additional constant. Since c is known as well as the meter can be realized, the contribution due to higher order terms in the Stark expansion will not reduce the accuracy of an α determination as long as they are not too large.

In view of the fact that the proposed precision measurement is based upon a perturbation expansion of the energy, the general convergence properties of the series are of particular importance. Several recent theoretical studies have considered the asymptotic nature of the perturbation expansion. Koch, using the tabulation of the terms of the expansion through 150$^{\mathrm{th}}$ order by Silverstone and co-workers [7], observes that if n^4F approaches unity, the lack of convergence or, in some cases, the convergence to the wrong value, is striking [8]. He notes that the use of Padé approximants allows the series to be resummed with improved convergence. Others have examined the strong field effects on hydrogen Stark states using exact numerical techniques [9]. To date, there have been few experimental checks of these calculations; however, the different theoretical approaches appear to be in reasonable agreement.

The lack of convergence of the perturbation expansion reflects the fact that Stark states are quasi-bound. In any given field there is a finite probability that levels will ionize spontaneously. In weak fields, levels ionize by a quantum-mechanical tunneling process, while in strong fields they ionize classically by rolling "over-the-barrier". In practice, ionization effects are important only for fields on the order of or larger than the classical threshold field, $F_c = (W^2/4) \simeq (1/16n^4)$ [10]. For fields less than F_c, decay is dominated by radiation with an average decay rate that decreases with increasing n as $n^{-4.5}$ [4, 10]. In other words, hydrogen Stark states in low fields are quasi-stable and should be well described by a perturbation approach. It is primarily for these reasons that we restrict our attention to the low field regime.

Due to the precise nature of the proposed measurements, the effects of fine and hyperfine structure, the Lamb shift and the ac Stark effect cannot be ignored. Even though we will not discuss these effects here, in the final analysis they will be taken into account.

The proposed experiment, which is described in the following section, requires the optical excitation of an extreme $n \simeq 35$ Stark state from a low lying $n = 2$ level. In order to determine the optically accessible states it is necessary to understand fully the symmetry of the states involved. It is well known that the degeneracy in hydrogen is broken by the Lamb shift. In sufficiently weak fields, that is, fields in which the Stark interaction is small compared with the degeneracy-breaking Lamb shift interaction, the spherical character of a given state is maintained. For the low lying $|2\,^2S_{1/2}>$, the electric interaction is important only above 300 V/cm. Below this field value, the state is nearly spherical. In contrast, highly excited levels assume a parabolic nature at very small fields. This situation in hydrogen at low fields in which the lower state is spherical and the upper states are parabolic, is analogous to the situation that exists in alkali metals in weak or strong fields. In the alkali case all parabolic Rydberg states of a given m are accessible from the spherical ground state [11]. We expect, likewise, that all hydrogen parabolic states, including the extreme ones, will be accessible in weak fields.

Recent experimental studies of the Stark effect have dealt primarily with hydrogen-like atoms. Littman and co-workers have measured the Stark structure of sodium using tunable lasers and atomic beams [12]. They have demonstrated the feasibility of selective pulsed laser excitation of each of the Stark states of a given azimuthal quantum number from a low-lying level. Koch has measured the position of several Stark states in atomic hydrogen using quite different methods [13]. In Koch's experiments a fast excited hydrogen beam is further excited to Rydberg levels using a fixed frequency CO_2 laser. The Doppler effect is used for tuning and field ionization is used for detection. Together these experiments

demonstrate that indeed one can use conventional techniques to excite and detect hydrogenic Stark states.

3. Experiment

The plans for measuring ν_1 and E_{NBS} are presented here. These two quantities are to be measured separately, with the ultimate goal being the determination of the quotient, ν_1/E_{NBS}, to 0.10 ppm or better. In the interest of brevity we consider only the more important aspects of the proposed experiment.

A schematic of our experimental configuration is given in Fig. 2. We require a thermal atomic beam source, a uniform field region, intense pulsed and cw lasers, and a field-ionization detection region.

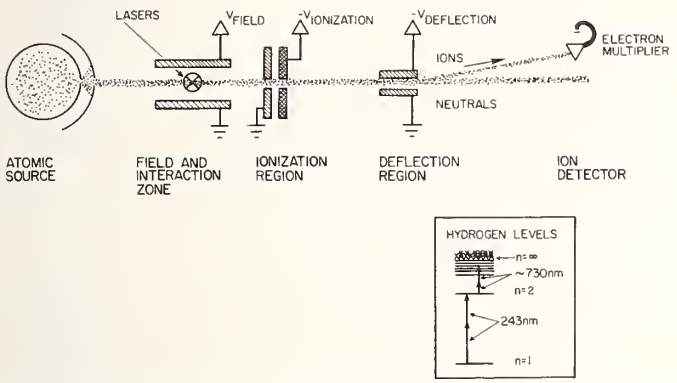


FIGURE 2. *Experimental arrangement. The excitation scheme is depicted in the inset.*

The atomic source will provide a ground state density of 10^{11} atoms/cm; the mean atomic velocity will be about 10^5 cm/sec. Excitation will occur in the static field in two two-photon steps, as shown in the inset in Fig. 2. We plan to excite Stark states having $n \simeq 35$ from the ground state using the perturbed $|2\,^2S_{1/2}>$ level as a resonant intermediate state. Both steps of excitation will take advantage of Doppler-free counter-propagating excitation. The first step, which serves to populate the intermediate level, will be initiated by a 25 MW/cm^2 burst of radiation at 243 nm (2.5 mJ/10 ns focused to 1 mm^2). For this excitation step we will use a laser system similar to that used by Bjorklund and Freeman in their work on three photon excitation of atomic hydrogen [14]. The second step of excitation will be driven by 30 W/cm^2 of continuous radiation at 730 nm from a ring dye laser (300 mW focused to 1 mm^2).

The transition linewidth is an important and limiting factor in the measurement procedure. The minimum linewidth expected as a result of transit time broadening in the cw laser will be about 100 kHz. (Because of their motion, the atoms sample the optical field only for about 1 μsec.) Additional linewidth broadening is expected as a result of the field-induced quenching of the $|2\,^2S_{1/2}>$ level. To prevent significant broadening we must restrict the external field to values below 40 V/cm. (At 40 V/cm less than 1% of the nearby rapidly-decaying $|2\,^2P_{1/2}>$ level is mixed.)

The restriction of a maximum field of 40 V/cm has a number of implications. First, the shift between the extreme high-n Stark components is limited to 8 cm^{-1}. Since the plan is to determine ν_1 to 0.10 ppm, the shift must be measured to about 20 kHz. This involves splitting the line by 1 part in 10. Second, corrections due to quadratic and higher order perturbation terms can be made reliably. Here the contribution of the higher order terms to the total shift is less than a few percent. Finally, this field region is well below the $n = 35$ ionization field of 250 V/cm, and thus ionization effects will not be important.

The difficult aspect of this experiment deals with the precise measurement of the field strength. To determine the field we must accurately measure an electrical potential difference and a distance. In addition, we must take into account a number of effects which would influence the strength of field that a given atom sees. Effects such as contact potential, the "patch" effect, fringing field, motional electrical field, and other external stray fields, such as those produced by photoions and photoelectrons created during laser excitation, can influence the precision of our determination. Fortunately, we have the flexibility to work with many different states in many different fields, so that we have natural internal consistency checks on our measurements. If necessary, we can even use the atoms to map out the field to determine local inhomogeneities.

The field region will consist of a vacuum-spaced capacitor and an associated precision voltage source as shown in Fig. 3. The field plates will be thin metallic films evaporated onto precision optical flats (λ/200). In addition, three symmetrically-placed, dielectric mirror pads will be deposited on each of the substrates. These mirror pairs will constitute high finesse Fabry-Perot interferometers which, when used in conjunction with a stabilized

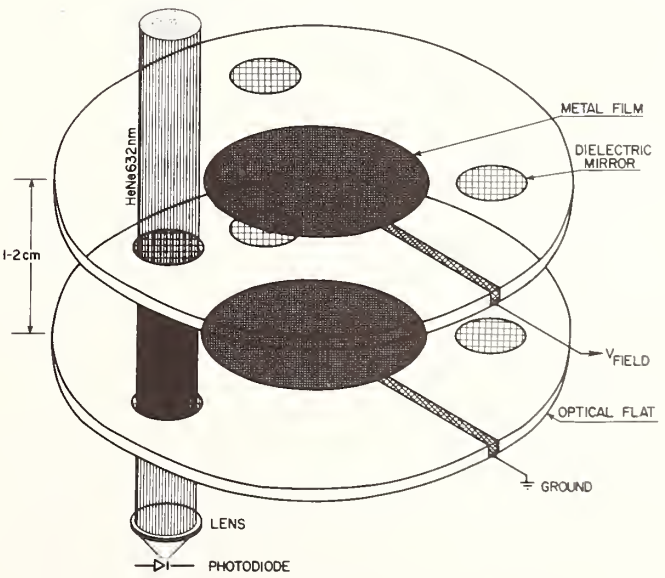


FIGURE 3. *Precision vacuum-spaced capacitor.*

helium-neon laser, will help determine surface parallelism and relative plate separation. Each of the optical flats will be attached to a long-range piezoelectrically-driven translator which will be used to effect the change in plate separation. The field measurement is to be made as follows: First, the lasers will be tuned to an atomic Stark resonance. Next, the plate separation will be increased carefully by about 1 cm. Finally, the electrical potential difference will be readjusted to bring the sytem back into resonance. The change in electrical potential divided by the change in plate separation is the field strength. We assume, of course, that fringing field effects can be taken into account to this level of precision. The differential measurement helps to eliminate problems of contact potential and difficulties in the absolute determination of surface positions. The interferometers will be used to determine the change in plate separation to better than 10 Å. A precision voltage divider will be used to determine the change in electrical potential in terms of the NBS voltage standard to better than 4 μV.

Field induced ionization will be used to detect the excited atoms [15]. Since the atoms will drift approximately 5 cm before entering the detection region, we can anticipate a loss of signal by a factor of 10 due to radiative decay and blackbody-induced diffusion of excited state population [16]. The remainder of the excited atoms, including those that have diffused to high-n states, will be detected with near unity efficiency.

We expect a signal of about 3 counts per laser shot. A random noise contribution due to several effects including multiplier dark noise of about 0.1 count per laser shot is expected. Given that the anticipated signal-to-noise ratio is 30, we should have no trouble in splitting the line to the needed factor of 10. Thus, ν_l / E_{NBS} should be able to be determined to the required precision of 0.10 ppm.

References

[1] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. **42**, 1575 (1979). (This is the best determination of α in which QED is not used for interpretation of the experimental results.)

[2] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Phys. Rev. Lett. 38, 310 (1977); and these proceedings.

[3] P. Vigoureux and N. Dupuy, Natl. Phys. Lab. Report DES 59 (January 1980).

[4] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One and Two Electron Atoms* (Academic Press, New York, 1957).

[5] B. F. Field, T. F. Finnegan, and J. Toots, Metrologia **9**, 155 (1973).

[6] J. E. M. Goldsmith, E. W. Weber, and T. W. Hänsch, Phys. Rev. Lett. 41, 1525 (1978).

[7] H. J. Silverstone, B. G. Adams, J. Cizek, and P. Otto, Phys. Rev. Lett. **42**, 1498 (1979).

[8] P. M. Koch, Phys. Rev. Lett. 41, 99 (1978); H. J. Silverstone and P. M. Koch, J. Phys. B **12**, L537 (1979).

[9] R. J. Damburg and V. V. Kolosov, J. Phys. B **9**, 3149 (1976); **11**, 1921 (1978); **16**, 2637 (1979); and references therein.

[10] M. G. Littman, M. M. Kash, and D. Kleppner, Phys. Rev. Lett. 41, 103 (1978).

[11] M. L. Zimmerman, M. G. Littman, M. M. Kash, and D. Kleppner, Phys. Rev. **20**, 2251 (1979).

[12] M. G. Littman, M. L. Zimmerman, T. W. Ducas, R. R. Freeman, and D. Kleppner, Phys. Rev. Lett. **36**, 788 (1976).

[13] P. M. Koch, Opt. Commun. **20**, 115 (1977).

[14] G. C. Bjorklund, C. P. Ausschnitt, R. R. Freeman, and R. H. Storz, Appl. Phys. Lett. 33, 54 (1978).

[15] T. W. Ducas, M. G. Littman, R. R. Freeman, and D. Kleppner, Phys. Rev. Lett. 35, 366 (1975).

[16] If n is greater than about 10, the apparent decay rate is modified by the presence of room temperature blackbody radiation, which serves to redistribute the excited population among many nearby levels; T. F. Gallagher and W. E. Cooke, Phys. Rev. Lett. **42**, 835 (1979); John W. Farley, private communication.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Time Resolved Sub-Natural Width Spectroscopy

William D. Phillips

Electrical Measurements and Standards Division, National Bureau of Standards, Washington, DC 20234

and

Harold J. Metcalf

Physics Department, State University of New York at Stony Brook, Stony Brook, NY 11794

A number of techniques exist in both practice and theory for achieving spectral signals narrower than the width imposed by the natural lifetime. We examine a simple mathematical model for sub-natural width spectroscopy, and identify two distinct characteristics of time resolved line narrowing. We discuss the conditions where line narrowing techniques may be valuable, and comment on some misconceptions concerning these techniques.

Key words: high resolution spectroscopy; precision measurement; spectral line narrowing; sub-natural linewidth spectroscopy; time-resolved spectroscopy.

1. Introduction

In 1950 Ramsey [1] introduced the separated oscillatory field method for atomic beam magnetic resonance spectroscopy. This method produces spectral features of width characterized by the total time the atoms spend in the apparatus, rather than the time spent interacting with the radiation. In 1960 Hughes [2] suggested that Ramsey's technique could also be used for short-lived states to produce spectral features that are narrower than the natural, decay broadened linewidth. This possibility was particularly attractive because the ultimate precision of many spectroscopic measurements is limited by the natural width imposed by the decay of one or both of the states involved in the transition being studied.

Since 1960, many specific experimental techniques have been proposed and used to achieve signals narrower than the natural linewidth [3–16]. An example of particular interest in the field of precision measurements and fundamental constants is the determination of the Lamb shift [5]. Successful applications of sub-natural linewidth spectroscopy span the electromagnetic spectrum from the Lamb shift [5] at 10^9 Hz to Mossbauer intervals [7] at 10^{18} Hz. All of the various techniques achieve line narrowing by using time-resolved techniques for preparation and observation of the system being studied and therefore must be amenable to temporal resolution. These techniques always involve sacrificing signal amplitude in favor of reduced spectral width. Usually a linear decrease in width results in an exponential loss of signal.

In earlier work [16] we numerically modeled the application of line narrowing techniques to a noisy signal. We concluded that it was advantageous to trade signal-to-noise ratio (S/N) for increased resolution in the case where the determination of the spectral line center was limited by the presence of an unknown asymmetry. Errors due to asymmetries are reduced in proportion to the amount of narrowing achieved, so that if there is sufficient S/N the accuracy of the line center determination is improved even if the precision is not. (The possibility of just such a systematic error is a major motivation for the use of line-narrowing techniques in Lamb shift [5] and level crossing [3] experiments.) If the line is known to be symmetric, or if the exact line shape is known, then line narrowing techniques do not improve the center determination.

In this paper we use our mathematical model for line narrowing to identify two distinct features of the process that we have called time-delayed interference (TDI) and decay envelope masking (DEM). We compare the relative advantages of these different methods for line center determinations and for interpretation of complex, unresolved spectra.

2. The Model

Consider a decaying oscillation given by

$$s(t) = s_0 \cos(\omega_0 t + \phi) e^{-\gamma t}. \tag{1}$$

This might represent, for example, the electric field radiated by an atom in free induction decay, or the oscillating portion of a quantum beat signal observed on a photomultiplier. The spectrum of this signal is obtained by taking the Fourier transform of $s(t)$:

$$S(\omega) = \int_0^\infty e^{-i\omega t} s(t)\, dt = \frac{S_0}{\gamma} \frac{e^{i\phi}}{ix - 1}, \tag{2}$$

where $x = (\omega_0 - \omega)/\gamma$ and we have ignored the "anti-resonant" terms. When $\phi = 0$ the real part of the spectrum is an absorptive Lorentzian with full width at half maximum (FWHM) of 2γ and the imaginary part is a dispersive Lorentzian. The power spectrum, $P(\omega)$, obtained by taking the absolute square of the spectrum $S(\omega)$, is a purely absorptive Lorentzian with FWHM of 2γ independent of ϕ. The signal, the real part of the

spectrum, and the power spectrum are shown in Fig. 1(a).

Time resolution, and thus line narrowing, is obtained by applying a "window function" $W(t)$ to the signal $s(t)$, and then performing the Fourier transform:

$$S(\omega) = \int_0^\infty e^{-i\omega t} W(t)\, s(t)\, dt \ . \tag{3}$$

Window functions that discriminate against the early part of the signal in favor of the later part in general will produce line narrowing.

In the following sections we examine the action of two different window functions: a rectangular window and an exponentially rising window. We shall consider how the line narrowing arises in each of these cases, and show that there are some fundamental, qualitative differences between the two methods under the same measurement conditions. The existence of two qualitatively different types of line narrowing has not been appreciated before.

3. Time Delayed Interference

Consider a rectangular window function $W(t)$ given by:

$$W(t) = \begin{matrix} 0, & t < t_0 \\ 1, & t \geq t_0 . \end{matrix} \tag{4}$$

Ignoring anti-resonant terms, the spectrum is

$$\begin{aligned} S(\omega) &= \int_0^\infty e^{-i\omega t} W(t)\, s(t)\, dt \\ &= \frac{S_0}{\gamma} \frac{e^{i\phi}}{ix - 1} e^{(ix-1)T} , \end{aligned} \tag{5}$$

where $T = \gamma t_0$. For $\phi = 0$, the real part of the spectrum is given by

$$\mathrm{Re}\{S(\omega)\} = \frac{S_0}{\gamma} \frac{e^{-T}}{1 + x^2} (x \sin xT - \cos xT) \ . \tag{6}$$

This function, shown in Fig. 1(b) for $T = 4$, is a damped oscillatory function, symmetric about $x = 0$. The central peak of the spectrum is considerably narrower than it would be without time resolution. For $T >> 1$ the FWHM of the central peak (with the maximum height of the peak measured from zero) is approximately $2/T$. The amplitude of the central peak is reduced by e^{-T} from that of the unnarrowed spectrum.

Note that the envelope of the spectrum is quite broad even though the central feature of the spectrum given by Eq. (6) is narrow. This contradicts the notion that the window function discriminates against that portion of the signal contributing to the wings of the spectrum in favor of that contributing to its center. To illustrate this point more clearly, consider the power spectrum for the rectangular window:

$$P(\omega) = |S(\omega)|^2 = \frac{S_0^2}{\gamma^2} \frac{e^{-2T}}{1 + x^2} . \tag{7}$$

Except for a reduction in amplitude, the power spectrum is unchanged by this windowing. Furthermore, the shape is independent of ϕ since the process of taking the absolute square destroys phase information. While one often hears that time resolved line narrowing techniques work because the detection process looks only at "long-lived" atoms that exhibit a narrower spectrum, it is clear from this analysis that late-arriving radiation shows exactly the same spectral distribution as the entire ensemble.

The narrow features of Eqs. (5) and (6) are not, therefore, due to the selection of late radiation alone, but rather to interference between two oscillations. The accumulated phase difference between these changes rapidly for small changes in x when T is large, and accounts for the narrow features. The narrowing is lost in the power spectrum because the phase information is lost. In the separated oscillatory fields method, the relevant phase difference is between the oscillating moment of the atoms under observation and the applied oscillatory field; in time resolved level crossing [3] the relevant phase information is preserved by a mixing between the states involved in the crossing. This description of the narrowing that occurs in such experiments has sometimes been appreciated but popular and erroneous explanations of the narrowing have also persisted in the literature.

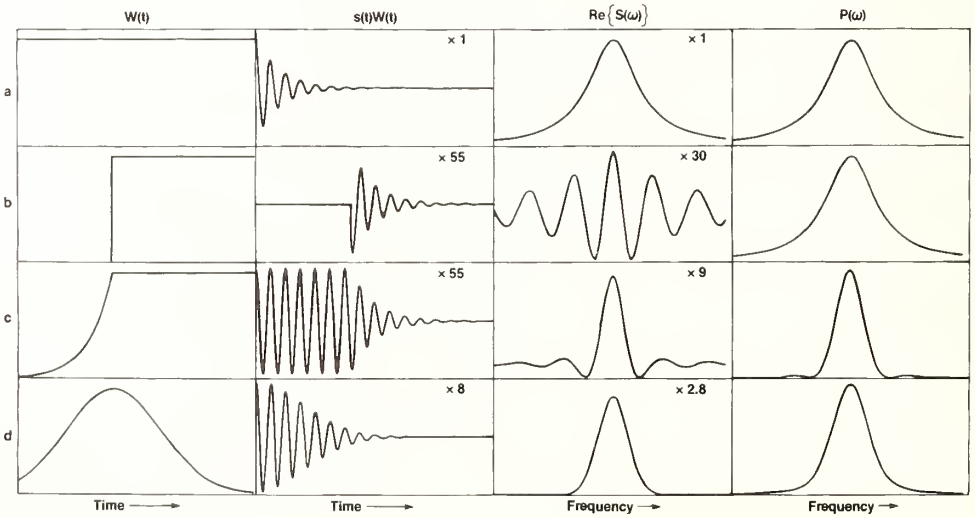


FIGURE 1. *Signals and spectra for a decaying oscillation observed through various windows. Window functions, $W(t)$; windowed signals, $s(t)W(t)$; real part of the Fourier transform, $Re\{S(\omega)\}$; and power spectrum, $P(\omega)$ are shown for no window (a), rectangular window (b), exponential window (c), and Gaussian window (d). Scale factors indicate the relative expansion of the vertical scale.*

As a further example to emphasize the importance of phase preservation along with time delay, consider measurement of the spectrum of a radiating atom with a slowly swept Fabry-Perot interferometer followed by a detector such as a photomultiplier. By repeated observations of the radiation described by Eq. (1), one would obtain the power spectrum of the emission, a Lorentzian with FWHM of 2γ. If a shutter, synchronized with the excitation of the atom, were placed between the atom and the spectrometer, and only opened after a long delay, the same power spectrum would be recorded, only with smaller amplitude. If, on the other hand, the shutter were placed between the interferometer and the detector, a spectrum similar to that of Eq. (6) would be obtained because light emitted early by the atom is stored in the interferometer, and interferes with light emitted later. The shutter does not select the light from atoms which emit late, but rather it selects that time when the interference between the early and late emissions is observed. Reference [17] gives a more detailed analysis of a very similar situation involving the Mossbauer spectrum of gamma rays passed through a filter and then detected with time resolution.

4. Decay Envelope Masking

We have seen in the previous section that the rectangular step window in Eq. (4) has no effect on the shape of the power spectrum of the signal given by Eq. (1). This is a special result for that particular window and signal; if the rectangular window function goes to zero at some time $t > t_o$, the power spectrum becomes broader but the spectrum of Eq. (6) still retains its narrow features. It is natural to ask whether one can narrow the power spectrum as well as broaden it. Any window function that causes the signal to decay more slowly than $e^{-\gamma t}$ will accomplish this. Consider the exponentially rising function given by

$$W(t) = \begin{cases} e^{-\gamma (t_o - t)} & , t < t_o \\ 1 & , t \geq t_o . \end{cases} \tag{8}$$

The signal and spectra resulting from this window function are shown in Fig. 1(c). We choose $\phi = 0$, ignore anti-resonant terms, and find that the real part of $S(\omega)$ is:

$$\mathrm{Re}\{S(\omega)\} = \frac{S_0}{\gamma} e^{-T} \left\{ \frac{\sin xT}{x} - \frac{\cos xT - x \sin xT}{1 + x^2} \right\}. \tag{9}$$

The second term in the brackets is identical to Eq. (6), and comes from $t > t_o$ where $W(t) = 1$. The center of the spectrum is near $x = 0$, and in this region the first term is dominant when $T >> 1$. Its FWHM is approximately $3.8/T$. If we ignore the second term in Eq. (9) and the similar term in the imaginary part, the power spectrum becomes:

$$P(\omega) = \frac{S_0^2}{\gamma^2} e^{-2T} \left[\frac{1 - \cos xT}{x^2} \right]. \tag{10}$$

Near the center of this spectrum its FWHM is about 5.6/T for the domain T >> 1. For any value of T, $\mathrm{Re}\{S(\omega)\}$ and $P(\omega)$ are each narrower than the corresponding raw spectrum obtained with no window function.

The narrowing of the power spectrum, $P(\omega)$ in this case, indicates that an interference effect is not responsible. In terms of the spectrometer analogy of the previous section, a variable attenuator with a transmission given by Eq. (8) between the radiating atom and the spectrometer would give a spectrum narrower than obtained without the attenuator. We might say that the variable attenuator has filtered out light in the wings of the emission spectrum and that this filtering is accomplished by modifying the decay rate of the radiation as seen by the spectrometer, *not* by choosing only the light that arrives later.

5. Comparison of the Two Types of Narrowing

It is clear that separated oscillatory fields and time resolved level crossing rely mainly on time delayed interference for the narrowing they achieve. Copley *et al.* [18] have analysed possibilities that offer an interesting hybridization of the two methods. If the time resolved level crossing signal is observed with a rectangular window equivalent to our $W(t)$ given in Eq. (4), a result essentially identical to Eq. (6) is obtained (see Eq. (2) of Ref. [18]). For some applications the extra oscillations in the wings may be undesirable and various "apodizations" have been proposed to remove them by modifying the window function. Copley *et al.* have studied exponential and Gaussian windows. Their Gaussian window has a mixed effect, resulting in both TDI and DEM narrowing. Using a Gaussian window proposed by Dodd and Series [11] which leads to no oscillations, we obtain the signal and spectra of Fig. 1(d). Since this narrowing is preserved in the power spectrum, we conclude that this type of window function leads to narrowing predominantly by DEM. Choice of a narrower Gaussian leads to oscillations in the spectrum, and the narrowing is then due more to TDI.

Some recently suggested techniques for line narrowing propose to populate an exponentially decaying state in an exponentially decaying manner, either by excitation from a decaying state [13] or by excitation with a decaying pulse of radiation [15]. It appears that these methods fall into the DEM category; the predicted line shapes are very similar to those obtained in level crossing with exponential window functions (compare Eq. (13) of Ref. [18] with Eq. (14) of Ref. [13a]).

Which of the two line narrowing mechanisms produces better results depends heavily on what the experimenter wants to learn. For example, to find the center of an isolated spectral line TDI is preferable because it suffers less signal loss for the same degree of narrowing than DEM. A rectangular window starting at 4 lifetimes gives a TDI spectrum with a central feature about 5 times narrower and 50 times smaller than the raw spectrum. To achieve the same width with an exponential window would reduce the height of the peak by another factor of 6. But to resolve blended spectral features, DEM might be better. Figure 2 shows two spectral lines of equal height and width, separated by half of the FWHM. Various types of windows with the same characteristic time (4 lifetimes) were applied in an attempt to resolve the structure. A Gaussian window chosen according to the prescription of Dodd and Series [11] does not quite resolve the structure. The rectangular window resolves the structure completely, but produces a complicated pattern that might be difficult to interpret in the absence of prior knowledge about the structure. The exponential window partially resolves the structure, with little confusion from the oscillations. The choice of the best narrowing method for a given situation will obviously depend on a number of factors, including the knowledge of the structure and the available S/N.

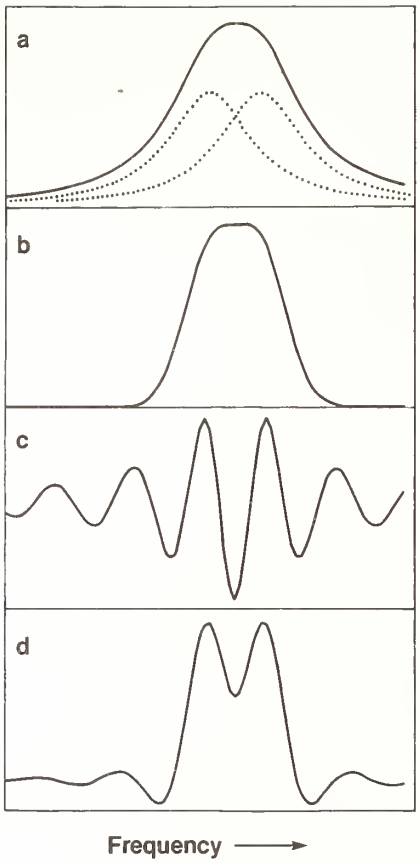


FIGURE 2. *Calculated spectra of two unresolved lines obtained with various window functions. Dotted curves in (a) are the individual lines. The solid curve is the no-window spectrum. Spectra for Gaussian window (b), rectangular window (c), and exponential window (d), are also shown.*

6. Summary

We have developed a simple model for time resolved line narrowing which is closely analogous to techniques such as time resolved level crossing spectroscopy. We have shown that within the context of this model two qualitatively different mechanisms for narrowing exist. We have compared, for some special cases, the relative utility of the two methods in applications to line center determination and resolution of complex spectra.

References

[1] N. F. Ramsey, Phys. Rev. 78, 695 (1950).
[2] V. W. Hughes, in *Quantum Electronics*, Ed. by C. H. Townes (Columbia University Press, New York, 1960), p. 502.
[3] P. Schenck, R. C. Hilborn, and H. J. Metcalf, Phys. Rev. Lett. 31, 189 (1973).
[4] J. S. Deech, P. Hannaford, and G. W. Series, J. Phys. B 7, 1131 (1974).
[5] S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. 34, 1368 (1975); *ibid.* 46, 232 (1981); and these proceedings.
[6] H. Figger and H. Walther, Z. Phys. 267, 1 (1974).
[7] J. E. Monahan and G. J. Perlow, Phys. Rev. A 20, 1499 (1979), and references therein describe the Mossbauer work.
[8] T. W. Ducas, M. G. Littman, and M. L. Zimmerman, Phys. Rev. Lett 35, 1752 (1975).
[9] P. B. Kramer, Phys. Rev. Lett. 38, 1021 (1977). Here and in Ref. [8] the phase of the transient state is preserved by a well-defined mixing with another state.
[10] T. Krist, P. Kuske, A. Gaupp, W. Wittman, and H. J. Andrä, Phys. Lett. 61A, 94 (1977).
[11] J. N. Dodd and G. W. Series, in *Progress in Atomic Spectroscopy*, Part A, Ed. by W. Hanle and H. Kleinpoppen (Plenum, New York, 1978), p. 639.
[12] D. E. Casperson *et al.*, Phys. Lett. 59B, 397 (1975).
[13] P. Meystre, M. O. Scully, and H. Walther, Opt. Commun. 33, 153 (1980); H. W. Lee, P. Meystre, and M. O. Scully, Phys. Rev. A 24, 1914 (1981).
[14] A. Denis and J. Desesquelles, Phys. Rev. A 21, 1602 (1980).
[15] P. L. Knight and P. E. Coleman, J. Phys. B 13, 4345 (1980); P. E. Coleman, D. Kagan, and P. L. Knight, Opt. Commun. 36, 127 (1981); P. M. Radmore, P. L. Knight, and D. Kagan, Phys. Lett. 82A, 288 (1981).
[16] H. J. Metcalf and W. D. Phillips, Opt. Lett. 5, 540 (1980).
[17] F. J. Lynch, R. E. Holland, and M. Hamermesh, Phys. Rev. 120, 513 (1960).
[18] G. Copley, B. P. Kibble, and G. W. Series, J. Phys. B 1, 724 (1968).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Spectroscopy of Atoms and Molecules in Gases: Corrections to the Doppler-Recoil Shift*

Mark P. Haugan

Center for Radiophysics and Space Research, Cornell University, Ithaca, NY 14853

and

Frank V. Kowalski

Department of Physics, Colorado School of Mines, Golden, CO 80401

We analyze the excitation by a monochromatic plane electromagnetic wave of a two-level atom or molecule moving through a rarefied gas. This yields a prediction for the Doppler-recoil shift which differs significantly from the familiar formula derived by studying isolated atoms or molecules in vacuum. The difference arises from an interplay between the collective and individual particle responses of a gas to an electromagnetic wave and is of fundamental importance for high-resolution laser spectroscopy. We propose an experiment using saturation spectroscopy to observe sodium molecular transitions near the atomic D lines that could easily verify our prediction for the first order Doppler shift of spectral features due to atomic or molecular absorption in rarefied gases.

Key words: Doppler and recoil effects; gases; high-resolution spectroscopy; macroscopic electrodynamics.

1. Introduction

Consider an atom or molecule with an energy difference Δe between two internal states. The natural frequency ω_0 of a transition between these states is given by $\Delta e = \hbar\omega_0$. However, because of the Doppler and recoil effects, the observed frequency of the transition depends on the velocity V of the atom or molecule relative to the laboratory and on its mass M. The difference between the observed frequency and the natural frequency of the transition is the Doppler-recoil shift.

Let ω_L represent the laboratory frequency of an electromagnetic wave that excites a transition in an isolated atom or molecule. The relativistic Doppler-recoil formula can then be written [1]

$$\omega_L = \omega_0 \left[\frac{(1 - V^2/c^2)^{1/2}}{1 - \hat{k}_L \cdot \mathbf{V}/c} \right] \left[\frac{1}{1 - \hbar\omega_0/2Mc^2} \right] , \quad (1)$$

where $\hat{k}_L$ is the wave's propagation direction and M is the mean rest mass of the atom or molecule in its initial and final states. We emphasize that this result is derived for an atom or molecule in isolation, in vacuum. The initial factor on the right-hand side of Eq. (1) is the Doppler factor, whereas the second factor represents the recoil effect. It accounts for the recoil kinetic energy, as measured in the frame in which the atom or molecule is initially at rest, that must be imparted when a photon is absorbed.

An understanding of the Doppler-recoil effect is crucial for the high-resolution spectroscopy of atoms and molecules in gases and its many applications. It is generally assumed that Eq. (1) gives the Doppler-recoil shift occurring when an atom or molecule within a rarefied gas is excited. We show that this assumption is incorrect. When an atom or molecule moves through a sample gas that is at rest in the laboratory, the Doppler-recoil shift is affected by an interesting interplay between the collective response of the gas to an incident electromagnetic wave and the wave's excitation of the atom or molecule. We find that the Doppler-recoil formula becomes

$$\omega_L = \omega_0 \left[\frac{(1 - V^2/c^2)^{1/2}}{1 - n(\omega_L)\hat{k}_L \cdot \mathbf{V}/c} \right] R , \quad (2)$$

where $n(\omega_L)$ is the refractive index of the gas at the laboratory frequency and R is a rather complicated recoil factor given in Eq. (11). This result is fully relativistic. It differs from Eq. (1) because the interaction between the excited atom or molecule and the surrounding gas has been accounted for. In section 3, we describe a proposal for an experiment on molecular transitions in sodium vapor which could easily detect the differences between the first-order Doppler predictions of Eqs. (1) and (2). The predictions can differ by several MHz. Equation (14) gives a simple approximation to Eq. (2) that is sufficiently accurate for most practical purposes.

Section 2 of this paper is devoted to an overview of the derivation of the Doppler-recoil formula, Eq. (2), and to a brief discussion of the physics that underlies it. The Doppler-recoil shift is understood in terms of the recoil kinetic energy, as measured in the laboratory frame, that is imparted to an atom or molecule which absorbs a photon. Interactions among the polarizations and polarization currents induced in the atoms and molecules of a gas by an electromagnetic wave influence the recoil momentum and, thus, the recoil kinetic energy that is imparted to an atom or molecule excited by the wave. Accounting for the effects of such interactions leads to the Doppler-recoil formula of Eq. (2). Our calculations treat the excitation of a two-level atom or molecule moving through a rarefied gaseous medium. Space limitations require that the details of these calculations and the discussion of a number of important points, including the response of a

*Supported in part by the National Science Foundation (PHY80-07957).

gaseous medium to the excitation of one of its molecules, appear elsewhere [2].

2. The Doppler-Recoil Shift for Absorption by Atoms and Molecules in Gases

2.1 Essential Ideas

A sample gas is at rest in our laboratory. We study the interaction between a molecule with velocity V and an electromagnetic wave with laboratory frequency ω_L propagating in the $\hat{k}_L$ direction. The object is to determine the frequency for which the wave resonates with a transition of the molecule having natural frequency ω_0. We make two assumptions in our derivation of this resonance condition. Fortunately, in many experimental settings they are valid assumptions.

First, the gas must act as a continuous medium. The wave we consider should satisfy the macroscopic Maxwell equations for some index of refraction $n(\omega_L)$ when propagating through the gas. Though there is absorption in the gas, the amplitude of the incident wave does not vary appreciably over the volume in which a molecule is localized. Absorption is, therefore, an inessential complication of the discussion of interaction between wave and molecule. We take the index $n(\omega_L)$ to be real for simplicity. We will suppose, to be definite, that the gas is nonmagnetic so that $n^2(\omega_L) = \epsilon(\omega_L)$, the permittivity of the gas. This is generally the case for real gases. The condition that the gas act as a continuous medium is, given the density of the gas and the dielectric properties of its molecules, a condition on the extent of the gas sample. If the index $n(\omega_L)$ characterizes propagation through an infinite gas, the same index also characterizes propagation through a finite sample so long as it is much larger than the extinction theorem [3] distance, $c[\omega_L \mid n(\omega_L) - 1 \mid]^{-1}$. For the gas we consider in section 3, the extinction theorem distance is roughly 0.05 cm at the frequencies of interest.

The second condition which must be satisfied by the gas is that it be sufficiently rarefied. Clearly, molecular collisions must be infrequent enough that collisional broadening of spectral lines does not mask the Doppler-recoil effect. Significantly, the condition guaranteeing that this does not occur leads to an essential simplification of the computation of the resonance condition. In fact, this simplification makes the problem tractable.

For collisional broadening to be negligible, the frequency of molecular collisions must be small compared to the molecular transition probability per unit time induced by the incident electromagnetic wave. When this condition is fulfilled, molecules are excited in a time interval so short that surrounding molecules are unable to approach them during the process. This fact, that for our purposes molecules have a mean separation large compared to their sizes, has two important consequences. First, in the absence of an electromagnetic wave the internal structure, e.g., energy levels, of molecules will be the same as if the molecules were isolated. Second, when an electromagnetic wave is present the molecules within the gas are influenced only by its macroscopic electric and magnetic fields. This follows because the microscopic fields within the gas deviate from the spatially-averaged macroscopic fields only in the vicinity of individual molecules. Since the molecules are well separated, the deviation produced by one molecule is unable to affect any other molecule.

We conclude that analysis of the excitation by an electromagnetic wave of internal transitions in the molecules of a collisionless gas reduces to the study of isolated molecules responding to the macroscopic fields of the wave in the surrounding medium. The macroscopic fields mediate all non-collisional effects of surrounding gas on an excited molecule as well as giving the effect of the incident wave. The effects of relatively infrequent molecular collisions can be accounted for after the analysis of the Doppler recoil effect for a collisionless gas.

2.2 The Nonrelativistic Doppler-Recoil Shift

In this section we briefly describe our semi-classical analysis of the excitation by an electromagnetic wave of a two-level molecule moving slowly through a collisionless gas. We will describe the analysis for molecules with arbitrary velocities in section 2.3. Since we are interested in the recoil of the excited molecule we include its center-of-mass degree of freedom in the calculation. This and the use of the incident wave's macroscopic fields are the simple differences between what follows and the familiar analysis of the excitation of an isolated two-level system.

The molecule we study has a pair of internal states, $\underset{\sim}{a}$ and $\underset{\sim}{b}$. Under the conditions described in the preceding section these states are indistinguishable from those of an isolated, but otherwise identical, molecule. We consider a molecule initially in a state $\underset{\sim}{A} \equiv (\underset{\sim}{a}, \mathbf{K})$, where $\mathbf{K}$ is the molecule's center-of-mass wavevector. The momentum and energy of this molecule are

$$\hbar\mathbf{K} = M\mathbf{V} \text{ and } \hbar\omega_{\underset{\sim}{A}} = e_{\underset{\sim}{a}} + \hbar^2K^2/2M \ , \tag{3}$$

where M is the molecule's mass, $\mathbf{V}$ its initial velocity, and $e_{\underset{\sim}{a}}$ its internal energy in state $\underset{\sim}{a}$.

Given a resonant incident wave, the wave-molecule interaction Hamiltonian will connect the molecular state $\underset{\sim}{A}$ to a state $\underset{\sim}{B} \equiv (\underset{\sim}{b}, \mathbf{K}')$. Typically, the dipole approximation to the interaction Hamiltonian will suffice,

$$H_I = -\mathbf{P} \cdot \mathbf{E}(\mathbf{X}, t) \ . \tag{4}$$

The symbol $\mathbf{P}$ denotes the molecule's electric dipole operator. This operator only depends on internal molecular coordinates. The field $\mathbf{E}(\mathbf{X}, t)$ is the macroscopic field of the incident classical wave. It depends on the molecule's center-of-mass coordinate, $\mathbf{X}$. For an incident monochromatic plane wave, the matrix element connecting $\underset{\sim}{A}$ and $\underset{\sim}{B}$ decomposes into a product of a momentum δ-function with the usual dipole matrix element, $\overline{P}$, that connects the internal states $\underset{\sim}{a}$ and $\underset{\sim}{b}$. The δ-function determines the recoil momentum $\hbar(\mathbf{K}' - \mathbf{K})$ that is imparted to the excited molecule. For an incident wave with macroscopic fields

$$\mathbf{E}(\mathbf{X}, t) = \hat{x}E_0 \cos(\omega_L t - \mathbf{k}_L \cdot \mathbf{X}) \ , \tag{5a}$$

$$\mathbf{B}(\mathbf{X}, t) = \hat{y}B_0 \cos(\omega_L t - \mathbf{k}_L \cdot \mathbf{X}) \ , \tag{5b}$$

where $\mathbf{k}_L = n(\omega_L)\omega_L c^{-1}\hat{z}$, $B_0 = n(\omega_L)E_0$, and the recoil momentum is $\hbar\mathbf{k}_L$. The particular choice for the polarization and phase of this wave does not affect our final conclusions.

The state of a molecule under the influence of the wave of Eq. (5) will be a superposition of states $\underset{\sim}{A}$ and $\underset{\sim}{B}$,

$$\Psi = C_{\underset{\sim}{A}}|\underset{\sim}{A}> + C_{\underset{\sim}{B}}|\underset{\sim}{B}> \ . \tag{6}$$

The system of equations

$$\dot{C}_{\underset{\sim}{A}} = (i\overline{P}E_0/2\hbar)C_{\underset{\sim}{B}}\, e^{-i(\omega_{\underset{\sim}{B}\underset{\sim}{A}} - \omega_L)t} \ , \tag{7a}$$

$$\dot{C}_{\underset{\sim}{B}} = (i\overline{P}E_0/2\hbar)C_{\underset{\sim}{A}}\, e^{+i(\omega_{\underset{\sim}{B}\underset{\sim}{A}} - \omega_L)t} \ , \tag{7b}$$

where $\omega_{\underline{BA}} \equiv \omega_{\underline{B}} - \omega_{\underline{A}}$, with initial conditions $C_{\underline{A}}(0) = 1$ and $C_{\underline{B}}(0) = 0$, determines the state of the two-level molecule we are studying. The condition for resonance is $\omega_L = \omega_{\underline{BA}}$.

The differences between Eqs. (7) and the two-state equations usually encountered are that $\omega_{\underline{BA}}$ includes the change of center-of-mass kinetic energy that accompanies the internal excitation of the molecule and that the recoil momentum imparted to the excited molecule is $\hbar \mathbf{k}_L = n(\omega_L)\omega_L c^{-1} \hat{k}_L$, rather than $\omega_L c^{-1} \hat{k}_L$. The former means that the resonance condition, $\omega_L = \omega_{\underline{BA}}$, accounts for the nonrelativistic Doppler-recoil shift. The latter explains the difference between the shifts for molecules excited within a gas and for molecules excited in isolation. Referring to Eq. (3), we write the resonance condition as

$$\hbar\omega_L = (e_{\underline{b}} - e_{\underline{a}}) + [\hbar^2(\mathbf{K} + \mathbf{k}_L)^2 - \hbar^2 K^2]/2M \ , \qquad (8)$$

with $\mathbf{K} = \hbar^{-1} M \mathbf{V}$, where $\mathbf{V}$ is the initial velocity of the excited molecule. The quantity $\hbar^{-1}(e_{\underline{b}} - e_{\underline{a}})$ is simply the natural frequency of the transition. To first order in Vc^{-1} and $\hbar\omega_0/2Mc^2$, which is as far as we can trust our nonrelativistic analysis, we find

$$\omega_L = \omega_0 (1 + n(\omega_L)\hat{k}_L \cdot \mathbf{V} c^{-1} + n^2(\omega_L)\hbar\omega_0/2Mc^2), \qquad (9)$$

in agreement with the general result of Eq. (2) to the order which we are working.

To conclude this section, it is worth remarking that the force responsible for imparting recoil momentum to an excited molecule is easy to understand. It arises from the interaction between the fields of the incident wave and the polarization and polarization current they induce in the molecule. Clearly, the semi-classical force on the molecule is

$$\mathbf{F}(t) = \langle \Psi \,|\, (\mathbf{P} \cdot \nabla)\mathbf{E}(\mathbf{X}, t) + c^{-1}\dot{\mathbf{P}} \times \mathbf{B}(\mathbf{X}, t) \,|\, \Psi \rangle, \qquad (10)$$

where Ψ is the solution, Eq. (6), to the two-state Eqs. (7). Since the fields in this expression are the macroscopic fields of the incident wave within a gaseous medium, it is not surprising that the recoil momentum imparted to the excited molecule should differ from that imparted to a molecule excited in vacuum. Using the Rabi [4] exact solution to the two-state equations for the case of a resonant incident wave, it is not difficult to verify that the impulse of the force given by Eq. (10) during the time interval in which the molecule flops from state $\underline{A}$ to $\underline{B}$, the recoil momentum imparted to the excited molecule, is $n(\omega_L)\hbar\omega_L c^{-1}\hat{k}_L$. This dependence of an excited molecule's recoil momentum on the properties of surrounding gas is the central reason for the differences between the Doppler-recoil formulae in Eqs. (1) and (2).

2.3 The Relativistic Doppler-Recoil Shift

We have argued that the analysis of molecular excitation by an electromagnetic wave propagating through a collisionless gaseous medium reduces to the analysis of an isolated molecule interacting with the macroscopic electric and magnetic fields of the wave. Since the physics of an isolated molecule interacting with electromagnetic fields is Lorentz invariant, the simplest way to analyze the excitation of a molecule that moves rapidly through the gas is to study the excitation process in the frame in which the molecule is initially at rest. In this frame, the analysis proceeds just as it would in the preceding section for a molecule initially at rest in the laboratory. The main difference is that the molecule interacts with electric and magnetic fields that are obtained by transforming the laboratory macroscopic fields of the incident wave to the molecular rest frame. The only other change is to employ relativistically correct relations between the velocity of the molecule and its momentum and kinetic energy in place of Eq. (3).

In this short paper we can only quote the results of this relativistic analysis. The Doppler-recoil formula of Eq. (2) is

$$\omega_L = \omega_0 \left[\frac{(1 - V^2/c^2)^{1/2}}{(1 - n(\omega_L)\hat{k}_L \cdot \mathbf{V}/c)} \right] \times \left[1 + \left[\frac{(1 - A)^2}{2A} \right] \frac{Mc^2}{\hbar\omega_0} (1 - \hbar\omega_0/2Mc^2) \right] , \qquad (11)$$

where

$$A \equiv \frac{(1 - \hbar\omega_0/2Mc^2)}{(1 + \hbar\omega_0/2Mc^2)} \left[1 + \frac{\hbar(\omega - ck)/Mc^2}{(1 - \hbar\omega_0/2Mc^2)} \right] , \qquad (12)$$

with

$$\omega = \gamma\omega_L\omega_L (1 - n(\omega_L)\boldsymbol{\beta} \cdot \hat{k}_L) \ , \qquad (13a)$$

$$k = n(\omega_L)\gamma\omega_L c^{-1} \left[(1 - \boldsymbol{\beta} \cdot \hat{k}_L)^2 + 2\left[(1 - n^{-1}(\omega_L)\right]\boldsymbol{\beta} \cdot \hat{k}_L - \left[1 - n^{-2}(\omega_L)\right]\beta^2 \right]^{1/2} , \qquad (13b)$$

the frequency and the magnitude of the wavevector of the incident wave as measured in the frame in which the molecule is initially at rest. Of course, $\boldsymbol{\beta} \equiv \mathbf{V}c^{-1}$ and $\gamma = (1 - \beta^2)$.

3. An Experimental Proposal to Verify the Doppler-Recoil Formula for Absorption by Atoms and Molecules in Gases

It is an interesting theoretical exercise to obtain the fully relativistic Doppler-recoil shift for molecular excitation in a gas. However, only the first few terms of the expansion of equation (11) in the Doppler parameter Vc^{-1} and the recoil parameter $\hbar\omega_0/2Mc^2$ will be of practical interest. Including only the first and second order Doppler shifts and the first order recoil effect, the Doppler-recoil formula of Eq. (11) reduces to

$$\omega_L = \omega_0 (1 + n(\omega_L)\hat{k}_L \cdot \mathbf{V}/c + n^2(\omega_L)(\hat{k}_L \cdot \mathbf{V}/c)^2 - V^2/2c^2 - n^2(\omega_L)\hbar\omega_0/2Mc^2) \ . \qquad (14)$$

We can imagine using this equation to predict the resonant laboratory frequency for a transition of given natural frequency ω_0 in a molecule of mass M moving with velocity $\mathbf{V}$, or we can imagine using it to reduce resonant frequencies observed in an experiment to a natural frequency for the transition being studied. Our point is that for either of these purposes the use of the usual Doppler-recoil relation rather than Eq. (14) when considering the excitation of molecules within a gas under reasonable experimental conditions can lead to significant error. In fact, certain experiments are particularly sensitive to differences between the Doppler-recoil formulae. For example, consider a measurement of the separation between spectral lines corresponding to two transitions with nearly equal natural frequencies ν_0 and ν_0'. We will discuss frequencies rather than angular frequencies in the remainder of this section. Observe the transitions of molecules in a velocity slice of the thermal

distribution having "velocity" $v = \hat{k}_L \cdot \mathbf{V}$. Equation (14) predicts the laboratory frequencies of the observed line centers,

$$\nu_L = \nu_0\{1 + n(\nu_L)vc^{-1}\}$$
$$\nu'_L = \nu'_0\{1 + n(\nu'_L)vc^{-1}\} . \quad (15)$$

We included only the first order Doppler effect because this is sufficient to illustrate the following general point: The observed separation between the lines, $\nu'_L - \nu_L$, depends on the difference in the refractive index between the frequencies ν_L and ν'_L as well as on the velocity v. Application of the usual Doppler-recoil formula rather than Eq. (14) in cases where dispersion makes this index difference significant can produce substantial errors. In our proposed realization of the preceding experiment, to be described below, fractional errors in the line separation of up to 10^{-6}, corresponding to an error of several MHz in frequency, can occur.

To verify at least the first-order Doppler shift prediction contained in Eq. (14), we propose an experiment on sodium molecular transitions which have a common ground state and which straddle the sodium D lines. The molecules are present in a sample of sodium vapor whose index of refraction is provided primarily by the D lines. Dispersion of the vapor medium will be important.

Using saturation spectroscopy on the molecular three level system (see reference [5]) a velocity slice of the thermal velocity distribution can be defined with the $1 \rightarrow 2$ transition. Light of frequency $\nu_L^{(12)}$ propagates in the $\hat{k}_L$ direction through the sodium vapor saturating the $1 \rightarrow 2$ molecular transition. The class velocity $v = \hat{k}_L \cdot \mathbf{V}$ of the selected molecules is

$$v = \frac{c}{n(\nu_L^{(12)})}\left[1 - \frac{\nu_L^{(12)}}{\nu_0^{(12)}}\right] , \quad (16)$$

where $\nu_0^{(12)}$ is the natural frequency of the $1 \rightarrow 2$ transition. We will consider v to be of the order of the root mean square thermal velocity of molecules in the sodium vapor when estimating experimental effects. The resonance condition for excitation by a counterpropagating beam, $\hat{k}_L$ direction, of the $1 \rightarrow 3$ transition in molecules of the chosen velocity slice is

$$\nu_L^{(13)} = \nu_0^{(13)}[1 - n(\nu_L^{(13)})vc^{-1}] . \quad (17)$$

Inserting the expression for v from Eq. (16), we can obtain the relation among observables

$$\frac{\nu_L^{(13)} - \nu_0^{(13)}}{\nu_0^{(13)}} = \frac{n(\nu_L^{(13)})}{n(\nu_L^{(12)})}\frac{\nu_L^{(12)} - \nu_0^{(12)}}{\nu_0^{(12)}} , \quad (18)$$

a relation between the Doppler shifts of the two molecular lines for the chosen velocity slice. The experiment we propose would verify that this relation holds.

For sodium vapor at roughly 700 K and one Torr monomer vapor pressure, and for sodium molecular transitions, ground $(X^1\Sigma_g^+)$ to excited $(A^1\Sigma_u^+)$ states, $(3, 13) \rightarrow (14, 26)$ and $(3, 13) \rightarrow (14, 25)$, $(v', J') \rightarrow (v'', J'')$, the deviation from equal Doppler shifts for the two lines predicted by Eq. (18) is two MHz. This can be doubled to four MHz by replacing observations of the $v = 0$ slice transition frequencies, $\nu_0^{(12)}$ and $\nu_0^{(13)}$, by observations of the transition frequencies for the velocity slice opposite the one selected above in the thermal profile.

The proposed experiment would require two dye lasers, one to saturate the $1 \rightarrow 2$ transition (the $1 \rightarrow 2$ laser); the other to probe the $1 \rightarrow 3$ transition (the $1 \rightarrow 3$ laser). A thermally stabilized cavity would be used as a frequency reference. By scanning the $1 \rightarrow 2$ laser and simultaneously recording the cavity transmission fringes one could determine where the zero velocity $1 \rightarrow 2$ resonance occurs relative to the frequency markers of the cavity. This laser can then be locked to the $1 \rightarrow 2$ transition while the $1 \rightarrow 3$ laser can be scanned to determine the zero velocity $1 \rightarrow 3$ transition frequency relative to the frequency markers. Having thus measured $\nu_0^{(12)}$ and $\nu_0^{(13)}$, the $1 \rightarrow 2$ laser would be locked to a cavity resonance about one GHz away from the zero-velocity $1 \rightarrow 2$ transition frequency. This laser would then be saturating a slice of the thermal distribution with velocity v comparable to the root mean square molecular velocity. The $1 \rightarrow 3$ laser could then be tuned to determine the $1 \rightarrow 3$ resonance frequency relative to the frequency markers. Finally, having measured $\nu_L^{(12)}$ and $\nu_L^{(13)}$, Eq. (18) could be put to the test since $n(\nu_L^{(12)})$ and $n(\nu_L^{(13)})$ can be either measured or calculated with sufficient accuracy.

4. Concluding Remarks

The Doppler-recoil shift is fundamental to our understanding of the interaction between atoms and molecules in gases and electromagnetic radiation. Our central result is a demonstration that interactions between gas particles as they respond to an incident wave can affect the Doppler-recoil shift. Equation (11), approximated in Eq. (14), accounts for such interactions in a rarefied gas. Our conclusions emerge from a detailed analysis of the excitation of a two-level molecule within a collisionless gas by a plane monochromatic wave.

To test our results, we propose an experiment on molecular transitions in sodium vapor that could detect the differences between the first-order Doppler shift predictions of Eq. (14) and the usual Doppler formula, valid for the excitation of isolated molecules. These differences can be made small by working with sufficiently rarefied gas. However, experiments are often performed under conditions where they cannot be neglected. Finally, the experiment we propose provides an example of the practical consequences of the Doppler-recoil formula of Eq. (14) in the important case where dispersion of a gas is significant.

References

[1] See for example J. L. Hall, C. J. Bordé, and K. Uehara, Phys. Rev. Lett. 37, 1339 (1976).
[2] M. P. Haugan and F. V. Kowalski, to be published.
[3] J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975) and references therein.
[4] I. I. Rabi, Phys. Rev. 51, 652 (1937).
[5] M. E. Kaminsky, R. T. Hawkins, F. V. Kowalski, and A. L. Schawlow, Phys. Rev. Lett. 36, 671 (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Implications of QED Theory for the Fundamental Constants*

G. P. Lepage and D. R. Yennie

Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, NY 14853

We review the current theoretical status of high precision measurements in quantum electrodynamics (QED). Theoretical predictions in QED fall into one of two general categories: perturbative, as for the magnetic moments of electrons and muons; and non-perturbative, as for the properties of atoms. We contrast these predictions and explore their implications for the fundamental constants of nature.

Key words: electron anomaly; hyperfine splitting; Lamb shift; muon anomaly; muonium; positronium; quantum electrodynamics.

1. Introduction

Quantum electrodynamics (QED) is the theory of the interaction of charged leptons with the quantized electromagnetic field. Given its tremendous success in predicting a wide range of experimental facts, from very refined details in atomic spectra to leptonic interactions in the multi-GeV range, it must represent in some measure the true nature of these interactions. It seems inconceivable that any theory which is conceptually less sophisticated could produce the same results. On the other hand, QED is almost certainly a part of a larger, deeper theory involving the weak, as well as the strong, interactions. Many aspects of the electroweak theory seem already well understood. It is even possible that the leptons (electron, muon, ...) are not elementary particles, but rather composites of still more fundamental entities. As yet there is no evidence for such structure, down to distances as small as 10^{-15} cm (i.e., up to energies of many GeV). In any case, QED should probably be regarded as a low energy approximation to some new fundamental theory—i.e., as a phenomenology. These ideas and other speculations are elaborated in Stanley Brodsky's contribution to this conference.

If it is "merely" a phenomenology, how can we account for QED's remarkable quantitative success as a theory of electromagnetic interactions? The answer is intimately related to one of the theory's most characteristic features, its renormalizability. Because it is renormalizable, all short distance, high energy effects in QED can be absorbed into a finite number of *measurable* masses and charges. Thus the high energy structure of the ultimate theory is irrelevant to the analysis of low energy phenomena except insofar as it determines the masses and charges of the particles involved. The renormalization program is also of tremendous practical importance, since the high energy behavior of pure QED leads to divergent integrals in perturbation theory. Whether or not this reflects the incompleteness of QED, or is rather just an artifact of perturbation theory, all divergent integrals are made convergent "simply" by expressing observable quantities in terms of physical masses and charges. The results of this intricate renormalization procedure have been outstandingly successful.

Here we will discuss the high precision, low energy features of QED—i.e., those quantities which have implications for, or require knowledge of, the fundamental constants, and particularly α, R_∞, and c. Rather than attempt to cover everything, we will concentrate on a small number of topics, in the hope of giving an accurate impression of the present status of theoretical work in this field. The most comprehensive review of QED in the past decade is the excellent paper by Lautrup *et al.* [1]. References to earlier work may be found there.

The subject naturally breaks down into quite different areas, each requiring its own special types of expertise:

(a) *Perturbative.* To compute magnetic moments or scattering amplitudes for leptons, it is necessary to evaluate a series of Feynman graphs. Each graph contributes to a definite order in α, and the final results are given as a (truncated) power series in α. A thorough mastery of the renormalization program, with its intricate subtractions of nested and overlapping divergences, is necessary.

(b) *Non-perturbative.* In bound state calculations, a definite graph does not correspond to a definite order in α. Often infinite sets of graphs are needed to evaluate a particular order. The method for organizing such a calculation is far from unique, and contributions can be readily shifted between the structure of the wave function and that of the interaction kernels. In brief, the formulation of the problem is more elusive than in perturbative analyses, while the actual computational work is in many ways substantially simpler.

These areas involve complementary aspects of QED, and both are of considerable importance in furthering our understanding of quantum field theories, renormalization theory, relativistic bound state theories, and so on.

2. Lepton Anomalies

The g-factors of the electron and muon are both very close to 2, the value predicted by the Dirac theory. To the present time, it appears that the deviations from this value (i.e., the anomalies) are explained by QED, together with very small corrections due to strong interac-

*Supported in part by the National Science Foundation.

tions. Measurements and calculations of $(g\text{-}2)/2$ have been prominent in the history of QED for almost 35 years. Progress in the two has been well matched. Advances in one have stimulated advances in the other, from Schwinger's famous calculation of the first order term, $\alpha/2\pi$, until the present stage where theory is being tested at the $(\alpha/\pi)^4$ level (i.e., four virtual photons).

These calculations provide the most stringent test of the renormalization program not only for QED, but for quantum field theories in general. The nature of this analysis is illustrated in Fig. 1. Circles are drawn around pieces of the graph which have divergent integrals. The renormalization program is a precisely defined procedure for removing these divergences. In practice, its realization is highly non-trivial, especially when numerical integration is necessary. Lengthy integrands must be subtracted in intricate ways to eliminate all combinations of nested and overlapping divergences. Another complication is that the separate graphs have spurious infrared behavior (i.e., sensitivity to very small virtual photon momenta), which cancels in the results when all diagrams are combined. The complexity of these subtractions, the dimensionality of the integrations, and the number of diagrams all increase with the order in (α/π). The $(\alpha/\pi)^2$ term (7 diagrams) has been known analytically for about 25 years. The $(\alpha/\pi)^3$ term has been reasonably well known for about 10 years, and work continues to improve its accuracy [2]. Fifty-one of the 72 diagrams have been calculated analytically; the remainder have been evaluated by a combination of analytic and numerical methods.

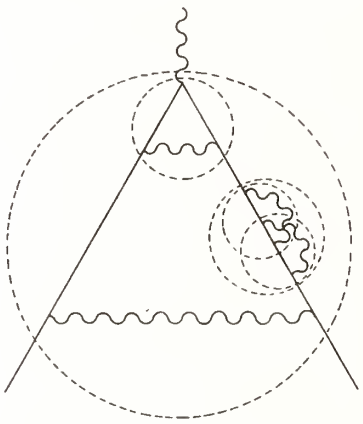


FIGURE 1. *An example of an eighth order vertex contribution. Each dashed circle encloses a divergent graph or subgraph. Note the nesting and overlapping of these divergences. If the figure is confusing, it conveys the correct impression!*

The theory of the electron's anomaly is summarized in Table 1, and compared with the current measurement of Dehmelt and his collaborators [3]. To give some indication of the recent progress in this area, we quote both the old (pre-1980) and the current estimates for the coefficients of $(\alpha/\pi)^3$ and $(\alpha/\pi)^4$. The measurements and their implications are further discussed by A. Rich in his contribution to this conference.

Four or five years ago, the order $(\alpha/\pi)^3$ calculations seemed to mark the limits to which theorists and computers could push such an analysis. However, the experimental accuracy was clearly destined to improve substantially. Anticipating these developments, Kinoshita studied the possibility of computing to order $(\alpha/\pi)^4$. He decided (he says to his regret) that it was feasible and embarked upon the project with Lindquist [4]. The scope of their calculation is enormous, even heroic. By clever techniques, they group several diagrams into single integrals, thereby reducing the 891 diagrams into "only" about 100 integrals. A typical integrand has 5000 to 15000 terms and can involve as many as 10 integration variables. Their results, presented here for the first time, must be regarded as very preliminary. More details can be found in Kinoshita's contribution to this conference.

TABLE 1. *Present status of the electron anomaly. Arrows indicate changes announced at this conference (see text).*

$$a_e^{\text{exp}} = 1\ 159\ 652\ 200(40) \times 10^{-12}$$

$$a_e^{\text{QED}} = \frac{1}{2}\frac{\alpha}{\pi} - 0.328\ 479\ 0\,(\frac{\alpha}{\pi})^2$$

$$+ (1.183\ 5(61) \rightarrow 1.176\ 5(13))\,(\frac{\alpha}{\pi})^3$$

$$+ (\ ?\ \ \ \ \ \ \rightarrow -\ 2.3\ (3.6))\,(\frac{\alpha}{\pi})^4$$

$$= 1\ 159\ 652\ 566 \times 10^{-12}$$

$$\rightarrow 1\ 159\ 652\ 411 \times 10^{-12}$$

$$\delta a_e(\alpha) = 127 \times 10^{-12}; \quad \delta a_e(\text{theory})^{a} = 76 \times 10^{-12} \rightarrow 106 \times 10^{-12}$$

$$a_e\ (\text{hadronic} + \text{muonic}) = 4 \times 10^{-12}$$

$$a_e^{\text{exp}} - a_e^{\text{th}} = (-12.6\,(5.3) \rightarrow -7.3\,(5.9))\,(\frac{\alpha}{\pi})^4$$

[a] The error in the older theory makes no allowance for the $(\alpha/\pi)^4$ term.

TABLE 2. *Present status of the muon anomaly. The experimental result is from Ref. [6]; the origin of the theoretical results may be found in Calmet et al. [5].*

$$a_\mu^{\text{exp}} = 1\,165\,923(12) \times 10^{-9}$$

$$a_\mu^{\text{QED}} = \frac{1}{2}\frac{\alpha}{\pi} + 0.765\,782\,(\frac{\alpha}{\pi})^2 + 24.45(6)\,(\frac{\alpha}{\pi})^3 + 135(64)\,(\frac{\alpha}{\pi})^4$$

$$= 1\,165\,852.2(2.0) \times 10^{-9}$$

$$a_\mu^{\text{had}} = 66.7(9.4) \times 10^{-9}$$

$$a_\mu^{\text{weak}} = 2 \times 10^{-9}$$

$$a_\mu^{\text{th}} = 1\,165\,921(10) \times 10^{-9}$$

A higher order correction is only useful once lower order coefficients have been computed analytically or at least to a high degree of accuracy. Our theoretical estimate of the $(\alpha/\pi)^3$ coefficient has been substantially improved by Levine and Roskies, using a combination of analytic and numerical techniques. See their contribution for more details.

Current results for the muon's anomaly are given in Table 2. Much of the analysis is identical for electrons and muons. However, the effects of electrons on the muon anomaly, via vacuum polarization, and vice versa are quite asymmetric. An electron, being much less massive, cannot readily create a virtual muon-antimuon pair, and hence muons have little effect on a_e. Muons, on the other hand, can easily create virtual electron-positron pairs, and indeed in second and higher order graphs containing electron loops dominate. For this same reason, strong interactions are much more important in a_μ than in a_e. Their origin is indicated schematically in Fig. 2; clearly their presence is demonstrated by the excellent agreement between theory and experiment. They are presently the most uncertain part of the theoretical analysis. A careful recent review of the theoretical contributions to a_μ is given by Calmet *et al.* [5].

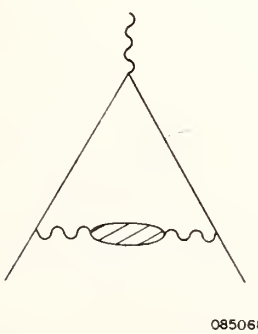


FIGURE 2. *The main hadronic contribution to the muon anomaly. The shaded blob represents the hadronic vacuum polarization; it can be determined experimentally from the process* $e^+ + e^- \to$ *hadrons.*

3. Bound States in QED

As mentioned in the Introduction, the technical problems encountered in bound state calculations are far different in character from those of the anomalies. To proceed, one must find a convenient, solvable approximation to the bound state problem, and then a systematic perturbation theory for identifying and computing corrections. The way this is to be done is certainly not unique for the relativistic two-body problem, and the optimum procedure is far from obvious. The calculations themselves are fairly laborious and full of spurious difficulties. One often finds that contributions from different perturbations cancel completely in a remarkable, but not necessarily obvious, way. An approximation made in one place to simplify the calculation of leading terms can result in a disaster at higher levels of approximation. To make the calculations relatively compact, and hence more likely to be error-free, requires considerable trial and error. Clearly, further elaboration of these methods is inappropriate here. Instead, we will try to illustrate the status of the work by examining some of the more important bound state quantities.

3.1 Hydrogen Fine Structure and the Lamb Shift

Figure 3 shows the $n = 2$ fine structure of hydrogen. All three separations have been measured to varying degrees of accuracy. The fine structure ΔE has the least theoretical uncertainty; it is given by (in frequency units)[1]

$$\Delta E = \frac{Z^4\alpha^2 R_\infty c}{16}\left\{[1 + \tfrac{5}{8}(Z\alpha)^2]\left[\frac{m_e}{m_R}\right]^{-1} - \left[\frac{m_e}{m_p}\right]^2\left[\frac{m_e}{m_R}\right]^{-3} + 2a_e\left[\frac{m_e}{m_R}\right]^{-2} - \frac{\alpha}{\pi}(Z\alpha)^2[\ln(Z\alpha)^{-2} + \delta_{FS}]\right\}$$

$$= 10969.044(3)\text{MHz}\ ,$$

[1]Although we are most interested in hydrogen ($Z = 1$), we frequently give the formulas for general Z. This distinguishes contributions related to binding effects (→ factors of $Z\alpha$) from those related to other virtual photons (→ factors of α).

where $m_e/m_R = (1 + m_e/m_p)$ and, according to an estimate by Erickson [7], $\delta_{FS} = 11/24 \pm 3/2$. The uncertainties in α and in δ_{FS} each contribute about 2 kHz to the uncertainty in ΔE. In searching for theoretical corrections previously overlooked, we found the only one of possible significance is of relative order $(Z\alpha)^2 m_e/m_p$. This should be calculated, if only to demonstrate that it is smaller than the previously mentioned uncertainties.

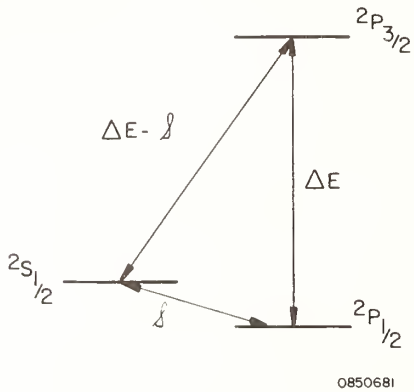


FIGURE 3. *The $n = 2$ fine structure of hydrogen (hyperfine structure is not shown).*

Figure 4 summarizes the more recent theoretical and experimental results for the Lamb shift ($\mathcal{S}$ in Fig. 3). Numerical results are given in Table 3. Two theoretical values are presented since the Erickson-Mohr discrepancy (see below) of 46 ± 10 kHz has not yet been resolved. The overall theoretical uncertainty is estimated to be roughly 20 kHz, based on order-of-magnitude estimates of the various, as yet uncalculated, terms to be discussed below.

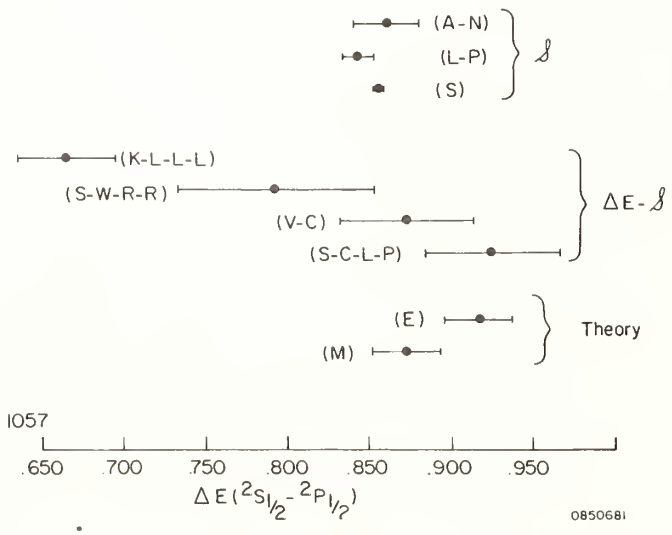


FIGURE 4. *Recent determinations of the Lamb shift. Direct experimental determinations are indicated by $\mathcal{S}$ and determinations from the large interval by $\Delta E - \mathcal{S}$. Theoretical values are also shown. References and numerical values are given in Table 3.*

TABLE 3. *Comparison of experimental and theoretical values of the Lamb shift.*

	MHz
Direct measurements [8]	
Andrews and Newton (A-N)	1057.862(20)
Lundeen and Pipkin (L-P)	1057.845(9)
Sokolov (S)	1057.8583(22)
From large interval measurements [9]	
Kaufman, Lamb, Lea, and Levinthal (K-L-L-L)	1057.664(30)
Shyn, Williams, Robiscoe, and Rebane (S-W-R-R)	1057.794(60)
Vorburger and Cosens (V-C)	1057.874(40)
Safinya, Chan, Lundeen, and Pipkin (S-C-L-P)	1057.927(41)
Theory	
Erickson (E) [10]	1057.918(20)
Mohr (M) [11]	1057.872(20)

The values of the Lamb shift determined from measurements of the large separation ($\Delta E - \mathcal{S}$) are far less accurate than the recent, more direct determinations, but it should be possible to reduce the errors to a similar level (Safinya *et al.* [9]).

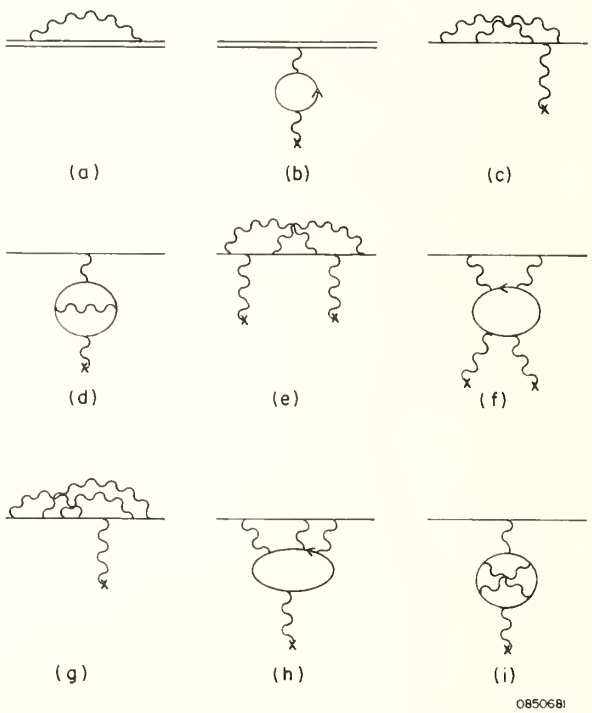


FIGURE 5. *Examples of QED contributions to the Lamb shift, as described in the text. Photon lines terminating in x represent individual Coulomb interactions. Doubled lines represent the complete Coulomb propagator or wave function.*

Various QED contributions to the Lamb shift are illustrated in Fig. 5. These are illustrative rather than inclusive. The double lines in Fig. 5(a) represent all numbers of Coulomb interactions—i.e., the figures represent an infinite number of Feynman diagrams, each of which can contribute to a given order in α. In Fig. 5(b), additional Coulomb interactions in the closed electron loop are unimportant for hydrogen [12], but the graph yields various powers of $Z\alpha$ because of the properties of the wave function. The result for diagrams (a) and (b) may be expressed as (E: Erickson [10]; M: Mohr [11]):

where $L = \alpha^3 R_\infty c/3\pi = 135.643\ 89(5)$ MHz. All terms are known exactly except $G(Z\alpha)$. In contrast to previous reviews, we advocate including the reduced mass factor in the order $\pi Z\alpha$ term. This reduces that contribution by 11 kHz, but this reduction may be offset by uncalculated recoil corrections.

The difference between the Erickson and Mohr calculations of $G(Z\alpha)$ is illustrated by Fig. 6. Erickson uses an expansion of the Lamb shift operator in which he ignores certain small terms and calculates the important integrals exactly, as functions of $Z\alpha$. His error estimate

$$\mathcal{S}^{(1)} = Z^4 L \left\{ [\ln(Z\alpha)^{-2} - 2.207909] \left(\frac{m_e}{m_R}\right)^{-3} + \frac{1}{8}\left(\frac{m_e}{m_R}\right)^{-2} \right.$$

$$+ 2.2962\ \pi Z\alpha \left(\frac{m_e}{m_R}\right)^{-3}$$

$$+ (Z\alpha)^2 \left[-\frac{3}{4}\ln^2(Z\alpha)^{-2} + 3.9184 \ln(Z\alpha)^{-2} \right.$$

$$\left. \left. + G(Z\alpha) \right] \right\}$$

MHz	
1050.560	
7.129	
	$-0.374(4)^E$
	$-0.420(9)^M$
	$1057.315(4)^E$
	$1057.269(9)^M$

is a guess of the size of the omitted terms. Mohr does a numerical integration of an exact expression derived from a partial wave expansion of the same operator. His integrals converge well for large Z but not for Z small. Consequently he must extrapolate to obtain $G(Z\alpha)$ for $Z = 1$. Mohr's extrapolation seems inconsistent with

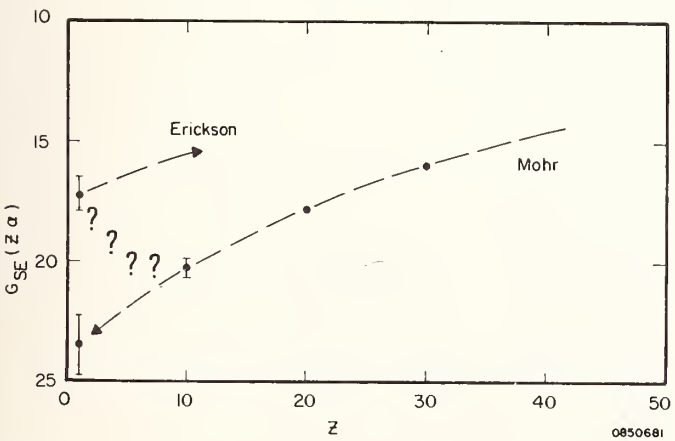


FIGURE 6. *Illustration of the Erickson-Mohr discrepancy. Values of $G_{SE}(Z\alpha)$ are plotted vs. Z; $G_{SE}(Z\alpha)$ is the part of $G(Z\alpha)$ associated with Fig. 5(a). Erickson evaluates the most important terms for $Z \approx 1$, while Mohr extrapolates from larger Z to $Z = 1$. For agreement, the extrapolation would have somewhat unexpected behavior for $Z \leq 10$ (indicated ???).*

Erickson's result for $Z = 1$, and his prediction of the small Z behavior. This conflict may ultimately be resolved in any one of several ways. Perhaps the terms neglected by Erickson add up to -6. Possibly $G(Z\alpha)$ has a more complex structure for $1 \leq Z \leq 10$. Clearly more work is needed on this important problem. [We know of such work in progress by J. Sapirstein. He expands the Lamb shift operator, dropping operators of higher order than $\alpha(Z\alpha)^6$. The remaining expressions are evaluated numerically using a closed form expression for the Coulomb propagator. His results are awaited with great interest.][2].

Figures 5(c) and (d) show sample terms of relative order α. All such terms have been calculated analytically, and they yield a Lamb shift contribution of

$$\mathcal{S}^{(2)} = 0.3215\frac{\alpha}{\pi}L = 0.101\ \text{MHz}\ .$$

Other QED corrections remain to be worked out. Coulomb corrections to diagrams (c) and (d), such as (e), and the photon scattering diagram (f) should be of order

$$\alpha(Z\alpha)L \sim 0.01\ \text{MHz}\ .$$

Finally, higher order vertices and vacuum polarization

[2]Sapirstein's calculation for the 1S state in hydrogen has been completed [Phys. Rev. Lett. 47, 1723 (1981)]; within errors, his result agrees with Mohr's.

diagrams, as in Fig. 5g-i, might contribute in order

$$\left[\frac{\alpha}{\pi}\right]^2 L \sim 0.001 \text{ MHz} .$$

Thus far we have described contributions for which the electron is assumed to move in an external Coulombic potential. Proton motion has been included only by introducing reduced mass factors $(m_e/m_R)^{-k}$, coming mainly from the radial scale of the wave function, which depends on m_R. It is important also to consider more dynamical recoil corrections. The most important of these arise when retardation in the photon propagator is accounted for. Examples of kernels contributing to recoil corrections are given in Fig. 7.

In Fig. 7(a), the virtual photon spans any number of Coulomb interactions. The physics here is related to the leading order Lamb shift in a simple way. For very soft photons ($k \lesssim Z\alpha m$), the photon interacts primarily with a particle's convection current, which is proportional to $e\boldsymbol{p}/m_e$ for an electron and $(-Ze)(-\boldsymbol{p}/m_p)$ for the nucleus. Thus for this range of photon momenta, the total current is enhanced by a factor $(1 + Zm_e/m_p)$ relative to leading order. This accounts for the relatively large coefficient in this correction, which is known analytically:

$$\mathcal{S}^{(3)} = 4.860Z\frac{m_e}{m_p}L = 0.359 \text{ MHz}$$

(the exact calculation reveals a coefficient which is smaller than suggested by the naive argument). The $^2P_{1/2}$ state also has a small shift which contributes to the Lamb shift:

$$\mathcal{S}^{(4)} = -\frac{Z^4\alpha^2 R_\infty c}{24}\left[\frac{m_e}{m_p}\right]^2 = -0.002 \text{ MHz}.$$

Some two-loop recoil corrections are suggested by Fig. 7(b). While these might possibly compensate the change in the $\pi\alpha ZL$ term discussed above, they should be computed. Their order of magnitude is

$$\pi\alpha Z^2\frac{m_e}{m_p}L \sim 0.002 \text{ MHz} .$$

Additional recoil corrections might come from diagrams like Fig. 7(c). These are probably somewhat smaller, being of order

$$\pi\left(\frac{\alpha}{\pi}\right) Z \frac{m_e}{m_p}L \sim 0.001 \text{ MHz} ,$$

but again they should be worked out.

Finally, there are corrections associated with the hadronic structure of the nucleon. The finite size correction illustrated in Fig. 7(d) is

$$\mathcal{S}^{(5)} = \frac{1}{6}Z^4\alpha^2 R_\infty c m_e^2 \langle R^2\rangle_p = 0.145(4) \text{ MHz} ,$$

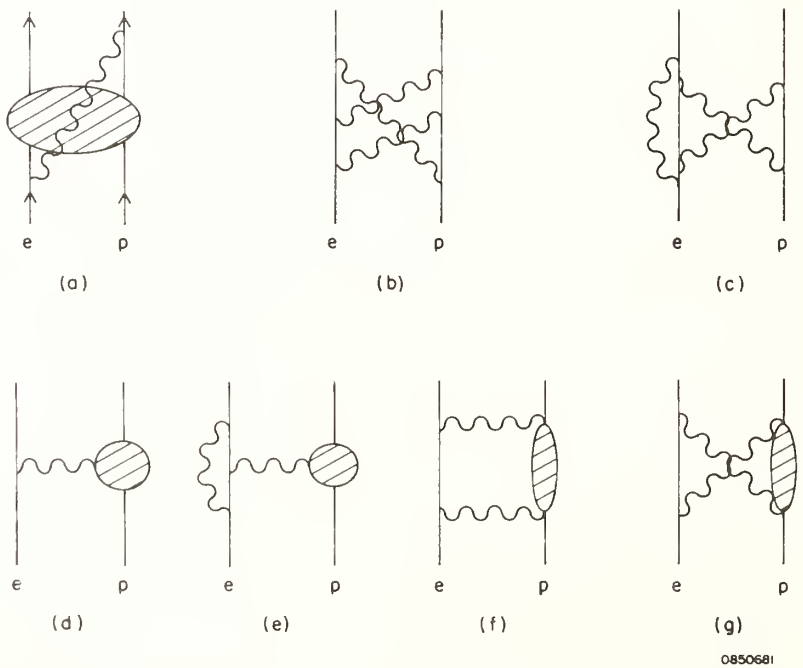


FIGURE 7. *Examples of various recoil and nucleon structure contributions to the Lamb shift, as described in the text. In (a), the blob represents any number of Coulomb interactions. In (d) and (e), the blob represents the static charge distribution of the proton, while in (f) and (g) it represents non-static features of the proton's structure.*

using the most recent value of $<R^2>_p$ as deduced from e-p scattering ($<R^2>_p^{1/2}$ = 0.862 fm). One should view the quoted uncertainty in this number with some caution, given the significant change in $<R^2>_p$ from the previously accepted value (which gave $\delta^{(5)}$ = 0.127 MHz). Radiative corrections to the finite size correction [Fig. 7(e)] are suppressed by an additional factor of order $\alpha(Z\alpha)$, making them negligible. Proton polarizability corrections, as in Fig. 7(f) and (g), should be completely unimportant for the Lamb shift because the two diagrams cancel for loop momenta much larger than m_e. As noted below, these last diagrams are important for the hydrogen hyperfine structure.

In summary then, several theoretical points must be cleared up before a meaningful comparison with experiment is possible at present levels of accuracy. Most important is to resolve the discrepancy between Erickson and Mohr. After that, several calculable QED contributions and recoil corrections should be worked out. At that point, the Lamb shift becomes a combined test of our knowledge of the proton radius and of fundamental QED.

3.2 Hyperfine Structure in Hydrogen, Muonium, and Positronium

The hyperfine separation in hydrogen is one of the most accurately determined quantities in physics [13]:

$$\Delta\nu^H_{\rm exp} = 1\ 420\ 405.751\ 766\ 7(9)\ \text{kHz}$$

Regrettably, theory is limited by nuclear structure and by radiative corrections to much less precision [14].

$$\Delta\nu^H_{\rm th} = 1\ 420\ 406(4)\ \text{kHz}$$

While uncertainties due to QED radiative corrections will be considerably reduced by current work on muonium, there is little hope that nucleon structure effects can be computed beyond 1 ppm without some radical improvement in our understanding of strong interactions.

Here we illustrate the theory more fully for muonium, a μ^+e^- bound state. Theory and experiment are summarized in Table 4. The theoretical terms known before this conference, and their contributions to $\Delta\nu$ are given by (Z = 1):

TABLE 4. *Status of muonium hfs.*

Exp[a]

$\Delta\nu$ = 4 463 302.88(16) (0.036 ppm) kHz

μ_μ/μ_p = 3.183 345 3(10) (0.31ppm)

Theory[b]

$$E_F \equiv \frac{16}{3}\alpha^2 R_\infty c \left[\frac{m_e}{m_R}\right]^{-3}\frac{\mu_\mu}{\mu_B}$$

= 4 459 034.5(1.8) (0.4 ppm)

$\Delta\nu = E_F \times$ corrections

= 4 463 304.4 (1.8, ~5) kHz

[a]Presented at this conference by V. W. Hughes.

[b]The first error in $\Delta\nu$ arises from the uncertainty in α and μ_μ; the second is a guess of the uncertainty due to missing theoretical contributions.

The first line is the splitting due to the interaction of two magnetic moments, without refinement. The error is due to uncertainty in measured values of μ_μ/μ_p, or equivalently in m_μ. The second line incorporates a relativistic correction due to Breit and the leading radiative corrections excluding those included in a_e, a_μ. The third contribution is again a radiative correction. The coefficient 18(5) was estimated by Brodsky and Erickson [15], and is analogous to $G_{SE}(Z\alpha)$ in the Lamb shift—i.e., the term disputed by Erickson and Mohr. Experience with the Lamb shift suggests that the error estimate for this coefficient is reasonable. However, it is most important that this uncertainty be substantially reduced. The fourth term is the leading recoil correction. It is illustrated by Figs. 8(a) and (b). An indication of the complications involved in such calculations is that diagram (b) also contains lower order contributions which must be removed to avoid double counting. Once this is done, the

$$\Delta\nu = E_F\Big\{1 + a_e \qquad 4\ 464\ 205.4(1.8)$$

$$+ \frac{3}{2}(Z\alpha)^2 + \alpha(Z\alpha)\left(\ln 2 - \frac{5}{2}\right) \qquad - 72.9$$

$$+ \frac{\alpha}{\pi}(Z\alpha)^2\Big[-\frac{2}{3}\ln^2(Z\alpha)^{-2} - 1.067\ln(Z\alpha)^{-2} + 18(5)\Big] \qquad - 31.5(2.8)$$

$$- 3\frac{Z\alpha}{\pi}\frac{m_R\ln(m_\mu/m_e)}{m_\mu - m_e} \qquad - 801.2$$

$$+ 2(Z\alpha)^2\frac{m_R\ln\alpha^{-1}}{m_e + m_\mu} \qquad + 11.2$$

$$\Big\{-2\frac{Z\alpha^2}{\pi^2}\frac{m_e}{m_\mu}\ln^2\frac{m_\mu}{m_e}\Big\} \qquad - 6.6$$

$$4\ 463\ 304.4(1.8,2.8)$$

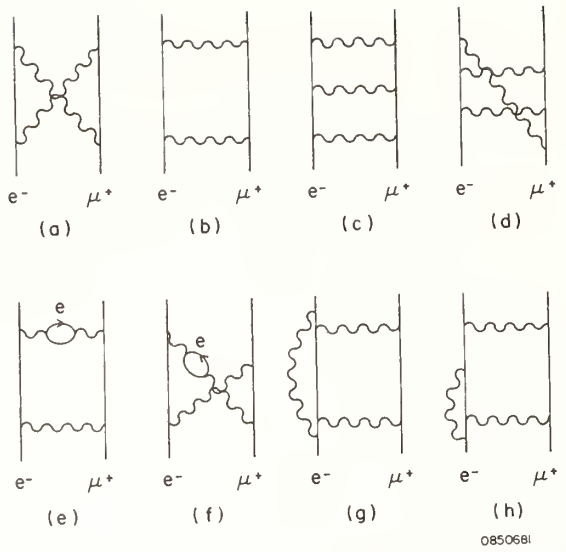


FIGURE 8. *Examples of various contributions to the muonium hyperfine structure, as described in the test. The same contributions occur in positronium, with μ^+ replaced by e^+.*

typical loop momenta range between m_e and m_μ, and consequently the "one-loop" correction is proportional to the square of the wave function at the origin. The fifth term is relatively recent [16] and comes from two-loop diagrams such as those in Figs. 8(c) and (d). Again lower order terms, including those from Figs. 8(a) and (b), must be removed, but here the resulting loop integrations range from $\sim \alpha m_R$ to $\sim m_e$. The last term is known as a "radiative-recoil" correction [17]. It results from radiative corrections to the kernels in Figs. 8(a) and (b), such as those illustrated in Figs. 8(e) to (h). In fact only the vacuum polarization diagrams [Figs. 8(e) and (f)] contribute to $\ln^2(m_\mu/m_e)$. Since the momenta in the recoil loop [Figs. 8(a) and (b)] tend to be large relative to m_e, the photon propagator containing the vacuum polarization develops the extra $\ln(m_\mu/m_e)$. The vacuum-polarization contribution has now been computed more completely with the result [18]

$$E_F(\frac{\alpha}{\pi})^2 Z \frac{m_e}{m_\mu}\left[-2\ln^2\frac{m_\mu}{m_e} - \underline{\frac{8}{3}\ln\frac{m_\mu}{m_e} - \frac{28}{9} - \frac{\pi^2}{3}}\right]$$

The new (underlined) terms yield a contribution of −2.4 kHz [2].

Further radiative-recoil corrections, coming from kernels as in Figs. 8(g) and (h), have also now been evaluated [18]. They give[3]

$$E_F(\frac{\alpha}{\pi})^2 Z\frac{m_e}{m_\mu}\left[\frac{15}{4}\ln\frac{m_\mu}{m_e} + ?\right]$$
$$= (2.3 + ? \times 0.12)\ \text{kHz}\ .$$

[3]The vacuum polarization contribution has been verified numerically by Lepage. He will also evaluate the electron radiative corrections numerically, including the non-logarithmic term.

Similar corrections, but on the muon line, have not yet been evaluated. To some extent these are incorporated in the value of μ_μ which appears as an overall factor in E_F, but there appear to be additional contributions as well.

There are terms, associated with Figs. 8(c) and (d) and the like, in which the $\ln \alpha^{-1}$ appearing above is replaced by $\ln m_\mu/m_e$ or by a constant. It has been shown that the $\ln(m_\mu/m_e)$ terms add to zero [19]. The additive constant is now being evaluated [20] and will soon be known. It could easily contribute at the level of a few kHz.

To summarize the situation for muonium: many of the remaining recoil terms will soon be computed. The uncertainty in the relative order $\alpha(Z\alpha)^2$ radiative corrections is of a different nature, but it is certainly feasible to reduce it considerably. Hopefully that challenge will be met in the not too distant future.

All terms contributing to the muonium hyperfine splitting also appear in the positronium splitting, but with m_μ replaced by m_e. However, positronium is more difficult to analyze to the same relative accuracy because recoil effects are much more important (i.e., terms with high powers of m_e/m_τ cannot be ignored), and because annihilation diagrams add many new contributions. Current theory and experiment are compared in Table 5. All

TABLE 5. *Status of positronium HFS.*

Experiment

$\Delta\nu$ = 203 387.0 (1.6) MHz [21]

203 384.9 (1.2) MHz [22]

Theory

$$\Delta\nu = \alpha^2 R_\infty c\left\{\frac{2}{3} + \frac{1}{2} - \frac{\alpha}{\pi}\left(\frac{16}{9} + \ln 2\right) + \frac{5}{12}\alpha^2\ln\alpha^{-1} + K\alpha^2 - K'\frac{\alpha^3}{\pi}(\ln\alpha^2)^2\right\}$$

$$= 203\ 399.1 + 10.8K - 2.1K'\ \text{MHz}$$

terms of relative order α and $\alpha^2\ln\alpha^{-1}$ have now been computed, and some progress has been made on the order α^2 corrections (Fig. 9). It is urgent that *all* terms of relative order α^2 be computed for a proper comparison between theory and experiment. As indicated in the table, terms of order $\frac{\alpha^3}{\pi}\ln^2\alpha^2$ are also likely and these too should be evaluated.

4. Summary

Our summary can be very brief. While it is expected that at some level—very high energies or high orders of perturbation theory—QED will be revealed to be incomplete, clear-cut evidence for this has not yet emerged. Further theoretical work is needed in most areas in order to do justice to the analysis of the fine results from many beautiful experiments. In the various cases, we have indicated which work is in progress and where further work is feasible and necessary. We hope theorists will accept the challenge to match or exceed the quality of the results from the experimentalists.

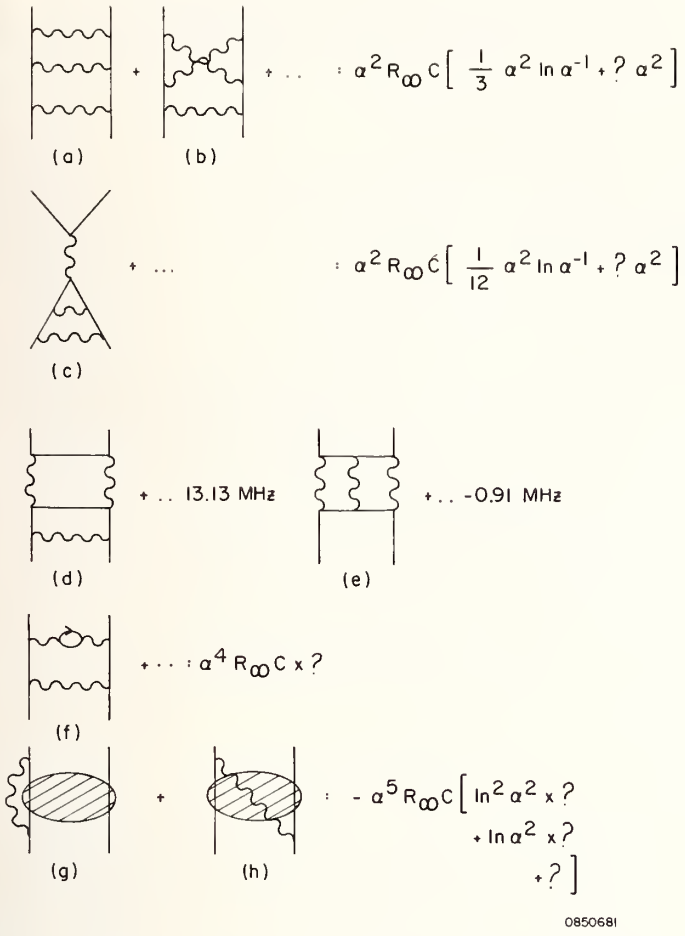


FIGURE 9. *Recent and future work on positronium hyperfine structure. (a-c): the logarithmic terms which have been obtained in the past few years [16, 23]. (d) and (e): Contributions proportional to* $\alpha^3 R_\infty c$, *worked out in Ref. [24] and Ref. [25], respectively. (f): To be evaluated in the near future [18]; also see footnote 2. (g) and (h): The blobs represent an arbitrary number of Coulomb exchanges.*

References

[1] B. E. Lautrup, A. Petermann, and E. de Rafael, Phys. Rep. **3C**, 193 (1972).

[2] M. J. Levine and R. Roskies, Phys. Rev. **D 14**, 2191 (1976), and references therein.

[3] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Phys. Rev. Lett. **38**, 310 (1977).

[4] T. Kinoshita and W. B. Lindquist, Phys. Rev. Lett. **47**, 1573 (1981); and Cornell preprints CLNS-424 and 426 (1979); and 508, 509, and 510 (1981).

[5] J. Calmet, S. Narison, M. Perrottet, and E. de Rafael, Rev. Mod. Phys. **49**, 21 (1977).

[6] J. Bailey *et al.*, Phys. Lett. **68B**, 191 (1977).

[7] G. W. Erickson, J. Phys. Chem. Ref. Data **6**, 831 (1977).

[8] D. A. Andrews and G. Newton, Phys. Rev. Lett. **7**, 1259 (1976); S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. **46**, 232 (1981); Yu. L. Sokolov, these proceedings.

[9] S. L. Kaufman, W. E. Lamb, Jr., K. R. Lea, and M. Leventhal, Phys. Rev. Lett. **22**, 507 (1969); T. W. Shyn, W. L. Williams, R. T. Robiscoe, and T. Rebane, Phys. Rev. Lett. **22**, 1273 (1969); T. V. Vorburger and B. L. Cosens, Phys. Rev. Lett. **23**, 1273 (1969); K. A. Safinya, K. K. Chan, S. R. Lundeen, and F. M. Pipkin, these proceedings.

[10] G. W. Erickson, Phys. Rev. Lett. **27**, 780 (1971).

[11] P. J. Mohr, Phys. Rev. Lett. **27**, 1050 (1975).

[12] E. H. Wichmann and N. M. Kroll, Phys. Rev. **101**, 843 (1956).

[13] H. Hellwig *et al.*, IEEE Trans. Instrum. Meas. **IM-19**, 200 (1970); L. Essen *et al.*, Nature (London) **229**, 110 (1971).

[14] S. D. Drell and J. D. Sullivan, Phys. Rev. **154**, 1477 (1967) and references therein.

[15] S. J. Brodsky and G. W. Erickson, Phys. Rev. **148**, 26 (1966).

[16] G. P. Lepage, Phys. Rev. **A 16**, 863 (1977); G. T. Bodwin and D. R. Yennie, Phys. Rep. **43**, 267 (1978).

[17] W. E. Caswell and G. P. Lepage, Phys. Rev. Lett. **41**, 1092 (1978).

[18] E. Terray and D. R. Yennie, in preparation.

[19] G. T. Bodwin, D. R. Yennie, and M. Gregorio, Phys. Rev. Lett. **41**, 1088 (1978).

[20] G. T. Bodwin, D. R. Yennie, and M. Gregorio, in preparation.

[21] A. P. Mills, Jr., and G. H. Bearman, Phys. Rev. Lett. **34**, 246 (1975).

[22] P. O. Egan, V. W. Hughes, and M. H. Yani, Phys. Rev. **A 15**, 251 (1979).

[23] W. E. Caswell and G. P. Lepage, Phys. Rev. **A 20**, 36 (1979).

[24] Vu. K. Cung, A. Devoto, T. Fulton, and W. W. Repko, Phys. Rev. **A 19**, 1886 (1979).

[25] Vu. K. Cung, A. Devoto, T. Fulton, and W. W. Repko, Phys. Lett. **68B**, 474 (1977).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Uncertainties in QED Fine Structure Calculations

Glen W. Erickson

Physics Department, University of California, Davis, CA 95616

The most general calculation of precise one-electron atomic energy levels includes uncertainties which are important in the determination of fundamental constants and in the comparison of QED theory and precision measurements. Besides the usual Lamb shift splitting S between $n\mathrm{S}_{1/2}$ and $n\mathrm{P}_{1/2}$ states, and the fine structure interval ΔE between $n\,\mathrm{P}_{3/2}$ and $n\,\mathrm{P}_{1/2}$ states, the large interval between $n\,\mathrm{P}_{3/2}$ and $n\,\mathrm{P}_{1/2}$ states has been directly measured to high precision, but the uncertainty in its calculated value is often mistakenly taken to be the combination of uncertainties in S and ΔE. This will be clarified, and the sources of these uncertainties in the low-Z calculations will be discussed. Comparisons will be made with independent calculations at higher Z and with experimental values. The overall picture is one of general agreement, but with small unresolved discrepancies. It will be seen that a tentative resolution may be obtained by discarding one of the calculations, the most recent proton and alpha particle size measurements, and 7 of the 27 most precise measurements of energy level difference.

Key words: atomic structure; electron structure; energy level shifts; fine structure; Lamb shift; proton size; quantum electrodynamics.

1. Introduction

Although the most precise calculations of quantum electrodynamical (QED) atomic energy level shifts have not changed since their results were published [1, 2], the interpretation has varied, partly due to the use of different nuclear sizes [3–6], and partly due to increasing precision of Lamb shift ($n\,\mathrm{S}_{1/2} - n\,\mathrm{P}_{1/2}$) and large interval ($n\,\mathrm{P}_{3/2} - n\,\mathrm{S}_{1/2}$) measurements [7–33]. The situation has mostly been considered as unresolved, especially since the results of the measurements have changed from agreement with one calculation to agreement with the other and to disagreement with both calculations. I would like to point out here that a tentative resolution exists, although it is not especially satisfying since it implies that there are errors, omissions and/or underestimates of uncertainties in one of the calculations (mine), in the two most recently published values of the sizes of the proton and alpha particle [4, 6], and in some of the most precise measurements of energy level differences [8, 15, 18, 19, 21, 27, 29]. The other uncertainty discussions here are not new, simply repeating selected parts of the comprehensive uncertainty discussion with my 1977 calculation of energy levels of one-electron atoms [1].

2. Interpretations of Discrepancies

The interpretations of comparisons between calculations and precision measurements of Lamb shifts and large intervals are highly dependent on the relative uncertainties involved. Because the most precise measurements in hydrogen [7–9] now have uncertainties of the same order of magnitude as those of the calculations, ±0.01 MHz = ±10 kHz, we must consider the different sources of those uncertainties before we decide if the measurements are testing QED theory, are testing the correctness of a calculational method, are testing higher order QED effects, are verifying relativistic quantum mechanics, are measuring the fine-structure constant, or are measuring nuclear size. Besides finding that all of these aspects are involved, I should point out that, after discrepancies (between different calculations, between different experiments, and between theory and experiment) are resolved, a precise agreement between theory and experiment allows useful limits to be placed on alternative theories.

Before getting into details, let me point out the difference between two apparently contradictory, but not necessarily inconsistent, interpretations of "discrepant" values: (1) they can indicate that both results support the underlying correctness of each other, and (2) they can indicate that at least one of the results is wrong. The second interpretation is the most obvious one, but depends on whether the "discrepancy" is large or small compared to the combined uncertainty. The first interpretation is valid if the "discrepancy" is small compared to the amount of "correctness" involved. Both interpretations can be valid simultaneously.

For example, my calculation [1] of a Lamb shift remainder coefficient G as -17.1 ± 0.6 and Peter Mohr's calculation [2] of the same remainder as -23.4 ± 1.2 differ by 6.3 ± 1.3. The first interpretation is that we, by using extremely different calculational methods to obtain essentially the same large negative value of a difficult remainder, have verified each others methods to an accuracy, ±6.3/2, which is an order of magnitude smaller than the remainder itself. The second interpretation notes that we differ by over four times our estimated uncertainty, so that we have either made an error or have under-estimated the uncertainty of our methods. Both interpretations are valid. The first allows our results to be used to test QED (or alternative theories) to the level of our 6.3 difference (46 kHz). The second points out the need to look for errors and/or to decide which result is more likely to be correct if it is to be used to the level of the estimated uncertainty (9 kHz).

The same set of interpretations can be made of the partially discrepant precision measurements of the $n = 2$ Lamb shift in hydrogen [7–11], which agree at the level of 20 kHz and disagree at the level of 10 kHz. The agree-

ment is sufficient to help decide which calculated value is preferred. The disagreement is sufficient to point out the possibilities of an error in a measurement [8] or of an inappropriately small uncertainty [7].

Depending on our purposes, we can treat slightly discrepant values in two distinctly different ways. The simplest is to use a weighted average to avoid missing the correct value by too much. Another way is to tentatively select one of the values as correct (and the other as incorrect) and then to test our hypothesis by using the selected value (to its claimed accuracy) and seeing if our selection will minimize discrepancies in related results. For example, instead of using a proton size obtained by adjusting scattering data normalization [4] in order to bring discrepant measurements into agreement at r_p = 0.862 ± 0.014 fm, we will consider the possibility that only the earlier measurements [3], r_p = 0.80 ± 0.02 fm or 0.805 ± 0.011 fm, are correct. This turns out to yield much better Lamb shift agreement than is possible with the larger proton size.

3. Comparison of Uncertainties

Let us now proceed to compare the sizes of the uncertainties in the calculations of the shift of an individual energy level. The largest contribution, and usually the largest uncertainty, is from the electron structure or self-energy contribution [1, 2]

$$\frac{4\alpha(Z\alpha)^4 mc^2}{3\pi n^3}\{[\ln(Z\alpha)^{-2} + \frac{11}{24}]\delta_{l0} + L_n + \frac{3}{8}\frac{C_{lj}}{2l+1} + H(Z\alpha)\} \quad (1)$$

whose higher order part has an expansion

$$H(Z\alpha) = C_5 Z\alpha + (Z\alpha)^2[C_{62}\ln^2(Z\alpha)^2 + C_{61}\ln(Z\alpha)^{-2} + G(Z\alpha)] \quad (2)$$

with leading coefficients C_5, C_{62}, and C_{61}, that have been calculated exactly. The remainder $G(Z\alpha)$ has not been calculated exactly or completely for arbitrary states, but the contributions of the dominant terms have been carefully accounted for by Erickson [1], who estimated the net size of the omitted terms for an uncertainty $\delta G(Z\alpha)$. Using a different approach, Mohr [2] has numerically evaluated the complete contribution for individual states and individual large values of Z, and has subtracted the known leading terms to obtain numerical values of $G(Z\alpha)$. Careful straightforward extrapolation to small Z yielded Lamb shifts for comparison with experiment and yielded values of $G(Z\alpha)$ for comparison with Erickson's calculations. These remainder calculations agree quite well (within small estimated uncertainties) for non-S states and, as noted before, agree fairly well (within a fraction of the remainder G) but differ by many times the estimated uncertainties for S states. For the $2S_{1/2} - 2P_{1/2}$ Lamb shift in hydrogen (S_H), the difference is 46 ± 9 kHz and scales roughly like Z^6/n^3 for other S states.

The other QED uncertainties for S_H are estimated to be ±6 kHz for uncalculated higher order parts of vacuum polarization, magnetic moment, and fourth order electron structure contributions, and ±3 kHz for uncalculated higher order parts of nuclear mass/recoil contributions. These uncertainties are smaller than the electron structure discrepancy in hydrogen and, because of their Z or m/M scaling, they are negligibly smaller than the electron structure uncertainty for other atoms.

The new aspect of the greatly improved precision of the Lamb shift measurements is that their uncertainties are now comparable to or smaller than even the ±6 kHz uncertainty due to proton size uncertainty ±0.02 fm in the nuclear size contribution,

$$\frac{<r^2>}{(\hbar/mc)^2}\frac{2(Z\alpha)^4 mc^2}{3n^3}. \quad (3)$$

This means that nuclear size can be determined by Lamb shift measurements as well as, or even better than, by electron scattering. The primary limitation is from the uncertainties in the QED calculations.

The other uncertainties (from Bethe logs, fine-structure constant, nuclear size effect in QED calculations, etc., as in Ref. [1]) are less than ±1 kHz in S_H and may be ignored for our purposes here. Even for the fine-structure interval ΔE or the large interval $\Delta E - S$, a 0.11 ppm α uncertainty causes only a 0.22 ppm uncertainty or ±2 kHz.

4. Hydrogen Comparison and a Tentative Resolution

To illustrate a number of points and to start to make sense out of the many inconsistent comparisons of theory and experiment, let us start by considering the Hydrogen Lamb shift S_H, where the measurements [7–11] lie below most calculated values. Since this favors smaller calculated S state shifts, it suggests that we consider using the smaller value of electron structure contribution [2] and the smaller proton size [3]. This turns out to give a reasonably consistent set of comparisons. The calculated S_H = 1057.867 MHz will be in acceptable agreement with the weighted average of all the measurements (1057.858 MHz) and with the new measurement with the smallest claimed uncertainty [7], 1057.8583 ± 0.0022 MHz. The discrepancies are the fewest possible: The electron structure contribution would be outside my estimated uncertainty [1], the proton size would be outside the Mainz uncertainty [4], and S_H would exceed the recently improved Harvard measurement [8], 1057.845 ± 0.009 MHz, by over twice the uncertainty of the latter.

If we consider the measurement of the large interval $2P_{3/2} - 2S_{1/2}$, there were three discordant values [13–15] and there is now a new value [12] somewhat below but in agreement with the lowest of the three. If we again use a calculated value based on the smaller proton size and the smaller upward S-state shift, the result depends slightly on the choice of fine structure constant used, but in any event lies close to the lowest of the three older values [13] and thus not far above the newest value [12], not far below the middle older value [14], and only in serious disagreement with the highest [15] of the older values. In summary, the four measurements by themselves are inconsistent and inconclusive, but our hypothesis requires no more discarding of data than does the data itself, and requires discarding of the same measurement with which all of the other three disagree. We will find similar results in the higher Z measurements.

Let me interrupt the evaluation of my hypothesis to briefly discuss the $P_{3/2} - S_{1/2} = \Delta E - S$ "large interval", which is often treated as the difference between the fine structure interval (ΔE) and the Lamb shift (S). However, the large interval calculation does not involve the $P_{1/2}$ level or its uncertainty (which is somewhat larger than for the $P_{3/2}$ level). Therefore, there are many disadvantages in trying to use a large interval measurement to obtain a value of either ΔE or S. Because of the size of

the large interval, its uncertainty due to the fine structure constant (α) uncertainty is larger than for the Lamb shift. Saying this another way, a large interval measurement could be used to determine the fine structure constant α (if the nuclear size and S state QED shift uncertainties were small). On the other hand, if the fine structure constant is accurately known, then the large interval has as much to say as does the Lamb shift S about nuclear size and S state QED shifts.

A Fine Structure interval (ΔE) measurement [16] tells us mostly about the fine structure constant α since the QED uncertainties are quite small (± 2 kHz or 0.2 ppm).

To continue the discussion of which electron structure calculation is less discrepant with precise measurements, let us consider deuterium measurements. The two precise Lamb shift measurements [17, 18] are discordant, as are the two large interval measurements [19, 20] and do not seem to prefer one calculated value more than the other. Either way, the smaller of each of the measurements [18, 19] are discordant.

These comparisons are probably most easily seen in the form given in Table 1, which considers the difference between the remainders $G(Z\alpha)$ as calculated by Mohr [2] and Erickson [1]. The differences between measurements and Erickson's calculated values are divided by the difference between Mohr's and Erickson's calculations,

$$\frac{4\alpha(Z\alpha)^6 mc^2}{3\pi n^3} \times (G_M - G_E) =$$

$$Z^6\left(\frac{2}{n}\right)^3 \Delta G \times 7.22 \text{ kHz}. \quad (4)$$

In that way, a value of 0 would indicate perfect agreement with Erickson's calculation and a value of +1 shows agreement with Mohr's smaller electron structure. Erickson's calculations use the smaller proton and alpha

TABLE 1. *Comparisons of Measurements and Calculations of Erickson [1] and Mohr [2]*

Z	$G_M - G_E$	$\frac{\text{Meas.} - \text{Calc.}(E)}{\text{Calc.}(M) - \text{Calc.}(E)}$(Meas. Unc.)	Quantity	Ref.
1	−6.3	1.16 ± 0.05	S_H	[7]
		1.45 ± 0.20	S_H	[8]
		1.08 ± 0.44	S_H	[9]
		0.24 ± 2.19	S_H	[10]
		1.15 ± 1.38	S_H	[11]
		−0.26 ± 0.90	$(\Delta E - S)_H$	[12]
		0.97 ± 0.92	$(\Delta E - S)_H$	[13]
		2.67 ± 1.38	$(\Delta E - S)_H$	[14]
		5.46 ± 0.57	$(\Delta E - S)_H$	[15]
		−0.21 ± 1.41	S_D	[17]
		6.07 ± 1.41	S_D	[18]
		−3.40 ± 1.23	$(\Delta E - S)_D$	[19]
		0.91 ± 2.07	$(\Delta E - S)_D$	[20]
2	−5.9	−0.58 ± 0.44	S_{He^+}	[21]
		1.61 ± 0.66	S_{He^+}	[22]
		1.61 ± 1.06	S_{He^+}	[23]
		1.51 ± 0.67	$n = 3S_{He^+}$	[24]
		0.79 ± 0.59	$n = 3(\Delta E - S)_{He^+}$	[25]
		1.61 ± 2.33	$n = 4(\Delta E - S)_{He^+}$	[26]
3	−5.64	−0.11 ± 0.71	$S_{Li^{++}}$	[27]
		−0.95 ± 2.36	$S_{Li^{++}}$	[28]
8	−4.43	−2.48 ± 0.89	$S_{O^{7+}}$	[29]
		0.23 ± 1.31	$S_{O^{7+}}$	[30]
9	−4.15	1.26 ± 2.20	$(\Delta E - S)_{F^{8+}}$	[31]
17	−3.20	1.28 ± 0.39	$S_{Cl^{16+}}$	[32]
18	−3.09	1.33 ± 0.78	$S_{Ar^{17+}}$	[33]

particle sizes [3, 5], so a value of +1 indicates perfect agreement with the hypothesis being tested.

The $n = 2$ $^4He^+$ Lamb shift measurements [21–23] are discrepant. My hypothesis of the small electron structure contribution [2] (and an older smaller nuclear size [5]) yields a Lamb shift, 14041.9 MHz, which indicates the discordance of only one of the three measurements [21] and of a new nuclear size calculation with a rather small uncertainty [6]. The opposite pair of hypotheses would favor the other Lamb shift measurement, but requires the discordance of both of the others.

The $n > 2$ $^4He^+$ measurements [24–26] of the Lamb shift and of the large interval definitely favor my hypothesis, although the other electron structure results only lie one or two standard deviations from the measurements.

For the $Z > 2$ measurements [27–33], some measurements prefer the larger electron structure, but there seems to be better agreement with the smaller electron structure [2]. Only two measurements [27, 29] are discordant with that hypothesis, and one of them [29] is discordant with either electron structure calculation.

In summary, only seven of the measurements are even partially discordant with the hypothesis that the smaller electron structure calculation [2] is correct, none of these measurements are corroborated by independent measurements of the same quantity, and three of them [18, 19, 29] even disagree with the other electron structure calculation [1].

5. Conclusions

If we accept the resolution given by the tested hypothesis, we can state that: QED seems to be verified to an accuracy of 13 kHz in hydrogen, but the calculation [2] has only been independently checked [1] to an accuracy of 46 kHz. It would be desirable to find errors in the discrepant calculation and measurements, but there can be serious danger of "intellectual phase-locking", such as my checking of omitted terms in my electron structure calculations [1] only until I find enough to add up to −46 kHz. I will avoid this, of course, and trust that experimenters will change their published values only if complete re-analysis demands a correction.

Even if the discrepancies were to be resolved (by verified corrections or by other independent checks to high precision) it may be seen that the next level of "ultimate" verification would be of the order of the 6 kHz due to the many other uncalculated higher order QED terms. This would prevent proton size determination better than ±0.02 fm from Lamb shift measurements, but could allow size determination for other nuclei.[1]

References[2]

[1] G. W. Erickson, J. Phys. Chem. Ref. Data 6, 831 (1977).

[2] P. J. Mohr, in *Beam-Foil Spectroscopy*, Ed. by I. A. Sellin and D. J. Pegg (Plenum Press, New York, 1976), p. 89; and Phys. Rev. Lett. 34, 1050 (1975).

[3] $r_p = 0.80 \pm 0.02$ fm: D. Frèrejacque, D. Benaksas, and D. Drickey, Phys. Rev. 141, 1308 (1966);
$r_p = 0.805 \pm 0.011$ fm: L. N. Hand, D. G. Miller, and R. Wilson, Rev. Mod., Phys. 35, 335 (1963);
$r_p = 0.81 \pm 0.04$ fm: J. J. Murphy, II, Y. M. Shin, and D. M. Skopik, Phys. Rev. **C 9**, 2125 (1974).

[4] $r_p = 0.87 \pm 0.02$ fm: F. Borkowski, G. G. Simon, W. H. Walther, and R. D. Wendling, Z. Physik **A 275**, 29 (1975);
$r_p = 0.862 \pm 0.012$ fm: G. G. Simon, Ch. Schmitt, F. Borkowski, and V. W. Walther, Nucl. Phys. **A 333**, 381 (1980).

[5] $r_\alpha = 1.63 \pm 0.04$ fm: U. Erich, H. Frank, D. Haas, and H. Prange, Z. Phys. **209**, 208 (1968).

[6] $r_\alpha = 1.674 \pm 0.012$: I Sick, J. S. McCarthy, and R. R. Whitney, Phys. Lett. **64B**, 33 (1976).

[7] $S_H = 1057.8583 \pm 0.0022$ MHz: Yu. L. Sokolov, these proceedings.

[8] $S_H = 1057.845 \pm 0.009$ MHz: S. R. Lundeen and F. M. Pipkin, these proceedings and Phys. Rev. Lett. 46, 232 (1981).

[9] $S_H = 1057.862 \pm 0.020$ MHz: D. A. Andrews and G. Newton, Phys. Rev. Lett. 37, 1254 (1976); G. Newton, D. A. Andrews, and P. J. Unsworth, Philos. Trans. R. Soc. London **290**, 373 (1979).

[10] $S_H = 1057.90 \pm 0.10$ MHz: R. T. Robiscoe and T. W. Shyn, Phys. Rev. Lett. **24**, 559 (1970).

[11] $S_H = 1057.859 \pm 0.063$ MHz: average calculated in Ref. [34] from transitions measured by S. Triebwasser, E. S. Dayhoff, and W. E. Lamb, Jr., Phys. Rev. **89**, 98 (1953).

[12] $(\Delta E - S)_H = 9911.117 \pm 0.041$ MHz: K. A. Safinya, K. K. Chan, S. R. Lundeen, and F. M. Pipkin, these proceedings and Phys. Rev. Lett. **45**, 1934 (1980).

[13] $(\Delta E - S)_H = 9911.173 \pm 0.042$ MHz: B. L. Cosens and T. V. Vorburger, Phys. Rev. **A 2**, 16 (1970).

[14] $(\Delta E - S)_H = 9911.250 \pm 0.063$ MHz: T. W. Shyn, T. Rebane, R. T. Robiscoe, and W. L. Williams, Phys. Rev. **A 3**, 116 (1971).

[15] $(\Delta E - S)_H = 9911.377 \pm 0.026$ MHz: S. L. Kaufman, W. E. Lamb, Jr., K. R. Lea, and M. Leventhal, Phys. Rev. **A 4**, 2128 (1971).

[16] $(\Delta E)_H = 10969.127 \pm 0.095$ MHz: J. C. Baird, J. Brandenberger, K-I. Gondaira, and H. Metcalf, Phys. Rev. **A 5**, 564 (1972); see also Ref. [34].

[17] $S_D = 1059.282 \pm 0.064$ MHz: B. L. Cosens, Phys. Rev. **173**, 49 (1968); see also Ref. [34].

[18] $S_D = 1058.996 \pm 0.064$ MHz: S. Triebwasser, E. S. Dayhoff, and W. E. Lamb, Jr., Phys. Rev. **89**, 98 (1953); see also Ref. [34].

[19] $(\Delta E - S)_D = 9912.607 \pm 0.056$ MHz: E. S. Dayhoff, S. Triebwasser, and W. E. Lamb, Jr., Phys. Rev. **89**, 106 (1953); see also Ref. [34].

[20] $(\Delta E - S)_D = 9912.803 \pm 0.094$ MHz: Ibid.

[21] $S_{He^+} = 14046.2 \pm 1.2$ MHz: M. A. Narasimham and R. L. Strombotne, Phys. Rev. **A 4**, 14 (1971).

[1] Sapirstein, Phys. Rev. Lett. **47**, 1723 (1981), finds a value of G for the 1S state in good agreement with that obtained by Mohr.

[2] Includes all measurements accurate to 0.1 MHz × $Z^6 (2/n)^3$.

[22] S_{He^+} = 14040.2 ± 1.8 MHz: E. Lipworth and R. Novick, Phys. Rev. **108**, 1434 (1957); see also Ref. [34].

[23] S_{He^+} = 14040.2 ± 2.9 MHz: G. W. F. Drake, S. P. Goldman, and A. Van Wijngaarden, Phys. Rev. **A 20**, 1299 (1979).

[24] $n = 3$ S_{He^+} = 4183.17 ± 0.54 MHz: D. L. Mader, M. Leventhal, and W. E. Lamb, Jr., Phys. Rev. **A 3**, 1832 (1971).

[25] $n = 3$ $(\Delta E - S)_{He^+}$ = 47844.05 ± 0.48 MHz: Ibid.

[26] $n = 4$ $(\Delta E - S)_{He^+}$ = 20180.6 ± 0.8 MHz: A. Eibofner, Z. Phys. **A 277**, 225 (1976).

[27] $S_{Li^{++}}$ = 62765 ± 21 MHz: M. Leventhal, Phys. Rev. **A 11**, 427 (1975).

[28] $S_{Li^{++}}$ = 62790 ± 70 MHz: D. Dietrich, P. Lebow, R. deZafra, and H. Metcalf, Bull. Am. Phys. Soc. **21**, 625 (1976).

[29] $S_{O^{7+}}$ = 2215.6 ± 7.5 GHz: G. P. Lawrence, C. Y. Fan, and S. Bashkin, Phys. Rev. Lett. **28**, 1612 (1972); *ibid.* **29**, 320 (1972).

[30] $S_{O^{7+}}$ = 2202.7 ± 11.0 GHz: M. Leventhal, D. E. Murnick, and H. W. Kugel, Phys. Rev. Lett. **28**, 1609 (1972).

[31] $(\Delta E - S)_{F^{8+}}$ = 68853.9 ± 35 GHz: H. W. Kugel, M. Leventhal, D. E. Murnick, C. K. N. Patel, and O. R. Wood, II, Phys. Rev. Lett. **35**, 647 (1975).

[32] $S_{Cl^{16+}}$ = 31.19 ± 0.22 THz: E. T. Nelson, O. R. Wood, II, C. K. N. Patel, M. Leventhal, D. E. Murnick, H. W. Kugel, and Y. Niv, these proceedings. See also Phys. Rev. Lett. **48**, 398 (1982).

[33] $S_{Ar^{17+}}$ = 38.0 ± 0.6 THz: H. Gould and R. Marrus, Phys. Rev. Lett. **41**, 1457 (1978).

[34] B. N. Taylor, W. H. Parker, and D. N. Langenberg, Rev. Mod. Phys. **41**, 375 (1969).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Sixth Order Contributions to *g*-2 of the Electron

M. J. Levine

Physics Department, Carnegie Mellon University, Pittsburgh, PA 15213

and

R. Z. Roskies

Physics Department, University of Pittsburgh, Pittsburgh, PA 15260

The contributions of 10 graphs to the anomalous magnetic moment of the electron in sixth order are presented to much greater accuracy than in previous evaluations. These results lead to a revised value for the entire sixth order contribution. When this is combined with the preliminary eighth order results, theory and experiment are again in reasonably good agreement.

Key words: fine-structure constant; g-factor; magnetic moment; quantum electrodynamics.

Of the 72 Feynman graphs contributing to the g factor of the electron in sixth order, 46 have been calculated analytically [1]. Five more have been reduced to one-dimensional integrals [2] and are known to eight decimal places. Another six (the light-by-light graphs) have been evaluated [3] to better than three decimal places. We report here a careful evaluation of 10 of the remaining graphs.[1]

The graphs are shown in Fig. 1 and the results in Table 1. We comment below on the significance of the fact that our value for the total is smaller than the best previous estimate [5]. Details of the calculation will be published elsewhere. It suffices here to explain that the contribution of each graph is transformed analytically to a three-dimensional integral, as in Ref. [2]. Because further analytic work seemed intractible, we resorted to numerical integration at this stage.

If we write

$$\frac{g-2}{2} = a_2\frac{\alpha}{\pi} + a_4\left[\frac{\alpha}{\pi}\right]^2 + a_6\left[\frac{\alpha}{\pi}\right]^3 + a_8\left[\frac{\alpha}{\pi}\right]^4 + \ldots$$

$$+ \text{weak interaction effects}$$

$$+ \text{strong interaction effects}$$

$$+ \text{effects due to } \mu\text{'s, } \tau\text{'s } \ldots$$

we know that

$$a_2 = 1/2$$

$$a_4 = \frac{197}{144} + \frac{\pi^2}{12} - \frac{\pi^2 \ln 2}{2} + \frac{3\zeta(3)}{4} = -0.328479\ldots.$$

[1]The astute reader will notice that Fig. 1 has only five graphs. In the counting scheme used in the text, a graph is distinguished from its mirror image.

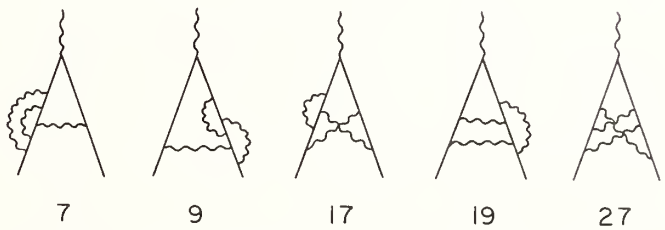


FIGURE 1. *The Feynman graphs for which new values are reported in this paper. Graph numbers follow Ref. [4].*

Combining our results with the previously quoted results, and a numerical evaluation [6] of the five remaining graphs we find

$$a_6 = 1.1765 \pm 0.0013\ ,$$

where the error is essentially all due to the five graphs of Ref. [6]. If we combine this result with Kinoshita's estimate reported elsewhere at this conference,

$$a_8 = -0.8 \pm 2.5\ ,$$

and use the value [7]

$$\alpha^{-1} = 137.035963 \pm 0.000015\ ,$$

we find

$$\left[\frac{g-2}{2}\right]_{\text{theory}} = (1\ 159\ 652\ 454 \pm 127 \pm 17 \pm 73) \times 10^{-12}\ ,$$

where the errors come from the errors in α, a_6 and a_8, respectively. (Weak and strong interaction effects, and the contribution of μ's and τ's amount to less than 5×10^{-12}.)

The best experimental value for $(g-2)/2$ is [8]

$$\left[\frac{g-2}{2}\right]_{\text{exp}} = (1\ 159\ 652\ 200 \pm 40) \times 10^{-12}\ .$$

Thus theory and experiment are in moderate agreement. Earlier discrepancies between theory and experiment have been reduced in two ways. First, our new value for the 10 graphs is smaller (by slightly more than

TABLE 1. *Contributions of the graph and its mirror image to a_6, in the Feynman gauge. Infrared divergences proportional to log λ or log^2 λ have been omitted.*

Graph No.	7	9	17	19	27	Total
Our New Values	−2.670546 ±0.000030	0.617727 ±0.000121	0.607660 ±0.000240	−0.334698 ±0.000071	1.861992 ±0.000240	0.082065 ±0.000362
			Previous Evaluations			
Levine and Wright [4]	−2.664 ±0.020	0.625 ±0.006	0.613 ±0.013	-0.330 ±0.013	1.854 ±0.013	0.0970 ±0.0230
Cvitanovic and Kinoshita [5]	-2.6707 ±0.0019	0.6189 ±0.0064	0.6097 ±0.0034	-0.3182 ±0.0072	1.8572 ±0.0086	0.0893 ±0.0060

one standard deviation) than the most accurate previous estimate. Secondly, Kinoshita's estimate for a_8 is negative. Both these effects operate in the same direction and reduce the disagreement between theory and experiment.

Another way of reporting the results is to assume the validity of QED, and to estimate α from the experimental value of $g-2$ and the theoretical calculations. The result is

$$\alpha^{-1} = 137.035993 \pm 0.000005 \pm 0.000002 \pm 0.000009 ,$$

where the errors come from experiment, a_6, and a_8, respectively. This estimate for α^{-1} agrees reasonably with that used above.

It is clear that the most pressing theoretical need is a better evaluation of a_8. If its error can be reduced significantly, and if better estimates for α can be obtained, as has been suggested in many ways at this conference, then we will be able to test QED to much greater precision.

Conversely, if the experimental precision in the determination of g_e increases as has also been suggested at this conference, a better evaluation of a_8 can provide us with an order of magnitude improvement in the present determination of α, assuming the validity of QED.

References

[1] M. J. Levine and R. Roskies, Phys. Rev. Lett. **30**, 772 (1973); Phys. Rev. D 9, 421 (1974); *ibid.* **14**, 2191 (1976); M. J. Levine, R. Roskies, and R. C. Perisho, *ibid.* **13**, 997 (1976); R. Barbieri, M. Caffo, and E. Remiddi, Phys. Lett. **57B**, 460 (1975); J. A. Mignaco and E. Remiddi, Nuovo Cimento A **60**, 519 (1969); R. Barbieri and E. Remiddi, Phys. Lett. **49B**, 468 (1971).

[2] M. J. Levine, E. Remiddi, and R. Roskies, Phys. Rev. D **20**, 2068 (1979).

[3] T. Engelman and M. J. Levine, unpublished.

[4] M. J. Levine and J. Wright, Phys. Rev. D 8, 3171 (1973).

[5] P. Cvitanovic and T. Kinoshita, Phys. Rev. D **10**, 40007 (1974).

[6] T. Kinoshita and W. B. Lindquist, Cornell preprint CLNS-374 (1977), unpublished.

[7] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. **42**, 1575 (1979).

[8] R. S. Van Dyck, P. B. Schwinberg, and H. G. Dehmelt, Bull. Am. Phys. Soc. 24, 758 (1979).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Calculation of the Eighth Order Anomalous Magnetic Moment of the Electron*

T. Kinoshita

CERN, 1211 Geneva 23, Switzerland**

and

Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, NY 14853

and

W. B. Lindquist

Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, NY 14853

We present a very preliminary result of our calculation of the eighth order QED contribution to the anomalous magnetic moment of the electron. Altogether 891 Feynman diagrams contribute to this term. By a method developed earlier, we have compressed them into about 100 integrals, which are evaluated using adaptive Monte Carlo integration routines. Our result is $-0.8\ (2.5)(\alpha/\pi)^4$. Combining this with the results of lower orders and using $\alpha^{-1} = 137.035\ 963(15)$ we find

$$a_e^{th} = 1\ 159\ 652\ 460\ (148) \times 10^{-12}.$$

The difference between experiment and theory is now $-251(154) \times 10^{-12}$.

Key words: anomalous magnetic moment; electron; fine-structure constant; quantum electrodynamics.

1. Introduction

The electron belongs to a special class of elementary particles whose interaction is predominantly electromagnetic. Since electromagnetic processes are the easiest to measure and can be done so with the greatest of accuracy, it has provided the main testing ground of quantum electrodynamics (QED). In particular, the magnetic moment anomaly of the electron has been studied both experimentally and theoretically over 30 years with ever-increasing precision. A very substantial progress has been made in the last few years by Dehmelt and his coworkers [1] whose published 1977 value for the magnetic moment anomaly $a_e = (g_e - 2)/2$,

$$a_e^{exp} = 1\ 159\ 652\ 410\ (200) \times 10^{-12}, \tag{1}$$

represents a factor of improvement of 20 over the best previous measurement. By 1979 [2] the error went down by a further factor of 5:

$$a_e^{exp} = 1\ 159\ 652\ 200\ (40) \times 10^{-12}. \tag{2}$$

In perturbation theory of QED, higher order radiative corrections to a_e can be written as a power series in α/π:

$$a_e = C_1(\alpha/\pi) + C_2(\alpha/\pi)^2 + C_3(\alpha/\pi)^3 + C_4(\alpha/\pi)^4 + \ldots \tag{3}$$

Thus far the first three coefficients have been calculated [3]:

$$\begin{aligned} C_1 &= 0.5, \\ C_2 &= -0.328\ 478\ 966\ \ldots, \\ C_3 &= 1.183\ 5\ (61). \end{aligned} \tag{4}$$

If one uses the 1979 value of the fine structure constant α [4]

$$\alpha^{-1} = 137.035\ 963\ (15), \tag{5}$$

the QED prediction, Eq. (4), gives the value

$$1\ 159\ 652\ 566 \times 10^{-12}. \tag{6}$$

To this one must add contributions from other sources. They include the contributions of the muon loop, the τ meson loop, and the hadronic effect, which are all small:

$$\begin{aligned} a_e\ (\text{muon}) &= 2.8 \times 10^{-12}, \\ a_e\ (\tau\ \text{meson}) &= 0.01 \times 10^{-12}, \\ a_e\ (\text{hadron}) &= 1.6\ (2) \times 10^{-12}. \end{aligned} \tag{7}$$

The effect of the weak interaction (according to the standard Weinberg-Salam model) is also very small [5]:

$$a_e\ (\text{weak}) \simeq 0.05 \times 10^{-12}. \tag{8}$$

Collecting Eqs. (6), (7), and (8) we obtain the best theoretical prediction available thus far:

$$a_e^{th} = 1\ 159\ 652\ 570 \times 10^{-12}. \tag{9}$$

Comparing this with Eq. (2) we see that theory and experiment differ by

$$\Delta a_e = a_e^{exp} - a_e^{th} = -370 \times 10^{-12} \tag{10}$$

*Supported in part by the U.S. National Science Foundation.

**1 April–31 July, 1981.

which is nearly 10 times as large as the experimental error quoted in Eq. (2).

In order to decide whether this discrepancy is significant or not, one must of course examine the errors in the theoretical value, Eq. (9). The uncertainty in the measurement of α quoted in Eq. (5) contributes an error of 127×10^{-12} to Eq. (9). The error due to that of C_3 in Eq. (4) amounts to 77×10^{-12} which results almost totally from the 15 integrals (out of 72) that have been evaluated only numerically. An additional six integrals (from diagrams containing photon-photon scattering subdiagrams) have recently had their errors reduced by several orders of magnitude by a combined analytical-numerical integration technique [6]. This method, currently being applied to the 15 integrals mentioned above, is expected to reduce the error of C_3 to an insignificant level [7]. The errors in Eqs. (7) and (8) are completely negligible at this stage. Actually, in view of the fact that

$$(\alpha/\pi)^4 \simeq 29 \times 10^{-12} \tag{11}$$

is almost as large as the experimental error in Eq. (2), the most serious theoretical uncertainty arises from the absence of knowledge about C_4 which makes comparison of theory and experiment rather tentative.

On the other hand, if we assume that the difference Δa_e in Eq. (10) is genuine and due entirely to the C_4 term, we obtain the "prediction" that

$$C_4 \simeq -13(6), \tag{12}$$

a surprisingly large value. Of course it is entirely possible that the discrepancy, Eq. (10), is mainly of experimental origin and fades away as measurements improve. Whether this turns out to be the case or not, however, there is no question that we must know the magnitude and sign of C_4 for a meaningful comparison of the theory with the present and forthcoming experiments.

2. Calculation of C_4

It is for this reason that we decided to calculate the quantity C_4. In the absence of any clever method which enables us to estimate it quickly and reliably, we had no choice but to calculate by brute force the values of all 891 Feynman diagrams that contribute to C_4. However, a substantial simplification has been achieved by a method developed earlier [8] which enables us to combine several integrals into one. In this manner we were able to reduce the number of integrals to be evaluated to slightly over 100.

The 891 Feynman diagrams fall naturally into the following five groups, each of which consists of one or more gauge invariant sets:

Group I. Second-order vertex diagrams containing vacuum polarization loops of second, fourth and sixth orders. This group consists of 25 diagrams. They are represented by 10 integrals.

Group II. Fourth-order vertex diagrams containing vacuum polarization loops of second and fourth orders. This group contains 54 diagrams and is represented by 8 integrals.

Group III. Sixth-order vertex diagrams containing a vacuum polarization loop of second order. There are 150 diagrams in this group. The number of independent integrals is 8.

Group IV. Vertex diagrams containing a photon-photon scattering subdiagram with further radiative corrections of various kinds. This group consists of 144 diagrams. The number of integrals is 13.

Group V. Vertex diagrams containing no vacuum polarization loop. This group is comprised of 518 diagrams and is represented by 47 integrals.

All integrands have been generated by an algebraic program SCHOONSCHIP. A typical integrand is a rational function consisting of up to 15,000 terms, each term being a product of up to 8 or 9 factors. The integration, over a hypercube of up to 10 dimensions, has been carried out by adaptive Monte Carlo subroutines RIWIAD and VEGAS.

The evaluation of the integrals of the first three groups was completed more than 2 years ago. The results are [9, 10]

$$\begin{aligned} C_4^I &= 0.076\ 6\ (6), \\ C_4^{II} &= -0.523\ 8\ (10), \\ C_4^{III} &= 1.419\ (16). \end{aligned} \tag{13}$$

The remaining groups are substantially more difficult to evaluate and the results have only recently become available. In view of difficulties encountered in carrying out some of the numerical integrations, the error (90 percent confidence limit estimated by the integration routines) is still very substantial and should not be fully trusted. Anyway, our very, very tentative results are [11]:

$$\begin{aligned} C_4^{IV} &= -0.78\ (48), \\ C_4^{V} &= -1.0\ (2.4). \end{aligned} \tag{14}$$

Combining Eqs. (13) and (14) we obtain

$$C_4 = -0.8\ (2.5). \tag{15}$$

The central value may still fluctuate considerably. The main significance of this result is that we now have finite bounds for C_4, although they may be rather soft. Anyway, it appears that the "prediction," Eq. (12), is not borne out by our calculations.

From Eqs. (4), (5), (7), (8), (15) and Ref. [7] we obtain

$$a_e^{th} = 1\ 159\ 652\ 460\ (148) \times 10^{-12} \tag{16}$$

This is consistent with the measurement, Eq. (2), as well as the new measurement of a_e for the positron [12]. Possible causes of the remaining discrepancy are underestimate of errors in the measurement of a_e and/or the measurement of α in Eq. (5) as well as theoretical errors. The latter includes not only purely computational errors in C_3 and C_4 but also physical assumptions we have implicitly made on the nature of the weak interaction and the (lack of) internal structure of the electron.

An important by-product of our calculation is that it enables us to determine α in a way least dependent on theoretical ambiguities. We give below the value of α determined from the measurement, Eq. (2), and one quoted in Ref. [12] and the new calculations of Ref. [7] and this report. For comparison we also list α determined from the muonium hfs [13] and the ac Josephson effect [4]:

$$\alpha^{-1}(a_e) = 137.035\ 993\ (10),$$
$$\alpha^{-1}(\text{muonium hfs}) = 137.035\ 989\ (47),$$
$$\alpha^{-1}(\text{ac Josephson}) = 137.035\ 963\ (15). \quad (17)$$

The error in α^{-1} (muonium hfs) is mostly theoretical and requires a considerable improvement of theory to take advantage of the very accurate measurements. Such work is in progress at present [14]. Since it is a pure QED process, comparison of $\alpha^{-1}(a_e)$ and α^{-1} (muonium hfs) is useful for checking the internal consistency of QED. On the other hand, α^{-1} (ac Josephson) is based on a purely solid-state physics phenomenon and is presumably not susceptible to higher order QED corrections [15].

3. Conclusion

As is seen from Eq. (17) the agreement of α determined by various means is satisfactory at present. But, irrespective of whether further work establishes the disagreement of different α's or not, it is about time to re-examine the basis of some of the underlying theories very closely. For instance, is the ac Josephson effect really capable of determining $2e/h$ to an accuracy better than 10^{-8}? What is the upper limit of the theoretical error and how does one estimate it?

Although the new and exciting way of determining α discovered by von Klitzing *et al.* [16] is not yet comparable in accuracy to that of the ac Josephson effect, a similar question about its theoretical foundation must be asked in order to make it a viable method for the high precision determination of α.

Finally, I call attention to the very far-fetched, but not completely crazy, possibility that the discrepancy of $\alpha(a_e)$ and other α's arises from the electron not being elementary but having in fact some internal structure which, for instance, may be described as a bound state of subquarks. The present agreement (or disagreement) of theory and experiment of a_e suggests that the mass of such subquarks must be at least of the order of 10^6 proton masses [17]. Should this turn out to be the case (which I must say is very unlikely), low energy particle physics and even solid-state physics have roles to play in determining the laws of Nature in the domain of ultra-high energy physics.

We should like to thank Brookhaven National Laboratory, the National Laboratory for High Energy Physics, Japan, and CERN for their generous support of our work.

References

[1] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Phys. Rev. Lett. 38, 310 (1977).
[2] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Bull. Am. Phys. Soc. 24, 758 (1979).
[3] C_3 is calculated from the results given by M. J. Levine and R. Roskies, Phys. Rev. D 14, 2191 (1976); T. Kinoshita and W. B. Lindquist, Cornell Preprint CLNS-374 (1977), unpublished, and papers quoted in Refs. [6] and [8].
[4] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. 42, 1575 (1979).
[5] Note that the Weinberg-Salam model has not yet been tested beyond the tree graph level. For discussions of other possibilities, see T. Kinoshita, in *New Frontiers in High Energy Physics—Studies in the Natural Sciences* (Orbis Scientiae), A. Perlmutter and L. F. Scott, Eds. (Plenum Press, New York, 1978), p. 127.
[6] T. Engelmann and M. J. Levine, private communication.
[7] M. J. Levine and R. Roskies, these proceedings. They have succeeded in reducing the error of 10 of 15 integrals by several orders of magnitude. They report the new value of C_3 to be 1.1765 (13), where the error is predominantly due to the 5 remaining integrals.
[8] P. Cvitanovic and T. Kinoshita, Phys. Rev. D 10, 4007 (1974).
[9] T. Kinoshita and W. B. Lindquist, Cornell Preprints CLNS-424 and 426 (1979), unpublished.
[10] T. Kinoshita and W. B. Lindquist, Cornell Preprint CLNS-508 (1981), unpublished.
[11] T. Kinoshita and W. B. Lindquist, Cornell Preprints CLNS-509 and 510 (1981), unpublished.
[12] P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt, these proceedings have succeeded in measuring α_e for a positron. Their result is $\alpha_e(e^+) = 1.159\ 652\ 222\ (50) \times 10^{-12}$.
[13] D. E. Casperson *et al.*, Phys. Rev. Lett. 38, 956, and 1504 (1977);
P. O. Egan *et al.*, 8th ICOHEPANS, Vancouver, Canada, 1979, abstract 1B10;
V. W. Hughes, these proceedings;
W. E. Caswell and G. P. Lepage, Phys. Rev. A 18, 810 (1977);
W. E. Caswell and G. P. Lepage, Phys. Rev. Lett. 41, 1092 (1978);
G. T. Bodwin and D. R. Yennie, Phys. Rep. 43, 267 (1978);
G. T. Bodwin, D. R. Yennie, and M. A. Gregorio, Phys. Rev. Lett. 41, 1088 (1978).
[14] D. R. Yennie and G. P. Lepage, these proceedings.
[15] D. N. Langenberg and J. R. Schrieffer, Phys. Rev. B 3, 1776 (1971); J. B. Hartle, D. J. Scalapino, and R. L. Sugar, Phys. Rev. B 3, 1778 (1971).
[16] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).
[17] S. Brodsky and S. Drell, Phys. Rev. D 22, 2236 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Experimental Determinations of the Anomalous Magnetic Moments of the Free Leptons*

Ralph Conti, David Newman, Arthur Rich, and Eric Sweetman

Physics Department, The University of Michigan, Ann Arbor, MI 48109

The ten year period since the First International Conference on Precision Measurements and Fundamental Constants has seen an improvement of approximately one hundred fold in the experimental determination of the (e^-, e^+) and (μ^-, μ^+) anomalous magnetic moments! Similar progress has been made in theoretical QED g-2 calculations as well as in the experimental determination of α.

The value of the electron g-factor anomaly ($a(e^-)$) is now known to 34 parts per billion (ppb) from measurements on electrons of typical energy 10^{-3} eV and may be compared with theory to 146 ppb. Theory and experiment are in agreement. The value of $a(e^+)$ has been compared to $a(e^-)$ to 10^4 ppb in an experiment done at 0.6 GeV and recent unpublished work on millielectron volt positrons has allowed a direct determination of $a(e^+)$ to better than 100 ppb. All results show that $a(e^+) = a(e^-)$ thus confirming CPT invariance. The values of $a(\mu^-)$ and $a(\mu^+)$ are each known separately to 10^4 ppb ($a(\mu^-) = a(\mu^+)$ to this accuracy) and may be compared with the theoretical value (QED plus hadronic contributions) which is known to about 8×10^3 ppb. Again theory and experiment are in agreement. In addition to providing the most direct and precise tests of QED, the g-2 experiments can also be used to test special relativity, set limits on possible fermion substructure and on electric dipole moments, and test other aspects of current theory.

The substance of this article will be a review of the experiments mentioned above, a discussion of new work now in progress, and an attempt to predict future possibilities in measuring lepton anomalous moments. The non-QED tests mentioned will also be discussed.

Key words: anomalous magnetic moment of leptons; lepton g-factors; quantum electrodynamics.

1. Introduction

Measurements of the anomalous magnetic moments of leptons have advanced remarkably in the ten years since PMFC-I. The uncertainty in the electron anomaly has been reduced by a factor of 180, that of the muon by a factor of 40, and that of the positron by a factor of 10^4. These measurements test quantum electrodynamics (QED), are sensitive to the effect of hadronic couplings, and allow limits to be placed on speculative interactions involving axion couplings or possible lepton substructure. Measurements of the lepton anomalies at differing velocities allow precise limits to be placed on the validity of special relativity and in addition, the muon experiments provide tests of special relativity involving the twin paradox and a direct test of time dilation and set new limits on the electric dipole moment (EDM) of the muon. Many of these developments (through 1978) have been reviewed in the comprehensive article by Field, Picasso and Combley [1], while a detailed treatment of earlier developments (through 1972) may be found in Ref. [2]. The vitality of research on the lepton magnetic moments is demonstrated by the fact that much of the work to be discussed in this article is sufficiently new so that it would not have been included in the 1978 Field *et al.* article. Finally we note that space limitations have restricted us to a review which is more a listing and referencing of new developments and their potentialities than an in-depth discussion of the experiments.

The g-factor of a particle is defined through the relation of its magnetic moment $\boldsymbol{\mu}$ to its spin S, i.e.,

$$\boldsymbol{\mu} = \frac{g}{2}\frac{e}{m_0 c}\mathbf{S} . \tag{1}$$

Deviations from the Dirac prediction of $g = 2$, due to radiative corrections and other effects, are called the "anomaly":

$$a \equiv (g - 2)/2 . \tag{2}$$

Since measurements of a are always made on leptons moving nearly perpendicular to a magnetic field, we now briefly discuss the well-known spin and orbital motion for a particle moving perpendicular to a uniform magnetic field. In the simple case, the relativistic energy levels for a single lepton are given by:

$$W(n, m) = m_0 c^2[(1 + (1 + 2n + 2m)\chi)^{1/2} + ma\chi] \tag{3}$$

where n is the orbital quantum number (0, 1, 2, . . .), m is the spin quantum number ($m = \pm 1/2$), $\nu_0 = eB/2\pi m_0 c$ and $\chi = h\nu_0/m_0 c^2 \ll 1$. The level diagram is shown in Fig. 1.

The cyclotron and spin precession frequencies ν_C and ν_S are determined by the energy difference between successive n and m quantum levels:

$$h\nu_C = W(n, m) - W(n - 1, m) \tag{4}$$

$$h\nu_S = W(n, 1/2) = W(n, -1/2) . \tag{5}$$

For $\chi \ll 1$ Eqs. (4) and (5) yield:

$$\nu_C = \nu_0/\gamma \tag{6}$$

$$\nu_S = \frac{g}{2}\nu_0 + (1 - \gamma)\nu_C = \nu_0\left[\frac{1}{\gamma} + a\right] . \tag{7}$$

Here γ is the usual relativistic factor $\gamma = (1 - \beta^2)^{-1/2}$, which for $\chi \ll 1$ may be written as $[1 + (2n + 1)\chi]^{1/2}$. Thus the quantum mechanical values for ν_C and ν_S are the

*The g-2 research at Michigan is supported by a grant from the Atomic, Molecular and Plasma Physics Program of the NSF.

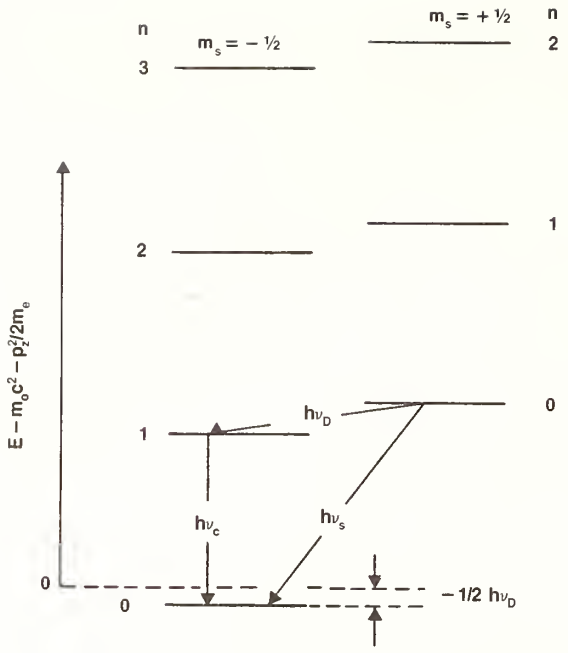


FIGURE 1. *Rabi-Landau levels of an electron in a uniform magnetic field.*

same as the classical values (independent of n). The term $(1 - \gamma)\nu_C$ is the Thomas precession. Both ν_C and ν_S are dependent on the velocity of the particle. However, the difference frequency $\nu_D \equiv \nu_S - \nu_C$, is independent of γ:

$$\nu_D = a\,\nu_0 \,. \tag{8}$$

This feature allows a precise measurement of the anomaly in experiments where uncertainty in the particle energy would otherwise limit the accuracy. Classically, the difference frequency is the frequency at which the spin of the lepton precesses relative to its velocity in a magnetic field. It corresponds quantum mechanically to the energy difference between quantum levels with a change of both m and n, or spin-cyclotron transitions:

$$h\,\nu_D = W(n,\ -1/2) - W(n\ -\ 1,\ +1/2) \,. \tag{9}$$

All of the precision measurements of lepton magnetic moments consist of direct measurements of ν_D and indirect evaluations of ν_0. The value of ν_0 is obtained either from Eq. (6) and a measurement of ν_C, or from a proton nuclear magnetic resonance measurement as described in sections 2.2 and 3.1.

2. The Electron and Positron *g*-2 Experiments

2.1 The University of Washington Resonance Experiment

The most accurate measurement of the electron anomaly a(e) to date is that done by the group at the University of Washington. Using $\sim 10^{-3}$ eV electrons, they obtain ν_D by driving spin-cyclotron transitions using externally applied rf electric fields. The cyclotron frequency is then determined by measuring resonant absorption for $\Delta m = 0$, $\Delta n = +1$ transitions at small values of n. Historically, a resonance technique employed at Washington was the first method (1958) to directly yield a result for the anomaly of a free particle.

The model of a uniform magnetic field with no electric field is not practical for a precision measurement since the electron must be confined to a small region for an extended period to obtain narrow linewidths. Confinement is achieved in a Penning trap (Fig. 2) by an electric quadrupole potential $V = (V_0/b^2)(r^2 - 2z^2)$ superposed on a uniform magnetic field $B_0\hat{z}$, where, in one version of the experiment whose parameters we refer to in this section, $V_0 = 9.1$ volts, $b = 0.47$ cm, and $B_0 = 18.3$ kG. In addition to its cyclotron motion, the electron now executes axial oscillations along the direction of **B** at frequency

$$\nu_z = (1/\pi)(\sqrt{V_0\,e/m_0\,b^2}) \simeq 59\ \text{MHz} \tag{10}$$

and exhibits magnetron motion (a drift of the cyclotron orbit center about the center of cylindrical symmetry) at frequency

$$\nu_m = \frac{\nu_0}{2}\left[1 - \left[1 - \frac{2\nu_z^2}{\nu_0^2}\right]^{1/2}\right] \simeq 34\ \text{kHz} \,. \tag{11}$$

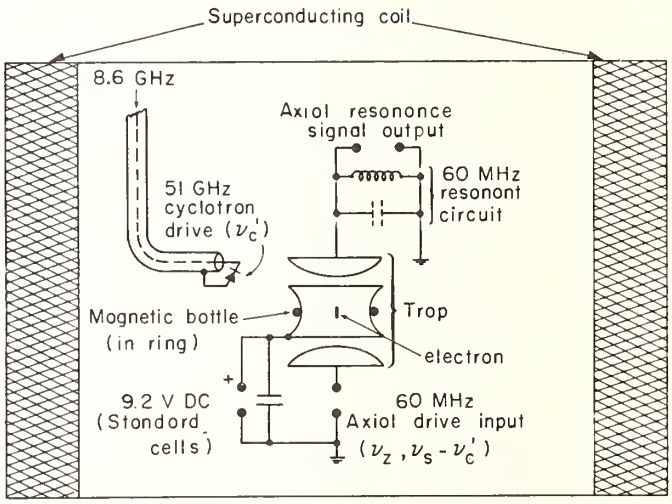


FIGURE 2. *Schematic of the University of Washington apparatus used to measure the electron anomaly. This apparatus allows the measurement of the cyclotron frequency ν_C and the spin-cyclogron beat (g-2) frequency $\nu_D = \nu_S - \nu_C$ on a single electron stored in a Penning trap at ~4 K.*

The axial oscillation is driven by an rf signal applied to one of the end caps of the Penning trap, and the motion of the trapped electrons is observed by detecting the current induced in the other end cap. The two signals are compared in a phase sensitive detector and the difference or error signal is integrated, lowpass filtered, and fed back to V_0 as a dc correction voltage, so that the axial oscillation frequency is locked to the drive frequency. This correction voltage is proportional to any shift $\delta\nu_z$ in the natural axial frequency. The amplitude of the detected signal is used to indicate the number of electrons in the trap. The *g*-2 measurements were made with only a single electron in the trap at a time, thus eliminating space charge electric fields.

In a perfect Penning trap the quadrupole electric field shifts the frequency ν_C by an amount equal to the magnetron frequency ν_m, i.e.,

$$\nu_C' = \nu_C - \nu_m \,. \tag{12}$$

In the non-relativistic limit, ν_S is unshifted by the electric field, yielding

$$\nu_D' = \nu_S - \nu_C' = \nu_D + \nu_m \tag{13}$$

as shown in Fig. 3. A useful expression relating ν_m to the

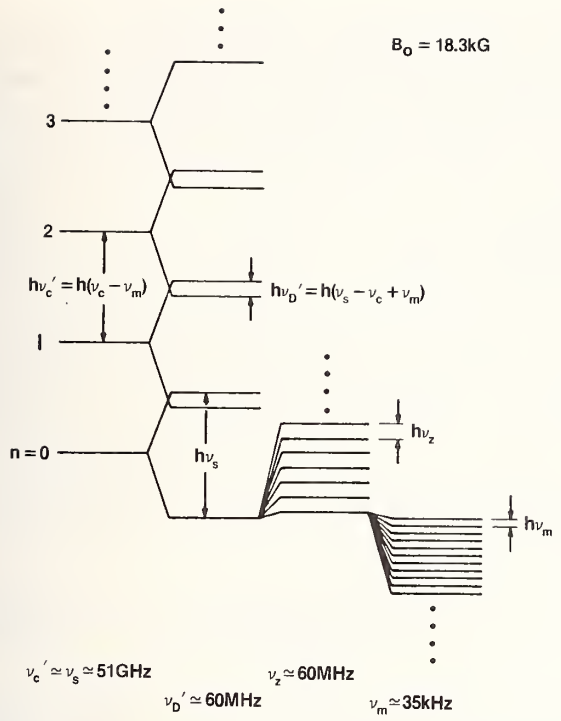


FIGURE 3. *Energy levels of an electron in a Penning trap.*

more easily measured quantities ν_z and ν_C' may be obtained from Eqs. (11) and (12) as

$$2\nu_m \nu_C' = \nu_z^2 \ . \tag{14}$$

The shift in ν_D [Eq. (13)] is typically several hundred ppm, four orders of magnitude larger than the final quoted error, and thus requires accurate measurement.

The cyclotron and difference frequencies are measured by means of their coupling to the axial motion, a coupling provided by the introduction of a small magnetic inhomogeneity of the form

$$B_z = B_0 + B_2 \left[z^2 - \frac{r^2}{2} \right] \tag{15}$$

with $B_2 \sim (120-300)$ G/cm^2. A nickel wire wound around the central electrode produces this field configuration. The inhomogeneity causes a shift in the axial frequency which is dependent on the spin, cyclotron, and magnetron quantum numbers (m, n, and q respectively):

$$\delta\nu_z = [m + n + 1/2 + (\nu_m/\nu_C)q\,]\delta \ , \tag{16}$$

with $\delta \simeq \mu_B\beta/2\pi^2 m_0 \nu_z \simeq 1$ Hz.

The cyclotron transition ($\Delta n = 1$, $\Delta m = 0$) is sufficiently strong so that changes in n, and thus in $\delta\nu_z$, are easily observed due to random thermal excitation even at the 4.2 K ambient temperature of this experiment. Similarly, the cyclotron frequency is measured by observing even larger changes in n and therefore shifts in ν_z when an external signal at ν_C' is applied (Fig. 2).

Spin-cyclotron transitions ($\Delta n = -1$, $\Delta m = +1$) are driven by applying an rf voltage to one end cap at frequency ν_{rf}. The electron moves axially at the driving frequency ν_{rf} while moving in cyclotron orbits at ν_C, so that it sees the radial magnetic field component associated with the axial inhomogeneity oscillating at the frequency $\nu = \nu_C \pm \nu_{rf}$. Since $\nu_S = \nu_C + \nu_D$, the spin flip transition is driven only when $\nu_{rf} = \nu_D$. The resulting change in $m(\Delta m = +1)$ does not produce an immediate change in $\delta\nu_z$ because of the simultaneous $\Delta n = -1$ cyclotron transition. The spin flip is observable, however, as a shift in the baseline of the cyclotron thermal noise output, since from Eq. (16), $\delta\nu_z$ increases by δ during a spin flip. To achieve narrow linewidths, the axial motion is driven alternately at $\nu_{rf} = \nu_D$ during the "transition" phase and then at ν_z during the "detection" phase.

The measured frequencies ν_D', ν_C' and ν_z [from which ν_m can be calculated using Eq. (14)] are then related to the anomaly by

$$a(e) = \frac{\nu_D'}{\nu_0} = \frac{\nu_D - \nu_m}{\nu_C + \nu_m} \ . \tag{17}$$

Shifts in $a(e)$ are caused by the non-uniform magnetic field, relativistic effects, misalignment of the electric and magnetic field axes, etc. These and other systematic affects have been considered by the Washington group but details have not yet been published. Their final result obtained for data runs at 18.6, 32.0, and 51.1 kG is [3]:

$$a(e^-) = 0.001\ 159\ 652\ 200(40) \ .$$

The precision quoted is noted to be due primarily to systematic error and not linewidth as was the case in the previous (200 ppb) result.

This resonance technique can be applied equally well to a positron, whose magnetic moment is expected to be exactly equal to that of the electron [$g(e^+) = g(e^-)$] if CPT is conserved. The principal new difficulty is the injection of positrons from an energetic radioactive source into a shallow electrostatic well. The Washington group has overcome this problem by the use of a separate trap in which the positrons are collected and cooled. Positrons are later transferred to the more precisely controlled measurement trap. The result

$$a(e^+) = 0.001\ 159\ 652\ 222(50)$$

has been obtained [4]. Since $a(e^+)$ agrees with $a(e^-)$ within the quoted error, the CPT theorem has been verified at the level of 0.05 ppb.

Improvements in the resonance technique, including the possible elimination of the magnetic bottle, are expected to yield even higher accuracy for both the electron and positron results in the future. Discussion of the new positron experiment and of the uniform field technique appear as contributed papers in these proceedings.

2.2 The Michigan Precession Experiments

The free precession technique is an alternative method for measuring the anomaly and was in fact the first method to yield the magnetic moment of a free particle (1953). In this technique, polarized electrons or positrons are trapped in a magnetic or electric well in a magnetic field (Fig. 4) and, to a first approximation, the spin precesses relative to the velocity at $\nu_D = \nu_S - \nu_C = a\nu_0$, as discussed previously. The spin direction is observed after a known trapping time in order to determine ν_D. The frequency ν_0 is then determined by measuring the proton NMR frequency in water ν_P' for the magnetic field in which the particles are trapped. This frequency has been compared to the electron spin-flip frequency in a hydrogen maser by Phillips *et al.* [5] who found:

$$\nu_0 = \nu_P'/(1.520993134(21) \times 10^{-3}) \ .$$

A detailed discussion of the precession experiments, including the various corrections to the formula $\nu_D = a\nu_0$ may be found in Refs. [1] and [2].

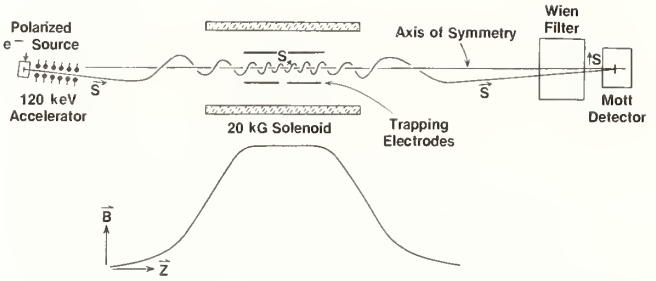


FIGURE 4. *Schematic of the proposed University of Michigan precession experiment to measure the electron anomaly. Polarized electrons produced by photoemission are accelerated along a helical path into the central field of a superconducting solenoid. An electrostatic pulse traps the electrons in a shallow magnetic well. After a predetermined time, the electrons are ejected and the variation of the longitudinal polarization with trapping time is measured with a Wien filter/Mott detector combination.*

A variation of this precession technique has been proposed by the Michigan group [6]. Electrons with polarization close to 100% are photoemitted from a cesium vapor (Fano effect) or from a magnetized nickel crystal and will be injected at 120 keV into an extremely shallow (20 to 200 ppb) magnetic trap in the 20 kG magnetic field of a superconducting solenoid (Fig. 4). After precessing for 10 to 20 msec (10^6 cycles of ν_D), the electrons are ejected and spiral out of the field. The trajectories of these particles have been computed using numerical integration. We note that the results are in agreement to within 1% with those obtained analytically on the assumption that the orbital magnetic moment $\mu_{ORB} \sim P_\perp^2/B$ ($P_\perp$ is the momentum perpendicular to the magnetic field direction) is an adiabatic invariant. These studies show that when the particles reach the position of the Wien filter, their momentum direction is within 6 degrees of the axis of the solenoid.

Since, if $g = 2$, the projection of spin onto momentum remains constant in a magnetic field, the longitudinal polarization of the electrons at the Wien filter is modulated at ν_D. This modulation is independent of the position of the particle in its cyclotron orbit. The effect of the anomaly ($g \neq 2$) is to cause a calculated 10% depolarization of the ejected beam as it spirals out of the magnetic field. The electrons then pass through the Wien filter which rotates the longitudinal component of spin by 90°, and their polarization (now modulated in the transverse direction) is measured by Mott scattering.

It is planned to trap approximately 100 electrons per cycle. With this number of electrons in the trap, ν_D is shifted by at most 3 ppb due to the space-charge electric fields. With these parameters and a repetition rate of 30 Hz, a statistical uncertainty of 10 ppb in ν_D may be obtained in only 30 minutes. The overall accuracy of the experiment will be limited to about 20 ppb primarily by systematic effects. The major anticipated systematic uncertainties are due to the current uncertainty in ν_P'/ν_0 (11 ppb), corrections for the magnetic well (10 ppb), and radial electric fields in the trapping region (5 ppb).

Major progress toward the realization of this experiment has already taken place. A ten kilogauss 60 cm diameter by 170 cm long superconducting solenoid, corrected to sixth order, has been constructed and used to contain positrons of 650 keV energy in a 60 ppm magnetic mirror trap for times up to 1.2 milliseconds. Using this solenoid, numerous tests relevant to the proposed electron experiment were completed. We conclude from these tests that the system necessary for a 20 ppb measurement of $a(e^-)$ can be constructed. In addition, an attempt was made to measure $a(e^+)$ using the previously described [6] combination of the resonance and precession techniques. No resonance was seen but the statistics were quite marginal and the time for a thorough search for the resonance, possibly with a new and stronger source, has not been taken because of emphasis on the electron work.

The importance of performing an electron precession experiment to this accuracy discussed above lies both in its value as a systematic check on the resonance experiment described previously and because, as a consequence of the relativistic electron velocity in this experiment, it constitutes a novel test of special relativity (Section 4.2). In addition it would provide the most accurate *direct* limit on a possible electron electric dipole moment (3×10^{-18} e-cm). This is not as precise as limits obtained from measurements on neutral atoms (10^{-24} e-cm) but has the virtue of being independent of the theoretical connection between an atomic EDM and the corresponding electron EDM.

2.3 The Novosibirsk Electron-Positron Comparison Experiment

As stated previously, the CPT theorem has been verified at the 5×10^{-11} level by separate measurements of the magnetic moments made on 10^{-3} eV electrons and positrons using the resonance technique. A much less precise but still quite interesting experiment has been performed by Serednyakov *et al.* [7], who compared but did not directly measure $a(e^+)$ and $a(e^-)$ for 625 MeV particles in the VEPP-2M storage ring at Novosibirsk. In the storage ring, the electron and positron beams became polarized parallel to the ring axis, with a time constant of order one hour, due to the emission of spin-flip synchrotron radiation. The beams are then depolarized by an oscillating longitudinal magnetic field in one section of the ring (Fig. 5). The applied frequency ν_r is chosen so that the relative phase between the particle spins and the depolarizing field remains unchanged after each orbit. A frequency which satisfies this condition is $\nu_r = \nu_S - 2\nu_C$. The change in polarization is then detected as a change in the spin-dependent (e^+e^+) and (e^-e^-) scattering which occurs within the beam bunches. Both the electrons and the positrons were seen to depolarize at the same applied frequency, from which the result is derived:

$$\left| \frac{a(e^+) - a(e^-)}{a(e)} \right| \leq 1.0 \times 10^{-5} \quad (95\% \text{ confidence})$$

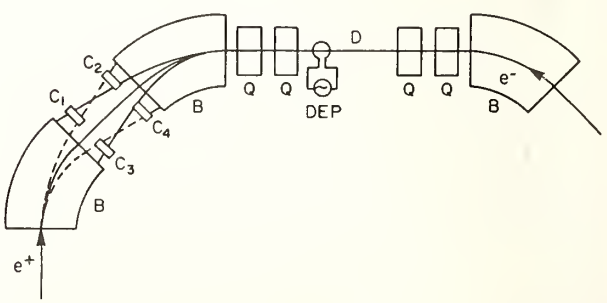


FIGURE 5. *The apparatus at Novosibirsk used for the simultaneous measurement of e^+, e^- depolarization in the VEPP-2M storage ring. B: bending magnets; Q: quadrupole lenses; DEP: depolarizer; $C_1 - C_4$: scintillation counters in 4-fold coincidence.*

2.4 The SLAC High Energy Experiment

A measurement of the electron anomaly at SLAC is of particular interest due to its high energy, 12 GeV ($\gamma = 2.5 \times 10^4$), and its resulting sensitivity as a test of special relativity [8]. This measurement is a by-product of the determination of the polarization of the SLAC longitudinally polarized electron beam [9].

As a polarized electron beam is deflected through a 24.5° bend (θ_c) in the SLAC switchyard magnetic field, its precession is observed. The spin rotates by an angle $\theta_D = a\gamma\theta_c$ (recall $\nu_D = a\gamma\nu_C$) relative to the momentum in this magnet. The longitudinal polarization of the beam after deflection is a sinusoidal function of the beam energy $E = \gamma m_0 c^2$, the magnitude of the initial polarization P_0, and the angle between the initial beam momentum and polarization ϕ_0, i.e.,

$$P(E) = P_0 \cos(\theta_D + \phi_0) = P_0 \cos(\pi E/E_0 + \phi_0) \ . \quad (18)$$

Here the parameter E_0, defined as $\pi m_0 c^2/a\theta_c$, is about 3.2 GeV. Figure 6 shows the longitudinal polarization measured by Moller scattering at values of γ between 1.27×10^4 and 3.81×10^4. These data points (three extrema and one zero-cross measurement) were fit to the above formula with E_0 (i.e., a) and P_0 as free parameters and with ϕ_0 assumed to be zero. The result $a(e^-) = 0.001\ 162\ 2(200)$, derived from the best fit value of E_0, is in good agreement with the more precise low energy measurements. A discussion of the g-2 experiments and this experiment in particular as a test of special relativity may be found in Section 4.2.

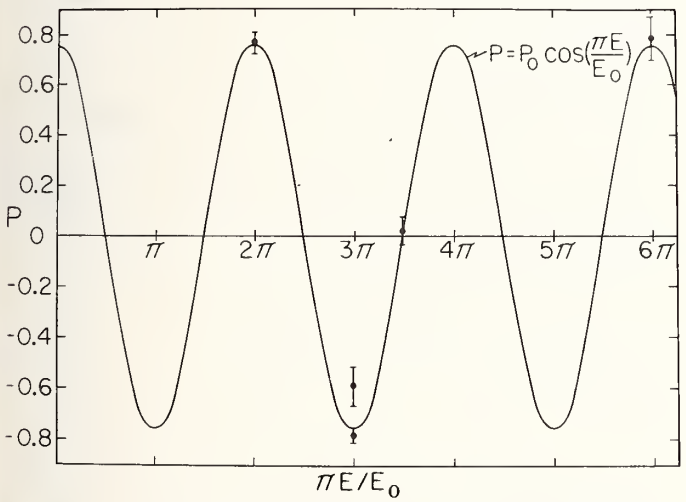


FIGURE 6. *The results of the SLAC high-energy electron g-2 measurement. The longitudinal component P of the electron beam polarization is plotted as a function of $\pi E/E_0$, the angle through which the spin precesses relative to the momentum during the 24.5° magnetic bend in the experimental area. The curve shown is the best fit to the data with a(e) and P_0 as free parameters.*

3. The CERN Muon g-2 Experiment

3.1 The g-2 Determination

The CERN group has performed a series of muon g-2 measurements resulting most recently in the separate determinations of $a(\mu^+)$ and $a(\mu^-)$ to 10 ppm. This work tests the theoretical QED prediction of the muon anomaly including an estimate of the eighth-order contribution, checks hadronic contributions calculated to be 67 ppm, sets new limits on the muon EDM, and tests special relativity with regard to Thomas precession and the twin paradox. In the CERN experiment, pions are injected into the 7-m storage ring (Fig. 7) with a momentum slightly greater than the 3.098 GeV/c design momentum of the ring. About 10% of the pions decay before exiting the storage volume. Muons produced in the top 1.5% of their allowed momentum range are then trapped.

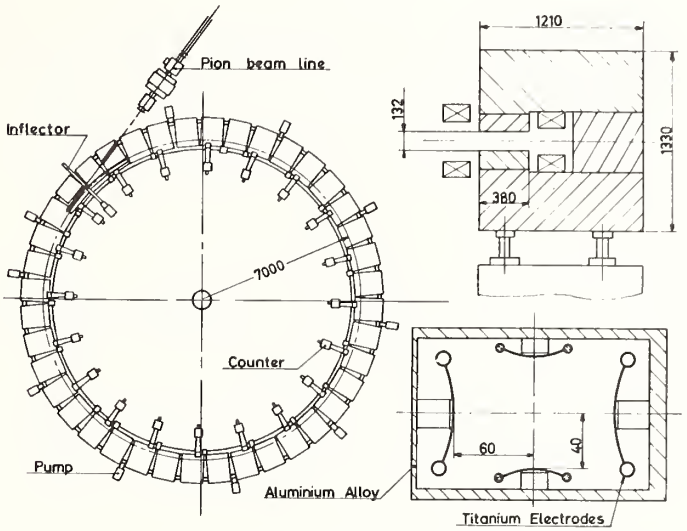


FIGURE 7. *Plan view of the second CERN muon storage ring with cross-sections of the magnets and focusing electrodes. (Dimensions in mm.)*

They are longitudinally polarized, due to the parity nonconservation in the pion decay, and at the forward angles, where efficient muon trapping occurs, a polarization of better than 90% is obtained. The muon spins precess around the magnetic field direction until the muons decay in the ring. The resulting electrons are detected in energy-sensitive shower counters and an energy cut selects those electrons of higher energy emitted in the forward direction with respect to the muon momentum. Since the electrons are preferentially emitted along the muon spin direction, the counting rate of high energy electrons will be modulated at the g-2 frequency as the muon spins precess with respect to their momentum at this frequency.

High precision was obtained, in part, by the use of a uniform magnetic field in conjunction with an electric field for vertical focusing. The electric field constitutes an electric quadrupole trap similar to that described for the resonance experiments. As in the electron experiment, the electric field $\mathbf{E}$ introduces a shift in $\boldsymbol{\omega}_D$ (the angular velocity of the spin vector relative to the velocity) given by

$$\boldsymbol{\omega}'_D = \boldsymbol{\omega}_D + \frac{e}{m_0 c}\left[\frac{1}{\gamma^2 - 1} - a\right](\boldsymbol{\beta} \times \mathbf{E}) \quad (19)$$

where $\boldsymbol{\beta} = \mathbf{v}/c$ ($\nu_D = |\boldsymbol{\omega}_D/2\pi|$). The CERN group eliminates the effect of $\mathbf{E}$ on $\boldsymbol{\omega}_D$ by setting the energy to 3.098 GeV so that $\gamma = (1 + 1/a)^{1/2} = 29.3$, and the electric field dependent term in Eq. (19) vanishes.

The decay data in Fig. 8 are fitted for ν_D and compared to the NMR frequency ν'_P for the magnetic field of the storage ring. After averaging over muon orbit distribu-

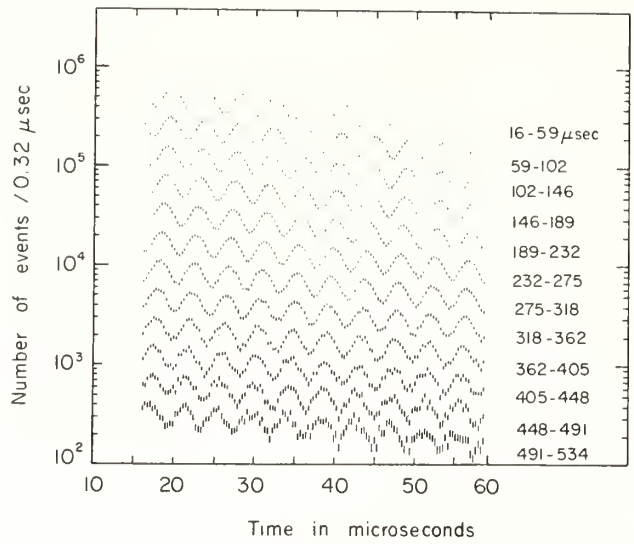


FIGURE 8. *Decay electron time distribution from the CERN muon g-2 experiment. The distribution contains 1.4 × 10⁸ events.*

tions and correcting for NMR probe shielding, calibration, and other effects, the result obtained is

$$R = \nu_D(\mu)/\nu'_P = 3.707213(27) \times 10^{-3} . \quad (20)$$

The ratio of the magnetic moment of the muon to that of the proton (λ) was found by averaging the results of Crowe *et al.* [10] and Casperson *et al.* [11] for the ratio of the muon Larmor frequency to the proton Larmor frequency (or muon moment to proton moment):

$$\lambda = \nu_L(\mu)/\nu'_P = 3.1833417(39), \quad (21)$$

so that using the relations $\nu_L(\mu) = (1 + a(\mu))\nu_0(\mu)$ and $\nu_D(\mu) = a\,\nu_0(\mu)$ we find $a(\mu) = R/(\lambda - R)$. The final result for μ^+ and μ^- data together yields [12]

$$a(\mu) = 0.001\ 165\ 924(8) \quad (22)$$

and for the individual particles

$$a(\mu^+) = 0.001\ 165\ 912(11)$$
$$a(\mu^-) = 0.001\ 165\ 938(12) .$$

These values are all in agreement with the theoretical prediction of

$$a(\mu\text{-theory}) = 0.001\ 165\ 920(10) .$$

The agreement between $a(\mu^+)$ and $a(\mu^-)$ constitutes a verification of CPT at the 10^{-8} level.

3.2 Limits on the Muon Electric Dipole Moment

The muon g-2 experiments also provide the best limits to date on the electric dipole moment (EDM) of the muon. (See Ref. [1] for a complete discussion of lepton EDM's and the theories in which they arise.)

A non-zero muon EDM would cause the muon spin to precess about both the electric and magnetic field directions with a resultant motion determined by the **E** and **B** field strengths and by the electric and magnetic moments. One result of this motion would be that the muon spin would oscillate above and below the plane of the storage ring with a frequency of ν_D and 90° out of phase with the g-2 motion, thus causing the emission of decay electrons or positrons in a direction along **B** to be modulated at ν_D. The data from scintillation counters placed above and below the median plane were analyzed for this effect, from which the limit [1] for the EDM (combining μ^+ and μ^- data)

$$D_\mu = -(3.7 \pm 3.4) \times 10^{-19} \text{ e-cm}$$

was derived.

A comparable limit of

$$D_\mu < 8 \times 10^{-19} \text{ e-cm} \quad (95\% \text{ confidence level})$$

can be independently obtained by ascribing any disagreement between the predicted and observed ν_D to the additional precession of the muon EDM caused by the electric field, i.e.,

$$\nu_D = \nu'_D \left[1 + \frac{f_\mu^2}{4a^2}\beta^2\right]$$

with f_μ defined by the relation

$$D_\mu = \frac{f_\mu}{2}\frac{e\hbar}{2m_\mu c} .$$

4. Theoretical Implications of the g-2 Experiments

4.1 QED and Other Contributions to the Muon and Electron Anomalies

The anomalous moment of the electron or muon is predicted as a power series in α, the fine structure constant, plus contributions from other known or speculative effects:

$$a = \Sigma A_n(\alpha/\pi)^n + H \text{ (hadron loops)} + W \text{ (weak interaction effects)} + X \text{ (axions)} \quad (23)$$

where the A_n represent only QED contributions including muon and tau loops. The coefficients in Eq. (23) are given in Table 1 with $\alpha = e^2/\hbar c = [137.035963(15)]^{-1}$ [13] used to evaluate a (theoretical).

TABLE 1. *Theoretic contributions to the electron and muon g-factor anomalies.*

Term[a]	Electrons	Muons	References
A_1	0.5	0.5	Schwinger [30]
A_2	−0.328478444	+0.7658582(7)[b]	Kinoshita [14] Calmet [24]
A_3	1.1765(13)	24.45(5)	Levine [18] Calmet [24]
A_4	−2.3(3.6)	128(70)	Kinoshita [19] Calmet [24]
H	$1.6(2) \times 10^{-12}$	$66.7(9.4) \times 10^{-9}$	Calmet [24]
W	0.05×10^{-12}	$2.1(2) \times 10^{-9}$	Kinoshita [14] Calmet [24]
X	0.1×10^{-12}	3.7×10^{-9}	Leveille [20] Kane [21]
a_{th}	0.001 159 652 416(165)	0.001 165 920(10)	Kinoshita [18] Field [1]

[a]See text for an explanation of symbols used and for discussion.
[b]Error due to uncertainty in τ mass.

In calculating the electron anomaly, the uncertainty comes primarily from the 109 ppb uncertainty in the measurement of α. The non-QED-dependent determination of α is limited by the uncertainty in γ'_P, the gyromagnetic ratio of protons in water [13]. Experiments currently under way are expected to improve this measurement by an order of magnitude in the next few years [15]. The value of α can also be determined from measurements of the hyperfine structure of muonium (QED-dependent) which are currently in progress [16]. A new method for obtaining a non-QED-dependent value for α from the quantized Hall resistance of a two-dimensional electron gas in a silicon metal-oxide-semiconductor has recently been suggested [17] and may be capable of producing precise results in the coming decade. A new result for $A_3(e)$, reported at this conference by Levine and Roskies [18] $A_3(e) = 1.1765(13)$, reduces the previous uncertainty from this source, 66 ppb, to only 15 ppb. Evaluation of $A_4(e)$, a monumental task involving a calculation of 891 Feynman diagrams, is in progress and is expected on completion to reduce the uncertainty from this term in $a(e)$ to the order of 1 to 10 ppb. A preliminary result [19] reported in these proceedings is $A_4(e) = -2.3(3.6)$. Thus, with these advances in both the QED calculations and reduction of the error in the value of α, it is expected that the total uncertainty in the predicted value of $a(e)$ will be reduced below the level of $(\alpha/\pi)^4 \simeq 30 \times 10^{-12}$ in the near future.

Contributions in Eq. (23) to $a(e)$ from muon and tau loops are known accurately. The contributions from the standard gauge model of the weak interaction (0.05 ppb) and for hadronic effects (1.6 ppb) are both below the resolution of electron g-2 experiments for the foreseeable future.

One can also consider the effect on $a(e)$ or $a(\mu)$ of more speculative interactions. For example, the effect of a light scalar or pseudoscalar neutral particle (axion) on the anomaly depends on the mass of the axion m_a and on the coupling assumed. One standard model (Leveille [20], Kane [21]) predicts in the limit $m_a \to 0$ that $a(e)$ will be shifted by 0.1×10^{-12} and $a(\mu)$ by 3.7×10^{-9}.

The prediction of the muon anomaly $a(\mu)$ includes sizable contributions from the $A_4(\mu)$ term [22] (3.7 ± 2.1 ppm) as well as from hadronic effects [23, 24] (66.7 ± 9.4 ppm) and weak interaction effects [24] (2.1 ± 0.2 ppm). Thus a comparison of the most precise experimental value of $a(\mu)$, derived from the weighted average of the measurements for μ^+ and μ^-, allows a check of the hadronic contribution to a lepton magnetic moment for the first time. The results are in good agreement both with QED and with CPT as summarized below:

$$\frac{a(\mu,\ \text{theory}) - a(\mu,\ \text{experimental})}{a(\mu,\ \text{theory})} = (-3 \pm 11)\ \text{ppm}$$

$$\frac{a(\mu^+) - a(\mu^-)}{\frac{1}{2}\left[a(\mu^+) + a(\mu^-)\right]} = (-22 \pm 14)\ \text{ppm} .$$

If a next generation of muon experiments is undertaken with error in $a(\mu)$ at the 1 ppm level it should be sensitive to the A_4 and weak interaction terms.

If leptons are not point-like, their substructure should affect $a(e)$ and $a(\mu)$ at some level. Using dimensional arguments, Brodsky and Drell [25] have shown that one could expect contributions to $a(l)$ linear in m_l/m^* ($\delta a(l) \sim m_l/m^*$), where m^* is the mass of the constituent and l refers to e or μ. Present g-2 measurements rule out electron and muon constituents with mass less than $\sim 10^6\ \text{GeV}/c^2$ and $\sim 2 \times 10^6\ \text{GeV}/c^2$ respectively, corresponding to an effective size no greater than $R \sim 10^{-20}$ cm for either particle. This limit on lepton size is four orders of magnitude smaller than the best current limits obtained from electron-positron collisions at a 32 GeV center of mass energy. Other models which give poorer limits on R are discussed in Ref. [25].

4.2 Tests of Special Relativity

The data obtained in the muon g-2 experiments may also be used to determine the lifetime τ of muons at the accurately known energy ($\gamma = 29.327$) with which the muons circulate in the storage ring. This is analogous to a test of the famous "twin paradox" and is a direct test of time dilation. In agreement with special relativity, the CERN group finds [26]:

$$\frac{\tau/\gamma - \tau_0}{\tau_0} = (-9 \pm 4) \times 10^{-4} .$$

where τ_0 is obtained from recent measurements of the rest lifetime accurate to 40 ppm [27].

More recently, new limits of very high precision have been placed on relativity violations by lepton g-2 measurements [8, 28, 29]. As shown in Eqs. (6) and (7), the spin and cyclotron frequencies depend on γ. As discussed in Ref. [28], if the γ from Thomas precession appearing in ν_S is not identical with the γ arising from electron dynamics (referred to as $\tilde{\gamma}$) in ν_C, then the difference frequency ν_D would no longer be independent of energy:

$$\frac{\nu_D}{\nu_0} = a + (1 - \gamma/\tilde{\gamma}) . \qquad (24)$$

Comparison of ν_D/ν_0 measurements at two different velocities then constitutes a test of special relativity which is independent of QED and which may be parameterized by the quantity Δ:

$$\Delta \equiv |[\nu_D/\nu_0]_1 - [\nu_D/\nu_0]_2| = |[\gamma/\tilde{\gamma}]_2 - [\gamma/\tilde{\gamma}]_1| . \qquad (25)$$

The results of the most recent Michigan measurement of the electron anomaly at $\beta = 0.57$, when compared with the low energy ($\beta \sim 10^{-3}$) Washington measurement, yields $\Delta = (5.3 \pm 3.5) \times 10^{-9}$. The current CERN muon g-2 measurements at $\gamma = 29$ can be similarly compared to the 1968 CERN result at $\gamma = 12$ to yield $\Delta = (236 \pm 310) \times 10^{-9}$. Finally, the agreement between the SLAC g-2 measurement with 12 GeV electrons and low-energy results constitutes an ultra-relativistic ($\gamma = 2.5 \times 10^4$) verification of relativity with $\Delta = (2600 \pm 20{,}000) \times 10^{-9}$.

An alternative approach to the problem of characterizing the relativity breakdown is presented in Ref. [29]. In both Refs. [8] and [29] the relativity violation is assumed to take a specific form, in which the parameter $(1 - \gamma/\tilde{\gamma})$ or its equivalent is expressed as a power series in $(\gamma - 1)$. The coefficient of the second term in this series then constitutes the figure of merit for the various experiments. Limits approaching 2×10^{-8} were obtained by Michigan and CERN, and 8×10^{-10} by SLAC for the coefficient in Ref. [8] while limits $1/a$ times larger are obtained using the model of Ref. [29]. The major conclusion to be drawn from the above work is that the g-2 experiments constitute a novel and sensitive test of special relativity.

5. Conclusions

The theoretical prediction of the electron and muon magnetic moments and their experimental determinations continue to constitute the most direct and precise tests of

quantum electrodynamics. Indeed, considering only the electron g-factor, the comparison of theory and experiment constitutes the most precise test of any extant physical theory. Although no new muon experiments are currently underway, the electron g-2 research is very active. We may conclude from a reasonable projection of what has been presented at this conference that the next several years could see well established comparisons of theory and experiment for the electron g-factor anomaly at the 10 ppb level. A direct electron-positron g-2 comparison by the resonance technique might even exceed this accuracy. Such results allow the study of a wide range of physical phenomena including theories with exotic interactions, relativity and symmetry tests, as well as providing a more stringent test of QED.

References

[1] J. H. Field, E. Picasso, and F. Combley, Sov. Phys. Usp. 22(4), 199 (1979).

[2] A. Rich and J. C. Wesley, Rev. Mod. Phys. 44, 250 (1972).

[3] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Bull. Am. Phys. Soc. 24, 758 (1979).

[4] P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt, Bull. Am. Phys. Soc. 26, 597 (1981); and these proceedings.

[5] W. D. Phillips, W. E. Cooke, and W. Kleppner, Phys. Rev. Lett. 35, 1619 (1975).

[6] D. Newman, E. Sweetman, R. Conti, and A. Rich, in *Atomic Masses and Fundamental Constants 6*, Ed. by J. A. Nolen, Jr., and W. Benenson (Plenum Press, New York, 1980), p. 183.

[7] S. I. Serednyakov, V. A. Sidorov, A. N. Skrinsky, G. M. Tumaikin, and Iu. M. Shatunov, Phys. Lett **66B**, 102 (1977).

[8] P. S. Cooper, M. J. Alguard, R. D. Ehrlich, V. W. Hughes, K. Kobayakawa, J. S. Ladish, M. S. Lubell, N. Sasao, K. P. Schuler, P. A. Souder, D. H. Coward, R. H. Miller, C. Y. Prescott, D. J. Sherder, C. K. Sinclair, G. Baum, W. Raith, and K. Kondo, Phys. Rev. Lett. 42, 1386 (1979).

[9] P. S. Cooper, M. J. Alguard, R. D. Ehrlich, V. W. Hughes, H. Kobayakawa, J. S. Ladish, M. S. Lubell, N. Sasao, K. P. Schuler, P. A. Souder, G. Baum, W. Raith, K. Kondo, D. H. Coward, R. H. Miller, C. Y. Prescott, D. J. Sherden, and C. K. Sinclair, Phys. Rev. Lett. 34, 1589 (1975).

[10] K. M. Crowe, J. F. Hague, J. E. Rothberg, A. Schenck, D. L. Williams, R. W. Williams, and K. K. Young, Phys. Rev. D 5, 2145 (1972).

[11] D. E. Casperson, T. W. Crane, A. B. Denison, P. O. Egan, V. W. Hughes, F. G. Mariam, H. Orth, H. W. Reist, P. A. Souder, R. D. Stambaugh, P. A. Thompson, and G. zu Putlitz, Phys. Rev. Lett. 38, 956 (1977).

[12] J. Bailey, K. Borer, F. Combley, H. Drumm, C. Eck, F. J. M. Farley, J. H. Field, W. Flegel, P. M. Hattersley, F. Krienen, F. Lange, G. Lebee, E. McMillan, G. Petrucci, E. Picasso, O. Runolfsson, W. von Ruden, R. W. Williams, and S. Wojcicki, Nucl. Phys. B 150, 1 (1979).

[13] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. 42, 1575 (1979).

[14] T. Kinoshita, paper presented at "Fundamental Interactions' Colloquium in honor of Antoine Visconti, Centre de Physique Theorique, CNRS Marseille-Luminy (1979).

[15] E. R. Williams, private communication (1980).

[16] V. W. Hughes, private communication (1980).

[17] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).

[18] M. Levine and R. Roskies, these proceedings.

[19] T. Kinoshita and W. B. Lindquist, these proceedings.

[20] J. P. Leveille, Nucl. Phys. B 137, 63 (1978).

[21] G. Kane, private communication (1980).

[22] J. Calmet, S. Narison, M. Perrottet, and E. de Rafael, Rev. Mod. Phys. 49, 21 (1977).

[23] V. Barger, W. F. Long, and M. G. Olsson, Phys. Lett. **60B**, 89 (1975).

[24] J. Calmet, S. Narison, M. Perrottet, and E. de Rafael, Phys. Lett. **61B**, 283 (1976).

[25] S. J. Brodsky and S. D. Drell, Phys. Rev. D **22**, 2236 (1980).

[26] J. Bailey, K. Borer, F. Combley, H. Drumm, F. J. M. Farley, J. H. Field, W. Flegel, P. M. Hattersley, F. Krienen, F. Lange, E. Picasso, and W. von Ruden, Nature **268**, 301 (1977).

[27] M. P. Balandin, W. M. Crebenyuk, V. G. Zinov, A. D. Konin, and A. N. Ponomarev, Sov. Phys. JETP **40**, 811 (1974).

[28] D. Newman, G. W. Ford, A. Rich, and E. Sweetman, Phys. Rev. Lett. **40**, 1355 (1978).

[29] F. Combley, F. J. M. Farley, J. H. Field, and E. Picasso, Phys. Rev. Lett. **42**, 1383 (1979).

[30] J. S. Schwinger, Phys. Rev. 73, 416 (1948); and *ibid.* 75, 898 (1949).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Preliminary Comparison of the Positron and Electron Spin Anomalies*

P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt

Department of Physics, University of Washington, Seattle, WA 98195

A new double Penning trap structure has been built using various techniques developed in our geonium experiment. Primary positrons are captured in a storage trap, centered, and then transferred into a well-compensated experiment trap where preliminary measurements have yielded the single positron g-factor anomaly $a(e^+) = (1159652222 \pm 50) \times 10^{-12}$. This value was obtained from four runs at a field of 50.8 kG with the $a(e^+)$ values extrapolated to zero spin flip power using the power dependence observed in the electron geonium experiment. The uncertainty is based on the resonance linewidths and an estimate of the remaining systematic errors. When compared to the electron spin anomaly, we obtain a positron/electron g-factor ratio of $1 + (22 \pm 64) \times 10^{-12}$.

Key words: compensated Penning trap; fundamental constants; matter-antimatter comparison; positron-geonium spin anomaly; positron trapping; precision measurements; single isolated positron; synchronous resonant detection.

1. Introduction

Recent success measuring the electron spin anomaly in a well-compensated Penning trap [1] points out the obvious potential of that device for measuring the positron anomaly to the same precision. Previously, the positron had been measured by Rich and Gilleland [2] to an accuracy of 1000 ppm and indirectly by Serednyakov *et al.* [3] to 10 ppm. In the first experiment, the anomaly is measured in a magnetic mirror machine by directly observing the difference between the precessing polarization of the spin of the positron and its cyclotron rotation and then calibrating the magnetic field using an NMR probe. However, the crucial weakness of that experiment was the need for an intense polarized positron source and an efficient positron polarimeter which did not exist at that time [4]. The other previous experimental result was obtained using intersecting electron/positron beams in a storage ring which allowed the anomaly frequency of each charge to be directly measured. When combined with the better known electron anomaly [1], the ratio $a(e^+)/a(e^-)$ yields the unknown positron anomaly.

In comparison, using well-compensated Penning traps, our resonance techniques can attain accuracies of a few parts in 10^8 for the single positron anomaly, with the magnetic field calibrated directly *in situ* with the positron's cyclotron frequency. In addition, the ratio of the positron and electron g-factors should be measurable to even greater accuracy since small systematic errors should be identical to first order. Thus, a very sensitive test of matter/antimatter symmetry should be possible. Any violation of that symmetry will be a violation of CPT invariance [5].

2. Experimental Loading Technique

The principal challenge in this experiment was to devise a means of continuously loading positrons into a Penning trap. After careful and tedious construction of this Penning trap, a static scheme that was previously proposed by us [6] was found to be adequate to trap many positrons each hour. This successful scheme was reported upon previously [7] and is summarized here.

A cross section of the Penning trap used to catch the positrons is shown in Fig. 1. The principal difference between this positron trap and the compensated Penning trap used in the electron magnetic moment anomaly experiment [1] is the addition of a sealed sodium-22 positron source and a field emission point (electron source) mounted off-axis in an endcap as shown. During the positron loading process, the positron source is negatively biased to approximately 300 V. This source bias produces the potential shown in Fig. 2 along the load-line, which is the line parallel to the z-axis that extends from the positron source into the trap through the center of the hole in the endcap.

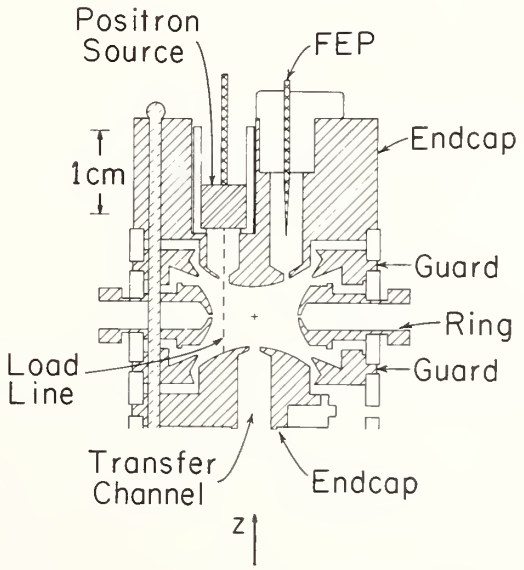


FIGURE 1. *Positron storage trap. The major changes in the basic compensated Penning trap design are an off-axis sealed sodium-22 positron source, an off-axis field emission electron source (for diagnostic studies) and a drift channel in the endcap opposite the source for transferring positrons into a very well compensated experiment Penning trap.*

*This research is supported by the National Science Foundation.

Positrons, emitted by the source with energies of approximately 100 keV (most of which is in the cyclotron motion), travel along the load-line into the trap guided by the axial magnetic field. Since the maximum potential along the load-line in the vicinity of the entry hole is less than the far endcap potential (zero), some of the injected positrons will turn around before hitting the opposite endcap. Due to the magnetron motion in the trap, the positrons are moved off the load-line as they travel through the trap. Hence they see a slightly larger potential (~1 mV) when they return to the entry hole, as illustrated by the broken line in Fig. 2 (greatly exaggerated). If the initial axial energy of the injected positron is low enough, the positron will be trapped for one magnetron period. During this time, enough axial energy can be extracted by a damping circuit (connected to an endcap) tuned to the relativistically shifted axial frequency ν_z, to permanently trap and subsequently thermalize the positron's z-axis motion. The positron's cyclotron motion is thermalized in a few seconds or less by synchrotron radiation. Finally, a motional sideband cooling technique [1] is used to radially center the positrons by applying an rf drive at $\nu_z + \nu_m$ to an asymmetrically placed probe (also called the SBE probe, shown in Fig. 3), where ν_m is the magnetron frequency.

Clearly, the loading technique relies on increasing the time available to damp the axial motion by temporarily trapping the positrons via the magnetron rotation. Since the positrons are trapped for at least one magnetron period, the axial damping is improved by a factor of ν_z/ν_m ($\geq$2000) over on-axis injection. As a result, the trapping rate obtained was $\approx$23 positrons/hr with a 0.5 mCi source, a 5 V deep trap, and a 51 kG magnetic field. Since the positron loading rate is very sensitive to collisions with the background gas during the energy damping phase of the trapping process, the apparatus is usually immersed in liquid helium to obtain pressures less than 10^{-11} Torr within the trapping volume.

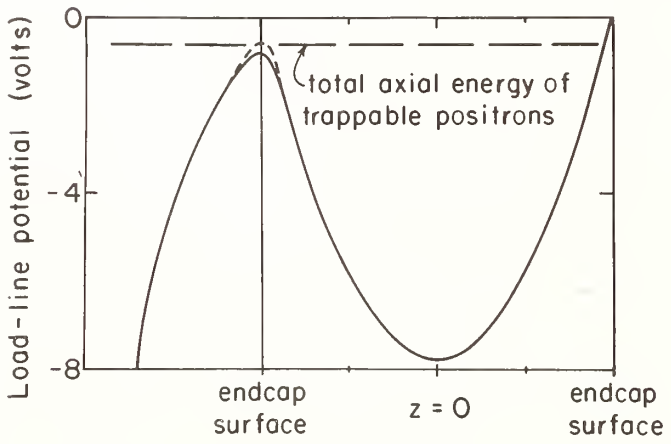


FIGURE 2. *Potential along the load-line. The source is biased at −300 V relative to the endcap in order to produce this potential variation along the line parallel to the z-axis that passes through the hole in the source endcap. The broken line greatly exaggerates the slightly higher potential (~1 mV) seen by a positron on the first return to the entry hole due to the magnetron motion.*

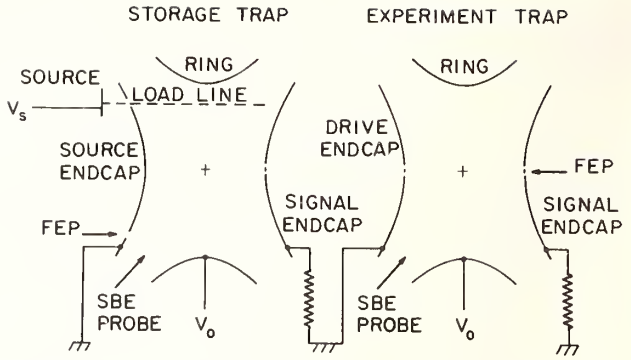


FIGURE 3. *Schematic of double trap configuration. The sideband excitation (SBE) probes are used in the radial centering process for off-axis loaded positrons and position stabilizing in the experiment trap. Signal endcaps are tuned to the axial frequency via an external inductor in order to observe the axial motion driven by an rf signal applied to the opposite endcap. The dc potential, V_0, is 8.3 V and 10.3 V respectively for the storage and experiment traps.*

3. Experiment Trap and the Transfer Process

A second trap has been carefully designed and compensated to make precision measurements possible. This trap is similar in detail to the loading trap (but without a positron source and with field emission point now located on-axis). It is virtually identical to past precision compensated Penning traps [1]. The double trap combination, shown only schematically in Fig. 3, was constructed because it was initially thought (and later demonstrated) that the holes required for positron trapping and centering would make it nearly impossible to compensate the storage trap enough to allow precision measurements to be made. By using this second trap, it then became necessary to transfer the positrons from storage into the second trap through a channel drilled between the two adjacent endcap electrodes of the double trap combination.

The transfer is accomplished by pulsing the two adjacent endcaps to the approximately common ring potential for a few microseconds. Pulses much longer than 10 μs tend to be less efficient since radial drifting can occur during the passage between traps (for instance via collisions with background gas). An accumulated drift more than 0.025 cm will cause the positron to hit the drive endcap of the experiment trap (see Fig. 3) since the hole in that electrode will be the smallest one encountered. Although the transfer is 25 to 50 percent efficient now, the efficiency of early attempts was greatly reduced by misalignment of the electric and magnetic axes in addition to higher than expected background pressure in the drift region between traps. The alignment of electric and magnetic axes is extremely critical since as little as one degree of misalignment can totally prevent any positrons from successfully transferring into the experiment trap.

Once the positrons have been transferred, they can be detected by using a large off-resonance axial drive, similar to that which is normally used when electrons are loaded into the trap, one at a time. This drive has been precalibrated in order to determine the number of positrons transferred, though not immediately after transfer since the positrons do not enter exactly along the axis of the experiment trap. If time is taken to apply sideband cooling at $\nu_z + \nu_m$ on the SBE probe, then, the calibra-

tion becomes reliable. After determining the number trapped, the excess beyond one are systematically ejected using intense rf pulses at $\nu_z + \nu_m$ again applied to the SBE probe. Figure 4 shows an example of this process for a case where four positrons were transferred into the trap and three are ejected. The rf amplitude in the pulse is carefully adjusted such that at least ten consecutive pulses are required in order to eject one positron from the cloud (i.e., less than 10% chance of driving out one per pulse).

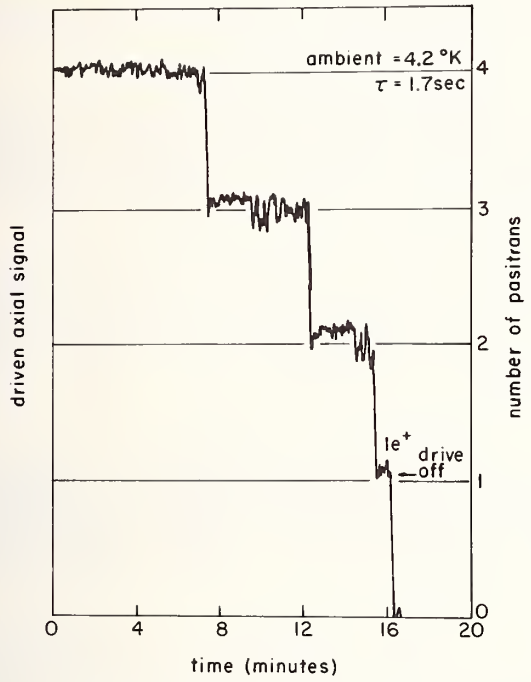


FIGURE 4. *Positron ejection record. An intense rf pulse at $\nu_z + \nu_m$ is applied to the SBE probe and the continuously monitored driven axial signal registers the loss of single positrons after numerous (> 10) consecutive pulses.*

4. Measurements

The signal shown in Fig. 4 is obtained using a preamplifier attached to one endcap and tuned to the axial frequency at $\nu_z \sim 60$ MHz in order to detect the image currents due to the positron's driven motion. Once a single positron is isolated, the much reduced drive signal, applied to the endcap opposite the signal endcap, is swept in frequency in order to obtain the axial resonance shown in Fig. 5. The compensation of this trap was quite effective with the result that resonances as narrow as 4 Hz can easily be resolved (resolution approaches 0.01 ppm).

Not shown in Fig. 3 are four symmetrically placed nickel screws, located in the central plane of the ring electrode within the experiment trap. This magnetic material will produce a very weak magnetic bottle which effectively couples the magnetic moment of the charge to the axial resonance in the form of a 1.3 Hz shift per unit change in the magnetic quantum level via the axial Stern-Gerlach effect [1]. By locking this axial frequency to a frequency synthesizer, the changes in the magnetic state are reflected in the lock-loop correction voltage, making it possible to measure both the cyclotron frequency, ν_c', and the spin anomaly frequency, ν_a'.

Figure 6 shows a typical positron cyclotron resonance using this coupling scheme. As typical of past electron

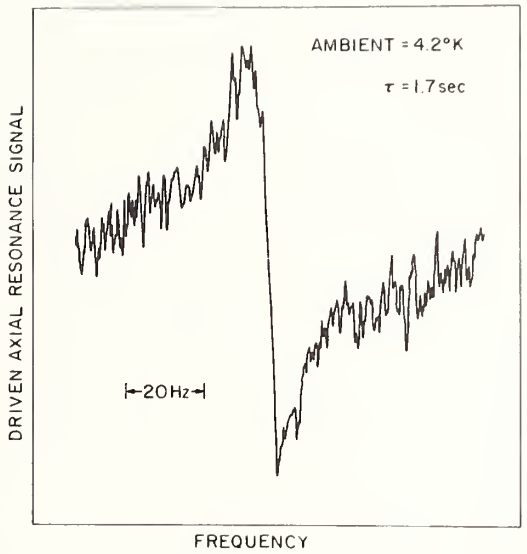


FIGURE 5. *Driven axial resonance of a single positron. The phase of the driven oscillator is carefully adjusted to yield this dispersion-type driven resonance as a function of frequency. The half-line-width (~2 Hz) represents a resolution that exceeds 0.03 ppm.*

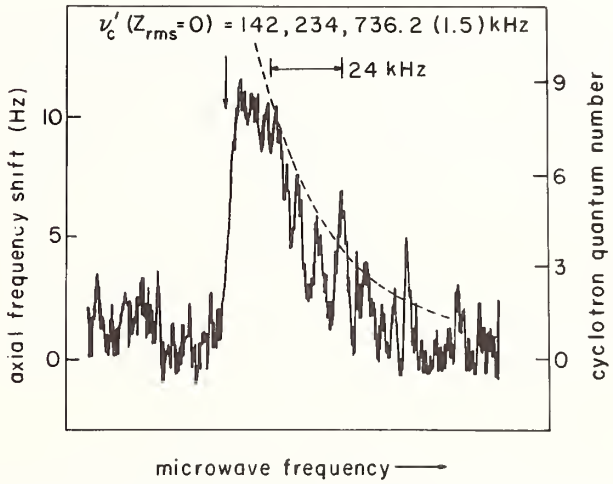


FIGURE 6. *Positron cyclotron resonance. The resonance is observed in the correction signals of the axial lock loop. The high frequency tail has an exponential decay linewidth of 24 kHz (see dotted curve) which is significantly larger than expected. However, the nearly vertical low frequency edge (z = 0) is resolvable to ≤1.5 kHz and is independent of linewidth.*

traps with magnetic bottles [1], the resonance shows a clear low frequency edge, corresponding to $Z_{rms} = 0$, and a high frequency exponential tail, indicative of the Boltzmann distribution of axial states. The dotted curve represents an exponential decay with a 24 kHz linewidth, which is four times greater than the narrowest electron cyclotron resonance obtained in previous traps [1, 8]. Nevertheless, the edge is resolvable to approximately 0.01 ppm and is independent of the resonance linewidth.

In order to flip the spin, the magnetic bottle is used again, but now to create a precessing magnetic field in the radial plane at the spin frequency

$$\nu_s = \nu_c' + \nu_a' .$$

This process has been described previously [1] and involves applying the anomaly drive to the endcaps yield-

ing a weakly driven axial motion at ν_a'. From the positron's frame of reference rotating at the cyclotron frequency ν_c', the axial modulation will yield sidebands with ν_s as the desirable upper sideband. Figure 7 shows the corresponding anomaly resonance taken by alternating the detection and excitation in order to yield the best resolution. The solid line represents an exponential decay with a 12.5 Hz linewidth. Again, the edge is nearly resolvable to 0.01 ppm.

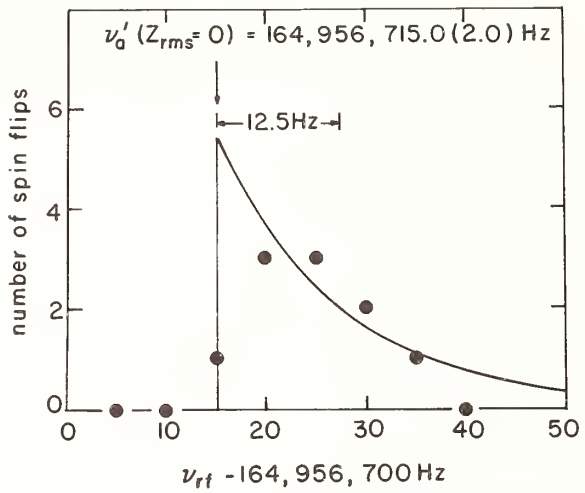


FIGURE 7. *Positron anomaly resonance. By alternating detection and excitation, the number of spin flips (out of a fixed number of attempts) is plotted versus frequency. The solid curve represents an exponential decay linewidth of 12.5 Hz. However, the low frequency edge (z = 0) is resolvable to ±2.0 Hz, which corresponds essentially to the same resolution found for the cyclotron resonance.*

5. Results

By measuring corresponding edge frequencies, the g-factor and its anomaly can be determined from the following relation:

$$a(e^{\pm}) = g(e^{\pm})/2 - 1 = [\nu_a' - \nu_z^2/2\nu_c']/[\nu_c' + \nu_z^2/2\nu_c'] .$$

This equation has been investigated and found to be highly accurate as long as ν_a', ν_c', and ν_z are the actual observed frequencies of motion in the Penning trap [9]. At present, four separate runs have been made using a single positron transferred into the highly compensated Penning trap. Time-studies of the cyclotron resonances indicate an effective magnetic field jitter on the order of 0.05 ppm. This may be due to collisions with some residual background gas which causes the radial position of the particle to vary in time within the magnetic bottle field. Future work will investigate and seal any possible leaks found in the trap-tube envelope.

At present, only one systematic effect in the anomaly (for both electrons and positrons) has been observed: the anomaly decreases with increasing applied anomaly drive. This may be due either to a change in the magnetron radius (change of 2–3 times the minimum radius observed) or to an uncorrected positive shift in the axial frequency associated with the added rf trapping potential produced by the strong anomaly drive. The effect was typically −0.03 ppm for the largest anomaly drive used in the older electron geonium experiments. Thus, from the four initial positron runs, a preliminary positron spin anomaly,

$$a(e^+) = 1159652222(\pm 50)\times 10^{-12} ,$$

is obtained from a weighted least-squares extrapolation using the systematic power dependence observed for the electron. Field jitter is incorporated in the individual anomaly errors, used to produce the weights in the least-squares adjustment. However, the error (0.043 ppm) represents primarily the uncertainty associated with the power extrapolation for the positron and only secondarily the error associated with the least-squares adjustment.

The result for the magnetic moment spin anomaly of a single electron [10] was reported earlier to be

$$a(e^-) = 1159652200(40) \times 10^{-12}$$

where the uncertainty was due to both a time varying magnetic field and the systematic dependence of $a(e^-)$ on the amount of spin power applied. These two results can be compared to theory [11]

$$a_e(\text{theor}) = 1159652569(150) \times 10^{-12}$$

where the uncertainty is almost entirely due to the uncertainty in the fine structure constant [12]. Finally, combining the results for the measured electron and positron anomalies we obtain a preliminary matter/antimatter comparison of

$$g(e^+)/g(e^-) = 1 + (22 \pm 64) \times 10^{-12} .$$

The resolution of this g-factor comparison will significantly improve when measurements of both positron and electron spin anomalies in the same apparatus are completed, without the pressure effects smearing out the magnetic field edge.

We wish to thank Don Russell for his excellent machine work on the traps and glassblower Bob Morley for his superb work.

References

[1] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, in *New Frontiers in High Energy Physics—Studies in the Natural Sciences* (Orbis Scientiae), A. Perlmutter and L. F. Scott, Eds. (Plenum Press, New York, 1981), p. 337.
[2] J. R. Gilleland and A. Rich, Phys. Rev. A 5, 38 (1972).
[3] S. I. Serednyakov, V. A. Sidorov, A. N. Skrinsky, G. M. Tumaikin, and Ju. M. Shatunov, Phys. Lett **66B**, 102 (1977).
[4] A positron polarimeter with three orders of magnitude more efficiency is reported by G. Gerber, D. Newman, A. Rich, and E. Sweetman, in Phys. Rev. D 15, 1189 (1977).
[5] J. H. Field, E. Picasso, and F. Combley, Sov. Phys. Usp. 22, 199 (1979).
[6] H. G. Dehmelt, P. B. Schwinberg, and R. S. Van Dyck, Jr., Int. J. Mass Spectrom. Ion Phys. 26, 107 (1978).
[7] P. B. Schwinberg, Ph.D. Thesis, University of Washington (1979); also P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt, Phys. Lett. 81A, 119 (1981).
[8] R. S. Van Dyck, Jr., P. B. Schwinberg, and S. H. Bailey, in *Atomic Masses and Fundamental Constants 6*, J. A. Nolen, Jr., and W. Benenson, Eds. (Plenum Press, New York, 1980), p. 173.
[9] G. Gabrielse and R. S. Van Dyck, Jr., Bull. Am. Phys. Soc. 26, 598 (1981); also Lowell Brown, private communication.
[10] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Bull. Am. Phys. Soc. 24, 758 (1979).
[11] Based on the best theoretical value of a_e to order α^3: $a_e(\text{theor}) = \alpha/2\pi - 0.328478445(\alpha/\pi)^2 + 1.1835(61)(\alpha/\pi)^3$ (T. Kinoshita, 1979, private communication).
[12] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. 42, 1575 (1979) [α^{-1} = 137.035963(15)].

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Geonium Without a Magnetic Bottle—A New Generation*

Gerald Gabrielse and Hans Dehmelt**

University of Washington, Department of Physics, FM-15, Seattle, WA 98195

Work is now underway to improve the accuracy of the electron magnetic moment measurements by 10 to 100; that is, to an accuracy of 1 part per trillion. At this level of accuracy, the magnetic bottle so crucial to previous measurements must be removed. In fact, intrinsic bottles due to the paramagnetism of the trap electrodes must be carefully canceled. A promising replacement for the magnetic bottle is the relativistic mass increase which acts like a magnetic bottle 20 times smaller than the one previously used but with no distortion of the magnetic field. The relativisitic mass increase has been observed for 0.5 eV electrons. We have also trapped electrons in a new style, compensated Penning trap (described here) which is simpler to construct, promises to be more reliable when cycled to liquid helium temperatures, and can be easily disassembled.

Key words: compensated Penning trap; geonium; intrinsic magnetic bottle; magnetic bottle; Penning trap; relativistic mass increase.

1. Introduction

The advent of single particle spectroscopy at the University of Washington began a string of very accurate measurements of properties of elementary particles. The magnetic moment anomaly of the electron has been measured to an accuracy of 34 parts per billion [1]. Elsewhere in these proceedings are reports of a comparison of the electron and positron magnetic moment anomaly to an accuracy of 64 parts per billion [2], and a measurement of the ratio of the proton to electron mass to 136 parts per billion with further increases in accuracy expected soon [3].

We judge there is a reasonable prospect of improving the measurements of electron and positron magnetic moments by a factor of 10 to 100. This paper is a progress report on the efforts in the past year and one half to realize this improvement. As in nuclear magnetic resonance spectroscopy (where resolutions of 10^{-10} have been achieved), we expect that geonium spectroscopy will have similar wide application if we can approach the NMR resolution. There is also considerable theoretical interest in the magnetic moments of the electron and positron. Calculations of the $(\alpha/\pi)^4$ corrections according to QED are presently underway (see these proceedings). If an independent, more accurate value for the fine structure constant also becomes available (some proposals are contained in these proceedings), then our measurement will provide a test of the QED calculations. Conversely, if one chooses to believe the calculation and the appropriateness of QED, then our measurement will provide the fine structure constant. To our knowledge, no other techniques can get within orders of magnitude of the accuracies we believe are realizable.

The new feature of the next generation geonium experiments is that we hope to rely on the relativistic mass shift rather than using the magnetic bottle which was crucial to previous experiments. The magnetic bottle is undesirable at this level of accuracy because it makes the magnetic field inhomogeneous.

*Supported by the National Science Foundation
**Chaim Weizmann Postdoctoral Fellow

We divide our program into three phases: the first is completed, the second is mostly completed, and the third is about to begin. First, we have demonstrated that we can observe the relativistic mass increase of an electron of energy 0.5 eV [4]. Second, we have constructed a simpler version of the compensated Penning trap which is demountable and promises to be more reliable as well. We have trapped electrons in it (24 hours before leaving for this conference) but much testing remains to be done including a demonstration that intrinsic magnetic bottles have been removed [5]. Third, we shall measure the electron's magnetic moment anomaly.

2. Observation of the Increase in the Relativistic Mass, γm_0, of a 0.5 eV Electron ($\gamma = 1.000\ 001$)

A slightly modified, compensated Penning trap made of OFHC copper was assembled using electron tube construction techniques which have been discussed elsewhere [6]. For our purposes we need not review the monoelectron oscillator formed except to observe that the electron's harmonic oscillation along the magnetic field lines has a frequency ν_z ($\simeq$60 MHz) which is inversely proportional to the square root of the electron mass. The observed cyclotron oscillation around the magnetic field lines has frequency ν'_c ($\simeq$21 GHz) which is inversely proportional to the electron mass.

The energy in the cyclotron motion is increased by sweeping a microwave drive through the cyclotron frequency. Both ν_z and ν'_c thus decrease because the electron's mass, γm_o, (where m_o is the rest mass) increases. The harmonic and cyclotron oscillation frequencies decrease by 60 Hz and 42 kHz, respectively, for a 1 eV excitation.

A typical measurement is summarized in Fig. 1. We monitored changes in the harmonic oscillation frequency (the ordinates of Fig. 1). The sensitivity was calibrated as in Fig. 1(a) by changing by 50 Hz the frequency of the synthesizer to which the electron's harmonic oscillation frequency was locked. This was done repeatedly to get an indication of uncertainty.

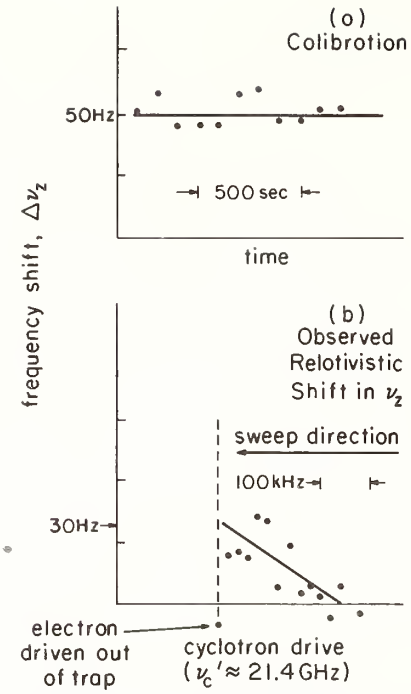


FIGURE 1. *(a) Calibration of sensitivity to changes in ν_z; (b) observation of relativistic mass increase of 0.5 eV electron.*

Sensitivity was never better than 20 Hz out of 60 MHz with 400 second signal averaging. This sensitivity was rather poor compared to the 1 Hz resolution reported for earlier geonium measurements with a one second time constant [7]. However, we were awaiting delivery of a superconducting NMR magnet with a liquid helium Dewar in its bore and thus were forced to use an old Varian electromagnet and immerse the Penning trap in liquid nitrogen rather than liquid helium. The new apparatus (discussed in the next section) will doubtlessly have much better sensitivity.

A typical mass shift measurement is plotted in Fig. 1(b). The ordinate is identical to Fig. 1(a). The cyclotron drive was swept down through resonance (right to left). Like the anharmonic oscillator, the cyclotron frequency is pulled along as the excitation proceeds. The strength of the microwave drive is very critical. A slightly stronger drive knocks the electron out of the trap, presumably via collisions.

A further complication accounts for the restricted frequency range and the abrupt loss of the electron at the left of Fig. 1(b). The cyclotron motion of an electron in a Penning trap is actually superimposed on a slow drift orbit (magnetron motion) of frequency $\nu_m \simeq 90$ kHz for our trap. As has been observed elsewhere [7], the electron can also absorb energy at $\nu_c' + \nu_m$ or $\nu_c' - \nu_m$, the former shrinking the magnetron radius, the latter increasing it. We thus sweep down in frequency. When we hit the $\nu_c' - \nu_m$ sideband our strong drive pushes the electron out of the trap radially.

2.1 New Demountable Penning Trap

We have successfully trapped electrons in a new style, compensated Penning trap which differs substantially from earlier traps. Previous Penning traps [6, 7] were constructed like glass vacuum tubes. A change in the trap, however slight, required that the whole trap be disassembled, reassembled on a new pin base, sealed by a glass blower and baked to 400 °C. When cycled between room temperature and liquid helium temperature, the glass-to-metal seals proved to be unreliable vacuum seals. An internal ion pump was thus added to partially compensate. Finally, each trap consisted of five separately machined metal electrodes separated by four tubular glass spacers each of which was precision ground. Much precision mill work was required so that construction and assembly times were measured in man-years.

We felt that the rigors of a parts per billion measurement would be greatly facilitated by a Penning trap which was simpler, more rugged, and easily accessible for modification. The glass vacuum tube technology was accordingly abandoned in favor of a tubular copper envelope, sealed to copper flanges with demountable indium seals. We dispensed with the ion pump entirely since it seems clear that cryopumping alone will easily keep the vacuum inside the trap to better than 10^{-15} Torr [8]. Porous glass cryopump cells were mounted on the upper copper flange to greatly increase the cryopumping surface area. The whole system was baked briefly to 100 °C to drive off water vapor from the electron surfaces and then a copper pump-out tube was pinched off at a vacuum of 5×10^{-8} Torr.

The trap itself is mounted on a copper pin base flange as shown in Fig. 2. Alignment is provided by a ring of tungsten pins brazed to the copper base. These pins align

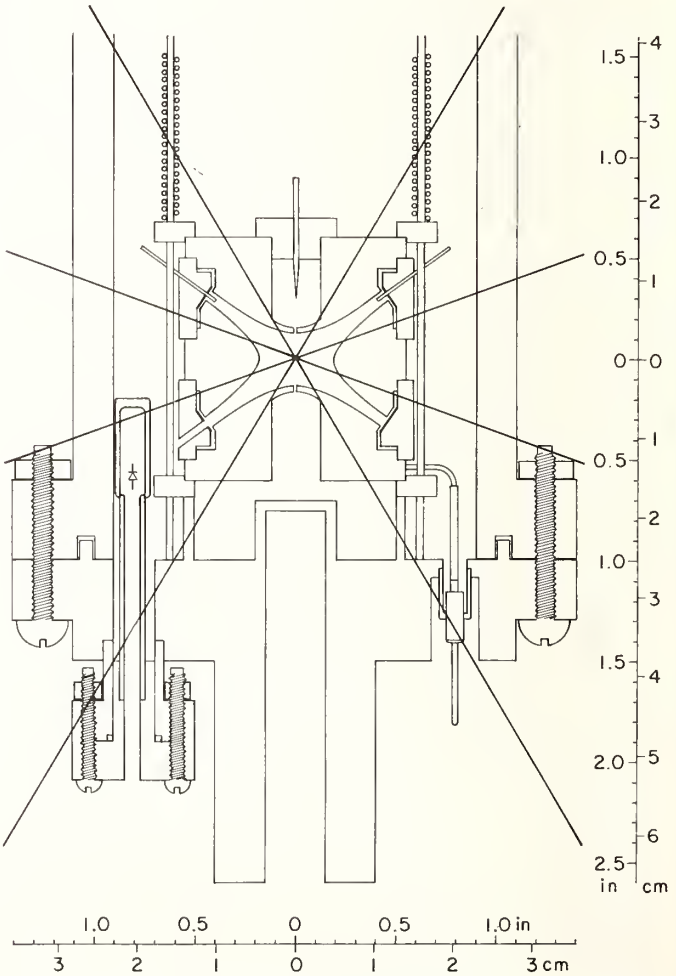


FIGURE 2. *Demountable Penning trap; $\theta = 30°$ and $\theta = 70°$ superimposed on Penning trap outline to indicate placement of magnetic compensation material.*

the trap elements by way of thin machinable glass ceramic (MACOR) rings on the outside of the endcaps and by springs. The two endcaps and the ring are made of molybdenum in this first version. These electrodes are simpler to construct than earlier traps in that no precision mill work is needed. The four ground glass spacers and the two metal compensation rings of previous traps have all been replaced by two spacers constructed from MACOR. Anharmonicity compensation is provided by a "thick-film" of silver metalized to the MACOR. Each thick film compensation ring is split to make a 1 turn coil which will be used to drive spin flip transitions (at $\simeq 180$ MHz). Holes through the spacers admit microwave drives and small probes are used to apply radio-frequency driving signals for cooling. MACOR can be machined using conventional metal working techniques and the thick film compensation electrodes can be added in several hours. Replacing six parts by two makes more precise alignment possible and also simplifies the machining.

Two demountable ports, also vacuum sealed by indium O-rings, are provided for admitting microwaves into the trap. In this first version microwaves will be generated by multiplying X-band up to 168 GHz in a multiplication diode inserted from outside the trap into a glass tube sealed to a copper stem. The electrical feed-throughs brazed into the copper base are commercially available. They are made from Constantan which is copper brazed to ceramic. The copper base has a hole on the trap axis which will allow inserting a ^{3}He NMR probe within 1.5 inches of the trap center. This probe will be used if necessary to stabilize the 60 kG magnetic field.

3. Bottleless Geonium

The small nickel ring, deliberately introduced into previous Penning traps, must be removed to improve our resolution of the anomaly beyond 10^{-8}. In addition, however, magnetic field inhomogeneities introduced by the paramagnetism of the trap electrodes must be removed as well [5]. A ring about the $\hat{z}$ axis with magnetism M and small cross sectional area A produces a magnetic field near the origin at (r, θ) and (ρ, z)

$$\mathbf{B} = \sum_0^\infty A_k\, r^k [P_k(\cos\theta)\hat{z} - (k+1)^{-1} P_{k+1}(\cos\theta)\hat{\rho}]$$

where P_{km} is an associated Legendre polynomial. In cgs units

$$A_k = 2\pi MA\,(k+1)(k+2)\rho\, r^{-k-3} P_{k+2}(\cos\theta)\,,$$

where now the coordinates locate the ring. A symmetrical Penning trap like this one, with $\rho_o = \sqrt{2}\, z_o$, thus has a magnetic bottle with $A_2 \simeq 30\ M$. For a molybdenum trap at 4.2 K and 60 kG this suggests an intrinsic bottle about 20% of that previously introduced deliberately.

The next trap will no doubt be made of copper as a result (i.e., ten times less susceptibility). In the meantime, compensation rings made of iron will be placed between $\theta = 31°$ and $\theta = 70°$ where P_4 is negative to severely reduce the size of the intrinsic bottle. Figure 2 indicates these angles superimposed on the trap surfaces. Final cancellation will be achieved using the second order shims of the superconducting magnet.

4. g − 2 in Cooled Relativistic Geonium

We have considered many schemes for measuring g − 2 in bottleless geonium. As might be expected for an attempted measurement accurate to better than a part per billion, none of these promises to be easy. In the next few months the new trap will be used to evaluate the feasibility of these schemes.

The most promising approaches make use of the relativistic perturbation to the cyclotron levels [9, 10], as diagrammed in Fig. 3. The quantum numbers n and m pertain to cyclotron and spin levels. At 4.2 K the electron spends roughly 90% of its time in $n = 0$ and 10% in $n = 1$. The g − 2 transition, between $(n = 1,\ m = -1/2)$ and $(n = 0,\ m = 1/2)$, will be driven via single-turn bucking coils made from the compensation electrodes. We hope to determine if $m = -1/2$ or $m = 1/2$ and thus whether or not the spin has flipped by ascertaining whether the first cyclotron absorption occurs at $(1 - R)\nu_c$ or $(1 - 3R)\nu_c$. As a first attempt the cyclotron drive will be swept towards lower frequencies, starting between $(1 - R)\nu_c$ and $(1 - 3R)\nu_c$. This starting point must of course be determined experimentally. The successful observation of the relativistic mass increase (Section 1) suggests that there will be no major problem in driving the electron's cyclotron energy high enough to observe it via the relativistic shift of the harmonic frequency, once the classical region of $n = 10$ or higher has been reached. Obtaining substantial excitation probability to $n = 10$ promises to be more difficult since the 0.1 sec lifetime will make it difficult initially to maintain both the excitation and the frequency resolution. Experiments to establish a feasible detection scheme begin as soon as this conference ends.

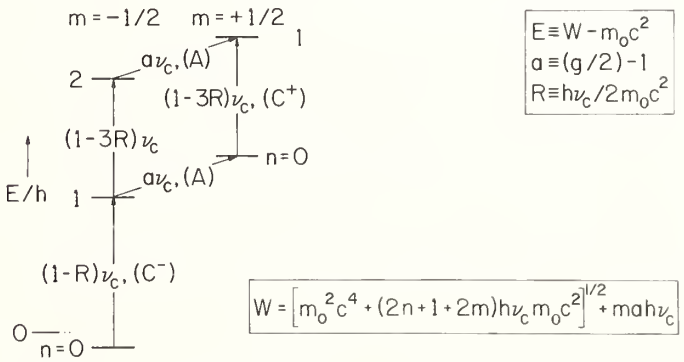


FIGURE 3. *Lowest cyclotron and spin energy levels for geonium.*

The new style trap was machined by Ralph Jochim. Technical assistance was provided by Kim Weimer.

References

[1] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Bull. Am. Phys. Soc. 24, 758 (1979).
[2] P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt, these proceedings.
[3] R. S. Van Dyck, Jr. and P. B. Schwinberg, these proceedings.
[4] G. Gabrielse and H. G. Dehmelt, Bull. Am. Phys. Soc. 25, 1149 (1980).
[5] G. Gabrielse and H. G. Dehmelt, Bull. Am. Phys. Soc. 26, 598 (1981).
[6] R. S. Van Dyck, Jr., D. Wineland, P. Ekstrom, and H. G. Dehmelt, Appl. Phys. Lett. 28, 446.
[7] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. C. Dehmelt, in *New Frontiers in High Energy Physics—Studies in the Natural Sciences* (Orbis Scientiae), Ed. by A. Perlmutter and L. F. Scott (Plenum Press, 1978), p. 159.
[8] W. Thompson and S. Hanrahan, J. Vac. Sci. Technol. 14, 643 (1977).
[9] H. Dehmelt, in *Atomic Physics 7*, Ed. by D. Kleppner and F. M. Pipkin (Plenum Press, N.Y., 1981), p. 337.
[10] H. Dehmelt and G. Gabrielse, Bull. Am. Phys. Soc. 26, 797 (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Magnetic Moment of Positive Muons in Units of the Proton Magnetic Moment*

E. Klempt, R. Schulze, and H. Wolf

Institut für Physik der Universität Mainz, F.R.G.

and

M. Camani, F. N. Gygax, W. Rüegg, A. Schenck, and H. Schilling

Laboratorium für Hochenergiephysik der ETH, Zürich, c/o SIN, 5432 Villigen, Switzerland

The magnetic moment of positive muons in units of the proton magnetic moment was determined. A value of $\mu_\mu/\mu_p = 3.1833441(17)$ was found.

Key words: diamagnetic shielding; muon magnetic moment; muon chemistry.

1. Introduction

We report on a measurement of the ratio of muon to proton magnetic moments, μ_μ/μ_p. A preliminary analysis of part of our data has been published in a previous letter [1], and a full account of the work has been submitted for publication [2]. The value of μ_μ/μ_p was derived from the relation shown in Table 1.

TABLE 1. *Basic relation from which μ_μ/μ_p is determined.*

$$\frac{\mu_\mu}{\mu_p} = \frac{\omega_\mu}{\omega_p}\,\frac{1+\sigma_\mu}{1+\sigma_p+\sigma_{BG}}$$

ω_μ	muon spin precession frequency
ω_p	proton spin precession frequency
σ_μ	diamagnetic shielding of muons
σ_p	diamagnetic shielding of protons
σ_{BG}	background correction

The muon spin rotation frequency ω_μ was measured in a magnetic field of 0.747 T at the SIN superconducting muon channel using a stroboscopic technique. Muons from the beam with polarization perpendicular to the magnetic field were stopped in a liquid Br_2 target. Their chemical state depends on the presence or absence of a small H_2O contamination in the Br_2 target: with no H_2O present muons form (μ^+e^-) Br molecules, otherwise a chemical reaction chain places the muons into MuHO molecules [1, 2]. The two diamagnetic shieldings σ_μ were determined from NMR measurements on HBr and H_2O molecules dissolved in liquid Br_2, and from quantum chemical calculations of the muon-proton isotope shifts [3]. The spin precession frequency of protons in H_2O in the same magnetic field was determined by conventional methods [4], the proton diamagnetic shielding in H_2O is known [5, 6].

Figure 1 sketches the principle of the stroboscopic method. It is based on the microstructure of the SIN muon beam. Due to the rf acceleration of the primary proton beam, the muon beam arrives in sequential beam bursts with a separation in time corresponding to the acceleration frequency $\Omega = 2\pi \times 50.633$ MHz. A magnetic field of 0.747 T leads to a muon spin rotation frequency

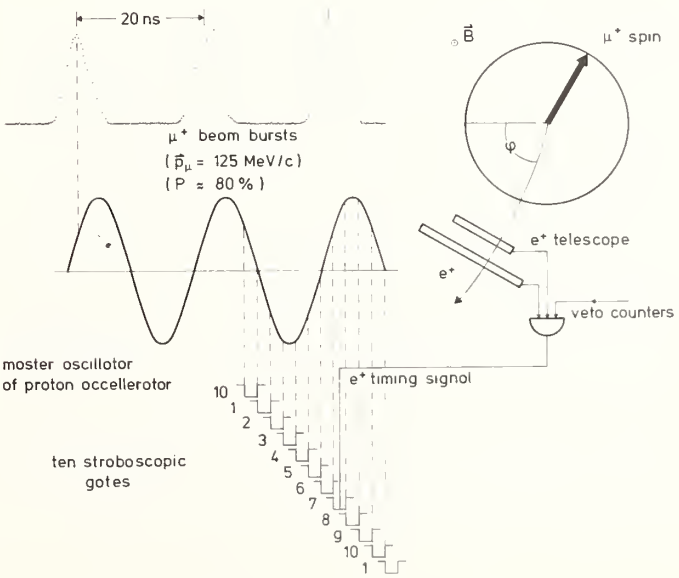


FIGURE 1. *Schematic representation of the stroboscopic method.*

*Supported by the Bundesministerium für Forschung und Technologie, FRG, and by the Schweizerisches Institut für Nuklearforschung, SIN.

of $\omega_\mu = 2\Omega$. At this field the polarization of muons from different beam bursts add up coherently. The polarization is small when muon spin precession frequency and beam burst repetition frequency are mismatched, i.e., when $|\omega_\mu - 2\Omega| \geq 1/\tau_\mu$. The polarization can be observed by gating—at the beam burst repetition frequency—the telescope for detection and identification of decay positrons. The gate corresponds to a fixed frequency "flash light" which detects the polarization in a rotating coordinate frame. The magnetic field has to be scanned in order to find a stationary polarization. Ten contiguous and nonoverlapping stroboscopic gates were set up covering one complete cycle of the rf period.

A line shape theory was developed which calculates the fractional number of positron counts, N_j/N_o, falling into a stroboscopic gate of length λ_j as a function of the magnetic field. For a periodic muon beam intensity and electronic gates of the same frequency, N_j is given by

$$N_j = N_o \lambda_j \{1 + A_{eff} \times \frac{\sin 2\pi\lambda_j}{2\pi\lambda_j} \frac{\cos \psi_j - x \sin \psi_j}{1 + x^2}\}, \quad (1)$$

with $x = (\omega_m - 2\Omega)\tau_\mu$, ψ_j as arbitrary phase, and A_{eff} and effective asymmetry of the muon decay. The finite width of the muon beam bursts results in an effective asymmetry of 0.16 compared to an asymmetry of 0.24 found by conventional muon spin resonance techniques at the same beam line.

2. Apparatus and Methodology

The experimental setup basically consists of a target positioned at the center of an iron magnet, a collimator, and scintillation counters. The diameter of the pole pieces is 30 cm with a separation distance of 8 cm. The target is surrounded by a box of plastic scintillators which serves to identify positrons from muons stopped in the target and to veto muons and unwanted positrons. The count rate of accepted positrons was 27000/s for a primary proton beam of 40 μA and a muon beam intensity of 2×10^6 μ^+ through the collimator.

The magnetic field was stabilized by a conventional NMR system. Maps of the magnetic field over the target volume of 14 cm^3 were taken regularly by a second NMR system. Figure 2 shows a histogram of eight different field maps measured in the course of six days and corrected for drifts of the magnetic field at the target center (typically 5×10^{-8}). The rms deviation of magnetic field values from the mean is 0.4 ppm. Different hypothetical μ^+ stop distributions lead to the same mean value of the magnetic field with an uncertainty of 0.08 ppm. At the end points of the magnetic field scan the homogeneity was typically 0.6 ppm. The homogeneity was achieved by a 10 current shim system, the stability of the field map by thermostating the magnet cooling water and the ambient air. The NMR measuring probe was calibrated by comparison with a series of unshielded calibration probes mounted inside an open pair of Helmholtz coils modulating the external field. The final corrections to the NMR frequencies are listed in Table 2.

Figure 3 shows the stroboscopic signal from 2 out of 10 stroboscopic gates and for a data taking period of 2 days. The data of all 10 gates were fit simultaneously by use of Eq. (1), but some small modifications for systematic effects were incorporated into the line shape theory:

(i) From incoming muons an electronic veto is derived. Positrons emitted within some 10 ns after a μ^+ stop are therefore not accepted. In 10 ns the muon polarization vector has precessed by an angle which depends on the magnetic field. Hence the phase ψ_j is different for different values of the magnetic field. This effect was implemented into the line shape theory. Test analysis routines proved that the effect is controlled and had no effect on ω_μ/ω_p. No free parameter was finally needed.

(ii) As muons arrive in bursts, there is a periodic chance that otherwise accepted positrons were vetoed. This effect was allowed for in the line shape theory and is described by five complex Fourier coefficients.

(iii) Positrons from muon decay have—due to the magnetic field—curved tracks. This leads to a magnetic field dependent low energy cut in the energy spectrum of accepted positrons. Thus the effective asymmetry A_{eff} is field dependent. This effect was shown to be negligibly small.

(iv) The curvature of positron tracks leads to a magnetic field dependent solid angle for positron detection. Such a field dependence of ψ_j was already introduced to describe the effect of prompt vetoes after μ^+ stops. The track curvature yields a much smaller effect and was therefore neglected.

(v) Phase jitter of beam bursts and stroboscopic gates

TABLE 2. *Corrections to the NMR frequency of the field measuring probe.*

Diamagnetic shielding, σ_{H_2O}	+25.64(3) ppm
Bulk susceptibility	− 1.51(4) ppm
Paramagnetic ions	+ 0.35(5) ppm
Calibration	+ 0.80(12) ppm
Field map	0.00(8) ppm
Total correction, σ_{NMR}	+25.28(16) ppm

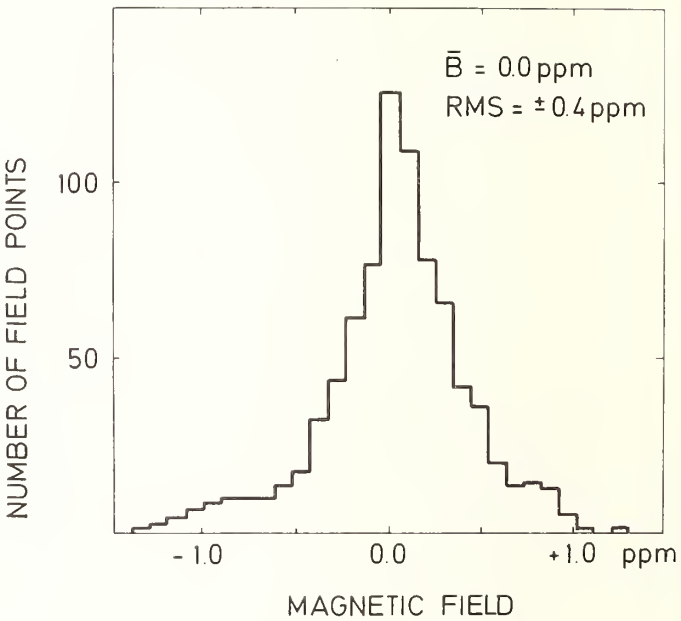


FIGURE 2. *Histogram of magnetic field values obtained from 8 field maps taken in the course of 6 days.*

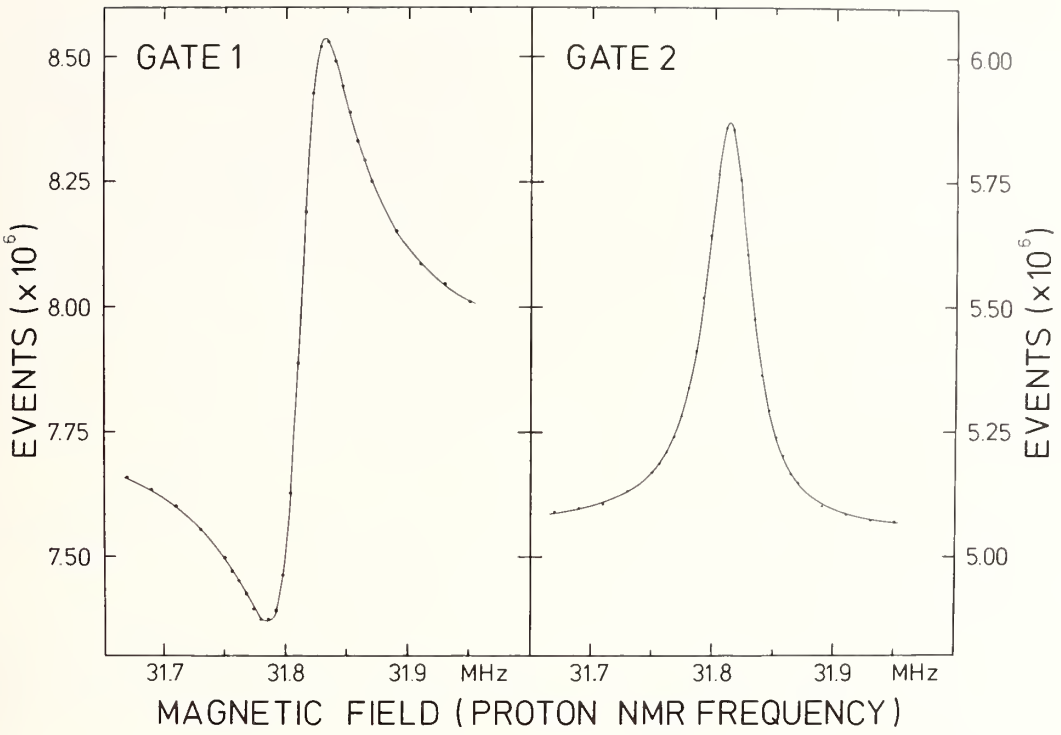


FIGURE 3. *Stroboscopic signal from 2 out of 10 gates. The dot radii correspond to the statistical error. The data were collected in 2 days.*

did not lead to systematic distortions of the line but to an increased variance of the number of positron counts in a stroboscopic gate. Therefore the errors from individual runs were scaled by $\sqrt{\chi^2/N_F}$, i.e., typically by 1.3.

(vi) The ratio of accepted positrons for target-full target-empty was 14. The target empty rate was dominantly attributed to muon stops in the quartz glass ampoule containing the liquid bromine, in the positron counters, or in the wrappings or surfaces of the scintillation counters. In target empty runs no stroboscopic background signal was firmly established. Therefore a background correction of $\sigma_{BG} = (-0.20 \pm 0.25)$ ppm was deduced from the largest stroboscopic background signal compatible with the background data.

For contiguous nonoverlapping gates, the 10 gate lengths λ_j and 10 phases ψ_j are linked by 10 constraints. Therefore a total of 22 parameters, ω_μ/ω_γ, A_{eff}, 9 gate lengths λ_j, 1 phase, and 5 complex Fourier coefficients were needed to describe 289 data points (N_o is fixed). From the fit, uncorrected values for ω_μ/ω_p were derived:

$$\begin{aligned} \omega_\mu/\omega_p &= 3.1833212\ (20) \quad (\mu^+e^-)\text{Br in } Br_2, \\ \omega_\mu/\omega_p &= 3.1833341\ (19) \quad (\mu^+e^-)\text{HO in } Br_2, \\ \omega_\mu/\omega_p &= 3.1833519\ (66) \quad (\mu^+e^-)\text{HO in } H_2O. \end{aligned} \tag{2}$$

The latter result was obtained by stopping muons in H_2O. It is only used for a comparison of chemical shifts, not for the final value of μ_μ/μ_p. Note that the error due to fluctuations of the magnetic field at the target center is contained in the above errors.

3. Muon Chemistry

The major advantages of using bromine as a target substance are the following:

(i) fast chemical reactions preserve the full muon polarization;

(ii) there is one unique chemical compound in which muons finally reside;

(iii) a small water contamination allows the final chemical compound to change;

(iv) the compounds allow a measurement of the chemical shift on analogous compounds containing protons;

(v) the isotopic shift of the diamagnetic shielding in these compounds can be derived from quantum chemical calculations; and

(vi) finally, liquid bromine has a high stopping power and is easy to handle at room temperature.

The discussion of the chemical reactions which μ^+ undergo when they are stopped in liquid bromine and of their diamagnetic shielding is summarized in Fig. 4. Positive muons stopped in liquid bromine neutralize to muonium which reacts with bromine molecules by forming (μ^+e^-)Br molecules, $(\mu^+e^-) + Br_2 \rightarrow (\mu^+e^-)\,Br + Br$, with an average time of 5×10^{-13} s. This time is very short compared to the muonium hyperfine structure period of 2×10^{-10} s. The polarization of the muon is therefore completely conserved. A possible H_2O contamination in a pure Br_2 target can be removed by adding P_2O_5 and by bubbling gaseous HBr through the liquid Br_2. The time constant for the reaction $(\mu^+e^-) + HBr \rightarrow (\mu^+e^-)H + Br$ is too long to lead to any sizable fraction of $(\mu^+e^-)H$ molecules.

One bromine target was contaminated with H_2O with a concentration of 0.1 mol/l. There the reaction chain

$$(\mu^+e^-)\,Br + H_2O \rightarrow (\mu^+e^-)\,H_2O^+ + Br^-,$$
$$(\mu^+e^-)\,H_2O^+ + H_2O \rightarrow (\mu^+e^-)\,HO + H_3O^+,$$

places the muon into a water molecule within one μs after thermalization.

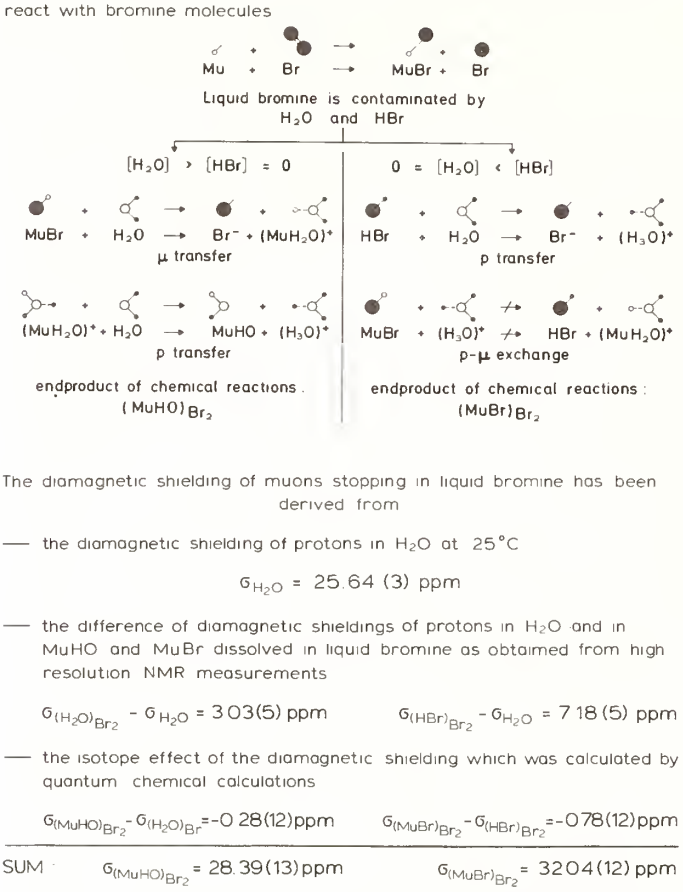


FIGURE 4. *The chemical reactions which μ^+ undergo when they are stopped in liquid bromine, and their diamagnetic shielding.*

The diamagnetic shielding of muons in (μ^+e^-)HO molecules in liquid Br_2 has been derived from the shieldings of HBr and H_2O molecules in liquid Br_2. The diamagnetic shieldings of free HBr and H_2O molecules are known; in liquid Br_2 they are subject to liquid association shifts of -1.92 ppm and -1.18 ppm, respectively. These were measured in a conventional high resolution NMR spectrometer. Liquid association shifts and the difference of the diamagnetic shielding when protons are replaced by muons in a chemical compound were derived from quantum chemical calculations [3, 4]. These isotope shifts of the diamagnetic shieldings—including the liquid association shifts—were found to be (-0.78 ± 0.12) ppm and (-0.28 ± 0.12) ppm. The calculation predicted liquid association shifts of HBr and H_2O in Br_2 of -1.93 ppm and -1.03 ppm, respectively, in striking agreement with the experimental values, and thus supporting the reliability of the calculation. The resulting diamagnetic shielding corrections for $\mu^{\mp}$ in (μ^+e^-)Br and (μ^+e^-)HO molecules and their contributions are listed in Table 3. An estimate given in Ref. [7] is assumed for the isotopic shift of the diamagnetic shielding of μ^+ stopped in liquid H_2O and forming (μ^+e^-)HO molecules.

4. Results

When the corrections listed in Table 1 are applied to the values of ω_μ/ω_p found in the fit the following results are obtained:

$$\begin{aligned} \mu_\mu/\mu_p &= 3.1833434\ (20)\ \text{MuBr in } Br_2, \\ \mu_\mu/\mu_p &= 3.1833447\ (19)\ \text{MuHO in } Br_2, \\ \mu_\mu/\mu_p &= 3.1833473\ (92)\ \text{MuHO in } H_2O. \end{aligned} \tag{3}$$

The three results are in excellent agreement after the corrections for the diamagnetic shieldings are applied. This supports the assumption that the chemical reactions derived from the chemistry of hydrogen atoms also take place in the case of muonium chemistry. The first two results are combined, and the systematic errors due to the calibration of the NMR frequency, due to the unknown muon stop distribution, and due to the stroboscopic background are added quadratically. Thus the final result obtained is:

$$\mu_\mu/\mu_p = 3.1833441(17) \quad (0.5\ \text{ppm}).$$

This value is consistent with previous determinations [7, 8] of μ_μ/μ_p and with the result of a recent high precision experiment on the muonium hyperfine structure in a strong magnetic field [9].

TABLE 3. *Isotopic changes of the chemical shift and total diamagnetic shielding corrections for μ^+ in a (μ^+e^-)Br or (μ^+e^-)HO molecule in liquid Br_2 (spherical probes), absolute and relative to protons in water.*

molecule	$\delta\sigma_{\text{isotope}} = \sigma_\mu - \sigma_p$ [ppm]		Shielding correction in the liquids	
+ environment	Free molecule	Molecule in liquid	$\sigma_\mu - \sigma_{H_2O}$ [ppm]	σ_μ [ppm]
$(MuBr)_{Br_2}$	$-0.70 \pm 15\%$	$-0.78 \pm 15\%$	$+6.40 \pm 0.12$	$+32.04 \pm 0.12$
$(MuHO)_{Br_2}$	$-0.46\ ^{+15\%}_{-110\%}$	$-0.19\ ^{+15\%}_{-110\%} = -0.28 \pm 0.12$	$+2.75 \pm 0.13$	$+28.39 \pm 0.13$
$(MuHO)_{H_2O}$	-2.0 ± 2.0[a]	-2.0 ± 2.0[a]	-2.0 ± 2.0[a]	$+23.64 \pm 2.00$[a]

[a]Values as assumed in Ref. [7] (only for comparison).

References

[1] M. Camani, F. N. Gygax, E. Klempt, W. Rüegg, A. Schenck, H. Schilling, R. Schulze, and H. Wolf, Phys. Lett. **77B**, 326 (1978).

[2] M. Camani, F. N. Gygax, E. Klempt, W. Rüegg, A. Schenck, H. Schilling, R. Schulze, and H. Wolf, Phys. Rev. D **25**, 652 (1982).

[3] M. Castro and J. Keller, "The Isotope Chemical Shift of μ^+ in HBr and H_2O," Report (UNAM, Facultad de Quimica, Mexico-City, 1980), and to be published.

[4] K. Borer, Nucl. Instrum. Methods **143**, 203 (1977).

[5] W. D. Phillips, W. E. Cook, and D. Kleppner, Metrologia **13**, 179 (1977).

[6] J. C. Hindman, J. Chem. Phys. **44**, 4582 (1966).

[7] K. M. Crowe, J. F. Hague, J. E. Rothberg, A. Schenck, D. L. Williams, R. W. Williams, and K. K. Young, Phys. Rev. D **5**, 2145 (1972).

[8] D. E. Casperson, T. W. Crane, A. B. Denison, P. O. Egan, V. W. Hughes, F. G. Mariam, H. Orth, H. W. Reist, P. A. Souder, R. D. Stambaugh, P. A. Thompson, and G. zu Putlitz, Phys. Rev. Lett. **38**, 956 (1977).

[9] V. W. Hughes, these proceedings.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Measured g_J-Factor Ratio of $^4He^+(1\,^2S_{1/2})$ and $^4He(2\,^3S_1)$*

H. G. Robinson**

Duke University, Durham, NC 27706

and

C. E. Johnson

North Carolina State University, Raleigh, NC 27607

The g_J-factor of the ground state hydrogen-like helium ion ($^4He^+$) relative to that of the $2\,^3S_1$ state of helium ($^4He^*$) has been measured using a ^{87}Rb optical pumping technique. The current experimental result is $g_J(^4He^+,\ 1\,^2S_{1/2})/g_J(^4He,\ 2\,^3S_1) = 1 - 29.95(10) \times 10^{-6}$. Using other previously determined experimental ratios, we find $g_J(^4He^+)/g(e^-) = 1 - 70.87(10) \times 10^{-6}$, in excellent agreement with theory. The linewidth budget for the $^4He^+$ Zeeman resonance shows the conspicuous absence of broadening due to charge exchange between the ion and ground state He even though the exchange rate is $>10^7$/s.

Key words: gyromagnetic ratio; helium; ion; metastable state; precision measurement.

1. Introduction

Precision measurements of the magnetic moment ratios of the simple systems, hydrogen, the free electron spin, and metastable $^4He^*(2\,^3S_1)$ have made possible exacting tests of the theory of the Zeeman effect for such species. Mass-dependent terms of the theory [1] for hydrogen-like systems have been checked to within an accuracy of 3 parts in 10^{11} through a measurement [2] of the ratio of g_J-factors of hydrogen and deuterium. Mass-independent terms were criticized to 1.3 parts in 10^8 by a measurement [3] of the ratio of g-factors of hydrogen and the free electron spin. The theory [4] for the two-electron systems, $^4He^*$, has been compared with an experiment [5] to a precision of 5 parts in 10^8. In this paper, we discuss results of a measurement [6] of the ratio of g-factors of $^4He^+$, and $^4He^*$. This permits a first precision determination of the g-factor of a simple atomic ion. It provides the potential for examination of the Z-dependence of the theory.

2. Experimental Technique

The experiments were performed in a 100 G, shielded, electrically shimmed, actively stabilized solenoid which has a long-term field drift permitting measurements to ~1 part in 10^9 and an unaveraged field homogeneity better than 1 part in 10^7 over a typical cell volume of 50 cm^3. Initial investigations toward achieving high resolution capabilities for the He^* Zeeman resonance used direct optical pumping of He (1.08 μm radiation). Although line widths of ~700 Hz (FWHM) were obtained, a major improvement in resolution was achieved by using ^{87}Rb optical pumping with a weak, pulsed, electrodeless rf discharge. A typical pulse width was 50 μs with a repetitive rate of 240 Hz. See Fig. 1 for a schematic diagram of the experimental apparatus. Both electron spin-exchange and spin-dependent destructive collisions between Rb and He^* contribute to the polarizing/sensing of the He^* resonance. Minimum linewidth for this system was 210 Hz in a cell filled with 14 Torr He.

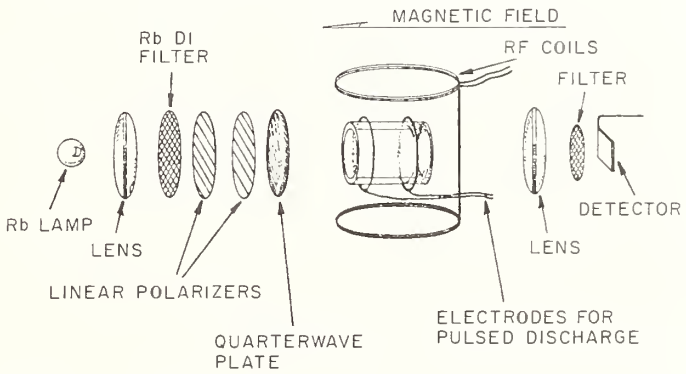


FIGURE 1. *Experimental apparatus. A typical ^{87}Rb optical pumping system is used together with a pulsed rf discharge to produce the $^4He^+$ and $^4He^*$ species. Zeeman resonances induced by the rf coils are detected as changes in the transmitted light intensity.*

Two effects conspire to degrade resolution when observing the $^4He^+$ Zeeman resonance: a lower cell pressure and a higher discharge intensity are required to produce and monitor the He^+ ion. The upper limit to usable cell pressure is determined by the three-body collision process

$$He^+ + 2He \rightarrow He_2^+ + He\,.$$

At a pressure $P = 2$ Torr, this reaction contributes ~85 Hz to the width of the He^+ Zeeman resonance. However, of even more critical importance is the effect this reaction has on the density of He^+. Calculation shows the equilibrium density ratio $[He^+]/[He_2^+]$ is proportional to P^{-3} with a ratio of 1:1 occurring at the surprisingly low

*Work supported by the National Science Foundation.

**1980-81 Visiting Fellow, Joint Institute for Laboratory Astrophysics, National Bureau of Standards and University of Colorado, Boulder, Colorado 80309.

pressure of about 2.8 Torr. At 6.0 Torr, this ratio is reduced to 1:10 showing how rapidly signal is lost with increasing pressure. Cells with pressures from 1–2 Torr have been used for the work reported in this paper.

The observed He^+ Zeeman resonance line width of 800 Hz is accounted for as follows: ~450 Hz, spin-exchange with electrons; 148 Hz, diffusion; 85 Hz, three-body collisions with He; 80 Hz, pulsed-discharge lifetime width; remainder due to field inhomogeneity, spin-exchange with Rb, and rf saturation. The known charge exchange cross section [7] for collisions between He^+ and He ground state atoms permits calculation of the charge exchange rate $\Gamma \sim 4 \times 10^7$/s. Since the major causes of observed linewidth have been accounted for without inclusion of charge exchange, we conclude that such collisions contribute to the linewidth at most a small fraction of the collision rate. This is attributed to the absence of a nuclear moment for the $^4He^+$ species.[1]

For the same reason, one finds the $^4He^*$ Zeeman resonance linewidth (400 Hz in a 2 Torr cell) is accounted for without inclusion of a component due to the known metastability exchange cross section [8]. The He^* exchange rate with ground state He is calculated to be ~10^7/s. The coherence of the rf transitions for the He^+ and He^* species in the presence of such exchange collision rates is impressive!

Figure 2 shows the resolved resonances of the free electron spin, $^4He^*$, and $^4He^+$ in a 100 G magnetic field (frequency ~280 MHz). In order to display the broad electron resonance, the rf level used saturated all resonances. Under conditions which optimize the electron spin resonance, linewidths of ~320 Hz have been obtained for the electron: in Fig. 2, the electron spin resonance linewidth is ~5 kHz.

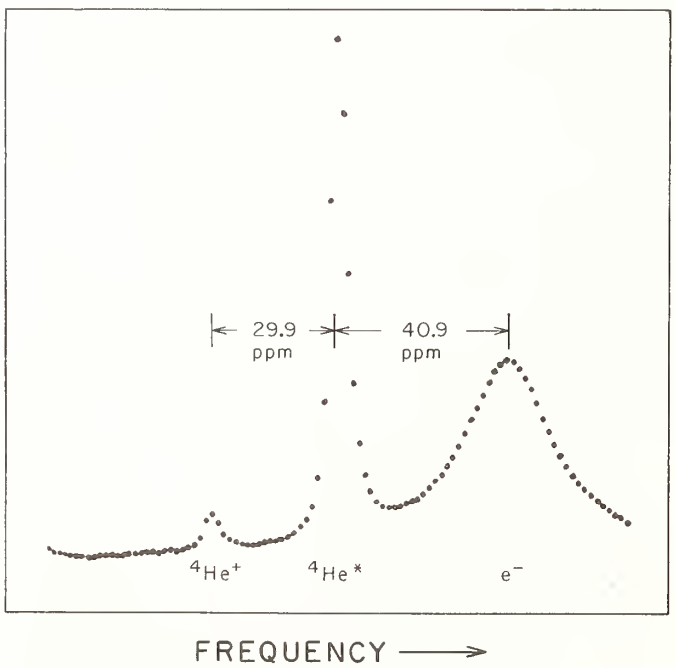


FIGURE 2. *Resolved Zeeman resonances for the $^4He^+$, $^4He^*$, and free electron spin (e^-) species. These resonances appear in the vicinity of 280 MHz in the 100 G applied magnetic field. A saturating rf level was used to enhance the broad e^- resonance. Separations in ppm are shown.*

[1]In Ref. [7] the Zeeman resonance of $^3He^+$ was observed with typical linewidths of 800 kHz at pressures of ~0.1 Torr.

In Fig. 3 a higher resolution scan shows the $^4He^+$ and $^4He^*$ resonances, again in a 100 G magnetic field. Data acquired for determining the ratio of g-values for these species were taken by sampling fifteen frequencies spread over four linewidths for each resonance. These thirty frequencies are interleaved so that each resonance is interrogated alternately. Several full cycles are used for signal averaging. A nonlinear least-squares routine fitting to a Lorentzian line shape with baseline and asymmetry yields the line-shape parameters. The ratio of g_J factors is obtained directly from the frequency ratio of the fitted resonance centers.

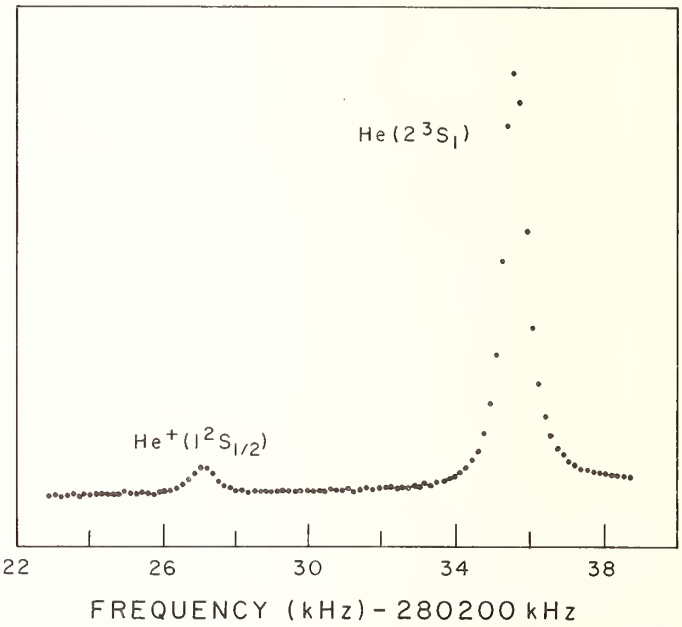


FIGURE 3. *High resolution scan of $^4He^+$ and $^4He^*$ Zeeman resonances in a 100 G magnetic field. The respective linewidths (FWHM) are 0.8 and 0.4 kHz and their separation is 8.37 kHz.*

3. Results

Two systematic effects have been identified which tend to limit the precision attainable in the measurement of $g_J(He^+)/g_J(He^*)$. The first is a shift in the He^+ Zeeman resonance frequency as a function of the polarization and intensity of the Rb pumping light. The resonance frequency for σ^- polarization differs from that for σ^+ by $\sim 1 \times 10^{-7}$. However, the average of the data for the two polarizations as a function of light intensity has zero slope. Thus such averaging appears to account correctly for this systematic effect. The shift is interpreted as due to spin-exchange. Its small size is partially attributed to the low polarization of all the species in the pulsed discharge afterglow.

The second systematic effect is associated with a small asymmetry of the He^* resonance. This has been a characteristic common to all cells used for our He^* work, even those with He pressure $\geq$ 5 Torr where a He^+ signal was not observed. Since the He^* resonance linewidth is a measure of the discharge intensity, an extrapolation of the values of $g_J(He^+)/g_J(He^*)$ versus He^* linewidth is used to suppress this effect. However, in the low pressure cells required to observe the He^+ resonance, the asymmetry appears to be smaller. This is consistent with the low polarization in such cells and a theoretical explanation [5] of the asymmetry as due to spin-exchange coupling between the nearly degenerate electron-spin resonance and the

He* resonance. No other resonances in the vicinity of He* have been found.

The current experimental result is

$$g_J(^4\mathrm{He}^+,\ 1\,^2S_{1/2})/g_J(^4\mathrm{He},\ 2\,^3S_1) = 1 - 29.95(10) \times 10^{-6}.$$

The error quoted is non-statistical and is estimated to cover possible remaining systematic effects.

By using the previously determined ratios $g_J(\mathrm{He}^*)/g_J(\mathrm{Rb})$ [5] and $g_J(\mathrm{Rb})/g(\mathrm{e}^-)$ [3], the following can be obtained:

$$g_J(^4\mathrm{He}^+,\ 1\,^2S_{1/2})/g(\mathrm{e}^-) = 1 - 70.87(10) \times 10^{-6}.$$

We stress that this is a purely experimental result found without the use of any theoretical ratios. The theoretical value [1] for this ratio is

$$\begin{aligned} g_J(\mathrm{He}^+)/g(\mathrm{e}^-) &= 1 - (Z\alpha)^2/3 + (\alpha/4\pi)(Z\alpha)^2 \\ &\quad + (m/2M)(Z\alpha)^2 + \ldots \\ &= 1 - (71.002 - 0.124 - 0.015 + \ldots) \\ &\quad \times 10^{-6} \\ &= 1 - 70.863 \times 10^{-6}. \end{aligned}$$

Thus, the experiment confirms the Breit term $(Z\alpha)^2/3$ to within 1.4 parts in 10^3.

We look forward to improvements in the apparatus including an increased magnetic field and laser pumping which should permit a precision of at least a few parts in 10^8. This will permit a determination of the radiative correction term $(\alpha/4\pi)(Z\alpha)^2$ to about 10%.

References

[1] H. Grotch and R. A. Hegstrom, Phys. Rev. A 4, 59 (1971).
[2] F. G. Walther, W. D. Phillips, and D. Kleppner, Phys. Rev. Lett. 28, 1159 (1972).
[3] J. S. Tiedeman and H. G. Robinson, Phys. Rev. Lett. 39, 602 (1977).
[4] P. G. Barkley and R. A. Hegstrom, Phys. Rev. A 14, 1574 (1976).
[5] G. M. Keiser, H. G. Robinson, and C. E. Johnson, Phys. Rev. A **16**, 822 (1977).
[6] C. E. Johnson and H. G. Robinson, Phys. Rev. Lett. 45, 250 (1980).
[7] M. Pinard and M. Leduc, J. Phys. (Paris) 35, 741 (1974).
[8] J. Dupont-Roc, M. Leduc, and F. Laloë, Phys. Rev. Lett. 27, 467 (1971).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Determination of the Neutron Magnetic Moment

G. L. Greene* and N. F. Ramsey

Harvard University, Cambridge, MA 02138

W. Mampe

Institut Laue-Langevin, 38042 Grenoble, France

J. M. Pendlebury and K. Smith

University of Sussex, Falmer, Brighton, BNI 9 QH, United Kingdom

W. B. Dress and P. D. Miller

Oak Ridge National Laboratory, Oak Ridge, TN 37838

Paul Perrin

Centre d'Etudes Nucléaires, 38042 Grenoble, France

The neutron magnetic moment has been measured with an improvement of a factor of 100 over the previous best measurement. Using a magnetic resonance spectrometer of the separated oscillatory field type capable of determining a resonance signal both for neutrons, and for protons in flowing H_2O, we find $\mu_n/\mu_p = 0.68497935(17)$ (0.25 ppm). The neutron magnetic moment can also be expressed without loss of accuracy in a variety of other units.

Key words: flowing water NMR; magnetic resonance; neutron magnetic moment; particle properties.

1. Introduction

The realization that the neutron, a neutral particle, possesses a non-zero magnetic moment has been of considerable importance in the development of nuclear and particle physics. Even before any explicit measurement of μ_n had been made, there was a strong indication that $\mu_n \neq 0$ from comparisons of the proton and deuteron magnetic moments μ_p and μ_d. If the neutron and proton combine in a pure $^3S_1$ state to form the deuteron, one would expect $\mu_d = \mu_n + \mu_p$. Early measurements of μ_p and μ_d gave, according to this relation, $\mu_n \sim -1.8\ \mu_N$ where μ_N is the nuclear magneton.

According to Dirac theory, which assumes structureless, spin 1/2 particles, $\mu_n = 0$ and $\mu_p = \mu_N$. The anomolous moments are defined as the differences between the actual moments and the Dirac moments. The first serious attempt to explain the anomalous moments of the nucleons was made by Frölich, Heitler and Kemmer [1] in 1938. Their "meson exchange theory" predicted equal magnitudes and opposite signs for the anomolous proton and neutron moments [2]. The actual magnitudes of the moments are much more difficult to obtain using this theory. Though subsequently refined by many others [3] the initial conclusions of Frölich, Heitler, and Kemmer adequately summarized the state of the theoretical understanding of the neutron magnetic moment until the introduction of the quark model.

The first explicit measurement of the neutron magnetic moment was made by Alverez and Bloch [4] in 1940. Using Rabi's magnetic resonance technique with a neutron beam obtained by deuteron bombardment of Be, they obtained a value of $|\mu_n| = 1.93 \pm 0.02\ \mu_N$. In general, no information about the sign of a magnetic moment is obtainable from a resonance experiment utilizing a purely oscillating magnetic field. (In 1949 Rogers and Staub [5] determined the sign of μ_n using a resonance technique that employed a rotating field. As had been thought, the sign was negative.)

The discovery of Kellog, Rabi, Ramsey, and Zacharias [6] that the deuteron has an electric quadropole moment implied that the ground state of the deuteron could not, in fact, be accurately described by a pure $^3S_1$ state but must have some admixture of D state. Thus the additivity of neutron and proton moments in the deuteron could not be exact. It was therefore of great interest to determine μ_n more accurately, so that it could be compared with the well known values of μ_p and μ_d.

In 1947 Arnold and Roberts [7], using a neutron beam from a reactor, determined μ_n in a resonance experiment with a single oscillatory field. They were able to obtain a more accurate value of $|\mu_n| = 1.913(12)\mu_N$ by using the then novel technique of NMR as a means of field calibration. About the same time, Block, Nicodemus, and Staub [8] also obtained a more accurate value for μ_n. They reported their result as the ratio of neutron to proton moment $|\mu_n/\mu_p| = 0.685001(30)$.

With these more accurate values for μ_n, it was possible to determine the $^3D_1$ admixture in the deuteron ground state by two independent means; through the deuteron electric quadropole moment, or by comparison between μ_n, μ_p and μ_d. Gratifyingly, these independent values were consistent.

*Current address: Gibbs Laboratory, Physics Department, Yale University, New Haven, CT 06520

With the introduction of the quark model an appealing explanation for the value of the ratio of the neutron moment to the proton moment emerged. Sakita [9] and Bég, Lee, and Pais [10] pointed out that this ratio should be $\mu_n/\mu_p = -2/3$ [11] if the internal symmetry SU(3) is broken *only* by electromagnetism. In this model, the ratios of all the baryon moments are uniquely determined and easily calculated. (Their absolute magnitudes are dependent on the quark magnetic moments. Determining these would require knowledge of the quark masses.) The agreement between theory and experiment for μ_n/μ_p is considerably better than for other baryon pairs and is striking for such a simple theory. The approximately 3% discrepancy between theory and experiment should not be viewed as a significant disagreement. SU(3) $\otimes$ SU(2) (i.e., 3 similar quarks with spin 1/2) is known to be an incomplete description for strongly interacting particles.

At the initiation of the work reported here, the uncertainty associated with theory far exceeded the experimental error of the best measurement of μ_n [12]. As a result, theoretical considerations did not provide the dominant motivation for the current work. Rather, the possibility of a substantial improvement in our knowledge of a fundamental particle property as well as the opportunity to demonstrate an elegant, new experimental technique prompted our effort.

The most accurate determination of μ_n, prior to the work reported here, was that of Corngold, Cohen, and Ramsey [12]. This experiment involved a thermal neutron beam and benefitted from the use of the separated oscillatory field method. A result of $|\mu_n/\mu_p| = 0.685039(17)$ was obtained. The major limitation to the accuracy of this measurement arose from the method used to calibrate the magnetic field. An NMR probe of small dimensions was employed to determine the field at a number of discrete locations. This field map was then used to give the appropriate field average by a pointwise integration. The discrete nature of this stepwise integration led to the dominant error in the experimental result.

The current work benefitted from several advances in experimental technique. Access to the high flux reactor of the Institut Laue-Langevin at Grenoble, France, allowed the use of an intense "cold" neutron beam. Such "cold" beams are characterized by low beam velocities. This led to an improvement of a factor of ~5 over the line width obtained by Cohen, Corngold, and Ramsey [12]. An additional advantage to the use of "cold" neutrons lies in the ability to employ neutron guides to reduce beam divergence and therefore increase the flux. Such guides also allowed a novel field calibration technique which provided the major improvement in experimental sensitivity.

Following a suggestion of Purcell [13] made some time ago, the field was monitored by obtaining a separated oscillatory field resonance signal with water flowing through the neutron guide. In this fashion, the field average taken by the protons in the water is the same as that taken by the neutrons. Since all the relevant diamagnetic corrections for protons in water are known to high accuracy, a direct measurement of the ratio of the neutron magnetic moment to the proton magnetic moment was obtained. From the result, the magnetic moment of the neutron can be expressed in a variety of other units.

2. Apparatus

The aim of the experimental technique was to measure the Larmor precession frequency for neutrons and protons (in water) in precisely the same volume as nearly simultaneously as possible. With the use of a neutron guide (a glass tube of circular cross-section) it was a straightforward matter to confine neutrons and water in the same volume. One simply let the water flow through the neutron guide. This procedure, however, precluded the possibility of determining the resonance frequency for neutrons and protons with exact simultaneity. To insure that no spurious effects resulted from drifts during the time required to make a change-over from neutron beam to flowing water, a second, parallel tube was installed in the spectrometer. Water was continuously sent through this tube to provide a field monitor. By alternating neutrons and protons in the principal tube while protons flowed continuously in the monitor tube, errors due to field drifts could be detected and reduced to an insignificant level. This technique has been discussed in detail elsewhere [14].

The cold source at the high flux reactor of the Institut Laue-Langevin in Grenoble provided the source of neutrons used in this experiment. The flux through the apparatus was typically 1.5×10^6 neutrons/s. the neutrons were polarized by glancing reflection from a magnetized mirror. A similar mirror was used to analyze the neutron polarization at the exit of the spectrometer. The neutrons were detected by a ^{6}Li loaded glass scintillator coupled to a photo-multiplier.

The protons in the flowing water were polarized by passage through a chamber placed in a high (~2 kG) prepolarizing magnet. As the protons spent more than one longitudinal relaxation time in this field, their polarization could be approximated by a simple Boltzmann distribution. While the degree of polarization so obtained was very small, the enormous flux (by molecular beam standards) led to an easily detected signal. The proton polarization at exit was detected in a second high field region (~4 kG) with slightly modified commercial NMR equipment.

The spectrometer was based on a high stability, high homogeneity, large volume electromagnet. The nominal field in the magnet was 18 G. The separated oscillatory field coils (each 3 cm long) were mounted in the magnet with a separation of 61 cm. Surrounding each of the separated coils were independently controllable trim coils to adjust the fields at the coils.

The magnet and spectrometer were mounted in a two layer cylindrical molypermalloy magnetic shield to reduce the effects of time dependent external fields. The entire assembly was mounted on a large rotating platform. This allowed the orientation of the spectrometer with respect to beam velocity to be reversed.

3. Results

Figures 1 and 2 show typical neutron and proton resonance lineshapes obtained in this experiment. The dramatic difference in line widths (~120 Hz vs ~0.8 Hz) for neutrons and protons is due to the great difference in particle velocities for the two cases. The neutron velocity was ~150 m/s while the proton velocity was ~1 m/s. The presence of many more side lobes in the proton resonance is indicative of a narrower relative proton velocity distribution.

The actual quantity measured in this experiment was the ratio between the Larmor precession frequencies for neutrons in vacuum and for protons in a cylindrical sample of water at 22 °C. To arrive at this quantity, corrections for several shifts and distortions had to be made.

The largest systematic effect resulted from the Bloch-Siegert effect [15]. This effect may be viewed as arising

from the counter-rotating component present in an oscillating field. It is a curious feature of the separated oscillatory field technique that the shift resulting from this effect is not strictly linear in oscillating field power [16]. However, this non-linearity is well understood and can be simply accounted for. Figure 3 illustrates the extrapolation technique which was used to account for the Bloch-Seigert shift.

The data shown in Fig. 3 were taken in the two possible machine orientations. These correspond to a reversal of the spectrometer with respect to the beam velocity. The shift seen between the two orientations is presumably due to geometrical imperfections in coil construction. It should be noted that the shifts seen in Fig. 3 arise entirely from the distortion of the neutron resonance. The much lower velocity of the protons, and the consequent reduction in oscillating power, implies negligible shifts in the proton resonance frequency.

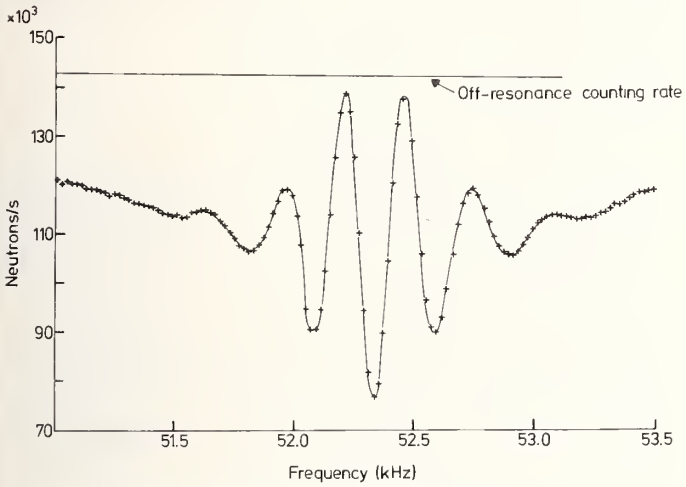


FIGURE 1. *Typical neutron resonance.*

Table 1 summarizes the significant shifts and errors in the measurement of the ratio of neutron moment to proton moment in a cylindrical sample of H_2O at 22 °C, ω_n/ω_p (cyl, H_2O, 22°). The errors quoted (1 σ) in Table 1 are independent and are added in quadrature to yield the error in ω_n/ω_p (cyl, H_2O, 22°). Our experimental result is

$$\omega_n/\omega_p \text{ (cyl, } H_2O\text{, 22°)} = 0.68499588(16) \text{ (0.24 ppm)} \quad (1)$$

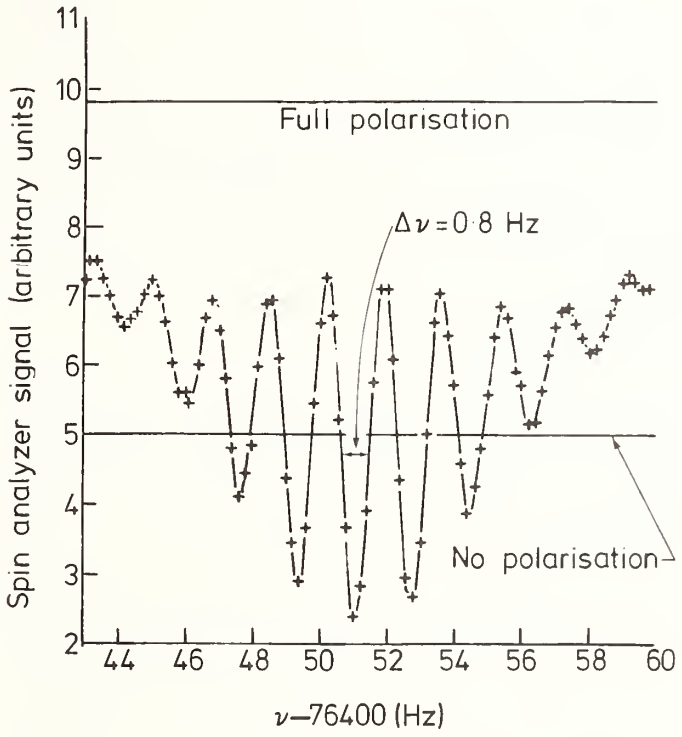


FIGURE 2. *Typical proton resonance.*

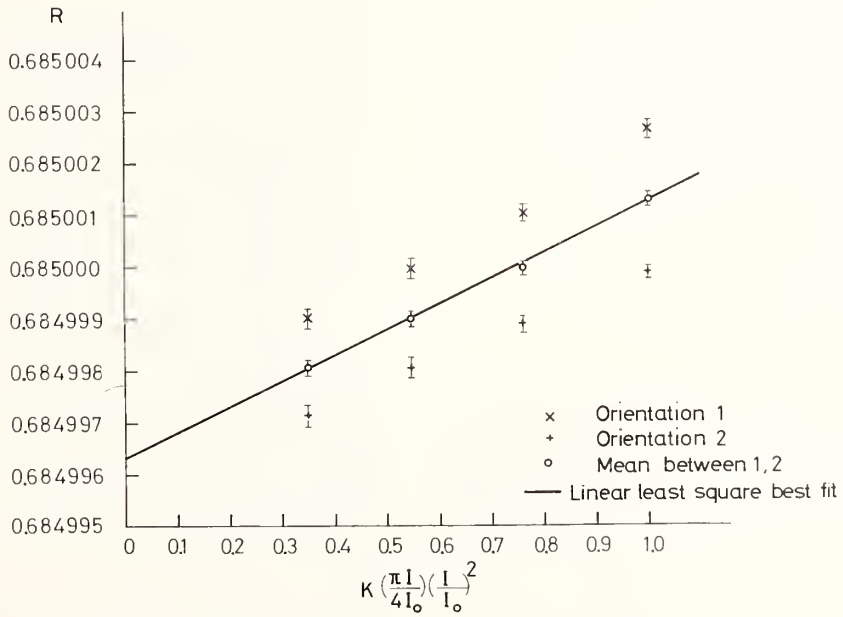


FIGURE 3. *Plot of neutron to proton resonance frequencies. The quantity* $\kappa(\pi I/4I_0)\,(I/I_0)$ *is an appropriately scaled oscillating power which yields a linear fit. (See references in Ref. [16] for a full description). The two orientations correspond to reversals of the machine axis with respect to the beam velocity.*

TABLE 1. *Significant shifts and errors in ω_n/ω_p (cyl, H_2O, 22°). Statistical error is included in the error assigned to the Bloch-Siegert effect and coil phase errors.*

Effect	Proportional shift	Proportional error
Bloch-Siegert effect	$+4.94 \times 10^{-6}$	13×10^{-8}
Coil phase errors	$\sim 1.6 \times 10^{-6}$	
Field inhomogeneity	$\lesssim 3 \times 10^{-7}$	$< 4 \times 10^{-8}$
Effects of finite velocity of distribution	$\sim 10^{-7}$	$\sim 10 \times 10^{-8}$
External diamagnetic effects of water	-4.8×10^{-7}	$\lesssim 3 \times 10^{-8}$

The value of the magnetic moment of the neutron can be expressed in terms of the Bohr magnetron μ_B with little loss of accuracy through the relation

$$\frac{\mu_n}{\mu_B} = \frac{\omega_n}{\omega_p(\text{cyl}, H_2O, 22°)} \times \frac{\omega_p(\text{cyl}, H_2O, 22°)}{\omega_p(\text{sph}, H_2O, 22°)} \frac{\omega_p(\text{sph}, H_2O, 22°)}{\omega_p(\text{sph}, H_2O, 35°)} \times \frac{\mu_p'}{\mu_B} \quad (2)$$

where the notation "sph" refers to a spherical sample, and μ_p' is the effective proton moment in a spherical sample of H_2O at 35 °C as measured by Phillips, Cook, and Kleppner [17].

The first ratio in the right hand side of Eq. 2 is our experimental result. The second ratio is given by $1 - 2\pi\kappa/3 = 1 + 1.505(2) \times 10^{-6}$ from the data summarized by Pople *et al.* [18] where κ is the volumetric susceptibility of water. The third ratio is taken to be $1 + 1.4(1) \times 10^{-7}$ from the results of Hindman as interpreted in a footnote in Phillips *et al.* [17]. The final ratio of μ_p'/μ_B is given by Phillips *et al.* [17]. From Eq. 2 we conclude.

$$\frac{\mu_n}{\mu_B} = -1.04187564(26) \times 10^{-3} \quad (0.25 \text{ ppm}). \quad (3)$$

The slight increase in experimental error arises from the temperature correction in Eq. 2.

Using the results of Winkler, Kleppner, Myint, and Walther [19], we can express μ_n in terms of the free electron moment μ_e and free proton moment μ_p by

$$\frac{\mu_n}{\mu_e} = 1.04066884(26) \times 10^{-3} \quad (0.25 \text{ ppm}) \quad (4)$$

$$\frac{\mu_n}{\mu_p} = -0.68497935(17) \quad (0.25 \text{ ppm}).$$

Our result can also be expressed in terms of the nuclear magneton μ_N. In this case, however, the accuracy is slightly degraded by the uncertainty in the electron to proton mass ratio. Using the value of m/M obtained by combining results from Phillips *et al.* [17] and Cohen and Taylor [18] we have

$$\frac{\mu_n}{\mu_N} = -1.91304184(88) \quad (0.45 \text{ ppm}) \quad (5)$$

Note added in proof: The recently reported value for the proton to electron mass ratio of Van Dyck and Schwinberg [these proceedings and Phys. Rev. Lett. 47, 395 (1981)] of M/m = 1836.15300(25) can be combined with the results reported here to give the value $\mu_n/\mu_N = -1.91304308(54)$ (0.28 ppm).

We wish to express our thanks to the director and staff of the Institut Laue Langevin for their assistance and hospitality during the course of this experiment. This work was supported in part by the Department of Energy and the National Science Foundation (U.S.A.) and Le Department de Recherche Fundamental du Commissariat à l'Energie Atomique (France).

References

[1] H. Frölich, W. Heitler, and N. Kemmer, Proc. R. Soc. London, Ser. A: 166, 154 (1938).
[2] E. Fermi, in his book *Nuclear Physics* (University of Chicago Press, Chicago 1951) gives a simple explanation of the meson exchange model, though he evidently did not think highly of the theory, calling it a "naive little fantasy" (p. 13).
[3] See G. L. Greene *et al.*, Phys. Rev. D **20**, 2139 (1979) for a more complete bibliography.
[4] L. Alverez and F. Bloch, Phys. Rev. **57**, 111 (1940).
[5] E. H. Roger and H. R. Staub, Phys. Rev. **76**, 980 (1949).
[6] J. M. B. Kellog, I. I. Rabi, N. F. Ramsey, and J. R. Zacharias, Phys. Rev. **57**, 677 (1940).
[7] W. Arnold and A. Roberts, Phys. Rev. **71**, 878 (1947).
[8] F. Bloch, D. B. Nicodemus, and H. R. Staub, Phys. Rev. 74, 1025 (1949).
[9] B. Sakita, Phys. Rev. Lett. 13, 643 (1964).
[10] M. A. Bég, B. W. Lee, and A. Pais, Phys. Rev. Lett. 13, 514 (1964).
[11] See also R. D. Young, Am. J. Phys. 41, 472 (1973) for a simple vector addition argument which gives the −2/3 result. This argument was first suggested by Bég *et al.*
[12] V. W. Cohen, N. R. Corngold, and N. F. Ramsey, Phys. Rev. 104, 283 (1956).
[13] E. M. Purcell, private communication (1952).
[14] G. L. Greene, Ph.D. Thesis, Harvard University (1977), available as Institut Laue Langevin Technical Report 77GR2065; G. L. Greene *et al.*, Phys. Rev. D **20**, 2139 (1979).
[15] F. Bloch and A. Siegert, Phys. Rev. 57, 522 (1940).
[16] J. H. Shirley, J. Appl. Phys. 4, 783 (1963); R. F. Code and N. F. Ramsey, Phys. Rev. A 14, 1945 (1971); and G. L. Greene, Phys. Rev. A 18, 1057 (1978).
[17] W. Phillips, W. E. Cook, and D. Kleppner, Phys. Rev. Lett. 35, 1619 (1975).
[18] J. A. Pople, W. G. Schneider, and H. J. Bernstein, *High Resolution Nuclear Magnetic Resonance* (McGraw Hill, N.Y., 1959).
[19] P. F. Winkler, K. Kleppner, T. Myint, and F. G. Walther, Phys. Rev. A 5, 83 (1973).
[20] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Precision Exotic Atom Spectroscopy

Vernon W. Hughes

Gibbs Laboratory, Physics Department, Yale University, New Haven, CT 06520

Precision measurements by microwave or laser spectroscopy techniques of the energy levels of exotic atoms containing particles such as positrons, muons, or pions not present in ordinary atoms allow the determination of the fundamental constants associated with these so-called exotic but basic particles. Moreover, the simplicity of some of these atoms, such as positronium (e^+e^-) and muonium (μ^+e^-), which consist only of structureless leptons, allows precise tests of quantum electrodynamics which is basic to our understanding of many of the fundamental constants, especially the fine-structure constant, α. This paper reviews recent work and work in progress on positronium, muonium, and simple muonic atoms relevant to the precision measurement-fundamental constants field.

Key words: muonic atoms; muonium; positronium; quantum electrodynamics.

1. Introduction

In the context of this "Second International Conference on Precision Measurement and Fundamental Constants" the topic of precision exotic atom spectroscopy can offer several contributions.

Firstly, since by definition an exotic atom contains particles such as positrons, muons, or pions not present in ordinary atoms, which consist of electrons and nuclei, study of exotic atoms will clearly provide the opportunity to determine the fundamental constants associated with these so-called exotic but basic particles. Hence, for example, the masses of the negative muon and the negative pion and the magnetic moments of the positive and negative muons are determined from exotic atom spectroscopy.

Secondly, the simplicity of some of the exotic atoms, such as positronium (e^+e^-) and muonium (μ^+e^-) which consist only of structureless leptons, provides an ideal opportunity for precise tests of the fundamental theory of quantum electrodynamics which is basic to our understanding of the fundamental constants. Muonic atoms can in general be considered as hydrogenic in the sense that to a good approximation the atom can be treated as a single μ^- bound to the nucleus because the outer electrons are relatively unimportant. Thus they provide a simple system for studying the basic interactions. An outstanding example is the truly hydrogenic atom $(\mu^-{}^4\mathrm{He})^+$ where a critical test of electron vacuum polarization has been made. Relatively low Z muonic atoms such as μ^- P and μ^- Mg have also been used to study electron vacuum polarization and to establish limits to exotic muon-nucleus interactions. Since the muon in a muonic atom has higher momenta than the electron in an electronic atom, tests of quantum electrodynamics can be extended to significantly higher momenta or correspondingly shorter distances with muonic atoms.

Finally, some of the experimental techniques used to study exotic atoms are quite different from those used for ordinary atoms. These include accelerators to produce the exotic particles, precision measurements of higher energy x-rays or γ-rays, and particle detectors. These different techniques broaden the field of precision measurement.

Since this conference is focused on the fundamental atomic constants and quantum electrodynamics, I will discuss principally the atoms most relevant to these topics which are positronium, muonium, and simple muonic atoms. However, brief mention of hadronic atoms and some other matters will be made.

2. Positronium

Since its discovery by Deutsch [1] in 1951, positronium has served as an important testing ground for elementary particle physics and fundamental interactions. These tests include precise spectroscopic measurements of fine structure intervals, measurements of annihilation rates, and studies of the symmetries associated with the electron-positron system. In this paper we treat only the precise energy level measurements; a recent review article by Rich [2] covers thoroughly the topics of annihilation rates and symmetries.

The energy levels of the $n = 1$ and $n = 2$ states of positronium, including its fine structure intervals, are indicated in Fig. 1. Theoretical values to order $\alpha^2 R_\infty$ are given by [3]:

$$
\begin{aligned}
E_{(n,L,S,J)} &= -\frac{R_\infty}{2n^2} \\
&\quad + \left[\frac{11}{32n^4} + \left[\epsilon_{LSJ} - \frac{1}{2L+1}\right]\frac{1}{n^3}\right]\alpha^2 R_\infty , \\
\epsilon_{(L,S=0,J)} &= 0 , \\
\epsilon_{(L,S=1,J)} &= \frac{7}{6}\delta_{L0} + \frac{1-\delta_{L0}}{2(2L+1)} \times \\
&\quad \begin{cases} \dfrac{3L+4}{(L+1)(2L+3)} & \text{if } J = L+1 \\ -\dfrac{1}{L(L+1)} & \text{if } J = L \\ -\dfrac{3L-1}{L(2L-1)} & \text{if } J = L-1 \end{cases} , \qquad (1)
\end{aligned}
$$

where α is the fine structure constant; R_∞ is the Rydberg constant for the electron; and L, S, and J are the quan-

tum numbers for orbital, spin, and total angular momentum, respectively.

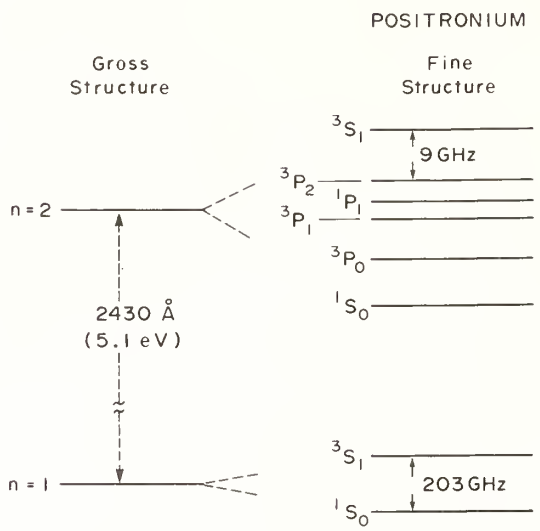


FIGURE 1. *Energy levels of the $n = 1$ and $n = 2$ states of positronium.*

For the $n = 1$ state the virtual radiative contributions to the hfs interval $\Delta\nu$ between 1^3S_1 orthpositronium and 1^1S_0 parapositronium have been calculated [4] through relative order $\alpha^2 \ln\alpha^{-1}$, as indicated in Eq. (2):

$$\Delta\nu = \alpha^2 c R_\infty \left\{ \frac{7}{6} - \frac{\alpha}{\pi}\left[\frac{16}{9} + \ln 2\right] + \frac{5}{12}\alpha^2 \ln\alpha^{-1} + K\alpha^2 - K'\frac{\alpha^3}{\pi}(\ln\alpha^2)^2 \right\} . \quad (2)$$

Some of the Feynman diagrams contributing to the terms with coefficients K and K' have been evaluated, but complete calculations of these terms remain to be done. Using values of the atomic constants given in Table 1, we obtain:

$$\Delta\nu = (203\ 339.1 + 10.8\ K - 2.1\ K')\ \text{MHz}. \quad (3)$$

For the $n = 2$ state virtual radiative contributions to the fine structure intervals of relative order α, or equivalently to the energy levels of order $\alpha^3 R_\infty c$, have been evaluated [5]. Hence for the only fine structure interval that has been measured, the theoretical value is given by:

$$E(2^3S_1) - E(2^3P_2) = \frac{23}{480}\alpha^2 c R_\infty (1 + 3.766\ \alpha)$$
$$= 8625.14\ \text{MHz} . \quad (4)$$

The measurement of the hfs interval $\Delta\nu$ in the ground state involves the Zeeman effect and utilizes the Zeeman transition indicated in Fig. 2 [6-10]. The energy of the $M = \pm 1$ states is independent of magnetic field H because of the equality of the magnitudes of the magnetic moments of e^- and e^+ as required by CPT invariance. The energies of the $M = 0$ triplet and $M = 0$ singlet states have approximately a quadratic dependence on H, and the frequency of the indicated transition between $M = 0$ and $M = \pm 1$ triplet states is given by

$$f_{01} = \frac{\Delta\nu}{2}[1 + x^2)^{1/2} - 1] , \quad (5)$$

TABLE 1. *Relevant atomic constants.*

R_∞	=	1.097 373 152 1 (11) × 10^5 cm^{-1} (0.001 ppm)[a]
c	=	2.997 924 580 (12) × 10^{10} cm s^{-1} (0.004 ppm)[b]
α^{-1}	=	137.035 963 (15) (0.11 ppm)[c]
m_{e^+}/m_{e^-}	=	1 ± 1.3 × 10^{-7} [d]
m_{μ^+}/m_{e^-}	=	206.768 259 (62) (0.3 ppm)[e]
μ_p/μ_B^e	=	1.521 032 209 (16) × 10^{-3} (0.01 ppm)[b]
μ_{μ^+}/μ_p	=	3.183 345 47 (95) (0.30 ppm)[f]
a_e	=	1 159 652 200 (40) × 10^{-12} g
a_μ	=	1 165 924 (8.5) × 10^{-9} [h]

[a]S. R. Amin, C. D. Caldwell, and W. Lichten, Phys. Rev. Lett. 47, 1234 (1981).
[b]E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[c]E. R. Williams and P. T. Olsen, Phys. Rev. Lett. 42, 1575 (1979).
[d]Ref. [16].
[e]V. W. Hughes and T. Kinoshita, in *Muon Physics*, Vol. I., Ed. by V. W. Hughes and C. S. Wu (Academic Press, New York, 1977), p. 11.
[f]Average value from following two references:
(i) E. Klempt, R. Schulze, H. Wolf, M. Camani, F. N. Gygax, W. Rüegg, A. Schenck, and H. Schilling, Phys. Rev. D 25, 652 (1982).
(ii) Ref. [22].
[g]H. Dehmelt, in *Atomic Physics 7*, Ed. by D. Kleppner and F. M. Pipkin (Plenum Press, New York, 1981), p. 337; also P. B. Schwinberg *et al.*, these proceedings.
[h]J. Bailey, K. Borer, F. Combley, H. Drumm, C. Eck, F. J. M. Farley, J. H. Field, W. Flegel, P. M. Hattersley, F. Krienen, F. Lange, G. Lebee, E. McMillan, G. Petrucci, E. Picasso, O. Runolfsson, W. von Ruden, R. W. Williams, and S. Wojcicki, Nucl. Phys. B 150, 1 (1979).

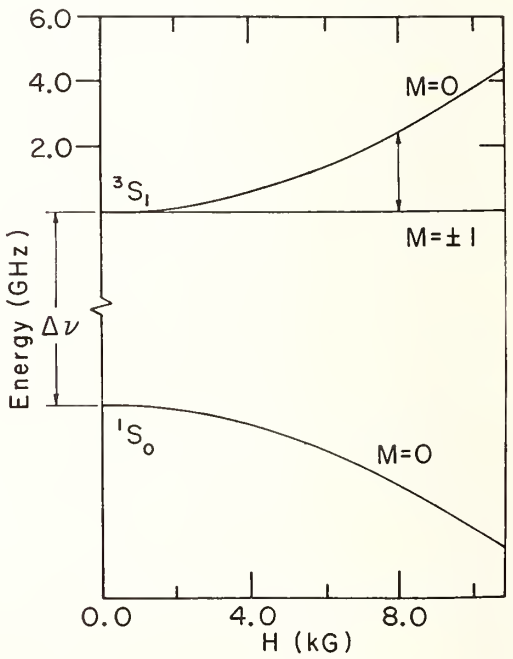


FIGURE 2. *Zeeman energy levels of positronium in its ground state.*

where $x = 2\mu_B^e g' H/(h\Delta\nu)$, in which g' is the g-value of the electron in positronium. The annihilation of the ground $^1S_0$ state results in the emission of two 511 keV photons and occurs at the rate $\lambda_p = 0.80 \times 10^{10}\ s^{-1}$, whereas the annihilation of the ground $^3S_1$ state results in the emission

of three photons whose energies sum to $2mc^2$ and occurs at the rate $\lambda_o = 0.72 \times 10^7\ s^{-1}$. In a magnetic field the $M = 0$ triplet state has an admixture of $M = 0$ singlet state and hence can decay either by two photon or three photon emission, and the annihilation rate is $\lambda_{10} \simeq \lambda_o + (x^2/4)\lambda_p$ in which the two photon part becomes important even at small x because $\lambda_p \gg \lambda_o$. Analogously, in a magnetic field the $M = 0$ singlet state has an admixture of $M = 0$ triplet state but the three photon annihilation mode is generally unimportant. The principle of the resonance experiment rests on the unequal population of the $M = 0$ triplet state and the $M = \pm 1$ triplet states in a magnetic field and on the increase in two photon annihilation which results from induced transitions from the $M = \pm 1$ to the $M = 0$ triplet states.

The experimental arrangement of the present Yale experiment [8, 10] is shown in Fig. 3. A ^{22}Na positron source of about 15 mCi is placed inside a microwave cavity resonant at 2323 MHz and filled with N_2 gas to a pressure between 0.25 and 3 atm. Eight NaI(Tl) detectors count in coincidence 0.5 MeV annihilation γ rays emitted at 180°. The magnetic field of about 8 kG is varied across the resonance line as indicated in Fig. 4. Analysis of the resonance line has been based on the Hamiltonian for positronium in a magnetic field $H_0 + H_1 \sin\omega t$ in which H_0 is the static magnetic field and $H_1 \sin\omega t$ is the time dependent microwave magnetic field of amplitude H_1 and angular frequency ω. The annihilation interaction has been treated by introducing decaying amplitudes for the states. Figure 5 displays the principal published values of $\Delta\nu$ since the original microwave resonance experiment of Deutsch and Brown, and Table 2 includes the two most recent measured values together with the theoretical value. Agreement between theory and experiment is satisfactory within the present accuracy of the theory. Clearly values of the coefficients K and K' are needed to allow a more sensitive comparison of theory and experiment.

Recently Rich [11] has pointed out that the neglect of off-diagonal matrix elements of the annihilation interaction may introduce significant error in the determination of $\Delta\nu$.

Further improvement in the experimental accuracy of determining $\Delta\nu$ should be possible by straightforward technical approaches together with improved statistics. Such an experiment is in the stage of final data analysis at Yale and a reduction of the error to about 3 ppm is anticipated.

Although the ratio of line center frequency to line width is the same for the Zeeman transition which has been studied as for a direct $\Delta\nu$ transition at zero magnetic field, it would nonetheless be valuable to measure the direct $\Delta\nu$ transition. The $\Delta\nu$ frequency of 203 GHz is inconveniently high, but adequate microwave power sources of the carcinotron type are now available.

The use of line narrowing techniques may offer a hope of an improved measurement, but as yet no practical experiment has been proposed.

The beautiful measurement of a fine structure interval in the $n = 2$ state of positronium by the Brandeis group [12-14] was made possible by their discovery that slow positrons (kinetic energy $\simeq$ 100 eV) striking certain solid surfaces in vacuum have an appreciable probability of rebounding as positronium in the 2S state in vacuum. Figures 6 and 7 show the Brandeis apparatus and relevant decay channels of the $n = 1$ and $n = 2$ states. Positrons of 30 eV are incident on a Cu target, and produce Ps in the 2S state. An electric dipole transition is induced by a microwave field from the 2^3S_1 to the 2^3P_2 state, which

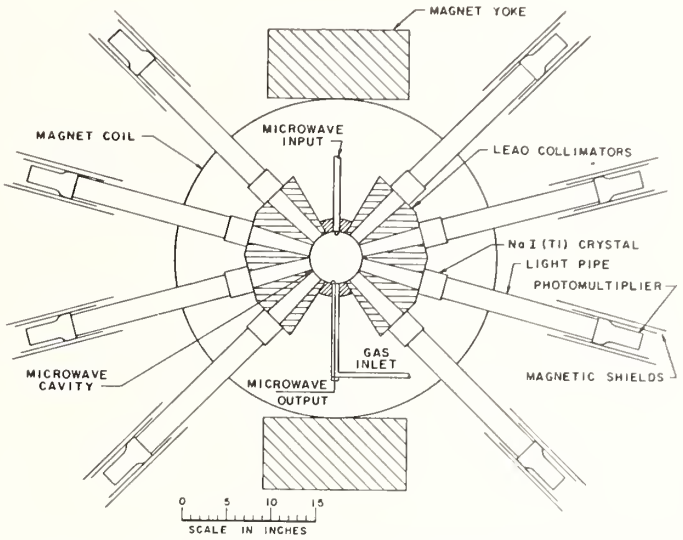


FIGURE 3. *Schematic diagram of the experimental apparatus.*

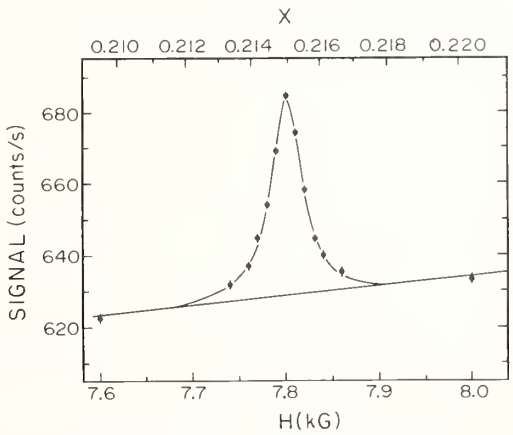


FIGURE 4. *Typical data for a resonance pass fit to the theoretical lineshape. f_{01} − 2323.364 MHz, 0.31 atm N_2.*

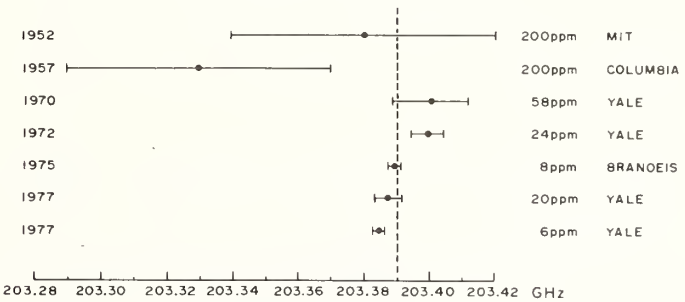


FIGURE 5. *History of measurements of $\Delta\nu$ of positronium.*

then emits a Lyman α photon of 2430 Å in the spontaneous transition to the 1^3S_1 state, followed by the three γ annihilation of this state. The signal shown in Fig. 7 is the increase in the Lyman α photon-annihilation γ delayed coincidence counts due to the microwave field. The experimental error in measuring the line center is largely statistical, but the Zeeman effect and motional Stark effect also contribute significantly to the uncertainty in determining the 2^3S_1 to 2^3P_2 interval. Table 2 gives the experimental and theoretical values of the 2^3S_1 to 2^3P_2 interval. Agreement is satisfactory and constitutes a test

TABLE 2. *Experimental results for positronium fine structure intervals and comparison with theory.*

$n = 1$ state; $\Delta\nu\,(1^3S_1 - 1^1S_0)$

Experimental value	Theoretical value
203 384.9 (1.2) MHz (6 ppm)[a]	203 400.3 MHz[c]
203 387.0 (1.6) MHz (8 ppm)[b]	

$$\Delta\nu_{th} - \Delta\nu_{exp} = 14.6(1.0)\ \text{MHz}$$

$n = 2$ state; $\nu(2^3S_1 - 2^3P_2)$

Experimental value	Theoretical value
(8 628.4 ± 2.8) MHz[d]	8 625.14 MHz[e]

$$\nu_{th} - \nu_{exp} = -3.3(2.8)\ \text{MHz}$$

[a]Ref. [9].
[b]A. P. Mills and G. H. Bearman, Phys. Rev. 34, 246 (1975).
[c]P. Lepage and D. Yennie, these proceedings.
[d]Ref. [13].
[e]Ref. [5].

Note Added in Proof (see Ref. [11]): Corrections to the $n = 1$ state $\Delta\nu$ values due to off-diagonal matrix elements of the annihilation interaction have been estimated by Mills. His weighted mean value of the two corrected measurements is $\Delta\nu$ = 203.388 5 (10) GHz (5 ppm).

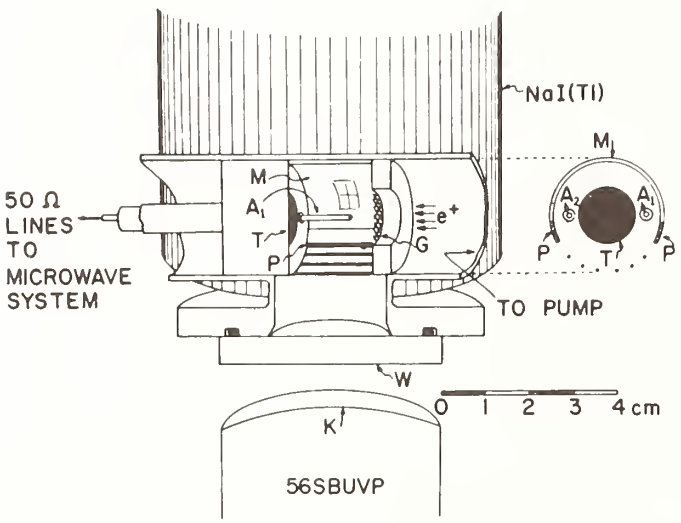


FIGURE 6. *Positron target chamber and microwave cavity. G, 95%-transmission tungsten grid; T, copper target; M, aluminized Suprasil quartz mirror; W, Suprasil quartz window; K, CsTe photocathode; P, support posts; A_1, input antenna; A_2, output antenna; NaI(Tl), annihilation detector coupled to an RCA 8575 phototube. Another NaI(Tl) detector (not shown) is placed above the chamber directly over the detector shown.*

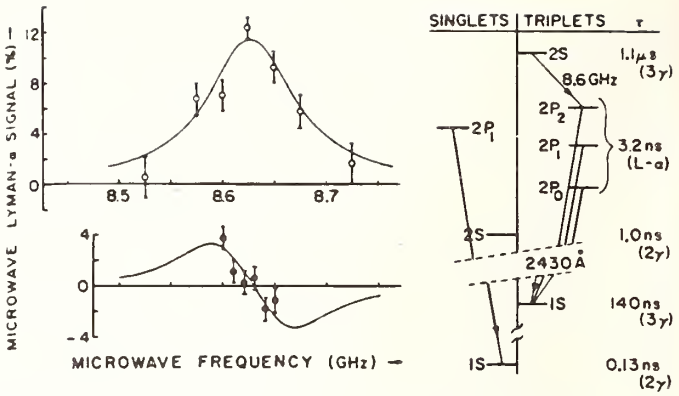


FIGURE 7. *The observed Lyman-α signal S (open circles) and logarithmic first-difference signal S' (solid circles) as a function of microwave frequency. The inset is the schematic term diagram for the $n = 2$ and $n = 1$ Ps states indicating the relevant transitions and the lifetimes τ for each level.*

of the $\alpha^3 R_\infty$ term to about 1%. An active program continues at Brandeis to improve the accuracy in the measurement of the fine structure of Ps in the $n = 2$ state.

Recently in several laboratories there has been an interest in the observation of the 1S-2S two photon transition in Ps. The theoretical value of the energy interval is [3, 15]

$$\nu(2^3S_1 - 1^3S_1) = \nu_{21} = \frac{3}{4}cR_p - \alpha^2 cR_p\left[\frac{719}{768}\right] + O(\alpha^3 R_p)\ , \quad (6)$$

in which the Rydberg constant for positronium, $R_p = R_\infty[m_{e^+}/(m_{e^+} + m_{e^-})]$, where $2\pi^2 me^4/h^3c$ is the usual Rydberg constant with $m = m_{e^-}$. The equality of the masses of electron and positron is expected on the basis of CPT invariance and has been established experimentally to 0.1 ppm [16]. The term $O(\alpha^3 R_p)$ is the lowest order radiative contribution and its principal part is the Lamb shift in the 1S state. A complete evaluation of this term has not yet been done.

The experiment would be similar in principle to the observation of the 1S-2S two photon transition in H by Doppler-free laser spectroscopy [17, 18]. A principal problem is to obtain enough Ps atoms during the short laser pulse. The experimental scheme being considered at Yale is indicated in Fig. 8. Detection would be achieved through observation of the e^+ resulting from photonionization of Ps(2S) in the laser beam.

Note Added in Proof: The $1^3S_1 - 2^3S_1$ transition in positronium has recently been observed using two-photon Doppler free excitation. The observed frequency agrees with theory to 0.5 GHz [S. Chu and A. P. Mills, Jr., Phys. Rev. Lett. 48, 1333 (1982)].

3. Muonium

Muonium is the hydrogen-like atom consisting of a positive muon and an electron. From the viewpoint of elementary particle physics and fundamental atomic physics, the motivations for studying muonium are to determine the properties of the muon and to measure the muon-electron interaction in this bound atomic state. Indeed muonium provides an ideal simple system for testing modern muon electrodynamics and for searching

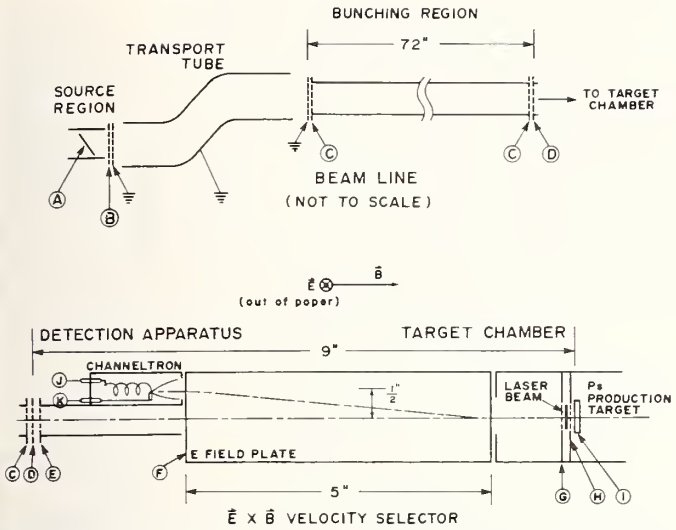


FIGURE 8. *Schematic diagram of the proposed Yale experiment to measure the 1S → 2S 2-photon transition in Ps by laser spectroscopy.*

for effects of weak, strong, or unknown interactions on the electron-muon bound state. Since its discovery in 1960 [19], made possible by the observation of parity nonconservation in the $\pi \rightarrow \mu \rightarrow$ e decay chain (Fig. 9), precision measurements have been made of the hyperfine structure interval $\Delta\nu$ in the ground $n = 1$ state and of the muon magnetic moment [20]. These studies of μ^+e^- have all confirmed that the muon behaves as a heavy electron [21].

Muonium is an ideal system to study the electromagnetic interaction of two different leptons and hence to test muon electrodynamics, which includes the viewpoint that the muon and electron are elementary, structureless particles; and also to determine the electromagnetic properties of the muon, including its spin, magnetic moment, and mass. In addition, a precise value of the fine structure constant α is determined from the hyperfine structure interval of muonium. Also certain aspects of the weak interactions of the muon and electron can be tested with muonium. A precise study of a basic system like muonium, for which the measured quantities can be calculated accurately, provides criteria for testing new speculative theories.

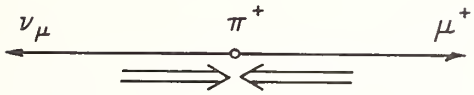
Important advances have recently been made in both our experimental [22] and theoretical [4] knowledge of the hyperfine structure interval $\Delta\nu$ in the ground $n = 1$ state of muonium. Closely related improvement has been achieved in the precision of measurement of the magnetic moment of the positive muon.

The present theoretical value for $\Delta\nu$ is given in Table 3. Values of the required atomic constants are given in Table 1. The value for μ_μ/μ_p is the weighted average of the value from the LAMPF experiment [22] and from the μSR measurement in liquid Br_2 at SIN [23]. There are two principal contributions to the estimated error in $\Delta\nu_{th}$. One is the inaccuracy in the calculation of the radiative correction term ϵ_3 and uncalculated radiative-recoil terms, which contribute an error of about 0.7 ppm. The second is the experimental uncertainty in μ_μ/μ_p of about 0.3 ppm.

The classic muonium microwave magnetic resonance experiment, which is possible because of parity nonconservation in the $\pi \rightarrow \mu \rightarrow$ e decay chain (Fig. 9), has recently been redone at the Los Alamos Meson Physics Facility (LAMPF) using a much more intense source of stopped muons obtained from a so-called "surface" muon beam [24]. The energy level diagram of the ground state of muonium in a static magnetic field is shown in Fig. 10. The general technique was the same as in an earlier experi-

PION-MUON DECAY

$$\pi^+ \longrightarrow \mu^+ + \nu_\mu$$



$$\mu^+ \longrightarrow e^+ + \nu_e + \overline{\nu}_\mu$$

$$I_{e^+}(\theta_{e^+,s_\mu}) \propto 1 + \tfrac{1}{3}\cos\theta$$

$\pi \rightarrow \mu \rightarrow e$ *decay chain.*

TABLE 3. *Theoretical value for muonium $\Delta\nu$ and comparison with experiment.*

$$\Delta\nu_{th} = \left[\frac{16}{3}\alpha^2 cR_\infty(\mu_\mu/\mu_B^e)\right][1 + (m_e/m_\mu)]^{-3} \times \left[1 + \frac{3}{2}\alpha^2 + a_e + \epsilon_1 + \epsilon_2 + \epsilon_3 - \delta'_\mu\right]$$

$$a_e = (g_e - 2)/2;\ \epsilon_1 = \alpha^2(\ln 2 - 5/2)$$

$$\epsilon_2 = -\frac{8\alpha^3}{3\pi}\ln\alpha\left[\ln\alpha - \ln 4 + \frac{281}{480}\right];\ \epsilon_3 = \frac{\alpha^3}{\pi^3}(18.4 \pm 5)$$

$$\delta'_\mu = \frac{3\alpha}{\pi}\frac{m_R}{m_\mu - m_e}\ln\frac{m_\mu}{m_e} + \alpha^2\frac{m_R}{m_\mu + m_e} \times \left[2\ln\alpha + 8\ln 2 - 3\frac{11}{18}\right] + \left(\frac{\alpha}{\pi}\right)^2\frac{m_e}{m_\mu}\left[2\ln^2\left(\frac{m_\mu}{m_e}\right) - \frac{31}{12}\ln\left(\frac{m_\mu}{m_e}\right) + \left(\frac{28}{9} + \frac{\pi^2}{3} - 1.9\right)\right]$$

where $m_R = \dfrac{m_e m_\mu}{m_e + m_\mu}$

$\Delta\nu_{th}$ = (4 463 303.7 ± 3.4) kHz (0.77 ppm)

$\Delta\nu_{exp}$ = (4 463 302.88 ± 0.16) kHz (0.036 ppm)

$\Delta\nu_{exp} - \Delta\nu_{th}$ = (−0.8 ± 3.4) kHz

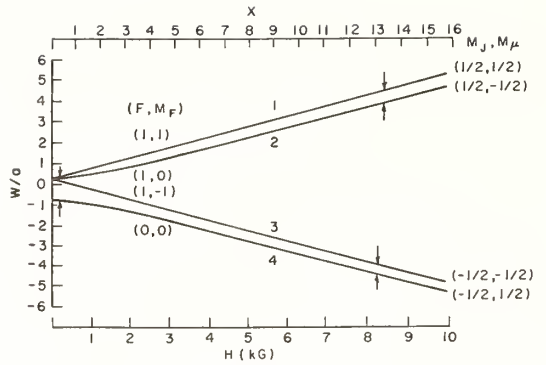


FIGURE 10. *Energy level diagram for muonium in its $1^2S_{1/2}$ ground state in a magnetic field.*

ment [25]. The two Zeeman transitions $(M_J, M_\mu) = (1/2, 1/2) \longleftrightarrow (1/2, -1/2)$ and $(-1/2, -1/2) \longleftrightarrow (-1/2, +1/2)$ were observed alternately at a strong static magnetic field of 13.6 kG. The very intense surface muon beam had a mean momentum of $p_\mu = 28$ MeV/c with a 10% momentum spread and a polarization of close to 100%, and provided μ^+ stopping rates of up to 10^6 s^{-1} average in a 0.5 atm Kr gas target 20 cm in length for a primary proton beam current of 300 μA. The experimental arrangement is shown in Fig. 11. The thin scintillation counter S1 monitors the incoming muon beam and the counters S2, S3 detect the decay electrons. A central element in this experiment is the high precision solenoid electromagnet which provides a magnetic field over the region of the microwave cavity which is homogeneous to several ppm and is stable to about 1 ppm. A typical resonance curve is shown in Fig. 12. It was obtained in a period of about 2 h by varying the magnetic field under computer control. For each field point the microwave frequency is changed from the ν_{12} to the ν_{34} frequency and also the microwave power is turned on and off with a modulation period of about 0.16 s. The fitted theoretical line shapes are shown. The result of our data analysis for the hfs interval $\Delta\nu$ and for the ratio μ_μ/μ_p of muon to proton magnetic moments is

$$\begin{aligned} \Delta\nu &= 4\ 463\ 302.88 \pm 0.16\ \text{kHz}\ (0.036\ \text{ppm}) \\ \mu_\mu/\mu_p &= 3.183\ 346\ 1 \\ &\quad \pm 0.000\ 001\ 1\ (0.36\ \text{ppm})\ , \end{aligned} \tag{7}$$

in which one standard deviation total errors are given. Figures 13 and 14 give the histories of the principal measurements of $\Delta\nu$ and of μ_μ/μ_p, and indicate that the published values are in reasonable agreement within their quoted errors.

The agreement of the theoretical and experimental values of $\Delta\nu$ is well within the error of the theoretical value. This agreement provides one of the most sensitive tests of modern muon electrodynamics.

Active theoretical work is now in progress on $\Delta\nu$ to evaluate further radiative and recoil contributions to the level of a few parts in 10^7. On the experimental side, the next important advance will probably involve the use of line-narrowing techniques in the measurement of strong field Zeeman transitions. At LAMPF this will require a pulsed muon beam with a time structure of the order of the muon lifetime of 2 μs, which could be obtained either by pulsing the primary proton beam, by pulsing the muon beam in the muon channel, or by using the proton storage ring now under construction at LAMPF [26].

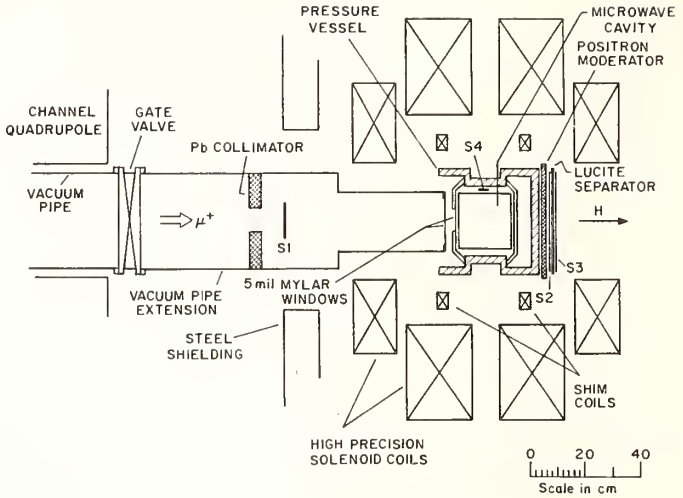


FIGURE 11. *Experimental setup. S4 is a 0.24-in. thick plastic scintillator. It was used for optimizing Kr pressure and beam degrader (not shown) thickness so that the muon beam stops in the middle of the cavity. S1—beam monitor, S2, S3—decay positron signature.*

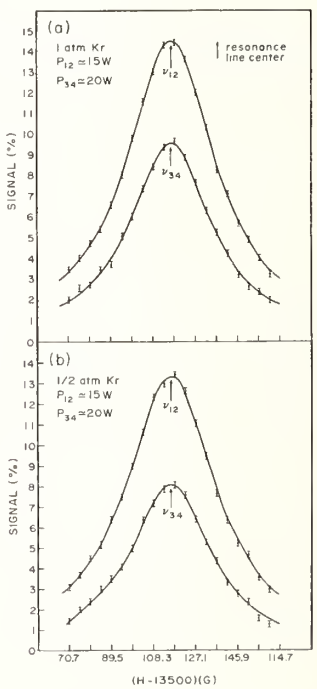


FIGURE 12. *Typical resonance lines and their fits (solid lines). Each pair of resonance lines was obtained in less than 2 h of data accumulation. Data were taken alternately on opposite sides of each curve to minimize signal asymmetries due to possible time dependent power drifts.*

Two further researches on muonium that would be of great interest [20] are, firstly, the measurement of the fine structure, Lamb shift, and hyperfine structure intervals in the $n = 2$ excited state; and secondly, a direct search for the conversion of muonium to antimuonium.

The energy level diagram for the $n = 1$ and $n = 2$ states of muonium is shown in Fig. 15. It would be particularly valuable to make a precise measurement of the Lamb shift, or $2^2S_{1/2}$ to $2^2P_{1/2}$ energy difference. Although the Lamb shift in H is known experimentally to high precision [27], its theoretical interpretation is limited by uncertainties associated with proton structure.

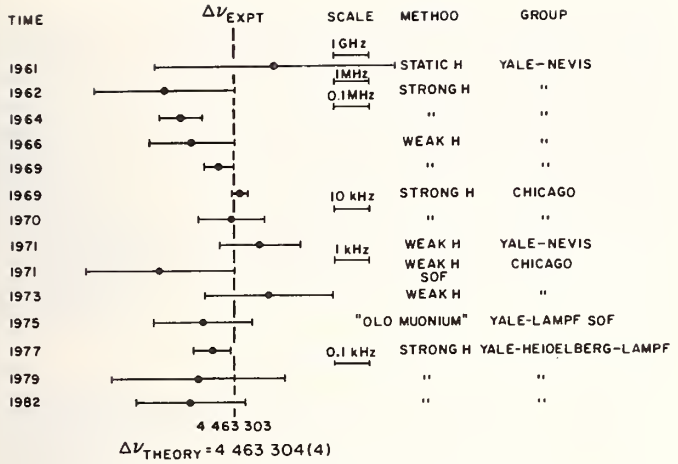


FIGURE 13. *History of measurements of $\Delta\nu$ of muonium.*

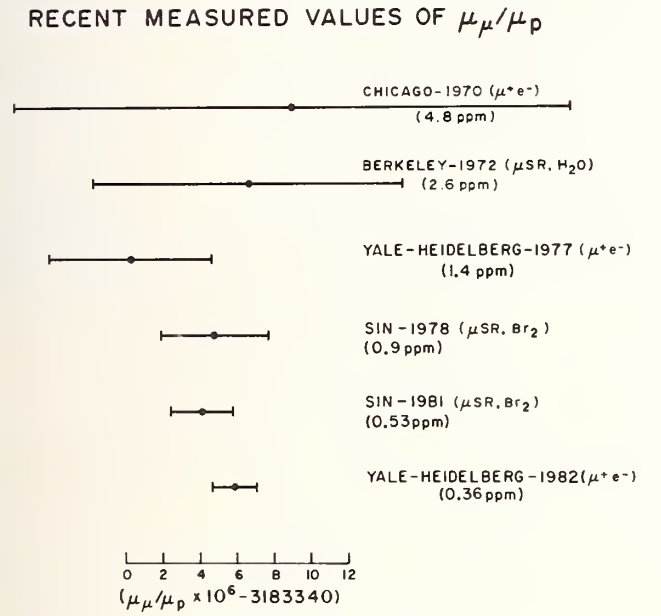


FIGURE 14. *History of μ_μ/μ_p values.*

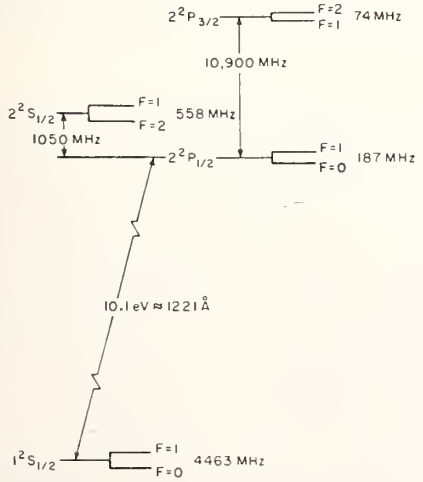


FIGURE 15. *Energy level diagram for muonium in its n = 1 and n = 2 states, including fine structure, Lamb shift, and hfs terms.*

Muonium is an ideal system in which to test the quantum electrodynamic theory of the Lamb shift since both the muon and the electron are believed to be structureless particles, unlike the proton which has a structure associated with the strong interactions.

A microwave spectroscopy experiment to measure the Lamb shift in the $n = 2$ state of muonium can be done in principle by detecting 2P-1S Lyman-α photons or a change in decay positron angular distribution associated with a change in muon spin direction resulting from an induced transition. Because of the large cross section for quenching of the 2S state of hydrogen (and muonium) in an atomic collision, such an experiment would seem to require that μ^+e^-(2S) be in vacuum. Recently it has been found that muonium in its ground state can be formed by passing μ^+ with energies of about 1 MeV through a thin foil into vacuum [28]. The experimental arrangement is shown in Fig. 16. The spectrum observed in the NaI detector is shown in Fig. 17 where the high energy portion indicates that muonium has stopped in the downstream target. Formation of the 2S state should also occur in the foil, and an experiment is being undertaken at LAMPF to observe the $2S_{1/2} \rightarrow 2P_{1/2}$ transition.

A search for the direct transition of muonium to antimuonium might also be undertaken with muonium in vacuum.

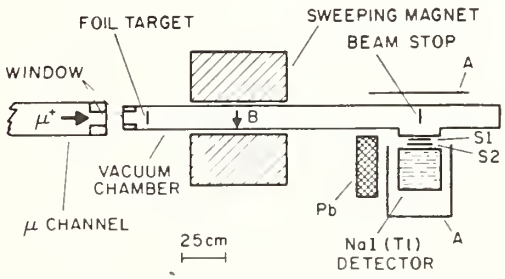


FIGURE 16. *Schematic diagram of the experimental apparatus; S_1, S_2, and A are plastic scintillation counters. The beam passes through a 125-μm Mylar window at the end of the μ channel, and a 50-μm Ti window entering the vacuum chamber, and is collimated to 7.5 cm diameter by polyethylene collimators.*

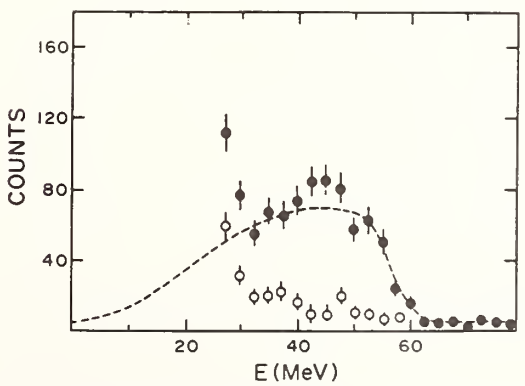


FIGURE 17. *Measured NaI spectrum for 25-μm Be foil target. Solid circles are with the apparatus evacuated; open circles are with 6 Torr helium. For each case 2.8 h of data are shown. The dashed curve is a fit to the data as described in the text.*

Some years ago the muonic helium atom ($^4He\mu^-e^-$) was discovered [29] from the observation of its characteristic, muonium-like, Larmor precession frequency. The Bohr picture of this amusing atom is indicated in Fig. 18. In a first approximation it can be considered to consist of an inner core or pseudo-nucleus consisting of $^4He\mu^-$ and of an outer electron. The pseudo-nucleus has a unit positive charge and a magnetic moment equal to that of μ^-. Hence the muonic helium atom is analogous to muonium and would be expected to have a hyperfine structure interval similar in magnitude to that of μ^+e^- but inverted because of the negative magnetic moment of μ^-.

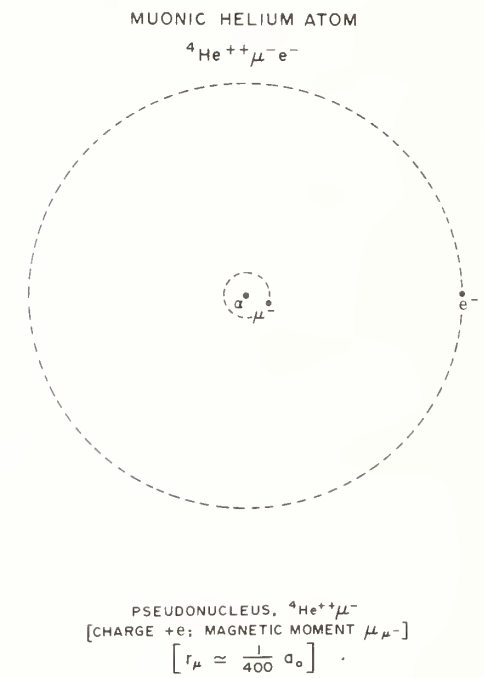


FIGURE 18. *Schematic diagram of the muonic helium atom. The orbits are not to scale. The radius of the "pseudonucleus" is* 1.3×10^{-11} *cm, while the electron orbit radius is* $r_{Bohr} = 5.3 \times 10^{-9}$ *cm.*

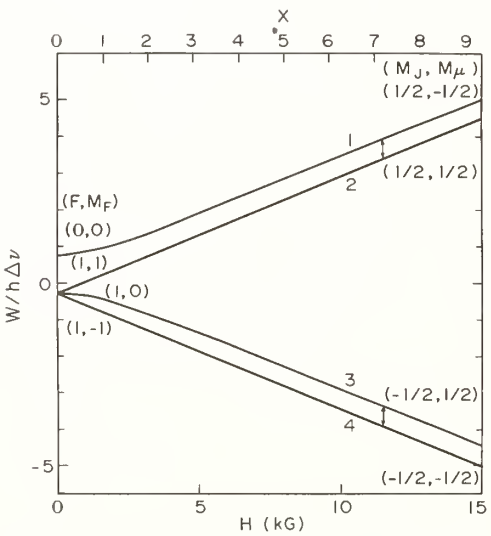


FIGURE 19. *Breit-Rabi energy level diagram of* $^4He\mu^-e^-$.

The hyperfine structure transition of $^4He\mu^-e^-$ has recently been observed at SIN in a microwave magnetic resonance experiment at zero static magnetic field [30]. The resonance signal is relatively weak because of the low residual polarization of μ^- in $^4He\mu^-e^-$. A strong magnetic field measurement of greater precision has also been done [31] to determine both $\Delta\nu$ and μ_μ/μ_p. Figure 19 shows the Breit-Rabi energy level diagram and Fig. 20 shows a resonance curve for the strong field transition. Table 4 gives the experimental values for $\Delta\nu$.

The theoretical value of $\Delta\nu$ for $^4He\mu^-e^-$ can be written [32] as:

$$\Delta\nu = \Delta\nu_F[1 + \delta^{rel} + \delta^{rad} + \delta^{rec}] , \qquad (8)$$

in which $\Delta\nu_F$ is the leading Fermi term, and δ^{rel}, δ^{rad} and δ^{rec} refer to relativistic, radiative, and recoil contribu-

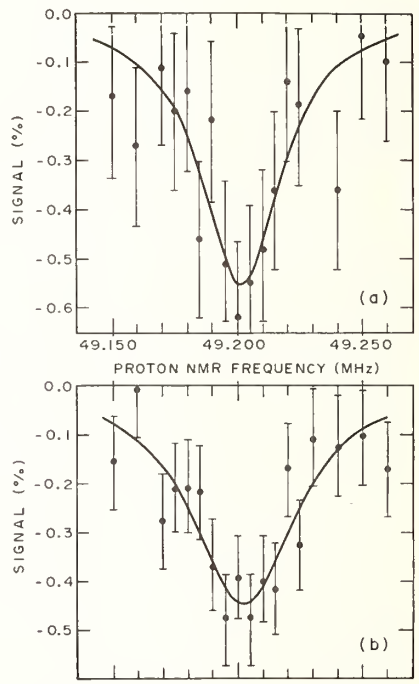


FIGURE 20. *Typical resonance curves for the* ν_{12}' *transition in* $^4He\mu^-e^-$ *obtained with the forward telescope at (a) 15 atm and (b) 5 atm. The data for these curves were obtained in (a) 24 h and (b) 100 h. For each curve obtained with the forward telescope there is a corresponding curve for the backward telescope.*

TABLE 4. *Comparison of theoretical and experimental values for the hfs interval in* $^4He\mu^-e^-$.

Method	$\Delta\nu$
Variation calculation[a]	4465.0(0.3) MHz
Perturbation theory[b]	4462.6(3.0) MHz
Born-Oppenheimer theory; global operator technique[c]	4460 MHz
Strong field experiment[d]	4465.004(29) MHz (6.5 ppm)
Weak field experiment[e]	4464.95(6) MHz (13 ppm)

[a]Ref. [32]; [b]Ref. [33]; [c]Ref. [35]; [d]Ref. [31]; [e]Ref. [30].

tions. Evaluation of $\Delta\nu_F$ with high accuracy is rather difficult because the most important aspect of the wavefunction involves the correlation of the muon and electron whereas the binding energy of the atom, the quantity to which a variational calculation applies, is determined principally by the muon part of the wavefunction. Theoretical values including the correction terms in Eq. 8 [32-35] are given in Table 4. The theoretical values are in satisfactory agreement with the experimental value within the relatively large theoretical errors. Improvement in the accuracy of the value of $\Delta\nu_F$ is particularly needed.

4. Muonic Atoms

Precision measurements in muonic atoms (ordinary nucleus plus μ^- plus electrons) have been of value for a sensitive test of electron vacuum polarization in QED and for a test of the existence of exotic interactions. It should be quickly remarked that no significant deviations from QED or exotic interactions have been observed [20].

The simplest muonic atom in which a significant precision measurement has been made is $(\mu^- {}^4\mathrm{He})^+$ in which fine structure intervals in the $n = 2$ state have been observed by laser spectroscopy. Figure 21 shows the energy levels of the $n = 1$ and $n = 2$ states [20]. Note particularly that the $2^2S_{1/2}$ level lies below the $2^2P_{1/2}$ level and that the energy interval $2^2P_{1/2}$ to $2^2S_{1/2}$ is much larger than the fine structure interval $2^2P_{3/2}$ to $2^2P_{1/2}$, due to the large negative contribution of electron vacuum polarization to the energy of the $2^2S_{1/2}$ state.

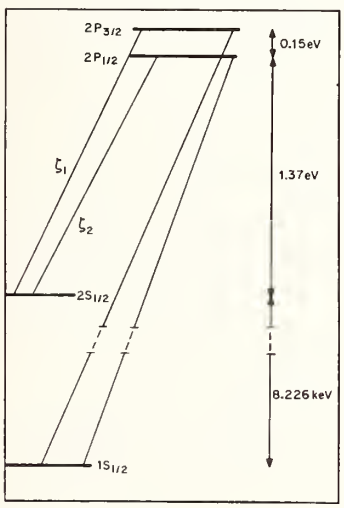


FIGURE 21. *Scheme of the lowest energy levels of the muonic helium ion* $(^4He\mu^-)^+$.

The experimental arrangement [36] is shown in Fig. 22. A fraction of the muons (~5%) stopped in He gas form $(\mu^- {}^4\mathrm{He})^+$ in the 2S state. A laser pulse at the transition wavelength from the $2^2S_{1/2}$ to the $2^2P_{3/2}$ state of about 8100 Å is then applied, and the transition is detected by observation of the 8.3 keV 2P-1S x-ray. Figure 23 shows an observed resonance line. Comparison of the experimental and theoretical [37] values is shown in Table 5. It is seen that the contribution of the nuclear size of ^{4}He is large compared to the experimental accuracy of 0.02%. The nuclear size is known from elastic $e^- - {}^4\mathrm{He}$ scattering and the uncertainty in this contribution to the interval is about 0.3%. The dominant contribution to the interval is electron vacuum polarization,

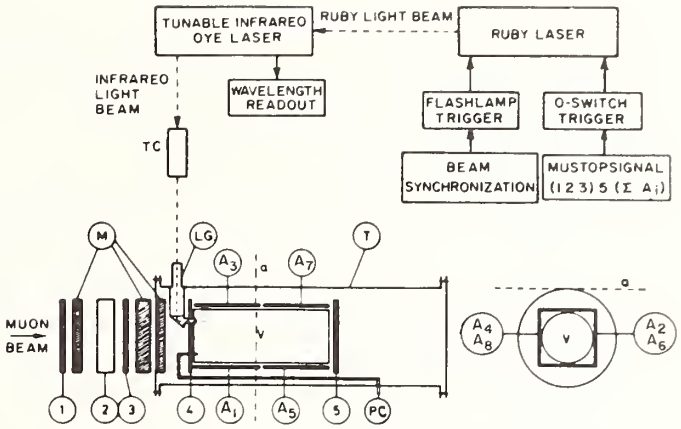


FIGURE 22. *Experimental setup for* $(^4He\mu^-)^+$ *fine structure measurement. M,* CH_2 *moderators; 1, 3, 4, 5, plastic scintillators; 2, anticoincidence Cerenkov counter; T, Invar steel tank; V, useful volume for stopping muons;* A_1-A_8, *NaI(Tl) counters; LG, light guide used to inject the infrared radiation into the target; TC, optical telescope; PC, optical fiber supplying the energy-monitoring signal.*

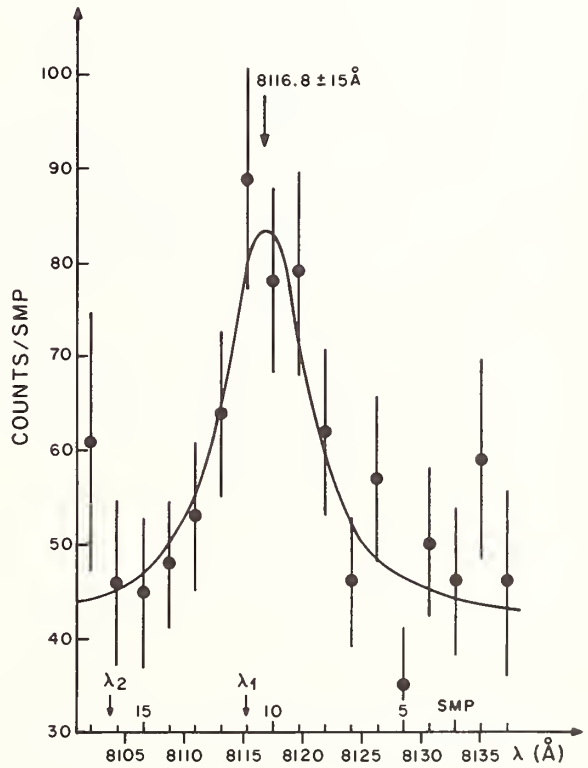


FIGURE 23. *The* $2S_{1/2} - 2P_{3/2}$ *resonance signal. Each datum point represents the number of events, normalized to the same number of stopped muons, per stepping motor position SMP; the scale in wavelength (shown below) is fixed by the calibration procedure* (λ_1 *and* λ_2 *are two lines of an Ar lamp.) The full-line curve drawn on the data is the result of a best-fit analysis* ($\chi^2 = 11$) *with a Lorentzian line (assuming for the 2P decay rate* Γ *the theoretical value* $\Gamma = 8$ *Å) plus a constant background.*

TABLE 5. *Muonic helium fine structure: theory and comparison with experiment: contributions[a] to the* $2^2P_{3/2} - 2^2S_{1/2}$ *transition energy[c]* $(\mu^- {}^4He)^+$

Fine structure	Electron vacuum polarization		Vertex corrections	Recoil	Nuclear size[b]	Nuclear polarization	Total (theory)	Total (expt.)
	$O(\alpha)$	$O(\alpha^2,\ldots)$						
145.7	1666.1	12.0	−10.7	−0.2	-288.9 ± 4.1 $-105.5\ r^2 + 1.40\ r^3$	3.1 ± 0.6	1527.1 ± 4.2	1527.5 ± 0.3

[a]Units in meV and fm.
[b]r^2 = mean square radius of ^{4}He = 2.802 fm^2
[c]The energy interval $2^2P_{1/2} - 2^2S_{1/2}$ has also been measured, and theory and experiment are in good agreement [36].

and agreement of theory and experiment to the accuracy of about 0.3% with which the ^{4}He size contribution is known confirms the radiative QED contribution. Alternatively, assuming that QED is correct, the experiment determines the most accurate value of the root mean square radius of ^{4}He.

In order to avoid the ^{4}He nuclear size effect which is associated with the S state, an experiment is being undertaken at the AGS at the Brookhaven National Laboratory to observe the 3D to 3P transition [38].

Similar measurements of muonic hydrogen (μ^-p) fine structure by laser spectroscopy would be useful to test electron vacuum polarization and also to study proton structure. Indeed, in principle a direct observation of the hfs transition in the 1S state is possible by observation of the accompanying change in the muon polarization. Figure 24 indicates the energy level scheme for the $n = 1$ and $n = 2$ states of μ^-p. For a fine structure measurement, formation of the 2S metastable state at low gas pressure is probably necessary because μ^-p (2S) is readily quenched in atomic collisions. Figure 25 shows observations of Lyman lines from μ^-p obtained in H_2 gas at pressures as low as 1 Torr [39]. These observations were achieved by using the low momentum μ^- from decay of π^- of about 40 MeV/c. However, attempts to observe μ^-p (2S) directly by collision-induced $2P \rightarrow 1S$ γ rays have not yet been successful [40, 41].

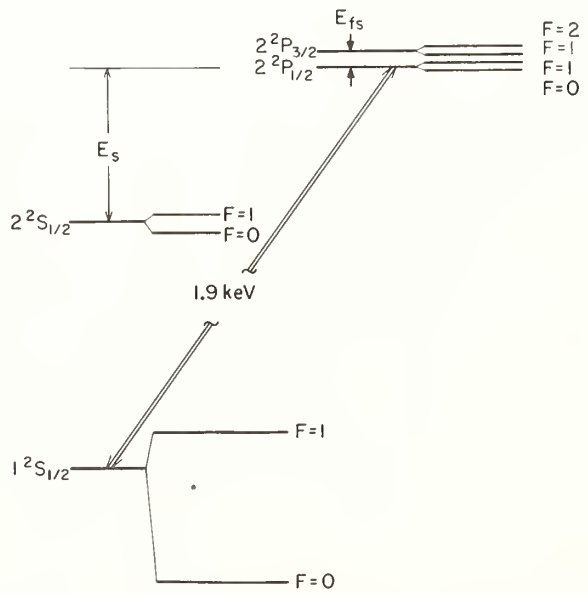


FIGURE 24. *Energy levels of $n = 1$ and $n = 2$ states of muonic hydrogen (μ^-p). Hyperfine structure interval $\Delta\nu$ in $1^2S_{1/2}$ state is ~4.4 × 10^{13} Hz.*

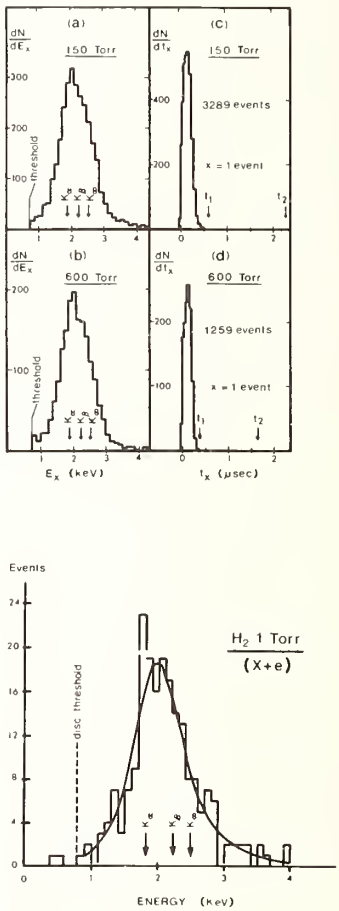


FIGURE 25. *X-ray energy distribution for μ^- stopping in H_2 at low pressures.*

In recent years precise measurements of spectra of medium Z muonic atoms have been made with a crystal diffraction spectrometer at SIN [42], where large μ^- stopping rates are achieved with a superconducting muon channel. Figure 26 shows the crystal diffraction spectrometer. Figure 27 displays some observed $3D \rightarrow 2P$ lines for $^{24}Mg\mu^-$. Figure 28 gives the fractional difference between standard QED theory and experiment. The good agreement can be interpreted as a measurement of electron vacuum polarization to about 0.3%. Alternatively, the result places a limit on non-QED muon-nucleus interactions; in particular, within the standard electroweak theory in which the Higgs boson mediates an interaction between muon and nucleus it places a lower limit of about 8 MeV to the mass of the Higgs boson.

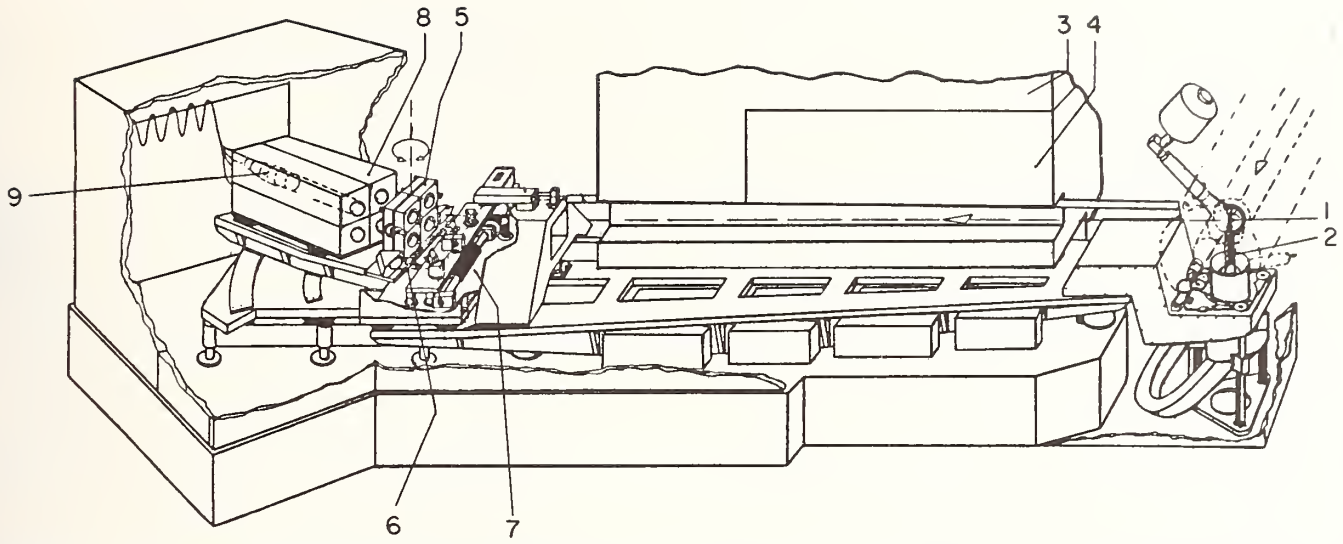


FIGURE 26. *Curved-crystal spectrometer for muonic X-rays at SIN, general view. 1. Muon target (X-ray source). 2. Cooled copper socket supporting the target holder. 3. Iron-loaded and normal concrete for shielding. 4. Iron shielding. 5. Bending blocks with (four) crystals. 6. Crystal turning table. 7. Threaded screw, part of the crystal-rotation system. 8. Tube collimator. 9. Phoswich detector.*

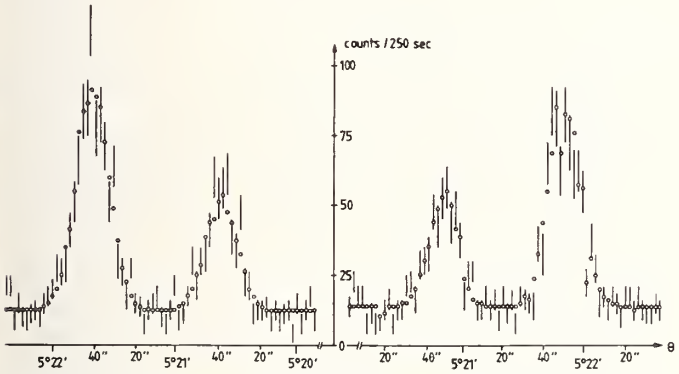


FIGURE 27. *Reflex pair (angular spectrum) from run 18 with $3D_{5/2} - 2P_{3/2}$ and $3D_{3/2} - 2P_{1/2}$ transitions. The sum of the angular spectra of the four crystals is shown. The bars represent the measured count rates. The circles are the values of the fit function corrected for the fluctuations in the muon stopping-rate.*

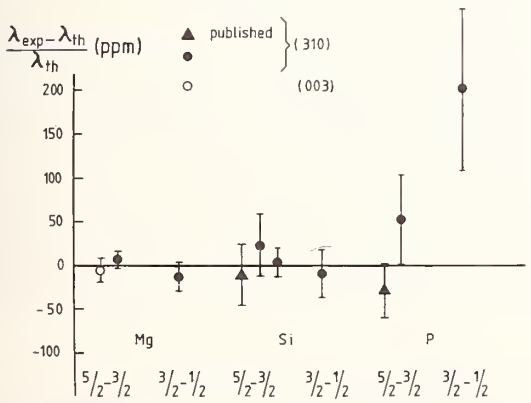


FIGURE 28. *Relative difference between experimental and theoretical (QED) values of transition wave-lengths (in ppm) for the individual runs and earlier published results. The values are grouped according to the six transitions measured ($3D_{5/2} - 2P_{3/2}$ and $3D_{3/2} - 2P_{1/2}$ transitions in ^{24}Mg, ^{28}Si and ^{31}P). Except for one case (open circle) diffraction from (310) planes at 3.2 mm thick quartz crystals was used. In run 15 diffraction was from (003) planes of quartz.*

In hadronic atoms the effects of the strong interaction between the hadron (π^-, K^-, $\bar{p}$, Σ^-, etc.) and the nucleus are important and have not been calculated from fundamental strong interaction theory. However, studies of the spectra of these atoms have yielded values for the masses and magnetic moments of these hadrons [43-46].

5. Conclusion

Exotic atoms consisting of pure leptons such as e^+e^- and μ^+e^- are ideal systems for studying QED and for measuring the properties of the leptons as well as for looking for exotic effects, because conventional theory involves no ambiguities. High precision measurements of energy intervals can be made by the methods of microwave and laser spectroscopy, limited ultimately by the lifetimes of the particles and atomic states. For exotic atoms involving hadrons the theoretical interpretation must consider hadron structure, just as for electronic atoms. The field of exotic atom spectroscopy is relatively new and can be expected to develop further and to continue to contribute to our knowledge of fundamental interactions and fundamental constants.

References

[1] M. Deutsch, Phys Rev. 82, 455 (1951); M Deutsch, Prog. Nucl. Phys. 3, 131 (1953).
[2] A. Rich, Rev. Mod. Phys. 53, 127 (1981).
[3] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Plenum Press, New York, 1977).
[4] G. P. Lepage and D. R. Yennie, these proceedings.
[5] T. Fulton and P. C. Martin, Phys. Rev. 95, 811 (1954).
[6] M. Deutsch and S. C. Brown, Phys. Rev. 85, 1047 (1952).
[7] E. D. Theriot, Jr., R. H. Beers, V. W. Hughes, and K. O. H. Ziock, Phys. Rev. A 2, 707 (1970).
[8] E. R. Carlson, V. W. Hughes, and I. Lindgren, Phys. Rev. A 15, 241 (1977); P. O. Egan, V. W. Hughes, and M. H. Yam, Phys. Rev. A 15, 251 (1977).
[9] A. P. Mills and G. H. Bearman, Phys. Rev. Lett. 34, 246 (1975).
[10] V. W. Hughes, in *Exotic Atoms '79*, Ed. by K. Crowe, J. Duclos, G. Fiorentini, and G. Torelli (Plenum Press, New York, 1980), p. 19.
[11] A. Rich, Phys. Rev. A 23, 2747 (1981). See also A. P. Mills, Jr., Phys. Rev. A 27, 262 (1983).

[12] K. F. Canter, A. P. Mills, Jr., and S. Berko, Phys. Rev. Lett. 34, 177 (1975).
[13] A. P. Mills, Jr., S. Berko, and K. F. Canter, Phys. Rev. Lett. 34, 1541 (1975).
[14] A. P. Mills, Jr., S. Berko, and K. F. Canter, in *Atomic Physics 5*, Ed. by R. Marrus, M. Prior, and H. Shugart (Plenum Press, New York, 1977) p. 103.
[15] T. Fulton, Phys. Rev. A 26, 1794 (1982).
[16] P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt, Phys. Lett. 81**A**, 119 (1981).
[17] T. W. Hänsch, S. A. Lee, R. Wallenstein, and C. Wieman, Phys. Rev. Lett. 34, 307 (1975); C. Wieman and T. W. Hänsch, Phys. Rev. A 22, 192 (1980).
[18] A. Gold, in *Quantum Optics*, Ed. by R. J. Glauber (Academic Press, New York, 1969), p. 397.
[19] V. W. Hughes, D. W. McColm, K. Ziock, and R. Prepost, Phys. Rev. Lett. 5, 63 (1960).
[20] V. W. Hughes and T. Kinoshita, in *Muon Physics*, Vol. I, Ed. by V. W. Hughes and C. S. Wu (Academic Press, New York, 1977), p. 11.
[21] V. W. Hughes, in *Exotic Atoms '79*, Ed. by K. Crowe, J. Duclos, G. Fiorentini, and G. Torelli (Plenum Press, New York, 1980), p. 3.
[22] F. G. Mariam, W. Beer, P. R. Bolton, P. O. Egan, C. J. Gardner, V. W. Hughes, D. C. Lu, P. A. Souder, H. Orth, J. Vetter, U. Moser, and G. zu Putlitz, Phys. Rev. Lett. 49, 993 (1982).
[23] E. Klempt, R. Schulze, H. Wolf, M. Camani, F. N. Gygax, W. Ruegg, A. Schenck, and H. Schilling, Phys. Rev. D 25, 652 (1982).
[24] H. W. Reist, D. E. Casperson, A. B. Denison, P. O. Egan, V. W. Hughes, F. G. Mariam, G. zu Putlitz, P. A. Souder, P. A. Thompson, and J. Vetter, Nucl. Instrum. Methods 153, 61 (1978); P. A. Thompson, V. W. Hughes, W. P. Lysenko, and H. F. Vogel, Nucl. Instrum. Methods 161, 391 (1979).
[25] D. E. Casperson, T. W. Crane, A. B. Denison, P. O. Egan, V. W. Hughes, F. G. Mariam, H. Orth, H. W. Reist, P. A. Souder, R. D. Stambaugh, P. A. Thompson, and G. zu Putlitz, Phys. Rev. Lett. 38, 956 (1977).
[26] "Muon Science and Facilities at Los Alamos," LA-9582-C (1982); LA-7490-MS (1978).
[27] S. R. Lundeen and F. M. Pipkin, Phys. Rev. Lett. 46, 232 (1981).
[28] P. R. Bolton, A. Badertscher, P. O. Egan, C. J. Gardner, M. Gladisch, V. W. Hughes, D. C. Lu, M. Ritter, P. A. Souder, J. Vetter, G. zu Putlitz, M. Eckhause, and J. Kane, Phys. Rev. Lett. 47, 1441 (1981).
[29] P. A. Souder, D. E. Casperson, V. W. Hughes, D. C. Lu, H. Orth, H. W. Reist, M. H. Yam, and G. zu Putlitz, Phys. Rev. Lett. 34, 1417 (1975); P. A. Souder, T. W. Crane, V. W. Hughes, D. C. Lu, H. Orth, H. W. Reist, M. H. Yam, and G. zu Putlitz, Phys. Rev. A 22, 33 (1980).
[30] H. Orth, K.-P. Arnold, P. O. Egan, M. Gladisch, W. Jacobs, J. Vetter, W. Wahl, M. Wigand, V. W. Hughes, and G. zu Putlitz, Phys. Rev. Lett. 45, 1483 (1980).
[31] C. J. Gardner, A. Badertscher, W. Beer, P. R. Bolton, P. O. Egan, M. Gladisch, M. Greene, V. W. Hughes, D. C. Lu, F. G. Mariam, P. A. Souder, H. Orth, J. Vetter, and G. zu Putlitz, Phys. Rev. Lett. 48, 1168 (1982).
[32] K.-N. Huang and V. W. Hughes, Phys. Rev. A **20**, 706 (1979); Phys. Rev. A **21**, 1071 (E) (1980); Phys. Rev. A 26, 2330 (1982).
[33] S. D. Lakdawala and P. J. Mohr, Phys. Rev. A **22**, 1572 (1980).
[34] E. Borie, Z. Physik A 291, 107 (1979).
[35] R. J. Drachman, Phys. Rev. A 22, 1755 (1980); J. Phys. B: At. Mol. Phys. **14**, 2733 (1981).
[36] G. Carboni, G. Gorini, E. Iacopini, L. Palffy, F. Palmonari, G. Torelli, and E. Zavattini, Phys. Lett. 73**B**, 229 (1978); G. Carboni, G. Gorini, G. Torelli, L. Palffy, F. Palmonari, and E. Zavattini, Nucl. Phys. A 278, 381 (1977); G. Carboni, U. Gastaldi, G. Neri, O. Pitzurra, E. Polacco, G. Torelli, A. Bertin, G. Gorini, A. Placci, E. Zavattini, A. Vitale, J. Duclos, and J. Picard, Nuovo Cimento A 34, 493 (1976).
[37] E. Borie and G. A. Rinker, Phys. Rev. A **18**, 324 (1978).
[38] E. Zavattini, A. M. Sachs, J. Fox, and R. Cohen, "An Improved Test of QED—An Experiment to Measure Vacuum Polarization in the 3d-3p Transitions in Muonic Helium," AGS Proposal #745, September, 1979.
[39] H. Anderhub, F. Kottmann, H. Hofer, P. LeCoultre, D. Makowiecki, O. Pitzurra, B. Sapp, P. G. Seiler, P. Schrager, M. Wälchli, and P. Wolff, Phys. Lett. **60B**, 273 (1976); F. Kottmann, in *Exotic Atoms*, Ed. by G. Fiorentini and G. Torelli (Erice School, Frascatti, 1977), p. 129.
[40] H. Anderhub, H. Hofer, F. Kottmann, P. LeCoultre, D. Makowiecki, O. Pitzurra, B. Sapp, P. G. Seiler, M. Wälchli, D. Taqqu, P. Truttmann, A. Zehnder, and Ch. Tschalär, Phys. Lett. 71**B**, 443 (1977).
[41] P. O. Egan, S. Dhawan, V. W. Hughes, D. C. Lu, F. G. Mariam, P. A. Souder, J. Vetter, G. zu Putlitz, P. A. Thompson, and A. B. Denison, Phys. Rev. A 23, 1152 (1981).
[42] B. Aas, W. Beer, I. Beltrami, P. Ebersold, R. Eichler, Th. v. Ledebur, H. J. Leisi, W. Ruckstuhl, W. W. Sapp, A. Vacchi, J. Kern, J.-A. Pinston, and R. Weber, Nucl. Phys. A 375, 405 (1982); B. Aas, W. Beer, I. Beltrami, K. Bongardt, P. Ebersold, R. Eichler, Th. v. Ledebur, H. J. Leisi, W. W. Sapp, J.-A. Pinston, J. Kern, R. Lanners, and W. Schwitz, Nucl. Phys. A 329, 450 (1979); B. Aas, R. Eichler, and H. J. Leisi, Nucl. Phys. A 375, 439 (1982). See also recent measurement on μ^{-12}C in W. Ruckstuhl, B. Aas, W. Beer, I. Beltrami, F. W. N. de Boer, K. Bos, P. F. A. Goudsmit, U. Kiebele, H. J. Leisi, G. Strassner, A. Vacchi, and R. Weber, Phys. Rev. Lett. 49, 859 (1982).
[43] D. C. Lu, L. Delker, G. Dugan, C. S. Wu, A. J. Caffrey, Y. T. Cheng, and Y. K. Lee, Phys. Rev. Lett. 45, 1066 (1980).
[44] S. C. Cheng, Y. Asano, M. Y. Chen, G. Dugan, E. Hu, L. Lidofsky, W. Patton, C. S. Wu, V. W. Hughes, and D. C. Lu, Nucl. Phys. A 254, 381 (1975).
[45] E. Hu, Y. Asano, M. Y. Chen, S. C. Cheng, G. Dugan, L. Lidofsky, W. Patton, C. S. Wu, V. W. Hughes, and D. C. Lu, Nucl. Phys. A 254, 403 (1975).
[46] G. Dugan, Y. Asano, M. Y. Chen, S. C. Cheng, E. Hu, L. Lidofsky, W. Patton, C. S. Wu, V. W. Hughes, and D. C. Lu, Nucl. Phys. A 254, 396 (1975).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Fundamental Tests and Measures of the Structure of Matter at Short Distances*

Stanley J. Brodsky

Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94305

Recent progress in gauge field theories has led to a new perspective on the structure of matter and basic interactions at short distances. It is clear that at very high energies, quantum electrodynamics, together with the weak and strong interactions, are part of a unified theory with new fundamental constants, new symmetries, and new conservation laws. A nontechnical introduction to these topics is given, with emphasis on fundamental tests and measurements.

Key words: quantum chromodynamics; quantum electrodynamics; strong interactions; weak interactions.

1. Introduction

In the past few years, there has been extraordinary progress in the understanding of the structure and interactions of matter at short distances. The most important theoretical progress has been in the area of non-Abelian gauge theories, which are now leading toward a unified description of the weak, strong, and electrodynamic interactions. It is now evident that quantum electrodynamics is just one manifestation of a larger unified theory. In this review, I will emphasize the areas where fundamental tests and precision measurements are crucial to the development of basic theory. Many previously believed conservation laws, such as baryon and lepton number conservation, are now open to question or in fact are predicted to be violated. High precision tests, including searches for very rare processes, are thus essential in order to definitively test the theories.

2. Quantum Electrodynamics

One of the most fundamental questions in physics is whether we have actually identified the fundamental constituents of matter [1]. In quantum electrodynamics, the leptons e, μ, and τ are elementary point-like carriers of the electromagnetic current, each with a Dirac coupling to the electromagnetic field. High energy ($\sqrt{s} = E_{\text{c.m.}} \leq 38$ GeV) measurements of the reactions $e^+e^- \to e^+e^-$, $\mu^+\mu^-$, $\tau^+\tau^-$, and $\gamma\gamma$ at the PETRA storage ring have placed severe limits on any deviation from the predicted Dirac structure or any internal lepton structure. For example, measurements of electron-positron annihilation into muon pairs by the Mark-J collaboration [2] lead to lower limits (95 percent confidence level)

$$\Lambda_- > 160 \text{ GeV}, \qquad \Lambda_+ > 120 \text{ GeV} \tag{1}$$

for modifications $1/Q^2 \to 1/(Q^2 \pm 1/Q^2 - \Lambda_\pm^2)$ of the photon propagator or the electron or muon vertex. Alternatively, this result demonstrates that the electron and muon are effectively point-like down to distances $R \sim \Lambda^{-1} \lesssim 2 \times 10^{-16}$ cm.

*Work supported by the Department of Energy under contract number DE-AC03-76SF00515.

Surprisingly, the strongest limits on possible internal lepton structure in some models come from the precise measurements of the gyromagnetic ratios of the electron and muon—measurements at the limit of zero momentum transfer. The most precise published value is [3]

$$\frac{1}{2} g_{e^-}^{\text{exp}} = 1.001\ 159\ 652\ 200\ (40)\,.$$

The QED prediction has now been computed through order $(\alpha/\pi)^4$ [ninth order in perturbation theory!] by Kinoshita and Lindquist [4]. The result is $g_e^{\text{th}}/2 = 1.001\ 159\ 652\ 504\ (182)$. The $(\alpha/\pi)^4$ calculations require the evaluation of 891 Feynman diagrams. The uncertainty reflects the limit of error on 10-dimensional numerical integrations, as well as the uncertainties in the determination of α. Since there is no *a priori* reason why a spin 1/2 system must have $g \sim 2$ in general (witness the nucleons), it is extraordinary that QED correctly predicts g_e to 10 significant figures!

Let us now consider the possibility that the electron is composite with an intrinsic radius R [1, 5]. The natural size of the magnetic moment of a charged extended system is $\mu \sim eR$, which would imply a contribution to the gyromagnetic ratio $\Delta g_e \sim m_e R$ (barring cancellations). Alternatively, one can compute the magnetic moment of any system from the general relationship between the anomalous moment $a = (1/2)(g-2)$ of a system and its excitation spectrum (the Drell-Hearn-Gerasimov sum rule). For a spin 1/2 system one has

$$a^2 = \frac{m^2}{2\pi^2\alpha} \int_{s_{\text{th}}}^{\infty} \frac{ds}{s} \left[\sigma_P^\gamma(s) - \sigma_A^\gamma(s)\right] \tag{2}$$

where $\Delta\sigma = \sigma_P^\gamma - \sigma_A^\gamma$ is the difference between the spin parallel and spin antiparallel total photoabsorption cross sections. Barring special cancellations, $\Delta\sigma \sim O(\alpha R^2)$ at energies where the compositeness of the target is manifest. For a composite electron, this again gives the estimate $\Delta a_e \sim m_e R$. The agreement between theory and experiment $|\Delta a_e| < 10^{-9}$ then implies that any internal size scale of the electron is limited to exceedingly small distances $R \lesssim 10^{-20}$ cm, $R^{-1} \gtrsim 10^6$ GeV. (The limits for the muon are comparable.)

It is, however, possible to construct specific models for composite leptons that give a smaller correction to the magnetic moment than the general estimate given above. For example, in a model in which the constituent fermions have mass m_F much less than the intrinsic momentum scale R^{-1} of the system, one obtains the quadratic relationship $\Delta a \sim m_e m_F R^2$. In fact, one can evidently conceive of the electron as a tightly bound composite system (radius $\lesssim 10^{-16}$ cm) of permanently confined but light mass ($m_F \lesssim 100$ MeV) fermion constituents, without violating any high energy or low energy constraint. The dynamics of such models are, however, far from clear. A light mass fermion constituent also seems to be required to understand the small mass of the electron. It is also interesting to note that a very complex theory can appear to be simple and renormalizable at low momentum scales $Q^2 << R^{-2}$, even though the particles are composite at short distances.

3. Weak Interactions and Quantum Electrodynamics

Before the advent of the Glashow, Weinberg, Salam SU(2) × U(1) theory [7] of the weak and electrodynamic interactions, there was no satisfactory way of computing the weak interaction corrections to QED predictions. Previous models, besides violating unitarity at high energies, gave logarithmically divergent corrections to the lepton magnetic moments and even quadratically divergent contributions to neutrino charge radii. It is, however, now clear that QED and the weak interactions are unified as part of a more general, completely calculable, renormalizable theory. In particular, the weak interaction contributions to the muon moment in the SU(2) × U(1) theory, as shown in Fig. 1, are readily calculable [9]:

$$\Delta a_\mu^{wk} \sim \alpha m_\mu^2 / M_W^2 \sim G_F m_\mu^2 \sim 2 \times 10^{-9}$$

compared with the present experimental uncertainty of $\sim 11 \times 10^{-9}$.

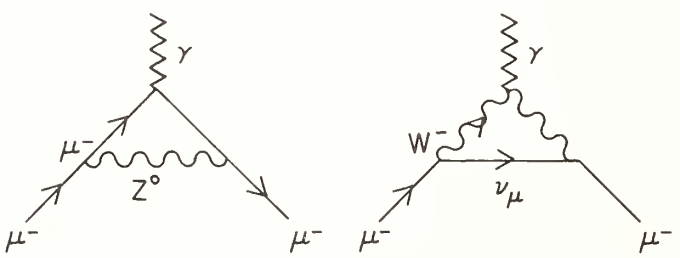


FIGURE 1. *Weak interaction contributions to the muon magnetic moment.*

Let us briefly review the main features of the GWS "standard model" [7]: In its initial stage, the theory begins with the assumption of an exact internal symmetry SU(2) analogous to isospin, with doublets of massless (negative helicity or "left-handed") leptons and quarks, and triplets of massless vector bosons; e.g.:

$$\psi_\ell \sim \begin{bmatrix} \nu_e \\ e^- \end{bmatrix}_L, \quad \psi_q \sim \begin{bmatrix} u \\ d \end{bmatrix}_L, \quad \mathbf{W} \sim \begin{bmatrix} W^+ \\ W^0 \\ W^- \end{bmatrix}. \qquad (3)$$

The interactions of this theory (see Fig. 2) generalize the

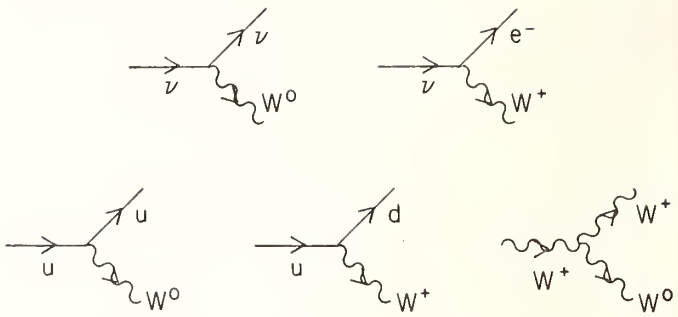


FIGURE 2. *SU(2) couplings of leptons and quarks.*

Dirac coupling of QED and preserve the SU(2) rotational invariance:

$$g_2 \bar{\psi}_\ell \mathbf{W} \psi_\ell + g_2 \bar{\psi}_q \mathbf{W} \psi_q = \cdots \qquad (4)$$

where $\mathbf{W}^\mu = \sum_{i=1}^{3} W_i^\mu \mathbf{T}_i$ and the $\mathbf{T}_i$ are the set of 2×2 traceless matrices. In fact, the entire theory, including the W self couplings, is invariant under rotations $\psi \rightarrow e^{i\Lambda(x)}\psi$, where $\Lambda(x)$ is an arbitrary 2×2 matrix function of space and time. This ("non-Abelian") local gauge invariance is a generalization of the (Abelian) local phase or U(1) gauge invariance of QED, and insures the renormalizability of the theory. One can also define an additional conserved "charge" Y (= −1/2 for leptons, 1/6 for quarks), which reflects the fact that the lepton and quark currents are separately conserved. The coupling of the weak hypercharge current $g_1 Y \bar{\psi} \gamma^\mu \psi$ to an additional zero mass vector boson B_μ then has an exact $\psi \rightarrow e^{i\Lambda(x)Y}\psi$ local U(1) gauge symmetry.

The theory discussed thus far has little resemblance to the observed weak interactions. However, if one introduces an extra doublet of interacting scalar bosons $\phi = \begin{bmatrix} \phi^+ \\ \phi^o \end{bmatrix}$ with non-zero expectation value in the ground state, then the exact SU(2) gauge invariance of the theory will be "broken," while still retaining the renormalizability of the theory in the ultraviolet domain. What emerges at low energies is (1) a massive charged vector boson $W^\pm$, (2) a massive neutral vector boson $Z_\mu^0 = \sin\theta_W B_\mu - \cos\theta_W W_\mu^0$, and (3) a massless neutral vector boson $A_\mu = \sin\theta_W B_\mu + \cos\theta_W W^0$ coupled to the electrodynamic current. One can then identify the electric charge $e = g_2 \sin\theta_W$ and the Fermi constant $G_F/\sqrt{2} = e^2/8 M_W^2 \sin^2\theta_W$, where the weak mixing angle is given by $\tan\theta_W = g_1/g_2$. The weak interactions at low momentum transfer $Q^2 << M_W^2$ then have the form:

$$\mathcal{L}_{\text{eff}}^{wk} = \frac{G_F}{\sqrt{2}} J_\mu^{\dagger Z} J_Z^\mu + \frac{G_F}{2\sqrt{2}} J_W^{\mu\dagger} J_\mu^W \qquad (5)$$

with

$$J_W^\mu = \sum_{q,\ell} \bar{\psi} \left[\gamma^\mu (1 + \gamma_5) T^+\right] \psi \qquad (6)$$

and

$$J_Z^\mu = \sum_{q,\ell} \bar{\psi} \left[\gamma^\mu (1 + \gamma_5) T^3 - 2q \sin^2\theta_W\right] \psi \qquad (7)$$

The form and relative normalization of the charged and neutral weak interactions predicted by Eq. (7) have now

been checked in many neutrino and weak/electromagnetic interference experiments—in many cases to better than 1 percent precision. One of the most precise experiments is the SLAC-Yale [10] measurement of parity violation in deep inelastic polarized electron e^-p scattering. The interference between the electromagnetic and weak neutral currents (see Fig. 3) leads to an asymmetry $(\sigma_L - \sigma_R)/(\sigma_L + \sigma_R)$ in the $e^-p \to e^-X$ cross section and thus to a determination of $\sin^2\theta_W$. A combined analysis of the various neutral current experiments gives (see Fig. 4) [11]:

$$\sin^2\theta_W = 0.229 \pm 0.009(\pm 0.005) \tag{8}$$

and

$$\kappa = \frac{\left[J_\mu^Z \; J_Z^\mu\right]_{\text{expt}}}{\left[J_\mu^Z \; J_Z^\mu\right]_{\text{predicted}}} = 0.992 \pm 0.017(\pm 0.011) \tag{9}$$

where the error in parentheses indicates errors due to theoretical uncertainties (radiative corrections, etc.).

Thus far, there is no discrepancy with the predictions of the SU(2) × U(1) model, although other models (such as those that are parity symmetric at large momentum transfer) are not ruled out. The most critical test of this model will be the experimental confirmation of the $W^\pm$ and Z^0 vector bosons at the predicted mass

$$M_W = 82.0 \pm 2.4 \text{ GeV}$$

$$M_Z = 93.0 \pm 2.0 \text{ GeV}. \tag{10}$$

The width of the Z^0 is particularly interesting since it signals the number of neutrinos. Confirmation of SU(2) × U(1) model will also require the identification of the Higgs scalars, the origin of symmetry-breaking in the theory. It is, however, possible that these particles are themselves composites [12] rather than new additional elementary-field degrees of freedom.

The fermions in the higher generations enter the SU(2) × U(1) multiplets in parallel to Eq. (2); i.e.:

$$\psi_\ell = \begin{bmatrix} \nu_e \\ e^- \end{bmatrix}_L, \quad \begin{bmatrix} \nu_\mu \\ \mu^- \end{bmatrix}_L, \quad \begin{bmatrix} \nu_\tau \\ \tau^- \end{bmatrix}_L$$

$$\psi_q = \begin{bmatrix} u \\ d \end{bmatrix}_L, \quad \begin{bmatrix} c \\ s \end{bmatrix}_L, \quad \begin{bmatrix} t \\ b \end{bmatrix}_L. \tag{11}$$

However, as first noted by Cabibbo, the quarks that appear in the weak interaction theory are not necessarily the mass eigenstates of the full theory, including strong interactions. In the case of three generations, the mass eigenstates

$$q_m = \begin{bmatrix} A_{mn} \; u_n \\ d_m \end{bmatrix}_L \tag{12}$$

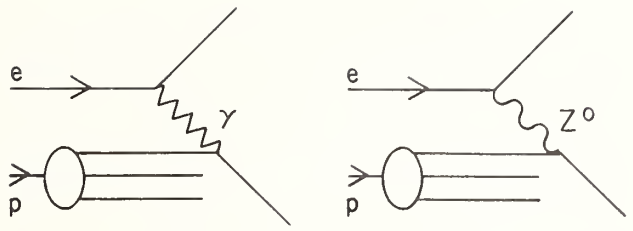


FIGURE 3. *Interfering weak and electromagnetic contributions to deep inelastic lepton-proton scattering.*

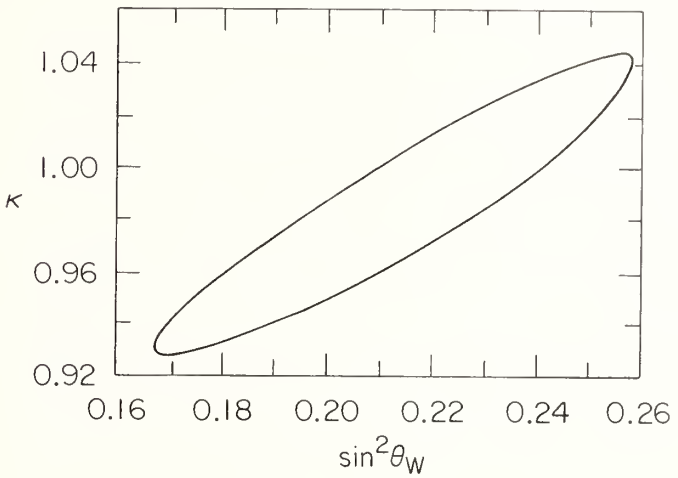


FIGURE 4. *Values of* $\sin^2\theta_W$ *and* κ *from a phenomenological analysis of neutral current data. The limit curve corresponds to a confidence level of 68%. From Ref. [11].*

are related by three rotation angles and one CP violating phase ϕ to the weak interaction eigenstates. The angle of the mixing between the first two generation quarks is the Cabibbo angle, which has been determined to considerable accuracy: $\theta_c \sim 13.17° \pm 0.64°$ from analyses of strangeness-changing weak interactions.

The fact that the mixing of three generations leads in a natural way to CP and T violation is a very interesting result [13]. By parameterizing present data for CP violating effects in terms of the phase angle ϕ, one can predict a non-zero value for the neutron electric dipole moment:

$$(D/e)_n \simeq 10^{-30} \text{ cm}$$

compared with the present limit $\sim 10^{-26}$ cm. If the e, μ, and τ neutrinos are massive, then one expects a similar mixing pattern for the lepton sector and a non-zero value for the electron electric dipole moment [9].

4. The Strong Interactions [14]

The successful application of local gauge theories to weak and electrodynamic interactions has led to an even more fundamental advance in the case of the strong interactions—the development of quantum chromodynamics. In QCD the fundamental degrees of freedom of hadrons and their interactions are the quanta of quark and gluon fields that obey an exact internal SU(3) "color symmetry." Each quark "flavor" $q = u, d, s, c, b \ldots$ is represented as a color triplet

$$\psi = \begin{bmatrix} q_R \\ q_Y \\ q_B \end{bmatrix}$$

interacting with an octet of gluon fields $\vec{G}^{\mu} = \sum_{a=1}^{8} G_a^{\mu} \vec{\lambda}_a$, where the $\vec{\lambda}_a$ are the set of 3×3 traceless matrices. The interactions

$$g_3 \bar{\psi} \vec{G} \psi$$

together with the gluon self couplings (see Fig. 5) preserve an exact local gauge symmetry for arbitrary rotations $\psi \rightarrow e^{i\vec{\lambda}(x)}\psi$ in the SU(3) color space. The gluons are massless and theory is renormalizable. The lowest energy states are the color-singlet baryons $|q_R q_Y q_B>$ and mesons $(1/\sqrt{3})|q_R\bar{q}_R + q_B\bar{q}_B + q_Y\bar{q}_Y>$. In addition, bound states of gluons $|gg>$ and $|ggg>$, as well as $|q\bar{q}g>$, states are predicted.

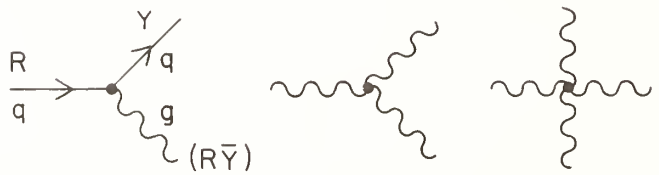


FIGURE 5. *Quark-gluon couplings in SU(3) color.*

It now seems possible that quantum chromodynamics is the *the* theory of the strong interactions in the same sense that quantum electrodynamics accounts for electromagnetic interactions. It is well known that the general structure of QCD meshes remarkably well with the facts of the hadronic world, especially quark-based spectroscopy, current algebra, the approximate point-like structure of large momentum transfer lepton-hadron reactions, and the logarithmic violation of scale-invariance in deep-inelastic reactions. The theory is particularly successful in predicting the features of electron-positron annihilation into hadrons: the magnitude and scaling of the total cross section, the production of hadronic jets with a pattern conforming to elementary quark and gluon processes, and heavy quark phenomena. The empirical results are consistent with the basic postulates of QCD, that the charge and weak currents within hadrons are carried by the quarks and that the strength of the quark-gluon couplings becomes weak at short distances (asymptotic freedom).

Although it is simplest to define the coupling constant in QED at zero momentum transfer, this choice is in a sense arbitrary since the one-photon exchange interactions can be computed at any Q^2 by including the vacuum polarization insertions to all orders:

$$\alpha(Q^2) = \frac{\alpha(Q_0^2)}{1 + \alpha(Q_0^2)\left[\pi(Q^2) - \pi(Q_0^2)\right]}. \qquad (13)$$

For $Q^2 >> m_e^2$, $\pi(Q^2) \sim -(1/3\pi) \log Q^2/m_e^2$ so the effective coupling $\alpha(Q^2)$ increases at large Q^2. In the case of QCD, the vacuum polarization diagrams involving gluon self-coupling (see Fig. 6) actually reverse the sign of $\pi(Q^2)$:

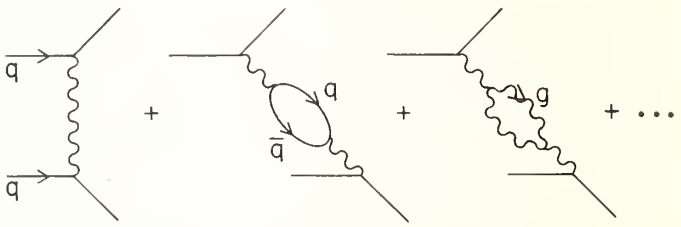


FIGURE 6. *Vacuum-polarization contributions to quark-quark scattering.*

$$\pi_{QCD}(Q^2) \simeq \frac{1}{6\pi}\left[\frac{33}{2} - n_f\right] \log Q^2 \qquad (14)$$

as long as the number of quark flavors (n_f) is less than 17. Thus, for very large Q^2, the strong interaction coupling constant $\alpha_s = g_3^2/4\pi$ is given by ($Q^2 \gg \Lambda^2$)

$$\alpha_s(Q^2) = \frac{4\pi}{\left[11 - \frac{2}{3}n_f\right] \log \frac{Q^2}{\Lambda^2}}. \qquad (15)$$

The fact that the coupling constant decreases at large Q^2 (see Fig. 7) implies that we can use perturbation theory for calculating hadronic processes at short distance. Con-

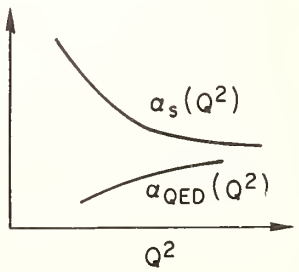


FIGURE 7. *Variation of the strong and electromagnetic coupling strengths with momentum transfer.*

versely, the increase of the effective coupling at large distances is consistent with the expectation that the theory strongly confines particles of non-zero color.

The fundamental parameters of QCD are the coupling constant α_s and the quark masses. The actual measurement of these quantities is complicated by the fact that only bound states of quarks and gluons are accessible, and there can be large and uncertain corrections from nonperturbative, binding, and other higher-order effects. In addition, the specification of $\alpha_s(Q^2)$ depends on the momentum scale chosen to express the leading order results and the choice of normalization scheme.

At this time, the most precise determination of the QCD coupling constant is given by an analysis of the decay of the upsilon, Y, the lowest energy bound state of b and b^- quarks with $J^{PC} = 1^{--}$ at $M_\Upsilon = 9.46 \pm 0.01$ GeV. The hadronic width of the upsilon can be computed in lowest order perturbation theory from the 3 gluon decay amplitude (see Fig. 8) in analogy to orthopositronium decay into three photons. The QCD prediction, including

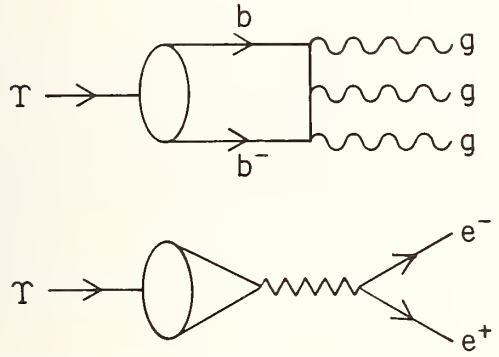


FIGURE 8. *Leading contributions to the hadronic and leptonic decays of the upsilon.*

first order radiative corrections as recently computed by Lepage and Mackenzie [15] is

$$\Gamma_{\Upsilon \to 3g} = |\psi_{NR}(0)|^2 \frac{\alpha_s^3(M_\Upsilon^2)}{M_\Upsilon^2} (\pi^2 - 9) \frac{160}{81} \left[1 + C_{3g} \frac{\alpha_s}{\pi}\right] \tag{16}$$

with

$$C_{3g} = 3.8 \pm 0.5 .$$

The lepton decay rate is

$$\Gamma_\Upsilon \to e^+e^- = |\psi_{NR}(0)|^2 \frac{16\pi \left[\frac{1}{3}\right]^2 \alpha^2}{M_\Upsilon^2} \left[1 - \frac{16}{3} \frac{\alpha_s}{\pi}\right] \tag{17}$$

The measured ratio for $\Gamma_{\Upsilon \to \text{had}}/\Gamma_{\Upsilon \to e^+e^-}$ then implies $\alpha_s(Q^2 = M_\Upsilon^2) = 0.14 \pm 0.01$, where the error indicates 1σ accuracy. The determination [2] of α_s from $e^+e^- \to q\bar{q}g \to 3$ jets also gives values in the range $\alpha_s(Q^2 \sim 1000\ \text{GeV}^2) \sim 0.15$ to 0.23, although there are large corrections from hadronization and higher order effects. It is also possible, in principle, to determine α_s from exclusive processes [16] such as ratios of meson form factors: $F_\pi(Q^2)/F^2_{\pi\gamma}(Q^2)$ and from radiative corrections to deep-inelastic lepton-hadron scattering.

Considering that new phenomena still continue to be discovered in QED and atomic physics, it is likely that we have only touched the surface of much more complex QCD phenomena. Thus far, there have been many semi-quantitative tests of the theory, including the verification of scale-invariant quark-quark interactions from the behavior of meson and baryon form factors at large momentum transfer. There is, however, only a rough qualitative understanding of the basic properties of hadrons, such as their masses, magnetic moments, and charge radii. Detailed answers to such questions require an understanding of the hadronic wavefunctions as well as their multiparticle Fock state structure (see Ref. [5]). Perhaps the most dramatic confirmation of QCD would be the observation of gluonium $|gg>$ and $|ggg>$ states. It should also be emphasized that the reported observation of fractional charge systems by Fairbank *et al.* [17] reopens the fundamental question of whether quarks are really confined in QCD—i.e., whether QCD is an exact local gauge theory or is spontaneously broken so that the gluons have a finite mass and the confinement potential has finite range. Alternatively, these results could signal the existence of color singlet fractional charged particles, or bound states of zero charge color anti-triplets $\bar{3}_c$ with quarks, or even fractionally charged subconstituents of the quarks themselves.

5. Unified Theories of the Strong, Weak, and Electromagnetic Interactions [1, 7]

Thus far, our discussion has led to no real understanding of the quark $\longleftrightarrow$ lepton parallelism, especially why the proton and electron charges are equal and opposite to within parts in 10^{20}. This could be understood if (a) quarks and leptons have common subconstituents and/or (b) quarks and leptons are in the same representation of a fundamental symmetry group. The latter possibility is the central motivation for many grand unified theories, as exemplified by the SU(5) gauge theory model of Georgi and Glashow [18]. In this model, the 15 fermions of the first generation (see Fig. 9) are identified members of the $\underline{5}$ and $\underline{10}$ representation of SU(5); e.g.:

$$\psi_{\underline{5}} \sim \begin{bmatrix} \bar{d}^R \\ \bar{d}^Y \\ \bar{d}^B \\ e^- \\ \nu_e \end{bmatrix}_L \tag{18}$$

and $\psi_{\underline{10}}$ is an antisymmetric 5×5 matrix with entries for ${e_L}^+$, $(u, d)_L$ and $(\bar{u})_L$. The gauge bosons correspond to the set of (24) 5×5 traceless matrices

$$\tilde{A} = \left[\begin{array}{c|cc} \text{gluons} & X^{R,Y,B} & Y^{R,T,B} \\ \hline \bar{X}^{R,Y,B} & Z^{0,\gamma} & W^+ \\ \bar{Y}^{R,Y,B} & W^- & Z^0 \end{array}\right] . \tag{19}$$

The X and Y vector bosons have color and fractional

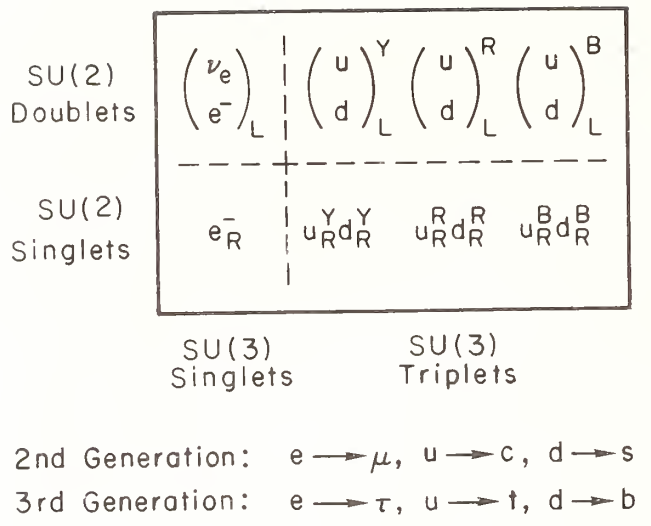


FIGURE 9. *The 15 fundamental fermions of the first generation. The subscript L and R indicates particle helicity.*

charge and couple of quarks to leptons! Since charge is a generator of SU(5), $\bar{\psi}Q\psi = tr\ Q = 0$ and the sum of the charges of the constituents in any representation must be zero—i.e., $3Q_d + Q_{e^+} = 0$—this naturally explains the equality of lepton and baryon charges.

The basic scenario of SU(5) is then as follows (see Fig. 10). At momentum transfers Q^2 much larger than M_X^2 and M_Y^2, all particles can be treated as massless and SU(5) is an exact gauge symmetry. At lower energies this symmetry is broken (by a Higgs scalar vacuum expectation value), leaving symmetries corresponding to the subgroups $SU(3)_C \times SU(2) \times U(1)$. Finally, at much lower energies, $Q^2 \gtrsim M_W^2$, $M_Z^2 \sim 10^2$ GeV^2, the theory is again broken (as in Sec. 3) and the exact gauge symmetries that remain are the $SU(3)_C$ (massless gluon octet) and U(1) symmetry of QED (massless photons). From analysis of observed mass scales and the fact that the U(1) coupling constant is increasing and the SU(2) and $SU(3)_C$ coupling constants are logarithmically decreasing, one can estimate that the grand unified scale (where the coupling constants coincide $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_5$) is $\sim 10^{15}$ GeV (see Fig. 11). (This also gives an estimate for M_X, M_Y.) In fact, $\alpha_{GUT}(Q^2 \sim 10^{15}\ \text{GeV}^2) \simeq 1/42$ now becomes the fundamental coupling constant, not $\alpha \simeq 1/137$.

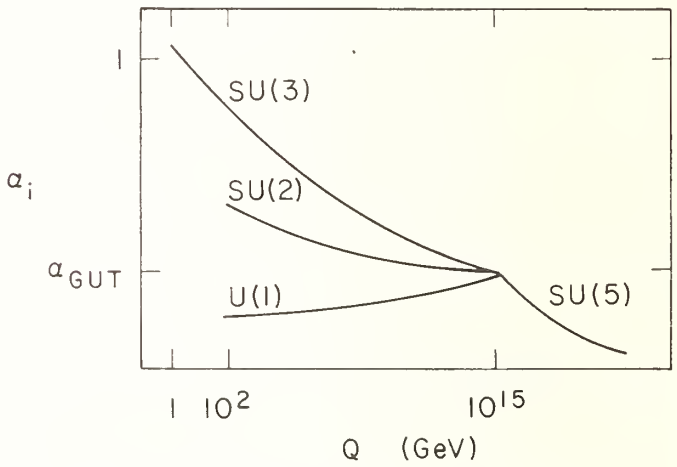


FIGURE 11. *Symmetry breaking pattern in SU(5).*

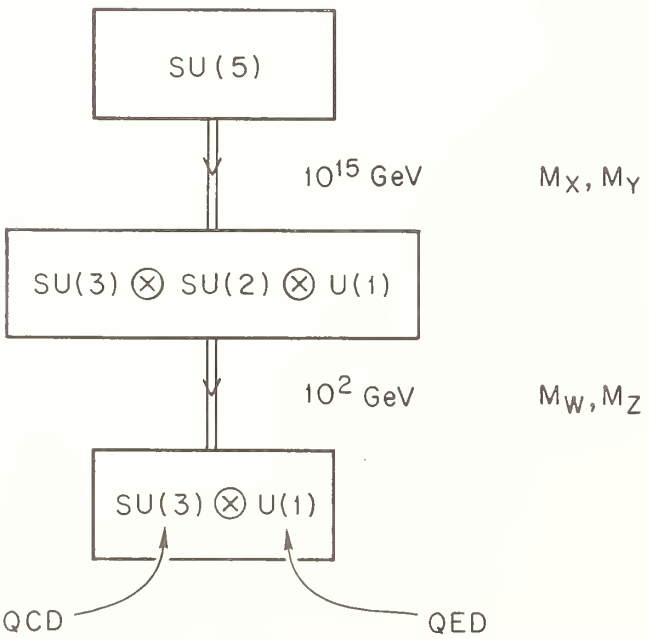


FIGURE 10. *The approach to equality of the SU(3), SU(2), and U(1) gauge couplings in the SU(5) grand unified theory.*

The SU(5) model leads to two critical predictions:

(1) The weak mixing angle is fixed to be $\sin^2\theta_W = 3/8$ at the grand unified scale. One can, however, use an extrapolation determined by the renormalization group (over 13 decades!) to compute $\sin^2\theta_W = 0.209 \pm 0.006$ at $Q^2 = M_W^2$. This is not far from the experimental value [10, 11] $\sin^2\theta_W = 0.229 \pm 0.009$ (± 0.005).

(2) Because of the X and Y couplings, baryon number is not conserved, although baryon number minus lepton number is still a conserved quantity in SU(5). The proton is thus unstable and will decay into channels such as $p \to e^+\pi^0$ (see Fig. 12). The predicted proton decay rate is proportional to the grand unified scale to the fourth power and thus has a large uncertainty: $\tau_p \sim 10^{31\pm2}$ years. The present experimental limit is $\tau_p \gtrsim 10^{30}$ years; in fact, four large-scale experiments now under construction will have a sensitivity to $\tau_p \sim 10^{32}$ years, so a decisive test of SU(5) appears possible. This is especially true since the most recent evaluation [15] of the QCD scale constant $\Lambda_{\overline{MS}} \sim 100$ MeV in Eq. (15) leads to the lowering of the grand unified scale mass and hence lower predicted values for τ_p.

In addition to SU(5), other grand unified theories have been proposed. All such models have profound implications for the evolution of the early universe [19], which

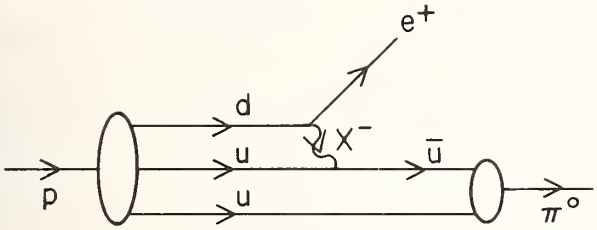


FIGURE 12. *Origin of baryon decay in SU(5).*

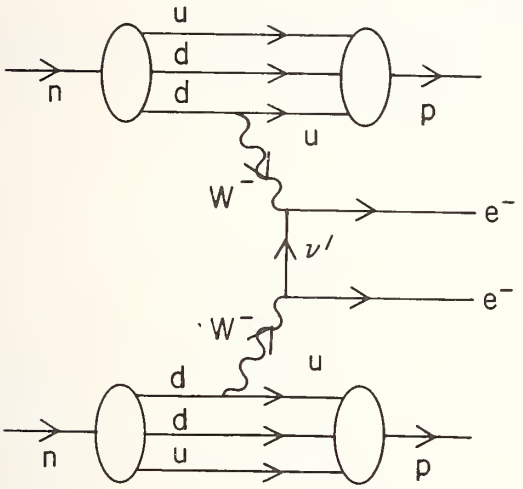


FIGURE 13. *Diagram leading to double* β*-decay in nuclei.*

would be expected to trace through the various symmetry phases, starting with a hot soup of massless quarks, gluons, W, X, etc. A possible connection of the baryon excess $n_B - n_{\bar{B}}$ in the universe to CP violation is also possible [20, 7, 11]. The grand unified models also evidently predict large rates for the production of free magnetic monopoles in the early universe.

In many grand unified theories, there is no natural or compelling reason why neutrinos should be massless, and the determination of the neutrino masses is now a topic of intense experimental activity [21]. In analogy to the quark case, the mass eigenstates ν_m may be linear combinations of ν_e, ν_μ, and ν_τ. This means that the neutrino ν_μ^L produced at time $t = 0$ in $\pi^+ \to \mu^+ \nu_\mu^L$ decay will at $t > 0$ be a linear combination of ν_μ^L, ν_μ^R (chirality nonconservation), and ν_e^L (neutrino "oscillations").

If a neutrino has mass, it also follows that it can have an anomalous magnetic moment [9]; diagrams such as Fig. 15 give contributions of order $a_\nu \sim m_\nu m_f G_F$. One can thus imagine inducing a neutrino helicity flip using external magnetic fields. If a neutrino is self-conjugate $C|\nu> = \pm|\nu>$ and there are charged right-handed currents so that $W^- \to e^- \nu_L$ and $W^- \to e^- \nu_R$ are possible, then double β-decay processes $nn \to pp\, e^- e^-$ in nuclei are possible (see Fig. 14). In some models $n \longleftrightarrow \bar{n}$ oscillations are also predicted.

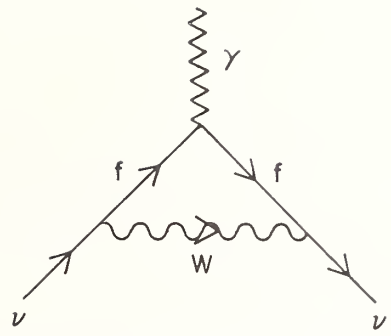


FIGURE 14. *Weak interaction contribution to the magnetic moment of a massive neutrino.*

6. Conclusions

The concept of a grand unified theory has led to an extraordinary new perspective on the structure of matter at short distances. It is clear that, at very high energies, QED, together with the weak and strong interactions, are part of a unified theory with new fundamental constants, new symmetries, and new conservation laws. The prediction that baryon and lepton numbers are no longer individually conserved should lead us to question all of presently accepted conservation laws and symmetries. This new gauge field theory perspective also makes it evident that the numerical value of parameters such as α_{QED} at zero momentum transfer is not significant in itself, considering the dependence of the coupling on resolution scale Q^2 and normalization scheme. Questions concerning the possible time variation of the fundamental constants such as α_{QED} are now seen to be connected with the evolution from the early universe.

Although gauge field theories and grand unified theories are leading to answers to many fundamental questions, one must acknowledge that countless others remain, such as the origin of a vast ratio of mass scales: $M_X/M_W \sim 10^{13}$, the origin of the fermion masses (e.g., why is $M_p < M_n$?), the possible compositeness of leptons and quarks, and the ultimate role of gravity ($M_{\text{Planck}} \sim 10^{19}$ GeV). It is clear that high precision experiments, searches for rare new processes, as well as detailed high energy measurements, will be crucial for continuing progress in these areas.

I wish to thank my colleagues at SLAC particularly Boris Kayser, for helpful discussions.

References

[1] For examples of fundamental theories with composite leptons and quarks, see J. G. Pati and A. Salam, Phys. Rev. D 10, 275 (1974); H. Harari, Phys. Lett. 86B, 83 (1979); and I. Bars, Yale preprint, YTP-81-19 (1981).

[2] D. P. Barber *et al.*, Phys Rev. Lett. 43 1915 (1979). For a review of the PETRA results, see G. Wolf, DESY 80/13 (1980).

[3] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Phys. Rev. Lett. 38, 310 (1977). For a review of anomalous moment measurements and theory, see J. H. Field, E. Picasso, and F. Combley, Sov. Phys. Usp. 22, 198 (1979).

[4] T. Kinoshita and W. B. Lindquist, these proceedings. For recent progress in the analytic evaluation of the anomalous moment through order $(\alpha/\pi)^3$, see M. J. Levine and R. Z. Roskies, these proceedings. A review of other precision measurements in QED is given by G. P. Lepage and D. R. Yennie, these proceedings.
[5] The analysis discussed here follows S. J. Brodsky and S. D. Drell, Phys. Rev. D 22, 2236 (1980). See also G. L. Shaw, D. Silverman, and R. Slansky, Phys. Lett. 94B, 57 (1980).
[6] S. Dimopoulos, S. Raby, and L. Susskind, Nucl. Phys. B 173, 208 (1980). I. Bars, Ref. [1].
[7] For reviews and references to electroweak theory, see P. Langacker, SLAC-PUB-2544 (1980); L.-F. Li, NSF-ITP-81-25 (1981); and the proceedings of the Cornell Z° Theory Workshop, CLNS-81-485 (1981).
[8] See, e.g., S. J. Brodsky and J. D. Sullivan, Phys. Rev. 156, 1644 (1967), and references therein.
[9] B. W. Lee and R. Strock, Phys. Rev. D 17, 2410 (1978) and references therein.
[10] C. Y. Prescott *et al.*, Phys. Lett. 84B, 524 (1979).
[11] J. Ellis, CERN TH.2942 (1980), and references therein.
[12] L. Susskind, Phys. Rev. D 20, 2619 (1979).
[13] M. Kobayashi and M. Maskawa, Prog. Theor. Phys. 49, 652 (1973).
[14] For recent reviews of QCD see, A. J. Buras, Rev. Mod. Phys. 52, 1 (1980), A. H. Mueller, CU-TP-192 (1981); E. Reya, DESY 79/88 (1979), S. J. Brodsky and G. P. Lepage, SLAC-PUB-2447 (1979).
[15] G. P. Lepage and P. Mackenzie, Cornell preprint CLNS/81-498 (1981).
[16] S. J. Brodsky and G. P. Lepage, Phys. Rev. D 22, 2157 (1980).
[17] C. S. Larue, J. D. Phillips, and W. M. Fairbank, Phys. Rev. Lett. 46, 967 (1981).
[18] H. Georgi and S. L. Glashow, Phys. Rev. D 6, 429 (1972).
[19] See, e.g., A. Guth and E. J. Weinberg, Phys. Rev. D 23, 876 (1981).
[20] A. D. Sakharov, JETP Lett. 5, 27 (1967).
[21] For reviews, see S. M. Bilensky and B. Pontecorvo, Phys. Rep. 41, 1 (1978); and B. Kayser, SLAC-PUB-2685 (1981), and references therein.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Experimental Limit for the Charge of the Free Neutron

R. Gähler,*+ J. Kalus,* and W. Mampe+

The neutron charge has been measured to be $q_n = |q_e|\,(-1.5 \pm 2.2) \times 10^{-20}$ (q_e = electron charge) at a confidence level of 90%. This value brings down the known limit by two orders of magnitude. In the experiment slow neutrons of 20 Å wavelength passed a strong electric field of 10 m length. The deflection of the neutron beam was measured with respect to reversal of the field. For an increase in sensitivity the beam was focused by a neutron lens to a sharp image in the detector plane. Over a long run time the deflection of the neutron beam due to the electric field was less than 2×10^{-2} μm. The result on the neutron charge is in agreement with the commonly accepted neutrality of the neutron.

Key words: neutron charge; neutron optics; particle properties.

1. Introduction

1.1 Background

Within the framework of modern unified theories even a tiny electric charge of the neutron q_n would lead to important consequences: assuming charge conservation as being valid, a finite neutron charge implies that a transition neutron-antineutron, as predicted by some theories [1] is not allowed; the electric charge has to change its sign in such a transition. Furthermore, if it were found that all baryons have charges slightly displaced from their usually accepted values by a common amount ϵ, then the conservation of baryons would follow from the conservation of electric charge rather than being an independent principle [2]. In this case the decay of the proton, predicted by several unified theories, is forbidden because the leptons into which it could decay carry a different charge.

With the present measurement, a new experimental limit for the charge of the electron-neutrino $q_{\bar{\nu}}$ can be set by regarding the β-decay of the neutron:

$$n \rightarrow p + e + \bar{\nu}.$$

The neutrality of the proton-electron system, i.e., the neutrality of the hydrogen atom has been measured to $(0.9 \pm 2.7) \times 10^{-21} \cdot q_e$ [3]. (q_e = charge of the electron.) Combined with the new value for the neutron charge, one gets:

$$q_{\bar{\nu}} = q_e(1.4 \pm 1.4) \times 10^{-20} \quad (68\% \text{ confidence}).$$

For these considerations charge conservation has been assumed to be valid.

Whereas newer theories assume the symmetry of the existing electric charges—this implies $q_n = 0$—there were some theories in the past predicting a finite neutron charge. The observed expansion of the universe would follow from a neutron charge of about $2 \times 10^{-18}\, q_e$ [4] and the earth's magnetic field was explained by a neutron charge of about 2×10^{-19} electron charges [5]. These theories can be refuted by our measurement.

1.2 Earlier Measurements

Most experimental limits on the neutron charge were derived in an indirect way by testing the charge neutrality of atoms or macroscopic bodies. For a review of these experiments see Dylla and King [3]. We have only to add more recent measurements, which dealt with the search for fractional charges in matter [6–9]. The sensitivity of all these measurements is in the region of 10^{-21} to 10^{-23} electron charges for the sum of $q_e + q_p + q_n$.

Direct measurements on the charge of the free neutron are all based on the deflection of a neutron beam in a strong homogeneous electric field. In the latest and most precise experiment of this type Shull *et al.* [10] got a limit on the neutron charge of $q_n = q_e(-1.9 \pm 3.7) \times 10^{-18}$. In that experiment not the lateral deflection of the neutron beam was sensed, but its angular deviation, using perfect Si-crystals in Bragg-condition. The authors also review all earlier charge measurements on the free neutron. It should be noted that the sensitivity of all direct measurements including ours is lower than that of indirect methods. But as mentioned by Shull *et al.* [10] "the possibility of a free charge being slightly different in magnitude, at this small level of charge difference when particles are amalgamated into an atom, does exist."

2. Measuring Principle

In the reported experiment a beam of slow neutrons (velocity $v = 200$ m/sec, wavelength $\lambda = 20$ Å) enters through a narrow slit S_3 into a strong transverse electric field of about 10 m length (see Fig. 1). A neutron charge q_n would lead to a deflection of the beam correlated to the direction of the periodically reversed electric field. The deflection y is given by:

$$y = \frac{q_n E_0 \ell^2}{4Q_0} \quad \text{(see section 3)} \qquad (1)$$

with E_0 the electric field applied over the length ℓ and Q_0 the kinetic energy of the neutron. After passage through the electric field the neutrons are counted behind slit S_4, which is positioned in the slope of the neutron beam. A change in count-rate for opposite field directions would indicate a beam deflection and hence a charge of the neutron. The sensitivity for measuring a

*Universität Bayreuth, 858 Bayreuth, Germany
+Institut Laue-Langevin, 38042 Grenoble, France

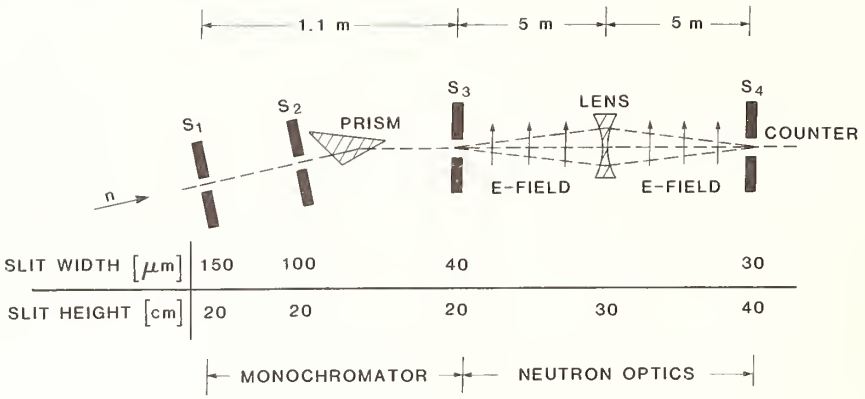


FIGURE 1. *Top view of the instrument (not to scale).*

deflection depends on the steepness of the slope and on the intensity. In order to obtain a narrow beam with high intensity, we studied different focusing systems and found a lens to be the most convenient device [11]. Similar to a lens in light optics, this neutron lens relies on the difference of the indices of refraction in the lens-material and in air. The severely chromatic lens (focal length $f \sim 1/\lambda^2$) asks for a fairly monochromatic beam. This is obtained by a prism-monochromator in front of slit S_3. The neutron lens is positioned 5 m downstream from the entrance slit S_3 which is 40 μm wide. The lens images the slit S_3 into a plane where slit S_4 is positioned. A detailed description of the optical system is given in a previous paper [11].

3. Sensitivity for the Charge Measurement

3.1 Deflection of a Charge q_n in the Electric Field

At first it will be shown that for our optical system the beam deflection is only half that without imaging by the lens. This result can be obtained by a wave optical calculation and even more simply from geometrical optics: A neutron starting parallel to the optical axis follows a parabola in the E-field leading to a deflection $y/4$ at the position of the lens. y is the total deflection in the case without the lens. Due to the focusing properties of the lens, the flight path is bent back parallel to the optical axis. The E-field downstream from the lens bends the flight path to an identical parabola as before, leading to an additional deflection $y/4$. The loss of a factor of 2 could be compensated by increasing the image-distance relative to the object-distance. But in this case the required precision for the lens—already a delicate problem—becomes even more severe.

In our case, the electric field E_0 is not applied over the full length ℓ_0 between S_3 and S_4—the necessary control of the optical components implied the need to divide the field into two halves, one between S_3 and the lens, the other between the lens and S_4. If the total length of the electric field is ℓ_1, it can easily be shown that the deflection y becomes:

$$y = \frac{q_n E_0 \ell_0 \ell_1}{8 Q_0} . \tag{2}$$

This is valid if the two electric fields are arranged symmetrically with respect to the optical elements. With the given parameters of the machine the deflection y for a hypothetical charge of $q_n = 1 \times 10^{-20}\, q_e$ becomes

$$y = 6.5 \times 10^{-9} \text{ m}$$
$$\left(E_0 = 11.9 \times 10^6 \frac{\text{V}}{\text{m}},\ \ell_0 = 10 \text{ m},\ \ell_1 = 9 \text{ m},\right.$$
$$\left. Q_0 = 2 \times 10^{-4} \text{ eV } (\lambda = 20 \text{ Å})\right)$$

3.2 Choice of the Instrumental Parameters

The parameters which determine the sensitivity of the apparatus for the charge measurement are the count rate N at slit S_4, the width of the image w_0, the wavelength λ (energy Q_0), the lengths ℓ_0 and ℓ_1, and the electric field E_0. The actual values of these parameters will be discussed now.

3.2.1 Count Rate N and Width of the Image w_0

The statistical error σ_N of the difference in count rate $N_1 - N_2$ for the two different directions of E_0, is given by:

$$\sigma_N = \sqrt{2N} \tag{3}$$

for $N_1 \approx N_2 = N$ and negligible background.

The measurable shift σ_y of the image is given by:

$$\sigma_y = \frac{\sigma_N}{dN/dy} \tag{4}$$

where dN is the change in count rate per shift dy;

$$\frac{dN}{dy} = N \frac{df}{dy} \tag{5}$$

with df/dy being the steepness of the slope of the image at the position of slit S_4. It follows that

$$\sigma_y = \frac{\sqrt{2}}{\sqrt{N}\ \dfrac{df}{dy}} . \tag{6}$$

In first approximation the image has Gaussian shape. At the steepest point of the slope $1/\frac{df}{dy}$ is proportional to w_0, the width of the image.

$$\sigma_y \sim \frac{w_0}{\sqrt{N}} . \tag{7}$$

In order to minimize σ_y, a small image w_0 with a high count rate N is desired. The count rate N increases with the width of the wavelength band and with the beam divergence. However, imperfections and aberrations of the focusing lens limit the achievable σ_y. A good compromise between high count rate N and small image width w_0 is obtained using the following argument: the

parallelism of the four axes of the lens system (see Ref. [11]) is about 10 μm. At the position of the image this leads to a broadening of about 20 μm for each point of the object, the slit S_3. It follows that a slit width smaller than about 40 μm is not useful for S_3. Further broadening of the image arises from the spherical and chromatic aberration. A broadening of each point to about 30 μm was accepted for each of the two aberrations. Concerning the spherical aberration, this allows a beam width of about 1.5 mm in the plane of the lens. For the chromatic aberration we obtain an acceptable wavelength band of $\Delta\lambda = \pm 0.3$ Å at a mean wavelength of 20 Å. In reality a wider wavelength band of $\Delta\lambda = \pm 0.5$ Å can be accepted because of the following reason: The monochromator in front of slit S_3 (see Fig. 1) delivers a neutron-spectrum which is asymmetric over the beam cross-section, because the deflection angle of the prism is proportional to λ^2. Therefore, by underfocusing the shorter and overfocusing the longer wavelengths, one slope of the image is broadened, the other becomes steeper (see Fig. 2). Only the steeper slope was taken for the charge measurement.

The steeper slope corresponds to a width of the image of about 80 μm. The width of slit S_4 was taken as 40 μm for optimal sensitivity. The count rate behind S_4 is about 50 n/sec.

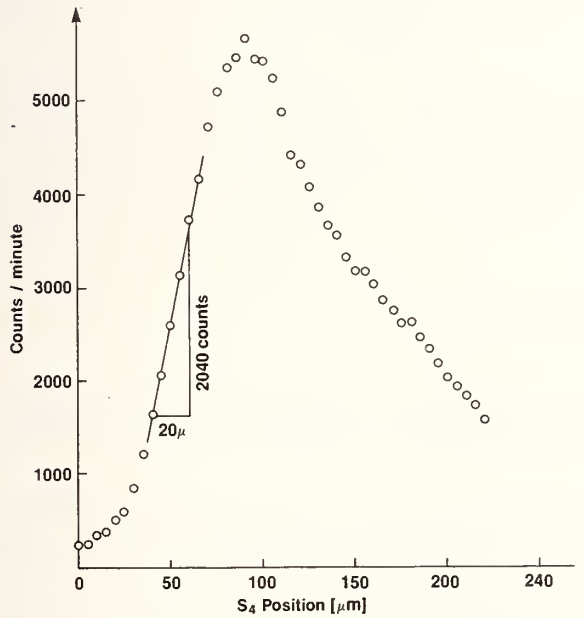


FIGURE 2. *Image of slit S_3 in the plane of slit S_4. The steeper slope was taken for the charge measurement.*

3.2.2 Mean Wavelength λ

The sensitivity of the apparatus to detect a neutron charge q_n is independent of the choice of the mean wavelength λ_0, as long as cold neutrons are used. This is due to the chromatism of the neutron lens and to the reactor spectrum.

The reactor spectrum $d\phi/d\lambda$ varies as λ^{-5} for cold neutrons. Integrated over $d\lambda$—the wavelength band accepted by the lens—we get $\phi \sim d\lambda/\lambda_0^5$. The focal length f of the lens is proportional to $1/\lambda_0^2$. It follows $df/f \sim d\lambda/\lambda_0$. Thus, for constant focal length f and constant acceptable width df resulting from chromatic aberration, the value $d\lambda/\lambda_0$ is constant. Therefore, the usable intensity becomes proportional to $1/\lambda^4$. The deflection of a possible charge q_n in the E-field is proportional to λ^2 and we get with Eq. (7) for σ_{q_n}, the precision for the measurement of a neutron charge:

$$\sigma_{q_n} \sim \frac{w_0}{\lambda_0^2 \sqrt{N}} . \tag{8}$$

Since, as shown above, the count rate is proportional to λ^{-4}, σ_{q_n} becomes independent of λ_0. The spectrum of the guide H18 is peaked at 20 Å with a width of about 10 Å. Shorter wavelengths are cut off because of the curvature of the guide. $\lambda = 20$ Å as mean wavelength was taken, because the high intensity simplifies the alignment of the optical components. The mean wavelength and the wavelength spread of ±5 Å selected by the monochromator was checked using the critical angle of total reflection from a Ni-mirror. During the charge measurement the mean wavelength was controlled by the width of the image. This control was possible because of the severe chromatism of the lens.

3.2.3 Lengths ℓ_0, ℓ_1

The length ℓ_0 of the apparatus was limited by the available space inside the reactor hall of the ILL. The best suited area was that at the neutron guide H18, where a free length of 13 m was available.

3.2.4 Electric Field

The total length of the electric field was 9 m. It reached from the entrance slit S_3 to slit S_4 in front of the detector and was separated by the lens into two sections of equal length. The beam height of the divergent neutron beam increased from 0.2 m at slit S_3 to about 0.38 m at slit S_4. This leads to rather large dimensions of the electrodes: The first one has an area of 0.32 × 4.5 m, the second one 0.4 × 4.5 m. For reasons of high electric field and for mechanical stability each electrode was a stainless steel tube with rectangular cross section. The surfaces which produced the electric field were milled. The waviness of each surface was about 0.1 mm, this being significantly lower than the 2.5 mm gap between the electrodes. The mean gap between the electrodes which determines the mean electric field E_0 was measured with a capacity meter. One electrode of each pair was connected to ground and the other one to a HV-supply with a maximum output of 30 kV at 2 mA. Each pair of electrodes was mounted inside a vacuum tube of 0.8 m diameter and 4.6 m length. A vacuum of better than 10^{-4} Torr was maintained in each tube by a turbomolecular pump. The tubes were supported independently from the optical bench and thus mechanically decoupled from the neutron optics.

The charge measurement was performed with an electric field of ±5.9 kV/mm. The field was limited to this value by breakdowns due to imperfect finishing and cleanliness of the electrode surfaces. Attempts at more efficient cleaning by glow discharge had to be abandoned because of damage to the insulators. Teflon had been chosen as the insulator material because of its shock absorbing qualities in the case of high voltage breakdowns.

4. Measurements and Results

4.1 Performance of the Measurement

At the beginning of each cycle the high voltage was raised in about 1 sec to its nominal value; after a waiting time of 2 sec the measurement, lasting 10 sec was

started. The count rate was recorded on magnetic tape and the high voltage was switched off. Then a measurement with reversed electric field was started. About 1000 to 3000 such cycles were combined in one run. At the end of such a run we regularly checked the steepness of the slope of the image in order to be sure that the sensitivity had not changed. The BF_3-detector was connected to two scalers, one for each polarity of the high voltage. Two monitor counters were used, one for monitoring the incoming flux in front of slit S_1 and one near slit S_4 in order to detect changes in background. In the case of a high-voltage-break-down (about 1 in 5 minutes) the measurement was stopped and the actual count rate was deleted.

4.2 Data Analysis

Each of the 54 runs was analyzed separately. We observed a change of the count rate of less than ±10% within one run, corresponding to a drift of the image of less than ±5 μm. This drift is probably due to thermal effects, though the reactor hall is temperature stabilized (within ±1 °C). This drift was small enough to take the slope dN/dy as constant. The difference in count rate for the two field directions ΔN was compared to the squareroot of the count rate $\sqrt{N}$. These values for all runs are shown in Fig. 3, calculated in terms of the neutron charge. A χ^2-test was made for ΔZ, the difference in count rate for every measurement cycle, i.e., each pair of count rates with opposite field ($\Delta Z = Z_+ - Z_-$), assuming $\Delta Z = 0$ as expectation value. For each run the value of $f(\chi^2) = \sqrt{2\chi^2} - \sqrt{2A - 1}$ [12] (A = number of cycles per run) was calculated. Considering only statistical errors (and no systematic shift of the image), the distribution of the values $f(\chi^2)$ should be a Gaussian centered at zero with standard deviation 1. The agreement with these expectation values was good; we got a mean χ^2 per run of $\chi^2 = 1.02$. This indicates that systematic errors are small compared to the statistical errors.

A second χ^2-test, performed with the results from all runs (Fig. 3), delivers a χ^2 of 1.2. Consequently the error bars from all runs were increased (not shown in Fig. 3) in order to obtain a value of 1.0 for χ^2.

The errors from the determination of the mean wavelength λ, the length l, and the electric field E can be neglected. We got a final result:

$$q_n = |q_e|(-1.5 \pm 1.4) \times 10^{-20} \quad (68\% \text{ confidence})$$
$$q_n = |q_e|(-1.5 \pm 2.2) \times 10^{-20} \quad (90\% \text{ confidence}).$$

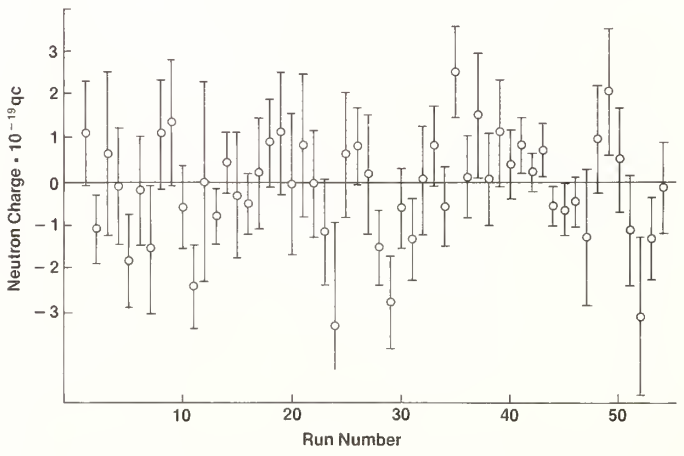


FIGURE 3. *Electric charge of the neutron for each run.*

4.3 Test Measurements

To test whether the apparatus produces a difference in count rate which is not due to the interaction of a hypothetical charge q_n with the electric field, we performed two other types of measurements:

(1) Seven runs were taken with high electric field E_0 for one direction but zero for the other. This test was done for the following reason: The electrostatic attraction of the electrodes changes the gap between them (the change was less than 20 μm). Neutrons striking the electrodes are reflected and can contribute to the count rate behind slit S_4. If the E-field would not be exactly the same for both directions ($|E_+| - |E_-| < 10^{-3}E_+$), the reflected intensities might be different for both directions of E. This could produce a misleading shift of the image. The measurement with zero field in one direction increases this possible effect by about 3 orders of magnitude. For these test measurements, we got, calculated in terms of the neutron charge:

$$q_n = |q_e|\ (0.1 \pm 4.6) \times 10^{-20}.$$

Therefore, this effect can be neglected for the charge measurement. The results with zero field in one direction were used together with the other charge measurements to obtain the final value.

(2) Eight long runs were performed without any electric field. The measured deflection of the image was

$$y = (-0.8 \pm 1.3) \times 10^{-2}\ \mu\text{m},$$

assuming negligible systematic errors. This result indicates that there is no significant beam deflection simulated by possible errors in the counting system.

The deflection from the measurements with electric field was:

$$y = (9.8 \pm 8.9) \times 10^{-3}\ \mu\text{m}$$

including systematic errors.

5. Systematic Errors Due to the Electric Field

Besides a possible charge q_n, the magnetic moment of the neutron and its inner strcture can give rise to a deflection of the neutron in an electric field. It is not possible to separate these forces experimentally from the force acting on a neutron charge, because most of them are coupled to the direction of E. In an extended paper [13], it was shown that the deflection caused by a neutron charge of $q_n = 1.10^{-20} \times q_e$—the sensitivity of the reported measurement—is large compared to all other deflections. The most convenient way is the wave-optical treatment where the mean potential energy along two symmetric paths is calculated. It was shown [13] that ΔQ, the mean energy difference between two symmetric paths, is much smaller for all competing effects than that resulting from a charge $q_n = 1.10^{-20} \times q_e$. The energy Q of a neutron in an electromagnetic field is given by [14]:

$$Q = q_n U + \boldsymbol{\mu}_n \cdot \mathbf{H} + \epsilon_0 \boldsymbol{\mu}_n \cdot \mathbf{E} \times \mathbf{v} + \frac{h\,\mu_n \epsilon_0}{2m} \nabla \cdot \mathbf{E}$$
$$+ C_1 \nabla \cdot \mathbf{E} + \frac{1}{2} C_2 E^2 + \mathbf{p} \cdot \mathbf{E}$$
$$= \text{I} + \text{II} + \text{III} + \text{IV} + \text{V} + \text{VI} + \text{VII}$$

I: charge q_n in a potential U
II: magnetic moment $\boldsymbol{\mu}$ in a magnetic field $\mathbf{H}$; $\mu_n = 1.21 \times 10^{-32}$ V·s·m;
III: "$\mathbf{E} \times \mathbf{v}$"-term [15]; $\mathbf{v}$ = velocity of the neutron
IV: Foldy-term [16]
V: intrinsic Foldy term [17]: $C_1 = \frac{1}{6} \int r^2 \rho(r) dr$
VI: Polarizability of the neutron;
VII: Electric dipole moment $\mathbf{p}$ [18].
Terms II-IV are caused by the magnetic moment μ_n. For an unpolarized neutron beam the terms II and III cannot give rise to a shift of the image but only to some broadening. Though the neutron beam was unpolarized it was shown [13] that even for a completely polarized beam the deflection is negligible. Interactions V-VII are due to a possible intrinsic structure of the neutron and give rise to negligible deflections.

We should like to express our sincere gratitude to Prof. E. Lüscher for the kind hospitality we received in his institute, to the workshop of the Fachbereich Physik of the Technische Universität München for most of the mechanical construction, to the directors and the staff of the ILL for efficient help and valuable discussions, and to Dr. A. Zeilinger for a critical reading of the manuscript.

This work was supported by the Bundesministerium für Forschung und Technologie.

References

[1] S. L. Glashow, in "Proceedings of Neutrino 79," Ed. by A. Haatuft and C. Jarlskog (Bergen, Norway, 18–22 June 1979), Vol. I, p. 518.
[2] G. Feinberg and M. Goldhaber, Proc. Natl. Acad. Sci. U.S.A. 45, 1301 (1959).
[3] H. F. Dylla and J. G. King, Phys. Rev. A 7, 1224 (1973).
[4] L. G. Chambers, Nature 191, 1082 (1961) and references therein.
[5] V. A. Bailey, J. and Proc. Roy. Soc. (N.S. Wales) 94, 77 (1960) and references therein.
[6] E. D. Garris and K. O. H. Zuck, Nucl. Instrum. Methods 177, 467 (1974).
[7] G. Gallinaro, M. Marinelli, and G. Morpurgo, Phys. Rev. Lett. 38, 1255 (1977).
[8] G. S. La Rue, W. M. Fairbank, and A. F. Hebard, Phys. Rev. Lett. 38, 1011 (1977).
[9] G. S. La Rue, W. M. Fairbank, and J. D. Philipps, Phys. Rev. Lett. 42, 142 (1979).
[10] C. G. Shull, K. W. Billmann, and F. A. Wedgwood, Phys. Rev. 153, 1415 (1967).
[11] R. Gähler, J. Kalus, and W. Mampe, J. Phys. E 13, 546 (1980).
[12] R. A. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh, 1958).
[13] R. Gähler, J. Kalus, and W. Mampe, Phys. Rev. D 25, 2887 (1982).
[14] H. Rauch, in *Neutron Interferometry*, Ed. by U. Bonse and H. Rauch (Oxford Univ. Press, Oxford, 1979), p. 161.
[15] J. Schwinger, Phys. Rev. 69, 681 (1948).
[16] L. L. Foldy, Rev. Mod. Phys. 30, 471 (1958).
[17] G. Dattoli, G. Matone, and D. Prosperi, Istituto Naz. Fisica Nucleare, Frascati, Report LNF-77/9(R) (1977).
[18] W. B. Dress, P. D. Miller, J. M. Pendlebury, P. Perrin, and N. F. Ramsey, Phys. Rev. D 15, 9 (1977).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

GAS CONSTANT, X-RAY INTERFEROMETRY, NUCLIDIC MASSES, OTHER CONSTANTS, AND UNCERTAINTY ASSIGNMENT

Methods for the Determination of the Gas Constant

A. R. Colclough

National Physical Laboratory, Teddington, Middlesex TW11 OLW, U.K.

Following a brief discussion of the role of the gas constant in physics, a review is made of the possible methods for its experimental determination and of values obtained in the past by direct measurements or by inference from other work.

Four practicable methods by which the gas constant might be accurately determined are considered. These are (a) the conventional method of the limiting density of a gas based on the virial equation of state, (b) the method of sound velocity measurement in a gas depending upon the normal law for the speed of sound in a compressible fluid together with the virial equation, (c) determination via the Boltzmann constant and Avogadro's constant by the measurement of the electrical noise in a resistance invoking Nyquist's formula, and (d) determination via the Stefan-Boltzmann and other constants by the measurement of the power radiated by a blackbody cavity relying on the Stefan-Boltzmann law. The four methods have very different sources of systematic error. Most of the historical determinations leading to the currently recommended value of the gas constant have been carried out by method (a). Some more recent determinations by method (b) can now be shown to be in agreement with them if the sound speed measurements are correctly interpreted and processed. An assessment of the sources of uncertainty characteristic of methods (a) and (b) is attempted and an indication given of how they might be reduced in the future. These uncertainties may be compared with those to be expected from the relatively little tried methods (c) and (d). The analysis of the four methods suggests that there is scope for a useful reduction in the uncertainty of the gas and the other thermal constants.

Key words: Boltzmann constant; fundamental constant; gas constant; limiting density; Stefan-Boltzmann constant.

1. Introduction

A physical constant is usually called "fundamental" if it represents the value of some property associated with a theoretically unanalyzable or irreducible entity or process (e.g., a fundamental particle or the propagation of light). Nontrivial dimensionless ratios of such constants, being pure numbers and so quite independent of any metrological conventions as to the size of physical units, are agreed to be even more fundamental [1]. The gas constant, R, falls into neither of these categories and when it is referred to as a fundamental constant it can only be a courtesy title, recalling a time when thermal processes had not been logically reduced to mechanical ones. It is now usually known simply as the "universal" gas constant in view of its role in the equations of state of virtually all substances. Nevertheless it remains a key constant in physics because, for sound metrological as well as historical or cultural reasons, we choose to retain temperature as a physical quantity *sui generis* ignoring the possibility of its reduction to mechanical concepts and assigning it a base unit in the International System of Units. Thus it must appear in all phenomenological equations where mechanical quantities are present together with thermal ones in order to maintain dimensional homogeneity. Similarly, its microscopic counterpart, Boltzmann's constant $k = R/N_A$ where N_A is Avogadro's constant, must often appear in the equations of statistical mechanics in an analogous role, for example to convert a numerical measure of molecular disorder into the thermal quantity, entropy.

Uncertainties in R and k are usually given as 31 and 32 parts per million (ppm) respectively (one standard deviation based on the 1973 adjustment of Cohen and Taylor [2]). That of Avogadro's constant is only 5 ppm so that, combining uncertainties in quadrature, a determination of R from a measurement of k will have the same uncertainty to within approximately one percent as the measurement, and vice versa.

A value for the ideal gas molar volume, V_0, defined under the conventional reference conditions of temperature and pressure will of course yield information on R of exactly the same fractional uncertainty and vice versa since they are related in Boyle's law, $P_0V_0 = RT_0$, by defined quantities. These are the temperature $T_0 \equiv 0\ ^\circ\mathrm{C} \equiv 273.15$ K and the pressure $P_0 \equiv 101325$ Pa. (But note that V_0 must be expressed in terms of length units of the same kind employed in the unit of pressure. Prior to the recent decisions of the CGPM (General Conference on Weights and Measures) to redefine the liter as equal to the dm^3, it differed from the dm^3 by about 28 ppm.)

R and k can in principle be determined also from a value for the Stefan-Boltzmann constant σ through the relation

$$\sigma = 2\pi^5 k^4/(15h^3c^2) , \qquad (1)$$

where h and c are Planck's constant and the velocity of light, respectively [3]. Again, h and c are relatively accurately known compared to the accuracy with which σ may be measured (see Section 5). Thus R and k could be determined to one quarter of the fractional uncertainty of a σ measurement. But in the past the accuracy with which σ has been measured has been no better than about 500 ppm [4] so that it has been the practice to use the best value of k to calculate a value for σ.

R and V_0, k, and σ form a small group of interrelated physical constants which, together with the gravitational constant G, contribute little to knowledge of the values of the fundamental constants as a whole because of the difficulties of measuring them accurately. Cohen and Taylor [2] group them under the unprestigious heading of "Other Less Precise Quantities." In order for them to contribute usefully to our knowledge of the values of the other constants [e.g., through Eq. (1)], their uncertainties would have to be reduced to about 1 ppm or less and this seems unlikely in the immediate future. What then is the use of reducing the uncertainties of R, V_0, and k to 10 ppm and σ to 40 ppm? This would have no bearing on any great question of physical theory or on practical energy calculations involving, say, efficient fuel use in industrial plant or energy losses from domestic dwellings.

The real need for new measurements of this order of accuracy arguably arises entirely from their role in primary thermometry. Often a primary thermodynamic thermometer has to be designed to operate in some region of temperature remote from the triple point of water where thermodynamic temperature is defined to fix the magnitude of the kelvin ($T \equiv 273.16$ K). This may have the consequence that the primary thermometer may not function well or even at all near 273.16 K so that it cannot be calibrated at the fundamental fixed point. A good example is afforded by acoustic thermometry in the liquid helium range where thermodynamic temperature is determined from measurements of sound velocity c in helium gas. One has for an ideal gas

$$c(T)^2 = \gamma RT/M \tag{2}$$

where γ is the ratio of the principal specific heats C_p/C_v and M is the molar mass of the gas. In this case experimental accuracy would be reduced at low temperatures if the instrument employed to measure c had to be designed to function at 273.16 K as well to yield $T = [c(T)/c(273.16\ \mathrm{K})]^2 \times 273.16$ K. But with an accurate knowledge of R this is not necessary and Eq. (2) can be applied directly. Thus a knowledge of the gas constant is equivalent to a water triple point calibration. Similar arguments can be applied in principle to other thermometric techniques, although sometimes, as in gas thermometry, it is useful to calibrate at the water triple point for other reasons. For instance they establish the need for accurate values of σ and k. Generally σ will be applicable to primary radiometric thermometry employing Stefan's law or Planck's law while k will apply to microscopic methods such as nuclear orientation thermometry or Johnson noise thermometry.

In what follows four practicable methods by which the gas constant might be accurately determined are considered. These are (a) the conventional method of the limiting density of a gas based on the virial equation of state; (b) the method of sound velocity measurement in a gas depending upon the normal law for the speed of sound in a compressible fluid together with the virial equation; (c) determination via the Boltzmann constant and Avogadro's constant by the measurement of the electrical noise in a resistance invoking Nyquist's formula; and (d) determination via the Stefan-Boltzmann and other constants by the measurement of the power radiated by a blackbody cavity relying on the Stefan-Boltzmann law. The four methods have very different sources of systematic error. Most of the historical determinations leading to the currently recommended value of the gas constant, $R = 8.31441\ \mathrm{J \cdot K^{-1} \cdot mol^{-1}}$ [2], have been carried out by method (a). Some more recent determinations by method (b) can now be shown to be in agreement with them if the sound speed measurements are correctly interpreted and processed. An assessment of the sources of uncertainty characteristic of methods (a) and (b) is attempted and an indication given of how they might be reduced in the future. These uncertainties are then compared with those to be expected from the relatively little tried methods (c) and (d). The comparison indicates that there is scope for a useful reduction in the uncertainty of the gas constant.

2. Determination of R by the Method of Limiting Density

The conventional method for determining R and V_0 is the misleadingly named method of limiting density. A quantity

$$L(p) = (P_0/p)\rho(p), \quad T = T_0 \tag{3}$$

is measured at successively lower pressures p where density $\rho(p)$ is determined by weighing the masses $m(p)$ of a constant volume V of gas at temperature T_0. Extrapolating the measured values to zero pressure yields

$$L(0) = MP_0/(RT_0) = M/V_0 \tag{4}$$

for $T = T_0$, where M is the molar mass of the gas. From this R and V_0 are easily determined.

In the past the directly measured quantities p, V, and m have been determined respectively by mercury barometry, the weighing of a glass flask when empty and filled with water, and weighing the flask evacuated and filled with the experimental gas. Most such determinations have used oxygen as the experimental gas since until 1961 this was assigned a molar mass of 32 exactly on the chemical atomic mass scale. Occasionally, the extrapolation of $L(p)$ to zero pressure was carried out with the aid of a value for the gradient of the $L(p)$ isotherm taken from the results of other workers. Usually the approximate range of measurements was from 0.25 to 1.0 atmosphere. Unfortunately, error analysis for the important historical determinations of R (shown in Table 1) is usu-

TABLE 1. *Some historical determinations of the gas constant by the method of limiting density.*[a]

Source	Gas	V_0 (dm^3/mol)		R_0 (J/K·mol)
Baxter and Starkweather, 1924 [5], 1926 [6]	O_2	22.4133 ±0.0002	*	8.31421 ±0.00007
Baxter and Starkweather, 1926 [7]	N_2	22.4126 ±0.0006	*	8.31395 ±0.00026
Moles *et al.*, 1934–1937 [8]	O_2	22.4135 ±0.0002	*	8.31429 ±0.00007
Batuecas, Casado, and Malde, 1939–1949 [9, 10, 11]	O_2	22.4142 ±0.0009	*	8.31455 ±0.00033
Pereira, 1978 [12]	N_2	22.4132 ±0.0007		8.31417 ±0.00026

[a]Values for V_0 in this Table marked * have been obtained from the last recalculation of Batuecas [11, 13] which incorporated a number of corrections for systematic effects not made by the authors and converted to standard gravity and to the physical scale of relative molecular masses. We have converted liters to dm^3. All quoted uncertainties are purely statistical ones arising from random scatter in a set of measurements. They represent one standard error.

ally rudimentary or lacking altogether. Corrections for several systematic effects are made and it is subsequently assumed that there are no other significant systematic errors and that no significant residual error remains arising from the uncertainty in the corrections. The error finally quoted is usually determined solely from the random error in a set of repeated determinations.

In order to obtain some conception of the likely magnitude of the error in a conventional determination, consider the expression for $L(p)$ in terms of measured quantities:

$$L(p) = \frac{P_0}{p}\frac{m(p)}{V} = \frac{P_0}{p}\frac{W_{f+g} - W_f}{(W_{f+w} - W_f)/\rho_w}, \quad (5)$$

where W_f is the mass of the evacuated flask, W_{f+g} the mass of the flask filled with gas, W_{f+w} the mass of the flask filled with water, and ρ_w the density taken for the water. The approximate fractional uncertainty in $L(p)$ will be given by

$$dL(p)/L(p) = (dW_{f+g} \pm dW_f)/W_g \pm (dW_{f+w} \pm dW_f)/W_w \pm dp/p \pm d\rho_w/\rho_w \pm dT_0/T_0 \quad (6)$$

where "±" here signifies linear or quadratic error combination according to one's preferred school of thought on the question and where W_g and W_w are the masses of gas and water respectively. In what follows, uncertainties which are arguably systematic or "Type B" (including values like ρ_w taken from other work) will be combined linearly with the root-sum-square values following in brackets. A plausibly estimated detailed error budget is given in Table 2. It applies to a good contemporary determination rather than to the historical ones. The notes following the table give the details of the assumptions made for each particular source of uncertainty.

Perhaps the most widely felt concern about this method, arising to some extent from the experiences of Guildner *et al.* [22] with gas thermometry, is that systematic effects related to gas sorption have been incorrectly assessed in the past. Consider the following sequence of operations and processes which occur in the determination of $m(p)$ through measurements of W_f and W_{f+g}:

(a) Evacuate flask and seal.

(b) Some gas (helium) diffuses into the flask from the surrounding atmosphere.

(c) Gas (probably H_2O, N_2, CO, O_2) desorbs from inner surface of the flask.

(d) The flask is weighed yielding a value for W_f.

(e) The flask is now charged with O_2 at T_0 and p is measured. Here three problems may arise:

(e1) Some diffused or desorbed matter is retained in the flask leading to an error, dM_1, in the value assumed for M (p is taken to be correct).

(e2) Some diffused or desorbed matter passes from the flask into the gas handling and barometer systems. This means that W_f as determined in (c) is no longer the appropriate value to subtract from W_{f+g} to obtain $m(p)$. We attribute the error somewhat arbitrarily to W_f inserting a corresponding value dW_{f1} into the error budget.

(e3) Some of the O_2 introduced into the flask will adsorb onto the wall where it will increase the observed value of W_{f+g} without contributing to the value of p. Again somewhat arbitrarily, this effect is expressed as an error in W_f rather than in p or W_{f+g}.

(f) The flask is sealed and weighed to determine W_{f+g}.

A study of the literature of gas-glass sorption reveals that sorption effects are very variable, depending on many generally ill-defined factors such as details of glass composition, gas or vacuum exposure times, gas species, and past chemical treatment. Nevertheless, some plausible guesses as to the nature of sorption mechanisms can be made and it can be shown reliably that the effect of helium diffusion will be negligible [23, 24]. Almost certainly the sorption of water will be far more important at ambient temperatures than that of any other species including oxygen, especially where flasks have been used repeatedly for water and gas weighings without being subjected to baking at high temperature (see e.g., Ref. [25]). The latter step would almost certainly affect their volumes at the ppm level of accuracy and would presumably be unthinkable practice between an initial volume determination and a subsequent gas weighing. If a fraction f of the desorbed gas is lost from the flask in process (e2), then the resulting error in the gas constant due to mechanism (e1) and (e2) will be

$$dR/R = n_2[f(2M_2 - M_1) + M_1 - M_2]/n_1M_1, \quad (7)$$

where n_2 and n_1 are the number of moles of desorbed and experimental gas respectively having corresponding molar masses M_1 and M_2 ($n_2 \ll n_1$). For $M_2 = 18$ (H_2O) and $M_1 = 32$ (O_2) it can be seen that this error is only weakly dependent on f so that a rough error of $n_2/(2n_1)$ ppm in R may be calculated for $f = 0.5$. Elsey [26] quotes figures for the outgassing rate of pyrex from 2.1×10^{-7} Pa·m·s^{-1} (for glass kept for 1 month in air prior to a ten hour high vacuum exposure) to 9.8×10^{-6} Pa·m·s^{-1} (for new glass following one hour's exposure to vacuum). If these rates were sustained for 24 hrs in a 1 L pyrex bulb, errors in the gas constant would result of 4 and 180 ppm respectively. The former figure shows that this source of error can probably be controlled with adequate pumping at ambient temperature and the latter shows that it is vital that it is.

Unlike water sorption, oxygen sorption effects have received considerable attention over the years [27, 28, 29], especially from Moles and collaborators at Madrid. Such work has led to corrections being made, occasionally retrospectively, based on Freundlich adsorption isotherms. In such work it is essential that adsorption measurements are quoted for the correct type of glass and for the treatment it would receive in use, especially with regard to the cycle of evacuation and filling [27, 29]. Fortunately, adsorption equilibrium is rapid, being completed in general in less than one hour provided that problems due to tap grease or surface bloom are absent. Hartley *et al.* [29] quote a figure of 4.1×10^{-5} cm^3/cm^2 of O_2 adsorbed on soda glass at 1 atmosphere and 15 °C, representing an error of about 20 ppm in a density determination made with a 1 L flask. With a silica flask, on the other hand, their figures suggest an error of only 0.1 ppm. It can be seen, therefore, that the effect of mechanism (e3), like that of mechanisms (e1) and (e2), needs careful control, but can be reduced to manageable proportions. Arguably the best practice to adopt would be to make repeated determinations of R with a high temperature bakeout preceding each gas filling and weighing, and with the volumetric determination being made last. In Table 2 it is assumed that this is done and that sorption errors are limited to 5 ppm. It is also of interest to note that Hickmott [30] has observed by the use of a mass spectrometer values of n_2M_2 for mercury equal to those for water in an unbaked pyrex system subject to pumping for two weeks at ambient temperatures. The significance of such an observation for the historical determinations of R can only be guessed.

TABLE 2. *Plausible error summary for a modern determination of R by the method of limiting density.*[a]

Measured quantity x, with its uncertainties dx and corrections, if made, with their uncertainties	Corresponding error in R (ppm)
p—mercury barometry	
dp_1: Hg temp. meast. (20 ± 0.01°C) & corrn. to 0 °C(3.6×10^{-3} p ± 2×10^{-6} p)	2
dp_2: Uncertainty in g ($\pm 10^{-5}$ m/s^2)	1
dp_3: Hg purity	2
dp_4: resolution	2
dp_5: capillary effect; dp_6: scale; dp_7: resolution	0
dp/p	7(4)
W_f—weigh evacuated flask (about 200 g)	
dW_{f1}: sorption (inc. dM_1—see main text)	5
dW_{f2}: air buoyancy of flask (see main text)	1
dW_{f3}: pressure distortion of flask (see main text)	2
dW_{f4}: weight calibration, dW_{f5}: weight buoyancy	2
W_{f+g}—weigh gas-filled flask (about 200 g)	
dW_{f+g1}: sorption (included in dW_{f1})	—
dW_{f+g2}: air buoyancy of flask	1
dW_{f+g3}: pressure distortion of flask	2
dW_{f+g4}: weight calibration; dW_{f+g5}: weight buoyancy	2
$(dW_{f+g} \pm dW_f)/W$ g, $Wg \simeq 1$g at 500 mmHg	15(7)
W_{f+w}—weigh water-filled flask	
dW_{f+w1}: sorption & buoyancy as in W_f and W_{f+g} above (see main text)	0
dW_{f+w2}: flask load distortion (see main text)	1
$(dW_{f+w} \pm dW_f)/W_w$, $W_w \simeq 10^3$ g	1(1)
ρ_w—density of water: $d\rho_w/\rho_w$	3(3)
T_0—filling temperature: $dT/T_0 = \pm 0.0005$ °C/273 °C	2(2)
$dL(p)/L(p)$	28(8)
Extrapolation to zero pressure: $dL(o)/L(o)$, 2 ppm	30(9)
M—molar mass (dM_1 included in dW_{f1})	—
dR/R	30(9)

[a]Generally it has been assumed that the work is carried out with the cooperation of a national standards laboratory, but not by such a laboratory. Thus accuracies are those obtainable from fine instruments calibrated against a primary standard within a year of use, say, rather than from the primary standards themselves. The figures estimated for the uncertainties do not necessarily relate to any particular past work, though occasionally such work has been used to provide a plausible value.

(dp_1). Assuming a water-jacketed barometer with temperature control and determined to ±0.01 °C using mercury-in-glass thermometers. The corrections to 0 °C (or to any other temperature for which the density of Hg is available) can be made to approximately 2 ppm [14, 15, 16].

(dp_2). Uncertainty in the local value of g will depend on where the measurement is made. At national standards laboratories g is often determined to parts in 10^8. In Table 2 it is assumed that g is known to 1 ppm through a comparison made with an accurately determined value of g using a transfer standard.

(dp_3). Cook [14, 15] has determined the density of seven pure Hg samples from different sources. Mean results of two methods agreed to 0.45 ppm and individual values differed by up to 1.7 ppm. We take ±2 ppm for density of any pure Hg likely to be used in the future.

(dp_4). It is assumed that the barometer can be read to about 0.2 Pa employing optical methods or that this precision is achieved by taking the mean of many readings.

(dp_5). It is assumed that large bore barometers are used making the capillary effect negligible. A 4 cm bore will have a capillary error of about 0.013 Pa or 0.2 ppm of pressure at 500 mmHg (68 kPa) [17, 18].

(dp_6). Scale calibration, alignment, and expansion errors can be made negligible with care. The same applies where interferometric methods are used.

(dW_{f4}, dW_{f5}, dW_{f+g4}, and dW_{f+g5}). It is assumed that weighings can be effected by substitution to 0.003 mg representing an uncertainty of 2 ppm on one liter of O_2 at STP.

(ρ_w). The density of water is known to about 3 ppm at 4 °C [19, 20, 21]. The effect of dissolved air and of the dependence of density on pressure and temperature are less important.

(Extrapolation). This will obviously depend on the range and quality of data. It is assumed here that extrapolation introduces errors of 20% of the rms deviation of the isotherm fit.

(M). This will depend on the gas chosen. It is no longer possible to use a gas with an exact defined molar mass. But Ne may be used with a measured isotopic composition enabling M to be defined with an uncertainty of less than 0.5 ppm.

Some similar problems to those outlined above will arise in the determination of V by measurements of W_{f+w} and W_f. But here all errors due to sorption are diminished according to Eq. (6) by a factor $W_w^{-1} \simeq 10^{-3}$ g^{-1} compared to $W_g^{-1} \simeq 1$ g^{-1} in the case of $m(p)$ and so are negligible compared to dW_{f1}.

It may be of importance, however, to consider how the weight of the water affects the volume of the flask on filling. If the flask is filled and sealed while suspended in air it will suffer forces tending to make it pear-shaped so changing its volume. If it is filled under water, resting on a flat surface or in a ring before sealing different changes will occur. Because the volume thus determined is not that applicable to the gas weighing we insert an error dW_{f+w2} in the error budget. Crude calculations indicate that the effect might amount to several ppm in unfavorable circumstances so that it calls for an error assessment. Unfortunately, reports of the historical measurements fail to do this. Nor do they usually give adequate details of the geometry, wall thickness and material of the flasks employed, or of the method of filling. In order to make a rough practical assessment of this error, a glass sphere of nominal volume 1 L with wall thickness approximately 1.5 mm was furnished with a long capillary tube and filled with colored water. A decrease of 1 ppm in volume was calculated to cause the water to rise in the capillary by approximately 1 cm. Different methods of supporting the sphere readily produced variation in its volume of 1 ppm.

Changes in flask volume would also be expected due to variations in the pressure of the experimental gas. Accompanying this would be a change in the buoyancy in air during weighing (this change is not to be confused with that arising from variations in air density between weighings—see below). The change in V from zero pressure to one atmosphere might typically be 100 ppm and could in principle be corrected for from a knowledge of the elasticity of the flask material and its geometry assuming that it retained its shape at all pressures. But in practice variations in wall thickness and details such as taps and support rings make this difficult. Some historical determinations (e.g., Ref. [28]) allow for this effect using a semiempirical formula due to Moles and Miravalles [31]. But the latters' work indicates that it is only accurate to about 7%, equivalent, very roughly, to an uncertainty of 7 ppm in R. The obvious way to control this effect is to increase the wall thickness of the flask. But this increases its weight and a point comes where delicate balances can no longer be employed to measure to 0.01 mg or less. Generally flasks weighing less than 200 g are used for this reason, implying for a volume of 1 L with a 50 g glass tap a wall thickness no greater than about 1.5 mm. In a modern determination this effect would presumably be investigated experimentally (e.g., by subjecting a water-filled flask furnished with a capillary as described above to external pressure changes). Any error from this source might be expressed as a weighing error dW_{f3} in W_f. We assume it can be controlled to 2 ppm.

Even with a perfectly rigid flask changes in buoyancy due to variations in air density from one weighing to another could have a serious effect. Usually a dummy counterpoise flask is employed to control this source of error, but even so, the error needs careful treatment. We suppose that the counterpoise has its mass accurately matched by the aid of mercury filling and that the external volume difference of the two flasks has been determined. It is difficult to determine changes in air density to better than about 100 ppm of initial density by pressure, temperature, and humidity measurements. This uncertainty in the air density, which arises mainly from uncertainties in air composition, will give rise to a comparable uncertainty in W_g. Thus in order to achieve control of this effect to 1 ppm, it needs to be established that the external flask volumes differ by no more than about 1%. We assume this is done, and that buoyancy errors in R due to all weighings amount only to 2 ppm (1 ppm each for W_f and W_{f+g}).

The final uncertainties arising from other aspects of weighing, pressure, and temperature measurement are more familiar and will not be discussed. Their effect is shown in Table 2 with its notes.

It is of considerable interest to note that an important projected refinement of the conventional method of limiting density is described in these proceedings by Guildner and Reilly. By using a piston displacement method to determine the volume of the experimental gas, they hope to avoid the key problems of sorption. The displaced gas is condensed into a small weighing vessel at 4.2 K so that problems of buoyancy are likely to be less troublesome too. Their preliminary assessment of errors suggests a fi-

nal uncertainty in a determination of R might be achieved of only 3 ppm (99% confidence limit with all uncertainties combined in quadrature). This compares well with a plausible uncertainty of 30(9) ppm by the conventional method based on our rough calculations.

3. Determination of R by Sound Velocity Measurement in Gases

In this method the velocity of sound $c(p)$ at a temperature close to 273.16 K is measured at increasingly lower pressures and an isotherm of $c^2(p)$ vs. p is plotted to enable the ideal gas value $c^2(p = 0)$ to be obtained by extrapolation to zero pressure. Then Eq. (2) is used to obtain the gas constant. (The velocity of sound in a real gas may be obtained from a virial expansion of pressure terms whose coefficients are related to those at the virial equation of state [32, 33].)

The uncertainties in a value of R determined in this way are principally those arising from the sound velocity measurements. Since temperature and pressure have to be accurately defined it is not possible to employ time-of-flight pulse methods which require long paths. Problems can also arise in identifying the edge of the pulse since the sharper this is, the worse will be the problems of dispersion. Continuous wave measurements are not used either due to problems of path accommodation and the elimination of reflections. Thus interferometric methods have to be employed. This presents the experimentalist with four rather fundamental choices. Shall he opt for (a) variable path/fixed frequency or variable frequency/fixed path interferometry; (b) audio or ultrasonic interferometry; (c) one or two transducer interferometry; and (d) cylindrical, rectangular, spherical (or other) resonator geometry? In what follows errors will be analyzed for a conventional interferometric determination of R much as they were for the limiting density method in Section 2. Some recent unconventional acoustic approaches will then be mentioned.

In the variable-path, fixed-frequency interferometer a linear cavity dimension is changed, e.g., by withdrawing an acoustic reflector, to bring the cavity from one order of resonance to another at a constant frequency. Thus one only needs to measure *changes* in length to obtain wavelength/2 and hence sound velocity. This is easier than accurately measuring absolute length as one has to in a fixed-path cavity though such variable-path cavities, having moving parts, are more complicated. More importantly, there are two further advantages of variable-path interferometry of an acoustic nature. Firstly, all the acoustic parameters which characterize the cavity (sound velocity with its small boundary layer changes, absorption, reflection losses, reflection phase changes, and hence effective cavity length, and the real and imaginary components of transducer impedance) are frequency dependent and so change from one order of resonance to the next with the variable frequency method. Indeed, to second order, they change over a single resonance. With the variable path cavity they remain constant at all orders of resonance since it is operated at constant frequency. This greatly simplifies the analysis of the acoustic resonances. The fact that with the variable-path instrument, absorption and reflection losses are constant confers a second very important advantage on the method. It means that they can be separated experimentally without recourse to theory by measuring the losses at two or more orders of resonance. Then one can check that loss mechanisms,. especially those arising from the acoustic boundary layer, are as predicted by theory. This provides an invaluable check on the operation of one's instrument and, if the instrument appears not to function normally, aids in diagnosis. Absorption coefficients can also be used directly to correct the measured sound velocity for the effect of the acoustic boundary layer (see below). Finally, if it is decided to operate at ultrasonic frequencies, commitment to a fixed-frequency variable-path instrument allows one to employ a quartz crystal transducer operated at resonance where its intrinsic impedance is effectively zero. Thus a measurement of its impedance under load is effectively a direct measurement of the impedance of the loading resonant cavity which greatly simplifies data processing.

Whether it is best to use audiofrequency or ultrasonic interferometry is a controversial matter. In discussing the closely related problems of acoustic thermometry, the author has argued strongly for acoustic frequencies [34]. He is still of this view, if less strongly, but the reasons have changed somewhat over the years. There used to be two main reasons why ultrasonic interferometry was thought to be problematic.

Firstly, there was no practical test one could make to ensure that only the plane wave (oo-order) mode propagated in the cavity. In fact it seemed quite probable that one would be measuring not the required plane wave phase velocity, but a group velocity corresponding to a group of unresolved higher modes. (Calculations of the likely magnitude of the resulting error in velocity and temperature measurement are given in Tables 4 and 5 of Ref. [34], but are based only on guesses about transducer behavior.) It was at one time assumed that the presence of higher modes might be detected from changes in the separation of the resonances with order or from measurements of acoustic absorption coefficients. This need not be the case, as detailed analysis shows. All that can be done is to assume that the transducer executes "plausible" vibrations and to calculate the resulting error. Then an uncertainty could be quoted covering this error several times. However, the error can be made quite small (about 200 ppm perhaps) if very high frequencies are employed so that the various phase velocities of the higher modes are close to each other. But then acoustic absorption, proportional to the square of the frequency, may increase rapidly and bring about a loss of sensitivity.

The second reason why ultrasonic interferometry was thought to be unreliable was that no theory existed giving the tolerances to which the cavity end faces needed to be in alignment. At high frequencies with very short ultrasonic wavelengths it is possible in principle for resonance widths to approach optical dimensions given ideal geometry. But if the cavity has geometrical errors of this order, could one be sure that one understood its behavior as is required of a primary instrument? This question was later treated by Colclough [35] who provided a realizable criterion for adequate cavity alignment in the ultrasonic interferometer and showed that optical cavity alignment is generally satisfactory.

However, other problems have become more important for ultrasonic interferometry since then. It has long been known that sound propagates more slowly in a tube or resonator than in the unbounded fluid due to viscous and conductive mechanisms in the acoustic boundary layer. Similarly, absorption and reflection losses are increased. These effects are smaller in the ultrasonic instrument than in the audio interferometer and have generally been assumed to behave as predicted theoretically so that the necessary corrections can be made for them. But no absolutely clear explanation has been found for Plumb's discovery [36] that the boundary layer effect can

disappear at low pressures. This can easily be checked by measuring the acoustic absorption (bearing out the above point on the importance of measuring one's losses) and raises the question as to how one should correct for boundary layer effect on the sound velocity. Again, a clear answer is needed for a primary measurement. Note too, that it was originally thought that the small boundary layer effect at ultrasonic frequencies constituted an important advantage of the method. But an ultrasonic boundary layer effect which is not corrected for or which is wrongly corrected for in its absence can introduce a velocity error even larger than that which remains in an audio measurement after making an approximate correction for the greater audio boundary layer effect.

However, it may be the case that in practice the extra work required to control the larger boundary layer effect at low frequencies, while unproblematic in principle, imposes a considerable burden on the low frequency method which results in less data and so poorer statistical control. And certainly the availability of quartz crystal transducers at ultrasonic frequencies is something that any audiofrequency worker will envy. To summarize: ultrasonic thermometry is experimentally more convenient and productive, but can raise serious problems of interpretation while the opposite is the case with audiofrequency thermometry. Thus which is the best to pursue is a difficult matter of judgment.

Finally, there is another serious systematic error which the author has encountered in his audiofrequency work which may occur in ultrasonic interferometry as well. This is due to the nonlinearity of the transducer used to excite the interferometer. This error has not been appreciated until recently [32, 37, 38] although it has been common knowledge for very many years that transducers, when driven sufficiently hard, are nonlinear. Now that its existence is known, it need not pose a serious threat either to ultrasonic or audiofrequency methods, but care is necessary to ensure that it is properly controlled.

With regard to the choice as to whether to employ one or two transducers, most workers in acoustic thermometry use one transducer both to excite the acoustic resonances and to monitor them. A few use one to excite the cavity and another to monitor it. The former is structurally simpler but the latter has the advantage that data processing may be facilitated since a pressure sensitive (as opposed to a velocity sensitive) transducer at the opposite end of the cavity to the exciting transducer produces symmetrical resonance peaks like the quartz crystal single transducer. (Single transducer interferometry was assumed in the preceding discussion.)

Concerning the question of geometry, any form of resonator can be employed in principle where one of its linear dimensions can be related to the acoustic wavelength. Needless to say the number of possible resonators where no change of this dimension is called for, i.e., variable-frequency resonators, is larger than the number of variable path resonators. For example, a spherical resonator (a Helmholtz resonator) such as that developed by Moldover and Mehl [39] might be employed—this has the advantage that there are no transmission boundary layer losses, but only reflection losses. (Such resonators are discussed in these proceedings by Moldover and Mehl.) However, if one wishes to operate with a fixed frequency and a variable path, then one needs a geometry with a linear axis of symmetry. Resonators both of rectangular or circular cross section have been widely employed. Manufacture of the cylindrical cavity is, of course, simpler. But in both cases great care needs to be taken to ensure excellent and stable end face alignment at ultrasonic frequencies and, at all frequencies, that the wall roughness is small compared to the expected acoustic boundary layer thickness. In the audiofrequency device it has been assumed throughout that the frequency is held below the first cut-off frequency of the cavity so that only the plane wave propagates. This requires that operating wavelengths are greater than 3.42 cavity radii for the cylindrical interferometer.

Table 3 shows a list of relevant systematic effects with some explanatory notes and references. Not all of the effects are problematic in that their influence can be made negligible. (These are marked with an asterisk.) But all need to be considered in a primary measurement. The errors are calculated for a typical audiofrequency and a typical ultrasonic interferometer.

The analysis presupposed in Table 3 is unlikely to be applicable in any simple way to the free-propagation ultrasonic Michelson interferometer under development by Sauder for the measurement of R nor to the spherical interferometer of Moldover *et al.* Transmission boundary layer errors clearly cannot exist in the absence of guiding boundaries so that one of the potentially most serious errors of conventional cavity interferometry is avoided *ab initio*. The interested reader is referred to these workers' contributions to these proceedings for further details.

From Table 3 it can be seen that the listed effects can introduce uncertainties into a measured value of R of approximately 35 ppm (16 ppm root-sum-square). Using a conventional audiofrequency instrument, Colclough, Quinn, and Chandler [32] determined a value of R for which they quoted a systematic root-sum-square uncertainty of 18 ppm, very close to the estimate of Table 3 and a random uncertainty also of 18 ppm (three standard errors).

Some confusion has arisen in recent years concerning the compatability of values of R determined acoustically and by the method of limiting density. This was due mainly to the systematic effect of transducer nonlinearity in the work of Quinn *et al.* [33] which caused isotherm curvature at low pressures. Misinterpretation of this curvature led to an excessively high value of R [32, 37]. At approximately the same time an ultrasonic determination of R by Gammon [44] yielded a value rather closer to the accepted value derived from limiting density measurements. However, when corrected for the effect of the acoustic boundary layer, Gammon's result appeared more compatible with that of Quinn *et al.* Subsequently the latters' value was corrected for nonlinearity and was shown, afterall, to be compatible with the accepted value [32]. Then an indirect audiofrequency determination by Colclough [37] by comparing the results of acoustic and gas thermometry in the liquid helium range (4.2–20 K) confirmed this. The status of Gammon's work now remains uncertain while two audiofrequency measurements agree with each other and the accepted value of R. These results are shown in Fig. 1.

4. Determination of $R = N_A k$ by Electrical Noise Measurement

Although at least one institute is considering the measurement of k through the Nyquist relation, no such determination has yet been completed. This is largely because the development of electronic components, most notably FETs, has only recently reached a point where there is a possibility of competing with conventional methods. Not surprisingly noise thermometry has re-

TABLE 3. *Plausible error summary for a determination of R by the sound velocity method.*

Source of error, size of correction if made		Resulting error in R (ppm)		Note
		Audiofrequency	Ultrasonic	
Higher modes propagate, $dR/R \simeq (X_{33}/bk_{00})^2$		0	200	1
Acoustic boundary layer, $dc/c = \alpha_{BL}/k \simeq 3000$ ppm (audio)		4	200	2,3
2nd order BL correction to Kirchhoff-Helmholtz 1st order BL correction, 30 ppm (audio)		0	0*	2,4
Molecular slip at wall, 2 ppm (audio)		0	0*	2,5
Temperature discontinuity at wall, 4 ppm (audio)		0	0*	2,6,
Finite thermal conductivity of wall, 1 ppm (audio)		0	0	2,6, 7
Wall smoothness	*	0	0	8
Wall elasticity, 3 ppm (audio)		0	0	6,7
Wall adsorption	*	0	0	9
Finite sound amplitudes	*	0	0	10
Non-uniform cavity bore	*	0	0	11
End face misalignment	*	0	0	12
2nd order correction for using 1st order theory $\simeq (\alpha/k)^2$	*	0	0	13
Blatt effect	*	0	0	14
Cavity length measurement (variable path assumed)		2	2	15
Resonance processing (geometrical processing of resonance circles or peaks assumed)		10	10	
Frequency measurement		0	0	
Pressure measurement (for isotherm plotting)		0	0	
Nonlinearity of transducer	*	10	10	16,2
Realization of T_0 ± 0.0005 °C/273 °C		2	2	
Isotherm extrapolation		2	2	17,2
Molecular weight, 5 ppm		5	5	18
TOTAL		35(16)		

*Signifies that the error is negligible or eliminable.

1. Error calculated using Colclough [34], paragraph (iii), p. 96. It represents only a plausible guess. $X_{33} = 14.54$, b = cavity radius, $k_{00} = 2\pi/$(wavelength of 00 mode). An error of about 200 ppm might be expected with Ne gas in a 3 cm diameter cavity at a frequency of 3 MHz.

2. The correction can be made from absorption measurements or from values of gas transport properties [33]. α_{BL} = absorption coefficient due to boundary layer, $k = 2\pi/$(wavelength).

3. See e.g., Ref. [33], p. 353, Eq. (6). It has been assumed that α_{BL} is measured to 0.05 percent at audiofrequencies and used to correct c.

4. See reference to Shields, Lee, and Wiley [40] in Ref. [33], p. 375.

5. See Ref. [40], and p. 402 of Ref. [33].

6. See Henry [41] and Weston [42].

7. This error extrapolates to zero with pressure and so does not affect isotherm measurements.

8. Protrusions and interstices on walls must be negligible compared to the boundary layer thickness; see p. 376, Ref. [33].

9. See Ref. [33], p. 354.

10. This is easily checked by measuring velocities at various amplitudes.

11. Second order effects due to changes in the boundary layer effect could in principle arise.

12. See Colclough [35], especially Eq. (46).

13. See Appendix B of Ref. [33]. But see also Colclough, Quinn, and Chandler [32], p. 128. One needs to think clearly whether to apply this correction.

14. See Colclough [37], p. 353. Also Quinn, Colclough, and Chandler [33], p. 410. Blatt [43] has argued that as the pressure falls along an acoustic isotherm, the ternary molecular collision frequency will approach the sound frequency so invalidating the normal (zero frequency) virial expansion for c^2. This point is discussed in the above references.

15. This assumes interferometric measurement with an unstabilized He-Ne laser.

16. 150 ppm corrected to within 20 ppm is typical of the author's experience [32, 37], but with foresight the error might be kept to 10 ppm.

17. This will obviously depend on the range and quality of data. It is assumed here that extrapolation introduces errors of 20% of the rms deviation of the isotherm fit.

18. This value is based on Ar as used in Ref. [33].

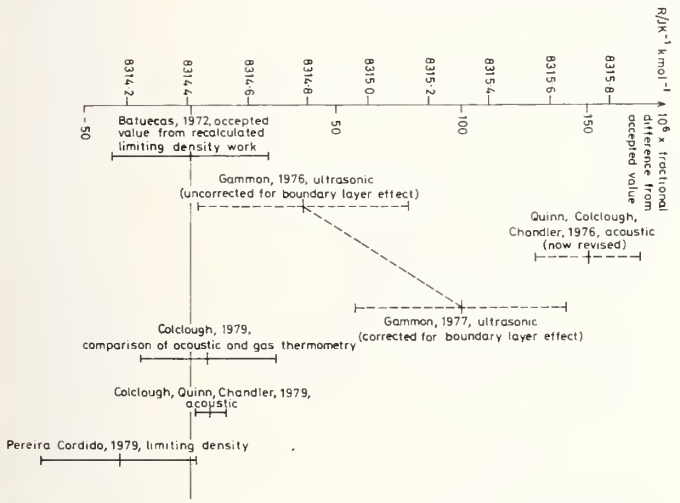


FIGURE 1. *Relationship between recent values of the gas constant determined acoustically and the accepted value. Error bars represent only the random scatter in a set of repeated measurements (1 standard error).*

ceived a higher priority than the determination of k and the recent literature of the former type of measurement provides a rich source of information by which to make a prima facie judgment of the potential of the latter. Only thermometry employing ambient-temperature discrete component electronics (as opposed to microwave or SQUID electronics [45]) need be of concern in this context.

In noise thermometry use is made of the classical Nyquist formula

$$<v_s^2> = 4kTrB = <V^2>B/\int_{f_1}^{f_2} |A(j\omega)|^2 df \quad (8)$$

to determine thermodynamic temperature T in terms of measured values of some instrument output voltage V. This is related to the mean square source noise voltage $<v_s^2>$ in bandwidth $B \equiv f_2 - f_1$ by the transfer function $A(j\omega)$ of the instrument. To determine k, the resistance r will be maintained at or close to the triple point of water ($\equiv 273.16$ K) to yield a value for k from $<V^2(273.16\ \mathrm{K})>$. Thermometry measurements for constant r can often be referred simply to some reference temperature, e.g., T_0, ($T/T_0 = <V^2(T)>/<V^2(T_0)>$) and there is no need to determine k just as in gas thermometry there is no need to determine R. Neither is there any need to determine $A(j\omega)$ except insofar as it is necessary to define any changes in bandwidth. But this is not the case with a determination of k where an absolute measurement of $<v_s^2>$ is required and hence of $A(j\omega)$. This demanding additional measurement (which corresponds to the additional gas weighing in the method of limiting density) is the main reason why determining k, superficially with the same apparatus, is harder than determining T.

There are four methods of noise thermometry of potential relevance to a k-determination. These are (i) the "direct" method; (ii) the correlation method; (iii) the Garrison-Lawson method [46]; and (iv) the Borkowski-Blalock method [47].

However, method (iii) need not be considered as a relevant option for determining k. Here T is determined by adjusting the reference resistance r_0 to achieve the operating condition where $r_1T_1 = r_0T_0$ so that test and reference measurements present the same source noise to the measuring system. This has the advantage that amplifier noise can be eliminated from the measured noise levels rather easily. However, it seems unlikely that anyone would choose a method for determining k which required measurement of a temperature T_1 to 1−10 ppm accuracy (say) as well as a realization of T_0. Similarly, method (iv) will not be considered. Here r is eliminated from the Nyquist relation by determining the noise power from the product of the measured current noise and the measured emf noise in the resistance. This would seem to exchange a relatively easy measurement for a difficult one and so appear to be a bad strategy for determining k. We therefore consider only (i) the direct method and (ii) the correlation method.

4.1 The Direct Method

Here the noise emf's of two nominally equal noise resistances r_1 and r_0 at T_1 and T_0, respectively, are compared by switching an amplifier between them, squaring or rectifying the outputs, and time averaging them to yield $<V^2(T_0)$ and $<V^2(T)>$.

In this method it is necessary to measure or otherwise determine the contribution v_e to V arising from amplifier noise, noise in the electrical leads to the resistors, and possibly switch noise. Similarly it is essential to ensure that gain and bandwidth are equal for both r_1 and r_0 though this is not unduly problematic since nominally $r_1 = r_0$. Neither is the stability of the amplifier, squarer, and averaging circuit since sufficiently rapid

switching between r_1 and r_0 removes drift errors in a comparison. Measurement of r_1 and r_0 is in principal simple to accuracies approaching 1 ppm, but where two-lead systems are employed, perhaps at high temperatures with ensuing problems of insulation leakage, it may become problematic.

With this method of noise thermometry it is not possible in practice to calculate v_e sufficiently accurately from the measured parameters of an amplifying device such as a FET. The obvious method to measure amplifier/lead noise (usually expressed as an equivalent input series resistance r_e at T_0, perhaps combined with a noise current source in parallel with the input) is by short-circuiting the input. But this usually causes the gain of the amplifier to change drastically and so yields an unsuitable value for correcting measured source noise. More promising is the method of injecting a known white noise current from a high impedance source through r_0 at T_0 which yields information enabling r_e to be calculated. Alternatively noise from two or more resistances at T_0 can be compared if a reliable model is available of the amplifier performance as a function of input impedance. This last approach was adopted by Crovini and Actis [48] when employing the direct method between 630 and 962 °C and affords the best available historical indication of what might be expected in a k-determination of this kind.

These workers analyzed their JFET amplifier in terms of an equivalent circuit of a sufficiently general kind in which the intrinsic amplifier noise is represented at the input by a series noise resistor r_a generating noise emf $\langle e_n^2 \rangle = 4kT_0 r_a B$, and a parallel resistor r_b generating a partially correlated noise current $\langle i_n^2 \rangle = 4kT_0 B/r_b$. The point of employing an equivalent circuit with two sources is to make the parameters of the equivalent circuit, r_a and r_b, independent of the thermometric source resistance r_s which is not possible with a single equivalent amplifier noise resistance model. The noise figure for the amplifier may be written as

$$F = 1 + r_a/r_s + r_s/r_b + 2Re(\gamma)\ (r_a/r_b)^{1/2}$$

where Re(γ) is the real part of the complex correlation coefficient between e_n and i_n. Correspondingly, if r_s is small compared to other parallel input impedances the measured output will be

$$\langle V^2 \rangle \sim 4kT_0 B r_s + \langle e_n^2 \rangle + \langle i_n^2 \rangle r_s^2 + 2Re(\gamma)\ (e_n^2 i_n^2)^{1/2} r_s, \qquad (9)$$

which is a quadratic in r_s. Using Van der Ziel's [49] JFET expressions $r_a = 2/3\ g_m$ and $r_b = 2kT/eI_g$, where g_m is the transconductance, e the electronic charge, and I_g the dc gate current, Crovini and Actis calculated that the term in r_s^2 would be negligible compared to $\langle e_n^2 \rangle$ for their amplifier. The same was true for the last term on the right hand side since the correlation coefficient γ will be entirely imaginary. Thus only $\langle e_n^2 \rangle$ remains as an error. These conclusions were checked by measuring $\langle V^2 \rangle$ for different values of r_s at T_0 relative to a fixed value of r_s at T_0. Such a $\langle V^2(r) \rangle$ isotherm shows firstly the reliability of the assumption that $\langle V^2 \rangle \sim 4kT_0 B r_s + \langle e_n^2 \rangle$ and secondly gives some indication of the statistical accuracy which is achievable in a k-determination by this method. Crovini and Actis found that in a seven point resistance isotherm (r_s = 300 − 1000 ohms approximately) an rms deviation of about 100 ppm from a linear fit of $\langle V^2 \rangle$ vs. r_s was obtained. No significant statistical improvement could be obtained with a quadratic fit and it was not possible to observe consistent quadratic coefficients in successive trials. Thus it could be concluded that the linear approximation for $\langle V^2 \rangle$ is adequate to within the resolution of the measurements. The statistical accuracy of the line fitted to the points will be approximately 100 ppm/$n^{1/2}$ where n is the number of degrees of freedom, four in this case. Thus noise measurements of an attainable statistical accuracy of approximately 50 ppm would be suggested by this work. This is comparable with the performance of the limiting density and acoustic methods. Doubtless in an experiment specially designed to measure k, this performance might be improved. The fractional statistical accuracy theoretically attainable in a noise measurement of this kind on a single resistance is given by $[t(f_2 - f_1)]^{-1/2}$ for an integration time t. Thus a 10 ppm statistical accuracy should readily be achieved in about 28 hours.

While the work of Crovini and Actis indicates that a statistical resolution comparable with that of the more conventional methods should be possible, there remains the problem of the absolute determination of $A(j\omega)$. This might well involve the evaluation of an attenuation vs. frequency curve to an accuracy sufficient to evaluate the area beneath the curve to < 10 ppm. The measurement of frequency contributes no appreciable error and the measurement of the curve-height at a given frequency should be feasible to an accuracy of about 1 ppm. Thus the crucial $A(j\omega)$ measurement should not present a source of error significantly larger than that expected statistically. It is important, however, that the instrumentation employed should remain stable at the 1 ppm level of accuracy throughout the whole integration period. Again, this is attainable with careful temperature control both of gain and bandwidth controlling components and also of those components in the first amplification stage determining amplifer noise.

It seems likely, therefore, that a determination of k by the direct noise method could yield a more accurate result than hitherto obtained by conventional means.

4.2 The Correlation Method

In this method the source noise voltage is fed via two separate pairs of leads to two identical amplifiers whose output signals are multiplied together, squared, and time-averaged [50]. Since the respective noise voltages in the amplifier input leads and first stages will be completely uncorrelated, they will not contribute to the output. The respective source resistance noise voltages, on the other hand, will be perfectly correlated and so will in principle yield $\langle V^2 \rangle$ without error. The transfer functions $A_1(j\omega)$ and $A_2(j\omega)$ of each channel have to be measured and remain constant as with method (i).

Storm [51] has undertaken a preliminary error analysis of this method prior to carrying out a determination of k and Klein, Klempt, and Storm [52] have described in detail correlation noise thermometry measurements in the liquid helium range. The latter work produced an isotherm analogous to that of Crovini and Actis where a noise temperature of about 2.1 K was plotted as a function of resistance in the range 2.5−15 kΩ. It was found that a statistical error of 0.25 mK was obtainable representing approximately 100 ppm resolution on a noise signal which was many times smaller than that of Crovini and Actis. The reason why Klein *et al.* plotted a noise isotherm was not, as with Crovini and Actis, to determine the amplifier noise which arose mainly from the FET channel noise. It was rather to control other noise sources which are much smaller, but nevertheless

significant at such low temperatures and which the correlation method does not eliminate. These can arise from current noise in the separate source resistance lead pairs which, because the amplifier input impedances, though large, are finite, generate currents in the source resistor. The resulting voltages are presented symmetrically to both amplifiers and so are not eliminated. These sources of noise were shown to be negligible by Crovini and Actis for their instrument operating at much higher temperatures. We may assume, too, that they would cease to be significant for a correlation measurement at 273.16 K. Were it possible to achieve the resolution of Klein *et al.* at this temperature a statistical uncertainty of about 1 ppm would be implied.

Storm's analysis assumes a measuring time of 5×10^6s (about 8 weeks) with a 20 kHz bandwidth implying a 3 ppm statistical noise measurement accuracy compatible with the figure suggested by the practical performance achieved by Klein *et al.* To make use of such a resolution it is necessary that amplifiers, attenuators, and other electronic units can be calibrated and proved to hold their calibration at a comparable level of accuracy for this length of time. Taking these problems into account and combining uncertainties in quadrature, Storm's analysis suggests an attainable uncertainty in k of only 8 ppm.

5. Determination of k and Hence R by the Radiometric Method

The history of direct radiometric determinations of the Stefan-Boltzmann constant suggests that it is the least accurate way of arriving at a value for R. The most recent and best completed determination is due to Blevin and Brown [53] who reported the value $\sigma = (5.6644 \pm 0.0075) \times 10^{-8}\mathrm{W \cdot m^{-2} \cdot K^{-4}}$ implying values for k and R some 230 ppm below those derived from limiting density or acoustic measurements. The corresponding uncertainty would be about 110 ppm (one standard error). This is considerably better than previous radiometric determinations, details of which can be obtained from the discussion of Blevin and Brown, but disappointing compared to other methods. However, interpreted differently, the experiment yielded useful thermometric information on the thermodynamic temperature of the gold point by using the values of the Stefan-Boltzmann constant implied by the accepted values of k and R. (See the introductory remarks of Section 1 where a knowledge of σ was argued to be equivalent to a water triple point calibration.)

These workers employed a near-black cavity source at the gold point because its high radiance at that temperature enabled more accurate intensity measurements to be made for a given radiometer sensitivity and rendered the parasitic effect of stray radiation less troublesome. Moreover at these temperatures most of the energy is concentrated in the near infrared where highly absorbing surfaces can be prepared more easily than for the far infrared. The radiant power arriving at the radiometer was limited by apertures to define its solid angle and measured by the substitution method in which an equivalent amount of electrical power is substituted for the radiant power. Measurement of the latter then yields a value for the former. The more important sources of uncertainty were those in the thermodynamic temperature of the gold point (equivalent to 0.06% in σ), in the realization of the gold point by calibrated thermocouple (0.06% and 0.03% in σ due respectively to calibration and thermal contact uncertainties), in losses from diffraction (0.06% in σ) and in the lateral non-uniformity of the receiver (0.05% in σ). These and other lesser uncertainties were combined in quadrature.

In a new primary radiometric measurement, Quinn and Martin working at the NPL are hoping to avoid many of the larger sources of uncertainty incurred in the above work and so to produce a measurement of σ which is competitive with other methods. They give a full description of their work elsewhere in these proceedings so that only the main points need be mentioned here. Firstly their radiant source is held at the triple point of water so avoiding uncertainties arising from primary thermometry, thermocouple calibrations, and thermal contact at very high temperatures. This naturally gives rise to problems of distinguishing between radiation entering the radiometer from the source blackbody and from other components at ambient temperatures. The method also makes stringent demands on the sensitivity of the radiometer since the radiance, varying as T^4, is much lower at 273.16 K than at the gold point (1337 K). The problem of stray radiation is met by providing the radiometer mouth with a thermally dead hemisphere except for the source aperture. This is achieved by using a surrounding black radiation trap held at liquid helium temperature. Sensitivity is provided by operating the radiometer at about 1 K above the temperature of a bath of superfluid helium at about 2 K to which it is connected by a thermal resistance. The temperature difference across this thermal resistance can be measured to a few microkelvins using germanium resistance thermometers. This is facilitated by the extreme temperature stability of the superfluid helium bath attributable to the anomalously high thermal diffusivity in the superfluid state. As with Blevin and Brown's instrument, radiant flux is measured by the substitution method, with the incidental advantage in this case that superconducting connections can be made to the radiometer's electrical heater, thus minimizing errors due to dissipation in the current leads.

Quinn and Martin hope by this means to measure the Stefan-Boltzmann constant to about 100 ppm equivalent to about 25 ppm in R and k. If this is achieved, their measurements will be comparable in accuracy to those obtained by the conventional method of limiting density.

Another radiometric determination of the Stefan-Boltzmann constant is being made by Ono adopting what is essentially the method of Blevin and Brown. He intends to avoid their larger thermometric errors by operating with a source at the 693 K zinc point whose exact thermodynamic temperature has recently been determined by Guildner *et al.* [22] by means of highly developed gas thermometry. Ono expects to achieve an accuracy comparable to that of Quinn and Martin. Again, interested readers are referred to Ono's contribution to these proceedings for details.

6. Current Best Values and Possibilities for the Future

The currently accepted "best" value for R, $8.31441\ \mathrm{J \cdot K^{-1} \cdot mol^{-1}}$, is that recalculated by Batuecas [2, 11, 13] from three sets of results (see Table 1). In this recalculation the results obtained by Baxter *et al.* and Moles *et al.* were processed in two ways to yield four values of R from two determinations. The first method of processing simply applied a linear fit to the $L(p)$ measurements to obtain $L(0)$ by extrapolation. The second used only $L(p)$ values taken near P_0 and corrected to zero pressure by employing a value for the second virial coefficient of oxygen critically evaluated by Cragoe [54]

TABLE 4. *Recalculation of the gas constant.*[a]

Source	Method	Value ($J \cdot K^{-1} \cdot mol^{-1}$)	Uncertainty (ppm)
Baxter *et al.* [5, 6]	LD, O_2	8.31463 *	27
Moles *et al.* [8]	LD, O_2	8.31465 *	24
Batuecas *et al.* [9, 10, 11]	LD, O_2	8.31518 *	66
Pereira [22]	LD, N_2	8.31417	75
Colclough *et al.* [32]	Acoustic	8.31448	25
Colclough [37]	Acoustic (indirect)	8.31447	81

[a]In the case of entries marked * an estimate was made for systematic uncertainty in each measured value of $L(p)$ from the considerations of Section 2. This was then combined in quadrature with the random uncertainty (3 std. deviations) and used with the extrapolation uncertainty to calculate a weighting factor for the value of $L(0)$ derived from each value of $L(p)$. A weighted mean of these $L(0)$ values then provided a value for R from each of the three determinations which arguably reflected the variation of accuracy with pressure in the original $L(p)$ measurements. No statistical errors were available for the work of Moles *et al.* and weights had to be obtained by interpolation between those estimated for Baxter *et al.* The uncertainties quoted above represent three std. deviations in R as calculated from the weighted $L(0)$ means. There is no significant difference between the weighted and unweighted means of the above six values of R, which have respective standard deviations of 9 ppm and 40 ppm.

from the literature of relatively high pressure PVT measurements. This value was expressed by Batuecas as $A_0^1 = (9.51 \pm 0.9) \times 10^{-4}$ in a form which gives the fractional correction for nonideality at P_0. It can be seen that its uncertainty is equivalent to about 90 ppm in R. With these four values he combined his own to provide a mean recommended value. The work by Baxter and Starkweather [7] employing nitrogen was rejected. This best value of R is open to the criticisms that it overweights the two earlier measurements compared to Batuecas's, that it underweights $L(p)$ values taken below atmospheric pressure, and that better values for A_0^1 are now available than that derived by Cragoe. We have recalculated a new "best" value using all the $L(p)$ values of the three sets of results as re-evaluated by Batuecas and correcting each to zero pressure using the value $A_0^1 = (9.89 \pm 0.13) \times 10^{-4}$ derived from Weber's critical evaluation of the PVT literature [55]. Weber's work also shows that the effect of the third virial coefficient is equivalent to a change of only about 1 ppm in R. To these three values we have added the new limiting density value of Pereira obtained with nitrogen at $p = P_0$ and using an external value of A_0^1 derived from high pressure measurements. Two acoustic values due to Colclough *et al.* and Colclough [32, 37] have been taken into account as well. In all cases an attempt has been made to estimate systematic and random errors including those due to extrapolation. Where past work does not enable this to be done easily, informed guesses have been made on the basis of the considerations of Section 2. The calcuation is summarized in Table 4 and the final answers are

$$R = (8314.60 \pm 0.22)\ J \cdot K^{-1} \cdot kmol^{-1}$$
$$V_0 = (22.41433 \pm 0.00059)\ dm^3 \cdot mol^{-1}$$
$$k = (1.380665 \pm 0.00037) \times 10^{-23}\ J \cdot K^{-1}$$
$$\sigma = (5.67020 \pm 0.00060) \times 10^{-8}\ W \cdot m^{-2} \cdot K^{-4}$$

In every case three "standard error" uncertainties have been quoted reflecting both random and systematic errors. It is of interest to note that a value of R some 32 ppm higher than Batuecas's best value would be obtained by applying our method of estimation to his three chosen determinations. It seems that personal judgment in such matters has a much larger effect on best values than the systematic effects discussed in Section 2. Our best value of R is 23 ppm higher than Batuecas's.

In Sections 2 to 5 plausible error analyses have been given for the four practicable methods of determining the gas constant. The method of limiting density which produces the currently accepted value was shown to be subject to systematic effects, notably due to sorption, which were not fully assessed in the reports of the historical determinations. Nor can they easily be assessed in retrospect. However, it has been shown that they could well be as small as 5 ppm or less and the agreement between values obtained from acoustic and limiting density measurements suggests that the accepted value is free from systematic errors greater than about 30 ppm. It is possible that this may be confirmed in the near future by radiometric measurements.

Appreciable reduction of the uncertainty in R will only come from considerably refined limiting density, acoustic, or noise measurements. Of particular interest in this connection are the projected limiting density measurements of Guildner *et al.* at the NBS, and the projected noise measurements of Storm at the Universitaet Muenster. Both parties anticipate root-sum-square uncertainties below 10 ppm. In the former case an advance uncertainty of 3 ppm has been estimated at the 99% confidence limit. This would be approximately equivalent to the 1 ppm standard deviation which was spoken of as the uncertainty limit at which these constants could begin to influence the values of the other fundamental constants through Eq. (1). However, before this can happen it will be necessary for a determination of the Stefan-Boltzmann constant to be made to about 15 ppm as well. (The current uncertainty in Planck's constant is 5.4 ppm and that in the velocity of light is 0.004 ppm.) As will have become apparent from Section 5, this is at the moment beyond the aspirations of even the most ambitious experimentalist. Thus it seems that the promotion of the thermal constants from the category of "other less precise quantities" must be deferred for the present.

References

[1] I. W. Roxburgh, in *The Encyclopaedia of Ignorance*, Ed. by R. Duncan and M. Weston-Smith (Pergamon Press, Oxford, 1977), p. 38.
[2] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[3] M. Born, *Atomic Physics*, Translated by J. Dougall, revised by R. J. Blin-Stoyle, (Blackie and Son Ltd., London, Glasgow, 1964), Chpt. VIII, p. 1.
[4] W. R. Blevin and W. J. Brown, Metrologia 7, 15 (1971).
[5] G. P. Baxter and M. W. Starkweather, Proc. Natl. Acad. Sci. U.S.A. 10, 479 (1924).
[6] G. P. Baxter and M. W. Starkweather, Proc. Natl. Acad. Sci. U.S.A. 12, 699 (1926).
[7] G. P. Baxter and M. W. Starkweather, Proc. Natl. Acad. Sci. U.S.A. 12, 703 (1926).
[8] E. Moles, "Les Determinations Physico-Chimiques des Poids Moleculaires et Atomiques des Gaz." Collections Scientifique d'Institute Internationale Cooperative Intellectuel, pp. 1–75 Paris (1939).
[9] T. Batuecas and G. Garcia Malde, An. Fis. Quim. 46, (1950).
[10] F. L. Casado and T. Batuecas, An. Fis. Quim. 48, 4 (1952).
[11] T. Batuecas, Rev. Real. Acad. Ciencias 66, 427 (1972).
[12] J. M. Pereira, private communication (1978).
[13] T. Batuecas, in *Atomic Masses and Fundamental Constants 4*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, New York, 1972), p. 534.
[14] A. H. Cook and N. W. B. Stone, Philos. Trans. R. Soc. London 250, 279 (1957).
[15] A. H. Cook, Philos. Trans. R. Soc. London 254, 125 (1961).
[16] J. A. Beattie, B. E. Blaisdell, and J. Kaye, Proc. Am. Acad. Arts Sci. 74, 389 (1941).
[17] G. W. C. Kaye and T. M. Laby, *Tables of Physical and Chemical Constants*, 14th edition (Longman, London, 1973).
[18] F. A. Gould and T. Vickers, J. Sci. Instrum. 29, 85 (1952).
[19] M. Menache, Metrologia 3, 58 (1967).
[20] International Union of Pure and Applied Chemistry, Pure and Appl. Chem. 45, 1 (1976).
[21] P. H. Bigg, Br. J. Appl. Phys. 18, 521 (1976).
[22] L. A. Guildner and R. E. Edsinger, J. Res. Natl. Bur. Stand. Sect. A: 80, 703 (1976).
[23] G. F. Weston, Vacuum 25, 469 (1975).
[24] R. J. Elsey, Vacuum 25, 299 (1975).
[25] L. Holland, *The Properties of Glass Surfaces* (Chapman Hall, London, (1964).
[26] R. J. Elsey, Vacuum 25, 347 (1975).
[27] E. Moles, Bull. Soc. Chim. Belg. 47, 405 (1938).
[28] E. Moles, T. Toral, and A. Escribano, Trans. Faraday Soc. 35, 1439 (1939).
[29] G. A. R. Hartley, T. M. Henry, and R. Whytlaw-Gray, Trans. Faraday Soc. 35, 1452 (1939).
[30] T. W. Hickmott, J. Appl. Phys. 31, 128 (1959).
[31] E. Moles and R. Miravalles, J. Chim. Phys. Phys. Chim. Biol. 21, 1 (1924).
[32] A. R. Colclough, T. J. Quinn, and T. R. D. Chandler, Proc. R. Soc. London, Ser. A: 368, 125 (1979).
[33] T. J. Quinn, A. R. Colclough and T. R. D. Chandler, Philos. Trans. R. Soc. London, Ser. A: 283, 367 (1976).
[34] A. R. Colclough, Metrologia 9, 75 (1973).
[35] A. R. Colclough, Acustica 36, 259 (1976/77).
[36] H. H. Plumb, private communication.
[37] A. R. Colclough, Proc. R. Soc. London, Ser. A: 365, 349 (1979).
[38] A. R. Colclough, Acustica 42, 18 (1979).
[39] J. B. Mehl and M. R. Moldover, J. Chem. Phys. 74, 4062 (1981).
[40] F. D. Shields, K. P. Lee, and W. J. Wiley, J. Acoust. Soc. Am. 37, 4 (1965).
[41] P. S. Henry, Proc. Phys. Soc. London 43, 340 (1931).
[42] D. E. Weston, Proc. Phys. Soc. London, Sect. B 66, 695 (1953).
[43] J. Blatt, J. Phys. A 9, 8 (1976).
[44] B. E. Gammon, J. Chem. Phys. 64, 2556 (1976).
[45] R. A. Kamper, in *Temperature—Its Measurement and Control in Science and Industry*, Ed. by H. H. Plumb (Instrument Society of America, Pittsburgh, 1972), Pt. 1, p. 349.
[46] J. B. Garrison and A. W. Lawson, Rev. Sci. Instrum. 20, 785 (1949).
[47] C. J. Borkowski and T. V. Blalock, Rev. Sci. Instrum 45, 151 (1974).
[48] L. Crovini and A. Actis, Metrologia 14, 69 (1978).
[49] A. Van der Ziel, Proc. IEEE 51, 461 (1963).
[50] A. Van der Ziel, *Noise: Sources, Characterization, Measurement* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1970).
[51] L. Storm, private communication.
[52] H. H. Klein, G. Klempt, and L. Storm, Metrologia 15, 143 (1979).
[53] W. R. Blevin and W. J. Brown, Metrologia 7, 15 (1971).
[54] C. S. Cragoe, J. Res. Natl. Bur. Stand. 26, 495 (1941).
[55] L. A. Weber, J. Res. Natl. Bur. Stand. Sect. A: 74, 93 (1970).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

An Ultrasonic Determination of the Gas Constant

William C. Sauder

Department of Physics, VMI, Lexington, VA 24450

Progress on a gas constant determination by means of ultrasonic interferometry is reported. The acoustic interferometer is a two arm instrument designed to operate in the range 0.1–1.0 MHz. An approximately half-scale model has been constructed in order to establish the data taking algorithm as well as to investigate aberrations. Electrostatic transducers have been designed for the experiment that will allow characterization of the acoustic field, a necessary step in extracting wave length measurements from the fringe data.

Key words: acoustic interferometry; gas constant; ultrasonic transducers.

1. Introduction

Prior to the middle of the last decade, the value of the gas constant had been established by applying data from limiting density measurements on oxygen to the ideal gas law. Then, in 1974, Quinn, Chandler, and Colclough [1] published initial results from their acoustical determination of the constant. In this case the velocity of sound in argon was measured at pressures between 0.03 and 1.3 MPa using a classical standing wave acoustic interferometer operated at low frequency (5.6 kHz).

In this paper we report progress on an acoustic determination of R to be carried out with a new form of interferometer operating in the ultrasonic domain. We will discuss the advantages and disadvantages of the method and describe the still-evolving design of the interferometer. Estimates of the expected performance will be set forth.

2. The Method

When the propagation of sound through an ideal gas is an exactly adiabatic process, the gas constant can be related to the sound velocity by the well-known relation

$$R = \frac{Mv^2}{\gamma T}, \tag{1}$$

where v is the sound velocity, M and T are the molecular weight and temperature of the gas, respectively, and γ is the ratio of specific heats.

We plan to measure v in a variety of rare gases using an acoustic instrument that rather resembles the Froome microwave version of the Mach-Zehnder interferometer [2]. A schematic drawing of the device is shown in Fig. 1. Two transducers, mounted back-to-back on a moving carriage, are excited simultaneously by a sinusiodal voltage at ultrasonic frequencies. Signals from stationary microphone transducers at either end of the instrument are combined electronically to form a fringe signal that varies sinusoidally with carriage position, having a periodicity of approximately $\lambda/2$ (where λ is the wavelength of sound in the gas). A plot of the position of the carriage at fringe peaks vs. the fringe number, when corrected appropriately for Fresnel diffraction effects and variations of temperature, will be a straight line having a slope of $\lambda/2$.

We will operate at ultrasonic frequencies in the range 0.1–1.0 MHz. Thus λ will be sufficiently short in comparison with dimensions of the instrument to ensure free space propagation of the acoustic waves. This will obviate the necessity of making significant boundary layer corrections to the measured velocities as would be necessary in low frequency interferometers characterized by a large wavelength to dimension ratio [3]. On the other hand, we will not operate at high enough frequencies to require the large dispersion corrections that would be necessary if the wavelength were comparable with the mean free path of the gas molecules [4].

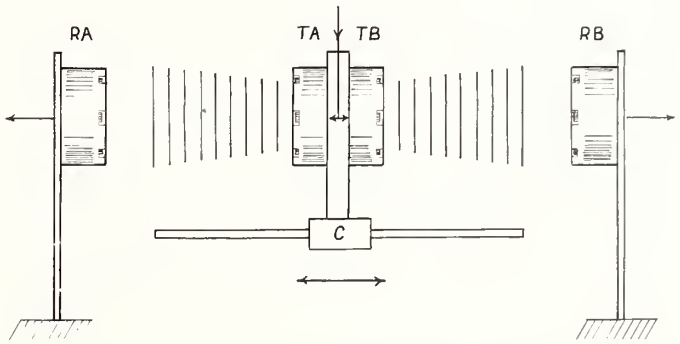


FIGURE 1. *Schematic drawing of the acoustic interferometer. TA, TB: sending transducers; RA, RB: receiving transducers; C: movable carriage.*

In order for us to extract the gas constant from velocity data as implied in Eq. (1), it will be necessary to make runs at several pressures and then extrapolate to zero pressure. The sound velocity will vary with pressure because of non-ideal behavior of the gas sample. In addition, dispersion effects resulting from viscosity and heat conduction are pressure dependent [5].

We will look for other systematic effects by varying experimental parameters, e.g., by using a variety of sample gases (including isotopically enriched gases), and by operating at several different frequencies in order to bring out the frequency dependence of the near field diffraction effects and of the virial coefficients that characterize non-ideality.

3. Progress of the Experiment

In order to study the characteristic behavior of the double arm acoustic interferometer, we have constructed an approximately half-scale model. The model interfer-

ometer operates in air rather than being enclosed in a vacuum-tight tank as will be the case for the final instrument. Displacement of the carriage is measured by counting turns of a lead screw, whereas in the final apparatus, carriage displacement will be determined by means of optical interferometry.

We anticipate that this model interferometer will be capable of wavelength (and therefore velocity) measurements with an accuracy of 100 ppm or better. Though we will use the instrument to make acoustic velocity measurements for air in order to confirm our planned procedure in the final experiment, the principal use of the model will be for studying diffraction effects and possible errors introduced by standing waves in the interferometer arms.

We can already discern some aspects of the design of the final interferometer. The instrument will be contained within a 1-meter long by 0.5 m diameter tank. The extent of carriage travel possible will allow scanning of 200 to 4000 fringes during a run, depending upon the frequency and the sample gas employed. The instrument will be submerged in an ice bath, so that temperatures will be stable to within a few mK. Carriage travel will be determined to within 0.01 fringe (for 6328 Å illumination of the optical interferometer). We expect standard deviations of the order of 1 ppm in the velocity determinations.

Because our apparatus employs free space propagation of acoustic waves, it is necessary that we establish with certainty the shape of the acoustical wavefronts. Otherwise, it would not be possible to translate the data on fringe spacing into a determination of wavelength. As a result, we are devoting considerable effort to the design of our transducers and the characterization of their vibration.

The transducers are of the electrostatic type investigated extensively by Kuhl, Schodder, and Schröder [6]. As shown in Fig. 2, the backing plate of our transducer is a gold coated infra-red diffraction grating having a grating spacing of the order of 0.1 mm. A 7 μm gold coated Kapton[1] polyamide film is made to adhere to the ridges of the grating by applying a dc bias voltage between the evaporated gold coating on the film and that on the grating. A high frequency voltage superimposed on the bias then produces mass-limited vibration of the rectangular strips of film between these lines of contact. The transducers are operated near resonance.

In order to ensure that the transducers are operating properly, we periodically will carry out frequency calibrations to establish the fact that the frequency response agrees with the theoretically predicted response. Furthermore, by incorporating a transducer into one arm of an optical Michelson interferometer by employing the transducer face as one of the interferometer mirrors, we will be able to examine the variation of phase of vibration over the face by studying the photoelectrically detected interferometer output signal as the optical beam is moved over the transducer face. Finally, knowing the vibration behavior of the transducer face, we will perform a computer integration of Kirchhoff's formula [7] in order to establish the shape of wave fronts in the acoustic wave field.

Because much information in addition to data on carriage position and the phase of the acoustical interferometer signals must be recorded, the instrument will be operated under microcomputer control. During a typical run, the computer will record the amplitude from each arm of the interferometer, the acoustic fringe signal, carriage position, temperature, pressure, and frequency. A block diagram of a portion of the microcomputer interface is presented in Fig. 3. At the completion of the run, these data, having been stored on magnetic floppy disk, will then be transmitted to our computer center via a telephone link and stored on magnetic tape. Subsequently, the data will be smoothed and then compared to the theoretically predicted fringe pattern in order to determine a velocity for the run.

[1] A trademark of E. I. du Pont de Nemours & Co., Inc.

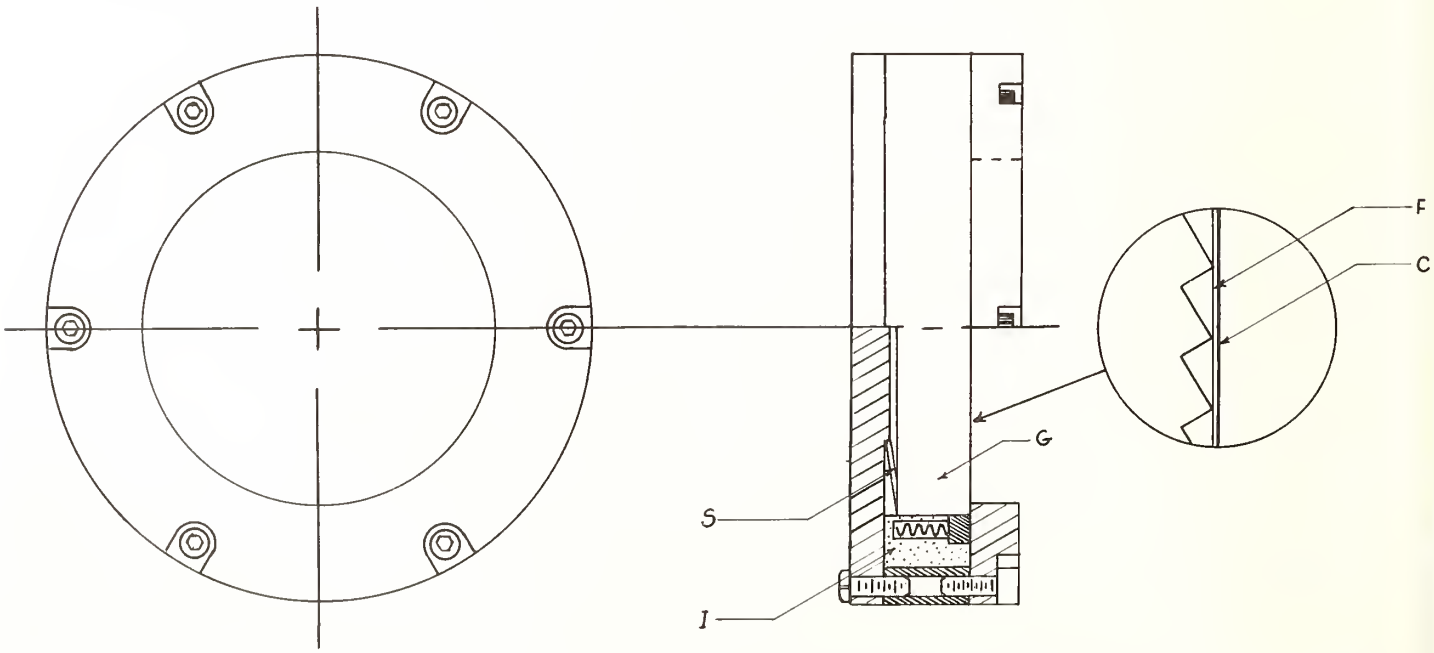


FIGURE 2. *Grating backed electrostatic transducer. G: grating; F: polyamide film; C: evaporated gold coating; S: spring (Schnorr washer); I: insulator.*

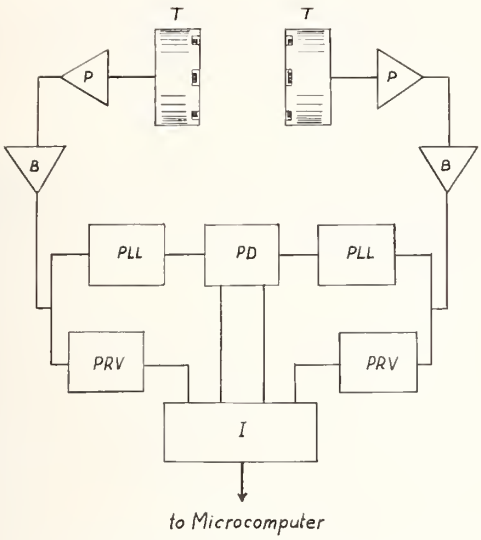


FIGURE 3. *Block diagram of a portion of the interface. T: transducer; P: preamplifier; B: band pass amplifier; PLL: phase-locked loop; PD: phase detector; PRV: peak-reading voltmeter; I: interface for microcomputer bus.*

Work on the model interferometer and its computer control as well as the development of the grating transducers has progressed sufficiently far that we would expect to be taking data within a year or two.

4. Conclusion

Because the gas constant is related to the body of atomic constants through the Avogadro constant and Boltzmann's constant, and since the latter is still a "weak link"[2], the only method available for uncovering systematic errors in gas constant determinations is to measure the constant by several methods. The acoustic determination is then a check on possible systematic errors in the old physical chemistry gas density determinations. The completion of the high frequency experiment described in this paper should establish the viability of the acoustic methods when compared with the results of Quinn *et al.* One would hope that then we could have considerable confidence in future "best values" of the gas constant.

This research has been and is being carried out with funding from a Research Corporation Grant and a National Bureau of Standards Precision Measurement Grant. I wish to acknowledge the help of my students, Ken Yates and Kim-Ying Chu, who, as undergraduates, are learning about research in general and about the joys and frustrations of precision measurements in particular.

References

[1] T. J. Quinn, T. R. D. Chandler, and A. R. Colclough, Nature **250**, 218 (1974).
[2] K. D. Froome, Proc. R. Soc. London, Ser. A: **247**, 109 (1958).
[3] D. H. Smith and R. G. Harlow, Brit. J. Appl. Phys. 14, 102 (1936).
[4] A. B. Bhatia, *Ultrasonic Absorption* (Oxford Univ. Press, London, 1967) pp. 59–64.
[5] A. B. Bhatia, *ibid.*, p. 61.
[6] W. Kuhl, G. R. Schodder, and F. K. Schröder, Acustica 4, 519 (1964).
[7] S. A. Schelkunoff, Phys. Rev. **56**, 308 (1939).

---

[2]But see the paper by T. J. Quinn and J. E. Martin in these proceedings.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Spherical Acoustic Resonators: Promising Tools for Thermometry and Measurement of the Gas Constant

M. R. Moldover and J. B. Mehl*

Thermophysics Division, National Bureau of Standards, Washington, DC 20234

We report progress in our study of the feasibility of using spherical acoustic resonators for primary thermometry and for measurement of the gas constant. Prototype resonators of differing sizes and materials have been tested. Under conditions of metrological interest (argon, 273 K, 0.1–0.5 MPa, 13 cm diameter resonator), the radially symmetric resonances have halfwidths, g_p, which are on the order of 2×10^{-4} of the resonance frequencies f_p. Instrumental errors in the measurement of f_p and g_p are on the order of $10^{-7}\ f_p$. The ratios of several measured resonance frequencies, extrapolated to zero pressure, are within $\pm 3 \times 10^{-6}$ of those expected from our theoretical model. For the same resonances, the measured values of g_p exceed those calculated using nonacoustic data by less than 5×10^{-6} f_p. These numbers suggest that an acoustic measurement of the thermodynamic temperature or the gas constant could be accurate to 10 parts in 10^6. Further improvements seem probable.

Key words: acoustic resonator; acoustic thermometry; gas constant; speed of sound; thermodynamic temperature scale; thermometry.

1. Introduction

Accurate speed of sound measurements in dilute helium have been used to study the relationship between the thermodynamic temperature scale and practical temperature scales in the 4–20 K range [1, 2]. Recently, speed of sound measurements in dilute argon have yielded a value for the gas constant, R, with an "overall uncertainty" of 25 parts in 10^6 [3]. In each of these measurements a cylindrical acoustic interferometer was used. In the present work, we explore the feasibility of applying a spherical acoustic resonator to similar problems in metrology. We shall present evidence that the errors in the theoretical description of real spherical resonators can be less than 5 parts in 10^6. This implies than an acoustic measurement of the thermodynamic temperature or the gas constant could be accurate to 10 parts in 10^6. Further improvement is very likely. We conclude that spherical resonators offer significant practical and theoretical advantages in comparison with cylindrical resonators for applying speed of sound measurements to problems in metrology.

Significant practical advantages arise from the exceptionally high Q of the radially symmetric acoustic resonances. For these modes, the gas motion is perpendicular to the sphere's walls; thus it is not subject to viscous damping at the walls. When our 12.7 cm diameter resonator is filled with argon at 273 K in the range 0.1–0.5 MPa, the halfwidths, g_p, of the first four radial resonances range from 0.5×10^{-4} to 2.0×10^{-4} of the resonance frequencies, f_p. These resonances are a factor of 10 narrower than the resonances of a cylindrical acoustic interferometer which were used in a recent measurement of the gas constant [3]. Thus the perturbations to the resonance frequencies arising from the gas-resonator wall interactions are an order of magnitude smaller in the spherical resonator than in the cylindrical resonator.

The strong signals associated with the radially symmetric modes permit the use of very small transducers which perturb the resonator geometry in a minor and predictable way.

The symmetry of the spherical resonator facilitates the detailed calculation of many perturbations. In particular one can show that the frequencies of the radial modes are insensitive to geometrical imperfections to first order in perturbation theory. Thus for practical purposes, one can construct a spherical resonator with modest tolerances and measure its volume; a detailed map of geometrical imperfections is not necessary. It is possible to model analytically the elastic response of the shell of the spherical resonator to the vibrations of the gas within it.

The application of spherical resonators to the measurement of the speed of sound in dilute gases has been discussed by Bancroft [4], Keolian *et al.* [5, 6], Moldover *et al.* [7], and by Mehl and Moldover [8]. The latter reference contains the most complete discussion of the theory of such resonators in the present context.

Much of the apparatus and techniques used for this work is identical with that described by Mehl and Moldover [8]. Here, we will emphasize the significant improvements in technique and the tests of our acoustic model that have been made subsequent to that work.

2. Experimental Techniques

Two resonators have been used in this work. The first is an aluminum sphere (designated A) with an inside diameter of 12.7 cm and a wall thickness of 1.27 cm. The second is a brass sphere (designated B) with an inside diameter of 6.35 cm and a wall thickness of 1.27 cm. Both resonators were contained within a single pressure vessel and thermostat; thus they were studied in identical gas samples at identical temperatures.

Resonator A is essentially the same one described in Ref. [8]. A variety of minor changes in its construction were made in an effort to improve its performance. (As a measure of resonator performance we customarily use the comparatively easily measured quantities $\Delta g_p / f_p$ where the Δg_p are the measured resonance halfwidths minus the calculated halfwidths). Polishing the interior of

*Physics Department, University of Delaware, Newark, DE 19711

the resonator did not significantly change its behavior. Neither did changing its support from a massive post to a weak spring. Lapping the two halves of the resonator together significantly improved the resonator's performance. The secondary source-detector coupling mentioned in Ref. [8] was drastically reduced by altering all three holes in the resonator. A new source transducer 0.35 cm in diameter was fabricated. This transducer was fitted in a snug hole through the resonator so that the active surface of the transducer was flush with the inner wall of the resonator. This eliminated the source transducer's coupling hole. The detector transducer was coupled to the resonator by a hole 0.08 cm in diameter and 0.64 cm long. (This required embedding the detector in the resonator wall.) Thus the resonant frequency of this coupling hole was pushed above the range of our measurements. Finally, the hole used to admit gas into the resonator was partially plugged so that it too had a diameter of 0.08 cm and an effective length of 0.64 cm.

Resonator B was also made in two halves. It was drilled with three holes spaced 90° apart. The outer portion of each hole was 0.31 cm in diameter and 0.95 cm long; the inner portion was 0.041 cm in diameter and 0.32 cm long. Transducers fit tightly in the outer portions of two holes; the third hole was used to fill the resonator with gas.

The data reported below have been taken at various temperatures between 15 and 25 °C. A platinum capsule thermometer indicated that the temperature of a specific point within the thermostat drifted less than 0.5 mK in one week. The temperature difference across the two resonators was less than 2 mK.

Both the argon and the propane were used as received from commercial suppliers. (The supplier claimed the minimum purity of the argon was 99.9999% and that of the propane was 99.94%.) We found that passing the argon through a molecular sieve trap cooled to −70 °C did not alter the resonators' performance.

To measure accurately f_p and g_p for each resonance, the resonator was excited successively at 11 distinct frequencies ranging from $f_p - g_p$ to $f_p + g_p$ in steps of $g_p/5$. At each frequency the in-phase voltage U and the quadrature voltage V at the receiving transducer were measured with a suitable tracking lock-in amplifier, scanner, and digital voltmeter under computer control. This process was then repeated with the frequency decreasing from $f_p + g_p$ to $f_p - g_p$. (The practice of taking data by stepping both up and down in frequency compensates for possible drifts in the gas temperature which are linear in time.) Then a function of the theoretically predicted form

$$U + iV = \frac{\overline{A}f}{(f_p + ig_p)^2 - f^2} + \overline{B} + \overline{C}f \qquad (1)$$

is fit to the data using a program described by Mehl [9]. Here the amplitudes $\overline{A}$, $\overline{B}$, and $\overline{C}$ are complex numbers. The last two terms in Eq. (1) represent contributions to the complex voltage $U + iV$ from the "tails" of all resonances other than the one of interest. Moldover *et al.* [7] demonstrated that the lowest 30 resonances in argon are fully resolved; hence $|\overline{B} + \overline{C}f|$ was usually less than a few percent of $|U + iV|$.

When the resonators are filled with propane the values of g_p encountered can be a factor of 10 smaller than those encountered with argon under similar conditions. For example, in propane at 287.7 K and 0.52 MPa the first radial mode of resonator A is at $f_1 = 2568$ Hz and has the halfwidth $g_1 = 0.0526$ Hz. Under these conditions we made an important check of the entire data acquisition system and the numerical methods used to fit the data. We excited this resonance, turned off the excitation, and then monitored the detected voltage as the resonator rang down. The value of g determined from the decay curve differed from that obtained by cw measurements by only 0.0004 Hz or $1.6 \times 10^{-7} f$.

3. Acoustic Model for Spherical Resonators

We will now review the acoustic model for obtaining the speed of sound, c, from the values of f_p and g_p which result from processing the raw data via Eq. (1). Important contributions to this model were made by Greenspan and will be published by him elsewhere [10]. (See also Ref. [8].) The starting point is the calculation of the eigenfrequencies of a lossless fluid confined in a rigid spherical enclosure by perfectly insulating walls. Then surface perturbation theory is used to calculate the complex corrections to the eigenfrequencies arising from four distinct sources. Finally, the bulk dissipation contribution to the resonance widths is considered. In the discussion of these phenomena we will provide numerical values for them which are appropriate for resonator A filled with argon at 273 K.

For the radial modes of a spherical resonator, the eigenfrequencies are found from the roots of the equation:

$$\frac{dj_0(2\pi fa/c)}{da} = 0. \qquad (2)$$

(Here a is the radius of the resonator and j_0 is the zeroth spherical Bessel function.) The frequencies of the first few solutions to Eq. (2) for resonator A with $c = 308$ m/s are:

p	f(Hz)
1	3469
2	5964
3	8418
4	10859
5	13294
6	15726

The first surface perturbation occurs because the resonator wall has both a high thermal diffusivity and a high heat capacity per unit volume compared with those of the gas. The thermal admittance of the walls leads to the perturbation

$$\Delta f_s + ig_s = (-1 + i)\frac{(\gamma - 1)}{2a}\sqrt{\frac{D_T f}{\pi}} \qquad (3)$$

(Here D_T is the thermal diffusivity of the gas which is equal to the thermal conductivity divided by the constant pressure heat capacity per unit volume, and γ is the ratio of the constant pressure heat capacity to the constant volume heat capacity.) For resonator A filled with argon, Eq. (3) is approximately

$$(\Delta f_s + ig_s)/f = 2.1 \times 10^{-4}(-1 + i)/\sqrt{P^* f^*}. \qquad (4)$$

(Here we choose to measure frequency in units of f_1 and pressure in units of 0.1 MPa, thus $f^* = f/3469$ Hz and $P^* = P/0.1$ MPa.)

Note that the perturbations given by Eq. (3) and subsequent equations are to be added to the tabulated "ideal" resonance frequencies to obtain predictions for the measured frequencies. This sign convention is opposite to that of Ref. [8].

A second perturbation occurs because the spherical

shell has a finite compliance. Greenspan [10] assumes that the shell vibrates in a perfectly elastic breathing motion in response to the radial vibrations of the gas within it. He finds for resonator A:

$$\frac{\Delta f_{el}}{f} = \frac{4.7 \times 10^{-6} \rho^* (c^*)^2}{1 - (f/20.2\ \text{kHz})^2} . \quad (5)$$

(Here ρ^* is the density of the gas measured in units of the density of argon at 273 K and 0.1 MPa and c^* is the speed of sound divided by the speed of argon at 273 K and 0.1 MPa.)

This expression diverges as the frequency approaches the frequency of the breathing mode of the shell, 20.2 kHz, where the approximations used to derive it are no longer accurate. Perhaps it is the failure of these approximations which leads to comparatively large pressure-dependent discrepancies between the present model and the data for the 7th and higher radial modes of resonator A (see below).

The third class of surface perturbations considered is that resulting from the holes drilled through the resonator wall. Such holes produce losses and frequency shifts which are complicated functions of frequency; however, the perturbations are very small unless the resonant frequency of the hole itself (considered as an open organ pipe) happens to coincide with one of the resonances of the spherical cavity. The result is:

$$\Delta f_h + i g_h = \frac{c r_0^2}{8 \pi a^3} \cot[2 \pi f L/c) + (1 + i) \alpha L] . \quad (6)$$

[(Here L is the effective length of the hole, r_0 is its radius, and α is the Kirchhoff-Helmholtz attenuation parameter. For argon we may use the approximation: $\alpha = 3.22\sqrt{f D_T}/(r c)$.]

As a final application of surface perturbation theory we consider geometrical imperfections in the shape of the resonator. It can be shown that the first order correction to the eigenfrequencies of the radial modes vanishes for all shape perturbations [11] which do not alter the resonator's volume. As an example of a typical second order perturbation we quote Greenspan's result for a spheroidal deformation such that one axis of the spheroid becomes $a(1 + \epsilon)$ while the other two axes become $a(1 + \epsilon)^{-1/2}$. The fractional frequency shifts are $1.2 \times (f^* \epsilon)^2$ for argon in resonator A. A value of $\epsilon = 2 \times 10^{-4}$ is attainable by conventional machine shop techniques; thus, this particular perturbation can be less than 10^{-6} for the first 5 modes.

Sound waves propagating in a gas are subject to attenuation throughout the gas arising from the nonvanishing thermal conductivity and bulk viscosity of the gas. This attenuation is analogous to dielectric losses in a microwave cavity. For a monatomic gas the bulk contribution to the resonance widths is approximately:

$$\frac{g_b}{f} = \frac{14 \pi f}{9 c^2} D_T = 3.5 \times 10^{-6} f^*/P^* . \quad (7)$$

The frequency perturbation associated with bulk losses is proportional to $(g_b/f)^2$; thus, it is negligible in these experiments.

4. Tests of the Understanding of Practical Resonators

Our understanding of resonators as realized in the laboratory has been subjected to four very different kinds of tests. First, we have measured the resonance frequencies of several radial modes with both argon and propane in resonators A and B. We found that under a range of conditions, ratios of these frequencies are within $\pm 3 \times 10^{-6}$ of the ratios predicted by the model outlined above. Of particular interest for potential metrological applications is the fact that the agreement between the data and the model improve as the pressure is reduced from 0.8 MPa to 0.05 MPa. For a second test of understanding, we have compared the measured values of the resonance halfwidths, g_p, with values of g_p which we have calculated using literature data for such quantities as the viscosity and thermal conductivity. We find that the measured values of g_p exceed the calculated values by less than $5 \times 10^{-6} f_p$ under an important class of conditions in both argon and propane. Again, the agreement between the model and the data improves as the pressure is reduced. A third test was designed specifically to examine the accuracy of the boundary perturbation theory for holes in the resonator's walls. We deliberately drilled an extra hole, 0.16 cm in diameter, through the wall of resonator A. The changes in g_p/f_p produced by this comparatively large hole were measured and calculated for 6 radial modes. In each case the difference between the measurement and the calculation was less than 2×10^{-6}. A final test has been provided by the published measurements of the speed of sound in argon [7] and ethylene [8] under conditions where the speed of sound is known from independent sources. In these meaurements it is demonstrated that the combined accuracy of the laboratory procedures and the acoustic model in use at that time was limited only by the accuracy with which the volume of resonator A was known, namely $\pm 0.02\%$. In the present work we have improved the laboratory procedures and the acoustic model; however, we have not made additional dimensional measurements. Thus we will not comment further on absolute speed of sound measurements.

Our first test is concerned with frequency ratios rather than the frequencies themselves because we are looking for systematic failures of the acoustic model at the level of a few parts per million. We wish to examine the model in ways which are not overwhelmed by our imperfect knowledge of the acoustic virials of the test gases and the present limitations in our ability to control gas purity and temperature.

In Fig. 1 we show an example of the predictions from our model for perturbations of the resonance frequencies produced by the thermal admittance [Eq. (3)] and mechanical admittance [Eq. (5)] of the walls of resonator A under conditions where the sum of these perturbations is rather large. The perturbations caused by the holes are too small to be seen clearly on this scale.

The raw data are analyzed by adding the predicted perturbations to the measured frequencies. Then certain imperfections in the model will be revealed by any apparent variation of the speed of sound with the frequency of measurement. Such apparent variations for some of the argon data at various pressures are shown in Fig. 2. The ordinate is

$$\frac{\Delta r}{r} = \frac{r_p^{\text{measured}} - r_p^{\text{calculated}}}{r_p^{\text{calculated}}} . \quad (8)$$

(Here, $r_p^{\text{measured}} \equiv f_p/f_1^{\text{A}}$ for resonator A and $r_p^{\text{measured}} \equiv f_p/f_1^{\text{B}}$ for resonator B. Also, f_1^{A} and f_1^{B} are the frequencies of the first radial mode at the pressure of interest.) The data for both resonators show that as the pressure is reduced, $\Delta r/r$ becomes quite small or, equivalently, that the frequency ratios approach their theoretical values. For each mode $\Delta r/r$ can be plotted as a function of pressure and extrapolated to zero pressure. The results of this extrapolation are shown in Fig. 3. *The data in*

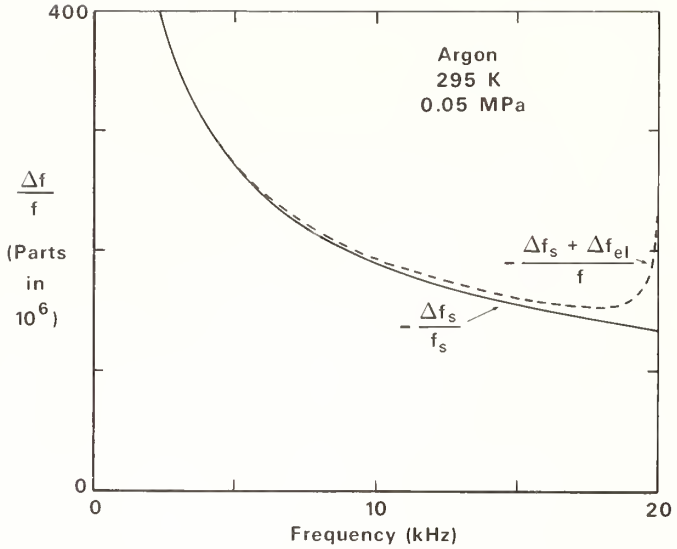


FIGURE 1. *Perturbations to resonance frequencies of resonator A with argon at 0.05 MPa. The lower curve is calculated for the thermal admittance perturbation [Eq. (3)]; the upper curve is the sum of [Eq. (3)] and the calculated elastic response of the shell [Eq. (5)].*

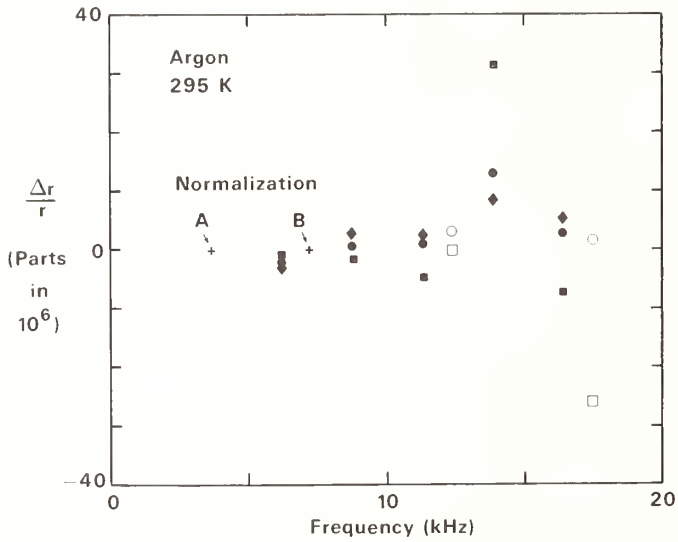


FIGURE 2. *Deviations in frequency ratios as a function of frequency at various pressures in argon. The ratios for each resonator are based on the first radial mode of that resonator (+ symbols). Solid symbols, resonator A: diamonds, 0.05 MPa; circles, 0.2 MPa; squares, 0.8 MPa. Open symbols, resonator B: circles, 0.1 MPa; squares, 0.8 MPa. As the pressure is reduced $|\Delta r/r|$ decreases, i.e., the measured frequency ratios approach their theoretical values.*

Fig. 3 indicate the level of internal consistency that one can expect in the zero-pressure speed of sound, c_0, if c_0 is obtained from several different modes of a single spherical resonator. The first 4 modes of resonator A yield zero pressure values of $\Delta r/r$ which are all within $\pm 1.9 \times 10^{-6}$ of their average. The first three modes of resonator B yield zero pressure values of $\Delta r/r$ which are all within $\pm 3.1 \times 10^{-6}$ of their average.

We have used the argon data for the first four modes of resonator A and the first three modes of resonator B to obtain an average ratio of the effective radii of these

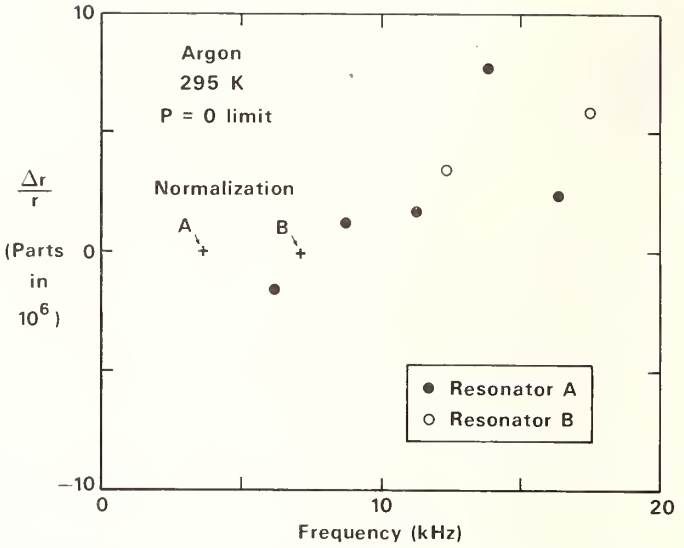


FIGURE 3. *Deviations in frequency ratios in argon as a function of frequency, extrapolated to zero pressure. The ratios for each resonator are based on the first radial mode of that resonator.*

resonators in the limit of zero pressure. The result is:

$$\frac{a_A}{a_B} = \left[\frac{\text{Volume of Resonator A}}{\text{Volume of Resonator B}}\right]^{1/3}$$

$$= 1.997156; \text{ argon, } P \to 0. \qquad (9)$$

It is very instructive to examine these frequency ratios with propane in the resonators. At a given pressure and frequency the thermal admittance perturbation to the resonance frequencies of propane is an order of magnitude smaller than that for argon, and the elastic response of the resonator wall to propane is half as large as that to argon. In Fig. 4 we examine the residual errors in the frequency ratios, $\Delta r/r$, for propane at 0.1 MPa. Here the frequency ratios for resonator A have been calculated as above ($r_p^{\text{measured}} = f_p/f_1^A$) while those for resonator B have been calculated with the definition $r_p^{\text{measured}} \equiv (f_p/f_1^A)\, a_B/a_A$. Upon comparing Fig. 4 to Fig. 3 mode by mode we see the same pattern of residual frequency dependence in $\Delta r/r$ for both gases. Of course the modes occur at different frequencies for the two gases. This suggests that the errors responsible for the residuals are related to the resonator geometry rather than instrumental noise or our imperfect knowledge of the properties of the two gases. If this conjecture is correct, improvements are quite feasible.

The analysis leading to Fig. 4 made use of the ratio of sphere radii determined from argon data. Thus the fact that the data for resonator B nearly coincide with those from resonator A demonstrates that this ratio is essentially independent of the gas used (at a low enough pressure). Indeed, we can use the propane data to obtain another measure of this quantity:

$$\frac{a_A}{a_B} = \left[\frac{\text{Volume of Resonator A}}{\text{Volume of Resonator B}}\right]^{1/3}$$

$$= 1.997155; \text{ propane, } P = 0.1 \text{ MPa}. \qquad (10)$$

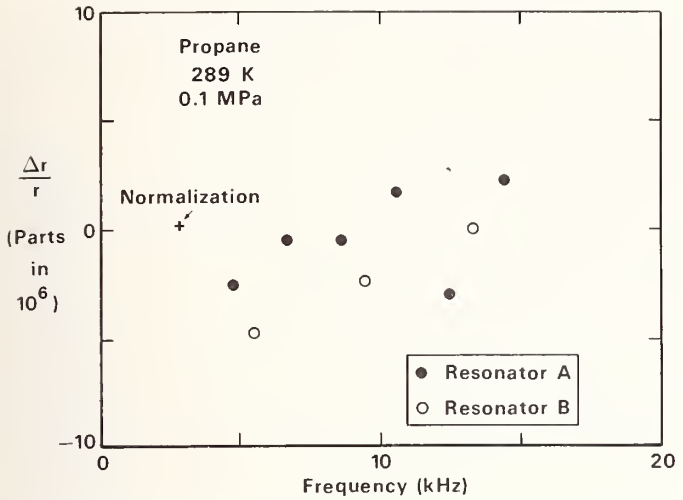


FIGURE 4. *Deviations in frequency ratios as a function of frequency in propane at 0.1 MPa. Here the frequency ratios are based on the first mode of resonator A and the ratio of the resonator diameters established with argon data (Fig. 3). The fact that the two sets of data coincide shows that the ratio of effective radii of the resonators is independent of the gas used in the measurement.*

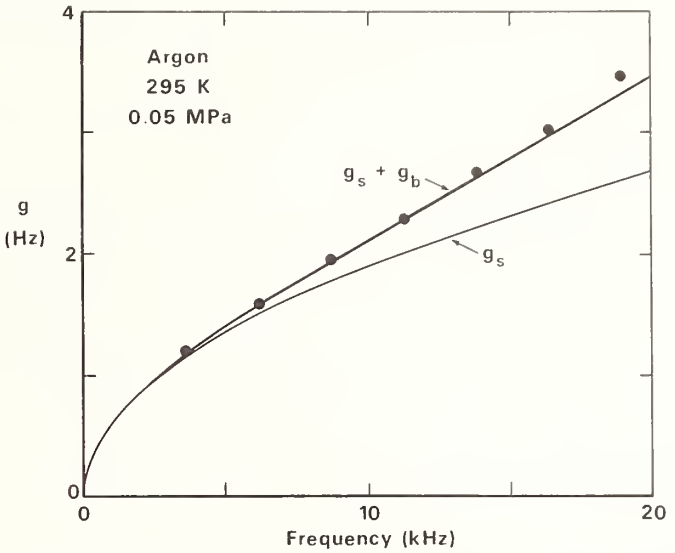


FIGURE 5. *Calculated and measured resonance halfwidths in argon at 295 K and 0.05 MPa. The lower curve is calculated for the thermal admittance perturbation [Eq. (3)]. The upper curve is the sum of the lower curve and the bulk losses [Eq. (7)].*

The agreement of Eqs. (9) and (10) must be considered somewhat fortuitous because the two sets of data were taken at temperatures 6 °C apart and adjusted using handbook values for the thermal expansions of the alloys used to fabricate the resonators. (The possibility of measuring the thermal expansion of the assembled resonators *in situ* using microwave resonances has been demonstrated by Moldover *et al.* [7].)

We now review the data for the halfwidths of the resonances. Figure 5 displays the predictions for the halfwidths of the first few radial resonances in argon at 0.05 MPa in resonator A. The dominant contributions to g_p come from the thermal admittance of the resonator wall [Eq. (3)] and the bulk losses in the gas [Eq. (7)]. As one might expect, there are mechanisms for energy loss which we have not modeled; thus the measured values of g_p exceed those calculated.

In Fig. 6 we compare our acoustic model with representative linewidth data for argon. The ordinate is

$$\frac{\Delta g}{f} = \frac{g_p^{\text{measured}} - g_p^{\text{calculated}}}{f_p} . \qquad (11)$$

It is clear that $\Delta g/f$ is less than 5×10^{-6} for the first six modes of resonator A when the pressure is less than 0.2 MPa. It is also clear that $\Delta g/f$ decreases for both resonators as the pressure is lowered. This is additional evidence that the present acoustic model is a sufficiently accurate description of the laboratory resonators to warrant serious consideration of their application to thermometry and/or a measurement of the gas constant.

As the pressure is increased, the coupling between the radial oscillations of the gas and the vibrations of the resonator shells increases. At frequencies where the resonator shells have mechanical resonances of their own (e.g., the breathing mode in resonator A at 20.2 kHz) this coupling will be especially strong. We conjecture the increase in $\Delta g/f$ with pressure and frequency is, in large part, related to such coupling. In future work we will discuss this in greater detail.

We have extensive data for $\Delta g/f$ when the resonators are filled with propane. For pressures ranging from 0.1

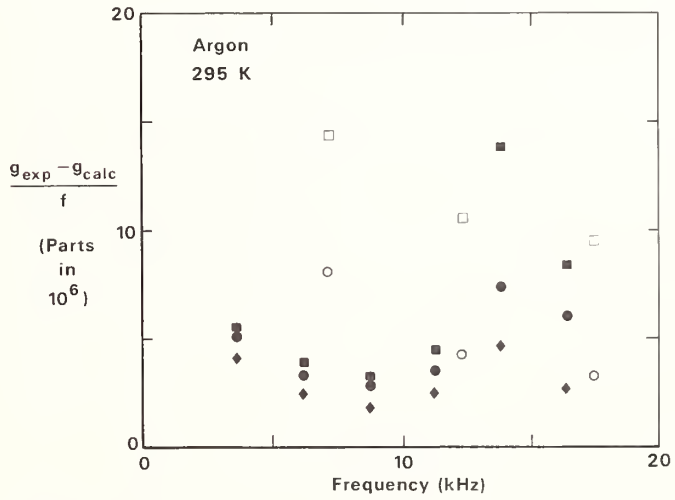


FIGURE 6. *The measured resonance halfwidth minus the calculated resonance halfwidth, divided by the frequency f, for argon at various pressures. Solid symbols, resonator A: diamonds, 0.05 MPa; circles, 0.2 MPa; squares, 0.8 MPa. Open symbols, resonator B: circles, 0.1 MPa; squares, 0.8 MPa. As the pressure decreases, the agreement between the measured and calculated values of g improves. There are no free parameters in the calculations.*

to 0.5 MPa we find that $\Delta g/f$ is less than 5×10^{-6} for the first 6 radial modes of resonator A. At higher pressures we find an anomalous increase in $\Delta g/f$. This precondensation phenomenon will be discussed in detail in a future publication [12].

In Fig. 7 we show the increase in the measured resonance halfwidths which occured upon drilling a hole 0.16 cm diameter in resonator A. The area of this hole greatly exceeds that of the other holes in the resonator. This hole is resonant near the 4th radial mode. The experiment confirms the perturbation calculation for the hole. The halfwidths of the modes other than the 4th are changed by less than $2 \times 10^{-6} f_p$.

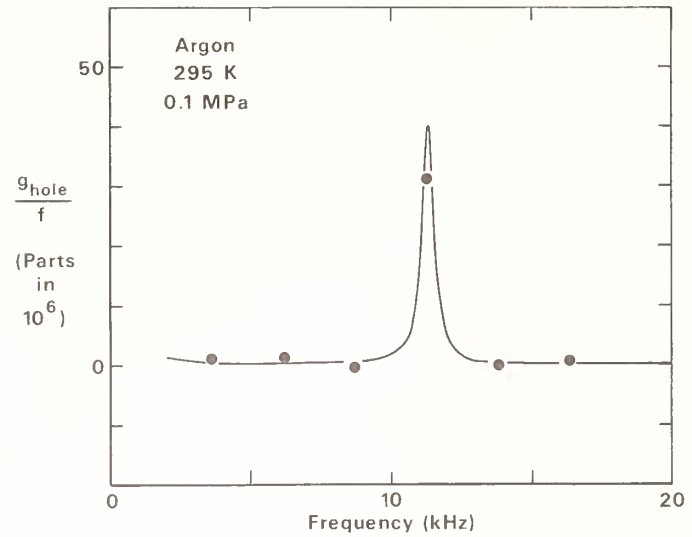


FIGURE 7. *Calculated and measured increase in g_p/f_p for a hole 0.16 cm in diameter in resonator A. There are no adjustable parameters in the calculation.*

5. Conclusions and Prospects

We have demonstrated our understanding of spherical acoustic resonators using two different resonators whose radii are in the ratio 2:1. The resonators were fabricated from different materials (aluminum and brass). Each resonator was used with two gases (argon and propane) whose acoustic and adsorption characteristics are quite different from each other's. The data indicate that our understanding of laboratory resonators is such that one could measure the speed of sound of a gas like argon in the limit of zero pressure with internal inconsistencies no greater than $\pm 3 \times 10^{-6}$. There is an indication that this number is related to the geometry of each resonator; hence it could be reduced should the control of the other experimental variables warrant it. We briefly consider the implications of this conclusion for thermometry and measurement of the gas constant.

The thermodynamic temperature, T, is related directly to a *ratio* of speeds of sound of a monatomic gas in the limit of zero pressure and zero frequency, namely: $T = T_0(c/c_0)^2$. (Here T_0 is 273.16 K and c_0 is the speed of sound in the gas at the temperature of the triple point of water.) Thus, for acoustic thermometry the primary measurements needed are resonance frequencies at T_0 and T as well as the thermal expansion (perhaps via microwave resonances) of the resonator between T_0 and T.

The error of any acoustic measurement of temperature must be at least twice the precision of the measurement of c/c_0. In this context it should be noted that errors in the acoustic model related to resonator geometry cancel out of speed of sound ratios to high order if a single mode or some specific selection of modes is consistently used to determine such ratios. It is quite likely that the acoustic contributions to the scatter among measurements of c/c_0 will be smaller than the scatter in Fig. 3 or Fig. 4.

In the course of an acoustic temperature measurement the average atomic weight of the gas must not change; however, it need not be precisely known. Auxiliary measurements such as the pressure, viscosity, and thermal conductivity of the gas and the volume of the resonator need only be made with accuracies in the range 0.1–1%. The spherical resonator's performance will not change substantially with temperature if the ratio of the mean free path in the gas to the diameter of the resonator is held constant. Thus results comparable to the ones above can be expected when acoustic thermometry is carried out at constant density rather than at constant pressure. The mechanical coupling between the gas and resonator is proportional to ρc^2; thus, it will become less important as the temperature is reduced. If ^{4}He were used instead of argon, the resonance widths at ambient temperature would remain unchanged if the pressure were increased by a factor of 2.7. The transducers and other materials we have used will perform well at all temperatures below 100 °C. Thus, the spherical resonator as it now exists is a promising tool for studying the temperature scale in the range 1 K–350 K.

An acoustic measurement of the gas constant will require the measurement of the value of the speed of sound of a monatomic gas at (or near) T_0 in the limit of zero pressure and zero frequency. The relation used is: $R = c_0^2\, 3M/(5T_0)$. Thus a measurement of R will require a very accurate measurement of the atomic weight of the gas used as well as the internal volume of the resonator after it is assembled. We imagine hydrostatic weighing is a feasible method for measuring the resonator's volume; however, we have not investigated this option yet.

References

[1] H. H. Plumb and G. Cataland, Metrologia **2**, 127 (1966); G. Cataland and H. H. Plumb, *Low temperature thermometry: interim report*, Natl. Bur. Stand. (U.S.) Tech. Note 765 (1972).
[2] A. R. Colclough, Proc. R. Soc. London Ser. A: **365**, 349 (1979); see also references therein.
[3] A. R. Colclough, T. J. Quinn, and T. R. Chandler, Proc. R. Soc. London Ser. A: **368**, 125 (1979).
[4] D. Bancroft, Am. J. Phys. **24**, 355 (1956).
[5] R. Keolian, S. Garrett, J. Maynard, and I. Rudnick, J. Acoust. Soc. Am. **64**(S1), 561 (1978).
[6] R. Keolian, S. Garrett, J. Maynard, and I. Rudnick, Bull. Am. Phys. Soc. **24**, 623 (1979).
[7] M. R. Moldover, M. Waxman, and M. Greenspan, High Temp. High Pressures **11**, 75 (1979).
[8] J. B. Mehl and M. R. Moldover, J. Chem. Phys. **74**, 4062 (1981).
[9] James B. Mehl, J. Acoust. Soc. Am. **64**, 1523 (1978).
[10] M. Greenspan, to be published.
[11] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw Hill, New York, 1953), pp. 1052–1053.
[12] M. R. Moldover and J. B. Mehl, J. Chem. Phys. **77**, 455 (1982).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Proposed Method for the Determination of the Molar Gas Constant, *R*

Leslie A. Guildner and Martin L. Reilly

Temperature and Pressure Measurements and Standards Division
National Bureau of Standards, Washington, DC 20234

A new, accurate measurement of the molar gas constant, R, is proposed. The volume occupied by 2 g of neon gas at 273.15 K is to be determined at pressures of 135, 81 and 40 kPa. An error analysis shows that the derived value of R should have a total uncertainty of 2.9 ppm at the 99% confidence level.

Key words: fundamental constants; molar gas constant; precision measurement; R.

1. Introduction

The molar gas constant appears as the proportionality constant in the equation of state for one mole of an ideal gas,

$$pV = RT , \qquad (1)$$

where p is the pressure, V the volume, and T the thermodynamic temperature. The present value of R is derived from measurements of the limiting density of pure oxygen gas which were carried out during the period from 1924 to 1941. Batuecas [1] has summarized and reviewed the relevant data from which he recommended a "most probable value" for V_m, the molar volume of an ideal gas at standard conditions. Accepting this value of V_m for their compilation, Cohen and Taylor [2] then calculated 8.31441 ± 0.00026 J/(mol·K) for R (where $R = V_m p_o/T_o$, p_o = 101.325 kPa and T_o = 273.15 K). The stated uncertainty is one standard deviation. This value of R has been officially adopted by CODATA [3].

Through the Boltzmann constant, k, (where $k = R/N_A$ and N_A is the Avogadro constant), R is related to two other fundamental physical constants. These are the second radiation constant, c_2, (where $c_2 = hc/k$, h is Planck's constant and c is the speed of light in vacuum) and the Stefan-Boltzmann constant, σ, (where $\sigma = 2\pi^5 k^4/15h^3c^2$). Of these four interdependent constants, R can be most accurately measured. As a consequence the values for k, c_2, and σ are derived from R and the appropriate atomic constants. The uncertainty in R (31 ppm) is presently so much larger than that associated with h, N_A, and c, that it constitutes almost all of the estimated error of k (32 ppm), c_2 (31 ppm), and σ (125 ppm).

The uncertainty in R also materially affects the treatment of good pVT data (in that V_m is significantly uncertain) and it is a major component of the uncertainty in the realization of thermodynamic temperatures by acoustic interferometry, isothermal gas thermometry, and noise thermometry.

New values for R have been determined from several recent experiments. The various methods used have been reviewed by Colclough [4]. The published values are: Quinn *et al.* (1976) [5], 8.31573 ± 0.00017; Gammon (1976) [6], 8.31479 ± 0.00035; Gammon (1977) [7], 8.31531 ± 0.00035; Colclough (1979) [8], 8.31447 ± 0.00022; Colclough *et al.* (1979) [9], 8.31448 ± 0.00007; and Pereira Cordido (1973, 1979) [10], 8.31417 ± 0.00027. While these results generally tend to confirm the accepted value of R, the important fact is that the reported estimates of uncertainty are still much larger than most other fundamental physical constants.

The importance of determining R with improved accuracy was recognized by the Comité Consultatif de Thermométrie, which specifically recommended to the CIPM at its 1976 meeting and again at its 1978 meeting that "measurements of the gas constant be carried out" [11, 12].

We propose a new, more accurate measurement of the molar gas constant by reliable, state-of-the-art techniques which are expected to be free of significant systematic uncertainty. The resulting value of R, with its anticipated total uncertainty of 2.9 ppm (at the 99% confidence level) would markedly reduce the uncertainty associated with the constants V_m, k, c_2, and σ, to a level comparable to that of the other fundamental constants.

2. Proposed Method

The proposed method provides very exact determination of pV/nT and the extrapolation to zero pressure to determine the molar gas constant, R, from the equation of state of neon gas,

$$pV/(nT) = R + Bp/T + \ldots . \qquad (2)$$

The volume, V, corresponding to an essentially constant number of moles, n, at a temperature, T, of 273.15 K, is to be determined for pressures, p, of 135, 81 and 40 kPa. The volume will be metered by multiple strokes of a piston whose displacement is accurately known. Approximately 2 g of gas will be transferred at constant pressure from a thermostated chamber into a detachable weighing flask held at liquid helium temperature. The combined relative accuracy of V and n is the same for each pressure and is optimized when the mass of gas used is 2 g. As a consequence, the extrapolated value of R will have much greater accuracy than is possible with the usual pVT methods.

A semi-schematic drawing of the proposed apparatus is given in Fig. 1. (The thermostated chamber and the weighing flask are shown to scale.) The piston (P) is shown at endstroke in the thermostated chamber (C). The piston has a nominal diameter of 7.62 cm and a

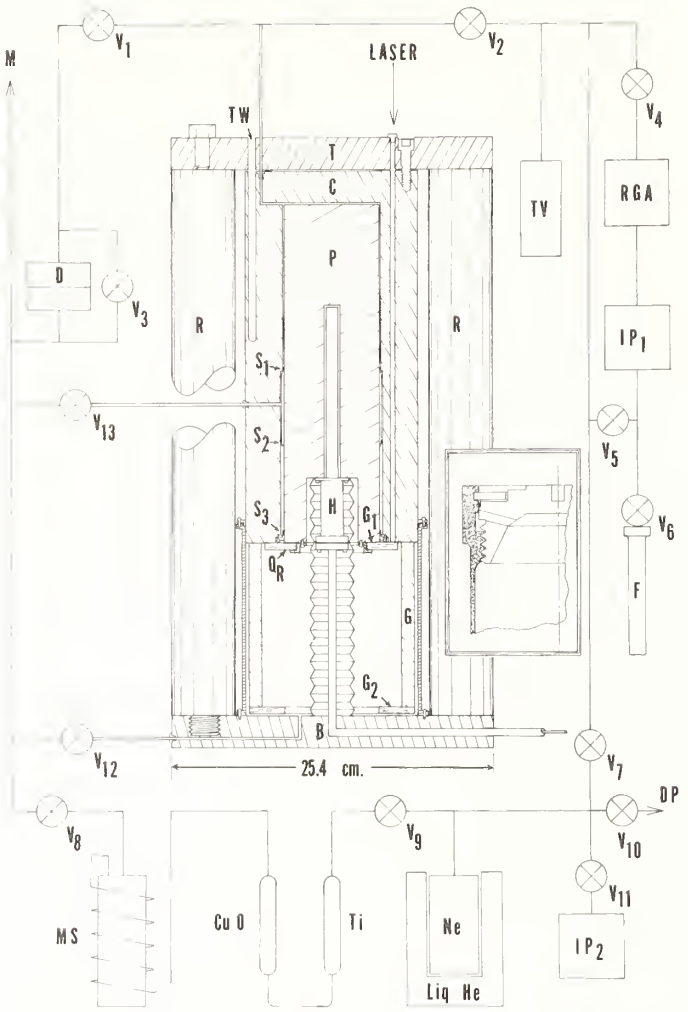


FIGURE 1. *Schematic of proposed apparatus for determining the gas constant.*

stroke of 12.3 cm. It will be sealed in the chamber by a series of three PTFE seals (S_1 to S_3) which will also guarantee its exact alignment [13]. The region between S_1 and S_3 is filled with neon at the manometer pressure. This arrangement reduces the danger of leakage across the main piston-chamber seal (S_1) and "preconditions" the piston in case there is any significant tendency for neon to adsorb on the surface. The motion of the piston will be controlled by a hydraulic system (H) which has stops to limit the stroke.

The surface of the lower end of the chamber (G_1) will be lapped flat so that the cylindrical "gauge block" (G) can be wrung to it. The annular disk (G_2) will be wrung to the other end of G. An annular quartz disk (Q_R) will be wrung to the lapped bottom surface of the piston. At the extremes of each stroke, the position of the piston will be established by the capacitance between guarded electrodes on the parallel upper or lower surfaces of Q_R and the grounded surfaces G_1 or G_2.

The chamber will be suspended from the top (T). The volume between C and T will be enclosed by a seal so that it can be filled with neon at the chamber pressure to avoid any effects from pressure dilation. The top will be supported from the base (B) by four symmetrically spaced vertical rods (R). The space between the bottom of the chamber and the base will be enclosed by a cylinder which will be sealed to the base and has a sliding seal around the outer circumference of the chamber. This volume will be also filled with neon at the manometer pressure. The length of this assembly will be about 50 cm, and its diameter will be about 25 cm. It will weigh about 100 kg. The entire assembly, including the diaphragm transducer (D) and valves (V_1, V_2, and V_3) will be placed in a bath of ice and circulating water.

The detachable weighing flask (F) will be nominally 1 cm in diameter, 10 cm high, and 0.04 cm in wall thickness. It will have an empty mass of 18 g with an interior volume such that the 2 g of neon gas will exert a pressure of about 12 MPa at room temperature. The valve (V_6) will be an integral part of the flask with interior threads and will seal against a gold gasket. It is designed to be operated at a temperature near 30 K. (Note details in inset.)

In a typical displacement cycle the chamber, with V_2 closed, will be filled with purified neon through V_1 and V_3 as the piston is slowly withdrawn from the position indicated in Fig. 1. After equilibration and adjustment of the pressure to balance that of the manometer (M), the null of the diaphragm transducer will be observed and V_3 closed. Valves V_1, V_2 and V_3 will be of "constant volume" design [14, 15]; negligible error will be produced by their action.

The neon gas will be transferred to the weighing flask in two stages. First, V_2 will be partially opened and, as the piston is advanced into the chamber, the neon will flow slowly into a "holding volume" bounded by valves V_2, V_4, and V_5, and V_7. This volume, which will be mainly that of the intermediate thermostatted vessel (TV), is to be slightly greater than the room temperature volume of the displaced neon. When the travel of the piston is complete and equilibrium reestablished, the pressure in the system will be once again balanced exactly against the manometer by adjusting the temperature of the gas contained in TV. After V_2 is closed all the neon in the "holding volume" can be quantitatively frozen into the weighing flask while the chamber is being refilled for another cycle.

To displace 2 g of gas, the practical limitations of both diameter and stroke of the piston make it necessary to use 3 cycles for a pressure of 135 kPa. Correspondingly, 6 and 10 cycles will be required for the other pressures. No degradation of accuracy is expected, however, because of the extra cycles. After the requisite number of cycles has been executed, valves V_5 and V_6 will be closed and the flask removed for weighing.

The peripheral facilities needed for evacuation, which include ion pumps (IP_1 and IP_2) and an oil diffusion pump (DP), for residual gas analysis (RGA), and for gas purification (from V_8 to V_{10}) are described elsewhere [16].

3. Measurements

The experimental measurements require great care and exactness. Sufficient experience exists for each element of the process so that it can be expected to function successfully.

3.1 Pressure

Pressures will be established with the NBS high precision manometer. The pressure arm of the manometer is to be filled with the high purity neon. The principles and

many details of its operation have been described earlier [17]. A main feature of the manometer is that the height of the mercury column is accurately determined with gauge blocks. The crowns of the menisci are accurately located relative to the gauge blocks by reproducing precisely the capacitance between the mercury surfaces and fixed electrodes which themselves are located accurately with respect to the gauge blocks.

Ten calibrated gauge blocks of nominally equal length (101.6 mm) will be used in various combinations of three to produce the lowest pressure, 40 kPa. Combinations of six blocks will be used to produce a pressure of 81 kPa and all ten will be used for a pressure of 135 kPa. One experimental sequence will consist of three measurements at the lowest pressure and one each at the two higher pressures. Two sequences constitute one "set".

3.2 Volume

The volume will be calculated as the product of the number of strokes and the calibrated piston volume adjusted by the product of the appropriate piston areas and the various endpoint displacements as determined by capacitance.

3.3 Mass of Neon Gas

All weighings will be made using a double-beam vacuum microbalance which has a sensitivity of < 1 μg at a load of 20 g. Two platinum weights will be used. A 20 g weight will serve as a tare for all weighings. A 2 g weight will be added when weighing the empty 18 gram flask.

The mass of the neon gas, M_{Ne}, will be obtained through transposition weighings of the flask filled with gas and of the empty flask plus the 2 g weight. Thus we have

$$M_f + M_{Ne} + M_1 = 20 \text{ g} \tag{3a}$$

and

$$M_f + 2 \text{ g} + M_2 = 20 \text{ g}. \tag{3b}$$

The quantity M_f is the mass of the weighing flask. The quantity M_1 is the mean of the small weight increments associated with ten weighings of the filled flask; M_2 is the mean of the ten corresponding weighings of the empty flask. Therefore the mass of the neon gas is

$$M_{Ne} = 2 \text{ g} + M_2 - M_1. \tag{3c}$$

3.4 Temperature

The temperature of the neon will be assumed to be the mean temperature of the chamber as measured by four long stem standard platinum resistance thermometers. The sensor of each thermometer will be at the mid-depth in one of the four thermometer wells (TW of Fig. 1) which will be arranged symmetrically about the body of the thermostated chamber. They will be calibrated at the triple point of water between each set of experiments.

4. Assessment of Imprecision of Measurements

Equation 2 will be fitted to the experimental data by the method of least squares. The value of R is the intercept of the line extrapolated to zero pressure. The data at 135 and 81 kPa will first be adjusted to compensate for the contribution of the third virial term. This is not greater than 1.0 ppm and can be made with a negligible error. The standard deviation of the extrapolated value reflects the combined standard deviations resulting from the imprecisions of the directly measured quantities. In addition, there are uncertainties resulting from the imprecision of the auxiliary measurements. Any errors produce a constant bias in the results, the limits of which are estimable in terms of the imprecisions determined by other experiments. (We denote the relative standard deviation of a directly measured quantity by s_A and that of an auxiliary quantity by s_B.)

4.1 Pressure

The only significant error in the measurement of the pressure is expected to be that from the manometer itself. Resolved into its components, the pressure has errors as follows: (1) The density of the mercury will be uncertain both in its value at 20 °C ($s_B = 0.33$ ppm) and in the measurement of its temperature ($s_A = 0.2$ ppm). (2) The relative standard deviation of the acceleration due to gravity as measured at the site of the manometer is $s_B = 0.2$ ppm. (3) The uncertainty in the height of the mercury column depends on the calibration of the gauge blocks for which $s_B \leq 0.2$ ppm.

4.2 Volume

The volume displacement of the piston when moved between the two fiducial limits will be determined by weighing the mercury so displaced. The calibration will be performed at the ice point under the conditions of the experiment. At the same time, the separation of the fiducial limits will be measured *in situ* by laser interferometry.

The piston diameter will also be very carefully determined at the locations corresponding to the extremes of its travel in order to account for small differences in the volume resulting from slight differences from its reference position. We anticipate that these adjustments do not involve significant error.

The significant components of error are: (1) The relative standard deviation associated with the absolute value, at 20 °C, of the density of displaced mercury, including the mass measurement, is $s_B = 0.4$ ppm. (2) The relative standard deviation in the density of the mercury as a result of the adjustment for the thermal dilation between 20 and 0 °C is $s_B = 0.3$ ppm.

4.3 Moles of Neon

The number of moles of neon gas will be determined from the ratio of the mass of gas to its molecular weight.

4.3.1 Mass of Neon Gas

The single largest uncertainty of the experiment arises from the mass assignment of the 2 g platinum weight. Its value will be derived in the most direct way possible from the National Prototype Kilograms. The standard deviation is estimated to be 1.5 μg ($s_B = 0.75$ ppm).

The relative standard deviation of the mass of neon will be the sum in quadrature of the relative standard deviation of the mean for each set of ten weighings of the full and empty flask. For each mean, $s_A = 0.35$ ppm, and for their difference in quadrature, $s_A = 0.5$ ppm.

4.3.2 Molecular Weight of Neon Gas

Naturally occurring neon has isotopes of mass 20, 21, and 22. Neon gas with a concentration >99.95% ^{20}Ne,

and a balance of mostly ^{22}Ne is formed by radioactive decay; it is readily available. The value of the atomic mass of each isotope is accurately enough known [18] that no significant error results. To specify the molecular weight of the neon gas used in the experiment within 1 ppm, the concentrations of the two rarer isotopes (which are present at a combined level of 500 ppm or less) need only be determined within 2%. The relative standard deviation associated with the molecular weight is s_B = 0.3 ppm.

4.4 Temperature

It is estimated that the standard deviation of the mean of the four values of the chamber temperature will be <0.05 mK, or correspondingly, the relative standard deviation in the thermodynamic temperature at the ice point will be s_A = 0.18 ppm.

A summary of these relative standard deviations is given in the following table:

QUANTITY	TYPE A		TYPE B	
	ppm	Source	ppm	Source
Pressure				
density of Hg	0.2	Temperature	0.3	Original measurement
g			0.2	Original measurement
h			0.2	Length calibration
Volume	0.2	Imprecision	0.4	Hg displacement calib.
			0.3	Thermal dilation of Hg
Moles of gas	0.5	Mass of neon	0.75	Calib. of 2 g weight
			0.3	Determination of MW
Temperature	0.2	Imprecision	0.2	Imprecision
$s^2_{pV/(nT)}$	0.37×10^{-12}		1.11×10^{-12}	

The error in R can be evaluated from the equation for the variance of the intercept when $pV/(nT)$ vs. p is extrapolated to zero. This is

$$s_R^2 = s^2_{pV/(nT)}\{1/z + \bar{p}^2/\sum_{i=1}^{z}(p_i - \bar{p})^2\}, \qquad (4)$$

where $\bar{p}$ is the mean of the individual pressures, p_i, and z is the number of points. For a "set" of measurements extrapolated vs. pressures in ratios of 3:3:3:6:10, the estimated variance of R is $s_R^2 = 0.37 \times 10^{-12} \times 0.398 = 0.15 \times 10^{-12}$, or for 5 "sets" $s_R^2 = 0.03 \times 10^{-12}$. The combined uncertainty is $s_{tot} = 1.07 \times 10^{-6}$, and the total uncertainty at the 99% confidence level for 5 "sets" of measurements is 2.71 times this number or 2.9 ppm.

5. Sources of Systematic Error

Numerous other sources of error have been examined (e.g., the uncertainty of the thermomolecular pressure, variations in the acceleration due to gravity, etc.). The expected errors from these sources will be insignificant in every case.

The largest potential source of systematic error could arise from sorption of neon gas on the interior of the thermostated chamber and on the piston. However, in contrast to constant volume gas thermometry, the neon chamber will not be subjected to variations of temperature. Hence it can be evacuated at a temperature high enough that, after cooling, no significant amount of desorption will occur. As is our practice in using helium for the thermometric fluid of the NBS gas thermometer, we intend to purify neon so well that no significant adsorption of impurity can occur. At levels of a few ppm the sorption characteristics of neon itself are uncertain. Such sorption as may occur will surely be very limited in extent and of the physical type (hence rapid). The experimental method is designed to evade any significant effect based on the fact that a stably sorbed layer of clean neon in the chamber will cause no error. Thus the guard section, which contains neon gas at the manometer pressure, will precondition the surface of the piston to the maximum extent possible before it enters the main chamber. With such precautions in design and procedure, it is our opinion that no significant sorption error will be encountered.

References

[1] T. Batuecas, in *Atomic Masses and Fundamental Constants 4*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Publ. Corp., New York, 1972), p. 534.
[2] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data **2**, 663 (1973).
[3] CODATA Bulletin No. 11 (Dec. 1973).
[4] A. R. Colclough, these proceedings.
[5] T. J. Quinn, A. R. Colclough, and T. R. D. Chandler, Philos. Trans. R. Soc. London, Ser. A: **283**, 367 (1976).
[6] B. E. Gammon, J. Chem. Phys. **64**, 2556 (1976).
[7] B. E. Gammon, (1977), private communication.
[8] A. R. Colclough, Proc. R. Soc. London, Ser. A: **365**, 349 (1979).
[9] A. R. Colclough, T. J. Quinn, and T. R. D. Chandler, Proc. R. Soc. London, Ser. A: **368**, 125 (1979).
[10] J. M. Pereira Cordido, cited by A. R. Colclough [4].
[11] Comité International des Poids et Mesures. Comité Consultatif de Thermométrie. 11e Session. {Rapports}. Pavillion de Breteuil, Sèvres, France, 1976.
[12] Comité International des Poids et Mesures. Comité Consultatif de Thermométrie. 12e Session. {Rapports}. Pavillion de Breteuil, Sèvres, France, 1978.
[13] R. E. Edsinger, L. A. Guildner, and R. L. Anderson, Rev. Sci. Instrum. **42**, 945 (1971).
[14] R. L. Anderson and L. A. Guildner, Rev. Sci. Instrum. **36**, 615 (1965).
[15] R. L. Anderson, L. A. Guildner, and R. E. Edsinger, Rev. Sci. Instrum. **41**, 1076 (1970).
[16] L. A. Guildner and R. E. Edsinger, J. Res. Natl. Bur. Stand. (U.S.) Sect. A, **80**, 703 (1976).
[17] L. A. Guildner, H. F. Stimson, R. E. Edsinger, and R. L. Anderson, Metrologia **6**, 1 (1970).
[18] A. H. Wapstra and N. B. Gove, Nucl. Data Tables **9**, 265 (1971).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A Radiometric Determination of the Stefan-Boltzmann Constant*

T. J. Quinn

Bureau International des Poids et Mesures, F-92310 Sèvres, France

and

J. E. Martin

National Physical Laboratory, Teddington, Middlesex TW11 0LW, U.K.

A new determination of the Stefan-Boltzmann constant is being made using a blackbody source of thermal radiation at 273.16 K and a detector in the form of a heat-flow calorimeter at 2 K. The blackbody is made of copper having an internal coating of 3M's C401 black paint. The beam of radiation entering the detector is defined by a pair of beryllium-copper apertures near 4.2 K. The calorimeter detector is a thin copper shell, also coated on the inside with 3M's black, made in the form of a blackbody and attached by means of a poorly conducting stainless-steel heat link to a reservoir of superfluid helium at about 2 K. Electrical power may be supplied to the calorimeter through a series of heaters wound on the outside and connected to an external source of power by superconducting leads. The temperature of the calorimeter is monitored by a germanium resistance thermometer.

On opening a shutter below the first aperture, blackbody radiation arrives at the calorimeter, raising its temperature by about 3 K. On closing the shutter, electrical power is applied to the calorimeter to maintain the same temperature. The applied electrical power is thus a measure of the thermal radiative power passing through the apertures. From a knowledge of the geometry of the aperture system and the thermodynamic temperature of the blackbody, the Stefan-Boltzmann constant may be deduced.

Key words: radiometer; Stefan-Boltzmann constant; thermal radiation.

1. Introduction

In this paper we describe briefly the design and operation of a calorimetric radiometer that has been built at NPL for the two-fold purpose of making a direct determination of the Stefan-Boltzmann constant and for the measurement of thermodynamic temperature in the range −50 °C to 230 °C. Preliminary results have been obtained for the Stefan-Boltzmann constant and work is in progress both to confirm these results and to make measurements of thermodynamic temperatures. Nothing further will be said here about the latter since they are the subject of another publication [1].

The Stefan-Boltzmann constant, σ, appears in the equation that describes the total radiant power, $E(T)$, emitted per unit area by a blackbody at a thermodynamic temperature T:

$$E(T) = \sigma T^4, \tag{1}$$

where, from a summation of Planck's equation over all wavelengths, σ is given by:

$$\sigma \equiv \left\{ \frac{2\pi^5 k^4}{15 h^3 c^2} \right\} \mathrm{W\,m^{-2}\,K^{-4}} \tag{2}$$

in which c is the speed of light, h is the Planck constant and k is the Boltzmann constant which is itself given by the ratio of the gas constant, R, to the Avogadro constant N_A. The accuracy with which σ can be established using Eq. (2) is limited principally by the relative uncertainty, of 25 parts in 10^6, in the most recent value of R.

Taking the most recent values for the constants, namely:

$c = 299\,792\,458$ m s^{-1} (conventional value recommended by the 15th CGPM 1975)

$h = 6.626\,176 \times 10^{34}$ J s, $N_A = 6.022\,045 \times 10^{23}$ mol^{-1} (1973 least-squares adjusted values)

$R = 8\,314.48 \pm 0.63$ J K^{-1} kmol^{-1} (Colclough, Quinn, and Chandler 1979 [2])

we find a value for the Stefan-Boltzmann constant of

$$\sigma = (5.670\,51 \pm 0.001\,7) \times 10^{-8}\ \mathrm{W\,m^{-2}\,K^{-4}}, \tag{3}$$

where the uncertainties in both R and σ are given at the level of three standard deviations.

A direct experimental determination of σ will be of interest, therefore, if a relative uncertainty approaching 3 parts in 10^4 can be achieved, since it would then provide a totally independent value for R with an uncertainty of about 70 parts in 10^6. This is the aim of the present work.

Previous experimental measurements of σ have, with the exception of that of Blevin and Brown in 1970 [3], given values that were very much larger than the theoretical value, sometimes by more than 2% and on average by about 1.5%. The work of Blevin and Brown demonstrated that, unless very specific measures were taken, errors due to diffraction and scattering could easily amount to a few percent. Although their work provides a landmark in absolute radiometry, the value of σ obtained by Blevin and Brown was subject to a number of uncertainties that were not radiometric in origin. In particular, their radiating source was at the temperature of freezing gold (t_{68} = 1064.43 °C), the thermodynamic temperature of which is uncertain by about 0.2 K, lead-

*Work carried out at the National Physical Laboratory.

ing to an uncertainty in σ of 0.06%. Further uncertainties associated with the temperature measurement of the blackbody at the gold point combined with this to give an overall uncertainty, related to the source temperature, of about 0.1%. The final value for σ obtained by Blevin and Brown was $(5.664\ 4 \pm 0.009) \times 10^{-8}\ \mathrm{Wm^{-2}\ K^{-4}}$. The uncertainty is equivalent to 0.16% at the level of three standard deviations, and the value differs from that given about in Eq. (3) by 0.11% (see Fig. 7).

It was in view of the uncertainties in Blevin and Brown's result, stemming from their use of a high temperature source, that we decided to design our method about a radiating source at the temperature of the triple point of water, for which the thermodynamic temperature is, by definition, 273.16 K exactly. Although in this way we eliminated many of the problems related to the temperature of the source, it was at the expense of generating others. These were related mainly to diffraction, scattering, and to the realization of sources which remain black at very long wavelengths. We were further influenced in our design by the work of Ginnings and Reilly [4] at the NBS. They had already demonstrated the feasibility of making measurements, with a sensitivity of a few parts in 10^5, of the ratio of the total energies emitted by blackbodies at the steam and ice points. They had used a low temperature calorimeter operating near 4 K as a radiation detector. Ginnings and Reilly were aiming at thermodynamic temperature measurement and made no attempt to make absolute measurements which might have led to a new value for σ. Unfortunately, their work was already well advanced before the work of Blevin and Brown had highlighted the problems of diffraction and it was subsequently found impracticable to modify their design.

2. Principles of Operation of Our Method

The equipment which we have used to determine σ is illustrated in outline in Fig. 1. The principle of operation of the system is as follows: A blackbody cavity held at a temperature close to 273.16 K irradiates an aperture system at a temperature near 4.2 K that permits a beam of radiation to enter a second blackbody cavity held initially at 2 K. The absorbing blackbody acts as a heat flow calorimeter and the incident radiation causes a rise in temperature until the radiant energy is balanced by the heat flow, down a poorly conducting heat link, to a heat sink maintained at a constant temperature close to 2 K. The temperature rise of the cavity, about 3 K for a radiating blackbody temperature of 273 K, is monitored by a germanium resistance thermometer attached to the calorimeter. A shutter, at a temperature near 4.2 K, is then closed and electrical power supplied to the calorimeter, via superconducting leads, to maintain the calorimeter at the same temperature as before. This electrical power is a precise measure of the radiant power previously present and, together with a knowledge of the geometry of the aperture system, it allows the Stefan-Boltzmann constant to be deduced. There are, of course, many conditions which must be satisfied before such a simple situation as that just described can exist. The more important of these we shall outline below.

3. Sources of Uncertainty

3.1 Diffraction and Scattering

From the very beginning we were in no doubt that the success of the work would depend upon whether or not we could overcome the problems of diffraction and

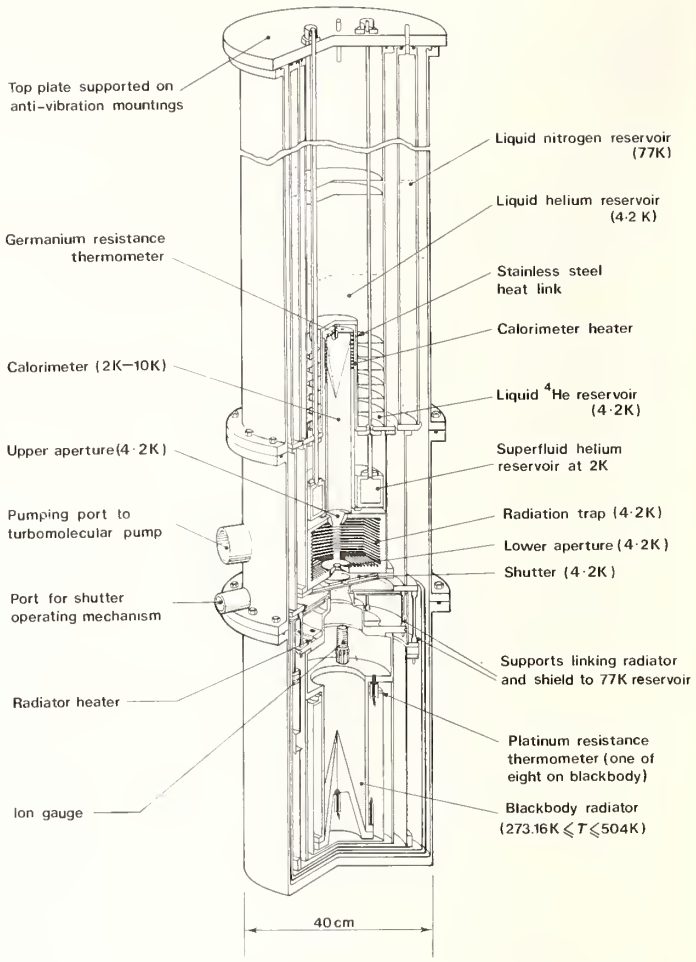


FIGURE 1. *Cut-away drawing of radiation calorimeter.*

scattering. To do this we tried to design a practical system which approached as closely as possible the nearly ideal system illustrated in Fig. 2(a). In the ideal system we require that the lower aperture, L, which limits the radiation coming from the radiating blackbody, the 'radiator', be illuminated over a solid angle of 2π steradian and that the absorbing blackbody, the 'calorimeter', subtend a solid angle of 2π steradian at the upper aperture, U, which limits the radiation entering it. If these conditions are satisfied then there will be no error in the measurement of the radiation from the radiator resulting from diffraction at the edges of the two limiting apertures L and U.

Problems of scattering arise in the space between the two apertures. The requirement here is that all of the radiation emitted by L, with the exception of that within the geometrical beam passing through U, must be absorbed. Since only about 1.6% of the radiation emitted by L passes directly to U, the design of the radiation trap (see Fig. 3) between L and U poses no mean problem and is crucial to the success of the experiment.

For purely practical reasons it is not possible to achieve complete hemispherical illumination of L nor complete hemispherical collection of the radiation leaving U. Nevertheless a close approach to these ideal conditions can be achieved by means of highly polished reflecting surfaces disposed in the manner shown in Fig. 2(b).

The fraction of the radiation entering the calorimeter which originates, due to diffraction, outside the geometrical beam and comes from the direction of the reflecting surfaces may be calculated following the method of

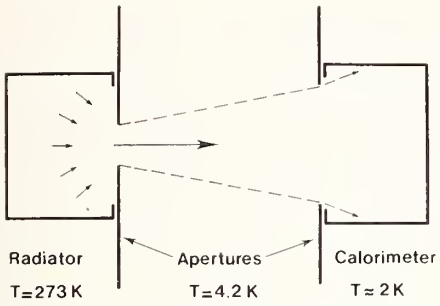


FIGURE 2(a). *Ideal geometry for a radiometric measurement of the Stefan-Boltzmann constant. The aperture adjacent to the radiator we call 'L' and that adjacent to the calorimeter, 'U'.*

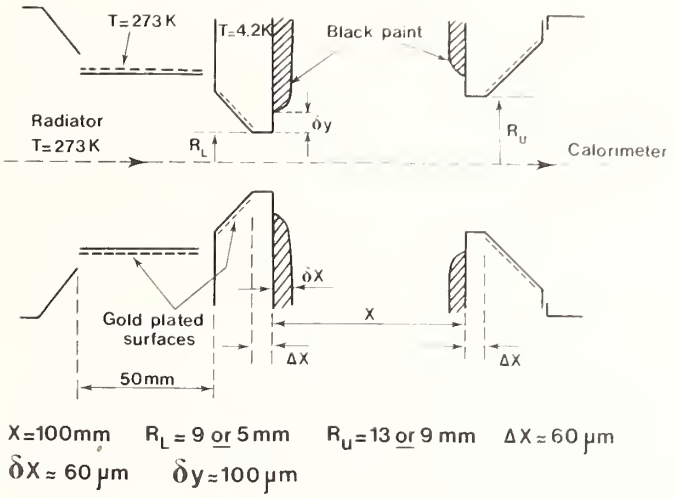


FIGURE 2(b). *The practical realization of the ideal geometry shown in Fig. 2(a).*

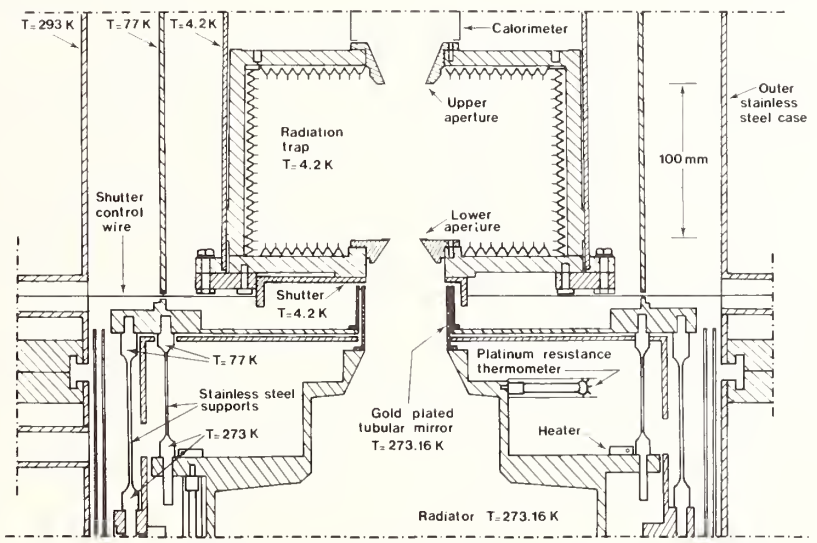


FIGURE 3. *Detailed drawing of the radiation trap and upper part of the blackbody radiator.*

Blevin [5]. Errors due to diffraction will only arise if the reflecting surfaces are at temperatures different from that of the blackbody and at the same time are not perfect reflectors.

The effects of diffraction are of course wavelength dependent. In a measurement of σ based upon 273 K blackbody radiation we have to take account of radiation having wavelengths extending from about 2 μm to 300 μm. Not all of these wavelengths are of equal importance, since 99% of the total energy is in the wavelength range below 85 μm, and 99.9% is in the range below 180 μm. A negligible amount is in the range below 2.5 μm. An effective wavelength, calculated to best represent diffraction losses, is about 19 μm for 273 K radiation. It is clearly desirable that the dimensions of the apertures be as large as possible compared to these wavelengths. The overall size of the apparatus is directly related to the size of the apertures and followed from the initial aim of determining σ with a relative uncertainty approaching 1 part in 10^4. This led us to choose a minimum aperture diameter of about 1 cm and a separation of the apertures of 10 cm. The dimensions of all the remaining components followed.

Of major importance were the optical and vacuum properties of the black paint with which the inside surfaces of the radiator, calorimeter, radiation trap, and other surfaces were coated. This paint was C-401 Black, made by the 3M Company, and its optical degassing properties were the object of a subsidiary investigation fully reported elsewhere [6].

3.2 The Radiator

The blackbody cavity, Fig. 4, that forms the radiator is made of copper and is coated on the inside with a layer, about 100 μm thick, of C-401 Black. The temperature of the blackbody is measured by eight capsule-type platinum resistance thermometers, specially made for us by the Tinsley Co. so that they could be bolted directly to the copper. The blackbody is suspended by thin stainless steel strips from the liquid nitrogen reservoir and controlled at a temperature near 273 K by a single ring-heater placed near the points of attachment of the steel strips. The emissivity of the blackbody was calculated [7] on the basis that the 3M's black behaved as a diffuse emitter of 273 K radiation having an emittance of 0.94 at angles up to about 70° from the normal. Beyond this angle the point exhibits a decrease in emittance with a consequent increase in reflectance which is almost wholly specular. The retroreflectance appears to be negligible at all angles of view [8].

3.3 The Calorimeter

The radiation absorber, or calorimeter, Fig. 5, is made in the form of a blackbody cavity similar, of course, in shape and internal dimensions to the radiator, and connected by means of a poorly conducting heat link to a constant temperature reservoir. The constant temperature reservoir is maintained at a temperature close to 2 K and just below the helium-4 lambda point where the thermal conductivity of superfluid helium is very high and good temperature stability can be achieved. The heat capacity of copper at these temperatures is about 10^{-4} times its value at room temperature. This enables adequate sensitivity to be obtained with a relatively large mass [~300 g] so that an efficient absorber can be constructed. And although the thermal conductivity of stainless steel near 2 K is about 1% of its room temperature value, the time constant of the calorimeter/heat link assembly is still only 1% of what it would be at room temperature for a calorimeter of this size. The measured time constant for small temperature changes at these temperatures is about 4 minutes. The thermal conductivity of copper near 2 K is about the same as it is at room temperature. Thus the thermal diffusivity is very high and so temperature gradients are insignificant for the power levels that we use.

The dimensions of the heat link were chosen so that the temperature difference between the calorimeter and the 2 K reservoir was 3 K, for the large set of apertures, when viewing 273 K radiation. Using a germanium resistance thermometer there is no difficulty in measuring the temperature rise of the calorimeter with a sensitivity of 1 part in 10^5 of its value. The requirements for temperature stability of the 2 K reservoir are, however, very stringent because of the large change in thermal conductivity of stainless steel as a function of temperature in this range. If the temperature of the 2 K reservoir is drifting, constancy of temperature difference between the calorimeter and reservoir does not indicate constancy of power input to the calorimeter. It is necessary to maintain the temperature of the 2 K reservoir within about 20 μK of a nominal value for the equivalence of successive power measurements to be good to better than 1 part in 10^5.

Further advantages of operating the calorimeter at very low temperatures are that the radiation exchange with the surroundings can be made negligible and the heater leads can be made of superconducting wire so that

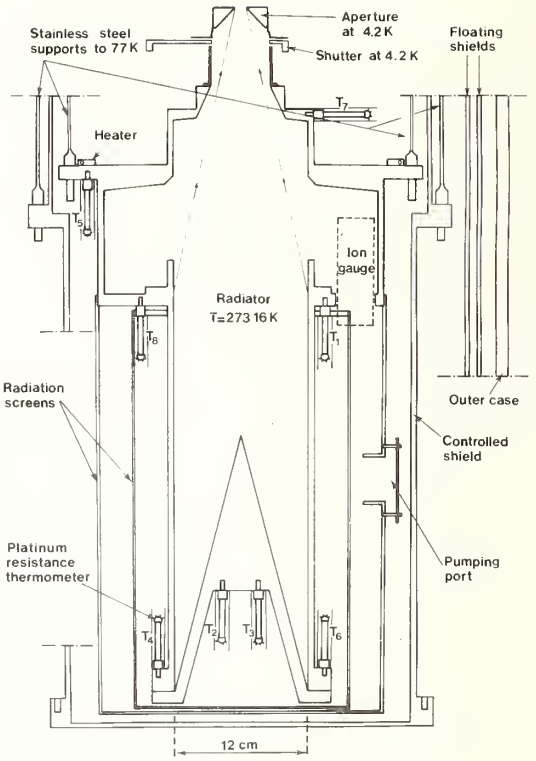


FIGURE 4. *Schematic of the blackbody radiator at 273.16 K showing positions of the eight capsule-type platinum resistance thermometers, each one enclosed in a cylindrical multi-layer radiation shield. The marginal rays of the radiation beam passing to the calorimeter are shown.*

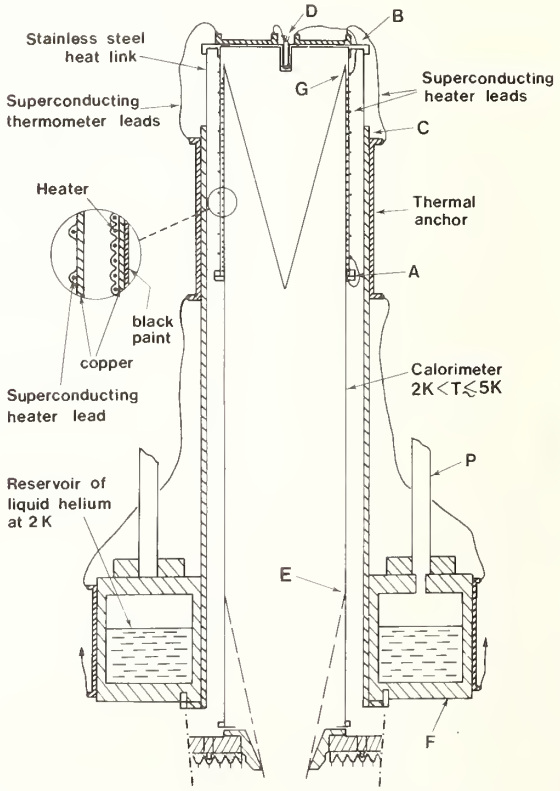


FIGURE 5. *The calorimeter, showing how it is supported and linked to the 2 K constant-temperature reservoir. The points marked A to G are also shown in Fig. 6.*

no heat is dissipated in them. Nevertheless, great care must be taken in the way that the leads to the heater and thermometer are thermally anchored to the calorimeter for a proper equivalence to exist between heating by radiant power and electrical power (Fig. 6). That this was achieved was confirmed experimentally by supplying the same electrical power to different heater windings placed at various positions on the calorimeter. The indications of the germanium thermometer were independent of the point of application of the power. The size and shape of the calorimeter were chosen so that the calculated absorption was greater than 0.999 for the large set of apertures. The departure from blackness of the calorimeter leads to a correction of some 3 parts in 10^4 to the measured value and is one of the largest sources of uncertainty in the final value of σ.

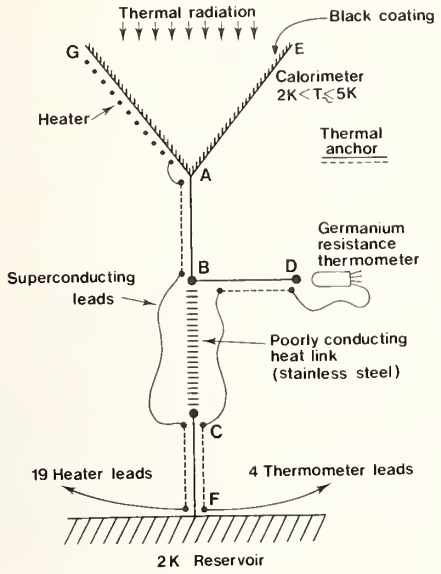


FIGURE 6. *A heat-flow diagram for the calorimeter showing the critical thermal anchoring points. The points marked A to G are also shown in Fig. 5.*

3.4 The Apertures and Geometrical Factor

Two sets of apertures are to be used so that errors originating in the aperture system and the blackbodies and also those related to diffraction and scattering, are likely to change and thus become apparent. The arrangement of the apertures is shown in Fig. 2(b). Each aperture is made from dispersion-hardened beryllium copper, from the same sample as a specimen for which the thermal expansion coefficient was measured between room temperature and 4.2 K [9]. The radiation trap which supports the apertures was made from OFHC (oxygen free, high conductivity) copper. The mean thermal expansion coefficients between 293 K and 4.2 K were taken as 3070×10^{-6} for beryllium copper [9] and 3257×10^{-6} for OFHC copper [10]. The diameters and distance apart of the apertures were measured using standard metrological techniques in the NPL length section. The edges of the apertures do not depart from mean circles by as much as 1 μm and scanning electron microscope photographs of each aperture allowed us to confirm the form and sharpness of the edges, and the widths of the cylindrical lands, ΔX in Fig. 2(b). The reflectivity of these lands is an important source of uncertainty in these results but will be rather less significant in those to be obtained using the small set of apertures which have lands only about half as wide.

The geometrical factor, g, that enters into the calculation of σ is, from standard photometric theory, given by:

$$g = 2\pi R_L^2 R_U^2 [\{R_L^2 + R_U^2 + D^2\} + \{(R_L^2 + R_U^2 + D^2)^2 - 4R_L^2 R_U^2\}^{1/2}]^{-1} \quad (4)$$

where R_L, R_U, and D are the radii of the lower and upper apertures and their distance apart, respectively.

3.5 The Electrical Measuring System

The electrical measuring system includes the controllers and power supplies for the 2 K reservoir; the blackbody radiator and its controlled shield; the stabilized power supply for the calorimeter heater and its current and potential measuring systems; the resistance measuring systems for the platinum resistance thermometers, germanium resistance thermometer on the calorimeter, and various other platinum and germanium thermometers placed elsewhere for monitoring purposes; together with all the ancillary electrical systems for pumps, pressure measurement, automatic liquid helium and nitrogen filling systems, etc. The resistance of the platinum thermometers on the blackbody was measured using a Tinsley ac resistance bridge Model 5840 and that of the germanium thermometer on the calorimeter by a Cryobridge of Automatic Systems Laboratories Ltd. The current and potential of the calorimeter heater were measured by means of a digital voltmeter, S.E. Laboratories Model SM215, connected in turn across the heater itself and a standard resistance arranged so that about 0.8 V was displayed by the voltmeter. The mean was always taken of the heater power measured before and after reversing the direction of the current through the heater. The differences obtained were small and constant amounting to about 5 parts in 10^5. Independent measurements were also made, as a check, using a Guildline Direct Current Comparator potentiometer Model 9930.

3.6 Residual Gas

For the energy received by the calorimeter to be a proper measure of the thermal radiation emitted by the blackbody, the energy transfer by gas molecules must be either negligible or corrected for. This energy transfer is made up of the kinetic energy appropriate to a gas molecule emitted from a blackbody at 273 K plus the energy of adsorption in the calorimeter at about 5 K. The energy of adsorption is solely that of physisorption, which does not exceed 36 kJ/mol, since the temperature of the calorimeter is much too low to allow subsequent chemisorption to take place. Elementary considerations based upon simple kinetic theory lead to the result that in our case the total energy transported to the calorimeter by residual gas, mainly composed of water vapor, at a pressure in the blackbody of 1.3×10^{-6} Pa (1×10^{-8} Torr), would be 1 part in 10^5 of the thermal radiation energy. Since the pressure measured inside the blackbody by an ion gauge was of the order of 4×10^{7} Pa (3×10^{-9} Torr) we considered the energy transport by residual gas to be negligible.

4. Results

A value of the Stefan-Boltzmann constant is obtained by measuring the electrical power, $P(T)$, required to balance the radiant power, $E'(T)$, received by the

calorimeter from the blackbody at a thermodynamic temperature, T, very close to the triple point of water. The equations used to calculate σ are as follows:

$$E'(T) = \epsilon(r)a(c)(1 - d)(1 - f)(1 + s)g\,\sigma T^4 \quad (5)$$

$$E'(T) = P(T) - \delta E' = V(c)V(s)/R(s) - \delta E', \quad (6)$$

where $\epsilon(r)$ is the emissivity of the radiator (0.999 92),

$a(c)$ is the absorptivity of the calorimeter (0.999 7),

d is the fraction of radiant energy lost due to diffraction (2.2×10^{-4}),

f is the fraction of radiant energy lost due to absorption at the lands of the apertures (2.8×10^{-4}),

s is the fraction of the energy added due to scattering in the radiation trap and at the apertures (1.3×10^{-4}),

g is the geometrical factor at 4.2 K (0.042 771 6),

$V(c)$ and $V(s)$ are the measured potentials (about 0.8 volts) across the heater and an external standard resistor, $R(s)$ (572 ohms), in series with the heater,

$\delta E'$ is the small change in radiant power entering the calorimeter from the radiation trap resulting from the cooling of the radiation trap (from about 6.5 K to 5.8 K) on closing the shutter.

The temperature T is deduced from the readings of the six platinum resistance thermometers placed at different points over the lower part of the blackbody viewed directly by the calorimeter. The mean effective temperature is a function of the different weights given to the radiation from different parts of the blackbody by the vignetting effect of the two apertures. The temperatures are evaluated on IPTS-68 and it is assumed that the differences between IPTS-68 and thermodynamic temperatures are less than 1 mK in the range from +2 °C to −2 °C.

The results of measurements made using the large set of apertures, R_L = 9 mm and R_U = 13 mm, are shown in Table 1 and Fig. 7. The size and associated uncertainty of the various correction factors appearing in Eqs. (5) and (6) are listed in Table 2.

The value of σ resulting from these preliminary measurements is

$$\sigma = (5.669\ 19 \pm 0.002\ 4) \times 10^{-8}\ \mathrm{Wm^{-2}\ K^{-4}} \quad (7)$$

where the uncertainty is given at the level of three standard deviations and is equivalent to 4.2 parts in 10^4. This differs from the value given above in Eq. (3) by 2.3 parts in 10^4. This is equivalent to a difference in the gas constant, R, of 58 ± 105 parts in 10^6. We consider this result to be satisfactory pending the repeat measurements with the small aperture set and further study of the principal sources of uncertainty given in Table 2.

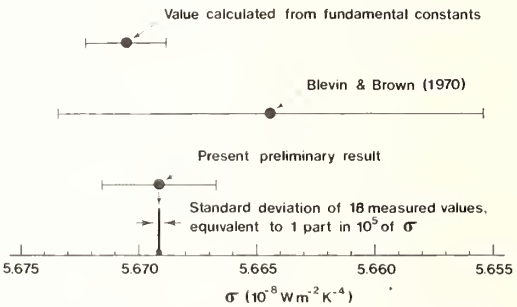


FIGURE 7. *A comparison of the present preliminary result, (5.669 19 ± 0.002 4) × 10^{-8} W m^{-2} K^{-4}, with that of Blevin and Brown, (5.664 4 ± 0.009) 10^{-8} W m^{-2} K^{-4}, and with the value of σ calculated from atomic constants using the most recent value for R, (5.670 51 ± 0.001 7) × 10^{-8} W m^{-2} K^{-4}. The uncertainty bars are drawn at the three standard deviation level.*

TABLE 1. *Results of eighteen measurements of the Stefan-Boltzmann constant, σ, made using the large set of apertures, $R_L \approx 9$ mm and $R_U \approx 12$ mm.* [a]

σ/ 10^{-8}W $m^{-2}K^{-4}$	T_{68}/K	σ/ 10^{-8}W $m^{-2}K^{-4}$	T_{68}/K
5.669 09	272.397 4	5.669 19	275.547 5
5.669 18	273.314 3	5.669 09	275.384 5
5.669 17	272.426 7	5.669 18	273.446 9
5.669 28	272.470 5	5.669 29	272.203 3
5.669 20	272.448 6	5.669 14	273.099 3
5.669 29	272.499 6	5.669 24	273.050 3
5.669 17	272.465 0	5.669 25	273.160 7
5.669 29	272.637 2	5.669 14	273.025 8
5.669 16	275.651 6	5.669 07	272.976 0

[a]The mean value is 5.669 19 × 10^{-8} W m^{-2} K^{-4} with a standard deviation, from the eighteen measurements, of 7 × 10^{-13} W m^{-2} K^{-4} equivalent to 1.2 × 10^{-5} of σ. The value of T_{68} is the mean temperature of the radiator at the time of each measurement.

TABLE 2. *Summary of corrections and associated uncertainties (evaluated at the one standard deviation level).*

Parameter	Correction to σ, parts in 10^4	Resulting uncertainty in σ, parts in 10^4
Emissivity of radiator, $\epsilon(r)$	+0.8	0.4[a]
Absorptivity of calorimeter, $a(c)$	+3.0	0.7[a]
Diffraction, d	+2.2	0.7
Absorption at aperture lands, f	+2.8	0.8
Scattering in radiation trap, s	−1.3	0.3
Geometrical factor, g	—	0.25
Measured electric power, $P(T)$	—	0.1
$\delta E'$	−0.1	0.1
Standard deviation of 18 measured values		0.1
Total correction	$+7.4 \times 10^{-4}\ \sigma$	
Combined standard deviation		$1.4 \times 10^{-4}\ \sigma$
Overall relative uncertainty at the three standard deviation level = 4.2×10^{-4}.		

[a]Although these two uncertainties are probably correlated, we have not taken this into account in combining them.

The authors are very pleased to acknowledge the many important contributions made in the early stages of this work by J. P. Compton. In addition we would like to acknowledge the able construction work carried out by E. Pinn, M. Rogers, and E. Charles of the NPL workshop, the aperture dimensional measurements made by S. Poole, G. Severn, and P. Kelley of the NPL length section, the measurements of the thermal expansion coefficient of beryllium copper carried out for us by C. A. Swenson of Iowa State University, the reflectivity measurements made on a model of the calorimeter by J. Geist and E. Zalewski of NBS, and finally the many helpful discussions which have taken place with members of the NPL temperature section.

References

[1] T. J. Quinn and J. E. Martin, in *Temperature: Its Measurement and Control in Science and Industry*, Vol. 5, Ed. by J. E. Schooley (American Institute of Physics, 1982), p. 103.

[2] A. R. Colclough, T. J. Quinn, and T. R. D. Chandler, Proc. R. Soc. London, Ser. A: 368, 125 (1979).

[3] W. R. Blevin and W. J. Brown, Metrologia 7, 15 (1971).

[4] D. C. Ginnings and M. L. Reilly, in *Temperature: Its Measurement and Control in Science and Industry*, Vol. 4, Part I, Ed. by H. H. Plumb (Instrument Society of America, 1972), p. 339.

[5] W. R. Blevin, Metrologia 6, 39 (1970).

[6] J. P. Compton, J. E. Martin, and T. J. Quinn, J. Phys. D 7, 2501 (1974).

[7] T. J. Quinn, High Temp. High Pressures 12, 359 (1980).

[8] The authors are indebted to J. Geist and E. Zalewski of NBS for measurements of the reflectance of a model of the calorimeter.

[9] The thermal expansion coefficient of beryllium copper was measured for us by C. A. Swenson of Iowa State University.

[10] F. J. Kroeger and C. A. Swenson, J. Appl. Phys. 48, 853–864 (1977).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

On the Radiometric Measurement of the Stefan-Boltzmann Constant at NRLM

Akira Ono

National Research Laboratory of Metrology, Umezono Niihari, Ibaraki 305, Japan

Progress at NRLM on the radiometric measurement of the Stefan-Boltzmann constant is described. In the method, the total radiant flux emitted from a blackbody source that is spatially defined by two limiting apertures is measured by an absolute radiometer. A newly developed absolute radiometer is described that has improved uniform responsivity over the surface of the receiver; the variation of responsivity is less than 0.2% over an area 15 mm in diameter. Emissivities of blackbody cavities are calculated by the Monte Carlo method taking into account partial specular reflection of radiation on the cavity walls. An approach to a perfect blackbody cavity is also presented. It is discussed how uncertainties of blackbody source temperature and of diffraction losses of radiant flux are to be minimized.

Key words: absolute radiometer; blackbody total radiation; emissivity of cavity; partially specular reflection; radiometric measurement; Stefan-Boltzmann constant.

1. Introduction

Since the Stefan-Boltzmann constant was precisely determined by a radiometric method at the National Measurement Laboratory (total uncertainty of 1300 ppm, 99% confidence level) [1], attention has been directed to possible improvements of the method at the National Research Laboratory of Metrology (NRLM). A radiometric measurement of the Stefan-Boltzmann constant involves a combination of radiometric and radiation thermometric techniques. The former is based upon the absolute measurement of radiant flux while the latter provides a blackbody radiation source maintained at a known temperature. Therefore a radiometric measurement of the Stefan-Boltzmann constant will contribute to our understanding of the quantitative consistency between radiometry and thermometry, as well as to a precise determination of the fundamental constant.

The principle of the method is based upon an absolute measurement of total radiance from a blackbody whose thermodynamic temperature is accurately known. Total radiant flux that is spatially defined by two limiting apertures is measured by a thermal detector sensitive over a wide spectral range. The radiation detector is called an absolute radiometer because it senses flux values on the absolute scale in units of power (watts). Major sources of error in the method result from uncertainties of the quality and temperature of the blackbody source, of the absolute measurement of radiant flux, and from the diffraction losses of radiant flux between the two limiting apertures.

The purpose of this paper is to describe recent progress at NRLM on the radiometric measurement of the Stefan-Boltzmann constant. First, an absolute radiometer is described that has improved uniform responsivity over the surface of the receiver. The total uncertainty of the radiant flux measurement has been reduced to less than 100 ppm, one standard deviation. Second, the evaluation of blackbody cavity emissivity is discussed. Since a major part of blackbody total radiation is in the infrared, specular reflection of radiation on surfaces can be important. A Monte Carlo method is applied to calculate apparent emissivity of the blackbody cavity taking into account partial specular reflection on the cavity walls. Finally, discussion is presented on how uncertainties associated with the blackbody source temperature and with the diffraction losses of radiant flux are minimized.

2. Experimental Apparatus

The experimental apparatus shown in Fig. 1 consists of three major components: the upper cell composed of the blackbody source, the lower radiometer section, and the middle aperture section which also provides for one of two connections to the vacuum system. A water-cooled shutter located between the blackbody source and the upper limiting aperture (source aperture) allows blackbody radiation to be incident upon the absolute radiometer. Otherwise, when the shutter is closed, the radiometer views the shutter which emits blackbody radiation at room temperature.

The absolute radiometer is a thermopile that provides an electrical power equivalent to the absorbed radiant power. The hemispherical mirror above the radiometer serves to attain a more uniform responsivity over the surface of the receiver, as well as to increase the apparent absorptivity of radiation. Since the two limiting apertures, both of 10 mm diameter, are spaced 400 mm from each other, this arrangement provides flux values of 500 μW, 160 μW, and 90 μW for blackbody source temperatures of the aluminum point (660 °C), the zinc point (420 °C), and the lead point (328 °C), respectively.

Since the entire system is evacuated, thermal insulation by radiation shielding permits the blackbody source and the source aperture to be in close proximity (150 mm). The aperture system (the lower half in Fig. 1) can be separated physically from the vessel of the blackbody source (the upper half in the figure). When the aperture system is removed, the angular distribution of irradiance at the location of the source aperture is measured. These facilitate reducing and evaluating diffraction losses of radiant flux between the limiting apertures.

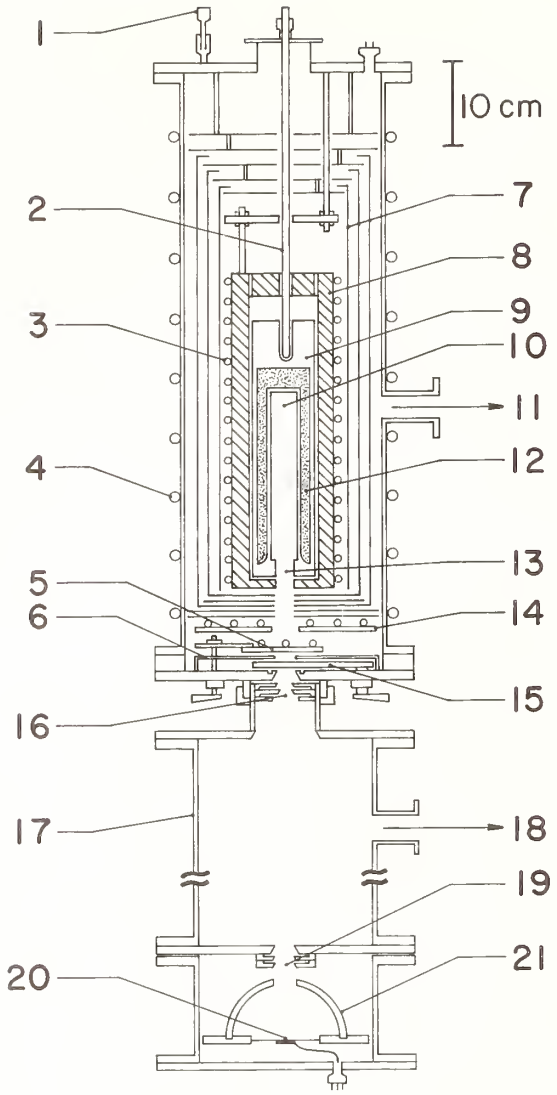


FIGURE 1. *Experimental apparatus. 1—vacuum gauge; 2—thermometer well; 3—sheath heater; 4—water cooling; 5—water-cooled shutter; 6 and 7—radiation shield; 8—copper tube; 9—graphite crucible; 10—cavity 30 mm in diameter and 200 mm deep; 11—vacuum pump; 12—fixed-point material; 13—cavity aperture 20 mm in diameter; 14—water-cooled shield; 15—gate valve; 16—source aperture 10 mm in diameter; 17—radiation trap; 18—vacuum pump; 19—detector aperture 10 mm in diameter; 20—absolute radiometer; 21—hemispherical mirror.*

3. The Absolute Radiometer

An exploded view of the absolute radiometer is illustrated in Fig. 2. The design principle assumes that uniform responsivity can be attained by permitting regular heat flow through the thermocouple legs but suppressing heat losses between the receiver and the surroundings [2].

Suppression of heat losses was achieved as follows. Heat losses by air convection and conduction are minimized by operating the radiometer in a vacuum. Heat loss by thermal radiation from the front surface of the receiver is minimized by placement of the hemispherical mirror above the receiver. The hemispherical mirror causes thermal radiation emitted from the front surface of the receiver to be reflected back onto the receiver itself. Heat loss by thermal radiation from the rear surface of the receiver is minimized by coating the surface with a low emissivity material (gold).

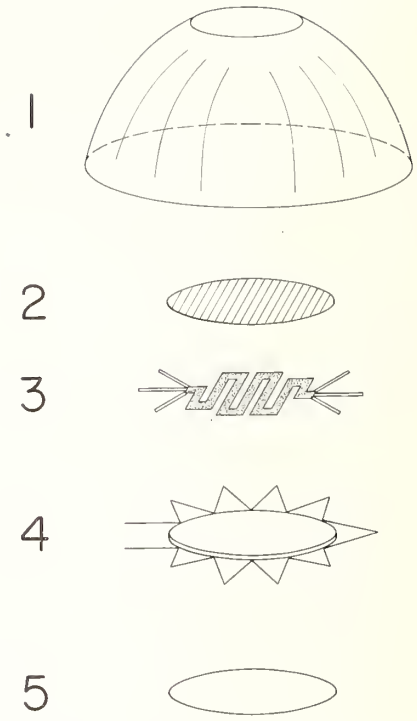


FIGURE 2. *Exploded view of the absolute radiometer. 1—hemispherical mirror; 2—black coating; 3—heater; 4—receiver disk with thermocouples; 5—gold coating.*

The responsivity distribution over the surface of the receiver has quadratic behavior with a minimum at the center and a maximum at the periphery. The responsivity variations were less than 0.2% over the receiver surface of 15 mm diameter [3]. The responsivity and time constant of response of the absolute radiometer are listed in Table 1 with construction details. It should be pointed out that a more uniform responsivity can be obtained by a thicker disk without serious increase of the time con-

TABLE 1. *Performance and construction details of the absolute radiometer.*

Responsivity	0.17 V/W
Variation of responsivity	0.2%
Time constant of response	14 s
Receiver	
area	15 mm in diameter
disk	aluminum 20 μm thick
black coating	3M velvet spray
heater material	gold evaporated film
heater resistance	54 Ω
insulating film	Araldite 985E
Thermocouple	
material	copper vs. constantan
number of junctions	24 pairs
copper wires	0.05 mm in diam. × 15 mm long
constantan wires	0.32 mm in diam. × 15 mm long
total resistance	6.4 Ω
Hemispherical mirror	
radius	60 mm
concave depth	55 mm
aperture	40 mm in diameter
surface	gold coating

stant [4]. An uncertainty of radiant flux measurement caused by the responsivity variation over the receiver surface can be confined to a few parts in 10^5. As such, the total uncertainty of the radiant flux measurement will depend entirely on the effective absorptivity of the absolute radiometer.

4. Apparent Emissivity of the Blackbody Source

Assuming that the emissivity and the absorptivity are equivalent for an isothermal opaque cavity, a Monte Carlo method has been applied to the calculation of the apparent absorptivity of partially specular cavities [5]. Radiative properties of cavity surfaces are characterized by the uniform specular-diffuse model in which radiation reflected on a surface consists of two components—specular and diffuse—and the directional hemispherical reflectivity has the same value at all points on the surfaces considered. Figure 3 illustrates a cylindrical cavity with a conical bottom where the cylindrical surface is diffuse and the conical surface is intermediate between diffuse and specular. This latter surface can be characterized by the uniform specular-diffuse model. Such modeling may be a good approximation for a real cavity where the cylindrical surface is spirally grooved or baffled and the conical surface is in the as-machined condition.

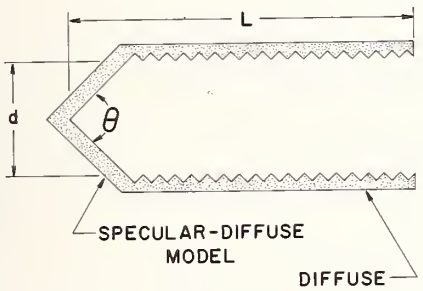


FIGURE 3. *Schematic of a cylindrical cavity with a conical bottom. The cylindrical surface is diffuse and the conical surface is intermediate between specular and diffuse.*

Normal apparent emissivities of cylindrical cavities with conical bottoms were calculated for various cavity depth-to-diameter ratios, apex angles of the cone, surface emissivities, and fractions of specular reflection component. It was shown that the apex angle of the cone was an important factor in the apparent emissivity behavior with partially specular reflection. For apex angles near 180°, 90°, and 60°, the apparent emissivities decreased rapidly as the fraction of specular component increased. This behavior is due to specular reflection effects on the conical surface. For apex angles of 70° and 50°, the apparent emissivities exhibited minima at intermediate fractions of specular component. This behavior is due to the combined effects of diffuse and specular reflections on the conical surface. For an apex angle of 120°, the apparent emissivity increased monotonically as the fraction of specular component increased.

Figure 4 shows apparent emissivities of a cavity having a depth-to-diameter ratio of 10, an apex angle of 120°, and a surface emissivity of the cylinder of 0.8. The apparent emissivity is plotted against the fraction of specular component on the conical surface for various surface emissivities of the cone. If the surface emissivity of the

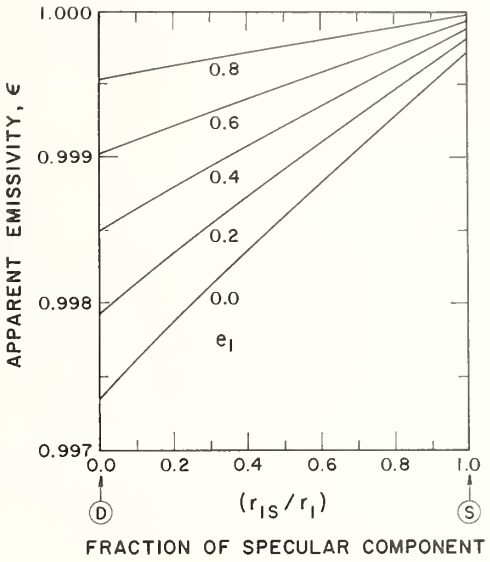


FIGURE 4. *Normal apparent emissivity of a cavity having a depth-to-diameter ratio of 10, an apex angle of 120°, and a surface emissivity of the cylinder of 0.8. It is plotted against the fraction of specular reflection component for different surface emissivities of the cone. Two arrows, D and S, on the abscissa indicate totally diffuse and completely specular reflection, respectively.*

cone and the fraction of specular component are estimated as $e_1 = 0.8 \pm 0.05$ and $r_{1s}/r_1 = 0.5 \pm 0.1$, respectively, then the apparent emissivity is evaluated as $\epsilon = 0.99975 \pm 0.0001$. It was shown that the surface emissivity of the cylinder affected the apparent emissivity much less than the surface emissivity of the cone. An effort to make the conical surface specular can be an approach to a perfect blackbody cavity even if the surface emissivity of the cone is not so high.

In closing, it should be noted that the effect of temperature distribution within a cavity has also been studied intensively at NRLM. A new concept of effective temperature of a cavity has been introduced that is convenient for the analysis of the radiant characteristics of non-isothermal cavities [6, 7, 8].

5. Temperature of the Blackbody Source

An uncertainty of the thermodynamic temperature scale affects the determination of the Stefan-Boltzmann constant according to the fourth-power law. The thermodynamic temperature scale has been realized with high accuracy at the National Bureau of Standards by gas thermometry [9]. The uncertainties are less than 10 ppm from 0 to 450 °C. It has been also reported that the scale is being extended to higher temperature, including the aluminum point (660 °C) [10]. Since our interest is in the aluminum, zinc, and lead fixed points as described later, an uncertainty resulting from the thermodynamic temperature scale will be confined within 50 ppm in the determination of the Stefan-Boltzmann constant.

In order to avoid uncertainties resulting from temperature measurement of the blackbody source, use has been made of blackbody sources at the fixed points on the International Practical Temperature Scale by packing the fixed-point materials around the blackbody cavities. A preliminary experiment using the zinc point was not successful because it was difficult to have a reliable airtight

crucible to prevent zinc vapor from passing out the crucible wall to the vacuum. Our interest is shifting to the aluminum and the lead points because these materials have vapor pressures much less than that of zinc at their freezing temperatures.

6. Uncertainties

Major factors in the uncertainty of the present measurement are: quality of the blackbody source, absolute measurement of the radiant flux, and diffraction losses. Uncertainties of the first two factors can be reduced to less than 100 ppm for each, one standard deviation, as reviewed here and analyzed in the earlier measurement [1]. The third factor is still difficult to estimate with confidence at this stage because data are lacking on the angular distribution of irradiance at the source aperture. If this factor could be evaluated with an uncertainty of less than 100 ppm, the total uncertainty in the present measurement of the Stefan-Boltzmann constant can be confined to 160 ppm, one standard deviation.

The author would like to express his sincere thanks to Dr. M. Morimura and Dr. S. Hattori of the National Research Laboratory of Metrology for their helpful discussions, and to Professor D. P. DeWitt and Dr. R. E. Taylor of Purdue University for their help in preparing the paper during the author's sabbatical leave at their laboratories.

References

[1] W. R. Blevin and W. J. Brown, Metrologia 7, 15 (1971).
[2] A. Ono, Metrologia 15, 127 (1979).
[3] A. Ono, Jpn. J. Appl. Phys. 18, 697 (1979).
[4] A. Ono, ibid. 18, 1995 (1979).
[5] A. Ono, J. Opt. Soc. Am. 70, 547 (1980).
[6] S. Hattori, Trans. Soc. Instrument and Control Engineers (Japan) 15, 847 (1979), in Japanese.
[7] S. Hattori, ibid. 16, 539 (1980), in Japanese.
[8] S. Hattori and A. Ono, to be published.
[9] T. J. Quinn, L. A. Guildner, and W. Thomas. Consultative Committee on Thermometry, Document Number 76–2, International Bureau of Weights and Measures, Paris (1976).
[10] T. J. Quinn, L. A. Guildner, A. Moser, and W. Thomas, ibid. Document Number 80–5, International Bureau of Weights and Measures, Paris (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Applications of X-Ray Interferometry

Richard D. Deslattes*

Division of Physics, National Science Foundation, Washington, DC 20550

This review begins by summarizing work at the PTB and NBS on optical interferometry of [220] repeat distances in samples of monocrystalline Si. Distribution of such an initial calibration to other samples and other species is briefly mentioned. The main emphasis is on subsequent applications of these crystals toward determination of fundamental constants, especially N_A and extension of the congruent electromagnetic scale to γ-rays as has so far been carried out at NBS. In the last mentioned case, applications emerge which include tests of QED in muonic atoms, determination of mass values for the pion and the kaon, and tests of relativistic self-consistent field calculations for inner vacancy states in atoms.

Key words: density; fundamental constants; gamma-ray standards; molar mass; x-ray interferometry.

1. Introduction

This review aims to give an overall perspective on x-ray interferometry with emphasis on its role in the determination of fundamental constants and in tests of basic theory. As is well known, these roles have been exemplified in the determination of the Avogadro constant and in the extension of the congruent electromagnetic scale to the γ-ray region. In each of these applications, x-ray interferometry alone is insufficient. Clearly it first needs to be combined with optical interferometry to obtain an optically based crystal lattice repeat distance. However this step is also insufficient to be of general interest since a crystal sample is a quasi-artifact. Other measurement technologies need to be added in order to reach manifestly invariant physical quantities. In the case of the Avogadro constant, these other technologies involve density determinations as well as absolute isotopic abundance measurements. In the case of γ-rays, angle measurements are required having both high precision and high accuracy. These additional steps needed to reach invariant results enter those results just as sensitively as the initial interferometric determination itself. For the present overview, it will be necessary to give attention to these other steps as well, at some cost to the level of detail with which the x-ray interferometry can be outlined.

Another kind of auxiliary resource, though not needed in principle, is a matter of considerable convenience. This is a crystal-to-crystal comparison method which permits transfer of the calibration initially embodied in one single crystal specimen to other specimens and other species. Once again the technique is required to be of high quality since its results directly influence numerical values for the invariant quantities ultimately produced. Such a technique also permits, as will be seen below, comparisons of work carried out in different laboratories at preliminary stages where invariant results are not available in all cases.

1.1 Historical Remarks

Present day work on x-ray interferometry began with the pioneering efforts of U. Bonse and M. Hart in the mid-sixties [1]. Their work initially explored fixed geometry arrangements of crystal diffracting elements which remained part of the original crystalline sample. Although such devices have had a considerable range of applications (see below), the main interest for this Conference attaches to those generalizations in which one crystal element has freedom of motion with respect to the others. This articulation, accompanied by suitable optical interferometry, permits measurement of the lattice repeat distance of the crystal in terms of an optical standard. Such structures were described at the first of these Conferences in 1970 and have been pursued with some diligence in several laboratories since that time [2, 3].

1.2 Applications of Fixed Geometry Interferometry

Aside from the initial demonstration of phase contrast microscopy, fixed geometry interferometers have found their main utility in the measurement of refractive indices. These have, of course, considerable interest near critical absorption edges where anomalous dispersion effects become large. Continuously wavelength scannable photons available from synchrotron radiation sources make their combination with x-ray interferometry a very natural and significant enterprise. Such work has been reviewed in a number of places especially by Hart [4], by Materlik [5], and by Bonse *et al.* [6]. It is also the case that the developments of fixed geometry x-ray interferometry presaged the corresponding work with deBroglie waves, i.e., neutron interferometry. This work has already been quite fruitful in demonstrating important fundamental properties of Fermions and their interactions with gravity and non-inertial motion. The overall situation of neutron interferometry for these purposes has been well summarized [7].

1.3 Systems Providing for Motion and Optical Readout

As is well known, when the last optical element in a three-crystal interferometer is moved uniformly with respect to the other two elements, x-ray intensity recorded in a detector, see Fig. 1, is modulated with the spatial periodicity of the lattice in a fashion that is substantially independent of the wavelength and characteristics of the source of x-ray illumination. Evidently optical

*Permanent Address: Center for Absolute Physical Quantities, National Bureau of Standards, Washington, DC 20234

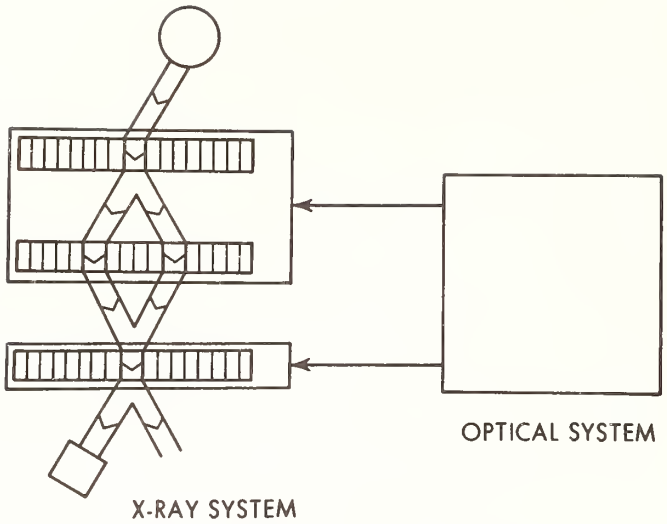


FIGURE 1. *Diagram indicating the x-ray and optical components of a lattice spacing determination.*

interferometry of a sufficiently refined character carried out simultaneously permits a reckoning of the crystal lattice repeat distance in terms of the visible wavelength illuminating the optical interferometer. If the source of optical radiation is well-connected with fundamental scientific or metrological standards, then these standards are effectively transferred to the crystalline lattice. There were a number of problems needing to be solved before such measurements became practical and even today they can hardly be considered routine. Most evidently, a provision for positional noise levels and optical interferometry near the picometer level and the maintenance of nanoradian tolerances in angular orientation, together with similar parallelism of motion, represented somewhat formidable challenges. Two such exercises have now been completed and will be discussed in some detail below.

1.4 Objectives and Cautions

Clearly the physical properties (including dimensions of the lattice) of a particular crystal, even if it were known to be perfect, are of very limited scientific interest since these properties are not calculable from first principles to any appreciable accuracy. Thus such calibrated crystals are interesting only as "stepping-stones" helping to bridge the large gap between relatively coarse visible wavelengths and the hundred- to million-fold smaller wavelengths in the x-ray and γ-ray region. Of course, if the particular crystal measured is otherwise well characterized or if the initial calibration is transferred to specimens of well characterized material, one has the additional hope of being able to determine the Avogadro constant. However, even in this case, our experience (see below) indicates that each sample has its own macroscopic density and its own isotopic abundance distribution so that only the formally invariant ratio of mean molar mass to density is effectively constant for a group of specimens coming from different sources. There is also the additional coupling of nuclidic mass to thermal vibration which leads, at least in high order, to a dependence of lattice parameter on mean molar mass so that, in the final analysis, the only truly invariant quantity for even a perfect crystal is the Avogadro constant itself.

When one enters the real world of crystalline samples produced by a variety of techniques and perhaps not fully characterized as regards density and isotopic abundance, the complications become even greater. Several explorations of groups of materials from different sources have shown lattice variability greater than 1 ppm from process to process and smaller than 1 ppm within a given manufacturing process [8, 9]. Of course, there can be, and often are, underlying changes in the manufacturing process which are not revealed by trade name designations. In summary, although the crystals with which we work are objects of considerable beauty, they should probably be treated with some of the cautions that were properly applied to various attempts at realization of the meter before it was fixed to be embodied in a single artifact.

2. Optical Interferometry of the 220 Period in Samples of Si

Measurement of a crystal lattice repeat distance evidently requires provision for parallel translation of one element of the x-ray interferometer with respect to the remainder together with closely coupled optical displacement sensing. The required delicacy of linear motion, the need for mutual orientation adjustment, and the essential parallelism of motion, can be obtained in a number of different ways. Several of these were described at the first Conference of this series [2] while others have been initiated subsequently [3]. Although each of these efforts includes interesting strategies, limitations of length and my lack of detailed current awareness suggest that this report should focus on the two efforts that have been carried to completion. One of these has recently been reported by a group at PTB [10] while the other earlier work was carried out at NBS [11]. Each result that has been obtained has an adequate precision to serve as a useful first step in a subsequent chain to invariant quantities. Since only the NBS work has led to determination of invariant quantities, statements regarding consistency of the x-ray/optical interferometry have to await direct sample-to-sample comparisons [12].

2.1 Translation Stages

Although Hart had earlier devised an elegant and ingenious means for providing parallel translation within a single block of silicon which included also the interferometer elements [13], both NBS and PTB have used flexure structures made of steel as carriages for separated interferometer elements. Figures 2(a) and 2(b) compare the structures used in the two measurements. The PTB structure, Fig. 2(a), is of highly symmetric design and is driven directly (in its final embodiment) by a piezoelectric transducer [10a]. The NBS structure, Fig. 2(b), is of a less symmetric design consisting of a flexure parallelogram internal to which there is a three-stage motion demagnifier. Input drive to the NBS structure (applied at point X) comes from a predominantly mechanical source, but one which includes a piezoelectric vernier [11d]. In each case sensitive tuning adjustments are provided so as to secure critical alignment. Considerable efforts were required in each case to provide a vacuum environment relatively free of mechanical disturbance. It is characteristic of the peripheral problems associated with these efforts that each began with a seismic survey

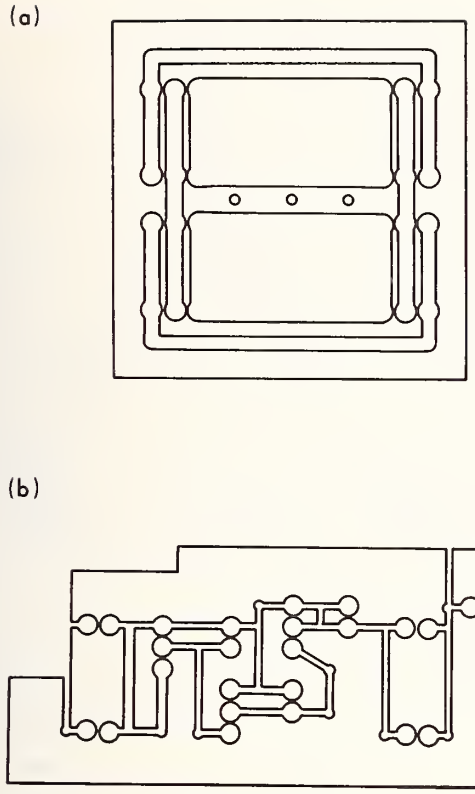


FIGURE 2. *Linear motion devices used in PTB (a) and NBS (b) measurements of Si 220 lattice periods.*

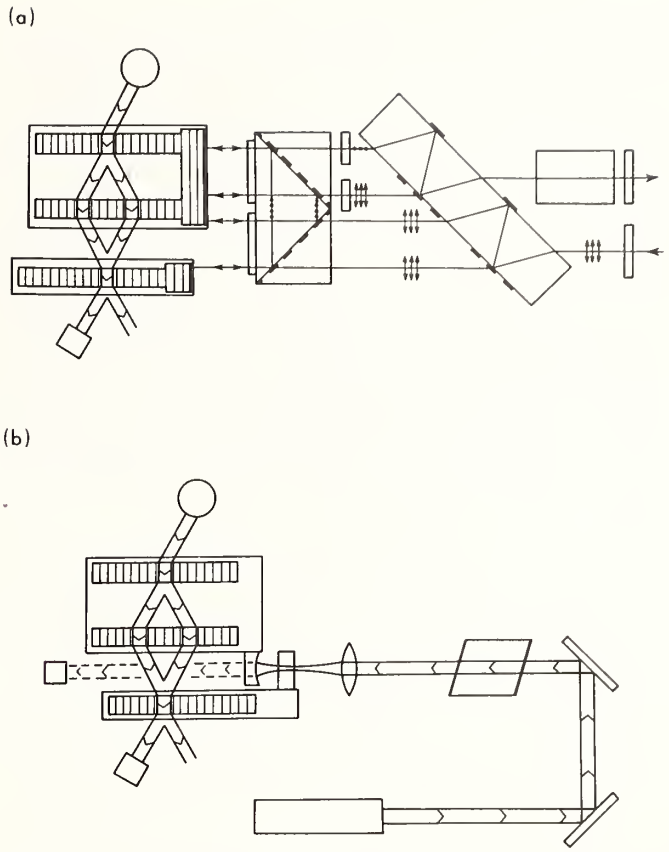


FIGURE 3. *Optical systems used in the PTB (a) and NBS (b) measurements. The system shown in (a) has polarization encoded two-beam optics while that in (b) is an amplitude (derivative) read high finesse Fabry-Perot interferometer.*

of available space on the respective sites! In addition to well-chosen environments, there were appreciable efforts required in the way of antiseismic supports and acoustic shielding.

2.2 Optical Interferometry

Although NBS and PTB measurements were each referred back to iodine stabilized lasers, the PTB effort used a Lamb dip stabilized laser periodically calibrated by beat frequency measurements with respect to a $^{127}I_2$ laser while the NBS work employed a $^{129}I_2$ laser directly. The calibration chain for the laser is in each case sufficiently well known as to introduce negligible uncertainty in the final results. For future reference, I note here the wavelengths assigned to the two lasers: λ(PTB) = 632 991.415 pm (9 ppb); λ(NBS) = 632 990.079 pm (4 ppb).

The optical interferometric techniques used in the two measurements were quite different. In its final version the PTB work has made use of the elegant polarization encoding scheme devised by Curtis *et al.* [2b]. This is shown in the diagram in Fig. 3(a). The NBS measurement employed, on the other hand, high finesse Fabry-Perot optics as indicated in Fig. 3(b) [11a, d].

Among the many advantages of the optical interferometer used in the PTB measurements are a strong set of invariance properties enjoyed by the optical system of Fig. 2(a). As was shown by Curtis *et al.* [2b], this structure is immune to bodily motions of the two crystals together and sensitive only to the difference in positions. As demonstrated in the detailed analysis by the PTB group, the relationship of optical and x-ray interference trajectories is sufficiently congruent that offset and cosine type errors introduce negligible corrections into the final result [10a]. On the other hand, one is here dealing with essentially two-beam fringes; use of polarization encoding enhances the sensitivity to a very great extent but the arrangement probably remains somewhat less sensitive than a high finesse Fabry-Perot cavity. Data-taking strategy involved summation of repeated scans through a region of fixed polarization azimuth while information from both optical and x-ray detectors was accumulated in a multichannel memory provided by a computer.

The optical plan of the NBS measurement shown in Fig. 3(b) offers certain advantages and certain disadvantages and was used with a somewhat different data acquisition strategy. First of all, the hemispherical Fabry-Perot cavity exhibited an effective optical finesse of approximately 1200. This meant that the full width at half maximum of an optical fringe was rather close in its shape to that of the (two-beam) x-ray fringe. These relatively desirable features came, of course, at the expense of having to mount separate optical elements which, on the present scale of refinement, exhibited appreciable drift with respect to the x-ray elements. In the end this

did not appear to be a serious limitation since, after suitable aging, drift became both slow and uniform. Thus, by averaging the results of scans taken in opposite directions, we obtained data equivalent to what would have been obtained in the absence of drift. Another difficulty with the NBS measurement scheme is evident also in Fig. 3(b). Clearly the trajectory along which optical interference takes place is not degenerate with the trajectory along which x-ray interference takes place. The offset, z, introduces a cosine error to the extent of $z\,\delta\theta/L$ for a traverse of length, L, accompanied by a directional change of $\delta\theta$. It was thus required to determine path curvature with appreciable refinement. In fact this problem entailed a correction to the measured lattice parameter of 0.27 ± 0.03 ppm [11b]. Finally one notes that the resonances of the hemispherical structures shown are not separated by $\lambda/2$ but contain an additional Fresnel phase shift whereby intervals between optical resonances are displaced by approximately 1.6 ppm [11a, d].

Data-taking in the NBS work did not benefit from the presence of a computer. Instead, incidental frequency modulation of the I_2 stabilized laser (as viewed through the hemispherical resonator) formed an input to a traditional lock-in amplifier system. At each lock point we determined x-ray intensities using the count-rate-meter technique or its digital equivalent. From these intensities and an estimate of maximum and minimum interference signals, a standard cumulative phase versus optical order number plot was used to determine the results.

Although, as described above, the mechanical and optical technologies employed in the NBS and PTB work were quite different, it would appear that each path was entirely adequate to obtain results of the stated precision and accuracy (see below). There may be, however, some subtleties especially in the optical interferometry that have escaped the rather diligent efforts expended by both groups.

2.3 Experimental Results

The output datum from an x-ray/optical interferometry exercise is an integer (in this case it is always 1648) plus a fractional part representing the size of the lattice period modulo the optical half-wavelength. This numeric translates, via the optical wavelength, to a (specimen dependent) lattice parameter. Robust comparison of the two results is thus not directly possible at this stage.

Histograms of measurements are shown in Figs. 4(a) and (b). Figure 4(a) gives the data from PTB while Fig. 4(b) gives the earlier results from NBS. Each of these must be corrected for the different visible wavelengths and then referred back to a common reference temperature. (Both groups used $\alpha = 2.56 \times 10^{-6}\,K^{-1}$ so that the temperature corrections do not introduce noticeable divergence.) The results for the two different samples are (referred to 22.5 °C):

$$\text{PTB: } d = 192\ 015.560 \pm 0.012 \text{ fm}$$

$$\text{NBS: } d = 192\ 015.902 \pm 0.019 \text{ fm}$$

The silicon sample used in the PTB work was obtained from Wacker Chemitronic and had low concentrations of C and O. The crystal used in the NBS work was obtained in the mid 1960's from Dow, is of high geometrical quality, and contains unmeasurably small concentrations of C and O. Earlier discussions of these results [10a, b] were based on the assumption that the two specimens shared a common lattice parameter so that the entire noted difference (1.8 ppm) was due to inconsistency between the two efforts. Subsequent efforts toward direct comparison [12] suggest that the two specimens differ in lattice parameter by less than 1 ppm. This preliminary work showed apparent secular change in the PTB sample which remains to be further investigated.

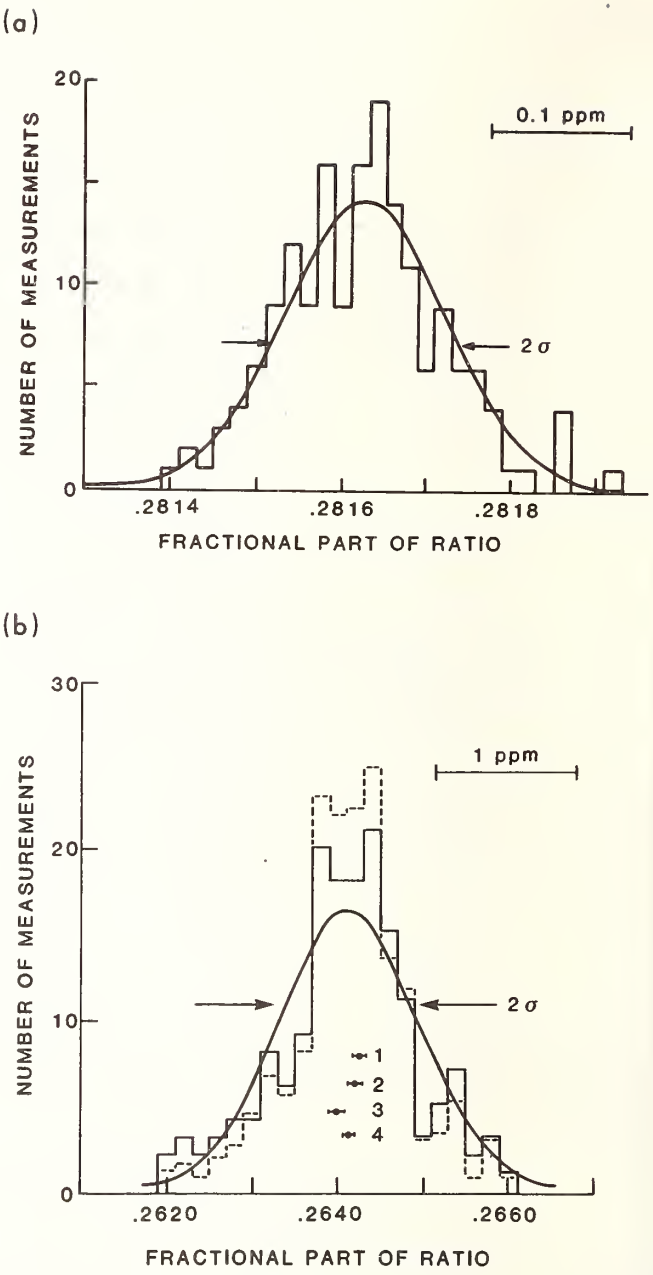


FIGURE 4. *Histograms of x-ray/optical interferometry results from PTB (a) and NBS (b). Note that because of different optical sources, the fractional scales are not directly comparable. In addition (see text), samples used had different sources and may not be directly comparable.*

3. Application to Determination of the Avogadro Constant

Given an $(hk\ell)$ lattice repeat distance obtained as described above, one can (for a cubic crystal) proceed to a unit cell volume, a_0^3 by means of $d_{hk\ell}(h^2 + k^2 + \ell^2)^{1/2} = a_0$. The mass contained in such a unit cell may be reckoned in two different ways. For a macroscopic density, ρ, the unit cell mass is evidently ρa_0^3. If there are $n = 8$ atoms in the unit cell and these have a mean molar mass, A, then there is a microscopic mass amounting to $8A(u)$ also associated with the same unit cell. Connecting the macroscopic mass ρa_0^3 with the microscopic one, $8A(u)$, is the numerical constant, N_A. Thus, N_A itself results from: $N_A = 8A/\rho a_0^3$.

In this section attention is directed toward determination of ρ and of A. These are exercises of more than passing difficulty because interest in a 1 ppm estimation of N_A requires a first-principles determination of density and absolute isotopic abundance measurements of the highest available quality. Efforts in both of these areas have been carried out at NBS while considerable progress in the case of density is reported from PTB at this Conference. New efforts toward absolute density measurements of high accuracy are also reported at this Conference by groups from IMGC (Italy) [13a] and CSIRO (Australia) [13b].

3.1 Density

There are several methods available for a first principles approach to crystal density determination. The most straightforward would appear to be to use dimensional metrology to fix the volume of an otherwise well-characterized Si sample; since mass determinations are rather easy even well below the 1 ppm level, one has thereby a quite adequate density estimate. Such an approach was proposed by Curtis *et al.* [2b] and is also being considered by the PTB group [10b]. Problems of this approach include difficulty in producing and maintaining regularity in the case of highly valued samples.

Another approach makes use of whatever regular solid object can be obtained conveniently. Such an approach is also being considered by the PTB group where it will involve using low expansion ceramic material both in the form of balls and cubes [10b]. By contrast, NBS made use of steel balls of high quality which were readily available as items of commerce. Volumes of these objects were determined by optical interferometry [14] after tests of sphericity showed that diameter measurements alone were adequate to yield volume estimates [15]. Unfortunately there remains an unexplained systematic offset of 1.6 ppm between the results obtained after two different cleaning procedures [16].

However one obtains a macroscopic artifact density standard (other than the Si crystal itself), it is still necessary to make density transfer measurements either directly to well-characterized Si samples or indirectly through the use of an intermediate collection of Si objects which need not be well-characterized. Such transfer measurements are most readily effected by hydrostatic weighing with the evident freedom to use immersion media other than water. The course to be followed at PTB in this regard is not yet clear.

In the NBS work, density values inherent in a group of steel spheres were transferred to a group of four (there were five but . . .) 0.2 kg Si samples by hydrostatic weighing [17]. These weighings were carried out in a fluorocarbon liquid having large density (1.8 gm/cm^3) and low surface tension (one-fifth that of water). Subsequently, these four transfer standards were compared with three well-characterized Si crystals whose individual isotopic abundances and contamination by C and O were also determined. The multiple samples, use of three spheres, permutation of groups of crystals (six groups of two in the sphere comparisons) together with the four secondary standards gave a significant overdetermination from which the precision of this step could be objectively assessed. The resulting values for the well-characterized crystals exhibited a transfer contribution to the final uncertainty of approximately 0.3 ppm. The large discrepancy in ball diameters obtained after different cleaning procedures remains unresolved forcing an expansion of the estimated uncertainty from near 0.5 ppm to about 0.95 ppm [16].

A summary of density values obtained for the three crystals used is shown in Table 1. For reasons already discussed, their relative values are considerably better known than the absolute values indicated. It is significant that the spread in data among these samples is quite large, ~3 ppm. As will be seen below, there is a corresponding variability of molar mass which correlates very closely with these density differences.

3.2 Isotopic Abundances—Mean Molar Mass

Evidently, one can obtain the quantity of microscopic mass in a unit cell by a fractional abundance weighted sum of isotopic masses. The NBS approach to such problems is well established but had to be pressed quite hard to obtain accuracies below 1 ppm. The rather long chain of steps has been described in detail elsewhere [18]. Mention of a few highlights is, however, in order.

A large Si crystal (not otherwise involved in our measurement) was tested for uniformity of its isotopic abundance ratios. These ratios were then established on an "absolute" basis by comparison with synthetic mixtures of separated isotopes. Complications are entailed by the

TABLE 1. *Density values for specimen crystals as observed and corrected for C and O.*

Sample	Observed density (g/cm^3)	Corrected density (g/cm^3)	σ_m (ppm)	σ_T (ppm)
D_1	2.3289942	2.3289935	0.22	0.97
D_2	2.3289931	2.3289924	0.44	1.04
D_3	2.3290028	2.3290028	0.95	1.33

TABLE 2. *Abundance and molar mass values for SRM-990 and samples.*

Isotope	^{28}Si	^{29}Si	^{30}Si
atomic %	92.22933	4.66982	3.10085
nuclidic mass	27.9769286	28.9764969	29.9737722
weight %	91.872710	4.817963	3.309326
A(SRM-990) = 28.0855258			
Sample			Molar Mass
D_1			28.085400
D_2			28.085379
D_3			28.085507

need to adequately characterize abundances in the "separated" isotopes and by the very small sample quantities of the less abundant species which were available at the time of our measurements. The large homogeneous boule was sliced into disks which are publically available through the Standard Reference Materials Program at NBS as SRM 990 [19]. This public availability invites testing of the reference abundances on which our molar mass values are based.

All subsequent measurements of abundances were carried out relative to those in SRM 990. These comparisons were made in a 15 cm, 60° magnetic sector mass spectrometer having a well characterized transmission function. This instrument uses gas phase samples, in the present case SiF_4, prepared by thermal decomposition of $BaSiF_6$. Initial samples of both separated isotopes and samples of crystalline material were prepared as fluosilicic acid which was converted successively to Cs_2SiF_6 and thence to $BaSiF_6$.

Results obtained for both SRM 990 and the three crystals whose densities were measured are shown in Table 2. Contributions to the estimate of σ in the samples came from chemical synthesis (0.3 ppm), analysis of abundances in the separated isotopes (0.3 ppm), and uncertainty in SRM 990 (0.5 ppm). Quadrature combination of these components yields a composite σ of 0.65 ppm. Table 2 also shows the weighted sum values for mean molar mass associated with each sample. It should be noted that the spread of values is quite appreciable but see below.

3.3 Evaluation of N_A

The three quantities required for obtaining N_A are the unit cell volume, a_0^3, macroscopic density, ρ, and molar mass, A [11e]. For the cell volume, a correction is required to take account of the fact that interferometry was done in vacuum while density was measured at atmospheric pressure. This correction was −1.02 ppm for the NBS measurements. Each of the crystals has, of course, its own cell dimension; these were, however, found to be sufficiently close together that the interferometer result could be used for all three samples. Two of the samples had appreciable concentrations of both C and O. The numerical values were, however, such that the expansion due to the latter was closely compensated by the contraction due to the former.

Concordance is not so evident in the case of the density and molar mass values given above. If it is assumed, however, that such density variation as is encountered follows entirely from the molar mass changes then the quantity A/ρ should exhibit a smaller range of variability. That such is indeed the case is evident from Table 3. The range of variation is, in fact, sufficiently small that we have used the weighted mean value also indicated in Table 3.

TABLE 3. *A/ρ values for the samples studied.*

Sample	A/ρ	σ_m (ppm)	σ_T (ppm)
D_1	12.0590290	0.52	1.19
D_2	12.0590257	0.65	1.25
D_3	12.0590268	1.05	1.48
weighted mean	12.0590274	0.45	0.94

The result of combining the above mean value for A/ρ and the cell volume (obtained in the NBS interferometry) is, as has been previously reported [11b, e]: $N_A = 6.0220978 \times 10^{26}$ u/kg. If the PTB result (after correction for the 0.5 ppm offset noted above) were used with the assumption of a common value for A/ρ, this N_A value would increase by 4 ppm. Pending a separate evaluation of A/ρ for the PTB crystal, there is no basis for further discussion.

4. Optically-referenced Gamma-ray and X-ray Measurements

Another area where crystals calibrated in the optical/x-ray exercise of Section 2 find application is in extension of the congruent electromagnetic scale to very short wavelengths [20]. To accomplish this extension, the calibration embodied in a single artifact crystal needs to be transferred to other specimens and species. That accomplished, these daughter crystals are useful for wavelength determinations via Bragg-Laue diffraction provided sensitive, absolute angle measurement techniques are available. Work at NBS has provided for both of these facilities, i.e., distribution and absolute goniometry. In these applications, contact with atomic reference values, e.g., R_∞, is more relevant than the metrological reference needed for N_A. Fortunately both concepts are well connected to molecular iodine stabilized He-Ne lasers.

The resulting instrumentation has been applied to a variety of problems of current interest [11d]. Among these I note that provision of new values for γ-ray reference lines for muonic atoms spectra has led to an improved test of QED. Also reference lines for pionic spectra have provided greater concordance among pion mass determinations. In the case of normal atom spectra, our work has revealed a suggestive pattern of systematic discrepancies between relativistic self-consistent field calculations and experiment. Incidental to such studies, we have participated in a new determination of the kaon mass by remeasurement on an absolute scale of the Er K absorption edge [21].

4.1 Crystal to Crystal Transfers

The particular specimen of Si crystal involved in the x-ray/optical interferometry is not of a size and shape suited to general diffraction applications. Furthermore, the large scattering power associated with the Ge crystals is of considerable value at γ-ray energies ≥ 0.2 MeV. It is evidently possible to make such transfers of calibration by direct remeasurement of γ-lines (x-ray lines being rather too broad). Unfortunately gamma-ray sources provide only low counting rates and have certain preparation and handling constraints. It is needed, therefore to have a procedure which uses the high flux available from a conventional x-ray source but which is not limited by the large widths and asymmetries of x-ray lines.

A family of suitable procedures was introduced by M. Hart and collaborators. We have made most use of a variant of these procedures shown in Fig. 5 [11d, f]. This

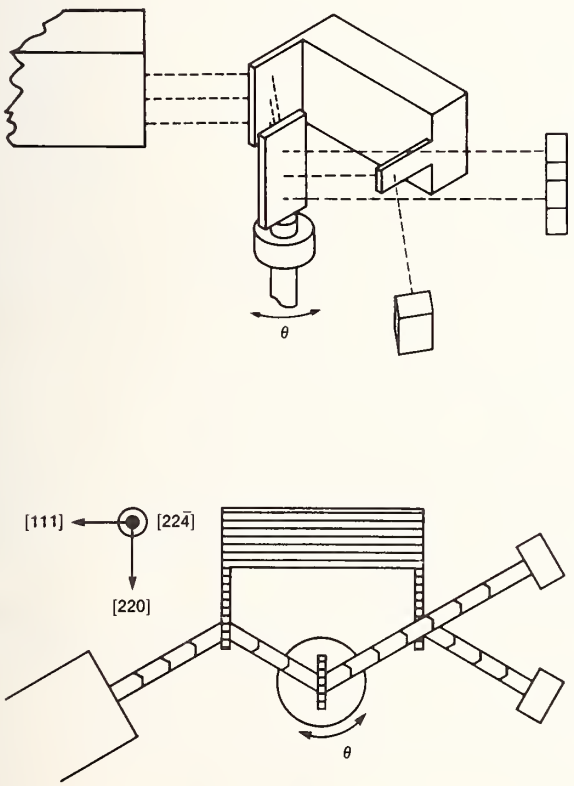


FIGURE 5. *Diagram of non-dispersive lattice comparison. Optically calibrated crystals and others needing such a reference are placed alternately in the second crystal position.*

geometry has small dispersion (zero in the case of "equal" crystals) and is thus insensitive to x-ray linewidth. The measurement cycle entails placing a calibrated and an uncalibrated specimen in the second crystal position. For all "equal" crystals, a small offset may obtain due to curvature or inhomogeneity in the long first crystal. Such an offset, however, will appear in scans with both standard and sample crystals. Thus such an offset will disappear in the difference of the differences. This last mentioned quantity is proportional to the difference in lattice parameters suggesting the almost null character of the comparison.

In the case of dealing with comparisons between different crystals, we have made use of approximate degeneracies which can often be found between (different) reflections in the distinct species. An example, used previously by Baker and Hart [22] which we have also used in an important way, is the approximate equality between the (800) spacing of Ge and the (355) spacing of Si; the difference is about 200 ppm. As long as equality or near equality can be maintained, angular measurements need sensitivity, but only limited accuracy as contrasted with what is required for the direct wavelength measurements described in the next section. Suffice it to say that we have been able to make all needed transfers to an accuracy not far from 0.1 ppm. Thus this step is not a limitation at the level of accuracy (~1 ppm) with which we are able to access invariant results.

4.2 Bragg-Laue Spectroscopy of X-rays and γ-rays

With crystals of known spacing direct wavelength measurements are possible provided sufficiently robust angle measurement schemes are at hand. Figure 6 indicates a scheme, particularly useful toward shorter

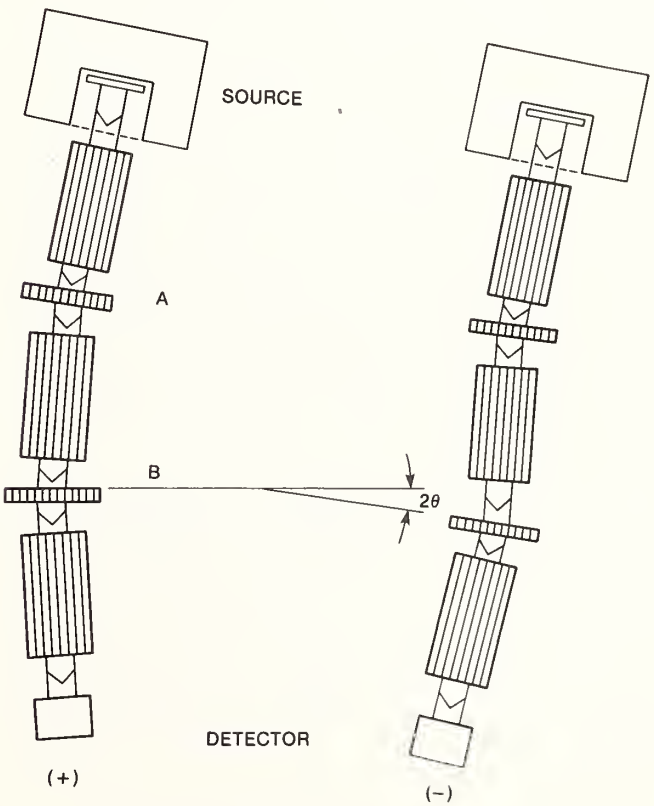


FIGURE 6. *Schematic of arrangement for transmission double crystal measurement of short wavelengths. With aid of sensitive angle interferometry calibrated from first principles, crystals with known spacings produce secondary γ-ray standards as well as accurate x-ray line positions.*

wavelengths, wherein in-plane diffraction obeys exactly the Bragg-Laue equation; finite vertical divergence gives an asymmetric contribution to the transmission function in the dispersive configuration but this is quite tractable if kept small.

Adequate angular sensitivity and dynamic range is readily available from two-beam polarization encoded angle interferometry. As is described in detail in our review [11d], one obtains a fully servoed system with noise below 0.1 milliarcsec and a range of either ±2.5° in an early version or ±15° in the latest model. Placing this precise angular scale on an absolute basis is a separate exercise. We use a sensitive null-pointing autocollimator together with a 72 or 24 sided optical polygon. With the autocollimator providing a reference direction, the angle interferometer is used to measure the intervals between all adjacent face pair normals. Since the sum of these external interfacial angles is constrained to be 2π, one obtains a calibration constant for the interferometer and, incidentally, a rather high quality polygon calibration. Resulting calibration constants vary slowly with time but by interpolating between prior and subsequent calibrations, a constant good to about 0.1 ppm is readily obtained.

Applications of such measurement technology has been in two main areas. First, as will be described next, there are a number of low energy γ-ray lines which have served either singly or in groups as reference values for muonic and pionic x-ray spectra. We have reported new values for most of these lines. A second area of application has emerged from some selective redetermination of atomic x-ray spectra themselves. Although these no longer have their traditional role as intermediates in chains to γ-rays, they appear to have regained a certain measure of intrinsic interest as will be seen below in Section 4.4.

4.3 Applications to Muonic and Pionic Spectra

Outer orbit muonic spectra particularly from medium to high Z atoms furnish a well-known testing place for QED. Shifts, for example due to vacuum polarization, are quite large so that even higher order terms become important. At one time, there appeared to be a significant pattern of discrepancies which were troublesome in that most approaches trying to account for them within the conventional framework appeared unpromising while certain exotic possibilities appeared implausible. Revision of the reference lines by the NBS γ-ray measurements combined with certain improvements in the muonic data have effectively removed the discrepancies.

Measurements of the corresponding outer orbit spectra of pionic atoms particularly toward lower values of Z are important sources of information on the pion mass. At one time, different routes to this mass using different reference lines were in some disagreement. When the various reference values were corrected to the new scale, such discrepancies are no longer significant. A related measurement of the mass of $K^{(-)}$ has now also been completed with results in agreement with alternative earlier procedures but not quite as accurate because of certain technological difficulties in the critical absorption measurement.

4.4 Application to Atomic X-ray Spectra

It is not at first glance obvious whether this venerable subject should retain interest. The earlier stage where splittings of $2p_{1/2}$ and $2p_{3/2}$ vacancy states (i.e., the $L_{II} - L_{III}$ interval) were used to reveal both nuclear size effects and Lamb-shifts [23] have now merely historical interest. Also, as noted above, x-ray lines as intermediate steps between optical and γ-ray wavelengths are no longer needed; this is a fortunate circumstance since the large widths, asymmetries, and unmodelable lineshape are severe metrological disabilities. What has happened, however, is that significant progress on the theoretical side has occurred making serious comparisons with term values and widely separated term differences (as contrasted with spin-orbit splittings) both practical and interesting provided data are adequate. This was not the case over the full range of Z which has led to significant new work which will be mentioned below.

On the theoretical side, progress both in principle and in terms of calculational rigor has been considerable especially within the past 5 years. What is involved are rather complete relativistic self-consistent field (SCF) calculations for the ground states and for atoms with inner-shell vacancies. Differences between total energies are then estimates of x-ray term energies. In some cases these terms can be approached directly from the experimental side, especially for free atoms. In an approximation whose limitations are difficult to quantify, x-ray line values, photoelectron thresholds, and work functions can be combined to give an estimate of free atom terms using data obtained from solid samples. A somewhat more robust procedure permitting use of solid samples to get estimates close to free atom properties is to focus on inner shell emission lines noting that condensed matter effects tend to perturb two inner holes equally. The loss of information entailed because only term differences appear can be compensated by systematic studies of discrepancy patterns as one vacancy is held fixed while the other changes character. Similarly, systematic trends in Z tend to be more believable than isolated points, especially if a "run" of Z includes some free atom data.

New experimental work has appeared from both the Institute Laue-Langevin (ILL) and NBS. The ILL work has concentrated on almost degenerate pairwise comparisons between high and mid-Z spectra brought into near coincidence by virtue of using different orders of diffraction. The NBS approach has aimed at obtaining a fairly uniform sampling of Z values limited only by the practicality of targets for electron bombardment and a desire to generate as few data as would suffice to reveal and confirm interesting trends.

What appears to emerge thus far is an approximately linear dependence of discrepancies with respect to Z which persists until fairly high Z values. The same general trend appears in $K\alpha_1$, $K\alpha_2$, $K\beta_1$, and $K\beta_3$ with less clear but not dissimilar results for $L\alpha_1$ and $L\alpha_2$. At high Z there seems a tendency for the K series discrepancies to become smaller. Also there appear to be cusps in the pattern near nuclear magic numbers. This investigation continues with further work on both the theoretical and the experimental sides.

Many collaborators were involved in the NBS work described above; most appear as co-authors in the cited references. In preparing this review I have turned to E. G. Kessler, Jr., and A. Henins who also carry out the work currently in progress. The cooperation of the PTB group, especially Drs. Seyfried, Becker, and Siegert, in sample exchange and discussions, is gratefully acknowledged.

References

[1] U. Bonse and M. Hart, Appl. Phys. Lett. 7, 99 (1965); Z. Phys. 188, 154 (1965). For a review, see U. Bonse and W. Graeff, in *X-Ray Optics*, H.-J. Quessar, Ed. (Springer-Verlag, Heidelberg, 1977), p. 93.

[2] *Precision Measurement and Fundamental Constants*, D. N. Langenberg and B. N. Taylor, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 343 (Aug. 1971), contains preliminary reports as follows:

a. R. D. Deslattes, p. 279.

b. I. Curtis, I. Morgan, M. Hart, and A. Milne, p. 285.

c. U. Bonse, E. teKaat, and P. Spieker, p. 291.

[3] Additional efforts are underway at the following institutions (personal names are suggested contacts):

a. Instituto Collonnetti, Torino, Italy; G. Zosi.

b. National Laboratory of Metrology, Tokyo, Japan; K. Nakayama.

c. VNIIM, Leningrad, USSR; Yu. V. Tarbeyev.

[4] M. Hart, Nucl. Instrum. Methods 172, 209 (1980).

[5] G. Materlik, in *Neutron Interferometry*, U. Bonse and H. Rauch, Eds. (Oxford Univ. Press, 1979), p. 431.

[6] U. Bonse, P. Spieker, J.-T. Hein, and G. Materlik, Nucl. Instrum. Methods 172, 223 (1980).

[7] J.-L. Staudenmann, S. A. Werner, R. Colella, and A. W. Overhauser, Phys. Rev. A 21, 1419 (1980).

[8] M. Ando, D. Bailey, and M. Hart, Acta Crystallogr. Sect. A: 34, 383 (1978).

[9] J. A. Baker, T. N. Tucker, N. E. Moyer, and R. C. Buschert, J. Appl. Phys. 39, 4365 (1968).

[10] Spacing measurements on a sample of Si at PTB is reported as follows:

a. These proceedings, contributions of P. Becker and H. Siegert; and H. Siegert and P. Becker. See also P. Becker, K. Dorenwent, G. Ebeling, R. Lauer, W. Lucas, R. Probst, H.-J. Rademacher, G. Reim, P. Seyfried, and H. Siegert, Phys. Rev. Lett. 46, 1540 (1981).
K.-J. Hanssen, G. Ade, W. Lucas, H. Siegert, and P. Becker, PTB-Berichte APL-14 (Physikalisch-Technische Bundesanstalt, Marz, 1981).
Overviews of the PTB effort on N_A are given at this Conference and previously.

b. P. Seyfried, these proceedings.

c. P. Becker, in *Neutron Interferometry*, U. Bonse and H. Rauch, Eds. (Oxford University Press, 1979), p. 416.

[11] Results of progressively more refined lattice spacing measurements have appeared in reports on x-ray wavelengths (a) and on N_A (b). The result was applied to gamma-ray reference lines in (c).

a. R. D. Deslattes and A. Henins, Phys. Rev. Lett. 31, 972 (1973).

b. R. D. Deslattes, A. Henins, H. A. Bowman, R. M. Schoonover, C. L. Carroll, I. L. Barnes, L. A. Machlan, L. J. Moore, and W. R. Shields, Phys. Rev. Lett. 33, 436 (1974); cf. Phys. Rev. Lett. 36, 898 (1976).

c. E. G. Kessler, Jr., R. D. Deslattes, A. Henins, and W. C. Sauder, Phys. Rev. Lett. 40, 171 (1978).

Reviews of this work have appeared focussing on x-ray and gamma-ray energies (d), on N_A (e), and on the overall program (f). Several earlier reviews were presented (g).

d. R. D. Deslattes, E. G. Kessler, W. C. Sanders, and A. Henins, Ann. Phys. (N.Y.) 129, 378 (1980).

e. R. D. Deslattes, in *Ann. Rev. Phys. Chem.*, Vol. 31, B. S. Rabinovitch, J. M. Schurr, and H. L. Strauss, Eds. (Ann. Rev. Inc., Palo Alto), p. 435.

f. R. Deslattes, in *Metrology and Fundamental Constants*, Proceedings of Course LXVIII, Summer School of Physics—Enrico Fermi, 1976, A. Ferro Milone, P. Giacomo, and S. Leschiutta, Eds. (North Holland, Amsterdam, 1980), p. 38.

g. E. G. Kessler, Jr., R. D. Deslattes, W. C. Sauder, and A. Henins, in *Neutron Capture Gamma Ray Spectroscopy*, R. E. Chrien and W. R. Kane, Eds. (Plenum Press, New York, 1979), p. 427.
R. D. Deslattes and E. G. Kessler, Jr., in *Atomic Masses and Fundamental Constants—6*, J. A. Nolen and W. Benenson, Eds. (Plenum Press, New York, 1980), p. 203.
R. D. Deslattes, Jpn. J. Appl. Phys. Suppl. 17-2, 1 (1978).
R. D. Deslattes, in *Atomic Masses and Fundamental Constants—5*, J. H. Sanders and A. H. Wapstra, Eds. (Plenum Press, New York, 1976), p. 552.

[12] Such a comparison was attempted by A. Henins at NBS shortly after this Conference. The PTB sample was kindly furnished by our colleagues from that institution.

[13] a. A. Peuto and A. Sacconi, these proceedings.

b. G. A. Bell and J. B. Patterson, these proceedings.

[14] J. B. Saunders, J. Res. Natl. Bur. Stand. Sect. C: 76, 11 (1972).

[15] D. P. Johnson, J. Res. Natl. Bur. Stand. Sect. A: 78, 41 (1974).

[16] For a discussion, see Refs. [11e and f].

[17] H. A. Bowman, R. M. Schoonover, and C. L. Carroll, J. Res. Natl. Bur. Stand. Sect. A: 78, 13 (1974).

[18] I. L. Barnes, L. J. Moore, L. A. Machlan, T. J. Murphy, and W. R. Shields, J. Res. Natl. Bur. Stand. 79A, 727 (1975).

[19] Standard Reference Materials are available through the Office of Standard Reference Materials, Room B311, Building 222, National Bureau of Standards, Washington, D.C. 20234.

[20] E. G. Kessler, Jr., L. Jacobs, W. Schwitz, and R. D. Deslattes, Nucl. Instrum. Methods 160, 435 (1979). See also Refs. [11c and g].

[21] G. K. Lum, C. E. Wiegand, E. G. Kessler, Jr., R. D. Deslattes, L. Jacobs, W. Schwitz, and R. Seki, Phys. Rev. D 23, 2522 (1981).

[22] J. F. C. Baker and M. Hart, Acta Crystallogr. Sect. A: 31, 364 (1975).

[23] R. L. Shacklett, Rev. Mod. Phys. 30, 521 (1958).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Work Related to the Determination of the Avogadro Constant in the PTB

P. Seyfried

Physikalisch-Technische Bundesanstalt, D 3300 Braunschweig, F.R.G.

The d_{220}-lattice spacing in a highly pure nearly perfect silicon crystal was measured and found to be d_{220} = 192 015.560 fm with an uncertainty of $\pm 6 \times 10^{-8} d_{220}$ in vacuum at 22.50 °C. Thus for this crystal the first of the three quantities, d_{220}, density ρ, and molar mass M, from which the Avogadro constant N_A can be derived, is known with sufficient accuracy. The d_{220}-value given is $-1.8 \times 10^{-6} d_{220}$ smaller than that reported by Deslattes *et al.* and used in his determination of N_A. Carbon and oxygen impurities in our crystal cannot explain this large difference. A possible change in N_A of $+5.4 \times 10^{-6}$ of its value should be taken into account.

Density standards in the form of cubes made of a ceramic material with trade name Zerodur have been prepared. For two (1 kg) standards an uncertainty of $\Delta\rho/\rho = \pm 7 \times 10^{-7}$ has been obtained. These standards will be used for density measurements of silicon samples found to be uniform ($\Delta d/d < 4 \times 10^{-8}$) by both high resolution double crystal and Moiré-topography.

Key words: Avogadro constant; density standards; lattice spacing; silicon.

1. Introduction

Stimulated by proposals of Bonse and Hart [1] a project was started in the PTB for the precision measurement of the d_{220}-lattice spacing in a silicon crystal. The work gained remarkable impetus by the publications of the extremely successful measurements of Deslattes and coworkers [2, 3] at the NBS. Recently, in an extended series of measurements, the lattice spacing d_{220} of a highly pure nearly perfect silicon crystal was determined with a relative uncertainty of 6×10^{-8} [4]. Thus, the first of the three quantities necessary to derive the Avogadro constant, namely d_{220}, the density ρ, and the molar mass M of a silicon monocrystal, is known with sufficient accuracy.

In a second project density standards are being developed in order to improve the accuracy of density measurements. On the way to a final uncertainty of less than $\pm 5 \times 10^{-7}$, a series of experimental devices has been investigated, and most recently two standards have been prepared, which will now be used for density measurements of silicon samples.

Close cooperation with external laboratories for mass spectrometry is necessary for the determination of molar masses.

2. Density Determination

The PTB-approach to density standards differs from that of Bowman *et al.* [5] at NBS in that we use cubes of a ceramic material (trade name Zerodur), instead of steel balls. Zerodur was chosen because it is a dielectric material (rather than a conductor), is insensitive to chemical effects, has a very low expansion coefficient ($\alpha < 0.05 \times 10^{-6}$ K^{-1}), and can be easily machined. Cubes were preferred instead of balls, as then the highly developed experimental methods and instruments used for measuring gauge blocks can be applied. Taking into account the known errors, the final uncertainty is estimated to be 6×10^{-7} ρ for a 1 kg (a = 73 mm) standard. Corrections due to deviations from the ideal shape are negligible if the following limits are not exceeded: unevenness $\epsilon < \lambda/10$, non-parallism $\pi < 2 \times 0.1''$, deviations from the right angle $\rho < 10''$. The missing volume should not exceed 2×10^{-7} of the volume.

Twelve rectangular blocks of mass M (called quaders) (a = 24 mm, b = 25 mm, c = 33 mm), six sharp cornered and six with bevels were made and investigated. After initial weighing, the deviations from ideal shape were investigated and found to be well within the limits given above (see Fig. 1 and Fig. 2). The distances at the centers of opposite faces a, b, c were determined both interferometrically and by means of a displacement comparator with mechanical contacting [6], and surfaces and edges were inspected by optical microscopy and scanning electron microscopy in order to estimate the missing volume. Additional weighing after each step ensured that no mass was lost during the process.

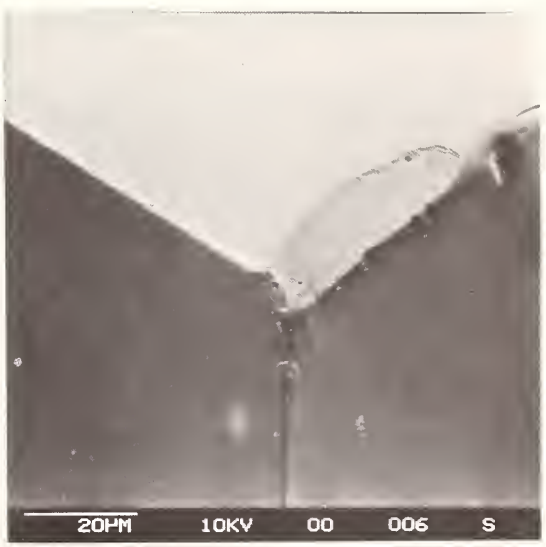


FIGURE 1. *Missing volume near a corner of a density standard estimated to be less than 2×10^{-6} mm^3, $\Delta V/V \approx 10^{-10}$; photograph taken from a scanning electron microscope.*

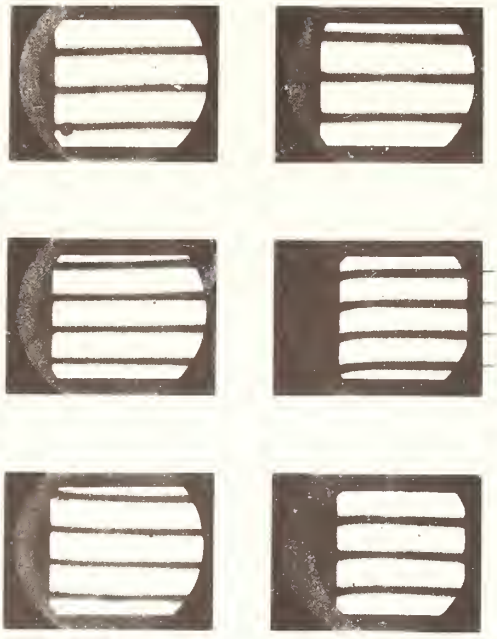


FIGURE 2. *Interference of equal thickness pattern obtained for the six faces of quader No. 8, unevenness below λ/13.*

Results obtained for two sharp cornered devices and two with bevels are summarized in Table 1. The indices m and a refer to interferometrical and mechanical measurements respectively, which differ by about 20 nm. The missing volume amounts only to $\Delta V = 2.5 \times 10^{-7} V$ and is not a restricting factor as was assumed at the beginning of the work. The following uncertainties have to be assigned to the various quantities:

$$s(m) = \pm 2 \times 10^{-7}\, m$$
$$s(a_m) = s(b_m) = s(c_m) = \pm 15 \text{ nm}$$
$$s(a_a) = s(b_a) = s(c_a) = \pm 20 \text{ nm}$$
$$s(\Delta V)_f = \pm 2.5 \times 10^{-6}\, V$$
$$s(\Delta V)_s = \pm 2.5 \times 10^{-7}\, V$$
$$s(V_m) = \pm 1.0 \times 10^{-6}\, V_m$$
$$s(V_a) = \pm 1.3 \times 10^{-6}\, V_a$$

$$s(\rho)_f = \pm 4 \times 10^{-6}\, \rho \qquad s(\rho)_s = \pm 2 \times 10^{-6}\, \rho\ .$$

The indices s and f refer to sharp cornered quaders and those with bevels, respectively.

As a consequence of these results, two sharp cornered cubes (a = 73 mm, m = 1 kg) were made and the density was determined with a final uncertainty of $s(\rho) = \pm 7 \times 10^{-7}\rho$.

Several samples will be prepared from sufficiently uniform silicon (examined by Lang-topography), with one from material of which the lattice parameter is accurately known. Their density will be determined in a collaboration between BIPM and PTB [7].

Separate investigations have begun to determine the density from the mass and the geometrical dimensions of a regularly shaped piece of silicon. A high quality cylinder (400 g) is being prepared and a hexagonal prism is under consideration. Missing volume and density changes in machined and polished faces are expected to limit the accuracy.

All of this work is timed to achieve an uncertainty in determining density of less than 5×10^{-7} by 1983.

TABLE 1. *Results of measurements on four rectangular blocks of Zerodur; indices m and a refer to interferometrical and mechanical measurements, respectively.*

Quantity		With Bevel		Sharp Cornered	
		No. 1	No. 3	No. 8	No. 10
m	[g]	50.220 669	50.221 167	50.222 750	50.222 609
a_m	[mm]	24.031 411	24.031 552	24.031 484	24.031 421
a_a	[mm]	24.031 413	24.031 565	24.031 509	
b_m	[mm]	24.927 630	24.927 642	24.927 114	24.927 138
b_a	[mm]	24.927 595	24.927 651	24.927 118	
c_m	[mm]	33.083 707	33.083 709	33.083 597	33.083 560
c_a	[mm]	33.083 737	33.083 726	33.083 613	
V_m	[mm^3]	19 818.666	19 818.793	19 818.250	19 818.195
V_a	[mm^3]	19 818.658	19 818.821	19 818.284	
$\bar{V}$	[mm^3]	19 818.662	19 818.807	19 818.267	19 818.195
ΔV	[mm^3]	−0.719	−0.610	−0.005	−0.005
V	[mm^3]	19 817.943	19 818.197	19 818.262	19 818.190
ρ	[g/cm^3]	2.534 101	2.534 094	2.534 167	2.534 167

3. Lattice Spacing in a Silicon Crystal

The d_{220}-lattice spacing in a highly pure, almost perfect silicon crystal, grown and purified by multiple float zone melting by Wacker-Chemitronik, has been measured using a scanning x-ray interferometer according to Bonse and Hart [1], combined with an optical polarization interferometer described by Curtis *et al.* [8]. The principle of the device is shown in Fig. 3. During a translation of the analyzer crystal A the signals shown in Fig. 4 are observed. A displacement $s = m\lambda/2$ is enclosed by two zero crossings of the optical signal. In this displacement, N full periods and a fraction f of the x-ray signal are observed; f results from the phase $\phi(m)$ and $\phi(0)$ of the x-ray signal at the limits of the displacement.

The n-fold ($n = N + f$) of d corresponds to a displacement $s = m\lambda/2$.

$$m\lambda/2 = nd\ ,$$
$$d = (n/m)^{-1}\lambda/2\ .$$

Here n/m is the quantity to be measured. A first approximation is obtained from a measurement over $\lambda/2 (m = 1)$. If m is increased in small steps N may be predetermined and the measured fraction f gives an improved value for n/m. It is then sufficient to measure f alone. In successive steps (e.g., m = 1, 2, 4, 8 . . . 128) the uncertainty is reduced until N may be absolutely predetermined for m = 120. Details are given in the contribution of P. Becker, these proceedings; see also Refs. [9, 10].

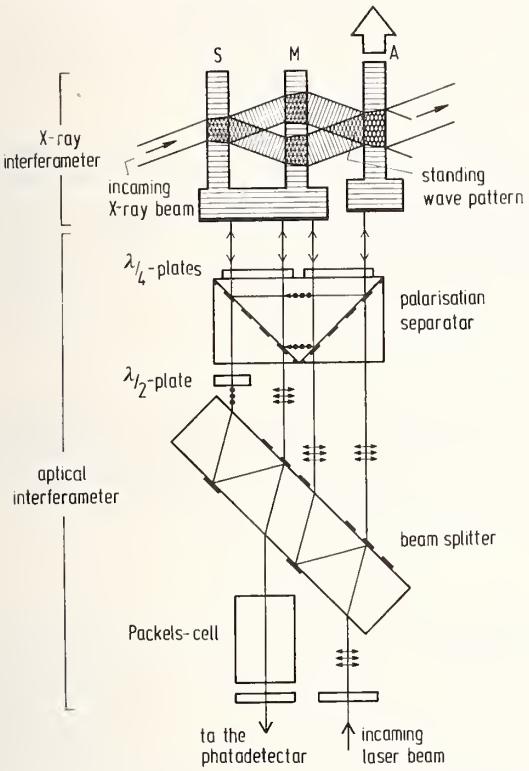


FIGURE 3. *Principle of the scanning x-ray interferometer of the PTB: LLL-x-ray interferometer according to Bonse and Hart [1] combined with an optical interferometer according to Curtis et al.* [8].

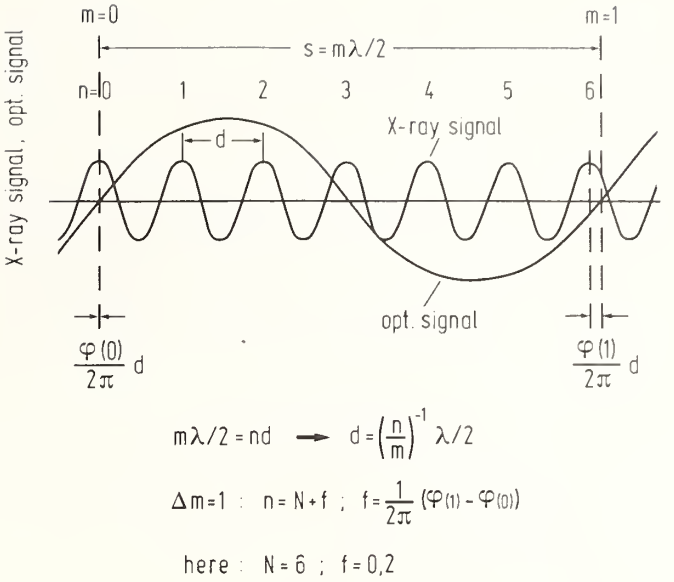


FIGURE 4. *Signals obtained in the scanning x-ray interferometer of Fig. 3 (schematically); modulation of the x-ray signal shown, average intensity subtracted.*

In a series of measurements in vacuum at temperatures between 22.42 °C and 22.50 °C a mean value $\overline{n/m}$ = 1648.281 626 (reduced to 22.50 °C) with a standard deviation $\sigma = \pm 5.8 \times 10^{-8}\ \overline{n/m}$ and a standard deviation of the mean $\sigma_m = \pm 4.3 \times 10^{-9}\ \overline{n/m}$ was obtained. Measuring results, corrections, and uncertainties are summarized in Table 2.

TABLE 2. *Summarized measuring results, corrections, and uncertainties of the d_{220}-lattice spacing measurements.*

Measuring results	
Number of measurements	170
Number of x-ray fringes in half the wavelength	n/m = 1648.281 626
Standard deviation	σ = 0.000 095
Standard deviation of the mean	σ_m = 0.000 007
Resulting lattice spacing	d_{220} = 192 015.560 fm
T = 22.5 °C. vacuum ($P \leq 10^{-6}$ bar.	λ = (632 991 415 ± 5) fm
Corrections and uncertainties	
Fresnel correction	$k_1 = (-3.8 \pm 0.2) \times 10^{-8}\ n/m$
Cosine-error	$k_2 = (-0.1 \pm 0.2) \times 10^{-8}\ n/m$
Abbe-error	$k_3 = (\ 0\ \ \pm 3.0) \times 10^{-8}\ n/m$
Temperature	$k_4 = (\ 0\ \ \pm 5.1) \times 10^{-8}\ n/m$
Total	$k = (-3.9 \pm 5.9) \times 10^{-8}\ n/m$
Uncertainty of wavelength	$\Delta\lambda = \pm 0.8 \times 10^{-8}\ \lambda$
Uncertainty of the lattice spacing	$\Delta d = \pm 6 \times 10^{-8}\ d$

The total correction amounts to less than the standard deviation for a single measurement. The largest contribution is given by the correction for the nonplanarity of the wavefronts of the laser beam (Fresnel phase shift) and is calculated from a beam radius in the waist $w_o = (0.521 \pm 0.016)$ mm. Only a very small cosine-correction is necessary as the normal to the mirrors, the direction of movement, and the direction of the laser beam are parallel to the normal to the lattice planes within a few seconds of arc, due to elaborate adjustment procedures. The largest uncertainty is given by a ±0.02 K uncertainty of the crystal-temperature, carefully calibrated thermistors being used. Details are given in the contribution of Siegert, these proceedings.

4. Discussion

The d_{220}-value reported here gives for the first time the opportunity to compare two lattice spacings of silicon crystals measured by combined optical and x-ray interferometry. This is done below:

$$
\begin{aligned}
d_{220}^{\mathrm{PTB}} &= 0.192\,016\,724\ \mathrm{nm} \pm 6 \times 10^{-8}\, d_{220} \\
-d_{220}^{\mathrm{NBS}} &= -0.192\,017\,07\ \mathrm{nm} \pm 10 \times 10^{-8}\, d_{220} \\
&= -0.000\,000\,346\ \mathrm{nm} \quad (\text{at } 25\ ^\circ\mathrm{C} \\
&= -1.8 \times 10^{-6}\, d_{220} \quad \text{and 1 bar})
\end{aligned}
$$

The difference exceeds the stated uncertainties by almost a factor of 20. There should of course be a difference in the lattice spacings of the two crystals. However, a difference as large as this has only been reported in the comparison measurements of Ando, Bailey, and Hart [11] for an "exotic" sample of silicon. The lattice spacings of all other samples of float zone silicon of different suppliers all over the world differed from the very pure PERFX by no more than $-6 \times 10^{-7} d$. This difference was assigned to residual carbon impurities.

PERFX samples are reported to differ by less than 10^{-7} [12]. A Lopex crystal was compared by Deslattes [13] to a PERFX crystal. Due to carbon and oxygen impurities of $n_C = 3.9 \times 10^{16}\ \mathrm{cm}^{-3}$ and $n_O = 6 \times 10^{16}\ \mathrm{cm}^{-3}$, respectively, a change in the lattice spacing $\Delta d/d = 4 \times 10^{-8}$ was predicted using $(\Delta d/d)_C = -6.5 \times 10^{-24}\, n_C$ and $(\Delta d/d)_O = +3.5 \times 10^{-8}\, n_O$. The difference was measured to be $(5 \pm 2) \times 10^{-8}$ and differs only by $(1 \pm 2) \times 10^{-8}$ from the predicted value.

From this we conclude that impurities of C and O are the main sources of lattice parameter variation. Since our Waso-silicon is very pure ($n_C = n_O < 1 \times 10^{16}\ \mathrm{cm}^{-3}$) and uniform ($\Delta d/d < 4 \times 10^{-8}$, measured using Moiré-topography and high resolution double crystal topography [14]), our material is expected to be closer to PERFX than to any other material used in the comparison of Ando, Bailey, and Hart.

If a crystal-to-crystal comparison confirms our assumption, there will be a considerable discrepancy indicating a systematic error in one of the measurements or both. On the basis of our d_{220}-lattice spacing, leaving fixed all other data as given by Deslattes *et al.* [15], the Avogadro constant would have to be changed by $+5.4 \times 10^{-6}$ of its value. Clearly, there is an urgent need for additional work.

I am indebted to Professor U. Bonse, University of Dortmund for scientific stimulation and critical discussions. I also wish to thank Dr. M. Kochsiek and other PTB colleagues for their fruitful cooperation.

References

[1] U. Bonse and M. Hart, Z. Phys. 188, 154 (1965).
[2] R. D. Deslattes and A. Henins, Phys. Rev. Lett. 31, 972 (1973).
[3] R. D. Deslattes, A. Henins, H. A. Bowman, R. M. Schoonover, C. L. Carroll, I. L. Barnes, L. A. Machlan, L. J. Moore, and W. R. Shields, Phys. Rev. Lett. 33, 463 (1974).
[4] P. Becker, K. Dorenwendt, G. Ebeling, R. Lauer, W. Lucas, R. Probst, H.-J. Rademacher, G. Reim, P. Seyfried, and H. Siegert, Phys. Rev. Lett. 46, 1540 (1981).
[5] H. A. Bowman, R. M. Schoonover, and C. L. Carroll, Metrologia 10, 117 (1974).
[6] Chr. Hoffrogged and R. Mann, Messtechnik 81, 1 (1973).
[7] Arrangement of M. Kochsiek (PTB) and G. Girard (BIPM).
[8] I. Curtis, I. Morgan, M. Hart, and A. D. Milne, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.) Spec. Publ. 343 (Aug. 1971), p. 285.
[9] H.-J. Rademacher, P. Becker, H. Siegert, W. Lucas, K. Dorenwendt, G. Reim, K.-J. Hanszen, R. Lauer, R. Probst, and P. Seyfried, PTB-Bericht APh-13 (Physikalisch-Technische Bundesanstalt, Braunschweig, July 1980).
[10] K.-J. Hanszen, G. Ade, W. Lucas, H. Siegert, and P. Becker, PTB-Bericht, APh-14 (Physikalisch-Technische Bundesanstalt, Braunschweig, March 1981).
[11] M. Ando, D. Bailey, and M. Hart, Acta Crystallogr. Sect. A: 34, 484 (1978).
[12] J. A. Baker, T. N. Tucker, N. E. Moyer, and R. C. Buschert, J. Appl. Phys. 33, 4365 (1968).
[13] R. D. Deslattes, private communication.
[14] We are indebted to Prof. Dr. U. Bonse, University of Dortmund, for these measurements.
[15] R. D. Deslattes, E. G. Kessler, W. C. Sauder, and A. Henins, Ann. Phys. (N.Y.) 129, 378 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Absolute Determination of the (220)-Lattice Spacing in Silicon

P. Becker and H. Siegert

Physikalisch-Technische Bundesanstalt, D 3300 Braunschweig, F.R.G.

The (220)-lattice plane spacing in a silicon crystal was measured using a combined x-ray and optical interferometer. The experimental set-up and important characteristic features of the interferometer crystals and the translation device are described in detail. The results of several individual measurements are discussed to explain the evaluation method applied. The results of 170 measurements are characterized by a standard deviation $1\sigma = 6 \times 10^{-8}$.

Key words: lattice parameter; silicon; two-beam polarization optical interferometer; two-crystal x-ray interferometer.

1. Introduction

In order to determine the lattice parameter of a silicon crystal by x-ray interferometry [1], the Physikalisch-Technische Bundesanstalt[1] adopted a course relying on the proposals of Bonse, te Kaat, and Spieker [2]. The experimental set-up differs significantly from that of the National Bureau of Standards, where Deslattes first succeeded in measuring the (220)-lattice plane spacing of a silicon crystal with high accuracy, see e.g., Ref. [3]. The basic requirement for performing these measurements is the moving of one part of a Laue-type x-ray interferometer with respect to the fixed crystal part in a direction normal to the reflecting planes.

2. The X-ray Interferometer

The two crystals of the x-ray interferometer were cut from a piece of silicon material (WASO), grown and purified by multiple float zone melting, manufactured by Wacker-Chemitronic. A drawing of the interferometer used in our experiment has been given previously, see e.g., Ref. [4]. It has the following special features: The front faces of the crystals are polished to form optical mirrors. By this provision, the mirrors used for the optical measurement of the displacement are a part of the crystal lattice itself. Both interferometer crystals, each with three steel balls attached to its base, rest on supports without being fixed to them, as shown in Fig. 1. Thus the crystals may be easily exchanged, and a direct comparison of crystal materials of different origin is facilitated.

Alignment of the lattice planes of the two crystals, parallel within 1×10^{-9} rad, is achieved by means of the adjusting element shown in the figure. Slight rotations about the ρ axis (axis of a rotational moire pattern) and about the ϑ axis (axis about the Bragg angle alignment) can be effected by two elastic pivots. Deflections of the leaf springs are made with the eccentric shown in the figure (rough adjustment), or with piezoelectric drives (fine adjustment). More details of the interferometer crystals and their alignment are described in Ref. [5], Sects. A and D.

[1]Interlaboratory project of the Physikalisch-Technische Bundesanstalt: X-ray interferometry: H. Siegert, P. Becker; Length measuring technique: K. Dorenwendt; Mechanics: H.-J. Rademacher; Electronics: P. Seyfried, G. Ebeling, W. Lucas.

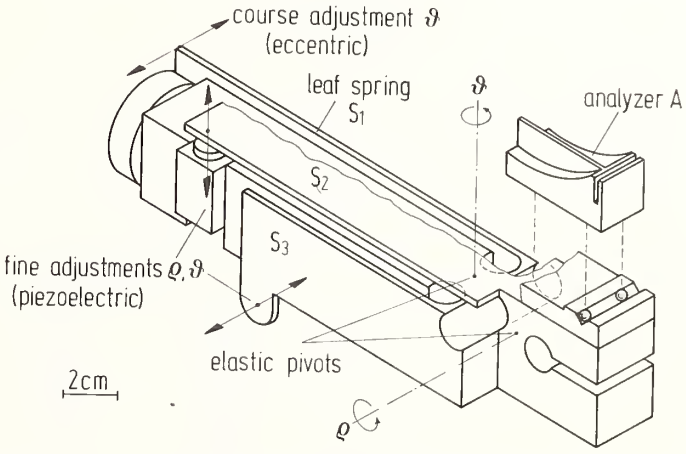


FIGURE 1. *Springs for adjustment of the analyzer crystal. A deflection of the leaf springs by ± 1 mm effects a turn of the crystal by* 5×10^{-5} *rad.*

3. The Translation Device

During the measurement, the analyzer crystal A must be moved in the direction normal to the lattice planes with a velocity of 10^{-2} nm/s to 10^{2} nm/s. Tilting angles of only 10^{-8} rad could be tolerated. For this purpose, translation stages with guiding errors of less than 10^{-8} rad are necessary. A translation stage fulfilling this condition and used in these experiments is shown in Fig. 2. It is in the form of a double parallel spring Ref. [5-E] manufactured from one plate of high-quality steel. Each of the four springs consists of two bars. One of them is connected to the external fixed frame, the other one to the movable central bar supporting the analyzer crystal A and the adjusting element. The frame rests on four supports, two of which are fixed to a steel plate of very high stiffness; the other two are equipped with a slide bearing. By this means, small deformations of the basic steel plate during movements of the central bar are avoided. The displacement of the central bar is achieved by a piezoelectric drive as indicated in the figure.

Guiding errors of the translation stage are caused by inexact mounting and inhomogeneities of the material. They are avoided by proper adjustment of the connecting rod with respect to the drive lever and finally, by means of the magnetic adjusting elements shown in Fig. 3. In

order to reduce or increase the elastic force of the mechanical springs during displacement, an additional magnetic force ΔF of a permanent magnet acts upon the springs of the translation device as indicated in the upper part of the figure. In order to eliminate the undesired

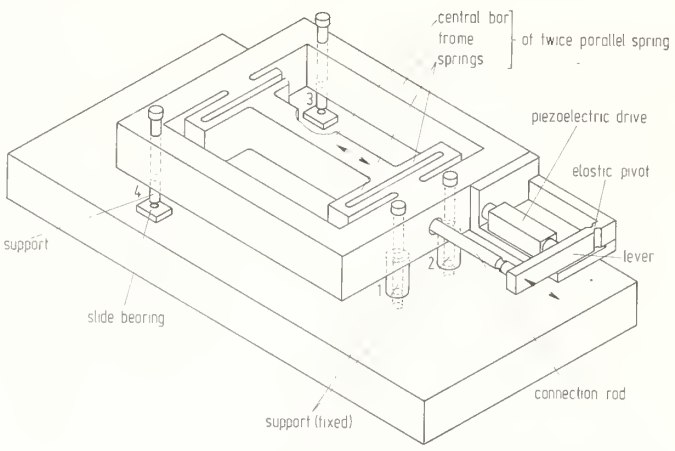


FIGURE 2. *Translation stage and drive. The piezoelectric drive acts upon a lever, which is directly coupled to the central bar of the double parallel spring. The whole drive is fastened to the frame.*

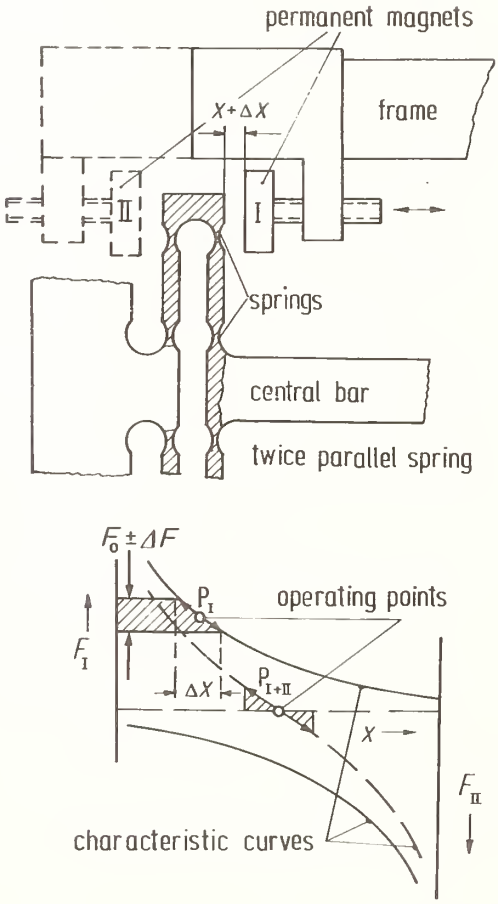


FIGURE 3. *Magnetic compensation of guiding errors. To begin with, the permanent magnet I is adjusted at a distance x to the spring of the translation stage and acts with a magnetic force F_o (see operating point P_I in the lower part of the figure). During measurements, x and F_0 are changed by Δx and ΔF as a function of deflection of the central bar.*

"static" force F_o of the permanent magnet, two identical magnetic elements I and II are used. In this way, the operating point P_1 on the characteristic curve of a single permanent magnet changes over to the point P_{1+11} on the curve corresponding to a pair of magnets, as shown schematically in the lower part of the figure.

4. The Optical Interferometer

The translation of the analyzer crystal is measured by the method of two-beam interferometry, using an optical polarization interferometer according to Curtis *et al.* [6]. The principles of the arrangement have been described and illustrated in Refs. [4, 5-J, 7, 8]. A Lamb-dip stabilized He-Ne laser is used as a light source. During measurements, the frequency of the laser light is repeatedly compared with that of a $^{127}I_2$ stabilized He-Ne laser. The waist of the laser beam is about 1 mm in diameter, therefore only a very small correction due to the nonplanarity of the wavefronts [9] (referred to as "Fresnel phase shift") is necessary. Optical feedback is avoided by the polarization optical components of the interferometer itself and, additionally, by a light trap of an attenuation of more than 30 decibels (for details see Ref. [5-H]).

5. Apparatus

The instrument described so far is arranged on a granite table of about 500 kg in weight. Measures for attenuating mechanical vibrations and stabilizing the temperature to better than 10^{-4} K/h in the vicinity of the interferometers have already been discussed in Refs. [4, 5-B, 5-F]. A schematic diagram of the apparatus is given in Ref. [10]. If the translation stage does not move linearly, and if the loci, where the translation is measured by optical and x-ray means are offset (see e.g., Fig. 1 in Ref. [11]), both interferometers will measure different displacements, namely, s_{opt} and s_{x-ray}. By measuring the relative phase of the optical and x-ray interferometer signals, the point of impact of the laser beam onto the crystal mirrors is optimized until the condition $s_{opt} = s_{x-ray}$ is fulfilled. Both interferometers and the translation device are mounted on the steel plate mentioned and inside a vacuum chamber with an aluminium hood. The temperature in the vicinity of the silicon crystals is determined by several thermocouples and NTC resistance thermometers which have been repeatedly calibrated with reference to a platinum resistance thermometer.

6. Measuring Procedure

The principle of the arrangement is given in Seyfried's contribution to these proceedings and this can be seen in Fig. 1 in Ref. [11]. During the translation of the analyzer crystal, periodic intensity variations indicate the lattice planes moving through the stationary x-ray wave pattern, as can be seen in Ref. [12]. The intensity variations of the laser light which occur simultaneously indicate the number of half of wavelengths λ of the laser radiation contained in the displacement.

A displacement of an integer number m of $\lambda/2$ corresponds to n x-ray fringes. Thus the lattice plane spacing d can be calculated according to

$$d = (n/m)^{-1}\lambda/2 \quad (1)$$

with

$$n = N_m + f_m . \quad (2)$$

Here, N_m is the integer part of n, while f_m is the residual fraction. By counting each x-ray fringe over pathlengths of $\lambda/2$, i.e., $m = 1$, we obtain the integer part of n/m to be 1648 and a first approximation of the fraction of the ratio n/m.

When the integer number has been fixed, it is—according to Deslattes' measurements—only necessary to measure the fringe fraction f_m between increasing optical intervals. Thus the accuracy of the lattice spacing measurement increases with larger displacements. The experimental data, obtained from each d-measurement enables us to determine the number of lattice planes contained in the whole translation length with an uncertainty of one thousandth of an x-ray fringe.

The last stage of our step-wise progression is shown in Fig. 4. In the preceding steps, the expected value $(n/m)_e$ was ascertained to be 1648.282. The output signals of the x-ray interferometer (upper part of the figure) and the optical interferometer near its zero level (lower part) are recorded simultaneously as a function of time. The record represents only the backward part of f_m determinations over a total displacement of 120 optical orders. We shall discuss here only the left-hand part of the drawing. In the interval I, the fraction f_{120} is determined by moving the analyzer crystal very smoothly, and simultaneously evaluating several x-ray fringes. After a fast translation of the crystal over a range of about 50,000 x-ray periods (interval II) and after a brief period of mechanical relaxation (interval III), the fractions f_{90}; f_{60} . . etc. are measured in the same way until the zeroth optical order is reached.

In order to determine the ratio n/m, it is advantageous to calculate the differences $\Delta n/m$ between the mean value (n/m) measured during a single run forward and backward, and the expected value $(n/m)_e$ according to

$$\Delta n = [n/m - (n/m)_e]m . \quad (3)$$

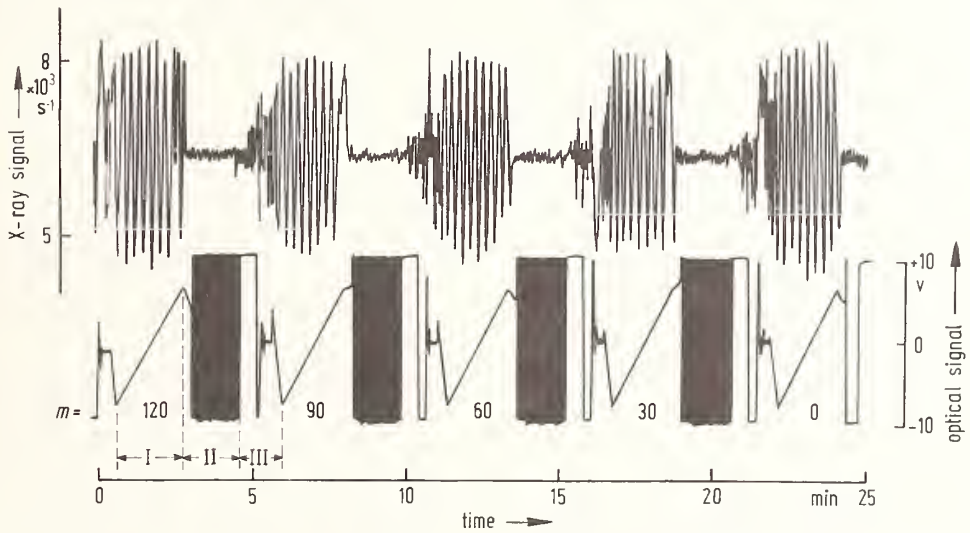


FIGURE 4. *Output signals of the combined x-ray and optical interferometer. The output signal of the optical interferometer is zero when the phase difference between the two beams is an integral number of half optical wavelengths.*

7. Results

In Fig. 5 the reduced values Δn of x-ray fringes are plotted versus the optical periods m. Five measurements with forward and backward translations are shown. They give an illustration of the reproducibility obtained. One hundred and seventy measurements have been carried out in a vacuum at temperatures between 22.42 °C and 22.50 °C. The maximum deviations of the measured values Δn, the crystal temperature and the individual ratio n/m, referred to a common temperature of 22.50 °C, have also been marked. In the upper part of the figure, for example the fractions $\Delta n(m)$ with m = 30, 60, 90, 120, 120, 90, 60, 30, 0 have been measured in comparison with the zeroth order. Continuing this way, forward and backward for each run, an almost straight line results.

Only in exceptional cases do the confidence limits of the measured fractions deviate from the plotted straight lines. From its slope, the ratio n/m is determined according to the equation given at the ordinate. In all 170 measurements, the deviation from each other of both branches—each for forward and backward translations—is smaller than six hundredths of one x-ray fringe, and the deviations of each measured value $\Delta n(m)$ are smaller than four hundredths of one x-ray fringe. The different mean slopes of all measured curves actually limit the reproducibility. More details with regard to the determination of the relative phase of the x-ray and optical signal have been previously reported [4, 10, 13].

A histogram of 170 measurements is shown in Fig. 6. A mean value of $\overline{(n/m)}$ = 1648.281 626 with a standard deviation of a single measurement $\sigma = 5.8 \times 10^{-8}$ and a standard deviation of the mean $\sigma_m = 4.4 \times 10^{-9}$ is calculated from the results displayed. The dotted line represents a least squares fit of a Gaussian curve. When the data represented by the histogram are divided into 17 groups according to the observation time, there is no significant change in the ratio $\overline{n/m}$ (as can be seen in Fig. 4 in Ref. [11]).

The systematic contributions in calculating the lattice spacing d together with their uncertainties are given in Siegert's contribution to these proceedings.

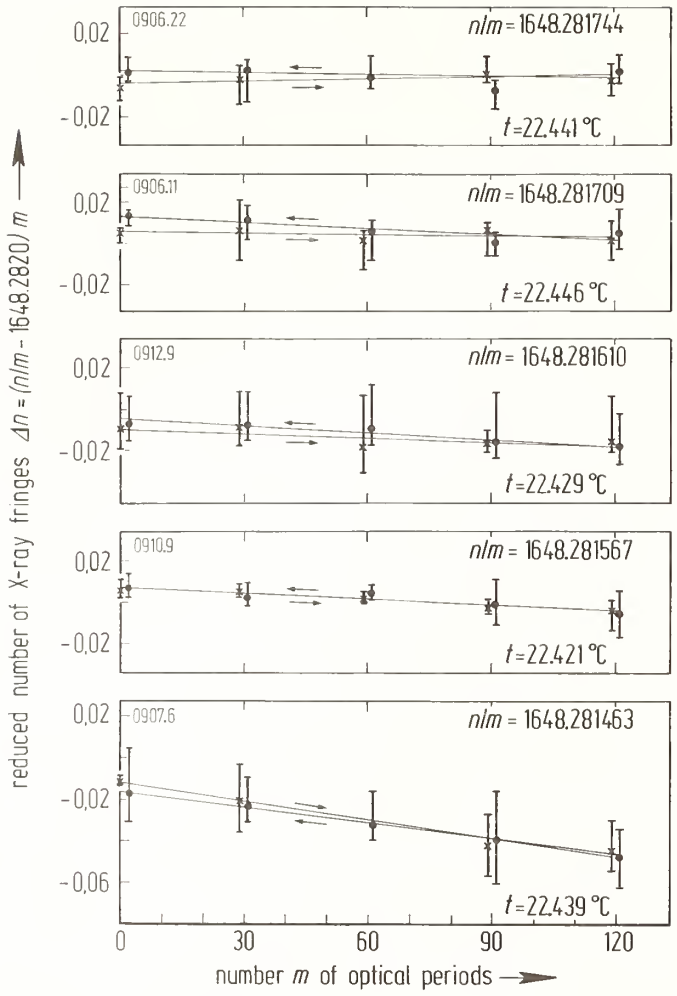


FIGURE 5. *Determination of the ratio n/m. Total displacement 120 optical orders.*

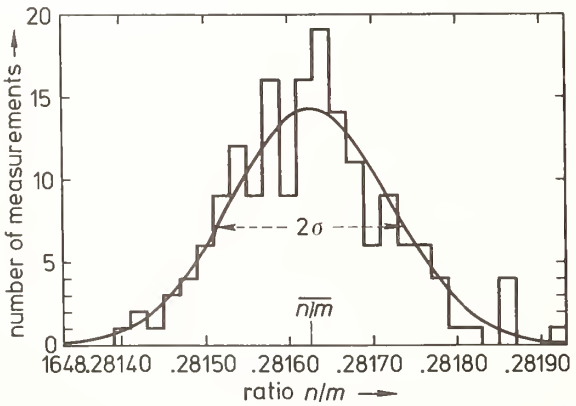


FIGURE 6. *Distribution of the ratio n/m of 170 measurements.*

Afterwards, by varying the locus onto the analyzer crystal A, where x-ray-interferences occur, the given value $\overline{(n/m)}$ was verified in more than 300 measurements. From these results it can be concluded that the standard deviation of a single measurement can be reduced by a factor of 3 or more. Nevertheless, with respect to the new determination of Avogadro's number and to the results given by Deslattes *et al.*, there is an urgent need to progress from the crystal under investigation to other Si-crystal materials and to compare crystal materials from different sources.

The authors wish to thank Professor U. Bonse, University of Dortmund, and Professor K. J. Hanszen for their fruitful discussions and scientific assistance. The financial support of the Federal Minister of Economics is gratefully acknowledged.

References

[1] U. Bonse and M. Hart, Z. Phys. 188, 154 (1965).
[2] U. Bonse, E. te Kaat, and P. Spieker, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.), Spec. Publ. 343 (Aug. 1971), p. 291.
[3] R. D. Deslattes *et al.*, Ann. Phys. (N.Y.) 129, 378 (1980).
[4] P. Becker, in *Neutron Interferometry*, Ed. by U. Bonse and H. Rauch (Oxford University Press, London, 1979), p. 416.
[5] H. J. Rademacher *et al.*, PTB-Bericht APh-13, Sects. A through P (Physikalisch-Technische Bundesanstalt, Braunschweig, July 1980).
[6] I. Curtis *et al.*, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.), Spec. Publ. 343 (Aug. 1971), p. 285.
[7] G. Reim, in PTB-Bericht Me-13 (Physikalisch-Technische Bundesanstalt, Braunschweig, 1977), p. 139.
[8] D. Kind, P.-V. Seances Com. Int'l. Poids Mes. Ser. 2, 45, 23 (1977).
[9] H. Kogelnik and T. Li, Appl. Opt. 5, 1550 (1966).
[10] P. Becker, in PTB-Bericht E-18, (Physikalisch-Technische Bundesanstalt, Braunschweig, Sept. 1981), p. 217.
[11] P. Becker *et al.*, Phys. Rev. Lett. 46, 1540 (1981).
[12] U. Bonse and M. Hart, Z. Phys. 190, 455 (1966).
[13] K. J. Hanszen *et al.*, PTB-Bericht APh-14 (Physikalisch-Technische Bundesanstalt, Braunschweig, March 1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Systematic Uncertainties in the Determination of the Lattice Spacing *d*(220) in Silicon

H. Siegert and P. Becker

Physikalisch-Technische Bundesanstalt, D 3300 Braunschweig, F.R.G.

The origin and magnitude of corrections and uncertainties connected with the absolute determination of the $d(220)$ lattice plane spacing in a silicon crystal are discussed. Contributions of crystal material, temperature, optical wavelength, alignment parameters, and guiding errors are estimated. A total correction of $-3.9 \times 10^{-8}\,\overline{n/m}$ of the mean value $\overline{n/m} = (\lambda/2d)$ is calculated. The total uncertainty, comprising random and systematic uncertainties, amounts to $\pm 6 \times 10^{-8}\,d(220)$. The uncertainty of the mean contributes by less than one percent to this value, while the main part results from the uncertainties of the crystal temperature and the Abbe error.

Key words: lattice parameter; precision length measurement; silicon.

1. Introduction

In order to achieve an accuracy level of less than 1×10^{-7} for the lattice spacing $d(220)$ obtained by the 170 measurements as reported in Ref. [1], a great number of subsidiary investigations were carried out.[1] Each particular correction and uncertainty was reduced below the 1σ error of a single measurement of the ratio $\lambda/2d$ (λ = optical wavelength).

2. Origin of Uncertainty Sources

According to the basic equation

$$d = (n/m)^{-1}\lambda/2\,, \qquad (1)$$

the lattice spacing d of a crystal is compared with a laser wavelength λ, known in meters, by simultaneous phase measurements $m\,2\pi$ and $n\,2\pi$ in an optical and in an x-ray interferometer.

Principal uncertainty sources in the measurement of d arise from a non-perfect crystal structure and from the uncertainty of the absolute crystal temperature. The main errors of the phase measurements may result from additional phase shifts proportional to the translation s of the analyzer crystal. Such phase shifts cannot be separated from the linear phase variation to be measured. Therefore, they lead to an erroneous value of the lattice plane spacing.

Due to Fresnel phase shift corrections, the wavelength within the interferometer requires special attention.

In the following, these effects will be discussed on the basis of the experimental data acquired at the Physikalisch-Technische Bundesanstalt (PTB) at the end of 1980.

[1]Interlaboratory project of the Physikalisch-Technische Bundesanstalt: x-ray interferometry: H. Siegert, P. Becker; Length measuring techniques: K. Dorenwendt; Mechanics: H.-J. Rademacher; Electronics: P. Seyfried, G. Ebeling, W. Lucas.

[2]Supplier: Wacker Chemitronic GmbH, Berghausen, Germany. Trade name: WASO (float-zoned) EPD-free $\rho \approx 1600\,\Omega$ cm, p-type material.

3. Magnitude of the Uncertainty Sources

3.1 Homogeneity of the Lattice Plane Spacing

The degree of homogeneity of the $d(220)$ lattice plane spacing within the silicon material[2] of the x-ray interferometer can be estimated from a moiré topogram [2] of the crystal lamellas. (For a schematic of the apparatus, see Ref. [3], p. 31.) The topogram is shown in Fig. 1. Over the whole crystal lamella, no moiré period smaller than $\Lambda = 5$ mm is detectable. The relative uncertainty $\Delta d/d$ of the crystal material, calculated from $\Delta d/d = d/\Lambda$, is therefore less than 4×10^{-8} ($d \approx 0.2$ nm).

FIGURE 1. *Moiré topogram of the x-ray interferometer (for details see Ref. [3], p. 31). As can be recognized, the lattice spacing varies by less than 4×10^{-8} in the given material. The region used in the experiments is marked.*

3.2 Temporal Drifts

Temporal drifts between the signals of the optical and the x-ray interferometer are unavoidable. (For details see Ref. [4], p. 75.) Figure 2 gives an example of the magnitude of the drifts occurring in a temporal interval of one day. For comparison, one single measurement is carried out in one hour. Intervals with linear drift, nonlinear drift, and no drift are indicated. But to a first approximation, the temporal drifts can be taken as linear and can be explained by a small temperature variation of

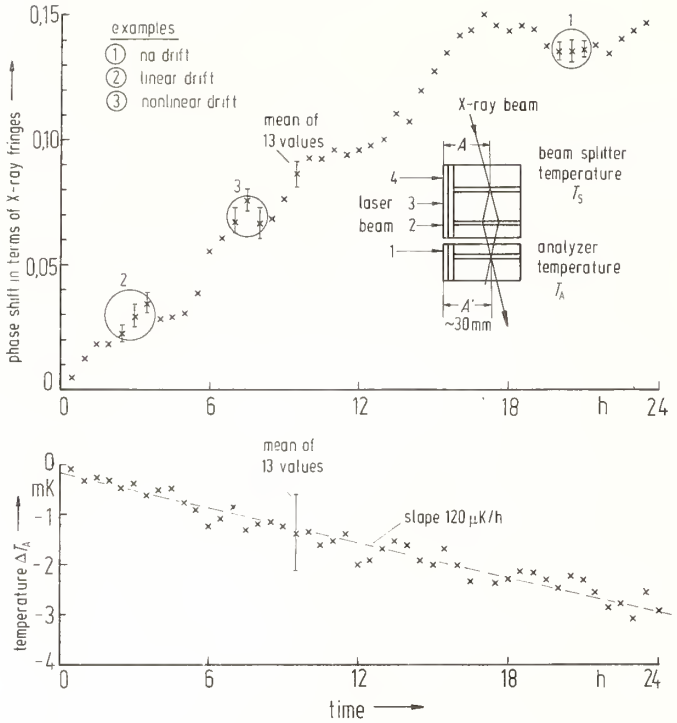


FIGURE 2. *Temporal drifts: (a) phase shifts between the signals of the x-ray and optical interferometers; (b) temperature variations of the analyzer crystal (those of the beam splitter are similar).*

the crystals. The linear part of the drift can be explained by assuming a difference of only $\Delta T_A - \Delta T_S = 25$ μK/h of a distance $A = 30$ mm between the measuring points of the two interferometers.

Great efforts were made to reduce all avoidable temperature gradients near the crystals. As shown by Fig. 2, the analyzer temperature varied by only 3 mK in 24 h. This was an essential requirement for the execution of the precision measurements of d. The uncertainty of the absolute temperature of the analyzer crystal actually amounts to ±20 mK, a value which will be reduced in the future.

3.3 Elastic Forces of the Translation Stage

Additional phase shifts proportional to the translation S of the analyzer are caused by the forces within the double parallel spring of the translation stage, see e.g., Ref. [5]. These forces vary proportionally to S. For measuring such displacements and tilts, we used additional optical means of high sensitivity (0.1 nm; 5×10^{-9} rad).

A special arrangement of the translation stage and the magnetic compensation of guiding errors, already given in Ref. [1], proved to be very effective. Within the limits given above we realized an almost perfect linear motion of the analyzer crystal.

3.4 Cosine Error of the Translation Measurement

The translation S of the analyzer as measured by the optical interferometer (S_L) and by the x-ray interferometer (S_X) is only of equal size if the directions of the lattice vector **h**, the translation vector **S**, the mirror normal **n**, and the direction of the laser beam $\boldsymbol{l}$ are collinear. The ratio (S_L/S_X) is proportional (see e.g., Ref. [4], p. 131) to an expression containing the deviation angles and their differences in arguments of cosine functions. Because all these deviation angles were reduced to values of less than 10^{-4} rad (see Fig. 3), the quadratic expansion of the cosine terms leads to a relative error $\Delta S/S < 0.5 \times 10^{-8}$ of the translation measurement.

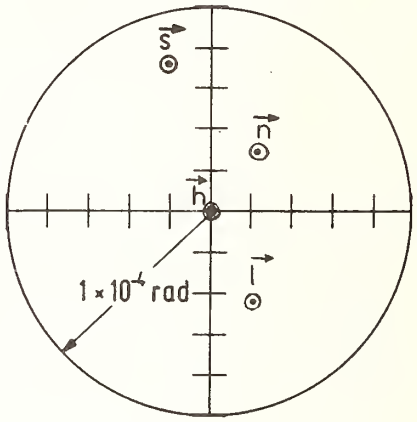


FIGURE 3. *Residual angular deviations of the translation vector S, the mirror normal* **n**, *and the laser beam* $\boldsymbol{l}$ *form the direction of the lattice vector* **h**.

3.5 Abbe Error of the Translation Measurement

A so-called Abbe error [6] of the translation measurement is produced by a combination of two effects;

(i) a tilt of the analyzer during motion, caused by an angular guiding error of the translation stage; and
(ii) an offset between the interferometric base lines of length measurement.

The product of the tilt angle and the offset results in the error given in the upper part of Fig. 4.

Errors of this kind are detected and eliminated by phase measurements between the interferometer signals, by varying the offset or tilt angle with respect to both the axis perpendicular to S, the axis usually called ϑ, and the ρ axis. In this way, the offsets $H\vartheta$ and H^ρ between the interferometric base lines were reduced to $H\vartheta = (40 \pm 20)$ μm and $H^\rho = (240 \pm 20)$ μm; and the guiding errors to the residual values $\Delta\vartheta$, $\Delta\rho = 0 \pm 5 \times 10^{-9}$ rad.

In the lower part of Fig. 4, the results of a new method for a quantitative elimination of the Abbe error are given. Both parameters, the offset and the tilt angle, were varied independently. The guiding error was changed by feeding an additional voltage to the piezoelement used for the fine adjustment of the analyzer crystal. The voltage was continuously increased from zero to a few volts in order to generate a tilt proportional to S. The point of intersection of the lines drawn in Fig. 4 gives the value $(\overline{n/m}) = 1648.281\,700$ where the Abbe error vanishes.

In applying this method to the complete elimination of the Abbe error with respect to the ρ axis, a new error not yet discussed in the literature has been observed. The additional error occurs when, during translation the lattice planes of the analyzer are tilted in relation to those of the beam splitter. This may be explained by a change of the intensity distribution within the moiré pattern, causing a small phase change of the x-ray signal

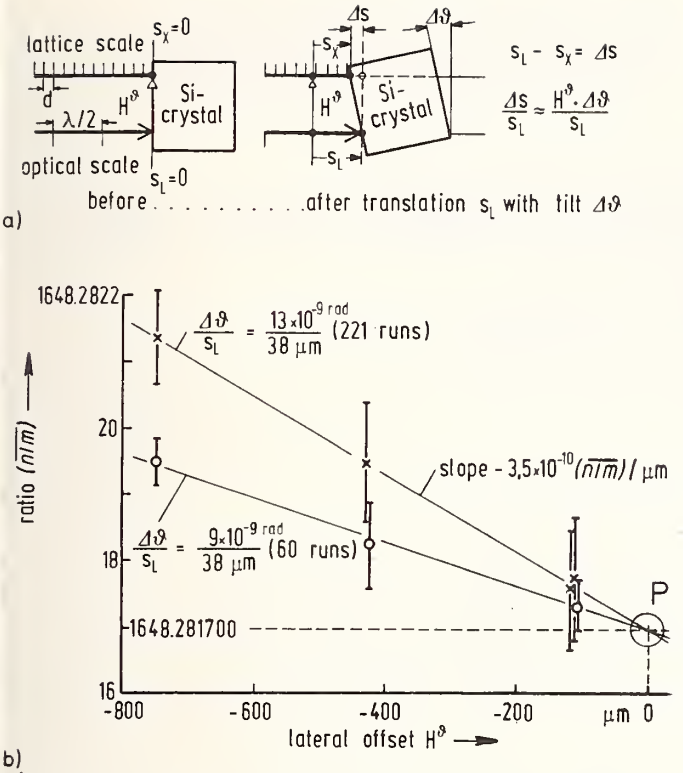


FIGURE 4. *Abbe error of the translation measurement due to an angular guiding error $\Delta\vartheta/S_L$ and an offset $H\vartheta$: (a) scheme; (b) influence of different values of $H\vartheta$ and $\Delta\vartheta/S_L$ on the measured ratio $(\overline{n/m})$.*

and thus simulating an additional translation. The effect would not be expected in the case of a perfect crystal structure and an ideal x-ray detector.

By keeping the guiding error $\Delta\rho_A/S_X$ close to zero, disturbing phase shifts of the x-ray signal were avoided. According to Fig. 5, the translation error is found to be proportional to the product $\rho_A\,\Delta\rho_A$. The unknown coefficient C in the relation

$$\Delta S/S_X = \Delta(\overline{n/m})/(\overline{n/m}) = C\,\rho_A\,\Delta\rho_A/S_X \qquad (2)$$

has been evaluated by means of Fig. 5. The result is C = 18 km.

This value is of the same order of magnitude as the minimum curvature radius R of the lattice planes, assuming the moiré period $\Lambda \geq 5$ mm and using $R = \Lambda^2/8d$. It is not yet clear if a direct relation exists between C and R.

The value $(\overline{n/m})$ obtained for the vanishing Abbe error (intersection point of the almost horizontal line with the vertical axis of Fig. 5) does not differ appreciably from the corresponding value $(\overline{n/m})$ $H\vartheta = 0$ (point P in Fig. 4). A small cosine error in the data of Fig. 4 and 5 has not yet been eliminated.

3.6 Wavelength Correction

The laser beam has a vacuum wavelength λ_0 = (632 991 415 ± 5) fm and a Gaussian beam profile. The beam radius in the waist is W_o = (0.52 ± 0.02) mm. Due to the non-planarity of the wavefronts [7], a small Fresnel phase shift correction is needed [8]. Its magnitude is given by

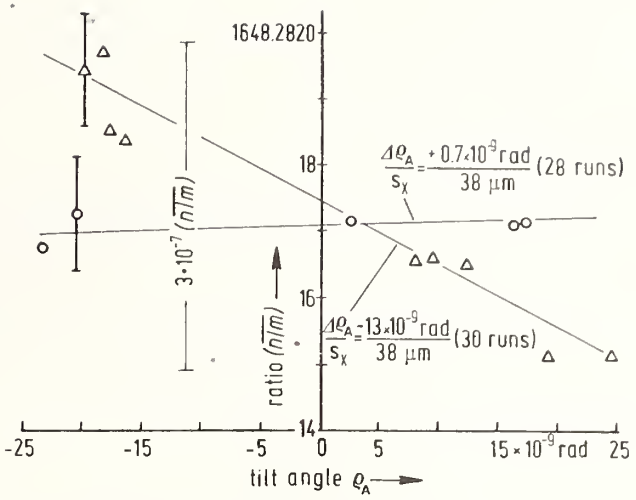


FIGURE 5. *Ratio $(\overline{n/m})$ as a function of the tilt angle ρ_A between the lattice planes of the crystals. The guiding error $\Delta\rho_A/S_X$ is the parameter. [For the meaning of S_X, see Fig. 4(a)].*

$$\Delta\lambda/\lambda = (\lambda/2\pi W_o)^2. \qquad (3)$$

We used a Lamb-dip stabilized HeNe laser (Spectra Physics Mod. 119) the frequency of which was controlled from time to time by a $^{127}I_2$-stabilized HeNe-reference laser [9].

4. Results and Discussion

The numerical results for the corrections and their individual uncertainties are given in Table 1. They are based on the above-mentioned series of 170 measurements of the ratio $(\overline{n/m})$ which are given in Ref. [1].

Of all the corrections, the wavelength correction, although small, is predominant. The Abbe error and the absolute crystal temperature contribute most of all to the uncertainties.

A total correction of -3.9×10^{-8} for the ratio $(\overline{n/m})$ is obtained by adding up the individual corrections arithmetically. The contributions of the uncertainties were regarded as independent quantities and summed in quadrature. The total uncertainty amounts to -6×10^{-8} on a relative scale.

The resulting value for the lattice spacing $d(220)$ of our silicon crystal at 22.50 °C in vacuum, given at the bottom of Table 1, has been confirmed under measuring conditions where Abbe errors were almost eliminated.

The authors wish to thank Professor U. Bonse, University of Dortmund, and Professor K.-J. Hanszen for fruitful discussions and scientific assistance. The financial support of the Federal Minister of Economics is gratefully acknowledged.

References

[1] P. Becker and H. Siegert, these proceedings.
[2] U. Bonse and M. Hart, Z. Phys. 190, 455 (1966).
[3] H.-J. Rademacher *et al.* PTB-Bericht APh-13 (Physikalisch-Technische Bundesanstalt, Braunschweig, July 1980).
[4] K.-J. Hanszen *et al.*, PTB-Bericht APh-14 (Physikalisch-Technische Bundesanstalt, Braunschweig, March 1981)

[5] Ch. Hoffrogge and H.-J. Rademacher, PTB-Mitteilungen 83, 79 (1973)
[6] H. Franke, *Lexikon der Physik*, Vol. 1 (Stuttgart, 1969), see "Abbe-sches Prinzip".
[7] H. Kogelnik and T. Li, Appl. Opt. 5, 1550 (1966).
[8] K. Dorenwendt and G. Bönsch, Metrologia **12**, 57 (1976).
[9] J. Helmcke, PTB-Bericht Me-13 (Physikalisch-Technische Bundesanstalt, Braunschweig, January 1977), p. 13.

TABLE 1. *Numerical results of the corrections and uncertainties for the measured ratio* $\overline{(n/m)}$. *Since the result of each measurement at an individual temperature (22.42 °C . . . 22.50 °C) is reduced to a reference temperature (22.50 °C), no temperature correction for the mean is required.*

result of 170 measurements	$\overline{n/m} = 1648.281\,626\,(1 \pm 5.8 \cdot 10^{-8})$ 22.5° C; vac.	
values $\frac{\Delta(\overline{n/m})}{(\overline{n/m})}$	rel. correction $\times 10^8$	uncertainty $\times 10^8$
statistical error of $(\overline{n/m})$	0	± 0.4
wavelength correction	- 3.8	± 0.2
cosine-error	- 0.1	± 0.2
Abbe-error	0	± 3.0
temperature	0	± 5.1
laser frequency	0	± 0.8
total correction of $(\overline{n/m})$	- 3.9	± 6.0
lattice plane spacing	$d_{220}^{\text{vac; } 22.5^\circ C} = (192\,015.560 \pm 0.012)\,fm$	

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

High Precision Studies of Pionic X Rays: Some Past Results and Future Prospects*

G. Dugan,† L. Delker,†† and C. S. Wu

Columbia University, New York, NY 10027

and

D. C. Lu

Yale University, New Haven, CT 06520

The development of high intensity pionic x-ray sources has allowed a high precision crystal spectrometer measurement of certain pionic x-ray energies to be made. These energy measurements have resolved the relativistic fine structure and have been used to provide an improved determination of the negative pion mass (accuracy 6.4 ppm). Substantial further increases in pionic x-ray source intensity have been shown to be feasible at LAMPF. These increases will make possible the use of double flat crystals and can lead to advances in the study of electron screening in pionic atoms, further high precision pion mass and pionic atom fine structure studies, and possibly also high precision muonic x-ray energy measurements.

Key words: bent crystal spectrometer; exotic atoms; pion mass; pionic x rays; relativistic fine structure.

1. Introduction

High resolution studies of pionic x rays using a bent crystal spectrometer have in the past been limited by the lack of an adequately intense source, which is required because of the relatively low efficiency of crystal diffraction instruments. However, the recent pioneering work of Marushenko *et al.* [1] illustrated how such an intense source could be provided by placing the pionic x-ray target directly next to the pion production target. We have used a pionic x-ray source of this general design at the Nevis synchrocyclotron in conjunction with a large-aperture bent crystal spectrometer of high intrinsic resolution. The high flux of pionic x rays and excellent crystal resolution were exploited to make high accuracy measurements of the energies of the π-Ti ($5g \rightarrow 4f$) and ($5f \rightarrow 4d$) transitions, and the π-P ($4f \rightarrow 3d$) transition, relative to the energy of a nuclear gamma ray transition in ^{99}Mo. Using the energy of the calibration ^{99}Mo line measured at the NBS [3], the absolute energies of the pionic x rays were established, leading to a determination of the pionic mass with an accuracy of 6.4 ppm [4]. The energy difference between the π-Ti ($5g \rightarrow 4f$) and ($5f \rightarrow 4d$) lines, which the crystal spectrometer was able to completely resolve, is due to the relativistic fine structure; the measured energy difference was found to agree within the error ($\sim 2\%$) with the predictions of the Klein-Gordon equation [5]. The details of these measurements are discussed below.

In an experiment under development at the LAMPF accelerator, we plan to utilize the 500 μA proton beam in an arrangement similar to that of the Nevis experiment, but with substantially increased signal rates. Results obtained during 1980, based on the observation of the π-Al ($4f \rightarrow 3d$) x ray at 30 keV, indicate that for the observation of a typical x-ray line, counting rates of 1000/min at a resolution of $\Delta E/E = 10^{-4}$ can be expected with the use of a bent crystal, or 10/min at $\Delta E/E = 10^{-5}$ with a double crystal set. With this performance, one may expect that significant improvements can be made in the precision of measurements of pionic x rays. The details of our preliminary results and some of the physics implications of such higher precision measurements are discussed below.

2. Nevis Experiment

2.1 Experimental Setup

The experimental arrangement is shown in Fig. 1. Negative pions were produced by the interaction of the 560 MeV, 2 μA extracted proton beam in a copper target of total mass 44 gm/cm^2. Some of these pions stopped in the x-ray target, which was placed adjacent to the pion production target as shown in Fig. 1. The x-ray target was composed of slabs of titanium or phosphorous. The atoms formed by the pions stopping in the x-ray target were the source of the pionic x rays observed in the experiment.

These pionic x rays traveled along a 6.4 m long hole in the shielding wall (the "neutral beam"), passed through a 0.91 m long tapered multislit collimator (which served as a background suppression device) and were incident on a quartz crystal deployed in the Cauchois configuration. The crystal was cut to the (3, 1, 0) orientation and bent to a radius of 3.51 m. By use of a new bending technique [2], the intrinsic angular resolution achieved over an open aperture of 9 cm by 9 cm was 6.6 sec of arc (FWHM). The diffracted radiation from the crystal

*Supported in part by the National Science Foundation and the U.S. Department of Energy.

†Present address: Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510.

††Present address: Riverside Research Institute, 80 West End Avenue, New York, NY 10023.

passed through a narrow (170 μm) slit at the image plane, and was detected in one of a pair of intrinsic Ge detectors. The crystal spectrometer lineshape was mapped out by scanning the slit (using a linear motion stage of ± 0.5 μm accuracy) across the angular range of interest under the control of a computer, which also supervised the collection of the x-ray data from the Ge detectors. The effective energy resolution of the crystal spectrometer with this slit thickness was 17 eV (FWHM) at 40 keV. The efficiency of the spectrometer in this arrangement was roughly 0.5×10^{-9}. The spectrometer was enclosed in a cave, as shown in Fig. 1, to reduce the neutron background.

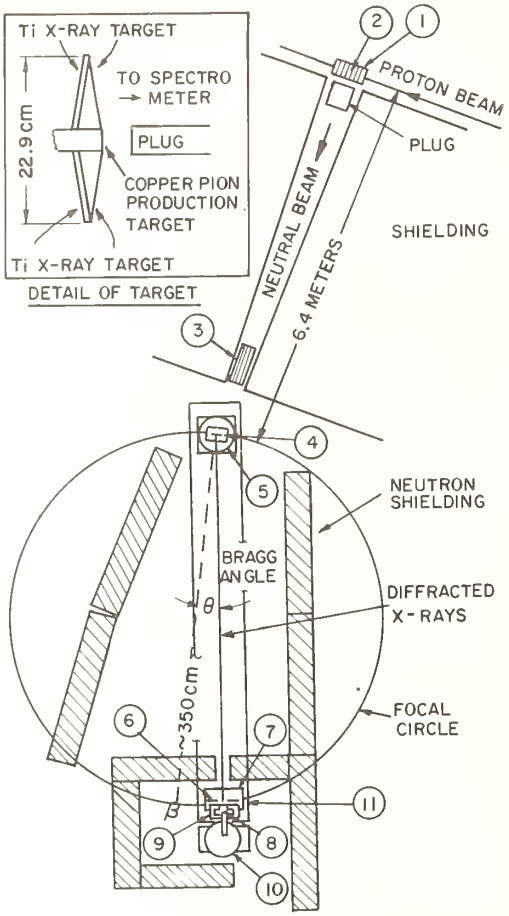


FIGURE 1. *Experimental layout of crystal spectrometer, collimator, and pion targets. Inset shows detail of pion production and x-ray targets. (1) x-ray target; (2) π^- production target; (3) collimator; (4) bent crystal; (5) turnable crystal mount; (6) thin slit; (7) slit table; (8) solid-state detector (Ge); (9) detector shield; (10) Dewar; (11) spectrometer frame.*

2.2 Data Collection

The experiment measured the difference in diffraction angle between the pionic x rays and a source of nuclear gamma rays (from the decay of ^{99}Mo). This angular difference was converted into an energy difference using a scale factor determined from measurements of the electronic x rays Sm($K\alpha_1$) and Eu($K\alpha_2$). The pionic x-ray energies were then determined from this energy difference by referring to a measurement of the absolute energy of the ^{99}Mo nuclear gamma ray using a double flat crystal at the NBS [3]. Consequently, the data collection involved measurements of electronic x rays and nuclear gamma rays as well as pionic x rays. In each case, the effective source geometry was arranged to be very similar to that of the pionic x rays.

The electronic x rays were produced by fluorescing powder oxide targets; the scale factor was determined from these lines (~0.0890 eV/μm at 40 keV) to an accuracy far better than required. The nuclear gamma rays were generated in a 1.5 Ci liquid source. Figure 2, top, shows the diffraction line from the nuclear gamma rays; since the natural linewidth is much less than 17 eV, this is essentially the spectrometer instrumental lineshape. Because of the large spectrometer vertical acceptance, the line profile shows a tail on the low-wavelength side. The nuclear gamma-ray peak position, which is indicative of the spectrometer stability, was found to be constant to within ± 1.5 μm.

During the collection of the pionic x-ray data, the lineshape was scanned typically 30 times during a 24-h run, to average out any x-ray intensity variations due to small shifts in the proton beam position on the production target. X-ray intensity variations due to proton beam intensity fluctuations were eliminated by determining the data collection time at each slit position by the number of protons on the production target. Finally, to provide additional background reduction, data from the Ge detectors were suppressed except during the beam spill, which was reduced from the usual 5 ms to 20 μs for this experiment.

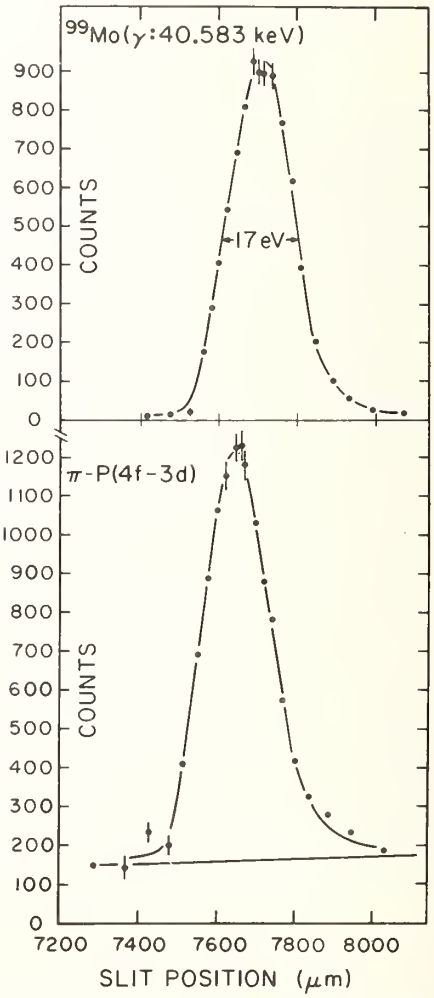


FIGURE 2. *Bent-crystal diffraction lines for the calibration γ ray from ^{99}Mo (upper panel) and the (4f → 3d) pionic x ray from π-P (lower panel).*

2.3 Data Analysis

For each slit position, a pulse height spectrum from the Ge detectors was obtained. Each of these spectra was fitted to a Gaussian lineshape (FWHM of 1 keV at 40 keV) plus a linear background. This background was subtracted and the area of the Gaussian for each slit position was extracted; this procedure removes the ambient background appearing in the Ge detector spectrum. The extracted areas for each slit position are plotted in Fig. 2, bottom, for a typical π-P ($4f \rightarrow 3d$) run, and in Fig. 3, for the π-Ti ($5g \rightarrow 4f$) and ($5f \rightarrow 4d$) transitions. The remaining background in these figures is due to a diffracted soft-photon continuum. The signal-to-background ratio for the π-P ($4f \rightarrow 3d$) data shown in Fig. 2, before the ambient background subtraction, was 2.2. The peak signal counting rates per proton charge on target were $12/\mu A \cdot$ min for the π-P data and $3/\mu A \cdot$ min for the π-Ti ($5g \rightarrow 4f$) data.

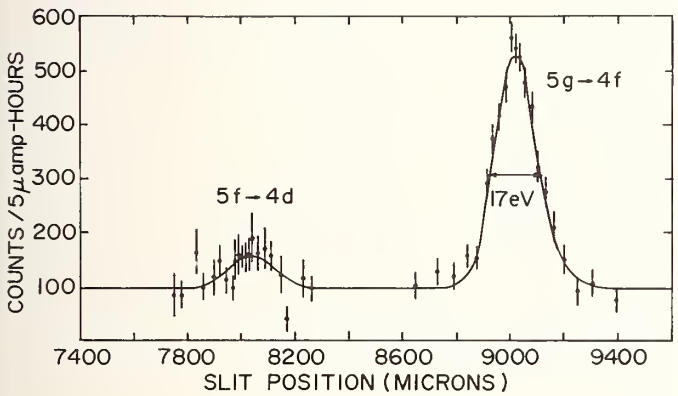


FIGURE 3. *Crystal spectrometer spectrum (with Ge background subtracted) of fine-structure components of the 5-4 transition in pionic titanium. Solid line is best fit to the data, with a χ^2 of 28.7 for 35 degrees of freedom.*

The line center positions for the various data sets were obtained by approximating the spectrometer instrumental lineshape by a Gaussian with an exponential tail, integrated over the slit width. The lineshape is dominantly instrumental for the pionic x-ray data as well as the nuclear gamma ray data. Since the pionic x-ray and nuclear gamma-ray lines are quite close in energy, any effects due to the line shape approximation essentially cancel out when the difference in line center position is computed. This line center difference is converted to an energy difference using the scale factor obtained from the electronic x-ray data. The energy differences between the pionic x-ray and nuclear gamma ray lines must finally be subjected to an ~ 0.8 eV correction associated with finite vertical aperture effects. The final results for the energy differences were:

$$E(\pi\text{-P}(4f \rightarrow 3d)) - E(\text{Mo}, \gamma) = 4.84 \pm 0.20 \text{ eV}$$

$$E(\pi\text{-Ti}(5g \rightarrow 4f)) - E(\text{Mo}, \gamma) = -118.4 \pm 0.28 \text{ eV}$$

$$E(\pi\text{-Ti}(5f \rightarrow 4d)) - E(\pi\text{-Ti}(5g \rightarrow 4f)) = 87.6 \pm 1.8 \text{ eV}.$$

The errors include both statistical errors and systematic errors associated with the vertical aperture correction.

2.4 Results

The measured energy differences between the π-P and π-Ti x-ray lines, and the ^{99}Mo gamma-ray line, were converted to absolute energies using the ^{99}Mo gamma-ray energy of 40583.468 ± 0.17 eV as measured at the NBS [3]. The resulting absolute pionic x-ray energies for the π-P ($4f \rightarrow 3d$) and π-Ti ($5g \rightarrow 4f$) transitions were then compared with the predictions of the Klein-Gordon equation, including corrections due to vacuum polarization, strong interaction, and electron screening effects. The pion mass is determined by requiring agreement between the measured energies and the theoretical predictions. Table 1 shows the results of these calculations and the resulting pion mass determination. The error in the mass includes both experimental and theoretical uncertainties. The major source of error in the predictions arises from uncertainties in the appropriate electron K and L shell populations at the time of the pionic transition, which affects the electron screening corrections. Estimates of these populations were established by a cascade calculation, including both radiative and non-radiative refilling processes to the K, L, and M electronic shells. The uncertainties in the estimates are reflected in the uncertainties in the electron screening corrections shown in Table 1.

The pion mass determination shown in Table 1 represents an improvement over the present world average [6]:

$$m_\pi = 139566.9 \pm 1.2 \text{ keV}$$

but is in agreement with it. When combined with the recent improved measurement [7] of the μ^+ momentum in pion decay, and with the assumption that $m_{\pi^+} = m_{\pi^-}$, it also allows a slight reduction in the upper limit on the muon neutrino mass to be made ($m_{\nu_\mu} < 0.52$ MeV at 90% confidence level).

TABLE 1. *Computation for pionic mass (all energies in eV unless noted).*

Transition	π-P ($4f \rightarrow 3d$)	π-Ti ($5g \rightarrow 4f$)
Measured energy	40588.31 ± 0.26	40465.07 ± 0.33
Corrections: vacuum polarization		
$\alpha(Z\alpha)$	−99.3	−82.6
$\alpha^2(Z\alpha)$	− 0.7	− 0.6
$\alpha(Z\alpha)^{3,5,7}$	0.1	0.2
Strong interaction	−0.23 ± 0.06	0.0
Electron screening	0.98 ± 0.20	3.99 ± 0.20
Klein-Gordon energy	40489.16 ± 0.33	40386.06 ± 0.38
Pionic mass (keV)	139567.8 ± 1.1	139567.1 ± 1.3
Average pion mass (keV)	139567.5 ± 0.9	

A comparison of the measured energy splitting between the π-Ti ($5f \rightarrow 4d$) and π-Ti ($5g \rightarrow 4f$) transitions with the prediction of the Klein-Gordon equation may also be made. (See also ref. [8].) This splitting is due primarily to the relativistic fine structure term in the Klein-Gordon equation. There are significant corrections

due to vacuum polarization, electron screening, and the strong interaction; a summary of the relevant numbers contributing to the transition energy difference is presented in Table 2, together with the measured result. Although the agreement between the calculations and the experimental measurement does verify the prediction of the Klein-Gordon equation for this splitting, it does not distinguish between that equation and the Schroedinger equation with relativistic corrections. This is because the prediction of the Klein-Gordon equation for the relativistic fine-structure splitting differs from that which one would calculate from relativistic corrections to the Schroedinger equation by considerably less than the experimental error.

TABLE 2. *Pionic titanium fine structure splitting* $E(5f \rightarrow 4d) - E(5g \rightarrow 4f)$, *in eV.*

Klein-Gordon splitting	59.3
Vacuum polarization:	
first order	25.6
higher order	0.2
Strong interaction	3.3 ± 1.1
Electron screening	−1.1 ± 0.1
Total predicted splitting	87.3 ± 1.1
Measured splitting	87.6 ± 1.8

3. LAMPF (Los Alamos Meson Physics Facility) Experiment

3.1 Experimental Setup

At the LAMPF accelerator, a beam of 800 MeV protons with a current of about 0.5 mA offers the opportunity for a several order of magnitude increase in the effective pionic x-ray source intensity compared to that available at Nevis. In order to investigate the feasibility of an experiment at LAMPF similar to that of our Nevis experiment, we have taken preliminary data in a relatively crude setup. Although these data do not constitute high precision energy measurements, they do indicate what future, more precise experiments can expect as regards signal-to-noise ratios and signal counting rates, which reflects directly upon what instruments can be used and the kinds of physics questions which can be addressed.

The preliminary investigations used a general setup similar to that shown in Fig. 1. The pion production target was a block of Mo, of total mass 20 gm/cm^2; the effective pionic x-ray target was an aluminum housing around the Mo block. The block was viewed at a 135° angle from the proton beam direction, through a 15 m long hole in the shielding wall. Because of initial uncertainties regarding the background, a 1.5 m long collimator with a 2.54 cm diameter hole in the center was located in the shielding wall aperture near the crystal; this severely limited the crystal viewing aperture. The crystal and Ge detectors were enclosed in a cave for neutron background shielding. To map out the pionic x-ray line, a 170 μm slit was scanned across the crystal image plane at the appropriate Bragg angle, as in the Nevis experiment.

3.2 Results

Figure 4 shows the diffraction line obtained with the setup described above. No ambient background subtraction has been made. The line width of 14 eV is dominantly instrumental. No precise energy calibration was performed. The signal-to-noise ratio of 2 and the peak counting rate of 12 counts/min are quite good for this crude preliminary setup. Extrapolating from this signal rate to that obtainable in a future bent crystal experiment with a larger collimator aperture, a thicker pion production target, and a more favorable x-ray target geometry, one can realistically expect signal rates of roughly 1000 counts/min. Because of various limitations present in the preliminary setup, we expect the background to scale up by considerably less than the signal, providing an extrapolated signal-to-noise ratio of 5 or better.

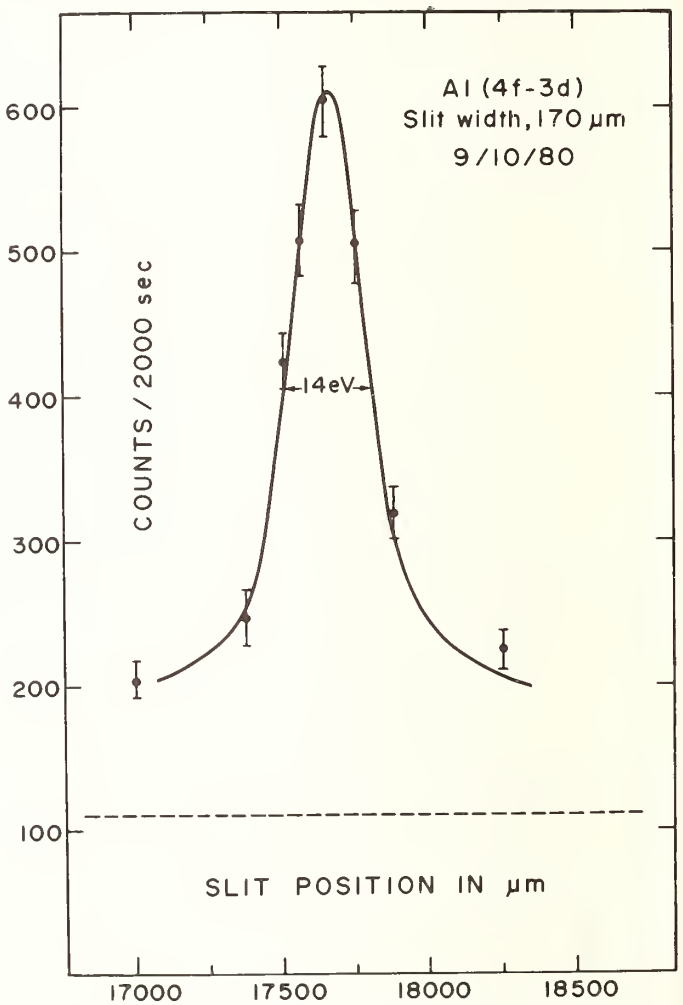


FIGURE 4. *Bent crystal diffraction line for the π-Al (4f → 3d) x-ray transition obtained at LAMPF.*

3.3 Future Outlook

The development of sufficiently intense sources of pionic x rays to allow the use of crystal diffraction spectrometers has now progressed to the point where, with counting rates as high as mentioned above, even very weak pionic x-ray lines can be studied with bent crystals. Additionally, the application to pionic x-ray studies of lower efficiency, higher resolution crystal diffraction instruments (such as the double flat crystal spectrometer)

appears quite feasible. With such intense sources of pionic x rays to be expected, one may also anticipate that the muons from the decay of these pions in flight could also provide a less intense but still useful source of muonic x rays which could be studied with crystal diffraction instruments.

With the resolution available in a double crystal pair, one may contemplate a direct measurement of the pionic x-ray line splitting due to missing K-electrons in pionic atoms, providing an experimental determination of the electron screening corrections and allowing the precision of pion mass measurements to be improved to about 1 ppm. Higher precision studies of the relativistic fine structure splitting can distinguish the Klein-Gordon equation from the Schroedinger equation with relativistic corrections. High precision absolute energy measurements, together with an accurate knowledge of the pion mass, can be used to place limits on the polarizability of the pion. Finally, if muonic x rays are observable in sufficient intensity, a high precision test of vacuum polarization effects in strong fields (high-Z muonic atoms) can be carried out.

References

[1] V. I. Marushenko *et al.*, Pisma Zh. Eksp. Teor. Fiz. 23, 80 (1976) [JETP Lett. 23, 72 (1976)].
[2] D. C. Lu, to be published.
[3] E. G. Kessler, R. D. Deslattes, and W. Schwitz, private communication (1980).
[4] D. C. Lu *et al.*, Phys. Rev. Lett. 45, 1066 (1980).
[5] L. Delker *et al.*, Phys. Rev. Lett. 42, 89 (1979).
[6] R. L. Kelley *et al.* (Particle Data Group), Rev. Mod. Phys. 52, 51 (1980).
[7] M. Daum *et al.*, Phys. Rev. D 20, 2692 (1979).
[8] K. C. Wang *et al.*, Phys. Rev. A 22, 1072 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Isotope Shifts of K X-Rays of Lead

G. L. Borchert, O. W. B. Schult, and J. Speth

Institut für Kernphysik, Kernforschungsanlage Jülich, D-5170 Jülich, West Germany

and

P. G. Hansen

Institute of Physics, University of Aarhus, DK-8000 Aarhus, Denmark

and

B. Jonson and H. Ravn

CERN-ISOLDE, CH-1211 Geneva, Switzerland

and

J. B. McGrory*

Oak Ridge National Laboratory, Oak Ridge, TN 37830, U.S.A.

During the last years progress in nuclear theory has allowed more detailed predictions for ground state properties of heavier elements especially in the vicinity of doubly magic nuclei. Experimentally the change of the mean square charge radius can be determined by a high resolution measurement of the K x-ray isotope shifts. Therefore we performed a study of the isotope shifts of all stable lead isotopes and compared them to very recent microscopic calculations.

Key words: high resolution crystal spectrometer; isotope shifts of stable lead isotopes; precision measurements of x-ray energies.

1. General Considerations

The energy of atomic electrons is determined by the Coulomb interaction with the central charge of the nucleus and with all neighboring electrons. Therefore a change of the nuclear charge distribution will be reflected in a change of the electron binding energies. As the wave function of the 1s electrons has the largest overlap with the nuclear volume the K x-ray energies show the most direct influence.

Therefore the shift of the x-ray energy when the neutron number changes, the isotope shift, yields information about the change of the mean square charge radius $\delta\langle r^2\rangle$ of the nucleus [1, 2, 3].

Of course, there are other nuclear and electronic effects that are known to shift the energies of K x-rays. Generally they are related to the mechanism used to create the K hole vacancy and to the chemical environment. For photoionization unresolved satellites due to multihole excitation shift the centroid of the K x-ray line [4, 5].

The electronic structure of an electron capture decaying element can partly survive during the lifetime of the K hole. Thus the change of the electronic Coulomb interaction (screening) causes an x-ray shift [6]. Statistically a similar effect arises from the chemical bond [7].

*Research sponsored by the Division of Basic Energy Sciences, U.S. Department of Energy, under contract W-7405-eng-26 with the Union Carbide Corporation.

The interaction of the nuclear magnetic moment with the electron spin of the K electrons splits the 1s level. For a statistical population, i.e., photoionization, only a broadening of the line width results [8], whereas in the case of allowed electron capture decay [9] or conversion decay [10], a nonstatistical population can occur that shifts the x-ray energy.

To study the isotope shift which is often smaller than these effects it is essential to aim for proper experimental conditions.

2. The Isotope Shift

Usually the isotope shift is expressed as a sum of three contributions:

$$\delta E^{AA'} = \delta E\,(\mathrm{Coul})^{AA'} + \delta E\,(\mathrm{Mass})^{AA'} + \delta E\,(\mathrm{Nuclpol})^{AA'}. \quad (1)$$

The last term describes the contribution of the nuclear polarization effects. It has been estimated to be negligibly small.

The second term gives the contribution of the normal and specific mass shift [11]. To a good approximation it can be expressed as

$$\delta E\,(\mathrm{Mass})^{AA} \sim -\,2/3\,\frac{A - A'}{1836\,A^2}\,E_x, \quad (2)$$

where E_x is the energy of the x-ray transition in keV. In the case of lead for $A - A' = 1$ it is only ~ -0.7 meV.

The first term describes the effect of the change of the volume of the nuclear charge distribution. Seltzer [3] found a rapidly converging expansion that connects it to the changes of nuclear moments:

$$\delta E^i(\text{Coul}) = \sum_n C_n^i \, \delta \langle r^{2n} \rangle \tag{3}$$

$$\delta \langle r^{2n} \rangle = \int \delta \rho(r) \, r^{2n} \, dv \Big/ \int \rho(r) dv. \tag{4}$$

It is common to write this expression as

$$\delta E^i(\text{Coul}) = C_1^i \sum_n \frac{C_n^i}{C_1^i} \, \delta \langle r^{2n} \rangle = C_1^i \, \lambda \,, \tag{5}$$

with the nuclear parameter λ. Neglecting the very small Coulomb shift of the $2p_{3/2}$ level, the $K\alpha_1$ isotope shift in Pb can be expressed numerically as

$$\delta E(K\alpha_1) - \delta E(\text{Mass}) =$$

$$1880 \, \delta \langle r^2 \rangle \left\{ 1 - 1.12 \times 10^{-3} \frac{\delta \langle r^4 \rangle}{\delta \langle r^2 \rangle} + \ldots \right\} \frac{\text{meV}}{\text{fm}^2} \,. \tag{6}$$

By means of the simple nuclear model of a homogeneously charged sphere with constant density the magnitude of the nuclear moments can be approximated by

$$\delta \langle r^{2n} \rangle \sim \frac{2n}{2n+3} R_0^{2n} \frac{A - A'}{A} \,, \tag{7}$$

$$\text{with } R_0 = 1.2 A^{1/3} \text{ fm} \,.$$

As the ratio C_n^1/C_1^i is largely independent of the type of electronic transition i, the same quantity λ can be determined in both x-ray and optical measurements.

As it is rather difficult to calculate to high precision for optical transitions [13] the C_1 coefficient and the mass shift, these parameters can be obtained from comparison with x-ray results [14].

Because of the problems in the analytical understanding of muonic x-ray data their comparison to electronic x-ray data seems to be too early [15, 16, 17].

The information deduced from elastic electron scattering experiments generally depends on the model used for the analysis [18]. Recently a new method was developed [19] which yields model-independent results for the nuclear parameter λ.

3. The Experiment

The simple estimate Eq. (7) shows that the shift of K x-ray lines for Pb due to the isotope effect is on the order of 10^{-6} of the transition energy or 10^{-3} of the natural line width. Therefore only a high resolution device like a crystal spectrometer [20, 21] is suited to detect such small energy differences. Even with such an instrument it is very difficult to perform absolute energy measurements with an accuracy at a level of about 10^{-7} [22].

For the study of the isotope shifts in Pb we used our DuMond-type bent crystal spectrometer. It is described elsewhere [23] and permits us to measure independently at essentially the same time (see Fig. 1) for each angular setting the intensity of two sources placed very close to each other at the focal point. In this way the difference of the x-ray energies is determined largely undisturbed by the uncertainty of the angular setting and reproducibility. In measuring the x-ray reflections at positive and negative Bragg angles an uncertainty due to a possible relative displacement of the sources in the direction of dispersion can be eliminated. An effective shift because of the vertical aberation can be corrected by measuring the two sources at interchanged positions and averaging the results.

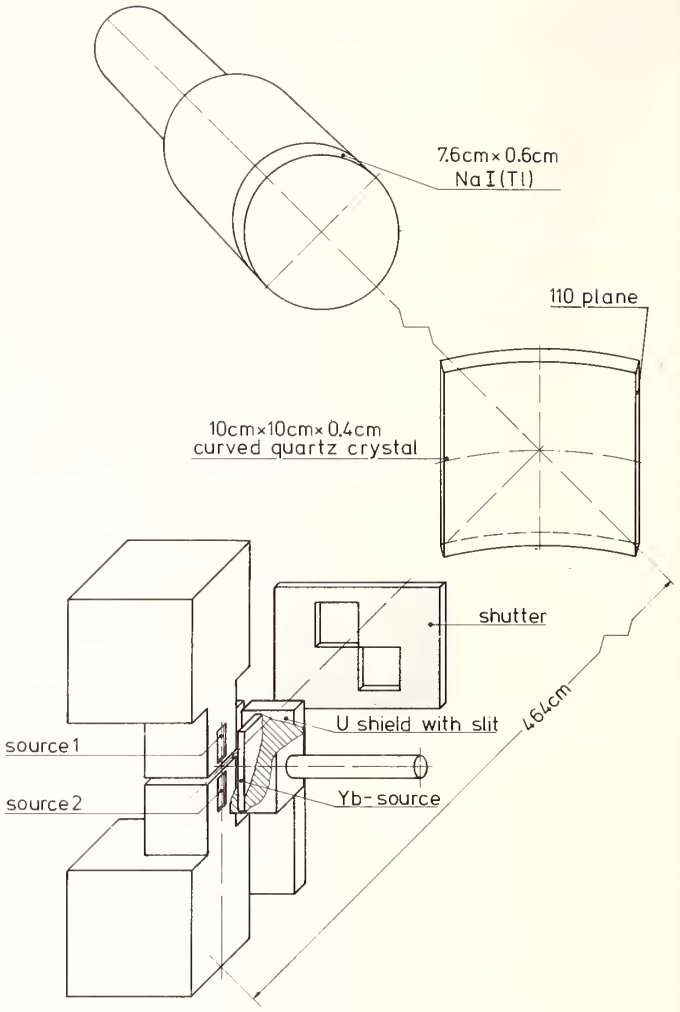


FIGURE 1. *Schematic view of the source arrangement of the crystal spectrometer. The inactive metal foils, source 1 and source 2, are mounted on aluminum blocks which can be adjusted separately in the direction of dispersion. The radioactive Yb source to excite the fluorescence is located behind a uranium plate with a slit to let pass the fluorescence radiation of the metal foils. Another movable uranium plate with two windows acts as a shutter, which allows the radiation of only one source at a time to reach the detector.*

As sources we used isotopically enriched metallic foils the composition of which is given in Table 1. The x-rays were induced through photoionization by means of a 50 Ci ^{169}Yb source placed very close to the Pb samples. Their dimensions, $4 \times 2 \times 0.12$ mm^3, were chosen such that after proper adjustment the line width of the reflections was essentially the natural one: $\Gamma \sim 1.3 \, \Gamma_{\text{nat}}$ [24]. A typical pair of reflections is shown in Fig. 2.

In this way the influence of all the competing effects mentioned above could be eliminated. The experimental results indicate that we measured the pure isotope shifts, the accuracy of which was essentially limited only by counting statistics.

We measured each combination of the lead isotopes ^{204}Pb, ^{206}Pb, ^{207}Pb and ^{208}Pb between 20 and 40 times. We used four different mathematical approaches to analyze the data:

1. A fit of the line shape and position.
2. An estimate of the line position from the center of gravity [25].

TABLE 1. *The composition of the sources in %*

Source	Isotope			
	^{208}Pb	^{207}Pb	^{206}Pb	^{204}Pb
^{208}Pb	96.6	2.1	1.3	—
^{207}Pb	5.5	92.4	2.2	—
^{206}Pb	2.9	6.7	90.4	—
^{204}Pb	10.0	6.4	12.7	70.9

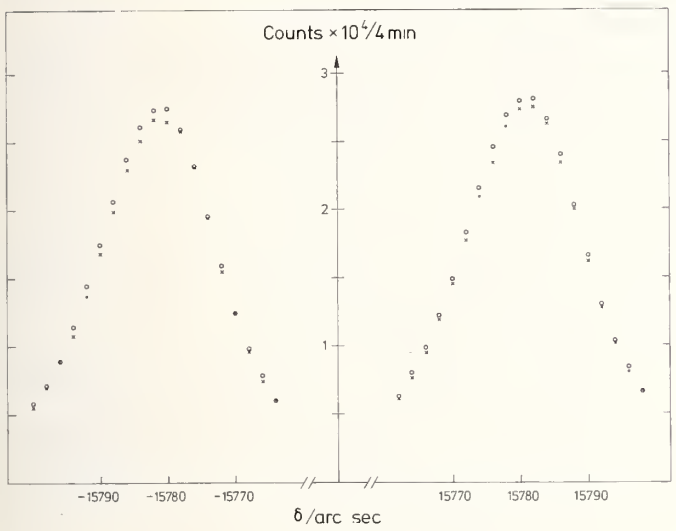


FIGURE 2. *A scan across the second order reflections of the $K\alpha_1$ transition in the experiment $^{204}Pb - ^{208}Pb$. The contents of counter 1 measuring the intensity of the ^{204}Pb source is denoted by circles, the contents of counter 2 measuring the intensity of the ^{208}Pb source is denoted by crosses. The relative geometrical displacement of the two sources is about 1 μm estimated from a mean distance of the line centers of about 40 milli-arcsec.*

3. An estimate of the shift from the displacement of the four corresponding lines.
4. Determination of the shift from the difference of the area of the overlapping lines.

As a typical example the fit results from the pair ^{204}Pb – ^{207}Pb are shown in Fig. 3. The results of the individual methods were found to agree within the uncertainties. As the final result the weighted mean was taken of the results of the four methods. The uncertainty is estimated as the average of the individual errors. In this

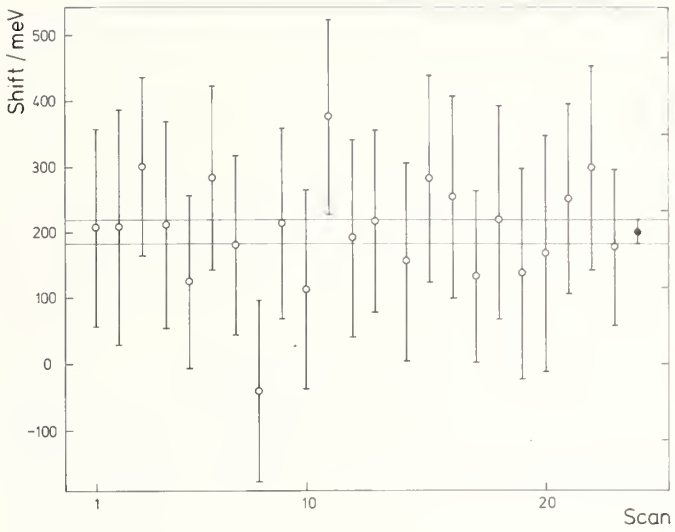


FIGURE 3. *The energy shift of the $K\alpha_1$ transition in the experiment $^{204}Pb - ^{207}Pb$. Displayed are the individual x-ray shifts from 23 scans as given by a fit of the line shape and position with the model of a Lorentz curve folded with a rectangular distribution which corresponds to the effective geometrical source width. The full circle and the horizontal lines denote the weighted average and its uncertainty, respectively.*

way for the six possible combinations of the four lead isotopes the measured isotope shifts were obtained that are given in the third column of Table 2. All quoted uncertainties are standard deviations. This overdetermined system was submitted to a level fit procedure taking into account the isotopical composition of the individual sources according to Table 1. The resulting best values for the measured shifts are listed in the fourth column of Table 2. From this system the Coulomb shifts for pure isotopes were computed. The results are shown in column 5 of Table 2. In column 6 the results of Ref. [26] are shown for comparison. The resulting nuclear parameter according to Eq. (6) is listed in column 4 of Table 3. In the third column we give the results of our theoretical calculation. For comparison in column 5 the x-ray results are shown from Ref. [26] and in column 6 very recent preliminary results from elastic electron scattering experiments [27]. The latter were found using the Bessel Fourier analysis according to Ref. [18] and they depend, therefore, on the applied model parameters.

4. Discussion

The isotope shifts determined in this experiment as given in Table 2, column 4, show a significant improve-

TABLE 2. *Energy ($K\alpha_1$, 1)—Energy ($K\alpha_1$, 2) in meV*

Isotope 1	Isotope 2	Experim.	Level Fit	Coulomb Shift	Ref [26]
204	208	324 (16)	325 (13)	414 (17)	—
206	208	200 (16)	209 (13)	228 (14)	186 (18)
207	208	153 (20)	138 (12)	150 (14)	136 (25)
204	207	186 (15)	187 (12)	264 (17)	—
206	207	81 (15)	71 (12)	78 (14)	50 (20)
204	206	135 (50)	116 (15)	186 (20)	200 (38)

TABLE 3. *The nuclear parameter λ in fm²*

Isotope 1	Isotope 2	λ (theor)	λ (x-ray)	λ (Ref. [26])	λ (electr. sc.)
208	207	0.061	0.080 (7)	0.072 (13)	0.0654 (20)
208	206	0.116	0.121 (7)	0.099 (10)	0.1283 (39)
208	204	0.226	0.220 (9)	0.205 (22)	0.2363 (71)

ment in accuracy as compared to the previous results [26]. The agreement is rather satisfactory.

When comparing the corresponding values of the nuclear parameters to those of the theoretical calculation, one has to keep in mind that the latter considers only the mean square radius. But it has been estimated that higher order moments should contribute only 7% so that the comparison is justified. Nevertheless the agreement is surprisingly good. As for the elastic electron scattering data one has to keep in mind that they still depend on the special choice of model parameters.

As a common feature of all the data, the nuclear parameter for the ^{208}Pb – ^{206}Pb pair is less than twice that of the pair ^{208}Pb – ^{207}Pb. This can be understood in terms of a state-dependent effective neutron charge [28], as there are contributions from $f_{5/2}$, $p_{3/2}$, $f_{7/2}$, $h_{9/2}$, and $i_{13/2}$ two neutron hole states to the ground state of ^{206}Pb. They all contribute with an effective neutron charge that has been estimated to be smaller than that of the $2p_{1/2}^{-2}$ state.

For the pair ^{208}Pb – ^{204}Pb there is very good agreement between all experimental results and the theoretical one.

References

[1] F. Boehm and P. L. Lee, At. Data Nucl. Data Tables **14**, 605 (1974).

[2] O. I. Sumbaev and A. F. Mezentsev, Sov. Phys. JETP **22**, 323 (1966); O. I. Sumbaev, A. F. Mezentsev, V. I. Marushenko, A. S. Ryl'nikov, and G. A. Ivanov, Sov. J. Nucl. Phys. **9**, 529 (1969).

[3] E. Seltzer, Phys. Rev. 188, 1916 (1969).

[4] G. L. Borchert, P. G. Hansen, B. Jonson, I. Lindgren, H. L. Ravn, O. W. B. Schult, P. Tidemand-Petersson, and the Isolde Collaboration, Phys. Lett. **65A**, 297 (1978).

[5] T. A. Carlson, C. W. Nestor, Jr., T. C. Tucker, and F. B. Malik, Phys. Rev. **169**, 27 (1968).

[6] G. L. Borchert, P. G. Hansen, B. Jonson, I. Lindgren, H. L. Ravn, O. W. B. Schult, P. Tidemand-Petersson, and the Isolde Collaboration, Phys. Lett. **66A**, 374 (1978).

[7] O. I. Sumbaev, in *Modern Physics and Chemistry, Vol. I* (Academic Press, London, 1976), p. 31; O. I. Sumbaev, Sov. Phys. Usp. **21**, 141 (1978).

[8] A. S. Ryl'nikov, A. I. Egorov, G. A. Ivanov, V. I. Marushenko, A. F. Mezentsev, A. I. Smirnov, O. I. Sumbaev, and V. V. Fedorov, Sov. Phys. JETP **36**, 27 (1973).

[9] G. L. Borchert, P. G. Hansen, B. Jonson, H. L. Ravn, O. W. B. Schult, P. Tidemand-Petersson, and the Isolde Collaboration, Phys. Lett. **63A**, 15 (1977).

[10] A. I. Egorov, A. A. Rodionov, A. S. Ryl'nikov, A. E. Sovestnov, O. I. Sumbaev, and V. A. Shaburov, JETP Lett. **27**, 483 (1978).

[11] R. B. Chessler and F. Boehm, Phys. Rev. **166**, 1206 (1968).

[12] S. K. Bhattacherjee, F. Boehm, and P. L. Lee, Phys. Rev. 188, 1919 (1969).

[13] A. Steudel, Z. Phys. **133**, 438 (1952).

[14] P. Brix, private communication.

[15] R. J. Powers, P. Barreau, B. Bihoreau, J. Miller, J. Morgenstern, J. Picard, and L. Roussel, Nucl. Phys. A **316**, 295 (1979).

[16] G. A. Rinker and J. Speth, Nucl. Phys. A **306**, 360 (1978).

[17] Y. Yamazaki, H. D. Wohlfahrt, E. B. Shera, M. V. Hoehn, and R. M. Steffen, Phys. Rev. Lett. **42**, 1470 (1979).

[18] H. Euteneuer, J. Friedrich, and N. Voegler, Nucl. Phys. A **298**, 452 (1978).

[19] L. Cardman, P. Leconte, D. G. Ravenhall, and S. E. Williamsson, Proc. Int. Conf. Nucl. Phys. with Electromagnetic Interactions, Mainz, 1979.

[20] J. W. M. DuMond, Rev. Sci. Instrum. **18**, 626 (1947).

[21] Y. Cauchois, J. Phys. Radium 3, 320 (1932).

[22] G. L. Borchert, W. Scheck, and K. P. Wieder, Z. Naturforsch Teil A: 31, 274 (1965).

[23] G. L. Borchert, P. G. Hansen, B. Jonson, H. L. Ravn, O. W. B. Schult, and P. Tidemand-Petersson, Nucl. Instrum. and Methods **178**, 209 (1980).

[24] G. C. Nelson, W. John, and B. G. Saunders, Phys. Rev. 187, 1 (1969); and Phys. Rev. A **2**, 542 (1970).

[25] P. G. Hansen, Nucl. Instrum. and Methods 154, 321 (1978).

[26] P. L. Lee and F. Boehm, Phys. Rev. C **8**, 819 (1973).

[27] G. Fricke, H. Miska, and D. Rychel, private communication.

[28] J. Speth, L. Zamick, and P. Ring, Nucl. Phys. A **232**, 1 (1974).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Measurement of Atomic Masses by Mass Spectroscopic Methods and a Role for Atomic Masses in the Determination of the Fundamental Constants

Walter H. Johnson

University of Minnesota, School of Physics and Astronomy, 116 Church St. S.E., Minneapolis, MN 55455

A description of atomic mass determinations is given with the emphasis on mass spectroscopic methods. A review of current techniques is made which includes conventional mass spectrometers and radio frequency instruments. Precision attained in these measurements is discussed together with the possibilities of improvements in precision. Finally, the use of atomic mass measurements as input data for the determination of fundamental constants is considered.

Key words: atomic masses; fundamental constants; mass spectrometer; rf mass spectrometer.

1. Introduction

One of the fundamental characteristics of an atom is its mass and as such the atomic mass has been a quantity subject to measurement for some time. I intend to review some of the important considerations dealing with the experimental determination of atomic masses, to discuss examples of current instrumentation for this purpose, to give a selection of recent results, and finally to comment on the possibility of experiments in which precise atomic mass information may contribute to the determination of fundamental constants.

2. Systematics of Atoms

The various atomic species may be classified by two parameters, N the number of neutrons and Z the number of protons in the nucleus or the number of electrons surrounding the positively charged nucleus of the neutral atom. Stable or near-stable atoms are represented in a distribution of N and Z values which range from $0 \leq N \leq \sim 160$ and $1 \leq Z \leq 92$. There are about 280 different stable or near-stable atomic species and perhaps five or six times as many unstable atoms which last long enough to be studied with present techniques.

Determination of the atomic masses of the stable atomic species has been the main emphasis of mass measurement experiments. These stable atom results supplemented by decay energy information from unstable atoms allows the construction of atomic mass tables which list the atomic mass $m(ZN)$ for most experimentally accessible combinations of N and Z. These masses are specified as the masses of neutral atoms in their atomic and nuclear ground state. Using these tables, one can calculate the binding energy of atoms with Eq. (1)

$$BE = [Zm_e + Zm_p + Nm_n - m(ZN)]c^2 \qquad (1)$$

in which m_e, m_p, and m_n are the masses of the electron, proton, and neutron, respectively. The binding energy of an atom may be divided into two parts, one involving the binding of the electrons about the nucleus and a much larger part involving the binding of the protons and neutrons in the nucleus. Because most of the binding energy results from the nuclear interactions, atomic masses have been extensively employed for studies of nuclear stability. In these studies, corrections sometimes were made to remove the effect of electron binding. In light atoms the corrections were very small, but in the heavy atoms the electronic binding energy can approach 0.1% of the total binding energy.

The masses employed in Eq. (1) are absolute masses, i.e., in kilograms as the equation is written. Experience has shown, however, that the use of a relative scale of atomic masses results in information of considerably higher accuracy and utility. From early in the history of the measurement of atomic masses, a relative scale was employed which sets the mass of a particular atom and then makes all other measurements relative to that atom. The current standard mass is the mass of the atom ^{12}C in its nuclear and atomic ground state which is assumed to be exactly 12 u. The conversion of atomic mass from relative to absolute units is given in Eq. (2).

$$\begin{aligned} m_{kg}(ZN) &= m_u(ZN)/N_A \\ &= m_u(1.660\ 5655 \pm 85 \times 10^{-27}\ \text{kg}) \end{aligned} \qquad (2)$$

where N_A is Avogadro's number [1].

Another useful conversion is to the electron-volt energy unit. This conversion is given in Eq. (3) in which F is the Faraday [1].

$$1\ \text{u} = \frac{c^2}{F} = 931.5016 \pm 26\ \text{MeV} . \qquad (3)$$

3. Measurement of Atomic Masses

Although this review will concentrate on the measurement of atomic masses by mass spectroscopic techniques, it is worthwhile at the onset to indicate that there is a second general technique for the determination of atomic masses. This technique employs nuclear reactions in which high velocity nuclei are caused to interact with other nuclei. Measurements of the energies and momenta of the colliding nuclei before and after the collision can be employed to determine the mass difference between the incident and the outgoing particles. Atomic mass difference measurements may therefore be accomplished with this technique. Discussion of this, however, is not within the scope of this review.

Mass spectroscopic determinations of the atomic mass of an atom employs a mechanical process in which a force is applied to the atom and the resultant change in path of the atom is dependent on its mass value. Because it is convenient to apply these forces by electric and magnetic fields, the atom must be converted to an ion and one

must bear in mind that the mass measured is now an ion mass rather than the neutral atom.

An ion of charge q and mass m moving with a velocity v perpendicular to a magnetic field B will move in a circular path of radius R such that

$$R = \frac{mv}{Bq} . \tag{3}$$

If the velocity of the ion is achieved by an acceleration through electric potential difference V, then

$$v = \sqrt{\frac{2qV}{m}} . \tag{4}$$

Substituting this value into Eq. (4) yields

$$R^2 = \frac{2V}{B^2} \cdot \frac{m}{q} . \tag{5}$$

This equation is the basic design equation for magnetic mass spectrometers. Another useful relationship is the cyclotron period T_c, of a ion of mass-to-charge ratio m/q in a uniform magnetic field B.

$$T_c = \frac{2\pi R}{v} = \frac{2\pi}{B} \cdot \frac{m}{q} . \tag{6}$$

This quantity as was the case for R^2 in Eq. (5) is directly proportional to the mass-to-charge ratio of the ion. Equation (5) is the basis for the design of a typical sector field mass spectrometer shown schematically in Fig. 1. Ions are formed in an ion source by collision with a fast beam of electrons. The ions are then accelerated through a potential difference V and into a sector magnetic field. Here ions of differing m/q will have different radii; shown here are two different m/q values. The two groups of ions emerge on two different paths with the beam of lighter mass ions being the inner path. The diagram also indicates the focusing ability of a sector field. Divergent ions emerging from the ion source are refocused by the sector field at the collector slit.

By varying B or V a series of ion beams are sequentially allowed to have the correct radius to pass through the collector slit and be measured as a collector current or counted as individually arriving ions. A mass spectrum is produced which is schematically shown in Fig. 2. The location of an ion group in this spectrum is related to its m/q, and the spacing between ion groups is related to the mass difference between the ions. Small fractional mass differences such as m_1 or m_2 in Fig. 2 and known as mass doublets have been the basis for mass determinations since the early history of these measurements.

In order to make accurate mass difference measurements a number of conditions are required.

(a) The ion beam shapes should be identical.

(b) The resolution of the instrument should be as high as possible to produce sharply defined location.

(c) The transformation from spacing to mass difference should be well known. This requires that the dispersion of the sector field can be determined accurately.

A great deal of effort has gone into developing ion-optical designs of mass spectroscopes to produce high resolution. The aberration of an optical system is usually expressed as a power series in the angular divergence of the ion beam emerging from the ion source and the velocity spread β such that the velocity of the emerging ion $v = v_0(1 + \beta)$ where v_0 is the mean velocity. Defining y as the image aberration, one can express y as a power series in α and β as indicated in Eq. (7):

$$y = a_m(B_1\alpha + B_2\beta + B_{11}\alpha^2 + B_{12}\alpha\beta + B_{22}\beta^2 + \ldots), \tag{7}$$

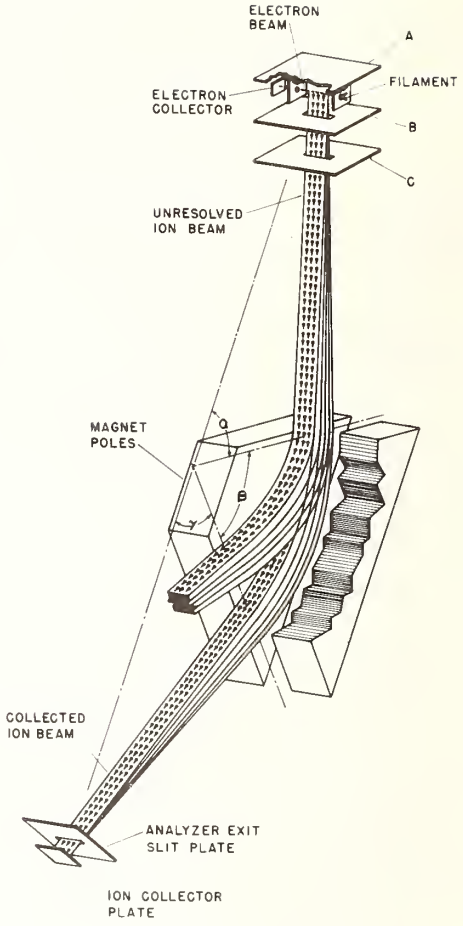


FIGURE 1. *A schematic of a simple mass spectrometer.*

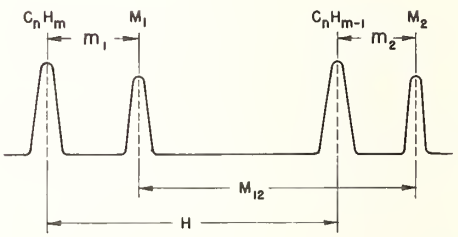


FIGURE 2. *A mass spectrum showing hydrocarbon ion peaks and element ion peaks. Doublets m_1 and m_2 measure the mass difference between the known hydrocarbon ion peak and the unknown element peak. The doublet H is a calibrating doublet for which the mass difference is one hydrogen mass.*

where a_m is the radius of curvature of the ion in the magnetic field. The design goal has been to arrange combinations of electric and magnetic fields in the proper way to cause as many of the B coefficients to be zero as possible. Causing the ion-optical system to have smaller aberrations not only will produce high resolution but also will cause the ion beam shapes to be less and less dependent on conditions in the ion source and thus more and more similar.

One of the limits to the accuracy of early mass difference measurements was caused by inaccuracy in knowledge of the dispersion characteristics of the ion-optical system. Mass spectra were detected using photographic plates located in the focal plane of the instrument.

Because ion beams of differing m/q had different paths and focused at different points along the plate, lack of homogeneity of the sector magnetic field and variations of the fringe field perturbed the ability to calculate dispersion accurately. An excellent review of the early history of atomic mass measurement has been written by Mattauch [2].

Let us next consider some specific examples of mass spectroscopic devices. I wish to start with an instrument design that was developed in the late 1940s by A. O. C. Nier and E. G. Johnson [3] at the University of Minnesota. The design of this instrument marked the emergence of electrical detection in mass measurement. Previous to that time, photographic plate detection was generally employed and the devices were known as mass spectrographs. This instrument was the first mass measurement instrument that properly could be called a mass spectrometer. The instrument was designed as a symmetrical 90° electrostatic analyzer followed by a 60° asymmetrical magnetic analyzer. The analyzer radii and magnetic sector asymmetry were chosen to produce a device in which the B_1, B_2, and B_{11} coefficients of Eq. (7) were zero for ions focused at the collector slit. A schematic diagram of the configuration of fields is shown in Fig. 3.

Mass difference measurements in this instrument were based on an ion optical theorem by W. Bleakney [4] or an earlier version by W. F. G. Swann [5]. Given that an ion of mass M pursues a certain trajectory in an arbitrary configuration of electric and magnetic fields; then if all the electric fields are changed by a factor A, a charged particle of the same charge having a mass M/A will pursue exactly the same trajectory as did the particle of mass M before the change in fields took place.

The various electric fields in the ion source and the electric field in the electrostatic analyzer are the two sets of electric fields which will influence the passage of an ion through the instrument. Because of the velocity focusing property $B_2 = 0$, the changes in the ion source electric fields are not influential in changing the location of the ion at the collector slit. Changes in the electrostatic analyzer are, however, crucial. The dispersion relationship may be derived from Bleakney's theorem by assuming two ions of mass m_1 and m_2 in the same charge state and associated electrostatic analyzer potentials V_1 and V_2:

$$\frac{m_2 - m_1}{m_1} = \frac{V_1 - V_2}{V_2} \text{ or } \frac{\Delta m}{m} = \frac{\Delta V}{V} . \qquad (8)$$

The measurement of a mass difference Δm is then measured in terms of an electrostatic potential ratio. The potential ratio $\Delta V/V$ can be determined by means of a resistance divider shown in Fig. 4. With $R = R'$, one can show that

$$\frac{\Delta m}{m} = \frac{\Delta V}{V} = \frac{\Delta R}{R} . \qquad (9)$$

The alternation of the group of ions at the collector slit is accomplished by switching the relay shown in Fig. 4 from position A to position B. Changes in the ion source potentials are made to follow approximately the $\Delta V/V$ of the electrostatic analyzer. Also the ion beam is swept across the collector slit by a small saw-tooth perturbing magnetic field. These wave forms are shown in Fig. 5. Timing of ion currents through the collector slit is shown in the line labeled electrometer. If this signal is viewed on an oscilloscope which is synchronized to begin its sweep at the beginning of the magnet sweep, first one and then the other ion current is recorded. Similarity of orbit is achieved when the two ion beam waveforms are

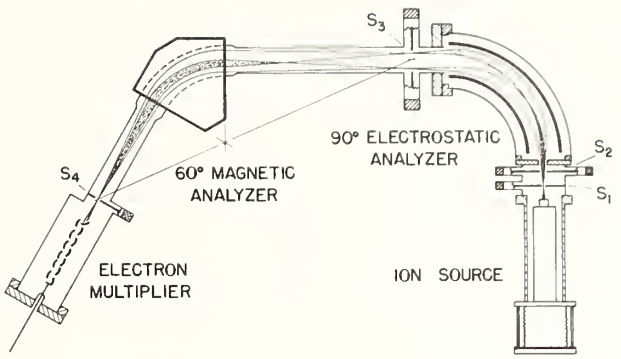


FIGURE 3. *A schematic of a Nier-Johnson geometry double-focusing mass spectrometer.*

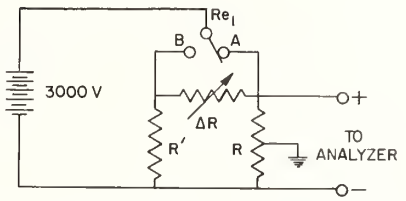


FIGURE 4. *The circuit employed to generate the electrostatic analyzer voltage wave forms.*

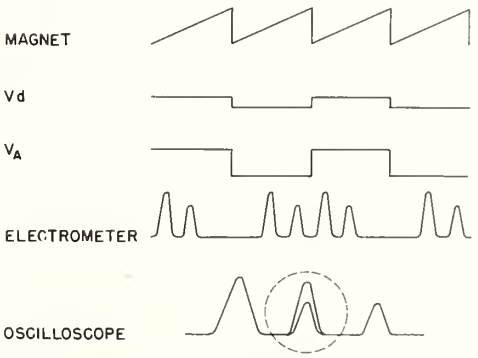


FIGURE 5. *Wave forms employed in peak matching.*

coincident. This is achieved by adjustment of ΔR. When the two ion beams shapes are coincident, the required value of ΔR can be used to calculate Δm. This technique, known as peak matching, was first used by Lincoln Smith [6] and adapted by Giese and Collins [7] to the Minnesota instruments.

Peak matching offers a precise method for locating the matched condition between the two ion beam waveforms. Smith [8] has claimed that matching accuracy can be as high as 1/2500 of the ion beam width. Typical photographic plate accuracies were usually no better than 1/50 of the ion beam width [9]. Thus, a very large increase in the accuracy of mass measurement was achieved.

A number of instruments have been constructed at Minnesota using this design. The largest had a magnetic radius of 40 cm and an ion path of about 4 meters. Full width at half maximum resolutions as high as 1 in 200,000 were achieved with this instrument. The instrument has been employed for mass measurements for more than 20 years, and in that time measurements of most stable isotopes have been made.

In 1959, Hintenberger and Konig [10] published a series of ion-optical designs for mass spectrometers

which were fully second order focusing, that is, all second order B coefficients were zero. A number of instruments have been constructed according to these designs. The instrument at the University of Manitoba described by Barber *et al.* [11] will be used as the example of this generation of instrument. The geometry of this instrument is given in Fig. 6. A 94.65° electrostatic analyzer with a mean radius of 100 cm is combined with a 90° magnetic analyzer of mean radius 62.74 cm. A non-normal entry and exit from the magnetic field is a significant feature of the design which through Z focusing produces a high transmission. The total path length of the ion beam in this design is 4.6 meters. With a source slit of width 2.2 μm, the expected base resolution is 1 part in 200,000.

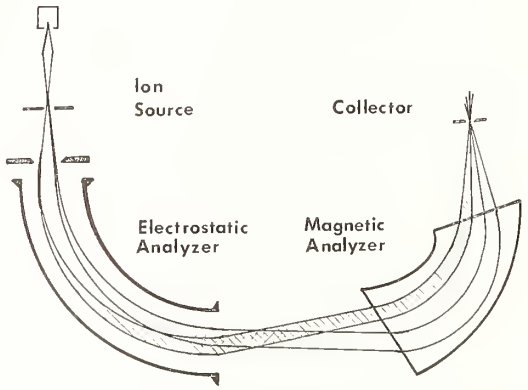


FIGURE 6. *Configuration of the University of Manitoba double-focusing mass spectrometer.*

Figure 7 is a schematic diagram of the instrument showing the vacuum system, ion source, and collector details. The various defining slits are individually adjustable in location and width. The instrument is mounted on a rigid table which is in turn supported by pneumatic mounts to aid in isolating the instrument from external vibrations. A plasma ion source of the von Ardenne type is employed. Ions are accelerated by an accelerating potential of 19.4 kV and focused on the source slit with a pair of electrostatic quadrupole lenses. The electrostatic potentials for the accelerator, quadrupole lenses and the electrostatic analyzer deflection plates can be incremented according to Bleakney's theorem requirements. Incrementing is accomplished by a timing signal which also starts other aspects of the measurements. Ions which pass through a collector slit are detected by an electron multiplier.

Mass differences are determined from the incremental electrostatic analyzer voltage ΔV needed to produce similar orbits. The circuit which provides the electrostatic analyzer potentials is shown in Fig. 8. The relay at the output of the ΔV supply alternately switches from one position to the other each time a timing signal is received. In one position the electrostatic analyzer voltage is V, and in the other $V \pm \Delta V$, where V is the sum of the series mercury battery voltages or about 780 V. Both V and ΔV may be precisely measured by the potentiometer shown. The batteries and potentiometer are housed in a temperature controlled box for stability. Ions are swept across the collector slit by a sawtooth magnetic field provided by a Helmholtz coil located at the point where the beam emerges from the magnet. Oscilloscope presentation of the beam shape is used for focusing and for direct view peak matching.

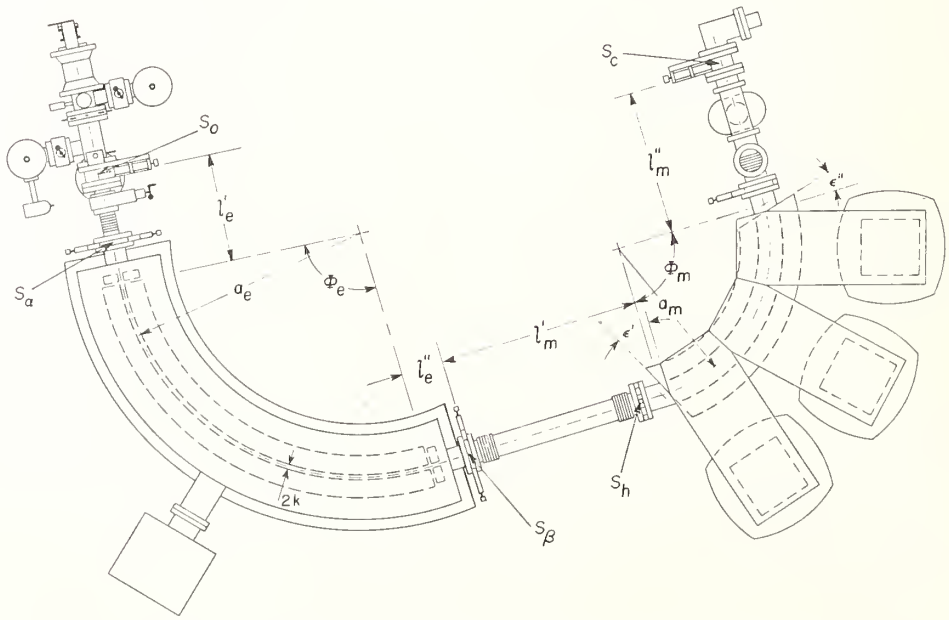


FIGURE 7. *Schematic of the University of Manitoba double-focusing mass spectrometer. Electrostatic analyzer radius $a_e = 1.00$ m; magnetic analyzer radius $a_m = 0.6274$ m; $\Phi_e = 94.65°$; $\Phi_m = 90°$.*

A recent mass spectrum of mercury chloride showing the performance of this instrument is illustrated in Fig. 9. A full width at half maximum resolution of 1 in 945,000 was achieved for this spectrum. Although this instrument is only about 50% larger than the Minnesota instrument, various advances, particularly the full second order focusing and improved transmission, lead to a working resolution almost 5 times higher than the Minnesota instrument.

I will next consider an entirely different mass measuring device which employs differences in transit time rather than spatial separation to form a mass spectrum. In order to describe the current instrument it is necessary to review some of the history of this form of mass spectrometer. In the early 1950s, Lincoln Smith at Brookhaven National Laboratory designed and constructed an instrument which was based on the dependence of the cyclotron frequency on the mass-to-charge ratio of the ion. Smith named this instrument the mass synchrometer [2]. A schematic of this instrument is shown in Fig. 10. In this instrument, ions were produced and accelerated in an ion source within a uniform magnetic field. After a half revolution they passed through a modulator which was driven by a radio-frequency voltage. The change in velocity of the ion produced by the modulator and consequent change in radius of the ion path cause a spacial modulation of the beam across slit S_5. Those ions which pass through S_5 continue and again pass through the modulator. If the phase of the rf voltage is correct when the group of ions again reaches the modulator, the group will again be decelerated and will continue and pass through slit S_6. The frequency condition is that the radiofrequency $f_{rf} = nf_0$ where n is the order integer (in practice) about 100 and f_0 is the cyclotron frequency. The process then is repeated for a third orbit in the magnetic field. Ions which pass through slit S_7 are collected. Perturbing the rf frequency by a sawtooth Δf will sweep the ion beam across the collector slit and produce an ion beam shape similar to more conventional mass spectrometers. Using a variation of Bleakney's theorem, one can measure a mass difference by measuring the change in f_{rf} needed to bring first one and then the other member of the mass doublet through the collector slit. Smith found three passes through the modulator were necessary to remove "ghost" signals at the collector. An instrument was constructed according to this design with a mean orbit diameter of 10 in. and a slit spacing of 2/3 in. This instrument produced a fullwidth at half-height resolution of up to one part in 32,000 and a mass measurement accuracy of one part in 2×10^7 and was competitive or exceeded the accuracy of more conventional instruments of that time. A summary of results from this instrument was published in 1958 [13].

The success of this instrument led Smith to develop a better design which was originally described in 1960 [14]. Design improvements included the following changes:

(a) The ion source and collector were moved outside of the magnetic field.

(b) The ion beam within the magnet was required to have a small axial velocity producing a helical path.

(c) The rf modulation was employed in a different manner than in the mass synchrometer in order to produce both better transmission of ions *and* higher resolution.

The design was refined and construction began at Princeton University. By the spring of 1966 most of the major parts of the instrument were completed and the remainder of that decade was occupied in testing and performing initial measurements [15].

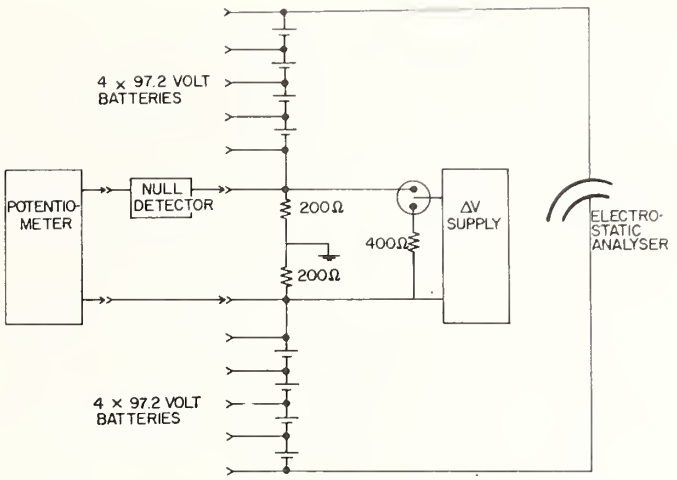


FIGURE 8. *Circuit used to generate the electrostatic analyzer voltage wave form for the University of Manitoba mass spectrometer.*

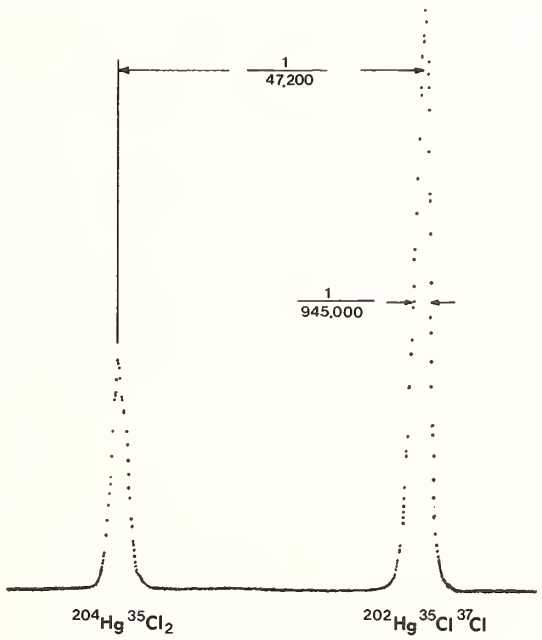


FIGURE 9. *Resolution performance of the University of Manitoba mass spectrometer.*

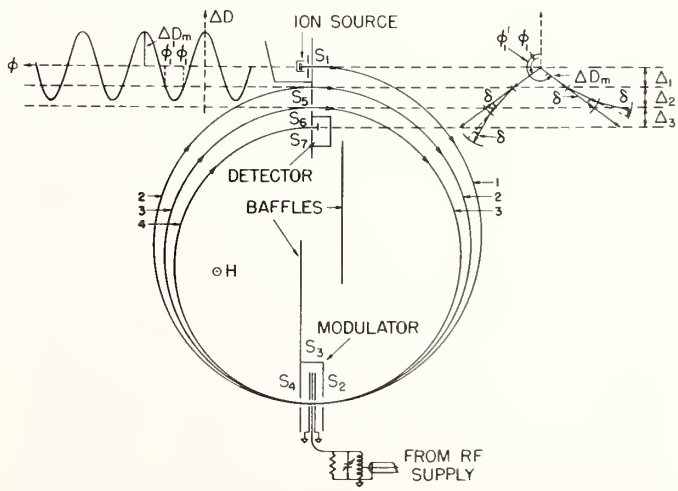


FIGURE 10. *A schematic diagram of the mass synchrometer.*

A schematic diagram of the magnetic field region of the instrument is shown in Fig. 11. Smith decided that instead of reducing the velocity of the ions on each passage through the modulator, as in the mass synchrometer, higher intensity could be achieved by a reduction of the velocity on the first pass and a compensating increase on the second pass through the modulator. The upper part of this figure shows an axial view of the ion beam path through the magnetic field. Ions first pass through an internal electrostatic deflector which changes their motion from radial to tangential in the magnet. The ions are also given a small axial velocity so that the subsequent helical path allows the introduction of various beam defining and focusing devices without any problem. After a half rotation, the beam passes through the modulator where its velocity and therefore diameter is modulated sinusoidally. After a second half rotation the beam passes through a phase defining slit and then back through a modulator another half rotation later. The change in diameter ΔD_1 at the first modulator is now removed by modulation ΔD_2 which is opposite to ΔD_1. If this occurs correctly, then the beam moves another half rotation and passes through a collector slit and is deflected out of the magnet by a second internal electrostatic deflector. A side view in the lower part of Fig. 11 indicates how the path in the magnet proceeds. Figure 12 shows the complete instrument. Emerging ions from an ion source pass into a pair of 90° toroidal electrostatic analyzers which form an image just outside the magnet. This image is used as the object for the internal deflector which forms an image on the entrance slit. The beam then proceeds through the path in the magnet. If the radio frequency is correct, the ion beam passes through the exit slit. An image of the exit beam is formed by the internal deflector which in turn is the object for another pair of 90° torroidal electrostatic analyzers. The detector is located at the focus of this analyzer.

Velocity and therefore diameter modulation is shown in Fig. 13. For the ion to pass through the exit slit, the velocity modulation V_1 on the first pass through the modulator must be balanced by the velocity modulation V_2 caused on the second pass through the modulator. In this diagram one sine wave describes V_1 and the other describes $-V_2$. Conditions are correct for transmission at times when the two sine waves cross. If, however, the two sine waves coincide, then all ions will be transmitted. This will occur when $f_{rf} = (n + 1/2)f_0$ where f_0 is the cyclotron frequency. By sweeping f_{rf}, a mass spectrum is produced as shown in the middle diagram of Fig. 13. Smith noted that the tails of the ion beam shapes were produced by ions which received large velocity changes. The tails could be removed by a phase defining slit that limits the change in velocity to a fraction of the maximum. This is shown by the band in the top view indicated by dotted lines in Fig. 13. The result of this limiting is the triangular ion beam shape shown in the bottom view.

As with the mass synchrometer, peak matching is employed in this instrument to determine mass differences. Two radio frequencies are generated, f_1 and $f_2 = f_1 + \Delta f$. The frequencies f_1 and f_2 are phase locked and linearly swept a small amount so as to present a time varying ion beam shape as the output of the detector. These frequencies cover the range from 50 MHz to 500 MHz. The frequency increment, Δf, is derived from a frequency synthesizer which can be set to 0.1 Hz. All voltages which produce dc fields in the system are alternated according to the Bleakney's theorem requirements.

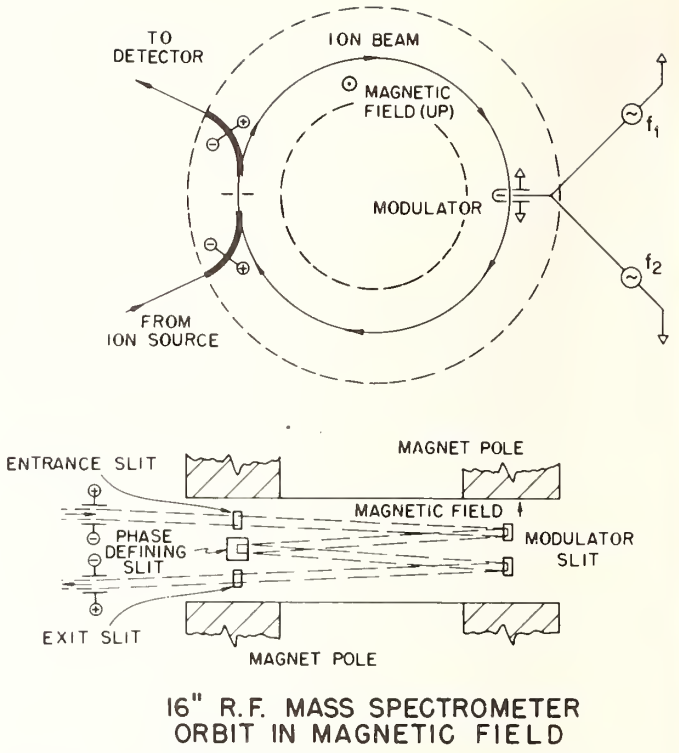


FIGURE 11. *A schematic diagram of the magnetic orbit for the rf mass spectrometer.*

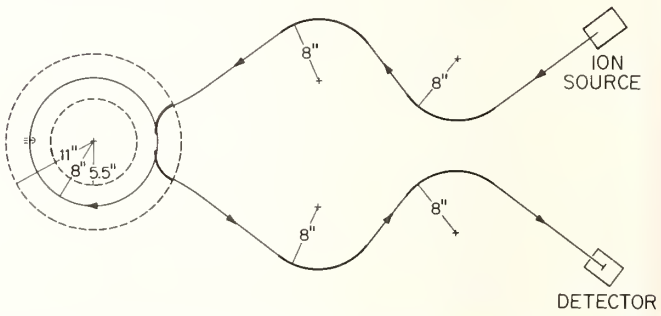


FIGURE 12. *A schematic diagram of the complete ion path in the rf mass spectrometer.*

Typical measurement conditions were:

Accelerating voltage	15 to 23 kV
order	$n = 800 - 1200$
frequency	140 to 360 MHz
rf voltage	100 to 150 V
resolution	one part in 200,000.

During the latter part of the 1960s, much effort was expended, mostly by Smith alone, to achieve the theoretical resolution of the instrument. Many of the difficulties he experienced were caused by electric fields resulting from charged surfaces, a difficulty which influences conventional mass spectrometers also. These fields were not switched as required by Bleakney's theorem and thus added systematic errors to the measurement. By use of baffles and insulator shielding, these effects were minimized to the point that measurements of narrow doublets could be made. These were reported in 1971 [16] in the paper "Measurement of Six Light Masses." During the spring of 1972 Smith's health began to fail and his last

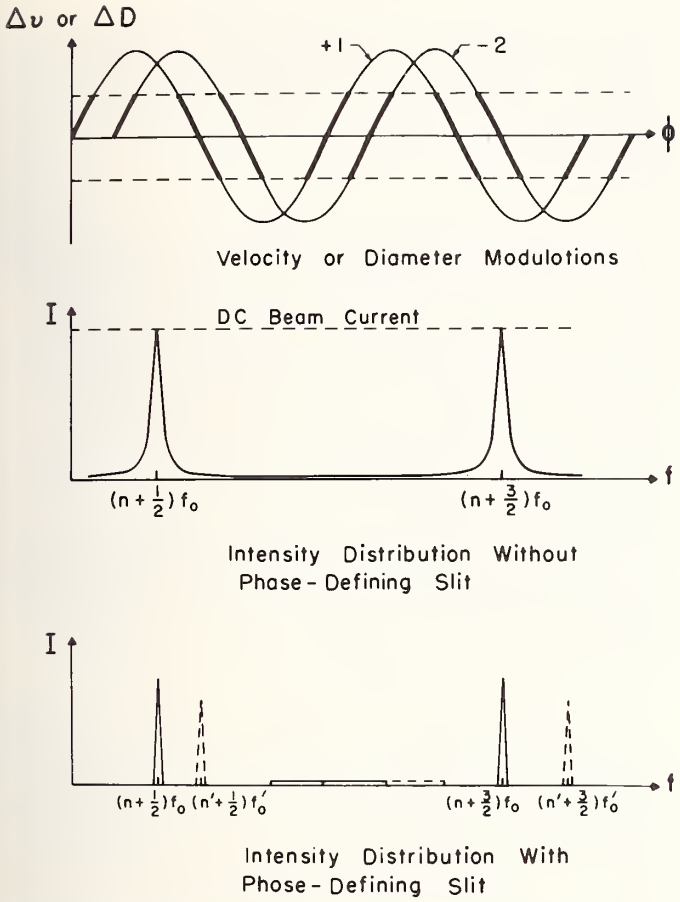


FIGURE 13. *Velocity and diameter modulation in the rf mass spectrometer.*

working days on the instrument were July 17 and 18. At this time he was repeating some of the light mass measurements, now with an instrument having a working resolution of one part in 400,000. Smith died in December 1972, leaving an instrument with a great deal of promise but with no program or no individuals who had recently worked with it.

Concern for the continuity of this project was great and finally a new location for the instrument was found at the Delft University of Technology in the Netherlands under the sponsorship and direction of A. H. Wapstra, J. B. LePoole, and E. Koets. The instrument was moved and was put back into operation officially on May 27, 1976. The Delft University of Technology was a logical choice for the new location because of previous ion optical experience in the design of electron microscopes. In the re-assembly of the instrument, a number of major improvements have been made:

(a) The instrument was re-aligned and a number of construction errors were eliminated.

(b) A higher-speed modern vacuum system was added. This improvement produced better vacuums and an oil-free system.

(c) Pump oil contamination of interior surfaces was eliminated. This was very likely a major cause of Smith's earlier charging problems.

(d) The vacuum housing and magnet were made more rigid to minimize vibration problems.

(e) Many sensing and focusing elements were added to study and adjust the beam for optimum passage through the system.

(f) A Telecentric Object Plane Selector was designed to study the beam cross-section. Any object plane from 0 to 2000 mm ahead of this device is focused with constant magnification on a channel plate array and a magnified image of the beam cross-section may be viewed on a screen. By inserting this system at various points in the orbit, detailed study and adjustment of the ion beam cross-section is possible.

(g) Improvements in the homogeneity of the magnetic field together with improvements in the stability of the magnet power supply were made.

(h) Higher rf voltages were used.

The reassembly and the various improvements have been described by Koets in several papers [17, 18]. The instrument now is functioning in a much more predictable manner. Charging effects appear still to contribute to instability but at a much lower level. The resolution achieved is as high as one part in 10^7 which is the highest currently reported by any method. I understand that a measurement program will begin sometime this year.

4. Recent Results

Because of the use of hydrocarbon molecular ions and fragments as known masses in doublets, the mass of hydrogen and a few other masses such as ^{16}O, ^{14}N, ^{13}C, ^{2}H, ^{35}Cl, and ^{37}Cl are considered as secondary standards and for this reason have received a great deal of attention. Precision attained in these measurements is high and thus they are a good example of what precision can be achieved with effort. The masses of selected secondary standard atoms can be determined from the measurement of an over-determined set of mass doublets. Smith [16] in 1971 measured the six masses ^{1}H, ^{2}H, ^{14}N, ^{16}O, ^{35}Cl, and ^{37}Cl in terms of 14 distinct doublets. The adjusted mass of ^{1}H for example was determined to be

$$1.007\ 825\ 029 \pm 5\ \text{u}\ .$$

The hydrogen mass can also be determined by direct comparison to ^{12}C. Smith [16] employed the doublet $C_9H_{10} - C_{10}H_8$ while Benson [19] used the doublet $C_{11}H_{22} - C_{12}H_{10}$.

In addition to the secondary standard measurements, a large number of doublet measurements have been reported in the past few years which extend from light atoms to the heaviest. Fractional error estimates, σ/M, for these mass measurements approach one part in 10^8 and for some recent measurements surpass this figure. Fractional errors for the doublets themselves, $\sigma/\Delta M$, for doublets with ΔM less than about 0.1 u have ranges from 10 to 20 ppm to those that approach the 1 ppm level.

It has been the practice in atomic mass measurements to quote an error which represents the precision of the measurement. Measurements of the same doublet are done a number of times and the error quoted is the standard deviation of the distribution of these measurements. Although estimates of possible systematic errors are not usually included as part of the quoted error, one can determine that certain forms of systematic error are small by means of several measurement procedures. One possibility is to measure a wide doublet which yields a known result such as $C_3H_8 - C_3H_7$. The ΔM here should be one hydrogen mass and thus any proportional error will be indicated by a departure from the known value. These proportional errors are often caused by surface charging within the instrument which produce unaccounted-for electric fields. Another means is to measure a set of masses by a series of doublets which

over-determine the results. Consistency can then be checked by using a χ^2 test.

The most extensive measure of consistency of mass doublet measurements has been done by Wapstra [20] in the preparation of comprehensive least-squares mass tables. These tables incorporate mass doublet information as well as nuclear reaction energy data in order to develop a best-value set of atomic masses. The tables, which have been revised periodically to incorporate new information, have become the accepted source of atomic mass information. The comparison of the subset of doublet measurements from one laboratory with output values can measure the consistency of the error estimation from that laboratory. Wapstra finds consistency factors from 2.5 to 5 for various mass spectrometer measurements [21].

A new feature in atomic mass determinations is the recent effort to apply the techniques developed for the measurement of stable atoms to the measurement of short-lived radioactive atoms. A group at Orsay, France, originally under the leadership of Rene Bernas and now Robert Klapisch, have made many mass spectroscopic measurements of various nuclear properties including the atomic mass. Atomic mass measurements have recently been reported for isotopes of francium [22], rubidium and cesium [23], and sodium [24]. Progress in the development of measurement techniques in this area has been reported by Barber [25] and by Smith [26].

5. Use of Mass Difference Measurements for Fundamental Constants

In the past, wavelength-energy and energy-mass relationships have been employed to transfer γ-ray and nuclear reaction information to equivalent mass differences. In this way a large and independent source of mass difference information could be added to conventional mass spectrometer doublet results. Most nuclear reaction and γ-ray energy information has been based on relative measurements, using a particular disintegration energy as a standard. For charged particle reactions, a variety of natural alpha-emitters have been employed as standards. Absolute α-energy measurments by Rytz [27] have provided energy standards with uncertainties in some cases as small as 7 ppm. In a similar manner, γ-ray energy measurements depend on a few γ-emitters as standards. The need for better γ-ray standards was recognized by the IUPAP Commission on Atomic Masses and Fundamental Constants in 1971 and a subcommittee was appointed to make standards recommendations. The report of the subcommittee (Helmer *et al.* [28]) bases γ-ray energy standards for energies up to 6.1 MeV on the 411 keV line in ^{198}Au. The uncertainty of the 411 keV line was 2.7 ppm based on the wavelength measurement of Kessler *et al.* [29] using a double-flat crystal interferometer. Most of the uncertainty in this energy results from the wavelength-energy conversion using the value from Cohen and Taylor [1] and only a small amount, 0.37 ppm, from the wavelength measurement. This work has been reviewed by Deslattes and Kessler [29] and they suggest that this form of measurement may be extended to the wavelengths of γ-rays with energies of several MeV.

The availability of this type of measurement suggests that the energy-mass transformation may be reversed and measurements of an absolute wavelength and the equivalent mass difference could be used together to yield a value for a combination of fundamental constants. This possibility has been suggested by Deslattes and Kessler [30] and by Taylor [31]. Suppose that λ is the wavelength in meters of a γ-ray which has an associated mass difference Δm measured in u. By assuming the Planck relationship and the Einstein mass-energy relationship, the combination of fundamental constants given in Eq. (10) is equal to the product of the experimental measurements:

$$\frac{N_A h}{c} = \lambda \Delta m \ . \tag{10}$$

In order to estimate the precision needed for a useful $\lambda \Delta m$ product, one can determine the uncertainty in the fundamental constant product from Cohen and Taylor [1]. Using the method outlined by Cohen and Taylor to determine product uncertainties, the uncertainty for this product is 1.7 ppm. Measurements of both Δm and λ to part per million levels appear to be certainly possible.

As an example of this application, consider the thermal neutron capture in ^{12}C which produces a γ-ray with an energy of about 5 MeV. Measurements of this γ-ray energy together with the thermal neutron capture γ-ray in hydrogen at 2.2 MeV can be employed for the wavelength part of the measurement. The mass difference needed can be found using a series of mass doublets investigated by Halverson and Johnson [32]. The two reactions are given in Eqs. (11) and (12)

$$^{12}\mathrm{C} + \mathrm{n} \rightarrow {}^{13}\mathrm{C} + \gamma_1 \tag{11}$$

$$^{1}\mathrm{H} + \mathrm{n} \rightarrow {}^{2}\mathrm{H} + \gamma_2 \ . \tag{12}$$

These equations can also be written as energy balance equations if all elements are stated in the same mass-energy unit:

$$m(^{12}\mathrm{C}) + m(\mathrm{n}) = m(^{13}\mathrm{C}) + Q_{\gamma_1} \tag{13}$$

$$m(^{1}\mathrm{H}) + m(\mathrm{n}) = m(^{2}\mathrm{H}) + Q_{\gamma_2} \ , \tag{14}$$

in which Q_{γ_1} and Q_{γ_2} are the total energies available in each reaction expressed in mass units. The masses employed in this equation are atomic masses with no corrections because the parent and daughter atoms in each reaction have the same electronic binding. The total energy for the reaction can be expressed in two parts, the γ-ray energy, E_γ, and a recoil energy, E_R, as in Eq. (15):

$$Q_\gamma = E_\gamma + E_R \ . \tag{15}$$

If one assumes that the recoiling atom is free, the recoil energy is given by

$$E_R = \frac{(hc/\gamma)^2}{2mc^2} \ . \tag{16}$$

For the ^{13}C emission, this recoil energy is about one part in 5000 of the total energy Q_γ. If Q_γ is written in atomic mass units as required by Eq. (13), we have

$$Q_\gamma = \frac{hc}{\lambda}\left[1 + \frac{hc/\lambda}{2mc^2}\right]\frac{N_A}{c^2} \ . \tag{17}$$

Although the correction for recoil contains fundamental constants, the correction can be made with nominal values because the correction is small.

The reaction energy, Eqs. (13) and (14), can be converted to a mass difference by rearrangement. Assume that all elements in Eq. (13) and Eq. (14) are given in unified mass units. First subtract Eq. (14) from Eq. (13):

$$^{12}\mathrm{C} - {}^{1}\mathrm{H} = {}^{13}\mathrm{C} - {}^{2}\mathrm{H} + Q_{\gamma_1} - Q_{\gamma_2} \tag{18}$$

or

$$^{12}\mathrm{C}^{2}\mathrm{H} - {}^{13}\mathrm{C}^{1}\mathrm{H} = Q_{\gamma_1} - Q_{\gamma_2} \ . \tag{19}$$

Equation (19) can be converted to available molecules by multiplying by six:

$$^{12}C_6{}^2H_6 - {}^{13}C_6{}^1H_6 = 6(Q_{\gamma_1} - Q_{\gamma_2}) \ . \qquad (20)$$

These two molecules in this mass doublet are both benzene and thus have the same atomic binding energy which cancels out of the mass difference. This doublet with C_6H_{12} is part of an over-determined mass triplet shown in Fig. 14. This triplet has been employed by Halverson and Johnson [32] to determine a value of the $^{12}C_6{}^2H_6 - {}^{13}C_6{}^1H_6$ doublet.

The doublet difference Δm, measured in u, is one part of the input to determine the fundamental constant product. The other part, $Q_{\gamma_1} - Q_{\gamma_2}$, can be simplified by assuming the recoil correction for reaction i to be $(1 + a_i)$. Then

$$Q_{\lambda_1} - Q_{\lambda_2} = hc \left[\frac{\lambda_2 - \lambda_1}{\lambda_1 \lambda_2} + \frac{a_1\lambda_2 - a_2\lambda_1}{\lambda_1\lambda_2} \right] \frac{N_A}{c^2} \ . (21)$$

Substituting this value of $Q_{\gamma_1} - Q_{\gamma_2}$ into Eq. (20) yields

$$\frac{\Delta m}{6} \left[\frac{\lambda_2 - \lambda_1}{\lambda_1\lambda_2} + \frac{a_1\lambda_2 - a_2\lambda_1}{\lambda_1\lambda_2} \right]^{-1} = \frac{hN_A}{c} \qquad (22)$$

The left side of Eq. (22) now involves measured quantities and the right side is the desired fundamental constant product.

In principle any Δm could be used for this comparison. There are, however, a number of practical considerations that make some choices better than others.

a. The neutron capture target should be an abundant isotope.

b. The mass doublet that results should be made up of molecular ions if possible.

c. The γ-ray energy should be linked to an absolute wavelength measurement as closely as possible.

d. The γ-ray energy should be large compared with the $^1H(n, \gamma)^2H$ reaction energy.

Table 1 lists several possible choices of atoms for this test. Included is the calculated recoil energy correction. The large $^{14}N(n, \gamma)^{15}N$ reaction energy results in a low fraction error for the Δm measurement. A doublet could be constructed similar to that found for the carbon based measurement using molecules of hydrazine. The large energy, however, would make the determination of λ for the emitted γ-ray more difficult.

Another possibility of choice for a reaction would be the decay of a long lived metastable state. The experience developed by groups measuring the masses of radioactive atoms would be useful for this form of measurement. Some possible choices are listed in Table 2. Of these the ^{92}Mo isomeric transition looks most favorable. The doublet has a spacing of one part in 39,000 and is wide enough so that a 1 ppm mass difference measurement is conceivable. This form of reaction would provide a more direct comparison of Δm and λ but will still have a small recoil correction. (The recoil correction for the ^{93}Mo decay is only 36 eV.)

In conclusion, it appears that from both the mass-difference side and the wavelength side, precisions can be achieved at present or in the near future to produce useful fundamental constants results.

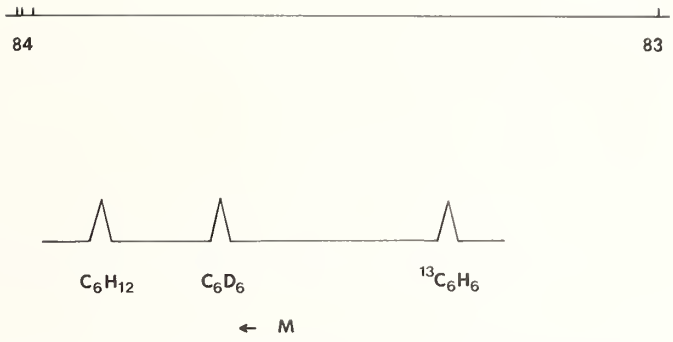


FIGURE 14. *The mass triplet employed to determine the $C_6{}^2H_6 - {}^{13}C_6{}^1H_6$ mass difference. The resolution shown in the lower portion is one part in 10^5.*

TABLE 1. *Possible reactions for fundamental constants tests.*

Reaction	E_γ in μu	Recoil energy μu
$^{12}C(n,\gamma)^{13}C$	5310.067 ± 23[a] (5 ppm) 5309.978 ± 63[b] (13 ppm)	1.084
$^{13}C(n,\gamma)^{14}C$	8777.746 ± 26[a] (3 ppm)	2.751
$^{14}N(n,\gamma)^{15}N$	11,629.936 ± 28[a] (2.5 ppm)	4.507

[a]See Ref. [33].
[b]See Ref. [32].

TABLE 2. *Possible isomeric transitions for fundamental constants tests.*

Isotope	Half-life	Energy (MeV)
^{89}Y	16 s	0.91
^{93}Mo	7 h	2.43
^{111}Cd	48 min	0.40
^{137}Ba	2.6 min	0.66

I wish to thank Robert Barber, University of Manitoba, and Erico Koets, Delft University of Technology, for their significant assistance in providing illustrations and other information for this talk.

References

[1] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[2] J. H. E. Mattauch, Adv. Mass Spectrom. 3, 1 (1965).
[3] E. G. Johnson and A. O. Nier, Phys. Rev. 91, 10 (1953).
[4] W. Bleakney, Am. Phys. Teacher 4, 12 (1936).
[5] W. F. G. Swann, J. Franklin Inst. 212, 439 (1931).
[6] L. G. Smith and C. C. Damm, Phys. Rev. 90, 324 (1953).
[7] C. F. Giese and T. L. Collins, Phys. Rev. 96, 823A (1954).
[8] L. G. Smith, Phys. Rev. C 4, 22 (1971).

[9] H. E. Duckworth, Rev. Sci. Instrum. **21**, 54 (1950).
[10] H. Hintenberger and L. A. Konig, Adv. Mass Spectrom. **1**, 16 (1959).
[11] R. C. Barber, R. L. Bishop, H. E. Duckworth, J. O. Meredith, F. C. G. Southon, P. van Rookhuyzen, and P. Williams, Rev. Sci. Instrum. **42**, 1 (1971).
[12] L. G. Smith and C. C. Damm, Rev. Sci. Instrum. **27**, 638 (1956).
[13] L. G. Smith, Phys. Rev. 111, 1606 (1958).
[14] L. G. Smith, in *Proceedings of the International Conference on Nuclidic Masses*, Ed. by H. E. Duckworth (Univ. of Toronto Press 1960), p. 418.
[15] L. G. Smith, in *Proceedings of the Third International Conference on Nuclidic Masses*, Ed. by R. C. Barber (Univ. of Manitoba Press, 1967), p. 811.
[16] L. G. Smith, Phys. Rev. C 4, 22 (1971).
[17] E. Koets, in *Atomic Masses and Fundamental Constants 4*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, New York, 1976), p. 164.
[18] E. Koets, J. Kramer, J. Nonhebel, and J. B. LePoole, in *Atomic Masses and Fundamental Constants 6*, Ed. by J. A. Nolen, Jr., and W. Benenson (Plenum Press, New York, 1980), p. 275.
[19] J. L. Benson and W. H. Johnson, Jr., Phys. Rev. **141**, 1112 (1966).
[20] A. H. Wapstra and K. Bos, At. Data Nucl. Data Tables **19**, 175 (1977).
[21] A. H. Wapstra and K. Bos, At. Data Nucl. Data Tables 20, 1 (1977).
[22] G. Audi, M. Epherre, C. Thibault, R. Klapisch, G. Huber, F. Touchard, and H. Wollnik, in *Atomic Masses and Fundamental Constants 6*, Ed. by J. A. Nolen, Jr., and W. Benenson (Plenum Press, New York, 1980), p. 281.
[23] M. Epherre, G. Audi, C. Thibault, and R. Klapisch, in *Atomic Masses and Fundamental Constants 6*, Ed. by J. A. Nolen, Jr., and W. Benenson (Plenum Press, New York, 1980), p. 299.
[24] C. Thibault, M. Epherre, G. Audi, R. Klapisch, G. Huber, F. Touchard, D. Guillemaud, and F. Naulin, in *Atomic Masses and Fundamental Constants 6*, Ed. by J. A. Nolen, Jr., and W. Benenson (Plenum Press, New York, 1980), p. 291.
[25] R. C. Barber, K. S. Kozier, K. S. Sharma, V. P. Derenchuk, R. J. Ellis, V. S. Venkatasubramanian, and H. E. Duckworth, in *Atomic Masses and Fundamental Constants 6*, Ed. by J. A. Nolen, Jr., and W. Benenson (Plenum Press, New York, 1980), p. 257.
[26] R. J. Smith, Ph.D. Thesis, Univ. of Minnesota (1977).
[27] B. Grennberg and A. Rytz, Metrologia 7, 65 (1971).
[28] R. G. Helmer, P. H. M. Van Assche, and C. vanderLeun, At. Data Nucl. Data Tables **24**, 39 (1979).
[29] E. G. Kessler, Jr., R. D. Deslattes, A. Henins, and W. C. Sauder, Phys. Rev. Lett. **40**, 171 (1978).
[30] R. D. Deslattes and E. G. Kessler, Jr., in *Atomic Masses and Fundamental Constants 6*, Ed. by J. A. Nolen, Jr., and W. Benenson (Plenum Press, New York, 1980), p. 203.
[31] B. N. Taylor, private communication (1981).
[32] J. E. Halverson and W. H. Johnson, Jr., Phys. Rev. C **17**, 1414 (1978); Errata, ibid., 18, 2450 (1978).
[33] L. G. Smith and A. H. Wapstra, Phys. Rev. C **11**, 1392 (1975).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Absolute Determination of the Threshold Energies of ^{7}Li(p,n), ^{10}B(p,n), and ^{14}N(p,n)

P. H. Barker, M. J. Lovelock, H. Naylor, R. M. Smythe, and R. E. White

Physics Department, Auckland University, New Zealand

The present status of the Heavy Ion Source System (HISS) for the measurement of nuclear reaction energies is discussed, and some results are presented. In particular, accurate values for the threshold energies of ^{7}Li(p,n), ^{10}B(p,n), and ^{14}N(p,n) are given.

Key words: heavy ion calibration; nuclear reaction energies; one volt standard; (p,n)-thresholds.

1. Introduction

The study of the theoretical implications of the ft values of super-allowed beta decays would be placed on a firmer basis if the decay Q-values (Q_β) were known more precisely. At present, they carry error bars which, on the average, are about 0.5 keV in several MeV. Our prime aim therefore has been to develop a system to improve this situation considerably, to show the reliability of the system, and to use it to determine as many of the Q_β as possible.

In the present work a Q_β is derived from measurement of the threshold of the associated (p,n) reaction. The proton beam from our folded tandem accelerator, AURA II, is momentum-analyzed by requiring it to pass through very narrow slots at the object and image positions of an Enge split-pole spectrograph. After the image slot, the beam is used to examine the yield of a particular (p,n) reaction near threshold by applying a variable offset voltage to the target. The shape of the yield curve is then analyzed in terms of the familiar $(V - V_0)^{3/2}$ energy dependency for a thick target, to give V_0, the offset voltage at threshold.

To measure the mean proton energy, a subsidiary surface ionization source is used to produce ions of a heavy element, usually thallium, which are then accelerated through a potential difference to pass down the same flight path through the spectrograph as the proton beam had done. This potential difference is measured using a resistive divider of our own construction and a one volt standard. The magnetic rigidity of the thallium ions, and hence of the proton beam, is then calculated from this measured potential difference, with small corrections due to contact potential and thermal effects. Thus the kinetic energy of the protons, E_p, is given essentially in terms of the thallium accelerating potential and the masses of a proton and a thallium ion.

The threshold energy, E_0 is then just given by $E_0 = E_p - V_0$. The system is described in detail in Ref. [1], with minor modifications and test results in Ref. [2].

2. The ^{7}Li(p,n) Threshold at 1.88 MeV

This reaction is of interest because the accepted threshold energy, $E_0 = 1880.51 \pm 0.08$ keV [3], carries a very small error and it is therefore a good test for our system. In addition, we will try to determine E_0 for ^{1}H(^{7}Li, n) to compare and look for possible atomic effects.

The yield curve of neutrons from ^{7}Li + p was examined six times with the spectrograph magnetic field set to make E_p about 1880 keV. A fresh ^{7}Li target was evaporated in situ for each run and the yield of neutrons was determined using a ^{6}Li − ZnS backed glass scintillator and photomultiplier with a perspex moderator. The beam intensity was monitored by observing the yield of the 480 keV prompt gamma ray from ^{7}Li + p. The magnetic rigidity of the proton path through the spectrograph was calibrated using rubidium ions of approximately 22 keV. Thallium could not be used because the required accelerating voltage is too low to be compared with our 1 volt standard.

In Fig. 1 is shown one of the neutron yield curves, the solid curve being the $(V - V_0)^{3/2}$ fit to the data points. The quoted errors for V_0 and E_p are typical for this experiment and are one standard deviation, as are all other errors quoted in this paper.

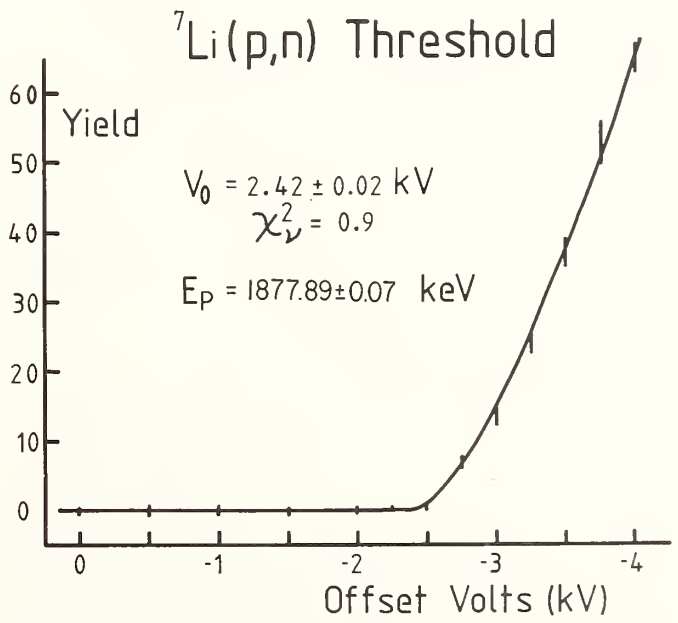


FIGURE 1. *A yield curve for ^{7}Li (p,n). The line is a theoretical least squares fit of the form $Y = a + b(V - V_0)^{3/2}$.*

The mean value for the six runs is $E_0 = 1880.41 \pm 0.05$ keV with $\chi^2_\nu = 1.0$. A small correction can be applied to take into account the proton energy distribution

(triangular with base width 0.4 keV). This brings the final threshold energy to 1880.42 ± 0.05 keV.

3. The ^{10}B(p,n) Threshold at 4.88 MeV

Although the decay of ^{10}C is one of the pure Fermi beta decays referred to in Section 1, the uncertainty in the ft value arises almost entirely from uncertainty in the branching ratio of the decay. Hence there is no strong reason to improve the measurement of the threshold energy. However, knowledge of the branching ratio may improve and, in addition, the two sub-keV measurements quoted in the literature, Refs. [4] and [5], disagree by more than 3 keV.

For this reaction, the yield of the 717 keV gamma ray which follows the ^{10}C decay was used as a measure of the reaction yield. The targets were of amorphous ^{10}B, evaporated with an electron gun on to thick tantalum backings. The experimental arrangement was such that the proton beam, after passing through the spectrograph defining slots, struck the target for 30 s, ($T_{1/2}{}^{10}C$ = 19 s), and then was interrupted upstream by a tantalum shutter. The target was pulled on a magnetically coupled plunger out of the spectrograph body, and the accumulated activity viewed with a Ge(Li) detector. The beam intensity was monitored by detecting protons scattered backwards from the target.

After a sufficient yield had been accumulated, normally after 10 beam cycles, the potential on the target was altered, and a yield curve like the one in Fig. 2 was taken from which a value for V_0 was extracted. The proton beam energy was calibrated using Tl ions of approximately 24 keV. The error in E_0 arises almost entirely from the error in V_0 due to the low yield of the reaction.

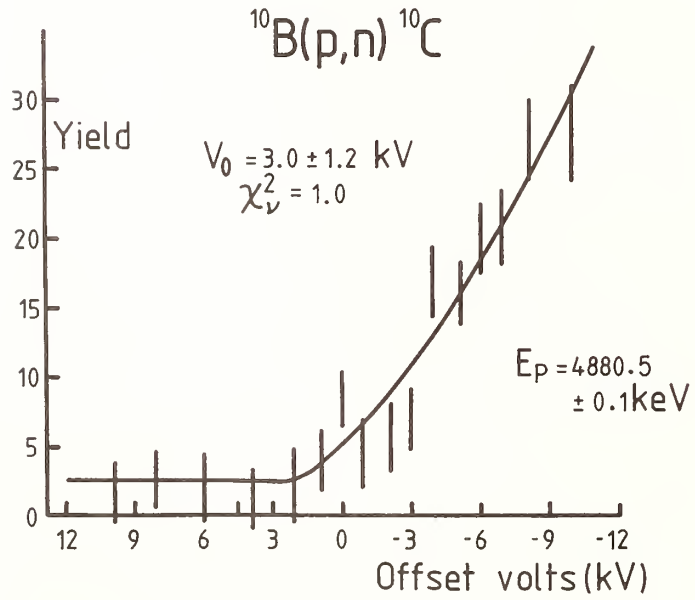


FIGURE 2. *A yield curve for $^{10}B(p,n)$. The line is a theoretical least squares fit of the form $Y = a + b(V - V_0)^{3/2}$.*

The mean of five such values was E_0 = 4876.4 ± 0.55 keV with χ^2_ν = 1.0. The correction for the beam energy shape (triangular with base width 0.8 keV) and for non-uniform proton energy loss in the target is +0.03 ± 0.03 keV, and so the final threshold energy is 4876.4 ± 0.6 keV, which is in good agreement with the value implied by the ^{12}C(p,t) Q-value of Ref. [4], 4876.4 ± 0.6 keV.

The error in the result of Ref. [5], the first Auckland (p,n) experiment performed with an earlier, more primitive set-up, we attribute to having used a target which was not freshly prepared, and which had indeed already been used by the Harwell group for an earlier determination. That target cleanliness and stability are crucial will become apparent in the following section.

4. The ^{14}N(p,n) Threshold at 6.35 MeV

This experiment is very similar to that of Section 3. The yield of the 2.3 MeV gamma ray following the decay of $^{14}O(T_{1/2}$ = 70 s) was measured. The proton beam energy was calibrated using Tl ions of approximately 31 keV. The beam-on/beam-off times were 100 s and 200 s respectively. Attempts were made to determine the ^{14}N(p,n) threshold energy using targets of melamine, nylon, tantalum nitride and titanium nitride. Of these, only melamine gave a yield which was sufficient, which did not diminish during the course of an experiment, and which did not have too large an associated contaminant activity.

In Fig. 3 are shown the results for the years 1979 and 1980 for those experiments for which no obvious reason could be found for their exclusion. They form a statistically inconsistent set. This we attribute to deterioration of the targets during the course of the runs. It is difficult to set up tests which are sensitive enough to detect these effects, and the best way seems to be to look for inconsistencies between runs done with the same target, and also between runs done with different targets.

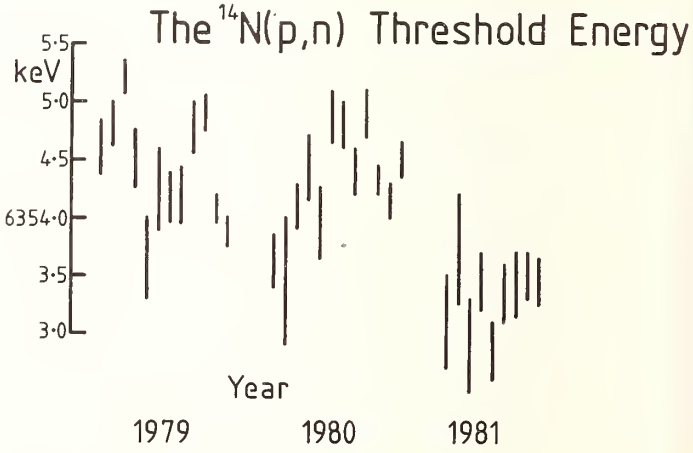


FIGURE 3. *Attempts in Auckland to measure E_0 for $^{14}N(p,n)$ in 1979, 1980, and 1981.*

The points in Fig. 3 which are indicated by the year 1981 were taken with targets of sputtered tantalum nitride on a tantalum substrate. We are confident that these do not deteriorate during use. Their yield is reproducible over periods of many hours. When judged visibly they look unaffected by the bombardment, and they give a self-consistent set of values for E_0.

In Fig. 4 is one of the nine ^{14}N(p,n) yield curves. As can be seen, the error in E_0 comes mainly from V_0, with a small contribution from E_p. In Table 1 are listed the individual results which combine to give a mean of E_0 = 6353.02 ± 0.08 keV, χ^2_ν = 0.8. After correction for the beam energy distribution, triangular with base width 1.05 keV, and for non-uniform proton energy loss in the target, a total of +0.02 ± 0.02 keV, the threshold becomes E_0 = 6.353.04 ± 0.08 keV.

This figure does not include any correction for atomic effects. Reference [13] recommends a correction of

−131 eV. However, consultation with one of the authors has indicated the figure is uncertain but should in any case be considerably smaller [14]. We have therefore chosen to leave it out.

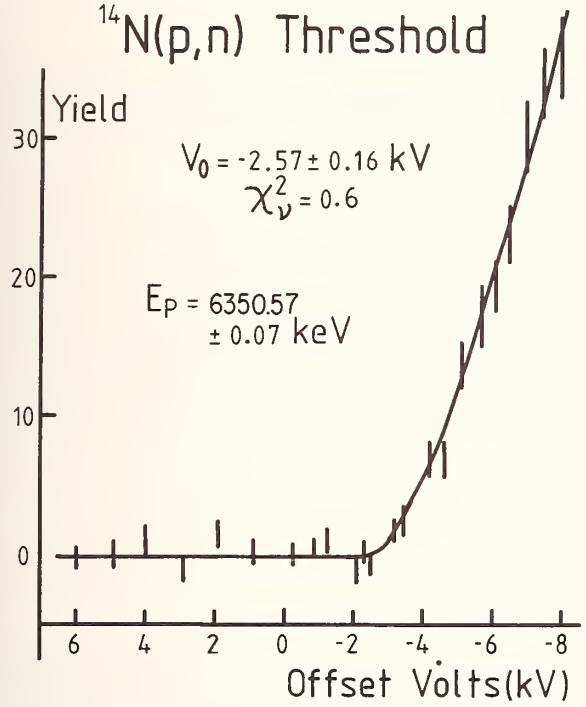


FIGURE 4. *A yield curve for $^{14}N(p,n)$. The line is a theoretical least squares fit of the form $Y = a + b(V - V_0)^{3/2}$.*

TABLE 1. *The set of nine threshold determinations for $^{14}N(p,n)$.*

E_p	V_0 (keV)	E_0 (keV)
6347.89(07)	−4.88(39)	6352.77(40)
6345.69(07)	−7.68(45)	16353.37(46)
6345.03(07)	−7.54(40)	6352.57(41)
6345.03(07)	−8.08(21)	6353.11(22)
6347.01(07)	−5.52(28)	6352.53(29)
6347.54(07)	−5.47(24)	6353.01(25)
6349.65(07)	−3.42(25)	6353.07(26)
6350.57(07)	−2.57(16)	6353.14(17)
6353.74(07)	+0.64(17)	6353.10(18)
	Mean (with $\chi^2_\nu = 0.77$)	6353.02(08)

Taking the excitation energy of the O^+, $T = 1$ state of ^{14}N to be 2312.81 ± 0.06 keV [6], and the latest value for the (n − ^{1}H) mass difference, 782.332 ± 0.017 keV [7], the Q_β for the superallowed decay of ^{14}O becomes Q_β = 1808.31 ± 0.10 keV. This is compared in Table 2 with other sub-keV measurements.

TABLE 2. *Q_β values for the ^{14}O superallowed decay.*

Q_β (keV)	Reference
1809.3 ± 0.7	8
1810.2 ± 0.5	9
1810.5 ± 0.5	10
1808.8 ± 0.4	11
1807.9 ± 0.8	12
1810.4 ± 0.6	4
1808.31 ± 0.10	Present result

References

[1] P. H. Barker, H. Naylor, D. P. Stoker, R. E. White, and W. B. Wood, in *Atomic Masses and Fundamental Constants 6*, Ed. by J. A. Nolen, Jr., and W. Benenson (Plenum, New York, 1980), p. 233.

[2] D. P. Stoker, P. H. Barker, H. Naylor, R. E. White, and W. B. Wood, Nucl. Instrum. Methods 180, 515 (1981).

[3] A. H. Wapstra and K. Bos, At. Data and Nucl. Data Tables 19, 177 (1977).

[4] P. H. Barker and J. A. Nolen, Tokyo Conference on Nuclear Physics, (1977).

[5] D. C. Robinson and P. H. Barker, Nucl. Phys. A 225, 109 (1974).

[6] F. Ajzenberg-Selove, Nucl. Phys. A 152, 1 (1970).

[7] R. C. Greenwood and R. E. Chrien, Phys. Rev. C 21, 498 (1980).

[8] J. W. Butler and R. O. Bondelid, Phys. Rev. 121, 1770 (1961).

[9] R. K. Bardin, C. A. Barnes, W. A. Fowler, and P. A. Seeger, Phys. Rev. 127, 583 (1962).

[10] M. L. Roush, L. A. West, and J. B. Marion, Nucl. Phys. A 147, 235 (1970).

[11] R. E. White and H. Naylor, Nucl. Phys. A 278, 333 (1977).

[12] H. Vonach, P. Glässel, E. Huenges, P. Maier-Komor, H. Rösler, H. J. Scheerer, H. Paul, and D. Semrad, Nucl. Phys. A 278, 189 (1977).

[13] J. M. Feagin, E. Merzbacher, and W. J. Thompson, Phys. Lett. 81B, 107 (1979).

[14] W. J. Thompson, private communication.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Preliminary Proton/Electron Mass Ratio Using a Precision Mass-Ratio Spectrometer*

R. S. Van Dyck, Jr. and P. B. Schwinberg

Department of Physics, University of Washington, Seattle, WA 98195

A new type of compensated Penning trap has its ring electrode split into equal quadrants in order to synchronously detect ion cyclotron resonances. Using this device, we have observed intense unshifted resonances with very small relative linewidths ($< 2 \times 10^{-9}$), comparable to that which is attainable in high resolution NMR studies. Thus, cyclotron frequencies of both protons and electrons have been measured in the same magnetic field (5 T) and the same trapping volume ($< 10^{-7}$ cm^3). From the ratio of these frequencies, a preliminary value of $m_p/m_e = 1836.15300(25)$ is obtained with a relative uncertainty of 0.14 ppm. This value agrees well with previous direct measurements within their experimental uncertainties.

Key words: compensated quadrupole trap; fundamental constants; ion cyclotron resonances; ion trapping; mass spectrometer; precision measurements; quadring Penning trap; synchronous resonant detection.

1. Introduction

The proton/electron mass ratio is a fundamental constant which appears often in specific theories of atomic or elementary particle physics in the form of a reduced mass effect or an isotope shift. An illustration of the need for precision in this mass ratio can be found in the recent comparison of theory with a precision measurement of the Lyman-α resonances of atomic hydrogen and deuterium [1]. Also, a 0.01 ppm determination of m_p/m_e will allow an unambiguous consistency check to be made upon many other precisely measured fundamental constants [2]. Previously, the most precise value of m_p/m_e was 1836.1518(7), indirectly determined from other precision constants [3]. This can be compared to direct measurements of 1836.1502(53) by Gärtner and Klempt [4] and 1836.1527(11) obtained by Gräff *et al.* [5] using a time-of-flight measurement in an external inhomogeneous magnetic field in order to detect resonant cyclotron excitation of an ion prior to ejection from a Penning trap. The only theoretical value of the mass ratio is a speculation by Wyler [6] that suggests that $m_p/m_e = 6\pi^5 = 1836.11811$ which is many standard deviations away from several other indirectly measured values of m_p/m_e [7] as well.

The compensated Penning trap [8] has been demonstrated to be a very precise resonance spectrometer in the measurement of the spin anomaly of a single electron [9]; thus, it should be possible to use this type of trap to isolate single ions and directly measure their ion cyclotron resonances [10, 11]. The simple result of comparing respective cyclotron frequencies then yields directly the mass ratio (assuming identical charges). However, the motion of a charged particle in a Penning trap slightly modifies this simple picture. The charge oscillates axially in the electric potential well with a frequency ν_z while rotating in the radial plane at the observed cyclotron frequency ν_c'. In addition, the crossed electric and magnetic fields produce a slow unbounded, but metastable, rotation of the center of the cyclotron orbit at the magnetron frequency ν_m. From the equations of motion for a single ion of mass m and charge e isolated within a Penning trap,

$$\nu_z = (1/2\pi)\,(eV_o/mZ_o^2)^{1/2},$$
$$\nu_c' = \nu_c - \nu_m,$$
$$\nu_m = \nu_z^2/2\nu_c',$$

where

$$\nu_c = eB_o/2\pi mc,$$

B_o is the applied magnetic field, $V_o/2$ is the well depth of the trap, and $2Z_o$ is the characteristic endcap separation (at $r = 0$). The magnetron frequency, used to correct the observed cyclotron frequency, can be either calculated using ν_z and ν_c' or measured using a motional sideband [12] such as $\nu_z + \nu_m$ or $\nu_c' + \nu_m$. Such cooling or centering resonances for the magnetron motion have been observed by us and described previously [9]. For a single electron in a recent trap, the measured and computed values have agreed to within a few ppm. Such agreement is crucial in order to justify using this type of resonance method.

2. Quadring Trap

The experimental apparatus, shown in Fig. 1, consists of a trap tube supported at the center of a 30 cm long, very stable (< 0.002 ppm/hr), and reasonably uniform (< 3 ppm over a 1-cm sphere) superconducting solenoid. The "quadring" Penning trap, shown only schematically in Fig. 1, is illustrated in detail in Fig. 2. In addition to the two endcap and ring electrodes with characteristic dimensions $Z_o = 0.112$ cm and $R_o = 0.159$ cm, the trap has two guard rings which compensate for the truncations of the main surfaces, one probe inserted into each guard for the centering drive, and four equal cuts in the hyperbolic ring electrode to form the "quadring." This design [11] allows the cyclotron resonance to be excited on two oppo-

*Supported in part by NBS Precision Measurements Grant No. G7-9023, in part by a M. J. Murdock Charitable Trust Grant of the Research Corporation, and in part by NSF Grant No. PHY-80-15328.

site quadrants and detected on the other pair which are externally tuned to the cyclotron frequency. The two axial frequencies, $\nu_z(e^-)$ and $\nu_z(p^+)$, are also detected using a tuned circuit resonant at the appropriate frequency and attached to one or the other endcap. Also shown in Fig. 1 is the Schottky multiplier diode which generates the microwave field required to excite the electron cyclotron resonance at $\nu_c'(e^-)$. Finally, the entire apparatus consisting of trap tube, microwave diode and tuned preamplifiers is constructed to operate in a liquid helium environment in order to yield the required signal-to-noise and high resolution.

Included in each endcap is a small nickel ring (volume $\sim 10^{-3}$ cm^3) and a source of either electrons or hydrogen. The electron source is a sharpened tungsten needle which

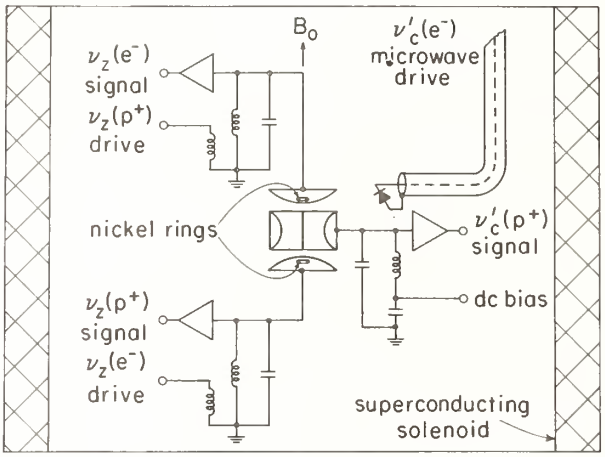


FIGURE 1. *Schematic of experiment. Both endcaps and the quadring have tuned preamplifiers attached to measure either ν_z or ν_c. rf axial drives are applied to the endcap opposite the detector and microwave excitation for $\nu_c(e^-)$ radiates from a Schottky multiplier diode.*

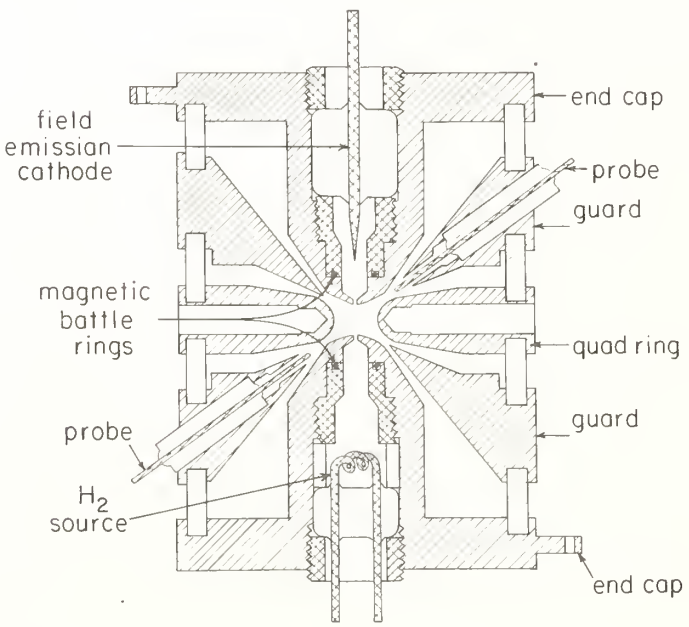


FIGURE 2. *Quadring Penning trap. The standard compensated Penning trap is altered by splitting the main ring electrode into four equal quadrants in order to synchronously detect ion cyclotron resonances directly.*

can be biased to field emit a primary electron beam through the trap and generate within the trapping volume both secondary electrons and ions from either background gas or hydrogen liberated from a hydrated titanium filament in the opposite endcap. The nickel rings produce a very weak magnetic bottle [9] which generates a change in the uniform magnetic field given by

$$B_2(z^2 - r^2/2)$$

where

$$B_2 \sim 350 \text{ G/cm}^2 .$$

The electrodes themselves may increase this coefficient by 50% due to an intrinsic bottle field formed by the paramagnetism of the molybdenum used to make the electrodes [13]. The net effect of this magnetic bottle is to couple the magnetic moment of the charge to the axial resonance in the form of a small frequency shift, 2–3 Hz per quantum level for electrons ($\nu_z \sim 86$ MHz) and 10–20 μHz per quantum level for protons ($\nu_z \sim 10$ MHz). Obviously, this bottle is useful only for detecting the electron's cyclotron resonance whose frequency is far too high ($\sim$ 140 GHz) to observe directly. However, this same bottle will also broaden [11] the cyclotron resonances of both ions and electrons via the thermal axial noise appearing in the z^2 term:

$$\Delta\nu_c / \nu_c = (B_2 Z_o^2 / B_o)(kT / eV_o) .$$

Thus, for a bottle with coefficient $B_2 = 500$ G/cm^2, $B_o = 50.5$ kG, $T = 4$ K, $V_o = 2$ V for electrons and 54 V for protons, $\Delta\nu_c / \nu_c$ is 2×10^{-8} and 8×10^{-10} respectively.

3. Measurements

Preliminary work using electrons stored in this quadring tap indicate that the chosen well depth is too shallow to fully compensate the trapping field and to allow narrow single electron resonances to be observed. Nevertheless, using approximately 10 electrons, the apparatus was able to yield the two cyclotron resonances shown in Fig. 3. These signals are proportional to axial frequency shifts and are produced by locking the axial resonance to a stable drive oscillator. As typical of past electron traps with magnetic bottles [9, 11], the two resonances show a clear low frequency edge, corresponding to $Z_{rms} = 0$, and a high frequency exponential tail, indicative of the Boltzmann distribution of axial states. Repeatability of these edges is typically better than 0.1 ppm for low applied microwave power.

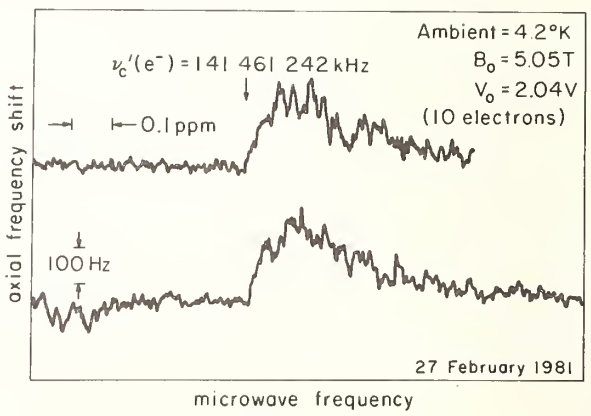


FIGURE 3. *Electron cyclotron resonances. Two separate traces show the repeatability of the low frequency edge ($Z_{rms} = 0$). As usual, the high frequency tail is due to the thermal axial states.*

Small numbers of protons have also been trapped and, for the first time, the proton cyclotron resonance has been rf driven and synchronously detected. An example of the high resolution available with this technique is shown in Fig. 4. Linewidths < 0.2 Hz can be reliably obtained out of an observed cyclotron frequency $\nu_c'(p^+)$ = 76.4 MHz (resolution ~ 0.001 ppm), limited primarily by observation time. From the tuned circuit damping, a single proton would have a linewidth of 0.005 Hz, though axial noise (via the magnetic bottle) will make such narrow resonances very difficult to observe. This single ion linewidth allows one to estimate an upper limit of 40 for the number of trapped protons in this example. Axial proton resonances are also synchronously detected but with approximately 100 times less resolution. However, their relative contribution to the accuracy of $\nu_c(p^+)$ is only 2%. In the future, a more stable electric potential well will be used to trap the protons in order to eliminate the contribution of the axial resonance to the uncertainty of the proton's cyclotron frequency.

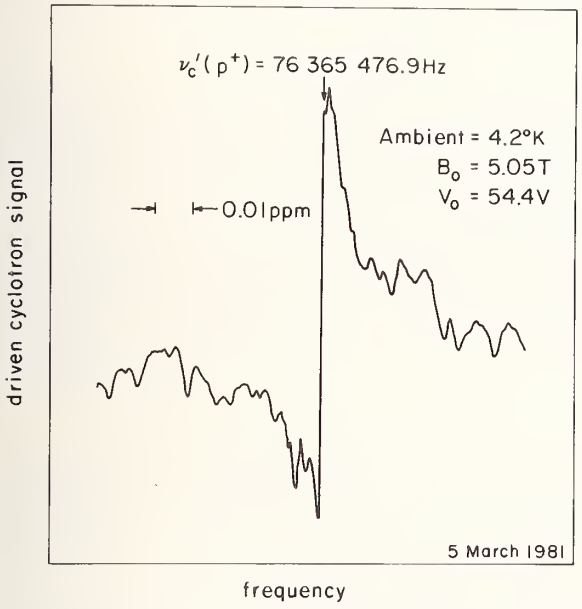


FIGURE 4. *Proton cyclotron resonance. The drive phase is adjusted to produce this dispersion curve during synchronous detection. The linewidth is limited primarily by observation time.*

4. Results

It is well known that any breakdown in ideal trap symmetry would lead to a cyclotron resonance which depends on the number of trapped ions [4, 5, 14]. This dependence does not result from space charge, unless the cloud contains dissimilar ions, because electrostatic interactions between like-ions in a cloud whose extent is small compared to the wavelength of the exciting rf field do not shift or broaden the cyclotron or axial resonances [15]. The procedure of applying intense rf drives at the axial frequency of the unwanted ions should remove them from the trap since their axial motion is undamped.

Figure 5 shows the mass ratio m_p/m_e plotted versus number of trapped protons. This dependence follows entirely from the number dependence of the proton cyclotron resonance $\nu_c(p^+)$ since no number dependence has been observed for $\nu_c(e^-)$ within the resolution available. The slope of a least squares fitted line through the data is -0.22×10^{-6}/ion. An earlier version of the quadring trap, with little compensation, yielded a number dependence 100 times as great and with opposite sign suggesting that this systematic error can be eliminated by proper adjustment of the compensation control. A plausible description of this number dependence which appears to be consistent with the above observations is made in appendix A.

The intercept of the linear fit in Fig. 5 corresponds to the preliminary proton/electron mass ratio m_p/m_e = 1836.15300(7). However, an analysis of bottle-related position dependence of the magnetic field, given in Appendix B, suggests that a systematic error as large as 0.1 ppm may exist if the two charge types do not have the same average position in space. Thus, a preliminary mass ratio of

$$m_p/m_e = 1836.15300(25)$$

reflects this uncertainty in magnetic field location. Future tests will specifically search for the magnetic saddle point [11] and vary B_2 in order to eliminate this uncertainty. Also, a deeper potential well will be used to trap and detect a single electron in order to yield $\nu_c(e^-)$ resolution which is typically parts in 10^9 for previous electron traps. This preliminary result agrees with values measured by Gräff *et al.* [5] and Gärtner and Klempt [4] within their experimental errors. However, we expect

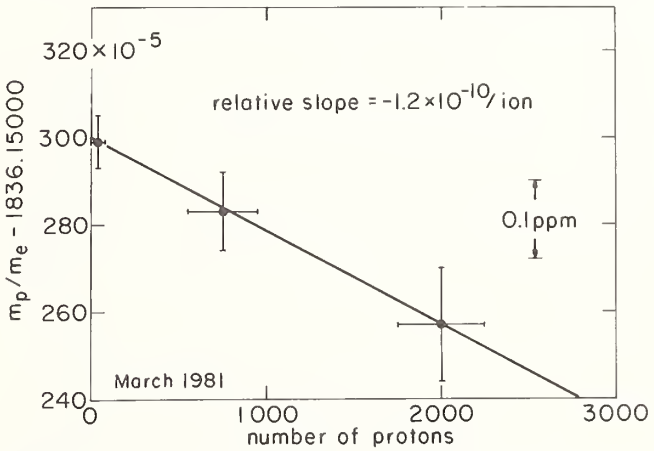


FIGURE 5. *Mass ratio versus proton number. This dependence is due entirely to the sensitivity of $\nu_c(p^+)$ to proton number and appears to depend on the compensation of the trap.*

our techniques, with appropriate systematic checks completed, will ultimately produce a mass ratio with a precision that exceeds 0.01 ppm. Other possible improvements include using double traps [16] and a variable magnetic bottle [17].

We wish to acknowledge our gratitude to machinists Don Russell and Ralph Jochim for their excellent machine work on the traps and to glassblower Bob Moerly for his superb and patient work.

References

[1] S. A. Lee, R. Wallenstein and T. W. Hänsch, Phys. Rev. Lett. 35, 1262 (1975).
[2] For instance, see the 1973 Least Squares Adjustment of the Fundamental Constants, E. Richard Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).

[3] See review article by B. N. Taylor, Metrologia **12**, 81 (1976), containing the values of $\mu_p' / \mu_n = 2.792\ 77406(103)$ {least-squares—adjusted value of quantities measured by B. A. Mamyrin, N. N. Aruev, and S. A. Alekseenko, Zh. Eksp. Teor. Fiz. 63, 3 (1972) [Sov. Phys. JETP 36, 1 (1973)], and by B. W. Petley and K. Morris, J. Phys. A 7, 167 (1974)} and $\mu_p' / \mu_B = 0.001\ 520\ 993\ 139(18)$ [temperature-corrected value of quantity measured by W. D. Phillips, W. E. Cooke, and D. Kleppner, Metrologia 13, 179 (1977)].
[4] G. Gärtner and E. Klempt, Z. Phys. A 287, **1** (1978).
[5] G. Gräff, H. Kalinowsky, and J. Traut, Z. Phys. A 297, 35 (1980).
[6] A. Wyler, C. R. Acad. Sci Ser. A 269, 743 (1969); 271, 186 (1971).
[7] See also D. O. Fystrom, Phys. Rev. Lett. 25, 1469 (1970); J. L. Luxon and A. Rich, Phys. Rev. Lett. 29, 665 (1972); H. Gubler, S. Munch, and H. H. Staub, Helv. Phys. Acta 46, 772 (1974).
[8] R. S. Van Dyck, Jr., D. J. Wineland, P. A. Ekstrom, and H. G. Dehmelt, Appl. Phys. Lett. 28, 446 (1976).
[9] R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, in *New Frontiers in High-Energy Physics* (Plenum, New York, 1978); also, H. Dehmelt in *Atomic Physics* 7, D. Kleppner and F. Pipkin, eds. (Plenum, New York, 1981) p. 337.
[10] D. Wineland, Bull. Am. Phys. Soc. 18, 1573 (1973).
[11] R. S. Van Dyck, Jr., P. B. Schwinberg, and S. H. Bailey, in *Atomic Masses and Fundamental Constants 6*, J. A. Nolen, Jr. and W. Benenson, eds. (Plenum, N.Y., 1980) p. 173.
[12] H. G. Dehmelt and F. L. Walls, Phys. Rev. Lett. **21**, 127 (1968).
[13] G. Gabrielse and H. Dehmelt, Bull. Am. Phys. Soc. **26**, 598 (1981).
[14] S. Liebes, Jr. and P. Franken, Phys. Rev. **116**, 633 (1959).
[15] D. Wineland and H. Dehmelt, Int. J. Mass Spectrom. Ion Phys. **16**, 338 (1973).
[16] P. B. Schwinberg and R. S. Van Dyck, Jr., Bull. Am. Phys. Soc. 23, 40 (1978) and 24, 35 (1979).
[17] P. B. Schwinberg, R. S. Van Dyck, Jr., and H. G. Dehmelt, Bull. Am. Phys. Soc. 24, 1202 (1979); P. B. Schwinberg and R. S. Van Dyck, Jr., Bull. Am. Phys. Soc., **26**, 598 (1981).

Appendix A

A plausible explanation for the observed dependence of the proton cyclotron resonance on number of trapped protons N can be found by noting the following result from a paper by Liebes and Franken [14], but adapted specifically to a Penning trap,

$$\nu_c' = \nu_c - \left[\left[\frac{1}{r} \frac{\partial \phi}{\partial r} \right]_o + \left[\frac{\partial^2 \phi}{\partial r^2} \right]_o \right] (\nu_m Z_o^2 / V_o)$$

where ϕ is the potential function which establishes the electrostatic field and the derivatives are to be evaluated at the guiding center of the magnetron motion. If fields from space charge are neglected (since they should not affect the observed cyclotron frequency), the potential function with the dominant anharmonic terms in cylindrical coordinates (r, z) is

$$\phi = (V_o / 4Z_o^2)(r^2 - 2z^2) + B_A (8r^4 - 24r^2 z^2 + 8z^4)$$

where $B_A \sim 0.2$ V/cm^4 in earlier uncompensated Penning traps [8]. Upon taking the appropriate derivatives, one finds that

$$\nu_c' = \nu_c - \nu_m [1 + (48 B_A Z_o^2 / V_o)(r^2 - 2z^2)_o]$$

where the average of $(r^2 - 2z^2)_o$ over the charge distribution will contain the observed number dependence.

This expression can be carried a step further if one assumes a uniform charge density of ions that will repel each other until the nearest neighbor distance is R, at which point radial spreading becomes insignificant or is balanced by the centering drive. In addition, if one assumes that no axial stacking of ions occurs (i.e., we can thus neglect the z^2 term), and that the area $(1/2)N\pi R^2$ can represent the average of $\pi(r^2)_o$, then

$$\nu_c' = \nu_c - \nu_m (1 + \gamma N)$$

where $\gamma = 24 B_A Z_o^2 R^2 / V_o$. The explicit appearance of the anharmonic coefficient B_A indicates that a well compensated trap should not depend on N. Also, since ν_m is independent of mass, the relative correction to $\nu_c'(e^-) \sim 140$ GHz is insignificant compared to the correction to $\nu_c'(p^+) \sim 76.4$ MHz. Finally, if one used the observed proton cyclotron number dependence of 0.0093 Hz/ion and $B_A \sim 0.2$ V/cm^4, one finds that $R \sim 0.0035$ cm. Of course, if the trap is 100 times better compensated, then R is 10 times smaller.

Appendix B

Since the nickel rings produce a magnetic field that depends on position in the trap, we must consider effects which will cause both charge types to see a different magnetic field.

Residual contact potentials on the electrodes, which are expected to be much less than 100 mV, can be resolved into constant electric fields in one of the three independent directions. On the average, such fields will give rise to an average relative magnetic field shift given by

$$\delta \nu_c / \nu_c = (B_2 Z_o^2 / B_o)(\Delta V / 2V_o)^2 ,$$

which for the parameters of our trap is much less than 7×10^{-8} and 1×10^{-10} respectively for electrons and protons. Implicit in this calculation is the assumption that the bottle and trap centers coincide. By construction, this is valid to ± 0.001 cm which reduces the corresponding shifts [11] to much less than 1×10^{-7} also.

The particles' axial and radial motion also causes average field shifts. The axial motion causes an average shift equal to

$$\delta \nu_c / \nu_c = (B_2 Z_o^2 / B_o)(kT / eV_o)$$

which for $T = 4$ K corresponds to 2×10^{-8} and 8×10^{-10} for electrons and protons respectively. For $\nu_c(e^-)$, the low frequency edge is independent of this effect. However, the cyclotron frequency of both protons and electrons depends on the magnetron radius R_m according to

$$\delta \nu_c / \nu_c = B_2 R_m^2 / 2B_o .$$

For the case of a well centered electron motion in a previous trap, $R_m < 0.002$ cm corresponds to $\delta \nu_c / \nu_c < 2 \times 10^{-8}$. If the centered magnetron radius for both protons and electrons is the same, then the mass ratio does not depend on R_m. However, in order to verify that m_p / m_e does not depend on any of these bottle related effects, mass ratio measurements will be made versus B_2 by adding an external magnetic ring to the present trap tube.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A Direct Determination of the Proton-Electron Mass Ratio

G. Gräff, H. Kalinowsky, and J. Traut

Institut für Physik der Johannes Gutenberg-Universität D-6500 Mainz, West Germany

The cyclotron frequencies of free protons and electrons have been measured in a magnetic field of 5.81 tesla with a superimposed electrostatic quadrupole field. The increase of energy connected with a transition at the cyclotron frequency is detected by the measurement of the time of flight through an inhomogeneous magnetic field. From the ratio of the measured cyclotron frequencies of both particles, the proton-electron mass ratio is deduced. The result, $m_p/m_e = 1836.1527(13)$, agrees within the limits of error (0.69 ppm) with the value of the indirect determination.

Key words: fundamental constants; proton-electron mass ratio; Penning trap.

1. Introduction

The determination of the proton-electron mass ratio is possible by both a direct and indirect method. The indirect method requires two different types of measurements: the determination of the magnetic moment of the proton in units of the Bohr magneton μ_p'/μ_B and the determination of the magnetic moment of the proton in units of the nuclear moment μ_p'/μ_N. The combination of both results leads to a proton-electron mass ratio $m_p/m_e = 1836.15152\ (70)$ [1].

The direct determination is based on the measurements of the cyclotron frequencies of proton and electron in the same magnetic field:

$$m_p/m_e = f_c(\text{Electron})/f_c(\text{Proton})$$

2. Apparatus

2.1 Ion Trap

Protons as well as electrons are trapped alternatively in a homogeneous magnetic field with a superimposed electrostatic quadrupole field, the potential of which is

$$V(x, y, z) = U/R_o^2(x^2 + y^2 - 2z^2).$$

In this configuration the ions behave within the nonrelativistic limit like a three-dimensional harmonic oscillator with the following frequencies:

$$f_z = (2eU/mR_o^2)^{1/2}/2\pi$$
$$f_B = f_c/2 + f_o$$
$$f_{EB} = f_c/2 - f_o$$

with $f_o = (f_c^2/4 + f_z^2/2)^{1/2}$ and the cyclotron frequency $f_c = eB/\gamma m 2\pi$, $\gamma = (1 - v^2/c^2)^{-1/2}$.

Under the conditions chosen in our experiment (B = 5.8 T, R_o = 6 mm, and U = 1 V) f_z is about 16 MHz (367 kHz), f_B = 163 GHz (89 MHz), and f_{EB} = 760 Hz (760 Hz) for electrons (protons). The trap is made of OHFL copper. Additional correcting electrodes provide a well-shaped quadrupole field. The central ring is split in the x, z plane to allow application of an rf-field perpendicular to the static magnetic field. At one side of the quadrupole trap, 40 cm away from it but near the magnetic field axis, a heated tungsten wire serves as the electron source. These primary electrons impinge partly on the interior trap surface through the entrance hole of the endcap (1 mm diameter) producing secondary electrons or protons by secondary emission or residual gas ionization, respectively. A few of them are trapped. Due to synchrotron radiation in the strong magnetic field the electrons loose their initial energy of a few eV and cool down to thermal energies of about 30 meV within a second.

2.2 Drift Tube and Detector System

The trap is installed in a 37 cm long copper drift tube of 30 mm diameter. The surface of the tube was cleaned by a glow discharge in an atmosphere of hydrogen. To obtain a vacuum of 10^{-9} Torr the complete apparatus was baked out at a temperature of 600 K for two days. At the open end of the tube a channel plate detector is mounted to count alternatively electrons or protons after ejection out of the trap (Fig. 1). The complete apparatus is installed within a electromagnetically shielded enclosure to avoid any disturbance of the measurement by unwanted electromagnetic fields. The data transfer between the apparatus inside the cabin and the data acquisition computer is made by fiber optic transmission lines.

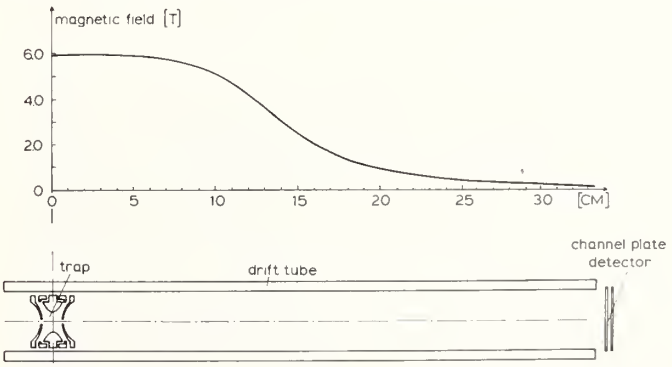


FIGURE 1. *The experimental arrangement for the time of flight measurement.*

2.3 Magnetic Field

A superconducting magnet is used in this experiment. Its maximum field strength is 6.4 tesla. Room temperature access is provided along a horizontal bore with

52 mm diameter. To ensure a sufficient homogeneity over the trap volume a series of coils is mounted on a vacuum enclosure. They allow a systematic correction of linear field gradients in three directions and of the quadratic term along the axis of symmetry. These coils are run at room temperature. The drift of the magnetic field was less than 0.02 ppm per day when the magnet was run at 5.28 T and about 1.0 ppm per day when run at 5.18 T (Fig. 2).

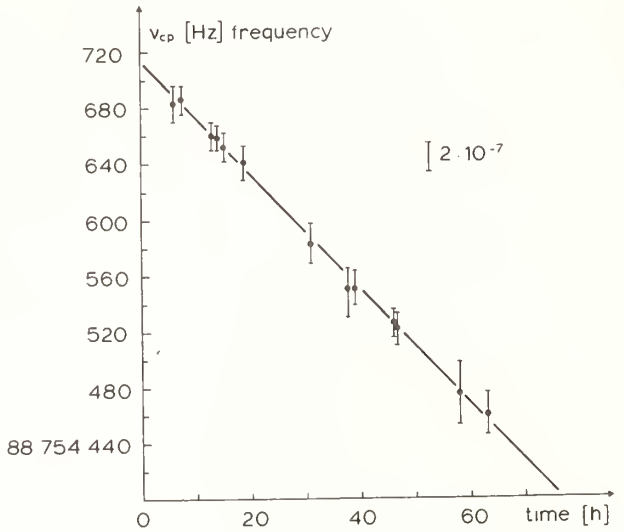


FIGURE 2. *Proton cyclotron frequency shift as a function of time caused by the magnetic field drift (error bars correspond to FWHM).*

3. Method

There are three possibilities to obtain the cyclotron frequency:

a) A measurement of f_B at different trapping potentials U and an extrapolation to zero electrostatic field strength.
b) A measurement of f_B and f_{EB}, the sum of which amounts to the cyclotron frequency f_c.
c) A direct induction of a transition at frequency $(f_B + f_{EB})$, which can be enforced at low electric field strengths.

We selected the latter method since this transition is most insensitive to a misalignment of the trap, to space charge effects, and to surface potentials. To detect the cyclotron frequency the following procedure is used which was proposed originally by Bloch [2]. After creation of a few electrons the potential is kept constant for one second during which the electrons cool down (Fig. 3). Then the cyclotron frequency is applied for about 500 ms. Finally the trap is cleared by a linear sweep superimposed by a sequence of pulses which define the starting time of the ejected particles. With the same pulses a series of consecutive gates is generated, which open 24 fast counters one after the other. By this method the time of flight spectrum is received. The time of flight is determined by the initial energy of the particles along the magnetic field axis (about 10 meV for electrons and some 100 meV for protons) and by their transverse energy incorporated in the cyclotron motion. The particles are accelerated by the inhomogeneous magnetic field proportional to their transverse energy. When the cyclotron

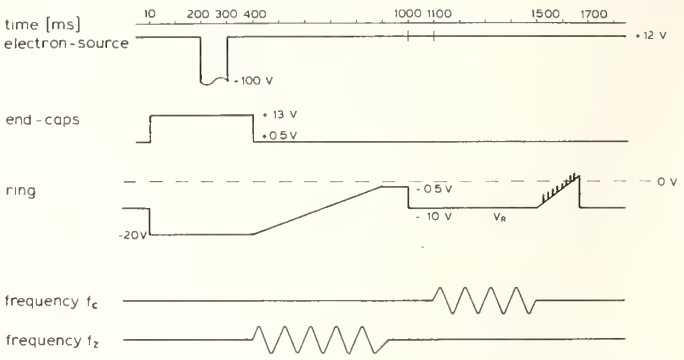


FIGURE 3. *Trapping potential.*

frequency is induced, the trapped particles gain transverse energy. The cyclotron orbit of the particle and consequently its magnetic moment increases as well. When these particles are expelled out of the trap they get an additional acceleration in the inhomogeneous part of the magnetic field corresponding to a decrease of their time of flight. Thus the cyclotron frequency can be detected.

4. The Measurement of the Cyclotron Frequencies

4.1 The Proton Cyclotron Frequency

The cyclotron frequency of the proton has been determined as outlined in section 3. Choosing an initial energy of 300 meV the mean time of flight is about 30 μs. At resonance it decreases by a few percent depending on the rf-amplitude (Fig. 4). At each frequency the whole cycle is repeated 30 times. The resulting time distribution stored in the 24 counters is then read by a PDP 11/20 computer which calculates the mean time of flight and the statistical error. To study the lineshape and possible line shifts the following parameters have been varied: the rf-amplitude, the trap position relative to the magnetic field, the magnetic field strength, the trap potential, and the space charge. In addition, the cyclotron frequency of the molecular ion H_2^+ was measured.

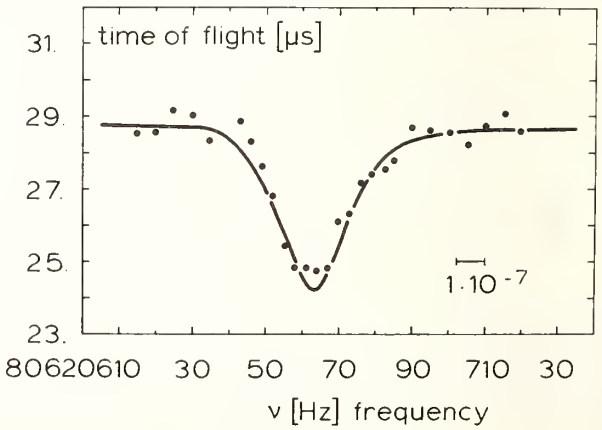


FIGURE 4. *Proton time of flight as a function of frequency.*

4.1.1 Electric and Magnetic Field Dependence

To the extent that a pure quadrupole field is realized, the transition frequency $(f_B + f_{EB})$ should not depend on the electric field strength. Within the statistical error

this appeared to be correct. At very low voltages (100 meV), however, when surface potentials become dominant, the center of ion motion was evidently shifted along the z-axis. This leads to observable shifts of the cyclotron frequency depending on the position of the trap and the magnetic field inhomogeneity. As expected the transition probability appeared to be inversely proportional to the electric field strength. Measurements were performed at 5.28 T and 5.81 T. The mass ratios m_p/m_e resulting from both experiments agree within their statistical errors.

4.1.2 Space Charge Effects

Depending on the charge distribution in the trap the two frequencies f_B and f_{EB} might be influenced differently. Therefore the cyclotron frequency f_c may also depend on the charge distribution. Indeed, the cyclotron frequency f_c is both shifted and broadened proportionally to the number of stored ions. Since the protons are generated by electron bombardment of the residual gas, other ions of heavier masses such as H_2O^+ and N_2^+ are also produced and stored inside the trap. These ions contribute, of course, to the space charge. Therefore the trap was cleared of all other ions before inducing the proton cyclotron frequency. This was achieved in the following way: After ion creation the trap voltage is decreased linearly down to 1 V. Simultaneously, a rf-field is applied, the frequency of which is chosen so that all ions heavier than the protons experience their resonant frequency f_z (Fig. 3). The amplitude of this rf-field is high enough to eject all unwanted ions. The cyclotron frequency was then measured for different numbers of trapped protons. Figure 5 shows linewidth and transition frequency as a function of the number of stored protons.

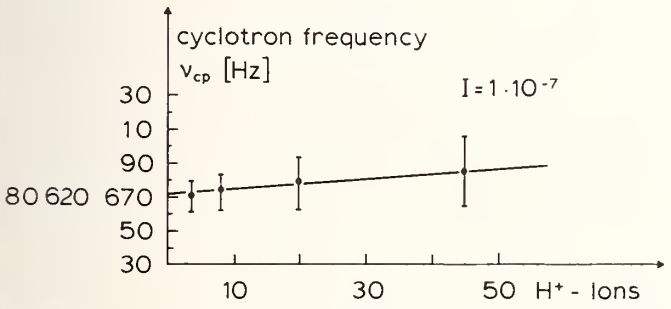


FIGURE 5. *Proton cyclotron frequency as a function of the number of trapped protons (error bars correspond to FWHM).*

4.2 The Electron Cyclotron Frequency

Within a second after their generation the trapped electrons are in a thermal state which shifts the electron cyclotron frequency by 0.05 ppm due to their relativistic mass increase. By applying a stepwise increasing potential at an electrode located at the end of the drift tube, the energy distribution of the electrons has been measured. Since the magnetic field strength is low at the position of this electrode this method yields the sum of the energies in the longitudinal and transverse degrees of freedom. As a result the width of the energy distribution appeared to be less than 50 meV. The mean value of the time of flight through the magnetic field for electrons is 3 μs. At resonance the time of flight decreases by about 10 percent. An increase in microwave amplitude results in an energy absorption by the trapped electrons, which leads to a relativistic mass increase and therefore to a shift and broadening of the cyclotron frequency (Fig. 6).

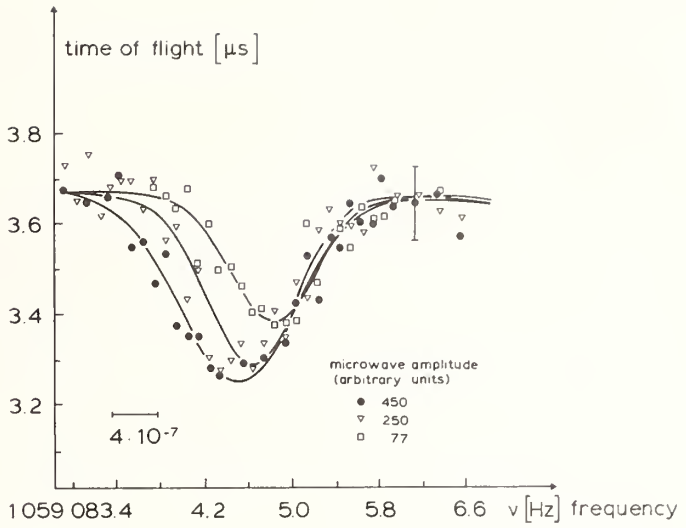


FIGURE 6. *Electron time of flight as a function of frequency and microwave amplitude.*

By means of measurements at different microwave amplitudes, this dependence was determined and extrapolated to zero energy absorption. Smaller linewidth can be obtained by increasing the sensitivity for the detection of energy changes [3]. The sensitivity is presently limited by surface potentials on the inner side of the drift tube. The transition frequency is influenced by power broadening, relativistic mass increase, and cooling effects due to synchrotron radiation. The resulting line shape could not be explained theoretically. Within the statistical error the line shape appeared symmetrical and was fitted by a Lorentzian function.

Again at very low trapping potentials a shift of the cyclotron frequency was observed. Space charge effects were negligible.

5. Result and Discussion

The measurement of proton and electron cyclotron frequencies was performed alternatively at two different magnetic field strengths of 5.28 tesla (run duration 120 h) and 5.81 tesla (run duration 60 h). The drift of the magnetic field during the measurement was determined by the variation with time of both the proton and the electron cyclotron frequencies. The values of the magnetic field drift calculated separately from the proton and the electron cyclotron frequencies agree within their statistical errors. The proton electron mass ratio m_p/m_e resulting from both experiments agree within their statistical error (0.38 ppm). Table 1 summarizes all corrections to the cyclotron frequencies and Table 2 the error contributions. The errors correspond to one-standard-

TABLE 1. *Corrections for the determination of the cyclotron frequencies of electron and proton.*

	f_c-electron	f_c-proton
Space charge effect	—	−0.05 ppm
Trapping voltage extrapolation	−0.19 ppm	+0.12 ppm
Zero kinetic energy extrapolation	+0.05 ppm	—
Zero microwave amplitude extrapolation	+0.09 ppm	—
Frequency generators difference	+0.14 ppm	—

TABLE 2. *Error contribution to electron and proton cyclotron frequencies and to proton-electron mass ratio.*

	f_c-electron	f_c-proton	m_p/m_e
Statistical errors			
Magnetic field drift extrapolation	0.07 ppm	0.02 ppm	0.09 ppm
Trapping voltage extrapolation	0.14 ppm	0.04 ppm	0.18 ppm
Zero microwave amplitude extrapolation	0.30 ppm	—	0.30 ppm
Frequency generators difference	0.01 ppm	—	0.01 ppm
Systematic errors			
Zero kinetic energy extrapolation	0.05 ppm	—	0.05 ppm
Magnetic field inhomogeneity	0.20 ppm	0.20 ppm	0.40 ppm
Zero microwave amplitude extrapolation	0.40 ppm	—	0.40 ppm
			0.69 ppm

deviations. Our final result is the mean value of both measurements:

$$m_p/m_e = 1836.1527(13) \quad (0.69 \text{ ppm}) \, .$$

The result of this direct determination of the proton-electron mass ratio agrees within the limits of error with the most recent least-squares adjusted value [1], $m_p/m_e = 1836.15152(70)$ (0.38 ppm).

References

[1] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[2] F. Bloch, Physica 13, 821 (1953).
[3] R. S. van Dyck, P. B. Schwinberg, and H. G. Dehmelt, Phys. Rev. Lett. 38, 310 (1977).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A New Determination of the Atomic Weight of Silver and an Improved Value for the Faraday

L. J. Powell, T. J. Murphy, and J. W. Gramlich

Center for Analytical Chemistry, National Measurement Laboratory, National Bureau of Standards, Washington, DC 20234

The atomic weight of a reference sample of silver has been determined by mass spectrometry with an uncertainty of one part in 10^6. Accurately known quantities of chemically pure ^{107}Ag and ^{109}Ag were mixed to produce standards of known isotopic composition for calibration of the mass spectrometer. The absolute isotopic ratio of the reference sample of silver is $^{107}Ag/^{109}Ag = 1.07638 \pm 0.00022$ yielding an atomic weight of 107.86815 ± 0.00011 [95 percent confidence level (CL)]. Recalculation of the Faraday using this atomic weight and the electrochemical equivalent of silver as determined by Bower and Davis leads to a value of (96486.18 ± 0.13) $A_{NBS} \cdot s \cdot mol^{-1}$ (68 percent CL).

Key words: atomic weight; chemical assay; Faraday; isotopic composition; mass spectrometry; silver.

1. Introduction

The redetermination of the atomic weight of silver was undertaken in conjunction with the determination of a more accurate value for the Faraday constant. The Faraday is directly related to other physical constants including the Avogadro constant, the proton gyromagnetic ratio, the magnetic moment of the proton in nuclear magnetrons, and the ratio of the NBS as-maintained ampere to the absolute or SI ampere. Over the years, a problem arose in assigning a recommended value for the Faraday due to apparent discrepancies between the Faraday determined from electrochemical experiments and the Faraday calculated from other fundamental constants. Unfortunately, earlier electrochemical experiments were not sufficiently precise [6.8 ppm, 68 percent confidence level (CL) estimate, for the best silver determination] to either prove or disprove the existence of this discrepancy. As a result, the Faraday was excluded from the 1973 CODATA least-squares adjustment of the fundamental constants [9].

In 1980, Bower and Davis [1], using the same source of silver as was analyzed in this work, NBS Standard Reference Material (SRM) 748, published a new value for the electrochemical equivalent of silver with an uncertainty of 1.28 ppm (68 percent CL). The calculation of the Faraday using this value and the atomic weight of silver determined in 1962 at NBS [2] indicated that significant differences between the various Faraday calculations might indeed exist. In order to assess the significance of these differences, however, it became necessary to significantly reduce the uncertainty in the atomic weight of silver.

2. Approach

The method used for the determination of atomic weights at NBS combines the techniques of high precision chemical assay with high precision mass spectrometry. The mass spectrometers to be used for the isotopic abundance measurements are calibrated for bias using synthetic mixes of known isotopic composition, prepared from nearly pure, separated isotopes. Extensive research has demonstrated that this bias is due to non-linearities in the measurement circuit and mass dependent isotopic fractionation [3]. In the case of isotopic ratios near one (as for silver), this bias reduces to isotopic fractionation. These measured biases are then used to calculate the absolute isotopic abundance ratio and, ultimately, the atomic weight of the reference sample. A block diagram characterizing the atomic weight method is shown in Fig. 1.

The chemical research requires the development of an assay procedure for the pure element or reference sample precise to at least one part in 10^4, which will ultimately be used to assay the separated isotope solutions. In addition, a procedure must be developed to purify the separated isotopes to eliminate impurities which could interfere with either the chemical or mass spectrometric procedures. Once these methods have been thoroughly

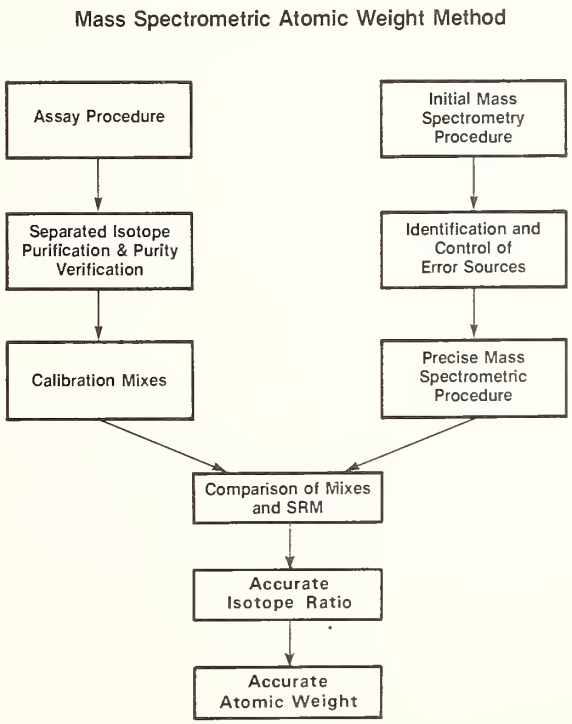


FIGURE 1. *Block diagram of the mass spectrometric atomic weight method.*

tested using the reference sample, the separated isotopes are purified and dissolved. Aliquots of each isotope are then taken to produce calibration mixes which are blended to bracket the natural isotopic abundance ratio of the element. In addition, aliquots of each isotope are taken for assay so that accurate solution concentrations can be ascertained. A knowledge of the solution concentrations combined with aliquoting data allows the actual ratios of the calibration mixes to be calculated.

The mass spectrometric research involves the development of a high precision method for the analysis of the relative isotopic abundance ratios of the reference sample, followed by a comprehensive study to identify and control any sources of bias which could affect the final measurements.

After the above tasks have been accomplished, two complete sets of analyses of the calibration mixes and the SRM are made by two different operators using two different mass spectrometers. The calibration mixes are randomly selected and are analyzed in an alternating pattern with the reference sample. A comparison of the relative isotopic abundance ratios obtained for the calibration mixes with the calculated or true value produces a calibration factor. This factor is applied to the relative isotopic ratio obtained for the SRM to produce the absolute isotopic abundance ratio. The atomic weight is then obtained by multiplying the fractional abundance of each isotope by its nuclidic mass[1] and summing the resultant products.

For the redetermination of the atomic weight of silver, a high-precision chemical assay procedure was developed so that accurately known quantities of ^{107}Ag and ^{109}Ag separated isotopes could be mixed to produce standards for calibration of the mass spectrometer. This method was based on the gravimetric determination of the major portion of the silver (>99.7%) as silver iodide and the determination of the soluble residual silver by isotope dilution mass spectrometry. The relative isotopic composition of the calibration mixes and the high purity natural silver were determined by surface ionization mass spectrometry using a platinum single-filament silica gel technique. This technique, which involves loading sequential layers of silica gel, sample, and phosphoric acid onto the filament surface, has not been traditionally used in high accuracy atomic weight determinations because the method of ionization enhancement is little understood. However, in the case of silver, it proved to be the key to achieving levels of precision not possible with other modes of ionization. Further details of these procedures and the associated data are published elsewhere [5].

3. Results

The atomic weight of the reference sample of silver was determined to be (107.86815 ± 0.00011) $g \cdot mol^{-1}$ (95% CL). This reference sample will be issued as both an isotopic and assay Standard Reference Material and will be designated as SRM 978a. Its isotopic ratio was found to be experimentally identical to SRM 978, which was based on the 1962 determination of the atomic weight of silver [2], the major difference being the uncertainty placed on the measurement.

[1]Nuclidic masses, as published by Wapstra and Bos [4], are known to an accuracy of a few parts in 10^8.

The redetermination of the atomic weight of silver has allowed the recalculation of the Faraday, yielding a value of (96486.18 ± 0.13) $A_{NBS} \cdot s \cdot mol^{-1}$ (68% CL) [6]. Figure 2 shows the value of the Faraday calculated from this work as well as other values of the Faraday. Point D is the present CODATA recommended value of Cohen and Taylor [9], which is calculated from other physical constants via least squares. It may be noted that not only has the overall uncertainty in the Faraday constant been reduced through this determination, but the value of the Faraday is now essentially the same as that calculated using the proton gyromagnetic ratio measured by Kibble and Hunt [11]. The ultimate impact, however, is that the Faraday will no doubt be included in the 1983 least-squares adjustment of the fundamental constants presently being carried out by Cohen and Taylor under CODATA auspices.

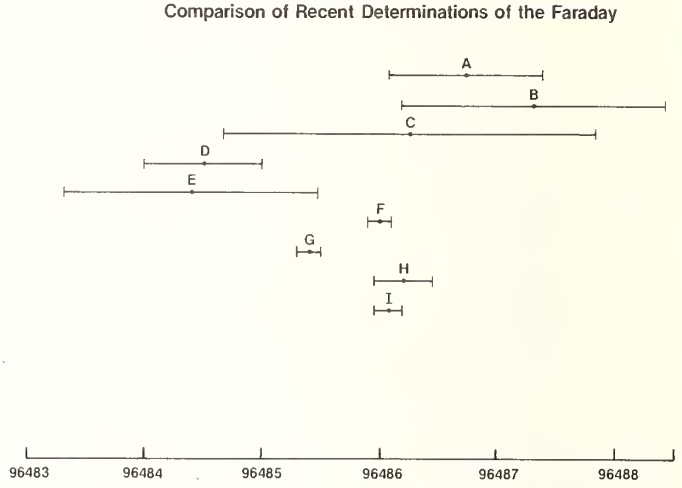


FIGURE 2. *Comparison of various values of the Faraday constant.* ***A****, Craig et al., [7];* ***B****, Marinenko and Taylor [8];* ***C****, Marinenko and Taylor [8];* ***D****, 1973 Recommended [9];* ***E****, Koch and Diehl [10];* ***F****, Kibble and Hunt [11];* ***G****, Deslattes et al., [12];* ***H****, Bower and Davis, 1960 Ag. [1];* ***I****, Bower and Davis, 1980 Ag [this work].*

References

[1] V. E. Bower and R. S. Davis, J. Res. Natl. Bur. Stand. 85, 175 (1980).

[2]. W. R. Shields, E. L. Garner, and V. H. Dibeler, J. Res. Natl. Bur. Stand. Sect. A: 66, 1 (1962).

[3] E. L. Garner, L. A. Machlan, and W. R. Shields, Natl. Bur. Stand. (U.S.), Spec. Publ. 260-27 (1971).

[4] A. H. Wapstra and K. Bos, At. Data Nucl. Data Tables 19, 177 (1977).

[5] L. J. Powell, T. J. Murphy, and J. W. Gramlich, J. Res. Natl. Bur. Stand. 87, 9 (1982).

[6] V. E. Bower, R. S. Davis, T. J. Murphy, P. J. Paulsen, J. W. Gramlich, and L. J. Powell, J. Res. Natl. Bur. Stand. 87, 21 (1982).

[7] D. N. Craig, J. I. Hoffman, C. A. Law, and W. J. Hamer, J. Res. Nat. Bur. Stand. Sect. A: 64, 381 (1960).

[8] G. Marinenko and J. K. Taylor, Anal. Chem. 40, 1645 (1968).

[9] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).

[10] W. F. Koch and H. Diehl, Talanta 23, 509 (1976).

[11] B. P. Kibble and G. J. Hunt, Metrologia 15, 5 (1979).

[12] R. D. Deslattes, A. Henins, H. A. Bowman, R. M. Schoonover, C. L. Carroll, I. L. Barnes, L. A. Machlan, L. J. Moore, and W. R. Shields, Phys. Rev. Lett. 33, 463 (1974); and these proceedings.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

High Resolution Magnetic Measurements on Rotating Superconductors to Determine h/m_e*

B. Cabrera, S. (Benjamin) Felch, and J. T. Anderson

Physics Department, Stanford University, Stanford, CA 94305

We have begun a new experiment to determine h/m_e (Planck's constant divided by the free electron mass) to an accuracy of several parts per million using rotating superconducting rings. This resolution will allow for the first time investigation of recently predicted relativistic corrections at a level of 100 to 200 ppm. Here we present initial experimental measurements of h/m_e. These include high precision measurements of the induced magnetic field from the rotating superconducting ring, showing an improvement in resolution by two orders of magnitude over previously reported observations, and a clear demonstration of flux quantization in the 50 mm diameter ring.

Key words: electron mass; flux quantization; fundamental constants; London moment; Planck's constant; superconductivity.

1. Introduction

A precise determination of h/m_e using a rotating superconducting ring can be based on several of the remarkable macroscopic quantum properties of the superconducting state. Becker [1] and co-workers first predicted that a resistanceless conductor rotated from rest would exhibit a magnetic moment proportional to its spin speed. The effect came to be known as the London moment after the phenomenological theory of F. London and H. London [2] was applied by F. London [3] to a rotating superconductor. London predicted a uniform magnetic field

$$\mathbf{B}_L = -\frac{2m}{e}\,\boldsymbol{\omega} \qquad (1)$$

within the rotating lattice. It was first observed by Hildebrandt [4] and has been verified several subsequent times to an accuracy approaching a few percent [5–9].

Flux quantization through a superconducting ring, also predictd by F. London [3] (as h/e instead of the correct $h/2e$), was first experimentally verified by B. S. Deaver and W. M. Fairbank [10] (observing $h/2e$) and R. Doll and M. Nabauer [11]. More recently, high precision measurements of the flux quantum $h/2e$ have been based on the ac Josephson effect [12–14]. A resolution better than 0.1 parts per million (ppm) has been achieved, which is by far the most accurate determination of any property of a solid state system. At that time several suggestions were made for measurements of h/m in rotating superconducting rings [15]. These are based on balancing the magnetic flux from an integral number of flux quanta $n(h/2e)$ against the London moment flux $2m\omega_n S/e$, where S is the area bounded by the ring. Thus, we obtain

$$h/m = 4\,S(\omega_n - \omega_{n-1}). \qquad (2)$$

Several experiments have measured h/m in this way [16, 17], with the most accurate result reported by Parker and Simmonds [17] in 1970 at PMFC-I with an accuracy of 400 ppm. It is in agreement with the accepted value for the free electron at rest obtained by other techniques.

Technological advances in magnetic shielding [18], SQUID magnetometry [19], cryogenic rotors [20], and high precision dimensional metrology [21] during the past decade, primarily associated with the Relativity Gyroscope Experiment [20], now allow an improvement in this determination to the several ppm level or better. For the first time this will allow investigation of relativistic corrections in a solid state system. Recent theoretical estimates by B. Cabrera, H. Gutfreund, and W. A. Little [22] predict corrections of 100–200 ppm in many superconducting materials.

In Section 2 we derive the exact non-relativistic expression given by Eq. (1) and summarize the relativistic results. Section 3 outlines the measurement scheme and describes the apparatus. Recent laboratory measurements including high resolution London moment and flux quantization data are presented in Section 4. We conclude in Section 5 with a discussion of the prospects for future corrected values below the parts per million level.

2. Theory

The non-relativistic basis for the determination of h/m can be understood in several ways. In Fig. 1 the Sommerfeld quantization rule is applied to the canonical

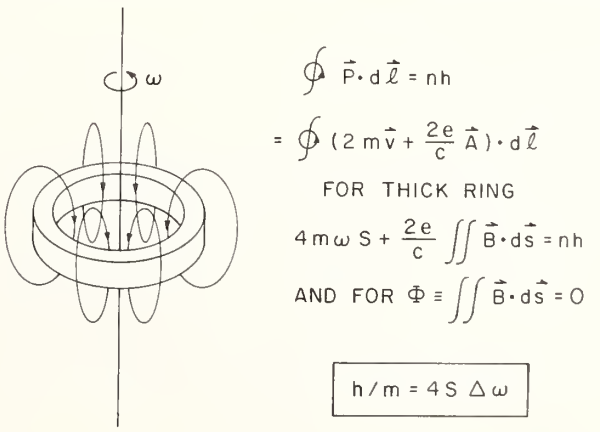


FIGURE 1. *The Wilson-Sommerfeld quantization rule applied to Cooper electron pairs in a rotating superconducting ring leads to a simple derivation of Eq. (2).*

*This work was supported by NBS Precision Measurement Grant GB-9026, NASA Contract NAS8–32355, and NSF Grant DMR 80–26007.

momentum of a Cooper pair around a superconducting ring. If the ring is many London penetration depths thick, a line integral can always be found where the supercurrent is negligible and thus $\mathbf{v}$ can be replaced by $\boldsymbol{\omega} \times \mathbf{r}$, the rigid body lattice rotation.

A more formal approach is obtained through the Ginzburg-Landau supercurrent [23]

$$\mathbf{j} = \frac{e^* \hbar}{2im^*}(\psi^* \nabla \psi - \psi \nabla \psi^*) - \frac{e^{*2}}{m^*}\psi^* \psi \mathbf{A}, \qquad (3)$$

where the mass m^* and charge e^* are equal to $2m$ and $2e$, respectively, and ψ is proportional to the superconducting order parameter with phase ϕ, $\psi = |\psi| e^{i\phi}$. Transforming Eq. (3) into a rotating coordinate system using the Galilean relation $\mathbf{v}' = \mathbf{v} - \boldsymbol{\omega} \times \mathbf{r}$, we obtain

$$\mathbf{j} = \frac{e\hbar}{m} |\psi|^2 \nabla \phi - \frac{2e^2}{m} |\psi|^2 (\mathbf{A} + \mathbf{A}_\omega), \qquad (4)$$

where we define an effective vector potential due to the rotation

$$\mathbf{A}_\omega = \frac{m}{e} \boldsymbol{\omega} \times \mathbf{r}. \qquad (5)$$

Integrating around a closed path Γ contained within the superconducting ring, which rotates at an angular velocity ω, we obtain

$$\frac{m}{e^2 n_s} \oint_\Gamma \mathbf{j} \cdot d\boldsymbol{\ell} = n \frac{h}{2e} - \int_{S_\Gamma} \mathbf{B} \cdot d\mathbf{s} - \frac{2m}{e} \boldsymbol{\omega} \cdot \mathbf{S}_\Gamma, \qquad (6)$$

where $n_s = 2|\psi|^2$ and S_Γ is the area bounded by Γ. As usual we have required ψ to be single-valued. In Fig. 2 we plot the flux through the path Γ given by the second term on the right of Eq. (6) against the angular rotation ω. For $\omega = 0$ Eq. (6) is exactly the fluxoid quantization equation, and for $n = 0$ it gives the London moment relation for the magnetic flux versus spin speed.

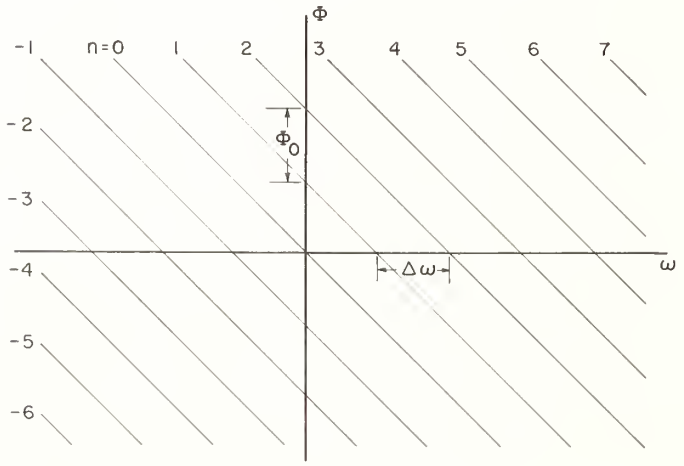


FIGURE 2. *The magnetic flux versus angular velocity for a rotating superconducting ring. The magnetic flux nulls, where an integer number of flux quanta cancels the London moment flux, are shown.*

If we assume the only source of $\mathbf{B}$ to be the super current $\mathbf{j}$ (any constant background field will cancel in Eq. (7)), then there exists an ω_n for each number of flux quanta n such that $\mathbf{j} = 0$ and $\mathbf{B} = 0$. These correspond to the $\phi = 0$ points on each parallel line in Fig. 2. We then obtain the simple relation

$$h/m = 4S\Delta\omega, \qquad (7)$$

where $\Delta\omega = \omega_n - \omega_{n-1}$ is the angular velocity difference between successive nulls. Thus, by measuring a macroscopic area and a frequency difference, a high precision determination of h/m is possible.

The theoretical analysis [22] for the relativistic corrections to Eq. (6) can be summarized to order $(v/c)^2$ by replacing m by an effective mass m' in Eq. (7) (not to be confused with the solid state effective mass), where now

$$m' = m\left[1 + \frac{<T>}{mc^2} - \frac{e\Phi_{in}}{mc^2}\right]. \qquad (8)$$

The first correction term is due to the mass-velocity shift and for an isotropic superconductor is given by the expectation value of the electron kinetic energy $<T>$ averaged over all Fermi surface states (actually states within Δ, the superconducting energy gap, of the Fermi surface). It is a factor of 5 or more larger than the second correction term, which depends on the inner crystal electrostatic potential Φ_{in}. Net mass increases of 100 ppm are predicted for many superconductors. Detailed solid state band structure calculations using existing techniques are needed and are expected to give 1% or better estimates for the correction terms in Eq. (8).

3. Measurement Scheme and Apparatus

The measurements are made with a cryogenic helium gas bearing constructed entirely of fused quartz (see Fig. 3). The rotor is a 25 mm radius sphere truncated 1 mm above its equator. A precision 25 μm wide by 50 nm (500 Å) thick superconducting ring will be deposited onto the equatorial plane using photoresist masking techniques. Gas bearings in the flat top and spherical bottom of the housing support the rotor with clearances of 15 μm. Helium spin-up channels are located in the sides of the housing, and spin speeds of up to 150 Hz in

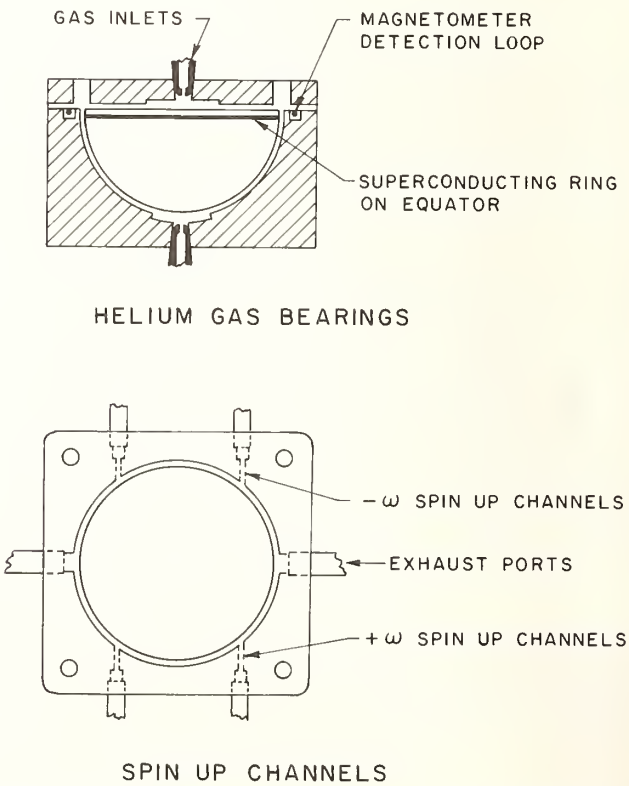


FIGURE 3a. *The all-fused-quartz helium gas bearing used for the h/m_e measurements shown schematically.*

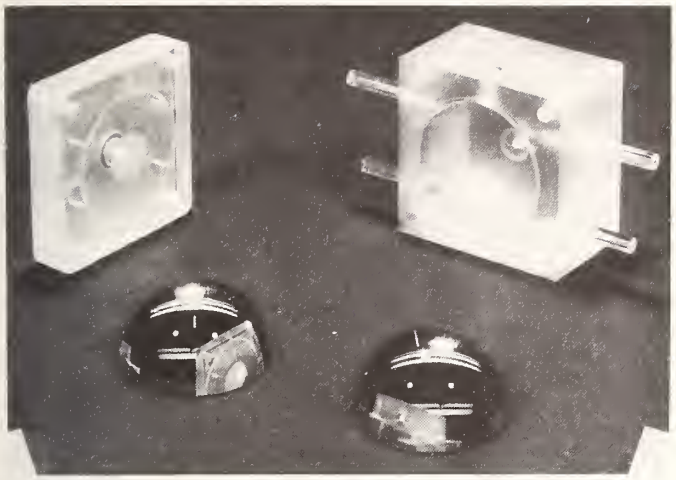


FIGURE 3b. *The all-fused-quartz helium gas bearing used for the h/m_e measurements photographed just prior to assembly.*

both directions will be possible. The rotor spin speed is measured using a fiber optic readout mounted through the bearing top flat. The intensity of the reflected light is modulated by a frosted radial line in the rotor top surface as it passes under the fiber bundle.

The magnetic flux nulls are measured using a superconducting pick-up loop placed in a groove cut into the top of the lower housing. This loop is coupled to a SQUID magnetometer and is capable of resolving 0.01 of a flux quantum through the rotating ring in a bandwidth of several hertz.

To obtain the necessary stability in the SQUID detection circuit, an ultra-low magnetic field environment is essential. A technique developed by one of us (Cabrera) uses expandable, superconducting, cylindrical shields [18]. In bootstrap fashion newly folded shields are carefully cooled inside previously expanded ones until the desired field reduction is achieved. Shields with absolute fields below 10^{-11} T are produced routinely. No observable changes occur in the field levels even after several years of use.

The determination of the frequency spacing between nulls $\Delta\omega$ is accomplished by counting the number of magnetic field nulls (integer and fractional parts) between the top spin speeds in the two opposite rotation directions (see Figs. 4 and 5). The spin speed is slowly varied in the pattern of Fig. 4 (upper) over several hours as the ring is continuously modulated through its transition temperature every 1 to 10 seconds (Fig. 4 (middle)) by controlling the spin-up gas temperature. Since for our large diameter ring $\phi_0^2/2L$ is smaller than one half kT, after many heatings and coolings a Gaussian distribution of quantum states results around the equilibrium value (Fig. 4 (lower)). The rms is about 3 ϕ_0. However, each time the ring becomes superconducting a single well-defined quantum state from the distribution is measured (Fig. 5). A change in frequency of 1 hertz corresponds to passage through 70 nulls. Thus, for our initially cautious spin speed maximum of $\pm$ 50 hertz, 7000 nulls will be included. Every null need not be observed; only a sufficient number so as to unambiguously determine the total (about every tenth one).

Our initial data acquisition scheme is shown schematically in Fig. 6 with the timing sequence given in Fig. 7. The bearing spin speed is controlled using two Reedair

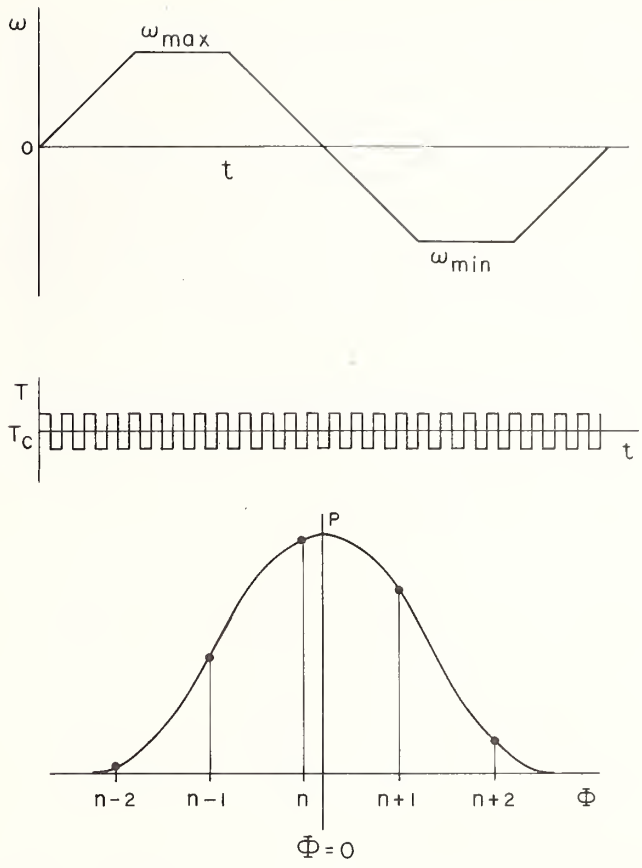


FIGURE 4. *The spin speed is slowly varied (upper) as the ring temperature is continuously modulated through T_c (middle). The large ring inductance causes a Gaussian distribution of states about $\phi = 0$ at any ω (lower).*

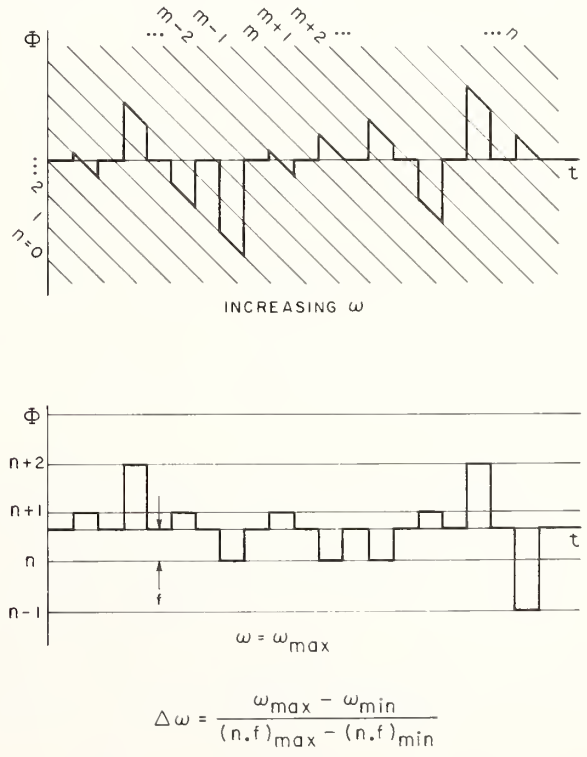


FIGURE 5. *During the ramping phase the nulls must be unambiguously counted (upper) with sufficient integration times at ω_{max} or ω_{min} for a 1% resolution between nulls (lower).*

valves placed in the room temperature portion of the spin-up gas lines. These valves are miniature, on-off, reed valves. The states of the two valves are complementary and are controlled by an oscillator with fixed frequency and variable duty cycle. The pulse width can be set using a 10-turn potentiometer or an external ramp generator. Since the valve operation is complementary, the total gas flow is constant. However, the distribution of gas between the $+\omega$ and $-\omega$ spin-up channels varies linearly with the pulse width, thus providing a straightforward method of spin rate control. The constant gas flow simplifies temperature stabilization.

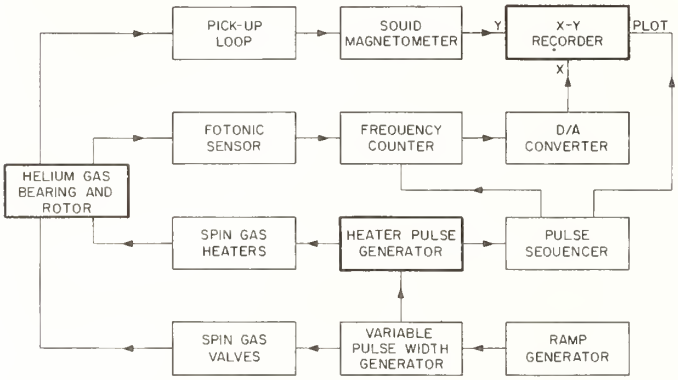


FIGURE 6. *Block diagram of data acquisition scheme.*

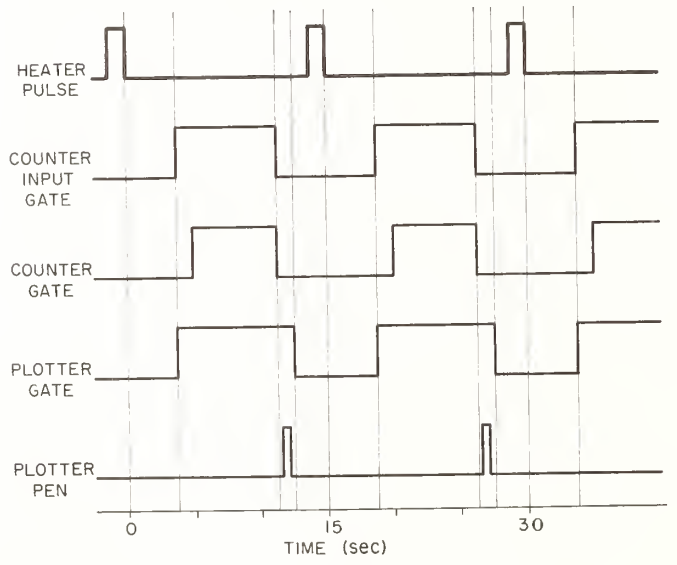


FIGURE 7. *Data acquisition timing sequence.*

Temperature modulation of the bearing is accomplished using manganin wire heater pellets located in the spin-up gas lines. Heater power is supplied by the heater pulse generator, which operates in a short-pulse mode or at a 50% duty cycle. The oscillator in this unit is the clock for the heating cycle (typically 15 seconds), as is shown in Fig. 7. The heater pulse generator provides heater power pulses of selected amplitude and length (up to one second) and an adjustable dc bias power level by which the base temperature can be set to just below the ring transition temperature. The heater power going to each spin-up line is apportioned according to the amount of gas flowing in that line. The heater has power ranges of 50 mW and 500 mW.

To determine $\Delta\omega$ in Eq. (7), the bearing spin speed and the magnetic flux in the ring are measured simultaneously (Figs. 5 and 6). The spin speed signal from the KD 100 Fotonic sensor fiber optic output is read using an HP 5316A frequency counter. The counter digital output signal is then converted to an analog voltage and plotted on the x-axis of an x-y recorder. At the same time, the magnetic flux is measured with a SQUID magnetometer coupled to a pick-up loop embedded in the bearing housing and is plotted on the y-axis of the recorder.

The data acquisition process is controlled with the pulse sequencer (Fig. 6). The clock period is given by the heater pulse generator (Figs. 6 and 7). A short time after the heater cycle begins (typically 2 seconds, allowing the return to thermal equilibrium) both the external counter input gate and the plotter gate open. The next Fotonic sensor pulse triggers the counter and opens its internal gate. An integral number of pulses are then timed; the first pulse that occurs after the nominal internal gate time has elapsed simultaneously closes the internal and external counter gates. During the next second the plotter pen is lowered and raised, simultaneously recording the frequency and magnetometer readings. The complete sequence repeats with the next heater cycle. When the 50% heater duty cycle is used, two data points are obtained during each heater period, alternately measuring the rotating ring in its normal and superconducting states, as shown schematically in Fig. 5.

The pulse sequencer has also been designed for operation under microcomputer control, and this will soon be implemented. The clock signal will then be supplied by the computer to the pulse sequencer. In addition to the x-y plotter, the data will be recorded by the microcomputer using the digital interfaces on the counter and the magnetometer.

Our frequency resolution of 1% between adjacent nulls together with spin speeds corresponding to 7000 nulls will allow a measurement of $\Delta\omega$ to an accuracy of 2 ppm. Since improvements in this accuracy to better than 0.1 ppm appear relatively straightforward, for the foreseeable future the dominant error in the experimental determination of h/m through Eq. (7) is the area measurement.

In a collaboration with the National Bureau of Standards in Gaithersburg, MD, we have determined the equatorial plane areas for our two rotors to a 1σ accuracy of 2 ppm. This work has been reported in detail elsewhere [24] and will only be summarized here. The thickness (50 nm) and width (25 μm) of the superconducting ring deposited on the equatorial plane of the rotors are chosen so that all possible cross-sectional areas bounded within the superconductor have the same value to a resolution of 2 ppm. Thus, a determination of the ring area to this accuracy requires only measurements in the equatorial plane of the bare rotor.

The area of a plane figure cannot be determined by any number of diametral measurements, since there exist figures of constant width other than a circle, all having the same circumference but varying in area [24]. To avoid this ambiguity the radius of the figure as a function of angle must be measured with respect to a fixed point. The area uncertainty in diametral measurements is eliminated by using the rotor sphericity data to correct one absolute diameter measurement to the true mean diameter.

A computer-aided Talyrond roundness measuring system is used to determine both the rotor profile and its least squares reference circle with a precision of better than 5 nm (50 Å). If matched by corresponding accuracy in the individual diametral measurement, the area could

be determined to 1 part in 10^7. However, at present, absolute diameter measurements are typically limited to 1 or 2 ppm, although significant advances seem likely in the near future. The NBS is assisting us in these measurements using a horizontal diameter measuring machine based on the HP Laser Interferometer.

To compensate for the expansion of the quartz rotor from 20 °C to 4 K (about 17 ppm), measurements made through an additional collaboration with the NBS will be taken on samples from the original rotor blanks.

4. Measurements of the London Moment, Flux Quantization, and h/m_e

Our inital cryogenic testing of the complete system as described in Section 3 has been very successful. The helium gas bearing performed well at STP with nominal gas flows of 500 ml per minute through the bearings. A spin gas flow of 2000 ml per minute easily allowed spin speeds up to 15 hertz, our self-imposed initial maximum. For these first runs, a 40 μm diameter niobium wire ring was glued to the top of the hemispherical rotor. Its nominal diameter of 49.49 mm allows absolute area measurements to only several parts per thousand, thus limiting our present determination of h/m_e through Eq. (2) to this level. However, this ring has allowed us to successfully test the measurement scheme and demonstrate resolutions for $\Delta\omega$ approaching parts per million. Using the SQUID magnetometer readout system and our superconducting ultra-low magnetic field shield (5×10^{-12} T), we were able to obtain high resolution data of the magnetic moment induced in a superconductor by rotation (the London moment, Eq. (1)) and to observe clearly flux quantization through the ring ($\omega = 0$ in Eq. (6)). These data, presented below, were taken with the semiautomatic data acquisition system described in Section 3.

Flux quantization through this 50 mm diameter ring is observed most convincingly by biasing the temperature within several millikelvin of the ring's transition temperature $T_c = 9.2$ K and observing the thermally induced transitions between neighboring quantum states. A time plot is shown in Fig. 8, where the temperature is slowly varied around T_c with about a one minute period. A one hertz bandwidth for the SQUID output is used. The quantized flux steps are shown and labeled on the right side of the figure, whereas the equivalent response at the SQUID is labeled on the left side. As far as we are aware, these measurements represent the largest area for which flux quantization has been demonstrated.

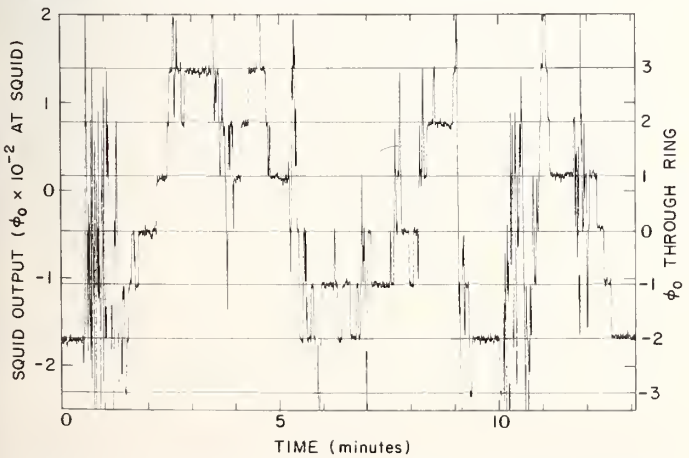


FIGURE 8. *Thermally induced transitions between quantum states clearly demonstrating flux quantization.*

Measurements of the magnetic moment of the niobium ring versus its spin speed were obtained at a constant temperature of 4.5 K. The heaters were disconnected from the heater pulse generator, which continued to provide the clock frequency for data acquisition. One data point was generated every 5 seconds. Figure 9 is a plot of the magnetic flux through the ring, in units of the flux quantum ϕ_0, versus spin speed. It was taken with the spin speed stabilized to within several millihertz using feedback. This feedback scheme was implemented by replacing the ramp generator signal to the variable pulse width generator by the sum of the ramp generator signal and the output of the frequency counter after it had been converted to analog and inverted for negative feedback (refer to Fig. 6). Note that the zero intercept of the plot is not arbitrary but was determined by recording the magnetometer reading at zero velocity.

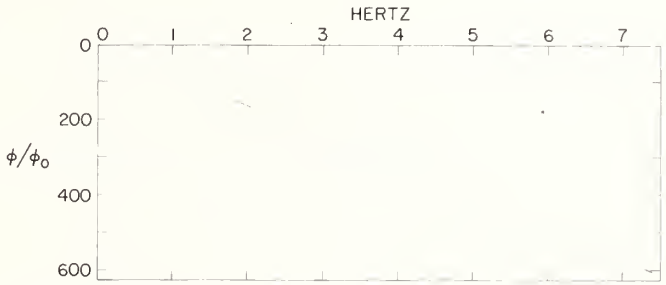


FIGURE 9. *Rotationally induced magnetic field (London moment) versus spin speed.*

The high resolution data of Fig. 10 shows both the flux quantization and the London moment and is exactly an experimental plot of Eq. (6) and Fig. 2. These data were taken in a one hertz bandwidth with the temperature held 100 millikelvin below T_c for each parallel line. After each line was recorded, repeated heatings and coolings were performed to obtain successive quantum states toward each end of the plot. From the data of Fig. 10 alone, $\Delta\omega$ of Eqs. (2) and (7) is determined to several parts per thousand. This resolution per null demonstrates a statistical error in $\Delta\omega$ of a part per million when 7000 null crossings are counted (± 50 hertz). The quantum state labels such as 208 are *not* arbitrary but are referred to the flux zero as defined when the ring is in its normal state. The ability to resolve several millihertz changes in the rotational velocity of a superconductor as seen in a magnetic flux readout (Fig. 10) represents an improvement by two orders of magnitude over all previously reported measurements.

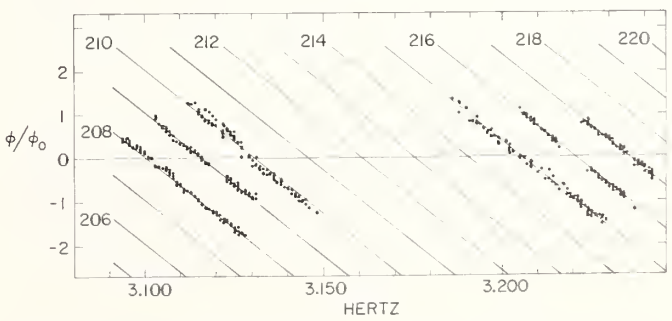


FIGURE 10. *Null crossings for several adjacent quantum states allowing a determination of $\Delta\omega$ to high precision.*

Our determination of h/m_e from these data (Fig. 10) is limited to 3 parts per thousand because of the difficulty in absolutely measuring the area bounded by our preliminary wire ring. At this level it is in agreement with the previous work of Parker and Simmonds and with the accepted value obtained from measurements on phenomena not involving superconductivity. We are now working on depositing a precision ring on the rotor equatorial plane to obtain accuracies at the several parts per million level.

5. Conclusions

In initial cryogenic measurements of the induced magnetic flux from a rotating superconducting ring (London moment) using a precision helium gas bearing, we have demonstrated an improvement in resolution by two orders of magnitude over previously reported observations. Magnetic moment signals from angular frequency changes of 3 millihertz were detected. We have also clearly shown the flux quantization levels for our 50 mm diameter ring. To our knowledge this represents the largest area in which quantization has been observed. The combined measurements determine h/m_e to 3 parts per thousand. The value is in agreement with the accepted value (known to several tenths of a part per million) and is presently limited by our preliminary wire ring area uncertainty. Using this apparatus with minor improvements, we will next make measurements on a precision niobium ring and expect to reach a resolution of 2 ppm in the determination of $\Delta\omega$ during the next year. Combining this result with our area determinations corrected to 4 K, we expect an overall uncertainty of 2 ppm on the right side of Eq. (7).

Through comparisons with the presently accepted value of h/m_e for the free electron at rest, we will for the first time be investigating relativistic corrections to the electron mass within the superconducting lattice. Detailed solid state calculations based on the relativistic theory of Cabrera, Gutfreund, and Little are expected to match the experimental accuracy. Various elemental superconductors will be measured to verify these calculations.

After the consistency of the calculations and measurements has been established, an independent value of h/m_e for the free electron can be obtained. Future improvements in the determination of areas together with more detailed calculations should provide a sub-ppm value, possibly approaching 0.1 ppm. Then the simple relation

$$\alpha^2 = \frac{2R_\infty}{c}\,(h/m_e)$$

would provide an independent value of the fine-structure constant α approaching an accuracy of 0.05 ppm, where the Rydberg constant R_∞ and the speed of light c are known to several parts per billion[1]. Then these measurements of h/m_e would provide a significant input datum for the next least-squares adjustment of the fundamental constants.

We greatly appreciate the assistance given to us by the National Bureau of Standards in performing the measurements of absolute diameters and for many helpful discussions. At Stanford previous work on the design of the helium gas bearing was contributed by G. B. Hess, now at the University of Virginia, and its construction was due to D. E. Davidson. The authors also wish to thank the Stanford Relativity Gyroscope group and, in particular, C. W. F. Everitt and W. M. Fairbank, for technical assistance and much encouragement.

[1]Refer elsewhere in these proceedings to the many excellent papers on these measurements.

References

[1] R. Becker, F. Sauter, and C. Heller, Z. Phys. 85, 772 (1933).
[2] F. London and H. London, Proc. R. Soc. London, Ser. A : 147, 71 (1935).
[3] F. London, *Superfluids*, Vol. 1 (John Wiley & Sons, Inc., New York, 1950).
[4] A. F. Hildebrandt, Phys. Rev. Lett. 12, 190 (1964).
[5] A. F. Hildebrandt and M. M. Saffren, in *Low Temperature Physics, LT9*, Ed. by J. G. Daunt *et al.* (Plenum Press, New York, 1965), p. 459.
[6] C. A. King, J. B. Hendricks, and H. E. Rorschach, ibid., p. 466.
[7] M. Bol and W. M. Fairbank, ibid, p. 471.
[8] N. F. Brickman, Phys. Rev. 184, 460 (1969).
[9] K. Oide and H. Hirakawa, J. Phys. Soc. Jpn. 43, 1087 (1977).
[10] B. S. Deaver and W. M. Fairbank, Phys. Rev. Lett. 7, 43 (1961).
[11] R. Doll and M. Nabauer, Phys. Rev. Lett. 7, 51 (1961).
[12] W. H. Parker, B. N. Taylor, and D. N. Langenberg, Phys. Rev. Lett. 18, 287 (1967).
[13] R. F. Dziuba, B. F. Field, and T. F. Finnegan, IEEE Trans. Instrum. Meas. IM-23, 264 (1974).
[14] V. Kose, IEEE Trans. Instrum. Meas. IM-25, 483 (1976).
[15] B. N. Taylor, W. H. Parker, and D. N. Langenberg, Rev. Mod. Phys. 41, 375 (1969).
[16] J. E. Zimmerman and J. E. Mercereau, Phys. Rev. Lett. 14, 887 (1965).
[17] W. H. Parker and M. B. Simmonds, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.), Spec. Publ. 343 (Aug. 1971), p. 243.
[18] B. Cabrera and F. van Kann, Acta Astronaut. 5, 125 (1978).
[19] B. Cabrera and J. T. Anderson, in *Future Trends in Superconductive Electronics*, Ed. by B. S. Deaver, Jr., *et al.*, AIP Conf. Proc. No. 44 (American Institute of Physics, New York, 1978), p. 161.
[20] C. W. F. Everitt, in *Experimental Gravitation*, Ed. by B. Bertotti (Academic Press, New York, 1973), p. 331; J. A. Lipa, Proceedings of the International School of General Relativity Effects in Physics and Astro-Physics: Experiments and Theory (3rd Course), Document MPI-PAE/Astro 138, Max Planck Institute for Physics and Astrophysics, Munich, November 1977, p. 129.
[21] G. J. Siddall, SME Tech. Pap. Ser. IQ78–167 (1978).
[22] B. Cabrera, H. Gutfreund, and W. A. Little, Phys. Rev. B 25, 6644 (1982).
[23] For example: P. G. deGennes, *Superconductivity of Metals and Alloys* (W. A. Benjamin, Inc., New York, 1966); also A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, San Francisco, 1971).
[24] B. Cabrera and G. J. Siddall, Precis. Eng. 3, 125 (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

High Precision Measurement of the Electron Compton Wavelength (h/m_e) Using Cryogenic Metrological Techniques

J. C. Gallop, B. W. Petley, and W. J. Radcliffe

Division of Quantum Metrology, National Physical Laboratory, Teddington, Middlesex TW11 OLW, England

This paper reports progress on the NPL measurement of the Compton wavelength of the electron, which relies on the unique properties obtainable in cryogenic environments. The method is to trap a magnetic field in a superconducting tube. The experimentally measured quantities are the spin precession frequency of the ^{3}He nuclei in the trapped field, and the cross sectional area of the tube. Novel cryogenic metrology is used for both of these measurements.

Key words: cryogenic metrology; free precession; fundamental constants; SQUIDS; superconducting cavity.

1. Introduction

This paper reports progress on an experiment to determine the Compton wavelength of the electron using a number of novel cryogenic techniques. It is one of a number of interesting fundamental constant determinations which rely essentially on the phenomenon of fluxoid quantization in superconducting rings [1]. These measurements involve a length determination at liquid helium temperatures, and the difficulty of carrying out such a measurement with high precision has so far inhibited serious attempts in this area. The work reported here attempts to use the liquid helium environment as an asset to the measuring process, rather than a hindrance. A measurement of h/m_e by a rotating superconductor method is being undertaken at Stanford University [2]. This involves an entirely different dimensional measuring technique and also a relativistic correction. If both measurements of h/m_e can be taken to sub-ppm accuracy the two results should allow an interesting comparison of their radically different methods.

2. Basis of the Experimental Technique

If a uniform axial magnetic field is applied to a superconducting tube of inner cross-sectional area A as it is cooled through its transition temperature and is then removed, the magnetic field will remain trapped in the tube. The single valued nature of the macroscopic wavefunction describing the superconducting electron pairs requires that the magnetic flux linking the tube must be quantized in units of $\Phi_o = h/2e$. If a sample containing polarized ^{3}He nuclei (gyromagnetic ratio γ_{He}) is inserted into the uniform field region at the center of the tube the free precession frequency ω_{He} of the nuclei can be measured [3]. It is related to the magnetic field B_o by the expression

$$\omega_{He} = \gamma_{He} B_o . \tag{1}$$

Now the flux linking the tube is related to B by the expression

$$\int \mathbf{B} \cdot d\mathbf{A} = n \Phi_o ,$$

where n is an integer. As a result of the variation of the circulating current density along the axis of the tube the section of the tube than would be the case for a uniformly wound solenoid of the same dimensions. This means that to a very good approximation the above expression may be replaced by the much simpler relationship

$$B_o = n \Phi_o / A . \tag{2}$$

One of the novel aspects of this work is in the method used to measure the cross sectional area A of the tube. The tube is converted, by the addition of niobium endplates, into a high Q superconducting microwave cavity. A measurement of a number of the resonant frequencies of the cavity allows both its leading dimensions and certain aspects of its irregularities in shape to be calculated with high precision. If the deduced cavity radius is r then

$$A = \pi r^2 = \pi c^2 g^2 \{f_{lmn}\} ,$$

where $g\{f_{lmn}\}$ is some function of the set of measured resonant frequencies $\{f_{lmn}\}$ [4]. Combining Eqs. (1) to (3) we have then

$$\gamma_{He} \Phi_o = \pi \omega_{He} (c^2 g^2 \{f_{lmn}\})/n . \tag{4}$$

The ratio γ_{He}/γ_p has been measured rather accurately [5], to about 0.02 ppm, and the left hand side of Eq. (4) may be written as

$$(\gamma_{He}/\gamma_p)\, \gamma_p\, \Phi_o = (\gamma_{He}/\gamma_p)(\mu_p/\mu_B)\, h/2m_e , \tag{5}$$

where μ_p/μ_B is the proton magnetic moment expressed in Bohr magnetons. Thus we have finally

$$h/m_e = 2\pi \omega_{He}/n\, (c^2 g^2 \{f_{lmn}\})\, [(\mu_p/\mu_B)\, (\gamma_{He}/\gamma_p)] , \tag{6}$$

where the parameters outside the square brackets are to be determined experimentally and those inside are already known with an accuracy much better than 1 ppm. Thus the complete experiment requires three main goals: (1) to produce a uniform magnetic field in a superconducting cylinder; (2) to observe nuclear magnetic resonance of ^{3}He in this uniform field; and (3) to determine the cross sectional area of the cylinder by a microwave resonance technique. The progress made towards fulfilling each of these requirements will be reviewed in the following three sections.

3. Observation of NMR in Liquid ^{3}He

The reason for choosing ^{3}He as the NMR probe to measure the trapped field is that it is one of a very few systems with non-zero nuclear magnetic moment which

are liquid or gaseous at superconducting temperatures and thus allows a narrow resonance line to be detected. The ^{3}He free precession frequency has been measured with the prototype system with a precision of about ±0.02 Hz. When the full field uniformity has been achieved this should be improved by at least a factor of 10^2, so that it seems reasonable to expect that the frequency will be measured to within ±0.0002 Hz, corresponding to an uncertainty in the magnetic field measurement of

$$dB = d\omega_{He}/\gamma_{He} = 5 \times 10^{-12} \text{ T}$$

A precision of 0.3 ppm in the frequency measurement would then require a trapped field of only 2×10^{-4} T. At a temperature of 3 K and using a SQUID as a low noise detector of the free precession signal, it is estimated that the number of spins required to provide adequate signal to noise ratio is not more than 10^{19}, corresponding to a sample volume of less than 1 mm^3.

A prototype system has been built which has been used to detect, for the first time, free precession signals from the polarized ^{3}He spin system in low fields, using a SQUID. A schematic diagram of the system is shown in Fig. 1. The ^{3}He sample is contained in a pyrex bulb, around which is wound the primary coil (of high conductivity normal metal). This coil is connected to a multiturn coil made of superconductor which serves as the signal coil for an rf SQUID, which has a bandwidth in excess of ω_{He}. The sample is then polarized in a transverse magnetic field of around 10^{-2} T provided by a coil arrangement external to the superconducting tube. When sufficient polarization of the ^{3}He spins has been attained this field is removed and the sample is moved axially into the uniform field region of the tube. At present the NMR signal has been observed in a magnetic field in the range of $10^{-5} - 10^{-4}$ T, trapped in the outer, lead, superconducting tube (see Fig. 1). The decaying free precession signal detected with a SQUID is shown in Fig. 2. This corresponds to a beat signal between the SQUID output signal and a clock frequency of 1.0 kHz. The decay time of the signal is believed to be set by the homogeneity of the trapped magnetic field, a T_2^* time of ~1 s., corresponding to a field uniformity of around 3×10^{-7} T/m. This is compatible with the value of the gradient component dB_z/dz, measured with a flux-gate magnetometer which could be moved along the axis of the tube.

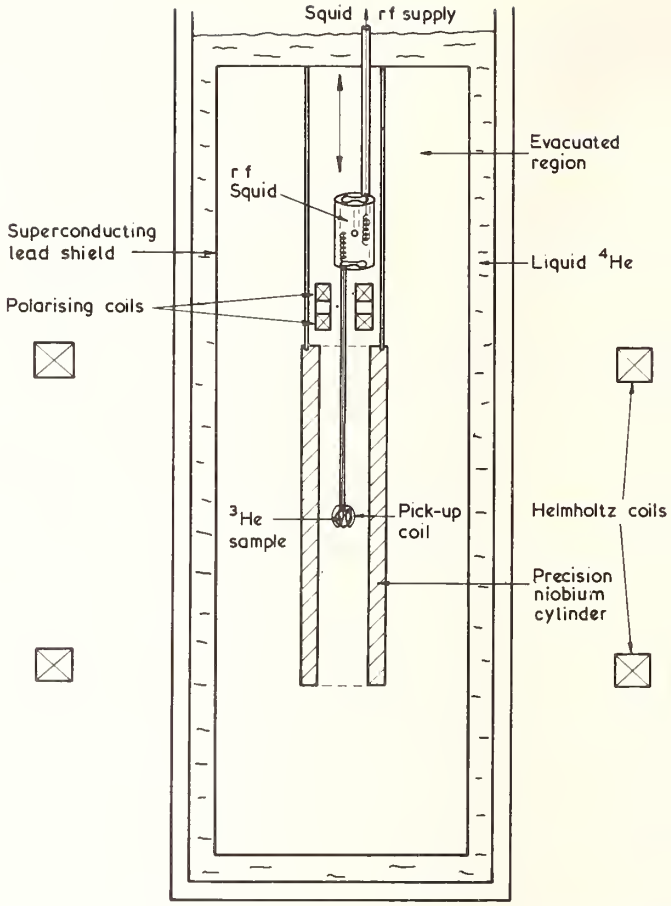


FIGURE 1. *A schematic diagram of the apparatus for detecting the ^{3}He free precession signal using a SQUID.*

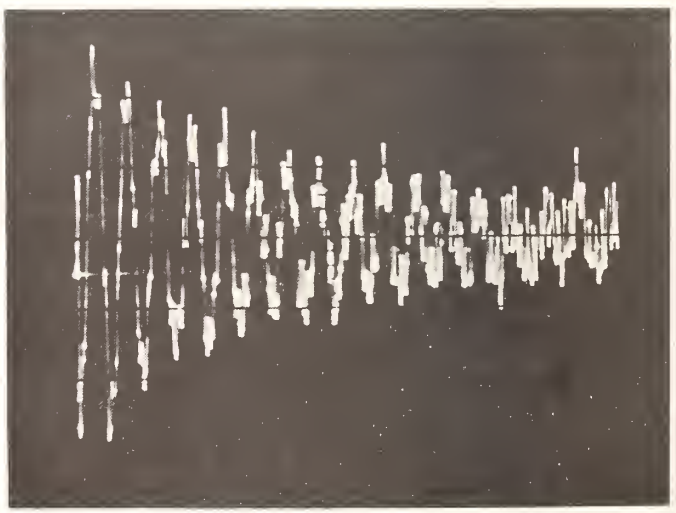


FIGURE 2. *Free precession signal from ^{3}He sample at 4.2 K in a field of 10^{-5} T, detected with a SQUID.*

4. Measurement of Tube Diameter by a Microwave Resonance Technique

The superconducting tube must be circular in cross section and of uniform diameter, at least over a region near the center of length comparable with the diameter. This requirement stems from the need for the trapped field to be uniform enough to allow resolution of the NMR signals corresponding to integer values $n\Phi_0$ and $(n + 1)\Phi_0$ of the quantized flux. The maximum diameter of the tube is set by the criterion that n must be determined for the minimum value of B_0 consistent with an accurate determination of ω_{He}. That is if the center of the resonance line can be determined to an accuracy of, say, 3×10^{-7} then $n \lesssim 10^6$. Taking $dB = 5 \times 10^{-12}$ T (section 3)

$$r \lesssim (\Phi_0/\pi dB)^{1/2} = 10 \text{ mm} .$$

Boring and lapping techniques available at NPL have allowed a circular hole of 23 mm diameter to be produced in a solid block of niobium (about 120 mm in length) with a tolerance of 0.2 μm over a 50 mm length. It is believed that at any point on the axis the deviations from circularity are an order of magnitude less than this. (A check on the circularity should be available from the measured frequencies of such microwave mode families as the TE_{31n}, which do not possess complete rotational symmetry about the tube axis [4].)

Already the general technique for measuring the mean

diameter of a tube at room temperature by this method has been demonstrated. It has also been possible to show that changes in mean radius of the tube with axial position can be measured. A brass cylindrical cavity was used, which had a step change of about 30 μm in radius over its central region [4]. Work is now proceeding on measuring resonant modes of the superconducting niobium cavity. About 30 microwave resonances in the range 7 to 20 GHz have been detected with this cavity, in vacuum, from room temperature down to about 2 K. These measurements indicate that the shape of the cylinder does not distort significantly as it is cooled down. The thermal expansion coefficient has been measured and compared with other experimental values, giving agreement to within about 10%. The highest Q values so far observed are of the order of 2×10^6, which is quite sufficient to allow the resonance line center to be determined to 1 part in 10^7, the required precision for an overall experimental accuracy of 0.3 ppm. In view of the large number of resonant frequencies existing in the microwave frequency range available to us (7–20 GHz) a semiautomatic system is being developed to determine frequency, linewidth, and amplitude of the resonances as a function of temperature.

5. Production of a Uniform Trapped Magnetic Field

One of the crucial requirements for the proposed high precision determination is that it should be possible to trap a uniform magnetic field through the superconducting tube such that the fluxoid quantization number is constant along the length of the tube. That is to say no flux lines must leave the tube through the superconducting wall. Preliminary experiments, in which a tube of lead was cooled through its superconducting transition in the earth's magnetic field by cooling it in a cryostat by immersion in liquid helium, showed that although rather uniform magnetic fields can be trapped in this way ($db_0/dz \sim 3 \times 10^{-7}$ T/m), this uniformity is about two orders of magnitude worse than would be expected if no flux lines left the tube through the wall. The experimental arrangement now being used consists of an outer lead tube, which is closed at both ends with non superconducting plates, which serves as a vacuum can as well as a superconducting shield. Inside this is suspended the precision niobium tube. The support structure is of low thermal conductivity and the niobium tube is supplied with a heater so that after the lead tube has become superconducting the niobium tube may be heated through its transition temperature before being allowed to become superconducting again in a controlled way via adjustment of the heater current. The uniformity of the trapped field can be measured (although not its absolute value) by moving the pick-up coil of a SQUID magnetometer along the axis of the niobium tube. In this way it has been shown that the "heat flushing" technique just described can improve the trapped field uniformity by at least two orders of magnitude and the effect of single flux lines leaving the tube through the walls can be clearly seen. A region of about 20 mm in length has been produced over which constant flux quantum number has been achieved, and the axial field gradient $dB_z/dz < 6 \times 10^{-10}$ T/m. The total change in trapped field over this length is less than 2 parts in 10^7. It is planned to extend this region over the whole length of the tube, by improved heat flushing or perhaps by high temperature annealing of the precision niobium tube.

6. Summary

A novel method is described here for measuring h/m_e (or equivalently the electron Compton wavelength $h/m_e c$) based on magnetic flux quantization and dimensional measurement using a high Q superconducting cavity. Although a number of questions remain to be answered it is hoped that the method will allow an ultimate precision of better than 1 ppm. At this level of precision a value for h/m_e would provide a useful input datum for the next evaluation of the fundamental constants. When combined with R_∞, the Rydberg, it provides a value for the fine structure constant α^{-1}, with an accuracy twice that of h/m_e. This constant also provides a direct link between $h/2e$, the unit in which the maintained volt is measured, and the unit in which the maintained ampere may be monitored. Such a measurement is of interest in its own right but perhaps more importantly it allows scope for investigating new techniques in cryogenic metrology. It is believed that the progress made so far with this work suggests there are a number of interesting applications of the new metrology involved. These include a SQUID free-precession magnetometer [6], a novel form of nuclear gyroscope [7], and the application at room temperature of the microwave dimensional measurement technique [4].

References

[1] B. N. Taylor, W. H. Parker, and D. N. Langenberg, Rev. Mod. Phys. 41, 375 (1969).

[2] B. Cabrera, S. Benjamin, and J. T. Anderson, these proceedings.

[3] J. C. Gallop, J. Phys. B 11, L93 (1978).

[4] J. C. Gallop and W. J. Radcliffe, J. Phys. E 14, 461 (1981).

[5] Y. I. Neronov and A. E. Barzakh, Sov. Phys. JETP 48, 769 (1978).

[6] J. C. Gallop and W. J. Radcliffe, Physica B + C, 107, 621 (1981).

[7] R. Adams, S. P. Potts, J. C. Gallop, and W. J. Radcliffe in *SQUID '80—Superconducting Quantum Interference Devices and their Applications*, Ed. by H. D. Hahlbohm and H. Lubbig (Walter de Gruyter, Berlin, 1980), p. 509.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Preliminary Determination of h/m_n

E. Krüger, W. Nistler, and W. Weirauch

Physikalisch-Technische Bundesanstalt, Bundesallee 100, D-3300 Braunschweig, F.R.G.

A precisely measured value of the ratio h/m_n (Planck constant divided by the neutron mass) would be a new input parameter for the least-squares adjustment of the fundamental constants. h/m_n can be determined by measuring the wavelength and the velocity of reactor neutrons. It is expected that this measurement can be carried out with a relative uncertainty of about 1×10^{-6}, using a method described in this paper. The method has been successfully tested in a preliminary determination of h/m_n and the results are reported.

Key words: fundamental constants; polarized neutrons; precision measurement.

1. Introduction

A measurement of the ratio h/m_n (Planck constant divided by the neutron mass) with a relative uncertainty of about 1×10^{-6} is being prepared in the Physikalisch-Technische Bundesanstalt. This quantity has not so far been precisely measured.

The last least-squares adjustment of the fundamental constants was carried out by Cohen and Taylor [1] in 1973. If h/m_n is calculated from the values of the constants recommended by these authors, the relative uncertainty of the result is 1.7×10^{-6}. This is about the same as expected for the measurement described here. Nevertheless, the experiment would be of great importance for altering the status of h/m_n from that of a deduced quantity to that of an additional input parameter of the least-squares adjustment. The reliability of the adjustment could thus be clearly improved.

The wavelength λ and the velocity v of a neutron are related by the de Broglie equation:

$$h/m_n = \lambda v \ . \qquad (1)$$

Hence, h/m_n can be determined by measuring both the wavelength and the velocity of neutrons.

Several methods have been proposed for this experiment, in particular by Stedman [2], Mezei [3] and Weirauch [3–7]. In all of them, λ is determined by Bragg reflection of the neutrons in a single crystal, and v is measured by a time-of-flight method. However, various techniques have been proposed for carrying out the velocity measurement with the desired high accuracy.

The method finally proposed by Weirauch [5–7] is briefly described in this paper. Furthermore, a preliminary measurement of h/m_n carried out to test the feasibility of the experiment is reported.

2. The Measuring Method

2.1 The Principle of the Measurement

For measuring the velocity of the neutrons, a beam of polarized neutrons is modulated by periodically changing the direction of the polarization vector. After a path of several meters, the neutrons are back-reflected by a silicon single-crystal. They thus pass the modulator for a second time, and the beam is again modulated. The modulator defines the beginning as well as the end of a flight path. The flight time follows from the phase difference between the two modulations.

The wavelength of the neutrons is determined by the Bragg reflection in the silicon crystal. Silicon should be used because almost perfect crystals can be grown from this material and its lattice parameter is known with the required small uncertainty [8–11].

2.2 The Beam of Polarized Neutrons

A polarized and monochromatic neutron beam is produced by Bragg reflection in a magnetized crystal of Heusler alloy (see Fig. 1, H1). The polarization of the reflected neutrons is characterized by the polarization vector **p**, which is antiparallel to the magnetic induction $\mathbf{B}_H$ in the crystal. In the ideal case of a completely polarized neutron beam, the spins of all the neutrons are parallel to **p**.

The polarized neutrons move in a magnetic guide field (magnetic induction $\mathbf{B}_0$) which is parallel to $\mathbf{B}_H$. The guide field is required in order to avoid a depolarization by magnetic stray fields.

The beam of the reflected neutrons and $\mathbf{B}_0$ define the y and z coordinates given in Fig. 1.

The neutrons traverse a flipping coil F which reverses **p**. It produces a constant magnetic induction $\mathbf{B}_F$. **p** precesses in the coil around the resulting magnetic induction $\mathbf{B}_r = \mathbf{B}_0 + \mathbf{B}_F$. $\mathbf{B}_F$ is chosen so that $\mathbf{B}_r$ is parallel to the x axis and has just the value which causes a precession by the angle π during the passage of the neutrons.

The neutrons then hit a second Heusler crystal H2, magnetized in the same direction as the first one. The probability that a neutron is reflected by this crystal depends on the direction of **p**. It is zero if **p** is parallel to $\mathbf{B}_H$ and maximal for the opposite direction. As **p** is now parallel to $\mathbf{B}_H$, the neutrons traverse the crystal without reflection.

Coil C1, which the neutrons pass next, produces a constant magnetic induction $\mathbf{B}_{C1}$ parallel to the x axis with $B_{C1} = B_0$. The resulting magnetic induction forms an angle of 45° with the x and the z axis. **p** precesses around

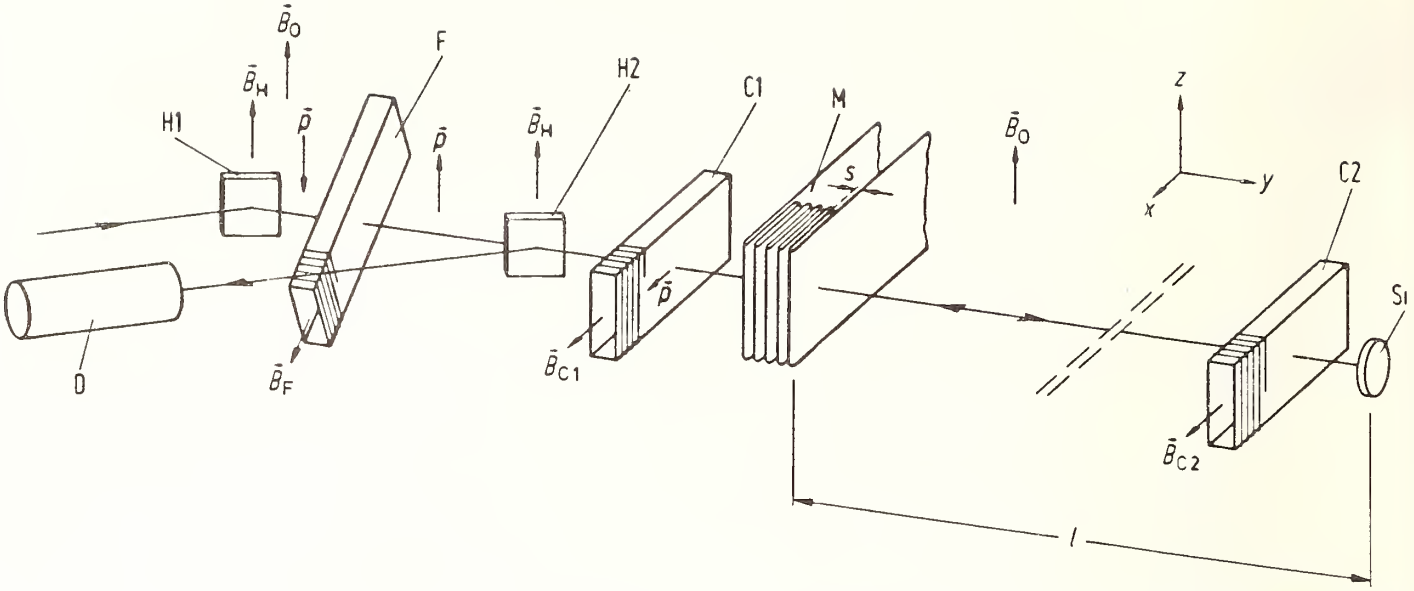


FIGURE 1. *Arrangement for measuring h/m_n. (H1, H2) Heusler crystals, (F) flipping coil, (C1, C2) $\pi/2$ coils, (M) meander coil, (Si) silicon crystal, (D) detector. $\boldsymbol{B}_0$ magnetic induction of the guide field, $\boldsymbol{B}_H$ magnetic induction magnetizing a Heusler crystal, $\mathbf{p}$ polarization vector if the meander coil is not operating.*

this direction. Just at the moment when it is parallel to the x axis, the neutrons leave the coil. Coils of this type were first used by Mezei [12]. They are called $\pi/2$ coils because they change the direction of $\mathbf{p}$ by the angle $\pi/2$.

The neutrons traverse the modulating assembly M and a second $\pi/2$ coil C2. Both these arrangements should be disregarded for the moment.

The neutrons are then reflected by the silicon crystal mentioned above and return to coil C1.

Along the entire path, from leaving coil C1 to reaching it again, $\mathbf{p}$ precesses in the guide field. But the distance between coil C1 and the silicon crystal is such that the entire precession angle is an integer multiple of 2π. Hence, $\mathbf{p}$ is again parallel to the x axis when the neutrons enter coil C1 for the second time.

In coil C1, $\mathbf{p}$ is turned back into the direction of the z axis. The neutrons are then incident on the Heusler crystal H2 for a second time. Because $\mathbf{p}$ is parallel to $\mathbf{B}_H$, the neutrons are again not reflected. Thus, they cannot reach the detector D.

2.3 Modulating the Neutron Beam

The assembly M used for modulating the neutron beam consists of an aluminum foil, bent to form a meander. It is therefore referred to as a meander coil in the following.

The meander coil forms a close sequence of single coils, each with one turn. If an electric current is passed through them, the magnetic inductions in neighboring coils are antiparallel. If alternating current is used, there is a phase difference of π between the magnetic inductions in neighboring coils. An arrangement of this type, supplied with direct current, was used by Drabkin *et al.* [13] for the construction of a velocity filter for polarized neutrons.

To modulate the direction of $\mathbf{p}$, the meander coil is supplied with high-frequency alternating current. The coil field is parallel to the guide field. The thickness s of the single coils is such that the neutrons pass each of them in half a period of the alternating current. As a consequence, the neutrons find the same variation of the magnetic induction in all the single coils.

$\mathbf{p}$ precesses in the meander coil due to two parallel magnetic fields, the guide field and the coil field. The first part of the precession has already been taken into account. Hence, only the precession caused by the coil field is considered in the following.

The magnetic inductions in the single coils are assumed to vary sinusoidally with time. As a consequence, the Larmor frequency of $\mathbf{p}$ also varies sinusoidally with time, while the neutrons traverse a single coil. The angle by which $\mathbf{p}$ precesses during the passage depends on the time at which the neutrons enter the coil.

The precession angle is the same in all the single coils. In the whole meander coil it adds up to the angle

$$\Phi(t) = \hat{\Phi} \cos(2\pi\nu t), \tag{2}$$

$$\hat{\Phi} = 4\mu_n \hat{B}_m N/h\nu \tag{3}$$

(ν frequency of the electric current, t time, μ_n magnetic moment of the neutron, $\hat{B}_m$ amplitude of the magnetic induction in the coil, N number of single coils of which the meander coil consists).

In principle, the modulation could be produced by a single coil. This case has been discussed in a previous paper [14] but following from Eq. (3), it cannot be applied here: in the proposed experiment $\hat{\Phi} \geq 0.61\pi$ and $\nu \approx 1$ MHz is required. Because $\hat{B}_m$ is limited for technical reasons to a few millitesla, these values can only be achieved with $N \gg 1$.

2.4 Measuring the Velocity of the Neutrons

When the neutrons leave the meander coil for the first time, the direction of $\mathbf{p}$ is modulated as given by Eq. (2). $\mathbf{p}$ is at all times in the xy plane (see Fig. 1). Returning from the silicon crystal, the neutrons traverse the meander coil for a second time. The direction of $\mathbf{p}$ is again modulated, this time by the angle

$$\Phi(t + 2l/v) = \hat{\Phi} \cos[2\pi\nu(t + 2l/v)]. \tag{4}$$

l is the distance between the meander coil and the silicon crystal.

The two modulations add up to the total modulation

$$\Phi_{tot}(t, l) = \Phi(t) + \Phi(t + 2l/v)$$
$$= 2\hat{\Phi} \cos(2\pi\nu l/v) \cos[2\pi\nu(t + l/v)]. \quad (5)$$

Φ_{tot} is a periodic function of both the time and the distance l.

After the second modulation, the neutrons traverse coil C1 in which **p** is turned from the xy plane into the yz plane. $\Phi_{tot}(t, l)$ is then the angle between **p** and the z axis.

The neutrons again hit the Heusler crystal H2. Because the direction of **p** varies, the probability of reflection varies, too. Hence, the time-dependent neutron current

$$I(t, l) = (I_0/2)\,[1 - \cos\Phi_{tot}(t, l)] \quad (6)$$

reaches the detector. I_0 is the neutron current if **p** is antiparallel to the z axis.

The mean neutron current $\bar{I}$ is measured in the detector. It is given by

$$\bar{I}(l) = (I_0/2)\,\{1 - J_0[2\hat{\Phi}\cos(2\pi\nu l/v)]\}. \quad (7)$$

J_0 is the zero-order Bessel function.

Figure 2 shows $\bar{I}(l)/I_0$ as calculated for $\hat{\Phi} = 0.61\,\pi$ and $\hat{\Phi} = 1.22\,\pi$. In the experiment, $\hat{\Phi}$ will probably be somewhere between these values.

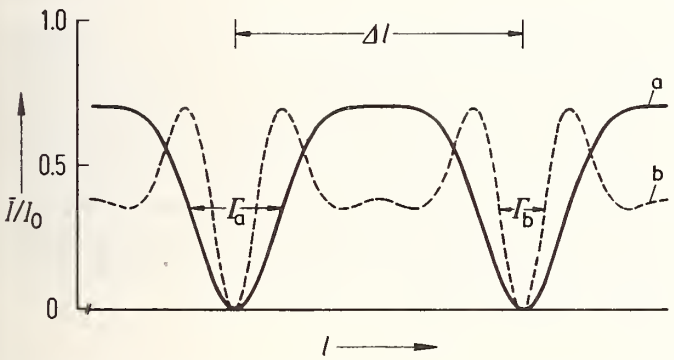


FIGURE 2. *Calculated ratio $\bar{I}/I_0$ as a function of the distance l. Parameter: (a) $\hat{\Phi} = 0.61\,\pi$, (b) $\hat{\Phi} = 1.22\,\pi$.*

The velocity of the neutrons is given by

$$v = 2\Delta l\,\nu, \quad (8)$$

as follows from Eq. (7). Δl is the difference between values of l belonging to neighboring primary minima of $\bar{I}(l)/I_0$ (see Fig. 2). The secondary minima which appear in curve b should not be used for determining v.

In order to determine v with high accuracy, an integer multiple of Δl must be measured. For this purpose, two minima of $\bar{I}(l)/I_0$ should be determined, the l values of which differ by several meters. To change l, the meander coil must be moved, not the silicon crystal; a displacement of the crystal would also change the angle by which **p** precesses in the guide field.

2.5 Eliminating Long-Term Instabilities of the Magnetic Guide Field

In section 2.2, it has been assumed that the polarization vector precesses by an integer multiple of 2π on the way from coil C1 to the silicon crystal and back again. If this condition is not met, the maximum and minimum values of $\bar{I}(l)/I_0$ differ in general by less than shown in Fig. 2. The precession angle and thus the guide field must therefore be constant during the entire measuring time, which will run into many weeks. This difficulty can be avoided by means of a technique which is similar to the spin-echo technique first used by Mezei [12].

To eliminate the instabilities of the guide field, the additional $\pi/2$ coil C2 is required. It is adjusted in the same way as coil C1.

On the way from coil C1 to coil C2, the polarization vector **p** precesses by the angle

$$\phi_1 = 2n\pi + \alpha \qquad (-\pi < \alpha \leq \pi) \quad (9)$$

due to the guide field. n is an integer. α gives the angle between **p** and the x axis. When the neutrons enter coil C2, **p** has the direction shown in Fig. 3a.

In coil C2, **p** is turned into the yz plane. Its component p_y is reversed, α becomes the angle between **p** and the z axis (see Fig. 3b).

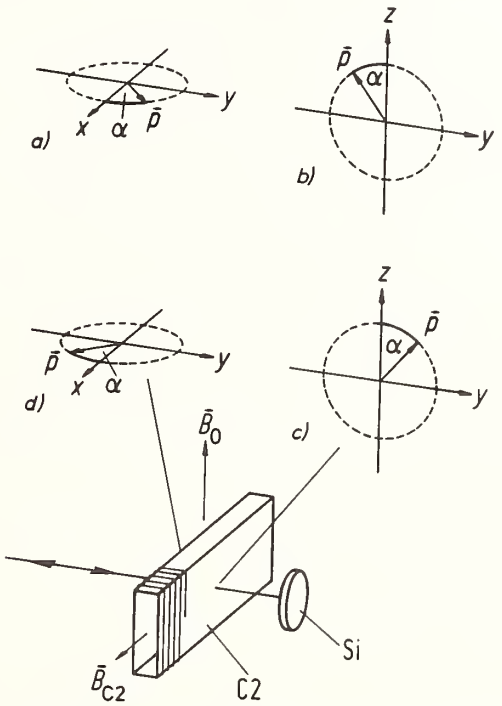


FIGURE 3. *Polarization vector **p** at the surfaces of the $\pi/2$ coil C2, (a, b) before the reflection in the silicon crystal Si, (c, d) after the reflection.*

Coil C2 is positioned at such a distance from the silicon crystal that **p** precesses by a small odd multiple of π while the neutrons traverse this distance twice. The precession results in the changing of the sign of α (see Fig. 3c).

While the neutrons traverse coil C2 for the second time, **p** is turned back into the xy plane. p_y is once more reversed (see Fig. 3d).

In Fig. 3d α has the opposite sign as compared with Fig. 3a. Hence, ϕ_1 has been replaced by

$$\phi_2 = 2n\pi - \alpha \quad (10)$$

On the way from coil C2 back to coil C1, **p** again precesses by ϕ_1. Between leaving and entering coil C1, it therefore precesses altogether by

$$\phi = \phi_2 + \phi_1 = 4n\pi. \quad (11)$$

Hence, **p** has the same direction at the beginning and at the end of the path, independent of α.

B_0 must be constant only during the time the neutrons need for twice traversing the distance between coil C1 and the silicon crystal. Long-term stability is needed only where the $\pi/2$ coils are positioned and in the space between coil C2 and the silicon crystal. However, p precesses in these regions solely by small angles. Thus, the absolute change of the precession angle caused by a deviation of B_0 from the correct value is also small. Because this is the only important quantity, the requirements on the stability of B_0 are rather low.

3. Preliminary Measurement

A preliminary determination of h/m_n was carried out at the reactor of the Physikalisch-Technische Bundesanstalt in Braunschweig. The arrangement was as schematically shown in Fig. 1.

The (331) reflection of silicon was used, corresponding to the wavelength $\lambda = 0.25$ nm.

The meander coil was made of aluminum foil, 60 μm thick and 90 mm wide. It consisted of $N = 99$ single coils, each of which was $s = 1.065$ mm thick.

From s and the value of the neutron velocity known beforehand, it followed that the modulation frequency had to be $\nu = 745$ kHz. Measurements showed that the meander coil actually operated best at this frequency. The amplitude of the electric current in the meander coil was 40 A. Under these conditions, the modulation amplitude is $\hat{\Phi} = 0.69\ \pi$, as determined in preparatory measurements.

The meander coil could be moved by 13 cm in order to vary l. The displacement was measured by a laser interferometer.

The guide field was produced by Helmholtz coils of rectangular cross-section. Its magnetic induction was $B_0 = 1$ mT.

Figure 4 shows the mean neutron current $\bar{I}$ measured as a function of the displacement L of the meander coil. Each point was measured for a period of 8 hours. 110 minima have been omitted between the two parts of the measurement, following from the previously known value of h/m_n.

A curve of the theoretically expected shape was adjusted to the measured values by means of a least-squares method. From the positions of their minima it follows that:

$$h/m_n = (3.9519 \pm 0.0040) \times 10^{-7}\,\mathrm{m^2\,s^{-1}}. \quad (12)$$

The given uncertainty is the standard deviation. In accordance with the preliminary character of the measurement, no systematic uncertainties have been considered.

The measured value of h/m_n is in good agreement with the more accurate one calculated from the recommended values of h and m_n [1], which is

$$h/m_n = (3.956034 \pm 0.000007) \times 10^{-7}\,\mathrm{m^2\,s^{-1}}. \quad (13)$$

4. The Final Experiment in Preparation

The preliminary measurement of h/m_n confirmed the feasibility of the experiment. However, to achieve the desired accuracy, a high-precision apparatus must be constructed.

In the final measurement, l will be varied by about 10 m instead of about 10 cm as in the preliminary one. The other data will probably not differ very much from those given above.

In order to measure h/m_n with a relative uncertainty of about 1×10^{-6}, the positions of the minima of $\bar{I}$ must be determined with an uncertainty of a few percent of the width Γ at half-minimum (see Fig. 2). This is possible, but the uncertainty in measuring $\bar{I}$ must be much smaller than in the preliminary experiment (see Fig. 4).

To lower the uncertainty of $\bar{I}$, more neutrons must be counted. The final measurements will therefore be carried out at the high-flux reactor of the Institut Laue-Langevin in Grenoble.

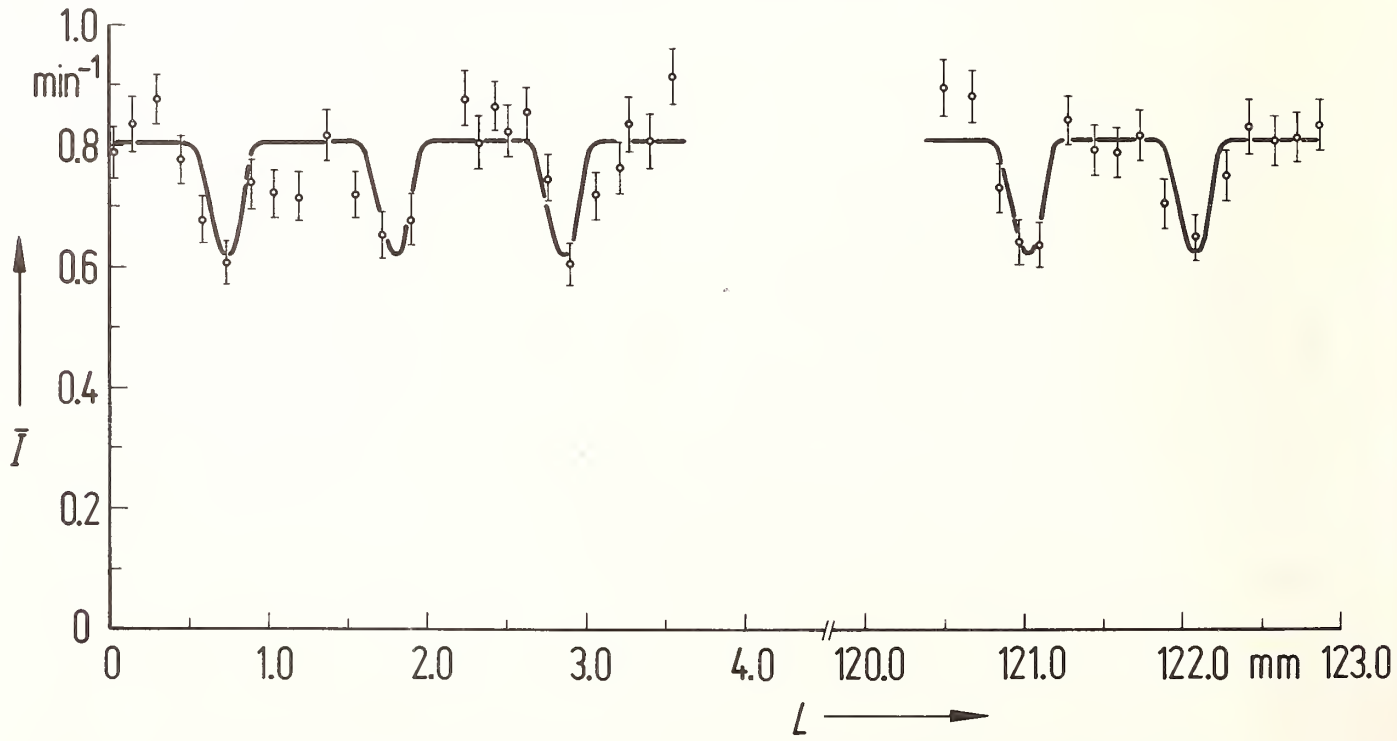


FIGURE 4. *Mean neutron current $\bar{I}$ measured as a function of the displacement L of the meander coil.*

5. Conclusion

A preliminary measurement of h/m_n has been carried out. It showed that there is a good chance of determining this quantity with a relative uncertainty of about 1×10^{-6}. However, a high-precision apparatus is required, and the measurement must be carried out at a high-flux reactor. Such an experiment is in preparation.

The authors would like to thank M. M. W. Berne and H. Schneider who helped to construct the apparatus for the preliminary measurement.

References

[1] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[2] R. Stedman, J. Sci. Instrum., Ser. 2 (J. Phys. E) 1, 1186 (1968).
[3] F. Mezei and W. Weirauch, research proposal No. 01-003R (1975), Institut Laue-Langevin, Grenoble.
[4] W. Weirauch, Nucl. Instrum. Methods 131, 111 (1975).
[5] W. Weirauch, in *Fundamental Physics with Reactor Neutrons and Neutrinos* (Inst. Phys. Conf. Ser. No. 42), Ed. by T. von Egidy (The Institute of Physics, Bristol and London, 1978), p. 47.
[6] W. Weirauch, E. Krüger, and W. Nistler, in *Neutron Spin Echo* (Lect. Notes Phys. 128), Ed. by F. Mezei (Springer, Berlin, 1980), p. 94.
[7] W. Weirauch, to be published.
[8] R. D. Deslattes and A. Henins, Phys. Rev. Lett. 31, 972 (1973).
[9] R. D. Deslattes, A. Henins, H. A. Bowman, R. M. Schoonover, C. L. Carroll, I. L. Barnes, L. A. Machlan, L. J. Moore, and W. R. Shields, Phys. Rev. Lett. 33, 463 (1974).
[10] R. D. Deslattes, A. Henins, R. M. Schoonover, C. L. Carroll, and H. A. Bowman, Phys. Rev. Lett. 36, 898 (1976).
[11] P. Becker, K. Dorenwendt, G. Ebeling, R. Lauer, W. Lucas, R. Probst, H.-J. Rademacher, G. Reim, P. Seyfried, and H. Siegert, Phys. Rev. Lett. 46, 1540 (1981).
[12] F. Mezei, Z. Phys. 255, 146 (1972).
[13] G. M. Drabkin, V. A. Trunow, and V. V. Runow, Zh. Eksp. Teor. Fiz. 54, 362 (1968) [Sov. Phys. JETP 27, 194 (1968)].
[14] K. Weise, A. Mehl, and W. Weirauch, Nucl. Instrum. Methods 140, 269 (1977).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

One has to remember that some errors are random for one person and systematic for another.
P. Vigoureux

The Assignment of Uncertainties to the Results of Experimental Measurements

Jörg W. Müller

Bureau International des Poids et Mesures, Pavillon de Breteuil, F-92310 Sèvres, France

After a brief review of some of the current ways of indicating in a quantitative manner the uncertainties which are associated with an experimental measurement, the main shortcomings of some of the usual practices are indicated. The present situation is unsatisfactory and results in frequent misinterpretations, rendering, for instance, a critical data evaluation a cumbersome task.

The traditional practice of distinguishing between "random" and "systematic" uncertainties is one of the main causes of confusion in the statement of uncertainties. In fact, this separation is in no way absolute since it is likely to be changed by a subsequent user. A subdivision, if required, should only depend on the way the numerical value of the uncertainty was arrived at. As for its effect on the final result, due care is automatically taken by a correct application of the well-known general law of error propagation, for which the only input data needed are the best available estimates for the variances and covariances of the measured quantities.

BIPM recently organized a meeting of experts, chosen by the national standardizing laboratories, for discussion of the relevant problems. This resulted in the establishment of five recommendations concerning the statement of uncertainties which are included in an appendix.

Key words: covariances; experimental uncertainties; variances.

1. Introduction

A conference on fundamental constants is clearly not the place where anybody will expect me to explain at length why the presentation of an experimental result should always include information on its supposed precision or accuracy. Indeed, from a scientific point of view a numerical result of a measurement which is reported without any indication concerning the claimed precision is nearly useless as it does not allow a meaningful comparison either with other determinations of the same quantity or with theory, where this is possible.

There are certainly many examples from daily life for which this strict rule can be relaxed with little or no loss of information. Thus, if we hear that 26 people have been killed in an air crash, nobody expects an uncertainty to follow the statement; on the other hand, if we are told that a Roman amphora is 2000 years old we will not take this number at its face value. However, these are limiting cases and they are in no way typical of the usual situation prevailing in physical measurements.

Since precise measurements, at least in a field such as astronomy, have been performed for more than 3500 years, as is evidenced, for instance, by the famous Venus tablets of Ammizaduga and the very sophisticated Babylonian theories for the apparent movement of the moon and the planets [1], it is quite astonishing that no explicit mention was ever made of the limited precision inherent in the observations. In fact, even the closely related basic notion of probability is completely absent in the whole thinking of Antiquity and (less surprisingly) in the Middle Ages, although the zeal of the populations for all sorts of dice-playing and similar random games is well attested. Historians may have a plausible explanation for this strange situation (perhaps religion), but the fact remains quite remarkable.

Obviously, the situation today is basically different—even if compared to the one prevailing, say, only 50 years ago when a scientist could still know personally most of the people working in his field and therefore base his judgment on the reliability of results on the appreciation of the personality of his fellow scientists. I ignore whether in practice this system always worked perfectly well, but in any case it could not be applied today when we may have to use data of hundreds of workers about whom we normally know nothing at all. It therefore has become necessary to replace the possible prestige of a name associated with the result of a measurement by an "objective" measure of its value, and this is done by the assignment of uncertainties. This practice is not necessarily an improvement, but rather an unavoidable development which we have to accept. The old system, where applicable, may still give useful hints.

2. Generalities

Unfortunately, the assignment of uncertainties has been plagued for many decades by a number of problems, some of which may be difficult to avoid, but to a large extent by others which just result from a lack of agreement on what we wish to describe and the manner in which we do so. The main trouble is that, although most experimenters maintain the practice of assigning an uncertainty to their measured values, it is often far from obvious to the reader what this number is exactly supposed to mean. This continues to cause serious headaches for any compiler or critical evaluator of data who is obliged to put the various indications on a common footing, for instance in order to assign appropriate statistical weights to data which he is comparing among themselves or with some theory. Misinterpretation of published error assignments is frequent (and often unavoidable for lack of in-

formation) and may result in the recommendation of doubtful "best values."

Let us first try to understand what the main problems are and where they come from. On the basis of this knowledge, we shall then suggest a possible way to overcome some of the difficulties, at least for the most current practical cases. It will be appropriate to mention right at the beginning that in a field such as this upon which so many words have already been expended, it is hardly possible to say anything which is really new; perhaps this is not even wanted, because there is no lack of a variety of well-intentioned advice. Rather, what we need is a process for sorting out signal from noise, possibly by using as guidelines such old stand-bys as common sense, simplicity, coherence and usefulness, while taking care not to violate basic laws of physics and statistics, where applicable. It is hoped that by proceeding in such a manner we can see more clearly the main issues and then, possibly, indicate some way out of a situation which seems, at present, inextricable. The literature on error statements is enormous, but usually quite unrewarding. Some of the more interesting opinions can be found in the proceedings of the first PMFC conference [2], especially in the informal discussions from which Vigoureux' quotation is borrowed.

Our task is rendered difficult by a number of circumstances which, one might think, are quite irrelevant to the real problems as they belong to psychology rather than to physics, but unfortunately this is not so. First we have to cope with the fact that everyone considers oneself to be an expert. Secondly, no one likes to change opinions and habits, especially if these are available in printed form—and they are numerous. Finally, there exist quite a number of recipes issued by groups of experts, most of which are clearly contradictory to each other and ill-founded, but sometimes backed by influential national or even international organizations. They have in common that they take a narrow and simplistic point of view, ignore the similar problems met in other fields, and arrive at descriptions which allow the evaluation of uncertainties with an amazing precision (at least three significant figures) even for the poorest data. Their world is invariably Gaussian and continuous. The fact that an evaluated uncertainty also has an uncertainty is commonly overlooked.

In view of this state of affairs, which was no secret to the well informed, it was certainly a bold initiative for Dr. E. Ambler, a member of the International Committee of Weights and Measures, to suggest that the International Bureau of Weights and Measures (BIPM) should make a serious attempt to clarify the situation and, if possible and with the help of the national laboratories, arrive at useful and generally acceptable rules for the statement of experimental uncertainties. Equally gratifying was the readiness of Dr. P. Giacomo, then director designate of BIPM, to undertake this mission. It was felt that the first need was to know the position of the national laboratories in this controversial field. This was done by setting up a "Questionnaire on uncertainties" where, after an introduction to the problems, eight questions were asked, each preceded by a comment which placed the problem in the appropriate context and recalled some of the difficulties involved. The main purpose of this enquiry was to choose carefully the subjects which could be usefully discussed at a later stage by an ad-hoc Working Group. Such a preliminary sorting out was considered to be unavoidable in view of the divergence of prevailing opinions. It was thought advisable to avoid purely philosophical or mathematical discussions as well as topics related to terminology, although the border line is often a vague and subjective one. In February 1978 the questionnaire was sent to 32 national standardizing laboratories inviting them to submit answers to the questions, and to 5 international organizations for information. By the end of the year BIPM had received responses from 21 countries, included in which were replies from all the large national laboratories. Their analysis has been the subject of a detailed report [3].

3. A Look at Some of the Main Problems at Issue

Instead of listing the questions asked and the replies received, which would be a rather tedious exercise, we prefer to mention some of the basic items under discussion. In addition, various details in the replies of the laboratories, as a result of subsequent internal discussion and better understanding, no longer correspond to their present point of view. Hence the document [3], although still of historical interest, is now superseded by later developments which it has no doubt initiated. The present situation will be discussed in Section 5.

It was clear from the beginning that the traditional subdivision of uncertainties (or "errors," as they used to be called) into "random" (or statistical) and "systematic" contributions would become one of the most controversial topics. This grouping is so strongly entrenched in the thinking of physicists and engineers that at first most of them may have considered it a waste of time and effort to question its good foundation. And yet a careful enquiry into the origins of the embarrassing confusion invariably pointed to this problem as a likely source of trouble.

In the prevailing opinion a systematic uncertainty (or "error") is characterized by the fact that its influence on the final result is independent of the number n of measurements taken. This is in clear contrast to the random contributions which diminish with the square root of n. This seems to be such a striking difference that a clear distinction between these two types of uncertainties appears to be a necessity. On the other hand, this view is somewhat difficult to reconcile with the observation that there also exist cases where the type of an error apparently changes, depending upon the use to which it is put. This is such a disagreeable situation in the eyes of an error-class purist, that the only way to cope with it was to ignore it. However, the fact exists and it is quite easy to find situations in which changes can take place in either direction. For those who believe in the need to combine uncertainties differently according to their type, this observation is disastrous. How does this come about and what can we do? The diagnosis of the trouble is simple: it is due to the fact that the traditional classification of uncertainties depends upon the further use we intend to make of them, and in general this cannot be known in advance. Any classification, if needed at all, should obviously only depend on the past, not on the future.

Although many people may be well disposed to accept the above arguments, they are not willing to change their habits. After all, the system worked quite well for most practical cases and how would we know whether or not one has to divide an uncertainty by the square root of n? This is an eminently practical question for which we have to find a convincing answer. It has been given before [4], but since this is a crucial point it may be justified to repeat the main arguments, perhaps best by means of a practical example. We assume the reader to be familiar with the general rules of error propagation, the elements of which are summarized in Appendix A.

Let us consider a series of weighings made with a balance, the numerical results of which are represented by the n results x_j. There may be some doubt as to the accuracy of the calibrated masses involved in these measurements. If such an error should exist, it would clearly affect all the measurements. Let us assume, for the sake of argument, that the effect is such that instead of the measured values x_j one should consider the corrected quantities

$$y_j = x_j + \alpha \ , \tag{1}$$

where α is a possible common shift (other types of corrections might in fact be more realistic, but this is the simplest situation). Since we assume that all known corrections have already been applied, the most probable value of α is zero, but there remains an uncertainty $s_\alpha > 0$.

The experimental uncertainty of a single measurement x_j is denoted by s_x and we assume all measurements x_j to be independent of each other.

For the "function f" (cf. Appendix A) we simply choose the mean value, hence

$$\overline{y} = \frac{1}{n}\sum_j y_j = \frac{1}{n}\sum_j x_j + \alpha \ . \tag{2}$$

The quantity looked for is the uncertainty of $\overline{y}$. This can be evaluated in various ways which depend on the choice of the variables.

Let us first put

$$\overline{y} = f_1(x_1, x_2, \ \ldots \ , x_n; \alpha) \ . \tag{3}$$

Since x_j and α are not correlated, all covariances vanish. As $\partial f_1/\partial x_j = 1/n$ and $\partial f_1/\partial\alpha = 1$, we readily obtain from Eq. (A3) for the variance of $\overline{y}$

$$s_{\overline{y}}^2 = \Sigma \left[\frac{1}{n} s_x\right]^2 + (1\, s_\alpha)^2 = \frac{1}{n} s_x^2 + s_\alpha^2 \ . \tag{4}$$

Another possible choice for the variable is

$$\overline{y} = f_2(y_1, y_2, \ \ldots \ , y_n) \ . \tag{5}$$

In this case the quantities needed in Eq. (A3) are the variances

$$s_j^2 = \mathrm{var}(y_j) = s_x^2 + s_\alpha^2$$

and the covariances

$$s_{jk} = \mathrm{cov}(y_j, y_k) \ .$$

For the latter it can be shown (see for instance Ref. [5]) that, since both E(α) and cov(x_j, x_k) vanish, we simply have

$$s_{jk} = s_\alpha^2 \ .$$

As furthermore $\partial f_2/\partial y_j = 1/n$, one finds by means of Eq. (A3)

$$s_{\overline{y}}^2 = \frac{1}{n^2}[n(s_x^2 + s_\alpha^2) + n(n-1)s_\alpha^2] = \frac{1}{n} s_x^2 + s_\alpha^2 \ ,$$

which is identical with Eq. (4).

Nobody will be astonished by this result as it corresponds exactly to what we would have expected. Yet, it is remarkable for several reasons. First, we note that the outcome does not depend upon the choice of the variables and that there was no need to decide whether an uncertainty is of a certain type. All we have to do is to describe correctly the experimental situation and then to apply the general rules of error propagation. This then yields automatically the correct answer [Eq. (4)]. Of course, in interpreting the result one can now say that s_x behaves as a "random" uncertainty and s_α as a "systematic" one, but this behavior is a result of the application made. If the experiment had consisted of a comparison of n balances of different types among themselves, the contributions s_α would have been treated as random uncertainties. An additional piece of information is contained in Eq. (4) and should not pass unnoticed, namely that both the "random" and the "systematic" contributions are added "in quadrature" (and not linearly or otherwise), and finally also that the "systematic error" s_α appears in a form which corresponds to a standard deviation. While all these things are obvious here and come out naturally, they were hotly debated items in those approaches that are not explicitly based on Eq. (A3). The above discussion reveals them as ill-founded. In fact, if variances and covariances are properly taken into account, there is no need to introduce different types of uncertainties.

4. On the Maximum Limit Concepts

There can be little doubt that there are cases where the numerical assessment of an uncertainty poses a real problem. This is in particular the case when it cannot be deduced from repeated measurements, because it is not possible to perform them. In such a situation of incomplete knowledge—which in fact is not an exceptional, but rather the normal case—the physicist has to make use of his experience and imagination. With the help of his knowledge of the theoretical basis and the practical realization of a given experiment he has to examine carefully all possible causes of error, to eliminate them or to apply corrections. Usually this can only be done in an approximate way; the possible residues have then to be taken as uncertainties. One always hopes that nothing of real importance was forgotten, but there is obviously no guarantee. As an alternative (but somewhat jocular) classification it has therefore been suggested to distinguish only between those uncertainties which have been taken into account (although perhaps not adequately) and the other ones which have been overlooked. If this were a realistic approach, it would be excellent. In practice, unfortunately, the second category can only be grasped at the very end, usually as a result of some disturbing discrepancy. Yet the idea should be kept in mind as it recalls that from a scientific point of view this is certainly the most interesting class: it permits one to uncover a previously unsuspected influence. It is to a large extent by the discovery of such errors that experimental physics advances. In any case, the challenge to resolve inconsistencies is a most important stimulus for new experimental or theoretical work, as has been illustrated with numerous examples by Taylor [6]. See also Ref. [7]. This also reveals the danger implied by the practice occasionally suggested of just enlarging a stated uncertainty in order to avoid possible difficulties later. But obviously, it is not by concealment that problems are solved.

In the general confusion on the assignment of uncertainties it has sometimes been taken for granted that "systematic errors" automatically belong to the class where an objective numerical value is not available, since the personal judgment of the experimenter comes in. As a subjective evaluation was ruled out as unscientific, it was claimed that the best and only thing to do was to assign "maximum" uncertainties which were chosen large enough to guarantee that the "true value" lies practically always within the corresponding limits. This is clearly a recipe which is simple to apply, but its usefulness is questionable. Since a "maximum limit" becomes "safer" when enlarged, it is an ill-defined quantity for which, for instance, no useful propagation rules can be

indicated. Nevertheless, repeated attempts have been made to save this concept, in particular by associating it with a specific probability density. By interpreting the upper and lower limits as bounds between which the "true" value must be located and by further assuming that, as a result of our ignorance, any point in this interval has the same chance to be the "good one," one arrives at a constant probability density between the limits. Since the standard deviation for a rectangular density of total width $2a$ is given by $a/\sqrt{3}$, it has been suggested that we consider this as the corresponding uncertainty. Indeed, some have taken this approach seriously enough to produce detailed numerical tables for the convolution of a rectangular density with a Gaussian function. Unfortunately, there is no sound basis for such an assumption, which is also known as the postulate of Bayes. On the contrary, thanks to Jaynes [8] some practical counter examples are now known for *a priori* densities corresponding to "complete ignorance," in particular the case of a binomial distribution with the probability p as unknown parameter [9]. The appropriate shape of the density is so far from a rectangular one that this approach is best abandoned.

5. The Present Situation

All the laboratories which had submitted a reply to the BIPM questionnaire mentioned in Section 2 were invited to send a participant to the meeting of a Working Group which had the task of discussing the most important problems related to the assignment of experimental uncertainties. Experts from 11 national laboratories attended the meeting at Sèvres from October 21 to 23, 1980. Dr. P. Giacomo, the Director of BIPM, acted as chairman.

Let us have a quick look at the results of this meeting. For more details the proceedings [10] should be consulted. The case of random uncertainties that can be evaluated by means of a series of repeated measurements raised no serious problem and the Working Group recommended that they be expressed in terms of estimated sample deviations s_i, adding the corresponding degrees of freedom ν_i. This does not imply any assumption on their distribution. For measurements which are expected to be correlated, one should also determine and indicate the respective covariances, provided that the importance of the problem justifies the additional work. Little need was felt for the evaluation of confidence intervals. If the population can be taken as normal, as it is frequently assumed to be, the correspondence is well known and its inclusion does not give any new information. Such an indication may be needed, however, for making a statistical decision or for comparison with legally imposed tolerance limits. More difficult is the situation where uncertainties cannot be obtained by statistical methods; in this case their estimation must rely on "other methods" which inevitably imply some element of subjective appreciation. However, it should be clearly recognized that this fact is neither new nor has it any connection with the previous error classification: an uncertainty obtained by "guesstimation" may affect a final result in the same way as was till now considered to be the effect of a "random" or a "systematic" influence. This depends entirely on the way the quantity under discussion enters our equations describing the physical situation. The resulting uncertainty is in any case of a random nature.

This has led the Working Group to suggest—for purely practical reasons and in a provisional way, especially as regards terminology—two groups or classes of uncertainty components, namely

—group A: those which are evaluated by applying statistical methods, and

—group B: those which are evaluated by other means.

This is neither a very strict nor a very profound classification, but it tries to come close to reality. Since, as we shall see in a moment, all contributions will be used from now onwards on an equal footing, a possible misgrouping is of no consequence. There is therefore little need to worry about further details concerning these groups. In fact, the difference between the groups is not of a "fundamental," but only of a practical nature and should be helpful to the experimenter. Whereas for group A the usual methods of evaluation normally lead in a straightforward way to a sample variance or an estimated standard deviation, this is usually not so with group B. However, this difficulty is not peculiar to the new approach; it has always existed but it used to be masked by concepts such as "maximum limits" or the like which then could not be usefully related to the other components of class A with a better-defined statistical significance.

It would be quite misleading to think that the new classification is, in essence, just the old one presented in another dialect. Surprisingly enough, it was indeed suggested in one of the replies to the questionnaire that replacing the word "error" by "uncertainty" would automatically solve all the major problems. We have somewhat less confidence in the power of words and think that the first concern should be the concepts, and that the words attached to them are of minor importance. The old and the new treatment of uncertainties differ in the basic approach and the transition therefore cannot be performed by means of a dictionary.

Which meaning do we wish to assign to the uncertainty components of group B? The answer comes from a look at the general law of error propagation [Eq. (A3)]. Since this is the only firm basis available, the choice is imposed. If we want to make use of it, we need for all uncertainties quantities which can be considered as valid approximations to the respective standard deviations. To claim that such an estimate is impossible to obtain leads to a deadlock, for this attitude implies that we deny any possibility of arriving at a value for the uncertainty of a final result which includes contributions from both groups A and B. This is obviously a possible point of view, but practical experience shows that it results in situations which are most unsatisfactory, for nobody really knows what to do with more than one uncertainty statement for a final result. It is then left to an evaluator to combine the various (and sometimes ill-defined) uncertainty components in some way, often to the irritation of the experimenter. It is our contention that this important step should be performed by the person who is best acquainted with the details of the measurements, and this is clearly the experimenter himself.

It should be clear, therefore, that if we want to give "useful" measures of uncertainty components, they have to be in such a form that they comply with the usual formalism. Hence we have to choose something which can be considered as the best available approximation to the corresponding standard deviations. This is the essence of the third draft recommendation (cf. Appendix B). We realize that the text restricts itself to indicating the goal to be achieved, without giving further advice on the important question of how this can be done in practical cases. This will be clearly an important topic for a later document which must be more concerned with the practical implementation of the new approach than with its background. As the number of possibilities is quite limited, we can already sketch a few of them. The choice

depends on the personal preference of the experimenter and, in particular, on his previous habits. Thus, some may be courageous enough to guess the required quantity, say u_j, directly; this requires a feeling for what might correspond to something like a 2/3 or 70% confidence interval. Others may be more at ease in trying to estimate a confidence interval of 50% (also known as "probable error"); then u_j may be taken as about 1.5 times this quantity. Finally, for those used to starting with a "safe" confidence interval, the problem is to decide to which multiple of the corresponding "standard deviation" they assume this to correspond, and to divide accordingly.

It may be objected that these correspondences depend on the form of the assumed probability density of the underlying population. In principle this is true, of course, but, for the modest precision aimed at, the assumption is usually not a critical one. In any case, it would clearly make little sense to replace our ignorance by a firm prescription. It seems best to leave this matter to the personal preference of the experimenters, some of whom may choose for themselves, as an intermediate personal guide, a Gaussian or a rectangular probability density, whereas others will feel able to dispense with such a construction.

Once the estimates are obtained for both classes, they can be readily transferred to the forms needed in Eq. (A3) by multiplying them with the corresponding partial derivatives. It is the squares of these new quantities which are, for the sake of simplicity, denoted in the draft recommendations by s_j^2 and u_j^2 for classes A and B, respectively. To this may be added the expressions for the estimated covariances, where required and available. We are now in a position to substitute these quantities into the general formula of error propagation. In the absence of correlations between the various n uncertainties this simply amounts to forming the sums

$$\sum_{j=1}^{n_1} s_j^2 + \sum_{j=1}^{n_2} u_j^2 , \quad \text{with } n_1 + n_2 = n ,$$

for obtaining the approximate "variance" of the required function $y = f(x_1, x_2, \ldots, x_n)$. The square root of this quantity has been called "combined uncertainty." This provisional terminology may provoke opposition, but the possible adoption of a more suitable name is left to later deliberations. However, the Working Group was certainly well advised not to use expressions like "standard deviation" or "variance" for quantities involving contributions from group B in order to avoid justified objections of mathematical statisticians who are used to attributing to these names a well-defined meaning which is no longer pertinent if, as in our case, subjective elements have to be included.

An important point concerns the relation between this combined uncertainty and the notion of a confidence interval. For the contributions from group A the link is usually established by adopting the assumption (which, even here, is often difficult to justify) that all contributions can be taken as a sample belonging to a normal population, permitting us thereby to use the corresponding Student's t factor for passing from a sample standard deviation to a confidence interval of a given probability level. For group B, and therefore for the combined uncertainty which includes such contributions, this is much more difficult to justify. However, it would be quite naive to think that this problem has been created by the new approach sketched above. Rather, the difficulty has always existed, but it used to be concealed by fuzzy concepts, whereas now the problem is clearly brought to light. If we want to be quite honest, we have to admit that a definite link with a confidence level cannot be established. Of course, one may assume (or rather hope) that for the contributions belonging to group B, the relation is approximately the same as the one which is valid for a normal population, but this cannot be assured and is possibly just wishful thinking. Invoking the Central Limit Theorem is normally of little help and seldom justified, at least for situations with few contributions.

Confronted with this difficulty, the Working Group found it necessary to adopt a somewhat different point of view. While an evaluation of confidence limits in the usual sense is not feasible, we are still free to choose an arbitrary multiple of the combined uncertainty for obtaining a wider range and therefore a higher confidence level. The only difference from the usual procedure now is that this multiplication factor, called k, has a conventional value, as for instance k = 2 or 3. There is little doubt that this has in fact often been done before, and not with more justification than now; the real difference therefore only consists in the explicit confession that this procedure is to some extent an arbitrary one. The quantity thus obtained has been termed an overall uncertainty, which is again to be taken as a provisional denomination, and recommendation 5 says that whenever its application is felt necessary, the multiplying factor should be clearly stated. The need to use such multiples (which obviously carry no more information than the combined uncertainty) hardly arises in scientific measurements, but is often imposed in industrial or legal metrology, where such requirements are demanded by people who believe that all random deviations necessarily follow the Gaussian law.

6. Final Remarks

The problems involved in the assignment of uncertainties to experimental results, as we have seen in the above analysis, are of a type which makes their treatment at the same time simple and difficult. Their ambiguous nature, it seems to me, is due to the fact that on the one hand they have an obvious relationship with such fields as mathematics and statistics where one is used to dealing with theoretically well-defined concepts, but on the other hand they are also intimately linked to the practical world of laboratories and workshops with all their hazards and compromises which are typical of the daily life of a practitioner. As uncertainties are indicated or used by virtually everybody and in very different fields, their statement has to be acceptable in a far larger domain than is usually required for more specialized quantities. In addition, the various users impose a number of constraints on format and content which have, as far as possible, to be taken into account. It is not surprising, therefore, that there exists a large variety of proposals for achieving this goal.

Instead of taking the various and often contradictory requirements as a mere nuisance, one can also try to consider them as guides permitting us to narrow down the number of useful proposals. Bearing in mind that we may choose only those quantities that are compatible with the conditions imposed by the general propagation law for uncertainties, the possibilities are so strongly restricted that there remains practically no place for arbitrariness. The result—and I hope the majority of you can agree with it—is the proposal outlined above and described in a more comprehensive form by the five recommendations set up by the Working Group. There is no doubt that various details are capable of improvement and may

eventually have to be changed (for instance notation), but taken as a whole the present proposal, which is believed to form a reasonably logical and coherent unity, offers little or no room for fundamental changes: if it is considered unacceptable, it will have to be dropped completely and the problem must be tackled again from scratch. I cannot but hope that this will not happen, but the decision is now with those who will try to apply this simple scheme for their own measurements. To my mind, the questions related to the assignment of uncertainties do not provide a useful playground for purists or fundamentalists. This is not the sort of problem which can always be solved by a yes or no answer. Some compromises and arbitrary conventions are difficult to avoid and should be accepted in the interest of a wider applicability. The new scheme described in the draft recommendations is already applied in several national laboratories and its practical usefulness is going to be tested during a trial period. If the results are considered satisfactory, these procedures may then be recommended for general use. In this way it is hoped to arrive, in due course, at an acceptable uniform way of assigning uncertainties to experimental measurements.

In the realm of fundamental constants, the main subject of this conference, uncertainties occupy just a very small corner. The more topical and interesting problems are on another level and concern the most appropriate way to actually perform the necessary least-squares adjustment when it is realized that both the mean value and the associated uncertainty, as indicated by the experimenter, may be in error. This important subject will not be discussed here at all (for recent reviews, see Refs. [7] and [11]), but it may be useful to bear in mind that it is essential in such an enterprise to dispose of reliable and realistic estimates for the uncertainties.

In the form outlined above, their treatment merely requires the use of some elementary and old-fashioned mathematics, and there is in all probability no need at all for anything much more sophisticated (although proposals to the contrary are not lacking). The rest is essentially common sense. It will be noted that special care has been taken to refrain from adopting any hypothesis which is not really needed, as this would have rendered the result less general and therefore narrowed its range of useful application.

Even if we adopt the optimistic point of view that in the future the reporting of uncertainties will follow essentially the proposals outlined above, it would no doubt be unrealistic to expect a rapid convergence towards uniformity. Habits are persistent and usually lifelong. This, therefore, reminds me of the story about the philosopher who was celebrating his 90th birthday. One of his young adherents expressed admiration for the fact that his opinions, once bitterly attacked, are now quite generally accepted and he asked the scholar how he had succeeded in convincing all his enemies. The old man gently smiled and then replied: "The truth is that nobody has ever changed his opinion, but it just happens that I have survived them all."

There are moments when I wonder if we are not in a somewhat similar situation.

Note added in proof: The final French version of the Recommendation INC-1 (1980), which corresponds to the draft version given here as Appendix B, has now been published in the document mentioned in Ref. [10]. In addition, the Recommendation 1 (CI-1981) of the CIPM which concerns the practical implementation of INC-1 (1980) is reproduced on page 26 of the same publication.

The author is very grateful to several staff members of the BIPM for helpful remarks made on a draft version. He alone, however, is responsible for possible controversial statements expressed in this review.

7. Appendices

A. *A Reminder on Error Propagation*

As it still forms the undisputed basis for any serious discussion on uncertainties, it may be useful to recall briefly some simple facts concerning the well-known "general law of error propagation," as it is called in the traditional terminology. This seems also justified by the conclusion, for which the underlying reasoning is sketched in Section 3, that its careful application essentially solves all of the much-discussed problems in this field which, to a large extent, have been artificially created by a simple neglect of some basic elements.

If y is a known function of n variables x_j, thus $y = f(x_1, x_2, \ldots, x_n)$, then small changes Δx_j affect the "result" y approximately by an amount

$$\Delta y \simeq \sum_{j=1}^{n} \frac{\partial f}{\partial x_j} \Delta x_j \ ,$$

if the terms of higher order in the Taylor expansion of the function can be neglected. The square of this quantity is given (again in first approximation) by

$$(\Delta y)^2 \simeq \sum_{j=1}^{n} \left[\frac{\partial f}{\partial x_j} \Delta x_j \right]^2 + \sum_{j=1}^{n} \left[\frac{\partial f}{\partial x_j} \Delta x_j \right] \sum_{\substack{k=1 \\ (k \neq j)}}^{n} \left[\frac{\partial f}{\partial x_k} \Delta x_k \right] . \quad \text{(A1)}$$

It is practical to characterize the "random displacements," or "errors," by their variances and covariances, which can be achieved by putting (E stands for mathematical expectation)

$$\mathrm{E}\{(\Delta x_j)^2\} = \sigma_j^2 \text{ and } \mathrm{E}\{\Delta x_j \, \Delta x_k\} = \sigma_{jk} \ . \quad \text{(A2)}$$

Then Eq. (A1) can be written in the more compact form

$$\sigma_y^2 \simeq \sum_j (f'_j \, \sigma_j)^2 + \sum_{j \neq k} f'_j \, f'_k \, \sigma_{jk} \ , \quad \text{(A3)}$$

with $f'_j = \partial f / \partial x_j$ and summations from 1 to n.

Equation (A3) is the well-known general formula for "error propagation." If one agrees to identify σ_{jj} with σ_j^2, then Eq. (A3) can be further condensed into

$$\sigma_y^2 \simeq \sum_{j,k} f'_j \, f'_k \, \sigma_{jk} \ . \quad \text{(A4)}$$

For most practical applications one has to replace the (unknown) population parameters σ_j^2 and σ_{jk} appearing in Eq. (A3) by their corresponding (unbiased) sample estimators. Hence, if we assume that for a quantity x_j we have available the m results of measurements x_{j1}, $x_{j2}, \ldots, x_{jm}$, and likewise for x_k, then instead of σ_j^2 we may use

$$s_j^2 = \frac{1}{m-1} \left[\sum_i x_{ji}^2 - \frac{1}{m} \left[\sum_i x_{ji} \right]^2 \right] ,$$

and instead of σ_{jk} we may use

$$s_{jk} = \frac{1}{m-1} \left[\sum_i x_{ji} \, x_{ki} - \frac{1}{m} \left[\sum_i x_{ji} \right] \left[\sum_i x_{ki} \right] \right] ,$$

where all summations over i go from 1 to m. There exist

equivalent forms which at times may be more practical; they can be found in any textbook on elementary statistics.

It is a fairly popular, but clearly unjustified belief that in most "practical cases" the various contributions to the uncertainty can be taken as independent and that the terms involving the covariances may therefore be dropped. This unfortunate prejudice has led to the rather confusing error classification which hides the fact that, for example, the so-called "systematic" uncertainties are in no way different from the "random" ones, except that they are correlated to the measurements in another way. This point is also discussed in Section 3.

B. *Draft Recommendations on the Statement of Uncertainties*

These recommendations are presented here in the form in which they were elaborated by the Working Group at the October meeting. They will be presented, possibly with minor changes, to the Comité International des Poids et Mesures. They have also been published in Ref. [12].

1. The uncertainty in the result of a measurement generally consists of several components which may be grouped into two categories according to the way in which their numerical value is estimated:

A—those which are evaluated by statistical methods,
B—those which are evaluated by other means.

There is not always a simple correspondence between the classification into categories A or B and the previously used classification into "random" and "systematic" uncertainties. The term "systematic uncertainty" can be misleading and should be avoided.

Any detailed report of the uncertainty should consist of a complete list of the components, specifying for each the method used to obtain its numerical value.

2. The components in category A are characterized by the estimated variances, s_i^2 (or the estimated "standard deviations" s_i), and the number of degrees of freedom, ν_i. Where appropriate, the estimated covariances should be given.

3. The components in category B should be characterized by quantities u_j^2, which may be considered as approximations to the corresponding variances, the existence of which is assumed. The quantities u_j^2 may be treated like variances and the quantities u_j like standard deviations. Where appropriate, the covariances should be treated in a similar way.

4. The combined uncertainty should be characterized by the numerical value obtained by applying the usual method for the combination of variances. The combined uncertainty and its components should be expressed in the form of "standard deviations."

5. If, for particular applications, it is necessary to multiply the combined uncertainty by a factor to obtain an overall uncertainty, the multiplying factor used must always be stated.

References

[1] B. L. van der Waerden, *Science Awakening II, The Birth of Astronomy* (Noordhoff, Leyden, 1974).
[2] *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.) Spec. Publ. 343 (Aug., 1971), p. 493.
[3] The BIPM enquiry on error statements, Rapport BIPM-80/3 (1980), 50 p. A French translation is being published in Procès-Verbaux Séances Com. Int. Poids Mes. 48 (69th session) (1980).
[4] J. W. Müller, Nucl. Instrum. Methods 163, 241 (1979).
[5] Les incertitudes de mesures, in *La Physique*, vol. 4 of the *Encyclopédie Scientifique de l'Univers* (Gauthier-Villars, Paris, 1981), p. 11.
[6] B. N. Taylor, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.) Spec. Publ. 343 (Aug. 1971), p. 495.
[7] E. R. Cohen, in *Metrology and Fundamental Constants*, Proceedings of the International School of Physics "Enrico Fermi," Course LXVIII, Ed. by A. Ferro Milone, P. Giacomo, and S. Leschiutta (North-Holland, Amsterdam, 1980), p. 581.
[8] E. T. Jaynes, IEEE Trans. Syst. Sci. Cyb. SSC-4, 227 (1968).
[9] J. W. Müller, Un nouveau regard sur les probabilités a priori, Rapport BIPM-80/6 (1980), 16 p.
[10] Report on the Meeting of the BIPM Working Group on the Statement of Uncertainties. The final French version will be published in Procès-Verbaux Séances Com. Int. Poids Mes. 49 (70th session) (1981).
[11] W. Wöger, PTB-Mitteilungen 89, 401 (1979).
[12] P. Giacomo, Metrologia 17, 69 (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

On the Statement of a Total Confidence Interval Based on the Concept of Randomization of Systematic Errors: Large and Small Sample Sizes

W. Wöger

Physikalisch-Technische Bundesanstalt, Bundesallee 100, D-3300 Braunschweig, F.R.G.

Interpreting the systematic error in a measurement series as the realization of a random variable, confidence intervals of the mean for the special cases of large and small sample sizes are constructed by statistical methods, assuming normal distributions. These confidence intervals combine the effects of "random" and "systematic" errors and are extensions of the commonly used formulae.

Key words: combination of random and systematic uncertainties; measurement uncertainty; randomization of systematic error; total confidence interval; uncertainty assignment.

1. Introduction

If a physical quantity Y is to be measured, then in the statement about the true value of this quantity the effects of systematic errors introduced by the measuring process has to be included. Apart from this problem still under discussion, the situation could get even more complex, when a conclusion is to be drawn on the true value of a quantity Z, which depends not only on Y, but also on some additional quantities, the determination of which are accomplished in separate experiments.

In the following it is assumed that a general relationship between Z, Y and the other quantities exists and can be linearized by an appropriate expansion, so that we start from

$$z = y + G(A_1, \ldots, A_N) . \tag{1}$$

For y we assume a series of n measurements, the values drawn from a normal distribution with variance σ_I^2. The quantities $A_1, \ldots, A_N$ may consist of influence quantities on the measurement of y, fundamental constants and auxiliary quantities. We assume further that we know, from estimation or other considerations, intervals within which the possible values of these quantities are situated. The values are constant, but unknown, for the above mentioned series of measurements with respect to y. Repeating the measurement series under the same nominal conditions, the function $G(A_1, \ldots, A_N)$ may very well take a different constant value, which again is unknown and unpredictable. This is especially clear for the influence quantities.

Therefore one is led to the idea to interpret the constant value of $A_j (j = 1, \ldots, N)$ in one special measurement series as the outcome of a random variable. Obviously the sample space for A_j is different from that for y. One can visualize an experiment to determine the true value of z in the following way: There are $N + 1$ urns, one urn for the y-value, one urn for the A_1 values and so on. The experiment then consists of drawing n values at random out of the urn for the y-values and simultaneously one value out of the A_1-urn, one value out of the A_2-urn and so on.

With this information at hand, mathematical statistics give the methods to handle the problem of construction of a confidence interval for the true value of z. This is shown in the following parts of the article.

2. Simplification of the Problem

To each component $A_j (j = 1, \ldots, N)$ corresponds an (estimated) interval

$$(A_l)_j \leqslant A_j \leqslant (A_u)_j .$$

Linearizing the function $G(A_1, \ldots, N)$ around $\overline{A}_j = \frac{1}{2}((A_l)_j + (A_u)_j)$ $(j = 1, \ldots, N)$ and calling:

$$y_c = y + G(\overline{A}_1, \ldots, \overline{A}_N)$$

one gets from Eq. (1):

$$y_c = z + \sum_{j=1}^{N} f_j = z + f ; \tag{2}$$

f_j takes values in the interval

$$-\Delta x_j \leqslant f_j \leqslant + \Delta x_j$$
$$\Delta x_j = |G'(\overline{A})|((A_u)_j - \overline{A}_j) .$$

We now assume that not only y_c, but also f is normally distributed, the expectation value being zero and the variance σ_s^2. This variance is assumed to be known from the width of the intervals for the A_j's, This means that σ_s^2 is some function of the Δx_j's. In this article we shall not discuss which function should be chosen but only quote the proposal of S. Wagner [1]:

$$\sigma_s^2 = \frac{1}{3} \sum_{j=1}^{N} (\Delta x_j)^2 . \tag{3}$$

The proposal is based on the assumption that all random variables f_j are stochastically independent. (This assumption may be weakened.)

If we measure several times y_c in an experiment with fixed but unknown f, then the scatter of these values is described by σ_I^2, the variance of the normal distribution.

If we build $\bar{y}_c$, the arithmetic mean of n measurements, and repeat the experiment again and again, then the scatter of the $\bar{y}_c$ values is given by σ_I^2/n, if for each experiment f takes the same fixed value. But this is not true: f also varies from experiment to experiment and its scatter is described by the variance σ_s^2. So the total variance of $\bar{y}_c$ is given by:

$$\sigma^2 = \sigma_I^2/n + \sigma_s^2 . \tag{4}$$

Note that σ_I^2/n usually is not known, but only the estimate

$$s_{\bar{y}}^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} (y_i - \bar{y})^2 , \tag{5}$$

which of course is independent of f.

From Eq. (2) we know that for each experiment, $\bar{y}_c$ is an estimate for $\mu + f$, where μ is the true value of z and f is fixed but unknown. Repeating the experiment again and again and averaging the resulting $\bar{y}_c$'s we will get the average μ, because the average of all the f's in the long run in zero. From this we conclude that $\bar{y}_c$ is normally distributed with the parameters μ and $\sigma^2 = \sigma_I^2/n + \sigma_s^2$. (The rigorous proof will be given elsewhere.) Therefore $(\bar{y}_c - \mu)/\sigma$ is normally distributed with mean zero and variance 1.

3. The Confidence Interval for the True Value μ of z

To draw a conclusion on the true value μ of z from the observed values $\bar{y}_c$ and $s_{\bar{y}_c}^2$, consider the random variable

$$T = \frac{\sqrt{n-1}\,(\bar{y}_c - \mu)/\sigma}{\sqrt{\frac{n(n-1)}{\sigma_I^2} s_{\bar{y}}^2}} = \frac{1}{\sqrt{1 + n\,\frac{\sigma_s^2}{\sigma_I^2}}} \cdot \frac{\bar{y}_c - \mu}{s_{\bar{y}}} \tag{6}$$

The variable T is Student-distributed with $(n-1)$ degrees of freedom. But this knowledge is of little help as long as σ_I^2 or the ratio σ_s^2/σ_I^2 is not known.

(a) If the ratio σ_s^2/σ_I^2 is known, then from Eq. (6) for a given confidence level P the confidence limits for μ are given by:

$$\bar{y}_c \pm t_p(n-1) \left[1 + n\,\frac{\sigma_s^2}{\sigma_I^2}\right]^{1/2} s_{\bar{y}} . \tag{7}$$

(b) If n is large, then σ_I^2/n may very well be approximated by $s_{\bar{y}}^2$. Therefore σ^2 can be regarded as known. Thus

$$T \to T_1 = \frac{\bar{y}_c - \mu}{\sigma} ,$$

with the resulting confidence interval:

$$\bar{y}_c \pm k_p \sqrt{\frac{\sigma_I^2}{n} + \sigma_s^2} \approx \bar{y}_c \pm k_p \sqrt{s_{\bar{y}}^2 + \sigma_s^2} .$$

Remark: k_p is defined by

$$P \overset{!}{=} \frac{1}{\sqrt{2\pi}} \int_{-k_p}^{+k_p} e^{-\frac{x^2}{2}}\, dx .$$

Inserting Eq. (2) for σ_s^2 yields Wagners' proposal [1].

(c) Small sample size n.

In this case $s_{\bar{y}}^2$ cannot be assumed to be an acceptable approximation for σ_I^2/n. If we formally let σ_I^2/n to be $s_{\bar{y}}^2$ in Eq. (6), then we have the variable

$$T_2 = \frac{\bar{y}_c - \mu}{s} \quad ; \quad s^2 = s_{\bar{y}}^2 + \sigma_s^2 ,$$

which is *not* Student-distributed.

However, using a procedure proposed by Welch [2], it can be shown that the distribution of T_2 may be approximated by a Student-distribution with an effective number of degrees of freedom:

$$\nu_{\text{eff}} = (n+1) \left[1 + \frac{\sigma_s^2}{s_{\bar{y}}^2}\right]^2 - 2 .$$

From this we get the confidence interval for a given confidence level P:

$$\bar{y}_c \pm t_p(\nu_{\text{eff}}) \sqrt{s_{\bar{y}}^2 + \sigma_s^2} .$$

4. Concluding Remarks

The confidence limits given above depend on the assumption of normal distribution of the random variables under consideration. This assumption may not always be true. But then the construction of confidence intervals is rather tedious, if possible at all. Nevertheless, the randomization of the "systematic" error, represented here by the function $G(A_1, \ldots, A_N)$, still is a useful concept, showing how the effect of several "systematic" error sources has to be handled, namely by summation of variances, and this is independent of the form of the distribution functions as long as there exist the second moments of the distributions.

It should be emphasized that the randomization of "systematic" errors is not just a trick, to solve a well known problem, but has a realistic background. This may be documented, for example, by the use of this concept for years in statistical quality control of measuring instruments. Physicists seem to have overlooked a part of the available information about their so called "systematic" errors. Of course, they are (at least nearly) systematic for one single experiment conducted under repeatability conditions, but their statistical nature immediately appears if one takes into account the impossibility to reproduce for example exactly the same conditions imposed on the experiment, when repeating it some time later.

Accepting this statistical behavior of the "systematic" error, one is left with the problem how to describe it. At the moment, for this one has to refer to plausible models like the assumption of a normal distribution with a variance given, for example, by Eq. (2). But this is by no means obligatory and is still under discussion.

References

[1] S. Wagner, PTB-Mitteilungen 89, 83 (1979).
[2] B. L. Welch, Biometrika 34, 28 (1947).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Measurement Assurance

A. F. Dunn

National Research Council of Canada, Ottawa, Canada K1A OR6

Measurement assurance involves combining intrinsic uncertainties of instruments and standards utilized in a measurement procedure with uncertainties associated with actual use of the instruments, in order to provide a meaningful statement of a total effective uncertainty of the measurement procedure.

Key words: accuracy; (equivalent) standard deviation; measurement assurance; uncertainty.

1. Introduction

The effective accuracy of an experimental determination is made up of contributions from a number of sources, such as:

Random uncertainties in the primary experiment and subsidiary experiments;

Systematic uncertainties from instruments used (calibrated and uncalibrated), reference standards used, and linearity and sensitivity of detectors; and

Influence quantity effects and measurements made to account for, or eliminate, the influencing effects.

Systematic errors due to design of measurements.

It is conventional to assess a random uncertainty of repeated measurements as the standard deviation of observed data and attribute to it a connotation of probability of occurrence. Because of the inherent indeterminacy of the calculation of a standard deviation from a relatively few (<50) occurrences, there is little to choose between an implication of 99.99% certainty ($\pm 3.9s$), 99.73% certainty ($\pm 3.0s$) or 99% certainty ($\pm 2.6s$).

As a working basis, a band of $\pm 2s$ (95% certainty) has been adopted as a statement of the "accuracy" of a repeated measurement.

In those cases where only random uncertainties of repeated measurements are to be combined, it has also been conventional to combine them in a root-sum-square (RSS) combination of standard deviation multiples which have similar meaning.

From statistical theory of random variations in a process which is a function of several variables

$$z = F(x_1, x_2, x_3, \ldots\ldots),$$

the combined variance (or standard deviation) is

$$\text{variance } (z) = \sum \left(\frac{\partial F}{\partial x_i}\right)^2 (s_{x_i})^2 + \sum\sum \frac{\partial F}{\partial x_p} \times \frac{\partial F}{\partial x_q} \text{ covariance } (x_p, x_q). \quad (1)$$

For independent variables, the covariances (x_p, x_q) are zero and the conventional root-sum-square combination remains, as

$$s_z^2 = \left(\frac{\partial F}{\partial x_1}\right)^2 (s_{x_1})^2 + \left(\frac{\partial F}{\partial x_2}\right)^2 (s_{x_2})^2 + \left(\frac{\partial F}{\partial x_3}\right)^2 (s_{x_3})^2 + \ldots\ldots\ldots \quad (2)$$

Thus to combine uncertainties it is necessary to find a system of assessing equivalent standard deviations to represent the contributing sources of uncertainty in a measurement.

In the absence of any other specific information one can usually assess an upper limit to the magnitude of an estimated uncertainty (from manufacturer's literature, calibration reports, experience, etc.) but generally nothing can be said about the frequency distribution within the band of possible values. It seems reasonable to assume that any value between the estimated limits is equally probable, leading to a rectangular frequency distribution of width $2a$ and height $1/(2a)$, from which the variance is $v_{eq}^2 = a^2/3$.

The experimental process of estimating the limits ($\pm a$) of the observed deviations is found to be easier than other techniques tried and, when combined with an assumption of equi-probability of occurrence, leads directly to an estimated equivalent standard deviation, $v_{eq} = 0.58\ a$. (v is the symbol adopted to indicate a derived equivalent standard deviation rather than a standard deviation calculated from repeated measurements.)

It is comforting for the experimentalist to realize that any departure from a rectangular distribution toward a Gaussian distribution means that the derived equivalent standard deviation is conservative.

An analytical approach to the problem of the combination or convolution of various types of distributions with each other and with more typical Gaussian distributions has been given by C. F. Dietrich [1]. In summary, he finds that for a combination of different types of uncertainties, using the equivalent standard deviations described, one can be satisfied that the probability of an uncertainty greater than $\pm 2s$ is always less than in the corresponding Gaussian case, although not much less.

The work of Dietrich has been accepted, and the combination of uncertainties in experiments has been based on this sort of interpretation of uncertainty bands converted to equivalent standard deviations and combined to give a total uncertainty of the experimental determination which is further assumed to have approximately the same probability connotations. It is also assumed that the ef-

fective accuracy of a measurement is represented by $\pm 2s_{\text{total}}$ with the anticipation that no more than 5% of the individual measurements will exceed this band.

2. Application

As an example, an assessment of the uncertainties in the Josephson-effect determination of voltage in the National Research Council of Canada is shown in Table 1.

In summary, the magnitudes of the estimated uncertainties for each of the five sub-divisions of the experimental procedure are:

$v_1 = 41 \times 10^{-9}$ [random observed s_m = 20 to 80]
$v_2 = 50 \times 10^{-9}$ [cell comparisons (TG-NRG)]
$v_3 = 26 \times 10^{-9}$ [cell comparisons (NRG)]
$v_4 = 29 \times 10^{-9}$ [cell value maintenance]
$v_5 = 96 \times 10^{-9}$ [voltage divider calibration]

The accuracy of a typical Josephson-effect determination of the emf of the cells in the local transfer group (TG) is

$$2s_{\text{TG}} = 2(v_1^2 + v_5^2)^{1/2} = \pm 0.20 \times 10^{-6} = \pm 0.20 \text{ ppm}$$

The accuracy of the maintenance of a cell in the national reference group (NRG) at its observed working temperature is

$$2s_{\text{NRG}} = 2(v_1^2 + v_2^2 + v_3^2 + v_4^2 + v_5^2)^{1/2} = \pm 0.24 \text{ ppm}$$

Figure 1 is a plot of the mean value of the national reference group obtained from Josephson determinations for a 3-year period. It is clear the group mean is decreasing with time, at the rate of 0.374 ppm/a from the data presented. The dashed lines indicate the boundaries of the accuracy of determination of the mean value, i.e. displaced ±0.24 ppm from the solid curve representing the best fit to the Josephson determinations. Of the 20 Josephson determinations plotted 15 lie within ±0.12 ppm, or 75% are within the band represented by $\pm s_{\text{NRG}}$, in good agreement with the intended interpretation of the equivalent standard deviations used in assessment of the total uncertainty.

References

[1] C. F. Dietrich, *Uncertainty, Calibration and Probability* (John Wiley & Sons, New York, 1973), Chapters 3 and 4.

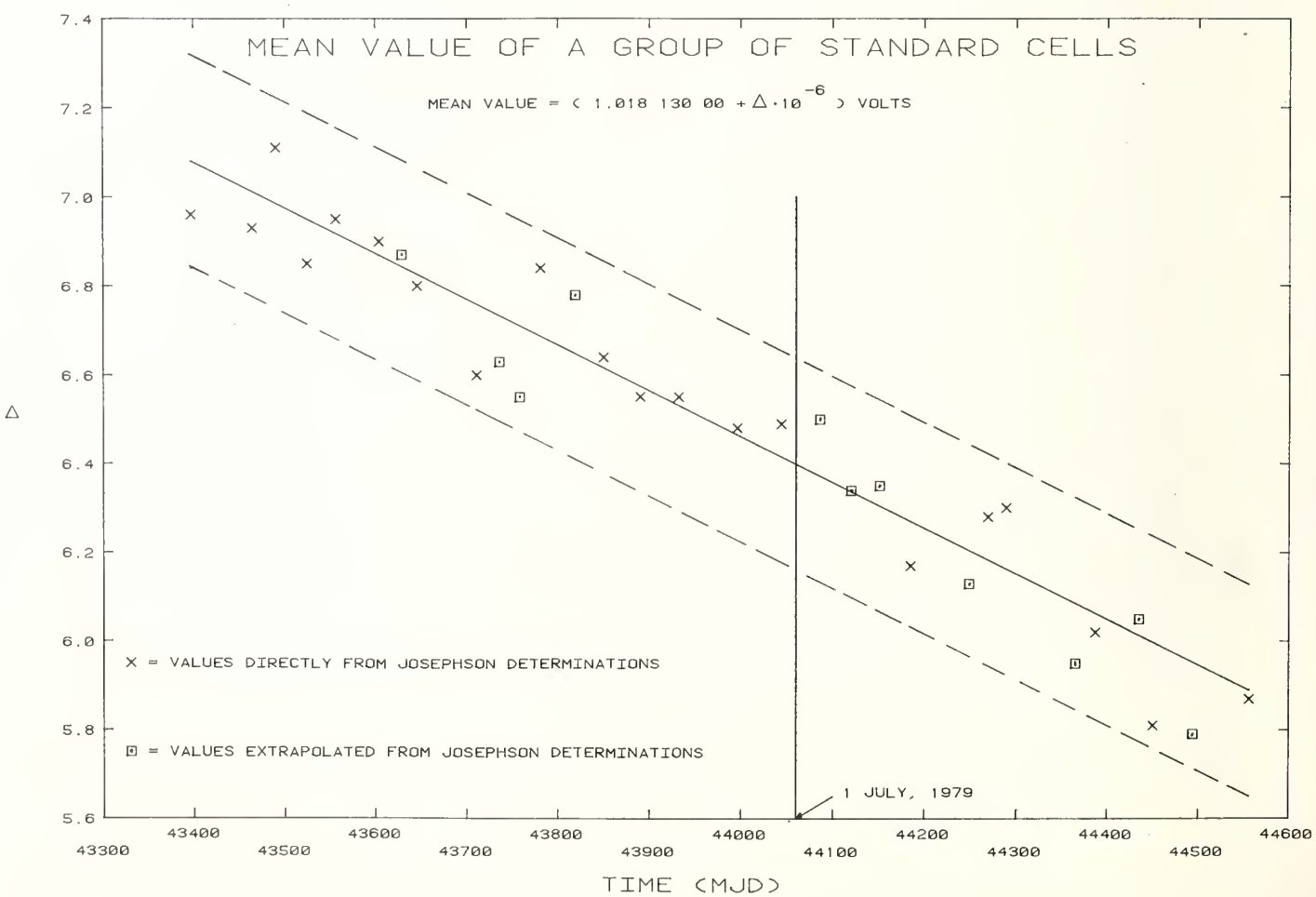


FIGURE 1. *Plot of the mean emf of the Canadian national reference group of standard cells as obtained from Josephson effect determinations over a three year period.*

TABLE 1. *Uncertainties in Josephson-volt determination*

Source of Uncertainty	Spread of Deviation ($\times 10^{-9}$)	Divisor (1 or $\sqrt{3}$)	Equivalent Standard Deviation
(1) Junction balanced against cells T1 or T2			
(a) Junction balance against 0.25-Ω resistor of voltage divider			
: stability of 10 GHz frequency observed on counter	±20	$\sqrt{3}$	12
: reference frequency for counter	±0.001	$\sqrt{3}$	—
: estimation of v_o from recorder trace with reversals of galvanometer and calibration of recorder deflection sensitivity ($\Delta D_1/\Delta v$)	<±40	$\sqrt{3}$	23
: estimated systematic error due to assumed linearity of galvanometer, pre-amplifier, and recorder deflections, equivalent to	± 5	1	5
: elimination of thermal emf in dc leads from junction at 4 K to resistor at 293 K by reversal of junction bias current, with N-R-N treatment of successive balances	<±30	$\sqrt{3}$	17
: voltage indeterminacy due to instability of junction bias current and non-vertical slope of current steps, equivalent to	± 3	$\sqrt{3}$	2
(b) Standard cell balance against 100-Ω resistor of voltage divider			
: estimation of Δe from recorder trace with reversals of galvanometer and calibration of recorder deflection sensitivity ($\Delta D_2/\Delta e$)	<±20	$\sqrt{3}$	12
: estimated systematic error due to assumed linearity of galvanometer, pre-amplifier, and recorder deflections, equivalent to	± 5	1	5
: temperature of TG cell enclosure (±0.2 mK) and calibration of recorder sensitivity, equivalent (-56×10^{-6}/K) to	±15	$\sqrt{3}$	9
: estimated systematic error due to assumed linearity of pre-amplifier and recorder deflections, equivalent to	± 5	1	5
: estimated systematic error due to assumed temperature coefficient of standard cells ($\pm 1 \times 10^{-6}$/K), equivalent to	<± 1	1	1
(c) Estimated systematic error due to assumed equivalence of voltage ratio and resistance ratio, i.e., the assumption that $I_{100} = I_{0.25}$	±10	1	10
(d) Voltage ratio divider			
: temperature change during measurement (<±10 mK), equivalent (1×10^{-6}) to	±10	$\sqrt{3}$	6
: calibration of temperature-measuring thermistor (±2 mK), equivalent (1×10^{-6}) to	± 2	$\sqrt{3}$	1
: measurement of temperature coefficient of voltage ratio ($\pm 1 \times 10^{-6}$/K) equivalent (<±10 mK) to	±10	$\sqrt{3}$	6
: estimated systematic error due to effect of barometric pressure (±0.7 kPa ≡ ±5 mm Hg) on resistance, equivalent ($\pm 20 \times 10^{-9}$/kPa) to	±15	1	15
v_1 = equivalent standard deviation (square root of sum of squares of individual components)			41
(2) Intercomparison of cells T1, T2, T3 with N1, N2			
: sensitivity of each cell-pair comparison (±0.02 μV), contribution to mean of sub-group of five cells	± 6 (1s)	1	6
: standardization of comparison potentiometer (±0.01%) equivalent (for emf differences of <±10 μV) to	<± 1	$\sqrt{3}$	1
: estimated systematic error due to assumed linearity of potentiometer (±0.01%), equivalent to	± 1	1	1
: thermal emf due to selector switching of cells	± 2	$\sqrt{3}$	1
: temperature of TG cell enclosure (±0.2 mK) and calibration of recorder sensitivity, equivalent (-56×10^{-6}/K) to	±15	$\sqrt{3}$	9
: estimated systematic error due to assumed linearity of pre-amplifier and recorder deflections, equivalent to	± 5	1	5

NOTE: In practice, the series of 6–8 measurements made on each standard cell gives a mean value with a standard deviation of the mean $s_m = s_1 = \pm(20 - 80) \times 10^{-9}$ which is in good agreement with the estimate obtained above. For each day's evaluation, the standard deviation of the mean value (s_1) is taken as the estimate of the random uncertainty which takes account of the actual combination of the various component uncertainties applicable to the junction-cell balance applicable to that set of measurements. For general summary reporting, the estimated value of v_1 is used.

TABLE 1. *Uncertainties in Josephson-volt determination (cont.)*

Source of Uncertainty	Spread of Deviation ($\times 10^{-9}$)	Divisor (1 or $\sqrt{3}$)	Equivalent Standard Deviation
: estimated systematic error due to location of temperature sensor and controlling sensor in standard cell enclosure—under stable ambient conditions estimated at ±0.2 mK but under slowly varying conditions estimated at ±0.5 mK, equivalent (-56×10^{-6}/K) to	±30	1	30
: estimated systematic error due to assumed temperature coefficient of standard cells ($\pm 1 \times 10^{-6}$/K), equivalent (1 mK) to	± 1	1	1
: temperature of NRG cell enclosure (±0.3 mK), equivalent (-56×10^{-6}/K) to	±17	$\sqrt{3}$	10
: estimated systematic errors for NRG, as for TG above			
—location of sensors	±30	1	30
—temperature coefficient of cells	± 1	1	1
: determination of values of N1, N2 from mean value of sub-group of five cells	± 6 (1*s*)	1	6
: stability of cells during 5-hour period from before to after junction-cell balance	±30	$\sqrt{3}$	17
: estimated systematic error due to 50 m of cable between two laboratory rooms, decreased by switching techniques used to eliminate thermal effects	±10	1	10
v_2 = equivalent standard deviation =			50
(3) Intercomparison of cells N1, N2, N3, N4			
: sensitivity of each comparison (±0.02 μV), contribution to mean of sub-group of four cells	± 8 (1*s*)	1	8
: standardization of comparison potentiometer (±0.01%) equivalent (for emf differences of <±10 μV) to	<± 1	$\sqrt{3}$	1
: estimated systematic error due to assumed linearity of potentiometer (±0.01%), equivalent to	± 1	1	1
: thermal emf due to selector switching of cells	± 2	$\sqrt{3}$	1
: temperature of NRG cell enclosure (±0.3 mK), equivalent (-56×10^{-6}/K) to	±17	$\sqrt{3}$	10
: estimated systematic error due to scaling of temperature sensor dial for reducing cell values to a fixed reference temperature of 30.0000 °C	±20	1	20
: estimated systematic error due to temperature coefficient of cells used for reduction to 30.0000 °C ($\pm 1 \times 10^{-6}$/K) for reductions of as much as 10 mK	±10	1	10
v_3 = equivalent standard deviation =			26
(4) Maintenance of NRG (8 cells)			
: sensitivity of each comparison (±0.02 μV), contribution to mean of group of eight cells	± 4 (1*s*)	1	4
: standardization of comparison potentiometer (±0.01%) equivalent (for emf differences of <10 μV) to	± 2	$\sqrt{3}$	1
: estimated systematic error due to assumed linearity of potentiometer (±0.01%) equivalent to	± 1	1	1
: thermal emf due to selector switching of cells	± 2	$\sqrt{3}$	1
: temperature of NRG cell enclosure (±0.3 mK), equivalent (-56×10^{-6}/K) to	±17	$\sqrt{3}$	10
: extrapolation of mean value by drift rate determined from previous Josephson determination ($(-0.374 \pm 0.04) \times 10^{-6}$/year), for periods as long as 3 months	<±10 (1*s*)	1	10
: conversions of cell values at reference temperature to working values at observed enclosure temperature (±0.2 mK) equivalent (-56×10^{-6}/K) to	±17	$\sqrt{3}$	10
: estimated systematic error due to scaling of temperature sensor dial for converting cell values from reference temperature to observed enclosure temperature	±20	1	20
: estimated systematic error due to temperature coefficient of cells ($\pm 1 \times 10^{-6}$/K) for reductions of as much as 10 mK	±10	1	10
v_4 = equivalent standard deviation =			29

TABLE 1. *Uncertainties in Josephson-volt determination (cont.)*

	Source of Uncertainty	Spread of Deviation ($\times 10^{-9}$)	Divisor (1 or $\sqrt{3}$)	Equivalent Standard Deviation
(5)	Calibration of 400:1 voltage divider [the divider is composed of two resistors (100 Ω and 0.25 Ω); resistance ratio is measured by a direct current comparator bridge in terms of a Hamon step-up transfer standard (100 Ω in series, 1 Ω in parallel) by a reversed 1:1 ratio measurement at 100 Ω and a 4:1 or 1:4 ratio measurement at 0.25 Ω]			
	(a) 100:100 ratio measurement			
	: mean value of R_{100}/R_s and R_s/R_{100} measurements	± 50	$\sqrt{3}$	29
	: estimated elimination of offset of comparison bridge	± 20	1	20
	(b) estimated systematic error in series/parallel ratio of step-up standard	± 20	1	20
	(c) 1:0.25 and 0.25:1 ratio measurement			
	: mean value of $R_p/0.25$ and $0.25/R_p$ measurements	±100	$\sqrt{3}$	58
	: estimated elimination of offset of comparison bridge	± 50	1	50
	(d) temperature measurement			
	: temperature measuring sensitivity (±1 mK), equivalent (1×10^{-6}/K) to	± 1	$\sqrt{3}$	1
	: calibration of temperature-measuring thermistor (±2 mK), equivalent (1×10^{-6}/K) to	± 2	$\sqrt{3}$	1
	: measurement of temperature coefficient of resistance ratio ($\pm 1 \times 10^{-6}$/K), equivalent (<±20 mK) to	<± 20	$\sqrt{3}$	12
	(e) estimated systematic error due to effect of barometric pressure on resistance (±2 kPa ≡ ±15 mm Hg), equivalent ($\pm 20 \times 10^{-9}$/kPa) to	± 40	1	40
	v_5 =	equivalent standard deviation =		96

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

An Extended-Least-Squares Treatment of Discrepant Data*

E. Richard Cohen

Rockwell International Science Center, Thousand Oaks, CA 91360

In the usual least squares analysis, the weighting of the data is proportional to the inverse of the assigned variance. Since this variance is itself the result of measurement, its value is uncertain. The observed residuals in the least squares adjustment provide *a posteriori* estimates of the variance. Linear, unbiased, minimum-variance estimators (LUMVE) are derived which provide a means for assigning weights to the experimental data. Using this estimator, data treatment algorithms can be formulated which are a significant improvement on the "traditional" Birge-ratio procedures.

Key words: adjustment of data; discrepant data; least squares; minimum variance estimators.

1. Introduction

In an earlier report [1] we attempted to include the uncertainty of the assigned errors of a set of data by including in the analysis an assumed distribution function for a universe of experimental standard deviations, and considered the assigned weight of an experimental result to be a random variable drawn from such a distribution, rather than an exact value. We then introduced a maximum likelihood criterion as the basis for determining "best" estimates of the variance of the data as well as "best" estimates of the parameters of the system.

The approach in this paper is to return to the basic concept of least squares as embodied in the Gauss-Markov theorem [2, 3] which defines the "best" estimate of any parameter as the estimate with minimum variance. The difference here is that an estimate will be sought for the variance of the input data as well as for the values of the output parameters.

2. Formulation of the Problem

As is the usual situation in a least squares analysis of the data bearing on the values of the fundamental physical constants, it is assumed that the experimental data consist of sets of quantities (y_i, s_i^2) which represent a numerical result of the i^{th} measurement and the best *a priori* estimated of its variance. The experimental measurement can be represented by the expression

$$y_i = y_i(x_\alpha) + \epsilon_i = y_i^0 + \sum_\alpha A_{i\alpha} x_\alpha + \epsilon_i \,, \qquad (1)$$

where i runs from 1 to M and α runs from 1 to N, with $M > N$, so that the system of equations is in fact overdetermined. In addition, the rank of A must be equal to N in order that a full solution may be found (if the rank of A is less than N only a projective subset of the x_α can be determined). In Eq. (1) the parameters y_i^0 and $A_{i\alpha}$ are fixed quantities, the set of parameters $\{x_\alpha\}$ are the unknowns whose values we wish to determine, and ϵ_i is the true error in the measurement.

The true error is assumed to have a variance σ_i^2 but one knows only an estimate of the variance, s_i^2, and a confidence parameter, ν_i, associated with that estimate. The basic *ansatz* of this development is to postulate the relations[1]

$$\langle \epsilon_i \rangle = 0 \,, \quad \langle \epsilon_i \, \epsilon_j \rangle = \sigma_i^2 \delta_{ij} \,,$$

$$\langle \epsilon_i \, \epsilon_j \, \epsilon_k \, \epsilon_\ell \rangle = \sigma_i^2 \sigma_k^2 \, \delta_{ij} \, \delta_{k\ell} + \sigma_i^2 \sigma_\ell^2 \, \delta_{ik} \, \delta_{\ell j} + \sigma_i^2 \sigma_j^2 \delta_{i\ell} \, \delta_{jk} \qquad (2.1)$$

$$\langle s_i^2 \rangle = \sigma_i^2 \,, \quad \langle \epsilon_i s_j^2 \rangle = 0 \,,$$

$$\langle s_i^2 s_j^2 \rangle = \sigma_i^2 \sigma_j^2 \left[1 + \frac{2}{\nu_i} \delta_{ij} \right] . \qquad (2.2)$$

In Eq. (2.2) a factor of 2 has been introduced into the definition of $\langle s_i^2 s_j^2 \rangle$ so that one may identify ν_i with the number of independent observations from which the mean and variance are deduced. However, it should be emphasized that ν_i need not be given that strict an interpretation; we use it here as a parameter (the "effective" number of independent observations, if one wishes to so identify it) defining the variance of our estimation of the variance, or less precisely as a parameter giving a measure of our confidence in the stated value, s_i^2, of the assigned variance of the i^{th} measurement. Equation (2.2) implies that the variance of the assigned variance is expressed as

$$\text{var}\,(s_i^2) = \frac{2}{\nu_i} \sigma_i^4 \qquad \text{or} \qquad \text{var}\,(s_i) = \sigma_i^2 / 2\nu_i \quad (2.3)$$

without specifically relating ν_i to the number of samples drawn from a Gaussian distribution.

*Work supported in part by NSF Grant PHY78-26467.

[1] The third component of Eq. (2.1) assumes that the distribution law for the errors is approximately Gaussian to the extent that the ratio of the fourth moment to the square of the second moment is the same as for a Gaussian, $(3\sigma_i^4)$. If the fourth moment is given by $(3 + \gamma_i)\sigma_i^4$ it will be necessary to add to this expression the additional term

$$\gamma_k \, \delta_{k\ell} \, \delta_{\ell m} \, \delta_{mn} \, \sigma_n^4$$

but we shall omit it in this discussion since in general we will not have the information required to evaluate the higher moments of the probability distribution functions. We note that this is the only point at which the specific form of a probability distribution, beyond the specification of its mean and variance, is required.

We adopt as the estimators of x_α and σ_i^2 the linear combinations

$$\hat{x}_\alpha = \sum_j C_{\alpha j}(y_j - y_j^0) \tag{3.1}$$

$$\hat{\sigma}_i^2 = a_i s_i^2 + \sum_{jk} B_{jk}^i (y_j - \hat{Y}_j)(y_k - \hat{Y}_k) \tag{3.2}$$

where

$$\hat{Y}_j = Y_j(\hat{x}_\alpha) = y_j^0 + \sum_\alpha A_{j\alpha}\hat{x}_\alpha \ . \tag{4}$$

The present approach to modifying the weights of discrepant data is embodied in Eq. (3.2). Other expressions for estimating σ_i^2 could certainly be written. The form adopted here is justified on the basis that it is a linear combination of the available estimators; s_i^2 is an *a priori* estimate while the discrepancy between the experimental value and its adjusted value is also a measure of the accuracy of the measurement. Although we could write

$$\hat{\sigma}_i^2 = a_i s_i^2 + b_i(y_i - \hat{Y}_i)^2 \ ,$$

the more general form adopted in Eq. (3.2) is to be preferred (and perhaps almost required) because the adjusted values $\hat{Y}_i$ are themselves dependent on *all* of the experimental data as expressed by Eqs. (4) and (3.1). Conversely, therefore, the estimate of the variance of the i^{th} measurement should depend on the observed variances of *all* of the experimental data, and since these are correlated quantities it is appropriate to include a full coefficient matrix B_{jk}^i rather than the scaler coefficient b_i.

In Eqs. (3.1) and (3.2) the vector a_i and the tensors $C_{\alpha j}$ and B_{jk}^i must be determined to satisfy the conditions that the estimators shall be unbiased, minimal variance estimators. The requirement that the estimators be unbiased places conditions on these coefficients. Since one must have $<\hat{x}_\alpha> = x_\alpha$ one is led to the conditions

$$<\hat{x}_\alpha> = x_\alpha = <\sum_j C_{\alpha j}\left[\sum_\beta A_{j\beta}x_\beta + \epsilon_j\right]>$$

$$= \sum_{j\beta} C_{\alpha j}A_{j\beta}x_\beta \tag{5}$$

or

$$\sum_j C_{\alpha j}A_{j\beta} = \delta_{\alpha\beta} \ . \tag{5.1}$$

Since we also require $<\hat{\sigma}_i^2> = \sigma_i^2$, one obtains the condition,

$$<\hat{\sigma}_i^2> = a_i\sigma_i^2 + \sum_{jkmn} E_{jm}B_{jk}^i E_{kn}<(y_m - y_m^0)(y_n - y_n^0)> \tag{6}$$

where

$$E_{jm} = \delta_{jm} - \sum_\beta A_{j\beta}C_{\beta m} \ . \tag{6.1}$$

To simplify this, one first writes

$$<(y_m - y_m^0)(y_n - y_n^0)> = \sum_{\gamma\epsilon} A_{m\gamma}A_{n\epsilon}x_\gamma x_\epsilon + \sigma_m^2\delta_{mn} \tag{7}$$

and notes that

$$\sum_m E_{jm}A_{m\gamma} = 0$$

by virtue of Eq. (5.1).

Equation (6) is then reduced to

$$\sigma_i^2(1 - a_i) = \sum_{jkm} B_{jk}^i E_{jm}E_{km}\sigma_m^2 = \sum B_{jk}^i T_{jk} \ . \tag{8}$$

The coefficients $C_{\alpha i}$ and B_{jk}^i are to be defined by minimizing the variances of $\hat{x}_\alpha$ and $\hat{\sigma}_i^2$; these are

$$\text{var}(\hat{x}_\alpha) = \sum_j C_{\alpha j}^2\sigma_j^2 \tag{9}$$

$$\text{var}(\hat{\sigma}_i^2) = 2a_i^2\frac{\sigma_i^4}{\nu_i} + \sum B_{jk}^i B_{pq}^i[<F_jF_kF_pF_q> \tag{10}$$

$$- <F_jF_k><F_pF_q>]$$

with

$$F_j = y_j - \hat{Y}_j = \sum_m E_{jm}\epsilon_m \ . \tag{10.1}$$

Equations (5.1) and (9) are the same as in the usual least-squares problem [2] and hence lead to the familiar result

$$C_{\alpha j} = \frac{1}{\sigma_j^2}\sum_\beta A_{j\beta}\overline{W}_{\beta\alpha} \ , \tag{11}$$

where the square symmetric matrix $\overline{W}$ is the variance matrix of the parameters $\{x_\alpha\}$, the inverse of the weight matrix, W.

$$W_{\alpha\beta} = W_{\beta\alpha} = \sum_i A_{i\alpha}A_{i\beta}/\sigma_i^2 \ , \tag{11.1}$$

$$\sum_\beta W_{\alpha\beta}\overline{W}_{\beta\gamma} = \delta_{\alpha\gamma} \ . \tag{11.2}$$

We now turn to the more difficult algebra, that of defining the coefficients B_{jk}^i so as to minimize the variance of $\hat{\sigma}_j^2$.

We obtain, directly from Eqs. (10.1) and (2.1),

$$<F_jF_k> = T_{jk} \tag{12.1}$$

$$<F_j\,F_k\,F_p\,F_q> = T_{jk}\,T_{pq} + T_{jp}\,T_{kq} + T_{jq}\,T_{pk} \tag{12.2}$$

With these expressions and Eq. (8), Eq. (10) becomes

$$\text{var}(\hat{\sigma}_i^2) = \frac{2}{\nu_i}(\sigma_i^2 - \sum_{jk} B_{jk}^i T_{jk})^2$$

$$+ 2\sum_{jk}\sum_{pq} B_{jk}^i B_{pq}^i T_{jq}\,T_{pk} \ . \tag{13}$$

Making use of Eq. (11) allows a simplification of these expressions. One introduces Eq. (11) into Eq. (6.1) and finds

$$E_{jm} = \delta_{jm} - \frac{1}{\sigma_m^2}\sum_{\alpha\beta} A_{j\alpha}A_{m\beta}\overline{W}_{\alpha\beta} \ . \tag{6.1a}$$

Then it follows,

$$T_{jk} = \sigma_j\,\sigma_k\,\delta_{jk} - \sum_{\alpha\beta} A_{j\alpha}A_{k\beta}\overline{W}_{\alpha\beta} \ . \tag{6.2a}$$

We define a new matrix, t_{jk},

$$t_{jk} = T_{jk}/\sigma_j\,\sigma_k = (\sigma_k/\sigma_j)E_{jk}$$

$$= \delta_{jk} - \frac{1}{\sigma_j\sigma_k}\sum A_{j\alpha}A_{k\beta}\overline{W}_{\alpha\beta} \ , \tag{6.3}$$

and note that t_{jk} is idempotent, i.e.,

$$\sum_n t_{jm}\,t_{mk} = t_{jk} \ .$$

Finally, we introduce a redefinition of the coefficient B_{jk}^i by writing

$$B_{jk}^i \equiv (\sigma_i^2/\sigma_j\sigma_k)\beta_{jk}^i$$

and (3.2) reduces to

$$\hat{\sigma}_i^2 = s_i^2 \left(1 - \sum_{m\ell} \beta_{m\ell}^i t_{m\ell}\right) + \sigma_i^2 \sum_{m\ell jk} \Delta_\ell t_{\ell j} \beta_{jk}^i t_{km} \Delta_m \ , \quad (14)$$

where

$$\Delta_m = (y_m - y_m^0)/\sigma_m \ ,$$

while (13) becomes

$$\mathrm{var}(\hat{\sigma}_i^2) = \frac{2\sigma_i^4}{\nu_i} \left[(1 - \sum_{jk} \beta_{jk}^i t_{jk})^2 + \nu_i \sum_{jk} \sum_{pq} \beta_{jk}^i \beta_{pq}^i t_{jq} t_{pk}\right] . \quad (13.1)$$

The derivative of this expression with respect to β_{jk}^i gives

$$(1 - \sum_{pq} \beta_{pq}^i t_{pq}) t_{jk} = \nu_i \sum_{pq} \beta_{pq}^i t_{jq} t_{pk} \ . \quad (15)$$

The trace of Eq. (15) gives

$$(1 - \sum_{pq} \beta_{pq}^i t_{pq}) \sum_j t_{jj} = \nu_i \sum_{pq} \beta_{pq}^i t_{pq} \ .$$

But

$$\sum_j t_{jj} = M - N = f$$

where f is the usual "degrees of freedom" of the adjustment. Then

$$a_i = \nu_i/(\nu_i + f) \quad (17.1)$$

$$\sum_{pq} \beta_{pq}^i t_{jq} t_{pk} = \sum_{pq} t_{jq} \beta_{qp}^i t_{pk} = t_{jk}/(f + \nu_i) \ . \quad (17.2)$$

With these expressions Eq. (14) becomes

$$\hat{\sigma}_i^2 = \frac{1}{\nu_i + f} \left[\nu_i s_i^2 + \sigma_i^2 \sum \left[\frac{y_m - \hat{Y}_m}{\sigma_m}\right]^2\right] = \frac{\nu_i s_i^2 + \sigma_i^2 \chi^2}{\nu_i + f} \quad (18)$$

and hence[2]

$$B_{jk}^i = \frac{\sigma_i^2 \delta_{jk}}{\sigma_j \sigma_k (\nu_i + f)} \ . \quad (19)$$

In this equation we express the estimate of the variance, $\hat{\sigma}_i^2$, in terms of the unknown variances σ_k^2, so that this as it stands is hardly a solution; Eq. (18) gives us an estimate of the variances only if we know the variances, but if these are known estimates are not needed. However, Eq. (18) can be reinterpreted as a consistency condition on the best values of the variances and used as an interactive algorithm.

3. The Adjustment Algorithm

Before we can go further it is necessary to define more clearly what is meant by ν_i; specifically, of the three quantities σ_i^2, ν_i, and $\mathrm{var}(s_i)$, which two are to be considered as independent estimates? If we consider σ_i^2 and ν_i to be independently estimated one should replace σ_i^2 in Eq. (18) by its estimate and write

$$\hat{\sigma}_i^2 = \nu_i s_i^2/(\nu_i + f - \chi^2) \ . \quad (18.1)$$

On the other hand, if we consider σ_i^2 and $\mathrm{var}(s_i)$ to be independently estimated, one must also replace ν_i in Eq. (18) by its estimate, $\hat{\sigma}_i^2/2\mathrm{var}(s_i)$, and hence

$$\hat{\sigma}_i^2 = s_i^2 + 2(\chi^2 - f)\mathrm{var}(s_i = s_i^2 [1 + (\chi^2 - f)/\nu_i] . \quad (18.2)$$

Since the estimates of σ_i^2 must of necessity be positive, it follows that using Eq. (18.1) would require $\chi^2 < f + \min(\nu_i)$ and this constraint has the disadvantage that the overall accuracy of the solution (as measured by χ^2) is controlled by the uncertainty assigned to the least-well-defined experimental error, independent of the value of the associated measurement and whether it is discrepant or in good agreement with the remainder of the data. Equation (18.2) also provides a constraint, $\chi^2 > f - \min(\nu_i)$, but one that is much more realistic since it states that in the presence of data uncertainties which are not well defined there is a minimum overall uncertainty that may be assigned to the adjusted values even if the data are highly consistent. In more cases than not, the situation is one of treating discrepant data, for which one expects $\chi^2 > f$.

Adopting Eq. (18.2) as the more appropriate choice, one then must carry out a least squares adjustment using weights

$$w_i = \frac{\nu_i}{s_i^2(\nu_i + \chi^2 - f)} \quad (20)$$

from which one determines the adjusted values $\hat{Y}_i$ and χ^2:

$$\chi^2 = \sum \frac{\nu_i (y_i - \hat{Y}_i)^2}{s_i^2(\nu_i + \chi^2 - f)} \ . \quad (21)$$

Equation (21) is an implicit equation for χ^2 which may be solved by iteration.

As it stands, however, Eq. (21) does not represent a convergent algorithm. Convergence can be achieved by using Newton's rule:

$$F(z) = z - \sum w_i(z)(y_i - \hat{Y}_i)^2 = z - \sum w_i(z) r_i^2 \quad (22)$$

with

$$w_i(z) = \nu_i/s_i^2(\nu_i - f + z) \ .$$

In carrying out the differentiation $F'(z)$ one may neglect the dependence of $\hat{Y}_i$ on z since it may be shown that, by virtue of Eqs. (11) and (5.1),

$$\partial \hat{Y}_i / \partial w_j = 0 \ .$$

Then, if z_n be an estimate of χ^2, an improved approximation may be found from

$$z_{n+1} = z_n - F(z_n)/F'(z_n)$$

or

$$z_{n+1} = \frac{\sum w_i(z_n) r_i(z_n)^2 + z_n \sum w_i(z_n)^2 s_i^2 r_i(z_n)^2/\nu_i}{1 + \sum w_i(z_n)^2 s_i^2 r_i(z_n)^2/\nu_i} . \quad (23)$$

At convergence, obviously, $z = \sum w_i r_i^2 = \chi^2$.

4. An Example of the Use of the Algorithm

As an example we shall look at a simple one-dimensional case: the spectroscopic data bearing on the value of α^{-1} used by Cohen and Taylor [4] in the 1973 adjustment of the physical constant. The data are given in Table 1.

[2] It is clear from Eq. (13.1) that only $\sum B_{jk} t_{tk}$ can be, or need be, determined; B_{jk} can not be uniquely defined since t_{jk} does not have an inverse. This definition for B_{jk} is the most convenient one.

TABLE 1: *Determinations of* α^{-1}.

Author	$(\alpha^{-1} - 137) \times 10^5$	σ/ppm	σ_A/ppm	ν_A	ν'	ν
Baird *et al.*	3544(54)	3.94	3.4	83	13.9	3.9
Kaufman *et al.*	3416(20)	1.45	1.3	2	2.7	2.7
Shyn	3508(46)	3.36	2.6	138	6.1	5.1
Cosens	3563(31)	2.26	2.1	4	4.8	4.8
Wesley, Rich	3563(42)	3.06	2.7	46	16.2	9.1
Muonuim hfs	3634(21)	1.53	0.2	27	1.04	1.04
H hfs	3597(22)	1.60	—	—	1	1

In order to define values of ν_i for these data we attempt to separate the total uncertainty into the so-called A and B components. The A-component is that portion of the variance which is based on a statistical analysis of repeated measurements and hence one to which the concept of degrees of freedom may be used in order to evaluate the accuracy with which the variance is determined. The B-component is then simply the remaining variance. The A-component is loosely identifiable with the "random" component and B with the "systematic" component. More precisely A refers to that component of the variance which is due to "statistical fluctuations only."

For the data of Baird *et al.*, Shyn, and Wesley and Rich the value of ν_A is the degrees of freedom of the observations; for Kaufman *et al.* and for Cosens the observations were statistically inconsistent and the data for different transitions had statistically discrepant means. In these cases the degrees of freedom were taken to be the number of transitions measured. The muonium data from Yale and Chicago were treated as a single set and the observations were analyzed for a single value for the hyperfine transition frequency and the density-shift coefficients of the buffer gases (see Ref. [4]). In this case ν_A is the degrees of freedom of that statistical analysis. For hydrogen hfs the observation uncertainties are completely negligible and the total uncertainty listed is that coming from the theoretical uncertainty in relating ν_{hfs} with α. The values of ν, the "effective" degrees of freedom of the observation, were calculated assuming that the B component of variance should be assigned 1 degree of freedom. Then we have

$$\frac{\sigma^4}{\nu} = \frac{\sigma_A^4}{\nu_A} + \frac{\sigma_B^4}{\nu_B};\ \sigma_B^2 = \sigma^2 - \sigma_A^2;\ \nu_B = 1; \qquad (24)$$

or

$$\nu = \sigma^4 \nu_A / [\sigma_A^4 + \nu_A(\sigma^2 - \sigma_A^2)^2]\ . \qquad (24)$$

The column labeled ν' in Table 1 is calculated from Eq. (24) using the given values of ν_A. It can properly be argued that the large values of ν_A do not realistically represent the precision of the observation and that mere consistency could be hiding a "systematic" error that was not considered and hence not represented in the estimation of the B component of the variance. On this basis the values of ν_A were arbitrarily restricted to the condition

$$\nu_A \leq 10 \qquad (25)$$

and the values of ν given in the last column of Table 1 are calculated with that constraint.

The iteration algorithm of Eq. (23) was applied to the data of Table 1 with an initial estimate $z_0 = f = 6$ and hence $w_i = 1/s_i^2$ from Eq. (22); the iterations to convergence are summarized below:

n	z	$\sum w_i r_i^2$	$x = (\alpha^{-1} - 137) \times 10^5$
0	6	66.318	3544
1	7.524	36.000	3527
2	10.018	21.141	3527
3	12.525	15.059	3526
4	13.458	13.614	3526
5	13.523	13.524	3526
6	13.523	13.523	3526

We see that adequate convergence has been achieved in only 4 or 5 iterations and the adjusted value is $\alpha^{-1} = 137.03526(21)$ with $\chi^2 = 13.52$ (in comparison with $\chi^2 = 66.32$ based on the initial uncertainties). Quoting an error by "external consistency" (i.e., applying the Birge ratio to this error) yields $\alpha^{-1} = 137.03526(32)$. This value of χ^2 for 6 degrees of freedom has a *pro forma* probability $P(\chi^2 > 13.52 \mid 6) = 0.04$. The data are shown in Fig. 1 where the triangular error-bars represent the assigned uncertainties of Table 1. The expansion factors, by which the

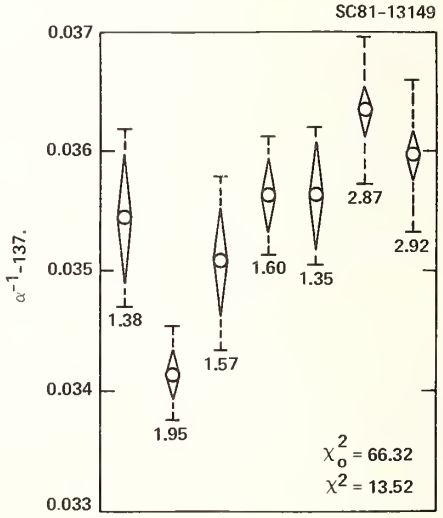


FIGURE 1. *The 1973 data pertaining to the fine-structure constant,* α^{-1} *(from Ref. 4). The data of Kponou, Hughes et al. have been omitted. See Table 1. Using the a priori variances gives* $\alpha^{-1} = 137.03545(34)$ *(external consistency). Using adjusted variances (expansion factors indicated) gives* $\alpha^{-1} = 137.03526(32)$.

algorithm modifies the errors, is given below each entry and the dotted lines indicate the magnitude of the extended errors. In comparison with this result, the fitting using ν' in place of ν gives a negligibly different value, $\alpha^{-1} = 137.03526(20)$ with $\chi^2 = 13.53$. An adjustment with the Kaufman *et al.* result omitted would lead to $\sigma^{-1} = 137.03583(16)$, $\chi^2 = 6.15$ with 5 degrees of freedom.

If one uses Eq. (18.1) as the condition for the estimation of the variances, Eq. (23) is replaced by

$$z_{n+1} = \frac{\sum r_i(z_n)^2/s_i^2 + f\sum r_i(z_n)^2/\nu_i s_i^2}{1 + \sum r_i(z_n)^2/\nu_i s_i^2}. \quad (23.1)$$

However, application of this algorithm leads to a nonphysical solution that implies negative weights (i.e., negative values for some of the $\hat{\sigma}_i^2$), and hence must be rejected.

5. Efficiency of the Estimators

The efficiency of an unbiased estimator in comparison with other unbiased estimators is measured by the reciprocal of its variance; therefore, we want to evaluate Eq. (13.1) and compare it with the *a priori* variance given in Eq. (2.3).

Substituting Eqs. (17.1) and (17.2) into Eq. (13.1) gives

$$\mathrm{var}(\hat{\sigma}_i^2) = \frac{2\sigma_i^4}{\nu_i}\left[\frac{\nu_i^2}{(\nu_i + f)^2} + \nu_i \sum_{jk} \frac{\beta_{jk}^i t_{jk}}{\nu_i + f}\right] = 2\sigma_i^4/(\nu_i + f). \quad (23)$$

This result is quite consistent with the interpretation of ν_i as the number of "effectively independent observations" which enter into the estimate of σ_i^2. The *a priori* estimate was attributed to ν_i effectively independent observations; when the residues $y_i - \hat{Y}_i$ are included as part of the information used in obtaining an improved estimate, the effective number of independent observations is increased by f (not by M, since the M residues are not independent statistical quantities).

6. Conclusions

We may conclude from Eq. (18) that this procedure will lead to the converged value $\chi^2 = f$ in only two cases: if $\nu_i = 0$ for all i (and the algorithm reduces to the Birge algorithm), or if the *a priori* estimates s_i^2, are equal to the converged values. In general, however, this algorithm will yield a converged value of χ^2 which will be between f, the degrees of freedom, and χ_0^2, the initial value calculated using the *a priori* estimates s_i^2 for $\hat{\sigma}_i^2$. This will be true whether the initial estimate χ_0^2 is larger or smaller than f. The algorithm will push χ^2 towards its expectation value and expand the variances of $\chi_0^2 > f$, $\chi_0^2 < f$. This is what a statistically effective algorithm would be expected to do. If the converged value of χ^2 is significantly different from f this information is important and should not be ignored, and the reasons for a statistically improbable result should be explored. It can come from several possible sources: (1) there may be discrepant data which should be rejected from the analysis because they are afflicted with unknown systematic errors, (2) the *a priori* estimate of the variance of an experimental result is in error and should be modified, (3) the accuracy of the estimate of *a priori* variance may be in error and the corresponding value of ν_i should be altered.

It is almost impossible to identify an individual measurement as being "too good," and it is only when χ_0^2 is improbably small that one would be justified in decreasing the variance measures or increasing the initial values of ν_i; on the other hand large discrepancies are more easily identified and an item which gives a large normalized residual, $|(y_m - \hat{Y}_m)/\hat{\sigma}_m|$, should be critically examined. Although its assigned variance can be increased in order to reduce the contribution of this measurement to χ^2, it will in general be found that an increase in $\hat{\sigma}_m$ which is sufficiently large to reduce the normalized residue to an acceptable value would reduce the effect of this measurement on the values of the adjusted parameters, $\hat{x}_\alpha$, to such a degree that very little would be lost by eliminating this measurement completely from the analysis.

References

[1] E. Richard Cohen, Maximum Likelihood Criterion for Adjusting Weights in a Least-Squares Analysis, Rockwell International Science Center, Report SCTR-78–9 (August 1978).

[2] E. R. Cohen, K. M. Crowe, and J. W. M. DuMond, *The Fundamental Constants of Physics* (Interscience, New York, 1957), pp. 237–240.

[3] E. R. Cohen, Rev. Mod. Phys. 25, 709 (1953); see also A. C. Aitken, Proc. R. Soc. Edinburgh 55, 42 1934–35; and R. L. Plackett, Biometrika 36, 458 (1979).

[4] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973) (Table 23.1, p. 697).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

GRAVITATIONAL ACCELERATION, MASS, AND ELECTRICAL QUANTITIES

Present Status of the Absolute Measurement of Gravitational Acceleration

A. Sakuma

Bureau International des Poids et Mesures, F-92310 Sèvres, France

This paper reviews the recent work on the absolute measurement of gravitational acceleration g, covering the last decade since PMFC-I in 1970. The single principle involved in the precise absolute measurement of g to better than 1 part in 10^8 is the observation of free fall in the gravity field: All the ten or so laboratories presently engaged in this measurement employ only this principle of free fall and no longer the traditional reversible pendulum. A large number of new absolute gravity stations (about 50 by the end of 1980) have been created in Europe, North America, Asia, and Oceania by transportable absolute gravity meters of several laboratories. These stations are aimed at improving the accuracy of the world gravity network, IGSN 1971, and also at monitoring in the future the secular variation of the network. Following the production of a successful prototype apparatus constructed jointly by IMGC (Torino) and BIPM, two examples of a modified commercial absolute gravity meter have been manufactured (1981) by Jaeger Industries S.A. (France). One is in operation at the Geographical Survey Institute, Japan, and the other at BIPM. These instruments employ the symmetrical free rise and fall of a cube-corner reflector together with a novel "multi-station" interferometric measurement procedure. An accuracy of better than 1 part in 10^8 is currently obtained. The principles together with an outline of the performance of these new instruments are described. Thus absolute gravimetry is now beginning to provide a new means for several metrological and geophysical studies, accessible to all who are interested in this field.

Key words: absolute gravimetry; acceleration due to gravity; earth tide; free fall; IGSN 71; secular variation.

1. Introduction

As I pointed out in my paper at the PMFC-I in 1970 [1], it was in the late 1960's that absolute gravimetry (absolute determination of acceleration due to gravity) became established as an important new means for the study of the earth's gravity field [2]. This new age of absolute gravimetry was opened by a remarkable improvement in the relative accuracy of this measurement from 1 part in 10^6 to 1 part in 10^8. This jump in accuracy had been achieved by several new experiments carried out in this period [3–5] with the methods of "free fall" using modern techniques, namely: interferometric length measurements with laser sources, high speed chronometry, and improved vibration control. Based on these new absolute measurements (ten results, of which eight were obtained by the first transportable absolute gravimeter of Hammond and Faller [5]), combined with a large number of gravity data collected during the previous twenty years by relative measurements (~1200 values by pendulums, ~24000 values by spring gravimeters), a new world-wide gravity network was established in 1971. It is called the International Gravity Standardization Net 1971, "IGSN 71" [6].

This IGSN 71 has been approved and adopted by the 15th General Assembly of the International Union of Geodesy and Geophysics: IUGG (Aug. 1971, Moscow) as the new international gravity reference, replacing the Potsdam gravity net. IGSN 71 contains 1854 gravity stations in 494 towns where the gravity values are determined by a single least-squares adjustment of the absolute and relative data. The accuracy of this gravity net is estimated to be better than ±1 part in 10^7 ($\pm 1\ \mu m \cdot s^{-2}$) at each station of the net.

Since the establishment of IGSN 71, the principal task of absolute gravimetry has been to check and improve the accuracy of this gravity net and also to monitor in the future its secular variation. All the ten or so laboratories that have been engaged in modern absolute gravimetry employ only the method of free fall which is the single method which can lead to an accuracy better than 1 part in 10^8 for g. Thus, the traditional pendulum method has ended its long career of three centuries (since the epoch of l'Abbé Picard in ~1670) with a final best accuracy of 0.3 ppm ($3\ \mu m \cdot s^{-2}$) [7, 8].

The present report is intended to summarize the evolution of absolute gravimetry since 1970, developed in ten laboratories throughout the world. Brief remarks are also given on each of the particular devices, and the advantages and disadvantages of the different methods are discussed. It is very encouraging for absolute gravimetry to note that the local values of gravity are required, in some laboratories of metrology, with more and more accuracy for electrical standards experiments and it is said [9] that in the future, the stability of the kilogram with respect to time may be monitored by using the stability of the Josephson voltage standard and the calculable capacitor with an accurate knowledge of the local value of the gravity field.

2. Transportable Absolute Gravimeters

Absolute gravimetry has been one domain of precise physical measurements which has been particularly favored by the recent development of the stabilized laser light source. The characteristics of such sources, accuracy of wavelength, high light intensity, and smallness of beam divergence, make them ideal for modern gravime-

try by the free fall method which requires an ultimate accuracy in the length (distance of fall) measurement in vacuum. This requirement is almost perfectly satisfied by a laser interferometer system in which the falling object is used as a reflector of the laser beam. In addition, the continuous observation of free fall by the laser fringe-counting method permits the establishment of a large number of intermediate "multiple stations," e.g., every 2048 fringes counts, to increase the accuracy of the measurements by using a statistical treatment of the large amount of data obtained. It also has need of a very sharp timing signal during the free fall which can be satisfied by modern digital electronics.

For reasons such as these all ten laboratories presently engaged in absolute gravimetry employ only the method of free fall combined with the Michelson type laser interferometer [Fig. 1a and b].

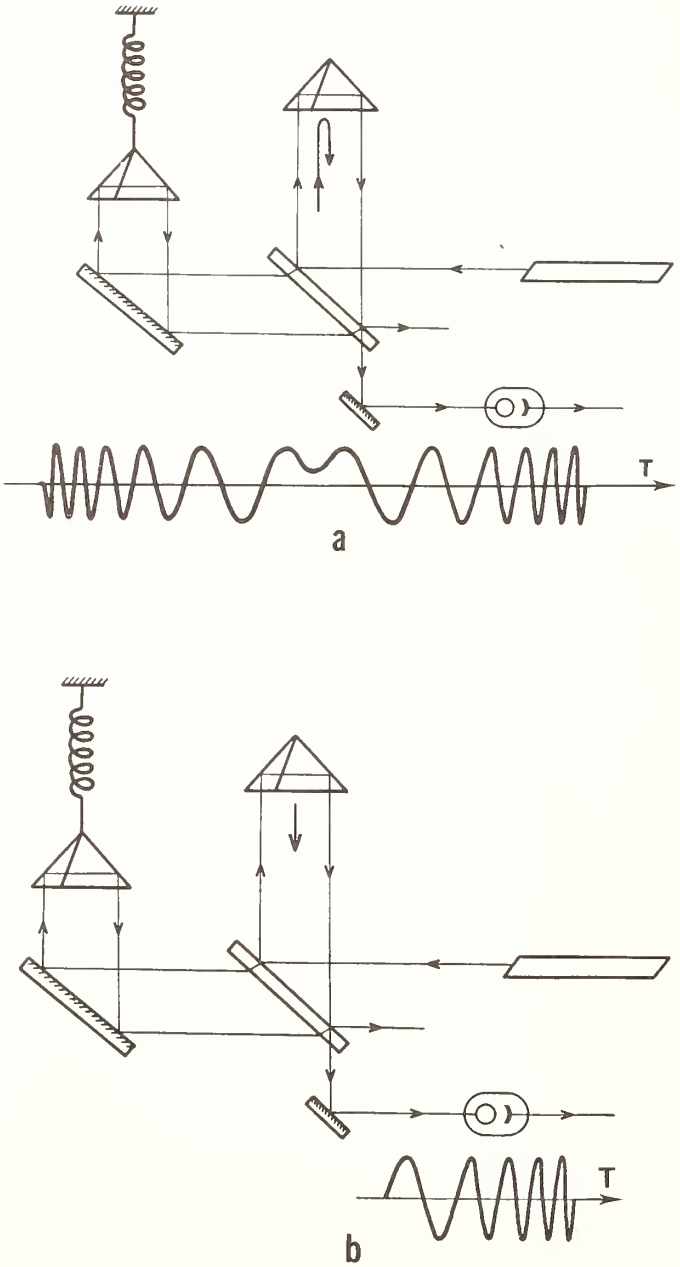


FIGURE 1. *(a) Michelson interferometer system with symmetrical free rise and fall method. (b) Same system with simple free fall method.*

2.1 Transportable Apparatus With Free Fall Method

(a) Air Force Geophysics Laboratory (AFGL), Hanscomb AFB, Mass. [10]. This apparatus has been called the second generation of transportable gravimeter developed from the first generation apparatus invented by Hammond and Faller and used at a number of international sites during the period 1968–69 [5], and which also greatly contributed to the establishment of the new world gravity net, IGSN 71. This is also the first apparatus which introduced the method of multiple stations in the free falling trajectory (60 cm) of a cube-corner prism in vacuum. A high-resolution time digitizer ($\leq$0.125 ns) combined with a minicomputer measures and memorizes the consecutive time intervals at a large number of stations (500 ~ 700) during the free fall and performs a least-squares average. This averaging by the multiple-station method has made it possible to determine the vertical gradient of gravity and also to detect a systematic in-phase vibration of a few nanometers amplitude of the apparatus due to the triggering of the falling object.

An apparatus using the simple free fall method is very sensitive to residual air resistance. This air resistance can be calculated and can also be measured; it is typically $\sim 1 \times 10^{-4}$ of g for a pressure of 1 Pa after a 1 m fall. Therefore a low pressure, of order 10^{-5} Pa, is necessary for the falling object if the error due to air resistance is to be less than 1×10^{-9} of g. A low pressure of this order is not so easy to realize, especially in the case of the gravimeter. Because the vacuum vessel contains some movable elements, among which is the falling object, the vacuum may change during the free fall. In order to reduce this resistance effect, the falling object is dropped with its container; Fig. 2 shows this in a schematic way. At the top of the chamber a soft iron block is attracted by an electromagnet and just before release of the block the ballast mass drops about 1 cm. Upon release, the springs pull the inner chamber and the ballast mass together, resulting in the cube corner assembly being roughly 6 mm above its resting place within the inner chamber. This configuration has resulted in approximately a 100 to 200 times reduction in the effect of air resistance. Thus with this "chamber-in-a-chamber technique," a vacuum of 10^{-3} Pa will become sufficient. One other advantage of this method is to have a free-fall object which has no magnetic material on it, while still employing the mechanically simple technique of releasing an object with an electromagnet.

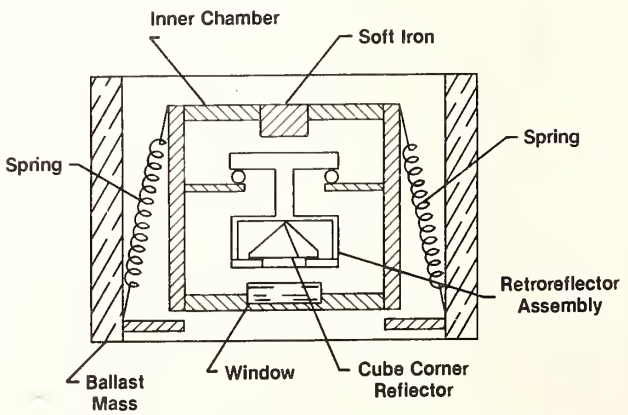


FIGURE 2. *Free-falling unit in AFGL second generation apparatus (schematic diagram).*

This apparatus is completely automated and the data are analyzed and corrected for gravity tide in real time. The apparatus has a total mass of about 700 kg when packed for air transport and is contained in ten boxes which can be handled by one or two people. It has been used at twelve new sites in the United States in the last few years and a final accuracy of a few parts in 10^9 is expected. The details of this apparatus will be presented at this conference.

(b) JILA/NBS, Boulder, Colo. [11]. This is another transportable laser-interferometer gravimeter which employs the simple free fall of a cube-corner prism and a final accuracy of a few parts in 10^{-9} is aimed at. In order to reduce the effect of air resistance, this apparatus also uses the chamber-in-a-chamber technique with a servo-controlled follower system. Figure 3 shows schematically the functioning of this falling system. The chamber of the falling cube corner is equipped with an optical position detector. This chamber is attached to a servo amplifier-motor system so that the vertical displacement of the chamber is controlled by this servo system. A sapphire sphere fixed on the falling cube corner serves as a collimator lens for the optical position detector. To start the free fall of the cube corner the chamber is displaced rapidly downward about 2 mm. Thus the sapphire sphere enters the horizontal light beam of the position detector to activate the servo amplifier-motor system so that the chamber follows the free fall of the cube corner.

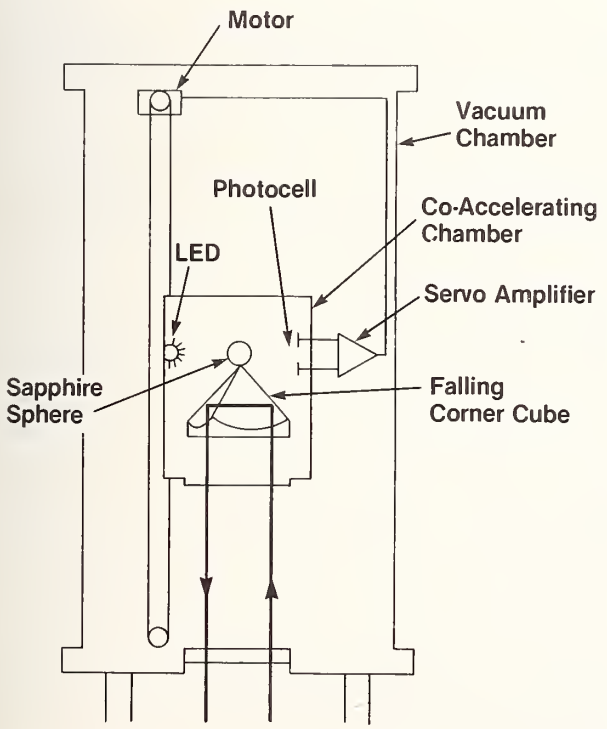


FIGURE 3. *Servo-driven "chamber-in-a-chamber" JILA free-falling unit (schematic diagram).*

The window in the bottom of the chamber is made of a very thin (~5 μm) transparent membrane called a pellicle. This membrane ensures that the slight inclination or rotation of the chamber will only negligibly vary the optical path length during the fall of the object. This falling system is used at present at a pressure of about 1 Pa and by means of small holes the inner chamber is also maintained at about the same pressure. Therefore, this chamber-in-a-chamber system must cancel out the effect of air resistance with a factor of about ten thousand in order to reduce the error of this effect sufficiently to assure a final accuracy of 3 parts in 10^9. Apparently it is not likely to attain such a high degree of cancellation so it will be necessary to determine experimentally the coefficient of this resistance by changing the pressure or by comparison with another type of absolute gravimeter to determine the correction to the value of gravity. This gravimeter is equipped with a long-period suspension system ($T \geq 40$ s) for the mechanical isolation of the reference cube corner from background disturbances. It is also completely automated and gravity measurements can be repeated quickly (150 drops in ten minutes). The recent results show that an accuracy is now obtained of better than 1 part in 10^8 and automatic measurements continued during several days permitted the form of the perturbing waves of the gravity due to the earth tides to be clearly recorded. The details of this apparatus will also be presented at this conference.

(c) Institute of Automation and Electrometry (IAE), Novosibirsk [12, 13]. This transportable gravimeter has been in operation since the beginning of the 1970's and also employs the simple free fall of a cube-corner prism with a stabilized-laser interferometer. The apparatus has no anti-vibration suspension. A set of measurements is carried out to obtain an arithmetical mean value of about 150 drops of the prism. The measurements are completely automatic and after each set of 150 drops (which takes about 30 minutes) the alignment and adjustment of the apparatus are checked. This apparatus uses a large electromagnet to reset the falling object after each drop so the uncertainty in the measured value of g due to this magnetic effect had to be evaluated in order to attain an accuracy of about 2 parts in 10^8. However, a direct comparison of this gravimeter at the BIPM gravity station carried out in September 1977 demonstrated that the uncertainty due to the magnetic effect had been overestimated (it was found to be less than a few parts in 10^9). The estimated total uncertainty of this apparatus is $\pm 1.6 \times 10^{-8}$ in which the three major contributions are: the error of vertical alignment and collimation of the interferometer: 1.0×10^{-8}; error of laser wavelength: $\pm 8 \times 10^{-9}$; and error of air resistance correction: $\pm 8 \times 10^{-9}$.

This apparatus has been used, since 1976, to determine the gravity at several international stations in Levodo, Potsdam, Sèvres, and Singapore. The measurements in Sèvres were aimed at finding the possible systematic errors of this apparatus. After about 4000 drops, made over a period of one week, the value of gravity by this gravimeter was in good agreement with the BIPM value (to within 4 parts in 10^9).

(d) National Institute of Metrology (NIM), Beijing [14]. This transportable gravimeter, constructed in 1974, also employs the simple free fall of a cube-corner prism, together with a Lamb-dip-stabilized laser interferometer. In this apparatus the falling object assembly (380 g) contains, at its top, a cylinder of soft iron (~45 g) required for the electromagnetic release of the object. In order to check the perturbing effect of this magnetic material on the gravity value, a similar object was made with a ferromagnetic material for which the magnetic moment is about 36 times stronger than that of the usual object. Experiments made with this object could not detect any evidence of systematic error. Therefore the error due to the soft-iron piece is considered to be less than 1 part in 10^8 of g.

In April 1980 this gravimeter was transported to Sèvres to compare with the BIPM apparatus. The results agreed to better than 1 part in 10^8 with an uncertainty of

$\pm 1.6 \times 10^{-8}$. Details of this apparatus will be given at this conference.

(e) Earthquake Research Institute (ERI), Tokyo [15, 16]. This transportable laser-interferometer gravimeter uses a "cat's eye" as the falling object instead of a cube corner. The rotation of a cat's eye during free fall causes a significant variation of the optical path in the interferometer. So in this apparatus the tolerance of rotation rate of the falling cat's eye is very small, only ± 0.3 mrad·s^{-1}. To realize a non-rotating free fall, a piezo-ceramic base is used to support the cat's eye. At the moment of the beginning of free fall the cat's eye stays momentarily fixed in space due to a rapid piezoelectric contraction of the ceramic base, and during this short time interval (~0.25 ms) the ceramic base is rapidly moved away from the fall path of the cat's eye. This device permitted the realization of a very small angle of rotation (10 sec of arc during 30 cm fall). The gravimeter employs also a multiple-station method, the interference fringes being sampled at millisecond intervals by means of a photographic recording. The uncertainty of the measured value of gravity is presently about $\pm 4 \times 10^{-8}$ of which the two major components are the microseismic perturbation: $\pm 2.6 \times 10^{-8}$; and the uncertainty of the He-Ne laser wavelength: $\pm 2.0 \times 10^{-8}$.

2.2 Transportable Apparatus With Symmetrical Free Rise and Fall

(a) Istituto di Metrologia G. Colonnetti (IMGC), Torino [17, 18]. The method adopted in this transportable gravimeter is the observation, from an inertial reference point, of the symmetrical free rise and fall motion in a vacuum.

This method [18] is the most promising due to its inherent high accuracy and its great freedom from systematic errors such as residual air resistance, timing errors, Doppler effect, and so on [1]. In this symmetrical method, a cube-corner reflector (mirror) is projected vertically upwards and crosses two defined horizontal stations whose separation, H, is known. Two independent time intervals corresponding to the upward and downward passage across the lower station T_l, and the upper station T_h, are measured and the value of gravity is obtained from: $g = 8H/(T_l^2 - T_h^2)$. This method is particularly advantageous in the case of a transportable apparatus due to the fact that the effect of the residual air pressure on the falling object is cancelled out to first order since on the upward portion the air resistance or drag acts in the same direction as gravity while during the falling part this air resistance force is in the opposite direction. Thus the vacuum system of the gravimeter can be made very simple: it consists only of a primary pump, of small capacity, which facilitates transportation and saves preparation time for the gravity measurements. This first transportable gravimeter using the symmetrical method was developed in 1974 at the IMGC, Turin after a long technical collaboration with the BIPM, Sèvres. The two essential parts of this apparatus are: a Michelson interferometer in a pressure-tight case and a vacuum chamber in which a cube-corner mirror, forming a reflector of the interferometer, is launched vertically by a catapult (Fig. 4a). The distance H and the time of flight of the reflector are measured simultaneously by a bi-directional fringe-counting method (Fig. 4b) using a Lamb-dip-stabilized He-Ne laser, the wavelengtn of which, 633 nm, is periodically calibrated

A specially-designed catapult (see next paragraph) throws the cube-corner mirror of 70 g about 50 cm in a

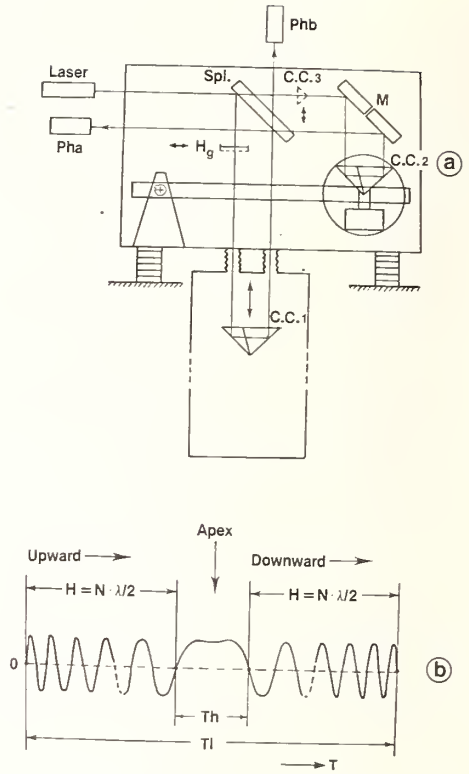


FIGURE 4. *(a) Optical scheme of the IMGC gravimeter: Ph: photomultiplier; Hg: mercury-pool (for levelling); Spl: beam splitter. (b) Distance and time determinations by laser interference fringes (see text).*

vacuum while giving it a rotation of less than 0.01 rad per second and a vertical deviation of the trajectory of about 10^{-4} rad maximum. Thus the acceleration due to the rotation and the horizontal velocity of the projectile remains within 5 parts in 10^9 of g. Two kinds of antivibration devices are employed in this apparatus to attenuate the perturbation of gravity due to ground motion and the mechanical shocks caused by launching the projectile. One is an inertial suspension of the reference cube corner by a long period ($\geq$20 s) vertical seismometer in a rigid pressure-tight case (Fig. 4a). The other consists of a set of active compensators of vertical vibrations, namely, three piezo-electric supports on which the interferometer is installed. By these means the perturbation of gravity is reduced over 20 times and the residual acceleration on the reference cube corner is normally less than 1 part in 10^8 of g. Because of the complexity of the manipulation of the launching device, it was difficult to completely automate this gravimeter. However, one operator can repeat the launching of the cube corner and can calculate a gravity value every 3 to 4 minutes, resulting in about 20 measurements during one hour and, owing to the high repeatability of each result, one set of 20 measurements is quite sufficient to determine a mean value of gravity accurate to 1 part in 10^8 after the application of the tidal correction. The coefficient of the residual air resistance was determined experimentally by changing the pressure in the vacuum chamber in the range 0.01 Pa to 100 Pa. The global coefficient of air resistance includes two other effects, namely, the effect of buoyancy amounting to about -1×10^{-8} of g per pascal in this case; and the effect of the diminution of the laser wavelength due to the refractive index of the residual air: $+3 \times 10^{-9}$/Pa; together these lead to -32×10^{-9}/Pa in a free rise and fall of 0.5 m. Practically, this global ef-

fect is almost negligible in this apparatus because during gravity measurements the vacuum is maintained normally at 0.02 Pa. It is worth noting that the global coefficient of air resistance in this symmetrical method is about 3000 times smaller than the coefficient of air resistance in the simple free fall method.

This apparatus has been used since 1976 to create a number of new gravity stations of 1×10^{-8} g accuracy, about 30 stations in Europe and about 10 in North America, and some 12 stations are projected in Asia during 1981. Not only have these expeditions contributed to an improvement in accuracy of the IGSN 71 but also the absolute value of gravity obtained at each station will permit the monitoring and detecting of future secular variation of gravity by means of the periodic repetition of the measurements.

The first absolute calibration of IGSN 71 was made at the 17 stations of the European calibration line in 1976–78 [18]. The results revealed that the uncertainty of IGSN 71, estimated to be 1 part in 10^7, was in general much lower ($\leq 5 \times 10^{-8}$) than expected.

(b) Jaeger Aviation Division, Levallois-Perret (France)/BIPM. Following the successful absolute gravimetry results obtained by the first prototype transportable apparatus just described, work has been continued at BIPM aimed at improving the performance of this type of apparatus. At the same time the Aviation Division of Jaeger S.A. (France) showed an interest in developing a commercial version of the improved transportable absolute gravimeter. The first such apparatus was manufactured in 1980–81 by Jaeger based on the studies carried out by BIPM. Of the two instruments manufactured so far, one is in operation at the Geographical Survey Institute, Tsukuba (Japan) and the other has recently been delivered to BIPM. This new absolute gravimeter employs the symmetrical free rise and fall of a cube-corner mirror with the multiple-station interferometric measurement procedure: during the free rise and fall, the position of the cube corner is observed continuously by means of an iodine-stabilized laser interferometer using a subnanosecond (≤ 0.1 ns) time digitizer. A large amount of data (a total of about 1300 measured relative positions and times) thus obtained is recorded and used to calculate the following values by a microprocessor integral with the gravimeter: a value of gravity at a well-defined height, the value of the vertical gradient of gravity throughout the trajectory, and the proportional factor that links deceleration force (due to residual pressure) to the velocity of the falling object. Figure 5 shows the general view of the apparatus: the total mass of all assembled components: ~400 kg; the overall dimensions in the operating state: 1.95 m high, 0.9×0.9 m at base level. This transportable gravimeter requires one day preparation for assembly and one half day for disassembly by two operators, and about 20 measurements of gravity can be made by a semi-automatic procedure during one hour. Because of the high accuracy and good repeatability of the measurements (≤ 1 part in 10^8), there is no need to make a large number of measurements. Normally, at one station a total of about 100 measurements made over one or two days, composed of five sets of 20 measurements at different phases of the tide correction, are quite sufficient to give an accuracy better than 1 part in 10^8 of g.

The most important advantage of the symmetrical free rise and fall is the possibility of the "self-checking" of the reliability of each measured value of gravity. This checking can be made by comparing the two different values of gravity obtained in the same launching; one is the value of g obtained during only the upward motion,

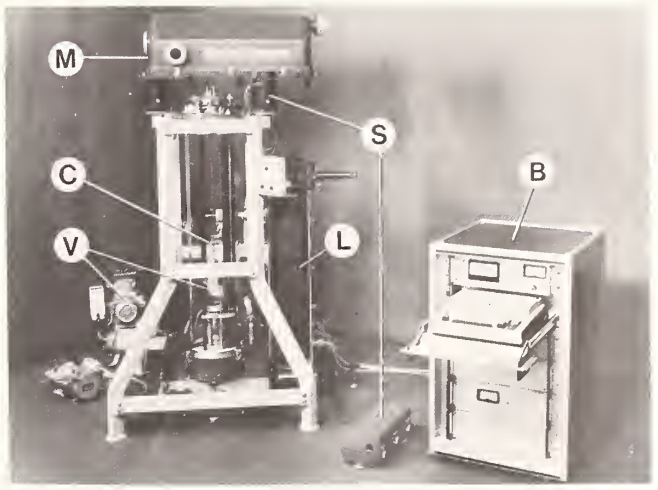


FIGURE 5. *General view of commercialized gravimeter "Jaeger." M: Michelson interferometer in a vacuum chamber; C: catapult mechanism and corner-cube reflector in a vacuum cylinder; V: vacuum-pumping system; S: anti-vibration suspension; L: iodine-stabilized laser; B: chronometer, computer, and printer.*

the other is the value during only the downward motion. This difference is influenced by the residual air resistance, Doppler effect, and electronic response of the chronometer, but it can be well determined experimentally and the mean of the two values of gravity must be equal to the integrated value of g obtained by both upward and downward motions. Therefore this self checking permits the determination of the statistical weight of each experiment and if a discrepancy is identified by this checking, the result must be rejected. Because of this self-checking feature, a gravimeter using this symmetrical method can be regarded as a veritable absolute apparatus which has essentially no need to be compared with other types of gravimeter. In spite of these important advantages, the symmetrical method has been utilized only at a few laboratories: NPL, Teddington [3] and NML, Sydney [20] for stationary gravimeters of 1×10^{-7} accuracy level. The reasons that the symmetrical method is less popular than the simple free fall, are: (i) the free-fall method is incomparably easier to realize than the symmetrical method; (ii) even with free fall, a final precision of a few parts in 10^9 can be obtained if the apparatus is compared with or calibrated by other symmetrical apparatus; (iii) once such a calibration is done, the free-fall apparatus in its highly-automated version will become a very handy field-usable apparatus (e.g., the JILA/NBS apparatus) replacing in future the spring-type relative gravimeters (this is a very important trend in modern absolute gravimetry); and (iv) in the symmetrical method, if the design of projectile and catapult is not quite satisfactory, the additional perturbing effects, e.g., mechanical shocks, outgassing, rotation, oscillation of the projectile, etc., may easily vitiate the inherent advantages of the method.

The catapult used in this commercialized gravimeter has been described [21] but it seems useful to recapitulate schematically with Fig. 6. The projectile (1) is a cube-corner reflector, 5 cm high with a mass of 70 g, designed so that the optical center of the reflector coincides with the center of gravity to within 10 μm in order to minimize the effects of any rotation which might occur. The three aluminized mirrors of the corner reflector are fixed on a duralumin support in which six holes of 6 mm diameter are made every 60 degrees around its

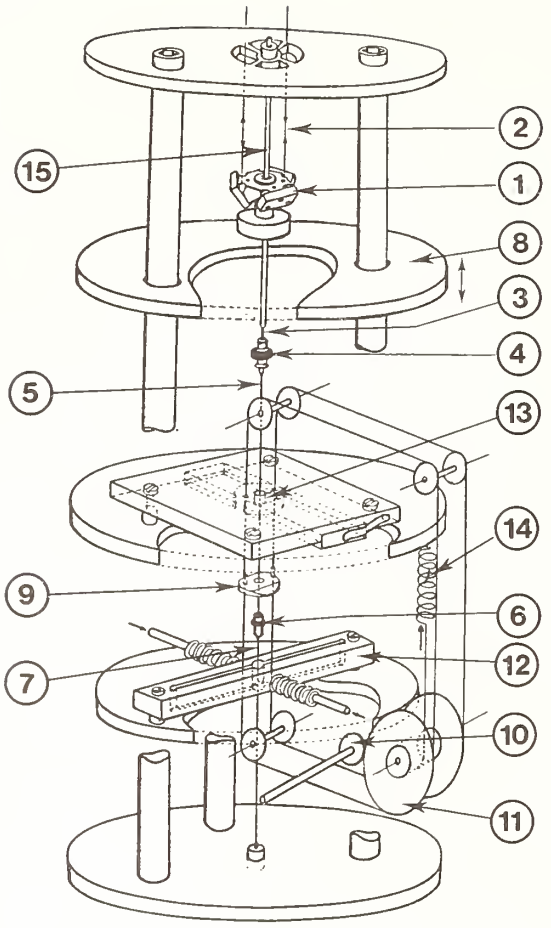


FIGURE 6. *Schematic drawing of the catapult with corner-cube reflector in flight.*

vertical center axis to let the laser beam (2) of the interferometer pass freely. This duralumin support also has a central hole of 11 mm diameter in the vertical direction and the top of this hole is slightly tapered. This tapered top of the reflector rests before launching on a conical piece (4) containing a shock absorbing device, attached at its top to the elastic cord (3) of the catapult and at its bottom to a cylindrical releasable piece (6) by a nylon wire (5). The piece (6) is also attached to a weak elastic cord (7) pulling it downward so that the release piece (6) moves always only in the vertical direction. Before launching, the reflector (1) sits on an elevator table (8) and is turned around its vertical axis so that the laser beam (2) falls correctly in a specified hole. Then the table (8) descends slowly and leaves the reflector (1) hanging on the conical piece (4). An annular piece (9) placed coaxially with the wire (5) can be displaced vertically by rotating a shaft with clutch (10) and by a pulley (11). By these mechanisms the piece (6) guided by the ring (9) is engaged into a hole of the releasing piece (12) by the pressure of compressed air. Then the ring (9) rises to be locked into a fixed stage (13) and the shaft (10) is declutched to liberate the pulley system (11) from mechanical friction. When the piece (6) is released, the reflector (1) is accelerated upwards by the traction of the elastic cord (3). After an accelerated rise of about 15 cm, the released piece (6) is stopped by the fixed stage (13). Likewise, the conical piece (4) which has accelerated the reflector is also stopped so that the reflector (1) begins its free rise. By the shock given by the released piece (6) to the fixed stage (13), the locked ring (9) is liberated and is pulled onto the piece (6) by the downward traction of the spring (14) which opposes and equilibrates with the upward traction of the piece (6) due to the elastic cord (3).

After free rise, the reflector (1) falls and sits again on the conical piece (4) so that the elastic cord (3) is stretched. During this braked fall, the two pieces (6) and (9) follow this downward motion of the reflector by means of the traction of the spring (14). During the decelerated motion, a friction disc in the pulley (11) absorbs the energy so that the reflector is stopped smoothly without bouncing. This catapult throws the reflector while giving it a rotation of less than 0.01 $rad \cdot s^{-1}$ and with a random horizontal velocity of 0.2 $mm \cdot s^{-1}$, resulting in a random perturbation on the value of gravity of ~2 parts in 10^9 if these quantities are not measured in each throw. The mass of the moving part of the catapult is only several grams and this mechanism is used as a catcher and a soft brake as well as the catapult. There is no other element moving or oscillating during the free rise and fall of the reflector and the effect of outgassing by this catapult is negligible. In addition, the catapult is mounted on a vibration absorber, therefore the mechanical perturbation due to the operation of the former is in general much smaller than the level of the ground-motion noise.

The elastic cord (3) generates some electrostatic charge by its deformation in vacuum. The effect of this on the reflector is eliminated by a metallic tube (15) placed around the elastic cord. This electrostatic shield tube (diameter 5 mm) maintains the capacitance (~10 pF) constant between the reflector and ground during the flight. Due to this constancy of the capacitance the effect of electrostatic charge on the reflector in flight, it was confirmed, was negligible, even when the reflector was charged artificially up to one hundred volts.

(c) Institut für Angewandte Geodäsie (IAG), Frankfurt [22]. Here, there is in progress another attempt at constructing a transportable gravimeter using the symmetrical rise and fall method. The projectile consists of a cube corner made of quartz while a second cube corner serves as a mass of a vertical seismometer. The cube corners are parts of a Michelson interferometer for measuring the distance. The times are measured by a 2 ns counter and the interference fringes due to the rise and fall of the projectile are counted simultaneously. A multiple station method will also be employed.

2.3 Remarks on the Transportable Apparatus

The above-mentioned eight transportable gravimeters can be classified into the following three types:

(i) Straightforward free-fall type (IAE, NIM, ERI). This type of apparatus has no protection against mechanical vibrations and therefore drop-to-drop scatter is very important (~5×10^{-7} g). Nevertheless, the mean value of a large number of drops is accurate up to about one part in 10^8 after application of several corrections due to systematic effects. For this kind of apparatus, automation of the measurements is indispensable to obtaining a sufficient number of drops.

(ii) Sophisticated free-fall type (AFGL, JILA). Use of the chamber-in-a-chamber technique, antivibration devices, and the multistation method permit one to obtain an accuracy of few parts in 10^9. Once compared with a symmetrical apparatus, this type will be a very useful apparatus for field operation.

(iii) Symmetrical free-fall type (IMGC, Jaeger, IAG). The advantage of this type is well known and its commercialization is very important. Future efforts are to be made for automation, or miniaturization for ease of transport, or for the improvement of accuracy beyond 10^{-9}.

3. Permanent Stations

In August 1975, at the 16th General Assembly of the IUGG in Grenoble, the International Association of Geodesy (IAG) adopted the following (partial) resolution [23] about the future extension of the permanent absolute gravimetry station net: "The IAG recommends that absolute gravity measurements should be made both in stable regions and in regions where secular or long-period gravity variations are expected and should be repeated after sufficiently long intervals. To this end: (a) About 10 permanent stations (or observatories) should be established where absolute gravity can be observed with microgal accuracy ($\sim 10^{-8}$ g) to investigate possible global variations and correlation in these long-term changes of g . . . ".

At that time, one station of this kind already existed, at BIPM Sèvres, having been in operation since 1967, while at the International Latitude Observatory (ILO), Mizusawa (Japan) construction of a similar apparatus had just started in collaboration with BIPM.

Since then, however, no new permanent station with a stationary absolute gravimeter has been established. The explanation of this fact is two-fold: first, the construction of such a permanent station takes, in general, a very long time—several years at least, with an important investment in personnel as well as in finance, and it is also necessary to obtain the technical assistance of experienced laboratories in this field. Second, it is during this interval that transportable absolute gravimeters have entered upon the scene (IMGC, AFGL, IAE, NIM) and demonstrated that a new absolute gravity station of $\sim 1 \times 10^{-8}$ accuracy can be created very easily within a few days. Thus, all laboratories newly interested in absolute gravimetry have adopted the transportable type to create their absolute stations and the measurement of gravity, although not in a permanent installation, can be repeated periodically. In such circumstances, the existing permanent station at BIPM as well as the future station of ILO, Mizusawa (presently in preliminary operation) are regarded as special stations where gravity measurements can be made with sufficient accuracy ($\sim 10^{-9}$ g) for the comparison and calibration of different kinds of transportable absolute gravimeters.

3.1 Earth Tide Measurements at an Absolute Station

The gravimetric earth tide is known as the most important (periodical) perturbing effect of the Earth's gravity field (~0.3 ppm, max. at mid-latitude). In order to improve the accuracy of the tide correction an earth-tide-recording spring gravimeter was constructed in 1973 at BIPM. This gravimeter employs a nulling method (the spring length is kept always constant by an electrostatic feedback system) to eliminate the effect of the spring's hysteresis.

This combination of the tide gravimeter (relative apparatus) with the absolute apparatus presents mutual advantages: the absolute apparatus permits one to evaluate the drift of the relative apparatus and to calibrate it, while the relative apparatus provides *continuously* for the absolute apparatus the true, observed correction of the earth tide which is needed to determine the mean value of gravity. The instrumental drift of this tide gravimeter was found to be very small (less than 5 parts in 10^{10} per day) and this permitted ready detection of perturbations of gravity connected with atmospheric pressure changes, the coefficient of which varies according to different meteorological conditions from -4 nm·s^{-2} to -2 nm·s^{-2} per millibar (100 Pa). Thus the recording of the earth tide is now becoming an indispensable means for accurate absolute gravimetry and reciprocally, in the near future, absolute gravimetry of ultimate accuracy ($\leq 10^{-9}$) will become the best way to calibrate the amplitude and the phase of the tides without perturbing the tide apparatus.

A new type of earth-tide gravimeter has been developed since 1972 [24]. In this apparatus, called a superconducting gravimeter, the mechanical spring is replaced by a magnetic field generated by the persistent current flowing in two superconducting coils. The extremely low drift of this gravimeter is a result of the inherent stability of a persistent current in a superconductor. The only moving part in the gravimeter is a superconducting hollow lead sphere of 3.6 g which is levitated by the magnetic field and the position of which is measured electronically. This gravimeter has been recently produced commercially [25] and an extremely low rate of drift, ±6 parts in 10^9 of g per year, is reported. It may not, of course, be prudent to use only this gravimeter for the study of the secular variation of gravity, due to possible instrumental perturbations, e.g., liquid helium refilling, but the combination of this superconducting gravimeter with an absolute gravimeter will become a powerful new means for the study of secular changes of gravity.

3.2 Long-Term Stability of Gravity

At the BIPM station, a significant variation of gravity has been observed: +4 parts in 10^8 during the last 14 years since 1967. A significant portion of this long-term variation of gravity was observed during relatively short periods in 1969–72 [26] but the causes have not been identified up to the present. One defect at this station is that, at present, there is no means to measure the variation of the underground water table. In order to clarify this kind of variation of gravity, it is projected to create several satellite stations around the BIPM station by means of the transportable absolute gravimeter. The first satellite was selected at the Geophysical Institute of BRGM at Orléans, about 130 km south of Sèvres, where the variation of the underground water table has been monitored by means of a piezometer. The difference in gravity of about 1.05 mm·s^{-2} between Sèvres and Orléans will be calibrated and its constancy will be monitored by the transportable absolute apparatus. Such gravity liaison with a satellite station will in the future supply us with information on the question of whether the observed variation is a local phenomenon or is occurring on a regional (even global) scale.

4. Conclusion

The present world-wide activities in absolute gravimetry have been summarized. Transportable absolute gravimeters have demonstrated their capabilities at the level of 1 part in 10^8 of g, corresponding to a resolution of 3 cm height change, by use of modern technologies: stabilized lasers, sub-nanosecond electronics, computer processing of data, and automation of measurements. A large number of new absolute gravity stations (about 50 up to 1980) has been established with international colla-

boration and has verified the accuracy ($\leq 1 \times 10^{-7}$) of IGSN 71. This work, that is, extension of a precision absolute world gravity net, will be continued with these gravimeters more intensively and more efficiently in the 1980's, covering new areas (e.g., Africa, Antarctica) where the gravity data are not yet sufficiently abundant. It is notable that, for the first time, a transportable absolute gravimeter has been manufactured commercially (Jaeger 1980–81) using the best principle, namely, the symmetrical free rise and fall of an interferometer reflector. The permanent gravity station at BIPM has been used frequently for comparisons and calibrations of various types of absolute and relative gravimeters. Because of its necessity and uniqueness, this permanent station has been maintained and improved with instrumental reinforcements: earth-tide recording gravimeter, iodine-stabilized laser, transportable absolute gravimeter (BIPM-Jaeger) for establishment of satellite stations.

During the 1980's the principal subjects of research in absolute gravimetry will be: (i) trials for the ultimate gravimeter working at the 10^{-9} level or below, e.g., by use of a multiple-beam reflection interferometer; (ii) numerical assessments of global secular variation of the gravity field; and (iii) studies of the relation between gravity variation and crustal movements including volcanic and earthquake phenomena.

References

[1] A. Sakuma, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.), Spec. Publ. 343 (Aug. 1971), p. 447.

[2] J. J. Levallois, Bull. Géodésique **99**, 111 (1971).

[3] A. H. Cook, Philos. Trans. R. Soc. London, Ser. A: **261**, 211 (1967).

[4] A. Sakuma, Procès-Verbaux Séances Com. Int. Poids Mes. **37** (58th session), 55 (1968).

[5] J. A. Hammond and J. E. Faller, J. Geophys. Res. **76**, 7850 (1971).

[6] C. Morelli *et al.*, Publ. Spec. No. 4, Assoc. Intl. Geod. (1974).

[7] T. Schüler *et al.*, Veröffentlichungen des Zentralinstituts Physik der Erde, No. 10, Potsdam (1971).

[8] E. Hytöten, Publication No. 75 of the Finnish Geodetic Institute (1972).

[9] P. T. Olsen, M. E. Cage, W. D. Phillips, and E. R. Williams, IEEE Trans. Instrum. Meas. **IM-29**, 234 (1980); and these proceedings.

[10] J. A. Hammond, Boll. Geofis. Teor. Appl. **20**, 401 (1978); and these proceedings.

[11] J. E. Faller, R. L. Rinker, and M. A. Zumberge, Boll. Geofis. Teor. Appl. **20**, 355 (1978); and these proceedings.

[12] G. P. Arnautov *et al.*, Sov. J. Quantum Electron. **9**, 333 (1979); Appl. Opt. **13**, 310 (1974).

[13] J. D. Boulanger, Publication No. 89 of the Finnish Geodetic Institute (1979).

[14] Fang Yong Yan *et al.*, Wu Li (Physics) **8**, 1 (1979) [in Chinese].

[15] I. Murata, Bull. Earthquake Res. Inst. Univ. Tokyo **53**, 49 (1978) [in Japanese].

[16] H. Hanada, J. Geodetic Soc. Jpn. **24**, 191 (1978).

[17] L. Cerutti *et al.*, VDI-Berichte No. 212, 49 (1974).

[18] L. Cannizzo, G. Cerutti, and I. Marson, Nuovo Cimento C **1**, 39 (1978).

[19] C. Volet, C. R. Acad. Sci. **224**, 1815 (1947).

[20] G. A. Bell, D. L. H. Gibbings, and J. B. Patterson, Metrologia **9**, 47 (1973).

[21] A. Sakuma, Publ. Internat. Sympos. on Earth's Gravitational Field and Secular Variation in Position, Sydney, Australia (1973), p. 674. See also Recl. Trav. Bur. Int. Poids Mes. **4**, Sect. 13 (1973-1974).

[22] R. Brein, Some Possibilities to Improve the Accuracy of Absolute Gravity Meters, presented at the 17th Gen. Assembly of the Intl. Union of Geodesy and Geophysics (IUGG), Canberra, Australia (1979).

[23] Resolution No. 16, Bull. Géodésique **118**, 377 (1975).

[24] W. A. Prothero and J. M. Goodkind, J. Geophys. Res. **77**, 926 (1972).

[25] The GWR Superconducting Gravimeter, GWR Instruments, San Diego, Calif. 92121 (1980).

[26] *The International Bureau of Weights and Measures 1875–1975*, Ed. by C. H. Page and P. Vigoureux, Natl. Bur. Stand. (U.S.), Spec. Publ. 420 (May 1975), p. 114. (Translation of the BIPM Centennial Volume *Le Bureau International des Poids et Mesures 1875–1975*).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A New, Portable, Absolute Gravimeter

M. A. Zumberge, J. E. Faller,* and R. L. Rinker

Joint Institute for Laboratory Astrophysics, National Bureau of Standards and University of Colorado, Boulder, CO 80309

We report on the performance of a new and easily portable apparatus for the absolute measurement of the acceleration of gravity. Rapid acquistion of data and high accuracy result from the use of a drag-free dropping chamber that descends with the falling object whose acceleration is measured interferometrically. Preliminary results indicate an absolute accuracy of 6 parts in 10^9.

Key words: absolute gravity meter; freely falling corner cube; stabilized laser interferometer.

1. Introduction

The absolute determination of g, the acceleration due to the earth's gravity, has long been a measurement of considerable importance. While g continues to play a significant role in the determination of certain physical constants and standards, with the accuracies obtainable today, this measurement now has broad applications to geophysics. The gravimeter that has been developed at JILA is based on the free-fall interferometric method which has undergone many refinements since it was first used in the 1960s [1–4]. A freely falling mass that contains a corner cube serves as the mirror in one arm of a Michelson-type optical interferometer. The times of occurrence of selected interference fringes give its position as a function of time, and from this the acceleration of gravity, g, can be calculated.

Over the surface of the earth the value of g varies by less than 0.5% from its nominal value of 980 cm/sec^2. This variation is due primarily to the earth's rotation which produces a centrifugal term causing g to decrease toward the equator. This rotation also causes the earth's shape to deviate from a sphere which in turn produces a further dependence on latitude. Local topography and density variations are two other important factors.

The vertical gravity gradient near the surface is approximately 3 μGal/cm (1 Gal ≡ 1 cm/sec^2). Because of this, cm-sized vertical motions of the earth's crust will show up as variations in g of a few parts in 10^9. Since such motions often occur very slowly, the long-term stability in any instrument—if it is to detect them—is of utmost importance. Relative gravimeters which are sensitive to μGal level changes in gravity have existed now for a decade or more; however, they must be operated continuously at a single site if unambiguous results are to be obtained. Furthermore the sensitivity of relative measurements is often limited by instrumental drifts that are not simply linear in time. For these reasons, a portable absolute instrument whose measurements depend on atomic standards and whose absolute accuracy approaches 3 parts in 10^9 would be a very useful tool in the study of tectonophysics. We have successfully built an instrument that appears to approach this accuracy while maintaining the highly portable character necessary for field applications.

*Staff Member, Quantum Physics Division, National Bureau of Standards.

2. Design

Figure 1 is a schematic representation of the system. A Michelson interferometer is arranged in such a way that one of the corner cubes is allowed to fall freely. An interference fringe results from each half wavelength of the laser light traversed in the dropped object's descent.

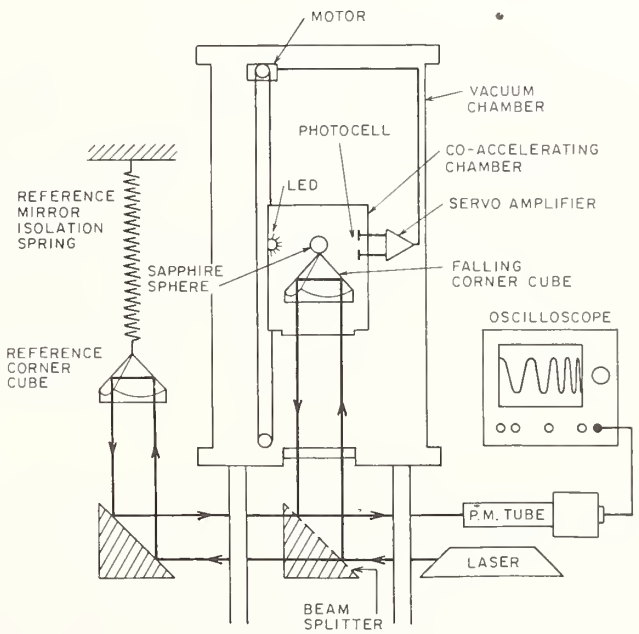


FIGURE 1. *Schematic of absolute gravity apparatus.*

Over the 20 cm length of free-fall during which the measurement is made, 600 000 fringes are generated. The zero-crossing time of every twelve-thousandth fringe is carefully measured and then stored in a computer. The resulting 50 data points are fitted to a quadratic curve which then gives the dropped object's acceleration. This process, which requires only a few seconds, can be repeated many hundreds of times in one hour.

The quantity actually measured is the second time derivative of the optical path difference between two optical elements, one of which is in some way attached to the earth's surface while the other (the falling one) is accelerated by gravity as well as by the vector sum of all other forces acting on it. The interpretation of this meas-

urement as the absolute value of local gravity requires that we fully understand the limitations set by all other forces and processes involved. This entails four fundamental considerations: one, non-gravitational forces on the dropped object must be eliminated or reduced to an acceptable level; two, the optical path difference must be translated into a physical length based on an accurate knowledge of the laser's wavelength and the index of refraction appropriate to the residual gas at the operating level of vacuum; three, the acceleration of the reference frame from which the measurement is made must be close to zero; and four, the electronic system that counts and times the occurrence of interference fringes must do so without introducing any systematic errors. The design of the instrument proceeded with these four considerations as the underlying guidelines.

Three recent innovations have contributed to the successful development of this instrument: One, a drag-free dropping chamber has been built to shield the dropped object from non-gravitational forces; it is described in detail below. Two, a long-period isolation spring developed at JILA [5] has been used to isolate the interferometer's reference corner cube from vertical motions in the laboratory that would otherwise accelerate the reference frame with respect to which the g measurement is made. [Note that with a freely falling corner cube and a vertical inertially suspended reference corner cube, the interferometer input is differential in that vertical motions of the rest of the instrument cancel out.] Three, a newly implemented Zeeman stabilization scheme also devised at JILA [6] has been used to stabilize the He-Ne laser that serves as the light source in the interferometer. Its wavelength is periodically (for example on a monthly basis) compared with an iodine stabilized laser developed and built for us by the Length and Mass Measurements and Standards Division of NBS. This absolute wavelength standard, taken together with a rubidium frequency standard (used as the basis of our measurement electronics), lets us base our gravity measurement directly on secondary length and time standards whose accuracy provides the requisite stability for the study of long-term changes in gravity resulting, for example, from tectonic deformation of the earth's crust in seismically active regions.

3. The Instrument

Air drag makes it impossible to approach any reasonable accuracy without dropping the corner cube in a vacuum. Some years ago at NBS, Tate [7] developed a technique in which the dropped object fell in a vacuum chamber that was dropped simultaneously, minimizing the force due to momentum transfer between the falling object and residual gas molecules in the chamber. To avoid the vacuum welding and other materials-related problems associated with ultrahigh vacuum systems, we have invoked a similar technique. In our instrument the dropped object is contained in a servo controlled motor-driven elevator which moves inside the main vacuum system. This elevator effects the release and then tracks the falling object—without touching it during the measurement. As a result, the object falls with the residual gas molecules rather than through them, while the differential motion between object and elevator is minimized. This results in a substantial reduction in air drag and allows the g determination to be carried out in a more moderate vacuum.

Figure 2 is a schematic representation of the drag-free dropping chamber. The dropped object rests in

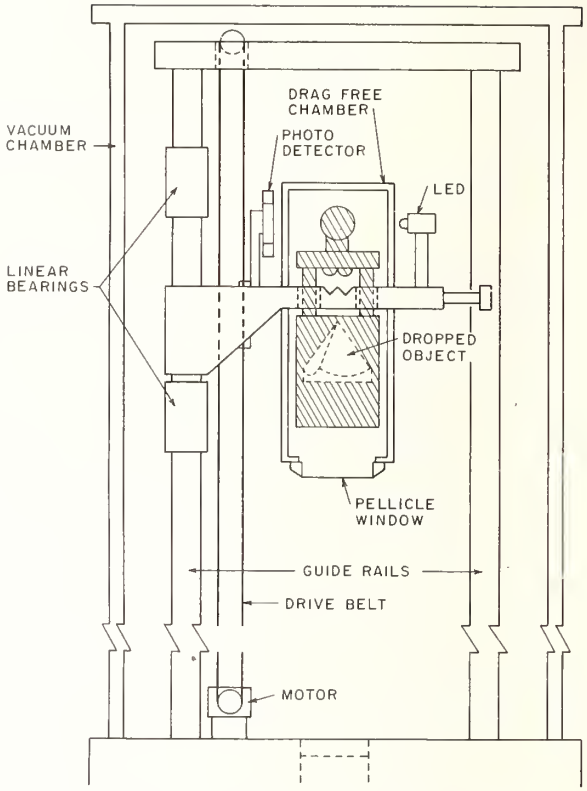


FIGURE 2. *Schematic of dropping system.*

kinematic mounts on a chamber that can be driven along vertical guide rails by a thin stainless steel belt connected to a dc motor. The position of the dropped object relative to this drag-free chamber is measured by focusing light from an LED through a lens attached to the dropped object onto a position sensitive photodetector. The error signal thus derived is used to control the motor that accelerates the chamber downward, leaving the dropped object floating inside. Near the bottom of the drop the chamber is first servoed to gently arrest the dropped object's fall, and then used to return the dropped object to the top of the track for the next measurement. Figure 3 is a photograph of this system.

In addition to shielding the dropped object from drag forces the falling chamber also serves to remove other non-gravitational forces. The chamber provides an electrically conducting shell to completely surround the dropped object so that external electrostatic fields do not affect the measurement. Also, the purely mechanical character of the release removes the necessity for having any sort of magnetic support or release mechanism that might subsequently result in an unwanted magnetic force. Finally, buoyancy effects are removed because a pressure gradient cannot exist in a zero-g environment.

The detection of the interference fringes using a photomultiplier results in a sinusoidal signal whose frequency is proportional to the falling object's velocity. A zero crossing detector and a digital scalar are used to convert this signal into a series of about 50 pulses, each of which corresponds to the dropped object having fallen 6000 wavelengths (12,000 fringes) or about 0.38 cm. The times of occurrence of these pulses referred to an arbitrary but common zero are measured to within 0.2 nsec by commercial electronics and stored in a minicomputer. A quadratic least-squares fit to these data determines g. Analysis of the residuals indicates that the length measurement errors are about 0.001 wavelength. The analysis

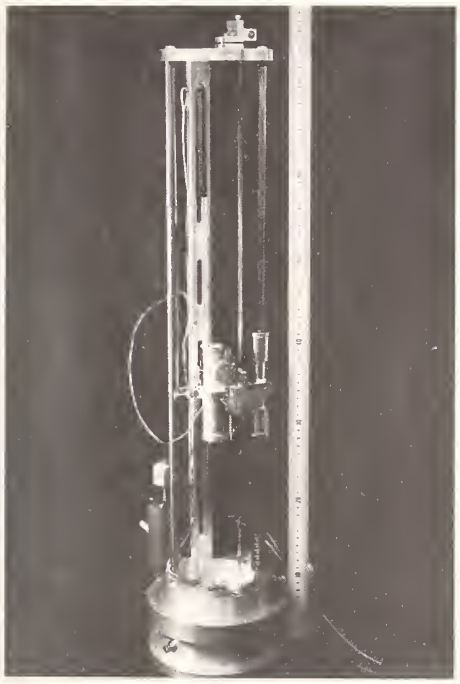


FIGURE 3. *The dropping system.*

currently requires about 4 sec so 150 drops can be made in 10 minutes.

Figure 4 is a photograph of the apparatus. The dropping mechanism is inside a vacuum chamber which is supported by three folding legs. Beneath this is a base that supports the long-period isolation spring [5] and contains the associated optical components that comprise the interferometer. The small computer and auxiliary electronics occupy two half racks. One person can disassemble and load the entire system into a small van in about one hour. Reassembly takes one to two hours. This ease of portability should facilitate rapid field measurements.

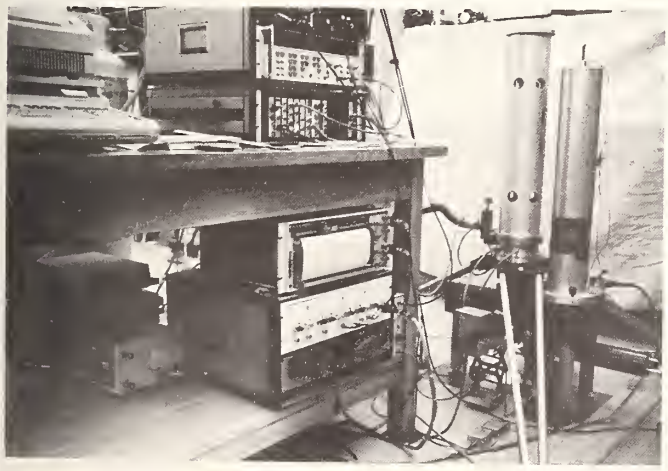


FIGURE 4. *Complete apparatus.*

4. Results

Despite the fact that the instrument is portable, it has been moved only a few times thus far in its development. Until now, the primary emphasis has been the detection, understanding, and elimination of systematic errors. Most of the data have been taken in the sub-basement of the laboratory wing of JILA on the University of Colorado campus in Boulder. Some data sets, however, have been taken in a building on the outskirts of the National Bureau of Standards grounds in Boulder. This latter site is significantly quieter in that it has less vibration than the JILA sub-basement, but is less convenient in terms of modifications. We plan to reoccupy the NBS site, as well as several others around the Western United States, in the near future.

Figure 5 is a histogram of 150 drops which comprises one 10-minute data set. The standard deviation in such sets varies from as low as 20 μGal to as high as 70 μGal during unusually noisy periods caused by poor weather or increased human activity nearby.

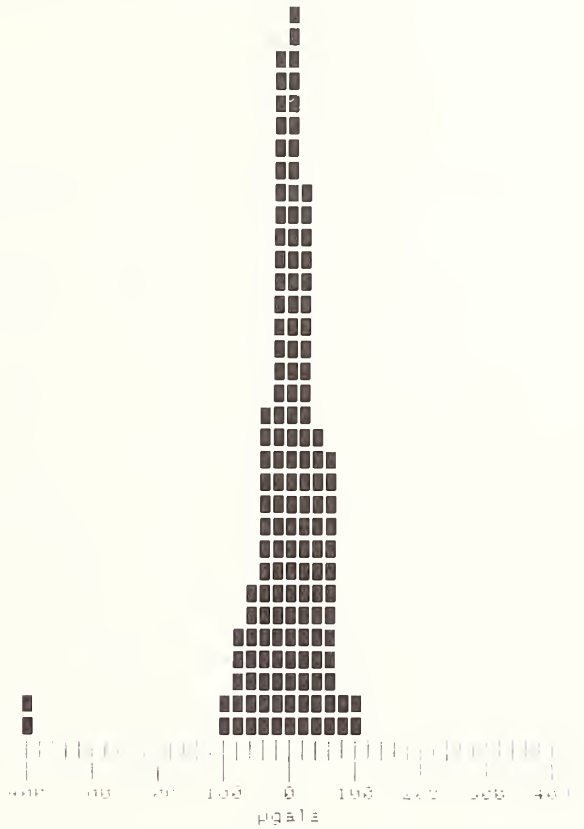


FIGURE 5. *Histogram of one run.* σ = *38* μ*Gal;* $\sigma/\sqrt{N}$ = *3.1* μ*Gal; start time = 5:15; stop time = 5:25.*

Figure 6 is the result of two days of continuous operation at about 70% of the maximum possible data acquisition rate. The tidal effects of the sun and moon can easily be seen. Subtraction of the theoretical tides results in an r.m.s. deviation of about 6 μgal. Meteorological effects, which can be significant at this level, have not yet been removed.

5. Errors

The high repeatability of a measurement (i.e., the precision) is unfortunately not always an indication of the accuracy. Opportunities for systematic errors abound in this type of measurement and only by varying certain (at first glance) inconsequential experimental parameters can their effects, if any, be determined with confidence. Even then it is always possible to overlook some significant processes. Because of space limitations we cannot give an in-depth error analysis here, but we will briefly

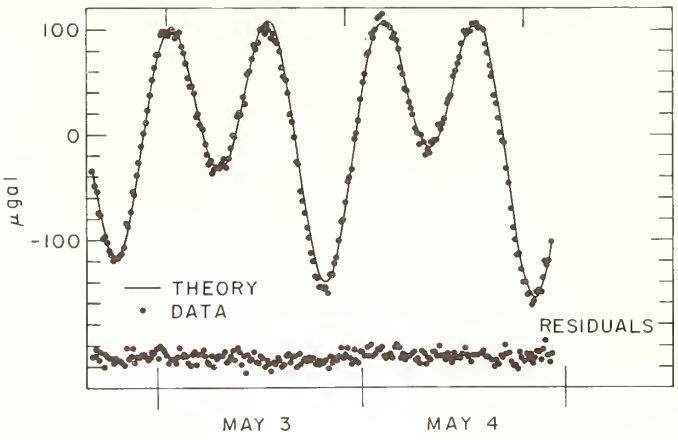


FIGURE 6. *Earth tides.*

discuss sources of error that have been considered to date.

The effects of possible magnetic and electrostatic forces on the dropped object are as yet undetectable. The addition of magnetic fields larger than those normally encountered by the dropped object have been found not to affect the value of g. The result of changing the separation of the dropped object and the co-falling chamber does not cause a shift in the measured value of g, indicating again that electrostatic or magnetic forces between them are not significant.

Initially we had hoped to be able to work at a pressure of 10^{-3} mm of Hg and were encouraged by measurements in which we purposely introduced a relative velocity between the falling chamber and the dropped object. No significant shift in g resulted from rather large relative velocities (2–4 mm/sec). However, onset of rapid temperature changes produced a several hundred μGal transient. Rough calculations of the magnitudes of forces due to temperature and pressure gradients across the dropped object as well as the effects of surface and cavity outgassing—all of which would occur during times of departure from thermal equilibrium—suggested that a lower vacuum was not only prudent but necessary. In our present operating vacuum range of 10^{-5} to 10^{-6} mm of Hg, the problems introduced by changing temperatures were found to be greatly reduced.

Errors that stem from rotation or translation of the dropped object [8], the effect of small rotations in optical elements in the interferometric path, the consequences of optical misalignment, and any effects of optical feedback on the laser, are all believed to be less than 10^{-9} of g. Other considerations, which have thus far yielded no indication that errors are present, include nonrandom mechanical vibrations introduced systematically by the measurement and frequency-dependent time delays in the electronics.

6. Conclusions

Although several of these points deserve further examination, we feel that most of the sources of systematic errors have been recognized and are well in hand. It is, however, always possible that further investigation may reveal other difficulties, but it seems unlikely that our quoted value will change by more than 10 μGal. To date, we have made 400 sets of 150 drops during which the system appeared to operate correctly (e.g., we were not subjecting it to special test conditions) over a one-month period, yielding a value (reduced to the floor using a gradient of 2.3 μGal/cm) of 979.608557 Gal. The standard deviation of the results of these 400 sets is 5.9 μGal or about 6 parts in 10^9 (see Fig. 7), while the standard deviation of the mean of a single 150 drop set ranges from 3 to 7 μGal. Recognizing the as yet uncompensated meteorological effects, the results are statistically consistent. We feel the one sigma uncertainty of 6 μGal is a conservative representation of the instrument's capabilities. Further tests of the instrument's accuracy are planned, including direct comparison with other absolute instruments.

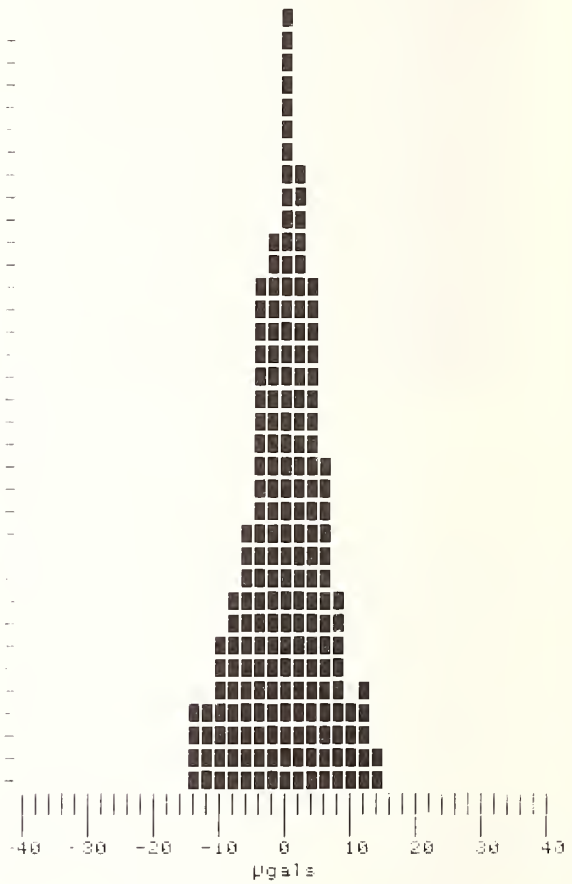


FIGURE 7. *Histogram of 419 data sets. Mean = 979.608557 Gal; σ = 5.9 μGal.*

This work was supported in part by the Air Force Geophysics Laboratory and in part by the National Bureau of Standards as part of its research program on improved precision measurement techniques for application to basic standards.

References

[1] J. E. Faller, J. Geophys. Res. 70, 4035 (1965).
[2] L. Cannizzo, G. Cerutti, and I. Marson, Nuovo Cimento C 1C, 39 (1978).
[3] J. A. Hammond and R. L. Iliff, in *Applications of Geodesy to Geodynamics*, Proceedings of Ninth Geodesy/Solid Earth and Ocean Physics (GEOP) Research Conference, Ed. by I. I. Mueller (1978), p. 245.
[4] A. Sakuma, in *Proceedings of the International Symposium on the Earth's Gravitational Field and Secular Variations in Position*, Ed. by R. S. Mather and P. V. Angus-Leppan (Dept. of Geodesy, Univ. of New South Wales, Sydney, 1973), p. 674.

[5] R. L. Rinker and J. E. Faller, these proceedings.
[6] T. Baer, F. V. Kowalski, and J. L. Hall, Appl. Opt. 19, 3173 (1980).
[7] D. R. Tate, J. Res. Natl. Bur. Stand. Sect. C: 72, 1 (1968).
[8] A. Sakuma, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.), Spec. Publ. 343 (Aug. 1971), p. 447.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

"Super Spring"—A Long Period Vibration Isolator

R. L. Rinker and J. E. Faller*

Joint Institute for Laboratory Astrophysics, National Bureau of Standards and University of Colorado, Boulder, CO 80309

We have devised a new mechanical isolating device which we call a "super spring." The super spring isolator makes use of the fact that a mass suspended by a long spring is effectively isolated (from vibrations) for all frequencies higher than the system's natural resonance. We have developed a method of electronically terminating a 30 cm-long spring in such a way that the mass suspended from it behaves as if the spring were one kilometer or longer in length. This permits us to provide isolation for frequencies as low as 0.02 Hz. We will discuss the principle, the results of shake-table tests, and the implications of this technique for measurement science.

Key words: inertial reference; long period seismometer; mechanical vibration isolation.

1. Introduction

This preliminary determination I generally make the night before the deflections and periods are determined, which in Oxford is best done on Sunday night between midnight and 6 a.m. The daytime, of course, is out of the question, owing to the rattling traffic on the stones in St. Giles', about a quarter of a mile away; and all nights except Sunday night the railway people are engaged making up trains and shunting, which is more continuous and disturbing to the steadiness of the ground than a passing train. Even these come through at intervals on a Sunday night, and this limits the accuracy with which the periods can be observed...

The traffic and the trains are not the only causes of disturbance. Wind, by pressing upon the building and neighboring trees, of course shakes the ground; but on Sept. 9–10, a particularly quiet night, I had to leave, owing to a sudden disturbance producing a pendular motion of 15 divisions, or 150 units, and for some time there was not quiet. As the motion was clearly produced by a lurch of the whole instrument and table carrying it, and was greater in amount than any traffic in the busiest part of the day had ever produced, and was moreover free from the high period tremor characteristic of human disturbance, I at once set it down to an earthquake...

I had hoped to have made a greater number of experiments under more widely differing conditions, but the strain which they entail is too severe, for not only have I had to give up holidays for the last three years, but to leave London on Saturdays and occasionally to sit up all Saturday and Sunday nights at the end of a week's work. The conditions, therefore, are too difficult for such an extended series as I should like to make to be possible, and I must after one or more effort, leave the problem to others who have leisure, and what is of far greater consequence, a quiet country place undisturbed by road and railway traffic, and who possess the knowledge and manipulative skill which the experiment requires...

Professor C. V. Boys "On the Newtonian Constant of Gravitation" Philosophical Transactions of the Royal Society of London for the Year MDCCCXCV. Volume 186.

Isolation of experiments from mechanical vibrations has been a long-standing problem for experimental physicists. Since the time of C. V. Boys, experiments have become increasingly more sophisticated and sensitive and this has served only to increase the problems resulting from vibration. And while measurements have increased in accuracy and complexity, the approaches to vibration isolation have remained at a rather simple level. Today, commercial isolation systems are available which provide isolation from frequencies greater than 2 Hz. However, for experiments sensitive to frequencies of 1 Hz and below, these systems are not helpful. In particular, today's state-of-the-art absolute gravity measurements require isolation at frequencies much lower than 1 Hz. Further, the low frequency limit of earth-bound gravity wave experiments will be limited by the character and the quality of the isolation that can be afforded. While at first glance it might seem that the problem of lowering the isolation regime ("corner" frequency of an isolator) is a simple matter of getting weaker springs and larger masses, there are very real practical limitations. A simple calculation shows that the resonant frequency of a mass m which stretches (by a distance x) a spring of spring constant k is equal to the frequency of a simple pendulum whose length l is equal to the amount of this stretch:

$$mg = kx = kl$$

$$\omega = \left[\frac{k}{m}\right]^{1/2} = \left[\frac{g}{l}\right]^{1/2}. \qquad (1)$$

Thus a spring with a period of 60 sec would have to be stretched about 1 km—which rules out its use for normal laboratory applications. Since the frequency of a mass-spring system depends on g and the stretch length of the spring, it is clear that in order to achieve low-frequency isolation one must do something other than attempt to straightforwardly utilize a simple spring system.

To overcome these practical limitations, we have developed what we call a "super spring"—a mass supported by a finite length spring but whose period—and isolation characteristics—is that of a spring very much longer. In order to understand how one might get a simple spring system of tractable size to behave as if it

*Staff Member, Quantum Physics Division, National Bureau of Standards.

were, say 1 km in length, imagine a mass hanging from a spring stretched 1 km. As the mass oscillates up and down all the coils of the spring oscillate with the same period. The coils near the bottom would have an amplitude nearly equal to the amplitude of the mass, while the coils near the top would have an amplitude that is small compared to the amplitude of the mass. At the half-way point, the amplitude would be exactly half. If we were to study the motion of a coil 30 cm above the mass as the rest of the spring was lengthened, we would find that the longer the spring got, the closer this coil would track the mass. In fact, if the spring were infinite in length, the coil's motion would track the mass exactly. It is this behavior of spring-mass systems that gave us the clue as to how to terminate the top portion of a spring at 30 cm so as to make the supported mass act as if it were suspended from a spring 1 km in length. If one were to grasp the coil 30 cm above the mass and then somehow continue to move it up and down in exactly the same way it was moving with a km of spring above it, one could then cut off the top km and still have a spring that would act if it were 1 km long. Since this involves causing a particular point on the spring to track the bottom, a servo approach can be used to transform an ordinary length of spring into a super spring.

Figure 1 is a schematic representation of the super spring. The system consists of a bracket supported by two springs which supply the dc force to support the bracket and the attached mass-spring system (or in terms of the previous discussion they supply the force that the balance of the 1 km of spring would have supplied). From this bracket, the main spring and mass are hung. The top of this spring is made to track the motion of its bottom utilizing a position detector which measures the length of the main spring with respect to the bracket. Light from a light emitting diode (LED) is focused by a ball (serving as a rotation insensitive lens) onto a split photodiode. The currents from the two halves of the diode are amplified, by a current controlled voltage source, and the resulting voltages are then differenced. This supplies an analogue voltage that is a measure of the position of the mass with respect to the bracket. This analogue voltage is passed through a servo compensating amplifier which drives a current through a voice coil centered in a loudspeaker magnet, which results in a force on the bracket. This forces the bracket (and thus the top of the main spring) to track the mass. The degree of tracking is simply related to the gain of the servo system. The higher the gain the better the tracking—i.e., the closer the coils track the motion of the bottom mass—and therefore the longer the system period.

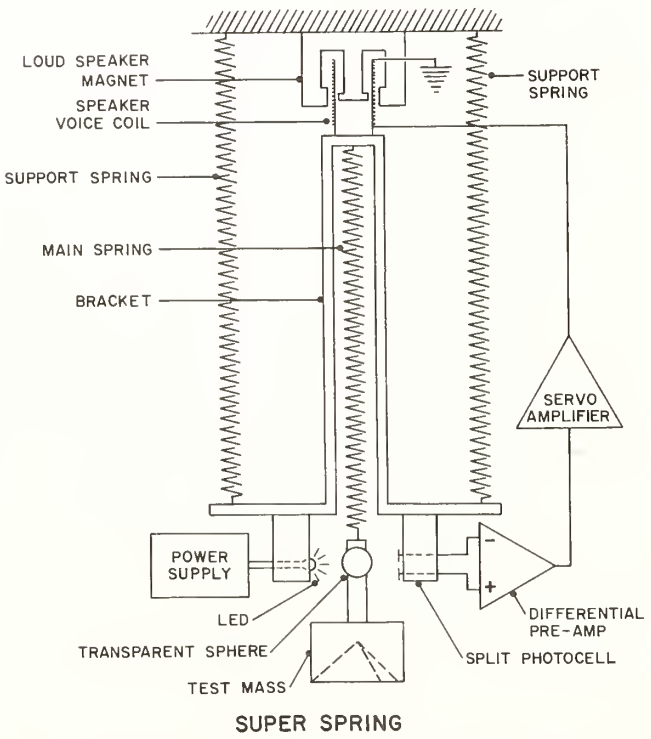


FIGURE 1. *Schematic diagram of the super spring.*

At this point a brief discussion of the isolation characteristics of a simple spring-mass system is in order.

2. Theory

Figure 2 shows two simple spring-mass circuits: (a) is called the absolute damping case, and (b) is the relative damping case. By writing the equations of motion for these systems and solving for $X_2(S)/X_1(S)$ (the system transfer function) we get for (a)

$$\frac{X_2(S)}{X_1(S)} = \frac{\omega_0^2}{S^2 + \frac{\beta}{m} S + \omega_0^2} \tag{2}$$

and for (b)

$$\frac{X_2(S)}{X_1(S)} = \frac{\omega_0^2 + \frac{\beta}{m} S}{S^2 + \frac{\beta}{m} S + \omega_0^2} \tag{3}$$

where S is the Laplace transform variable ($\sigma + j\omega$), β is the system's damping coefficient, k is the spring constant, m is the mass, and $\omega_0^2 = k/m$. The isolation characteristic is found by evaluating the magnitude of the transfer function for $S = j\omega$. For (a) we get

$$\left|\frac{X_2(j\omega)}{X_1(j\omega)}\right| = \frac{\omega_0^2}{\left[(\omega_0^2 - \omega^2)^2 + \frac{\beta^2}{m^2}\omega^2\right]^{1/2}} \tag{4}$$

and for (b)

$$\left|\frac{X_2(j\omega)}{X_1(j\omega)}\right| = \left[\frac{\omega_0^4 + \frac{\beta^2}{m^2}\omega^2}{(\omega_0^2 - \omega^2)^2 + \frac{\beta^2}{m^2}\omega^2}\right]^{1/2}. \tag{5}$$

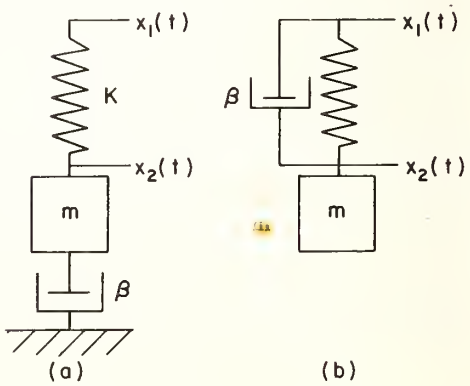


FIGURE 2. *Spring-mass circuits (a) absolute damping case; (b) relative damping case.*

Since spring-mass systems isolate only for frequencies ω that are greater than ω_0, the two expressions can be simplified for $\omega > \omega_0$. Examination of these asymptotic forms reveals a basic difference between the two systems. For (a), the "absolute" damping case, we get

$$\left|\frac{X_2(j\omega)}{X_1(j\omega)}\right| \sim \frac{1}{\omega^2} \quad \omega \gg \omega_0 ; \tag{6}$$

for (b), the "relative" damping case, we get

$$\left|\frac{X_2(j\omega)}{X_1(j\omega)}\right| \sim \frac{1}{\omega} \quad \omega \gg \omega_0 . \tag{7}$$

This difference in the asymptotic behavior leads one to the conclusion that, for an isolation device, the "absolute" damping case is preferable. The problem, however, is that the body of the dash pot has been assumed to be connected to an inertial reference. This being available, however, assumes that we have already solved the problem.

The simple spring-mass circuit that models the super spring is shown in Fig. 3 where K_1 is the spring constant of the support springs, β_1 is the damping coefficient of the support springs, m_1 is the mass of the bracket, K_2 is the spring constant of the main spring, β_2 is the damping coefficient of the main spring, m_2 is the mass of the test mass, and $f(t)$ is the force supplied by the voice coil and speaker magnet.

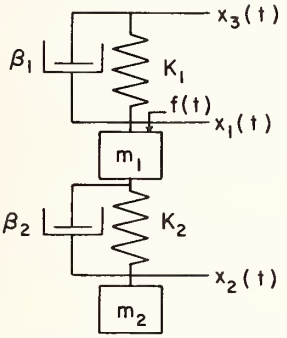


FIGURE 3. *Spring-mass circuit for super spring.*

The equations of motion are two coupled second-order linear differential equations:

$$m_1\ddot{x}_1 = K_1(x_3 - x_1) + \beta_1(\dot{x}_3 - \dot{x}_1) + K_2(x_2 - x_1) + \beta_2(\dot{x}_2 - \dot{x}_1) + f(t) , \tag{8}$$

$$m_2\ddot{x}_2 = K_2(x_1 - x_2) + \beta_2(\dot{x}_1 - \dot{x}_2) . \tag{9}$$

The key to the modification of the system's behavior is contained in the term $f(t)$. The exact form of $f(t)$ is determined by the transfer function of the electronics. For this discussion we will use a simplified form for $f(t)$:

$$f(t) = G(x_2 - x_1) + \gamma_S(\dot{x}_2 - \dot{x}_1) . \tag{10}$$

This is the statement that the servo action is proportional plus derivative. The most convenient method for the analysis of the system is to work in the Laplace transform domain.

Substituting Eq. (10) into Eq. (8) and then taking the Laplace transform of Eqs. (8) and (9) we get

$$m_1X_1(S)S^2 = (K_1 + \beta_1S)[X_3(S) - X_1(S)] + [K_2 + G + (\beta_2 + \beta_S)S] \times (X_2(S) - X_1(S)) , \tag{11}$$

$$m_2X_2(S)S^2 = (K_2 + \beta_2S)(X_1(S) - X_2(S)) , \tag{12}$$

where $X_1(S)$, $X_2(S)$, and $X_3(S)$ are the Laplace transforms of $x_1(t)$, $x_2(t)$, and $x_3(t)$, and S is the complex transform variable $(\sigma + j\omega)$. These two equations can be solved for the transfer function of interest:

$$\frac{X_2(S)}{X_3(S)} = \frac{(\gamma_1S + \omega_1^2)(\gamma_2S + \omega_2^2)}{S^4 + S^3(\gamma_1 + \gamma_2 + \gamma_{12} + \gamma_S) + S^2(\omega_1^2 + \omega_2^2 + \omega_{12}^2 + \omega_S^2) + S(\gamma_1\omega_2^2 + \gamma_2\omega_1^2) + \omega_1^2\omega_2^2} , \tag{13}$$

where $\gamma_1 = \beta_1/m_1$, $\gamma_2 = \beta_2/m_2$, $\gamma_{12} = \beta_2/m_1$, $\gamma_S = \beta_S/m_1$, $\omega_1^2 = K_1/m_1$, $\omega_2^2 = K_2/m_2$, $\omega_{12}^2 = K_2/m_1$, $\omega_S^2 = G/m_1$. For this particular system we can approximate some of these terms for high servo gain values:

$$\gamma_1 \ll \gamma_S \qquad \omega_1^2 \ll \omega_S^2$$
$$\gamma_2 \ll \gamma_S \quad \text{and} \quad \omega_2^2 \ll \omega_S^2$$
$$\gamma_{12} \ll \gamma_S \qquad \omega_{12}^2 \ll \omega_S^2$$

and Eq. (13) becomes

$$\frac{X_2(S)}{X_3(S)} = \frac{(\gamma_1S + \omega_1^2)(\gamma_2S + \omega_2^2)}{S^4 + S^3\gamma_S + S^2\omega_S^2 + S(\gamma_1\omega_2^2 + \gamma_2\omega_1^2) + \omega_1^2\omega_2^2} . \tag{14}$$

If we factor the quartic in the denominator into two quadratics of the form

$$\frac{X_2(S)}{X_3(S)} = \frac{(\gamma_1S + \omega_1^2)(\gamma_2S + \omega_2^2)}{(S^2 + AS + B)(S^2 + CS + D)} , \tag{15}$$

for large values of ω_S and γ_S we get

$$A = \gamma_S , \quad B = \omega_S^2 , \quad C = \frac{\gamma_1\omega_2^2 + \gamma_2\omega_1^2}{\omega_S^2} - \frac{\gamma_S\omega_1^2\omega_2^2}{\omega_S^4} ,$$

$$D = \frac{\omega_1^2\omega_2^2}{\omega_S^2} .$$

The quantity of interest is of course the isolation characteristic $|X_2(j\omega)/X_3(j\omega)|$. For the frequency range $\omega \ll$ (larger of ω_1/γ_1, ω_2/γ_2) we get

$$\left|\frac{X_2(j\omega)}{X_3(j\omega)}\right| = \frac{\omega_1^2\omega_2^2/\omega_S^2}{\left[\left(\frac{\omega_1^2\omega_2^2}{\omega_S^2} - \omega^2\right)^2 + C^2\omega^2\right]^{1/2}} . \tag{16}$$

By comparison to Eq. (4) we see this is exactly the isolation characteristic of a simple spring with resonance $\omega_0 = \omega_1\omega_2/\omega_S$ and damping constant C. This analysis shows that one can in fact lower the "corner" frequency of a mechanical system with feedback, without the dimensions of the apparatus getting out of proportion. There is, however, a difficulty with this technique. Recall that

$$C = \frac{1}{\omega_S^2}(\gamma_1\omega_2^2 + \gamma_2\omega_1^2) - \gamma_S\frac{\omega_1^2\omega_2^2}{\omega_S^4} . \tag{17}$$

Since we started with a high "Q" mechanical system, γ_1 and γ_2 are small, and since the servo gain is large the negative term in the expression for C can dominate. When this is the case the damping factor for the low frequency pole in Eq. (15) $[S^2 + CS + (\omega_1^2\omega_2^2/\omega_S^2)]$ is negative. This leads to a transient response that is an exponentially growing sinusoid with frequency ω_0, which is clearly not acceptable.

The obvious means of damping the long-period motion is to connect a dash pot from x_2 to x_3. In the development of super spring, two types of dash pots were tried. The first was a 4×10^{-3} cm tungsten rod immersed in a light silicone oil. While this damped the system's motion

quite well, the creep of the fluid up the rod increased the mass of the test mass to a degree that the system simply continued to move down until the rod hit the bottom of the oil container. Remember that even at dc the spring behaves as a very long one (and is therefore very soft): The addition of only a few milligrams of mass to the 500 g mass results in a substantial lowering of the equilibrium position.

To eliminate the creep problem associated with the fluid, a magnetic dash pot was tried. The test mass was placed in a magnetic field with the result that as the mass moved, eddy currents were induced giving rise to damping forces. This method, though quite successful for low servo gain settings, suffered as the servo gain was increased because the interaction of the fringing field of the magnet and the magnetic spring tended to add a negative spring constant, which either canceled or dominated the positive spring constant, making the system unstable.

The damping method we finally decided on modifies the low frequency ($\omega < 0.628$) response of the servo in such a way as to damp the long-period motion. In order to understand how this method works we will consider the simple spring-mass system shown in Fig. 4 where K is the spring constant of the spring, β is the damping coefficient of the system, m is the mass, and $f(t)$ is a time-dependent force which is derived electronically from the displacement $(x_1 - x_2)$.

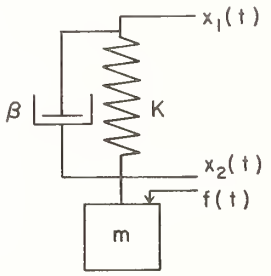


FIGURE 4. *Simplified spring-mass circuit for the damped super spring.*

In the Laplace domain the equation of motion can be written as

$$mS^2X_2(S) = K[X_1(S) - X_2(S)] + \beta S[X_1(S) - X_2(S)] + G(S)[X_1(S) - X_2(S)] , \quad (18)$$

where $G(S)$ is the transfer function of the electronics. Equation (18) can be solved for the transfer function $X_2(S)/X_1(S)$

$$\frac{X_2(S)}{X_1(S)} = \frac{\omega^2 + \frac{\beta}{m}S + \frac{1}{m}G(S)}{S^2 + \frac{\beta}{m}S + \omega^2 + \frac{1}{m}G(S)} ,$$

$$\omega = \left[\frac{K}{m}\right]^{1/2} . \quad (19)$$

From this transfer function we can see that the time domain behavior of the system depends on the mechanical parameters β and ω and also on the transfer function of the electronics $G(S)$. If we choose

$$G(S) = \gamma S , \quad (20)$$

which is equivalent to saying the force is proportional to the time derivatives of $(x_1 - x_2)$, then Eq. (19) becomes

$$\frac{X_2(S)}{X_1(S)} = \frac{\omega^2 + \frac{(\beta + \gamma)}{m}S}{S^2 + \frac{\beta + \gamma}{m}S + \omega^2} . \quad (21)$$

This is easily identified as the transfer function of a simple spring with a variable damping coefficient $(\gamma + \beta)$. If $\gamma < 0$, the "Q" of the system is increased. If $\gamma > 0$, the "Q" is decreased. The case $\gamma > 0$ is an example of using electronic feedback to damp a mechanical system.

While the above analysis demonstrates the concept of electronic damping, there is a practical limitation that is not taken into account in Eq. (20). The limitation comes in the fact that we are not free to choose $G(S)$ in a totally arbitrary manner. The constraints on our choice of $G(S)$ stem from the fact that $G(S)$ is the transfer function of an electronic system, which must be realized using amplifiers having a finite band width. We can at best achieve a transfer function $G(S)$ that acts as a differentiator for a restricted frequency range. The optimum $G(S)$ that can be realized is the transfer function of a second-order band pass filter

$$G(S) = \frac{\gamma S}{S^2 + AS + \omega_B^2} . \quad (22)$$

This is simply the transfer function of a perfect differentiator with two poles added. One pole must be added for reasons of op-amp stability, the second pole is added to reduce high frequency noise. Using this $G(S)$ in Eq. (19) we get

$$\frac{X_1(S)}{X_2(S)} = \frac{\left[S^2 + AS + \omega_B^2\right]\left[\frac{\beta}{m}S + \omega^2\right] + \frac{\gamma}{m}S}{\left[S^2 + \frac{\beta}{m}S + \omega^2\right]\left[S^2 + AS + \omega_B^2\right] + \frac{\gamma}{m}S} . \quad (23)$$

To understand how the response of this system is changed as we increase the value of γ, we utilized root locus analysis. Figure 5 is a rough sketch of the root locus of the transfer function Eq. (23) as a function of γ. As γ is increased we see that the mechanical poles move toward the σ axis indicating a decrease in the "Q" of

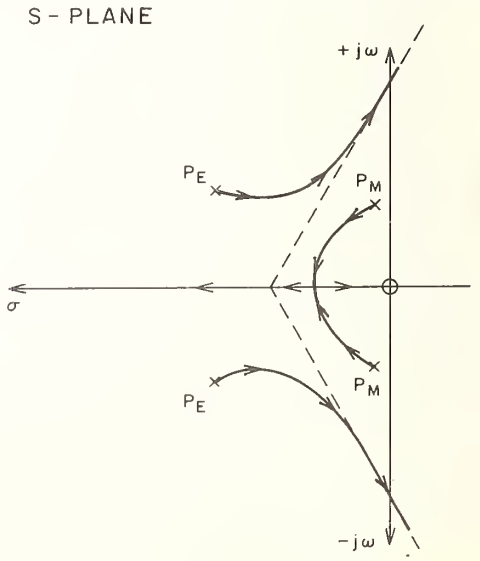


FIGURE 5. *Sketch of the root locus of the transfer function, Eq. (23), as a function of* γ.

this resonance. But as γ is increased, the electronic poles move toward the $j\omega$ axis, which indicates an increase in the "Q" of the electronic poles. If γ is increased enough the electronic poles will go into the right half plane, at which point the system becomes unstable. While these poles are electronic in nature, they will nevertheless cause the mechanical system to oscillate. The total effective damping one can achieve is determined by the tradeoff between the increase in the "Q" of the electronic poles and the decrease of the "Q" of the mechanical poles.

In the derivation of Eq. (13) we used for the feedback force $f(t)$

$$f(t) = G(x_2 - x_1) + \gamma_S(\dot{x}_2 - \dot{x}_1) \ . \tag{24}$$

This is equivalent to saying the transfer function of the servo is

$$G(S) = (G + \gamma_S S) \ . \tag{25}$$

This of course cannot be the transfer function of any real system, since all real systems have finite gain and band width. When we assume a general form for the transfer function $G(S)$ of the electronics, then Eq. (13) becomes

$$\frac{X_2(S)}{X_3(S)} = \frac{(\gamma_1 S + \omega_1^2)(\gamma_2 S + \omega_2^2)}{(S^2 + \gamma_1 + \omega_1^2)(S^2 + \gamma_2 S + \omega_2^2) + S^2 \frac{1}{m_1} G(S)} \ . \tag{26}$$

For the servo transfer function used in the super spring, which takes into account the finite band width of the servo amplifiers and the low frequency differentiator (used to damp the long period motion), $G(S)$ has the form

$$G(S) = G \left[\frac{\omega_0^2 + \alpha 2\eta_0\omega_0 S}{S^2 + 2\eta_0\omega_0 S + \omega_0^2} + \rho \frac{2\eta_L \omega_L S}{S^2 + 2\eta_L \omega_L S + \omega_L^2} \right] . \tag{27}$$

Here ω_0, η_0 are the band width and damping ratio for the servo amp, α is the gain of the second order bandpass filter used as the differentiator to damp the servo, and ω_L, η_L and ρ are the break point, damping ratio, and gain of the second-order bandpass filter used as the low frequency differentiator which damps the long period motion.

The complete analysis of Eq. (26) with this $G(S)$ involves finding the roots of, and eighth-order polynomial for, various values of G, α, ρ, η_0, ω_0, η_L and ω_L. This analysis has been done using a digital computer and a FORTRAN program. The results were in no way profound, so they will not be discussed at length. It is sufficient to say that the optimum values for ω_0, ω_L, α, and ρ were found to depend on the value of G. This value of course is determined by the desired period of the "long spring." Since we wanted this to be an easily variable parameter the electronics was designed so that ω_0, ω_α, α, and ρ were also easily varied.

3. Mechanical

There are three basic parts to the mechanical system; the bracket assembly, the flexures which constrain the bracket to have only a vertical degree of freedom, and the main and support springs.

The design requirements on the bracket are that its mass must be kept to a minimum while its stiffness be as high as possible. The low mass requirement results from the fact that the "gain" in Eq. (23) of the electronic transfer function $G(S)$ is divided by the mass of the bracket m. Stiffness minimizes the tendency at high servo gain of exciting the mechanical modes of the bracket. One can also address this problem by applying damping material (e.g., Dux Seal) to lower the "Q" of these modes and thus making them harder to excite. The bracket (see Fig. 6) consists basically of a long tube which is centered in the external housing by means of a flexure at each end. These flexures constrain the bracket to move in only the vertical direction. The housing containing the position detection system, which consists of a light emitting diode and a split photodiode, is attached to the bottom of the tube. The voice coil is attached at the bottom of this housing. The main spring hangs down through the center of the tube which results in the test mass being centered in the position detection housing. The weight of this assembly (tube, housing, voice coil, main spring, and test mass) is carried by the three auxiliary support springs. The voice coil fits into the gap of the speaker magnet—the latter being attached to the external housing. The external housing rests on three leveling screws.

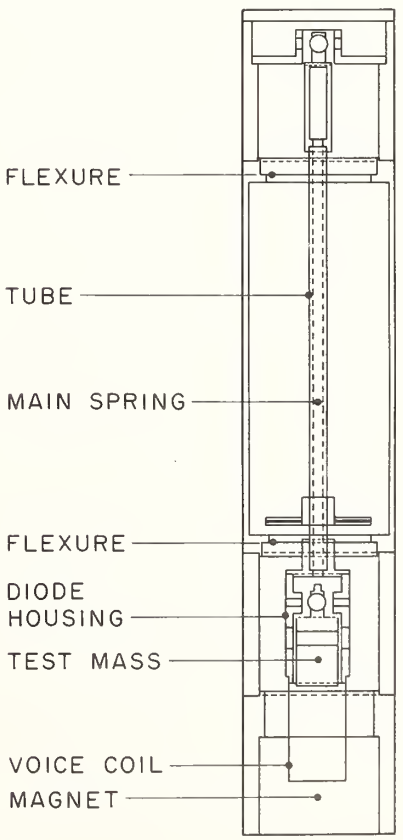


FIGURE 6. *Schematic diagram of the overall super spring system.*

The flexures are one of the most critical parts of the mechanical system. They must be soft to vertical motion, still to sideways motion, and precise enough so as not to cause the tube to tilt as it moves up and down. Several flexure designs were tried. Figure 7 shows the design that ultimately worked the best. They are first machined from 17.5×10^{-3} cm thick, 10.8 cm diameter disks of beryllium copper, and then heat treated. These flexures give about 2 arcseconds of tilt for a bracket motion of 0.6 cm.

FIGURE 7. *Photograph of one of the flexures.*

Since the "long" spring acts exactly like a spring 1 km long, minimizing the temperature coefficient of the spring material is very important. The effect of a change in temperature comes in chiefly through the temperature dependence of Young's modulus rather than through the thermal expansion of the material. Springs made from beryllium copper or fused silica have Young's modulus temperature coefficients on the order of 10^{-3}. Our springs are made from Ni-Span-C alloy 902 which is manufactured by Huntington Alloys. Ni-Span-C has the unique property that cold working drives the thermal elastic coefficient negative while heat treating drives it positive. With proper treating of springs made from Ni-Span-C one can get a very low coefficient. The springs we use have a thermoelastic coefficient of about 6 ppm.

Light from a light emitting diode (LED—Texas Inst. TTES16 series) is passed through a cylindrical lens. This results in a rectangular image. This image is focused by the transparent sphere attached to the test mass (see Fig. 1) onto a split photodiode (UTD Pinspot 2DJ). As a result of this geometry the amount of light on the two halves of the diode changes when the test mass moves up or down. This changes the amount of current from the diode halves. These currents, amplified by a current-to-voltage converter, are differenced to provide an analogue signal proportional to the displacement of the test mass. This signal, processed by a servo-compensating amplifier, drives a loudspeaker voice coil which in turn supplies the needed force to cause it to (nearly) track the motion of the test mass. With an LED current of 200 mA the vertical position sensitivity of this detector is 0.7 V μm^{-1}. The noise level in a 5.0 kHz bandwidth is 5 mV P-P while the noise level in a 1 Hz bandwidth is about 0.5 mV P-P.

4. Test Results

Once the construction of super spring was completed we tested its isolation characteristics. To do this, we constructed a "shake table" on which we could place the spring. This table could produce an amplitude of 25×10^{-3} cm at frequencies of less than 5 Hz, and about 2×10^{-3} cm at 25 Hz. The table surface, driven by a system of levers and a speaker magnet-voice coil system, is constrained so as to tilt less than one arcsecond—the accomplishment of which proved to be a nontrivial task. The power amplifier used to drive the coil could deliver ±12 A into the coil. An LED photodiode position detector was used to monitor the table motion. The isolation measurements were made using a HP 3582A Spectrum Analyzer. The noise source from the 3582A was used to drive the table. The output from the table's position detector and the position of the test mass with respect to the floor ("inertial space") were applied to the two inputs of the spectrum analyzer which computed the transfer function. Figure 8 is an example of such a transfer function. This particular one was taken with the spring period set at 12 s (lower than usual) so that the 1/12 Hz resonance could be seen.

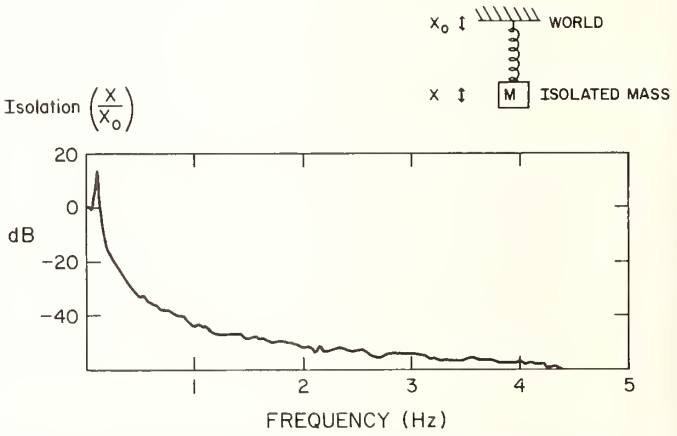


FIGURE 8. *Transfer function obtained during "shake table" testing of super spring.*

Further evidence that the spring indeed isolates is seen when the test mass is used to hold the reference cube corner in a free falling absolute gravimeter (Zumberge *et al.*, in these proceedings). Figure 9 shows two histograms of 150 g measurements (one set with the inertial reference and one without) which demonstrate a reduction in scatter by a factor of 20.

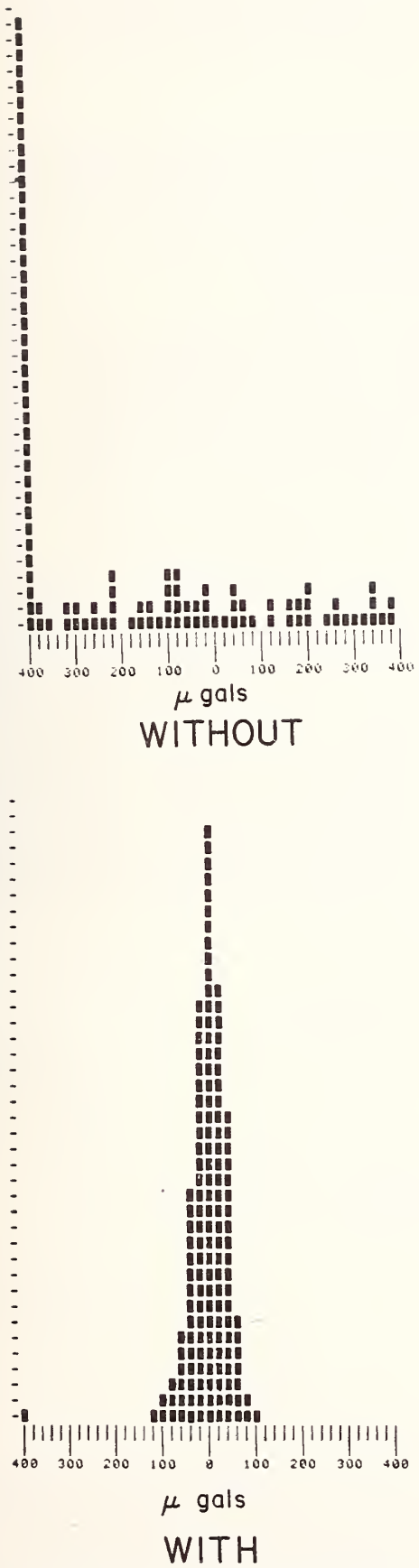


FIGURE 9. *Histogram of measurements of the gravitational acceleration g without and with a super spring isolated reference corner cube.*

5. Conclusions

Toward the end of the same article quoted in part at the beginning of this paper, C. V. Boys states: "I would strongly urge that in such a case, a room more uniform in temperature than the one at Oxford should be employed. . . I do not think any ready-made room is likely to be found available. A disused adit, at a great distance from existing mining operations, would be perfect." The use of the super spring concept as we have applied it to isolate in the vertical, or as it could be adapted to isolate in other directions, offers an alternative to Boys' recommendation. Clearly the ideal situation would be to both isolate and operate at a quiet site. In many cases this simply is not practicable. However, in all cases the principles of the super spring can still be used to effectively get new labs for old.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Transportable Gravimeter for the Absolute Determination of Gravity

Guo You-guang, Huang Da-lun, Li De-xi, Zhang Guang-yuan, Gao Jing-long, Fang Yong-yuan, and Huang Cheng-qing

National Institute of Metrology, Beijing, P.R.C.

At the National Institute of Metrology in Beijing, a transportable gravimeter using the method of free fall has been constructed. The instrument consists of an optical interferometer illuminated by light from a stabilized He-Ne laser, in which one of the mirrors, a corner-cube reflector, falls freely. The time standard is obtained from a highly stabilized rubidium clock. The methods of time and distance measurement are described. The effect of the verticality or collimation is discussed. A positive correction has to be included. Some recent (1979–1980) results are presented. The accuracy achieved with this apparatus is about two parts in 10^8.

Key words: acceleration due to gravity; coincident method; transportable gravimeter.

1. Introduction

High precision absolute determinations of gravity contribute to our knowledge about the earth's gravity field and its secular variation. It is of considerable significance for research on earthquake prediction and is important in geophysics and geodesy [1].

At the end of 1977 we developed a transportable gravimeter. It is based on the original system [2] but with improved electronic, vacuum, and mechanical components. The apparatus consists of a Michelson interferometer illuminated by light from a stabilized He-Ne laser locked to the Lamb-dip. A diagram of the optical system utilized is illustrated in Fig. 1. The apparatus cycles automatically once started and makes 50 drops per hour.

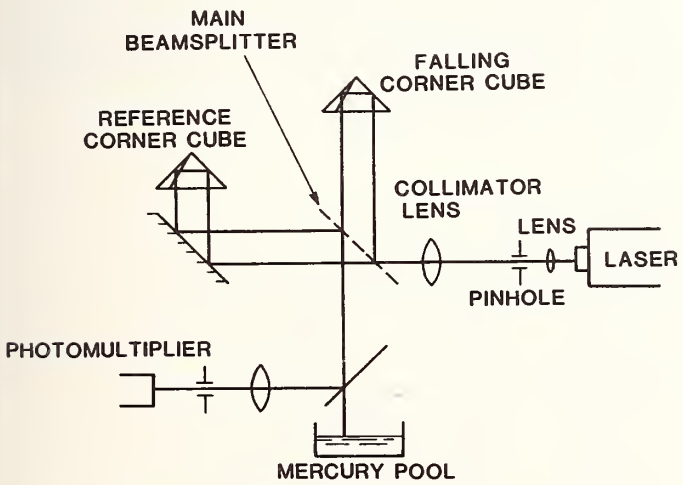


FIGURE 1. *Diagram of the optical system.*

2. Measurement of Distance and Time

The principle of simple free-fall has been adopted. The initial velocity and position of a freely falling object are difficult to determine precisely in this method. In order to avoid the measurement of initial velocity, two successive times of passage of the object, T_1 and T_2, must be measured through three horizontal stations h_1, h_2, and h_3, for which the vertical separations H_1 and H_2 are known. The value of g is obtained as follows [3, 4]:

$$g = \frac{2}{T_2 - T_1}\left(H_2/T_2 - H_1/T_1\right). \qquad (1)$$

The two time intervals T_1 and T_2, and the distances H_1 and H_2, are precisely measured by a coincidence method, which avoids the problem of measuring fractional clock cycles or interferometer fringes. A special electronic system is designed for this purpose. Figure 2 shows a block diagram of the system.

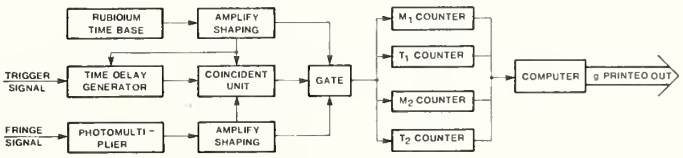


FIGURE 2. *Block diagram of the electronic system.*

The width of input pulses in our instrument is 4 ns. The selected times for the three horizontal positions h_1, h_2, and h_3 are about 0.09 s, 0.24 s, and 0.39 s after release of the falling object at the top of the chamber. The determination of accurate time intervals depends on coincidences between the pulses of the interference fringes and the signals of the frequency standard. Theoretical calculations and computer models have shown that the probability of such coincidences is sufficient; our experiments have corroborated these calculations.

The measured distance H is directly dependent on the accuracy of the laser's wavelength. Our laser wavelength is periodically calibrated in terms of the NIM iodine stabilized He-Ne laser. A relative accuracy for the wavelength of 10^{-9} is thus reached. The time standard is a rubidium frequency standard made in China whose accuracy is better than 1×10^{-10}.

3. Effect of Verticality or Collimation

If the laser beam is at an angle, the distance measurement will be in error. The effect is to lengthen the distance corresponding to one fringe and to make the measured value of g lower than the true value. This can be shown by the variation in the actual optical path when

the corner-cube prism moves. It can be deduced directly from the following viewpoint. Only the motion of the optical center of the corner-cube in the direction of the light beam will produce a variation of the optical path difference. We assume that when an object falls freely in the plane M, which consists of the vertical line and the main light beam passing the optical center A of the corner-cube (Fig. 3), there will be a small angle α between the main light beam and the vertical line. The projection of the trajectory of the optical center of the falling object on the plane M is the curve AB, and the error between the vertical distance AC (h) travelled during time T and the distance AB′ measured by the interferometer is Δh. Thus the corresponding error in g is

$$\Delta g = \frac{d^2(\Delta h)}{dt^2} \simeq -\frac{1}{2}\alpha^2 \frac{d^2h}{dt^2} - \alpha \frac{d^2d}{dt^2} = -\frac{1}{2}\alpha^2 g \qquad (2)$$

or

$$\frac{\Delta g}{g} = -\frac{1}{2}\alpha^2. \qquad (3)$$

From this we see that the error is independent of the translation value d, as the translation acceleration is always very small in a free fall experiment.

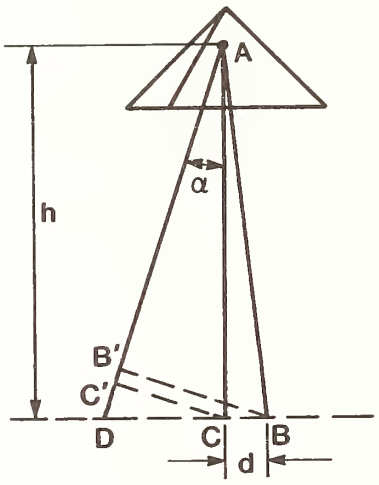


FIGURE 3. *Diagram used to analyze the effect of a nonvertical laser beam.*

An experiment was carried out to verify the effect of this error. By making the light beam deviate from the vertical by the small angle 147 seconds of arc, the difference between the correct value and the experimental result is found to be

$$\Delta g = -0.29 \pm 0.03 \text{ mGal},$$

which agrees with the value calculated by the above formula:

$$\Delta g \simeq -0.25 \text{ mGal}.$$

Therefore, based on a careful measurement of the angles, we must include a positive correction.

4. The Results

Table 1 presents a compendium of results obtained with the transportable instrument during the last two years.

Table 2 is a summary of the corrections and estimates of uncertainty for the various possible systematic effects, both natural and instrumental.

Figure 4 is a typical histogram of the data obtained at a quiet site. It shows the results of 51 drops taken on 26 October 1979 at Shiang Shan laboratory which is located in the suburbs of Beijing. There, the microseismic-noise level is very low. The standard deviation of a set of measurements, which is 18 drops, is usually about 70 μGal.

Figure 5 shows the results of five sets of data taken in the time span 19–20 September 1979 at Shiang Shan Laboratory. Each point represent 17–26 drops with the value obtained by averaging. Each data point corresponds to a 1/2 hour time span. The results of the five sets of data are compared with the computed earth tide adjusted for a best fit.

In April 1980 we determined the gravitational acceleration at Sèvres A3, BIPM. The result obtained with this transportable gravimeter was 901 925 914 ± 16 μGal. At the same time, the value given by A. Sakuma was 980 925 908 ± 8 μGal. The difference between the two results is within their assigned measurement uncertainties.

TABLE 1. *Summary of data obtained with this NIM transportable gravimeter (all values in μGal)*

Site	Date	Number of data sets	Standard error	Net correction	Estimated error	Final value
NIM-1 (Beijing)	June-Sept. 1979	29	12	−299	22	980 125 837 ± 25
NIM-1 (Beijing)	January 1980	5	20	−157	14	980 125 830 ± 24
Sèvres A3 (Paris)	April 1980	29	10	−360	13	980 925 914 ± 16

TABLE 2. *Corrections and systematic effects and their estimated uncertainty (for NIM measurements at the BIPM site Sèvres A3)*

	Source	Correction (μGal)	Estimated uncertainty (μGal)
1.	Laser wavelength	0	4
2.	Electronic circuit	−22	5
3.	Electrostatic and magnetic	0	8
4.	Air resistance	0	8
5.	Velocity of light	+25	0
6.	Gravitational gradient	−363	3
	Net correction	−360	13

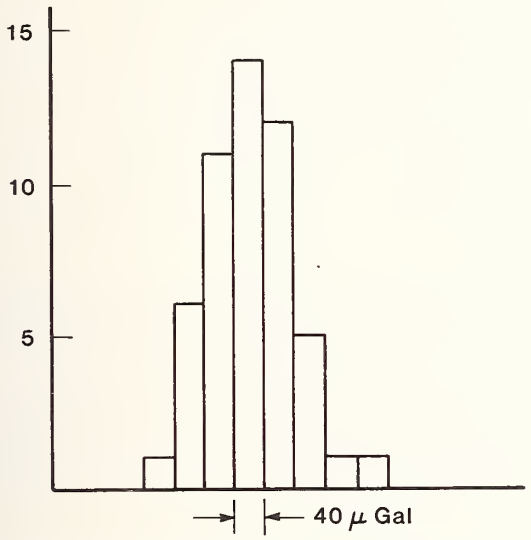


FIGURE 4. *Histogram of data from 51 drops. The standard deviation is 61 μGal.*

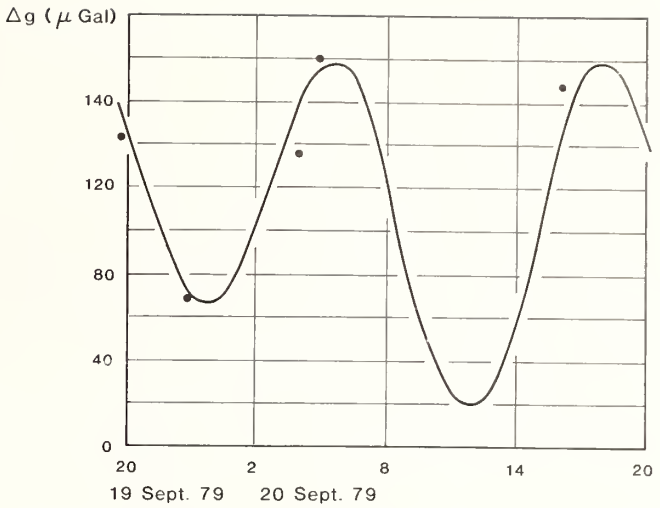


FIGURE 5. *Results obtained at Shiang Shan Laboratory. The points are measured g values, the curve the computed earth tide.*

References

[1] Fang Chun, *Measurement of Gravity and the Geography of the Earth* (in Chinese), 1965.

[2] Fang Yong-yuan *et al.*, Physics (in Chinese) 8, 53 (1979).

[3] A. Sakuma, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.) Spec. Pub 343 (Aug. 1971), p. 447.

[4] J. A. Hammond and J. E. Faller, ibid., p. 457.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

New Techniques for Absolute Gravity Measurement

James A. Hammond, Robert L. Iliff, and Roger W. Sands

Air Force Geophysics Laboratory, Terrestrial Sciences Division, Hanscom AFB, MA 01731

In the 11 years since PMFC-1 a number of new techniques have been put into practice in the Air Force Geophysics Laboratory's transportable system for measuring the acceleration of gravity. The improved system in use at the present time incorporates an earlier vacuum chamber with some modifications and includes new electronics, data analysis, and optical subsystems. The electronics now produce time measurement at a large number (500) points during the free fall of the object. These time values are analyzed with a least-squares fit to a second-order polynomial to obtain the average acceleration. The correction for air resistance is now made by monitoring the pressure and making a correction based on extrapolation from high pressures to the low operating pressures. The use of an iodine-stabilized laser as a reference for the length measurement has significantly reduced the uncertainty due to the wavelength of the Lamb-dip stabilized laser.

Key words: absolute gravity; acceleration of gravity; gravity; measurement of gravity; precision measurement.

The Air Force Geophysics Laboratory (AFGL) has long been involved in the development and support of instruments for the absolute measurement of gravity. Until 1975 we were known as the Air Force Cambridge Research Laboratories (AFCRL) and as such supported the development of the first transportable absolute gravity system (see Fig. 1) which contributed greatly to the establishment of a new international gravity standardization network. This instrument and the results obtained with it have been discussed in the literature [1–4].

Since that time AFCRL supported the development of a second generation instrument which used a chamber-in-a-chamber technique. This system produced some reasonably good data [5, 6], but several systematic effects presented problems which encouraged us to backtrack somewhat. We decided, then, to modify the first generation system to use the same electronics as were developed for the second generation system. Thus, at the beginning of 1978 we started assembling the system which exists at this time. Most of the "new techniques" which we will discuss in this paper are the result of an innovation first put into operation with the second generation system. The electronics system developed by JILA (Joint Institute for Laboratory Astrophysics), with AFCRL support, performed many time measurements during the fall of the object, and these time measurements were used to compute g. This is in contrast to the three-position technique used in the first generation system.

When the decision to use the mechanical parts of the first generation system was made, we decided to reduce the height of the chamber by about 0.5 m to simplify the transportation and operation of the system. The first data obtained with this new system have been described earlier [7]. There was reasonably good agreement between the new measurement and previous measurements at the AFGL site.

The instrument, shown in Fig. 2, was used in 1979 for a series of measurements along a calibration line from Great Falls, Montana to Ft. Davis, Texas. Data with an estimated accuracy of 0.010 mGal (one mGal = 10^{-5} m/s^2) were obtained at most sites, and the operation time required for seven sites was 28 days. At this time and until early in 1980 our timing electronics made a direct measurement of the time of occurrence of 350 fringes during the drop. Then a least squares fit of the time and distance values to the second order equation describing the distance as a function of time was made. The resulting g value was the measurement of the drop. The data were acquired with an on-line computer, and a g value was calculated for each drop. Data were averaged together for up to 24 hours of total operating time at each site. The appropriate corrections were made to arrive at the final value.

This measurement technique enables a unique method to determine the value of the corrections required because of two phenomena which are important in this kind of measurement. These two phenomena are the vertical gradient of gravity and the finite value of the velocity of light.

Both effects produce errors in the measured value when the calculation assumes uniform acceleration. Since the least-squares method is an averaging process, the result is some kind of average value of gravity. A spatial gradient becomes a time-dependent value of g because the falling body is moving in space and the least-squares process will obtain this time-averaged value. If the variation in g is small compared with its magnitude, the effective point at which the acceleration is measured is then the time-averaged position of the falling body. That point can be determined analytically in a straightforward manner by computing a relatively simple integral.

Alternatively, we have used an essentially numerical method to determine this effective measuring point. We calculate a set of synthesized times using the formula:

$$t = (2X/g_o)^{1/2} - (5/12)A\ (2X/g_o)^{3/2}, \qquad (1)$$

where A is the gravity gradient, X is the distance the object has fallen, and g_o is the value of gravity at the top of the chamber. This expression may be derived by solving the differential equation describing the motion of a body

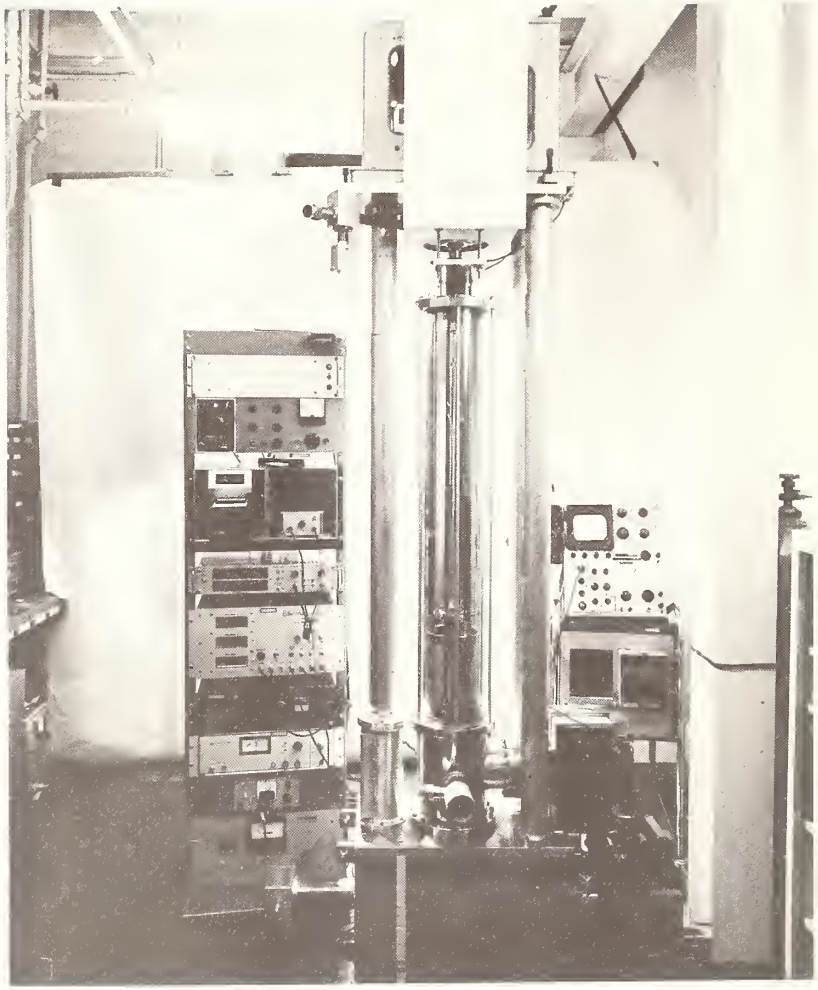


FIGURE 1. *AFCRL Absolute Gravity System—1968.*

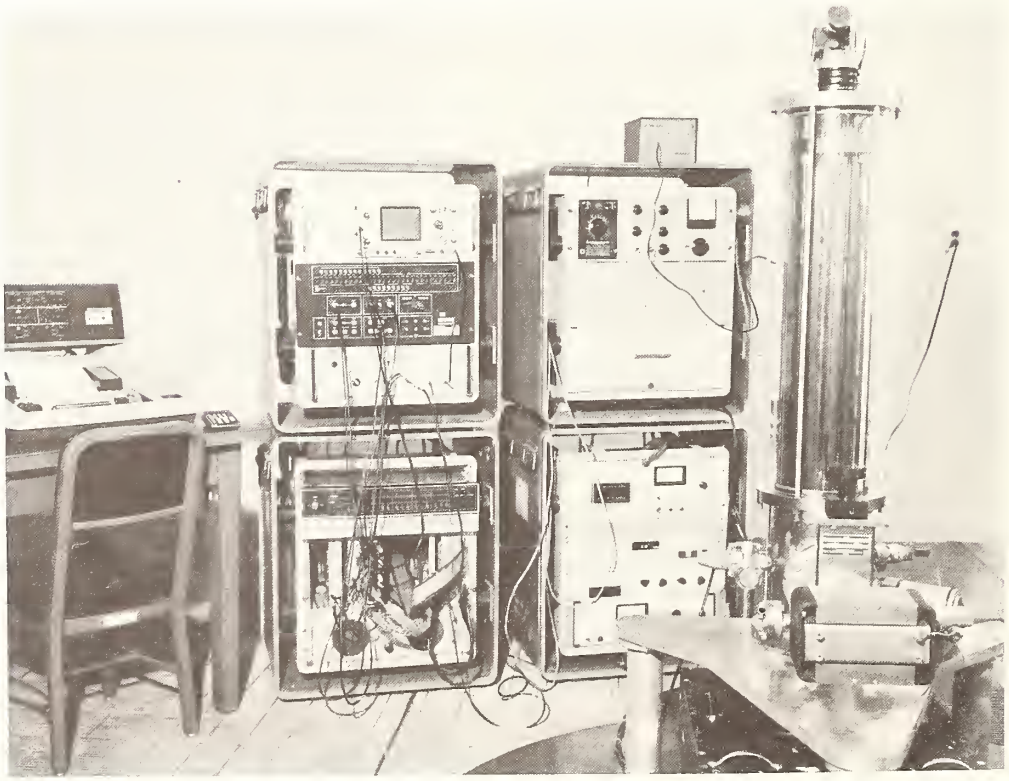


FIGURE 2. *AFGL Absolute Gravity System—1980.*

in a gradient field and expanding the result assuming the gradient is small compared with g. The result of fitting these time values is to produce a measured value of g_m which is different from g_o. That difference can be used to calculate the effective measuring point X_m:

$$X_m = (g_m - g_o)/A \qquad (2)$$

In our case (0.075 m fall before the 0.575 m measurement interval begins) the result is an effective measurement position of about 0.32 m from the top of the chamber.

The effect of the velocity of light being finite is to require corrections to the time values which depend upon where the falling body is in the chamber. A set of corrected time values can be calculated and then analyzed by the least-squares technique and the effect on the measured g value thereby determined. The result of such an exercise is a correction of 0.024 mGal.

The residuals obtained from our least squares fit would be expected to show effects which could indicate how well the system is operating. The residuals should show the face that the gradient is not included in the curve fitting. Thus, in plots of residuals we show a more or less smooth line which results from the fit of the synthetic data described above with an assumed gradient of 0.3 mGal/m. The actual residuals should more or less track the smooth curve as shown in Fig. 3. The extent that it

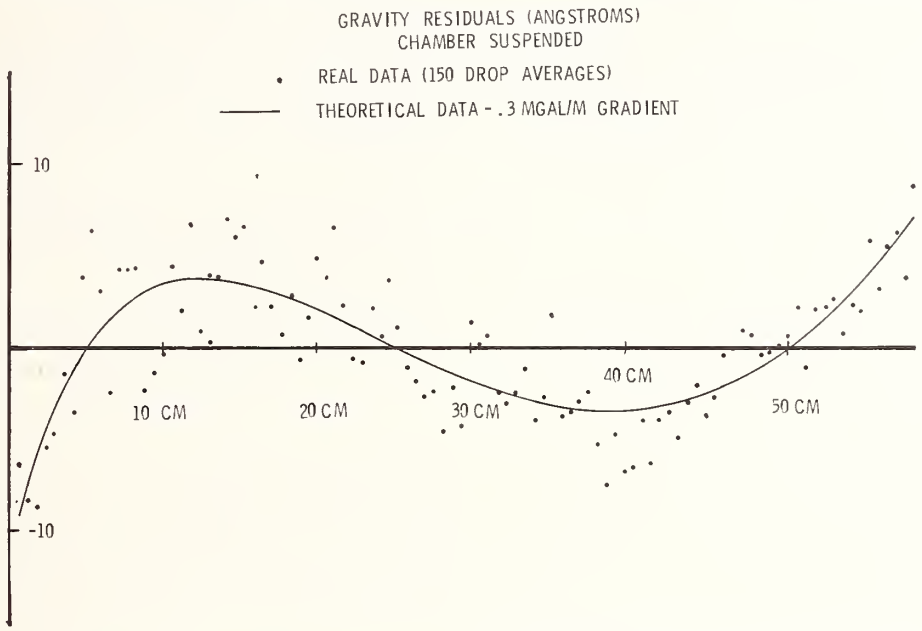


FIGURE 3. *Residuals in position for least squares fit of real and synthesized data.*

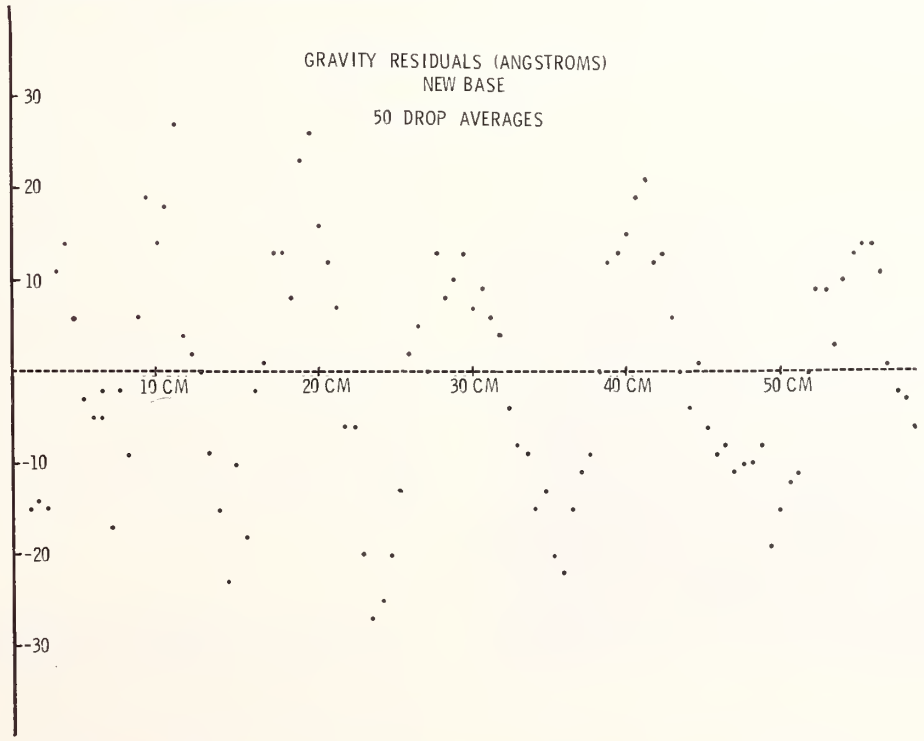


FIGURE 4. *Residuals showing systematic vibration of reference reflector.*

does not indicates vibrations or other effects that are influencing the measurement. The earliest such data showed a strong vibration (Fig. 4) which was removed by suspending the dropping chamber above the optics box. More recently such graphs show vibrations of 60 and 120 Hz, the exact source of which has not been determined.

Other corrections must be made for the effect of air resistance on the falling body and for the difference between the assumed (for the purposes of computation) wavelength of the stabilized laser light used as the length standard and its actual value. Air resistance corrections are determined periodically by taking data in the pressure range of about 2×10^{-7} to about 5×10^{-4} Torr and extrapolating the variations to zero pressure. The resulting error at 2×10^{-7} Torr appears to be only a few microgals, and thus the linearity is not checked closely. We have been periodically checking the wavelength of our laser against iodine-stabilized lasers built by the National Bureau of Standards. We have been continuously using the same Lamb-dip stabilized laser for almost 3 years and we have observed the rather typical drift of the wavelength of about one part in 10^8 per year.

A considerable number of absolute measurements of gravity have been made over the last few years and (up until about a year ago) they have generally corroborated previous measurements. Recently, some systematic effects have perturbed the results and produced some rather disappointing comparisons with previous measurements. It is hoped that the sources of these problems can be discovered and that the AFGL system will continue to be significant in establishing and maintaining gravity standards for precision measurements, geodesy, and geophysics.

References

[1] J. A. Hammond and J. E. Faller, IEEE J. Quant. Electron. **QE-3**, 597 (1967).

[2] J. A. Hammond, A Laser Interferometer System for the Absolute Determination of the Acceleration of Gravity, JILA Report No. 103, Joint Institute for Laboratory Astrophysics, Boulder, CO (1970).

[3] J. A. Hammond and J. E. Faller, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.), Spec. Publ. 343 (Aug. 1971), p. 457.

[4] J. A. Hammond and J. E. Faller, J. Geophys. Res. **76**, 7850 (1971).

[5] J. E. Faller and J. A. Hammond, A New Portable Absolute Gravity Instrument, Bulletin D'Information, International Gravimetric Bureau, No. 35 (November 1974), p. I-43.

[6] J. A. Hammond, Bollettino Di Geofisica Teorica ed Applicata Vol. **XX** (December 1978).

[7] J. A. Hammond and R. L. Iliff, The AFGL Absolute Gravity System, Proceedings of the 9th GEOP Conference, October 2–5, 1978, Dept. of Geodetic Science Report No. 280, The Ohio State University, Columbus, Ohio (1979).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Mass Unit "Kilogram," Precision Measurement of Mass, Attainable Uncertainties, and Possibilities of a New Definition

M. Kochsiek

Physikalisch-Technische Bundesanstalt, 3300 Braunschweig, F.R.G.

In this review, the present state of the dissemination of the unit of mass and of the uncertainty attainable in mass determination is described. The establishment of a mass scale (hierarchy of mass standards, multiples and submultiples), requirements for mass standards, important designs of weighing machines, and the predominant influencing parameters such as air density, in mass determination are discussed. Lately it has become possible to determine air density with smaller uncertainty by way of calculation and experiment. Possibilities are shown for a new definition on the basis of fundamental constants.

Key words: dissemination of mass unit; influences and uncertainties of mass measurement; mass standards; possibilities of new definition; prototype balance.

1. Introduction

In physics mass determinations are carried out over a large range following various measuring procedures. Within the framework of this Conference weighings only are of interest and these are carried out mainly in the range from 10^{-13} kg to 10^{6} kg. Weighing means using a measuring device, or balance, to compare an unknown mass with the base-unit "Kilogram," represented by mass standards or weights.

In the effort to base the definitions of the base units of the SI system on fundamental constants, that is, on constants occurring in nature, the values of which are ideally independent of the experimental arrangement, the unit of mass is an exception. According to available historical data, to this day masses have been compared using balances with a particular mass defined as the primary standard. The linkage of the mass unit to a fundamental constant has not as yet been achieved with the required precision. In this sense the history of the unit of mass has at no time been spectacular. Over the last hundred years, however, there has been a notable development towards better precision (stable primary standard, better balances, improved weighing procedures), so that the smallest relative uncertainties achievable today for mass comparisons are of the order of 10^{-8} to 10^{-9}, thus being worthy of belonging to the best in metrology. This is all the more astonishing as the physical phenomenon "mass" is still not yet completely understood. So mass metrology has developed pragmatically without full insight into the theoretical problems associated with mass [1, 2].

2. The Unit of Mass and the Hierarchy of Mass Standards

The unit of mass is the kilogram, which was defined by the first General Conference of Weights and Measures in 1889 as the mass of the International Kilogram Prototype, maintained at BIPM (Bureau International des Poids et Mesures) in Sèvres, Paris:

> The International Kilogram Prototype is a cylinder 39 mm in height and 39 mm in diameter of a 90% platinum with 10% iridium alloy (Pt-Ir); its density is about 21,500 kg·m^{-3}.

This means that the unit of mass can never be transferred with a precision better than permitted by the mass comparison with the International Prototype at BIPM. The consequence is a hierarchy of mass standards to guarantee the dissemination of the unit of mass with the highest possible accuracy. Since 1889 more than 60 platinum-iridium cylinders of the same dimensions and quality have been distributed as national prototypes, Fig. 1.

FIGURE 1. *National Kilogram Prototype No. 52 at the PTB, FRG, kept under two bell-jars. In the background are further primary and reference standards.*

At the top of the hierarchical ladder for the dissemination of the unit of mass stands the International Kilogram Prototype at the BIPM, Fig. 2. Using a prototype balance, the national prototypes are compared with primary standards of the BIPM which are themselves compared with the International Prototype. Thus the International Prototype only had to be used for comparisons in the years 1889, 1939 and 1946 and has, therefore, been nearly protected against wear and tear and possible damage.

Most national laboratories disseminate the unit with the aid of primary standards, which are made of stainless steel (density around 8000 $kg \cdot m^{-3}$). The comparison of these primary standards with the national prototype is the most significant from the point of view of measuring technique since for the necessary transition from density 21,500 $kg \cdot m^{-3}$ (Pt-Ir) to about 8000 $kg \cdot m^{-3}$ (Fe), the present uncertainty of the air buoyancy correction is larger than the uncertainty of the balance and of the other influencing parameters. The reference standards of the verification boards are then compared with the primary standards, as well as the reference standards of firms and other institutions which can be assigned to the best classes of mass standards of International OIML-Recommendation No. 20.

The prototypes, and primary and reference mass standards are standards of the highest precision; handling means taking the risk each time of a change in the mass (wear and tear, contamination) and of possible damage. Hence the time intervals between comparisons should, on the one hand, be as long as possible, but nevertheless short enough to allow timely recognition of mass variations. The right-hand column of Fig. 2 gives some indication of the time interval between two comparison measurements as well as the measurement site. Since working

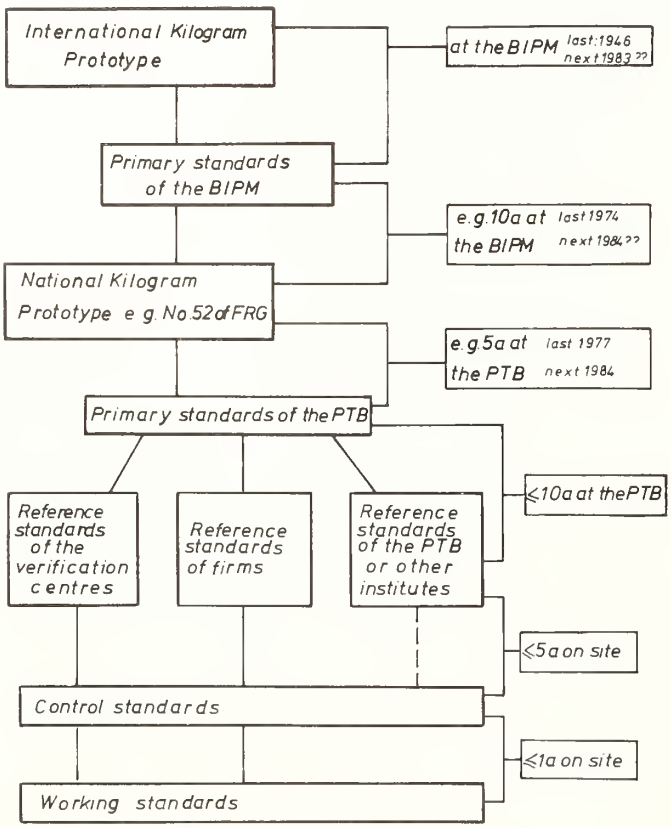


FIGURE 2. *Example from PTB, FRG, of a hierarchy of mass standards.*

standards and verification standards can be replaced more easily, they should be changed depending upon their frequency of use [3].

The next step—starting from a 1-kg standard mass—is to realize submultiples and multiples of the unit in the form of mass standards in such a way that the mass of any arbitrary body can be determined. For example 150.012 g using standards of 100 g, 50 g, 10 mg, and 2 mg. This is done by realizing the values 1 to 10 in every decade, which necessitates at least four standards for each decade: For example, the subdivision 1,2,2,5 is commonly used. That is, for the decade from 100 g to 1 kg the standards have nominal values 100 g, 200 g (two weights), and 500 g. Of these weights in the 100 g to 1 kg decade, two suitably chosen groups of mass standards of equal nominal sum value are all compared with each other. In this way the well-known overdetermined system of linear equations is obtained. Many schemes for the intercomparison of sets of weights are available [4–7, 44].

Since an optimum scheme (few weights with small uncertainty) depends on the balances available, the subdivision of the mass standards, the correlation between weighings, and the effort put into weighing, many schemes have been developed. Therefore an optimum scheme cannot be given.

Figure 3 shows a simple scheme. The 1-kg standard is known; one 500 g, two 200 g, and two 100 g standards are unknown. For five unknown standards seven determinations are carried out. According to the scheme on one side we have the 1-kg (positive sign), and on the other side 500 g + 2 × 200 g + 100 g (also of sum 1 kg with negative sign). For the seven series of weighings we obtain observations x(1) to x(7), from which the unknown masses may be calculated.

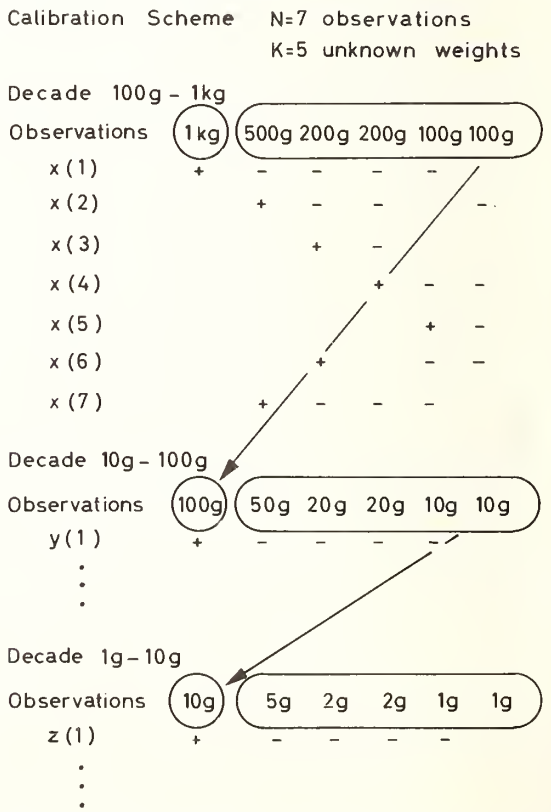


FIGURE 3. *Simple calibration scheme with 5 unknown masses (K) and 7 weighings (N).*

In the next decade from 10 g to 100 g the 100 g standard is known and the 50 g, 20 g, and 10 g standards are unknown. In this way every decade, and finally every set of standards, can be consistently determined.

3. Mass Standards

Mass standards must guarantee long-term stability of their mass. This can be achieved by the choice of corrosion resistant, non-magnetic material, suitable geometry, and specified procedures for handling, maintaining, and cleaning the mass standards [8–10]. Forces due to magnetic fields and electrostatic charging may not act on the standard. This is achieved for primary standards when the magnetic permeability is smaller than 1.01, the surface resistance is smaller than 10^{12} ohm, and the corrosion resistance is smaller than $10^{-3}\ g \cdot m^{-2} \cdot A^{-1}$ [11]. The geometry should be chosen for best handling and cleaning; for example, a cylinder according to OIML-IR No. 20 [12]. The surface should be polished with roughness value $R_a = 0.1\ \mu m$, according to ISO-R 468. Two examples will show what else one has to consider:

In the BIPM it was found that the mass of national prototypes increases after cleaning by 2 to 5 μg per year [13]. This effect was also observed in standards of other institutes. After special measurements it is possible to take this into account by correction [14].

Extensive investigations showed that sorption layers on the surface should be considered when the environmental conditions change [15–17]. For example, the change in the water sorption layer of a 1-kg mass standard for an increase in humidity from 40% to 80% is about 20 μg. Figure 4 shows the specific mass of the water sorption layer calculated using this value ($0.14\ \mu g \cdot cm^{-2} \times 134\ cm^2 = 19\ \mu g$).

So one sees that it is important to carry out mass comparisons under fixed environmental conditions. According to all these criteria, besides platinum-iridium, high-alloy steels are also particularly useful for mass standards. Composition, density, attained magnetic permeability, and roughness of a stainless steel standard are given as an example in Table 1 [18].

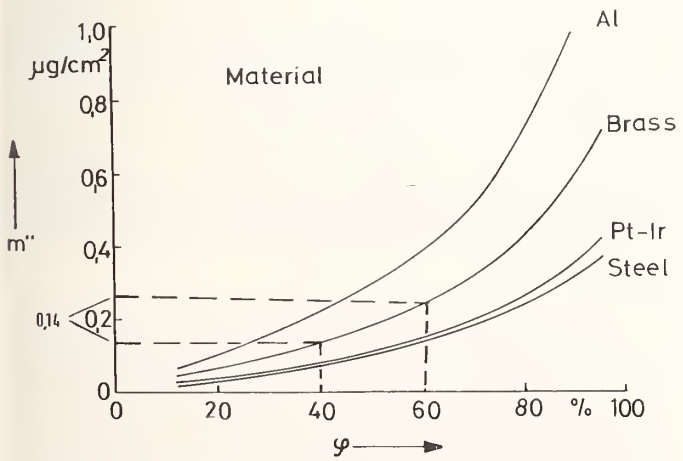


FIGURE 4. *H_2O-sorption layer m'' of different materials as a function of relative humidity. Surface: polished, $R_a = 0.2\ \mu m$; cleaning was with alcohol in an ultrasonic bath.*

TABLE 1. *Stainless steel for mass standards (example).*[a]

1. Chemical composition	
Carbon (C)	0.060%
Magnesium (Mn)	1.200%
Phosphorus (P)	0.021%
Sulphur (S)	0.004%
Silicon (Si)	0.530%
Chromium (Cr)	19.900%
Nickel (Ni)	24.900%
Molybdenum (Mo)	2.850%
Copper (Cu)	0.100%
Iron (Fe)	50.600%
2. Density at 20 °C	$\rho = 7998\ kg \cdot m^{-3}$
3. Roughness (ISO-R 468)	$R_a < 0.11\ \mu m$
4. Magnetic Permeability	$\mu < 1.01$

[a]From Ref. [18].

4. Prototype Balances

Prototype balances and precision balances with resolution (i.e., maximum load divided by scale division) higher than 10^7 are today still chiefly being built as beam balances, where either the knife-edge and flat bearing or a flexible strip acts as a central component. The quality of all these balances depends on the construction, setup, and adjustment accuracy, and in particular on storage and sensitivity as well as on auxiliary apparatus such as the locking device, the device for placing the masses on the pans, and the arrangement to compensate for small mass differences. In addition, the experience and skill of the operator is of no little importance. Figure 5 shows a typical prototype balance of the Rueprecht firm of about 1960, built as an equal-arm balance with three knife-edges and two pans, which is still used in numerous state institutes. The standard deviation achieved is up to 15 μg for a maximum load of 1 kg. Based on this construction, further developments over the last 20 years were the following:

During a weighing series the knife-edge and pan are not separated [19, 20], the oscillation of the beam is followed photoelectrically [21, 22], the balance can be loaded with up to four standards for substitution weighing, and it can be remotely controlled [23] (attainable standard deviation as small as 1 μg).

A new design was developed by the National Bureau of Standards (NBS), namely a single-pan beam balance with two knife-edges and short, unequal arms: see Figs. 6 and 7. The knife-edges and pans are not separated, and up to six standards can be compared with each other in a single series [24, 25] (attainable standard deviation as small as 1 μg).

Praiseworthy is the further development of commercial balances, e.g., when the beam of a knife-edge balance is brought into reference position by magnetic compensating forces [26] (see Fig. 8; attainable standard deviation as small as 10 μg).

Currently, as far as I know, there are two new developments in the prototype balance field:

In the National Physical Laboratory (NPL), Teddington, and BIPM, Sèvres, a beam balance is being

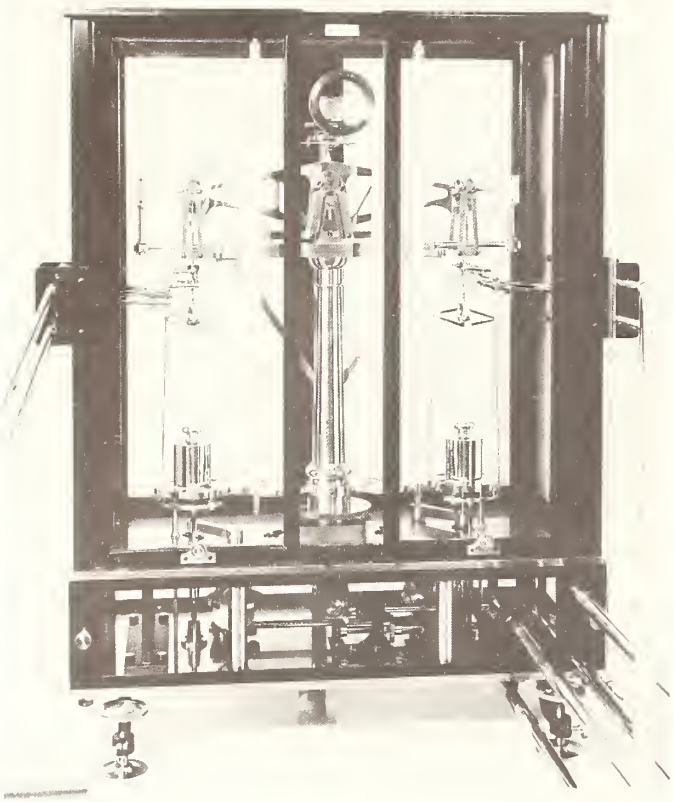


FIGURE 5. *1-kg Rueprecht prototype balance. 1: mass standard; 2: beam; 3: mechanism for loading additional weights; 4: device for changing the standards.*

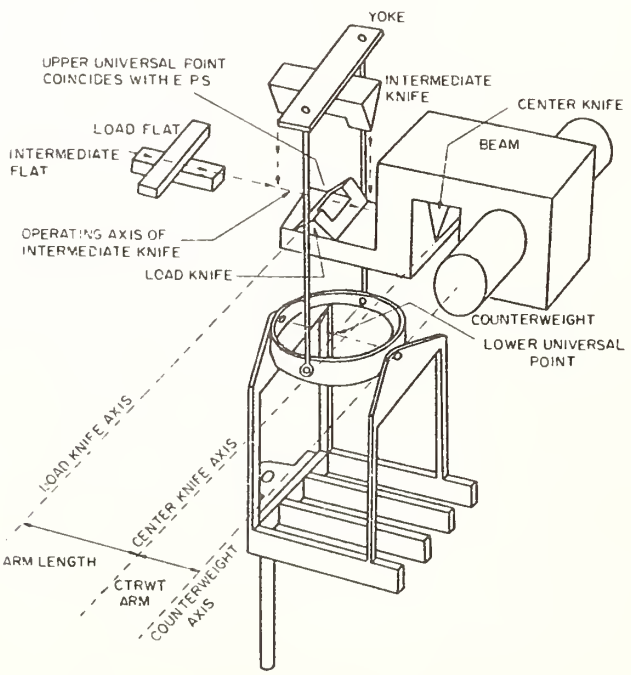


FIGURE 6. *Exploded sketch of the oscillating system of NBS 2 [24].*

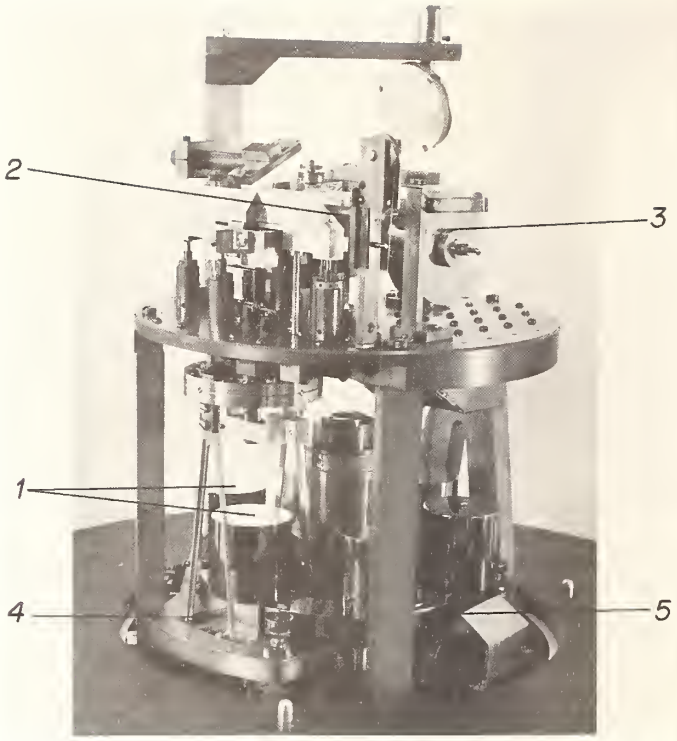


FIGURE 7. *1-kg NBS prototype balance. 1: mass standard; 2: beam with counterweight 3; 4: suspension with pan; 5: roundabout for mass standards. (Source, NBS.)*

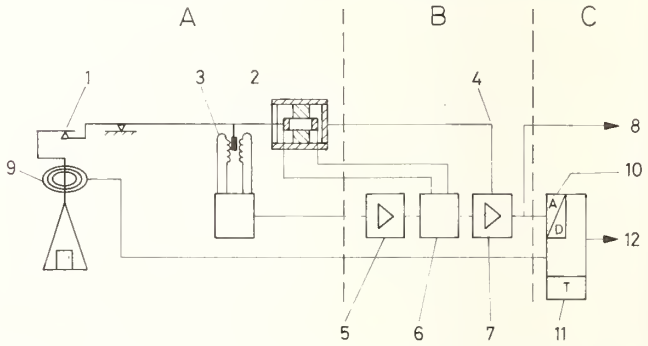


FIGURE 8. *Block diagram of a magnetic force compensation balance [26]. A: balance; B: control instrument; C: digital readout unit; 1: weighing system; 2: magnetic compensation system; 3: inductive beam scanning; 4: temperature compensation for sensitivity; 5: variable-gain amplifier; 6: range selector; 7: measurement amplifier; 8: analog output signal; 9: digital detection of dialed weights; 10: digital readout; 11: automatic tare; 12: digital output signal.*

developed with flexible strips instead of a knife-edge and flat [27]. Figure 9 shows in an exploded sketch the beam, 1, the flexible strips, 2, and the suspension, 3, of the pan.

At the PTB, Braunschweig, a liquid mass comparator is under construction, where differences in mass are determined by hydrostatic force compensation (Fig. 10). The buoyant force on a ring-shaped body, 1, submerged in a liquid, 2, is transmitted via three thin submerged rods, 3, to the suspension, 4, 5, of the balance which is located underneath the liquid container, 6. Owing to the change in upthrust due to the submersion depth of the rods, the vertical equilibrium depends on the applied load, 8. Using a weight changing device,

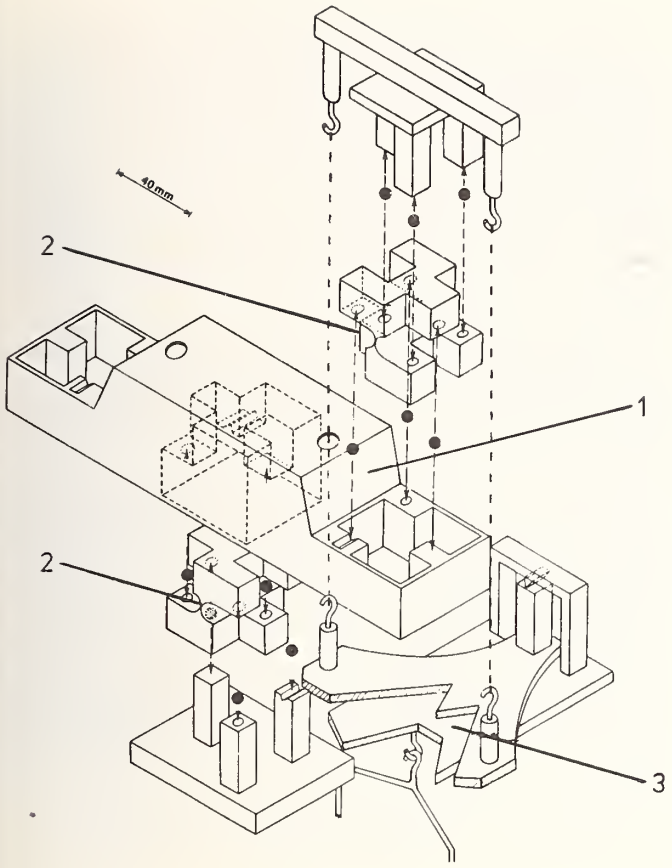


FIGURE 9. *Enlarged sketch of the beam and flexible strips of BIPM balance under construction [27]. 1: beam; 2: flexible strip; 3: pan suspension.*

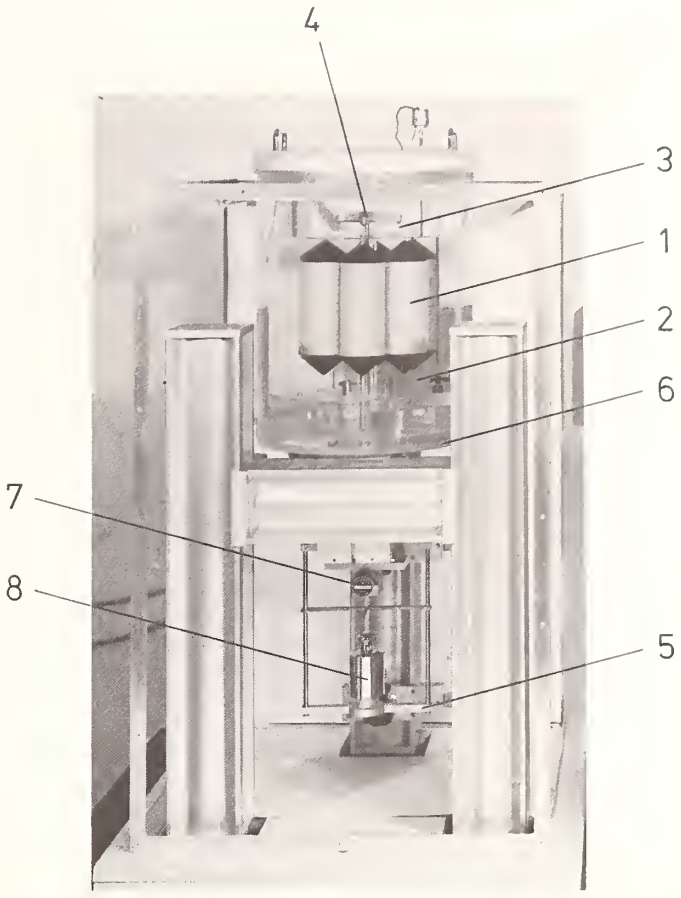


FIGURE 10. *Liquid mass comparator at PTB. 1: ring shaped floater; 2: liquid; 3: rods; 4: upper part of suspension; 5: suspension with load pan; 6: liquid container in water bath; 7: laser interferometer; 8: mass standard.*

up to six 1-kg weights can be placed consecutively in suspension, whereby the latter is kept constant by locking. The vertical difference in upthrust for the various weights is measured using an optical interferometer, 7, and with the aid of appropriate small mass standards the determined sensitivity is calculated as a mass difference. At this moment first measurements show a sensitivity of 10 μg.

5. Mass Hierarchical Linkup and Determination of Mass

According to the basic axiom of the 3rd General Conference of Weights and Measures in Paris, 1901, the weight of a body represents a quantity of the nature of a force and is the product of the mass of the body and the local acceleration due to gravity [28]:

$$F = mg_{\mathrm{loc}} .$$

This is the fundamental formula underlying the comparison of masses with one another. We regard the masses of two bodies to be equal when they exert the same force due to weight for the same value of the acceleration due to gravity (i.e., at the same place). Nevertheless, it is clear that we can only compare weight forces in vacuum. For weighing in air, as is generally the case, the sum of weight forces and buoyant forces are compared with each other.

To prove the equality of the masses of two bodies, we utilize the balance. When forces act at both ends of a pivoted lever of an equal-arm balance, the lever remains at rest if the resultant torques sum to zero. The equation of weighing in air is

$$L_{\mathrm{L}}(m_1 g_{\mathrm{loc}} - V_1 \rho_{\mathrm{L}} g_{\mathrm{loc}}) = L_{\mathrm{R}}(m_2 g_{\mathrm{loc}} - V_2 \rho_{\mathrm{r}} g_{\mathrm{loc}}) ,$$

where

m_1, m_2 = mass of bodies 1 and 2;
V_1, V_2 = volumes of bodies 1 and 2;
ρ_1, ρ_2 = density of bodies 1 and 2 from m/V;
ρ_{L} = density of air during weighing;
g_{loc} = local acceleration due to gravity; and
L_{L}, L_{R} = length of the two effective lever arms.

If $L_{\mathrm{L}} = L_{\mathrm{R}}$ and if g_{loc} is considered to be constant within the immediate vicinity of the balance, then with $\rho = m/V$, we obtain

$$m_2 = m_1 \frac{1 - \frac{\rho_{\mathrm{L}}}{\rho_1}}{1 - \frac{\rho_{\mathrm{L}}}{\rho_2}} .$$

This formula shows that for a comparison of mass standards the densities of the test object and the standard (if $\rho_1 \neq \rho_2$) as well as the momentary air density have to be

taken into account. In addition, there are several other influencing parameters:

5.1 Influences Due to the Mass Standard

A mass standard which satisfies the conditions already mentioned, protected from dust under bell-jars, of which the contamination and sorption layers are known and special forks or tongs are used for handling, has a mass stability of a few micrograms over several years in the case of a 1-kg standard.

5.2 Influences Due to the Balance

The uncertainty values achievable with prototype balances, mainly a few micrograms, have already been mentioned in Section 4. Table 2 shows the relevant data on five chosen balances, including the achievable uncertainty.

Three balances suitable for comparing mass standards have relative standard deviations of 10^{-7} to 10^{-8}, but only a few prototype balances like the NBS-2 reach 10^{-9}. One of the smallest known balances is a balance with which mass variations up to 20 pg are measured by the frequency variations of a resonating quartz crystal. A typical railway scale for special purposes in steel mills has a maximum capacity of 1800 t, which is taken up by 16 load cells.

5.3 Influences Due to the Weighing Process

The precision weighing described here takes place in air and so the air buoyancy has a greater effect the more the density of measurand and standard differ. For the comparison of 1-kg standards of platinum-iridium with those of steel, the air buoyancy correction is around 100,000 μg. The density of air can only be determined with a relative uncertainty of 1 to 2×10^{-4}, because the universal gas constant R (or equivalently, the density of dry air) and the composition of the air are not known to the required precision.

In 1979 at an international meeting on mass determination, agreement was reached over a uniform formula to calculate the density of moist air as well as tables and auxiliary formulae to be used for the application of this formula [29].

Changing environmental conditions further influence the balance and mass standards during weighing. Here one should take note of temperature fluctuations (Δt = 0.02 K); atmospheric pressure variations (Δp = 20 Pa); changes in the relative humidity (Δh = 1%); the composition and purity of the air; inhomogeneities of the gravitational field; stability of the balance site against gradients and vibration; effect of forces due to air turbulence, electrostatic charging, and magnetic fields between the balance, standard, and surroundings; and change of the temperature gradient along the balance arm.

TABLE 2. *Data on some balances of high accuracy.*

Weighing Instrument	Maximum Load	Standard Deviation
Mass Standards		
Electronic balance with additional weights	5g	0.2 μg ($S_{REL} = 4 \times 10^{-8}$)
Prototype balance NBS 2	1 kg	1 μg ($S_{REL} = 1 \times 10^{-9}$)
Mechanical equal arm balance	5 t	100 mg ($S_{REL} = 2 \times 10^{-7}$)
Research		
Resonating crystalline quartz balance	2 mg	20 pg ($S_{REL} = 1 \times 10^{-8}$)
Industry		
Railway truck scale with 12 load cells	1800 t	200 kg ($S_{REL} = 1 \times 10^{-4}$)

Here are 2 examples:

If the temperature of one entire lever-arm changes by 10^{-3} K compared with the other, then the result is a weighing error of 20 μg for a 1-kg weight and a 200 mm-long brass lever-arm.

Let the center of gravity of a 1-kg mass lie 50 mm higher on one side of the balance than the center of gravity of the standard on the other side. The correction for a 50 mm difference in the distance from the center of mass of the earth is about 15 μg.

5.4 Influences Due to the Weighing Procedure

Certain effects such as unequal lever arms or an inhomogeneous gravitational field are eliminated when tare weighing according to Borda or exchange weighing according to Gauss are carried out. Whereas both weighing procedures can be applied using the three knife-edge balance such as the Rueprecht (Fig. 5), only tare weighing is possible with the other balances such as the two knife-edge balance. As an example of the problem of unequal arms, we note that for an effective armlength of 200 mm, a maximum load of 1 kg, and a specified standard deviation for the balance of 1 μg, differences of only 0.2 nm in armlength will affect the result. Under optimum conditions, using the best prototype balance to compare two 1-kg standards of equal density, one can achieve a standard deviation as small as 1 μg.

6. Attainable One Standard Deviation Uncertainties

Taking the quadratic sum of the component uncertainties, namely the uncertainty of the national prototype—compared at BIPM: 8 μg; the uncertainty of the mass stability of the standards: 5 μg; the uncertainty of the air buoyancy correction: 10 μg; and of the weighing procedure itself: 2 μg; we obtain for a 1-kg standard of corrosion-resistant, non-magnetic steel of density 8000 kg·m^{-3} an empirical one standard deviation of uncertainty of about 15 μg (Table 3). Based on this value, a mass determination of a measurand with a density of 8000 kg·m^{-3} can be performed with the uncertainties shown in Figure 11.

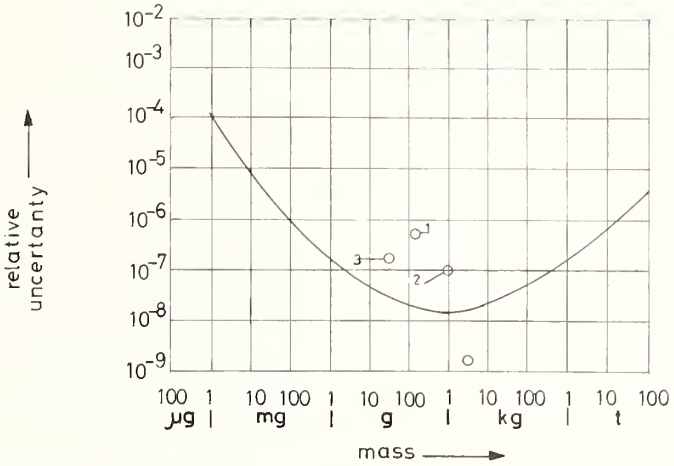


FIGURE 11. *Attainable uncertainties (one standard deviation) for mass determination.*

7. Mass Determination for the Realization of Other Units or Fundamental Constants

Three examples demonstrate the demands made on mass determinations:

For the determination of the Newtonian gravitational constant according to the setup by de Boer [30], weights of mass 125 g and density 2500 kg·m^{-3} must be determined with an uncertainty of less than 10^{-6} which, even taking into account the air buoyancy correction and the standard deviation of the balance, presents no problems (point 1 in Fig. 11).

For the determination of the Avogadro constant the density of silicon is required and for this the mass determination should have an uncertainty of less than 1×10^{-7} [31, 32] (mass of the measurand 1 kg, density 2400 kg·m^{-3}, point 2 in Fig. 11). With a certain amount of effort this requirement can be satisfied.

For the realization of the unit of electric potential using the voltage balance, with a dead load of about 3 kg, a mass determination of 30 g with an uncertainty of less than 2×10^{-7} is required. Whereas the mass

TABLE 3. *Attainable uncertainty in the mass of a 1-kg mass standard of density 8000 kg·m^{-3}*

Sources of Uncertainty	One Standard Deviation Uncertainty	
	At Present (μg)	Near Future (μg)
National standard	8	2
Instability of mass standards	5	2
Air buoyancy	10	2
Prototype balance, weighing procedure	2	2
Root sum square	14	4

determination is possible (point 3 in Fig. 11), the realization of a standard deviation of 2×10^{-9} for the voltage balance presents considerable difficulties [33].

8. Present Activities in the Area of Mass Determination

Three International working groups under the leadership of BIPM are currently concerned with the analysis and reduction of sources of error in mass measurement arising, for example, from determination of air density, volume of mass standards, sorption layers, cleaning, and handling. The first result of the work is a report on the air buoyancy correction with formulae and tables, etc., to calculate the air density [29]. A reduction of uncertainty in air buoyancy correction is equivalent to a more precise determination of the air density. A relative uncertainty of a few parts in 10^5 should be aimed for here [13].

In particular, the reference density of dry air, the actual composition of the air, and the degree of contamination due to suspended particles should be determined more accurately [34]. Recent measurements by Schoonover *et al.* [28], Prowse [35], and Kobayashi [36], show that a fundamental determination of the air density with a relative uncertainty of about 10^{-5} is possible. The method used is the measurement of the apparent mass difference of two samples, each of similar surface and mass but differing in volume. Further work in progress includes investigation of sorption layers on mass standards and the specification of cleaning procedures and environmental conditions for the maintenance and comparison of standards.

New developments concerning prototype balances have already been discussed in Section 4. However, weighing in vacuum is a different problem. As soon as the question has been clarified for which state the mass of the prototype has been defined, and further knowledge on the adsorption and desorption of mass standards and balances have been obtained, weighings in vacuum will recommence.

With the activities described here it may be possible over the next several years to achieve a relative uncertainty of less than 4×10^{-9} for relating a 1-kg standard with density 8000 kg·m^{-3} to one of density 21,000 kg·m^{-3}; see Table 3, righthand column.

9. New Definition of the Mass Unit "Kilogram"

A favorite daydream of metrology, analogous to the current representation of length, is to break away from the embodiment of the mass unit by the International Prototype and to define and realize it using fundamental constants. The following three possibilities are frequently discussed in the literature:

The kilogram is defined by a fixed number of atoms or molecules, for instance 2.741×10^{35} particles of the orange-red radiation of krypton 86 for the transition between $2p_{10}$ and $5d_5$ states [37]. However, since the authors never mention how this can be realized, this and other definitions based on counting particles need not be discussed at this time.

A second possible definition comes from the Avogadro constant if known with sufficient accuracy and using the following relation: The kilogram is

$$\frac{(N_A)\ \text{kmol}^{-1}}{12}$$

times the mass of an atom of ^{12}C [38]. The Avogadro constant itself is measured with a relative uncertainty of 1×10^{-6} using a silicon single crystal via density, lattice constant, and amount of substance [39, 40]. Compared with the present system where the relative uncertainty in the kilogram is 2×10^{-8} the Avogadro constant would have to be determined more precisely by 2 powers of 10; only then would a new definition be metrologically justified.

A third possibility is to determine the kilogram by electrostatic forces using a voltage balance [41]. The three mechanical base units meter, second, and kilogram are related to the unit of potential and to the electric constant by way of the force due to an electric field. In this case the kilogram would be metrologically defined by the electrostatic forces of a voltage balance. All base units, the kilogram too, could be derived from quantum transitions. However, this would only be a transposition of the least accurately realizable unit from the ampere to the kilogram. The relative uncertainty of the realization to date would be about 2×10^{-6}, which is also 2 powers of 10 worse than the present arrangement. Another problem is the influence of the gravitational acceleration (see Fig. 12).

These considerations show that a new definition cannot be expected in the near future. Nevertheless, the search should continue for ways of realizing a new definition using fundamental constants, like the Avogadro constant, in a consistent set of units and constants (Fig. 13).

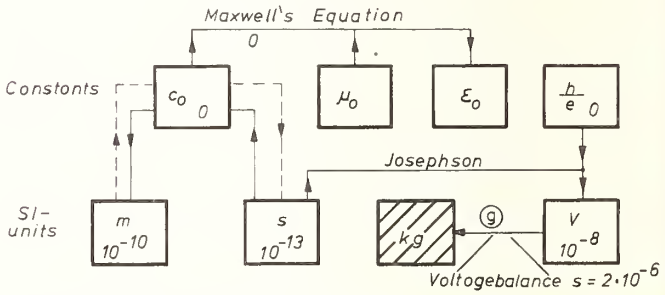


FIGURE 12. *New definition of the unit of mass based on a voltage balance [41].*

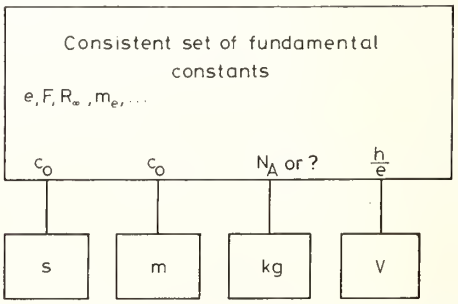


FIGURE 13. *New definition of the unit of mass based on a consistent set of fundamental constants.*

References

[1] F. S. Sawelski, *Die Masse und die Messung* (Verlag Harri Deutsch, Frankfurt/Main, 1977) [original in Russian].
[2] L. E. Howlett, Z. Angew. Phys. 15, 463 (1963).
[3] S. German and M. Kochsiek, Bull. OIML No. 68, 18 (1977); No. 69, 7 (1977) [Organisation Internationale de Métrologie Légale, Paris, France].
[4] D. B. Prowse and A. R. Anderson, Metrologia 10, 123 (1964).
[5] P. E. Pontius, Natl. Bur. Stand. Tech. Note 288 (1966).
[6] M. Grabe, PTB Mitteilungen 87, 223 (1977).
[7] J. M. Cameron, M. C. Croarkin, and R. C. Raybold, Natl. Bur. Stand. Technical Note 952 (1977).
[8] A. Bonhoure, Microtecnic 6, 151 (1952).
[9] H. E. Almer, Weight cleaning procedures, Natl. Bur. Stand. Report 9683 (Feb. 15, 1968).
[10] G. Girard, BIPM Report RM76/5 (Bureau International des Poids et Mesures, 1975).
[11] M. Kochsiek, Wägen u. Dosieren 9, 4 (1978).
[12] OIML-IR 20: Poids des classes de precision E_1, E_2, F_1, F_2, M_1 des 50 kg a 1 mg (Organisation Internationale de Métrologie Légale, Paris, France).
[13] Rapport sur la réunion concernant les masses, BIPM-CIPM/77−1, p. A3 (Bureau International des Poids et Mesures, 1977).
[14] M. Kochsiek, PTB Mitteilungen 89, 421 (1979).
[15] Y. Kobayashi, these proceedings.
[16] M. Kochsiek, PTB Mitteilungen 87, 478 (1977).
[17] T. Yoshimori *et al.*, Trans. Jpn. Inst. Met. 14, 396 (1973).
[18] ASMW, Germ. Dem. Rep., Communication CCM/81−13E Consultative Committee on Mass (Bureau International des Poids et Mesures, 1981).
[19] F. A. Gould, Proc. R. Soc. London, Ser. B: 62, 817 (1949).
[20] R. Spurny, Communication CCM/81−22, Consultative Committee on Mass (Bureau International des Poids et Mesures, 1981).
[21] W. Lotmar and J. G. Ulrich, Metrologia 12, 159 (1976).
[22] M. Kochsiek *et al.*, Bull, OIML No. 70, 7 (1978) [Organisation Internationale de Métrologie Légale, Paris, France].
[23] Y. Kobayashi and K. Uchikawa, INSYMET (1974), Bratislava.
[24] H. A. Bowman and H. E. Almer, J. Res. Natl. Bur. Stand. Sect. C: 67, 227 (1963).
[25] H. E. Almer, J. Res. Natl. Bur. Stand. Sect. C: 76, 1 (1972).
[26] Mettler Instrument Corporation, specifications for model H315 spec.
[27] Rapport du directeur sur l'activité et la gestion du BIPM, corrections et additions, 30 Sept. 1980, document CIPM/80−11 (Comité International des Poids et Mesures).
[28] *The International Bureau of Weights and Measures 1875–1975*, Ed. by C. H. Page and P. Vigoureux, Natl. Bur. Stand. (U.S.), Spec. Publ. 420 (May 1975). [English translation of the BIPM Centennial Volume *Le Bureau International des Poids et Mesures 1875–1975*.]
[29] BIPM Report 79/10 (Bureau International des Poids et Mesures, 1979).
[30] H. de Boer, these proceedings.
[31] R. D. Deslattes *et al.*, Phys. Rev. Lett. 33, 463 (1974); Ref. [40]; and these proceedings.
[32] P. Seyfried, these proceedings.
[33] V. Sienknecht, PTB-Bericht E-12 (Physikalisch-Technische Bundesanstalt, 979), p. 47.
[34] R. M. Schoonover *et al.*, J. Res. Natl. Bur. Stand. 85, 27 (1980).
[35] D. B. Prowse, these proceedings.
[36] Y. Kobayashi, Communication CCM/81−12B, Consultative Committee on Mass (Bureau International des Poids et Mesures, 1981).
[37] S. K. Mukherjee and A. Choudhury, ISI Bull. 24, 288 (1973) [Indian Standards Institution, India].
[38] P. Draht, PTB Mitteilungen 85, 44 (1975).
[39] Dimensions/NBS 58, 219 (1974) [Dimensions is the former Natl. Bur. Stand. monthly information magazine].
[40] R. D. Deslattes, in *Metrology and Fundamental Constants*, Proc. of the International Summer School "Enrico Fermi," course LXVIII, Ed. by A. Ferro Milone, P. Giacomo, and S. Leschiutta (North Holland, Amsterdam, 1980), p. 38.
[41] H. J. Schrader, PTB Mitteilungen 85, 29 (1975).
[42] S. I. Toropin and V. C. Snegov, Izmer. Tekh. 1975, No. 12, 75 (1975). [English transl.: Meas. Tech. 18, 1847 (1975).]
[43] W. F. Koch *et al.*, J. Res. Natl. Bur. Stand. 83, 407 (1978).
[44] M. Romanowski and G. Mihailo, New developments in the metrology of mass standards, unpublished internal report, National Research Council, Canada.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Measurement of Air Density for High Accuracy Mass Determination

D. B. Prowse

CSIRO Division of Applied Physics, Sydney, Australia 2070

A description is given of the air density balance developed for the fundamental determination of the density of air. The method consists of measuring the apparent mass difference between a sphere and a ring of similar mass and large known difference in volume. The two-arm balance used has been modified to give a continuous electrical readout of the air density. The uncertainty (3 standard deviations) in air density given by the method is 3.5×10^{-5} kg · m^{-3}. Preliminary results indicate that the air density balance agrees with the values obtained by the equations developed to calculate air density to 2 parts in 10^4. Rapid fluctuations up to 1 in 10^4 of the air density are observed in an air-conditioned laboratory.

Key words: air density; apparent mass difference; continuous electrical readout; sphere and ring; two-arm balance.

1. Introduction

All measurements of mass are based on Pt-Ir prototype kilograms. For practical use these prototypes need to be compared with stainless steel kilograms and the special standardizing balances used in these comparisons usually have a discrimination of 1–2 μg. Because the masses have different volumes (approximately 80 cm^3 disparity) the density of the air must be accurately known to enable corrections to be made for the difference in air buoyancy on the masses. Recent work by Jones [1] has shown that current state of the art measurements for pressure, temperature, relative humidity, and composition give an overall fractional uncertainty in air density of 3×10^{-4}. This corresponds to an uncertainty of 29 μg when comparing kilogram artifacts of Pt-Ir and stainless steel.

To be able to transfer the mass unit from the Pt-Ir to the stainless steel artifacts with an uncertainty approaching the discrimination of the balances, work at the National Measurement Laboratory has been aimed at a direct measurement of the air density with an uncertainty approaching 1–2 $\times 10^{-5}$.

The method used is the measurement of the apparent mass difference between a glass sphere and ring, each of similar surface area and true mass but differing in volume by just over 1000 cm^3. The sphere is fitted with a small teflon sealed valve to enable it to be evacuated. Because the sphere is not permanently sealed it is possible to measure the necessary properties to enable the air density to be determined without any assumptions or further calibration.

2. Volume Measurements

Two spheres were made from 1 liter flasks by removing the neck and adding a small teflon-sealed vacuum valve. The ring was made from glass tubing of diameter and thickness chosen so that the surface areas of the ring and spheres were as similar as possible—within about 10%. No effect of the small difference in surface area has been observed.

If the arms of the balance used for measuring the air density are exactly equal only the difference in volume, and mass, between the sphere and ring would need to be measured. Because the arm lengths are not equal, the individual volumes and masses must be known.

The volumes were calculated from weighings in air and water. The density of water in equilibrium with air is known to at least 2 in 10^6 [2]. If the density of air can be calculated to 2 parts in 10^4 using standard equations [1, 3] then, except for weighing the sealed spheres in air, the largest uncertainty in any individual weighing due to the air density is 4 parts in 10^6.

Sufficient mercury was added to the spheres so that they just sank in water. The spheres were then evacuated—a necessary procedure because the change in volume on evacuation was found to be 1 part in 10^4. In use they are always evacuated to a pressure of less than 1 Pa.

The weighings were made with a hydrostatic balance. The vessel containing distilled water was placed in a water bath which was stirred, but not temperature controlled as the temperature remained sufficiently constant and uniform during each measurement.

By weighing the ring and two spheres in air and water and against each other, 10 equations in 6 unknowns of the form

$$M_{s1} - V_{s1}\rho = m_1(1 - \sigma/\Delta) \qquad (1)$$

$$M_r - V_r\rho - M_{s1} + V_{s1}\rho = m_5(1 - \sigma/\Delta)\,, \qquad (2)$$

were obtained.

In matrix notation these may be written

$$\begin{bmatrix} 1 & & & -\rho & & \\ 1 & & & -\sigma & & \\ & & 1 & & -\rho & \\ & & 1 & & -\sigma & \\ -1 & & 1 & \rho & -\rho & \\ & 1 & & -\rho & & \\ & 1 & & -\sigma & & \\ & -1 & 1 & & \rho & -\rho \\ 1 & -1 & & -\sigma & \sigma & \\ 1 & -1 & & -\rho & \rho & \end{bmatrix} \begin{bmatrix} M_{s1} \\ M_{s2} \\ M_r \\ V_{s1} \\ V_{s2} \\ V_r \end{bmatrix} = \begin{bmatrix} m_1(1 - \sigma/\Delta) \\ '' \\ '' \\ '' \\ '' \\ '' \\ '' \\ '' \\ m_{10}(1 - \sigma/\Delta) \end{bmatrix} . \qquad (3)$$

where

M_{s1}, M_{s2} = mass of sphere 1 and sphere 2, respectively;

V_{s1}, V_{s2} = volume of sphere 1 and sphere 2, respectively;

M_r = mass of the ring;

V_r = volume of the ring;

Δ = nominal density of stainless steel masses;

ρ = density of water at the time of each weighing;

σ = density of air at the time of each weighing (calculated from measurements of pressure, temperature and relative humidity); and

$m_1 \ldots\ldots m_{10}$ = stainless steel masses required to balance each weighing.

These 10 equations were solved by the method of least-squares for the 6 unknowns, of which only V_{s1}, V_{s2} and V_r are relevant. From the variance-covariance matrix a standard deviation of 3.5×10^{-3} cm^3 (3.5 parts in 10^6) was obtained for the volume difference $V_s - V_r$.

3. Mass Measurements

By adjusting the nichrome wire suspensions the true mass of the spheres and the ring were made equal to within a few mg. The mass difference between the material in the ring and the spheres was determined initially by weighing them in evacuated aluminium cans. This however gave an uncertainty of 0.5 mg, due mainly to the size of the cans and the difficulty in manipulating them on the balance. The ring and the spheres were then compared in air with the valve on each sphere open. Static electricity was found to cause discrepancies of up to 1 mg for both sets of weighings. When this was eliminated the standard deviation of the mass difference between the spheres and the ring was found to be 10 μg. Within the uncertainties the values were the same for weighings made in air and in vacuum.

4. Air Density Balance

The actual measurement of air density is made on a two-arm hydrostatic balance which has been modified by the method described by Kibble [4] to give electrical readout. A second order filter was added to smooth out noise and any fluctuations caused by air turbulence. The sphere and ring were suspended beneath the balance in an enclosure to protect them from air currents. A special mechanism enabled them to be lifted clear of the pans without arresting the balance. This permitted the balance zero to be easily determined so that small drifts due to temperature, etc., could be eliminated. The beam compartment of the balance was thermally insulated in an attempt to eliminate temperature gradients, but the balance still has a small zero drift of about 80 μg/K. The standard deviation of the balance with the Kibble system is 5 μg, which in terms of air density is equivalent to 5×10^{-6} kg $\cdot$ m^{-3}.

Tests to determine the sensitivity of the balance itself to changes in air density were made by filling the balance case with argon and then helium. Within the uncertainty of measurement the balance appears to be unaffected by changes in the ambient atmosphere.

The coil in the system was adjusted to give the balance a range of ± 20 mg which is adequate to cover the change of air density under ambient conditions in the laboratory. However, with such a large range, the response of the system is not quite linear so that as the mass difference measured by the balance increases, the light spot on the null detector moves further off center. This can be overcome in two ways: working near the balance null by adjusting the rider associated with the balance, or preferably incorporating an integrator in the feedback circuit of the operational amplifier to reduce the feedback current at null to zero. The first method is being used while the apparatus is being modified to incorporate the second method.

If the balance is assumed to have equal arm lengths then the air density is given by

$$\sigma = \frac{M_s - M_r + W - (a - b)s}{V_s - V_r + W/\Delta}, \qquad (4)$$

where W is a known stainless steel mass of approximately 1.2 g associated with the sphere; a, b are the reading and zero reading of the balance, respectively; and s is the sensitivity.

For a balance in which the arm lengths are unequal, the air density is given by an equation of the form:

$$\sigma = \frac{(M_s L_1 - M_r L_2) + (m_L L_1 - m_R L_2) + W L_1 - (a - b)s}{(V_s + m_L/\Delta + W/\Delta)L_1 - (V_r + m_R/\Delta)L_2}, \qquad (5)$$

where m_L, m_R are the mass of the left and right hangers, respectively (these hangers are lifted off with the sphere and ring when the zero reading is determined); L_1, L_2 are the left and right arm lengths, respectively, of the balance; and L_1 and L_2 are determined by weighing two known stainless steel masses of the same mass as the spheres and ring (250 g). Equation (5) includes the term $(M_s L_1 - M_r L_2)$, however, this can be transformed so that

$$M_s L_1 - M_r L_2 = M_r(L_1 - L_2) + (M_s - M_r)L_1. \qquad (6)$$

As $L_1 - L_2$ is small, M_r needs only to be known to 0.1 g. Equations (5) and (6) assume that the sphere is on the left hand pan of the balance.

5. Conclusion

The uncertainty in air density (3 standard deviations) given by this method is 3.5×10^{-5} kg $\cdot$ m^{-3}.

The equations developed by Jones [1] and the BIPM [3] for calculating the air density from measurements of pressure, temperature, relative humidity, and composition agree with each other to 2 parts in 10^5. The results from these equations have been compared with the density as given by the air density balance for an assumed standard composition of the air. In general these agreed to 2 parts in 10^4 with any closer agreement being limited by the uncertainty in calibration of the barometer. Future work will improve the pressure measurements and allow for variation in the composition of the air.

The balance shows that the air density in a normal air-conditioned laboratory fluctuates too fast for the level of accuracy obtained from the balance to be used effectively. Rapid fluctuations in air density with periods ranging from seconds to minutes and of magnitude up to 1 in 10^4 have been observed. These fluctuations are mainly due to pressure changes, for example, opening the door of the room causes a density change of $2.3 \times$

10^{-4} kg · m^{-3}, and are confirmed by observations with a commercial gas density meter (see Fig. 1). Thus for measurements of mass of the highest precision, either the balance case or the balance room should be sealed against rapid pressure changes and the air-conditioning turned off.

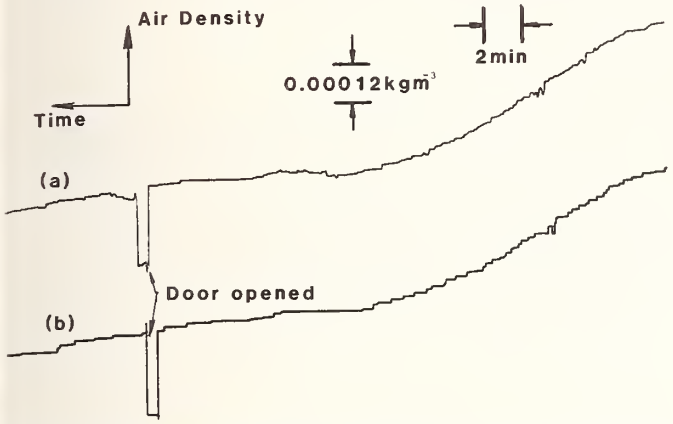


FIGURE 1. *Recorder trace of the output of the air density balance (a) and a commercial gas density meter (b).*

One obvious application of the air density balance is to monitor the output of both the standardizing and the air density balance with a small computer so that each turning point of the standardizing balance swing can be corrected for the value of the air density. Another possible application is to obtain a value for the gas constant. If the balance case and sphere and ring enclosure are sealed, then by filling them with pure gas it should be possible to measure the gas constant R to a few parts in 10^5.

I would like to thank Mr. G. A. Bell and Dr. N. Bignell for their contribution to this work.

References

[1] F. E. Jones, J. Res. Natl. Bur. Stand. 83, 419 (1978).
[2] G. A. Bell and J. B. Patterson, these proceedings.
[3] Bureau International des Poids et Mesures, Rapport BIPM 79/10, October, 1979.
[4] B. P. Kibble, Metrologia 11, 1 (1975).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

On a More Precise Correction for Buoyancy and Gas Adsorption in Mass Measurement

Y. Kobayashi

National Research Laboratory of Metrology 1-4, 1-Chome, Umezono, Sakura-Mura, Niihari-Gun, Ibaraki 305 Japan

Using Pt-Ir Kilogram Prototypes, mass standards made of other materials, for example stainless steel, are usually calibrated in the atmosphere. In this case, the precision is limited by various conditions of the measuring environment. In this study, the increased precision of the correction for buoyancy and the introduction of a correction for water vapor adsorption made it possible to decrease, by one order of magnitude, the errors caused by the measuring environment.

Keywords: buoyancy correction; gas adsorption; kilogram prototype; mass standards; precision weighing.

1. Introduction

It is essential to be able to compare a Pt-Ir Prototype with standard weights made of other materials such as stainless steel, etc. In the International System of Units (SI), the International Prototype Kilogram defines the unit of mass, which is disseminated as a practical quantity through National Prototype Kilograms. In this case, comparisons are carried out in the atmosphere in order to prevent any change in mass due to diffusion of the gases included in the Prototypes and other such effects. Consequently, in order to ensure an exact link of practical mass standards to the SI unit, one must take into account appropriate corrections arising from the difference in materials, shapes, and surface conditions between the Prototype and standard weight. Additional corrections for the difference between the weighing environment at the time the National Prototype was calibrated by the Bureau International des Poids et Mesures (BIPM) and the environment at the time of the calibration of the standard weight in terms of the Prototype must also be considered.

It is therefore necessary to overcome the problems which limit calibration precision and, ideally, to eliminate them. These problems can be classified into two categories: precise correction for buoyancy effects; and possible corrections for gas adsorption.

2. Precise Correction for Buoyancy

When weighing is carried out in the atmosphere, the measured value is an apparent mass difference, which includes an influence due to the conditions of the weighing environment. Apparent mass must be converted to true mass by means of a correction for air buoyancy. This correction can be determined on the basis of the pressure, temperature, and humidity of the air in the vicinity of the weighing (e.g., in the balance case). Such a determination carries with it 30 to 50 μg of uncertainty in the buoyancy correction when comparing a Pt-Ir Prototype and a 1-kg stainless steel weight due both to insufficient precision in characterizing the ambient air, and to the difference between the composition of the air in the vicinity of the weighing and that of "standard air". It is impossible to improve this precision by using the above conventional method.

To obtain better than 10 μg accuracy for the buoyancy correction, the National Research Laboratory of Metrology (NRLM) has developed a direct measurement method for air density and for the buoyancy correction itself.

2.1 Direct Measurement of the Air Density

When the true mass difference $m \equiv (m_A - m_B)$ between two sinkers A and B, made of stainless steel, and the volume difference $V \equiv (V_A - V_B)$ are known, the density of air, d_a, is determined by the measurement of the apparent mass difference M in air between the two sinkers as follows:

$$d_a = (m - M)/V \ . \tag{1}$$

Figure 1 shows the forms of the sinkers, made so that their surface areas are similar and their volume difference large. Their mass difference was measured by weighing under vacuum with a precision balance having knife-edge bearings [1]. The volume difference and the coefficient of volume expansion were determined by hydrostatic weighing.

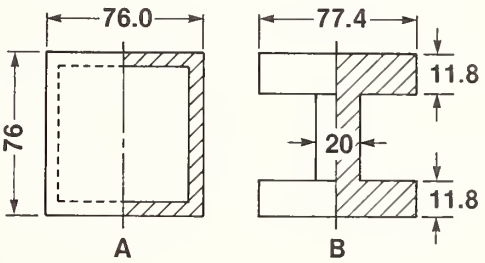


FIGURE 1. *Forms of the sinkers A and B (side view). The dimensions are given in mm and they are cylindrical.*

Table 1 shows the measured value of the volume and the coefficient of expansion of the sinkers, and Table 2 shows their mass difference.

The experiment was performed by a substitution weighing method using the sinkers. Table 3 shows the result of a N_2 density measurement which was carried out as a test of the measuring apparatus. The measured value of N_2 density was in agreement with accepted

TABLE 1. *Measured volumes and expansion coefficients of the sinkers.*

Sinker	Volume at 20 °C (cm^3)	Expansion coefficient	Relative uncertainty of volume
A	343.408	44.7×10^{-6}/°C	6×10^{-6}/°C
B	126.890	44.9×10^{-6}/°C	9×10^{-6}/°C

TABLE 2. *Mass difference of sinkers A and B determined by weighing in vacuum.*

Time under vacuum (days)	Mass difference (mg)
83	266.417
93	266.412
99	266.416

TABLE 3. *Measured values of the density of N_2*

Run No.	Conditions			Density (kg/m^3)	
	p (mmHg)	t (°C)	H (%)	As measured	Converted to STP
1	776.53	16.05	54	1.20243	1.25063
2	769.00	16.02	—	1.19515	1.25061

values and had a precision better than 1×10^{-4}. These results suggested the capability of performing a highly precise measurement of air density, without relying on measurements of the state of the air. Table 4 shows the results of the air density measurements. Repeated measurements of the same air had a precision of 1×10^{-4}.

TABLE 4. *Measured values of the density of air.*

Run No.	Conditions			Density (kg/m^3)	
	p (mmHg)	t (°C)	H (%)	As measured	Converted to STP
1	763.79	16.86	32	1.22114	1.29321
2	764.72	17.34	29	1.22085	1.29328
3	757.29	17.41	55	1.20602	1.29296
4	751.31	17.28	52	1.19743	1.29310

2.2 Direct Correction of the Buoyancy

When the apparent mass difference a and b between two combinations of the prototype K, stainless steel weight W, and auxiliary pans P and Q are measured under stable air conditions, a and b are obtained as follows:

$$a = (W' + Q') - (K' + P')$$
$$b = (W' + P') - (K' + Q'), \quad (2)$$

where W', K', P', and Q' are the apparent mass of each of the weights and auxiliary pans. If the true mass difference $m \equiv (P - Q)$ and the volume difference $V \equiv (V_P - V_Q)$ between the two auxiliary pans are known, and the volume difference $v \equiv (V_K - V_W)$ is also known, then the mass of W is determined by

$$W = K + (a + b)/2 + [(a - b)/2 + m](v/V). \quad (3)$$

According to this method, it is possible to carry out a correction for buoyancy which is more precise than 1×10^{-4} under the condition that $v < V$ and the surface area of the auxiliary pans is the same as that of the present sinkers.

3. Correction for Gas Adsorption

Although the mass variation of mass standards caused by gas adsorption can not be neglected, no method has been developed to cope with it. In order to achieve higher accuracy, it is necessary to develop a method of correction for gas adsorption so as to reduce its influence. It appears that mass variation due to gas adsorption is mainly caused by water vapor. Figure 2 shows the relation between adsorbed gas and humidity as determined by using Ni foil. These measurements, as shown in Fig. 2, support the claim that humidity is the only important variable. In this case, adsorbed amount per unit

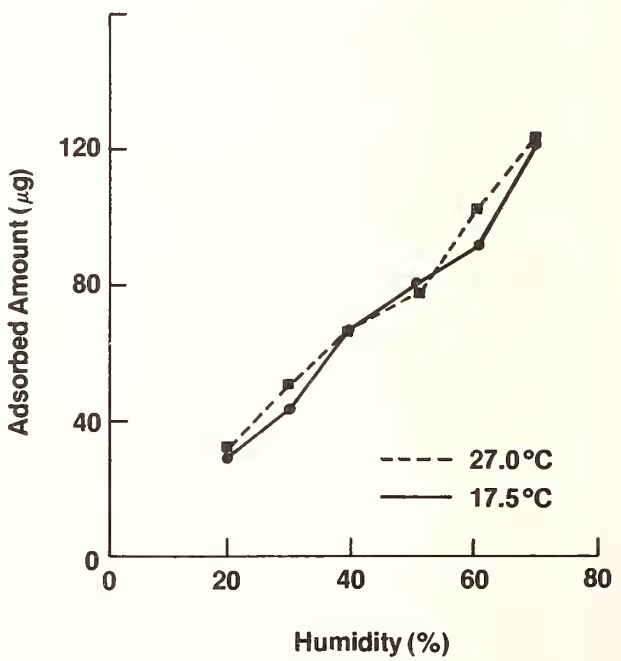


FIGURE 2. *Relation between the amount of gas adsorbed and relative humidity.*

area, ζ, is described as follows:

$$\zeta = \eta H + C \,, \tag{4}$$

where H is relative humidity in percent and C and ζ are constants. Table 5 shows the result of tests on a Ni sample having a surface area of 200.0 cm^2 which was measured between limits of humidity of 20% to 70%.

TABLE 5. *Coefficients of Eq. (4).*

Sample	t (°C)	η ($\mu g/cm^2$)	C ($\mu g/cm^2$)	H (%)
Ni	17.5	0.009	−0.03	20 to 70
	27.0	0.010	−0.06	

Several measurements were carried out on a Pt-Ir Prototype and a stainless steel weight. They were performed in vacuum as well as in air of various conditions. The results are as follows:

$$\begin{aligned} &(S2-1) - (E59) = 95.464 \text{ mg} \ldots\ldots \text{ in vacuum} \\ &(S2-1) - (E59) = 95.489 \text{ mg} \ldots\ldots \text{ in air} \,, \end{aligned} \tag{5}$$

where (S2-1) is a stainless steel 1-kg weight and (E59) is a Pt-Ir Prototype. The vacuum measurement was made after 30 days at a pressure of less than 67 Pa; the values in air (H = 52%) were corrected for buoyancy by means of the sinkers described above. The difference between the result in air and in vacuum is 25 μg. It seems that this discrepancy is due to the influence of adsorption. The equation obtained for the correction due to gas adsorption, derived from the mean values of several measurements, is as follows:

$$A_c = (0.01)S_K - (0.0092H - 0.10)\,\Delta S \tag{6}$$

where A_c is the correction in micrograms, S_K is the surface area of the prototype in square centimeters, H is the relative humidity in percent, and ΔS is the surface area difference between the prototype and the stainless steel weight in square centimeters.

4. Conclusion

The increased precision of the correction for buoyancy and the introduction of a correction for gas adsorption have made it possible to decrease by one order of magnitude the error caused by the weighing environment.

References

[1] Y. Kobayshi and K. Uchikawa, Bull. Natl. Res. Lab. Metrol. (Tokyo) 32, 6 (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Density Standards—The Density and Thermal Dilatation of Water

G. A. Bell and J. B. Patterson

CSIRO Division of Applied Physics, Sydney, Australia 2070

A hydrostatic weighing experiment has been done to measure the density of water samples of differing isotopic composition from which values have been derived for the density of SMOW (standard mean ocean water). Measurements are also being made of the dilatation of water in the temperature range 0 to 40 °C.

Key words: density standards; water; weighing.

1. Introduction

Pure water became the standard of density when the kilogram was first defined as the mass of one cubic decimeter of water at its temperature of maximum density under a pressure of one atmosphere. However the mass of the kilogram as fabricated by Lavosier was not equal to the mass of a cubic decimeter of water and this fact was demonstrated by the very careful work at the BIPM of Guillaume, Chappuis, and of Macé de Lépinay, Buisson, and Benoit. All of this work is reported in the Travaux et Memoires du BIPM [1].

The dilatation of water was measured in the first decade of this century in the temperature range 0° to 40 °C by Chappuis of the BIPM [2] and by Thiesen, Scheel, and Diesellhorst of the PTR Berlin [3]. These two sets of results differed considerably above 15 °C and have been the subject of numerous re-workings, the first of which was published in 1937 and the latest in 1975 [4] [5].

Although the experimental work of both Chappuis and Thiesen was of very high quality, it is of little value today as the isotopic composition of the water used is not known and the relationship of the temperature scales used by them with the modern IPTS is not known with any certainty. By the end of the 1960's it was clear that something would have to be done to improve the density standard and two different approaches have been made.

In one approach, measurements are being made with the object of defining the density standard by means of a single crystal of a solid, viz. silicon. Whether single crystals of silicon can be produced sufficiently free from impurities, vacancies and dislocations to serve as a natural standard is not yet established.

The other approach aims at establishing pure air-free water of known isotopic composition, at its temperature of maximum density, as the density standard, and one experiment to this end is described here.

2. Principle of the Experiment

The measurement of the density of water involves the measurement of the mass of water displaced by a body of accurately known volume. In order that the volume of the immersed body be readily determined to a high order of accuracy from its linear dimensions, a body of exact geometrical form must be used. Practical considerations limit the choice to a sphere, a rectangular prism or a cylinder, and for this experiment a sphere was chosen as it is much less susceptible to damage.

The sphere, which is hollow and made of a low expansion glass, has been described by Bell and Clarke [6].

The experimental set-up is shown in Fig. 1. The sphere is suspended in a cage from the pan of a two knife substitution balance and provision is made for lifting the sphere from the suspended cage.

3. Measurement of the Sphere

Measurements using a "Talyrond" roundness measuring instrument indicated that the departure from sphericity did not exceed 100 nm in any measured diametral plane.

Using a He-Ne stabilized laser calibrated in terms of the krypton 86 standard as a light source, the diameter

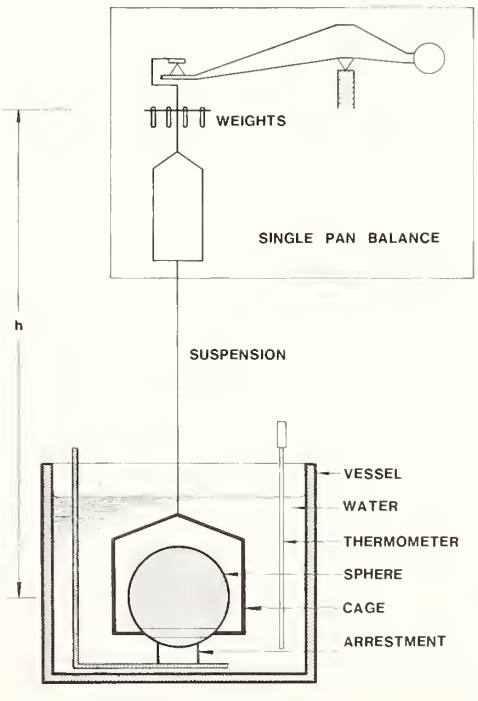


FIGURE 1. *Principle of experiment.*

of the sphere was measured by means of an interferometer based on the design of Saunders [7]. The sphere was placed between the parallel plates of a Fabry-Perot interferometer and the radial separation of the spherical surface from each plate was measured using Newton's rings. The interferometer plates were fixed to two steel plates, the separation of which was fixed by three invar spacing rods. Parallelism of the plates was achieved by deforming the ends of the rods with a differential screw mechanism. The sphere and parallel plates were mounted in a temperature controlled vacuum chamber with provision for making all the necessary adjustments externally.

The sphere could be rotated to any orientation and was able to be lowered away from the interferometer path to measure the separation of the plates.

Twenty small spots of aluminium were evaporated on to the surface of the sphere at the extremities of ten diameters to enhance the visibility of the Newton rings and to eliminate unwanted interference in transmitted light. The interferometer is shown diagramatically in Fig. 2.

The correction for phase loss on reflection and for the thickness of the aluminium film was determined by measurement in an interference microscope.

The compressibility of the sphere was determined by using the differences in diameter values at both atmospheric pressure and under vacuum conditions, at a temperature of 20 °C.

In all, three complete sets of measurements of the sphere diameter were made, between two of which the interferometer was taken down and reassembled. Measurements were made at a number of temperatures in the range 6 to 40 °C and the results fitted to an expression quadratic in temperature. The results are shown graphically in Fig. 3.

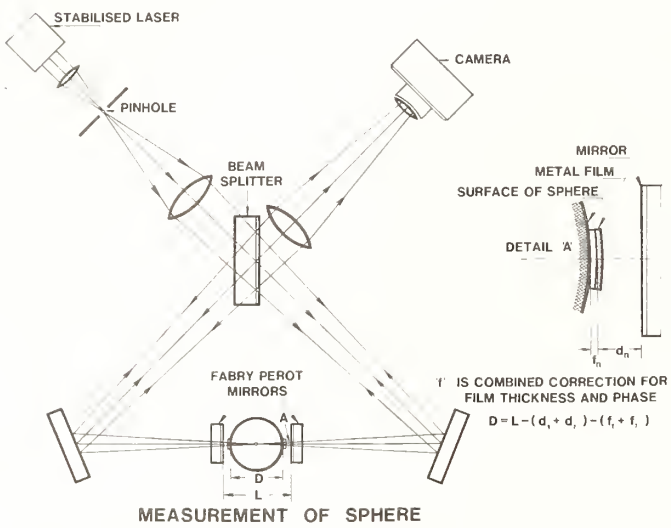


FIGURE 2. *Measurement of sphere.*

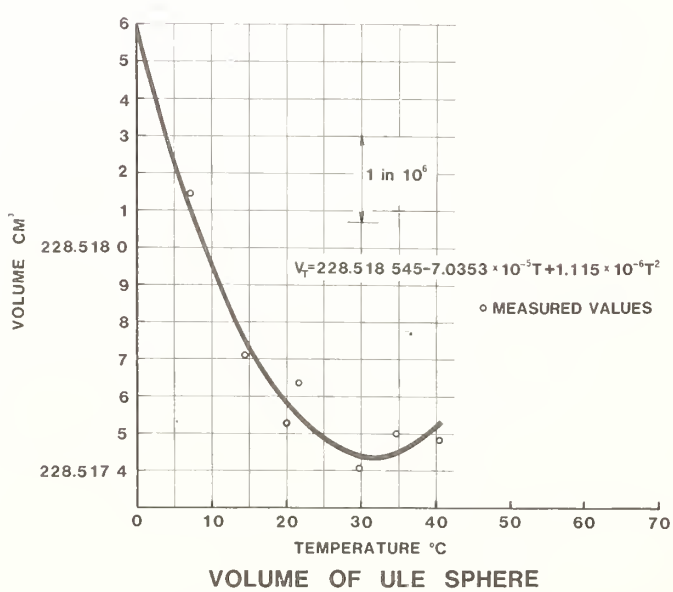


FIGURE 3. *Volume of ULE sphere.*

4. Measurement of the Mass of the Sphere

The mass of the sphere was measured on a 1 kg capacity balance of which the measurement uncertainty is known to be not greater than 50 μg. The mass was measured using the Gauss double weighing procedure with standards known to within ±1 part in 10^7 or ±5 μg. The sphere was cleaned with AR grade tri-chloro ethane and allowed to stand in air for some hours prior to the weighing. The air buoyancy correction amounts to approximately 230 mg and was determined by measuring the temperature, pressure, and relative humidity of the air, to calculate its density [8].

The initial value of the mass obtained in 1975 was 329.616 5 g. Subsequent measurements showed a slow increase in mass but the value appears to have stabilized at a value of 329.617 80 g at which it has remained constant over the past 3 years.

5. The Balance

The balance used for the hydrostatic weighings was a Mettler Model H51 having a maximum load of 160 grams and a readability of 10 μg. The minimum increment in mass attainable with the mechanically loaded weights as supplied was 100 mg, corresponding to a change in the depth of immersion of the suspension of several millimeters.

To reduce to a minimum the change in the depth of immersion of the suspension between weighings of the loaded and unloaded cage the balance was modified to incorporate an additional two decades (1-10 mg) and (10-100 mg) of mechanically loaded weights. By suitable choice of the additional weights the balance was tared to produce the same rest point (within 1 mg) for both conditions of loading.

The balance was calibrated over its full range using mass standards of which the values were known with an uncertainty of ±1 in 10^7 or ±5 μg. It is estimated that the uncertainty associated with the corrected value of any mass reading does not exceed ±5 in 10^7 or ±20 μg whichever is the greater.

6. Temperature Measurement

The sample of water being measured was contained in either of two glass vessels, one in which the water was in equilibrium with the ambient air, and the other from

which the air could be extracted. The vessel used was immersed in a 50 liter tank of water, the temperature of which was contolled by an electronic proportional controller with a resistive temperature sensor. The water in the outer tank was mixed with a propeller type mechanical stirrer. The temperature of the water sample is measured with a platinum resistance thermometer and an ac resistance bridge designed by Thompson [9]. The thermometer was calibrated in terms of the IPTS and regular measurements were made of the resistance at the triple point of water. The temperature of the water sample in the inner vessel remained constant within ±1 mK over a period of several hours corresponding to a change in the density of water of 2×10^{-4} kg/m^3 at 20 °C and 4×10^{-4} kg/m^3 at 40 °C.

7. Hydrostatic Weighing Procedure

The vessel containing the water sample was placed in the large water tank, the temperature of which had been set at the required value. After a suitable time for stabilization of temperature, the balance readings were made in the following sequence:

1. Balance scale reading with the cage unloaded, 2 readings S_1, S_2.
2. Balance scale reading with the sphere loaded on the cage, 4 readings S_3, S_4, S_5, S_6.
3. Balance scale reading with the cage unloaded, 2 readings S_7, S_8.
4. Reading of the water temperature with the IPTS thermometer.

The weight of the sphere in water (W) is then calculated using the equation

$$W = (W_1 - W_o + m) + (S_3 + S_4 + S_5 + S_6)/4 - (S_1 + S_2 + S_7 + S_8)/4$$

where

W_1 = tare weight on balance for sphere, cage, and suspension

W_o = tare weight on balance for cage and suspension

m = increment weight required to equalize the balance scale readings for both load conditions.

The next set of readings was taken using the values S_7, S_8 as new values for S_1, S_2 so that any tendency for zero drift of the balance was compensated. At least nine sets of readings were obtained at each temperature, during which three sets of readings were made of the temperature, pressure, and humidity of the air.

8. Results

The density ρ_T at a temperature T °C was calculated using the equations

$$\rho_T = [M - W(1 - \sigma/D)(g_W/g_s)]/V_T ,$$

where

M is the mass of the sphere and V_T its volume at T °C,

W is the observed weight of the sphere in water,

σ is the air density and D the density of the weights,

g_s is the gravitational acceleration at the level of the sphere,

g_W is the gravitational acceleration at the balance weight level,

ρ_T is the measured density of water at temperature T °C,

and

$$\rho_{T_o} = \rho_T[1 + \alpha_T(T - T_o)][1 + \beta_T(P_o - P)] ,$$

where

T_o is the reference temperature and $(T - T_o) \not> 0.01$°C,

α_T is the thermal dilatation of water at temperature T,

β_T is the compressibility of water at temperature T,

P_o is the reference pressure (101325 Pa), and

P is the measured pressure (Pa).

Measurements were carried out on various samples of water at different temperatures, either in equilibrium with the air, or air free, and of different measured isotopic compositions.

After adjusting the absolute values to SMOW (standard mean ocean water) values by using the data of Menache [10], the results below were obtained.

The temperature at which the density of water is a maximum under a pressure of 101 325 Pa (760 mm Hg) is

$$3.989 \pm 0.004 \text{ °C}.$$

The density of SMOW at the temperature of maximum density and a pressure of 101 325 Pa is

$$999.974\ 5 \pm 0.0010 \text{ kg/m}^3 .$$

The equation for the density of SMOW in kg/m^3 as a function of temperature at 101 325 Pa is

$$\rho_t = 999.974\ 5 - \frac{(t - 3.989)^2(t + 335.1596)}{558.7666(t + 72.2553)} .$$

It is estimated that in the temperature range 0° to 40 °C the uncertainty ℓ_t associated with the value of ρ_t is given by

$$\ell_t = \pm(0.001 + 0.000\ 03 \times |t - 4|) .$$

These results are quoted as provisional results and the uncertainties quoted for the density of SMOW at its temperature of maximum density and for the value at temperature t make allowance for the possibility of a systematic error in the diameter of the sphere. A new determination of the diameter of the sphere is in progress and the results will be reviewed when this work is completed.

References

[1] C. E. Guillaume, Bur. Int. des Poids et Mesures Trav. et Mém. 14, 1 (1910).

[2] M. P. Chappuis, Bur. Int. des Poids et Mesures Trav. et Mém. 13, D1 (1907).

[3] M. Thiesen, Physik Tech. Reichs, Wiss. Abhandl 3.1; 4.1 (1900).

[4] L. W. Tilton and J. R. Taylor, J. Res. Natl. Bur. Stand. 18, 205 (1937).

[5] G. S. Kell, J. Chem. Eng. Data 20, 97 (1975).

[6] G. A. Bell and A. L. Clarke, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Pub. Corp., New York, 1976), p. 615.

[7] J. B. Saunders, J. Res. Natl. Bur. Stand. Sect. C: 76, 11 (1972).

[8] F. E. Jones, J. Res. Natl. Bur. Stand. 83, 419 (1978).

[9] A. M. Thompson and G. W. Small, Proc. IEE 118, 1662 (1971).

[10] M. Menache, C. Beauverger, and G. Girard, Annales Hydrographiques 6, 41 (1978).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Precision Measurements on Solid Artifacts for a Redetermination of the Density of Water

A. Peuto, A. Sacconi, R. Panciera, W. Pasin, and M. Rasetti

Istituto di Metrologia G. Colonnetti, Strada delle Cacce,73 -10135 Torino, Italy

Mass and volume of four spheres made of low thermal expansion ($\alpha \simeq 10^{-8}$/°C) glass-ceramic were determined. Mass was measured against stainless steel standards, with $\pm 2 \times 10^{-7}$ relative uncertainty. The volume of two of the spheres was obtained through correlation of measured diameters and roundness data. Diameters were measured with a two-step interferometric method. Roundness data were taken on nine sections 20° apart. Estimated uncertainty of volume values is $\pm 1 \times 10^{-6}$. The volume of all four spheres, at 20 °C, was also measured by hydrostatic weighing in water with $\pm 3 \times 10^{-6}$ relative uncertainty. The comparison between volumes determined through both methods yields results in agreement within 1.5×10^{-6}. Volume ratios agree within 4×10^{-7}.

Key words: density of water; density standard; hydrostatic weighing; interferometric measurements; spherical volume; volume measurements.

1. Introduction

The inadequacy of water as a density reference standard has stimulated research work in order to establish a solid density standard [1–3]. Characterizing such a standard implies interferometric measurements of its linear dimensions, precise mass determination and hydrostatic weighing. The solid chosen should possess simple geometrical shape, mass dimensional stability, low coefficient of thermal expansion, and mass and density close to the reference standards. In order to find artifacts meeting the above, often contrasting, requirements, a compromise must be made. The material selected was Zerodur, a glass-ceramic with $\alpha \simeq 10^{-8}$/°C and inert in water and air. Four spheres of Zerodur were used (henceforth designated as SP, S1, S2 and S3) of 90 mm dia. and 1 kg mass each.

All data were taken at a temperature within ± 1.5 °C of 20 °C and no correction was made for thermal expansion of the spheres. The difference between the phase change on reflection from Zerodur and that from the etalon plates (quartz) was considered to be negligible. The lasers used were wavelength calibrated against an iodine ($^{127}I_2$) stabilized He-Ne laser with an uncertainty better than $\pm 1 \times 10^{-7}$. All individual uncertainties, expressed as standard deviations, are combined in quadrature, unless otherwise specified.

2. Volume Determinations through Dimensional Measurements

For dimensional measurements only SP and S2 were examined. In order to determine the diameter, interferometric techniques were used. Measurements were performed in two steps. First, diameters were measured by interferometric counting and mechanical contact (subsequently referred to as "first step" measurements) giving the exact integral part of the fringe order number. In the second stage the corresponding fringe fraction was obtained from single-wavelength interferometry. The initial estimate of the diameter was thus corrected to obtain a value with a smaller uncertainty, the correction depending on the initial estimate and the measured fractional fringe. Repeating the same procedure for different diameters, a mean diameter can be obtained and the volume determined. This method suffers from tedious alignment procedures and stability problems characteristic of a triangular interferometer. To avoid these and also to have a better knowledge of the shape contribution, a method resorting to roundness measurements was adopted in this work. Roundness measurements yield the variations in the diameter rather than its absolute value. Nevertheless it is sufficient to know a single diameter of the explored profile for obtaining a mean diameter representing that profile. Then from the measured diameter and the roundness data for a family of planes passing through this diameter, the average diameter representing the sphere can be obtained.

First step measurements were made using a Universal Measuring Machine (Moore M18) equipped with a laser interferometer (HP 5500). A linear variable differential transformer (LVDT) bidirectional transducer (CARY I-DIM, with a resolution of 0.01 μm) was used as a null detector. To limit systematic errors arising from the mechanical contact of the feeler, its constant (spherical tip dia. and bending of the stylus) was taken into account by measuring, in a symmetrical cycle, the length of a dummy gage (interferometrically calibrated gage block). The reproducibility in a set of 20 measurements was represented by a standard deviation $s = 0.02$ μm and the systematic uncertainties (due mainly to: the dummy gage-block calibration, geometrical and mechanical corrections, instrumental errors in pressure-temperature-humidity measurements, and air composition) by an $s = 0.03$ μm.

The fringe fractions were measured with an interferometer (Fig. 1) similar to that used by Saunders [4] but the fringes were scanned and detected with a rotating mirror and a PMT [5]. The interference ring diameters were then least-squares fitted to yield the fringe fraction. In the relation $D = E - (d_1 + d_2)$, given the integral part, I_D, of the order number of D, the fringe fraction, f_D, can be determined from the fractional fringes f_E, f_1, and f_2 only, as $f_D = f_E - F(f_1 + f_2)$,

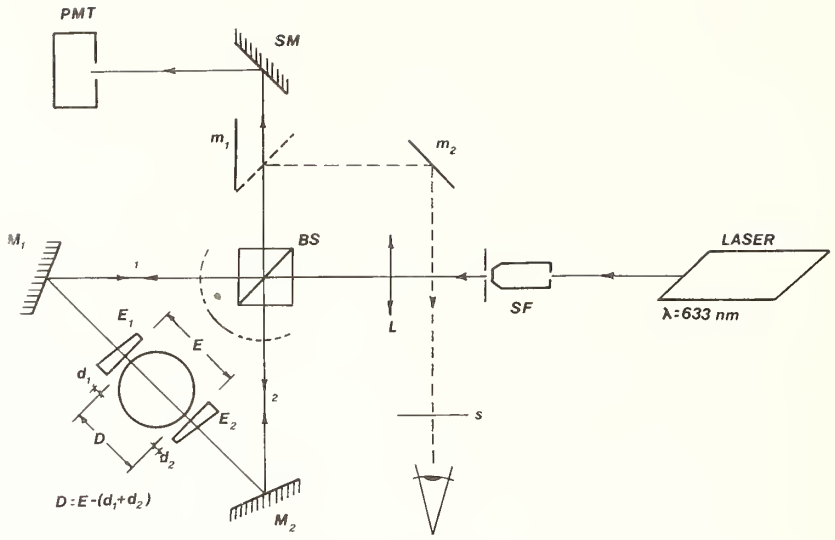


FIGURE 1. *Optical layout of the interferometer for fringe fraction measurements. PMT, photomultiplier tube; SM, scanning mirror; BS, beam splitter; M_1, M_2, mirrors; E, etalon length; D, diameter under examination; L, focusing lens; SF, spatial filter; s, screen; m_1, m_2, auxiliary mirrors.*

where F indicates the fractional part, the knowledge of the integer I_E being unnecessary.

Each series of measurements consisted of eight runs. A sequential scan in which beams 1 and 2 were interrupted in the order 2, 1, 1, 2 was adopted. After measuring f_1 and f_2, the sphere was removed and the same procedure repeated for the etalon alone. The f_E's were then reduced to the same environmental conditions as for f_1 and f_2 and a single D computed for each series using the relation $D \doteq (I_D + f_D)\lambda_a/2$, λ_a being the laser wavelength in ambient air. In order to randomize the systematic errors arising from the optical alignments of the interferometers and of the sphere, the system was aligned for each series. The results are summarized in Table 1. Estimated systematic uncertainty was ± 0.01 μm. Repeatability tests ($N = 20$) on f_E and on $(f_1 + f_2)$ yielded an s of 0.003 and 0.01 fringe fractions, respectively.

TABLE 1. *Results of the diameter measurements obtained from the two-step method.*

Sphere	Measured diameter D (μm)	Standard deviation s (μm)	Number of meas. N
SP	89 703.158	0.030	18
S2	90 389.326	0.026	9

It is worth mentioning that first step measurements show a systematic difference from these results of about 0.1 μm, which exceeds the previous estimated uncertainties. This discrepancy may be due to some unidentified source of error. However the root-sum-square combination of uncertainties might be inadequate for assessing the reproducibility of measurements performed by different methods.

For roundness measurements the previously mentioned Universal Machine was used. The sphere was rotated around the measured diameter in equal steps by means of a specially designed mechanism equipped with an incremental encoder. A set of nine profiles, spaced 20° apart, was investigated: Fig. 2 is a typical profile with a radial magnification of 100 000. Eccentricity terms were calculated and subtracted from the transducer signal using the limacon approximation as criterion for best fit [6]. Each profile was tested for eight revolutions and time averaged to smooth accidental errors. Time sampling of the profile was taken over 625 angular points per revolution. Analysis of the data yielded the mean diameter for each section. A reproducibility test carried out for eight measurements of eight runs each, resulted in a standard deviation of 0.01 μm for a single measurement of the mean diameter, the std. dev. for maximum diameter variations (ΔD) being 0.006 μm. Systematic uncertainty is due mainly to the spindle errors which, when evaluated by the reversal method [7], had an s of 0.014 μm on ΔD evaluation.

The average diameter D_{av}, i.e., mean diameter averaged over the nine profiles, is given by:

$$(D_{av} - D_{meas})_{SP} = +0.054 \text{ μm}$$
$$(D_{av} - D_{meas})_{S2} = -0.184 \text{ μm}.$$

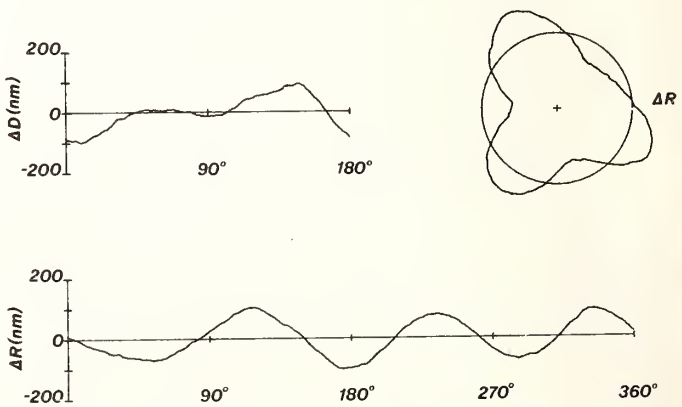


FIGURE 2. *A typical roundness profile. ΔR and ΔD are maximum radial and diametral variations respectively. Both polar and cartesian plots are given.*

From these relations and Table 1, the average diameters of SP and S2 are 89 703.212 μm and 90 389.142 μm, respectively. Random uncertainty of each D_{av} is given by $s = 0.03$ μm; the systematic uncertainty is given by $s = 0.02$ μm. The corresponding volumes are given in Table 3.

It is worth mentioning that the standard deviations of the distributions of the nine profiles giving the mean values ($D_{av} - D_{meas}$) are unequal for SP and S2: $s_{SP} = 0.02$ μm and $s_{S2} = 0.05$ μm. This indicates different geometrical quality (i.e., deviation from perfect spherical shape) of the two spheres and can be compared with the max. diametral variation (0.25 μm and 0.8 μm, respectively) measured on single profiles.

As a further check of the method outlined above, a second family of nine profiles around another measured diameter was investigated for the same two spheres. This should yield the same average diameter as that obtained previously. Moreover, from the roundness measurements the difference between the two measured diameters (ΔD_R) can be obtained and compared with the difference measured directly (ΔD_M). For the latter we had only the data from first step measurements, as only one diameter of each sphere was measured with the triangular interferometer. This however poses no problem as only the differences are of interest here. This comparison is substantially unaffected by any differences in the geometrical quality of the spheres. The maximum deviation ($\Delta D_R - \Delta D_M$) of 0.03 μm obtained in our case indicates a good agreement between the roundness and direct measurements in addition to the validity of the former.

3. Mass Measurements and Hydrostatic Weighings

The mass of the sphere SP was 959 g, that of S1, S2, and S3 being 979 g each. The masses were measured by comparison against a 1 kg stainless steel standard, calibrated with an uncertainty of ± 100 μg. For this comparison a polished 20 g Zerodur support (calibrated within ± 20 μg) was used for the spheres. Stainless steel standards were then added to equalize loads. A single-pan damped balance of 1 kg max. capacity was used. Its sensitivity, approximately 10 div/mg, was determined at each weighing. Care was taken so that the temperature in the balance case did not vary by more than 10 mK during weighings. The standard deviation over one weighing cycle was typically 50 μg.

The air buoyancy effect on the spheres (density 2 500 kg · m^{-3}) differs from the effect on the 1 kg standard by about 300 mg. Air density thus becomes critical; it was evaluated using the relation given in Ref. [8]. The uncertainty of the air buoyancy correction is estimated[1] to be ± 180 μg; the overall uncertainty of the mass measurements amounts to 200 μg.

The volumes and the ratios between volumes of all four spheres were determined by hydrostatic weighing, taking water as a reference standard. The balance was identical to the one used for weighings in air, modified so that the load could be suspended under the pan. A stainless steel device allowed three spheres to be weighed in the same water bath. It could rotate and move vertically, letting the sphere under test rest on a hook-shaped suspension attached to the balance. The suspension wire (dia. = 0.25 mm) was made of stainless steel covered with platinum black. The water was de-ionized and bi-distilled in a quartz apparatus and subsequently degassed at low pressure. Tests for conductivity and dissolved oxygen were carried out before and after each series of weighings. For volume computation (not for ratio determinations) those weighings conducted in water of conductivity greater than 0.3 mS · m^{-1} were disregarded. The temperature of the bath was measured with two platinum resistance thermometers and an ac double bridge (Automatic Systems Lab., 7 decades). Bigg's dilatation tables [9] were adopted for the evaluation of the density of water. Corrections for isotopic composition [10] and dissolved gases were applied according to the relations of Ref. [11].

Four series of weighings, each involving three spheres, were performed, following different weighing sequences. Each sphere was usually weighed eight times per series. The standard deviation over one series was typically 150 μg (4×10^{-7} on volumes). The masses and volumes of the four spheres are summarized in Table 2. Typical reproducibility ($N = 20$) is represented by an $s = 180$ μg (5×10^{-7}). The systematic uncertainty of the volume measurements was estimated to be ± 1.3 mm^3 (3.3×10^{-6}).

TABLE 2. *Results of mass measurements and hydrostatic weighings.*

Sphere	Mass (g)	N	s (μg)	Volume (cm^3)	N	s (mm^3)
SP	958.567 66	8	100	377.939 22	15	0.16
S1	979.351 39	8	150	386.521 87	10	0.20
S2	979.682 15	7	80	386.675 59	19	0.07
S3	978.886 91	8	140	386.356 51	19	0.18

Volume ratios obtained from weighing sequences S_i, S_j, S_i in the same water bath do not differ from ratios computed from average volumes by more than 4×10^{-7}.

The comparison between volumes obtained through dimensional measurements and hydrostatic weighings is shown in Table 3. The agreement of the results lies well within the uncertainty of the hydrostatic weighings, the difference being 1.5×10^{-6}. The ratios V_{SP}/V_{S2} obtained by the two methods differ by only 3.5×10^{-7}, account being taken only of those volume ratios obtained from direct hydrostatic comparisons, i.e., from comparisons in the same bath.

TABLE 3. *Comparison of volumes obtained from diameter measurements with those determined by hydrostatic weighings.*

Sphere	Volume from dimensional meas. V_D (cm^3)	Difference from V_H (cm^3)	$\frac{V_D - V_H}{V_H}$
SP	377.939 78	+ 0.000 56	$+ 1.5 \times 10^{-6}$
S2	386.676 18	+ 0.000 59	$+ 1.5 \times 10^{-6}$

From the mass values (Table 2) and dimensionally determined volumes (Table 3) the densities of SP and S2 are found to be 2536.301 kg · m^{-3} and 2533.750 kg · m^{-3}, respectively. The difference between these values is not surprising since the spheres come from different lots of

[1] To check this estimate of air buoyancy uncertainty, several weighings were performed using different instruments for measuring environmental variables. The experimental standard deviation was found to be 150 μg.

Zerodur. It should be noted that the aim of this experiment was not to pursue a "natural" standard. In our case only the objects and not the material were to be characterized as standards. The uncertainty of the density values is $\pm 3 \times 10^{-3}$ kg · m^{-3} (1.1×10^{-6}).

4. Conclusions

The reliability of a roundness-based representation of the volume of the spheres was demonstrated experimentally. The intercomparison between dimensionally and hydrostatically determined volumes was shown to yield results in quite good reciprocal agreement. Since different measurement techniques were involved in this intercomparison, it can be concluded that (within $\pm 1 \times 10^{-6}$ limits) there exist no significant systematic errors in the adopted methods. Furthermore, the excellent agreement in volume ratios between the two methods assures the feasibility of a dissemination program based on hydrostatic weighings of a series of secondary density standards.

References

[1] H. A. Bowman, R. M. Schoonover, and C. L. Carrol, J. Res. Natl. Bur. Stand. Sect. A 78A, 13 (1974).
[2] G. A. Bell and A. L. Clarke, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, New York, 1976), p. 615.
[3] K. Iizuka *et al.* in *Acta IMEKO 1979*, Ed. by Gy. Striker (Akademiai Kiado, Budapest, 1980), Vol. 3, p. 827.
[4] J. B. Saunders, J. Res. Natl. Bur. Stand. Sect. C **76C**, 11 (1972).
[5] A. Sacconi and S. Fontana, Metrologia **11**, 33 (1975).
[6] D. J. Whitehouse, J. Phys. E. **6**, 921 (1973).
[7] R. R. Donaldson, CIRP Ann. **21**, 125 (1972).
[8] BIPM Report 79/10, Sèvres (1979).
[9] P. H. Bigg, Br. J. Appl. Phys. **18**, 521 (1967).
[10] G. Guzzi, Centro Comune di Ricerca, Euratom, Ispra, private communication.
[11] M. M. Menaché, M. C. Beauverger, and M. G. Girard, Ann. Hydrogr. **6** (3), 37 (1978).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Influence of Dissolved Air on the Density of Water

Georges Girard and Marie-Josette Coarasa

Bureau International des Poids et Mesures, F–92310 Sèvres, France

A knowledge of the density of water to an accuracy of 1×10^{-3} kg·m^{-3} (1 part in 10^6) has been of metrological importance for a long time. The basic work on the density of water dates from the beginning of the century and at the time included the absolute value itself together with the variations of density as a function of temperature and of amount of dissolved air.

In recent years a number of laboratories, BIPM among them, have once again taken up this work. At BIPM, studies have been undertaken of the influence of variations in isotopic composition on the density. A relation has been derived which is valid for all samples of water likely to be encountered in any of the various national metrological laboratories. Recent work at BIPM has been devoted to the question of the effect of dissolved air on the density. The density of samples having different levels of saturation of dissolved air has been measured at various temperatures between 4 and 22 °C. The levels of saturation of dissolved oxygen were within the range 0.1 to 0.95.

Key words: atmospheric dissolved gases; compressibility; density of water; fundamental value; hydrostatic weighing; isotopic composition; thermal expansion.

1. Introduction

Pure water is used for precision measurements of density and volume. The knowledge of its density depends upon a number of factors, the majority of which are not known with sufficient precision:

(a) the fundamental value: the accepted value, 999.972 kg·m^{-3}, is the density of air-free water at 4 °C under a pressure of 101 325 Pa, determined between 1895 and 1905, a time when the existence of isotopes was not suspected;

(b) the thermal expansion coefficient: the available data between 0 °C and 40 °C are not consistent, particularly above 16 °C;

(c) dissolved atmospheric gases: their influence is not sufficiently well known;

(d) isotopic composition: this factor was not studied until recently;

(e) compressibility: the only factor that is known, around a pressure of one atmosphere, with sufficient precision.

2. Redetermination of the Density

In 1967, Menaché drew attention to the fact that for modern requirements, particularly in physical oceanography, there was a need for a better knowledge of the density of water. At that time it was not known to better than 0.01 kg·m^{-3} due to the cumulation of uncertainties stemming from different parameters.

In 1969, the Joint Panel on Oceanographic Tables and Standards (sponsored by the International Association for the Physical Sciences of the Ocean (IAPSO), the International Council for the Exploration of the Sea (ICES), UNESCO, and the Scientific Committee on Oceanic Research (SCOR)), prepared a Recommendation that was adopted by IAPSO. The International Union for Geodesy and Geophysics (IUGG) made a recommendation for a new study of the density of water, of which the text was passed to the Comité International des Poids et Mesures (CIPM).

In parallel, Commission 1.4 (Commission on Physicochemical Measurements and Standards) of the International Union of Pure and Applied Chemistry (IUPAC) had been looking at the problem since 1965. In the course of successive meetings, in particular those of 1969 and 1971, this Commission accepted concrete proposals from J. Terrien, at that time Director of BIPM. In 1973 it prepared a Recommendation on the re-determination of the density of water and the study of the various determining factors, which was sent to a certain number of national standards and oceanographic laboratories. In about 1972 preliminary work began in various laboratories and in 1975 IUGG approved the Recommendation of IUPAC.

2.1 Fundamental Value—Thermal Expansion Coefficient

At present, work on the fundamental value and thermal expansion coefficient is in progress in Australia, Italy, Japan and the Federal Republic of Germany.

Briefly, it consists of a determination of the volume of a body by means of geometrical measurements, weighing the body in air, and then, by hydrostatic weighing, determining the density of the sample of water in which it is immersed. If these hydrostatic measurements are repeated (using the same sample of water) at different temperatures, the thermal expansion coefficient can be obtained. This same parameter can also be studied by using a solid body having a very well known thermal expansion coefficient, and carrying out hydrostatic weighings at a number of temperatures in the chosen range.

2.2 Compressibility

For the compressibility of water we already have an equation that is perfectly satisfactory [1].

2.3 Isotopic Composition

The influence of the isotopic composition on the density has been studied theoretically by Menaché who, in 1967,

published a formula relating the density to its content of oxygen-18 and deuterium. A later, more complete equation was published in 1971 [2].

The isotopic composition of ^{18}O, ^{17}O and D of a sample is compared with that of a reference water which is taken, by the specialists, as a standard of isotopic composition: Standard Mean Ocean Water, now called Vienna-SMOW (V-SMOW). With respect to V-SMOW, the relative departures of the density, δ_{18} and δ_D, for a given sample of water are defined by the following expressions:

$$\delta_{18} = \frac{\frac{[^{18}O]}{[^{16}O]}\,(\text{sample}) - \frac{[^{18}O]}{[^{16}O]}\,(\text{V–SMOW})}{\frac{[^{18}O]}{[^{16}O]}\,(\text{V–SMOW})} \cdot 10^3 \quad (1)$$

$$\delta_D = \frac{\frac{[D]}{[H]}\,(\text{sample}) - \frac{[D]}{[H]}\,(\text{V–SMOW})}{\frac{[D]}{[H]}\,(\text{V–SMOW})} \cdot 10^3 \quad (2)$$

where $[^{18}O]$, $[^{16}O]$, [D] and [H] are the concentrations (amount of substance divided by volume) of the different nuclides present in the samples.

The effect of the isotopic composition on the density of natural waters is given by the following formula, valid between 0 °C and 30 °C:

$$(\rho_{\text{sample}} - \rho_{\text{V–SMOW}}) \times 10^3/(\text{kg·m}^{-3}) = 0.233\ \delta_{18} + 0.016\ 6\ \delta_D \pm 0.1 \quad (3)$$

where δ_{18} and δ_D are defined above in Eqs. (1) and (2).

This theoretical formula has been confirmed by experimental work at BIPM, the preliminary results of which were presented at the IUGG General Assembly in Grenoble in 1975. The measurements were made by weighing a body, of known thermal expansion coefficient, in the various samples of water of known isotopic composition, over a range of temperatures and pressures. From these measurements the buoyancy and hence the density of the water may be deduced.

The samples of water used in these measurements came from: (1) melted Antarctic ice (one of the lightest of the natural waters); (2) the Mediterranean; (3) BIPM tap water. After double distillation, the difference between the maximum and minimum densities was found to be 19×10^{-3} kg·m^{-3}.

3. Influence of Dissolved Atmospheric Gases

Marek [3], in 1891, was the first person to attempt to measure the difference in density between a sample of air-free water and the same water saturated with air at the same temperature. Between 0 and 14 °C he found that this difference was of the order of 3×10^{-3} kg·m^{-3} and diminished at higher temperatures. Chappuis [4], in 1910, made similar measurements between 5 and 8 °C

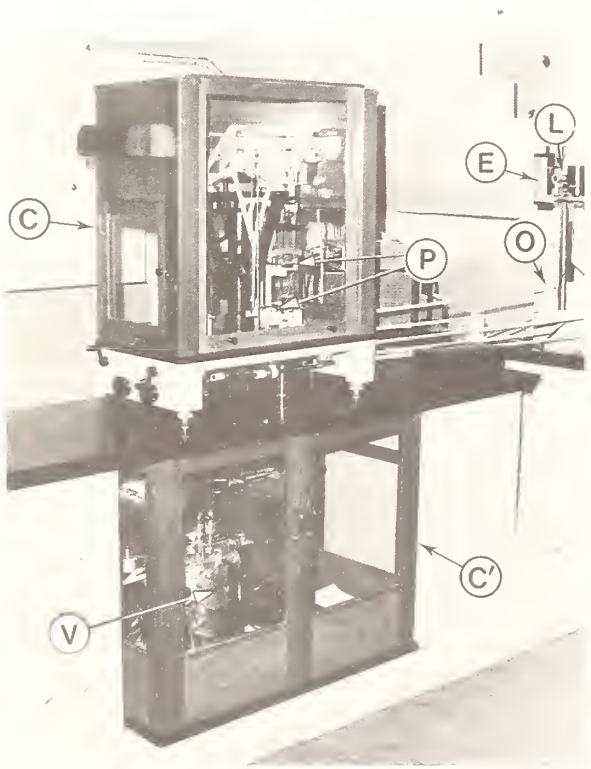


FIGURE 1. *The BIPM hydrostatic balance. C, upper enclosure containing the balance; C', lower enclosure containing the hydrostatic bath V; P, two stage pan; E, scale; L, projector; O, control console.*

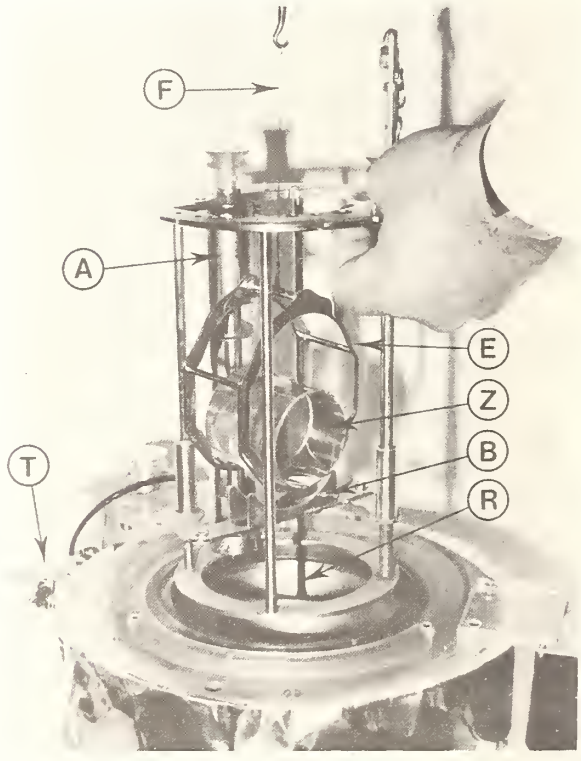


FIGURE 2. *Detail of the hydrostatic bath and the suspension and measurement mechanism of the cylinder in the water sample. Z, cylinder of Zerodur; E, stirrup; F, stirrup suspension wire; B, cradle; A, stirrer; T, entry of temperature-controlled water; R, recipient.*

and obtained results very close to those of Marek.

The difficulties with Marek's results are that: first, they refer to air-free and saturated states that were almost certainly far from being realized; second, they provide only the difference in density between two extreme conditions; and third, they imply the notion of "dissolved air" that was not sufficiently well defined.

We have therefore repeated the work of Marek, but taking all the precautions, and using the equipment (Figs. 1 and 2), that was developed during our work on the variation of density with isotopic content. The method is similar: the variation in the buoyancy is measured for a body totally immersed in water for which the quantity of dissolved atmospheric gas is varied. These measurements were repeated at various temperatures. The quantity of dissolved gas was determined from samples taken from the water just after each hydrostatic weighing.

3.1 Hydrostatic Measurement of the Density of a Sample of Water

Figure 3 shows the two phases, A and B, of a hydrostatic weighing. In Fig. 3(a), the tare mass t balances the standard of mass m_2 and the ensemble of the stirrup and suspension wire of mass K. In Fig. 3(b), the equilibrium with the tare mass is re-established by the body of mass m_1, replacing the standard of mass m_2, the immersed cylinder of Zerodur (mass m, volume V, density ρ), now being suspended on the pan of the balance.

The equations at equilibrium are as follows:

$$t = m_2(1 - a/\lambda) + K \tag{4}$$

$$t = m - V\rho + m_1(1 - a/\lambda) + K \tag{5}$$

from which

$$m - V\rho = (m_2 - m_1)(1 - a/\lambda) \tag{6}$$

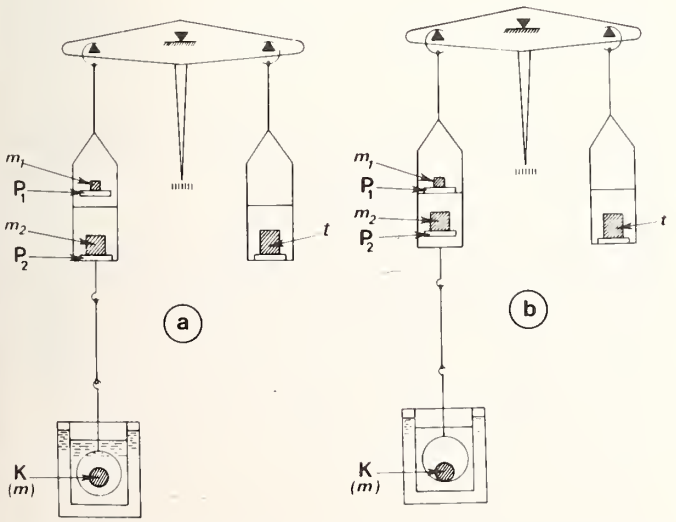


FIGURE 3. *Schematic showing the use of the hydrostatic balance. The masses of the auxiliary pans P_1 and P_2 are eliminated by a second weighing in which masses m_1 and m_2 are associated respectively with P_2 and P_1.*

hence

$$\rho = [m - (m_2 - m_1)(1 - a/\lambda)]/V \tag{7}$$

where a is the density of air and λ that of the bodies of mass m_1 and m_2. A complete weighing comprises three weighings under conditions B and two under conditions A. The temperature of the water, at the level of the Zerodur cylinder, is measured by means of a platinum resistance thermometer.

Particular care is taken in the degassing of the water. This is carried out by maintaining the doubly-distilled water for three hours at a temperature of about 68 °C under the corresponding partial pressure, about 28 kPa, so as to eliminate all traces of air. In preparing for the weighing, with the Zerodur cylinder and all its supports in place, the pressure in the water bath is reduced to a few hundred pascals so that the air-free water can be transferred to the water bath without air being added. This low pressure is subsequently maintained for many hours in order to remove all traces of air from the walls. The following morning, after having let air into the bath, the absence of air bubbles is checked, a process made easy by the transparency of Zerodur. The first weighings are made as soon as the temperature is stabilized (i.e., two hours later) with only a very small amount of air dissolved in the water. The measurements are then pursued for a number of days, the quantity of dissolved air slowly increasing, either naturally or with external assistance.

3.2 The Measurement of the Quantity of Dissolved Air

In a study of the solubility of the principal atmospheric gases in sea water (pure water can be considered simply as sea water having a zero salinity), Menaché *et al.* [5] show that for the four atmospheric gases, O_2 N_2, Ar, and CO_2, argon and carbon dioxide have very low solubilities for which the sum does not exceed 4% by volume of the solubilities of the four together.

In a parallel study on the rate of solution of the same gases, made under conditions similar to those existing in our hydrostatic weighings, Menaché *et al.* [5] show that between 4 and 22 °C the levels of saturation $x(N_2)$ and $x(O_2)$, as a function of time of exposure to air, are practically identical. We can thus write, with a good accuracy:

$$x(N_2) = x(O_2) \tag{8}$$

from which:

$$\frac{C(N_2)}{C(O_2)} = \frac{s(N_2)}{s(O_2)}, \tag{9}$$

where for a given sample $C(N_2)$ is the content of nitrogen and $s(N_2)$ is the solubility of nitrogen or the content of nitrogen corresponding to the partial pressure $p(N_2)$ of the gas at the time of the experiment.

It is therefore possible, as in the present case when the sample is not saturated, to limit the analysis to a chemical titration for oxygen. The level of saturation of oxygen alone may then be calculated, which will represent the level of saturation of the sample for dissolved atmospheric gases. The chemical titration of oxygen is carried out by the Winkler method as improved by Carpenter [6]. This titration is conveniently made with the 50 cm^3 samples taken at the end of each weighing. It was possible to deduce that when air was let into the hydrostatic bath, the level of saturation was of the order of 0.03.

4. Results[1]

From the analysis it was possible to associate with each hydrostatic weighing, which gave a relative value for the density of water, a value for the quantity of dissolved atmospheric gas. Measurements were made at temperatures of 4, 10, 16, and 22 °C. Account was taken of the variation in the isotopic content of the samples during the course of the weighings. Figures 4 and 5 give examples of the result obtained at 4 and 22 °C. One can deduce that the variation of density is proportional to the quantity of dissolved gas at all temperatures studied.

Figure 6 shows the variation of the difference in density between air-free and fully saturated water as a function of temperature. It is planned to carry out further measurements at other temperatures. The values of Marek and the mean value of Chappuis are also given in the Figure.

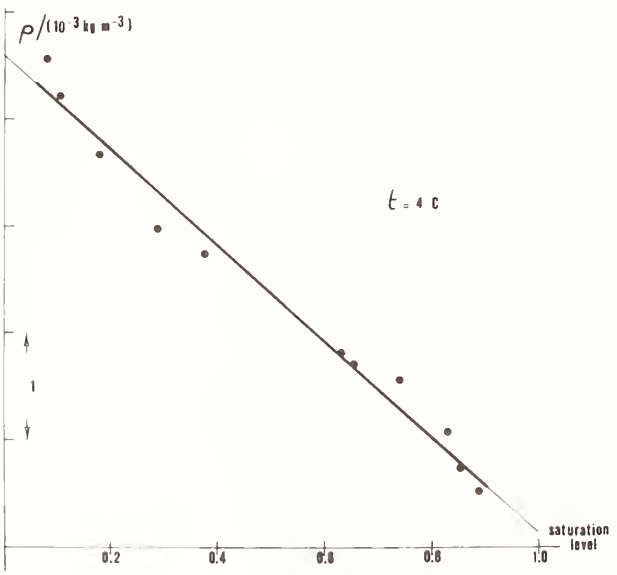


FIGURE 4. *Variation, at 4 °C, of the density of a sample of water as a function of dissolved atmospheric gas.*

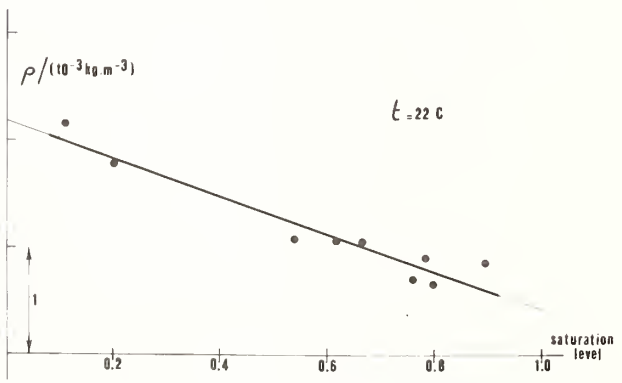


FIGURE 5. *Variation, at 22 °C, of the density of a sample of water as a function of dissolved atmospheric gas.*

[1]Note added in proof: In view of certain defects subsequently found in the platinum resistance thermometer used to measure the temperature of the water in the hydrostatic bath, the results given here at 22 °C should have twice the uncertainty given in the text.

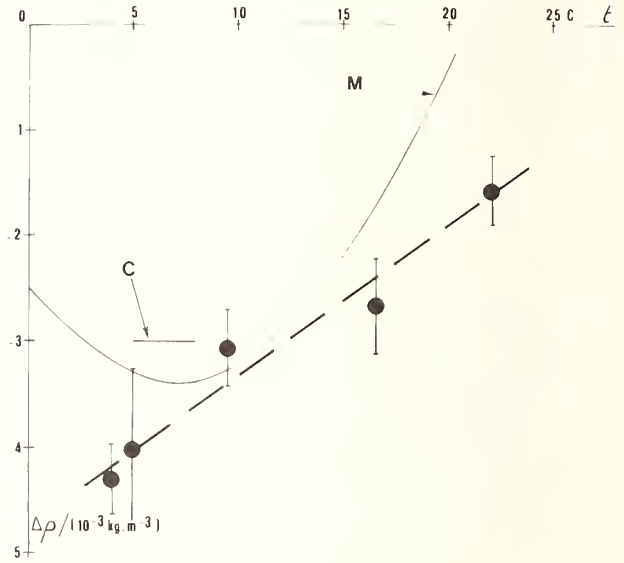


FIGURE 6. *Variation, as a function of temperature, of the difference in density between air-free and fully saturated water. •, this work (mean value with uncertainty shown at the 3 σ level); M, curve given by Marek; C, mean value of Chappuis.*

We would like to thank M. Menaché for helpful advice.

References

[1] G. S. Kell, J. Chem. Eng. Data 15, 119 (1970).
[2] G. Girard and M. Menaché, Metrologia 7, 83 (1971).
[3] W. J. Marek, Ann. Phys. Chem. 44, 171 (1891).
[4] P. Chappuis, Trav. Mem. Bur. Int. Poids Mes., 14, D1 (1910).
[5] M. Menaché, C. Beauverger, and G. Girard, Ann. Hydrog. 6, 37 (1978).
[6] J. H. Carpenter, Limnol. Oceanogr. 10, 135 (1965).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Helium Melting Curve and the Linkage of Fundamental Constants, Pressure, Density, and Mass

Craig T. Van Degrift

Temperature and Pressure Measurements and Standards Division, National Bureau of Standards, Washington, DC 20234

Recent progress made toward the calculation of the ground states of liquid and solid helium from first principles suggests that a future pressure standard might be based on the calculated phase diagrams of ^{3}He and ^{4}He. The absolute reference pressures of the zero temperature intercepts and minima of the melting curves of these systems would be transfered to room temperature gages and provide a new connection between fundamental constants and density or mass. Already, present experiments suggest that the helium melting curves can provide pressure reference points which are precise within 1 ppm. A summary of the current state of theoretical work is given followed by a discussion of the practical difficulties in realizing the linkages between the melting curve and room temperature pressure, mass, and density.

Key words: density; fundamental constants; helium; many-body theory; mass; pressure; quantum fluids; quantum solids.

1. Introduction

The helium melting curve is uniquely well-suited to be used as a set of pressure reference points. Figure 1 shows the region below 1.5 K of the melting curves of the two stable isotopes of helium—^{3}He and ^{4}He. Both curves exhibit minima (B and D) and have non-zero intercepts (A and C) at zero temperature. For a given isotope, the region above the curve is solid and the region just below is liquid. Not shown in the figure are the superfluid-ordinary fluid and liquid-vapor curves. There is no ordinary triple point, as the liquid-vapor equilibrium curve never intersects the melting curve. The liquid-solid curves are so easily measured by low temperature physicists that the region between A and B on the ^{3}He curve is now used for thermometry [1]. Pressure transducers, necessary for these measurements, work particularly well at low temperatures because mechanical creep, thermal expansivity, and thermal relaxation times of materials are generally many orders of magnitude smaller than at room temperature.

In this paper, we discuss the key linkages which must be mastered to fully utilize the potential standards applications of these melting curves. We first discuss the most difficult linkage—the theoretical calculations of these pressure points from fundamental constants. We then describe the relatively easy task of transferring the melting curve pressures to a room temperature piston gage. Finally, we briefly summarize the present limitations in manometry and piston gage technology and hence in the linkage of pressure to density or mass.

It is perhaps optimistic to expect to achieve 1 ppm accuracy in all linkages by the turn of the century; theoretical calculations of helium melting pressures are currently in error by as much as 7% but would contribute to pressure standard work only when performed at the 1–5 ppm level. However, the transfer between the melting curve and room temperature can be expected at the 1 ppm level with straightforward improvements in present technology. The linkage between pressure and density of mercury is expected to be done at NBS at the 2 ppm level within a year, while the linkage to mass may reach 2 ppm within a decade when foreseeable improvements are made in dimensional metrology and the theoretical modeling of piston gage physics. Thus, rather substantial improvement in the accuracy of theoretical calculations is needed to provide absolute helium pressure standards.

Even in the absence of an accurate theory, however, the linkage between the melting curve and room temperature piston gages can already be useful for transferring and maintaining the accuracy of pressure measurements. Although the best current piston gage calibration [2], performed at 0.1 MPa against the density of mercury, has an inaccuracy of 5 ppm, the best working standards are two to five times poorer in accuracy. At present, the helium melting curve and associated low temperature transducers can provide two reference pressures (points B and C in Fig. 1) with an overall precision of about 1 ppm. These

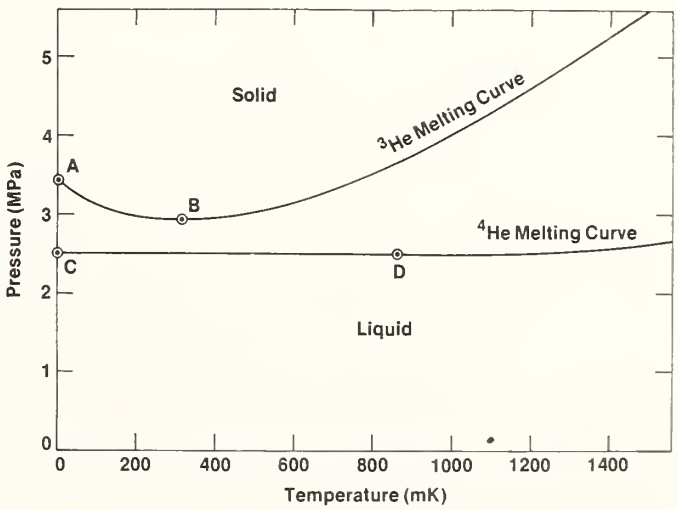


FIGURE 1. *The melting curves of ^{3}He and ^{4}He below 1.5 K. Points A and C are the zero temperature intercepts and points B and D are minima.*

points can thus be used immediately to maintain piston gage reproducibility at these pressures. Such a calibration method would help to compensate for the fact that piston gage precision is better by more than a factor of ten than its accuracy [3].

2. Linking Fundamental Constants to the Helium Melting Curves

The accurate calculation of the thermodynamic properties of a condensed substance from first principles should no longer be considered impossible. Helium-4, being a noble gas with no nuclear spin, may be the object of the first such calculation with the zero-temperature intercept of its melting curve serving as a convenient test point. The problem naturally separates into two similar many-body problems: (1) the calculation of inter-atomic potentials from the interaction of pairs or small clusters of alpha particles with each other and their electrons; and (2) the calculation of the thermodynamic properties of an ensemble of atoms interacting via those interatomic potentials.

The theoretical calculation at zero temperature will give the pressure of point C in Fig. 1 primarily in terms of the following equivalent combinations of microscopic constants

$$P \propto \frac{e^2}{\epsilon_0 a_0^4} \propto \frac{\alpha hc}{a_0^4} \propto hc \frac{R_\infty^4}{\alpha^3}, \qquad (1)$$

where e is the charge on an electron, ϵ_0 is the permittivity of free space, a_0 is the Bohr radius, h is Planck's constant, c is the speed of light, α is the fine structure constant, and R_∞ is the Rydberg.

The first expression in Eq. (1) may be readily recognized as the electrostatic energy per unit volume. The calculation also requires quantum mechanical corrections of 20% involving the ratio of zero point kinetic energy to the binding energy and thereby bringing in the electron-to-alpha particle mass ratio [4]. At elevated temperatures, evaluation of kT will be required with quantum corrections proportional to the ratio of the thermal wavelength, $h/\sqrt{2\pi MkT}$, divided by the cube root of the mean atomic volume, $\sim a_0$, where M is the atomic mass.

In recent years considerable progress has been made toward the formulation of a practical mathematical algorithm which is capable of accurately calculating thermodynamic properties [5]. The method is called the Green's function Monte Carlo (GFMC) method. For a given interaction potential, it is claimed to have an accuracy limited only by computational effort. The rapidity of convergence of the algorithm depends upon the choice of a function called the "importance function," which approximates the many-body wave function. An imperfect choice results in a slower convergence to the same final solution.

The determination of inter-atomic potentials is presently the weakest part of the calculation. The helium-helium potential traditionally used is the Lennard-Jones (12–6) potential with its adjustable coefficients fitted to the second virial coefficients of helium gas. Aziz *et al.* [6] have recently proposed an improved potential less dependent on experimental data. In calculating the second virial coefficients and thermal conductivity for helium gas, they achieved agreement nearly within experimental error. Potentials describing the interaction of more than two helium atoms, important in the condensed phases, are just beginning to receive attention.

Whitlock, Ceperley, Chester and Kalos [7] have used the GFMC method with a Lennard-Jones potential to calculate the pressure dependence of the energy, density and velocity of sound at zero temperature for liquid and solid ^{4}He. They calculated the melting pressure to be 70% above the experimental value. The subsequent use of the Aziz potential by Kalos, Lee, Whitlock, and Chester [8] improved the pressure calculation to just 7% above experiment. The results were very sensitive to the precise shape of the potential. Addition of an approximate three-body potential (Axilrod-Teller) did not improve the agreement. The determination of the proper interaction potential *ab initio* may require the direct application of the GFMC method to an ensemble of alpha particles and electrons.

Nevertheless, one can expect improved *ab initio* calculations of the melting pressure of ^{4}He. For a given importance function, the algorithm has an accuracy which increases with the square root of the computational effort. Sheer increase in available computational capacity over the next decade may yield a 100-fold improvement in accuracy [9]. Monte Carlo calculations are particularly well-suited for computers with parallel processors. Improvements in the importance function based on physical insight would be similarly advantageous. The task of calculating correct helium-helium interaction potentials from alpha particles and electrons and that of calculating the thermodynamics of condensed helium from helium atoms are very similar and, if not done as one large computation, can both benefit from machine and algorithm improvements.

The problem of calculating the zero temperature intercept of the melting curve of ^{3}He is made more difficult by its nuclear spin. Lee, Schmidt, Kalos, and Chester [10] have recently reported on their extension of the earlier ^{4}He work and calculated the correct ground state binding energy of condensed ^{3}He within 12%.

Calculations of thermodynamic properties at non-zero temperatures (specifically the minimum of the ^{3}He melting curve) are expected to be relatively straight-forward once the ground state calculation can be performed.

It is clear from the foregoing discussion that a great deal of progress must be made before theoretical calculations of the helium melting pressures reach a level of accuracy to be useful for standards work in pressure. Nevertheless, the path to progress is clear, and there are useful goals to achieve.

3. Linking Measurements of the Helium Melting Curve to Room Temperature Piston Gages

The transfer of the melting curve pressures to room temperature piston gages is best performed in two steps: (1) the measurement of the melting curve with an in situ (low temperature) pressure transducer; and (2) the calibration of the transducer at a higher temperature (typically 4 K) against the room temperature gage.

When using the transducer to measure the melting curve, there is no pressure head correction because the transducer is in situ. Furthermore, the cubic structure of solid helium, the rapid annealing of strains, and the self-regulating nature of a phase boundary make the melting curve relatively easy to measure, given the proper cryogenic apparatus. Impurity effects are unimportant since ^{3}He and ^{4}He may be made isotopically pure to 1 ppm. Isolation of the melting curve measurement from the filling line can be obtained by the installation of a low temperature valve, but even without a valve, a

solid plug of helium fills the line when measurements are made at temperatures below the melting curve minimum.

Recent measurements by Greywall and Busch [1] exemplify the state-of-the-art in measuring the melting curve with a transducer. They present a relation describing the temperature dependence between 8 and 330 mK of the melting pressure of ^{3}He relative to the pressure at point B of Fig. 1. The residuals of their data have a standard deviation of 30 ppm. These are completely dominated by uncertainty of the temperature scale. Their pressure measurements have a resolution of 0.2 ppm and an accuracy of 100 ppm limited only by the accuracy of their room temperature piston gage.

Earlier workers [11, 12] measuring the minimum of the melting curves were similarly limited by the accuracy of room temperature pressure gages. We are presently [13] measuring the melting curve minima of ^{3}He and ^{4}He using room temperature gages of better accuracy. Aside from the inaccuracy of the room temperature gage, the most serious source of error in transferring pressure accuracy between a room temperature piston gage and a low temperature transducer is in the pressure head correction. In our initial work, we simply used a vertical tube to transmit pressure between low and room temperatures. The pressure head correction was 680 ± 20 ppm. Fortunately, the fill line may be brought out of the dewar horizontally and the pressure head errors thereby reduced by a factor of 1000. Another possible error source, the thermomolecular pressure effect, presents no difficulty. For our present 0.075 cm diam. capillary, its effect is smaller than 0.1 ppm for pressures greater than 240 kPa.

4. Linking the Piston Gage to Density or Mass

Peggs, Elliott, and Lewis [2] have made an evaluation of the accuracy of a piston gage at 0.1 MPa. To link pressure to density, they used a manometer filled with Cook's mercury [14] and attained an inaccuracy of 4.6 ppm. The primary error sources were in the measurement of mercury column height (2.9 ppm) and temperature (3.1 ppm). Guildner and Welch [15] are presently performing a similar experiment with the goal of making the dominant error that of Cook's absolute density determination (1 ppm). Peggs *et al.* also measured the dimensions of their pistons and cylinders to link pressure to mass with an error of 15 ppm.

The transfer of these calibrations from 0.1 MPa to the pressures of the helium melting curves produces little loss of accuracy. Uncertainties in the pressure dependence of the geometry of the gage become greater than 0.1 ppm only at pressures above 10 MPa [16].

Although piston gage technology [17] has a long history and in recent decades has advanced by less than a factor of two in accuracy per decade, an order of magnitude improvement is possible for low pressure gages during the next decade. When used to link pressure to mass, present piston gages are limited by uncertainties in the piston and cylinder areas and by an imperfect understanding of the detailed kinetics of fluid flow in the gap between the piston and cylinder [2, 16, 18, 19]. Foreseeable advances in measurement technology could allow the uncertainty of the piston and cylinder areas to be pushed below 2 ppm in the next decade [20]. More elaborate modeling of the fluid dynamics in the gap, tested against fall-rate meaurement and varied piston and cylinder configurations will be required to properly assign a value to the effective piston area for imperfect, but well-measured geometries. Ultimately, less tractable effects such as those associated with surface roughness and adsorbates may inhibit progress beyond the 2 ppm level.

The availability of a stable low-temperature pressure transducer to be used in conjunction with the helium melting curve to provide pressure fixed points can help to reduce the uncertainty of piston gage calibrations in the standard laboratory. Furthermore, such reference standards can assist the pressure metrologist in understanding the limitations in precision of pressure transducers. Finally, there is hope that, one day, helium melting curve pressure data derived from first principles calculations can compare in accuracy with those derived from artifact standards.

The author is grateful for numerous discussions with interested persons within NBS concerned with pressure and fundamental constants and with theorists working on the helium many-body problem. Particularly valuable was the advice of L. A. Guildner which resulted in major revisions. Also greatly appreciated were conversations with V. E. Bean, R. J. Hocken, J. C. Houck, M. R. Moldover, R. D. Mountain, W. D. Phillips, J. F. Schooley, and B. E. Welch of NBS and theorists C. E. Campbell, D. M. Ceperley, M. H. Kalos, and R. A. Guyer.

References

[1] D. S. Greywall and P.A. Busch, J. Low Temp. Phys. 46, 451 (1982).
[2] G. N. Peggs, K. W. T. Elliott, and S. Lewis, Metrologia 15, 77 (1979).
[3] John E. Daborn, Meas. and Control 10, 377 (1977).
[4] D. M. Ceperley, private communication.
[5] For a review of Monte Carlo methods as applied to the many-body problem see M. H. Kalos, Nucl. Phys. A 328, 153 (1979).
[6] R. A. Aziz, V. P. S. Nain, J. S. Carley, W. L. Taylor, and G. T. McConville, J. Chem. Phys. 70, 4330 (1979).
[7] P. A. Whitlock, D. M. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B 19, 5598 (1979).
[8] M. H. Kalos, M. A. Lee, P. A. Whitlock, and G. V. Chester, Phys. Rev. B 24, 115 (1981).
[9] M. Kalos, private communication.
[10] M. A. Lee, K. E. Schmidt, M. H. Kalos, and G. V. Chester, Phys. Rev. Lett. 46, 728 (1981).
[11] A comprehensive review of the melting curve of ^{3}He was made by W. P. Halperin, F. B. Rassmussen, C. N. Archie, and R. C. Richardson, J. Low Temp. Phys. 31, 617 (1978).
[12] E. R. Grilly, J. Low Temp. Phys. 4, 615 (1971).
[13] C. T. Van Degrift and W. J. Bowers, Jr., to be published.
[14] A. H. Cook, Philos. Trans. R. Soc., London, Ser. A: 254, 125 (1961).
[15] L. A. Guildner and B. E. Welch, private communication.
[16] R. S. Dadson, R. G. P. Greig, and Angela Horner, Metrologia 1, 55 (1965).
[17] P. L. M. Heydemann and B. E. Welch, in *Experimental Thermodynamics, Vol II: Experimental Thermodynamics of Nonreacting Fluids*, Ed. by B. LeNeindre and B. Vodar (Butterworths, London, 1975), p. 147.
[18] A. H. Bass, J. Phys. E. 11, 682 (1978).
[19] C. M. Sutton, Report No. 562, Phys. and Eng. Lab, New Zealand (1977).
[20] R. J. Hocken, private communication. These improvements will come from advances in diamond turning machine technology such as described by R. R. Donaldson, Lawrence Livermore Lab. Report # UCRL-528-12, Sept. 1979.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Realization of the Electrical SI Units

B. P. Kibble

Division of Electrical Science, National Physical Laboratory, Teddington, England

The ampere is unique in the set of SI units in that the accuracy with which it can be realized is still insufficient for accurate measurements of the fundamental constants. Possible ways of improving this situation are described. In contrast, the ohm is in good shape, and a new cryogenic phenomenon concerning a quantized Hall effect in a MOSFET structure enables a more accurate non-QED measurement of the fine structure constant to be made in terms of it. A change of base units to eliminate the artifact-based definition of the kilogram is a more speculative possibility.

Key words: ampere; fine structure constant; kilogram; moving coil experiment; ohm; quantized Hall effect.

1. Introduction

Although the title might imply that the realization of all electrical units is to be discussed, in fact we need only be concerned with the ampere and the ohm, with its associated impedance units, the farad and the henry. The magnetic units, the tesla and ampere/meter are dealt with indirectly in the paper by Williams *et al.* [1] because magnetic flux densities are measureable with all the precision desired by the usual NMR techniques once the gyromagnetic ratio of the proton is known in SI units. The volt and watt are directly derivable from the ampere and ohm, and the SI base units of length and time lead to the coulomb and units of electric field strength.

2. The Ampere

At the time of the last conference ten years ago the ampere was thought to be realized by various current balances in different laboratories with sufficient accuracy for measuring the closely related fundamental constants such as the Faraday F, the gyromagnetic ratio of the proton γ'_p and Avogadro's number N_A. Later more accurate work on the measured values of these constants revealed discrepancies in relationships between themselves and other constants which were greater than the uncertainty with which the ampere was thought to have been realized [2]. A further current balance measurement at the National Physical Laboratory, England (NPL) [3] and refinements and additions to the constants measurements have now turned this healthy situation on its head; the realization of the ampere is now so uncertain that instead of serving the needs of measurement of the fundamental constants, measurements of F [4, 5], γ'_p [1, 3] and N_A [6–10] are being used to realize the ampere, with a consequent loss of physical information.

It has become usual [2] to express the results of ampere realizations as a ratio, K, of a maintained unit of current to the SI unit. The present situation is summarised in Fig. 1. In ascending order in the diagram are plotted the results of:

1) The latest NPL current balance realization [3].

2) A mean of previous current balance determinations [2].

3) The result from an attracted disc electrometer [11, 12].

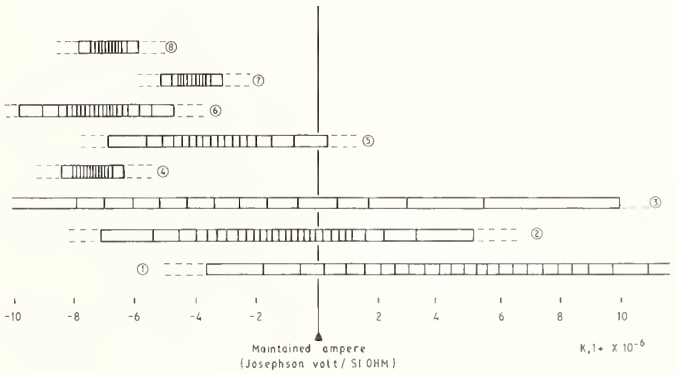


FIGURE 1. *The approximate present state of the direct and indirect determinations of the ampere (with two-sigma uncertainties).*

4) The geometric mean of the NPL strong-field γ'_p and National Bureau of Standards, USA (NBS) weak-field γ'_p measurements. This combination is equivalent to the use of a current balance, and also yields γ'_p in SI units [2].

5) The same procedure, applied to the results from the National Institute of Metrology, Beijing, China [13].

6) A realization based on a relationship amongst the constants where the least accurately known member is F [4, 5].

7) Similarly for N_A, based on the result of an NBS measurement [6, 7].

8) Partly based on results towards a N_A measurement obtained at the Physikalisch-Technische Bundesanstalt, W. Germany (PTB) [8–10].

The customary error bars, which have the disadvantage of visually suggesting a limit of error, have been replaced by shading whose density at a given abscissa represents the probability, according to the realization in question, of the corresponding value of K being correct. The shading is terminated at points two standard deviations away from the mean, as hindsight applied to the results of previous precision measurements suggests that a subsequent much more accurate result lies within two, rather than one, standard deviations of previous measurements. Even with this generous latitude however it is clear in the present instance that there are inconsistencies in that no single vertical line can be drawn which intersects all results, even between the $\pm 2\,\sigma$ limits. More seriously, the

true result might lie anywhere within a range of more than five, or more than twenty, parts in a million depending on whether or not one discounts the "direct" current balance determinations. The use of a current balance does not seem favorable for improved accuracy. Electrometers [11, 12, 14] of various kinds are a possibility, but the pace of ever-increasing accuracy of measurement of the constants associated with the ampere means that the target is receding.

A conventional current balance is prevented from achieving part per million accuracy by the smallness of the force generated between the fixed and suspended coils. Moreover, disturbing random convective forces are generated by the considerable heat dissipated in the coils in order to generate even this small force, so there is a several parts per million random component in the results which makes it difficult to seek systematic errors at or below this level. Furthermore it is an exceedingly difficult and tedious instrument to use, particularly with regard to the many hundreds of sub micrometer mechanical measurements which must be made to establish the relative positions of the conductors between which the force is generated and even then assumptions must be made about how these measurements relate to the actual mean current paths.

A possible alternative approach [15] which has been much discussed during the past five years and which is being implemented in at least two laboratories (NPL and NBS) [16, 17] uses a coil suspended from a balance in a non-uniform magnetic field. The force F newtons (equal to an opposing mass times the gravitational acceleration) can be measured when a current I flows in the coil. If, in a separate measurement, the coil is moved so that it has velocity v meters per second at the instant it passes through the position it occupied for the weighing operation an emf V is then generated. The equality of mechanical and electrical watts implied in the ampere definition and electromagnetic theory ensures that the relation

$$VI = Fv$$

holds exactly. Because of the careful SI ohm determinations carried out by several laboratories possessing calculable capacitors [18–21], resistances R whose values are accurately known in SI ohms are available, so V can be equated to the potential drop produced by I in R and hence

$$I = \sqrt{(Fv/R)} \quad \text{SI amperes}$$

By shaping the coil and magnetic field distribution appropriately the force and emf can be made substantially constant over a moderate distance of uniform travel of the coil, and the problem of measuring an instantaneous velocity is replaced by the easier requirement of obtaining an average of the force and emf. The photograph (Fig. 2) shows the NPL version of this idea. It is based on a very large permanent magnet which provides a field of almost 0.7 T in a 58 mm gap between 300 mm square pole faces. In this gap an astatic coil (wound as a figure of eight) of 3362 turns having a mean distance of 240 mm between its vertical sides, is suspended from one end of a long balance beam. The total travel of the coil, obtained by allowing the beam to rotate, is ± 25 mm centered on the beam's horizontal position. Voltage measurements will be made during the traversal of the central ± 20 mm with the velocity servo controlled to better than 10 ppm.

Force measurements will be obtained by weighing at several of the positions the coil passes through, by elevating the whole platform on which the balance case rests. A current of 10^{-2} A, dissipating a power of 10^{-2} W

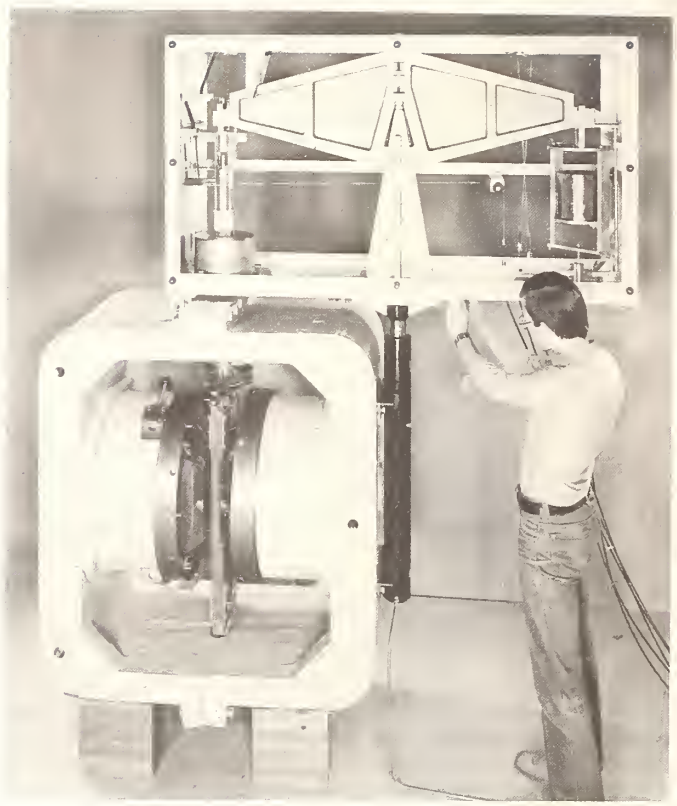


FIGURE 2. *The prototype NPL moving coil absolute ampere apparatus.*

in the coil will generate just over 1 kg weight of force change when its direction is reversed and a velocity of about 2 mm per second will produce 1 V of emf. Figure 2 gives an impression of the whole apparatus before the thermostatted cabinet which will enclose it is erected.

The moving and weighing processes will be remotely operated to avoid the thermal effects of a human operator, and will be largely under automatic computer control, so that the tedious repetition of obtaining results in the vital search for systematic errors will be bearable. We expect to obtain results with this apparatus in 1982. The final NBS version of this idea [16, 17] is based on the coil geometry used in Campbell mutual inductors and the Ayrton-Jones current balance, where the uniformity of dM/dz of a coil in a region of the midplane of symmetry of two opposed cylindrical coils is exploited to give the desirable uniform generation of emf with rate of displacement.

At NBS, preliminary trials and measurements have also been carried out using the geometry of the Pellat electrodynamometer where the torque and induced emf on rotation of a cylindrical coil pivoted about an axis in the plane of symmetry in a uniform magnetic field are measured in turn [16]. The torque was determined from the product of an opposing mass, the gravitational acceleration and the length of the balance arm. The rotational rate was found from the linear velocity of an interferometer mirror at the end of the same arm so that the arm length was eliminated when the results of the two measurements were combined.

In yet another approach to what is essentially the same fundamental idea, K. Hara *et al.* [22] are pursuing a cryogenic method first suggested by Sullivan and Frederick [23] of equating the change of electromagnetic energy of a superconducting body levitated in the magnetic field of a current carrying coil to the change in gravitational

potential energy when the body is moved. The situation could be thought of as the inverse of a moving coil measurement as the coil is now fixed and the field threading it is modified by the movement of the levitated body. The use of cryogenic equipment makes it convenient and attractive to measure the currents directly in terms of an SI resistance and a Josephson effect voltage to provide a determination of the Josephson effect constant in SI units but the method suffers from difficulties with vibration, in common with room temperature moving coil measurements, and in addition would have the problem of ensuring that the levitated cryogenic mass was correctly related to mass standards at room temperature.

3. The Ohm

The farad having been realized in SI units by finding the SI value, C, of a fixed capacitor from a calculable capacitor, the ohm can in turn be set up by using frequency-dependent ac bridges which realize the relationships $R = 1/\omega C$ or $L = 1/\omega^2 C$ to find the SI resistance R in ohms of a resistor at an angular frequency ω or the SI inductance L in henries of an inductor [13, 18–21]. Such realizations are, at any rate by comparison with the ampere, in good shape, with at least five different laboratories' determinations agreeing to within 0.2 parts in a million for the ohm.

As the number of laboratories in various countries possessing Thompson-Lampard capacitors has increased greatly of late, we may shortly expect yet more results at about this level of accuracy. At present, it would appear to be very difficult to improve the accuracy much beyond a few parts in 10^8, but the discovery of the quantized Hall effect in two-dimensional electron flow in MOSFET's cooled to liquid helium temperatures [24] does make it worthwhile to pursue greater accuracy.

Figure 3 shows the device schematically; for a suitable combination of strong magnetic field B and gate voltage V_G the potential drop V down the device is vanishingly small. Under these conditions the Hall resistance R_H which is the ratio of the transverse Hall voltage V_H to the current I through the device, appears to be very closely equal to $e^2/nh \approx 1/n \times 25813$ ohms, where n is an integer. The full potential of this discovery is not yet certain, but the maintained standard of resistance generated by this effect could be useful at any one of three possible levels.

1) Even if R_H depends slightly on the type of device and its environment, it would still be very useful for maintaining a resistance unit locally by specifying the conditions and type of device.

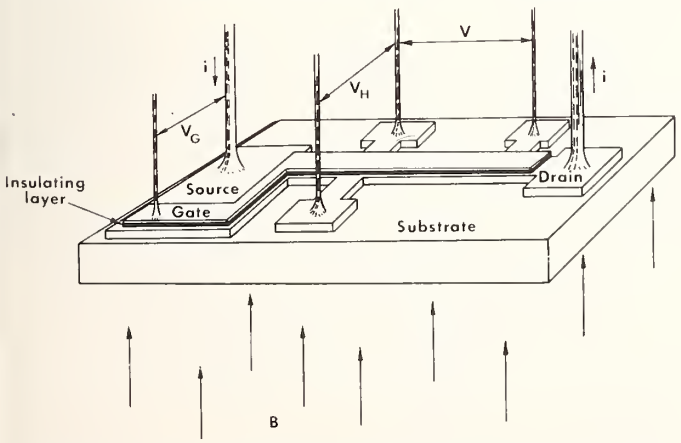


FIGURE 3. *A schematic diagram of a MOSFET device for exhibiting quantized Hall resistance effects.*

2) If R_H is device, and within limits, environment independent, it could maintain a unit universally so that resistors need no longer be continually exchanged between laboratories to compare their results.

3) If R_H, without significant correction (or the corrections are calculable and subject to experimental investigation) is exactly an integral fraction of e^2/h, it provides a potentially very accurate determination of the value of the fine structure constant.

There is therefore an urgent need to make theoretical and experimental investigations to find out at which level the device is useful, and the contributions relating to this effect are especially welcome [24–31].

4. Possible Developments in the SI Units

Examination of history shows that the SI units and their predecessors are not a fossilized entity; the definitions of the base units have been changed several times as part of a natural evolution towards higher accuracy. Naturally, care was taken that, at each redefinition, continuity with the old units was ensured to within their reproducibility. One can also, more radically, contemplate a different choice of base units, and it is interesting to see whether in any sense one could make an improvement, bearing in mind that the sole purpose of a unit is that the realization of its definition should be more reproducible, both from time to time and place to place, than any measurement we might seek to make in terms of the unit. We have seen that the definition of the ampere at present fails this test by two or more factors of ten, and moreover although the definition quite properly specifies the realization of the unit without regard to the correctness of any particular physical theory (in this case electromagnetic theory), unfortunately, to make any sort of practical apparatus that theory has to be used.

The kilogram is also unsatisfactory in a different sense, as there has been a rapid evolution in recent years towards units based on universal physical properties which can be shown to high accuracy to be free from secular change, and are freely available to any laboratory that cares to realize their definition. If the quantized Hall effect and the moving coil experiment fulfil all their early promise they could be key links in a system of units whose basis is completely free of material standards. A possible basis would be to define four units somewhat as follows:

Whizz. 1/299792458 of the velocity of plane electromagnetic radiation in free space. Accuracy of realization (AR), 10^{-3} ppm.

Second. As at present (and implying the hertz). AR 10^{-7} ppm.

Volt. That potential step which would correspond to radiation of 4.83594×10^{14} Hz in a Josephson junction. AR 10^{-2} ppm.

Ampere. The current flowing through a quantized Hall effect device which would give rise to a Hall potential of 6453.2 V.

From these, the kilogram would be a derived unit, with a definition somewhat as follows:—

Kilogram. That mass which, when moving with a velocity of one whizz, possesses mechanical energy equal to an electrical energy of one half volt-ampere-second.

Although this scheme has the merit of providing universal units, it will only become attractive if it can be

shown to yield units at least as reproducible as the present ones. From an electrical point of view there is no problem because, as we have seen, the reproducibility of the ampere would certainly be improved. No fundamental constants such as the permeability of free space, μ_0, need have their values defined, which is arguably as it should be, and some apparatus such as the moving coil apparatus would realize the kilogram. To approach the present reproducibility of the kilogram would require uncertainties of a few parts in 10^8. It may be pointed out that there are some problems at this uncertainty level connected with the present kilogram. Presumably the definition implies that only the platinum-iridium material is to be included in the mass, in which event the prototype kilogram should be compared with others with all the surface films removed and kept absent, in vacuo. Ordinary cleaning, as at present practiced gives rise to ambiguities of a few parts in 10^8.

Clearly we do not know now whether the target of a few in 10^8 is attainable, but if past progress is a reliable guide to the future, the next ten years should give us a very good idea.

References

[1] E. R. Williams, P. T. Olsen, and W. D. Phillips, these proceedings.
[2] B. N. Taylor, Metrologia 12, 81 (1976).
[3] P. Vigoureux and N. Dupuy, Nat. Phys. Lab. (NPL) Rep. DES 59, (1980).
[4] V. E. Bower and R. S. Davis, J. Res. Natl. Bur. Stand. 85, 175 (1980).
[5] L. J. Powell, T. S. Murphy, and J. W. Gramlich, these proceedings.
[6] R. D. Deslattes, Annu. Rev. Phys. Chem. 31, 435 (1980).
[7] R. D. Deslattes, these proceedings.
[8] H. Siegert and P. Becker, these proceedings.
[9] P. Becker and H. Sieger, these proceedings.
[10] P. Seyfried, these proceedings.
[11] N. Elnékavé et al, Rapports Laboratoire Central des Industries Electrique (France), No. 749-677-629-583 (c. 1979).
[12] N. Elnékavé and A. Fau, these proceedings.
[13] Wang Zhu-Xi, these proceedings.
[14] G. J. Sloggett, W. K. Clothier, D. J. Benjamin, M. F. Currey, and H. Bairnsfather, these proceedings.
[15] B. P. Kibble and I. A. Robinson, NPL Report DES 40, (1977).
[16] P. T. Olsen, W. D. Phillips, and E. R. Williams, these proceedings.
[17] P. T. Olsen, W. D. Phillips, and E. R. Williams, J. Res. Natl. Bur. Stand. 85, 257 (1980).
[18] R. D. Cutkosky, IEEE Trans. Instrum. Meas. IM23, 305 (1974).
[19] A. M. Thompson, Metrologia 4, 1 (1968).
[20] B. P. Kibble *et al.* (unpublished), (1979).
[21] ETL News, No. 328 (1977) (in Japanese).
[22] K. Hara, F. Shista, and T. Kubota, these proceedings.
[23] D. B. Sullivan and N. V. Frederick, IEEE Trans. Magn. MAG-13, 396 (1977).
[24] K. V. Klitzing, these proceedings.
[25] E. Braun, P. Guttmann, G. Hein, P. Warnecke, and S. Xue, these proceedings.
[26] C. Yamanouchi, K. Inagaki, J. Moriyama, S. Baba, S. Kawaji, T. Endo, M. Koyanagi, K. Murakami, T. Igarishi, and A. Nakamura, these proceedings.
[27] M. E. Cage, R. F. Dzuiba, B. F. Field, C. F. Levine, and R. J. Wagner, these proceedings.
[28] A. Hartland, these proceedings.
[29] A. C. Gossard and D. C. Tsui, these proceedings.
[30] L. Bliek and G. Hein, these proceedings,
[31] S. M. Girvin and R. W. Rendell, these proceedings.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

An Absolute Determination of the Volt at LCIE

N. Elnékavé and A. Fau

Laboratoire Primaire Electricité-Magnétisme, Laboratoire Central des Industries Electriques, 92260 Fontenay-aux-Roses, France

The absolute determination of the volt using Kelvin's electrometer which was carried out in 1978 is now repeated with an improved volt balance. The new instrument includes a larger active electrode and a reduced gap between it and the guard ring. Second order corrections due to gap effects have been calculated by finite element methods and checked through rheographic mapping. It is estimated that random errors associated with the use of the new instrument will amount to less than 4 ppm.

Key words: absolute volt; electrical standards; International System of Units; Kelvin electrometer; precision measurements.

1. Introduction

Determination of the SI volt is one of the main problems of present electrical metrology. We have chosen to make such a determination with a Kelvin-type plane electrometer, thus relating the value of emf standards (such as standard Weston cells) to the SI units: meter, kilogram, and second (Fig. 1). From such a determination we can deduce the fundamental constant ratio e/h (e the charge of the electron, h Planck's constant) related to emf and frequency by the Josephson effect.

The absolute value (U_a) of the applied difference of potential to the electrometer is evaluated in terms of four main parameters which are: direct capacitance C; distance of electrodes; mass m that equilibrates the electrical attraction between the plates; and the gravitational acceleration g:

$$mg = -\frac{1}{2} U^2 C/d \ .$$

A slight correction should be made to the calculated force, to take account of the airgap effect (see proof below) [1].

An electrometer built in 1975–1978 was used for a first determination of the volt, yielding its absolute value with an estimated uncertainty of 9 ppm. The force on the electrometer was 0.035 N. A new electrometer with enlarged electrodes provides now a higher electrostatic force (approximately 0.05 N) with the same applied voltage (8.4 kV) and electrode distance (1 cm).

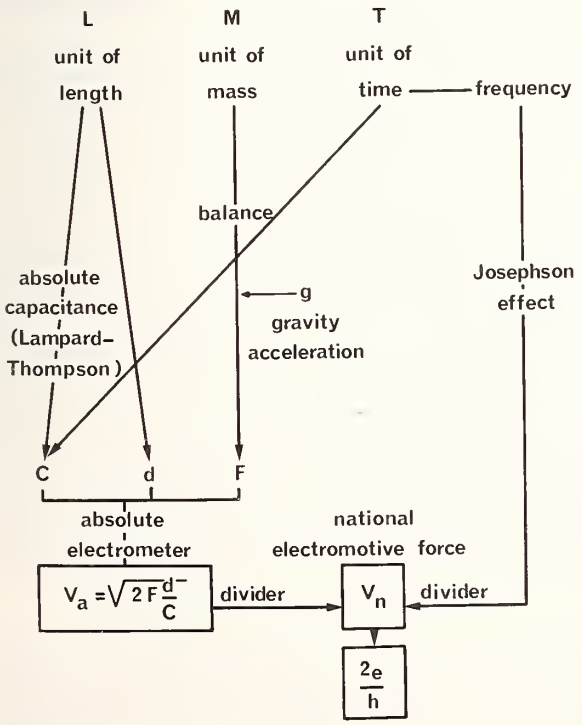


FIGURE 1. *Absolute determination of the volt and $2e/h$ and the role of the SI units.*

2. The Absolute Electrometer

The electrometer (Fig. 2) is of the balance-type, built according to Kelvin's guarded electrode design [2]. It is

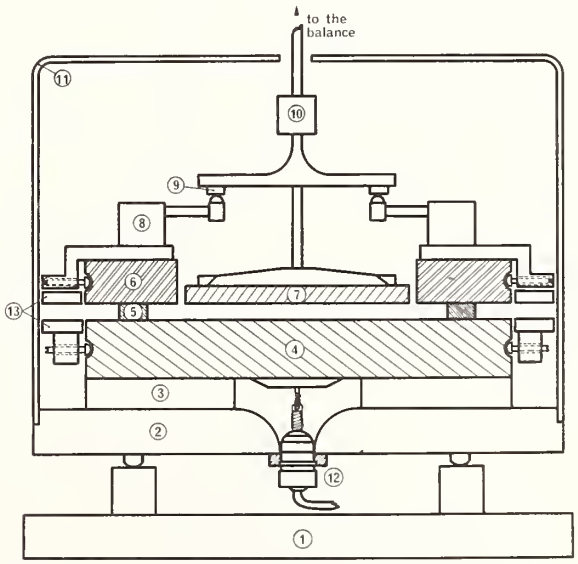


FIGURE 2. *Schematic of the LCIE Kelvin electrometer: (1) x-y table; (2) base plane; (3) glass slab; (4) fixed electrode; (5) silica spacers; (6) guard ring; (7) movable electrode; (8) micrometer screw; (9) corindon planes; (10) gimbal suspension; (11) screening; (12) high voltage terminal; (13) additional ring.*

in fact a 3-terminal plane condenser whose guarded electrode is suspended from the balance. Electrodes are thick, gold-chromium plated silica discs (total metal thickness 0.4 μm). Flatness is within λ/20 for the movable and high voltage electrodes and within λ/10 for the guard ring (λ ≃ 0.6 μm). The movable electrode diameter is 14 cm and the airgap is reduced to 0.16 mm on the new instrument. The edges of the airgap are rounded with a radius of 0.05 mm and 0.09 mm on the movable and guard electrodes, respectively.

The distance between the electrodes is defined through four silica spacers, with an approximate height of 10 mm (working voltage 8.4 kV) or 12 mm (working voltage 10 kV). The movable electrode suspended from the balance beam is maintained coplanar with the guard through the use of three stops.

2.1 Adjustment

Parallelism between the high voltage and guard electrodes is achieved by the equal height of the spacers. A mechanical system can then act on the stops to provide the necessary displacements of the movable electrode in order to bring coplanarity of the latter with the guard.

The first adjustment is to make the guard electrode horizontal. This is done with reference to an already horizontal plane by auto collimation. A mirror fixed above the guard disc will henceforth be used in conjunction with the auto collimator as a horizontal reference for the electrometer structure.

Coplanarity of the guard and movable electrodes is checked by means of a Michelson interferometer (Fig. 3). A first test with red light (He-Ne laser) is made for parallelism. A second test with white light (the unique black fringe on the movable electrode should be in line with the black fringe on the guard electrode) is then made to prove exact coplanarity of the surfaces (Fig. 4).

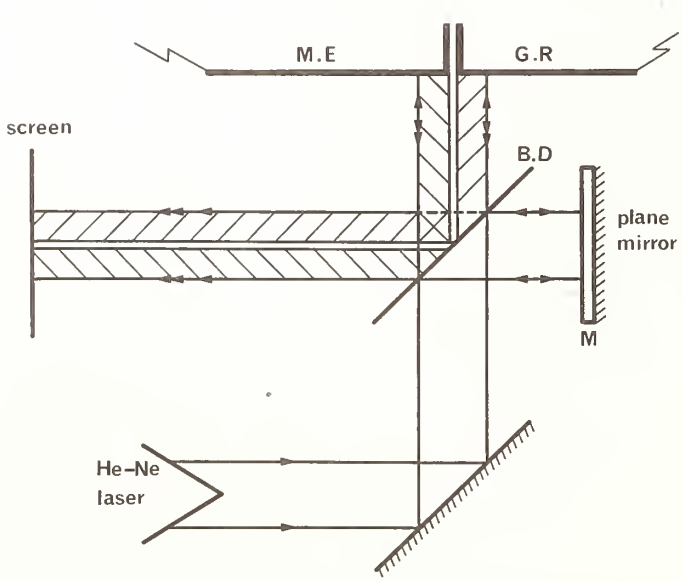


FIGURE 3. *Michelson interferometer arrangement used to achieve coplanarity of the guard and movable electrodes.*

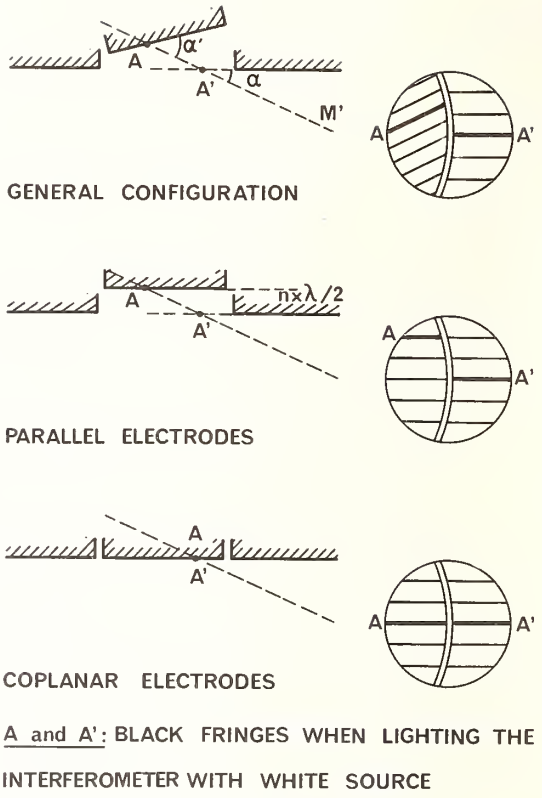


FIGURE 4. *Fringe pattern obtained from the Michelson interferometer of Fig. 3.*

3. Voltage Measurement With Electrometer

Equilibrating the balance with no voltage applied is the first step in the measurement. Applying the voltage (8.4 kV or 10 kV) to the electrometer and withdrawing the calculated weight on the movable electrode side of the balance would then restore equilibrium. It is known, however, that no stable equilibrium can thus be obtained with a quasi-linear restoring force. Stops are therefore provided which limit the motion of the movable electrode in the downward direction. Decreasing the voltage slightly results in an upward motion of the movable electrode.

Electrometer balance is effected with a linear controlling force, through the action of a current carrying coil system (fixed) on a magnet on one of the balance pans. The whole balance system is within a nitrogen filled enclosure at a pressure slightly above atmospheric. Measurements of capacitance and force are thus made in identical conditions. Capacitance measurements are made with an inductive-ratio transformer bridge at a frequency of 795 Hz. Standards used are calibrated with respect to a 5-cylinder, Lampard-Thompson type, cross capacitor.

4. dc Voltage Divider

The divider is made up, on its high voltage side, of one hundred 100 kΩ resistors in series, and on its low voltage side of a single 1000 Ω resistor. The parallel setting of the 100 high-voltage resistors results in a 1000 Ω resistor which is compared to the low voltage side resistor by a substitution method.

5. Results of the 1978 Determination

An absolute determination of the volt was performed in August 1978 at the LCIE yielding for the as-maintained French unit of potential V_F

$$1\ V_F = (1 - 4.5 \times 10^{-6})\ V\ ,$$

where V stands for the SI unit. The French volt being defined in terms of the Josephson effect with an adopted value of $2e/h$ equal to 483 594.64 GHz/V_F, the corrected SI value for the combined constant is

$$2e/h = 483\ 596.82\ \text{GHz/V}.$$

The total uncertainty was estimated not to exceed 9 ppm.

6. Expected Accuracy of the New Electrometer

The following are the estimated uncertainties related to the independent evaluations which are included in the absolute measurement:

Electrode separation	2.5×10^{-6}
Capacitance	1×10^{-6}
Gap correction	0.5×10^{-6}
Mass	3×10^{-6}
Acceleration (gravitational)	0.2×10^{-6}

The total uncertainty with the new instrument is thus expected to be smaller than 4 ppm.

7. Appendix

7.1 Gap Correction for Axially Symmetrical Electrometer

Notation:

- R: inner radius of guard ring
- r: outer radius of movable electrode
- d: distance between fixed and movable electrodes
- ϵ: permittivity of gaseous medium
- C: measured direct capacitance between fixed and movable electrodes
- $\sigma(x)$: charge density on fixed electrode as a function of radius x

The airgap is limited by two coaxial cylindrical surfaces (radii R, r). Slightly rounded edges join these cylindrical surfaces with the plane surfaces of the movable and guard electrodes (Fig. 5).

A so-called "neutral" tube of force separates those lines of force which end on the movable electrode from those ending on the guard electrode. In between the cylindrical surfaces, this tube of force tends to be cylindrical with a radius equal to $\sqrt{Rr}$. The intersection of the neutral tube and the fixed electrode plane is a circle with radius R_0.

Charge density $\sigma(x)$, total charge Q, capacitance C, and potential difference U are related by

$$Q = CU = 2\pi \int_0^{R_0} \sigma(x)\, x dx \equiv \pi R_0^2\, \bar{\sigma}\ . \qquad (1)$$

An average charge density $\bar{\sigma}$ is thus defined on the area πR_0^2 of the circle of radius R_0.

The electrostatic force between the fixed and movable electrode is thus

$$F = \frac{\pi}{\epsilon} \int_0^{R_0} \sigma^2(x)\, x dx\ . \qquad (2)$$

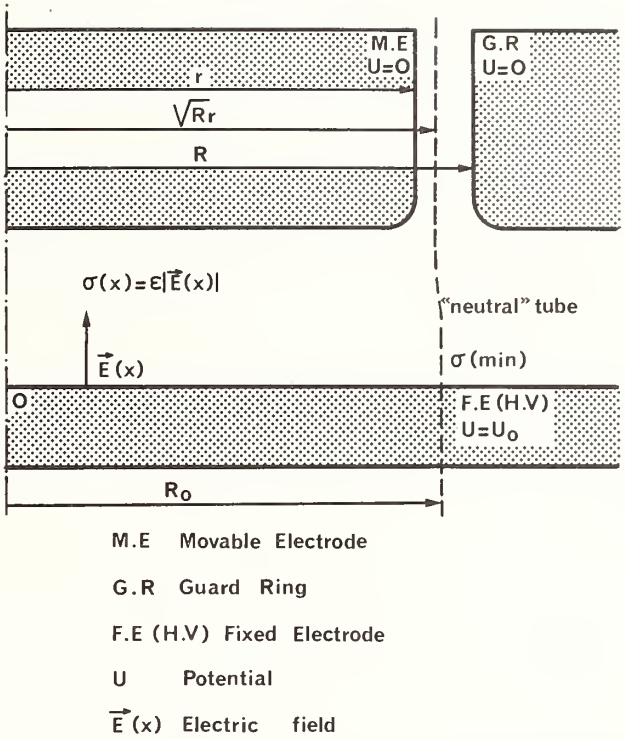


FIGURE 5. *Schematic of the electrometer showing the quantities which enter the calculation of the gap correction.*

Let us define a function $\delta(x)$ such that

$$\sigma(x) \equiv \bar{\sigma}[1 + \delta(x)]\ . \qquad (3)$$

The above relation implies

$$\int_0^{R_0} \delta(x)\, x dx = 0\ . \qquad (4)$$

It is then easily shown that

$$F = \frac{\pi R_0^2 \bar{\sigma}^2}{2\epsilon} \left[1 + \frac{2}{R_0^2} \int_0^{R_0} \delta^2(x)\, x dx\right]\ . \qquad (5)$$

Let

$$C_0 \equiv \frac{\epsilon \pi R_0^2}{d}\ .$$

Then

$$F = \frac{U^2 C}{2d} \frac{C}{C_0} \left[1 + \frac{2}{R_0^2} \int_0^{R_0} \delta^2(x)\, x dx\right]\ . \qquad (6)$$

Let

$$\beta \equiv \frac{2}{R_0^2} \int_0^{R_0} \delta^2(x)\, x dx\ , \qquad (7)$$

$$\gamma \equiv 1 - \frac{C}{C_0}\ . \qquad (8)$$

The force F can then be written

$$F = (1 + \beta)(1 - \gamma) \frac{U^2 C}{2d}\ . \qquad (9)$$

The radius R_0 and the corrections β and γ can be calculated in case of cylindrical gaps with sharp edges [3, 4]. For rounded edges a 2-dimensional rheographic plot, or a 3-dimensional finite element calculation that takes into account the cylindrical symmetry of the electrometer, have been developed.

References

[1] C. Snow, Bur. Stand. J. Res. (U.S.) 1, 513 (1928).

[2] W. Thomson (Lord Kelvin), *Reprint of Papers on Electrostatics and Magnetism*, 2nd edition (Macmillan & Co., London, 1884), Sect. 360.

[3] C. Moon and C. M. Sparks, J. Res. Natl. Bur. Stand. **41**, 497 (1948).

[4] H. Abraham and J. Lemoine, J. Phys. Theor. Appl. **4**, 463 (1895).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The CSIRO Absolute Volt Project

G. J. Sloggett, W. K. Clothier, D. J. Benjamin, M. F. Currey, and H. Bairnsfather

CSIRO Division of Applied Physics, Sydney, Australia, 2070

A liquid electrometer of unique design has been constructed to provide a precise absolute voltage standard. In this progress report the instrument is described and critical aspects of its performance are assessed. Known sources of uncertainty are consistent with a volt determination at or below the 1 ppm level. The principal areas of remaining work are discussed.

Key words: absolute voltage standards; interferometry; liquid electrometer; mercury; precision electrical measurements; SI units.

1. Introduction

The volt may be maintained with a precision of about a part in 10^7 by means of the ac Josephson effect [1], but neither the volt nor the ampere is known in absolute measure to better than a few parts in 10^6. Taylor [2] showed in 1976 that the best available values for the absolute ampere differed by 4 to 10 in 10^6 from three indirect values calculated from fundamental constants, and subsequent ampere [3, 4] and volt [5] determinations have not resolved these disagreements. In 1979 the 16th Conférence Générale des Poids et Mesures recommended "the continuation and intensifying of research on both the direct realization of the electrical units and their indirect realization through the determination of physical constants" [6].

In the CSIRO Division of Applied Physics a direct determination of the dc volt is being undertaken using a liquid electrometer [7]. The method entails elevation of a body of conducting liquid by electrostatic attraction towards an electrode at a high potential V a distance d above it (Fig. 1). The elevation h near the center of the raised plateau is determined solely by the balance of electrostatic and gravitational forces, and in principle the applied voltage V is calculable given a knowledge of h, d, and certain other physical properties of the system. Surface effects are best eliminated by using at least two voltages V_1 and V_2 with corresponding spacings d_1 and d_2 chosen so that h remains nearly constant. One obtains

$$V_2 h_2^{-1/2} - V_1 h_1^{-1/2} = (2\rho g/\epsilon_0 \epsilon_r)^{1/2} (d_2 - d_1) \quad (1)$$

where

$$h_1 \approx h_2 \approx h \quad (2)$$

and ρ is the density of the liquid, g the acceleration due to gravity, ϵ_0 the permittivity of free space and ϵ_r the relative permittivity of the medium. If these constants and the ratio V_2/V_1 are known, the absolute determination of voltage reduces to the measurement of a few linear dimensions.

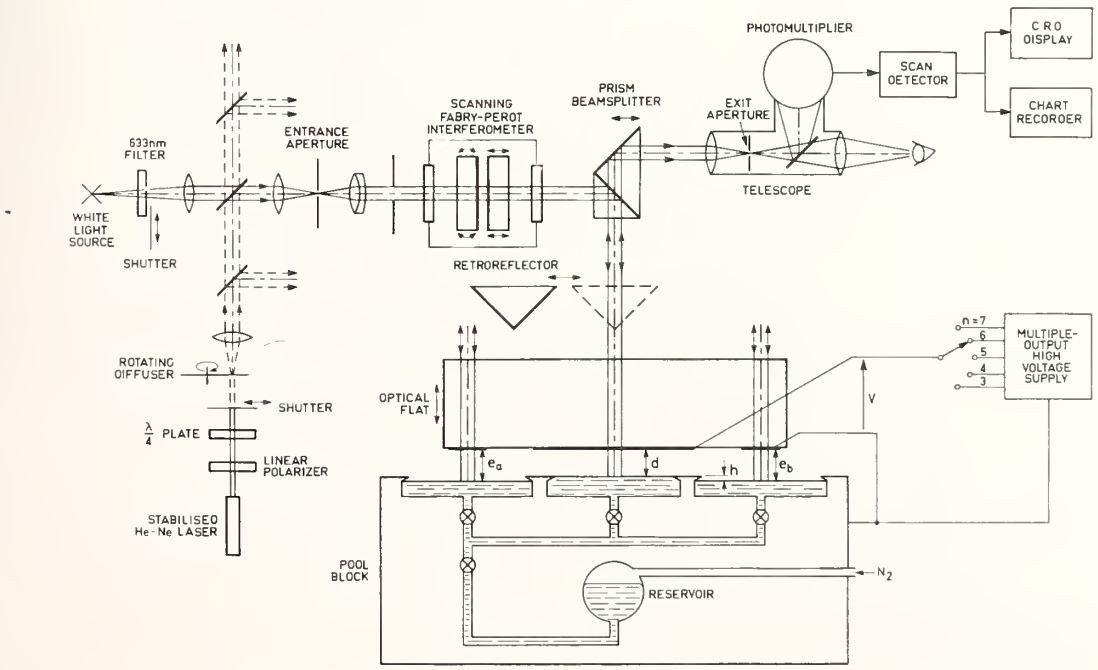


FIGURE 1. *General schematic of the liquid electrometer and its optical measuring system.*

2. Electrometer

The chosen liquid is mercury, whose chief advantages are a high reflectance and a density known to about 1 in 10^6, contributing an uncertainty of 0.5 ppm to a volt determination. If the overall uncertainty of the experiment warrants it, a more accurate density determination will be made, most likely by comparison with a sample held in this laboratory whose density was determined by Cook with an uncertainty of 3 or 4 in 10^7 [8]. Of the remaining constants in Eq. (1), g has been determined at the site of the experiment to 1 in 10^8 by gravimeter transfer from the site of an absolute g determination in this laboratory [9]. ϵ_o is known to 1 in 10^8, and ϵ_r should be within 10^{-8} of unity if the experiment is performed, as planned, in a vacuum of order 0.1 Pa.

The linear dimensions required for Eq. 1 are obtained from three interferometric spacing measurements. As shown in Fig. 1, the mercury is contained in three interconnected pools, above which is located an optical flat with three sections of thin chromium film on its lower surface. With the measurement voltage V applied to the central section of film and the outer sections grounded, the mercury level in the central pool is elevated relative to that in the outer pools. The three sections of film are optically semi-transparent and of such a thickness that reflection fringes of good contrast are obtained in conjunction with a mercury surface. Spacings e_a, d and e_b are measured at the three equispaced pool centers, and a raw value for the elevation h is calculated using

$$h = (1/2)(e_a + e_b) - d. \tag{3}$$

This value must be corrected for aplanarity of the three sections of film and for the gravitational curvature of the interconnected liquid surfaces. The necessary correction is determined from spacing measurements made with zero applied electric field.

The approximate value of h is 0.662 mm and, for reasons related to the interferometry, it is convenient to make spacing d an integral multiple of h:

$$d \approx nh, \tag{4}$$

where, for stability, n must exceed 2 [7]. A range of spacings d corresponding to values n from 3 to 7 is available by vertical movement of the optical flat. In all, ten distinct measurements may be made with the electrometer (five spacings for each of two voltage polarities), considerably in excess of the minimum number (two) required for the elimination of surface effects. This additional data should be useful in the reduction of measurement uncertainty.

A general view of the electrometer is shown in Fig. 2. The three 75 mm diameter pools are located in the upper surface of a stainless steel block, and may be interconnected and/or connected to a reservoir within the block by a system of pneumatically-operated valves. The rate of filling or draining of the pools is controlled by the pressure of N_2 gas admitted to the reservoir.

The pool diameters are sufficient that capillary depression effects are negligible. For a sessile drop of diameter 75.6 mm (a conservative model for mercury in a vertical-walled pool of diameter 75 mm) the capillary depression, calculated from Blaisdell's expressions [10], is 0.05 nm at the pool center and 0.24 nm at a radius of 5.7 mm. These effects are further reduced by means of meniscus rings fitted to the pool rims. These have a triangular cross-section with an internal angle (32° to the horizontal) which is within the range of contact angles between mercury and stainless steel [11], so that the sur-

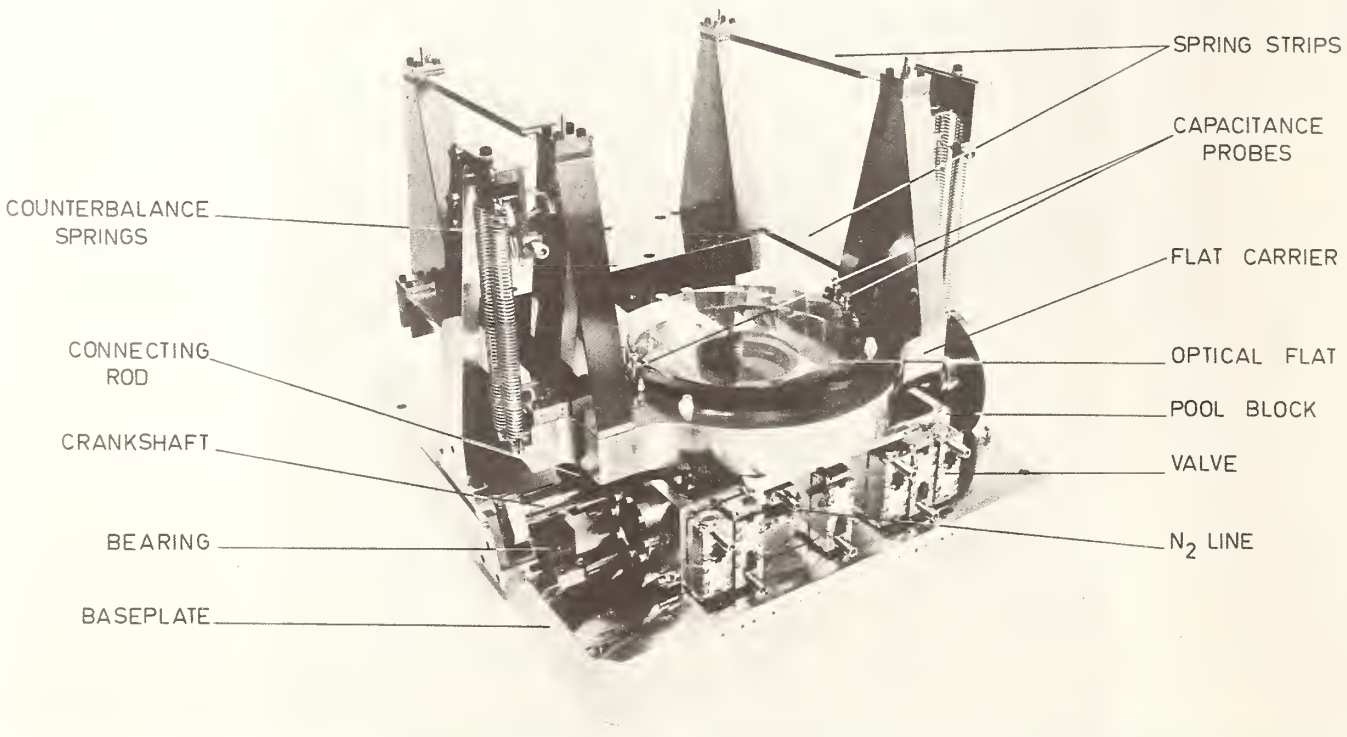


FIGURE 2. *Absolute liquid electrometer.*

faces can, in principle, be adjusted to be entirely free of meniscus curvature. In the case of the center pool, however, some curvature is to be expected at the periphery due to the weakening of the electric field, and there may also be a modification (most likely a reduction [12]) of the surface tension due to surface charges.

The optical flat is supported in a carrier which may be moved vertically by means of a precision crankshaft. Parallelism between the optical flat and the mercury surfaces is maintained to within about 10^{-6} rad over the range of vertical movement by accurate matching of the crankshaft eccentrics, by differential adjustment of spring counter-balancing systems at either end of the carrier, and by the geometry of four spring strips connecting the flat carrier to pedestals fixed to the baseplate of the instrument. The chief components of this system are made of Invar.

The instrument is housed in a vacuum chamber which may be tilted hydraulically as a means of fine adjustment of the parallelism of the optical flat and the mercury surfaces, and stands on an antivibration table (see Fig. 3). The table has three fine vertical legs each of which carries a compressive load almost sufficient to cause buckling. This system, properly damped, attenuates horizontal components of vibration above 0.5 Hz. Lead weights in elevated trays raise the center of mass of the table to the plane of the pool surfaces, so that any rotational vibration of the assembly does not cause substantial wave motion if, as mentioned, the mercury surfaces are horizontal at their edges. The use of separate pools for the three spacing measurements, each having local minima of wave motion near its center, and each fitted with a perforated damping plate a few mm below its surface, gives further attenuation of vibration.

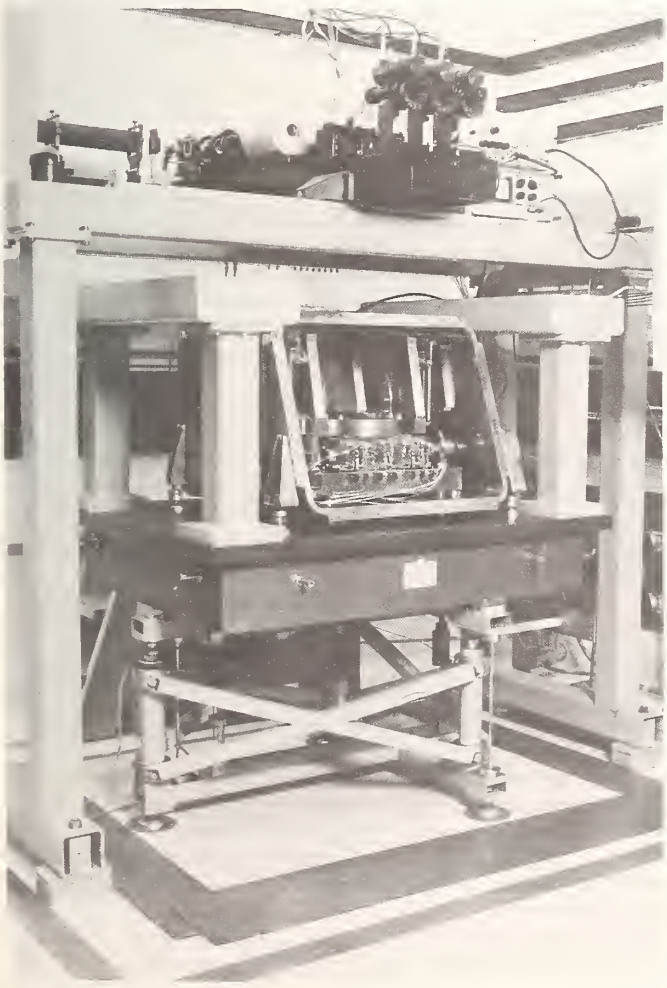


FIGURE 3. *General view of apparatus, showing (from top) the optical bench, the vacuum chamber (front cover removed) containing the electrometer, and the antivibration table.*

Figure 4 shows typical noise and drift in the interferometric spacing measurements, obtained with the three pools interconnected. The drift is principally due to the evaporative loss of mercury from the pools, and the noise to vibration on the pool surfaces. The noise level for spacing d (central pool) is satisfactory but spacings e_a and e_b exhibit a high level of antiphase noise, apparently due to the vibrational interchange of mercury between the outer pools. A fourth trace in Fig. 4 shows that the noise in the mean of e_a and e_b, which is the quantity required for the determination of h via Eq. (3), is similar to that for d. Preliminary results obtained under zero-field conditions indicate that h may be measured with a repeatability of about 0.3 nm, or 0.5 ppm. This type of random uncertainty can be reduced by repeated measurements.

3. Interferometry

Under the condition of Eq. (4) all three spacings take values which are close to integral multiples of h. Figure 1 shows one of the three identical interferometry systems. A spacing is measured in two steps. First, the unknown spacing is compared by optical multiplication with the spacing (nominally h) of a Fabry-Perot reference interferometer and, second, the Fabry-Perot interferometer is calibrated against a stabilised He-Ne laser. The Fabry-Perot interferometers, which are of critical importance in these procedures, have been described in detail elsewhere [13]. Their spacings are electromagnetically adjustable over a range of $\pm 0.05\%$, the amount of the adjustment being precisely calculable from the adjusting current. In the optical multiplication step a white light source is used and the Fabry-Perot spacing adjusted to set on a superposition fringe of order n (for spacing d) or $n + 1$ (spacings e_a and e_b). To calibrate the interferometer a retroreflector is substituted for the electrometer and the interferometer spacing is adjusted to set on the nearest monochromatic fringe (of order 2093). If the adjustments in the calibration and optical multiplication steps are f_c and f_m fringes respectively, the electrometer spacing d (for example) is given by

$$d = \frac{n\lambda}{2}(2093 - f_c + f_m), \tag{5}$$

where λ is the laser wavelength (633 nm).

The chief advantage of this method over one in which the electrometer spacings are measured directly in monochromatic light is that the precision with which the spacings need be adjusted is greatly relaxed. The present requirement of a few parts in 10^4 is readily achieved with the aid of capacitance probes mounted in the optical flat carrier (see Fig. 2). The Fabry-Perot interferometers also provide a convenient means of scanning, which is essential for fringe setting to an accuracy of 10^{-3} fringe. Some of the systematic effects limiting the interferometric precision have been discussed elsewhere [13] and the results of recent work, on the errors due to superposition fringe asymmetry, will be published shortly. Total uncertainty in the interferometric measurement of any spacing should not excede 10^{-3} fringe (0.3 nm), or about 5 in 10^7 of h.

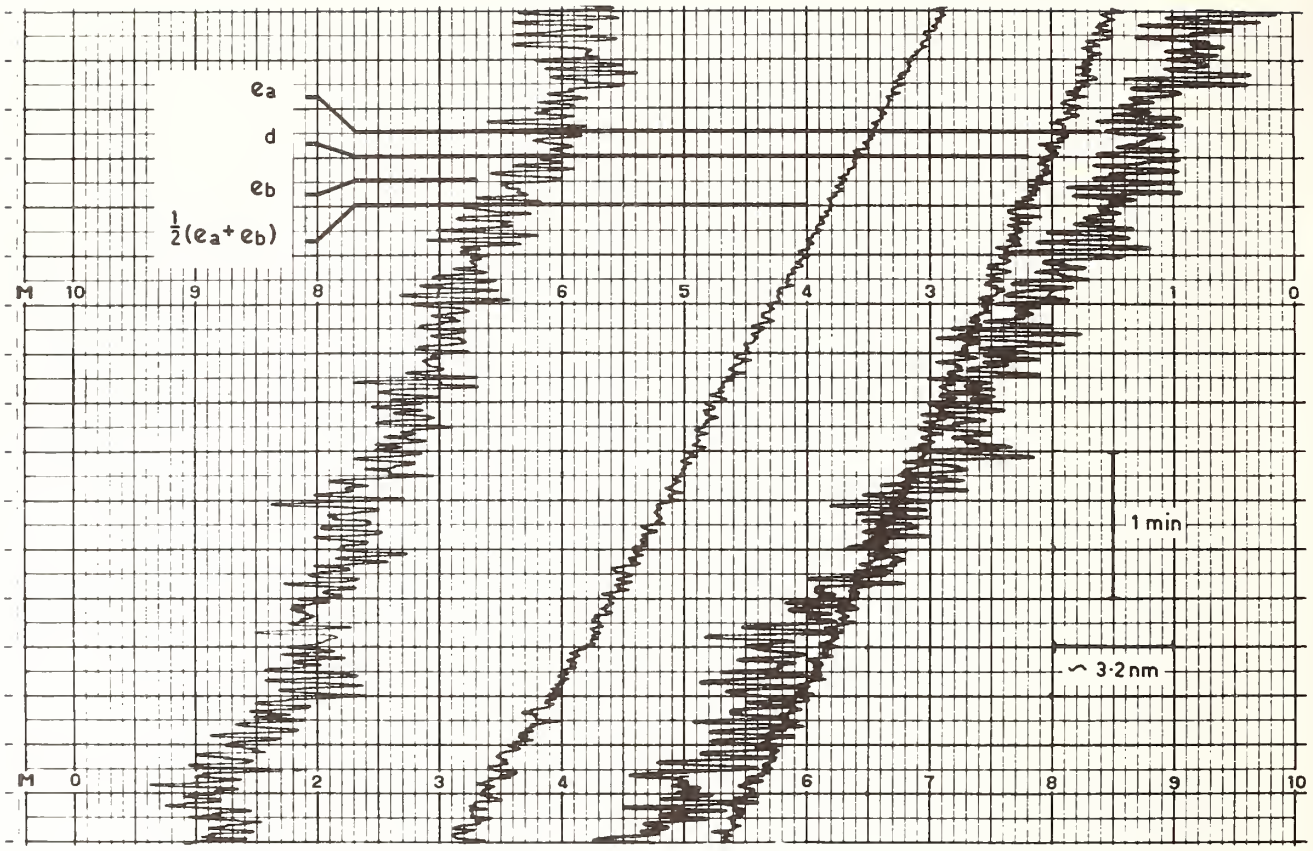


FIGURE 4. *Drift and noise in electrometer spacings. The multi-pen recorder displaces the traces by one small time-scale division (10 s), in the order indicated by the labelling.*

The components of the optical system are mounted on a triple optical bench supported on a bridge over the vacuum chamber, as shown in Fig. 3.

4. Electrical

For constant h the voltage V must increase in proportion to d, and, for the same value of n as in Eq. (4), should satisfy

$$V \approx nV_o \tag{6}$$

for some constant voltage V_o. V_o is uniquely determined by h and has the value 2953 V, which is 290 times the voltage of ten standard cells in series. A multiple-output power supply using standard cells as a voltage reference has been constructed along the lines indicated in Fig. 5. The outputs of 14 transformer-rectifier modules, each producing a voltage $V_o/2$, are connected in series and used to supply current to a 14-section resistive divider. The electrometer is connected to the $(2n)$th divider tap (at voltage nV_o) via a potentiometer control which enables the electrometer voltage to be adjusted smoothly, the selected tap being shorted to the corresponding power supply output to ensure a low output impedance. The first section of the divider has an additional tap at 10.18 V, and the amplified difference between this voltage and that of the reference cells is used to regulate the ac voltage supplied to the transformer-rectifier modules. With a loop gain of 10^6 the power supply is stable to within a few parts in 10^8 over several hours, determined essentially by the stability of the cell e.m.f.s and the divider ratios. A small known balance voltage may be inserted in series with the cells for fine adjustment of the output voltage.

The standard cells, in a portable temperature-controlled enclosure, will be used to transfer the absolute volt to the CSIRO as-maintained voltage standard [1]. The ratio of the $(2n)$th divider output to the voltage at the 10.18 V tap, nominally $290n$, is determined using two bridges which are housed with the divider in a temperature-controlled airbath. A series-parallel bridge uses two sets of 12 resistors to establish the ratio (approximately 145 or $12^2 + 1$) of the first section to its 10.18 V tap. A movable equal-arm bridge is then used to determine the ratios between adjacent divider sections, and from the 13 balances so obtained cumulative ratios are calculated. Study of the performance of the two bridges indicates that the overall divider calibrations should be accurate to within 1 or 2 in 10^7.

5. Discussion

With minor exceptions construction of the apparatus is complete and performance testing is well advanced. We are currently investigating certain systematic effects in the operation of the electrometer.

The electrometer responds as a manometer to pressure differences between the inner and outer pools, and differences as small as 10^{-4} times the saturation vapour pressure of mercury (0.16 Pa at 20 °C) may be significant in the determination of h. Preliminary measurements suggest that such a pressure difference exists and, as might be expected, has a magnitude dependent on the spacing d and the pumping history of the vacuum system. A decision on whether to attempt to eliminate this effect or merely to correct for it has been deferred until more reliable data on its magnitude and repeatability are obtained.

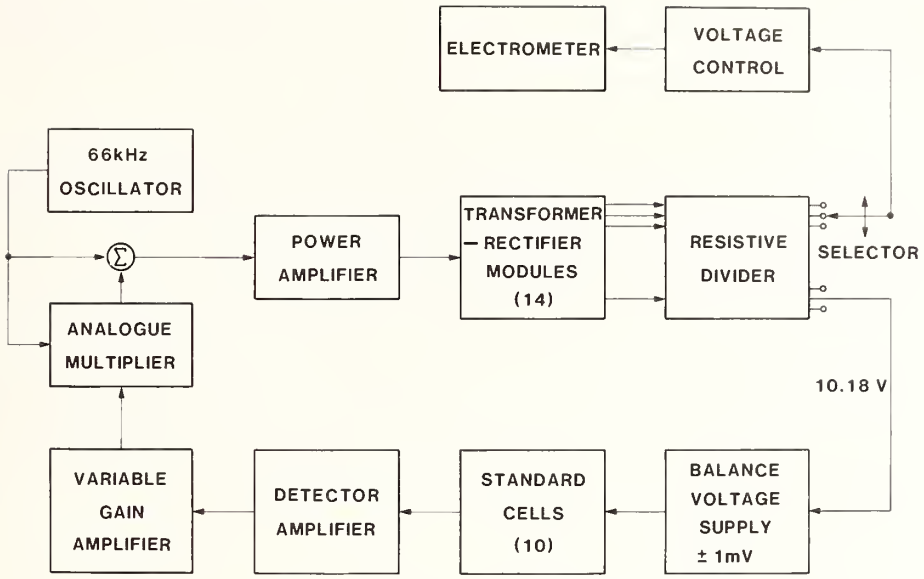


FIGURE 5. *Schematic of power supply.*

The rate of evaporation of a free mercury surface into vacuum is sufficiently high as to make accurate interferometry impossible, and causes unacceptable cooling of the liquid. The rate is reduced by using a low pumping speed and by forming oxide films on the pool surfaces, under which conditions the rate of surface movement is of the order shown in Fig. 4. These measures also tend to reduce the vapor pressure gradients across the pools. A possibly significant systematic error arises from the oxide films. If the films on the outer pools differ in thickness from that on the central pool, as could be caused by stretching or compression during adjustment of the mercury levels, the resultant differences in surface effects may be significant in the determination of h. This question will be investigated experimentally.

Another small correction to h is yet to be determined. This arises from the absence, in the zero-field measurements, of the deflection of the optical flat normally caused by the electric field.

The thermal stability of the pool block is of great importance and remains to be fully evaluated. At 20 °C mercury has a temperature coefficient of expansion of $1.81 \times 10^{-4}\,K^{-1}$ and its vapour pressure varies by 0.014 Pa K^{-1}, hence the pool temperatures must be equal to within a few mK. Temperature uniformity is achieved by completely excluding significant heat sources from the vacuum chamber and, if necessary, by surrounding the chamber and antivibration table with an insulating enclosure. Thermal gradients in the apparatus will be monitored by thermocouples and the mercury temperature measured using two platinum resistance thermometers installed in holes in the pool block.

Many members of this Division have contributed significantly to this project. We particularly wish to thank D. Bicknell, F. J. Bramall, M. Jenkins, and H. W. King for their participation in the design or manufacture of the apparatus.

References

[1] I. K. Harvey, J. C. Macfarlane, and R. B. Frenkel, Metrologia 8, 114 (1972).
[2] B. N. Taylor, Metrologia 12, 81 (1976).
[3] W. Chiao, R. Liu, and P. Shen, IEEE Trans. IM-29, 238 (1980).
[4] P. Vigoureux and N. Dupuy, in *CPEM Digest 1980* (abstracts of the Conference on Precision Electromagnetic Measurements, Braunschweig, 1980, IEEE cat. no. 80CH1497-7 IM), p. 47.
[5] A Fau, Bull. Bur. Nat. Métrologie (Paris) 10 (36), 9 (April 1979); N. Elnekavé and A. Fau, these proceedings.
[6] P. Giacomo, Metrologia 16, 55 (1979).
[7] W. K. Clothier, Metrologia 1, 181 (1965).
[8] A. H. Cook, Philos. Trans. R. Soc. London, Ser. A: 254, 125 (1961).
[9] Yu. D. Boulanger, private communication.
[10] B. E. Blaisdell, J. Math. Phys. (Cambridge, Mass.) 19, 186 (1940).
[11] D. J. Wright, Proc. Phys. Soc. London, Sect. B 68, 297 (1955).
[12] N. K. Adam, *The Physics and Chemistry of Surfaces* (Oxford University Press, London, 1941), p. 336.
[13] W. K. Clothier, G. J. Sloggett, and H. Bairnsfather, Opt. Eng. 19, 834 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Status of the Measurement of the NBS Ampere in SI Units

P. T. Olsen, W. D. Phillips, and E. R. Williams

Electrical Measurements and Standards Division, National Bureau of Standards, Washington, DC 20234

We report on the status of a preliminary measurement of the NBS as-maintained ampere in terms of the SI, or absolute, ampere by a method which avoids the usual determination of dimensions of current-carrying coils. As a result, a major source of uncertainty is removed from the experiment. The preliminary work reported here has a statistical uncertainty on the order of 1 ppm. Systematic errors have not been fully evaluated, but we believe them to be on the order of 30 ppm or less.

Key words: absolute ampere; current balance; fundamental constants; NBS ampere; Pellat balance; SI ampere.

1. Introduction

Electric current is routinely measured in the laboratory in terms of the voltage it produces across a standard resistor. The United States practical or NBS ampere is that current which would cause a voltage drop of one NBS volt across one NBS ohm. By contrast, the SI ampere is defined in terms of forces between current carrying conductors. Since electrical experiments are invariably performed using the practical electrical units, the ratio, K_A, of the practical ampere to the SI ampere is an important, albeit artificial, fundamental constant [1].

Values of K_A obtained indirectly from measurements of other fundamental constants are in significant disagreement with each other and with direct measurements using current balances [2, 3]. A major improvement in the accuracy of a direct determination of K_A would therefore shed light on the possibility of errors in the present determinations of such constants as the gyromagnetic ratio of the proton, the Avogadro constant, and the Faraday constant, all of which can be used for indirect determinations of K_A.

The various direct and indirect values of K_A differ by about 10 ppm and the 1973 recommended value of K_A may be in error by that much [2]. The measurement reported here has a precision of about one ppm and is itself only the first step in a program which may ultimately yield K_A to an accuracy of 0.1 ppm. This effort, then, is an attempt at improving the present knowledge of K_A by as much as two orders of magnitude.

2. Theory

In all previous direct determinations of K_A (absolute ampere experiments[1]), the largest contribution to the uncertainty was from the measurement of the geometrical dimensions of some physical object, most commonly the diameter and pitch of a coil. A new approach [4] avoids such measurements and may be envisioned as a comparison of electrical work measured in NBS electrical units to mechanical work measured in absolute or SI units.

In practice this comparison is realized by a two part experiment.

Part one: A coil in motion in a magnetic field generates a voltage across its terminals resulting from the time rate of change of magnetic flux in the coil ($d\phi/dt$). The coil motion is kinematically constrained to follow a fixed path. A laser interferometer measures its z displacement. An active servo system forces $d\phi/dt$ to be nearly constant from time t_1 to t_2 enabling the magnitude of the voltage to be measured accurately in terms of the NBS electrical standards:

$$\mathcal{E}_{NBS} K_A K_\Omega = d\phi/dt \tag{1}$$

where K_Ω is the ratio of the NBS ohm to the SI ohm.

Part two: If at a given position z the coil carries a current, expressed in NBS units as $I_{NBS}(z)$, it will experience a force in the z direction which can be measured in terms of a known mass m, and the local gravitational acceleration g. The relevant expression is:

$$F = mg = K_A I_{NBS}(d\phi/dz) . \tag{2}$$

If the motion of the coil system is such that time t_1 occurs at position z_1 and t_2 occurs at position z_2, then from Eqs. (1) and (2)

$$\Delta\phi = K_A K_\Omega \int_{t_1}^{t_2} \mathcal{E}_{NBS}(t)dt = \left[\frac{mg}{K_A}\right] \int_{z_1}^{z_2} \frac{1}{I_{NBS}(z)} dz , \tag{3}$$

or

$$K_A = \left\{ \frac{mg \int_{z_1}^{z_2} \left[\frac{1}{I_{NBS}(z)}\right] dz}{K_\Omega \int_{t_1}^{t_2} \mathcal{E}_{NBS}(t)dt} \right\}^{1/2} . \tag{4}$$

3. Experiment

3.1 Apparatus

The apparatus used is the Pellat-type "electrodynamometer" or current balance shown in Fig. 1. A central coil (B) with its axis vertical is mounted on a balance beam. The coil and balance are placed in a uniform horizontal magnetic field (in this case, the magnetic field of a long solenoid). The magnetic field direction (H), the axis of the coil, and the edge of the balance knife (C) are

[1] A summary of recent absolute ampere experiments is given in Ref. [1].

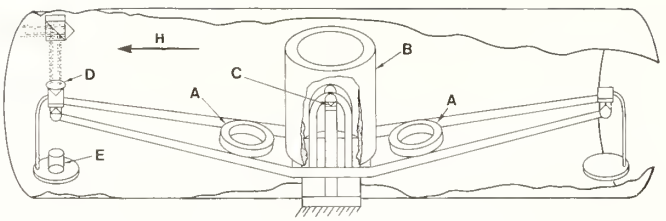


FIGURE 1. *Schematic representation of Pellat-type ampere balance. A: Torquing coils; B: Main coil; C: Central knife edge; D: Corner reflector (moving arm of interferometer); E: standard mass; H: Applied stationary magnetic field (supplied by surrounding solenoid).*

approximately mutually orthogonal. When free to swing, the balance rotates in a vertical plane as does the axis of the central coil, so that the flux in the coil varies as the sine of the angular displacement from the vertical. A current in the coil produces a torque on the balance which is proportional to $I(\mathbf{j} \times \mathbf{H})$ where $\mathbf{j}$ is the coil axis vector and $\mathbf{H}$ is the applied magnetic field vector. The torque produced is then countered by an appropriate mass E placed on the scale pan of the balance. The torque due to this mass m is proportional to $m(\mathbf{l} \times \mathbf{g})$ where $\mathbf{l}$ is the balance arm vector connecting the central knife to the scale pan knife and $\mathbf{g}$ is the acceleration of gravity. Thus, if the central coil and the solenoidal field are properly aligned and the solenoid field is uniform, the mass needed to balance the electromagnetic torque will be independent of the balance angle.

A corner reflector attached to the scale pan assembly is part of a laser interferometer system that is used to measure the vertical displacement z of the scale pan as the balance rotates. The interferometer axis is adjusted to within 10^{-4} rad of vertical by autocollimation from a mercury pool.

3.2 Evaluating the Voltage Integral

A block diagram of the system for the voltage measurement is given in Fig. 2. When the balance is in motion, the voltage generated in the central rotating coil is compared to a 10 mV reference voltage. The voltage difference, or error signal V_e, is amplified by a null detector and a current proportional to V_e is applied to three other

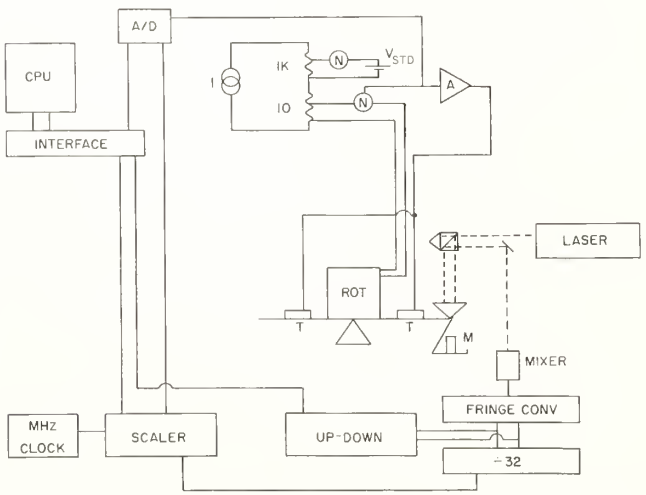


FIGURE 2. *Block diagram for voltage integral determination. T: Torquing coils; M: standard mass: ROT: Main rotatable coil; N: Null detectors; A: servo amplifier.*

coils attached to the balance beam.[2] These three torquing coils apply a torque to the balance which changes the rotational velocity, completing the servo loop so that the generated voltage is nearly equal to the reference voltage. The torquing coils [two of which are seen as A in Fig. 1, the third being inside B] are arranged so that currents in them induce no flux into coil B.

The reference voltage is generated from the voltage drop across a 10 Ω resistor (V_{10}). A constant current supply delivers a current which passes through a 1 kΩ resistor and the 10 Ω resistor. The voltage drop across the 1 kΩ resistor, V_{1k}, is compared, using a null detector, to the voltage of a calibrated standard cell, V_{std}, and the current is adjusted to null ($V_{std} - V_{1k}$). By knowing the ratio of the two resistors and measuring the error voltage V_e, the voltage generated in the rotatable coil can be determined in NBS units. When the scale pan gets near the end of its travel, an optical detector switch reverses the direction of the current through the 1 kΩ − 10 Ω resistor chain thus reversing the polarity of the reference voltage. The servo then reverses the rotation direction of the balance, and data is recorded for the opposite sense of rotation. The total distance between turn-around points is about 2 cm.

The laser interferometer, mixer, and fringe converter system produces a pulse for each λ/4 (0.15 μm) of vertical travel of the scale pan. The pulse appears on one of two channels depending on whether the pan is moving up or down. An up-down counter is used to determine the relative position of the pan. A divide-by-32 counter provides a trigger once every 32 pulses which initiates the reading of a 1 MHz-clock-driven scaler and a reading of the error voltage by the analog-to-digital converter. The computer, interfaced to the measurement instrumentation involved, typically stores over 100 averages of such readings. Each average value of $V_e(i)$, $z(i)$, and $t(i)$ is the result of 50 individual measurements. One voltage integral determination consists of 10−12 consecutive repetitions of such measurements, with alternating sense of rotation, with the region near the turn-around points (where the balance rotation changes direction) discarded. At a later time the voltage integral is computed numerically from the stored averages $V_e(i)$ and $t(i)$.

3.3 Evaluating the Force Integral

The torque produced by ~3.3 mA in the rotatable coil will support a mass of ~5 g on the scale pan. Thus, when the current is in such a direction as to produce a downward force on the scale pan with the corner reflector on it, the opposite scale pan must have a 5 g counter mass on it for equilibrium. When the current in the rotatable coil is reversed, a calibrated mass of ~10 g is placed on the pan with the corner reflector. The weight of the 10 g mass gives us the force mg in Eqs. (3, 4).[3] It remains to determine $I_{NBS}(z)$, the change of current in the rotatable central coil which balances the change of a 10 g mass on the scale pan.

As shown in Fig. 3 the computer reads from the up-down counter the position of the corner reflector attached to the scale pan. This is compared to a preset position, z_i, and the computer generates, via the digital to analog converter, an analog correction signal which is

[2]For stable servo operation, currents proportional to the time integral and derivative of V_e are also applied.

[3]The acceleration of gravity is measured by a portable absolute gravimeter. See the paper by J. A. Hammond *et al.* in these proceedings.

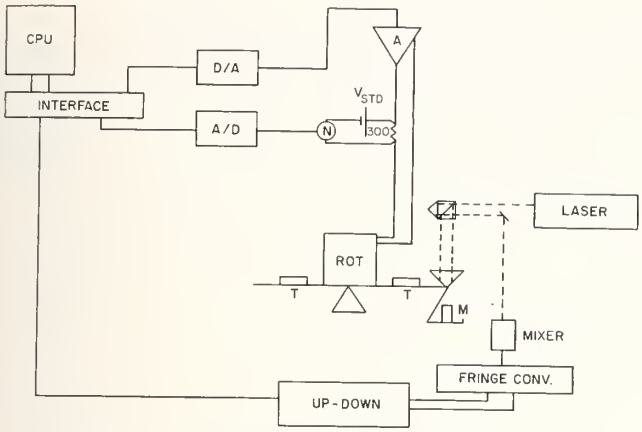


FIGURE 3. *Block diagram for force determination.*

used to drive current through the rotatable coil, completing a servo loop which holds the scale pan position constant to about λ/4. The current in the coil is measured by comparing the voltage drop it produces in a 300 Ω resistor with a standard cell voltage V_{std}. The difference voltage is read by the calibrated null detector (N) and recorded by the computer. The 10 g mass is removed, the current reverses to maintain the scale pan at the preselected position z_i, and the current is measured again. Ten to eleven current reversals are accomplished in 30 minutes and constitute one "force measurement."

In principle one should determine the balancing current $I_{NBS}(z)$ for a number of positions so as to calculate the integral of Eq. (4). We have found it more convenient, however, to infer the z dependence of the current from the z dependence of the voltage integral, evaluated over segments shorter than the full path length $z_2 - z_1$. Thus, at the present time, we directly determine $I_{NBS}(z)$ at one point only.

4. Analysis, Results, and Discussion

The voltage integral measurements for both senses of rotation and the force-current measurements for mass on and off are combined so as to cancel the effects of linear drifts (for example, drifts in magnetic field or rotating coil dimensions), thermal emf's, and null detector zero offsets. This procedure yields a set of alternating voltage integral and force-current determinations taken over a period of several hours on any one day. These are further combined to eliminate the effects of drift and used to calculate K_A according to Eq. (4).

Figure 4 shows 16 such determinations of K_A, plotted with respect to their unweighted mean, taken on four different days. (We specifically do not quote a value for K_A because of the possibility of a large systematic error as discussed below.) Each point represents about an hour of running time. The individual uncertainties are based on the standard deviation of the measurements (after correction for drift and offsets) from which each value of K_A was calculated. The dotted line shows the one standard deviation limits for the entire set—less than 2 ppm, and consistent with the individual uncertainties. These results indicate that a statistical uncertainty better than 1 ppm is easily achievable in a reasonable period of time. In addition, some of this scatter is probably due to instability in the current sources and ambient temperature, and to poorly damped swinging of the balance pans, problems which can easily be corrected in future work.

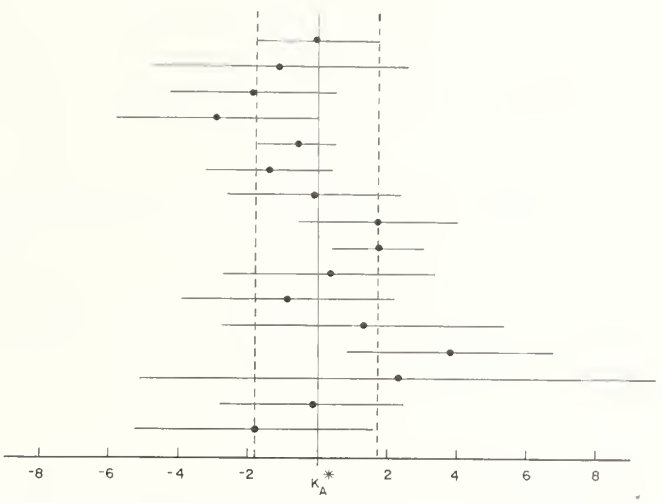


FIGURE 4. *Individual values of $K_A^* = K_A - K_A^{ave}$. Solid error bars indicate standard deviations of individual points; dotted lines indicate the standard deviation of the entire set. Points are plotted in chronological order.*

Systematic uncertainties due to calibration of the mass, resistors, standard cell, laser wavelength, time standard, as well as uncertainties in the determination of the acceleration of gravity and the buoyancy correction are all smaller than the statistical scatter and can all be made smaller than 0.1 ppm. The largest source of systematic uncertainty in the present work is in the correction for the bending of the balance beam under the load from the 10 g mass. This effect has been considered in previous ampere determinations [5]. The bending results in an error because the gravitational force of the 10 g mass is applied at the balance pan, while the magnetic force is applied near the center of the balance. The distortion of the balance is different in the two cases, so that a weighing error results. We have used data from earlier measurements [5, 6] on the balance distortion to estimate this. The correction was on the order of 20 ppm, but it is difficult to assess an uncertainty since the mass and mass distribution of the present balance are quite different from those existing when the original distortion measurements were made. The corrected value for K_A is within about 30 ppm of any of the previous high precision direct and indirect measurements as summarized in Ref. [2]. Future measurements of the ampere by this method will be designed so that a direct substitution of weight for electromagnetic force is carried out, thus eliminating the distortion error described above.

5. Conclusions

We have obtained preliminary results from a new, dynamic method for realizing the ampere. The statistical scatter of a single determination requiring about one hour of running time is less than 2 ppm. The statistics indicate that a realization of the ampere with sub-ppm accuracy is achievable. We are now designing [7] and constructing [8] an ampere balance using superconducting coils to generate the magnetic fields. The geometry will be such that the major part of the balance distortion problem will be eliminated. Forces and voltages will be larger by a factor of 100, so that relative statistical scatter due to measurements of these quantities, based on the experience with the present balance, should be at the parts in 10^8 level. We believe that this superconducting version of the experiment will lead to a final uncertainty in K_A on the order of 0.1 ppm.

References

[1] B. N. Taylor, W. H. Parker, and D. N. Langenberg, Rev. Mod. Phys. 41, 375 (1969); and E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[2] B. N. Taylor, Metrologia 12, 81 (1976).
[3] B. P. Kibble, these proceedings.
[4] B. P. Kibble and I. Robinson, NPL Report DES No. 40 (National Physical Laboratory, 1977).
[5] R. L. Driscoll, J. Res. Natl. Bur. Stand. 60, 287 (1958).
[6] R. L. Driscoll and P. T. Olsen, unpublished (1968).
[7] P. T. Olsen, M. E. Cage, W. D. Phillips, and E. R. Williams, IEEE Trans. Instrum. Meas. IM-29, 234 (1980).
[8] W. Y. Chen, J. R. Purcell, P. T. Olsen, W. D. Phillips, and E. R. Williams, presented at the 1981 CEC/ICMC Cryogenic Engineering Conference, August 10–14, 1981, San Diego, California.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A Feasibility Study of an Absolute Determination of the Magnetic Flux Quantum*

Ko Hara, Fuyuhiko Shiota, and Teruhiro Kubota

Faculty of Engineering, University of Tokyo, Bunkyoku, Tokyo, Japan

The principle and an experimental study of a new method to determine the magnetic flux quantum ϕ_0 is described. Magnetic energy proportional to ϕ_0^2 is substituted for and measured in terms of gravitational potential energy. A superconducting magnetic levitation system consisting of a persistent current coil and a superconducting floating body is employed for this energy substitution.

Key words: absolute determination; cryoelectronic measurement; magnetic flux quantum.

1. Introduction

The magnetic flux quantum ϕ_0 is essential in defining the unit of voltage in terms of the Josephson effect, yet to date its value is only known to an accuracy of a few ppm. This situation is one of the major difficulties in the adjustment of the fundamental physical constants [1]. An entirely new method to determine ϕ_0, hopefully with ~0.1 ppm accuracy, is proposed and the authors now have under way an experiment to investigate its feasibility.

A superconducting magnetic levitation system consisting of a persistent current mode coil L and a superconducting floating body M is an energy conserving system. Such a system with a current source i_s and a superconducting switch Sw is shown in Fig. 1. The conservation relation is

$$E(z) = E_m(z) + mgz + \frac{m\dot{z}^2}{2} = \text{constant}. \qquad (1)$$

Here, z is the position of the center of mass of M along the direction of gravity, $E_m(z)$ is the magnetic energy of the system and depends on z, m is the mass of M, and g is the acceleration of gravity.

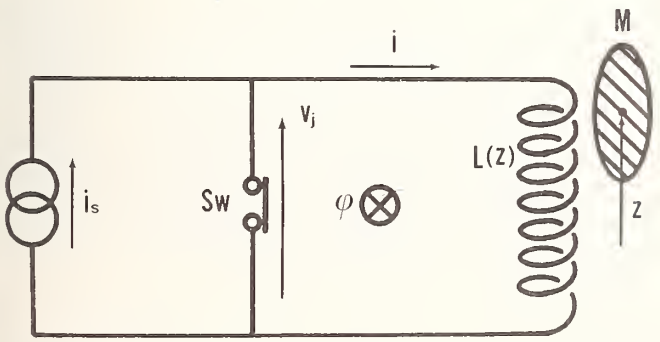


FIGURE 1. *Schematic diagram of flux-up and superconducting magnetic levitation in a constant magnetic flux mode.*

*Work supported by a Grant in Aid for Superconducting Quantum Electronics from the Ministry of Education.

When Sw is opened during the time interval 0 to t with the initial flux zero, the magnetic flux ϕ linking the system is

$$\phi = \int_0^t v_j\, dt\,, \qquad (2)$$

where v_j is the voltage across the coil terminals measured in terms of the Josephson voltage reference. ϕ is conserved after Sw is closed. Equation (2) describes the "charge up" of the system or what may be called the "flux-up" process. The current i through the coil is measured by a SQUID ammeter throughout this flux-up period and also in the persistent current state.

The equilibrium condition of M at ϕ is

$$\frac{d\,E_m(z_e)}{dz_e} = \frac{\phi}{2}\frac{d}{dz_e}\,i(z_e) = \frac{\phi^2}{2}\frac{d}{dz_e}\left[\frac{1}{L(z_e)}\right] = -mg\,. \qquad (3)$$

Here, z_e is the equilibrium z value of M and Eq. (3) can be understood as a functional relation between ϕ and z_e. Thus, this magnetic levitation system in equilibrium ($\dot{z} = 0$) can be used to substitute magnetic energy for gravitational energy.

2. Measurements with Josephson Electronics

Josephson electronics are used to make two important measurements:

(1) v_j is measured in terms of a Josephson voltage reference and is expressed as

$$v_j = f\,\phi_0\,, \qquad (4)$$

where f is obtained from the ratio (measured v_j/(defined ϕ_0 value). Thus, its uncertainty merely depends on that of the Josephson voltage standard, i.e., <0.1 ppm.

(2) The current i is measured by a SQUID ammeter and is expressed as

$$i = kn\,\phi_0\,. \qquad (5)$$

Here, it is assumed that the SQUID is calibrated by a Josephson voltage reference and a resistor calibrated absolutely. The integer n is the counted order of the SQUID and k is the current to flux conversion factor of the ammeter.

Combining Eqs. (2) and (4), ϕ fluxed-up over a time interval T is

$$\phi = \int_0^T v_j(t)\mathrm{d}t = \phi \int_0^T f(t)\mathrm{d}t = N\phi_0, \qquad (6)$$

Where N is the number of flux quanta linking the system. The self inductance of the system is thus simply $L = \phi/i = N/kn$ and the magnetic energy is $E_m = \phi i/2 = knN\phi_0^2/2$.

3. Methods of ϕ_0 Determination

With the above preparation, the following three methods to determine ϕ_0 are at least possible.

(1) Oscillation method: When the body M is in an oscillatory motion with upper and lower stationary points z_2 and z_1, respectively, where $\dot{z} = 0$, we have

$$E_m(z_1) - E_m(z_2) = \frac{Nk\,\phi_0^2}{2}[n(z_1) - n(z_2)]$$

$$= mg(z_2 - z_1),$$

and hence

$$\phi_0^2 = \frac{2mg(z_2 - z_1)}{Nk[n(z_1) - n(z_2)]}. \qquad (7)$$

The underlying principle of this method is easy to understand but it suffers from damping due to the viscosity of the ambient gas and also to eddy current losses in normal metals close to the system.

(2) Energy injection method: Prior to the experiment, the system is assumed to be in an equilibrium state characterized by $\phi_1 = N_i\phi_0$ and $i_i = k\phi_0 n(z_i)$. At $t = t_i$, Sw is opened and the system is fluxed-up until $t = t_f$ and the system reaches the state $\phi_f = N_f\phi_0$ and $i_f = k\phi_0 n(z_f)$. During the flux-up period, $v_j(t) = \phi_0 f(t)$ and $i(t) = kn(t)\phi_0$ are measured simultaneously and the energy injected into the system is $E_{inj} = \int_{t_i}^{t_f} v_j(t)i(t)\mathrm{d}t$. The total energy difference between the initial state (energy E_i) and the final state (energy E_f) is equal to E_{inj}, provided that the motion of M is so slow that the kinetic energy contribution during this period is negligible. Thus ϕ_0^2 is determined by

$$\phi_0^2 = \frac{mg}{k}\frac{z_f - z_i}{\int_{t_i}^{t_f} f(t)n(t)\mathrm{d}t + [N_i n(z_i) - N_f n(z_f)]/2}. \qquad (8)$$

This method is thus exact only at the limit of slow flux-up, but is more convenient at the present preliminary stage of the experiment.

(3) Equilibrium trajectory method: Integrating Eq. (3) over z_{e_1} to z_{e_2}, we find after some manipulation

$$\phi_0^2 = \frac{2mg\int_{z_{e_1}}^{z_{e_2}} \frac{\mathrm{d}z_e}{N^2 z_e}}{k\left[\frac{n(z_{e_1})}{N(z_{e_1})} - \frac{n(z_{e_2})}{N(z_{e_2})}\right]}. \qquad (9)$$

This method requires knowledge of $\phi(z_e)$ to sufficient accuracy that the accuracy of the numerical integration in Eq. (9) is adequate. It is, however, not bothered by dissipation.

One thing must be added about the uniqueness of $\phi(z_e)$. M has six degrees of freedom of motion, among which rotation about the vertical axis has no stable equilibrium when the system is axially symmetric. When it is slightly anti-symmetric, there may be multiple equilibrium positions separated by low barriers. In this case, equilibrium trajectories of M can interchange due to a slight shock and the $\phi(z_e)$ relation shows multiple values.

4. The Coil and Floating Body

The coil and the body are illustrated in Fig. 2 with their dimensions in mm. The conic body is made of Al coated with Pb. Inside the cone, a solid corner cube reflector is mounted for interferometry. The mass of M is 25.24 g.

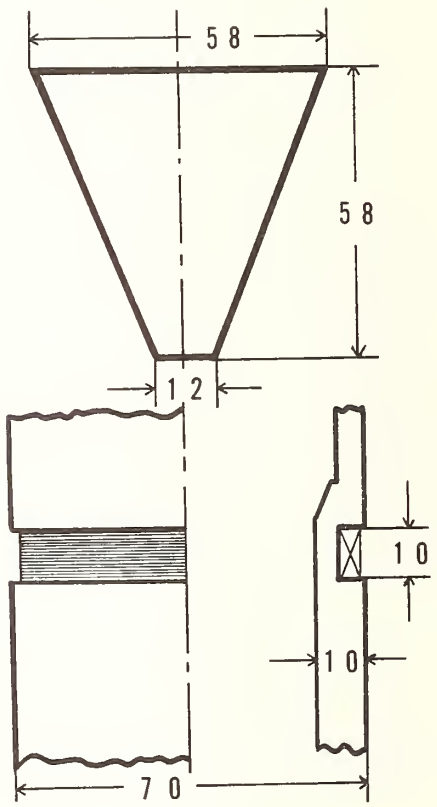


FIGURE 2. *Coil and floating body.*

The coil contains 720 turns in 18 layers of 0.25 mm diameter Nb wire on a phenol resin form, the bottom end of which is closed to prevent liquid He from getting inside it. The self inductance of the coil without M is 48.6 mH.

The fluctuating current through the coil in the persistent current mode without M is a measure of the ambient magnetic field fluctuations. This was measured using the SQUID and is shown in Fig. 3. The fluctuations are on the order of 10 ppm of the working current which is about 1 A.

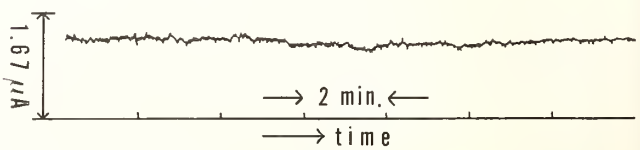


FIGURE 3. *Current fluctuation due to ambient magnetic field fluctuation.*

With M floating and a persistent current of about 1 A, the fluctuation current was measured and is shown in Fig. 4. It is about 100 ppm and corresponds to a vibration of M due to external vibrations on the order of 1 μm. This test is a good means to examine the Meissner state of both the coil and the material of which the levitated body is composed.

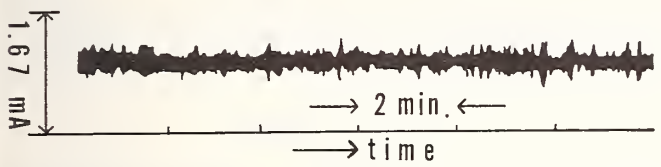


FIGURE 4. *Current fluctuation due to ambient mechanical vibration.*

5. Experiments

(1) Experimental set up, Figure 5 shows the experimental arrangement. The voltage v_j across the terminals and the current i through the coil were measured by a microvoltmeter (2% accuracy) and a SQUID ammeter, respectively. The coil was operated both in the flux-up mode and the persistent current mode by switching the thermal superconducting switch Th Sw which was shunted by a resistor R of 0.97 mΩ and which had a normal resistance of 30 mΩ. Accordingly, the flux-up time constant was about 30 s. The amount of flux was obtained by graphically integrating the recorded v_j. The SQUID ammeter was operated in digital and analog modes simultaneously and its sensitivity was 0.916 mA/count which could be interpolated to 10^{-3} counts. The laser interferometer was operated in both digital and analog modes, and its sensitivity was 2.56 μm/count in digital and 0.16 μm in analog.

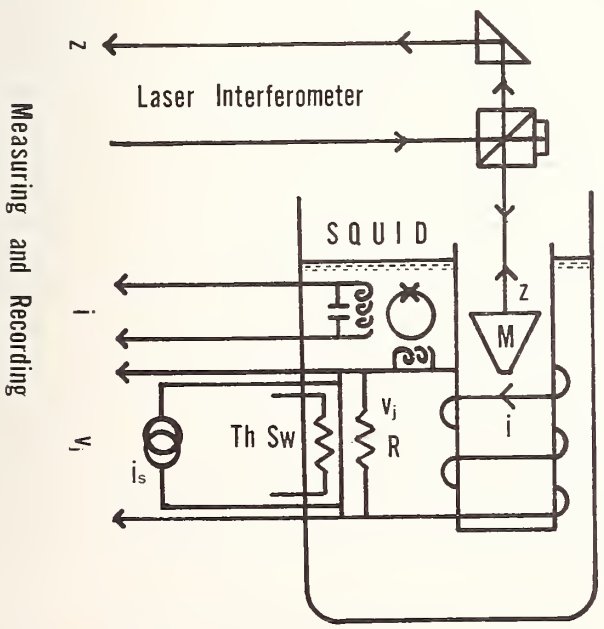


FIGURE 5. *Schematic diagram of experimental set-up.*

(2) Uniqueness of the equilibrium floating state. The equilibrium floating state can be specified by a set of values of ϕ, z_e, and i_e. Figure 6 shows the uniqueness of the floating equilibrium for one down and up round trip.

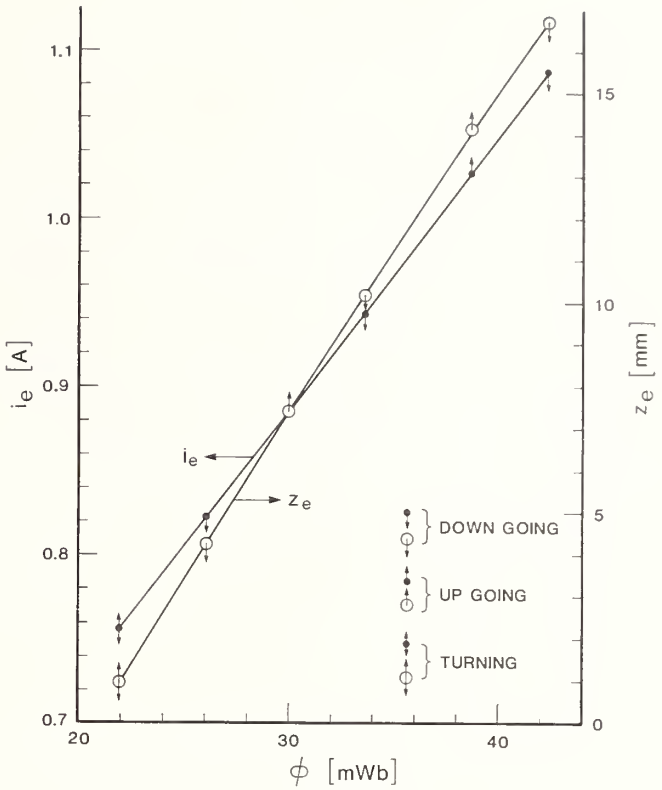


FIGURE 6. *Equilibrium floating state dependence on* ϕ.

In addition, the uniqueness of the state was examined incrementally with $\Delta i \sim 20$ μA and $\Delta z \sim 2$ μm. In this examination, the system was operated in a constant current mode and a small ac current was superposed. The resultant oscillation of M and the accompanying oscillation in i were observed to be uniquely correlated to the limit of the ambient fluctuations.

(3) Energy balance in energy injection method. The stored magnetic and gravitational energy during flux-up are denoted by ΔE_m and ΔE_g, respectively. The ratio $(\Delta E_m + \Delta E_g)/\Delta E_{inj}$ for the successive equilibrium points was computed for the data in Fig. 6 and was found to be 93.5, 93.7, 94.4, 97.9, and 96.4%, respectively. In this test, it was estimated that the kinetic energy loss was negligible.

(4) Estimation of ϕ_0. To confirm the overall consistency of the method, ϕ_0 was computed by the equilibrium trajectory method using the data of Fig. 6. The result, $\phi_0 = (2.07 \pm 0.04) \times 10^{-15}$ Wb, is in good agreement with the presently recommended value.

6. Discussion

According to the above, the authors believe the new method to be a hopeful way for measuring ϕ_0 to high accuracy. The main advantages of the method are:

(1) There is no need to machine the floating system precisely nor to measure its dimensions precisely. Only its stability during the experiment is required. In this sense, the system does need a unique floating state characterized by a set of values ϕ, z_e, and i_e. Accordingly, the system may be flexible as long as the uniqueness of these parameters is satisfied. This feature comes from utilizing the unique properties of Josephson electronics in contrast to Sullivan's proposal [2] in which precise measurement of the self inductance of the floating system is necessary.

(2) In this floating system, there is no source of static friction and so it is ideal for substituting magnetic energy with gravitational energy.

(3) In analogy with an ampere balance, the method can be viewed as a weber balance but with a mass far greater than that used in the ampere balance. This is a positive feature which should overcome the difficulty pointed out by Petley [3], if it exists.

There are many problems to be solved to reach the goal of an accuracy of a few tenths ppm;

(1) The most important is to find a superconducting material of ideal high H_{c_1}.

(2) A means should be provided to protect the apparatus from ambient magnetic and mechanical fluctuations.

(3) The reference corner cube reflector should be attached to the coil form. At present, it is attached to the top of the cryostat.

(4) The electronic measuring system should be improved. It is especially important to introduce a Josephson voltage reference and a resistance standard calibrated against a cross-capacitor.

Finally, the attainable error limit will be discussed briefly. Among the relevant quantities, the limiting factor is the resistance measurement which is restricted to 0.1 ppm. In principle, it is possible to exceed this limit for the other quantities. However, measuring the mass of M may be a great source of trouble because M has a large surface area which makes it a rather poor mass standard.

References

[1] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).

[2] D. B. Sullivan and N. V. Frederick, IEEE Trans. Magn. **MAG-13**, 396 (1979).

[3] B. W. Petley, NPL Report Qu52 (National Physical Laboratory, March 1979).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Work Done at the Mendeleyev Research Institute of Metrology (VNIIM) to Improve the Values of the Fundamental Constants

Yu. V. Tarbeyev

All-Union Research Institute of Metrology (VNIIM), Moskovsky prospect 19, 198005, Leningrad, U.S.S.R.

The work on fundamental constants carried out at VNIIM is a part of the overall effort in metrology aimed at improving measurement standards and systems. The results of work on the Avogadro constant, the gyromagnetic ratio of the proton, the Rydberg constant, and in associated research fields are reported. Improvements in the techniques for adjusting the values of fundamental constants as well as nuclear spectroscopy reference lines are discussed.

Key words: Avogadro constant; discrepant data; fundamental constants; gyromagnetic ratio of the proton; least-squares adjustment of the constants; Rydberg constant.

1. Introduction

The transfer from a system of standards based on the SI units to a system of standards based on the fundamental physical constants, i.e., to natural standards, should be considered as a major problem in the progress of metrology. Among the many tasks essential to the solution of the above problem, the decisive one is the task that involves the determination and improvement of the values of the fundamental constants. On the one hand the accuracy of these constants characterizes the depth of our understanding of physical phenomena and the laws governing their interrelation; and on the other hand, the development of new standards based on fundamental constants is the best way to carry out a qualitatively new realization of the units for physical quantities.

Advances in modern physics have already made it possible to develop standards for the second and the meter based on the method of frequency counting. Thus, a new standard for the meter based on a He-Ne laser stabilized using saturated absorption in $^{127}I_2$ was developed and put into operation at VNIIM in 1980. Additionally, the high accuracy of the experiments for determining the velocity of light in vacuum has enabled the metrology community to raise the question of how best to use this accurately known constant and of fixing its value by international agreement. This could eventually permit the adoption of new and more accurate definitions of the meter and the second based on the wavelength and the period of the same radiation. At present the problem of developing a "natural" standard for the kilogram is on the agenda. Indeed, in recent years such a possibility has opened up as a result of precision experiments on the determination of the Avogadro constant.

2. Improvement in the System of Standards for Electrical Quantities

Primary emphasis is being given at VNIIM to the improvement of the system of standards for electrical units [1] which are essential not only for the metrological assurance of electrical instruments, but also for refining the methods and measuring instruments used for a number of non-electrical quantities by converting them into electrical ones.

The precision of the absolute realizations of the major electrical units depends to a great extent on the accuracy of the determination of absolute values of the fundamental physical constants. For example, the accuracy of the USSR national primary standard for the volt based on the Josephson effect which has been newly developed at VNIIM depends mainly on the accuracy of the flux quantum $h/2e$ as determined at VNIIM in absolute (SI) units on the basis of its physical connection with other fundamental constants of a "mechanical" nature. Values of these constants are determined in terms of the basic SI units, with no electrical units used. The relevant equation is:

$$\left[\frac{h}{2e}\right]_{SI} = \frac{c}{4}\left[\mu_0 \frac{m_e}{m_p} \frac{\alpha}{R_\infty} \frac{M_p}{N_A}\right]^{1/2}, \qquad (1)$$

where c is the velocity of light in vacuum; m_e/m_p is the electron-to-proton mass ratio, equal to $(\mu_p/\mu_B)/(\mu_p/\mu_N)$ (μ_p is the proton magnetic moment, μ_B is the Bohr magneton, and μ_N is the nuclear magneton); M_p is the atomic mass of the proton; α is the fine structure constant; N_A is the Avogadro constant; and R is the Rydberg constant. The accuracy of $h/2e$ depends on that of the above constants with the value of the magnetic constant $\mu_0 = 4\pi \times 10^{-7}$H/m following from the definition of the base SI unit for current, the ampere.

In recent years theoretical and experimental studies have been carried out at VNIIM for improving the above fundamental constants, with the result that the magnetic flux quantum $h/2e = 2.0678410(10) \times 10^{-15}$ Wb was obtained for the realization of the absolute volt using Eq. (1), the relative standard deviation in this case being 5.6×10^{-7}. The new value of $h/2e$ defined at VNIIM differs from the previous one adopted by BIPM on the basis of the adjusted (CODATA 1973) value of $h/2e = 2.0678506 \times 10^{-15}$ Wb with a greater uncertainty (standard deviation), namely 2.6×10^{-6}.

The differences in the values of $h/2e$ adopted in the different countries which use the Josephson effect to realize their units of voltage is reflected in the differences in the

various national units of emf and voltage. A comparison of the primary standards maintained at the NBS and VNIIM (carried out at VNIIM in 1975) showed the mean difference to be 4.25 μV. The result agreed with the difference in the national units of the USSR and USA obtained from the 1973 international comparison at BIPM ($V_{NBS} = V_{VNIIM} - 4.22$ μV). The latest comparisons between the VNIIM standard cells and the NBS standards (carried out at NBS in 1976) gave a greater difference, i.e., 9 μV. Such a large discrepancy in the 1976 results was probably due to unfavorable conditions of transport of the VNIIM cells to the USA. The comparison of the VNIIM standard cells carried out at BIPM in 1980 gave a much better result: the mean difference in the emf values obtained at VNIIM and BIPM was 2.15 μV. It is obvious that further theoretical and experimental studies for the improvement and adjustment of the major fundamental physical constants should be conducted to provide international uniformity for the electrical and magnetic units. The importance of this work, which is to be carried out on a world-wide scale, is confirmed by the recommendations (E1 1976 and E1 1978) of the BIPM Consultative Committee on Electricity put forward by the Soviet delegation. The work carried out in this field at VNIIM is briefly outlined below.

3. Work on the Improvement of the Values of the Fundamental Constants

In order to improve the system of USSR national standards, work on the redetermination of a number of fundamental constants is currently underway at VNIIM.

3.1 The Proton Gyromagnetic Ratio, γ'_p

In 1980 a cycle of measurements for a new determination of γ'_p using the low magnetic field method was completed. The new data were obtained as a result of the work to determine γ'_p conducted from 1958 to 1968 [2, 3].

The proton gyromagnetic ratio γ'_p is determined from the ratio of the proton free precession frequency ω'_p for a spherical sample of distilled water placed in the magnetic field of a standard current carrying coil, to the magnetic induction of this field. The magnetic induction is found using the value of the coil constant K as determined by measuring the diameters of and spacings between the coil windings as well as from the strength of the current I flowing through the coil. The current I is measured by balancing the emf of a standard cell E against the voltage drop across a standard electrical resistor R connected in series with the windings of the standard coil. Thus, the value of γ'_p is found from the expression

$$\gamma'_p = \frac{\omega'_p R}{KE} .$$

For the redetermination of γ'_p over the period 1977 to 1980, a new apparatus was developed. It is comprised of a new type of standard magnetic induction coil which permits the magnetic induction to be increased by a factor of 5 to 10, the volume of the region with a magnetic field of the required uniformity by a factor of 50, and the effect of the turns diameter measurement uncertainty to be more than halved, in comparison with the Helmholtz coils used before.

New equipment was also developed for the contact measurement of the turn-to-turn spacing of the standard coil, for stabilizing and measuring the strength of the current I for the nuclear precession frequency measurements, and for compensating the earth's magnetic field. All this made it possible to reduce the corresponding components of the uncertainty in the determination of γ'_p by a factor of 3 to 5 in comparison with the data obtained in the previous VNIIM γ'_p experiment.

The new standard magnetic induction coil [4] which comprises a four-section solenoid on a 300 mm diameter, 500 mm long cylindrical quartz form sets up a 5.7×10^{-4} T magnetic field with a current through the coil of 1 A. The experimentally determined temperature coefficient of the constant K is $0.56 \times 10^{-6} I/°C$ and the nonuniformity of the magnetic induction within the working volume (a 55 mm diameter sphere) is no more than 1×10^{-6}. The coil turn-to-turn spacing was measured by a laser length measuring device for each pair of turns symmetrical with respect to the coil center along eight generating lines. The systematic error bound is no more than $0.2L$ μm, where L is the measured distance in meters. The measurements were computer-controlled with 100 observations made at each point. The systematic error in the turn diameter measurements which were performed over each section for each turn is no more than 0.8 μm. The wire diameter measurement error is no more than 0.1 μm. Temperature corrections to the standard coil constant K were determined from resistance measurements of auxiliary windings wound on the quartz form.

The nuclear magnetic resonance frequency was measured by the free nuclear precession method using separated receiving and polarizing coils. The spherical sample of distilled water was 40 mm in diameter. The time required for a single measurement was 2 s and the relative standard deviation of the observations was no more than 2×10^{-7}. The effect of the magnetic properties of the nuclear precession converter components, of the attenuation of the nuclear precession signal amplitude, and of the screening effect of the windings and forms on the accuracy of the precession frequency measurements, was determined experimentally. The results of the analysis of the above error components in the determination of γ'_p and other components estimated to be 1×10^{-7} or larger are given in Table 1. The proton gyromagnetic ratio obtained in terms of the USSR national volt and ohm units is found to be

$$\gamma'_p = 267512.57 \times 10^3 \, s^{-1} \, T_{VNIIM}^{-1}$$

The total error bound Θ of the proton gyromagnetic ratio determined in terms of the USSR national units with a confidence level of $P = 0.99$ is no more than

$$\Theta = 1.4 (\Sigma \Theta_i^2)^{1/2} = 1.5 \times 10^{-6} .$$

Estimated in terms of a standard deviation ($P = 0.67$), the same error is 0.6×10^{-6}.

3.2 The Avogadro Constant, N_A

One of the constants governing the accuracy of $h/2e$ as obtained from Eq. (1) is the Avogadro constant. The latest advances in the measurement of N_A, which made it possible to reduce the measuring error by a factor of over 10, hold much promise for the further refinement of its value. Impressive progress in atomic mass measuring techniques and a new approach to the solution of the density problem reduced the uncertainty of these measurements to the level of $\sim 5 \times 10^{-7}$. These achievements as well as the development of an x-ray-optical interferometer, which permitted the monocrystal lattice constant to be measured with an accuracy of better than 1.5×10^{-7}, provided the basis for the precision measurement of N_A. Using this procedure, the Avogadro constant is determined from

TABLE 1. *Error bounds for the proton gyromagnetic ratio (P = 0.99).*

Contributing factors	Relative uncertainty ($\times 10^{-7}$)
1. Standard coil turn-to-turn spacing measurement	3
2. Standard coil turn diameter measurement	6
3. Standard coil wire diameter measurement	1
4. Uncertainty in the shape of the coil winding	3
5. Correction for the current distribution over the wire cross-section	1
6. Correction for positioning the connecting leads and return lead	2
7. Magnetization of standard coil form	2
8. Standard coil temperature measurement	2
9. Measurement of precession decaying signal frequency	2
10. Correction for screening effect of windings	2
11. Calibration of standard cells	2
12. Instability of standard cells	2
13. Standard cell temperature measurement	2
14. Spurious emf's	1
15. Standard resistance calibration	3
16. Standard resistance calibration measurement	2
17. Correction for load dependence of standard resistance	2
18. Correction for the influence of the standard coil on the earth's magnetic field compensation system	1
19. Undercompensation of the earth's magnetic field	1

$$N_A = \frac{Mf}{\rho d_0^3 \Phi} ,$$

where M is the atomic mass of the substance (silicon), f is the number of atoms per unit cell, ρ is its density, d_0 is the crystal lattice constant, and Φ is the unit cell geometric factor which is unity for cubic crystals such as silicon.

Higher accuracy in determining the atomic mass and density can be achieved with monocrystal silicon enriched by ^{28}Si up to 99.99 percent and grown by a special process that produces silicon free from defects, dislocations, and impurities. The development of new density standards produced from monocrystal type substances featuring a stable structure is the basis for higher accuracy measurements of ρ.

Solving these problems as well as obtaining a higher accuracy in the optical and x-ray measurements will lead to a decrease in the (relative) uncertainty of the Avogadro constant to about 1×10^{-7}. A still higher accuracy in the determination of N_A can be obtained from the accumulation of data gained in equally precise experiments carried out in different laboratories and by different methods. This will make it possible in the final analysis to reduce the uncertainty in N_A by approximately one more order of magnitude. To solve the problems specified above, work is being carried out at VNIIM and a number of other institutions on the development of monoisotope-monocrystalline silicon, x-ray optical interferometers, and density standards based on samples perfect in shape (sphere, cube) and made from quartz or glass. It is anticipated that the use of monocrystalline structures and an x-ray optical interferometer will make it possible to eventually transfer to a natural standard for the unit of mass.

3.3 The Rydberg Constant, R_∞

Though further refinement of the Rydberg constant is not immediately required in order to use Eq. (1), it is of great importance for improving our knowledge of the whole set of fundamental physical constants. The Rydberg constant, R_∞, describes the binding energy between an electron and the atomic nucleus and as such is a major fundamental constant. A promising method of precisely measuring R_∞ is the two-photon absorption technique used with the $1S \rightarrow 2S$ transition in hydrogen [5] in a standing wave field. The elimination of first order Doppler line broadening enables one to obtain narrow resonances whose relative width is $\sim 10^{-11}$. This opens up the way for a measurement accurate to $< 10^{-10}$ (the uncertainty of the present value of R_∞ is 1×10^{-9}).

In order to make such an experiment possible, a number of technical problems must be solved: (1) development of a 243 nm ultraviolet continuous-wave tunable laser featuring a single frequency and fair radiation power; (2) development of a system for frequency measurement over the optical range corresponding to the transition selected; (3) development of a hydrogen atomic beam source featuring a high intensity in the region of interaction with the radiation, a greater extent of dissociation, and as low noise as possible; and (4) development of a system for recording metastable hydrogen atoms in the $2^2S_{1/2}$ state which is essential to obtain a maximum signal-to-noise ratio.

The above method for measuring R_∞ is being employed at VNIIM and the Institute of Thermal Physics (Siberian Branch of the Academy of Sciences of the USSR). A thermal-type hydrogen beam source whose intensity is $\sim 10^{15}$ atoms/(cm^2·s) at 10 to 15 cm and a degree of dissociation of about 95 percent at some 2700 K, has been developed at VNIIM. Such parameters for the source make it possible to reduce contributions to the two-photon resonance line width due to the second-order Doppler effect and line broadening due to the time of interaction between the beam atoms and the laser radiation. This enables us to obtain the resulting relative resonance linewidth of about 8×10^{-10}. A further increase in the atomic beam density over the interaction region using multi-channel shapers (capillary arrays) and through a decrease in the effective temperature of the beam will

make it possible to obtain a two-photon absorption linewidth of no more than 3×10^{-11}. This will open up the way for the determination of the Rydberg constant with an accuracy of the same order.

3.4 Magnetic Flux Quantum, $h/2e$

To improve the value of the magnetic flux quantum $h/2e$, a levitated superconducting mass method which involves a comparison between mechanical and electrical energy is being experimentally studied [6, 7]. An electromechanical system under development at VNIIM is an attempt to develop this method for measuring $h/2e$ in absolute units to an accuracy of about 10^{-6}.

The levitated body, a thin-walled niobium cone [(1) of Fig. 1)] with a vertex angle of 15°, is mounted on the thin-walled niobium tube guide (2). The cone can move together with its guide in the direction preset by the axes of the two superconducting magnetic bearings [(3) and (4)]. The bearings comprise superconducting coils wound on niobium split forms. The bearing axes are vertically oriented and the 300 mm spacing between the bearings ensures that the vertical deflection of the guide is no more than one minute of arc. The upper bearing and part of the guide are furnished with special slots designed to prevent the body from rotating around the vertical axis. The windings of the support coil (5) located inside the cone are composed of Bitter-type disks. The disks, made of 50 μm thick niobium foil, are insulated from each other with mica spacers. The coil outside the conical surface has the same cone angle as the levitated body. Located inside the upper part of the guide tube is a corner-cube reflector of a laser interferometer, and its lower part accommodates an electrodynamic device made up of a fixed feedback coil inserted in the gap of a permanent magnet. This apparatus enables the experiment to be carried out by the method put forward by B. P. Kibble [8].

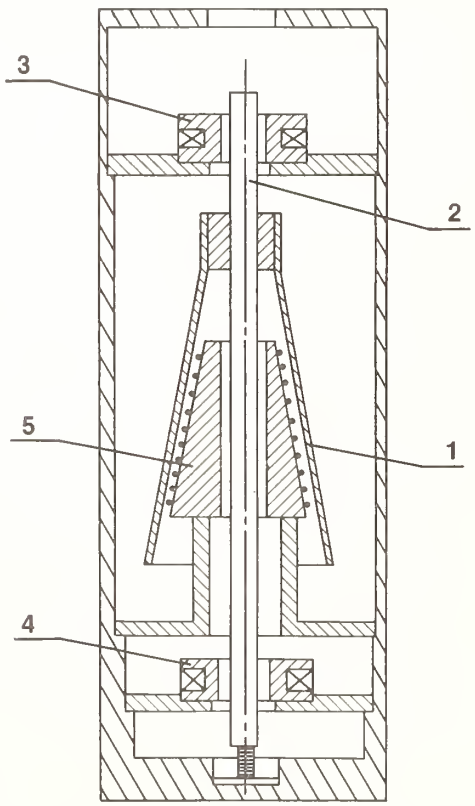


FIGURE 1. *Schematic diagram of the levitated superconducting body portion of the apparatus for the new VNIIM experiment to determine the flux quantum h/2e in SI units. 1: levitated body of thin-walled niobium; 2: thin-walled niobium guiding tube; 3 and 4: superconducting magnetic bearings; 5: support coil winding.*

The electromechanical system is placed in a niobium superconducting shield located in an evacuated chamber inside a helium cryostat which is installed on a shock-mounted platform. The levitated mass is about 1 kilogram and the support surface area is ~180 cm^2. Current from a SQUID-regulator and the feedback current from the laser interferometer electronic circuit are fed to the support coil, which makes it possible to fix the levitated mass at distances from the support coil that are an integral number of interferometer fringe spacings.

A standard resistance is switched into the support coil circuit and the voltage across its terminals is measured in terms of $hf/2e$ using a Josephson voltage standard. With the levitated mass moving, the emf across the support coil is measured in the same fashion. The principal uncertainty components have been discussed in Refs. [7] and [8].

3.5 An Improved Method for Measuring the Combination of Fundamental Constants $\gamma'_p(\mu'_p/\mu_N)$ Using a Static Mass Spectrometer

The procedure under development at VNIIM is based on the mass spectrometric method for measuring the above constants worked out at the Ioffe Physico-Technical Institute (PTI) [9]. Unlike the PTI method, where a calibrated increment of energy was imparted to ions and the corresponding change in the diameter of their orbit was measured, the VNIIM procedure involves the measurement of the total energy of the ions and the total diameter of their orbit. In the method, measurement of these basic parameters is based on full deceleration of ions in a gap with a calibrated potential difference taking into account the distribution of ions over the width of the analyser slits. Calculations show that the method under development provides a more accurate measurement of the above parameters in comparison with the PTI method [9], and will allow measuring the set of fundamental constants, $\gamma'_p(\mu'_p/\mu_N)$ to better than 0.5×10^{-6}. Since to date the proton magnetic moment in nuclear magnetons, μ'_p/μ_N, has been measured to 0.43×10^{-6}, the method under development will make it possible to obtain a new value of γ'_p accurate to 0.66×10^{-6}. The units for γ'_p obtained in this method correspond to those of γ'_p as obtained in the low magnetic field method while phsyically the experiment is carried out in a strong field. This distinguishing feature of the mass-spectrometer method should help to eliminate some of the discrepancies concerning the proton gyromagnetic ratio data [10] when adequate accuracy is attained.

4. Refinement of the Data Handling Procedure for the Adjustment of the Fundamental Constants

The growing importance of the fundamental constants for many problems in science and technology, especially for metrology, places more stringent requirements on the reliability of a self-consistent set of their values. Considerable attention has always been given to this particular problem and it was extensively discussed at the First International Conference on Precision Measurement and

Fundamental Constants [11]. It is obvious that the reliability of the adjusted values of the fundamental constants is determined not only by that of the results obtained in the different experiments which provide the basis for the adjustment, but also by the procedure used for handling the data that culminates in a self-consistent set of values for these constants. A distinguishing feature of the data handling procedure employed for the fundamental physical constants is the necessity to solve the problem of their inconsistency. Investigations [12] show that the reliability of a final set of adjusted values for the fundamental constants can be increased appreciably by the choice of an optimum way for solving this problem. Analytical studies carried out at VNIIM of the generalization of the methods for eliminating inconsistencies among the data on the fundamental constants (methods which have been improved over the years by R. T. Birge [13], and E. R. Cohen and J. W. M. DuMond [14], B. N. Taylor, W. H. Parker, and D. N. Langenberg [15], and E. R. Cohen and B. N. Taylor [16]), have made it possible to draw a number of conclusions.

There are several methods at present for eliminating inconsistent data, among which two main approaches may be distinguished. One method consists of increasing the variances for the whole set of data by the same general scale factor equal to the Birge ratio [13]:

$$R_B^2 = \chi_{\text{cal}}^2 / F \ , \qquad (2)$$

where χ_{cal}^2 is the calculated value of χ^2 and F is the number of degrees of freedom for the set of data. The use of the Birge ratio implies that the data variances are assumed to be known with equal accuracy (i.e., all are equally reliable), and that the general scale factor is equally well known.

The other method consists in the analysis of variance [14] which leads to the choice of a particular sub-group of consistent data out of the entire set of input data. In the 1973 adjustment a combination of these methods was used, that is, part of the data was excluded from the treatment and for individual data groups the uncertainties were expanded by the proper Birge ratio. Introduction of a set of expansion factors $R_i^2 = \sigma_i'^2/\sigma_i^2$ for the *a priori* uncertainties σ_i^2 is common to the above methods. These expansion factors are introduced so as to meet exactly or approximately the condition

$$\chi^2(\sigma_2'^2) = F \qquad (3)$$

for the new uncertainties $\sigma_i'^2$.

Since there is no objective way of selecting specific values for the factors R_i^2 (at that stage of the adjustment when the data have already been subjected to preliminary analysis and are ready for statistical treatment), all values should be considered as more or less equally probable. Thus the main idea underlying the improvement in the adjustment procedure consists in the randomization of a set of possible values for the factors R_i^2 or equivalently, the *a posteriori* variances $\sigma_i'^2$.

If no rigid assumptions are made as to the statistical mechanism governing the measured values χ_i and σ_i^2, the method for their generalization becomes evident when their common distinctive features are considered. This method can be expressed by Eq. (3) and by

$$\prod_i \Phi(R_i^2) = \max ,$$

where $\Phi(R_i^2)$ is the probability density function of different values for R_i^2. Choosing Φ in the form of different distributions, one can easily derive both the usual adjustment methods and their generalizations. For example, the method characterized by Eq. (2) corresponds to the uniform distribution; the analysis of variance method to the delta function shaped distribution; and so on. The algorithm

$$\sum_i (1 - \sigma_i'^2/\sigma_i^2)^2 = \min \qquad (4)$$

corresponds to the assumption (most natural from the physical point of view) that within the body of data under treatment, lower probabilities correspond to greater relative errors in the *a priori* variances [17], with the normal distribution adopted for Φ.

The results obtained using this algorithm [i.e., Eqs. (3) and (4)] and some modified versions have been studied at VNIIM and NBS within the framework of their joint cooperative program in metrology. It was shown that the application of this algorithm to the input data of the three recent adjustments of the constants (1963, 1969, and 1973) changes the adjusted values by more than the standard deviations specified for them, with the changed values in good agreement with the improved adjusted values found by taking into account the most recent experimental results. More specifically, it was found that the relative deviations of the adjusted values calculated using Eqs. (3) and (4) from the improved (i.e., most recent) values of the constants (also calculated from these equations) are considerably less and have a smaller spread than those obtained by the analysis of variance (Fig. 2). Thus, the results of the studies carried out at VNIIM show that the suggested algorithm, which generalizes the usual methods for adjusting the constants, substantially increases the reliability of the final set of the adjusted values for the fundamental physical constants.

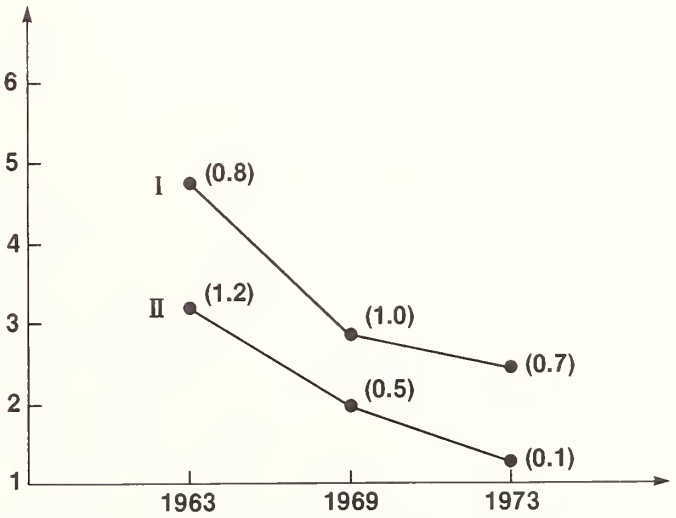


FIGURE 2. *Average relative deviations* $t = |\bar{A}_{adj} - \bar{A}_{impr}| \sigma_{adj}$ *of the adjusted values of the constants. Here* $\bar{A}_{adj}$ *is the adjusted value of the constant;* σ_{adj} *is its one standard deviation uncertainty resulting from the adjustment; and* $\bar{A}_{impr}$ *is the current improved adjusted value resulting from the application of Eqs. (3) and (4) (i.e., the VNIIM algorithm) to the best data currently available on the fundamental constants. The curve marked I is for* $\bar{A}_{adj}$ *taken to be the results of the 1963, 1969, and 1973 adjustments, respectively [14, 15, 16]. The curve marked II is for* $\bar{A}_{adj}$ *taken to be the results of the application of the VNIIM algorithm to all of the data available at the time of the 1963, 1969, and 1973 adjustments, respectively. The numbers in parentheses give the spread in t.*

5. Nuclear Spectroscopic Reference Lines

The last five years have been marked by a substantially increased precision in the determination of energies and relative line intensities of nuclear electromagnetic radiation. Therefore, the problem of choosing the energy reference lines (γ-ray standards) needed to make the relative measurement process convenient and sufficiently precise has become particularly pressing. However, too great a number of reference lines could make numerous intercomparisons necessary, which would lead to difficulties in evaluating the situation at any instant of time. An obvious way out is to establish a system of reference lines as first suggested by Associate Member of the Academy of Sciences of the USSR, B. S. Dzhelepov in his book *Methods for Development of Complex Decay Schemes* (1974). In such a system, a particular reference line is taken as the main line and other reference lines are compared primarily to the main one.

A classification of energy reference lines developed at VNIIM is based on the principle that the main metrological line (presently ^{86}Kr, $\lambda = 6058$ Å) should be the first order reference line. Draft rules for introducing the second, third, and fourth order reference lines for nuclear spectroscopy have also been proposed. Simultaneous analysis of all of the requirements which a reference line of the second order should satisfy, has led to the conclusion that the best choice is the 411.8 keV, ^{198}Hg γ-ray line though there are a number of γ-ray transitions of ^{169}Yb, ^{192}Ir, etc., which are now measured with nearly the same accuracy. However, reasons why the x-ray line $WK\alpha 1$ should be temporarily regarded as the reference line of the second order have been discussed.

Reference lines of the third order should be determined by direct comparison with those of the first and second order. They should be compared at the current level of precision by methods based on well known physical laws. Concurrently they can be introduced using the two reference lines of the third order already adopted and Ritz's rule: $E_1 + E_2 = E_3$. Based on these rules, 148 transitions were classified as reference lines of the third order in 1980.

Rules for introducing reference lines of the fourth order have made it possible to introduce 87 reference lines, and this number is growing rapidly. The main body of information is provided by Ge(Li) spectrometers used for measuring small energy differences.

Another area of interest is reference ratios of intensities. In nuclear spectroscopy there are methods that allow the spectral line intensity ratios to be measured with an accuracy of 2–3 percent. However, there is an indirect method which permits these ratios in some cases to be determined to 2×10^{-4} percent. At VNIIM 46 reference ratios have been calculated to better than one percent [18].

The experience gained at VNIIM shows that the scope of work carried out on the fundamental physical constants makes it necessary to pose the question of coordinating these studies within the framework of the various international metrological organizations. First of all, we would like to suggest for discussion a system of energy reference lines (γ-ray standards) developed in the USSR with a view to working out an international system of γ-ray standards.

In conclusion, the author would like to express his gratitude to the following researchers at VNIIM who prepared material for this review: K. A. Krasnov, A. B. Mamyrin, T. B. Rozhdestvenskaya, N. V. Studentsov, and S. A. Shestopalova.

References

[1] Yu. V. Tarbeyev, T. B. Rozhdestvenskaya, I. N. Krotkov, and O. P. Galakhova, in *Measurement for Progress in Science and Technology, ACTA IMEKO 1979*, Vol. 3, Ed. by Gy. Striker, J. Solt, and T. Kemeny (North Holland, Amsterdam, 1980), p. 795.

[2] B. M. Yanovskii and N. V. Studentsov, Izmeritel. Tekhn. 6, 28 (1962). [English transl.: Meas. Tech. 1962, 482 (1962).]

[3] N. V. Studentsov, T. N. Malyarevskaya, and V. Ya. Shifrin, Izmeritel. Tekhn. 11, 29 (1968). [English transl.: Meas. Tech. 1968, 1483 (1968).]

[4] N. V. Studentsov and V. N. Khorev, Abstracts of Reports for the All Union Scientific and Technical Session on Problems of Magnetic Measurements and Magnetic Measuring Instruments, VNIIM, Leningrad, 1972 [in Russian].

[5] L. S. Vasilenko, V. P. Chebotayev, and A. V Shishayev, Pis'ma Zh. Eksp. Teor. Fiz. 12, 161 (1970). [English transl.: JETP Lett. 12, 113 (1970).]

[6] D. B. Sullivan and N. V. Frederick, IEEE Trans. Magn. MAG-13, 396 (1977).

[7] E. T. Frantzouz, Metrologiya (USSR) No. 1, 43 (1979) [in Russian].

[8] B. P. Kibble and I. Robinson, NPL Report DES 40, (National Physical Laboratory, Sept., 1977).

[9] B. A. Mamyrin, N. N. Aruyev, and S. A. Alekseenko, Zh. Eskp. Teor. Fiz. 63, 3 (1972). [English transl.: Sov. Phys. JETP 36, 1 (1973).]

[10] B. A. Mamyrin, Metrologiya (USSR), No. 4 (1981) [in Russian].

[11] *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.), Spec. Publ. 343 (Aug. 1971), pp. 493–525.

[12] V. S. Touninsky, Metrologiya (USSR), No. 1, 13 (1978) [in Russian].

[13] R. T. Birge, Phys. Rev. 40, 207 (1932).

[14] E. R. Cohen and J. W. M. DuMond, Rev. Mod. Phys. 37, 537 (1965).

[15] B. N. Taylor, W. H. Parker, and D. N. Langenberg, Rev. Mod. Phys. 41, 375 (1969).

[16] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).

[17] V. S. Touninsky and S. V. Kholin, Metrologiya No. 8, 3 (1975) [in Russian].

[18] B. S. Dzhelepov and S. A. Shestopalova, *Nuclear Spectroscopic-Ray Standards* (Atomizdat, Moscow, 1980) [in Russian].

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Realization of a Josephson Potentiometer

M. Koyanagi, T. Endo, and A. Nakamura

Electrotechnical Laboratory, 1-1-4 Umezono, Sakura-Mura, Niihari-Gun, Ibaraki, 305 Japan

Preliminary experiments have been done on a new Josephson potentiometer. The emf of a standard cell is determined with an uncertainty at the 10^{-8} level by using a Josephson voltage of 100 mV which is generated by multiple Josephson junctions.

Key words: ac Josephson effect; Josephson junction array; Josephson voltage standard; microwave strip line; tunnel junction.

1. Introduction

The constant voltage produced by the ac Josephson effect has two remarkable features: (1) universality as a reference voltage; and (2) its variability. Therefore, it is possible to make a high precision potentiometer using the ac Josephson effect. In fact, the Josephson voltage standard systems used at present in several countries are based on these characteristics [1]. However, as the Josephson voltage used in these systems is in the range from 1 to 10 mV, a voltage divider which has a fixed divider ratio is necessary for obtaining the standard voltage of about 1 V. If we can generate a Josephson voltage as high as 1 V, we will be able to make a high precision potentiometer without depending on the divider which may be the main origin of the uncertainty for such a high precision potentiometer. Such a potentiometer will also serve as a Josephson voltage standard system with higher precision than those standard systems used at present because its uncertainty will simply be attributable to the Josephson voltage and the voltage sensitivity of the null detector.

A few attempts have been reported on the generation of Josephson voltages where the aim was 100 mV or higher [2 to 5]. However, none of them were completed as a system. In the present paper we report preliminary experiments on determining the emf of a standard cell by using a Josephson voltage of 100 mV. In the experiments the divider ratio is calibrated automatically in the determination of the standard cell voltage.

2. Experiment

The present Josephson potentiometer consists of the Josephson voltage generator of 100 mV, a voltage divider which is composed of ten resistors of 45 Ω, a constant current supply, and two photo-cell type galvanometers as null detectors. In order to understand the principle of the potentiometer, let us explain the method for determining the emf of a standard cell (see Fig. 1). At first, using galvanometer 1, the current from the constant current supply is adjusted so that the voltage across the divider becomes equal to the unknown voltage such as the emf of a standard cell of nearly 1 V. The standard cell voltage is simply given by the sum of the voltages produced in each resistor of the divider. The voltage across each resistor (about 100 mV) can be determined in terms of the Josephson voltage by adjusting the microwave frequency. The uncertainty of the present potentiometer would be attributed to the Josephson voltage of 100 mV, the voltage sensitivity of the galvanometers, and the stability of the voltage across the divider (dependent upon the stability of the constant current supply) in the time interval necessary for determining the voltage. In this system, the divider ratio is given automatically in the course of the voltage measurements as long as the current through the divider is kept constant. An unknown voltage lower than 1 V is determined by using a smaller number of resistors with the same procedure as mentioned above.

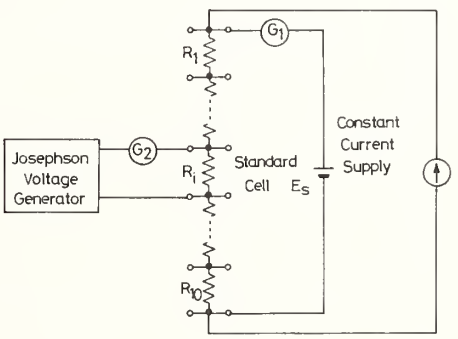


FIGURE 1. *Schematic diagram of the Josephson potentiometer for determining the emf of a standard cell.*

The principal part of the Josephson voltage generator is an array of twenty Josephson junctions connected in series for generating 100 mV. Microwaves are supplied to the twenty junctions simultaneously from a single microwave generator, although we need twenty dc current sources for the independent biasing of each junction. The biasing system contains a monitoring system which has an oscilloscope, an ac current supply, and a digital voltmeter. The step structure around a dc bias point of any one of the twenty junctions is monitored on the oscilloscope, and the total dc voltage is read by the digital voltmeter. The monitoring system is disconnected when the comparison is made between the Josephson voltage and the voltage across a resistor.

The junctions of Pb(In, Au)-oxide-Pb(Au) are fabricated on a silicon wafer by the procedures of photolithography, vacuum evaporation, rf-sputter cleaning, and rf-oxidation. The typical size of a junction is 0.9×0.05 mm^2, where the 0.9 mm length corresponds approximately to half of the wave length at 10 GHz.

Microwaves of 9.3 GHz are fed through a 50 ohm strip line made of evaporated superconducting film, and a copper plate attached to the waveguide wall as shown in Fig. 2. The array of twenty Josephson tunnel junctions is incorporated into the superconducting film. One of two ends of the strip line is connected to an antenna inserted into the X-band waveguide. Another end is terminated to ground through a resistor of 50 Ω so that the reflected

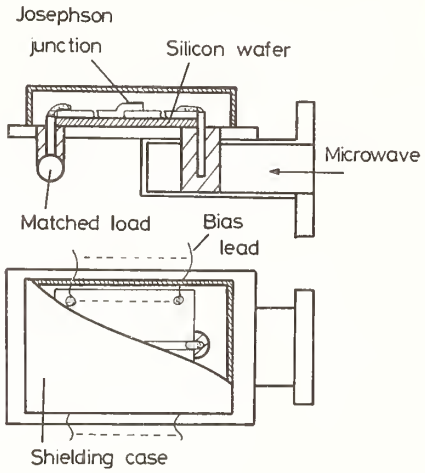


FIGURE 2. *Sample mount.*

microwave power from the end of the strip line becomes small and all junctions are irradiated with nearly equal microwave power without a dependence on junction position. The X-band strip line filter is inserted between the junction and the terminal for the dc lead as shown in Fig. 3. With this filter, mutual interactions are considerably reduced between the strip line for the microwaves and the leads for the dc supply.

For the potentiometer experiment, one needs large widths of the constant voltage steps at reasonably high voltages. In Fig. 4, the width and the voltage of the step used are shown for each junction at various positions along the strip line when the incident power to the cryostat is 210 mW and the total dc voltage is 101.9 mV. The closed circles and solid bars show the bias voltages and the current widths. Two junctions, No. 17 and 18, are missing because they showed poor coupling to the microwaves.

In spite of our intent of irradiating all the junctions with uniform microwave power, the microwave standing wave in the strip line prevents such uniform irradiation. This may be due to the existence of constrictions in the superconducting strip line near the Josephson junctions.

The divider, which consists of ten resistors with values of resistance within 14 ppm, is placed in an oil bath which is temperature regulated within 0.005 °C at 23 °C. The temperature coefficient of the resistors is 0.6 ppm/ °C. The self heating effect of the resistors is negligible. The current supply to the divider is made by a method similar to that of Finnegan and Denenstein [6] and has a drift of 2 to 3 ppm/hr.

The working standard cell used is contained in a temperature regulated air bath which has temperature stability of 0.001 °C at 30 °C as long as the room temperature is kept at (23 ± 0.5)°C. The air bath is placed in another shielded room. The comparison between the voltage across the divider and the voltage of the standard cell is performed through a cable of about 15 m. Before and after the experiments, the standard cell is calibrated against three other standard cells in the same air bath.

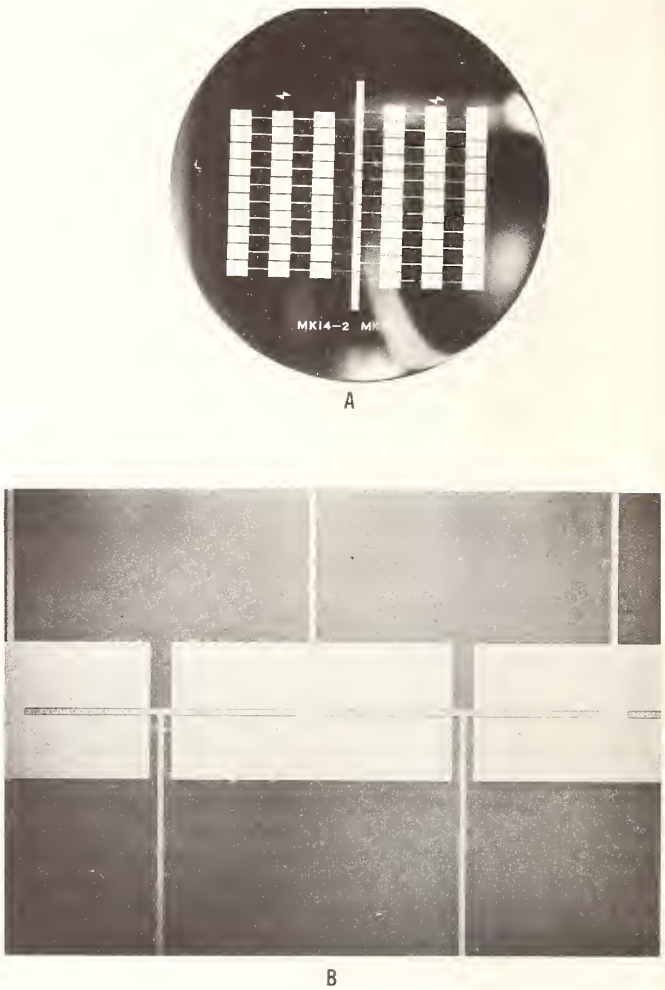


FIGURE 3. *The Josephson device pattern for 100 mV generation. (A) Photograph of the pattern on the wafer; (B) microscope photograph of the Josephson junctions.*

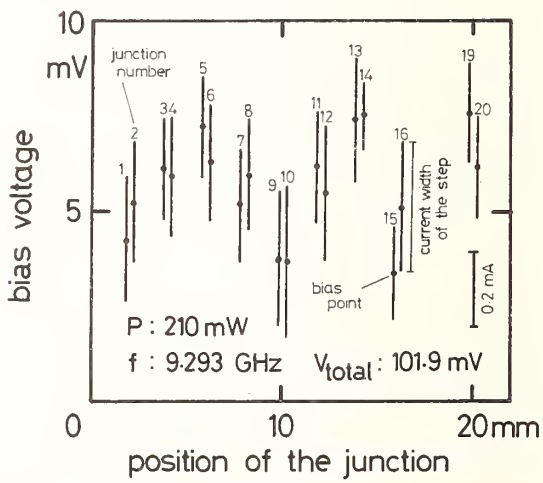


FIGURE 4. *The bias voltages and the current widths of the constant voltage steps used for voltage generation. The closed circle and solid bar show bias voltage and current width, respectively.*

The microwave frequency is read on the frequency counter along with the rubidium frequency standard. The drift of the present rubidium frequency standard is below 10^{-11}/month.

For determining the voltage across each resistor of the divider, four comparisons are usually made with the current direction through the divider and the junction + − − + in order to compensate for thermal emf's and the drift of the observed voltage difference. Therefore, one run of the experiment consists of 40 comparisons because we have to determine the voltages across ten resistors for the measurement of the standard cell voltage.

As the present constant current supply to the divider is not stable enough to determine the emf of the standard cell with an uncertainty at the 10^{-9} level, the preliminary experiment is performed assuming that the voltage of the standard cell is constant through the experimental time interval which is about ten hours, and the constant current supply is adjusted in each comparison so that the voltage across the divider is balanced to the working standard cell.

3. Results

Sources of uncertainty in the determination of the emf of the standard cell are given in Table 1. Random uncertainty for determining the voltage of one resistor in the divider is estimated from the scatter for seven runs. The random uncertainty is about 4.5×10^{-8}. This uncertainty reduces to 1.5×10^{-8} for the determination of the emf of the standard cell because ten equivalent measurements are contained in that determination. However, the constant voltage steps of some junctions have small but non-zero resistance on the steps which may be due to external noise. By taking into account the uncertainty due to this non-zero resistance of the constant voltage steps, the uncertainty of the present experiment would be at the 10^{-8} level, though each junction is biased near the center of the step.

TABLE 1. *Sources of uncertainty in the determination of the emf of a standard cell.*

Sources	Relative uncertainties (one standard deviation)
(1) Standard cell comparison	
Temperature fluctuation of voltage divider	0.3×10^{-8}
Leakage resistance	0.1
Resolution of standard cell balance	0.5
Calibration signal of standard cell balance	0.5
(2) Josephson junction comparison	
Resolution of frequency counter	0.01
Resolution of junction balance	0.3
Calibration signal of junction balance	0.3
(3) Drift of the junction bias current combined with inclination of the step	3.0
(4) Random uncertainty for determining the voltage of one resistor in the divider	1.5
Total (RSS)	3.5×10^{-8}

In Table 2 we give values of the average emf of three standard cells in the same air bath mentioned above, which were determined by the present preliminary experiments.

TABLE 2. *The emf of a standard cell determined in the present experiment.*[a]

Date (1981)	E_s (V)
April 30	1.01816963
May 2	1.01816965
May 7	1.01816962
May 14	1.01816955

[a]One standard deviation is 0.035 ppm.

In the course of determining the emf of the standard cell, the resistance ratios (R_i/R_1, i = 2 to 10) are measured automatically. Comparisons between resistance ratios determined by the present Josephson potentiometer $[(R_i - R_1)/R_1]_{J.P.}$ and by the current comparator type potentiometer $[(R_i - R_1)/R_1]_{C.P.}$ are given in Table 3. They agree within the uncertainty of 0.1 ppm.

TABLE 3. *Comparisons of the resistance ratio between two measurements, one by the Josephson potentiometer and one by a current comparator type potentiometer.*[a]

Number of resistor	$[(R_i - R_1)/R_1]_{J.P.}$ (ppm)	$[(R_i - R_1)/R_1]_{C.P.}$ (ppm)
1	—	—
2	5.56	5.54
3	11.85	11.87
4	7.71	7.79
5	5.76	5.71
6	−1.92	−1.84
7	3.71	3.67
8	−0.99	−0.99
9	2.54	2.48
10	3.34	3.32

[a]Standard deviations are below 1×10^{-7} for both measurements.

4. Conclusion

As a demonstration of the Josephson potentiometer, the emf's of standard cells were determined with an uncertainty at the 10^{-8} level by using a Josephson voltage of 100 mV. Further improvement of the uncertainty will be possible by using a stable constant current supply with an uncertainty at the 10^{-9} level instead of the present current supply used, and by removing the external noise which may be the cause of the non-zero resistance of the constant-voltage steps.

The authors would like to express their thanks to Toshiaki Sakuraba for measuring the temperature of the air bath.

References

[1] B. W. Petley, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, New York, 1976), p. 450.

[2] M. Koyanagi, T. Endo, and A. Nakamura, in *Future Trends in Superconductive Electronics*, Ed. by B. S. Deaver, Jr., C. M. Falco, J. H. Harris, and S. A. Wolf, AIP Conference Proceedings No. 44 (American Institute of Physics, New York, 1978), p. 187.

[3] R. Joly, E. C. Hannah, J. E. Opfer, and L. S. Cutler, IEEE J. Solid-State Circuits **SC-14**, 685 (1979).

[4] R. L. Kautz, Appl. Phys. Lett. **36**, 2289 (1980).

[5] R. L. Kautz and G. Costabile, IEEE Trans. Magn. **MAG-17**, 780 (1981).

[6] T. F. Finnegan and A. Denenstein, Rev. Sci. Instrum. **44**, 944 (1973).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A Transportable Josephson Voltage Standard

K. Lahdenperä, H. Seppä, and P. Wallin

Helsinki University of Technology and Technical Research Centre of Finland, Otakaari 5, 02150 Espoo 15, Finland

A transportable cryogenic Josephson effect emf standard has been developed in which all precise measurements are carried out in a helium bath (4.2 K). This standard uses a superconducting current comparator and 0.1 Hz measurement current for the calibration of a resistive potential divider. The cryogenic emf standard has a measurement uncertainty of 2×10^8 (one standard deviation).

Key words: ac Josephson effect; Josephson voltage standard; superconductivity.

1. Introduction

Standards of emf are nowadays maintained in a number of laboratories using the ac Josephson effect, which relates potential to frequency by the equation

$$V = \frac{h}{2e} f \ , \qquad (1)$$

in which h is Planck's constant and e the electron charge [e.g., 1, 2]. The development of Josephson voltage standards has lately split into two branches. On the one hand the goal is better repeatability, on the other hand reliability and easy use. The former goal leads often to complex solutions, and difficulties with reliability in operation, while to reach the latter requires compromises with regard to accuracy.

The components of Josephson voltage standards developed in various countries have been described in several publications. We describe here the main features and novel parts of our instrument. The aim of this work was to develop a transportable, small sized Josephson voltage standard.

2. Apparatus

A cryogenic Josephson effect emf standard has been developed in which precise measurements are carried out in a helium bath (4.2 K). The cryogenic system includes the Josephson junction (a Nb-Nb point contact or a Nb-oxide-Pb junction), a cryogenic voltage divider which is calibrated during measurements using a cryogenic current comparator, a SQUID as a null detector, and a superconducting switch. The diameter of the cryogenic unit is 48 mm.

The principle of the instrument is shown in Fig. 1 and a photograph in Fig. 2. The photograph does not show the microwave frequency source for the Josephson junction. In the circuit of Fig. 1 a standard cell is calibrated using a Josephson junction. The comparison is made with a potentiometer (resistors 200 Ω and 0.4 Ω). The null detector of the potentiometer is a SQUID. The measurement of voltage is made with the switch S in position 2, and calibration of the resistors in the potentiometer with the switch in position 1. The signal obtained from the SQUID controls the current source that balances the potentiometer. The galvanometer used in the measurement of the emf of the standard cell is balanced by adjusting the 10 GHz microwave frequency.

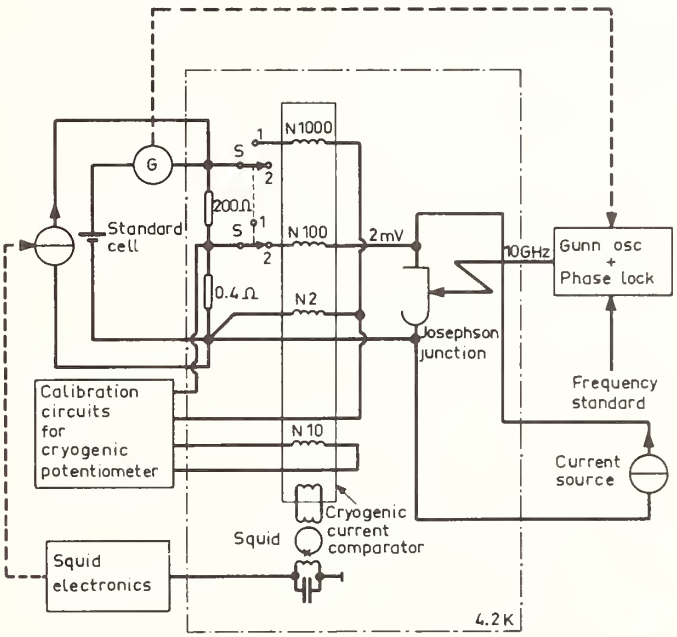


FIGURE 1. *Schematic diagram of the transportable cryogenic Josephson voltage standard.*

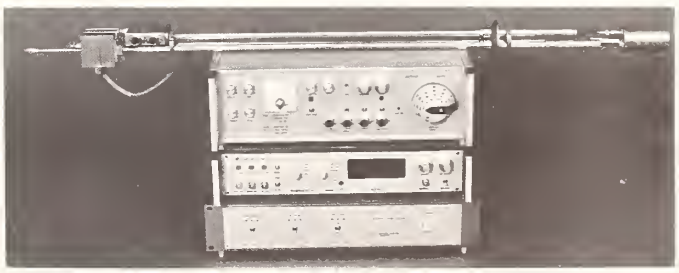


FIGURE 2. *Photograph of the voltage standard.*

New methods of coupling microwave radiation to Josephson junctions have been developed. The junctions are shown in Fig. 3 and the microwave frequency source in Fig. 5. In the case of a thin film tunnel junction the microstripline coupling technique was used and a mismatch of 11 dB was obtained between the signal source and the junction at x-band. The coupling was

made using a linear taper. Because of the high impedance difference between the signal source (50 Ω) and the resonant type tunnel junction (0.1 Ω), the linear taper was made using both a coplanar stripline and an ordinary stripline. In the case of a point contact the mismatch was 6 dB. The point contact was set on the end of the coaxial line and a 0.57 λ transformer was used to match the 50 Ω-line to the junction [impedance $(0.1 + 2j)\Omega$]. Voltage steps at the level of 2 mV were used and current steps were about 50 to 250 μA (point contact) and about 20 μA (tunnel junction). The SQUID uses a thin film Nb-Pb tunnel junction, of which the resolution is about $6 \times 10^{-5}\Phi_0\sqrt{\text{Hz}}$.

A voltage divider resistance ratio of 1:500 was used. A block diagram of the resistance ratio measurement system is illustrated in Fig. 4. The resistance bridge is based on a cryogenic current comparator [3]. A low frequency alternating current was used to avoid the disturbing influence of 1/f-noise from the SQUID. The form of the ac-current can be seen in Fig. 4. There is a transient in the output of the SQUID (see Fig. 4) because of the large inductances of the cryogenic current comparator. The effect of this transient is avoided by using a sample and hold circuit. The sample and hold circuit is in the "hold" mode and both the switches in the phase sensitive detector are open when the measurement current is

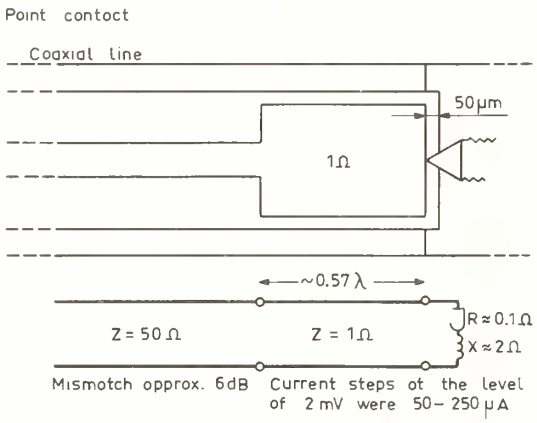


FIGURE 3(a). *Schematic of scheme used to couple microwave radiation to point contact junctions.*

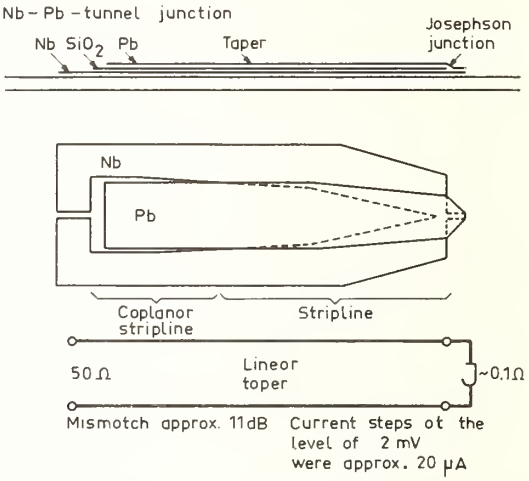


FIGURE 3(b). *Schematic of scheme used to couple microwave radiation to thin film tunnel junctions.*

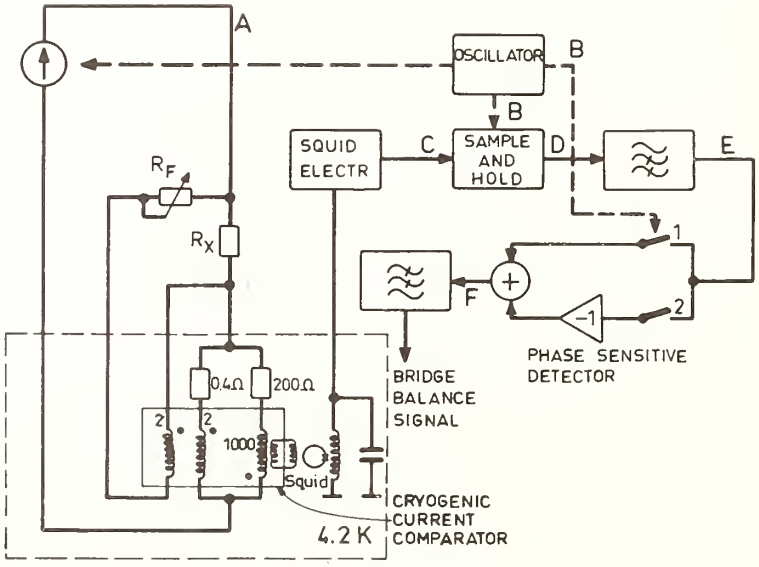


FIGURE 4(a). *Calibration circuit for the cryogenic voltage divider.*

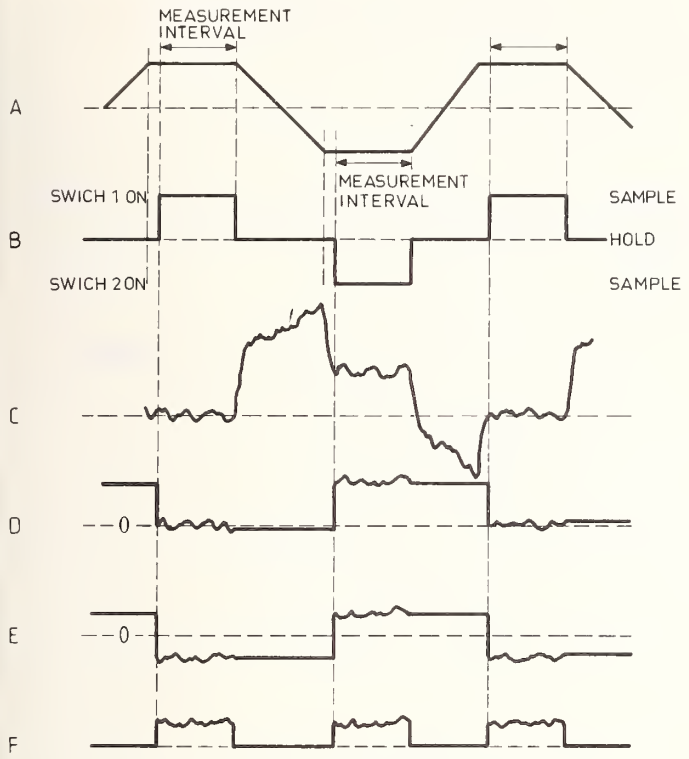


FIGURE 4(b). *Waveforms of signals.*

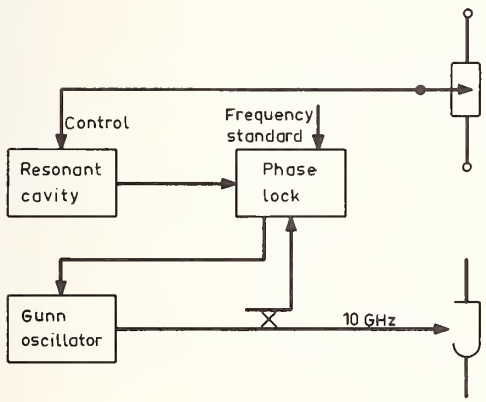


FIGURE 5. *Microwave frequency source. The signals from the Gunn oscillator and from the stable tunable resonant cavity are fed to the harmonic mixer, The intermediate frequency is phase-locked to the frequency standard. The frequency is monitored using a microwave counter.*

changing. Balance of the bridge is achieved by variation of the resistor R_F. The resistance ratio can be calculated by determining the values of R_F and R_X. More details can be found in Ref. [4].

The voltage divider was calibrated using different currents and the result of the ratio determination was obtained by extrapolating the result to zero current, thus minimizing the effect of self-heating. This measurement is laborious because of the requirement for many measurements. However, this part of the measurement is quite easy to automate.

3. Measuring Results

The uncertainty of the measurements made using this voltage standard is about 2×10^{-8} (standard deviation), due to the uncertainty concerning the extent of voltage divider self-heating during calibration.

A comparison between the Josephson standard and a standard cell group calibrated at the BIPM has been carried out. The accuracy of the comparison was limited by the instability of the standard cells. The measured difference was 2×10^{-7}.

The Josephson standard was later compared to the PTB (Federal Republic of Germany) voltage standard at Braunschweig. The mean values of the results obtained using the two standards differed by about 1×10^{-8} and the standard deviation was 5×10^{-8}. The comparison was made through an intermediate working group of standard cells. Their instability limited the measurement accuracy. The systematic uncertainty due to this limitation and due to an unreliable contact was estimated to be 1×10^{-7}. The measurement results are seen in Fig. 6.

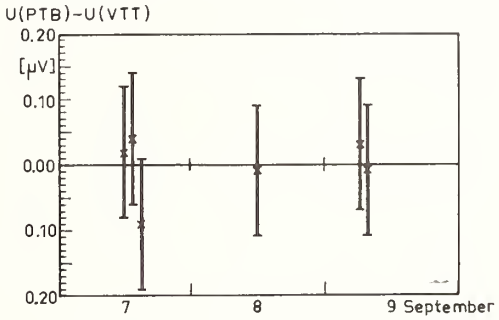


FIGURE 6. *A comparison between U(PTB) and U(VTT) in September 1980.*

4. Reliability

The Josephson thin film tunnel junction and the SQUID are very reliable. During several years of use no service has been needed. The Josephson point-contact has proved to be the most unreliable component in the instrument. It has to be tuned at intervals of about six months. Between the tunings the junction is reliable and requires no adjustments.

References

[1] B. F. Field, T. F. Finnegan, and J. Toots, Metrologia 9, 155 (1973).
[2] I. K. Harvey, Metrologia 12, 47 (1976).
[3] D. B. Sullivan and R. F. Dziuba, Rev. Sci. Instrum. 45, 517 (1974).
[4] H. Seppä, CPEM Digest 1980, IEEE Catalog No. OCH 1497-7 IM. p. 172 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Proton Gyromagnetic Ratio in H_2O—A Problem in Dimensional Metrology*

E. R. Williams, P. T. Olsen, and W. D. Phillips

Electrical Measurements and Standards Division, National Bureau of Standards, Washington, DC 20234

The last 10 years of work on the proton gyromagnetic ratio in H_2O, γ'_p, is reviewed. Results from both the low and high field methods are summarized along with short descriptions of several experiments. However, the emphasis is on the results from the low field method and the discrepancies among them. The values of the fine structure constant, α, obtained from the low field experiments are compared with several values of α obtained from various atomic physics experiments which to varying degrees require quantum electrodynamic theory (QED) for their interpretation. There is good agreement between the values of α obtained from two γ'_p determinations and these QED experiments as well as with the values of α obtained from quantized Hall resistance measurements. Two other γ'_p determinations disagree drastically: one differs by (5 ± 0.6) ppm and the other by (7 ± 0.5) ppm. We have no explanation for a disagreement this large.

Key words: fine structure constant; fundamental constants: proton gyromagnetic ratio; QED tests.

1. Introduction

The past decade has been very eventful for proton gyromagnetic ratio experiments. On the one hand, the best value of γ'_p(low) has improved by a full order of magnitude with a corresponding improvement in the resulting value of the fine-structure constant, α. This value for α is consistent with that obtained from the anomalous electron moment using quantum electrodynamic (QED) theory. On the other hand, the value of α obtained from two other γ'_p(low) experiments completely disagrees with the QED-consistent value. The high field measurements were not very consistent at the beginning of the decade, but significant improvements have been made in both the quoted accuracy and the consistency among different laboratories.

When we first started this review paper, we felt that much of the paper would be spent describing the successes (i.e., the tests of physics) which have come about over the past decade, in part due to the increased accuracy of measurements of γ'_p. Indeed, we will show data which are in excellent agreement and represent some of the most precise tests of QED, of the ac Josephson effect, and of a new effect, the quantized Hall resistance. But it has become clear that there also exists a large discrepancy among various low field γ'_p determinations.

Because there exist these large discrepancies among recent values and because the low field value has a dramatic impact on the value of the fine-structure constant, the low field γ'_p experiment requires the most discussion. There are four recent values of γ'_p(low); the spread in values is over 8 ppm, a discrepancy of over 10 standard deviations (see Table 1). At the same time, the value of γ'_p(low) claiming the highest accuracy is in remarkable agreement with QED predictions and with the latest quantized Hall resistance results.

TABLE 1. *Low field γ'_p measurements since 1970.*

Authors (year), and Laboratory	γ'_p (low) − 26751.00 × 10^{-11} s^{-1} T_{BI-76}^{-1} [a]	Uncertainty[b] (ppm)
Olsen, Driscoll (72) NBS[c]	0.384(54)	2.0
Olsen, Williams (75) NBS[d]	0.354(11)	0.4
Vigoureux, Dupuy (78) NPL[e]	0.178(13)	0.5
Williams, Olsen (79) NBS[f]	0.3625(57)	0.2
Chia, Shen (80) NIM[g]	0.391(21)	0.8
Forkert, Schlesok (80) ASMW[h]	0.55(13)	5.0
Tarbeyev (81) VNIIM[i]	0.228(16)	0.6

[a]All values have been expressed in terms of the BIPM 1976 unit of current.
[b]One-standard deviation (68% confidence level) estimates.
[c]Dimensions of solenoid measured mechanically; see Ref. 1.
[d]This value is by the same technique as in footnote f (i.e., not an independent value). Ref. [2].
[e]Refs. [3, 4]. [f]Ref. [5]. [g]Refs. [6,7]. [h]Refs. [8, 9]. [i]Ref. [10].

*Supported in part by the Department of Energy

The proton gyromagnetic ratio in H_2O, γ'_p, is defined to be the proton precession frequency ω'_p of a spherical sample of H_2O at 25 °C when placed in a magnetic field B, divided by the magnitude of the field if the sample were not present:

$$\gamma'_p = \omega'_p/B \ . \tag{1}$$

In both the low and high field methods, ω'_p is measured by conventional NMR techniques. In the low field method, B is calculated from the measured dimensions of a current carrying precision solenoid or a precision Helmholtz coil: $B = \mu_o I_{LAB} G$, where μ_o is the permeability of free space, I_{LAB} is the current measured in some laboratory unit, and G is the coil geometry factor which has units of inverse length. Equation (1) becomes

$$\gamma'_p(\text{low})_{LAB} = \omega'_p/(\mu_o I_{LAB} G) = K_A \gamma'_p \ , \tag{2}$$

where $K_A \equiv A_{LAB}/A$ is the ratio of the laboratory unit of current to the SI unit of current. As we shall see, the value of γ'_p(low) in laboratory units is particularly important to the fundamental constants field because I_{LAB} can be realized in terms of a voltage using the Josephson effect and a resistance which can be realized in SI units using the calculable capacitor. The measurement of γ'_p in SI units is limited to the precision with which K_A is presently known. In the high field method of determining γ'_p, one calculates B from the force equation $d\mathbf{F} = \boldsymbol{I}(d\boldsymbol{l} \times \mathbf{B})$ where $d\mathbf{F}$ is the elemental force experienced by a current element $\boldsymbol{I}\,d\boldsymbol{l}$ in the presence of a magnetic induction $\mathbf{B}$. In a uniform field, with a wire of length L perpendicular to the field, one can measure the force, F, perpendicular to both and thus

$$\gamma'_p(\text{high})_{LAB} = \omega'_p \, I_{LAB} \, L/F = \gamma'_p/K_A \ . \tag{3}$$

The distinguishing feature between the high and low field methods is not the field that is used, but that the two quantities are measured in different units. For example, to compare a high and low field measurement one must introduce another experimental quantity which involves a mass measurement (i.e., an absolute ampere, a Faraday, or an Avogadro's constant experiment). Alternatively, Eqs. (2) and (3) can be combined to obtain K_A or γ'_p in SI units from the ratio or product of γ'_p(low) and γ'_p(high):

$$\gamma'_p = \left[\gamma'_p(\text{high})_{LAB}\,\gamma'_p(\text{low})_{LAB}\right]^{1/2} , \tag{4}$$

$$K_A = \left[\gamma'_p(\text{low})_{LAB}/\gamma'_p(\text{high})_{LAB}\right]^{1/2} . \tag{5}$$

2. The High Field Results

The feature that distinguishes the high field method is that a force measurement is required in order to obtain the desired quantity. In all experiments to date, a rectangular coil is placed on one side of a balance with the upper end of the coil in zero magnetic field and the lower end in a strong uniform field provided by an electro- or permanent magnet. One must map out the strong field using an NMR probe and thus obtain the average value of γ'_p(high). The width of the coil must be measured especially accurately in the region of high field gradients. The force on the coil is measured by finding the mass (and knowing the acceleration of gravity) that just counters the force change when the current in the rectangular coil is reversed.

The three high field measurements carried out since the 1970 PMFC Conference are given in Table 2. All three techniques are similar so we will describe the most precise determination, that of Kibble and Hunt [11, 12]. These workers developed a method to eliminate a major source of systematic error. The problem is that horizontal forces and torques acting on the movable coil can cause "fictitious" vertical forces and thus an error. They eliminated these errors by introducing a flexure near the middle of the long arm that was attached to the rectangular coil. Thus this long arm and coil formed a double pendulum. They used an autocollimator to detect any bending of the double pendulum, thus detecting the unwanted torques and forces. They then adjusted their coil position and balance to minimize these effects. This source of error was probably the cause of many of the discrepancies in the earlier strong field measurements. However, the latest measurements are in reasonable agreement.

The Kibble and Hunt experiment had two other distinguishing features worthy of mention. First, their technique to shim the permanent magnet to produce a uniform field is simple and clever. They simply painted magnetic paint in the appropriate places to accomplish the final shimming. Second, the critical measurement of the width of the coils was made by two methods. First, they mechanically measured the wire position; then they used an inductive technique similar to the one we will describe later in this paper to locate the current in the wire. The two methods were in reasonable agreement.

At present, the major area of impact for the strong field experiments is in providing a value of the ampere, K_A. There exist significant discrepancies in the determination of the ampere, but this is discussed by B. Kibble in another review paper in these proceedings. Until reasonable agreement about the value of K_A can be reached, the high and low field results cannot be directly compared. Because the high field method has far fewer discrepancies at the present time, most of this paper will concentrate on the low field experiments.

TABLE 2. *High field γ'_p measurements since 1970.*

Authors (year) and Laboratory	γ'_p(high) − 26751.00 × 10^{-11} s^{-1} T_{BI-76}^{-1}	Uncertainty (ppm)
Kibble, Hunt (79) NPL[a]	0.689(27)	1.0
Chiao, Liu, Shen (80) NIM[b]	0.572(95)	3.5
Forkert, Schlesok (80) ASMW[c]	0.32(41)	16.0

[a]Refs. [11, 12]. [b]Refs. [6, 7]. [c]Refs. [8, 9].

3. The Low Field Experiments

The ac Josephson effect has had a major impact on the low field experiments. Before the laboratory unit of voltage could be maintained by the Josephson effect and related to the constant $2e/h$, the low field γ'_p experiment seemed destined to play the role of monitoring the ampere, and thus the volt, since the ohm can be accurately determined via the calculable capacitor. But at the same time that γ'_p(low) lost this job, the Josephson effect created a new role for it in the area of fundamental constants. The uncertainty in γ'_p(low) soon became the most important limitation in the determination of α without the use of QED theory. The pertinent equation is

$$\alpha^{-2} = [(c/4R_\infty)/(\Omega_{LAB}/\Omega)] \times (\mu'_p/\mu_B)(2e/h)_{LAB}/\gamma'_p(\text{low})_{LAB} . \quad (6)$$

Here c is the speed of light in vacuum; R_∞, the Rydberg constant for infinite mass; Ω_{LAB}/Ω, the ratio of the laboratory as-maintained ohm to the absolute or SI (International System) ohm; μ'_p/μ_B, the magnetic moment of the proton in units of the Bohr magneton (again, the prime indicates protons in a spherical sample of pure H_2O at 25 °C); $(2e/h)_{LAB}$, the ratio of twice the elementary charge to the Planck constant measured in terms of the LAB as-maintained volt using the ac Josephson effect; and $\gamma'_p(\text{low})_{LAB}$, the gyromagnetic ratio of the proton measured in terms of the LAB as-maintained ampere by the low field method.

With the motivation provided by Eq. (6), a number of laboratories set out to accurately measure γ'_p(low). Note that it is γ'_p(low) in as-maintained units that is of greatest interest, not the absolute value of γ'_p given in Eq. (3). The results reported by the various laboratories are summarized in Table 1 and can best be described as being in complete disagreement with each other, falling roughly in two sets. But there is no common parameter that seems to distinguish the two sets, other than the value of γ'_p(low).

The Chinese group of Chiao, Liu, and Shen from the National Institute of Metrology (NIM) has reported a value of γ'_p(low) [6] which is also reported in these proceedings by Prof. Wang [7]. In the NIM experiment, the dimensions of a precision single layer Helmholtz coil were measured by comparison with end standards. The end standards were in turn measured via an interferometer. The impressive part of the NIM work is the precision obtained in the NMR part of the experiment, 0.07 ppm, considering the low field, about 2×10^{-4} T, produced by the Helmholtz configuration. The final 0.8 ppm accuracy was primarily limited by the dimensional measurements.

The VNIIM (USSR) group has reported a new value of γ'_p(low) in these proceedings [10]. They use a four segment solenoid and mechanically locate the wires, using a laser interferometer to measure the dimensions. They use a two coil free induction decay method to measure the NMR in an approach similar to that used by the NPL.

P. Vigoureux and N. Dupuy at the NPL (UK) have spent many years obtaining their results [3, 4]. In measuring the pitch of their solenoid they use a unique technique to locate the wire position. Their solenoid is bifilarly wound and the pitch is measured by placing a voltage between the two windings. A metal sphere that is attached to one end of an interferometer is moved along a generator of the solenoid, shorting out the double helix as it goes. The pulse that results from this short triggers the recording of the position of the sphere by the laser interferometer. This information is recorded for different generators by rotating the solenoid. These data are then combined to describe the average position of each turn of the solenoid.

Because of the large discrepancy among low field γ'_p experiments and because our technique for measuring the pitch of the solenoid differs significantly, we will describe the NBS measurement in greater detail than those of the other laboratories. The reader will note that this obviously represents a bias on the part of the authors, but we hope it will be informative.

While all other experiments locate the surface of the wire at many positions and then calculate where the current will flow from these measurements, the NBS has developed a method to locate the average position of an entire current loop using the magnetic field produced by that current [5]. Coils A and A' in Fig. 1 form a linear differential transformer which locates the axial position

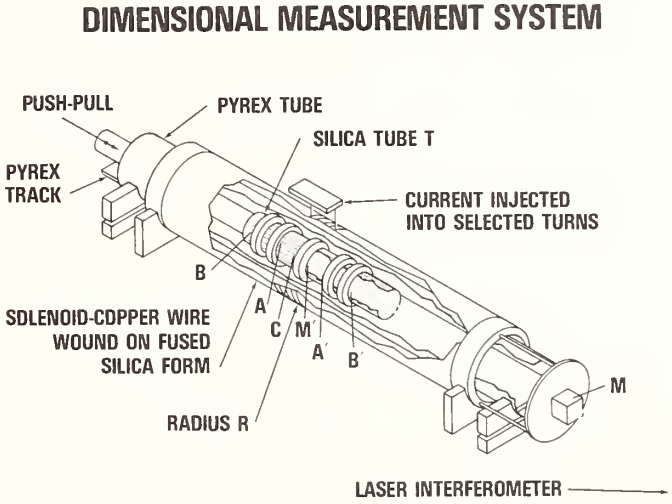


FIGURE 1. *Solenoid dimensional measurement system used to determine the axial position and radial variations of the wires. The five coils A, A', B, B', and C are attached to a silica tube T and can be pushed or pulled along the axis of the solenoid. Coils A and A' locate the axial position of the injected current and coils B, B', and C form a diameter-to-voltage transducer. Mirrors M and M' are part of a laser interferometer. The Pyrex tube is evacuated.*

of the alternating current injected into selected turns of the solenoid. Coils A and A' are connected so that their output voltages cancel when centered on the activated turns of the solenoid, and a servo system locks the coil assembly to the null point with a precision of better than 0.05 μm. A mirror (corner cube M') located in the center of the coil assembly is part of a laser interferometer system, with the reference mirror (M) connected to the end of the solenoid. Thus, with the aid of a computer-automated system, the relative position of successive turns can be recorded. We chose to activate 10 turns at one time, and then move the current injector to a successive 10 turns until information about all 1000 turns of the solenoid was obtained. Coils B, B', and C form a radius-to-voltage transducer which measures the variations in the radius of the injected current. The voltage induced in coil C is inversely proportional to the radius of the activated turns of the solenoid. To detect small changes we first "buck out" most of the voltage in coil C by using the two coils B and B'. At the same time that we are bucking out the voltage in coil C, we are also increasing

the sensitivity of this three-coil system to changes in the radius R of the solenoid, because the voltage in coils B and B' increases when the radius R increases. This three-coil radius-to-voltage transducer is then calibrated by having a few turns of wire on both ends of the solenoid that are 25 μm larger at one end and 25 μm smaller at the other end. With this system, the axial position and the radius variations of the turns of the solenoid were measured to an accuracy of 0.05 μm. No turn deviated from that expected of a perfect solenoid by more than 2 μm.

Figure 2 shows the pitch variation (axial position) and the variations in the radius as a function of the wire position. These data are one measurement set, where the central wires from +230 mm to −230 mm were measured four times and the outer wires measured twice. The total height of the error bars represents twice the rms scatter of the measurements recorded during this measurement set, and it takes about 4 hours to complete such a set. The field at the center is then calculated using the appropriate weighting functions which are also shown in Fig. 2. Our final value of γ'_p was calculated from 13 sets of pitch variation and 22 radius variation measurements, all very similar to Fig. 2. The statistical scatter was small compared to other systematic uncertainties, such as the measurement of the mean diameter. From these diameter and radial variations, we calculated the gradient in the magnetic field along the axis of the solenoid to be 0.2 ppm/cm. Using the NMR data we measured the residual gradient after compensation and found good agreement with that predicted from the dimensional measurements.

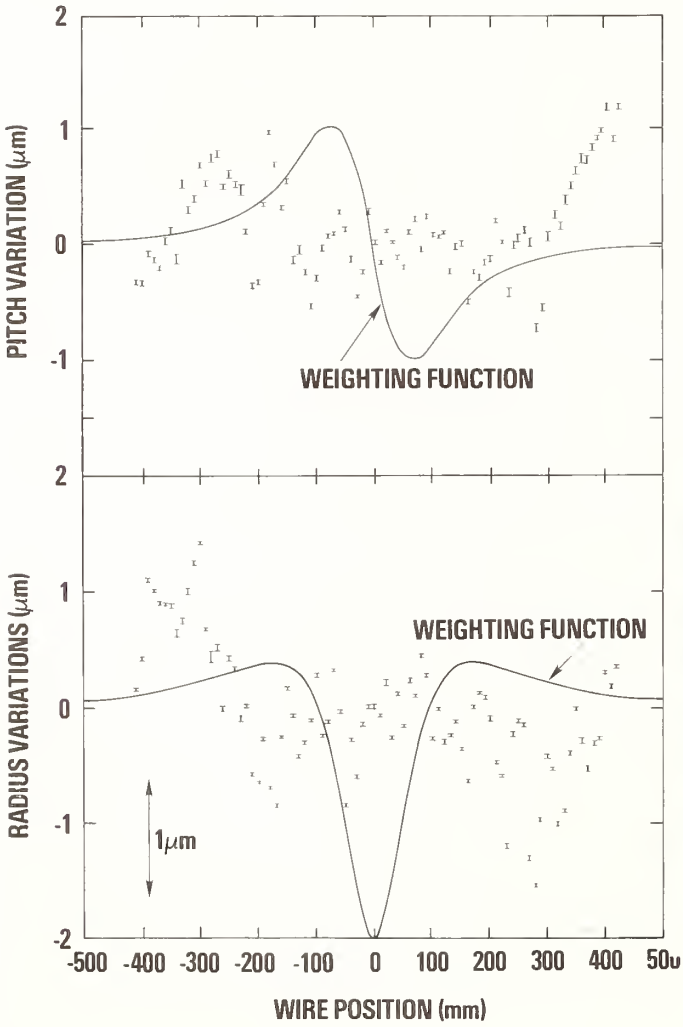


FIGURE 2. *One measurement set of the axial position (pitch variation) and radial variation as a function of distance from the center wire. For the upper plot a pitch of 1.000087 mm/turn has been subtracted from the measured positions. The total height of the error bars is twice the rms value of the measurements taken on each wire during this measurement set. The structure is reproduced from day to day within the error bars. The curves are weighting functions which must be multiplied by the measured variations to calculate the field at the center of the solenoid.*

One source of systematic error is very important and its elimination is of interest. If the injected current is a sinewave, then an error we call an end effect is introduced. Because this current induces a voltage in the other windings of the solenoid and because there is some capacitance across these turns, some current flows in these turns thereby producing an error signal in the detector coils [13]. This effect varies inversely with the square of the frequency. We eliminate it by employing the special wave form for the injected current shown in the upper trace of Fig. 3. The output of the detector coil is just the derivative of the injected current. The second trace of Fig. 3 results when the probe is +2 mm off null, and the bottom trace when it is −2 mm off null. In the next to last trace we are locked to the null. Using a lock-in amplifier with a flat frequency response, we can look at the integral of the output voltage from time A to time B minus the integral from B to C. This integral is proportional to the mutual inductance or total flux change in the output coil. The transient signals, however, integrate to zero. The effectiveness of this method is demonstrated by either showing that the null position is independent of frequency or that it is independent of the phase of the lock-in. We find an error of less than 0.04 μm for a phase shift of up to 90°.

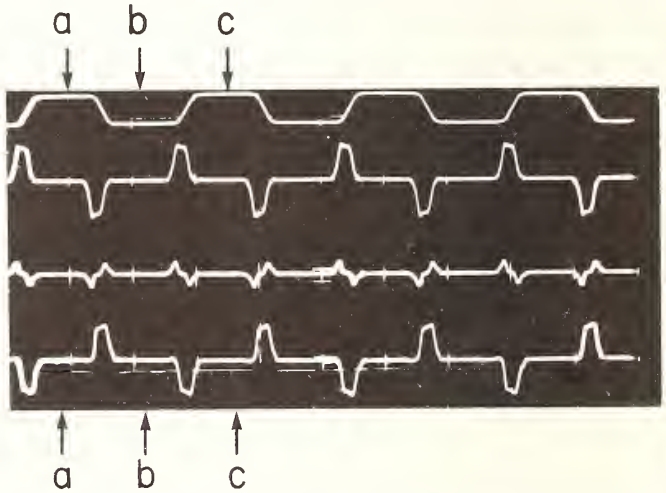


FIGURE 3. *Injection current and resulting signal. Upper trace: current injected into a portion of the solenoid. Second trace: voltage observed across the pickup coils when the probe is displaced +2 mm from the center of the activated loop. Third trace: probe is centered. Bottom trace: probe is displaced −2 mm. Letters indicate various instants of time along the oscilloscope trace.*

Our solenoid is 1 m long and 0.280 m in diameter. The diameter was compared to a quartz end standard that was measured with a laser interferometer. The 0.8 μm (2.9 ppm) uncertainty in the diameter measurement produced a 0.25 ppm error in γ'_p. With the high accuracy of

the radius variation and pitch measurements, the diameter measurement became the limiting factor in the experiment. The basic problem in determining the diameter is that even if we make a perfect measurement of the position of the surface of the copper wire, it is still difficult to know where the current flows in the wire. Because of our concern over a possible systematic error in measuring the diameter, a second approach was developed. The technique is to find a coil geometry which produces a magnetic field that is nearly independent of the average diameter. While a longer solenoid is one solution, we were able to find a suitable configuration for the existing solenoid in which current could be taken out of selected turns. Using a computer, we searched for a configuration where the sum of the diameter-variation weighting function is small and the second, fourth, and sixth-order gradients are compensated. The final configuration (see Fig. 4) uses five 1 A current sources, one of which is used to pass current through the entire solenoid in the normal fashion. The other four are connected to selected turns on the opposite side from the main return lead so that the net current in these selected turns is zero. This five-current system produces a 0.8 mT field (uniform to 0.1 ppm over the H_2O sample), instead of the 1.2 mT field of the conventional solenoid, but this 0.8 mT field is 8.6 times less sensitive to the average diameter of the solenoid.

The 1 A current placed in the solenoid is measured by comparing the voltage across a 1 Ω precision resistor in series with the solenoid, with a saturated standard cell calibrated in terms of the United States legal volt, which in turn is maintained via the Josephson effect. A cable is used to transfer this current to the NBS nonmagnetic facility, thus providing us with the closest possible tie to the Josephson voltage standard [14]. The observed scatter in the NMR measurements (0.02 ppm) reflects the precision we obtained in making the voltage comparison. At the nonmagnetic facility a rubidium magnetometer system is used to reduce Earth's magnetic field fluctuations.

The two values of γ'_p which we obtained from the two different current configurations agreed within their respective uncertainties and were combined together in a least-squares fit to give our final value and 0.21 ppm uncertainty. The final accuracy of this determination was primarily limited by the diameter measurement either directly in the conventional configuration or indirectly by the uncertainty for the corrections due to the compensation currents which reduced the diameter sensitivity in the five-current configuration.

4. Comparison of Results

The major purpose of a new proton gyromagnetic ratio measurement is to help test the validity of the conven-

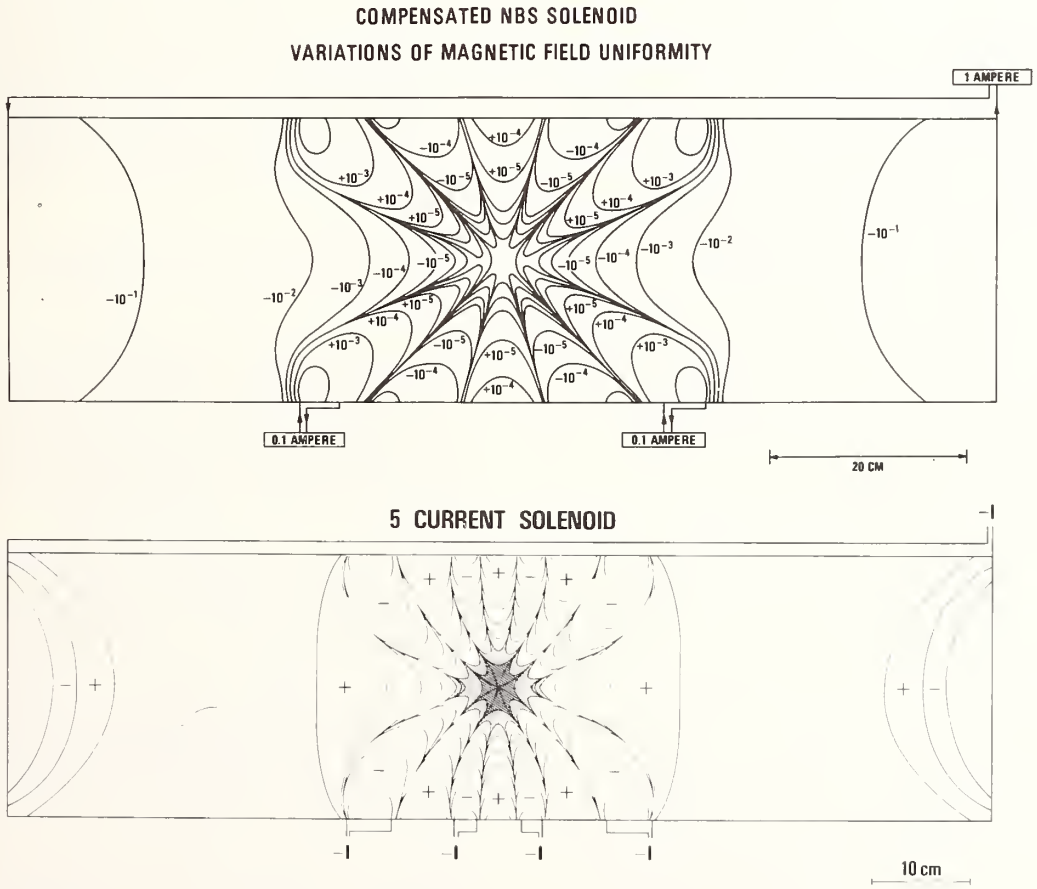


FIGURE 4. *The two current configurations. The top solenoid is compensated with two 0.1 A current sources for maximum uniform field. The field is uniform to better than 0.1 ppm over a center 7.5 cm sphere. The bottom configuration has four one ampere compensation currents so that the resulting field is independent of the mean diameter of the solenoid. The field is uniform to 1.0 ppm over the light shaded region and 0.1 ppm over the dark region. The 4 cm H_2O sample just fits within the 0.1 ppm region. All compensation currents are injected on the opposite side of the solenoid. Leakage between current sources must be carefully controlled.*

tional QED theory. The important role of γ'_p in testing QED comes from the need for a value of the fine-structure constant, α, which is essentially independent of QED (see Eq. 6). The importance of Eq. (6) was first demonstrated in the late 1960's when the value of α derived from it was used to resolve the discrepancy between the theoretical and experimental values for the ground-state hyperfine splitting in hydrogen. Subsequently, the Josephson-effect measurement of $2e/h$ advanced quickly, and by the early 1970's, the primary limitation for obtaining a more accurate value of α from Eq. (6) was the uncertainty associated with the experimental determination of the proton gyromagnetic ratio. At that time, the most accurate measurement of γ'_p(low) had an uncertainty of two ppm and was obtained in our laboratory [1].

Figure 5 shows a comparison of some values of α which are derived from various atomic physics experiments plus the appropriate theories. (These values were kindly provided by B. N. Taylor and were calculated in preparation for a new least-squares adjustment of the constants.) Each of these QED quantities was discussed by various authors at this conference. The curve plotted is simply a visual guide which points out discrepancies and is called an ideogram [15]. We have included two theoretical corrections for the electron anomalous magnetic moment reported at PMFC-II, one by Levine and Roskies and the other by Kinoshita and Lindquist. The uncertainty in α_{a_e} is now limited by the uncertainty in the eighth order correction terms that Kinoshita and Lindquist reported. Larger, faster computers are all that are needed to reduce this uncertainty significantly. The difference $(\alpha_{a_e} - \alpha_{\gamma'_{p\mathrm{NBS}}})/\alpha = (0.22 \pm 0.13)$ ppm represents a remarkable success for atomic and solid state physics and QED theory.

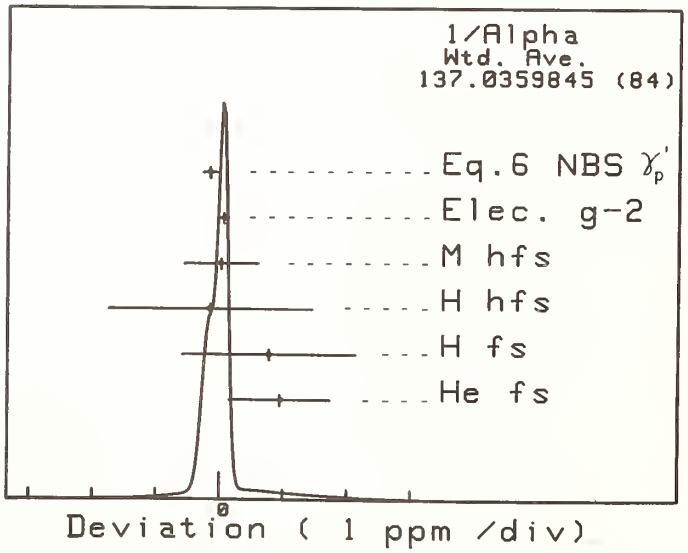


FIGURE 5. *Values of the fine-structure constant derived from QED theory and various experimental determinations. The horizontal axis is the deviation from the mean of the measurements. (The values have been provided by B. N. Taylor.) At the top we have the value obtained without QED from the NBS determination of γ'_p using Eq. (6). Then in descending order we have values from the anomalous moment of the electron; the muonium hyperfine splitting; the hydrogen hyperfine splitting; the hydrogen fine structure; and the helium fine structure measurements. The curve is provided as an aid to the reader. It is a composite of Gaussian lines with the value and width of each Gaussian corresponding to each measurement. The area of each Gaussian is inversely proportional to the uncertainty.*

Two recently reported values of γ'_p(low) sharply disagree with this pleasing situation [4, 10]. This disagreement is apparent in Fig. 6 which shows values of α obtained from the various values of γ'_p(low) and for comparison, the value α_{a_e}. Also included in the figure are two values of α which were derived from a new effect described by von Klitzing and others in these proceedings, called the quantized Hall resistance, R_H. The pertinent equation used to obtain α from R_H is:

$$\alpha^{-1}_{\bar{R}_H} = [2/(\mu_0\, c)\,]\,(\Omega_{\mathrm{LAB}}/\Omega)\;\;(R_H)_{\mathrm{LAB}}\;\;. \qquad (7)$$

Equations (6) and (7) can be combined to eliminate the quantity $\Omega_{\mathrm{LAB}}/\Omega$:

$$\alpha^{-1}_{(\bar{R}_H\cdot\gamma'_p)} = \{\,[\,1/(2\mu_0\, R_\infty)\,]\,(\mu'_p/\mu_B)\,[\,(R_H)\times (2e/h)/\gamma'_p\,(\mathrm{low})\,]_{\mathrm{LAB}}\}^{1/3}\;\;. \qquad (8)$$

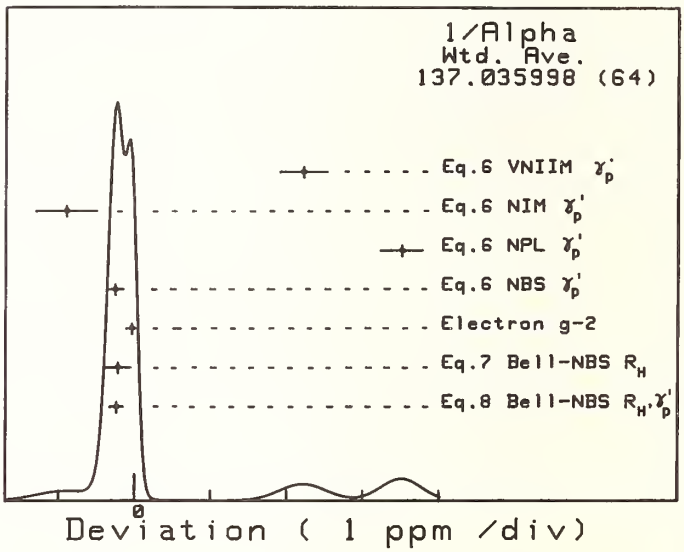


FIGURE 6. *Values of the fine-structure constant from various non QED sources. The four uppermost values use Eq. (6) and the various values of γ'_p from Table 1. For comparison, the value from the anomalous moment of the electron of Fig. 5 is included. The last two values are based on a new Bell Laboratories and NBS value of the Hall resistance and Eqs. (7) and (8), respectively. The three NBS values are not independent.*

Equation (8) is obviously not independent of Eq. (7), but represents an alternate path to obtain α from the Hall resistance, and the cube root dependence helps reduce the uncertainty in $\alpha^{-1}_{(\bar{R}_H\cdot\gamma'_p)}$. Of course, all three results could be combined together to obtain a least-squares value for α, but for now we choose to show the results from Eqs. (6), (7), and (8) separately. (The reader should keep in mind that the three values are not independent.) The preliminary value and uncertainty shown for R_H were obtained in a collaborative effort between Bell Laboratories and NBS [16]. This value is in agreement with less precise values reported by the German and Japanese groups [17, 18]. The two discrepant values of α are in disagreement with the value obtained from a variety of different physical phenomena; however, these two measurements should not be ignored because if they are correct, the implications for physics would indeed be dramatic. For example, if α derived from the NPL value of γ'_p(low) is correct, then α_{a_e} is in error by 35 standard deviations.

5. Experiments and Ideas in Progress

The need for new γ'_p determinations is obvious. For the high field experiments the major motivation is to resolve the inconsistencies in the absolute ampere measurements. We know of one experiment in progress at ETL in Japan, in which the field of 2.8 T is produced by a superconducting magnet. Because of the extra strong field, they can use thin film techniques to deposit the conductor on fused silica to produce the movable rectangular coil and still have a force of sufficient magnitude. They are presently limited by magnetic field inhomogeneities.

In the low field experiments, two new ideas are discussed in these proceedings [19, 20]: one by K. Weyand of the PTB (FRG); the other by G. Greene of Yale University. The PTB experiment is under construction and nearly ready for experimental testing. Greene's paper discusses a new idea that employs the Ramsey separated oscillating field technique.

Continued progress in low field experiments at VNIIM (USSR), ASMW (DDR), and NIM (China) is expected. At ETL (Japan), a new non-magnetic facility is being constructed and a new precision solenoid is ready to be measured. At NBS we plan to extend our accuracy to a few parts in 10^8. A new 0.3 m diameter, 2.4 m-long precision solenoid has been constructed and will be used to reduce the major sources of uncertainty in our last experiment. We also plan to incorporate a new technique using Ampere's law for an independent method of measuring the solenoid's dimensions [21].

The quantized Hall effect has added a new dimension to the electrical standards and fundamental constants field. Figure 7 is a geometric representation of the significance of this new technique. The tetrahedron is a way of summarizing our options for the present and future measurements of γ'_p and other electrical quantities. Currently, we maintain the electrical units of resistance, voltage, and current in laboratory units by two of these four electrical quantities, namely the calculable capacitor and $2e/h$. From that triangular base we now have three paths to obtain α. The $-R_H$ path (meaning "without" Hall resistance) is formed from the well known equation, Eq. (6). "Without" ($2e/h$, γ'_p) is the new Hall equation, Eq. (7); and we can also obtain α "without" the calculable capacitor experiment $K_\Omega \equiv \Omega_{LAB}/\Omega$, using Eq. (8). The degree to which one obtains closure on this tetrahedron contains a great deal of physics (i.e., validity of the Josephson effect and the quantized Hall effect) and verifies the quality of our electrical units.

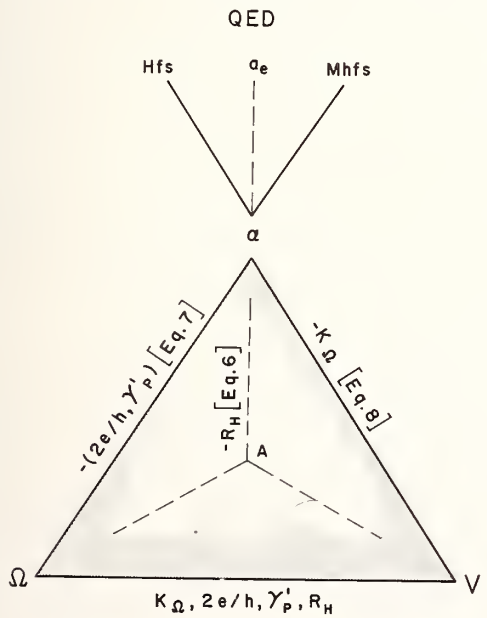


FIGURE 7. *This pyramid is simply a visual representation of the over-determination of the electrical quantities by the four measured constants. The base of the pyramid is formed by the three maintained electrical quantities: ohm, volt, and ampere. The three lines that rise to form the apex represent the three paths for obtaining α from Eqs. (6), (7), and (8). The degree of closure in this pyramid contains a great deal of physics and is indicative of the quality with which one maintains the electrical standards. The quantity ($-K_\Omega$) means "without" the absolute ohm experiment, K_Ω.*

Beyond the internal self-consistency of this tetrahedron, the values obtained for α from various QED experiments provide additional physics and constraints on the validity of the tetrahedron itself. Presently, the NBS group is showing closure at the 0.2 ppm level in α, and the possibility of extending this to a few parts in 10^8 in the near future seems realistic.

γ'_p(low), $2e/h$, R_H, and the absolute ohm experiments are now joined, through their relation to α, in a common effort to achieve the highest possible accuracy. The realization of this goal by all these experiments will provide not only an exacting test of QED, but also a crucial insight into the validity of our maintenance of electrical standards.

References

[1] P. T. Olsen and R. L. Driscoll, in *Atomic Masses and Fundamental Constants 4*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Publ. Corp., New York, 1972), p. 471.

[2] P. T. Olsen and E. R. Williams, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Publ. Corp., New York, 1976), p. 538.

[3] P. Vigoureux and N. Dupuy, NPL Report DES44, (1978).

[4] P. Vigoureux and N. Dupuy, NPL Report DES59, (1980).

[5] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. 42, 1575 (1979).

[6] W. Chiao, R. Liu, and P. Shen, IEEE Trans. Instrum. Meas. **IM-29**, 238 (1980).

[7] Z. Wang, these proceedings.

[8] W. Schlesok, IEEE Trans. Instrum. Meas. **IM-29**, 248 (1980).

[9] J. Forkert and W. Schlesok, Metrlogischen Abhandlungen des ASMW, Heft 3 (1981).

[10] Y. V. Tarbeyev, these proceedings.

[11] B. P. Kibble, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Publ. Corp., New York, 1976), p. 545.

[12] B. P. Kibble and G. J. Hunt, Metrologia 15, 5 (1979).

[13] E. R. Williams and P. T. Olsen, IEEE Trans. Instrum. Meas. **IM-21**, 376 (1972).

[14] E. R. Williams, P. T. Olsen, and B. F. Field, IEEE Trans. Instrum. Meas. **IM-23**, 299 (1974).

[15] A. H. Rosenfeld and P. R. Stevens, in *Proceedings of the Fifth Biennial International CODATA Conference*, Ed. by B. Dreyfus (Pergamon Press, New York, 1976), p. 19.

[16] D. C. Tsui, A. C. Gossard, B. F. Field, M. E. Cage, and R. F. Dziuba, Phys. Rev. Lett. 48, 3 (1982). Information in Fig. 6 has been revised to reflect the data which became available after the conference.

[17] E. Braun, P. Gutmann, G. Hein, F. Melchert, P. Warnecke, S. Q. Xue, and K. v. Klitzing, these proceedings.

[18] C. Yamanouchi *et al.*, these proceedings.

[19] K. Weyand, these proceedings.

[20] G. L. Greene, these proceedings; and Metrologia 17, 83 (1981).

[21] E. R. Williams and P. T. Olsen, IEEE Trans. Instrum. Meas. **IM-27**, 467 (1978).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Development of Precision Measurement and Fundamental Constants in China

Wang Zhu-xi

Peking University, Beijing, P.R.C.

In China there is some experimental work on precision measurement and fundamental constants, such as cesium and hydrogen frequency standards, iodine and methane stabilized lasers, the determination of the proton gyromagnetic ratio γ_p' and the gravitational acceleration g, the realization of the SI electrical resistance unit, and so on.

Key words: cesium and hydrogen frequency standards; electrical resistance; gravitational acceleration; iodine and methane stabilized lasers; proton gyromagnetic ratio.

1. Introduction

Great advances have been made in precision measurement and fundamental constants since PMFC-I in 1970. Although this work has only been done for a short time in China, there has been some good results. I am very pleased to give a brief review of these results to the specialists participating in this conference.

In China, the research work which has had good results includes the cesium beam frequency standard, the hydrogen maser as a frequency standard, iodine and methane stabilized He-Ne lasers as wavelength standards, the determination of the proton gyromagnetic ratio γ_p', the absolute measurement of the gravitational acceleration g, and the realization of the SI electrical resistance unit.

2. The Cesium Beam Frequency Standard

Two primary cesium beam frequency standards, CsII and CsIII, have been constructed at the National Institute of Metrology (NIM), Beijing, China.

The cesium beam apparatus is shown in Fig. 1. The beam tube is made of stainless steel; the total length is about 5.3 m and the diameter 290 mm. The active region of the cavity is about 3.7 m and the line width is about 42-46 Hz with a 150 °C oven temperature.

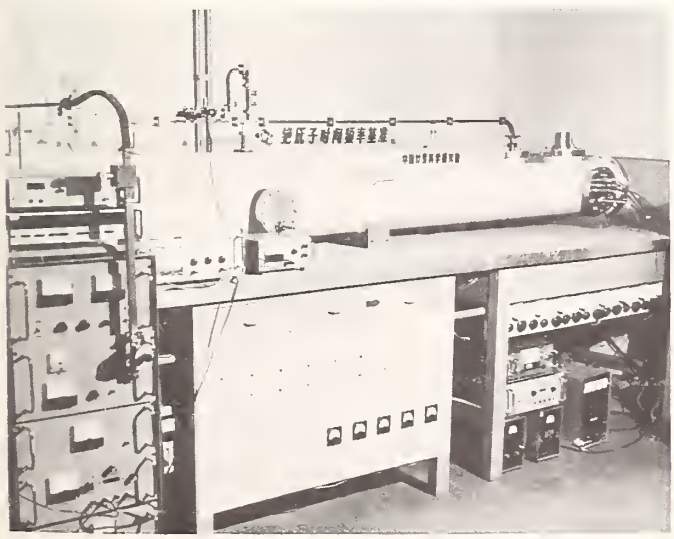


FIGURE 1. *NIM cesium beam frequency standard.*

In December 1980, CsIII was operated for 50 h to carry out a frequency comparison, using HP5061A cesium oscillators with a high performance tube as stable reference standards. The resulting frequency stability data is given in Table 1. There are many factors which affect frequency uncertainty, such as pulling of near lines, cavity phase difference, spectral purity, second order Doppler effect, and so on. The RSS of the total uncertainty is 4.5×10^{-13}.

TABLE 1. *Frequency stability data of CsIII.*

1s	10s	100s	1000s
1.8×10^{-11}	4.2×10^{-12}	1.3×10^{-12}	4.0×10^{-13}
1h	2h	3h	5h
2.8×10^{-13}	2.5×10^{-13}	2.3×10^{-13}	2.4×10^{-13}

3. Hydrogen Maser

Hydrogen maser experiments have been done since 1973 at the Shanghai Bureau of Metrology (SBM). A photograph of the SBM hydrogen maser is shown in Fig. 2.

The main accuracy limitation of a hydrogen maser is the uncertainty in the determination of the wall shift. When the hydrogen atoms collide with the coating of the storage bulb, an additive phase shift is introduced in the hydrogen atom wavefunction, so that the maser frequency is offset from the free oscillator frequency of the hydrogen atom. The frequency offset value is given by an empirical formula as follows:

$$\Delta f_w = \frac{k(t_1)}{D}[1 + \alpha(t_1)(t - t_1)],$$

where k is a wall shift coefficient, D the bulb diameter, α the temperature coefficient in $°C^{-1}$, and t_1 and t the

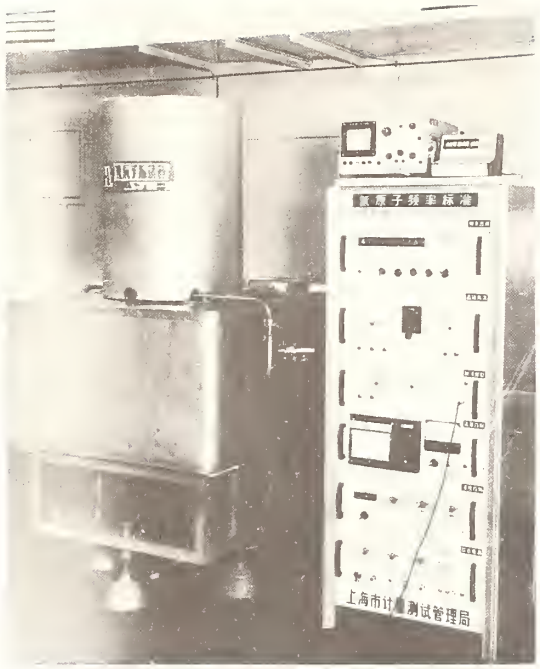


FIGURE 2. *SBM hydrogen maser.*

reference temperature and temperature of the bulb, respectively. The parameters k and α are obtained from our experiments on the hydrogen maser wall shift. In these experiments, we used five different size quartz storage bulbs in turn, which were all coated in the same way with a locally made F4 coating. The results of the experiments are [1]

$$k(40\ ^\circ\mathrm{C}) = (-293 \pm 17)\ \mathrm{mHz} \cdot \mathrm{cm}$$

and

$$\alpha(40\ ^\circ\mathrm{C}) = (-17 \pm 2) \times 10^{-3}\ ^\circ\mathrm{C}^{-1}\ .$$

The frequency stabilities for different sampling times are 4×10^{-13}/1s, 6×10^{-14}/10s, 9×10^{-15}/100s and 2.5×10^{-14}/day, respectively.

The value of the unperturbed hyperfine frequency of the ground state of the hydrogen atom f_0, which we obtained with respect to TAI, is

$$f_0 = 1420\ 405\ 751.768 \pm 0.002\ \mathrm{Hz}\ .$$

This value is in close agreement with the average value of f_0 published by other authors after 1970. The total uncertainty is 2.5 mHz.

4. Iodine and Methane Stabilized He-Ne Lasers

Iodine and methane stabilized He-Ne lasers have been constructed at NIM. The third harmonic lock servo control circuit was designed by NIM and Peking University. The He-Ne laser tubes and absorption cells are all made in China. An iodine stabilized laser is shown in Fig. 3.

The frequency stability of the iodine and methane stabilized lasers is about $2 \times 10^{-11}/\sqrt{\tau}$ and $3 \times 10^{-12}/\sqrt{\tau}$, respectively. We have carried out an international intercomparison of these two stabilized NIM lasers with similar BIPM lasers at BIPM in April 1980. The frequency stability of the NIM and BIPM iodine stabilized lasers NIM2 and BIPM2 relative to each other was 6×10^{-13} from 900 s to 2700 s. The difference in frequency between NIM2 and BIPM2 was 13.8 kHz and between

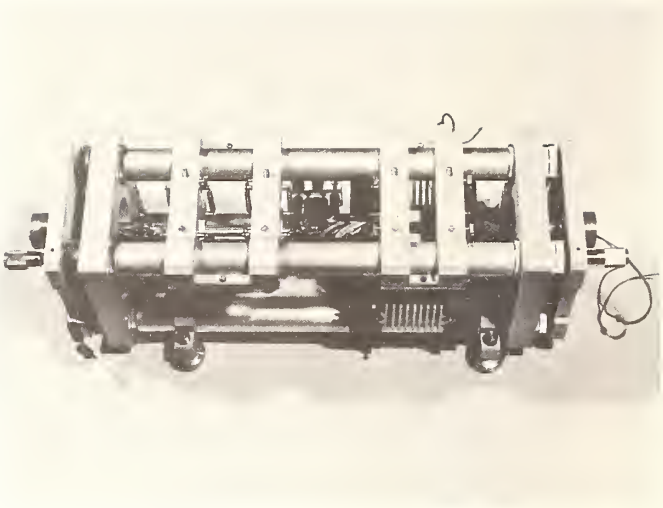


FIGURE 3. *NIM iodine stabilized He-Ne laser.*

$NIMCH_41$ and $BICH_46$, 0.56 kHz. These differences correspond to relative differences of 2.9×10^{-11} and 6.3×10^{-12}, respectively [2].

The comparison of iodine stabilized lasers was done 13 times from April 2 to 17, 1980. The standard deviation was 4.5 kHz, corresponding to 0.95×10^{-11}. Various frequency shifts of NIM2 were also measured at the BIPM. The pressure shift and modulation shift were found to be $-(9 \sim 11)$ kHz/Pa and -6.6 kHz/$\mathrm{MHz_{pp}}$, respectively, and the power shift -0.05 kHz/μW, which is very small. This is good for frequency reproducibility. The success of the intercomparison of NIM and BIPM iodine stabilized lasers is significant because the NIM2 He-Ne laser tube and iodine cell are different from those of BIPM2 as well as from those of lasers built by other laboratories.

5. The Absolute Measurement of the Gravitational Acceleration

A three point, freely-falling body system for determining the gravitational acceleration has been built at NIM. The experimental work began in 1970. This stationary laser gravimeter has been used for the absolute measurement of gravity since 1975. The standard deviation of g measured with the apparatus is 100 μGal. Since the great earthquake in Tangshan on July 28, 1976, we have constructed a new, precise transportable gravimeter for earthquake forecast investigations and other scientific research. The transportable gravimeter is shown in Fig. 4.

The object with the corner-cube falls in a high vacuum chamber. The dropping distance is determined by a Michelson interferometer. The He-Ne laser light source is stabilized by locking to the Lamb-dip. The time base is obtained from a highly stabilized rubidium frequency standard which is compared with the NIM cesium frequency standard.

Some absolute gravity stations have been established in China from 1979 to 1981, for example, at NIM, Xiangshan in Beijing, and some other stations in Yunnan Province in southwest China.

We have also determined the gravitational acceleration at Sèvres A3, BIPM. The value found was 980 925 914 $\pm$ 16 μGal. The difference between this value and that of A. Sakuma is only 6 μGal.

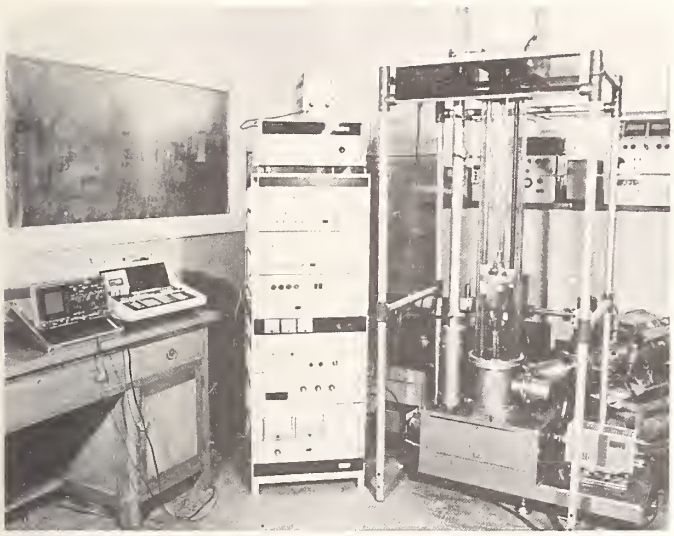


FIGURE 4. *NIM transportable gravimeter.*

6. The Determination of the Gyromagnetic Ratio of the Proton γ'_p

The gyromagnetic ratio of the proton γ'_p, was determined at NIM in 1977 by both the low and high field methods using NMR [3]. The average of the experimental results from 44 groups of γ'_p(high) measurements obtained in 1977 was

$$\gamma'_p(\text{high}) = 2.675\ 163\ 7(95) \times 10^8\ A_{\text{NIM-77}} \cdot s \cdot kg^{-1} ,$$

with a total RSS uncertainty of 3.5 ppm.

The apparatus for the low field determination of γ'_p is shown in Fig. 5. The work was done at the NIM nonmagnetic laboratory in a suburb of Beijing. In this experiment, the earth's residual magnetic field was compensated automatically and reduced to less than 50×10^{-10} T; its variable component could be kept to less than 5×10^{-10} T. The average value of γ'_p(low) obtained was

$$\gamma'_p(\text{low}) = 2.675\ 132\ 6(22) \times 10^8\ s^{-1} \cdot T_{\text{NIM}}^{-1} ,$$

with a total RSS uncertainty of 0.8 ppm.

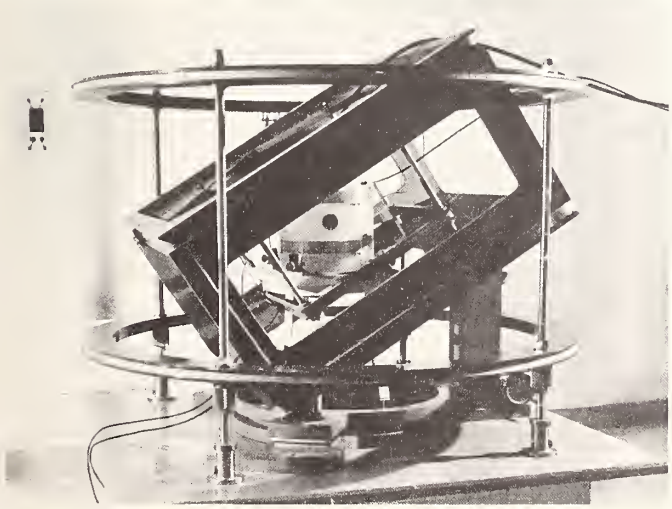


FIGURE 5. *NIM apparatus for measuring the gyromagnetic ratio of the proton in a low magnetic field.*

From the results for γ'_p(high) and γ'_p(low), the value of K, which is called the ampere conversion factor, and the value of γ'_p(abs), may be obtained:

$$K = 1 - (5.81 \pm 1.80)\ \text{ppm} ,$$

and

$$\gamma'_p(\text{abs}) = 2.675\ 148\ 1(48) \times 10^8\ s^{-1} \cdot T^{-1} .$$

This value of K implies that the ampere maintained at NIM is smaller by 5.81 ppm than the absolute or SI ampere.

7. The Absolute Determination of the Ohm

An absolute determination of the ohm by a cross-capacitor was carried out at NIM in 1978. The uncertainty of the cross-capacitor and the 1 Ω absolute determination are ± 0.35 ppm and ± 0.48 ppm, respectively. The cross-capacitor is shown in Fig. 6 and is different from those of other laboratories. The null electrode is a fused quartz tube with its surface covered with a composite metal film of chromium-aluminum-chromium. Two turns of fine lines are made by photo-etching of the film to provide an insulating air gap with the distance between the two line centers determining the capacitance value. (The nominal value is 0.5 pF.) The advantage of this construction is that it keeps the capacitance constant, so that it is a convenient standard for operation and measurement.

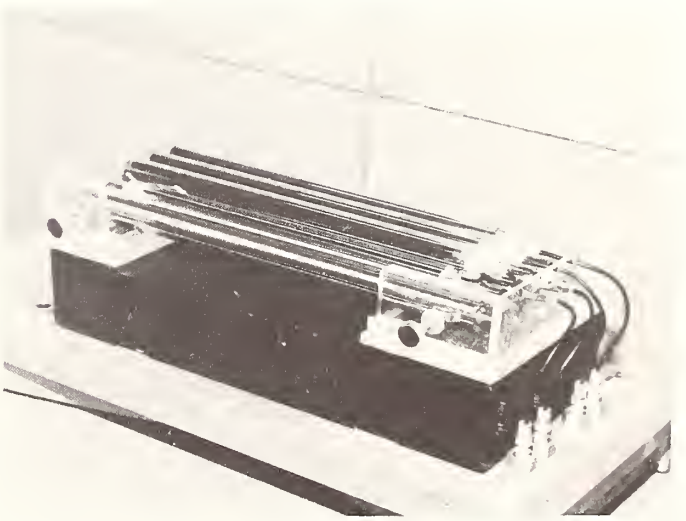


FIGURE 6. *NIM cross-capacitor.*

The ohm standard has been measured via the cross-capacitor periodically since 1978 with the result

$$\Omega_{\text{NIM}} = \Omega_{\text{ABS}}(1 - 0.36 \times 10^{-6}) \pm 0.48\ \mu\Omega . \quad (1)$$

The international resistance intercomparison between NIM and BIPM [4] allows Eq. (1) to be expressed as

$$\Omega_{\text{BI-78.9}} = \Omega_{\text{ABS}}(1 - 0.91 \times 10^{-6}) \pm 0.48\ \mu\Omega . \quad (2)$$

Equation (2) is in good agreement with the results of other countries. In addition, the NIM cross-capacitor was compared to those of other countries by NIM's participation in a cycle of the international intercomparison of 10 pF fused quartz capacitors organized by BIPM in August 1980. The value obtained at NIM was only 0.2 ppm larger than that given by BIPM.

A cross-capacitor with a movable guard rod is currently under investigation at NIM. A new method for compensating the end effect has been proposed and a further improvement in the precision of the NIM cross-capacitor can be expected.

References

[1] Y. M. Cheng *et al.*, IEEE Trans. Instrum. Meas. **IM-29**, 316 (1980).

[2] BIPM Report, 1980 (unpublished).

[3] W. Chiao, R. Liu, and P. Shen, IEEE Trans. Instrum. Meas. **IM-29**, 238 (1980).

[4] Cross Capacitor Group, NIM, Acta Metrological Sinica (in Chinese) 1, 16 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds.,
Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The γ_p-Experiment at PTB

Kurt Weyand

Physikalisch-Technische Bundesanstalt, D 3300 Braunschweig, Bundesallee 100, F.R.G.

This paper will give a brief summary of the mechanical apparatus and the electronic set-up which have been constructed to determine the gyromagnetic ratio of the proton by observing free precession in a low magnetic field. A novel scheme is used to determine the coil constant of a multilayer field coil system consisting of four winding packages by measuring a quantity proportional to the magnetic flux density at a number of points on the axis of the set of coils carrying a stabilized current.

Key words: coil constant; fundamental constants; nuclear magnetic resonance; precision measurement; proton gyromagnetic ratio; SQUID magnetometry.

1. Introduction

In order to determine the gyromagnetic coefficient of the proton (γ_p) by the low-field method, the precession frequency of protons is observed in a magnetic field of known flux density, produced by an electric current in a coil of known dimensions. That means the experiment has to be performed in a magnetically neutral surrounding, to avoid errors in determining the actual magnetic field strength during the measurement of the precession frequency. Therefore a nonmagnetic building has been erected about 300 m away from other laboratories. Before starting the construction in 1978 the available terrain was magnetically surveyed. A place was found where the spatial gradient in the earth's magnetic field shows a maximum value of about 0.25 nT/m [1]. Alternating magnetic fields at that place were found to be of the order of 4 nT peak-to-peak at 50 Hz and 0.5 nT at 150 Hz respectively.

In order to compensate the earth's magnetic field, a large tridimensional Braunbeck-coil system was constructed consisting of four octagonal coils in each direction [2, 3]. By careful adjustment of the currents through these coils the main field in a spherical volume, 80 cm in diameter, at the center of the system has been reduced to values of not more than 1 nT. Variations of the earth's magnetic field are observed by means of a triaxial fluxgate magnetometer giving a sensitivity of 100 mV/nT up to 0.3 Hz. The sensor is placed halfway between the nonmagnetic building and an observation station. (The observation station, which contains all the electronic equipment, is about 80 m from the nonmagnetic building.) The variations are compensated by means of additional currents through the coils, resulting in an instability of the residual field of 0.1 nT in the day and 0.01 nT at night, as measured by means of a SQUID-magnetometer.

There are two main problems in doing the γ_p-experiment: to obtain a relaxation time as close as possible to the natural relaxation time for protons in water and to determine the coil constant, i.e., the value of flux density produced by a geometrical arrangement of turns carrying a current I.

2. Detection of the Precession Signal

If a NMR experiment using water as the source of protons and done in the free precession mode does not give the natural relaxation time, there may be a gradient in the magnetic field over the probe volume. With a field coil arrangement, which should provide an adequate field homogeneity ($\Delta B = 3$ nT/cm), a gradient may nevertheless be caused by extraneous materials near the probe, such as those used for the probe supports or for coil forms. Impurities in the water may cause additional relaxation. Whatever the source of the relaxation, the determination of the precision frequency becomes more uncertain.

Relaxation due to impurities in the sample itself may be identified by performing a spin echo experiment [4]. In order to reduce extraneous material near the probe as much as possible the free precession experiment is done using a single coil with about 400 turns for polarizing and energizing the nuclear moments as well as for detecting the precession signal. The probe itself is a spherical sample 4 cm in diameter, filled in an artificial N_2-atmosphere with triple distilled water by means of a special distilling apparatus [5].

Figure 1 shows the block diagram of the signal detecting circuit, Fig. 2 indicates the time sequence of the measuring procedure in the upper part and shows the observed signal trace in the lower part. On the starting pulse, the control unit initiates the polarizing current source and afterwards the burst oscillator for the rf-90°-pulse. To avoid spikes the polarizing current is not switched abruptly but is slightly smoothed. Some milliseconds after disabling the burst oscillator the signal amplifying path is enabled and the precession signal is fed through a preamplifier acting as a line driver to a high gain tuned amplifier. Its output voltage is mixed with the frequency f_r of a synthesizer by means of a phase sensitive detector PSD. That frequency is set some hertz lower than the expected signal frequency f_0, thus the PSD is acting as a narrow band filter. Its output signal, consisting of a frequency $f_0 - f_r$, is fed through a low-frequency bandpass filter to a transient recorder for observing the signal and to a computing counter for measuring the difference frequency.

As may be gathered from Fig. 2, the relaxation time T_2 amounts to 2.2 s for the observed precession signal, which shows a signal to noise ratio better than 50 for a 8-Hz-bandwidth of the PSD.

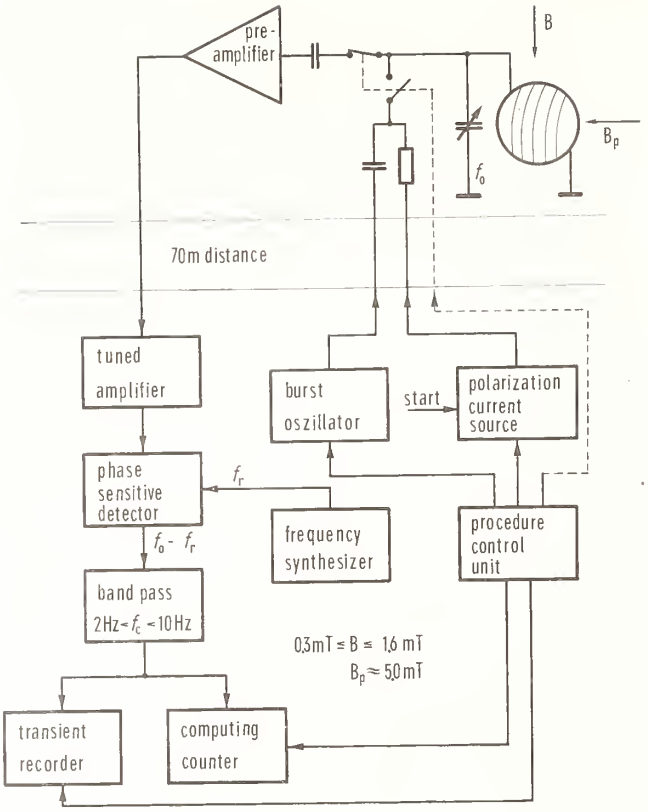
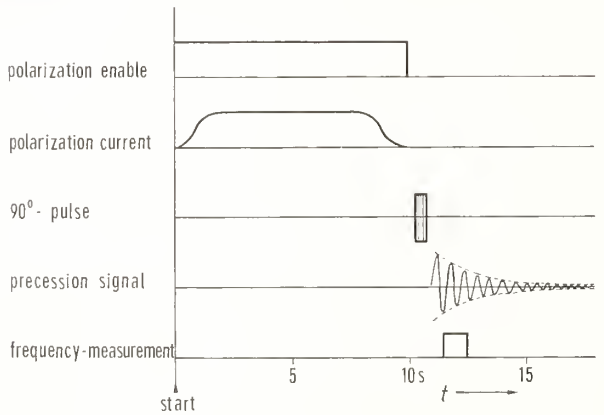


FIGURE 1. *Block diagram of the precession signal detection circuit.*



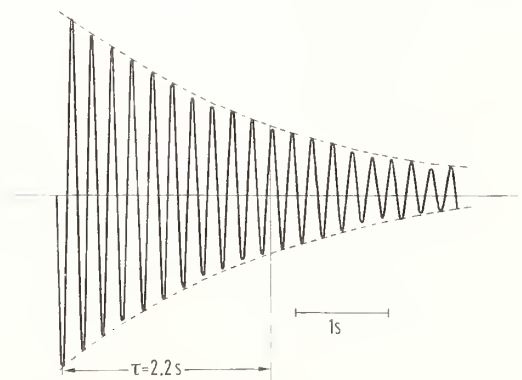


FIGURE 2. *Time sequence of signal detecting procedure and observed precession signal.*

3. Determination of the Coil Constant—Theoretical Background

Up to now the dimensions of singlelayer coils have been measured by mechanical and optical methods [6, 7], sometimes in combination with electromagnetic indicator systems [8], and from the results of those measurements the flux density at the center of the field coil has been evaluated.

Novel schemes have been investigated and one of them will be used, in order to find the coil constant of a system of coils, i.e., as it shall be understood here, the value of flux density at the center of the system as a function of the current through the coils [9, 10]. By means of a SQUID-magnetometer a quantity proportional to the magnetic flux density will be measured in arbitrary units at a number of points along the axis of the coil system when a current is maintained in it. If the flux profile along the axis of any coil system is calculated in terms of its as yet unknown dimensions, the comparison of the magnetometer readings with the calculated profile should in principle yield a set of effective dimensions, from which the coil constant may be evaluated again. Thus in general the coil constant of any arrangement of turns may be determined in this way leading to a more or less complicated set of unknown values. But in practice measurements by means of two flux gate sensors connected in a differential bridge and thus rendering a resolution of 1 nT have shown, that it will be nearly impossible to record the flux profile of a long solenoid [11]. This is due to the scatter of the angle between magnetic axis and sensor axis while one of the sensor heads is displaced. For that reason a quite different coil arrangement has been investigated, which may itself be shifted precisely whereas the magnetometer probe is fixed.

Assuming a single circular turn with radius a as shown in Fig. 3—then the field along the x-axis is given by the Biot-Savart law according to

$$B_x = \frac{\mu I}{2a}[1 + (x/a)^2]^{-3/2}. \tag{1}$$

Thus the only unknown geometric dimension a may be evaluated by measuring the field B_0 in the center and B_x at any other point on the axis of that turn at the distance x and by forming the field ratio $V_x = B_x/B_0$ as given by

$$a = x[1/V_x)^{2/3} - 1]^{-1/2} \tag{2}$$

where all B values may be measured in arbitrary units.

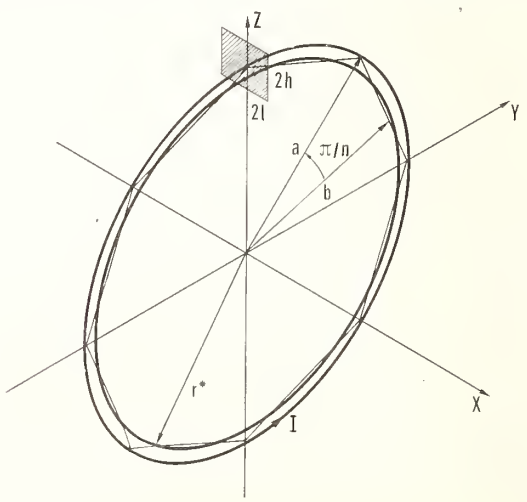


FIGURE 3. *The different kinds of current paths.*

If the current path differs from a circular path, it can be assumed that there exists an effective circular path and using the method above an equivalent radius r^* can be obtained. Supposing a polygonal current path, where n is the number of sides and a is now the radius of the circumscribed circle, the field on the axis follows the equation

$$B_x^p = \frac{\mu I}{2}\,\frac{n}{\pi}\,\frac{a^2\cos\pi/n\,\sin\pi/n}{(x^2 + a^2\cos^2\pi/n)(x^2 + a^2)^{1/2}}\,. \qquad (3)$$

By using the ratio of the field values at the positions $x = 0$ and $x = a$, produced by the polygon, the radius r^* of the effective circular current loop amounts to

$$r^* = a\left[\left[\frac{\sqrt{2}(1 + \cos^2\pi/n)}{\cos^2\pi/n}\right]^{2/3} - 1\right]^{-1/2} \qquad (4)$$

producing a field at the center point $B_0^* = \mu I/2r^*$. The fractional difference from the real field is drawn as a function of n, as line (1) in Fig. 4, according to

$$\frac{B_0^* - B_0^p}{B_0^p} = \left[\left[\frac{\sqrt{2}(1 + \cos^2\pi/n)}{\cos^2\pi/n}\right]^{2/3} - 1\right]^{1/2}\frac{\pi}{n}\,\frac{\cos\pi/n}{\sin\pi/n} - 1. \qquad (5)$$

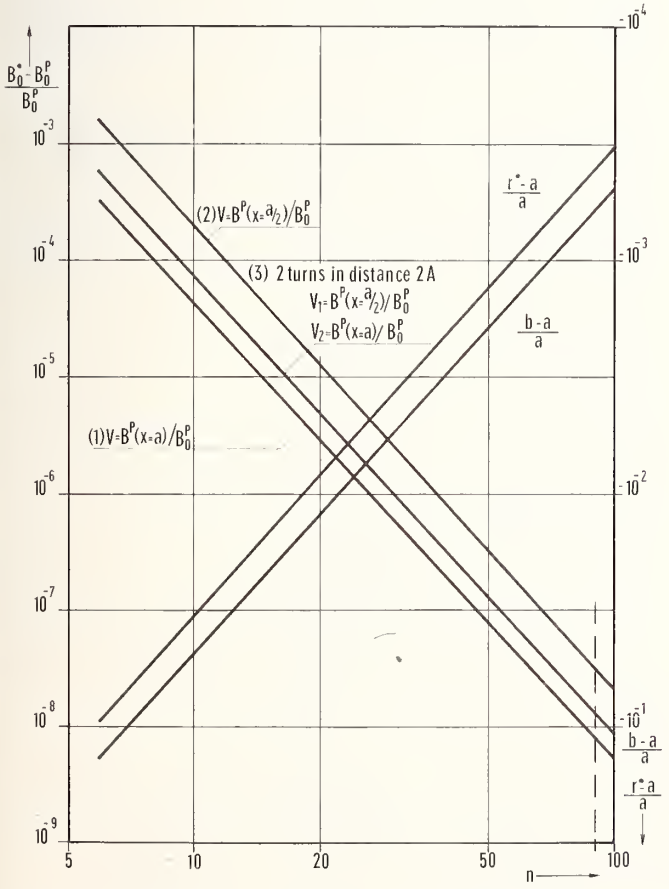


FIGURE 4. *Deviations of the center field, when polygonal current paths are described by effective circular loops, and the effective radius of such a loop, both as a function of the number of polygon sides.*

If the field value at $x = a/2$, instead of that at $x = a$, is taken to perform the field ratio the difference in the center field becomes about six times higher, and is given by line (2) of Fig. 4 according to

$$\frac{B_0^* - B_0^p}{B_0^p} = 2\left[\left[\frac{\sqrt{5}(1 + 4\cos^2\pi/n)}{8\cos^2\pi/n}\right]^{2/3} - 1\right]^{1/2}\frac{\pi}{n}\,\frac{\cos\pi/n}{\sin\pi/n} - 1. \qquad (6)$$

In order to simplify the measuring procedure for determining the field ratio function it is desirable to make use of two current loops connected in series and axially displaced by $\pm A$; then by measuring additionally the flux profile when both the loops are counter-connected the center point $x = 0$ may be found more precisely. For two loops the field function must be changed containing now two position dependent terms $(x + A)/a$ and $(x - A)/a$. Thus a system of two nonlinear equations has to be solved using two field ratios at different positions to evaluate numerically the effective geometric values r^* and A^*. Assuming again both loops as polygonal and using the field values at $x = a/2$ and $x = a$ related to that at $x = 0$ in the case $a/A = 4$, the error in evaluating the center field is given by line (3) in Fig. 4. For more clarity, the geometric proportions from which the presented field differences result are also shown in that figure. Looking to the righthand ordinate one may find the differences of the effective radius r^* and that of the inscribed circle $b = a\cos\pi/n$ from that of the circumscribed circle a, both related to a.

In order to achieve a center field strength which is high enough to do the γ_p-experiment, winding packages must be used instead of two singular loops. This is indicated in Fig. 3. Then a current NI flows through the cross section $4lh$, where N is the number of turns, and produces a field on the axis according to the much more sophisticated equation

$$B_x = \frac{\mu NI}{8lh}\left\{(x + l)\ln\left[\frac{a + h}{a - h}\left[\frac{1 + \left[1 + \left[\frac{x + l}{a + h}\right]^2\right]^{1/2}}{1 + \left[1 + \left[\frac{x + l}{a - h}\right]^2\right]^{1/2}}\right]\right] + (x - l)\ln\left[\frac{a - h}{a + h}\left[\frac{1 + \left[1 + \left[\frac{x - l}{a - h}\right]^2\right]^{1/2}}{1 + \left[1 + \left[\frac{x - l}{a + h}\right]^2\right]^{1/2}}\right]\right]\right\}. \qquad (7)$$

For two packages displaced by $\pm A$, as before, x must be replaced by $x + A$ and $x - A$ leading again to two equivalent position dependent terms. A set of about 360 equations of this type, each representing a field ratio at a different position x, is used to evaluate a set of at least five effective geometrical values, which are able to describe the field function in the observed axis interval $x \leqslant \pm a$.

4. The Coil System

A one-pair-winding coil system does not yield sufficient homogeneity over the sample volume, so the new coil system consists of two pairs of nominally identical coils

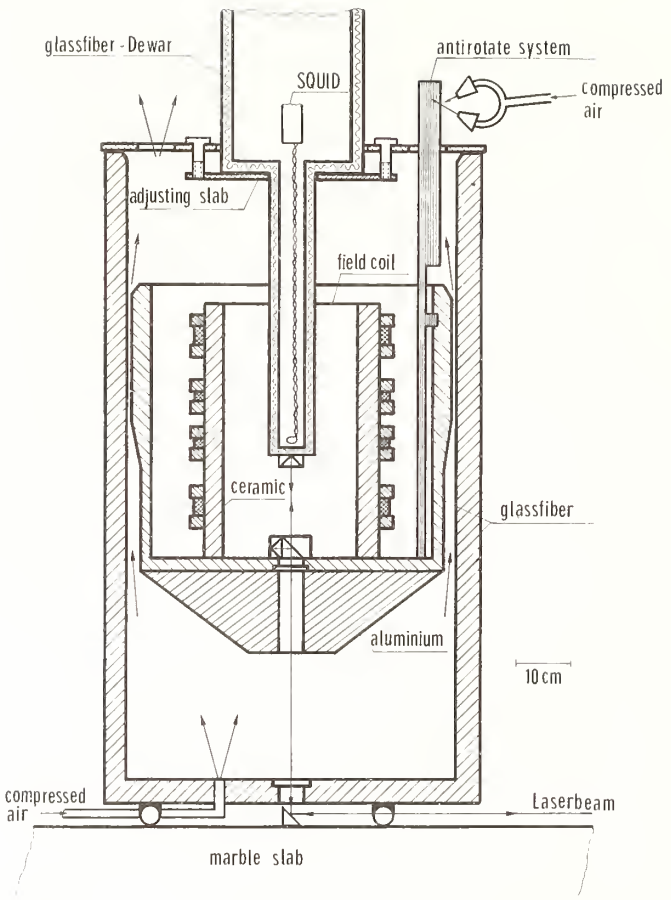


FIGURE 5. *Pneumatic field coil displacing apparatus.*

as shown by Fig. 5 and in a more detailed manner by Fig. 7. All the coils are of the same nominal radius (160.85 mm) and are wound between flanges on a cylinder of fused silica. The flanges, which are made of glass fiber, are glued carefully on the cylinder to ensure coaxiality. The number of turns (1295 and 2916), the length of the windings (13.6 mm and 31.4 mm) and the mean distances of the coils (79.5 mm and 307.8 mm) are so chosen, that the coefficients up to the seventh order of the magnetic scalar potential expressed in zonal harmonics vanish. Because of the symmetrical arrangement of the coils the even terms vanish in any case, thus the inhomogeneity over a sphere of the NMR-sample size at the center should be less than 10^{-7} theoretically, while the coil constant amounts to 17.5 mT/A.

It has turned out to be difficult to wind the multilayer coils in a regular fashion by centering one turn on another by means of a fitting nylon thread wound between adjacent turns—with increasing number of layers the winding began more and more to "float." Therefore the coils have been wound in the usual way but the crossing lines from one layer to the next have been staggered by a distance corresponding to the circumference of the cylinder divided by the number of layers. This technique has yielded stable coils showing variations of outer diameter at different generators not larger than 0.2 mm.

5. Determination of the Coil Constant—Hardware

In order to measure the flux profiles due to the coils a pneumatic coil carrying apparatus has been constructed as shown in Fig. 5. The shifting device is made of glass fiber and consists of a piston 600 mm in diameter and 550 mm in length inside of a cylindrical tube of 1200 mm length. By means of a conical aluminum block mounted under the bottom of the piston the center of gravity is lowered. The small air-gap of about 0.1 mm between piston and tube enables a shift procedure having an angular scatter of not more than 0.001°. A two-nozzle-system prevents the piston from rotating around its axis.

Thus the coils carrying a current equal to that which is to be used for the measurement of the precession frequency are moved in the axial direction past a SQUID-probe. The magnetometer is a somewhat modified commercial type using a one turn, 6 mm diameter flux transformer coil and thus giving a resolution of 0.003 nT. It is fixed inside of a special type of superinsulated Dewar, which is made of glass fiber, and which rests on a triaxially adjustable slab, allowing the probe to be positioned on the magnetic axis of the field coils. During the motion the flux linking the probe is recorded, as well as the readings of a laser interferometer, which gives the relative displacement of coil form and probe.

Three different measurements are necessary to determine the coil constant of the whole system. First the center of the inner pair is found by recording the flux profile when both the windings are counter connected, as mentioned before. Then the flux profile of these windings connected in series will be recorded and from that record the set of dimensional values will be found just representing the coil constant of the inner pair. Finally the total coil constant of the four packages is found by energizing the outer pair instead of the inner one and by comparing the flux produced at the center point with that due to the inner pair. Figure 6 gives the block diagram

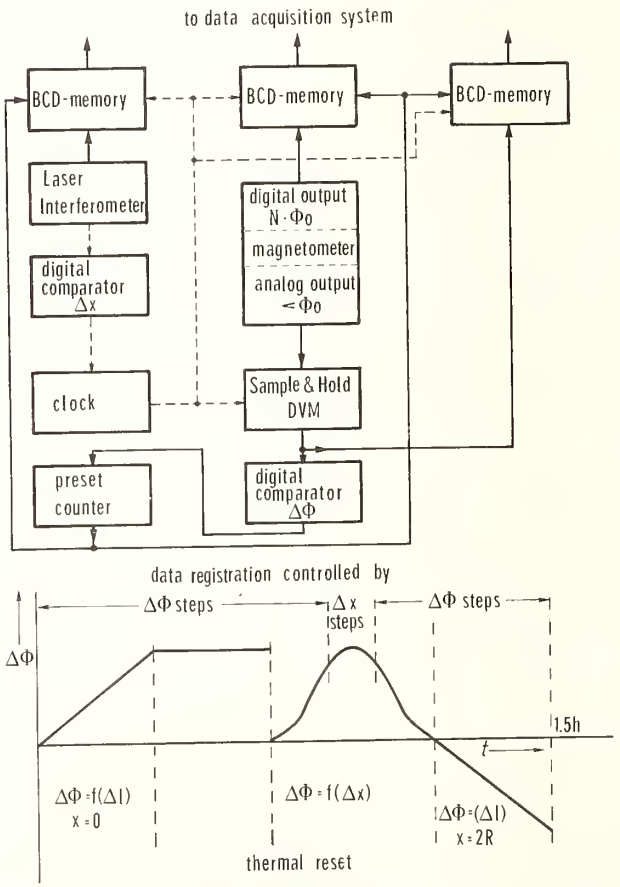


FIGURE 6. *Block diagram of flux profile recording circuit and time sequence of such a procedure.*

of the flux profile measuring circuit and the lapse of time of one of the field measuring procedures. The magnetometer readings are split in two parts; one part represents by means of a digital flux counter a value in terms of an integer multiple of the flux quantum ϕ_0, while the other part is an analog signal proportional to the fractional part of ϕ_0. This signal is observed by a fast acting sample and hold digital voltmeter, from which a command for data registration is initiated by comparing its output voltage with a predetermined value, for example, when a preset number of flux quanta $n\,\phi_0$ is encountered. Thus a record of correlated magnetometer and interferometer readings will be obtained in nearly constant steps of $n\,\phi_0$. In the range close to the center point the recording mode is modified; here correlated readings are recorded in equidistant x-steps.

6. Energizing the Coils

The flux profile measuring procedure takes more than 1 hour, meanwhile the current through the coils should not change. A block diagram of the complete field coil energizing set-up is given in Fig. 7. The left side of that figure shows the circuit components of the regulated main current source, essentially a digital to analog converter as a digitally presetable reference voltage source, the chopper stabilized control amplifier V_1, and the regulating unit T_1 connected to the feedback resistor R_v. By means of the other components in this branch—the counter circuits, the comparator, and the oscillator—a linearly rising or decaying current with selectable time constant may be generated. This is desirable for the beginning of a profile measuring run in order to avoid high flux slewing rates to the magnetometer and for comparing the coil constant of the outer pair of coils to that of the inner one. When the main current I_0 has reached the desired value, a superposed control loop is closed, thus providing a current stability in the order of 10^{-7} for currents in the range $20\text{ mA} < I_0 < 100\text{ mA}$. The loop consists of a high gain null detector, which compares the voltage drop at the standard resistor R_N with the Zener diode reference voltage U_R, and a proportional-integral-differential (PID) regulating stage followed by the voltage to current converter (V_2 and T_2) biased to about 30 μA. Instead of the shown superposed control loop based on voltage comparison a more sophisticated one using a SQUID-gradiometer system may be applied, resulting in a stability almost two orders of magnitude higher [12]; but this stability is not needed given the present state of the experiment. By means of two other current sources, each of them connected to one of the outer windings, additional currents up to 100 μA may be adjusted in order to optimize the field homogeneity at the center of the field. These current sources have to be floating because of the voltage drop at the windings due to the main current I_0, as indicated by the voltages U_{f1} and U_{f2}.

7. Systematic Errors

There are some sources of error limiting the accuracy in determining a coil constant by the presented method. By measuring mechanically the outer diameter of the four winding packages at 18 different generators, differences of about 0.2 mm have been found. Using this value in constructing a polygonal current path would result in a polygon with 90 sides, as indicated by the dashed line in Fig. 4, and that would cause, theoretically, a worst case relative error of nearly 10^{-8}. This error is small compared with those assigned to the magnetometry. Due to the area of the flux transformer loop the measured field values have to be corrected by considering the flux density over the loop area as a function of displacement x. This correction function may be derived from the vector potential expressed in terms of the flux density and its derivatives along the axis of the two current carrying loops:

$$\frac{\Delta B}{B} = \frac{3}{8}\left[\frac{\rho}{a}\right]^2\left[\frac{1-(2y_1)^2}{(1+y_1^2)^{7/2}} + \frac{1-(2y_2)^2}{(1+y_2^2)^{7/2}}\right] \times \left[(1+y_1^2)^{-3/2} + (1+y_2^2)^{-3/2}\right]^{-1} \qquad (8)$$

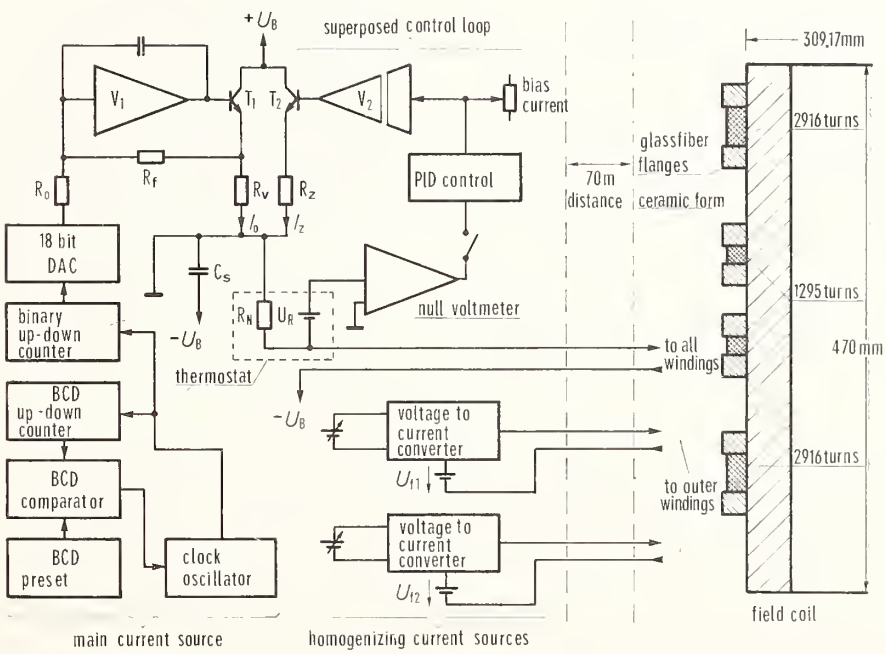


FIGURE 7. *Current sources and field coil system.*

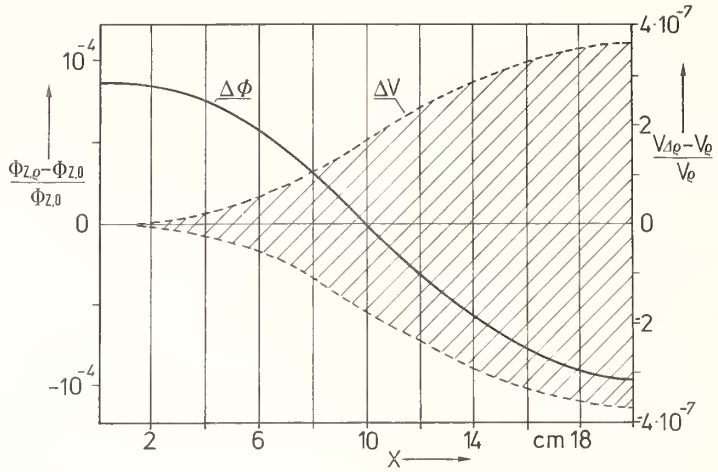


FIGURE 8. *Error in flux measurement due to the size of the flux transformer loop (lefthand ordinate), and resulting error in field ratio assuming an uncertainty in the loop radius of* $\Delta\rho/\rho = \pm 0.001$ *(righthand ordinate), both as a function of displacement x.*

with $y_1 = (x + A)/a$ and $y_2 = (x - A)/a$ [3]. Higher order terms of the evaluation need not be taken into account, because they are correlated to higher order terms of the ratio of radii of the sensor loop and of the field coil loop ρ/a. Equation (8) is illustrated in Fig. 8 with respect to the dimensions of the coils used in our experiment. Assuming an error in diameter of the sensor coil not larger than $\pm 10^{-3}$, the uncertainty in field ratio depending on position is given by the hatched area in Fig. 8, where it is to be noted that the right hand ordinate is to be used.

By examination of the correction function an adequate alignment of the SQUID-probe to the axis of the field coil is suggested by energizing the field coils by an ac-current and by looking for a minimum value of flux near the center and for a maximum value beyond the points of inflexion. Thus the error due to misalignment of the axis should be not larger than a few parts in 10^7. An error of the same order will occur due to the nonlinearity of the magnetometer.

Nevertheless, by means of some additional experiments it should be possible to determine the coil constant with a total relative uncertainty of less than 10^{-6}.

The author would like to thank Prof. Dr. Capptuller for many helpful discussions, Prof. Dr. Cammenga and his co-workers for preparing the NMR-samples, Dr. Theile and co-workers for doing the magnetic measurement in order to find the best place for the non-magnetic building, Dr. Vigoureux for the examination of the coil constant measuring method [14], and last but not least, the co-workers of the project for their assistance.

References

[1] B. Theile *et al.*, Magnetische Vermessung im Südgelände der PTB, private communication.
[2] W. Braunbeck, Z. Phys. 88, 399 (1934).
[3] A. L. Bloom and D. J. Innes, Appl. Phys. **36**, 2560 (1965).
[4] K. Weyand, PTB-Annual Report 1978, 3.2.24, p. 155.
[5] H. J. Petrick and H. K. Cammenga, Ber. Bungenges. Physik. Chem. 8, 1105 (1977).
[6] P. Vigoureux and N. Dupuy, NPL Report DES 49 (1980).
[7] W. Chiao *et al.*, IEEE Trans. Instrum. Meas. **IM-29**, 238 (1980).
[8] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. 42, 1575 (1979).
[9] K. Weyand, PTB-Mitt. 85, 206 (1975).
[10] K. Weyand, PTB-Annual Report 1978, 3.2.24, p. 155.
[11] K. Weyand and E. Simon, PTB-Annual Report 1977, 3.2.23, p. 146.
[12] K. Weyand, IEEE Trans. Instrum. Meas. **IM-29**, 324 (1980).
[13] W. Smythe, *Static and Dynamic Electricity* (McGraw-Hill, New York, 1950).
[14] P. Vigoureux, PTB-report-E-16, 1980.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A New Method for the Determination of the Proton Gyromagnetic Ratio

G. L. Greene

Gibbs Laboratory, Physics Department, Yale University, New Haven, CT 06520

A method for the determination of γ'_p is proposed which involves the separated oscillatory field magnetic resonance technique with flowing water. An association between the volume integral of a magnetic field in a cylinder (determined by a resonance frequency) and the line integral along its axis (determined by Ampere's law) is made. An accuracy of a few parts in 10^8 may be attainable with this technique.

Key words: flowing water NMR; fundamental constants; magnetic resonance; precision measurements; proton gyromagnetic ratio; Ramsey technique.

1. Introduction

The principal interest in the measurement of the proton gyromagnetic ratio (γ'_p) lies in its application as a very sensitive test of QED. The importance of γ'_p arises in its role in the determination of the fine structure constant through the relation [1]:

$$\alpha^{-2} = [(c/4R_\infty)(\Omega/\Omega_{LAB})(\mu'_p/\mu_B)(2e/h)]/\gamma'_p, \qquad (1)$$

where $2e/h$ and γ'_p are measured in terms of as-maintained electrical units and Ω/Ω_{LAB} is the ratio of the S.I. to as-maintained ohm. Such a determination of α is not dependent on QED and can therefore be compared with the results of $g-2$ experiments to provide a sensitive test of the theoretical predictions of QED. It shoud be noted that the uncertainty in α as determined by Eq. (1) is dominated by the uncertainty in γ'_p. Thus, a more accurate determination of γ'_p leads directly to an improved knowledge of α.

The major limitation to the determination of γ'_p has been the accurate absolute determination of a magnetic field [2]. It has been suggested [3] that a field determination technique based directly on Ampere's law may allow a significant improvement in the accuracy of the field determination. This suggestion has been incorporated into an experimental proposal which involves a movable SQUID magnetometer or pickup loop [3] to take a pointwise average of a magnetic field along a path through a solenoid. This technique, while extremely promising, has the disadvantage that it requires a pointwise integration. Such pointwise integrations are difficult to obtain to high accuracy and can lead to unforeseen systematic errors. An alternative technique involving the Faraday effect [4] has problems related to variations in the index of refraction.

It has recently been demonstrated that Ramsey's method of separated oscillatory field magnetic resonance, using flowing water, can be employed as a sensitive field averaging magnetometer [5–7]. This technique can be applied to the measurement of γ'_p in such a way that there is no need for a highly accurate pointwise integration [8]. With such a technique, the determination of γ'_p is reduced to a frequency measurement, a measurement of a current, and a measurement of a single length. An accuracy in the determination of γ'_p of one part in 10^8 may be possible.

2. Experimental Technique

A schematic outline of the apparatus can be seen in Fig. 1. To simplify the explanation of the technique the following assumptions are made (the validity of these assumptions will be discussed later):

a) Only longitudinal fields (parallel to the axis of the water tube) will be considered;
b) Demagnetization and chemical shifts are known;
c) The tube is aligned on the solenoid axis;
d) The tube is uniform along its length;
e) Field inhomogeneity shifts, Bloch-Siegert effect, etc., are negligible.

It is further assumed that, as suggested by Williams and Olsen [3, 4], the windings of the solenoid are arranged in such a way that the field falls off rapidly outside the solenoid. In particular it is assumed that the field at, and beyond, the separated Ramsey coils is insignificantly small (or at least accurately calculable).

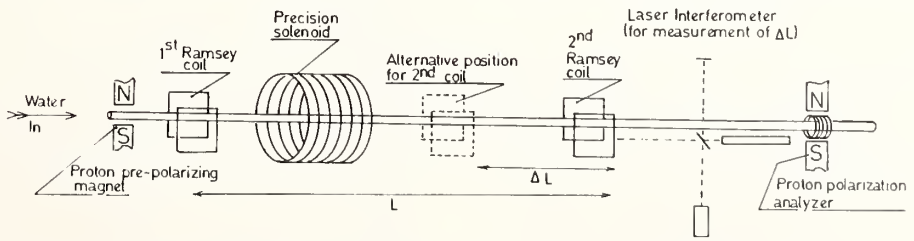


FIGURE 1. *Schematic outline of γ'_p apparatus. The second Ramsey coil can be shifted and its change in position accurately determined with the laser interferometer.*

The resonance frequency ω in the Ramsey arrangement will be given by [5, 6] (with assumptions b-e)

$$\omega = \frac{\gamma'_p}{L} \int_0^L |\mathbf{B}| \, d\ell. \tag{2}$$

where L represents the Ramsey coil separation. Using the assumption that the tube and the field are parallel, it is possible to express Eq. (2) as

$$\omega = \frac{\gamma'_p}{L} \int_0^L \mathbf{B} \cdot d\,\ell. \tag{3}$$

By Ampere's Law it is known that for any path through the solenoid:

$$\int_{-\infty}^{\infty} \mathbf{B} \cdot d\,\ell = \mu_0 N I \, . \tag{4}$$

Now using the assumption that the field at and beyond the Ramsey coils (i.e., for $X < 0$ and $X > L$) is negligible, Eqs. (3) and (4) can be combined to yield

$$\omega = \frac{\gamma'_p \mu_0 N I}{L} . \tag{5}$$

It should be noted that in general it will *not* be possible to determine L to high accuracy.

The experiment is repeated with the coil separation reduced by an amount ΔL. This change in coil separation can be measured very accurately by means of the laser interferometer. The resonance frequency ω' for this case will be given by

$$\omega' = \frac{\gamma'_p \mu_p N I}{L - \Delta L} . \tag{6}$$

Combining Eqs. (5) and (6) and solving for γ'_p one obtains

$$\gamma'_p = \frac{\Delta L \, \omega' \, \omega}{\mu_0 N I (\omega' - \omega)} . \tag{7}$$

Note that all the quantities on the right of Eq. (7) can ultimately be determined to very high accuracy (perhaps a few parts in 10^8). This allows a very sensitive measurement of γ'_p to be made by (in principle) a determination of:

1) The two frequencies ω, ω'
2) The distance ΔL,
3) The current I.

This promises to be a greatly simplified method for the accurate determination of γ'_p.

3. Systematic Effects and Possible Sources of Error

Several assumptions were required to lead to the result of Eq. (7). These are dealt with in turn. All of the effects, as well as other sources of systematic error, are discussed in detail elsewhere [8].

3.1 Effects of Transverse Fields

Transverse fields may arise from external sources or from the solenoid. In the case of external fields, the spurious effect may be examined by a systematic variation of the current I. Such a variation of I can also be used to account for stray longitudinal fields. Transverse fields from the precision solenoid may either be calculated or examined experimentally by changes in tube diameter. If the field on the axis of the solenoid is reasonably well known, the difference between the volume integral of $|\mathbf{B}|$ and the line integral of $\mathbf{B}$ can be calculated sufficiently well so that no significant spurious effect results [8]. For a reasonably sized apparatus, the field on axis need be known to ~1% in order for this shift ($\sim 10^{-6}$) to be known to approximately one part in 10^8.

3.2 Diamagnetic Effects

The proton gyromagnetic ratio γ'_p is defined as the gyromagnetic ratio for protons in a spherical sample of water (at a temperature of 25 °C). In the procedure here, a measurement of the gyromagnetic ratio for protons in a cylindrical sample of H_2O aligned with the magnetic field is made. In this geometry, the field seen by the protons will be altered from that experienced by protons in a spherical sample (assuming the same field) due to the diamagnetism of water.

This shift will be given by $(1 + 4\pi\kappa/3) = -3.010(4) \times 10^{-6}$ where κ is the volumetric susceptibility of water (at 25 °C). This implies an uncertainty in the final measurement due to the correction for the bulk diamagnetism of water of about 4 parts in 10^9.

There is a variation in this correction with temperature (due primarily to bulk expansion in water) of approximately -2×10^{-8} per °C. This implies that the water temperature must be kept to within ±0.5 °C for a measurement to a part in 10^8.

3.3 Tube Alignment

It is vitally important that the axis of the tube be parallel with the magnetic field of the solenoid. Otherwise, one will not have equality between Eqs. (2) and (3). If the tube is "canted" by an angle θ, one will expect the $\int |\mathbf{B}| d\ell$ to differ from $\int \mathbf{B} \cdot d\ell$ by approximately $(1 - \cos\theta)$. To insure accuracy to 1 part in 10^8, alignment must be accurate to 10^{-4} radians. This may be difficult, but it is certainly possible. It should be noted that $\int \mathbf{B} \cdot d\ell$ is a lower bound for $\int |\mathbf{B}| d\ell$. Thus any adjustment which minimized ω is correct. Sagging or deformation of the tube must be controlled to a like degree, or calculated.

3.4 Tube Uniformity

If the tube's cross-sectional area varies along its length, then the field average taken by the protons will not be a uniformly "weighted" average. This problem poses very serious constraints on the tube uniformity. The effect of variations in tube cross section can be accounted for in a variety of ways which involve shifting the tube within the coils. While it is possible to correct for the systematic shifts due to such nonuniformity, this effect will probably pose a major problem in the successful application of this technique.

3.5 Other Systematic Effects

There are a number of small effects which can give rise to shifts and errors in any "beam" type resonance experiment. Among these are coil phase errors, Bloch-Siegert effect, and effects due to field inhomogeneities. None of these effects is expected to give a significant contribution to the error of a final measurement. However, it will be necessary to employ a series of trim coils to eliminate effects due to field inhomogeneities [8].

In addition to those mentioned in the above assumptions, attention must be paid to a variety of other effects which need to be well understood. These include details of fluid flow and mixing, field averaging, field inhomogeneity, T_2 relaxation, determination of current I, deter-

mination of ΔL, line center determination, and signal to noise ratio. It appears [8] that the errors which would be associated with these effects can be kept at the level of a few parts in 10^8.

4. Conclusion

It is possible that a significant improvement over the current uncertainty of 0.21 ppm in γ'_p will be very difficult without the application of a new technique. The heart of such a new technique may well lie in the direct application of Ampere's law to the determination of the magnitude of a magnetic field as suggested by Williams and Olsen [3, 4]. It is suggested that this approach may be greatly simplified by the use of the Ramsey separated oscillatory field magnetic resonance method using flowing water. An ultimate accuracy of a few parts in 10^8 may be attainable.

I wish to thank Dr. E. R. Williams, Dr. W. D. Phillips and Dr. B. N. Taylor of the Center for Absolute Physical Quantities, National Bureau of Standards, for several suggestions and comments.

This work was supported in part by NSF contract PHY 78–25655.

References

[1] B. N. Taylor, W. H. Parker, and D. N. Langenberg, Rev. Mod. Phys. **41**, 375 (1969).
[2] E. R. Williams and P. T. Olsen, Phys. Rev. Lett. 42, 1575 (1979).
[3] E. R. Williams and P. T. Olsen, IEEE Trans. Instrum. Meas. **IM-27**, 467 (1978).
[4] E. R. Williams, W. D. Phillips, and P. T. Olsen, unpublished (1978).
[5] G. L. Greene, Ph.D. Thesis, Harvard University (1977).
[6] J. M. Pendlebury, K. Smith, P. Unsworth, G. L. Greene, and W. Mampe, Rev. Sci. Instrum. 50, 535 (1979).
[7] G. L. Greene, *et al.*, Phys. Rev. D 20, 2139 (1979).
[8] G. L. Greene, Research Proposal to NRC (1979), unpublished; and G. L. Greene, Metrologia 18, 93 (1982).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Quantized Hall Resistance in Two-Dimensional Systems

K. von Klitzing, H. Obloh, and G. Ebert

Physik-Department, Technische Universität München, D-8046 Garching, F.R.G.

and

J. Knecht and K. Ploog

Max-Planck-Institut für Festkörperforschung, D-7000 Stuttgart, F.R.G.

After the Josephson effect, a second solid state quantum effect—the quantized Hall resistance—seems to be useful to determine fundamental constants. We will demonstrate that electrons at the interface between two semiconductors or at a semi-conductor-insulator interface may form at low temperatures a degenerate two-dimensional electron gas with discrete energy levels (Landau levels), if a strong magnetic field is applied perpendicular to the interface. Under experimental conditions where an integer number i of Landau levels is fully occupied with electrons, the value of the Hall resistance R_H (ratio between Hall voltage and current through the sample) becomes quantized with values which depend only on fundamental constants: $R_H = h/e^2 i$ (h = Planck constant, e = elementary charge). Our experiments on (100) silicon MOSFETs (Metal-Oxide-Semiconductor Field-Effect-Transistors) and GaAs-Al_xGa_{1-x}As heterostructures show that the measured value of the quantized Hall resistance is independent of device parameters. The analysis of high precision measurements on different MOSFETs yields a value of $h/e^2 = (25812.79 \pm 0.04)\ \Omega$, which is about 1 ppm smaller than the recommended value. Calculations show that in our experiments corrections to the value of the quantized Hall resistance, which can be further reduced by increasing the magnetic field or decreasing the temperature, should be smaller than 10^{-7}.

Key words: fine structure constant; Hall effect; magneto-quantum-transport; MOSFET; resistance standard; two-dimensional electron gas.

1. Introduction

Recent measurements of the Hall resistance R_H (Hall voltage U_H divided by the current I through the sample) on a two-dimensional electron gas have shown that at low temperatures ($T < 4$ K) and strong magnetic fields ($B > 10T$) this resistance is quantized in units of $h/e^2 i \approx 25813/i\ \Omega$ (h = Planck constant, e = elementary charge, $i = 1, 2, 3, \ldots$) [1]. Since the present uncertainty in the realization of the SI unit Ohm is less than 10^{-7} [2], measurements of the quantized Hall resistance can, in principle, be used to determine directly in SI units with high accuracy the ratio h/e^2 and therefore the fine-structure constant α:

$$\alpha = \frac{\mu_o c}{2} \frac{e^2}{h} \tag{1}$$

with the permeability of vacuum $\mu_0 \equiv 4\pi \times 10^7$ H/m and the velocity of light $c = 2.99792458 \times 10^{-8}$ m/s. On the other hand, the quantized Hall resistance can be used to develop a resistance standard based on fundamental constants, if the value of h/e^2 is known.

The new quantum effect is based on the fact that the energy levels of a two-dimensional electron gas (2DEG) in a strong transverse magnetic field B becomes fully quantized with the same number of states $N = eB/h$ (degeneracy factor) for each level. Under experimental conditions, where an integer number of energy levels is occupied, we will demonstrate that the Hall resistance depends only on fundamental constants and not on the magnetic field, device geometry, or other parameters. Such a result is not expected because the Hall resistance R_H of a three-dimensional degenerate electron gas depends on inaccurate parameters like the thickness d of the sample, the strength of the magnetic field B, and the carrier concentration n_c:

$$R_H = \frac{B}{e n_c d} . \tag{2}$$

However, for a 2DEG the parameter d is incorporated in the two-dimensional carrier density n_{inv} and it will be shown that the ratio n_{inv}/B becomes quantized in a strong magnetic field in integer multiples of e/h. This leads to a quantization of the Hall resistance with resistance values of

$$R_H = \frac{h}{e^2 i} \quad i = 1, 2, \ldots \tag{3}$$

In this paper we will, after an introduction to the transport properties of a 2DEG in a strong magnetic field, summarize our experimental results obtained on two different systems of a 2DEG—the electrons in the inversion layer of a (100) silicon MOSFET (Metal-Oxide-Semiconductor Field-Effect-Transistor) and the electrons at the interface of a GaAs-Al_xGa_{1-x}As heterostructure.

2. The Quasi Two-Dimensional Electron Gas

A quasi two-dimensional electron gas (2DEG) is formed, if electrons can move freely within a plane (x-y plane) but are trapped in the z-direction. Such a 2DEG can be realized at the surface of a semiconductor, if a sufficiently large electrostatic field F_z is applied normal to the interface, which causes a drop in electron potential towards the surface. If the width of this potential well is small compared to the wavelengths of the carriers, the energy levels of the electrons are grouped in so-called electric subbands E_j corresponding to quantized levels for motion in the z-direction, the direction perpendicular to the surface, and a continuum for motion in the plane parallel to the surface [3]. Experimentally the existence of a 2DEG was demonstrated by Fang and Howard [4]. They measured the channel conductance of a silicon MOSFET in strong magnetic fields at helium temperatures and found that the observed quantum oscillations originate from a 2DEG. Figure 1 illustrates the situation when such a 2DEG (inversion layer) is formed on a p-type semiconductor. The surface electric field F_z is built up by a positive voltage (gate voltage V_g) at the metal side of the metal-oxide-semiconductor capacitor. The downwards bending of the conduction band originates

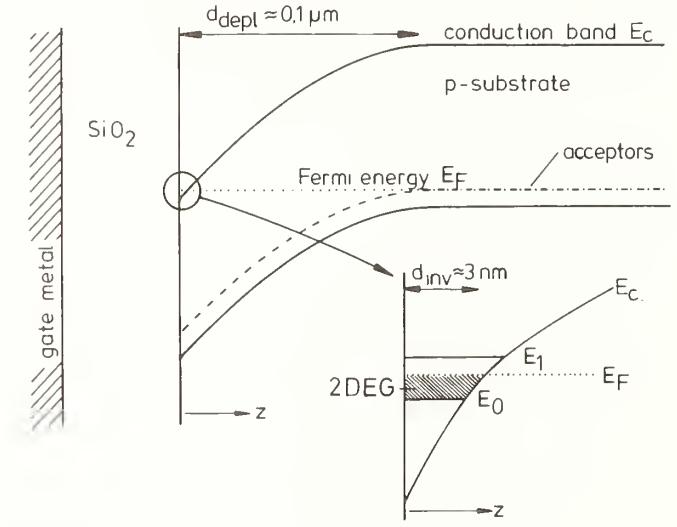


FIGURE 1. *Formation of a two-dimensional electron inversion layer at the surface of a p-type semiconductor.*

mainly from fixed negatively charged acceptors in the depletion layer close to the surface. At gate voltages above a certain threshold voltage V_{th} the depletion layer d_{depl} is thick enough to bend the conduction band near the surface below the Fermi level E_F forming a narrow nearly triangular potential well in which the 2DEG is bound. The concentration n_{inv} per unit area of electrons in this potential well can be calculated from the capacitance of the surface capacitor C_{ox}:

$$n_{inv} \approx \frac{C_{ox}}{e} (V_g - V_{th}) . \tag{4}$$

The energies of the electric subbands E_j (j = 0, 1, 2, . . .) can be estimated by calculating the energy eigenvalues of an electron in a triangular potential with an infinite barrier for $z < 0$ and a constant electric field $F_s = eN_G/\epsilon_o \epsilon_s$ for $z \geq 0$ (N_g = concentration per unit area of positive charges at the gate, $\epsilon_o \epsilon_s$ = static dielectric constant of the semiconductor). The solutions are approximately [3]:

$$E_j = \left[\frac{h^2}{2m_z}\right]^{1/3} \left[\frac{3}{2}\pi e F_s\right]^{2/3} \left[j + \frac{3}{4}\right]^{2/3} \quad (j = 0.\ 1,\ 2,\ \ldots) , \tag{5}$$

where m_z is the effective mass for the motion perpendicular to the surface. For silicon with a (100) surface orientation, two of the six valleys of the conduction band have an $m_Z = 0.916\ m_o$ and four valleys have an $m_z = 0.19\ m_o$. The lowest electric subband E_o belongs to the two valleys with the higher effective mass m_z and an effective mass for the motion parallel to the surface of $m_x = m_y = 0.19\ m_o$; the other four valleys form a series of electric subbands E'_j (j = 0, 1, 2, . . .) starting at higher energies. For a more quantitative calculation of the energies of the electric subbands it is necessary to solve the Schrödinger equation for the actual potential V'_z which changes with the distribution of the electrons in the inversion layer:

$$\left\{ -\frac{\hbar^2}{2m_x}\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m_y}\frac{\partial^2}{\partial y^2} \right\} \left\{ -\frac{\hbar^2\partial^2}{2m_z\,\partial z^2} + V'_z \right\} \psi = E\psi . \tag{6}$$

In the effective-mass approximation, the electronic wavefunction for the j^{th} subband can be characterized by an envelope function

$$\psi = \phi_J(z) \exp(ik_x x + ik_y y) , \tag{7}$$

and the motion of the electron can be separated into a part parallel to the interface and a part perpendicular to it with an energy

$$E(k_x, k_y, j) = \frac{\hbar^2 k_x^2}{2m_x} + \frac{\hbar^2 k_y^2}{2m_y} + E_j . \tag{8}$$

If only the lowest electric subband E_o is occupied, the electrons form a strictly two-dimensional electron gas with a constant density of states per unit area

$$D(E) = 2\pi \sqrt{m_x m_y}\ /h^2$$

The effective potential V'_z in Eq. (6) contains not only the electrostatic potential due to the electrons in the inversion layer and the charges in the depletion layer (Hartree potential) but also many-body effects [5-7] (exchange and correlation potential). Different authors have calculated numerically the energies of the electric subbands.

Experimentally, the separation between electric subbands can be measured by analyzing the resonance absorption of electromagnetic waves with a polarization of the electric field perpendicular to the interface [8]. Actually the resonance absorption energy is not identical with the difference between the one-electron energies $E_1 - E_o$, because the interband excitation induces a polarization in the inversion layer and the electron-hole interaction leads to an exciton shift. However, the total correction is relatively small because the exciton shift cancels to a certain extent the polarization shift [9]. Figure 2 shows measurements of the resonance absorption energy for transitions between the electric subbands E_o and E_1 as a function of the electron concentration n_{inv} in the surface channel of a (100) silicon MOSFET [10]. The results are plotted for two different substrate bias voltages V_{SB} – the externally applied voltage between

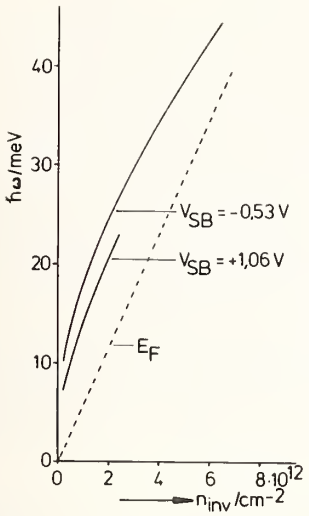


FIGURE 2. *The resonant positions of intersubband transitions $E_1 - E_o$ as a function of the inversion carrier concentration n_{inv} for a (100) silicon MOSFET. The results are plotted for two different substrate bias voltages V_{SB} [10]. The dotted line corresponds to the Fermi energy E_F relative to the lowest subband E_o.*

the surface channel and the p-type substrate. Such a substrate bias voltage changes the thickness of the depletion layer

$$d_{depl} = \sqrt{2\epsilon_o \epsilon_{Si}(V_i - V_{SB})/e\,(N_A - N_D)}\,, \quad (10a)$$

and therefore the charges Q_d in the depletion layer,

$$|Q_d| = \sqrt{2\epsilon_o \epsilon_{Si} e\,(N_A - N_D)\,(V_i - V_{SB})}\,. \quad (10b)$$

The built-in voltage V_i for silicon at low temperatures is about 1.1 V corresponding to an energy gap of 1.1 eV. Since the contribution of the depletion charge to the electric field at the surface is proportional to Q_d, the surface electric field can be varied without changing the electron concentration in the surface channel. Therefore the energy distance between the electric subbands becomes larger with increasing negative substrate bias voltage as shown in Fig. 2. The lowest subband separation at a fixed n_{inv} is obtained at $V_{SB} = +1.06$ V corresponding to a vanishingly small depletion charge. The Fermi energy $E_F - E_o$ (dotted line in Fig. 2) is for surface carrier concentrations less than $n_{inv} = 6 \times 10^{12}$ cm^{-2} and is always smaller than the energy separation between the electric subbands. Therefore only the lowest electric subband E_o is occupied at low temperatures (electric quantum limit). The question, whether the higher electric subband E_1 or a new subband series E'_o becomes occupied at extremely high surface carrier concentrations, has not been solved unambiguously [11, 12]. At least at inversion carrier concentrations below 2×10^{12} cm^{-2} or for measurements with a negative substrate bias voltage, the condition for the electric quantum limit and therefore for a strictly two-dimensional electron gas is fulfilled.

A two-dimensional electron gas can not only be formed at the semiconductor-insulator interface but also at the interface between two semiconductors [13]. Figure 3 illustrates the situation for a GaAs-$Al_xGa_{1-x}As$ heterostructure. In this case the narrow potential well for the 2DEG originates from the difference in the energy gaps of the two semiconductors and the depletion layer on the GaAs-side. The electron concentration in the interface

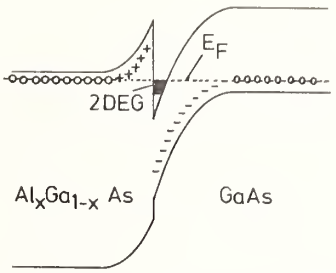


FIGURE 3. *Schematic energy level diagram of a GaAs-$Al_xGa_{1-x}As$ heterostructure. The p-type GaAs is weakly doped compared to the doping level of the n-type $Al_xGa_{1-x}As$.*

channel is mainly determined by the concentration of ionized donors in the depletion layer of the highly doped $Al_xGa_{1-x}As$. Without interface states, the slope of the conduction and valence band at the interface is the same for both semiconductors. For the 2DEG in GaAs only one series of electric subbands E_j is expected, because the conduction band of GaAs has only one minimum in the first Brillouin zone with an isotropic mass of $m_x = m_y = m_z \approx 0.07\ m_o$.

Contrary to the MOS system, the carrier concentration of the 2DEG in a heterostructure cannot be varied over a wide range. Small changes in n_{inv} up to $\delta n_{inv} \approx 4 \times 10^{11}$ cm^{-2} are observed, if the device is illuminated with intrinsic light [14] or if a voltage between a Schottky contact at the surface of the heterojunction and the interface channel is applied [15]. Since leakage currents across the Schottky barrier may falsify experimental results, our high-precision measurements are performed on devices without a Schottky contact and without intrinsic illumination. GaAs-MOSFETs are not available up to now because of a large number of interface states located energetically within the energy gap. Top views and cross-sections of typical silicon MOSFETs and GaAs-$Al_xGa_{1-x}As$ heterostructures used in our experiments are shown in Fig. 4. For our MOSFETs a SiO_2 layer is thermally grown with a thickness of typically 0.1μm – 1μm on a single crystal of p-Si. A metal (usually Al) on the top of the oxide forms the gate. Highly doped n^+ regions at the semiconductor surface are used as ohmic contacts to the inversion layer. The depletion layer between the n^+ contacts or the n-type inversion layer and the p-type substrate separates electrically the surface channel from the bulk. The arrangement of the n^+ contacts shown in Fig. 4 allows measuring in the presence of a magnetic field either the conductivity com-

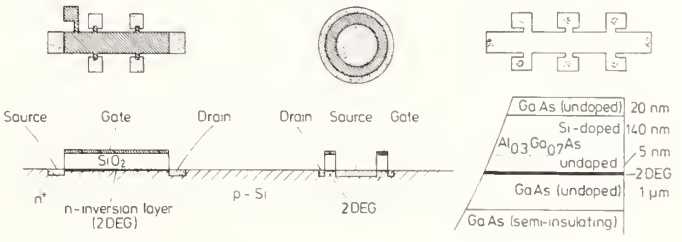


FIGURE 4. *Typical geometries and cross-sections of devices used in our experiments. From left to right:*
(a) long-silicon MOS-device for Hall effect measurements with a typical length of 0.5 mm;
(b) circular MOS-device for conductivity measurements with a diameter of about 100 μm;
(c) cross-section and top view of a GaAs-$Al_{0.3}$ $Ga_{0.7}As$ heterostructure used in our experiments.

ponent σ_{xx} (circular device) or the resistivity components ρ_{xx} and ρ_{xy} (long device).

For the GaAs-$Al_xGa_{1-x}As$ heterostructures, the different layers are grown using MBE (molecular beam epitaxy) techniques [16]. Basically the heterostructure consists of a silicon doped $Al_xGa_{1-x}As$-layer ($x \approx 0.3$, $N_{Si} = 3 \times 10^{17}$ cm^{-3}) on top of an undoped GaAs-layer as shown in Fig. 4. The resulting carrier density of the 2DEG is about 5×10^{11} cm^{-2}. The Hall geometry of the device is obtained by a mesa-etching process. Indium metal was alloyed into the $Al_xGa_{1-x}As$-layer to make electrical contacts to the 2DEG for Hall effect and resistivity measurements.

3. Quantum Transport of a 2DEG in Strong Magnetic Fields

3.1 Density of States

A strong magnetic field B with a component B_z normal to the surface causes the inversion layer electrons to move in cyclotron orbits parallel to the interface.

As a consequence of the orbital quantization the energy levels of a 2DEG in a strong magnetic field can be written schematically in the form [silicon(100)-inversion layer]:

$$E_{n,s,v} = E_o + (n + \frac{1}{2})\hbar\omega_c + sg\,\mu_B B + vE_v\,. \quad (11)$$

where $s = \pm 1/2$ (spin quantum number), g is the Landé g-factor [17], $v = \pm 1/2$ (valley quantum number), E_v is the valley splitting [18, 19], and $\hbar\omega_c = \hbar(eB_z/\sqrt{m_x m_y})$ is the energy difference between Landau levels with quantum numbers n and $n + 1$. For a 2DEG in GaAs the last term in Eq. (11) is zero. The Landau splitting $\hbar\omega_c$ depends on B_z, the magnetic field component perpendicular to the x-y plane of the 2DEG, whereas the spin splitting depends on the total magnetic field B. This allows one to change their energies independently by changing the orientation of the magnetic field relative to the interface.

The wavefunction of a 2DEG in a strong magnetic field may be written in a form where one of the center coordinates of the cyclotron motion is a good quantum number [20]. The degeneracy of each level $E_{n,s,v}$ is given by the total number of center coordinates of the system. Since only one center exists in a region of area $\pi\ell^2$ [21] ($\ell = \sqrt{\hbar/eB}$ is the radius of the ground Landau orbit) the degeneracy factor per unit area is given by

$$N = \frac{eB}{h}\,. \quad (12a)$$

This degeneracy factor for each energy level $E_{n,s,v}$ can also be calculated from the number of states (for $B = 0$) in the energy range $\hbar\omega_c$, since the condition of total state conservation leads to a condensation of the states within this energy range on one Landau level

$$N = D(E)\hbar\omega_c = \frac{eB}{h}\,. \quad (12b)$$

This degeneracy factor (number of states per unit area for each energy level) plays an important role in the analysis of the quantized Hall resistance because the surface carrier concentration for fully occupied Landau levels becomes quantized in units of eB/h.

3.2 Magneto-Conductivity σ_{xx}

The modification of the density of states in a strong magnetic field leads to a drastic change in the conductivity. Ando has calculated the level broadening and the transverse conductivity σ_{xx} of a 2DEG in strong magnetic fields [20, 22] and finds that the density of states of each level $E_{n,s,v}$ is not a delta function, but has an elliptic form in the self-consistent Born approximation (SCBA) with energy gaps between the levels. Additional tails at the edges of the density of states appear in a higher approximation [23]. The conductivity σ_{xx} at low temperatures is mainly determined by the square of the density of states at the Fermi energy and has peak values (if one level is half filled) of $\sigma_{xx}^{max} = (e^2/\pi^2\hbar)(n + 1/2)$, if the scatterers are of short-range. In the case that the scatterers are of finite range, this peak value is reduced [20].

The characteristic features of the conductivity can be understood by a simple diffusion of the center coordinate of the cyclotron orbit, where the conductivity is proportional to the product of the diffusion constant and the density of states at the Fermi energy. Without scattering, the electrons cannot diffuse in the direction of the electric field (they are drifting perpendicular to the electric and magnetic field). Therefore an observed vanishing conductivity σ_{xx} in strong magnetic fields corresponds to a vanishingly small scattering rate.

Figure 5 shows a typical result for the conductivity σ_{xx} (measured with a circular MOS device) as a function of the gate voltage at two different magnetic field values of B. The oscillations are related to the variation of the density of states at the Fermi energy. At gate voltages where σ_{xx} drops to zero, the density of states $D(E_F)$ is zero. The valley splitting is resolved only for the Landau levels up to $n = 1$, whereas the spin splitting is visible for Landau levels up to $n = 3$. The period in the gate voltage is constant within 1% and increases linearly with increasing magnetic field within the experimental accuracy of 0.5%. This result is compatible with Eqs. (4) and (12), which means that the inversion carrier concentration increases proportionally to the gate voltage V_g, and the degeneracy of each level increases proportional to the magnetic field.

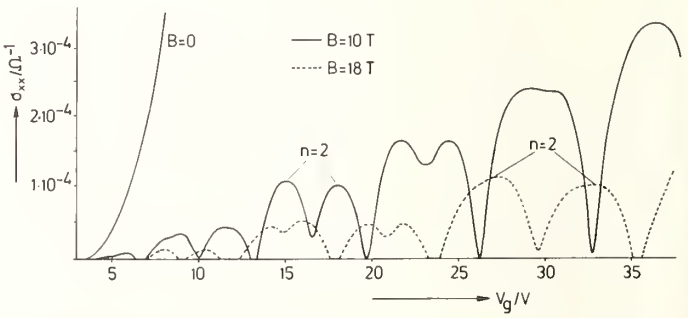


FIGURE 5. *The conductivity σ_{xx} as a function of gate voltage V_g measured with a circular silicon MOSFET at different magnetic field values B(T = 1.5 K).*

At gate voltages where σ_{xx} drops to zero, an energy level $E_{n,s,v}$ is fully occupied. This means that exactly at these gate voltages the inversion carrier concentration is a multiple of the degeneracy N of a Landau level

$$n_{inv} = iN = i\frac{eB}{h} \quad (i = 1, 2, 3, \ldots)\,. \quad (13)$$

At these gate voltages a Hall resistance

$$R_H = \rho_{xy} = \frac{B}{en_{inv}} = \frac{h}{e^2 i} \quad (i = 1, 2, 3, \ldots) \quad (14)$$

is expected. For high-precision measurements of this quantized resistance an accurate adjustment of the gate voltage V_g corresponding to fully occupied Landau levels is necessary.

Such an adjustment of V_g with an uncertainty of less than 10^{-6} seems to be impossible due to instabilities in both the gate voltage and the threshold voltage. In addition, the σ_{xx} curve close to σ_{xx}^{min} is strongly asymmetric (Fig. 6, see also Ref. [24]) and the position of the minimum is not well defined. Moreover, at very high magnetic fields (Fig. 5, $B = 18$ T) σ_{xx} becomes immeasurably small ($\sigma_{xx} < 10^{-9}\ \Omega^{-1}$) within a certain region ΔV_g, usually referred to as the localized region. Therefore, the gate voltage position of the minimum in σ_{xx} cannot be determined. Kawaji and Wakabayashi investigated this localization for silicon MOSFETs in detail and they found that for Landau quantum numbers $n \geq 1$, approximately a fraction $1/(2n+1)$ of a Landau level is localized [25]. Different theories (localization due to lattice imperfections [26-28], electron-electron interaction [29, 30], or simply an exchange of electrons between the inversion channel and an electron reservoir in the SiO_2- or the $Al_xGa_{1-x}As$-layer [31], are able to explain the observed "localization" of electrons.

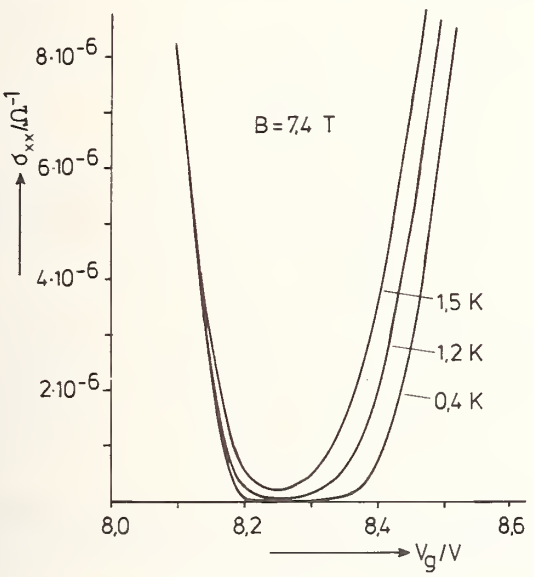


FIGURE 6. *Temperature dependence of σ_{xx} (measured at B = 7.4 T) close to the gate voltage of the fully occupied Landau level $n = 1$.*

Up to now, it is not clear at which gate voltage V_g within the localized region all states of the energetically lower lying Landau levels are occupied. The results in Fig. 6 ($T = 0.4$ K) seem to indicate that in the localized region states of the higher energetic Landau level become occupied (in agreement with calculations for a two-dimensional disordered electron system in strong magnetic fields [28]), but measurements by other authors [32] or on other samples do not confirm this result.

The uncertainty in the location of the gate voltage for which Eqs. (13) and (14) are fulfilled, seems to prevent a highly accurate adjustment of $n_{inv} = ieB/h$ and therefore an accurate determination of the quantized Hall resistance. Fortunately, Hall effect measurements show that the Hall resistance remains extremely constant in the gate voltage region where localization is observed and has a value corresponding to a complete occupation of an integer number of energy levels [33].

3.3 Hall Conductivity σ_{xy}

Experimentally, the Hall conductivity σ_{xy} cannot be determined from measurements with a circular device, because the azimuthal component of the current $j_\phi = \sigma_{xy} E_{SD}$ cannot be measured (E_{SD} = source-drain field). However, for a long device both the current along the long axis (source-drain current) and the electric field perpendicular to the current (Hall field) can be measured. If the current has only one component ($j = j_x$), the potential drop along the current direction is proportional to the component ρ_{xx} of the resistivity tensor, and the Hall field is directly proportional to ρ_{xy} [34]. It should be noted that in two-dimensional systems ρ_{xy} is given directly by the ratio of the measured quantities U_H (Hall voltage) and I_x (source-drain current) and is (apart from small corrections) independent of the dimensions of the sample. Therefore, the experimentally determined quantities $R_H = U_H/I_x$ (Hall resistance) and R_x (resistivity of the sample measured between potential probes) are identical with ρ_{xy} and ρ_{xx}, if $R_x = 0$. The corrections $(R_H - \rho_{xy})/\rho_{xy}$ and $(R_x - \rho_{xx})/\rho_{xx}$ for $R_x \neq 0$, which will be calculated in Section 3.4, are unimportant for the following discussion and we can set $R_H = \rho_{xy}$ and $R_x = \rho_{xx}$.

The corresponding components of the conductivity tensor σ_{xx} and σ_{xy} can be calculated from the equations ($\rho_{xx} = \rho_{yy}$, $\rho_{xy} = -\rho_{yx}$):

$$\sigma_{xx} = \rho_{xx}/(\rho_{xx}^2 + \rho_{xy}^2) \quad (15)$$

$$\sigma_{xy} = -\rho_{xy}/(\rho_{xx}^2 + \rho_{xy}^2) \quad (16)$$

In the limit of fully occupied Landau levels ($\sigma_{xx} = 0$) eqs. (15) and (16) reduce to $\rho_{xx} = 0$ and $\rho_{xy} = -1/\sigma_{xy}$.

3.3.1 Measurements on Silicon MOSFETs

A typical result for $R_x \approx \rho_{xx}$ and $R_H \approx \rho_{xy}$ as a function of the gate voltage at a constant magnetic field is shown in Fig. 7. The oscillations in $\rho_{xx}(V_g)$ are mainly determined by the oscillations in $\sigma_{xx}(V_g)$, because Eq. (15) reduces to $\sigma_{xx} = \rho_{xx}/\rho_{xy}^2$ in strong magnetic fields ($\rho_{xy} > \rho_{xx}$).

Similar to the measurements on circular device (Fig. 5) "localized" regions ($R_x = 0$) around $V_g = 5$ V, 10V, and 21 V are visible. In these regions a constant Hall resistance is measured corresponding to a value of an integer number of fully occupied Landau levels. Fig. 8 shows high-precision measurements of R_x and R_H close to gate voltages where the lowest Landau level $n = 0$ is occupied. Within the experimental uncertainty of 3 ppm for these measurements, the R_H-curve is flat in the gate voltage region $11.65\ V < V_g < 12.15\ V$ and has a value which agrees with the recommended value [35] $h/4e^2 = (6453.204 \pm 0.005)\ \Omega$. This experimental result is inconsistent with the idea that the Hall resistances R_H should decrease with increasing carrier concentration and therefore with increasing gate voltage V_g [Eq. (4)]

$$R_H \approx \rho_{xy} \sim \frac{1}{n_{inv}} \sim \frac{1}{(V_g - V_{th})} \quad . \quad (17)$$

The unexpected result that R_H remains constant

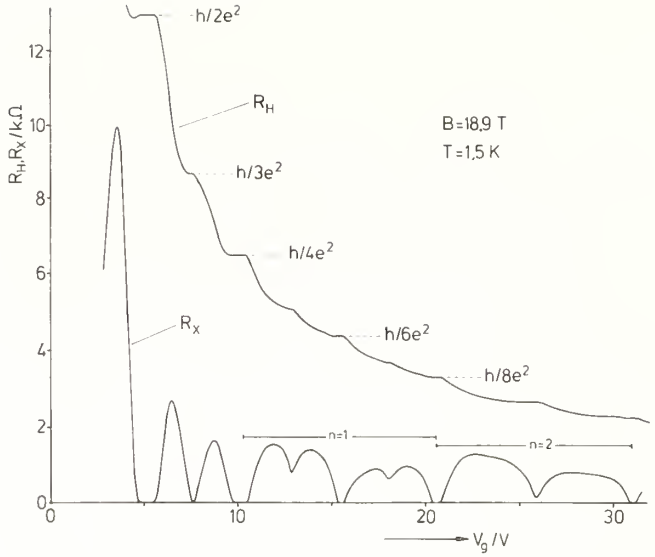


FIGURE 7. *Gate voltage dependence of the measured Hall resistance R_H ($\approx \rho_{xy}$) and resistivity R_x ($\approx \rho_{xx}$) at $B = 18.9$ T for a long silicon MOS device (length to width ratio $L/W = 8$, distance between potential probes $\Delta = L/3$).*

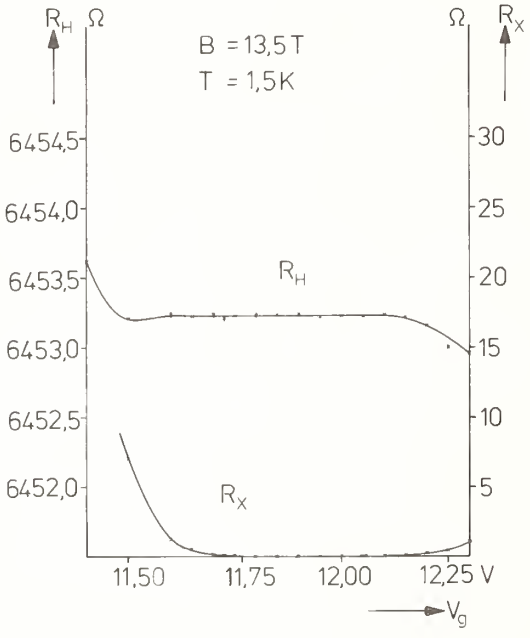


FIGURE 8. *High-resolution measurements of the Hall resistance R_H and the resistivity R_x for a silicon MOSFET at gate voltages close to the occupation of the $n = 0$ Landau level ($B = 13.5$ T, $T = 1.5$ K).*

($\Delta R_H/R_H < 3 \times 10^{-6}$), even if the gate voltage is changed by $\Delta V_g/V_g = 5\%$, can be explained in different ways:

(a) the assumption that the carrier concentration in the inversion layer changes linearly with the gate voltage [Eq. (4)] is not correct. A change in the gate voltage in the region of the plateaus leads to the occupation of states *not* connected with the inversion layer (e.g., states in the depletion layer, in the SiO_2 etc.). This process seems to be important in GaAs-$Al_xGa_{1-x}As$ heterostructures [31] and will be discussed later.

(b) the carrier concentration in the inversion layer changes proportionally to the gate voltage, but in the region of $\rho_{xx} \sim \sigma_{xx} \approx 0$ localized states in the tails of the Landau levels are occupied. Prange [36] as well as Aoki [37] calculated the contribution of localized states to the ρ_{xy} component and found that localized states do not contribute to σ_{xx} but contribute to ρ_{xy} in such a way that a gate voltage independent value corresponding to a fully occupied Landau level is expected. A summary of the theoretical calculations is shown in Fig. 9. This result agrees very well with the experimental data but is inconsistent with a former publication [38] which stated that the Hall conductivity is determined by the number of *mobile* carriers below the Fermi energy E_F.

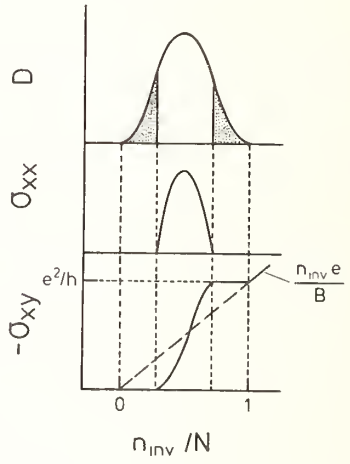


FIGURE 9. *The density of states D, dc conductivity σ_{xx}, and the calculated Hall conductivity σ_{xy} [37] as a function of the normalized inversion carrier concentration n_{inv}/N.*

Up to now it was not possible to decide which one of these two explanations applies to our devices. All theoretical calculations published after the discovery of the quantized Hall resistance [36, 37, 39, 40] show that independent of the origin of the plateaus, the correct value $\rho_{xy} = h/e^2 i$ is expected as long as the condition $\rho_{xx} = \sigma_{xx} = 0$ is fulfilled.

Since σ_{xx} is never exactly zero, the most important correction to the Hall resistance seems to be the finite scattering time τ ($\sigma_{xx} \neq 0$). The value of the conductivity σ_{xx} (or resistivity ρ_{xx}) at gate voltages where n_{inv} corresponds to completely filled Landau levels can be increased by increasing the temperature or decreasing the magnetic field. Figure 10 shows measurements on the same sample as shown in Fig. 8, but at a lower magnetic field and a higher temperature. The plateau in R_H disappears, but the value of R_H and $V_g = V_g(\rho_{xx}^{min})$ remains unchanged (uncertainty $\pm$ 0.05 Ω), even if the resistivity is increased by more than a factor of 100. If we assume that relative to the measurements shown in Fig. 8 minima in ρ_{xx} are observed at gate voltages corresponding to $n_{inv} = iN$ (which is not correct if an overlap between adjacent Landau levels with different lineshapes is present), and the correction to the Hall resistivity $\Delta\rho_{xy}$ is proportional to ρ_{xx}^{min}, we can estimate $\Delta\rho_{xy} < 0.1\ \rho_{xy}^{min}$. Therefore, the correction to R_H at the minimum of R_x in Fig. 8 should be smaller than 10^{-7}. This correction is still smaller if, as assumed by Laughlin [39], ΔR_H changes

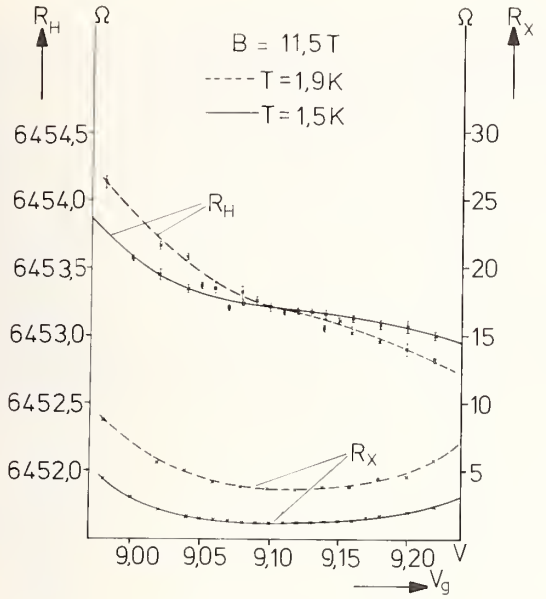


FIGURE 10. *Measurements of the Hall resistance R_H and the resistivity R_x for the same sample as used in Fig. 8 but for a lower magnetic field (B = 11.5 T) and a higher temperature T = 1.9 K.*

with the square of the resistivity R_x. Our high-precision measurements on two devices at a fixed gate voltage in the middle of the plateau between the Landau levels $n = 0$ and $n = 1$ yield a value $\rho_{xy} = h/4e^2 = (6453.198 \pm 0.009)\ \Omega$ [33]. This value should be corrected by about -0.6 ppm due to the finite input resistance of our voltmeter [46].

The comparison between Fig. 8 and Fig. 10 demonstrates that a strong magnetic field is necessary for high-precision measurements of the quantized Hall resistance. Figure 11 shows a series of measurements at different magnetic fields B. With increasing B the plateau shifts to higher gate voltages and the width ΔV_g of the "localized region" increases drastically. ΔV_g seems to be proportional to the effective energy gap ΔE between two Landau levels [34], but a quantitative analysis is not possible without a clear definition of the width ΔV_g. The depth of the minimum on the left side of the plateau in

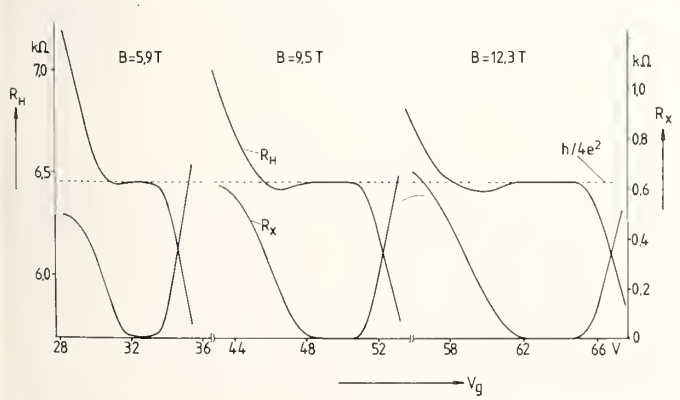


FIGURE 11. *Measurements of the Hall resistance R_H and the resistivity R_x as a function of the gate voltage V_g at different magnetic field values B. The curves are plotted for gate voltages close to the fully occupied Landau level n = 0 (aspect ratio of the MOS device L/W = 1, distance between potential probes Δ = 0.6 L).*

ρ_{xy} depends on the length to width ratio of the device and originates mainly from the shorting of the Hall voltage at the source and drain contact. This correction will be discussed in Section 3.4.

3.3.2 Measurements on GaAs-$Al_xGa_{1-x}As$ Heterostructures

After the observation of the quantized Hall resistance on silicon MOSFETs different groups started to measure this effect on GaAs-$Al_xGa_{1-x}As$ heterostructures [41, 42, 15], although earlier measurements on this system did not show steps in the Hall resistance [43]. But recent measurements on high-quality samples demonstrate that this system too is suitable for the observation of magneto-quantum effects discussed in Section 3.3.1 for silicon. Figure 12 shows an experimental result of the resistivity $\rho_{xx}(B)$ and the Hall resistance $\rho_{xy}(B)$ of a GaAs-$Al_xGa_{1-x}As$ heterostructure with a carrier concentration of $n_{inv} \approx 5 \times 10^{11}\ cm^{-2}$. Since the carrier concentration is fixed, the population of the Landau levels is changed by varying the magnetic field and therefore the degeneracy $N = eB/h$ of each level. The Hall resistance is not a linearly increasing function as expected from the simple theory of the Hall effect, but shows a step-like behavior. At magnetic field values where ρ_{xx} goes to zero (fully occupied Landau levels), ρ_{xy} becomes magnetic-field independent with values equal to h/e^2i (i = number of filled levels). The origin of the plateaus in ρ_{xy} can be explained, if an exchange of electrons (tunneling) between the 2DEG and the donors in the $Al_xGa_{1-x}As$ is assumed. Baraff [31] calculated under this assumption self-consistently the ratio n_{inv}/B as a function of the magnetic field and found that this ratio remains constant within certain magnetic field regions. This behavior originates from the fact that the Fermi energy E_F (strictly speaking the energy bands relative to E_F) cannot jump from one Landau level to the next lower one, if the magnetic field is raised, but moves monotonically through the energy gap between two Landau levels,

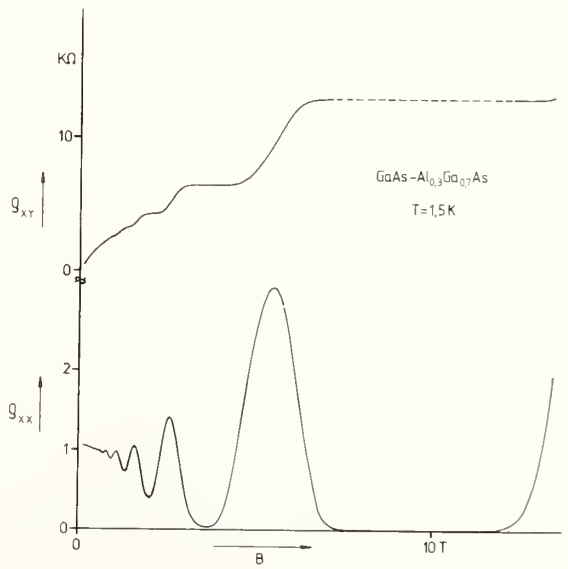


FIGURE 12. *ρ_{xy} and ρ_{xx} measurements on a GaAs-$Al_{0.3}Ga_{0.7}As$ heterostructure as a function of the transverse magnetic field. The inversion carrier concentration is about $5 \times 10^{11} cm^{-2}$. The magnetic field region where a "contact" resistance higher than 1 MΩ is observed, is characterized by the dotted part of the $R_H(B)$-curve.*

because a jump would be connected with an increase in the width of the $Al_xGa_{1-x}As$ depletion layer (= step-like increase in n_{inv}). Therefore, within a certain magnetic field range an integer number i of Landau levels remains completely filled which leads to a constant Hall resistance $R_H = h/e^2 i$ as a function of the magnetic field. This means that the plateaus in ρ_{xy} and ρ_{xx}, which look very similar to the corresponding measurements on silicon MOSFETs, can be explained without assuming a localization of carriers in the tails of Landau levels.

The dotted part of the curve in Fig. 12 characterizes the region where the relatively high "contact" resistance of the sample used in this experiment leads to an error in our potential measurements. Even with an input resistance of the volt-meter of 10^{10} Ω the uncertainty in our measurements is larger than 10^{-4} and therefore highly accurate measurements are not possible. The "contact" resistance oscillates with the magnetic field B and has maxima at $R_x(B) = R_x^{min}$. This result demonstrates that systematic errors in high-precision measurements due to "contact" resistances are possible. For most of our silicon MOSFETs the contact resistance is negligibly small (< 100 Ω).

3.4 Corrections Due to Sample Geometry and Problems in Accurate Hall Resistance Measurements

Generally the measured Hall resistance R_H^{exp} is always smaller than the theoretical value $R_H^{theor} = \rho_{xy}$ expected for an infinitely long sample:

$$R_H^{exp} = G\ R_H^{theor} \qquad G \leqslant 1\ . \tag{18}$$

The factor G, which depends on the Hall angle Θ ($\tan\Theta = \sigma_{xy}/\sigma_{xx}$) and the aspect ratio L/W, has been calculated using the conformal mapping published by Lippmann and Kuhrt [44]. This result agrees with recent calculations of Girvin and Rendell [45]. Figure 13 shows results for the correction $1 - G$ for different values of L/W. In these calculations the position ℓ of the Hall electrodes is exactly in the middle between source and drain ($\ell/L = 1/2$).

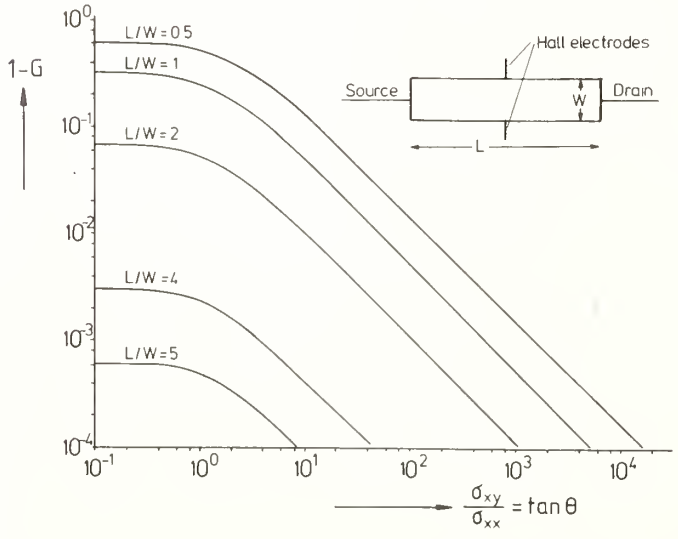


FIGURE 13. *Calculations of the correction term G in Hall resistance measurements due to the finite length to width ratio L/W of the device (Hall electrodes are assumed to be exactly in the middle between source and drain).*

If the position of the Hall electrodes in asymetric ($\ell/L < 1/2$), the correction $1 - G$ becomes much larger. Fig. 14 shows results for a sample with an aspect ratio $L/W = 5$ and ℓ/L values ranging from $\ell/L = 1/4$ to $\ell/L = 1/13$.

For a strongly asymmetric position of the Hall electrodes ($\frac{\ell}{L} \ll \frac{1}{2}$), the correction $1 - G'$ can be found from Fig. 13 using an aspect ratio $2\ell/W$:

$$1 - G'\left(\frac{L}{W}\ ,\ \frac{\ell}{L}\right) = \frac{1}{2}\left[1 - G\left(\frac{2\ell}{W}\ ,\ \frac{\ell}{L} = \frac{1}{2}\right)\right]. \tag{19}$$

This correction is especially important outside the plateaus as may be seen in Fig. 15. The reduction of the Hall resistance relative to the ρ_{xy} value originates from the shorting of the Hall voltage at the source and drain

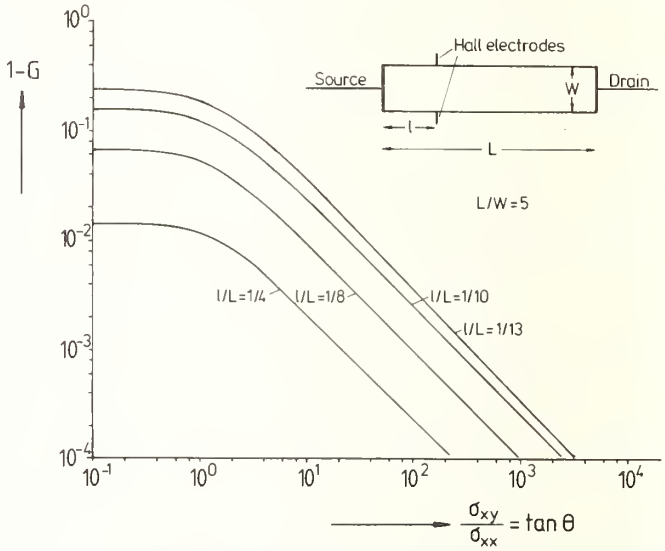


FIGURE 14. *Calculations of the correction term G in Hall resistance measurements for a device with an aspect ratio L/W = 5 and different positions ℓ/L for the Hall electrodes.*

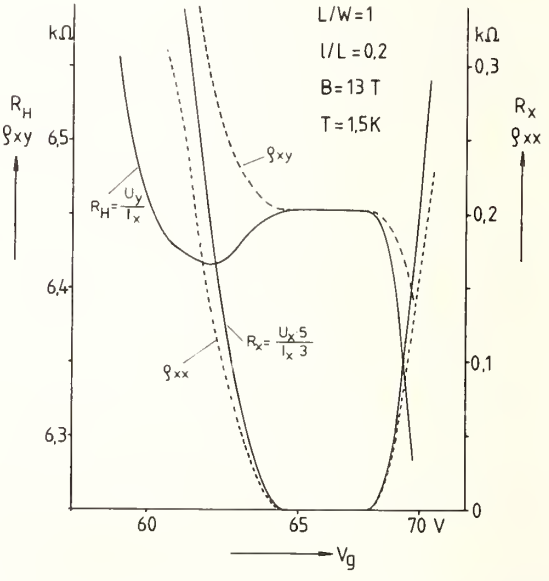


FIGURE 15. *Comparison between the measured quantities R_H and R_x and the corresponding resistivity components ρ_{xy} and ρ_{xx}, respectively. The silicon MOS device has a length to width ratio L/W = 1 and the positions of the Hall electrodes (potential probes) are ℓ/L = 0.2 (see Fig. 14).*

contact. However, for a Hall angle of 90° the influence of the contacts on ρ_{xy} and ρ_{xx} measurements is zero. For typical minimal values of $\rho_{xx}^{min}/\rho_{xy} = 10^{-6}$ and aspect ratios of $L/W > 3$ the correction should be smaller than 10^{-7}. This correction is unimportant because, at present, the uncertainty in the value of the quantized Hall resistance is determined by experimental uncertainties and we would like to summarize some problems which arise in measurements of the quantized Hall resistance.

(a) The Hall voltage is relatively small because the upper limit for the current through the sample (typically 10 μA for a width of the device of 200 μm) is limited by heating effects of the electron gas. Since the current lines for a long device in strong magnetic fields converge at corners of the device, the upper limit for the current is not a simple function of the width of the sample. Moreover, the inhomogeneous heating of the electron gas close to the source and drain contacts leads to an inhomogeneous conductivity within the surface channel. Under this condition the correction G (Fig. 13, Fig. 14) which can be calculated only if the conductivity is spatially constant, is inaccurate.

(b) MOS devices are extremely sensitive to high frequency pick-up. Rectifying metal-semiconductor contacts may lead to dc voltages even without an applied source-drain voltage. The pick-up of spikes may lead to an injection of charges in the oxide or the substrate connected with a shift in the threshold voltage V_{th} (relaxation time up to some hours).

(c) In order to eliminate the influence of the thermoelectric power, the difference in the Hall voltage for different current directions is usually measured. However, a change in the current direction leads automatically to a change of the potential within the surface layer relative to the source. This means that the gate voltage and the substrate bias voltage relative to the surface channel changes because these voltages are applied relative to the source contact. For MOS devices with a thick oxide and weakly doped substrate material, the influence of the voltage drop across the device becomes less important.

In spite of the problems discussed above we believe that measurement of the quantized Hall resistance using cryogenic current comparators [46, 47] should allow one to determine h/e^2 with an uncertainty of about 10^{-7}.

4. Conclusion

Measurements of the Hall resistance R_H of a two-dimensional electron gas, realized with a silicon MOS field effect transistor or a GaAs-Al_xGa_{1-x}As heterostructure, demonstrate that under the condition of fully occupied Landau levels R_H is (within our uncertainty of less than 10^{-5}) independent of external parameters. We believe that the value of R_H is given by the equation $R_H = h/e^2 i$ (i = number of fully occupied energy levels) as expected for an ideal two-dimensional electron gas, since all known corrections (finite scattering rate, interaction between Landau levels, shorting of the Hall voltage at the contacts) should change with magnetic field, temperature or sample geometry, which is not observed. The silicon MOSFETs have the advantage that the channel conductivity can be changed and becomes zero for negative gate voltages, which allows determining the leakage current of the experimental set-up. On the other hand, GaAs-Al_xGa_{1-x}As heterostructures can be used for measurements of the quantized Hall resistance at magnetic fields as low as 4 T and at temperatures of $T = 4.2$ K, which is much more pleasant than a magnetic field of 14 T and a temperature of 1.5 K necessary for silicon MOSFETs. If we can establish that the value of the quantized Hall resistance measured on GaAs-Al_xGa_{1-x}As heterostructures is independent of device parameters (no leakage currents, no contact resistances), this system seems to be (at present) the most useful one for the practical application as a standard resistor or for the determination of h/e^2.

For our experiments on silicon MOSFETs we used a large number of different devices obtained from different sources. We would like to thank M. Pepper (Cavendish Laboratory, Cambridge and Plessey Company), G. Dorda (Siemens Forschungslaboratorien, München), D. Tsui (Bell Laboratories, Murray Hill), and R. J. Wagner (Naval Research Laboratory, Washington) for providing us with samples. The high-precision measurements were not possible without the help of the PTB, Braunschweig, and we would like to thank E. Braun, V. Kose and F. Melchert for their assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 128) and the Bundesministerium für Forschung und Technologie of the Federal Republic of Germany.

References

[1] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).
[2] R. D. Cutkosky, IEEE Trans. Instrum. Meas. IM-23, 305 (1974).
[3] For a review see:
F. Stern in Crit. Rev. Solid State Sci. 4, 499 (1974);
G. Landwehr, in *Festkörperprobleme* (Advances in Solid State Physics), H. J. Queisser, Ed. (Pergamon-Vieweg, Braunschweig, 1975), Vol. XV, p. 49;
Conference Proceedings, Electronic Properties of Two-Dimensional Systems: Surf. Sci. 58, 1–340 (1976); Surf. Sci. 73, 1–571 (1978); Surf. Sci. 98, 1-616 (1980).
[4] F. F. Fang and W. E. Howard, Phys. Rev. Lett. 16, 797 (1966).
[5] B. Vinter, Phys. Rev. 13 4447 (1976); *ibid.*, 15, 3947 (1977).
[6] K. Nakamura, K. Watanabe, and H. Ezawa, Surf. Sci. 73, 258 (1978).
[7] T. Ando, Phys. Rev. B 13, 3468 (1976).
[8] J. F. Koch, in *Festkörperprobleme* (Advances in Solid State Physics), H. J. Queisser, Ed. (Pergamon-Vieweg, Braunschweig, 1975), Vol. XV, p. 79.
[9] T. Ando, Solid State Commun. 21, 133 (1977).
[10] C. C. Hu, J. Pearse, K. M. Cham, and R. G. Wheeler, Proc. Second Int. Conf. on Electronic Properties of Two-Dimensional Systems, Berchtesgaden 1977 (unpublished); see also Surf. Sci. 73, 207 (1978).
[11] D. C. Tsui and G. Kaminsky, Phys. Rev. Lett. 35, 1468 (1975).
[12] W. E. Howard and F. F. Fang, Phys. Rev. B 13, 2519 (1976).
[13] R. Dingle, in *Festkörperprobleme* (Advances in Solid State Physics), H. J. Queisser, Ed. (Pergamon-Vieweg, Braunschweig, 1975), Vol. XV, p.21.
[14] J. L. Störmer, R. Dingle, A. C. Grossard, W. Wiegmann, and R. A. Logan, in *Physics of Semiconductors 1978*, B. L. H. Wilson, Ed. (Institute of Physics, Bristol, No. 43, 1979), p. 557.
[15] S. Narita, S. Takeyama, W. B. Luo, S. Hiyamizu, K. Nanbu, and H. Hashimoto, to be published in Jpn. J. Appl. Phys.
[16] K. Ploog, Crystals 3, 73 (1980).

[17] Th. Englert, K. von Klitzing, R. J. Nicholas, G. Landwehr, G. Dorda, and M. Pepper, Phys. Status Solidi B 99, 237 (1980).
[18] H. Köhler and M. Roos, Phys. Status Solidi B 91, 233 (1979).
[19] R. J. Nicholas, K. von Klitzing, and Th. Englert, Solid State Commun. 34, 51 (1980).
[20] T. Ando and Y. Uemura, J. Phys. Soc. Jpn. 36, 959 (1974).
[21] R. Kubo, S. J. Miyake, and N. Hashitsume, in *Solid State Physics*. Vol. 17, 269 (1965), F. Seitz and D. Turnball, Eds. (Academic Press, New York, 1965).
[22] T. Ando, Y. Matsomuto, Y. Uemura, M. Kobayashi, and K. F. Komatsubara, J. Phys. Soc. Jpn. 32, 859 (1972).
[23] T. Ando, J. Phys. Soc. Jpn. 37, 622 (1974).
[24] R. J. Nicholas, E. Kress-Rogers, F. Kuchar, M. Pepper, J. C. Portal, and R. A. Stradling, Surf. Sci. 98, 283 (1980).
[25] S. Kawaji and J. Wakabayashi, Surf. Sci. 58, 238 (1976).
[26] M. Pepper, Philos. Mag. B 37, 83 (1978).
[27] H. Aoki and H. Kamimura, Solid State Commun. 21, 45 (1978).
[28] H. Aoki, J. Phys. C 11, 3823 (1978).
[29] H. Fukuyama, Solid State Commun. 19, 551 (1976).
[30] M. Tsukada, J. Phys. Soc. Jpn. 42, 391 (1977).
[31] G. A. Baraff and D. C. Tsui, Phys. Rev B 24, 2274 (1981).
[32] D. C. Tsui, Solid State Commun. 21, 675 (1977).
[33] E. Braun, E. Staben, and K. von Klitzing, PTB-Mitteilungen 90, 350 (1980).
[34] Th. Englert and K. von Klitzing, Surf. Sci. 73, 70 (1978).
[35] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[36] R. E. Prange, Phys. Rev. B 23, 4802 (1981).
[37] H. Aoki and T. Ando, Solid State Commun. 38, 1079 (1981).
[38] H. Aoki and H. Kamimura, Solid State Commun. 21, 45 (1978).
[39] R. B. Laughlin, Phys. Rev. B 23, 5632 (1981).
[40] D. C. Tsui and J. S. Allen, Phys. Rev. B 24, 4082 (1981).
[41] D. C. Tsui and A. C. Gossard, Appl. Phys. Lett. 38, 550 (1981).
[42] Th. Englert, private communication.
[43] D. C. Tsui, H. L. Störmer, A. C. Gossard, and W. Wiegmann, Phys. Rev. B 21, 1589 (1980).
[44] H. J. Lippmann and F. Kuhrt, Z. Naturforsch. Teil A: 13, 462 (1958); Z. Naturforsch. Teil A: 13, 474 (1958).
[45] S. M. Girvin and R. W. Rendell, these proceedings.
[46] E. Braun, P. Gutmann, G. Hein, P. Warnecke, S. Xue, and K. von Klitzing, these proceedings.
[47] A. Hartland, these proceedings.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Hall Effect in Silicon MOS Inversion Layers for h/e^2 Determination

C. Yamanouchi, K. Yoshihiro, J. Kinoshita, and K. Inagaki

Device Fundamentals Section, Electrotechnical Laboratory
Sakuramura, Niiharigun, Ibaraki 305 Japan

and

J. Moriyama, S. Baba, and S. Kawaji

Department of Physics, Gakushuin University,
Mejiro, Toshimaku, Tokyo 171 Japan

and

K. Murakami and T. Igarashi

Electrical Standards Section, Electrotechnical Laboratory
Sakuramura, Niiharigun, Ibaraki 305 Japan

and

T. Endo, M. Koyanagi, and A. Nakamura

Quantum Metrology Section, Electrotechnical Laboratory
Sakuramura, Niiharigun, Ibaraki 305 Japan

The quantized Hall conductivity has been measured in n-channel inversion layers of silicon metal-oxide-semiconductor field effect transistor devices at temperatures below 1.4 K using magnetic fields up to 15 T with improved accuracy and resolution. This work aimed to reexamine a new method for h/e^2 determination proposed by von Klitzing, Dorda, and Pepper. The result, which corresponds to $h/4e^2$, and its one standard deviation uncertainty are $-1/\sigma_{xy}$ = (6453.1969±0.0046) Ω_{ETL} or (6453.1969±0.0056) Ω_{ABS}. This result agrees with that obtained by other methods to 1 ppm.

Key words: electric quantum limit; inversion layer; precision measurement; quantized Hall conductivity; silicon MOSFET; strong magnetic field.

1. Introduction

A new method for h/e^2 determination has been proposed by von Klitzing, Dorda, and Pepper [1]. The method consists of precise measurements of the Hall emf and the source to drain current flowing in an n-channel inversion layer on a silicon (100) surface at low temperatures under a strong magnetic field normal to the interface. Appropriate surface carrier concentrations for the measurements are attained by adjusting the gate voltage. At these concentrations, the Fermi level lies halfway between the highest spin state of the Nth Landau state and the lowest spin state of the $(N+1)$th state. The electron system is effectively two-dimensional, degenerate, and fully quantized in orbital motion.

The new method is based on the fact that plateau regions occur in the Hall conductivity σ_{xy} versus gate voltage curves and the values of σ_{xy} at the plateaus are given by

$$\sigma_{xy} = -ie^2/h \tag{1}$$

when $\omega_c\tau \gg 1$ and interactions between different Landau states are absent, and when the Fermi level lies near the tail of Landau states, i.e., $\sigma_{xx} \simeq 0$ [1,2] as predicted theoretically by Ando, Matsumoto, and Uemura [3]. Their theory has shown that the plateau region of σ_{xy} arises from immobile carriers, i.e., immobile holes in the upper edge of the lower Landau states and immobile electrons in the lower edge of the higher Landau states. Here, i is the number of Landau levels below the Fermi level, $\omega_c = eB/m^*$ the cyclotron frequency, and τ the zero field relaxation time.

The experimental results [4] show that the value of the Hall resistance, the ratio of Hall emf to channel current,

and its one standard deviation uncertainty are $-1/\sigma_{xy}(i = 4) = 6453.198 \pm 0.009\ \Omega$, in surprising agreement with the corresponding value $h/4e^2 = 6453.204 \pm 0.005\ \Omega$ based on the recommended value for the fine-structure constant, $1/\alpha = 137.03604 \pm 0.00011$. This fact strongly suggests that the simple relationship Eq. (1) holds almost exactly in the real experiment.

The experimental conditions available are such that $\omega_c\tau \simeq 15$, $\hbar\omega_c \simeq 9$ meV, and the broadening of a Landau subband $\Gamma \simeq 46(B/\mu)^{1/2} = 2 \sim 3$ meV [5,6,7], when the zero field mobility value $\mu \simeq 10^4$ cm^2/volt·sec and $B = 15$ T are used. It has been shown in theoretical studies [8,9] that σ_{xy} should be exactly an integral multiple of $-e^2/h$ at $T = 0$ when the Fermi energy lies in the localized regime and the magnetic field is strong enough. However, for finite magnetic field strengths, no theoretical basis to evaluate the level of precision has been obtained as yet.

The object of the present experiment is to test the validity of Eq. (1) making similar measurements to Ref. [1] by use of different measuring systems with improved accuracy and resolution, in the hope of finding a many body effect which is supposed to quench the effect of interactions involved in the two-dimensional disordered system and/or establishing confidence in the use of the quantized Hall resistance to determine h/e^2.

2. Experimental Procedure

Measurements were made on n-channel (100) inversion layers of silicon MOSFET devices at temperatures below 1.4 K with magnetic fields up to 15 T. The magnetic fields were supplied by an Intermagnetics General Corp. 150M superconducting magnet at the ISSP, University of Tokyo.

The samples used were obtained from A. Yagi of the Research Center, Sony Corporation. The shape and probe arrangement of the samples are shown in Fig. 1. Table 1 lists the samples, their oxide thickness and dimensions, and the maximum value of the Hall mobility μ_H, the peak value in the Hall mobility versus gate voltage curves.

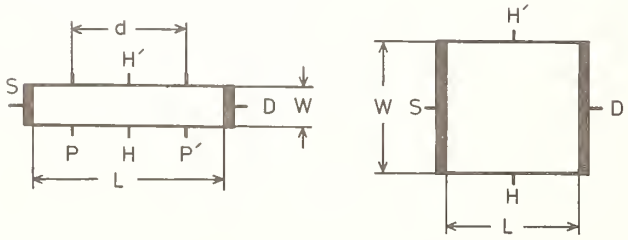


FIGURE 1. *Shape of samples.*

TABLE 1. *Sample geometry and Hall mobility.*

Sample	Oxide thickness (nm)	L (μm)	W (μm)	μ_H max. (cm^2/V·s)
72-17H53-19-L1	200	600	100	13,600
72H57-1-13	200	1600	1600	11,000
62H53-17-1-L1	200	600	100	10,300
62H53-17-1-L2	200	600	100	10,300
44-4-11-L	800	500	100	11,850

The channel current source consists of eleven 1.35 V mercury batteries and a series of step-variable resistors (380 kΩ at maximum current) with two selector switches and a current reversing switch. These switches and emf selector switches are mounted on acrylic plates and operated via teflon rods to minimize the electrical leakage and heat flow from the operator to the contact points. The channel current source and a Yokogawa Electric Works 2781 standard 10 kΩ resistor were immersed in an oil bath to minimize the effect of changes in room temperature. The leads inside the Dewar are Formvar-insulated, 0.16 mm diameter copper wire, and are doubly enclosed in vacuum tight cupronickel cylinders, which extend from 30 cm above the top flange of the Dewar down to just above the sample holder, in order to keep a fixed temperature distribution along the leads to minimize changes in thermal emfs, and also to protect against possible moisture condensation during liquid He transfer. The leakage resistance of the whole circuit was more than $10^{12}\ \Omega$.

The Hall emf V_H and the potential drop V_{std} across a standard 10 kΩ resistor are respectively balanced against a Guildline Instruments Model 9930 Direct Current Comparator Potentiometer to obtain a null reading on a Guildline Instruments Model 9460A Nanovolt Amplifier without feedback and a strip chart recorder. The current sensitivity of the amplifier (galvanometer) and recorder system is $(3 \sim 5) \times 10^{-12}$ A, and the voltage sensitivity of the system is 20 ~ 50 nV for the present measurements.

For each measurement of V_H and V_{std}, eight pairs of readings were taken with opposite directions of the channel current so as to eliminate the thermal emfs, their uniform drifts, and long term drifts of the channel current. A typical run takes about 30 minutes to complete 16 readings. Another typical run consists of 24 readings in which both measurements of V_H and V_{SD} are included, as is shown in Table 2. Here, V_{SD} is the potential drop across the source and drain electrodes.

We refer in this paper to $R_H \equiv R_{std}\ \Sigma V_H/\Sigma_{std}$, $R_{SD} \equiv R_{std}\ \Sigma V_{SD}/\Sigma V_{std}$, and $R_{PP'} \equiv R_{std}\ \Sigma V_{PP'}/\Sigma V_{std} = (d/W)\ \rho_{xx}$, where $V_{PP'}$ is the potential drop between two probes P and P', d the distance between P and P', $(d/W) = 4$ for the samples used, R_{std} is the resistance value of the standard 10 kΩ resistor, and Σ denotes a summation over the corresponding readings in a run.

TABLE 2. *Example of a typical run. Date: 23 March 1981. Sample: 72H57-1-13. B = 14.71 T. T = 1.4 K.*

	Position of emf selector switch	Channel current polarity	emf (×10 nV)	Time (am)	Gate voltage (V)	$T_{\text{std.res.}}$ (°C)	P_{He} (Torr)
						23.9	2.8
	V_G				14.951		
1	std	+	9888253	5:05.0			
2	SD	+	6381046	08.5			
3	H	+	6381046	10.0			
4	H	−	6381008	12.0			
5	SD	−	6381006	15.0			
6	std	−	9888231	17.5			
7	H	−	6381009	5:19.0			
8	SD	−	6381001	21.5			
9	std	−	9888229	25.0			
10	std	+	9888237	27.5			
11	SD	+	6381035	29.5			
12	H	+	6381036	31.5			
13	std	−	9888225	5:34.0			
14	SD	−	6381001	36.5			
15	H	−	6381001	38.0			
16	H	+	6381035	40.5			
17	SD	+	6381033	43.5			
18	std	+	9888226	45.0			
19	H	+	6381032	5:48.0			
20	SD	+	6381037	50.5			
21	std	+	9888231	52.5			
22	std	−	9888224	54.5			
23	SD	−	6381001	56.5			
24	H	−	6381000	58.5			
	V_G				14.945		
						23.9	2.3

3. Experimental Results and Assignment of Uncertainties

It has been found during the course of the experiments that the Hall resistance R_H and the source-to-drain resistance R_{SD} are independent of the channel current in the range up to 10 μA at gate voltages where $R_{PP'}$ $(V_G) <$ 0.01 Ω and the lowest four Landau subbands ($N = 0$) would be fully occupied, within the accuracy of the present work. When the current exceeds ~15 μA, $R_{PP'}$ increases abruptly and the Hall resistance begins to deviate from a constant value.

The effect of geometrical configuration and electrodes was also checked. For "wide" samples with $L/W = 0.1$ and $L/W = 1.0$, both R_H and R_{SD} were measured. When $\sigma_{xx} = 0$, the relationship $R_{SD} = R_H = -1/\sigma_{xy}$ should hold if the spread resistance of the source and drain electrodes is not present. In this case R_{SD} can be measured in place of R_H in a bridge configuration, and the samples could be connected in series or parallel (or both) to get various, probably very stable resistance values.

Figure 2 shows the result of measurements for a sample with $L/W = 1.0$ obtained at $B = 14.21$ T, $T = 1.4$ K, and with a channel current of $I_{SD} = 9.46$ μA. No

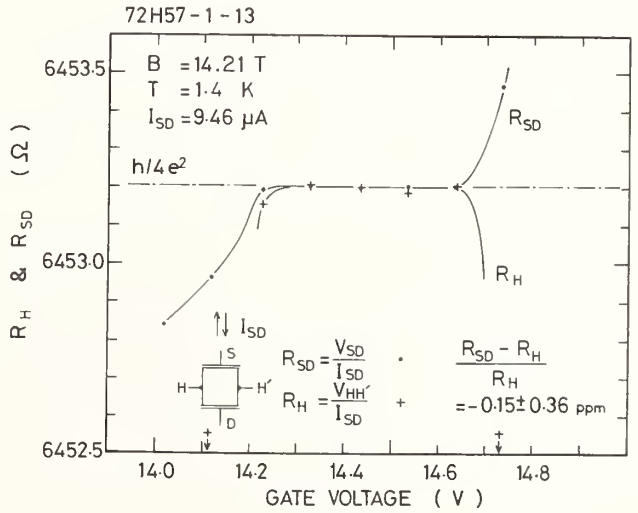


FIGURE 2. *Hall resistance R_H and source to drain resistance R_{SD} of a "wide" sample as a function of gate voltage. At gate voltages between 14.3 V and 14.6 V, the average values of R_H and R_{SD} are almost equal, i.e., $(\overline{R}_{SD} - \overline{R}_H)/\overline{R}_H = -0.15 \pm 0.36$ ppm, which is within the random uncertainty of the present measurement.*

difference is found within the resolution of the present work between these R_H and R_{SD} data in the gate voltage region between 14.3 V and 14.6 V.

For "long" samples with $L/W = 6$, R_H and $R_{PP'}$ were measured. A result obtained at $B = 14.21$ T, $T = 1.4$ K, and with $I_{SD} = 9.46$ μA is shown in Fig. 3. In the gate voltage region between 14.4 V and 14.6 V, R_H passes through a plateau. Little differences are found again between the R_H data of the "long" sample at the plateau and those of the "wide" sample.

In order to ascertain whether the measurement result is independent of the method of measurement, a high input impedance ($>10^{10}$ Ω) digital voltmeter, Datron Electronics Ltd. Model 1071, was also used instead of the Guildline potentiometer and galvanometer system. Results of $R_{PP'}$ measurements obtained by use of the different methods are shown in Fig. 4. It is noted that the direct reading of potentials with a high impedance digital voltmeter could cause large and qualitative errors in the $R_{PP'}$ measurements, supposedly due to a loading effect caused by transient imbalance in the feedback loop of the digital voltmeter. These errors might produce some difficulties in measuring R_H values or in choosing the gate voltage value for R_H measurements.

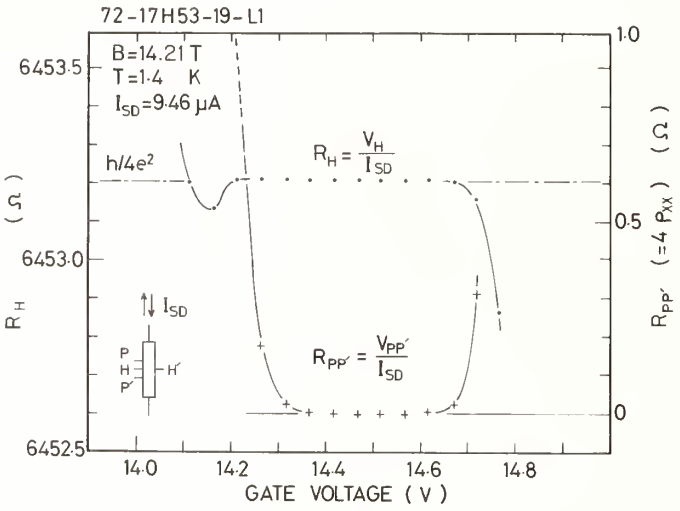


FIGURE 3. *Hall resistance R_H and transverse resistance $R_{PP'}$ of a "long" sample as a function of gate voltage. At gate voltages between 14.4 V and 14.6 V, $R_{PP'}$ passes through a minimum ($\simeq 5 \times 10^{-4}$ Ω) and R_H passes through a plateau of value $R_H = 6453.2024 \pm 0.0008$ Ω. This result was not taken into account in the final result because this measurement was made without the oil bath.*

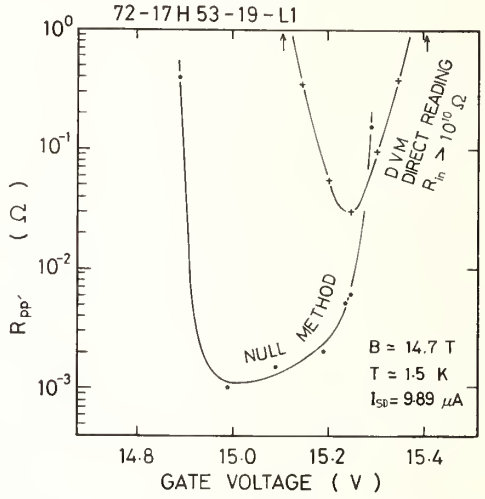


FIGURE 4. *Values of $R_{PP'}$ obtained in different ways are plotted against gate voltage.*

The results of the present measurements are summarized in Table 3, in which average values of R_H and R_{SD} with their respective one standard deviation uncertainties are listed.

The final value of R_H obtained at $B = 14.71$ T and its one standard deviation uncertainty are

$$R_H = (6453.1969 \pm 0.0046)\ \Omega_{ETL} \quad (0.72\ \text{ppm}),$$

or

$$R_H = (6453.1969 \pm 0.0056)\ \Omega_{ABS} \quad (0.88\ \text{ppm}).$$

Table 4 lists the sources of uncertainty associated with the experimental system when it is operated at $I_{SD} \simeq 9.5$ μA and $V_H \simeq 60$ mV, and their one standard deviation estimates.

TABLE 3. *Result of precision measurement.*

Sample	R_H or R_{SD} (Ω_{ETL})	$\sigma_n - 1$ (Ω_{ETL})	$R_{PP'}$ (Ω)	B (T)
72-17H53-19-L1	R_H=6453.1966	±18	0.0015	14.71
62H53-17-1-L1	R_H=6453.1969	±23	0.0005	14.71
62H53-17-1-L2	R_H=6453.1982	±26	0.0056	14.71
72H57-1-13	R_H=6453.1958	±15		14.71
	R_{SD}=6453.1948	±18		14.71
44-4-11-L	R_H=6453.1974	±31		10.50
	$\overline{R}_H$=6453.1969	±10 Ω_{ETL}	Fluctuation of Mean Values	
		±20	Random Uncertainty	
		±41	Systematic Uncertainty Estimated from Table 4	
	$\overline{R}_H$=6453.1969	±0.0046 Ω_{ETL}	Final Result	

TABLE 4. *Sources of Uncertainty.*

Sources of Uncertainty	One Standard Deviation Uncertainty in ppm
Random Uncertainty of R_H around the Mean of Each Sample	0.31
Fluctuation of the Mean Values of Each Sample (Possible sample dependent uncertainty)	0.15
Root Sum Square	0.34
Sources of Systematic Uncertainty	
(a) Thermal emfs variation Less than 50 nV/2 hr and almost linear function of time. Measuring sequence can cancel the effect if linear. A single run takes ~ 30 min.	0.1
(b) Variation of measuring current (1) Potentiometer current (from spec.) (2) Mercury battery current Caused by emf change due to variation of oil bath temperature, less than 0.1°C/2 hr and almost linear in time. Measuring sequence can cancel the effect if linear.	0.1
(c) Errors in potentiometer (1) Linearity of dials (from spec.) (2) Power coefficient of emf generating resistors -0.1 ± 0.05 ppm for comparing 10 kΩ and 6.5 kΩ in 0.2 V range.	0.05 0.05
(d) Insulation Normally greater than 5×10^{12} Ω.	0.05
(e) Standard resistor value (1) Temperature uncertainty (2) Aging and effect of transportation -1.5 ppm/year $\pm$ 0.6 ppm. Compared to the same type resistor kept at rest at 20°C. (3) Calibration including $\Omega_{ETI}/\Omega_{ABS}$	0.1 0.6 0.5
Root Sum Square except e-3	0.64
Root Sum Square	0.81
RSS (Random + Systematic except e-3)	0.72
RSS (Random + Systematic)	0.88

4. Discussion

4.1 Self Test of the Relation $-\sigma_{xy}^{-1}(i) = h/ie^2$

As is seen in Table 3 and Table 4, the difference in R_H values corresponding to $h/4e^2$ between the measured samples is of no statistical significance within the resolution of the present work. However the uncertainty associated with the standard resistor value makes it difficult to test the relation up to the order of, for example, 0.1 ppm or higher in precision, and to check whether many body effects determine the precision of the relation. A relief to this situation would be obtained by measuring $\sigma_{xy}^{-1}(8)$ and/or $\sigma_{xy}^{-1}(2)$ as well as $\sigma_{xy}^{-1}(4)$ on the same sample. If the relation $\sigma_{xy}^{-1}(2) = \sigma_{xy}^{-1}(4) \times 2 = \sigma_{xy}^{-1}(8) \times 4$ is verified to hold experimentally up to the highest precision, the many body effect on the unit of quantization of Hall conductivity in Si-MOS systems would be ruled out, at least up to the order of the precision, since many body effects depend primarily on the carrier concentration.

A preliminary measurement with a magnetic field of 14.7 T on sample 72-17H53-19-L1 reveals that $2\sigma_{xy}^{-1}(8)/\sigma_{xy}^{-1}(4) = 0.9999984 \pm 0.0000010$ at a channel current of 2.46 μA [10]. However, higher magnetic fields should be used to get higher precision since the plateau region of R_H was narrow and the highest limiting current was low at $B = 14.7$ T. This kind of measurement will provide a method of self-test of the quantized Hall conductivity phenomenon independent of the calibration accuracy of the standard resistor.

4.2 Experiment With Different Materials

In the case of e/h determination via the ac Josephson effect, an experiment with different materials was essential to establish confidence in the use of the effect. Use of different materials is aimed at changing carrier concentrations and/or screening properties. In the case of MOSFET devices, these changes are realized by simply adjusting the gate voltage and magnetic field. The carrier concentration can be varied at least a factor of four and even a qualitative difference in screening properties is expected to occur by changing the magnetic field: the carriers in the lowest Landau subband, which are considered to be localized, become delocalized at magnetic fields above 25 T [11], for example. This fact suggests that the role of material dependence measurements would be less important to establish confidence in the e^2/h determination by means of the quantized Hall conductivity in two dimensional systems.

The authors would like to express their sincere thanks to Professor S. Tanuma and Dr. I. Oguro of the ISSP, University of Tokyo, for extending to us the use of their high magnetic field facilities, and to Dr. A. Yagi of the Sony Corporation for providing us with samples. In particular, the authors wish to acknowledge the stimulating discussions with Professor T. Kinoshita of Cornell University, Professor T. Ando of Tsukuba University, Professor Y. Uemura of the University of Tokyo, Professor N. Miura of the ISSP, and Dr. S. Iijima of ETL.

References

[1] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).
[2] S. Kawaji and J. Wakabayashi, in *Physics in High Magnetic Fields*, Ed. by S. Chikazumi and N. Miura, Proc. Oji International Seminar, Hakone, 1980 (Springer, Berlin, 1981), p. 284.
[3] T. Ando, Y. Matsumoto, and Y. Uemura, J. Phys. Soc. Jpn. 39, 279 (1975).
[4] E. Braun, E. Staben, and K. v. Klitzing, PTB-Mitteilungen 90, 350 (1980).
[5] T. Ando and Y. Uemura, J. Phys. Soc. Jpn. 36, 959 (1974).
[6] J. Wakabayashi and S. Kawaji, J. Phys. Soc. Jpn. 48, 333 (1980).
[7] S. Kawaji, Surf. Sci. 73, 46 (1978).
[8] H. Aoki and T. Ando, Solid State Commun. 38, 1079 (1981).
[9] R. E. Prange, Phys. Rev. B 23, 4802 (1981).
[10] K. Yoshihiro, J. Kinoshita, K. Inagaki, C. Yamanouchi, J. Moriyama, and S. Kawaji, Surf. Sci. 113, 16 (1982).
[11] N. Miura, Y. Iwasa, T. Itakura, and G. Kido, J. Phys. Soc. Jpn. 36, 1228 (1982).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Cryogenic Method for the Determination of the Fine-Structure Constant by the Quantized Hall Resistance

E. Braun, P. Gutmann, G. Hein, F. Melchert, P. Warnecke, and S. Q. Xue*

Physikalisch-Technische Bundesanstalt, 3300 Braunschweig, F.R.G.

and

K. v. Klitzing

Physik Department der Technischen Universität München, 8046 Garching bei München, F.R.G.

Recently a determination of the fine-structure constant with a relative uncertainty of 1.3 parts in 10^6 has been carried out. Since then attempts have been made to construct an experimental setup based on a potentiometric method using modern cryogenic measurement techniques, which should lead to a significant decrease of the uncertainty.

Key words: cryogenic current comparator; cryogenics; fine-structure constant; flux transformer; fundamental constants; Hall resistance; SQUID.

1. Introduction

As shown by v. Klitzing *et al.* [1] the Hall voltage U_H of a MOSFET in high magnetic fields and at low temperatures shows steps at which U_H is given by

$$U_H = \frac{1}{i} \frac{h}{e^2} I \qquad (1)$$

$$(i = 1, 2, 3 \ldots),$$

where h is the Planck constant, e the elementary charge, and I the source drain-current. The ratio of the Hall voltage to the current is the so called Hall resistance R_H:

$$R_H = \frac{1}{i} \frac{h}{e^2}. \qquad (2)$$

It has to be pointed out that the Hall resistance at the steps is independent of magnetic field, temperature and any geometric dimension of the sample. Therefore, it is possible to determine the ratio of two fundamental constants h/e^2, by measuring the Hall resistance.

Moreover, by use of

$$\alpha^{-1} = \frac{2}{\mu_o c} \frac{h}{e^2}, \qquad (3)$$

where μ_o is the permeability of vacuum and c the speed of light in vacuum—presently known with a relative uncertainty of 4×10^{-9}—it is possible to determine the fine-structure constant α with about the same uncertainty as h/e^2. It has already been shown [2] that with a rather simple experimental setup it is possible to determine h/e^2 and thus α with a relative uncertainty of about 1 part in 10^6. Since the fine-structure constant is of great importance to quantum electrodynamics and for the realization of electrical units, great efforts are undertaken to decrease its uncertainty. Up to now the uncertainty of other experimental methods for the determination of α is one part in 10^7 while the currently recommended CODATA value which is based on the 1973 least-squares adjustment of the fundamental constants has a relative uncertainty of 8×10^{-7} [3].

One further aspect of interest in the quantized Hall resistance is the fact the h/e^2 has the dimensions of a resistance. One may hope to decrease the experimental uncertainty to a value which allows the use of this new method for the maintenance and the reproducibility of the SI unit of resistance, the ohm, similiar to the use of the Josephson effect in connection with the volt.

In the following we shall at first give a short discussion of the initially used experimental setup and then describe the new method.

2. Method of Voltage Comparison

The MOSFET and a standard resistance of 10 kΩ were connected in series and fed by a current of 5 μA to 10 μA as shown in Fig. 1. The Hall voltage at the MOSFET U_H and the voltage drop at the standard resistance U_N were measured by means of a high impedance digital voltmeter (DVM). The Hall resistance is then given by

$$R_H = \frac{U_H}{U_N} R_N. \qquad (4)$$

In the result given in Ref. [2] a systematic error had been corrected with the wrong sign. The correct results based on the same experimentally obtained data of Ref. [2] is

$$\frac{h}{4e^2} = (6453.194 \pm 0.009)\ \Omega \qquad (5)$$

and

$$\alpha^{-1} = 137.03584 \pm 0.00018. \qquad (6)$$

*On leave from National Institute of Metrology, Peking, People's Republic of China.

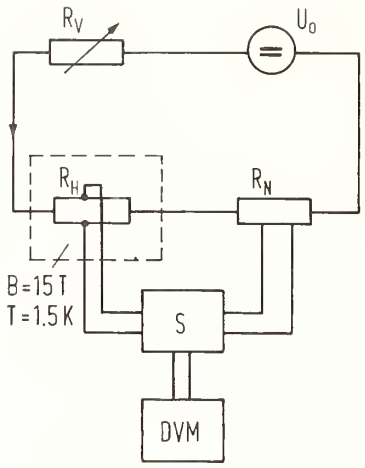


FIGURE 1. *Experimental setup for the method of voltage comparison. R_H is the quantized Hall resistance, R_N the standard resistance. S is a two-way selector switch.*

The uncertainty of this method is limited for various reasons:

(a) the resolution of the voltmeter;
(b) possible deviation from linearity of the voltmeter (for a current of 10 μA, $U_H \approx 64$ mV and $U_N \approx 100$ mV); and
(c) drifts of thermal emf's in between different readings.

The potentiometric method described in the following should overcome these difficulties and should lead to a smaller uncertainty.

3. Cryogenic Potentiometric Method

3.1. Principle

The Hall voltage U_H at the MOSFET is balanced by an equal voltage U_N across a standard resistance R_{NC}. The MOSFET and the standard resistance are in two separate circuits and the ratio of the two currents has then to be known with high accuracy. In order to decrease thermal emf's and thermal noise, most of the components of the two circuits are at liquid helium temperature. The standard resistance for instance will be a cryoresistor made of a copper-phosphor alloy.

The setup is in principle shown in Fig. 2. The ratio of the two currents I_1 and I_2 in the two circuits is kept constant by a cryogenic current comparator (CCC) in connection with a SQUID (SQ1) [4]. Each of the currents passes through the CCC via a coil with N_1 and N_2 turns, respectively, and wound in opposition.

Any deviation of the condition

$$N_1 I_1 - N_2 I_2 = 0 \tag{7}$$

is coarsely controlled through feedback by master (MCS) and slave (SCS) current sources. The fine regulation is automatically done by the SQUID.

3.2 Balancing Condition

Since the two-voltage drops U_H and U_N cannot be expected to be equal at the beginning of the measurement, there has to be the possibility to change I_2 into I'_2 in order to get

$$U_H = 0 \; U'_N = R_{NC} I'_2 . \tag{8}$$

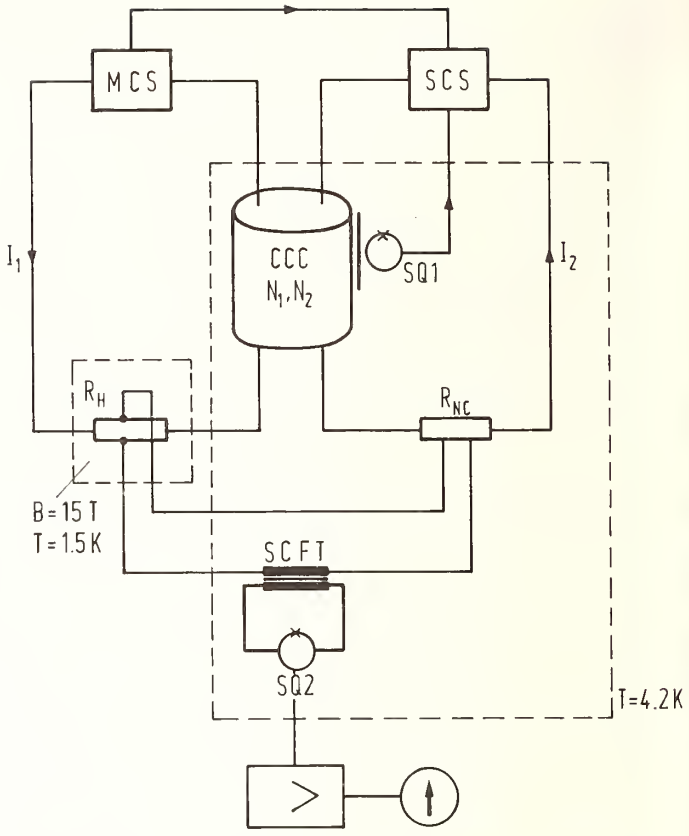


FIGURE 2. *Principle of the cryogenic potentiometric method.*
R_H is the quantized Hall resistance,
R_{NC} the cryogenic standard resistance,
CCC the cryogenic current comparator,
MCS the master current source,
SCS the slave current source,
SCFT the superconducting flux transformer,
SQ1 and SQ2 are SQUIDS.

If for instance R_{NC} is smaller than R_H, I_2 has to be increased for achieving a balance. In order to keep the automatic feedback by the CCC and SQ1 still working, a fraction I_3 of I'_2

$$I_3 = \eta I'_2 , \tag{9}$$

is fed through an additional coil of the CCC with N_3 turns and opposite sense of winding in comparison with N_2 as shown in Fig. 3.

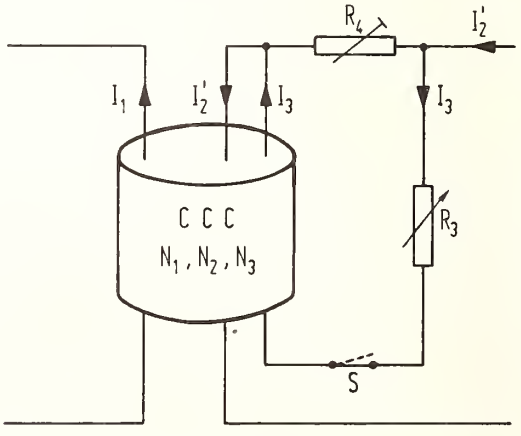


FIGURE 3. *Scheme of the additional circuit to achieve the balance at SQ2 by adjustment of the resistance R_3 and R_4.*

Now, as the switch S is closed, the condition

$$N_1 I_1 - N_2 I_2' + N_3 I_3 = 0 \tag{10}$$

is satisfied by means of the CCC. As seen from Fig. 3, η is given by

$$\eta = \frac{1}{\left[1 + \frac{R_3}{R_4}\right]}. \tag{11}$$

The ratio R_3 to R_4 has now to be adjusted until the balance

$$R_H I_1 = R_{NC} I_2' \tag{12}$$

is achieved. As balance indicator a further SQUID (SQ2) in connection with a newly developed superconducting flux transformer (SCFT) with a 1200 to 1 turns ratio will be used [5]. Together with Eqs. (9) and (10) the unknown Hall resistance is now given by

$$R_H = \frac{N_1}{N_2} R_{NC} \frac{1}{1 - \eta \frac{N_3}{N_2}}. \tag{13}$$

The same setup with either different windings or even another CCC can be used for the comparison of any two other resistances, for instance to compare R_{NC} with a known standard resistance.

By assuming thermal noise from a resistance of about 10 kΩ in the balancing circuit the current resolution of our system is given by 2×10^{-13} A at 1-Hz-bandwith [5]. Measuring at a current of 10 μA means that a determination of the fine-structure constant to a few parts in 10^8 should be possible if R_{NC} is known with a corresponding uncertainty.

We would like to thank Dr. Pepper from Plessey Co. and Dr. Dorda from Siemens A.G., who provided us with samples and all our colleagues who assisted us. Prof. Kose stimulated the work with fruitful discussions.

References

[1] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).
[2] E. Braun, E. Staben, and K. v. Klitzing, PTB-Mitteilungen 90, 350 (1980).
[3] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[4] S. Q. Xue, P. Gutman, and V. Kose, to be published.
[5] S. Q. Xue, P. Gutman, and V. Kose, Rev. Sci. Instrum. 52, 1901 (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Status of the NBS-NRL Determination of the Fine-Structure Constant Using the Quantized Hall Resistance Effect

M. E. Cage, R. F. Dziuba, and B. F. Field

Electrical Measurements and Standards Division, National Bureau of Standards, Washington, DC 20234

and

C. F. Lavine* and R. J. Wagner

Electronics Technology Division, Naval Research Laboratory, Washington, DC 20375

Measurements of quantized Hall steps have been made on (100) Si MOSFET devices using a potentiometric method. In this approach the quantized Hall resistance at a step is compared to a nominally equal room temperature reference resistor. The standard deviation of a single observation is 2 parts in 10^7 at 10 μA source-drain current. The measurement system is described, along with a report on some of the problems encountered.

Key words: fine-structure constant; fundamental constants; inversion layer electrons; precision measurement; quantized Hall resistance; silicon MOSFET; superconducting magnet.

1. Introduction

The paper of von Klitzing *et al.* [1] on precision Hall resistance measurements of metal-oxide-semiconductor field-effect transistors (MOSFETs) suggests that this method can be used to determine the fine-structure constant and to establish an absolute resistance standard. When a positive voltage is applied to the gate electrode of an n-channel MOSFET, a thin layer of electrons (called the inversion layer) is induced beneath the oxide. The layer approximates a two-dimensional electron gas. A magnetic field $B > 10$ T applied perpendicular to the inversion layer constrains the electron motion in the layer to quantized orbits with discrete energy levels (Landau levels). At sufficiently low temperatures, $T < 2$ K, the Fermi level can be adjusted so that a given Landau level is filled while higher levels are empty, so that an inversion layer electron cannot be scattered if the spacing between Landau levels is sufficiently large. Under these ideal conditions, the Hall resistance (the ratio between the Hall voltage and source-to-drain current I_{SD}) is given to high accuracy by

$$R_H = \frac{V_H}{I_{SD}} = \frac{\mu_0 c}{2\alpha i} \approx \frac{25,813}{\mathrm{i}}\ \Omega,$$

where μ_0 is the permeability of vacuum (exactly equal to $4\pi \times 10^{-7}$ H/m), c is the speed of light in vacuum, α is the fine-structure constant, and i is an integer.

Since resistances can be determined in SI units to within a few hundredths of a ppm (part per million) by means of the calculable capacitor, and since the present uncertainty in the speed of light is only 0.004 ppm, measurements of the quantized Hall resistance of MOSFET devices can, in principle, be used to determine the fine-structure constant. Conversely, the effect can be "inverted" and used to establish a known resistance standard. The ultimate goal is to determine the fine-structure constant to an accuracy of a few parts in 10^8 and to develop an absolute, highly reproducible resistance standard based on fundamental constants of nature.

2. Experimental Procedure

Measurements have been made on Si MOSFET devices using the 15 T superconducting magnet at the Naval Research Laboratory's High Magnetic Field Facility. Our measurement approach is indicated in Fig. 1. A constant current is applied to the MOSFET, cooled to 1.5 K, and to a room temperature reference resistor. Reference resistors have been constructed with values nominally equal to that of the Hall resistance, i.e., 6, 453.20 Ω for $i = 4$, and 12, 906.40 Ω for $i = 2$. Therefore $V_H \approx V_R$. The potentiometer voltage is made almost equal to the voltage drop across V_H or V_R, and an electronic detector amplifies the difference voltage signal. The isolated detector output is either displayed on an X-Y recorder or is digitized for further processing by a computer. X-Y recorder tracings of a Hall step are shown in Fig. 2. The

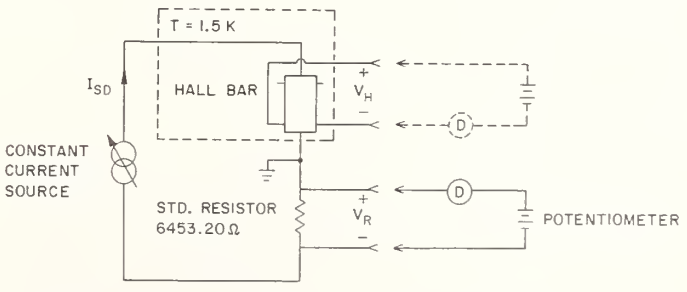


FIGURE 1. *A simplified schematic of the measurement circuit used to compare the Hall Resistance of a MOSFET to a nominally-equal room temperature reference resistor via a potentiometer adjusted to within a few ppm of V_H. The entire system is battery-operated, including the electronic detector, D.*

*Present address: Department of Physics, Saint John's University, Collegeville, MN 56321.

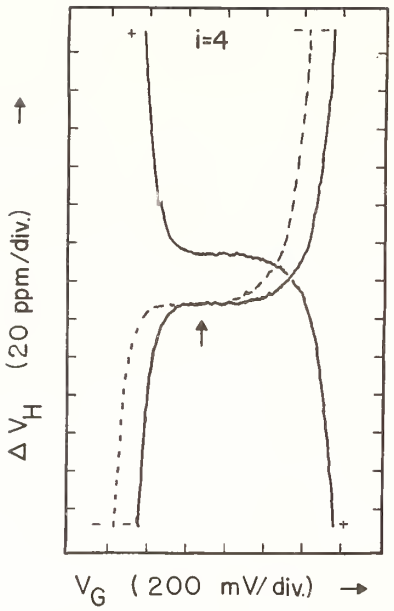


FIGURE 2. *High sensitivity X-Y recordings of an i = 4 Hall step for both current polarities. The dashed line shows the shift in gate voltage for the step upon reversing the current direction instead of interchanging the source-drain connections. This is our only "perfect" step; all other steps have structure on one or both sides. The arrow indicates the gate voltage position at which the voltage drop along the channel, V_x, is a minimum. The Hall voltages are displaced vertically by an arbitrary amount for clarity.*

current source, potentiometer, and electronic detector are all battery-operated.

The potentiometer does not require calibration in this arrangement; it need only remain sufficiently stable during the measurement period. Thermally induced emfs and linear drifts in the current source and potentiometer are cancelled by reversing the current through the MOSFET and reference resistor. A series of reversals in the order $+--+$, $+--+$ is made for each of two measurements of V_R which bracket in time one measurement of V_H. It requires 5 min to complete this sequence in order to obtain a single data point, and typically between 10 and 20 data points are obtained in a single run.

The detected voltage signal has short-term peak-to-peak variations of 2 ppm for a 10 μA source-drain current, and is the same for both V_H and V_R. The standard deviation for a data point is typically 2 parts in 10^7 at 10 μA. A low standard deviation of a single observation is required to study current-dependency effects, the structure of the Hall steps, and the flatness of the steps.

3. Problems Encountered

3.1 Systematic Errors

The measurement system is checked for sources of systematic error by replacing the MOSFET at the end of the probe and the reference resistor with two calibrated 10 kΩ resistors that differ by about 10 ppm. The two resistors are compared using the measurement sequence discussed in the last section. The two resistors are then interchanged and the sequence repeated. Systematic errors are immediately obtained by noting the deviation from the mean value of the two measurement sets. For example, rectification of external high frequency signals by the constant current source (with resonant frequencies at 490 kHz, 4 MHz, and 6 MHz) caused a 2 ppm systematic error. After introducing a radio-frequency interference filter at the current source output, the systematic error dropped to 0.2 ppm. Measured leakage resistances in the probe and the circuit account for this residual error. Obviously, the sources of error depend on the particular measurement system; but it is *vital* that a comparable method for testing systematic errors be incorporated into the measurement procedure.

3.2 Current Reversal in the MOSFET

If the polarity of V_H is changed by reversing the current direction then the negative polarity Hall step occurs at lower gate voltages. This effect is shown by the dashed line of Fig. 2, in which the step is shifted by 0.13 V when I_{SD} = 15.5 μA and V_H = 100 mV. The reason for the shift can be explained by studying the equal-potential lines and electron flow patterns for the two current directions as shown in Figs. 3 and 4 for I_{SD} = ±5 μA. The inversion layer electron potential distribution and the gate voltage are referenced to circuit ground at the source. The average electron potential is +16 mV for +5 μA and −16 mV for −5 μA. Therefore, the gate voltage required to just fill the i = 4 quantum level shifts by 32 mV for I_{SD} = ±5 μA, or by 100 mV for I_{SD} = ±15.5 μA. The observed shift, however, was 130 mV for I_{SD} = ±15.5 μA because the potential distribution also changes relative to the grounded substrate, and this change depends upon the acceptor atom doping concentration.

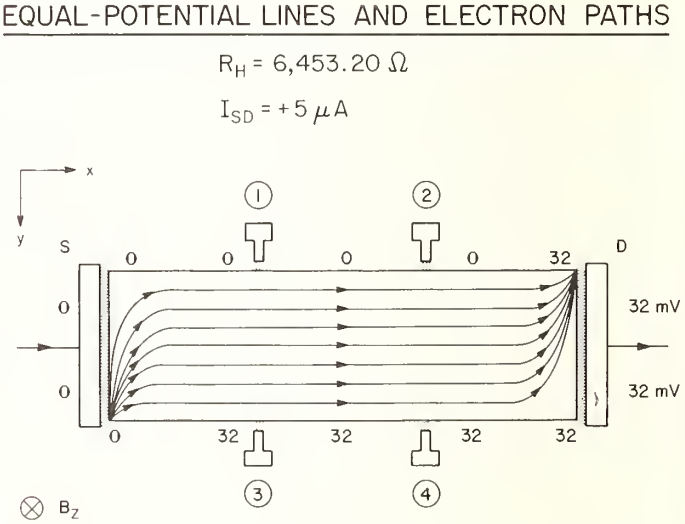


FIGURE 3. *The resulting electron flow pattern and the potential distribution for I_{SD} = +5 μA when the gate voltage is adjusted so that the inversion layer electrons completely fill the four N = 0 Landau quantum states, but none of the N = 1 states.*

This shift is unacceptably large for high currents, but it can be corrected by either changing the gate voltage the appropriate amount on current reversal, or effecting current reversal by interchanging all the connections to the source and the drain. We are using the latter method as demonstrated in Fig. 2. (In the example shown in Fig. 3, the 32 mV and 0 mV values are all interchanged on reversal.)

Although this method has the advantage that the Hall steps "line up" with respect to gate voltage when the current is reversed, it suffers from a lack of cancellation of errors that result from leakage currents between the

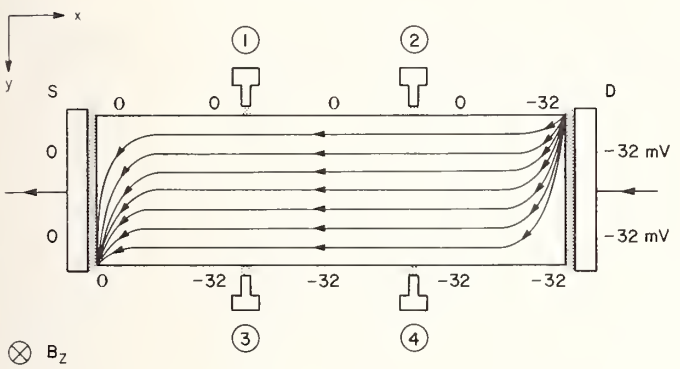


FIGURE 4. *The electron flow pattern and potential distribution when the four $N = 0$ Landau quantum states are just filled and $I_{SD} = -5$ μA.*

gate and the source or drain. In the present apparatus, a $3.3 \times 10^{11}\,\Omega$ leakage resistance exists between the formvar-insulated gate, source, and drain leads. This results in a systematic error of 1 ppm for each 10 V applied to the gate when $I_{SD} = 15\ \mu\text{A}$ ($V_H \approx 100$ mV) and 2 ppm when $I_{SD} = 7.5\ \mu\text{A}$) ($V_H \approx 50$ mV). Because of this error, we have been unable to obtain accurate results with MOSFETs. We are correcting the problem by replacing the source, drain and potential leads with twisted pairs of teflon-coated wire, and the gate and substrate leads with coaxial cable.

4. Conclusions

We have developed a measurement system of sufficient stability and precision to enable us to study, in detail, the structure and flatness of Hall steps. After correcting the leakage resistance problems discussed in the last section, it is expected that measurements of $(R_H)_{SI}$ and α can then be made with a 1 σ total uncertainty of less than 0.2 ppm.

The authors wish to acknowledge the encouragement of B. D. McCombe, NRL, B. N. Taylor, NBS, P. J. Stiles, Brown University, the efforts of the staffs of the Research Device Facility and the High Magnetic Field Facility at NRL, and the financial support of the Office of Naval Research.

References

[1] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Use of a Cryogenic Current Comparator to Determine the Quantized Hall Resistance in a Silicon MOSFET

A. Hartland

Division of Electrical Science, National Physical Laboratory, Teddington, Middlesex, U.K.

A measurement system based on a cryogenic current comparator bridge is described. The system has the capability of being able to determine the ratio of a 6.453 kΩ and 100 Ω resistors, at 295 K, with an uncertainty (1 σ) of 2 parts in 10^7 when the larger resistor is energized with a current of 10 μA. The suggested application of this technique to the determination of the quantized Hall resistance in silicon MOSFETs is described and sources of possible error are assessed.

Key words: cryogenic current comparator; quantized Hall resistance; resistance ratio bridge; silicon MOSFET.

1. Introduction

It has been demonstrated [1] that the Hall resistance, R_H, measured in silicon MOSFETs in high magnetic fields, ($\gtrsim$14 tesla) and low temperatures ($\lesssim$1.5 K) for particular values of the voltage applied to the gate electrode can have the quantized values h/ie^2, where i is an integer, h is Planck's constant, and e is the charge on the electron, e.g., when $i = 4$ then $R_H = 6.45320$ kΩ. R_H has been measured [2] with a claimed uncertainty of 1.4 ppm using standard digital voltmeter (DVM) techniques. In order to assess whether R_H is invariant to geometric, material, or other physical variables associated with 2-dimensional electron systems in semiconductors, and, consequently would be suitable as a resistance standard, it is essential for the measurement system to have the capability of relating R_H to the national standard of resistance (the ohm) with an uncertainty smaller than 1 part in 10^7. This uncertainity is comparable to that achievable [3] using a calculable capacitor in conjunction with a combination of ac and dc ratio bridges.

The basis of the system which has been developed is a cryogenic current comparator [4] used in a circuit to control the ratio of two dc currents, the smaller one of which flows through the MOSFET and the larger through a standard resisitor maintained at a temperature of 295 K. The Hall potential is nulled against the potential difference developed across the standard resistor using a bandwidth limited flux transformer coupled to a SQUID (Superconducting Quantum Interference Device) as a current detector. The metrological problems are presented in the next section, followed by a description of the measurement system and its application to the determination of a 6.453 kΩ resisitor at 295 K substituted for the MOSFET, and finally, an estimate is given of the limiting uncertainties to be expected for a determination of R_H.

2. Metrological Considerations

In order to examine the conditions required for a high precision measurement of R_H, a preliminary study has been undertaken of the characteristics of silicon MOSFETs at a temperature of 1.45 K and in a magnetic field of 7.5 tesla. These devices which are nominally equivalent to some of those employed in the original determination of R_H [1] have conduction channel dimensions 400 μm long by 50 μm wide, and mobilities, μ, ranging from $0.8 - 2.0\ T^{-1}$ ($m^2V^{-1}s^{-1}$) determined in the absence of a magnetic field at the particular value of that gate voltage, V_g, corresponding to the Fermi level lying between the first and second Landau levels. The Hall voltage, V_H, was measured with a Datron 1075 averaging DVM and the source-drain current, I_{SD}, likewise determined by allowing it to flow through a standard resistor in series with the device. Despite the relatively high mobility of the electrons in these MOSFETs it was apparent that for $B_z = 7.5$ T plateaus in the V_H versus V_g plots do not exist and it is not possible to measure a unique value of R_H. The value of R_H ($i = 4$), obtained when R_{xx} ($=V_{xx}/I_{SD}$, where V_{xx} is the voltage measured between electrodes separated by 133 μm along the conducting channel) is a minimum, was within 5 parts in 10^4 of $h/4e^2$, with a repeatability of 3×10^{-5} for $I_{SD} = 1$ μA. This lack of agreement between R_H and $h/4e^2$ is expected, since the maximum value of the ratio (σ_{xy}/σ_{xx}), where σ_{xy} and σ_{xx} are the Hall and transverse conductivities respectively, is 4.7×10^3, which, based on the available theroretical and experimental evidence is an estimated factor of 100 less that that necessary to ensure $R_H = h/ie^2$ without systematic corrections.

There are two distinct relationships:

$$\sigma_{xy} = ie^2/h \tag{1}$$

and

$$R_H = 1/\sigma_{xy}\ , \tag{2}$$

both of which only become exact in the limit $\sigma_{xy}/\sigma_{xx} \to \infty$. The degree to which Eq. (1) is obeyed depends on the totality of quantization of the 2-D electron system, with all mobile electrons confined to an integral number of Landau levels. For a given mobility, as B_z is increased and T reduced, the width of the energy gap [5, 6] between Landau levels increases, thermal excitation of electrons from occupied to unoccupied levels is reduced, and Eq. (1) is expected to become more exact. One of the main objectives for the technique described in this paper is to examine the conditions for the breakdown of Eq. (1).

The classical theory of the Hall effect for electrons in homogeneous, isotropic media in low values of B_z relates the x and y components of the electric fields, $E_{x,y}$ and

current densities, $J_{x,y}$ to σ_{xx} and σ_{xy} as follows:

$$J_x = \sigma_{xx} E_x + \sigma_{xy} E_y , \quad (3)$$

$$J_y = -\sigma_{xy} E_x + \sigma_{xx} E_y . \quad (4)$$

However, when B_z is large there is considerable electric field distortion in the vicinity of the electrodes where the currents are introduced and it is necessary to solve the total field distribution by the techniques of conformal mapping. This has been done [7] and the results, as quoted by Kawaji [8], have been used to derive expressions which relate V_{SD}, the voltage measured between the source and drain electrodes, and V_H, to I_{SD} and I_H, the Hall current, by

$$V_{SD} = Z_{XX} I_{SD} + Z_{XY} I_H \quad (5)$$

$$V_H = -Z_{XY} I_{SD} + Z_{XX} I_H , \quad (6)$$

where the impedance parameters Z_{XX} and Z_{XY} are given by

$$-Z_{XY}/Z_{XX} = h(\sigma_{xy}/\sigma_{xx}) \quad (7)$$

$$1/Z_{XX} = \sigma_{xx}\, g(\sigma_{xy}/\sigma_{xx}) . \quad (8)$$

The quantities h and g are both functions of $(\sigma_{xy}/\sigma_{xx})$, and L/W and P/L, where L and W are the length and width, respectively, of the conducting channel, and P is the potential electrode contact length. If the device is allowed to supply Hall current to an external resistance, R_e, then

$$I_H = -V_H/R_e$$

and

$$R_H = (V_H/S_{SD}) = h/1/(R_e + \sigma_{xx} g) . \quad (9)$$

Consequently, the Hall resistance is a complex function of the device geometry, σ_{xx}, and R_e. However, in the limit $(\sigma_{xy}/\sigma_{xx}) \to \infty$, $h \to 1$ and $g \to \sigma_{xy}/\sigma_{xx}$, and from Eq. (9)

$$R_H \to 1/(1/R_e + \sigma_{xy}) ,$$

i.e., $R_H = 1/\sigma_{xy}$ when $I_H = 0$. However, if Hall currents flow then $1/\sigma_{xy}$ can still be derived from measurements of (V_H/I_{SD}) versus R_e. This has been done for devices at $B_z = 7.5$ T and $T = 1.45$ K and it was found necessary to include an additional device dependent resistance, $2r_p$, ranging in value from 1.7–3.0 kΩ in series with the Hall potential electrodes before agreement with Eq. (9) was achieved. Clearly, $\sigma_{xx} \to 0$ is an essential pre-requisite for an accurate determination of h/ie^2 and from the available experimental evidence this occurs for Si MOSFETs when $B_z \gtrsim 14$ T and $T \lesssim 1.7$ K.

In order to determine V_H by a null technique the equivalent circuit shown in Fig. 1 is appropriate. Direct currents I_{SD} and NI_{SD} are supplied to the MOSFET and a standard resistor, R_s, respectively, such that $N \simeq (h/ie^2)/R_s$. Additional emf's $r_p I_B$ due to a possible leakage current, I_B, flowing through r_p, and, V_{th}, caused by thermal emf's are included since these can cause errors in the intercomparison of V_H and V_{R_s}. The current detector is a bandwidth limited dc flux transformer coupled to a SQUID. The noise sources of emf are also indicated and it has been assumed that the principal noise generator, $\langle v_{n1}^2 \rangle^{1/2}$, is due to the conversion of shot noise, generated by electron injection into the conducting channel, into voltage noise by the Hall effect. The equivalent noise resistance, r_n, has been calculated so that the total circuit noise is equivalent to the Johnson noise generated in r_n at a temperature of 295 K.

When the bridge is balanced the minimum current to be observed for a 1 part in 10^n resolution in V_H is

$$\Delta I_{min} = 10^{-n} (I_{SD}/\sigma_{xy} R) , \quad (10)$$

where $R = (1/\sigma_{xy} + 2r_p + R_s)$. The current noise $\langle i_n^2 \rangle / \sqrt{Hz}$ flowing in the primary winding, inductance L, for $R \gg r_L$ is

$$\langle i_n^2 \rangle^{1/2} = (2/R)(2kT\, r_n/\pi)^{1/2} \quad (11)$$

for a bandwidth $(r_L/4L)$. If G is the current transformation ratio of the flux transformer then for a given bandwidth the optimum value of G is

$$G \langle i_n^2 \rangle^{1/2} = \langle i_{ns}^2 \rangle^{1/2} , \quad (12)$$

where $\langle i_n^2 \rangle^{1/2}$ is the intrinsic SQUID equivalent current noise/$\sqrt{Hz}$ at its input coil. Also the current signal/noise ratio is

$$\Delta I_{min}/\langle i_n^2 \rangle^{1/2} = 10^{-n}(I_{SD}/\sigma_{xy})(\pi/8\, kT\, r_n)^{1/2} \quad (13)$$

in a 1 Hz bandwidth. Since for the model chosen r_n depends mainly on I_{SD}, $\Delta I_{min}/\langle i_n^2 \rangle^{1/2}$ is proportional to $I_{SD}^{1/2}$. Unfortunately, I_{SD} cannot be increased arbitrarily since electron heating causes the effective temperature

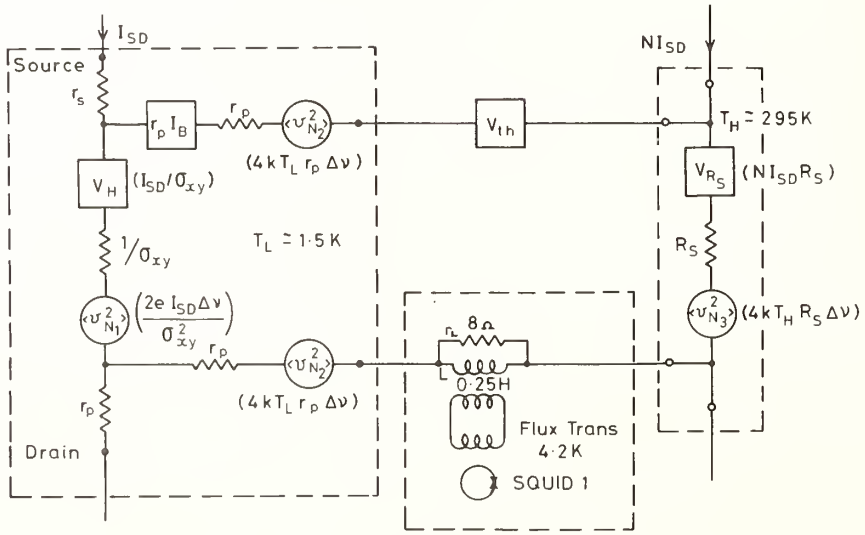


FIGURE 1. *Equivalent circuit for the null detection of V_H.*

to rise resulting in a decrease in the value of $(\sigma_{xy}/\sigma_{xx})$ with a consequent reduction in accuracy in the determination of h/ie^2. In the currently available silicon devices in magnetic fields $B_z \simeq 14$ T serious electron heating occurs when $I_{SD} \gtrsim 10\ \mu A$.

3. Cryogenic Measurement System

An outline circuit diagram of the measurement system is shown in Fig. 2. The primary current, I_{SD}, is generated by an integrated circuit current source (LM334Z) in a temperature controlled oven and has a short term (ten minutes) stability of better than 5 parts in 10^6 at 10 μA. NI_{SD} is generated by the operational amplifier with MOSFET inputs (CA3140) driving a booster amplifier and operating as a voltage follower. The ratio of the two currents is controlled by a cryogenic current comparator in conjunction with SQUID 2 (SHE, System 330) sensing the unbalance signal in the comparator, which is then fed back to stabilize the ratio. The cryogenic current comparator [9, 10, 11] was constructed by making a "rope" consisting of 67 twisted strands of insulated, copper clad, single core, niobium-titanium superconducting wire (Niomax S, IMI Ltd), having an overall diameter of 0.33 mm. The rope was wound onto an insulating former and sixty-six windings of 109 turns and one winding of 58 turns each had their ends separately brought out and labelled. The windings were then covered by 2 1/2 overlapping turns of lead shielding (0.5 mm thick) with PTFE "plumbers" tape used as insulation between the turns. A sense winding of 6 turns of Nb-Ti wire was wrapped around the outer-most layer of the shielding and connected to the input coil of SQUID 2. The whole assembly was then surrounded with lead sheeting. The quality of the comparator was determined by passing 10 mA through the windings joined successively in series opposition in a binary build-up arrangement. Figure 3 shows the result of doing this. The most serious error occurred for the thirty-two winding set (position number 8) which indicated a maximum error of about 5 parts in 10^{10} in the final ratio. The short winding of 58 turns could not be tested, but it is assumed to have coupling errors no worse than any of the other 66 major windings. Sixty-four of the 109 turn windings were then joined in series with the 58 turn winding, producing a comparator with an effective ratio N of 64.5321101. Since the interconnections between windings had non-superconducting joints it is important that there be no leakage between windings greater than 100 kΩ if the effective ratio of the comparator is not to be degraded below the intrinsic 5 parts in 10^{10}. At liquid nitrogen temperature interwinding resistance exceeded 10 MΩ. The SQUID 2 output was fed back to the 100 Ω standard resistor and the N winding set of the comparator via an isolation amplifier which effectively decoupled the bridge circuit from the earth of the SQUID system. SQUID 2, which has an intrinsic equivalent input current noise of about 10 pA/$\sqrt{\text{Hz}}$ was operated in the internal feedback mode and the output gain and bandwidth were set at ×100 and 100 Hz, respectively. The stability of the control circuit, which had a loop-gain of approximately 10^6, is shown in Fig. 4. This demonstrates that for a current $I_{SD} = 10\ \mu A$ the ratio of the two currents drifts by about 2 parts in 10^8 in a period of ten minutes. The drift is probably caused by changes in the temperature of the liquid helium bath giving rise to changes in the SQUID's zero. The input bias current, I_b, of the CA3140 is about 2 pA for a 5 volt supply voltage and it is assumed that this changes by $\Delta I_b \simeq 0.2$ pA as the input voltage is increased from 0 to 0.2 volt (a sufficient voltage to supply NI_{SD} to the standard resistor). Consequently, the change in the error voltage, $r_p \Delta I_b$, as the comparator is energized is expected to be no worse than 0.5 nV or 1 part in 10^8 of V_H. A known and variable fraction of NI_{SD} can be diverted through a single 109 turn winding of the comparator to provide a calibration signal.

The flux transformer coupled to SQUID 1 was constructed in a similar manner to the current comparator with approximately 7000 turns of Nb-Ti wire wound on a nylon former which was then almost completely surrounded (leaving a 1 mm gap) by lead shielding. Six turns of Nb-Ti wire wound on the shielding acted as a secondary winding which was connected to the input of SQUID 1. An 8 Ω resistor (phosphor bronze wire) was connected in parallel with the primary winding (inductance 0.25 H) resulting in an input bandwidth of 8 Hz.

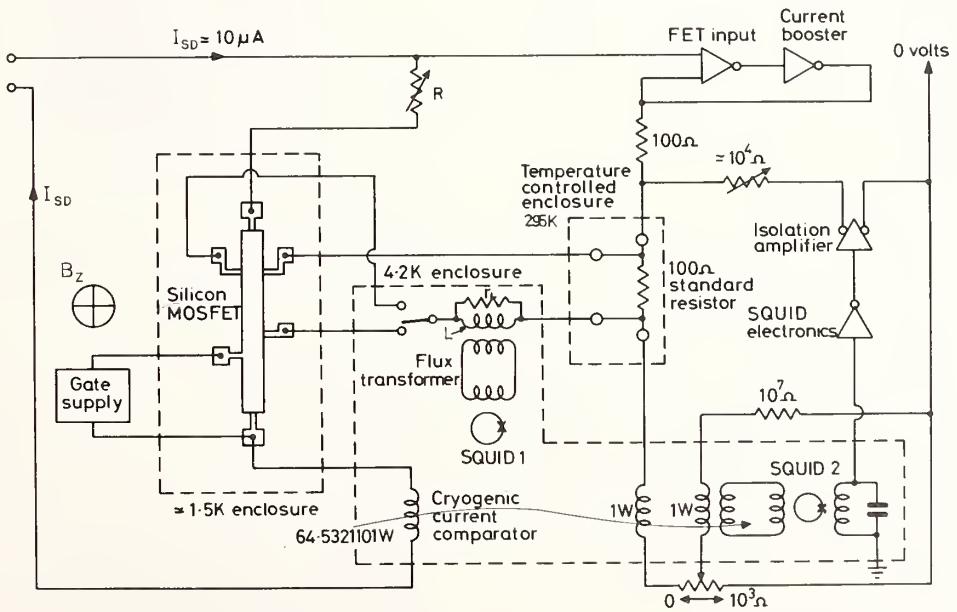


FIGURE 2. *Outline circuit diagram of the cryogenic measurement system connected to the MOSFET.*

The current amplification factor, G, of the transformer-SQUID input coil combination was measured to be 32.5, which is more than adequate to satisfy Eq. (12), thus ensuring that the detection system was not detector-noise limited.

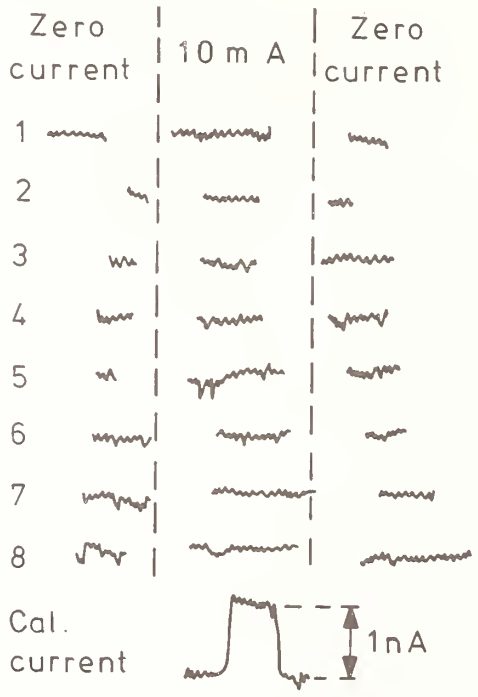


FIGURE 3. *Determination of errors in the cryogenic current comparator ratio. The numbers indicate switch positions for connecting the various windings in series opposition in a binary build-up manner, e.g., position 1 connects a single winding $1W_1$ in oppositon to $1W_2$; position 4 connects $1W_1 + 1W_2$ in opposition to two windings, $2W$, in series; and position 8 connects $1W_1 + 1W_2 + 2W + 4W + 8W + 16W$ in opposition to $32W$.*

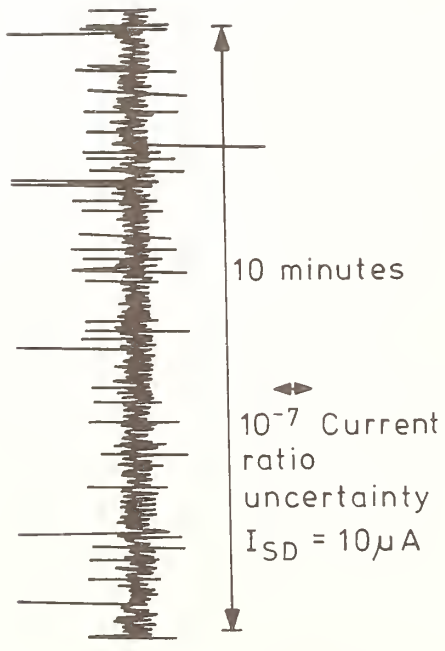


FIGURE 4. *Output of SQUID 2 showing drift in the maintained current ratio, N.*

Figure 5(a) shows an overall view of the cryogenic apparatus, and Fig. 5(b) is an enlarged view of the current comparator and the flux transformer. Great care was taken to use shielded, insulated twisted pairs of wires conveying currents and potential to and from the cryostat. It is particularly important to avoid stray leakage paths for I_{SD} and NI_{SD}.

FIGURE 5(a). *Overall view of cryogenic apparatus. The Dural block at the top is an electromagnetic and thermal shield for the potential leads, and the niobium can surrounding the comparator and flux transformer is a superconducting shield.*

4. Measurement Procedure

Initially, the measurement system has been used to relate the value of a 6.45 kΩ resistor to a 100 Ω standard resistor. These resistors together with two others of 10 Ω and 1 Ω, all of a standard ac/dc construction [12], are maintained at a constant temperature of 295 K in an air enclosure [13] with a long term temperature stability better than 1 mK. Eventually, the intention is to determine R_H in terms of the 1 Ω standard using the bridge described here and, also, a similar bridge having a 10:1 ratio.

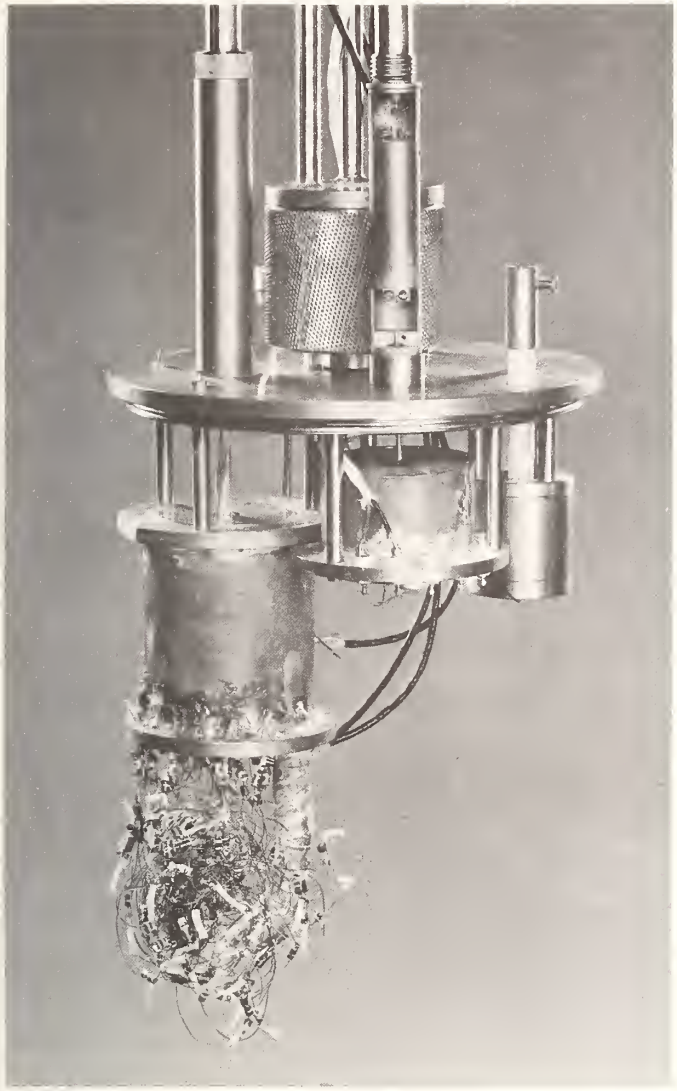


FIGURE 5(b). *Enlarged view of the comparator (on the left), flux transformer, and SQUID sensors.*

The measurement proceeds by closing the switch in the potential circuit and monitoring the drift on a recorder connected to the output of SQUID 1. I_{SD} is then increased from zero to about 10 μA after ensuring the NI_{SD} tracks with it by adjustment of the variable resistance R. If the resistance ratio is not equal to N then a deflection is observed. A calibration signal (1 part in 10^5) is injected into the potential circuit and the further deflection monitored. Finally both calibration signal and I_{SD} are returned to zero. The sequence of events is shown in Fig. 6. The measurement is repeated four times and the 1σ standard error of the mean determined from the chart recordings. The repeatability of the resistance ratio (1σ) is $\pm 2 \times 10^{-7}$ for a current I_{SD} = 10 μA. The equivalent current noise observed with a post detection bandwidth of 1 Hz is in excess of the anticipated circuit noise, Eq. (11), probabily due to instability in the current ratio control circuit and/or fluctuating thermal emf's arising in the potential leads from the liquid helium bath to room temperature. For the eventual measurement of R_H there are two such connection paths, but, unless the thermal emf's have a significant second-order time variation, they are eliminated by the measurement procedure. However, it will probably be essential that themal emf's are kept less than 100 nV. The following is a provisional

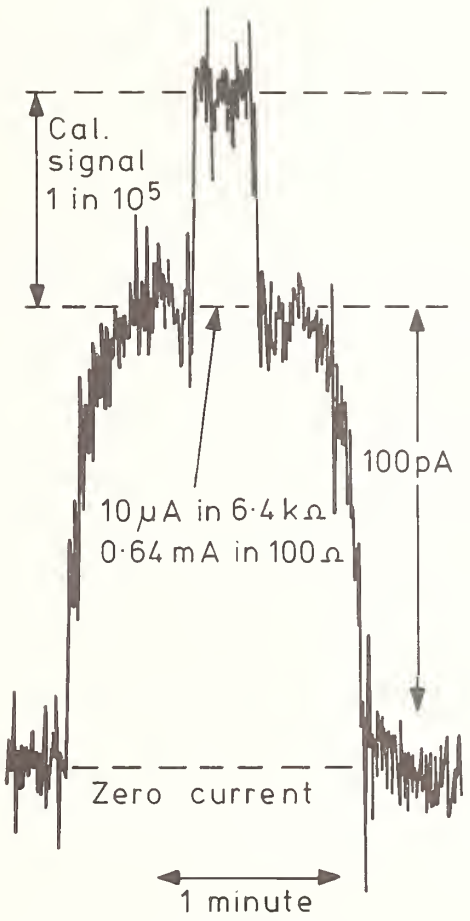


FIGURE 6. *Recorder output of SQUID 1 showing a measurement cycle with the null detector circuit complete.*

estimate of the major error contributions in the determination of R_H in terms of a 1 Ω standard.

Comparator ratio	5×10^{-10}
Current ratio (during measurement)	1×10^{-8}
Emf $\Delta I_b r_p$ due to bias current	8×10^{-8}
Thermal emf	1×10^{-8}
Measurement uncertainty	2×10^{-7}
Determination of 1 ohm: 100 ohms	2×10^{-8}
RSS Total (1σ)	2.2×10^{-7}

5. Conclusions

Preliminary investigations of the circuit parameters in a silicon MOSFET with attached Hall potential probes have been undertaken to assess the problems of determining $R_H(=h/4e^2)$ with an accuracy of the order 1 part in 10^7 in terms of the national standard of resistance. A current ratio bridge, operating at liquid helium temperatures, has been constructed and used to determine the resistance ratio of two resistors of 6.453 kΩ and 100 Ω situated in a 295 K temperature regulated enclosure with an uncertainty of 2.2 parts in 10^7. A major limitation towards reducing this uncertainty is created by the restriction that the source-drain current must be less than about 10 μA. It may be possible to alleviate this by using devices with larger conducting channel areas or by connecting a number of them in parallel. If this approach is satisfactory, then the limiting uncertainty in the measurement will arise from the change in the emf, $r_p \Delta I_b$, in-

duced in the potential circuit as I_{SD} is increased from zero. This could be eliminated, but with some additional experimental complexity, by the complete isolation and individual stabilization of the currents I_{SD} and NI_{SD}.

Thanks are expressed to Dr. M. Pepper of the University of Cambridge for providing the MOSFET devices and to Mr. P. Scrimshaw for technical assistance.

References

[1] K. von Klitzing, G. Dorda, and M. Pepper. Phys. Rev. Lett. 45, 494 (1980).
[2] E. Braun, E. Staben, and K. von Klitzing, PTB-Mitteilungen **90**, 350 (1980).
[3] R. D. Cutkosky, IEEE Trans. Instrum. Meas. **IM-23**, 305 (1974).
[4] I. K. Harvey, Rev. Sci. Instrum. **43**, 1626 (1972).
[5] R. J. Nicholas, R. A. Stradling, and R. J. Tidey, Solid State Commun. **23**, 341 (1977).
[6] Th. Englert and K. von Klitzing, Surf. Sci. **73**, 70 (1978).
[7] R. F. Wick, J. Appl. Phys. **25**, 741 (1954).
[8] S. Kawaji, Surf. Sci. **73**, 46 (1978).
[9] D. B. Sullivan and R. F. Dziuba, IEEE Trans. Instrum. Meas. **IM-23**, 256 (1974).
[10] A. Hartland, in *Euromeas 77* (IEE Conference Publication 152, 1977), p. 10.
[11] F. Delahaye, IEEE Trans. Instrum. Meas. **IM-27**, 426 (1978).
[12] F. J. Wilkins and M. J. Swan, Proc. IEE **117**, 841 (1970).
[13] J. J. Hill, S. Harkness, and J. J. Denton, IEEE Trans. Instrum. Meas. **IM-25**, 501 (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A Resistance Standard Using the Quantized Hall Resistance of GaAs-$Al_xGa_{1-x}As$ Heterostructures

A. C. Gossard and D. C. Tsui

Bell Laboratories, Murray Hill, NJ 07974

We have observed quantization of the Hall resistance of the two-dimensional electron gas in GaAs-$Al_xGa_{1-x}As$ heterostructures. The quantized Hall resistance is given by $\rho_{xy} = h/[2\,e^2(N+1)]$ where N is the quantum number of the Landau level immediately below E_F. Our results show that the quantized ρ_{xy} as given by this relation does not depend on temperature T, magnetic field B, and the electron mobility μ of the sample, which in this experiment is varied from $\mu = 1.8$ m^2/V · s to 6.9 m^2/V · s. In other words, B, T, and μ must suffice to reach the quantum regime, which is evidenced by the vanishing of the parallel resistance ρ_{xx} ($<0.1\,\Omega/\square$), but the value of quantized ρ_{xy} is independent of them. In our samples, this quantum regime is reached for $B \gtrsim 3$ T at $T = 1.2$ K and for $B \gtrsim 4$ T at $T = 4.2$ K. This modestly low magnetic field requirement makes this system a practical primary resistance standard.

Key words: fine-structure constant; GaAs heterostructures; Hall effect; quantized Hall resistance; resistance standard.

1. Introduction

It has been known for several years that at sufficiently high magnetic field ($B \approx 15$ T) and low temperature ($T \approx 1.5$ K), the gate voltage (V_g) dependence of the transport coefficients of the inversion layer in (100) Si-MOSFETS (metal-oxide-semiconductor field effect-transistors) shows anomalous behavior [1]. More specifically, the magneto-resistance (ρ_{xx}) vanishes and the Hall resistance (ρ_{xy}) goes through a plateau in a finite range of V_g, which corresponds to the situation that the Fermi level (E_F) lies between Landau levels.

In a quantizing magnetic field, B, the energy spectrum of a two-dimensional, 2D, electron gas is a series of discrete Landau levels, each having a degeneracy $\beta = \gamma eB/h$. (Here, h/e is the flux quantum and γ stands for the spin- and valley-degeneracies.) Scattering removes the orbital degeneracy and broadens each level into a band of states. The electrical transport properties of the system are determined primarily by the position of the Fermi level E_F in relation to these "Landau subbands" [2]. When a "subband" is completely filled, the presence of an energy gap between the filled and the empty Landau levels inhibits scattering and causes ρ_{xx} to vanish at $T = 0$. The off-diagonal Hall resistance then becomes $\rho_{xy} = B/ne = h/ie^2$, where n is the electron density and i is the number of filled Landau levels [3].

Recently, von Klitzing *et al.* [4] made accurate measurements in the Si-inversion layer and showed that at the plateau

$$\rho_{xy} = \frac{h}{e^2 i} = \alpha^{-1}\mu_0 c/2i \qquad (1)$$

where i is the number of filled Landau levels, α the fine structure constant, μ_0 the vacuum permeability, and c the velocity of light. The width of the Hall resistance plateau is apparently the result of localized states between the Landau levels. They may be due to impurities or surface imperfections.

Von Klitzing *et al.* demonstrated the use of this quantization of the Hall resistance for a high-accuracy determination of α and noted the possible use of it as a resistance standard. However, to reach this quantization regime, in the case of Si inversion layers, requires $B \gtrsim 15$ T and $T \lesssim 2$ K. Such high B and low T requirements will clearly limit its use as a standard resistor. However, $B \sim 8.5$ T at 4.2 K or ~10 T at 2 K can be obtained cheaply with NbTi superconducting magnets. Since this quantization is characteristic of 2D carriers, it should be possible to lower this requirement to $B < 10$ T and $T \sim 4.2$ K, by using the 2D electrons in GaAs-$Al_xGa_{1-x}As$ heterostructures, whose cyclotron mass is approximately three times lighter ($m^* = 0.068\ m_o$ [5]) than that of electrons in the (100) Si inversion layer. We recently prepared modulation-doped GaAs-$Al_xGa_{1-x}As$ heterojunctions in which the scattering rate of the 2D electrons is comparable to that in Si-MOSFETs. We measured ρ_{xx} and ρ_{xy} vs. B and observed quantization of the Hall resistance at $T = 4.2$ K and at B as low as 4.2 T. These observations demonstrate the use of GaAs-$Al_xGa_{1-x}As$ heterostructures as a practical primary resistance standard [6].

2. Sample Preparation

The GaAs-$Al_xGa_{1-x}As$ heterostructures were prepared by using molecular-beam-epitaxy techniques on Cr-doped semi-insulating GaAs substrates of (100) orientation which show no observable conductivity at 4.2 K. Each sample consists of a 1 μm thick undoped GaAs layer, a 0.07 μm thick n-type Si-doped $Al_xGa_{1-x}As$ layer, and a top cap layer of 0.02 μm thick undoped GaAs. The interface between the top GaAs and the $Al_xGa_{1-x}As$ is depleted of free carriers by the surface potential, and the 2D electrons, which result from ionized donors in the $Al_xGa_{1-x}As$ layer, are confined to the interface between the $Al_xGa_{1-x}As$ and the GaAs underneath it. The electrons are confined by the potential well formed by the conduction band discontinuity at the interface ($\Delta E_c \sim 300$ meV) and the band bending in the GaAs [5]. Consequently, the electronic motion perpendicular to the interface is quantized and the electrons, still free to move parallel to the interface, constitute a 2D system.

The specimens were made into standard "Hall bridges", each having six side-arms symmetrically placed to facilitate measurements of magnetoresistance and the Hall effect [7]. The width of the bridges was either 0.1 mm or 0.38 mm and the separation between two neighboring side-arms was 0.3 mm for the narrow ones and 1.0 mm for the wide ones. Electrical contacts were made by alloying indium into the epilayers at 400 °C in hydrogen atmosphere. The quality of the contacts and the homogeneity of the specimens were checked by comparing the I-V characteristics measured across various combinations of the voltage and current contacts.

3. Measurements and Results

Figure 1 shows the magnetic field dependence of ρ_{xx} and ρ_{xy} for a sample with a real electron density of $4.2 \times 10^{11}/\text{cm}^2$ and low-field mobility of 7.9×10^4 $\text{cm}^2/\text{V} \cdot \text{s}$ at 4.2 K. The data were obtained by measuring the voltage between appropriate side-arms with a high input impedance voltmeter and a current source of 10 μA. The

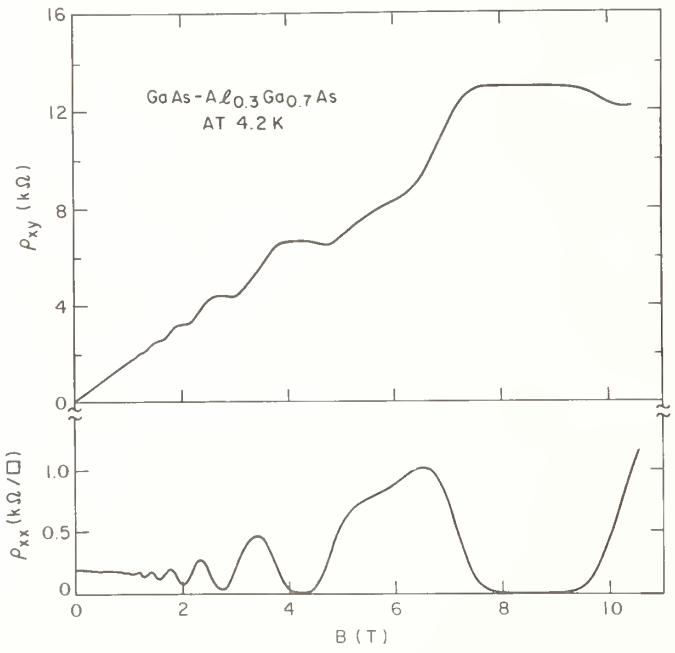


FIGURE 1. ρ_{xy} *and* ρ_{xx} *vs.* B *taken from GaAs-*$Al_xGa_{1-x}As$ *heterostructure at 4.2 K.* B *is perpendicular to the surface.*

magnetic field, which was generated by a superconducting solenoid at 4.2 K and calibrated to better than 3%, was applied perpendicular to the plane of the GaAs-Al_x Ga_{1-x} As interface. The Shubnikov-deHaas oscillations at $B < 4$ T are periodic in $1/B$ and the electron density was deduced from the period [7]. The 2D nature of the electrons is confirmed by the fact that this period depends only on the perpendicular component of B and no oscillatory effect was observed with B applied parallel to the surface. The structure at $B \sim 6$ T depends strongly on T and it develops into two well resolved peaks at $T = 1.2$ K. We attribute it to spin-splitting of the $N = 1$ Landau level. It should be noted that the effective g-factor of electrons in GaAs is $g^* = 0.522$ [8] and the splitting at $B \sim 6$ T is $\Delta E = 0.18$ meV. In view of $kT \sim 0.36$ meV at 4.2 K and $h/\tau = 0.25$ meV, its observation is not expected and exchange enhancement of the splitting of spin-up and spin-down states is appreciable [9, 10].

The most striking feature in Fig. 1 is the vanishing of ρ_{xx} and the constant value of ρ_{xy} (at ~12 kΩ and ~6 kΩ) in a finite range of magnetic field (ΔB) about $B = 8.4$ T and 4.2 T. At these values of B, E_F lies between the $N = 0$ and $N = 1$ and between the $N = 1$ and $N = 2$ Landau levels respectively. As T decreases, ΔB of these two regions increases and at $T \sim 1.2$ K a new region becomes apparent at $B \sim 2.7$ T, corresponding to having E_F between the $N = 2$ and $N = 3$ Landau levels. These data resemble those observed in ρ_{xx} and ρ_{xy} vs. V_g in Si inversion layers and manifest the quantization of Hall resistance of 2D electrical carriers. In this case, the carrier density is fixed and filling of the Landau levels is accomplished by varying B, which varies the orbital degeneracy of the Landau level. Because of a smaller cyclotron mass in GaAs, the quantization is observable at much lower B and higher T in a sample having an electron scattering time ($\tau \sim 2.7 \times 10^{-12}$ s) comparable to that in high mobility Si inversion layers.

Initial measurements were made by connecting a calibrated resistor in series with the sample and comparing the Hall voltage with the voltage developed across it using a null detection scheme. The calibrated resistor was a decade resistor made by General Radio Co. (model no. 1433F), which in the range of resistance of interest here, was accurate to 0.01%. Within this accuracy, the Hall voltage was strictly a linear function of current to $I = 10$ μA, the range checked in this experiment. The Hall resistance at 4.2 T was (6453.4 ± 0.7) Ω and at the 8.4 T plateau was (12907.0 ± 1.3) Ω.

These values of ρ_{xy} in the observed plateau regions agreed within the accuracy of the decade resistor with that given by Eq. (1) [but with $i = 2(N + 1)$], using accepted values for α, μ_0 and c. Measurements on three additional samples showed that, in addition to being independent of B and T, the quantized ρ_{xy} is independent of mobility, which varied from $\mu = 1.8$ $\text{m}^2/\text{V} \cdot \text{s}$ to 6.9 $\text{m}^2/\text{V} \cdot \text{s}$ for the samples used. In other words, B, T, and μ must suffice to reach the quantum regime, which is evidenced by the vanishing of ρ_{xx} (<0.1 Ω), but the value of the quantized ρ_{xy} is independent of B, T, or μ of the sample. Clearly, pathological problems with electrical contacts to the 2D electron layer and inhomogeneities which destroy the 2D character of the sample would affect ρ_{xy}. Our experience indicates that such problems also manifest as differences in voltage drop across equivalent pairs of side-arms, lack of linearity and symmetry in I-V, and irregular structure in ρ_{xx} vs. B. None of these were detected in these samples.

More recently, a preliminary measurement of ρ_{xy} to a much higher precision has been made on one of our modulation-doped GaAs-Al_x Ga_{1-x} As heterostructures in collaboration with M. E. Cage, R. F. Dzuiba, and B. F. Field [11]. This more recent measurement yielded a value of $\rho_{xy} = 6453.2057(10)\,\Omega_{\text{NBS}}$ at the $i = 4$ plateau which would yield a value of the fine structure constant $\alpha^{-1} = 137.035965(12)$. The precision obtained in these measurements seems encouraging for further development of these structures for resistance standards and for measurement of α.

4. Conclusion

In summary, we have observed the quantization of the Hall resistance of GaAs-Al_x Ga_{1-x} As heterostructures at

4.2 K and at B as low as 4.2 T. The B and T requirement is sufficiently simple to warrant its practical use as a primary resistance standard. Multi-layer or superlattice structures which have m identical 2D electron layers electrically connected in parallel might also be used for resistance standards. The quantized Hall resistance would then be given by $\rho_{xy} = h/e^2 im$, and the value of the resistance could be tailored by varying the number of layers. It should also be noted that the quantum regime might be reached with even lower B by using narrow gap semiconductors such as InSb, InAs, PbTe, or $Hg_xCd_{1-x}Te$ if junctions of suitable quality could be formed in such materials.

We thank S. J. Allen, A. Y. Cho, P. M. Platzman, J. M. Rowell, and H. L. Störmer for discussions, and G. Kaminsky and W. Wiegmann for their most valuable technical help.

References

[1] S. Kawaji and J. Wakabayashi, Surf. Sci. 58, 238 (1976).
[2] T. Ando and Y. Uemura, J. Phys. Soc. Jpn. 36, 959 (1974).
[3] T. Ando, Y. Matsumoto, and Y. Uemura, J. Phys. Soc. Jpn. 39, 279 (1975).
[4] K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).
[5] H. L. Störmer, R. Dingle, A. C. Gossard, W. Wiegmann, and M. D. Sturge, Solid State Commun. 29, 705 (1979).
[6] D. C. Tsui and A. C. Gossard, Appl. Phys. Lett. 38, 550 (1981).
[7] D. C. Tsui and R. A. Logan, Appl. Phys. Lett. 35, 99 (1979).
[8] W. Duncan and E. E. Schneider, Phys. Lett. 7, 23 (1963).
[9] T. Ando and Y. Uemura, J. Phys. Soc. Jpn. 37, 1044 (1974).
[10] D. C. Tsui, H. L. Störmer, A. C. Gossard, and W. Wiegmann, Phys. Rev. B 21, 1589 (1980).
[11] D. C. Tsui, A. C. Gossard, B. F. Field, M. E. Cage, and R. F. Dziuba, Phys. Rev. Lett. 48, 3 (1982).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A Quantitative Theory for the Determination of h/e^2 from the Hall Effect in Two-Dimensional Conductors

L. Bliek and G. Hein

Physikalisch-Technische Bundesanstalt, D 3300 Braunschweig, F.R.G.

In experiments on two-dimensional conductors, steps proportional to e^2/h have been observed in the quotient of the electric current and the Hall voltage. They can be explained on the basis of the well-established theory of the Shubnikov-De Haas effect, provided the dependence of the scattering lifetime on the density of states is consistently taken into account. Results of numerical calculations of the Hall voltage and of the electrical resistance agree closely with published experimental data.

Key words: fine-structure constant; h/e^2 determination; quantitative theory of resistance and Hall effect; quantized Hall resistance; two-dimensional conductors.

1. Introduction

Both in the Hall voltage vs. gate voltage dependence of Si MOS-transistors [1, 2] and in the Hall voltage as a function of magnetic induction [3] for GaAs-$Al_xGa_{1-x}As$ heterostructures, steps have been observed experimentally in high magnetic fields and at low temperatures. At the step positions, R_H, i.e., the quotient of the Hall voltage and the current, is given, within the experimental accuracy, by

$$R_H = h/(e^2 j) \tag{1}$$

with $j = 2, 3, 4, 5 \ldots\ldots$. Whenever R_H assumes such a constant value, the samples have no measurable electrical resistance.

If the steps are to be used for a more precise determination of h/e^2 than currently available, theory has to prove that Eq. (1) is sufficiently accurate or even exactly correct, and not just true within the present experimental accuracy. Previous calculations, assuming a uniform, energy-independent scattering time had predicted steps in the number of electrons—plotted against the Fermi energy—corresponding to Eq. (1) with small, but significant, deviations [4]. Somewhat later Ando *et al.* [5], making different assumptions about the electron scattering process, arrived at the conclusion that, to the accuracy of the approximations they made, the Hall conductivity should assume values corresponding exactly to Eq. (1).

In the following, it is shown that the well-established theory of the Shubnikov-De Haas effect, describing the quantum effects in the high-field magnetoresistance, predicts the exact validity of Eq. (1) for $T = 0$, provided the dependence of the scattering lifetime on the density of states is consistently accounted for. With the help of this theory, the influence of a finite temperature can be calculated. Moreover, it gives a good quantitative description of the experimental resistance curves.

2. Fundamental Equations for the Resistance and for the Hall Voltage

Resistance theory provides relations between electric fields and current densities in a conductor. If a magnetic induction B is present, directed in the z-direction, perpendicular to a two-dimensional conductor, there will be two electric fields corresponding to a current density i in the direction of x: E_x and the Hall field E_y. The resistivity tensor is then given by

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} \rho_{xx} & \rho_{yx} \\ \rho_{xy} & \rho_{yy} \end{bmatrix} \cdot \begin{bmatrix} i \\ 0 \end{bmatrix} \tag{2}$$

and consequently

$$\rho_{xx} = E_x/i \tag{3}$$

and

$$\rho_{xy} = E_y/i \ . \tag{4}$$

For the conductivity σ, on the contrary,

$$i/E_x \neq \sigma_{xx} \tag{5}$$

and

$$i/E_y \neq \sigma_{xy} \ . \tag{6}$$

Hence, theoretical relationships between E and i are resistivity formulae rather than conductivity formulae, a fact which is sometimes overlooked.

Keeping this in mind, we can obtain the following two relations from the literature [6]:

$$\rho_{xx} \sim \sum_k \sum_l \int_{-\infty}^{+\infty} A_{k,l} \times D_k(E)\, f(E)\, D_l(E)[1 - f(E)]\, dE \tag{7}$$

$$\rho_{xy} = B/(e\, n_0) \, . \tag{8}$$

Here $f(E)$ is the Fermi function, $A_{k,l}$ is the transition probability from the quantum state k to the quantum state l, where k and l stand for the complete set of quantum numbers designating a magnetic energy level, $D_l(E)$ is the density of states associated with the level l, and n_0 is the electron density. The density of states is concentrated near the quantum energies $E_{n,s,v} = (n + 1/2)\,\hbar eB/m + s\,\Delta_s + v\,\Delta_v$, where $n = 0, 1, 2, 3 \ldots$, $s = +1, -1$, $v = +1, -1$, m is the cyclotron effective mass, Δ_s is the spin splitting, and Δ_v is an additional splitting, observed for Si, but absent in GaAs-$Al_xGa_{1-x}As$. Equation (7) holds for elastic scattering. If the scattering is treated in the Born approximation, one finds $A_{n,s,v,\ n',s',v'} = n + 1/2$. There are correction terms to Eq. (8), but they vanish in the regions of interest be-

cause there the resistivity vanishes. The missing factor in Eq. (7) is found from the condition that ρ_{xx} should reach its zero magnetic field value as $B \rightarrow 0$.

3. Hall Voltage and Resistance in the Absence of Scattering-Induced Broadening of the Quantum States

Without any broadening of the quantum states, one has

$$D_{n,s,v}(E) \sim \delta(E - E_{n,s,v}). \tag{9}$$

For $B \rightarrow 0$, the total density of states has to reach its zero magnetic field value $D_0(E) = 8\pi m/h^2$ which leads to

$$D(E) = (eB/h) \sum_{n,s,v} \delta(E - E_{n,s,v}) \tag{10}$$

and hence

$$n_0 = (eB/h) \sum_{n,s,v} f(E_{n,s,v}). \tag{11}$$

Inserting Eq. (11) into Eq. (8), in the limit $T \rightarrow 0$, leads to Eq. (1), j being equal to the number of states n,s,v, with $E_{n,s,v}$ below the Fermi energy ζ, i.e., $R_H(\zeta)$ shows steps as given by Eq. (1).

Experimentally, one does not measure $R_H(\zeta)$, but rather $R_H(B)$ or $R_H(U_G)$, where U_G is the gate voltage applied to the MOS transistor and is proportional to the total number of induced electrons. If this number were equal to n_0, $R_H(U_G)$ would be a simple hyperbola just as $R_H(B)$ should be a straight line, given by Eq. (8). The experimental observation of steps in both curves shows that, apart from the n_0 mobile electrons considered up to now, there are additional electrons in bound states, giving no contribution to R_H. Therefore, Eq. (11) has to be replaced by

$$n = n_0 + n_b = (eB/h) \sum_{n,s,v} f(E_{n,s,v}) + \int_{-\infty}^{+\infty} D_b(E) f(E) dE, \tag{12}$$

where n is the total electron density, n_b is the density of electrons in bound states, and $D_b(E)$ is the density of bound states. This equation provides for steps in $n_0(n)$ at constant B, as observed experimentally with Si MOS-transistors, and in $n_0(B)$ at constant n, as observed in the GaAs experiments.

It should be noted that in all the experiments ζ remains almost constant, apart from small oscillations around ζ_0, the zero magnetic field value. This is obvious for the GaAs experiments, but it is also true for the Si MOS work because, on changing the gate voltage, the number of electrons is changed, not by changing ζ, but by changing the position of the conduction band edge relative to ζ. Consequently, only the density of states $D_b(E)$ near ζ_0 has an influence on the width of the steps and the number of states actually involved is small. Moreover, it should be noted that on passing several steps, the same bound states are occupied by electrons and emptied again several times.

For non-zero temperatures, the steps in n_0 are not completely flat. If ζ is at a distance δ from the central position between two adjacent energy levels, which are Δ apart (the influence, even of next-nearest neighbors, is completely negligible), n_0 deviates from the ideal value $h/(je^2)$, by an amount $[2h/(je^2)] \exp(-\Delta/2kT) \sinh(\delta/kT)$. For the Si MOS experiments, e.g., a distance of 0.1 times the observed width of the $j = 4$ step corresponds to a relative change of 10^{-6}.

4. Influence of an Energy-Independent Scattering Time

While Eq. (12) closely fits the observed Hall effect data, the underlying Eq. (10), for the density of states, inserted into Eq. (7), does not fit the experimental resistance data at all. Rather than almost δ-shaped peaks, one observes broad maxima. According to the uncertainty principle, a δ-shaped density of states corresponds to an infinite lifetime. Therefore, since the electrons are scattered on the average after a scattering time τ, broadening is not unexpected. In fact, the energy levels should show Lorentzian broadening [7], and Eq. (10) has to be replaced by

$$D(E) = \frac{eB}{h} \sum_{n,s,v} \frac{\hbar/(\pi\tau)}{(E - E_{n,s,v})^2 + \hbar^2/\tau^2} + D_b(E). \tag{13}$$

The broadening parameter τ is normally assumed not to depend on energy [6]. Figures 1 and 2 show examples of the results one then obtains [9]. Compared with the experimental curves, the calculated peaks in the density of states, and the corresponding ones in the resistance, are too narrow, and the values at the minima are far too high. The steps in the electron density, and the corresponding ones in the Hall voltage, appear at the correct positions, but they are not nearly as flat as the experimental ones.

The explanation for the deviations of the calculated results from the experimental ones, lies in the long tails

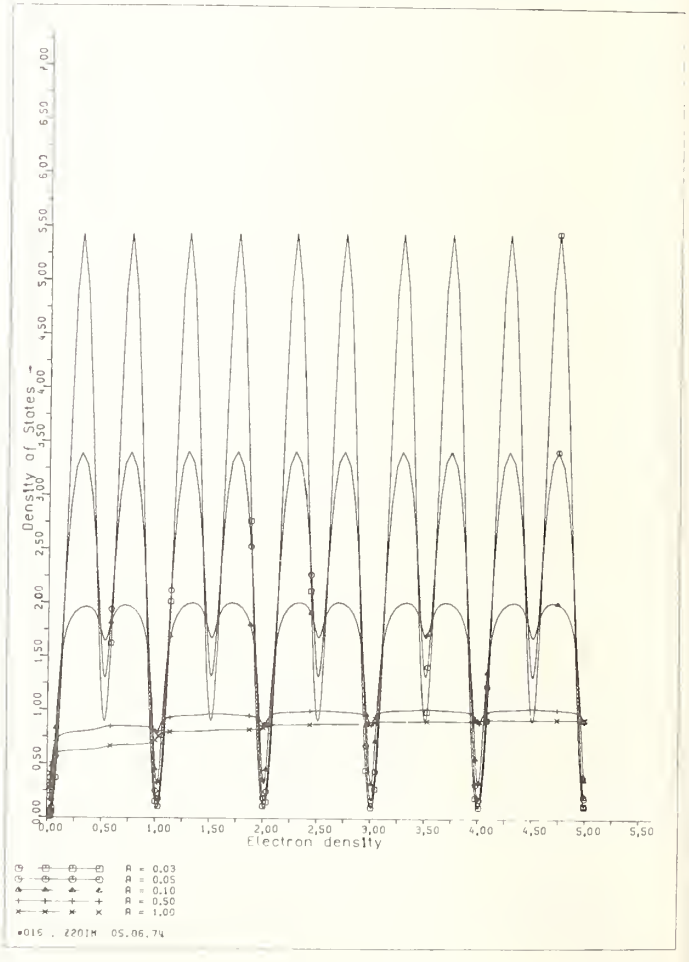


FIGURE 1. *Density of states at the Fermi energy in multiples of $D_0(\zeta)$, as a function of the electron density in multiples of $2eB/h$, for a spin-splitting of $0.2\,\hbar\omega$, $\Delta_v = 0$, and several different constant values for the broadening parameter $A = 1/\omega\tau$.*

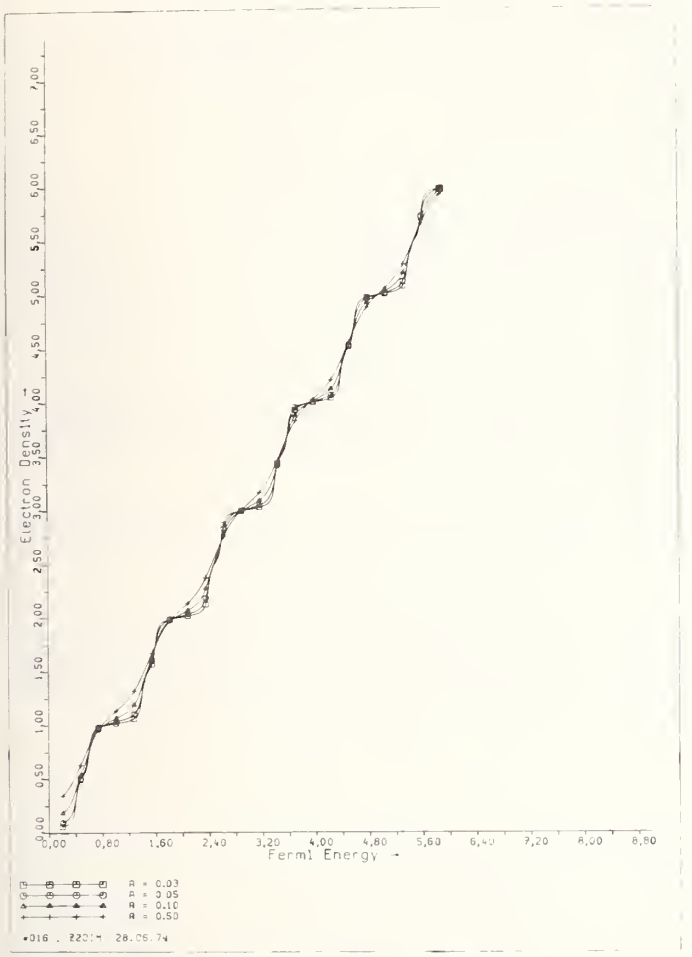


FIGURE 2. *Electron density in multiples of $2eB/h$, as a function of the Fermi energy in multiples of $\hbar\omega$. Parameters as in Fig. 1.*

of Lorentzian functions for constant τ. Such a constant value of τ is not consistent with Eq. (7) which implies that τ^{-1} is proportional to the number of states an electron can be scattered into. Therefore, the assumption of a constant broadening parameter is only justified if $D(\zeta)$ is almost constant. This is true for metals, where it has successfully been used to explain details of the De Haas-Van Alphen effect, but it is not so for the present case and for others where discrepancies have previously been explained using $\hbar/\tau \sim D(\zeta)$ [8].

5. Hall Effect and Resistance with Consistent Assumptions for the Energy Dependence of the Scattering Lifetime

A consistent calculation of the resistance and the Hall voltage has to start from the following equation for the dependence of $\hbar/\tau$ on $D(E)$ which underlies Eq. (7):

$$\frac{\hbar}{\tau} = \frac{2[D(E) - D_b(E)][1 - f(E)]}{D_0(E) - D_b(E)} \frac{\hbar}{\tau_0}, \qquad (14)$$

where $D_0(E)$ and τ_0 are the zero magnetic field values.

Inserting Eq. (14) into Eq. (13), one finds that $D(E) - D_b(E) = 0$ is a possible solution. Neglecting contributions of other states, one finds for the density of states belonging to the energy level $E_{n,s,v}$ at ζ:

$$\frac{D_{n,s,v}(\zeta)}{D_0(\zeta)} = \begin{cases} 0 & \text{for } (\zeta - E_{n,s,v})^2 \geq \dfrac{\hbar^2\omega}{2\pi\tau_0} \\ \sqrt{\dfrac{\omega\tau_0}{2\pi}} \cdot \sqrt{1 - \dfrac{(\zeta - E_{n,s,v})^2}{\hbar^2\omega/(2\pi\tau_0)}} & \text{for } (\zeta - E_{n,s,v})^2 \leq \dfrac{\hbar^2\omega}{2\pi\tau_0} \end{cases} \qquad (15)$$

which is essentially the result of Ando and Uemura [10].

Neglecting the contributions of other states is, however, not justified under the conditions prevailing in the experiments. There the states clearly overlap, and for such a situation Eq. (15) is not applicable. Instead, $D(E)$ has to be calculated numerically, starting from Eq. (13) and Eq. (14), and then has to be inserted into Eq. (7) and, after multiplication with $f(E)$ and integration to give n_0, into Eq. (8). Results, obtained by such numerical calculations are shown in Figs. 3 and 4. The calculated curves agree closely with the experimental ones. Only the amplitudes of the various resistance maxima relative to one another show considerable deviations from experiment. Here, the results might have been improved by allowing for different transition probabilities for different types of scattering processes. The average amplitude of the calculated maxima differs by only 2.2% from the experimental value.

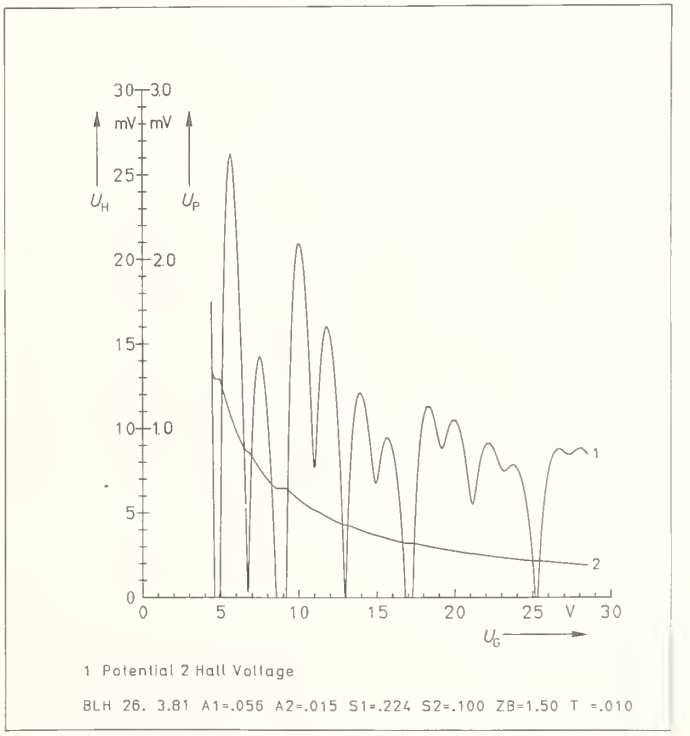


FIGURE 3. *Potential drop in the direction of current flow, U_P, and Hall voltage, U_H, for a Si MOS transistor with a ratio of active length to width of 2.6, for $B = 18$ T and a current of 1 μA, as functions of the gate voltage, U_G. The Hall voltage was calculated for $T = 1.5$ K. Because of the small influence of temperature near 1.5 K, the resistance curve was calculated for $T = 0$. With this exception, the parameters have the experimental values [1].*

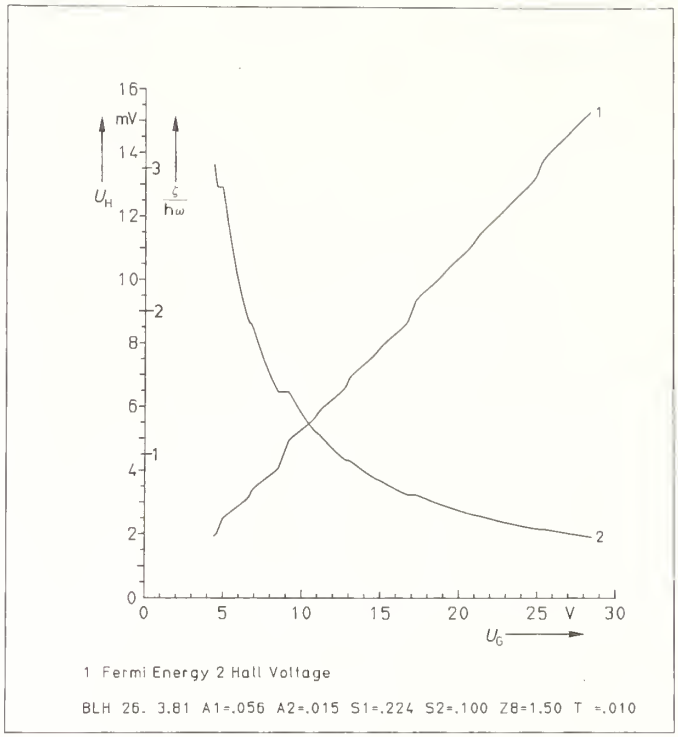


FIGURE 4. *Hall voltage, U_H, and separation of the Fermi energy ζ from the conduction band edge, in units of $\hbar\omega$, as functions of the gate voltage U_G, for a Si MOS transistor. Parameters as in Fig. 3.*

By fitting the calculated curves to the experimental Si MOS data, the following results were obtained for the parameters:

(1) The two level splittings were found to be equal to $0.224\ \hbar\omega$ and $0.10\ \hbar\omega$.

(2) $\mu = e\ \tau_0/m$ was found to depend on n in the manner shown in Fig. 5.

(3) $D_b(\zeta)$ was obtained as 7.16×10^{13} /eV cm^2 which means that, on changing U_G, at most 2% of the electrons go into or come out of bound states.

(4) d_{ox}/ϵ_r, i.e., the thickness of the oxide layer, divided by its relative permittivity, was found to be 26.0 nm, or, with $\epsilon_r = 3.78$ [11], $d_{ox} = 98$ nm.

(5) U_G depends linearly on n, as it should, but there is an offset, such that U_G is not zero for $n_0 = 0$. The corresponding number of bound electrons is four times as large as the number corresponding to the step width.

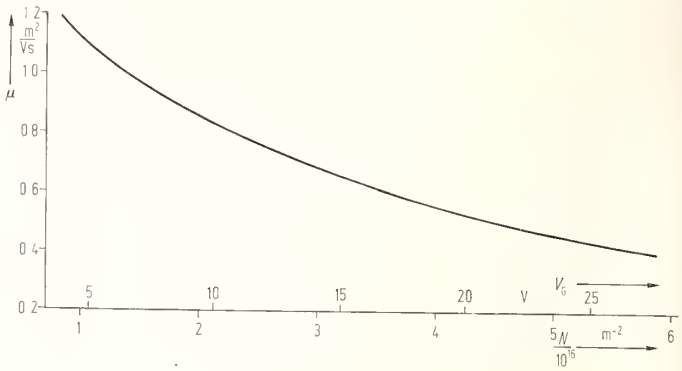


FIGURE 5. *Mobility, μ, as a function of the gate voltage, V_G, and the density of conduction electrons, N, for a Si MOS transistor. Parameters as for Figs. 3 and 4.*

The authors have the pleasure to acknowledge helpful discussions with Dr. K. von Klitzing and Dr. H. Köhler, University of Würzburg, Würzburg, Germany.

References

[1] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).
[2] E. Braun, E. Staben, and K. v. Klitzing, PTB-Mitteilungen 90, 350 (1981).
[3] D. C. Tsui and A. C. Gossard, Appl. Phys. Lett. 38, 550 (1981).
[4] L. Bliek, results presented at the 12th International Conference on the Physics of Semiconductors, Stuttgart, 1974.
[5] T. Ando, Y. Matsumoto, and Y. Uemura, J. Phys. Soc. Jpn. 39, 279 (1975).
[6] See for example L. M. Roth and P. Argyres, in *Semiconductors and Semimetals, Vol. 1*, Ed. by R. K. Willardson and A. C. Beer (Academic Press, New York, 1966), p. 159.
[7] R. B. Dingle, Proc. R. Soc. London, Ser. A: 211, 517 (1952).
[8] L. M. Bliek, G. Landwehr, and M. v. Ortenberg, in *Proceedings of the 9th International Conference on the Physics of Semiconductors, Moscow*, Ed. by S. M. Rybkin, Acad. Sci. USSR (Nauka, Leningrad, 1968), p. 710.
[9] L. M. Bliek, in *Proceedings of the 12th International Conference on the Physics of Semiconductors, Stuttgart*, Ed. by M. H. Pilkuhn (Teubner, Stuttgart), p. 729.
[10] T. Ando and Y. Uemura, J. Phys. Soc. Jpn. 36, 959 (1974).
[11] Several different values for ϵ_r are given in the literature. The present one was measured at low temperatures by P. Thoma, PTB (private communication).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Quantum Hall Effect: Role of Inversion Layer Geometry and Random Impurity Potential

R. W. Rendell*

Electron Physics Group, National Bureau of Standards, Washington, DC 20234

and

S. M. Girvin

Surface Science Division, National Bureau of Standards, Washington, DC 20234

The recently discovered quantum Hall effect offers the potential for a new precision determination of the fine-structure constant and establishment of a quantum standard of resistance. We present here a discussion of possible sources of error associated with finite sample size. In addition, we present a preliminary report on an investigation of the density of states between Landau levels in the presence of a model impurity potential which is based on a non-perturbative variational calculation using recently developed field theoretic techniques.

Key words: fine-structure constant; impurity scattering; measurement errors; quantum Hall effect; resistance standard.

1. Introduction

The recently discovered quantum Hall effect [1] has attracted a great deal of interest because of the possibility of obtaining an improved value of the fine-structure constant and development of a quantum standard of resistance. It is well known that the Hall coefficient of a two dimensional electron gas (e.g., a MOSFET or heterojunction inversion layer) oscillates with magnetic field because of quantization of the electronic states into Landau levels. At low temperatures and high magnetic fields, the carrier scattering rate between Landau levels becomes extremely low. When the Fermi level is adjusted to fill i levels, current flow becomes nearly lossless and the Hall resistance goes through a plateau assuming the universal value

$$R = h/e^2 i \ , \qquad (1)$$

apparently to extreme accuracy.

Two types of questions are of great current interest. The first is: What are the possible sources of experimental error in a high precision measurement of h/e^2 using a quantum Hall effect device? We present here a discussion of errors introduced by end effects in MOSFET devices of finite length. Within a simple model of the inversion layer we formulate expressions for the current and voltage distributions and for the (very small) deviation of the Hall resistance from its ideal value. Limitations of the model and other possible sources of error will also be briefly discussed.

The second question of interest is: What are the fundamental physical limitations of Eq. (1)? One of the most important topics is the effect of disorder and the consequent localization of some of the carriers [2–4]. As a small step toward developing a quantitative understanding of the role of impurity scattering, we have investigated the density of states between Landau levels in the presence of a random Gaussian white noise potential. We present here preliminary results of a non-perturbative variational calculation of the density of states using recently developed field theoretic techniques.

2. End Effect Errors

We have analyzed [5] a standard model [6, 7] in which the inversion layer system is assumed to consist of a sharply defined (ideally two dimensional) rectangle within which the Hall coefficient is constant. The source and drain electrodes are attached to the top and bottom of the rectangle and the Hall probes are taken to be point contacts at arbitrary locations on the sides of the device. It is assumed that no current is drawn from the Hall probes.

In order to calculate the current and voltage distributions within the sample it is necessary to determine what the appropriate boundary conditions are. Because the source and drain are heavily doped metallic regions, they have a very small Hall coefficient and will short out any electric field parallel to the boundary at the top and bottom of the device. The assumption that the Hall field is completely shorted out at the ends slightly overestimates the error, since the Hall coefficient in the electrodes, though small, is non-zero and has the same sign as in the inversion layer. The boundary condition on the sides is taken to be no current flow out the sides. We may take advantage of the assumption of a uniform value of the Hall coefficient to note that the current makes a fixed angle δ with respect to the electric field. Thus since the current at the edges is necessarily parallel to the edges, the electric field must make an angle δ with respect to the sides.

Assuming steady state conditions one now has a well defined electrostatics problem which can be solved employing the techniques of conformal mapping [5–7]. Our method of using a direct conformal mapping is consider-

*NRC/NBS Postdoctoral Fellow

ably simpler than previous approaches and the results are more readily evaluated in the limit $\delta \rightarrow \pi/2$ appropriate to the quantum Hall regime. Since the Hall angle is so close to $\pi/2$ it is convenient to define the quantity ϵ via

$$\delta = \frac{\pi}{2}(1 - \epsilon) . \tag{2}$$

Because of the sudden change in the boundary conditions at the corners the current is forced to enter the device almost entirely from one corner and to exit from the opposite corner [8]. In fact within this simple model having sharply defined corners, the electric field and current density both have power law singularities of the form

$$f(y) \sim (T/y)^{1-\epsilon} , \tag{3}$$

where y is the distance above the corner and T is the length of the device. The equipotential lines for this situation are shown qualitatively in Fig. 1.

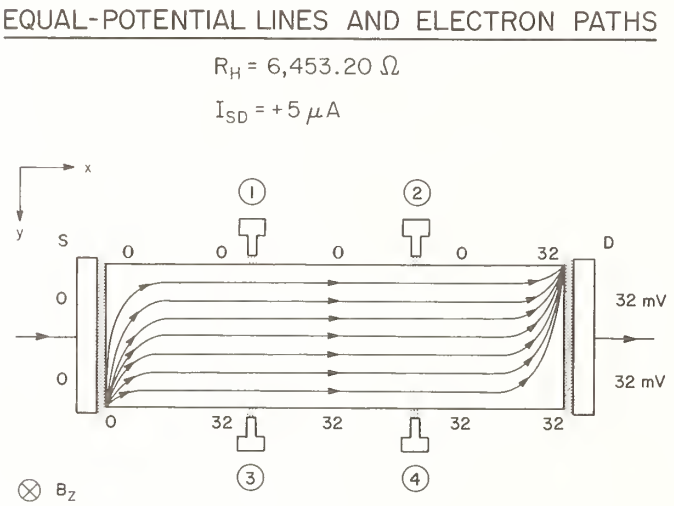


FIGURE 1. *Qualitative picture of equipotential lines in the inversion layer. The numbers on the sides indicate the approximate potential distribution when the source drain current is 5 μA and the Fermi level is at the i = 4 Hall step.*

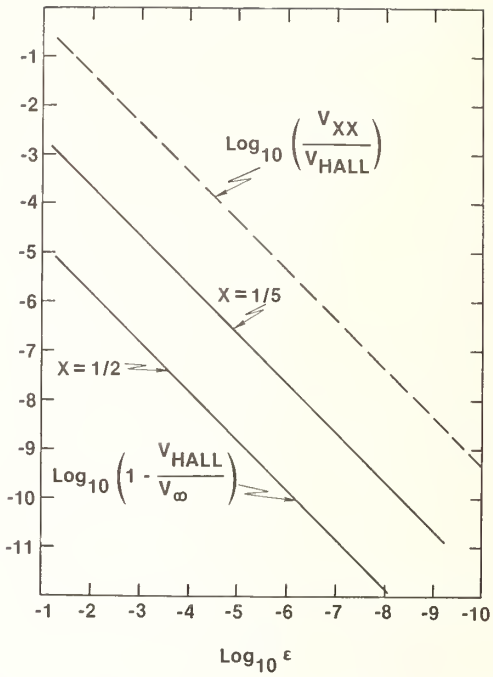


FIGURE 2. *Graph of error in measured Hall voltage due to the shorting effect. See text for a detailed discussion.*

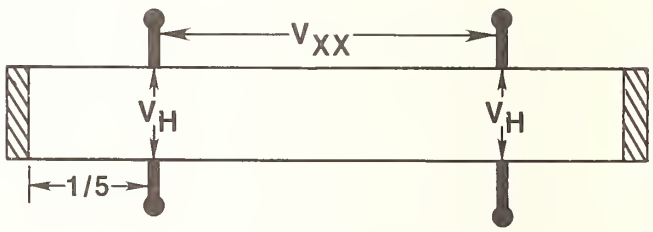


FIGURE 3. *Geometry of NRL MOSFET showing the location of the four potential probes.*

By integrating the calculated electric field [5] on a path connecting the two Hall probes we have determined the ratio of the actual Hall voltage to the ideal Hall voltage for an infinitely long device. These results are displayed graphically in Fig. 2. In order to understand this figure it is useful to refer to Fig. 3 which shows the geometry of the devices used in the Naval Research Laboratory (NRL)—National Bureau of Standards (NBS) collaboration. V_{xx} refers to the voltage measured between two probes on the same side and is a measure of the carrier scattering rate and the resistivity ρ_{xx}. V_H refers to the Hall voltage measured across the device. In the NRL geometry the probes are located 1/5 of the way along the edge from the ends. Returning now to Fig. 2 one sees a dashed line from which the quantity ϵ may be determined after V_{xx}/V_{Hall} has been measured. Given ϵ one may then determine the error from the quantity

$$\Delta = \log_{10} \frac{(1 - V_{\text{Hall}})}{V_\infty} , \tag{4}$$

where V_∞ is the ideal Hall voltage for an infinitely long device. The solid line labelled $x = 1/5$ in Fig. 2 gives this error for Hall probes located as in the NRL geometry. The line labelled $x = 1/2$ is for the case of Hall probes located at the midpoints of the sides. These curves were calculated using the aspect ratio of 6.5 appropriate to the NRL geometry. The smallest attainable value of ϵ is not known but preliminary measurements [1, 9] give an upper limit of 5×10^{-7} and it is likely that ϵ is much smaller than this. From Fig. 2 one sees that the error due to the shorting effect is less than 3×10^{-10} for Hall probes at the center of a device with aspect ratio 6.5.

Within the present simple model it appears that the shorting effect can be neglected as an important source of error. It is clear however that this simple model has several limitations. It will certainly fail in the singular field regions in the corners but this should not have a serious effect in sufficiently long devices. Also we have implicitly assumed that the electrons scatter in the inver-

sion region so that it is meaningful to speak of a fixed Hall angle. It is quite possible that the scattering rate is so low that this is not valid. Boundary scattering effects in this problem are generally poorly understood at the moment.

3. Random Impurity Potential

In the presence of a random impurity potential the sharp Landau levels will be broadened into a distribution of energies. This broadening has been subjected to perturbative analysis [10–14] but even self-consistent perturbation theory is ill-suited to situations involving localized states [15]. We present here preliminary results of a non-perturbative variational calculation of the density of states in a two dimensional white noise potential in the presence of a magnetic field. No assumption of a smoothly varying potential is made as in Ref. [15].

Haughton and Schäfer (HS) [17] have developed a variational scheme for calculating the density of localized states at asymptotically deep energies. We have modified their method to take into account the magnetic field and use this to find the density of states in the tails of the Landau levels. The probability distribution for the potential is taken to be

$$P(V) = \exp(-S(V)), \tag{5}$$

where the action S is given by

$$S = \frac{1}{2\sigma^2} \int d^2x V^2(x) \tag{6}$$

and σ is a measure of the disorder. If one is interested in the density of states at some energy Ω then one would like to choose V to minimize S subject to the constraint that the Schroedinger equation with potential V have a solution with eigenvalue Ω. HS show that the wave function associated with this extremal potential satisfies a non-linear Schroedinger (instanton) equation. In the presence of a magnetic field B with vector potential

$$\mathbf{A} = \frac{B}{2} \hat{z} \times \mathbf{r} \tag{7}$$

the HS result is readily modified to yield

$$-\frac{1}{2}\nabla^2\psi + \frac{1}{2}R^2\psi \pm \psi^3 = \Omega\psi . \tag{8}$$

This is the equation for a two dimensional harmonic oscillator with a self-interaction. We work here with energy units $(\hbar\omega_c/2)$ where ω_c is the cyclotron frequency and use a length scale $\beta^{-1} = (2\hbar/m\omega_c)^{1/2}$. The minimum action solutions of this non-linear equation are believed to be rotationally symmetric and so we have dropped an angular momentum term from (8). Following HS we obtain for the extremal action.

$$S = KS_o \tag{9}$$

where $K = \frac{\pi}{4}\left[\frac{\hbar\omega_c}{\sigma\beta}\right]^2$ and

$$S_o = \int_0^\infty dRR\left[\psi(R)\right]^4 . \tag{10}$$

Evaluation of the fluctuations about the extremal solution is a non-trivial exercise which is rendered more difficult by the presence of the magnetic field. Hence, as a first crude approximation we neglect these fluctuations and using Eqs. (5) and (9) approximate the density of states by

$$\rho(\Omega) = \exp(-KS_o) . \tag{11}$$

The procedure used for evaluation of this equation is as follows. The ± sign in Eq. (8) was chosen according to whether a self-attractive or repulsive interaction was required to pull a state out of the nearest Landau level and Eq. (8) was then solved numerically. The results are shown in Fig. 4 which displays a graph of S_o versus Ω. The curve centered on $\Omega = +1$ is for zero node solutions

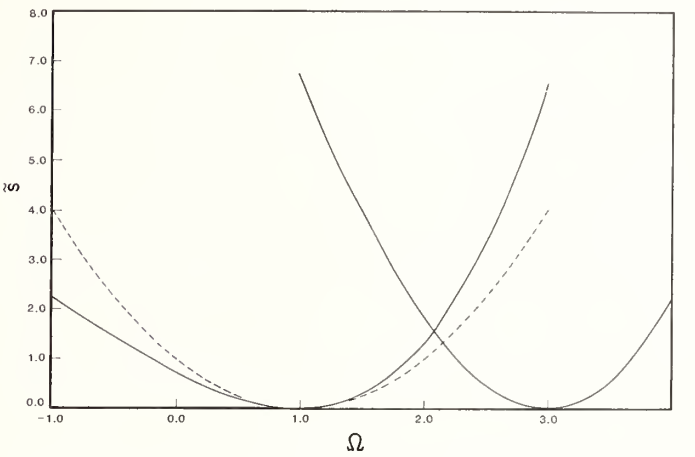


FIGURE 4. *Graph of the extremal action appearing in the variational theory of the density of states. The dashed line is a perturbation theory result. The solid lines are based on numerical solution of Eq. (8).*

to Eq. (8) whose primary parentage is the lowest Landau level (at $\Omega = 1$). The curve centered on $\Omega = 3$ is for one node solutions to Eq. (8) whose primary parentage is the second Landau level (at $\Omega = 3$). The dashed line is the action corresponding to a lowest order perturbation theory solution of Eq. (8):

$$S_o = (\Omega - 1)^2 . \tag{12}$$

We see that the exact solution matches this closely for Ω close to one but deviates further away. In the limit $\Omega \to \infty$, the exact S_o becomes linear in Ω and independent of the magnetic field. This occurs because the size of the localized state becomes much smaller than the Landau radius.

The approximate density of states given by Eq. (11) with $K = 4$ for the lowest Landau level is shown in Fig. 5. The dashed line is a simple Gaussian obtained from Eq. (12) while the solid line uses the numerical result for the action. One sees that the numerical result is skewed toward lower energies relative to the Gaussian. This is a result of the Landau levels repelling each other under the influence of the perturbing potential. The numerical result crosses over from a Gaussian near the peak to a slower exponential falloff in the low energy tail. This is consistent with the known result at low energies in the absence of a magnetic field [16].

We now need to consider the range of energies in which this variational approach is valid. We have seen that the variational approximation agrees with the perturbation result close to the peak. However, both the perturbation theory and the variational calculation assume rotational symmetry which becomes a poor

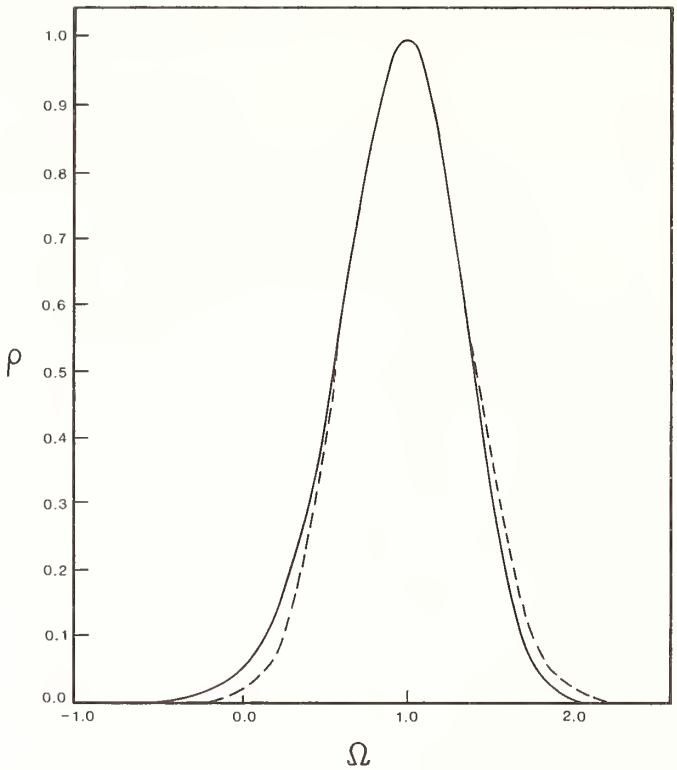


FIGURE 5. *Approximate density of states for the lowest Landau level calculated in the variational theory. The dashed line is a Gaussian approximation based on lowest order perturbation theory. The solid line is the numerical result. Note the asymmetry.*

approximation close to the peak due to the large number of degenerate levels. In the absence of a magnetic field, one needs to be well outside the band of unperturbed states in order for the variational approximation to be valid. This would suggest that the present result is valid only in the tails of the density of states—the region inaccessible to ordinary perturbation theory.

It is possible in principle to experimentally determine the density of states by measuring the MOSFET differential capacitance. More detailed calculations of the density of states are underway and will be presented elsewhere [18]. Another interesting and important quantity is the location of each of the mobility edges. It is beyond the present capability of theory to reliably estimate the positions of the mobility edges but the asymmetry we have observed in the density of states suggests that the mobility edges will not be symmetrically placed about the nominal Landau level position.

4. Summary

We have briefly reported on investigations into the inversion layer voltage distribution in the quantum Hall regime and on the density of states between Landau levels in the presence of a model impurity potential. The quantum Hall phenomenon presents a great variety of new physics puzzles which challenge our understanding of basic ideas of transport. Further study of the many questions of current interest should prove fruitful for some time to come.

References

[1] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).
[2] R. E. Prange, Phys. Rev. B 23, 4802 (1981).
[3] R. B. Laughlin, Phys. Rev. B 23, 5632 (1981).
[4] H. Aoki and T. Ando, Solid State Commun. 38, 1079 (1981).
[5] R. W. Rendell and S. M. Girvin, Phys. Rev. B 23, 6610 (1981).
[6] R. F. Wick, J. Appl. Phys. 25, 741 (1954).
[7] H. J. Lippman and F. Kuhrt, Z. Naturforsch. Teil A: 13, 462 (1958); 13, 474 (1958).
[8] S. Kawaji, Surf. Sci. 73, 46 (1978).
[9] M. E. Cage, R. F. Dziuba, B. F. Field, R. J. Wagner, and C. F. Lavine, private communication.
[10] T. Ando and Y. Uemura, J. Phys. Soc. Jpn. 36, 959 (1974).
[11] T. Ando, J. Phys. Soc. Jpn. 38, 989 (1975).
[12] C. S. Ting, S. C. Ying, and J. J. Quinn, Phys. Rev. B 16, 5394 (1977).
[13] S. Fujita and M. Prassad, J. Phys. Chem. Solids 38, 1351 (1977).
[14] S. Das Sarma, Solid State Commun. 36, 357 (1980).
[15] John L. Cardy, J. Phys. C 11, 1321 (1978).
[16] Masaru Tsukada, J. Phys. Soc. Jpn. 41, 1466 (1976).
[17] A. Houghton and L. Schäfer, J. Phys. A 12, 1309 (1979).
[18] R. W. Rendell and S. M. Girvin, unpublished.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

GRAVITY AND RELATIVITY

Experiments Relating to the Newtonian Gravitational Constant

H. de Boer

Physikalisch-Technische Bundesanstalt (PTB) D-3300 Braunschweig, F.R.G.

The Newtonian gravitational constant G is one of the natural constants, the exact knowledge of which might give us important answers to cosmological questions and contribute to the support of a theory of gravitation. In a survey, the problems are set forth which are related to the gravitational constant, appertaining measurements are described, and the uncertainties of measurements obtained so far are compiled.

Key words: dependence of G on time; dependence of Newton's law on R; determinations of the gravitational constant; Newtonian gravitational constant.

1. Introduction

In all general lectures, we are accustomed to the significant progress achieved in the respective field of metrology being discussed, a reduction of the measuring uncertainty, and a theoretical interpretation of the problems in the field. With regard to the Newtonian gravitational constant G it is not, however, possible to report any pioneering success, as the progress made in this special field is slow and the problems facing us are far from being solved experimentally. The interest of astronomers and physicists in the exact knowledge of G, and in the questions bound up with this, has grown with the initial findings concerning the composition and interaction of the mass distribution in the cosmos. New gravitational theories have been developed, and questions are being raised in support of the various theories which are increasingly difficult to answer.

In this review paper we summarize the experiments carried out to determine the Newtonian gravitational constant, and also briefly deal with some experiments which are known to us under the following key-words: inertial mass; active and passive gravitational mass; R-dependence of the gravitational constant; and time dependence of the gravitational constant. At the end we report on the experiment carried out in the PTB to determine G.

In its simplest form, the law of gravity formulated by Newton in 1687 states that two spherical bodies with mass m_1 and m_2 separated from each other by a distance R attract each other with the force

$$F = G \frac{m_1 m_2}{R^2},$$

where G is the Newtonian gravitational constant.

To give an impression of the magnitude of the forces occurring, Fig. 1 shows the forces as a function of the distance between two spherical bodies of either platinum, lead, and glass of a fixed radius (unbroken lines), and as a function of the distance with a specified mass (broken lines). The forces which occur are within the 10^{-9} N range.

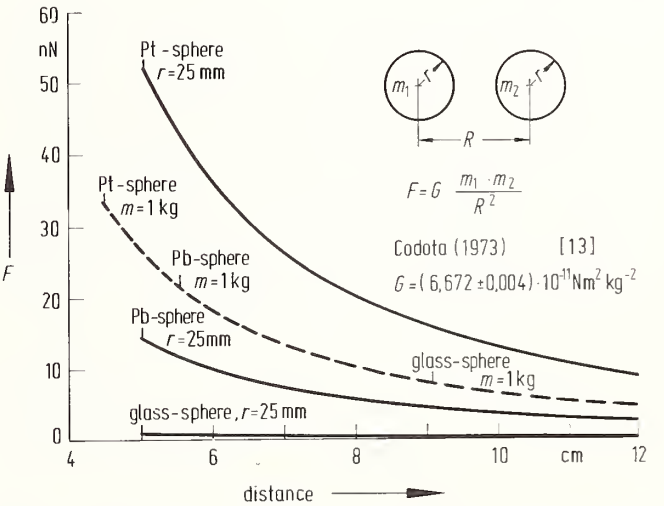


FIGURE 1. *Forces as a function of distance between spherical bodies of platinum, lead, and glass at a fixed radius (unbroken lines) and as a function of distance with specified mass (broken lines).*

More or less the same force of attraction would be produced by a voltage of about 10 V between the spheres separated from each other by a minimum distance. As can be seen from this simple example, contact voltages possibly occurring between the attracting masses at the relative measuring uncertainty to be aimed at of $\Delta G/G = 10^{-6}$ must be less than 10^{-5} V.

If one of the interacting bodies is the Earth and the other a 1 kg weight at the Earth's surface, a force of attraction of 10 N is encountered. The most precise balances for the comparison of forces due to weight have a relative uncertainty of $\Delta F/F = \Delta m/m \simeq 10^{-8}$. Thus, when using such a balance to compare two 1 kg weights and when bringing a 1 kg weight closer from the top or

from the bottom to one of the weights placed on the beams of the balance, the balance would not show any reaction as a result of the gravitational effect produced.

The task of an experimenter who wishes to measure G consists in the measurement of very small forces to be carried out, as we shall see below, in the presence of a great many disturbing forces.

2. G Experiments and Results

Experiments hitherto carried out can be roughly grouped as follows:

(1) measurements in which the Earth and mountains or parts of the Earth's crust are involved as reaction masses (so-called large-scale experiments);
(2) measurements with Jolly beam balances; and
(3) measurements with torsion balances.

2.1 Large-Scale Experiments

The measurements within the scope of the first group aimed primarily at determining the mean density of the Earth. The possibility of calculating the value of the gravitational constant from these measurements is only a secondary result. Three methods served to determine the Earth's mean density (Fig. 2):

(1) pendulum measurements at sea level and, subsequently, on an elevated plane the geometry and density of which are measured;
(2) determination of the deviation of the direction of the acceleration due to gravity from the radial direction of the Earth at two sides of a mountain the geometry and density of which must again be known; and
(3) determination of the mean density from measurements of the acceleration due to gravity, for example with pendulums, at various depths of a mine. To calculate G, the mass of the spherical shells is related to the whole mass of the Earth.

Measurements according to the first two methods were carried out for the first time in 1738 by Bouguer and de la Condamine on a suitably shaped mountain in Peru. Measurement results obtained according to the third method were published in 1828 by Airy and his collaborators who had carried out measurements in a mine 400 m in depth. The G value which would at present be the most precise and which was published by Stacey, Tuck and others [57] who had carried out measurements in an Australian mine according to the last method is $G = (6.712 \pm 0.037) \times 10^{-11}\ \mathrm{N \cdot m^2 \cdot kg^{-2}}$.

Stacey and Tuck furnish another G value of $(6.730 \pm 0.025) \times 10^{-11}\ \mathrm{N \cdot m^2 \cdot kg^{-2}}$ which they calculated from comparable measurements by McCulloh in a mine 700 m in depth. Both values deviate from the 1973 CODATA value: $G = (6.672 \pm 0.004) \times 10^{-11}\ \mathrm{N \cdot m^2 \cdot kg^{-2}}$ [13], and is a possible hint that the $1/R^2$ dependence of Newton's law of force is not obeyed for short distances, as is supported by other authors. We will return to this point later.

Figure 3 shows a summary list of measured values obtained in the large-scale experiments. We see that the spread of the measured values is very large and deviates from the CODATA value, the uncertainty range of which has been entered as well. Hence it follows that measurements of this kind are not suitable for the precision determination of G.

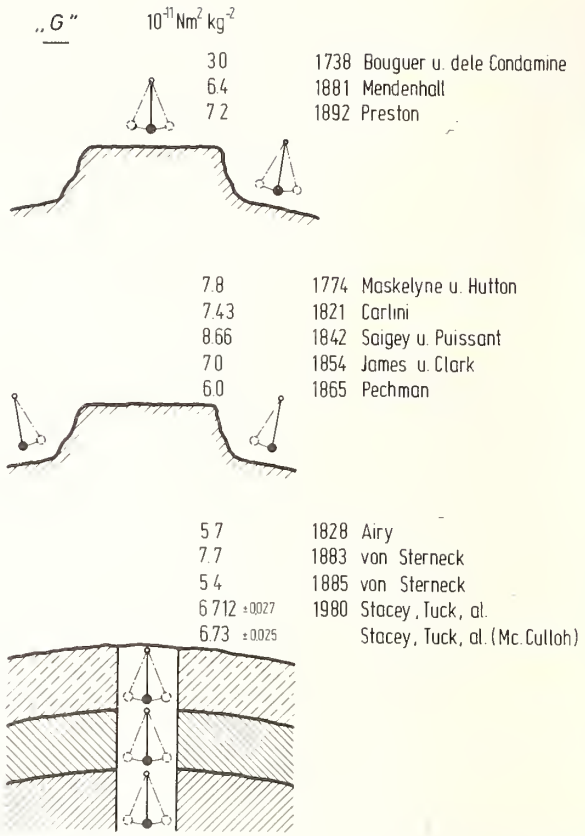


FIGURE 2. *Methods and results of "large-scale measurements" of G.*

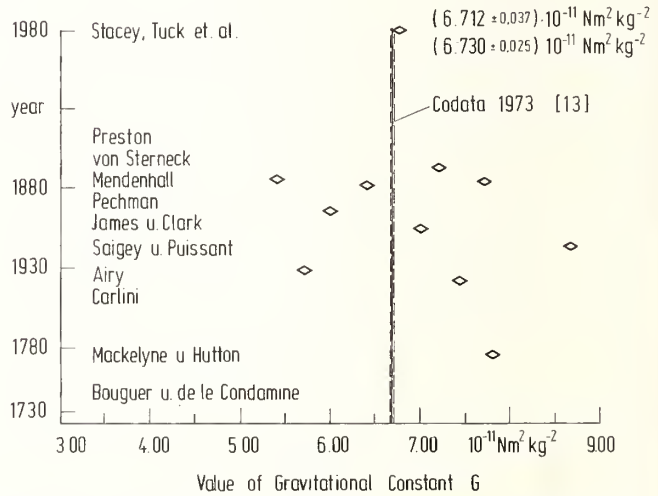


FIGURE 3. *Compilation of G values attained by "large-scale measurements."*

2.2 Measurements with Jolly Beam Balances

In the introduction it was said that it is not at present possible for the most precise balances to reveal the gravitational force between two 1 kg weights separated by a minimum distance. This can naturally be remedied by using greater masses and countermasses. Figure 4 shows a schematic drawing of the measuring arrangement used by Richarz and Krigar-Menzel [48, 49]. These authors used a balance with a 23 cm long beam on the ends of which pans were mounted at a vertical distance of 200 cm. Lead weights of 0.67 kg were weighed in the

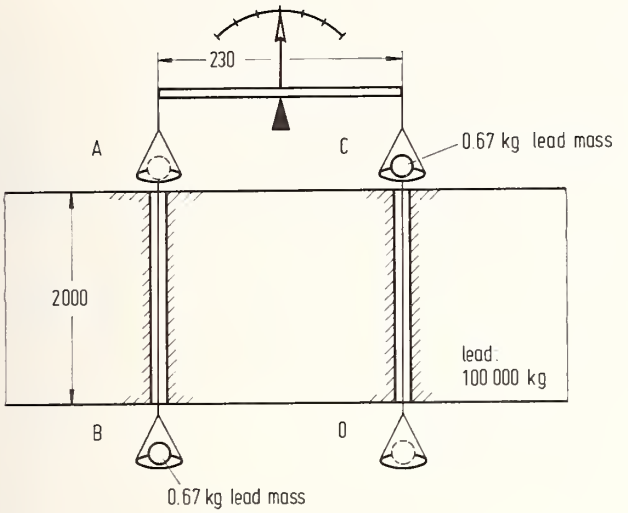


FIGURE 4. *Jolly beam balance arrangement used by Richarz and Krigar-Menzel (schematic). The beam length is 230 mm and the pan separation about 2000 mm (2 m).*

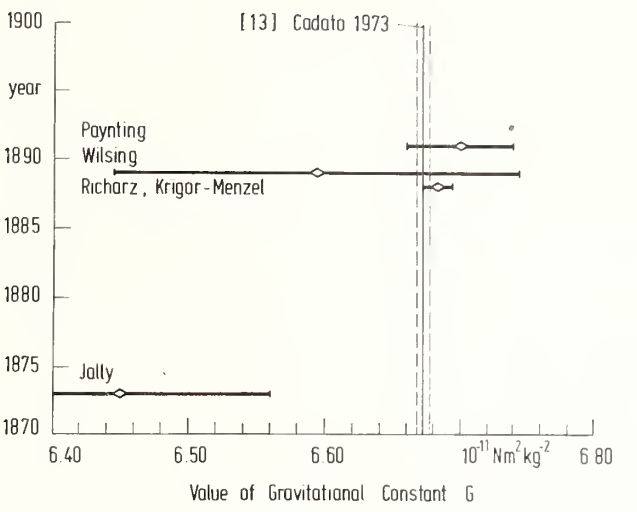


FIGURE 5. *Compilation of G values attained by Jolly beam balance measurements.*

pans A,D and C,B without the lead mass of about 100 000 kg in order to be able to correct for the difference of the acceleration due to gravity between the planes AC and BD. Subsequently, the same measurement was repeated with the lead block. These measurements allowed a G value of $(6.683 \pm 0.011) \times 10^{-11}$ $N \cdot m^2 \cdot kg^{-2}$ to be determined. Figure 5 shows a summary list of the best-known measurements of this kind with the estimated uncertainties. The relative uncertainties of the measurements are within the 10^{-2} to 10^{-3} range. The limit of this method is probably almost reached with the result obtained by Richarz and Krigar-Menzel. These authors indicate the following disturbance influences:

(1) air convection as a result of temperature gradients;
(2) non-homogeneity of the density of the sets of weights;
(3) influences by the knife-edge bearing of the balance beams; and
(4) disturbances of the ground.

Table 1 gives a summary about the most important balance measurements. After the measurements by Krigar-Menzel, and Wilsing, i.e., after 1850, no further G measurements were carried out by the weighing method.

2.3 Measurements with Torsion Balances

The most successful laboratory experiments to determine the gravitational constant have been carried out with torsion balances. Figure 6 shows the fundamental design of this balance type. Cavendish [12] was the first scientific researcher to use this measuring arrangement in 1758 for the determination of G. The arrangement con-

TABLE 1. *Determinations of the gravitational constant by means of Jolly beam balances (from L. Stegena and M. Sagitov [58])*

Experiment: author, year of publication, reference, and place of experiment	Beam length and height difference between masses m and M (cm)	Attracted masses M: material, shape, weight	Attracting masses M: material, shape, weight	Constant of gravitation G and rms error (10^{-11} $N \cdot m^2 \cdot kg^{-2}$)
Jolly, 1873 [31, 32], Munich, Germany	$L = 60$ $h = 2100.5$	Mercury in a sphere-shaped glass vessel; 5 kg each	Lead sphere; 5775.2 kg	6.447 ± 0.11
Poynting, 1891 [43, 44], Birmingham, England	$L = 123.26$	Lead-antimony alloy, sphere-shaped; 21.6 kg each	Lead-antimony alloy, sphere-shaped; 153.4 and 76.5 kg	6.70 ± 0.04
Richarz, Krigar-Menzel, 1888, [48, 49], Spandau citadel	$L = 23.32$	Lead, sphere-shaped; 0.67 kg each	Lead, parallel-piped 211.1 × 211.1 × 209.5 cm; 100536.8 kg	6.683 ± 0.011
Wilsing, 1889 [62, 63, 64] Potsdam, Germany	$L = 100$	Copper and lead-antimony alloy; 540 g each	Lead circular cylinders; 325 kg each	6.594 ± 0.15

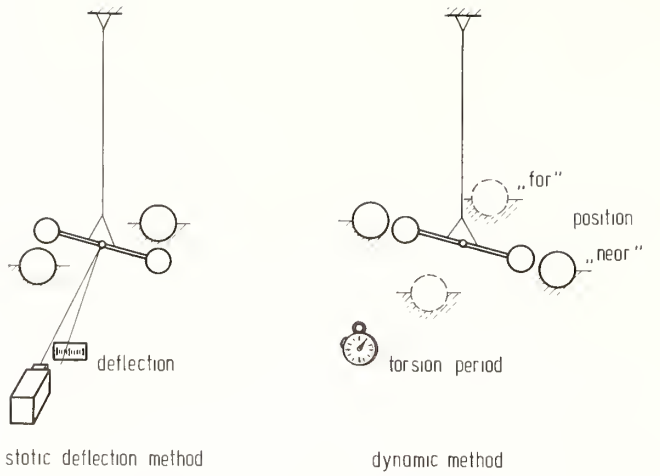


FIGURE 6. *Principles of torsion beam balance measurements.*

sists of a balance beam with weights fixed to the end. The balance beam is suspended from a torsion wire so that the direction of the wire and the direction of the gravitational acceleration are the same. The outer reaction mass arrangement is such that the reciprocal force of attraction lies in the plane perpendicular to the acceleration due to gravity. When G is determined from the mass attraction with the aid of the angle of deflection of the torsion balance, as is shown on the left of Fig. 6, this method is referred to as the static method. In this case, the torsional moment of the wire must be known for the calculation of G; it is generally calculated on the basis of a measurement of the period of oscillation with the moment of inertia being known. The measuring method is referred to as dynamic when the torque and, thus, G are determined on the basis of the mass attraction by two measurements of the period of oscillation of the torsion balance, the great masses being arranged first in the "near" position and then in the "far" position. This method is shown on the right of Fig. 6. It was used for the first time in 1896 by Eötvös [17, 18].

Figure 7 and Table 2 show a summary of the results of the most significant measurements with torsion balances since the first measurements carried out by Cavendish. As can be seen from Figure 7, the G values spread over a wide range until 1920. After this date, the spread decreases. If we wish to know the most precise G values,

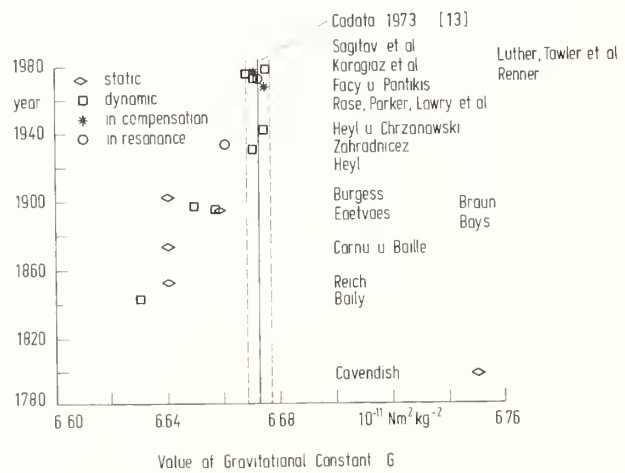


FIGURE 7. *Compilation of G values attained by torsion beam balance measurements.*

we will have to consider the values obtained after 1920. Figure 8 shows the values obtained after 1920 together with the simple standard deviations stated by the authors. The reports do not state in every case exactly how the calculation was made, so that a direct comparison of the values stated requires caution. In most cases, the authors state measured values for series of measurements, obtained successively when making measurements under varying conditions, as, for example, using different materials for the masses or different torsion wires.

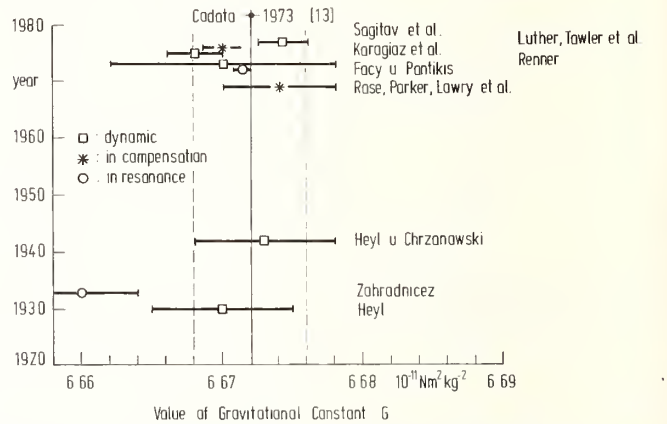


FIGURE 8. *Compilation of G values and appropriate one standard deviation uncertainties attained by torsion beam balance measurements after 1920.*

2.3.1 Measurements by Heyl (1930) and Heyl and Chrzanowski (1942)

Heyl and, at a later date, Heyl and Chrzanowski [24, 26] used the dynamic method. Heyl was the first to use large, cylindrical masses since these could be manufactured and dimensioned with a higher precision. He had, however, to pay for this advantage with a great deal of mathematical calculation at a time when programmable computers were not yet available. The large 66 kg masses were made of tool steel. The small masses Heyl used were spherical bodies weighing about 50 g and made of gold, platinum, and optical glass. The results obtained with the different materials differed from one another. Heyl found that the gold masses increased in size during the test because mercury from the pressure measuring system combined with the gold. Heyl obtained the following result from his 1930 measurements:

$$G = (6.670 \pm 0.005) \times 10^{-11}\,\mathrm{N \cdot m^2 \cdot kg^{-2}}.$$

In 1942, Heyl and Chrzanowski reported on another measurement. They intended to improve Heyl's measuring apparatus to allow a better G value to be obtained. They carried out the following improvements:

(1) Extension of the balance beam from 20 cm to 80 cm in length. It subsequently failed because the deflection of the beam exceeded the permissible value. An increase of the beam section resulted in the mass increasing to such an extent that it could not be handled by the torsion wire.
(2) Heyl and Chrzanowski made an attempt to use a disc-shaped balance beam offering the advantage of rotational symmetry with respect to the gravitational forces. In this case too, difficulties arose concerning the carrying capacity of the torsion wire.
(3) Furthermore, they attempted to replace the torsion

TABLE 2. *Determination of the gravitational constant by means of torsion beam balances (from L. Stegena and M. Sagitov [58], updated).*

Author, year of publication, place, reference	Beam length (cm)	Material and weight of attracted experimental masses	Material, shape and weight of attracting experimental masses	Material, diameter of filament (μm)	Air pressure (mm Hg)	Techn. s-stat. d-dyn.	Constant of gravitation and rms error (10^{-11}N·m²·kg^{-2})
1. H. Cavendish, 1798 [12]	182	lead, 730 g	lead spheres, 158 kg	silvered copper	760	s	6.75 ± 0.05
2. F. Reich, 1838, 1852, Freiburg,Germany [45]	202	lead, zinc, bismuth, iron, 484 g	lead sphere, 45 kg each; cast iron spheres, 39 kg each	iron and others	760	s	6.64 ± 0.06
3. F. Baily, 1843, Tavistock Place, England [2]	196	platinum, lead, zinc, glass, brass ivory, 116–1538 g	lead spheres, 172.4 kg each	copper, silk, iron, brass	760	s d	6.63 ± 0.07
4. A. Cornu, J. Baille 1872, 1873, Paris, France [15, 16]	50	copper, 109 g	mercury in sphere-shaped vessels, 12 kg	annealed silver	760	s	6.64 ± 0.017
5. C. V. Boys, 1895, Oxford, England [7]	2.3	gold, 1.32 g 2.65 g, 3.98 g	lead spheres, 7.4 kg	quartz	760; unsuccess-full with hydrogen	s	6.658 ± 0.007
6. R. Eötvös, 1886, Budapest, Hungary [17]			lead parallelepi-peds, 600 kg		760	d	6.657 ± 0.013
7. C. Braun, 1897, Marienstein, Austria [9]	24.6	brass, 54 g	mercury in sphere-shaped vessels, 9 kg	brass	4 17	s d	6.649 ± 0.002
8. G. Burgess, 1902, Paris, France [10, 11]	24	lead, 2 kg	lead spheres, 10 kg	quartz	760	s	6.64
9. P. Heyl, 1930, Washington, USA [25]	20.6	gold, platinum	steel circular cylinders, 66 kg	tungsten, 25 μm	2	d	6.670 ± 0.005
10. J. Zahradnićek, 1933, Brünn, Czechoslovakia [65]	40.6	lead, 94 g	lead spheres, 11.3 kg	platinum-iridium, 18 μm	760	d	6.66 ± 0.04
11. P. Heyl, P. Chrzanowski, 1942, Washington, USA [26]	20	platinum,	steel circular cylinders, 66 kg	hard drawn tungsten; 30 μm annealed tungsten; 35 μm	0.5	d	6.673 ± 0.003
12. R.D. Rose *et al.*; R. A. Lowry *et al.*; G. G. Luther *et al.*, 1969, 1976, Charlottesville, VA, NBS [52, 38, 39, 3]	3.8	cylinder, dumbbell aluminum alloy	tungsten spheres, 10 kg	fused quartz, 25 μm	helium	compen-sation	6.6699 ± 0.0014
13. L. Facy, C. Pontikis, 1972, Yveline, France [19, 20, 21]	6.0	spheres; silver, copper, brass, 4.6 g	spheres; silver, copper, brass, mercury, lead, bronze	platinum; 20 μm tungsten; —	1×10^{-7}	reso-nance	6.6714 ± 0.0006
14. J. Renner, 1973, Budapest, Hungary [46, 47]	—	spheres; copper, 16.0 g	4 equal hollow cylinders, 2 filled with mercury, 17.4 kg	tungsten; platinum-iridium;	1 to 2	d	6.670 ± 0.008
15. O. V. Karagioz, 1976, Moscow, USSR [33]	15.6	spheres; copper, 0.3 g	spheres; steel brass, copper dural, 4.2 kg	tungsten; 5 μm	1×10^{-8}	d	6.668 ± 0.002
16. M. U. Sagitov, 1977, Moscow, USSR [53, 54, 58]	37.0	cylinders; copper, 29 g	cylinders; steel, 39.7 kg	tungsten; 35 μm	5×10^{-5}	d	6.6745 ± 0.003

wire by two wires or a bifilar suspension, to achieve a higher carrying capacity and at the same time, a prolonged period of oscillation of the torsion balance. The attempts to achieve a multiple filament suspension failed, as they did not succeed in loading the filaments uniformly. The 1942 tests used hard-drawn tungsten wire 0.030 mm in diameter and annealed tungsten wire 0.035 mm in diameter; thus the modifications in comparison with the 1930 tests were not very significant.

The results obtained with the different wires varied slightly. The overall result was in agreement with the results obtained in 1930. These workers held the view that with their final result of

$$G = (6.673 \pm 0.005) \times 10^{-11}\,\mathrm{N \cdot m^2 \cdot kg^{-2}},$$

they had reached the metrological limits of their apparatus.

2.3.2 Measurements by Rose, Beams and Others (1969) and by Luther and Others

Chronologically, the G measurements by Rose [52], Beams [3], and others in 1969 were the next to be carried out; they were continued by Luther [38, 39] and others at a later date.

Figure 9 is a schematic representation of the measuring arrangement. On an air-supported rotary table, two

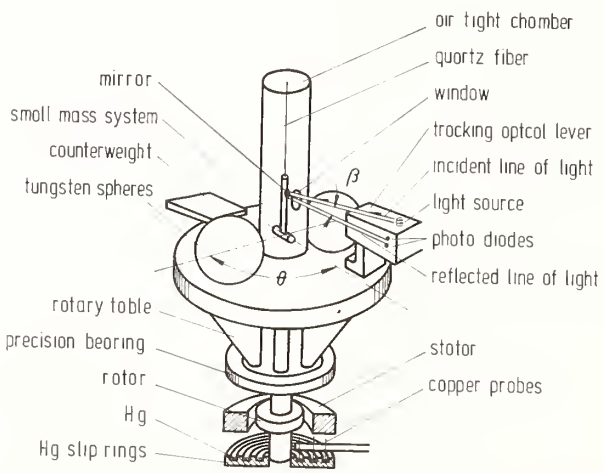


FIGURE 9. *Measurement arrangement used by R. D. Rose et al. [52].*

tungsten spheres are fixed as large masses, which together with a torsion balance can be rotated about the balance axis. When placing the large masses at an appropriate angle to the small masses, the torsion balance is deflected. This deflection is measured with a spot of light by means of photodiodes. From this a signal is formed which controls a motor driving the rotary table. The rotary movement of the table is such that the deflection of the torsion balance is just compensated for by the forces of inertia. This interaction of the forces of inertia and gravitation results in an ever-accelerating rotation of the table, from which G can be determined. What is remarkable and new about this arrangement is the following:

(1) The measurement of the gravitational forces is not only based on geometric measurements but also on a frequency measurement (or, to be more precise, on the measurement of a frequency change). There is no deflection or oscillation of the torsion balance, so this method can be referred to as a compensation method. The measurement can be carried out by integrating over a prolonged period of time.

(2) By the rotation of the table, influences of gradients of the gravitational field of surrounding masses are averaged out to first order.

(3) During the measurement, the coordinates of the reaction masses remain fixed and can therefore be estimated with a smaller uncertainty.

2.3.3 Measurements by Facy and Pontikis

In 1972, Facy and Pontikis [19, 20, 21] published the final results of their G measurements, using the resonance method (Figure 10). The large masses fixed on a motor-driven rotary table circle the small masses attached to

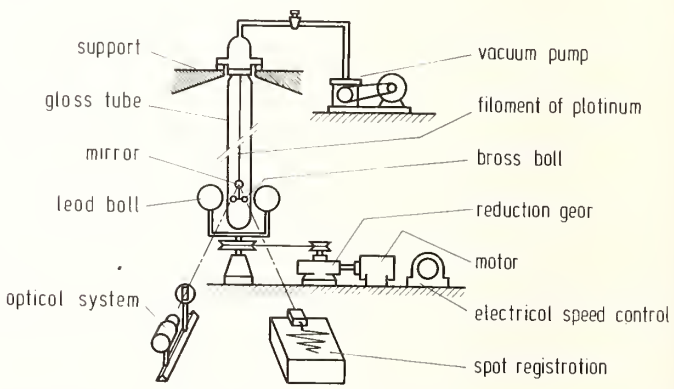


FIGURE 10. *Measurement arrangement used by Facy and Pontikis [19].*

the torsion balance. At certain rotational frequencies of the large masses, there is a resonant excitation of the oscillation of the torsion balance. The movements of the torsion balance are recorded. The measurements are evaluated using a computer, the respective geometric data and the oscillation values actually measured being used in the theoretical differential equation of oscillation from which the G value best suited for the measurements is calculated. As other authors had previously done, Facy and Pontikis also used different materials in their tests, i.e., silver, copper, bronze, lead, and brass. Unfortunately, the documentation of the results is obscure and incomplete; it is not clear, for example, which material combinations were selected for the large and small masses to obtain the results published in 1972. The authors announced a final, detailed report which, however, to my knowledge was not published.

2.3.4 Measurements by Renner

Another G measurement was reported on in 1973 by Ya. Renner, Budapest [46, 47]. He applied the dynamic measuring method in which the period of oscillation in the near and in the far position of the large masses was determined in the customary way. The large masses con-

sisted of four identical hollow cylinders mounted on a table which could be rotated about the axis of the torsion balance. Two cylinders arranged opposite each other were filled with mercury.

2.3.5 Measurements by Karagioz and Others

A preliminary measurement by Karagioz [33] and others in 1976 was the next to be published. The authors used the dynamical method. They did not, however, measure in the near and in the far position of the large masses but used only one large mass which was shifted in the direction of the torsion beam.

The mounting for the large mass was provided with boreholes 4 mm in diameter and each at a distance of 12 mm, into which the spherical mass was placed. The large masses were made of steel, copper, brass, and duraluminium and the small masses of oxygen-free copper. Only with a mass of steel did it prove possible to carry out measurements with a small deviation of the measured values. The unroundness of the spherical masses resulted in a positional uncertainty of the mass centers of about 50 μm. Karagioz states that the mass center for a relative uncertainty of smaller than 10^{-6} must be positioned to better than 3 μm.

2.3.6 Measurements by Sagitov and Others (1977)

In 1977 a measurement was carried out by M. Sagitov [53, 54, 58] and others. Sagitov too uses the dynamic method and like Karagioz, he shifts the large masses in the direction of the balance beam. As a reason for this arrangement, he states that it allows the geometries to be measured more easily and precisely. Sagitov uses cylindrical masses made of steel and copper. His device for detecting the oscillations of the torsion balance is so arranged that for periods of oscillation in the range of approximately 2000 s to 2300 s, he measures the instantaneous angular deflection every 20 seconds with a relative uncertainty in the time measurement of 1.5×10^{-6} and a resolution of 10^{-5} rad. He thus obtains a reliable picture of the angle vs. time behavior of his oscillation. This knowledge enables him to recognize the nonlinear influences upon the oscillation by comparison of the experimental values with the respective differential equation of the oscillation and to take them into account in the evaluation.

3. Influences on the Measurement Uncertainty

Looking at the measurements which are at present the best, one notices that the relative uncertainties hitherto achieved are in the range of some 10^{-4}. When considering the measurements more closely, one discovers that the uncertainties represent the deviations of the measured values in a given experiment. No publication states an estimate of the possible systematic error due to disturbances of the measurements by influencing parameters. This certainly cannot be explained by the fact that the authors did not think of these but that most of the parameters still exerted an influence upon the measurement although their influence had been eliminated to the greatest possible extent. The influence of the disturbances consists of the interaction of all of the disturbing parameters so that in the end, the overall influence can be described by statistical data at best.

Describing the measuring uncertainty by simple statistical means, however, is allowed only if stationary processes are involved which are independent of time. Measurements for the determination of G are decidedly long-term measurements, some of the individual series of measurement taking several months. Between the measurements, the measuring apparatus is usually reconstructed or modified so that in many cases, a stationary behaviour of the influencing parameters cannot be assumed. An indication of the validity of this objection is given by the measured values themselves: In the reports of many authors, it is noticeable that the mean values and the scatter of various series of measurements are not consistent.

3.1 Inconstancy of the Torsional Moment of Suspension

In the torsion balance experiments hitherto successfully carried out, the balance beam with the masses fastened to it is held by a torsion wire. The torsion wire has thus to fulfil two functions:

(1) It must carry the weight of the torsion balance.
(2) It must supply the torsional moment.

To obtain a very sensitive torsional balance, it is desirable to select a very small wire diameter and a very long wire. However, since it must support the weight of the beam, a lower limit is imposed on the wire diameter by the yield and breaking point of the wire material. In the usual laboratory arrangement, the length of the torsion wire is limited to a maximum of 1 m by the temperature constancy required for the measuring arrangement. In the selection of the wire material, other aspects play an important role, for example, the dependency of the torque upon the magnetic field and temperature. To date, materials such as tungsten, platinum, platinum-iridium, steel, quartz, and carbon have been used. Almost all authors report difficulties encountered with the various wire materials. They report a null drift of the torsion balance, slow changes of the torsional moments, and erratic modifications when relieving or re-loading the wire.

Burgess in 1902 [10, 11] and Tangl in 1926 [59] attempted to find a way out by freeing the wire from the load by using the buoyancy of a body in mercury or water (Fig. 11). Burgess described great difficulties with the mercury surface. With time, impurities in the mercury result in a layer forming on the surface making the meas-

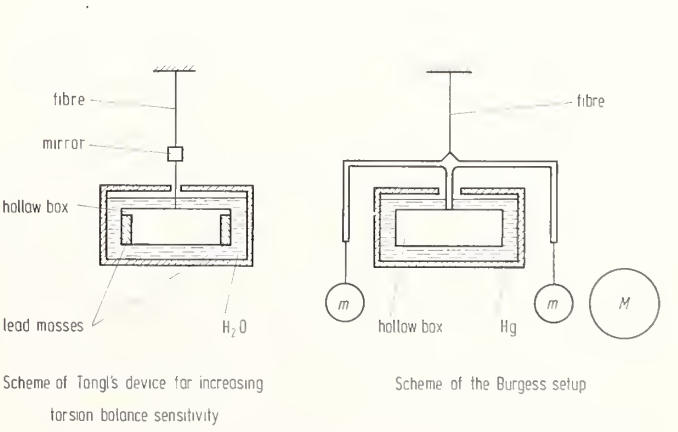


FIGURE 11. *Measurement arrangement with buoyancy bearings [58].*

uring conditions undefinable. Because of the complex frictional conditions of the body floating in the liquid, this method can be used only for static measurements. Another attempt to achieve a suspension free from friction and torque using a magnetic suspension also encountered difficulties. It was not possible to eliminate the influences of torque in the operation of the suspension.

3.2 Disturbances of the Ground

All measurements suffer from disturbances and tilting of the ground, which vary locally and manifest themselves in disturbing oscillations of the torsion pendulum. The frequency range of ground disturbances extends to far below 1 Hz so that a decoupling of the measuring arrangement by mechanical filtering is not possible. Owing to nonlinearities of the oscillating system, oscillations of the balance beam about other axes of inertia can be coupled to the oscillation of interest. One could resort to an attenuation by using suitable gaseous atmospheres, but in dynamic measurements this possibility is precluded by the fact that the frictional behavior and the adsorption of gas on mobile components would result in systematic errors. The influence of ground disturbances can be reduced by using sufficiently long averaging times for the measurements (long compared with the period of the disturbances).

3.3 Influence of Temperature

As in almost all precision measurements, temperature exerts a considerable influence upon the measuring result. Most critical is the variation of geometrical dimensions with temperature. For G measurements in particular, tolerances within the range of 10^{-6} and 10^{-7} m must be complied with and conserved over a prolonged period of time; this requires a temperature constancy over this time of 10^{-2} to 10^{-3} K. Keeping the temperature stable is not the only requirement; there must not be any temperature gradients, either. In measurements with sensitive torsion balances, between the upper and the lower part of the measuring arrangement, for example, temperature gradients of 10^{-1} to 10^{-2} K produce convection currents of the filling gas and, as a result, interfering torques at the level of the torques in the experiment. It might be considered that such difficulties can be avoided only by operating the torsion balance at a reduced pressure.

3.4 Inhomogeneity of the Masses

As is shown in Table 2, the various workers carried out measurements with widely ranging masses. Heyl used the largest masses, 66 kg, and Facy and Pontikis the smallest, 1.5 kg. The advantage of a greater torque on the torsion balance produced by the larger mass, and thus, an improved signal-to-noise ratio between the effective signal and spurious signals due to spurious forces on the torsion balance is, however, cancelled out by the phenomenon that the inhomogeneity (density variation) generally increases with the mass. This results in an uncontrollable systematic error in the measurement result. It is very difficult to obtain information on the density distribution of the masses by experiment—information that could be used later as a correction in the evaluation.

The selection of the masses also calls for compromises. Theoretically, it would be advantageous, for example, to manufacture the masses from materials with a high density such as platinum or tungsten. In practice, it has been shown that knowledge of how to process these materials in order to achieve homogeneous density is insufficient, so the theoretical advantage of the high density cannot be fully utilized for practical purposes.

3.5 Gradients in the Gravitational Field, Changes of the Surrounding Field

In laboratory measurements, it is fundamentally impossible to block off the influences exerted by surrounding masses. In the measuring arrangement area, the gravitational field always shows gradients which in a dynamic measurement result in a disturbance of the torsional oscillation difficult to correct by calculation. As already mentioned, G measurements are long-term measurements with measuring periods sometimes covering several months. Figure 12 clearly shows to what proximity an object coming from infinity can be brought to the measuring arrangement with the influence upon the measurement result remaining smaller than the relative value of 10^{-5}. Figure 13 shows by what distance the position of an object can be changed with respect to the torsion balance so that the influence remains smaller than 10^{-5}.

3.6 Magnetic and Electrical Influences

When setting up a measuring arrangement, it should also be ensured that all the materials used are as free as possible from ferromagnetic impurities which can be

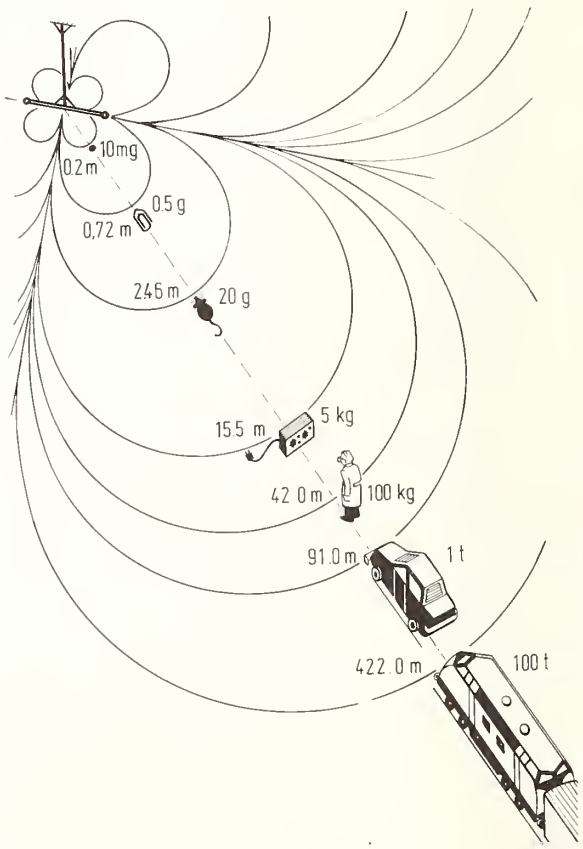


FIGURE 12. *Influence of masses approaching a torsion balance from infinity.*

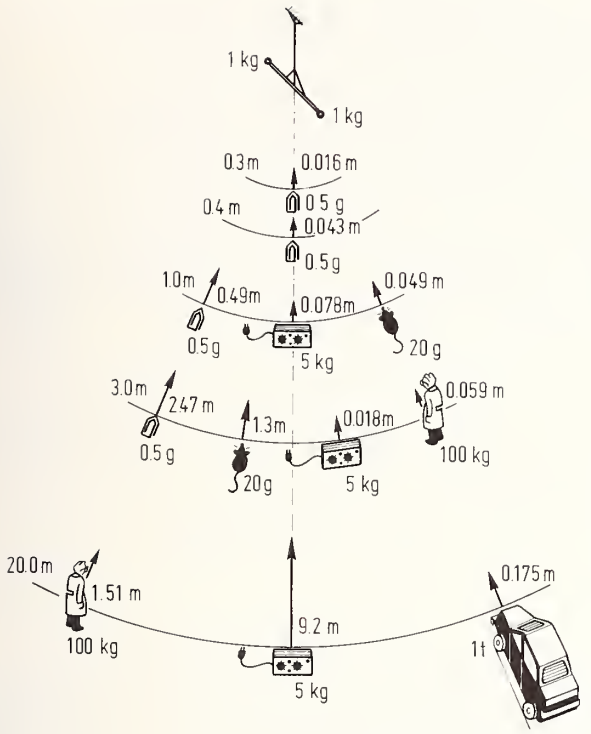


FIGURE 13. *Influence of masses changing their distance relative to the torsion balance.*

caused by unclean tools during production. Particularly in the case of large arrangements, it is difficult to achieve a complete shielding of the magnetic field of the earth. In addition to the magnetic disturbances, electrical disturbances are to be anticipated. Here, electrostatic charges of electrically insulated components and contact potentials can build up if different materials are used to set the arrangement up.

4. Perspectives

In the foregoing, a number of disturbing factors were given which must be taken into consideration when designing a measuring assembly. It is not possible for each disturbing factor to be dealt with independently in order to find a way to eliminate it. This task becomes a problem of optimizing several parameters simultaneously, with the further difficulty that analytical knowledge of the influences of the parameters is very inexact. When the question is asked: "What are the possible limits of an attainable measuring uncertainty?", it is difficult to find an answer. On the basis of his results which he published together with Chrzanowski in 1942 and which he had obtained with a relative uncertainty of 4.7×10^{-4}, Heyl took stock, saying: "The conclusion may be drawn that the limit of the possibility of the torsion balance has been reached". According to estimates which take into account the well-known disturbing influences, it should be possible to attain relative uncertainties in the range of 10^{-5}. To achieve this, it will however be necessary to carry out several tests with different test assemblies to improve our knowledge of the disturbances.

The next question that could be put is: "Are there any suggestions for G determinations which are based on completely new measuring methods and which promise substantial progresses with respect to the expected reduction of the measuring uncertainty?". All experts are obviously in agreement that the laboratory experiments which promise to be successful can be carried out only with torsion balances, whereby a number of good, new ideas would certainly result in a further reduction of the uncertainty in determining G.

I do not know of any new laboratory experiments. Some suggestions have been made by several authors [4, 5, 6, 14, 60]; David Berman, Robert Forward, and John Vinti suggest that G experiments should be carried out in space, that is, on board a satellite in a force-free environment. They suggest two possible measuring methods (Figure 14):

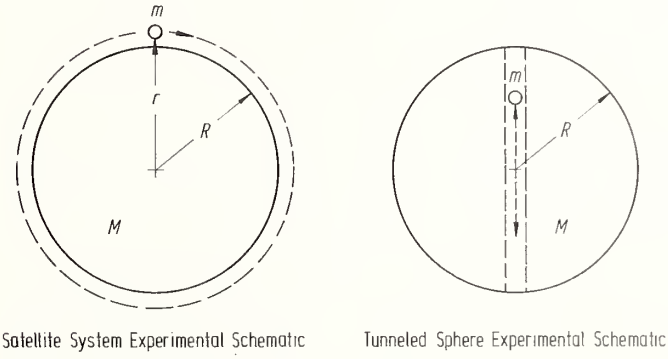


FIGURE 14. *Proposals for G experiments on board force-free satellites [4].*

(1) In the force-free space of a satellite, a small test mass circles around a large mass. G can be determined on the basis of the known masses and the orbit data.
(2) In a tunnel inside a large mass, a small mass oscillates around the center of the large mass. The determination of G is made possible by knowledge of the masses and the oscillation period.

The authors believe that such tests are feasible. Great demands are made, however, on the symmetry of the satellite's mass distribution and internal equipment as well as on a force-free orbit.

5. Other Issues Related to G and Newton's Law

Up to now, we have dealt with laboratory experiments for the determination of G. There are still some other interesting problems which are directly related to the Newtonian equation and thus to G. Let us take a look at these problems which can be defined as follows:

(1) active and passive gravitational mass, inertial mass;
(2) possible deviations from Newton's $1/R^2$ law; and
(3) the time dependence of G.

5.1 Active and Passive Gravitational Mass, Inertial Mass

Fundamentally and in the strictest sense, one may identify three types of mass:

(1) the inertial mass which appears in Newton's second law as the factor of proportionality between the force on and the acceleration of a body;
(2) the passive gravitational mass which is the mass of a body on which a gravitational field acts; and
(3) the active gravitational mass which is the factor of proportionality for the gravitational field emanating from a body.

It is assumed that metrologically, the three properties of a body just described can be defined by the same quantity. In order to prove this, however, it must be shown experimentally that these quantities are in agreement.

In the well-known Eötvös [18] experiment, gravitational and inertial masses of bodies made from different materials are compared with one another by comparing the gravity acting upon them in the earth's gravitational field with the centrifugal force produced by the earth's rotation with the aid of a torsion balance. Eötvös found that the inertial and the gravitational masses are in agreement with a relative uncertainty of 10^{-9}. In an experiment carried out at a later date by Roll, Krotkov, and Dicke [51] who effected measurements by comparing the influence of the gravitational force of the sun and the centrifugal forces due to the earth's revolution around the sun, an agreement between inertial and gravitational mass could be indicated with a relative uncertainty of 10^{-11}.

Tests directly comparing passive and active gravitational mass are rare; I know only of an experiment by Kreuzer [34, 23] (Figure 15). In his experiment, he compares the influence of different materials which have,

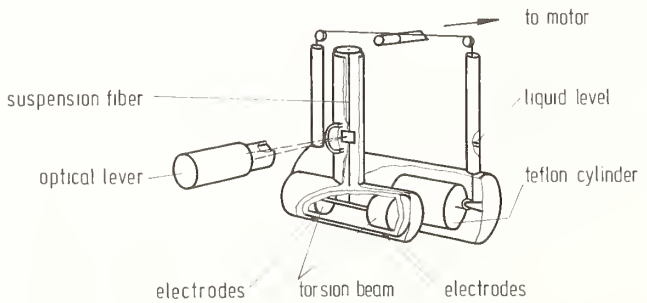


FIGURE 15. *Set-up of Kreuzer to measure the equivalence of active and passive gravitational mass [34].*

however, the same density, on a torsion balance. A Teflon cylinder immersed in a liquid of more or less the same density is moved periodically to and fro and the signal caused by the gravitational force is recorded. By altering the temperature, the density of the liquid can be adjusted to that of Teflon so that—with the measuring signal disappearing and the density of the material remaining the same—proportionality between passive and active gravitational mass can be assumed within the scope of the attained relative measuring uncertainty of 10^{-5}. The result is not as convincing as the comparison between gravitational and inertial mass.

5.2 R-Dependence of the Gravitational Constant

Another problem which has stimulated experimenters in the last few years to carry out relevant experiments is the presumed deviation of Newton's law from the $1/R^2$ form. From the consideration of a general scalar tensor theory of gravitation, Wagoner [61], Fujii [22], and O'Hanlon [24] derived a gravitational potential which is extended by a Yukawa term:

$$V(R) = -\frac{G_\infty m}{R}(1 + \alpha e^{-\mu R}),$$

where G_∞ is the Newtonian gravitational constant at infinite separation, and α and μ are constants. We may thus write

$$F(R) = G(R)\frac{m\,m'}{R^2}$$

with

$$G(R) = G_\infty[1 + \alpha(1 + \mu R e^{-\mu R})].$$

Fujii surmises that $\alpha < 1/3$, and $10\text{ m} < \mu^{-1} < 1000$ m. If this is correct, $G(R)$ would have the value of $G_0 = (4/3)G_\infty$ over the short range (i.e., for small separations as used in the laboratory). In this case, all cosmic mass determinations resulting from assuming $G_\infty = G_0$ would have to be corrected accordingly. In the event, however, of μ^{-1} being smaller than 0.01 m, the G values measured in the laboratory would be in agreement with the value G_∞.

Long's statement for $G(R)$ over the short range is somewhat different [35, 36, 37]:

$$G(R) = G(R_0)[1 + \lambda \ln(R/R_0)],$$

with $\lambda = 0.002$, $R_0 = 1$ cm, R_L = upper limit for validity, and $R < R_L$.

Measurements carried out by Panov and Frontov [42] in 1979 on a torsion balance in which masses of different weight acted upon the torsion balance from a distance of between 0.4 and 10 m did not show any deviation from the $1/R^2$ form within the range of the measuring uncertainty.

Hsou-Tao Yu [29, 30] and others carried out $G(R)$ measurements at a distance of 2 m to 15 m from an oil tank 54 m in diameter and 15 m in height, which was sometimes filled and at other times empty. For these measurements they used a Worden gravimeter. The result they obtained was $G(15\text{ m}) = G_0(1 \pm 0.18)$.

In 1980, Hirakawa [27] and others made measurements in the range of $2\text{ m} < R < 4.2$ m by measuring the dynamic gravitational field of a rotating steel bar using a mass quadropole antenna. They did not find any deviation within their measuring uncertainty.

Spero and others [56] made an attempt to demonstrate a deviation from the $1/R^2$ law by a null method. If the $1/R^2$ law is valid, there must be no gravitational field within an infinitely long hollow cylinder, due to the walls of the cylinder. If this law is deviated from, it must be possible for a gravitational field to be detected which depends upon a lateral shift. In terms of Long's $G(R)$ representation, the limit obtained for λ was $\pm 7 \times 10^{-5}$. This result is at variance with that obtained by Long. However, Long points out that for physical reasons null methods of this kind are perhaps not suitable for finding the deviation from the $1/R^2$ law.

Another experiment which can provide significant information because of its very wide R range is that of Stacey and others who carried out gravitational acceleration measurements at different depths of a mine 950 m deep. On the basis of exact density measurements of the respective spherical shells, they state a value of $G = 6.712 \pm 0.037 \times 10^{-11}$ N·m^2·kg^{-2} which is higher than the laboratory values and tends in the direction of the $1/R^2$ dependence found by Long.

All these measurements show that the measuring uncertainty hitherto achieved does not suffice to give clear answers to the questions posed. It does, however, seem possible to further reduce the measuring uncertainties, so that an answer can be found in the foreseeable future to the physically relevant question of a deviation from the $1/R^2$ force law.

5.3 Time Dependence of the Gravitational Constant

In theories of cosmic development, it is presumed that G is not constant with time. The relative change $\dot{G}/G$ is in the range of some 10^{-11}/year or, accordingly, some 10^{-18}/s.

The proof of changes of G with time of this order requires the availability of processes or measuring devices which are G-dependent and which are stable with respect to all other influences over the necessary observation period, this stability being greater than the anticipated effect produced by G. The attempts to demonstrate a time dependence of G which are the most promising up to now, are based on exact observations of the orbits of planets of the solar system and of the moon [1, 41, 55]. The present state of this field and the measuring uncertainties hitherto attained are discussed by R. C. Vessot, T. C. van Flandern, and R. C. Ritter in these proceedings.

As yet, no successful laboratory tests can be reported. V. B. Braginskii and V. L. Ginzburg [8] have submitted proposals for a new measurement. Provided that the change in gravity behaves in the same way as the change of the gravitational constant with time, they suggest that experiments should be made with special pendulums especially designed for long-term stability. The change of G with time is to be demonstrated in these experiments by means of the resulting change of the oscillation period with time.

Another possibility of determining changes of G within the scope of laboratory tests results from the assumed relationship between a change of the gravitational constant and the formation of new matter which can be detected by sensitive measuring devices. The state of these tests will be reported on by R. C. Ritter and G. T. Gillies [50].

6. PTB Experiment to Determine the Gravitational Constant

In conclusion, I wish to report briefly on the principle on which the determination of G is based in experiments carried out in the PTB. Figure 16 shows the most important design features. We have selected a static measuring method with a torsion balance. The balance beam with the masses fixed to it is supported by a mercury buoyancy bearing whose special design produces a centering effect in the direction of the torsional axis. This design allows the wire suspension to be dispensed with. The torque produced by the gravitational forces is compensated by an electrostatic quadrant electrometer. The masses are cylindrical in shape and are made, at present, of glass. The countermasses are supported by prisms and can be shifted in an axial direction.

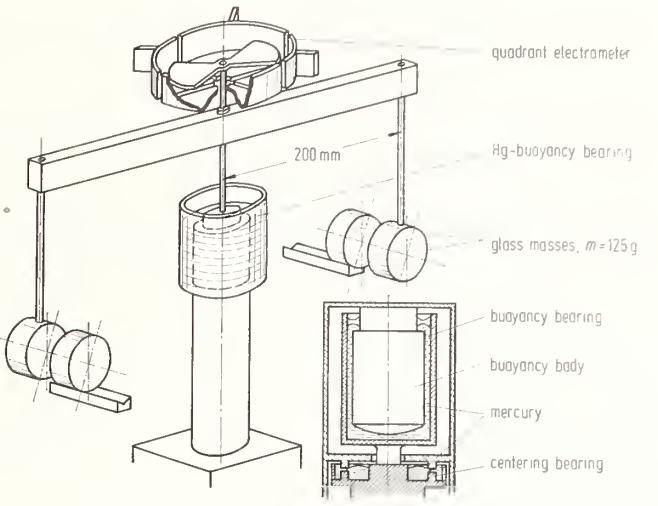


FIGURE 16. *PTB set-up for the determination of the gravitational constant (schematic).*

References

[1] C. O. Alley *et al.*, Science 167, 458 (1970).
[2] F. Baily, Mem. Astron. Soc. 14, 1 (1843).
[3] J. W. Beams, Physics Today No. 5, 35 (May 1971).
[4] D. Berman and R. L. Forward, *Exploitation of Space for Experimental Research* Vol. 24, Science and Technology Series (American Astronautical Society, Tarzana, Calif., 1963), p. 95.
[5] U. Bleyer and R. W. John, Gerlands Beitr. Geophys., 86, 11 (1977).
[6] U. Bleyer, R. W. John, and D. E. Liebscher, Gerlands Beitr. Geophys. 86, 148 (1977).
[7] C. V. Boys, Philos. Trans. R. Soc. London, Ser. A: 186, 1 (1895).
[8] V. B. Braginskii and V. L. Ginzburg, Sov. Phys. Dokl. 19, 290 (1974).
[9] C. Braun, Naturwiss. Rundschau 12, 273 (1897).
[10] G. K. Burgess, C. R. Hebd. Seances Acad. Sci. 129, 407 (1899).
[11] G. K. Burgess, Phys. Rev. 14, 247 (1902).
[12] H. Cavendish, Philos. Trans. R. Soc. London 88, 469 (1798).
[13] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[14] A. H. Cook, Contemp. Phys. 9, 227 (1968).
[15] A. Cornu and J. Baille, C. R. Hebd. Seances Acad. Sci. 75, 954 (1873).
[16] A. Cornu and J. Baille, C. R. Hebd. Seances Acad. Sci. 86, 699 (1878).
[17] R. v. Eötvös, Ann. Phys. Chem. (Leipzig), 59, 354 (1896); Gesammelte Arbeiten, Budapest (1953).
[18] R. v. Eötvös, D. Pekar, and E. Feteke, Ann. Phys. (Leipzig) 68, 11 (1922).
[19] L. Facy and C. Pontikis, C. R. Hebd. Seances Acad. Sci. Ser. B: 270, 15 (1970).
[20] L. Facy and C. Pontikis, C. R. Hebd. Seances Acad. Sci. Ser. B: 272, 1397 (1971).
[21] L. Facy and Pontikis, C. R. Hebd. Seances Acad. Sci. Ser. B: 274, 437 (1972).
[22] Y. Fujii, Nature Phys. Sci. (London) 234, 5 (1971).
[23] J. J. Gilvarry and P. M. Muller, Phys. Rev. Lett. 28, 1665 (1972).
[24] J. O'Hanlon, Phys. Rev. Lett. 29, 137 (1972).

[25] P. R. Heyl, J. Res. Natl. Bur. Stand. 5, 1243 (1930).
[26] P. R. Heyl and P. Chrzanowski, J. Res. Natl. Bur. Stand. 29, 1 (1942).
[27] H. Hirakawa *et al.*, Nature (London) 283, 184 (1980).
[28] F. N. Hooge and J. A. Poulis, Appl. Sci. Res. 33, 191 (1977).
[29] Hsou-Tao Yu *et al.*, Chinese J. Phys. (Taiwan) 16, 201 (1978).
[30] Hsou-Tao Yu *et al.*, Phys. Rev. D 20, 1813 (1979).
[31] Ph. v. Jolly, Ann. Phys. Chem. (Leipzig) 5, 112 (1878).
[32] Ph. v. Jolly, Ann. Phys. Chem. (Leipzig) 14, 331 (1881).
[33] O. V. Karagioz, V. P. Ismaylov, N. L. Agafonov, E. G. Kocheryan, and Yu A. Tarakanov, Phys. Solid Earth 12, 351 (1976).
[34] L. B. Kreuzer, Phys. Rev. 169, 1007 (1968).
[35] D. R. Long, Phys. Rev. D 9, 850 (1974).
[36] D. R. Long, Nature (London) 260, 417 (1976).
[37] D. R. Long, Nuovo Cimento B 55, 252 (1980).
[38] R. A. Lowry, W. R. Towler, H. M. Parker, A. R. Kuhlthau, and J. W. Beams, *Atomic Masses and Fundamental Constants 4*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, London, 1972), p. 521.
[39] G. G. Luther, W. R. Towler, R. D. Deslattes, R. Lowry, and J. W. Beams, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, New York, 1976), p. 592.
[40] D. R. Mikkelsen and M. J. Newman, Phys. Rev. D 16, 919 (1977).
[41] R. R. Newton, J. Geophys. Res. 73, 3765 (1968).
[42] V. I. Panov and V. N. Frontov, Sov. Phys. JETP 50, 852 (1979).
[43] J. H. Poynting, Proc. R. Soc. London 28, 2 (1878).
[44] J. H. Poynting, Philos. Trans. R. Soc. London, Ser. A: 182, 565 (1891).
[45] R. Reich, Ann. Phys. Chem. (Leipzig) 85, 189 (1852).
[46] Ya. Renner, Communications of GAISR, No. 167 (Publ. House of the Moscow University, Moscow, 1970).
[47] Ya. Renner, Determination of the gravitational constant in Budapest, in *Determination of Gravity Constants and Measurement of Certain Fine Gravity Effects*, Ed. by Yu. D. Boulanger and M. U. Sagitov (Nauka Press, Moscow, 1973). [English transl.: NASA-TT-F-15722, 1974.]
[48] F. Richarz and O. Krigar-Menzel, Sitzungsber. K. Preuss. Akad. Wiss. (Berlin), p. 48 (1896).
[49] F. Richarz and O. Krigar-Menzel, Abh. K. Akad. Wiss. Berlin (1898).
[50] C. Ritter *et al.*, Nature (London) 271, 228 (1978).
[51] P. G. Roll, R. Krotkov, and R. H. Dicke, Ann. Phys. (N.Y.) 26, 442 (1964).
[52] R. D. Rose, H. M. Parker, R. A. Lowry, A. R. Kuhlthau, and J. W. Beams, Phys. Rev. Lett. 23, 655 (1969).
[53] M. U. Sagitov, Sov. Astron. 13, 712 (1970).
[54] M. U. Sagitov, Vermessungstechnik 19 Heft 2, 59 (1971).
[55] I. I. Shapiro *et al.*, Phys. Rev. Lett. 26, 27 (1971).
[56] R. Spero, J. K. Hoskins, R. Newman, J. Pellam, and J. Schultz, Phys. Rev. Lett. 44, 25 (1980).
[57] F. D. Stacey, Geophys. Rev. Lett. 5, 377 (1978).
[58] L. Stegena and M. Sagitov, *The Constant of Gravitation* (Lorand Eötvös University. Distribution: Akadémiai Kiadó Budapest, 1979).
[59] K. Tangl, Math. Naturwiss. Anzeiger Akad. Wiss. Budapest 43 (1926).
[60] J. P. Vinti, Celestial Mechanics 5, 204 (1972).
[61] R. V. Wagoner, Phys. Rev. D 1, 3209 (1970).
[62] J. Wilsing, Sitzungsber. K. Preuss. Akad. Wiss. (Berlin), p. 13 (1885).
[63] J. Wilsing, Publ. d. Astrophys. Obs. zu Potsdam 6, 35 (1887).
[64] J. Wilsing, Publ. d. Astrophys. Obs. zu Potsdam 6, 133 (1889).
[65] J. Zahradnicek, Phys. Z. 34, 126 (1933).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Redetermination of the Newtonian Gravitational Constant "G"

Gabriel G. Luther

Center for Absolute Physical Quantities, National Bureau of Standards, Washington, DC 20234

and

William R. Towler

Department of Nuclear Engineering and Engineering Physics, University of Virginia, Charlottesville, VA 22901

The universal Newtonian gravitational constant, G, has been redetermined at the National Bureau of Standards using the time-of-swing method, in which the period of a torsion pendulum in the form of a dumbbell weighing approximately 5 g is modified by the presence of two 10.5 kg tungsten balls. The difference in the squared frequencies with and without the balls is proportional to G. In this experiment the difference was approximately 3%. The apparatus used, albeit with extensive modification, is the same as used in the previous redetermination done at the National Bureau of Standards, using the method of constant angular acceleration. The apparatus is described, improvements are discussed, data is presented and evaluated. The resulting value of G is: $6.6726 \pm 0.0005 \times 10^{-11}\ m^3 \cdot s^{-2} \cdot kg^{-1}$, which is in agreement with the generally accepted CODATA value but with significantly smaller uncertainty.

Key words: gravitation; Newtonian gravitational constant; universal gravitational constant.

1. Introduction

The presently accepted value of G, defined by Newton's equation

$$F = G M_1 M_2 / R^2$$

and listed by Cohen and Taylor [1], was determined mainly from measurements made by Paul R. Heyl and Peter Chrzanowski in 1942 [2]. Of the several attempts to improve this measurement made during the past 39 years [6, 7, 8], none seems significantly to have increased the precision of the internationally accepted value of $G = 6.6720 \pm 0.0041 \times 10^{-11}\ m^3 \cdot s^{-2} \cdot kg^{-1}$.

This work is an extension of the work done at NBS [4] using Beams's method of constant angular acceleration; however, it has been reconfigured to use the time-of-swing method of C. V. Boys [3] and Heyl [2]. The salient modification is the addition of an autocollimator incorporating a diode array, suggested by R. D. Deslattes, which allows for the continuous monitoring of the position of the small mass system. This additional information results in a more precise determination of the frequency, amplitude, noise, drift, and linearity of the torsion constant of the small mass system.

2. Theory

For a torsion balance with moment of inertia I and angular frequency ω_f, supported by a fiber with torsion constant K_f, the angular frequency squared is given by:

$$\omega_f^2 = K_f / I \text{ or } K_f = I \omega_f^2 . \quad (1)$$

If the large masses are placed near this torsion pendulum, a small "fictitious" torsion constant K_g due to the gravitational attraction is added to the torsion constant of the fiber and the frequency of the oscillations becomes

$$\omega_{f+g}^2 = (K_f + K_g)/I \text{ or } (K_f + K_g) = I \omega_{f+g}^2 . \quad (2)$$

Subtracting Eq. (2) from Eq. (1) yields

$$K_g = I(\omega_{f+g}^2 - \omega_f^2) = I \Delta(\omega^2) ,$$

where $K_g = d^2 U / d\theta^2$ evaluated at $\theta = 0$ and U = gravitational potential of the small mass system in the gravitational field of the large masses calculated by integrating

$$U = G M_1 \int \frac{\rho dv}{R}$$

over the volume of the small masses.

The detailed calculations of the potential U of a cylinder in the gravitational field of a spherical mass, are given in reference [9]. The expression for K_g is proportional to G, and is a function of the magnitude of the masses and the geometry of the apparatus. Thus, G can be expressed as

$$G = \Delta(\omega^2) I / k_g \text{ where } k_g = K_g / G .$$

Note that $\Delta(\omega^2)$ is determined by measuring the frequency of the torsion pendulum and that I and k_g are calculated from the measurements of the appropriate dimensions, angles, masses, and densities of the apparatus. The above is true of an undamped harmonic oscillator. However, the frequency of a harmonic oscillator is changed when damped. This frequency change is second order in the damping constant for small values of the damping constant. Because the damping constant in our case is approximately 10^{-4}, the frequency change resulting from this damping is negligible.

Evidence of anharmonicity was sought by looking for the second and third harmonics of the natural frequency. None was found.

3. Apparatus

The apparatus is shown in Fig. 1 and consists of the following:

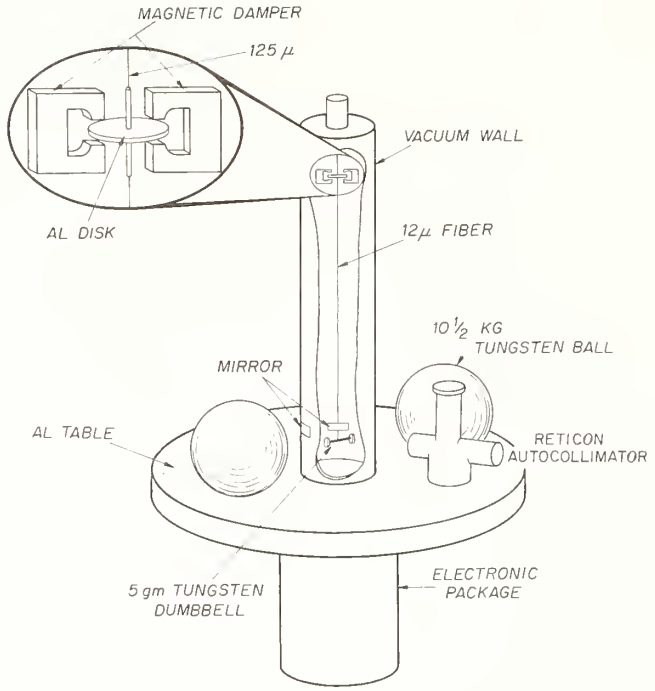


FIGURE 1. *Diagram of the apparatus with insert showing detail of the damper.*

3.1 The Large Masses

The large masses are the same tungsten spheres used by Beams at the University of Virginia and in our earlier measurements [10]. They were fabricated from sintered tungsten by the Y-12 Plant of the Union Carbide Corporation at Oak Ridge, TN. They are 10.165072 and 10.165108 cm in diameter, and their masses are 10.489980 and 10.490250 kg, respectively.

3.2 The Small Mass System

This system consists of two tungsten discs mounted in a dumbbell configuration. The discs are 2.5472 mm thick by 7.1660 mm in diameter mounted on the ends of a 1.0347 mm diameter centerless-ground tungsten rod 28.5472 mm long. The mass of the entire assembly is approximately 7 g.

3.3 Suspension

The small mass system is supported by an approximately 12 μm diameter quartz fiber 40 cm long. The fiber was plated with chromium and gold to make it conductive. The period without the large masses in place is approximately 6 minutes. The change in the period due to the large masses is a few percent.

3.4 Damper

The damper is shown in the insert in Fig. 1. It consists of a small circular aluminum disc suspended in a strong magnetic field approximately 10 cm from the top of the vacuum chamber by a 125 μm diameter quartz fiber. The spindle through the center of the disc serves as the support for the much smaller 12 μm diameter fiber which holds the small masses. The larger top fiber is quite flexible with respect to motions of the small masses in the pendulous mode which are introduced by seismic disturbances. These vibrations are damped in a few seconds. The torsion constant of the larger top fiber is much greater (10^5) than that of the smaller fiber, i.e., the top fiber is rigid, compared to the bottom fiber, with regard to torsion motions: therefore, the damping of the torsion mode is negligible. The tranquilizing effect of the damping is evident in Fig. 2.

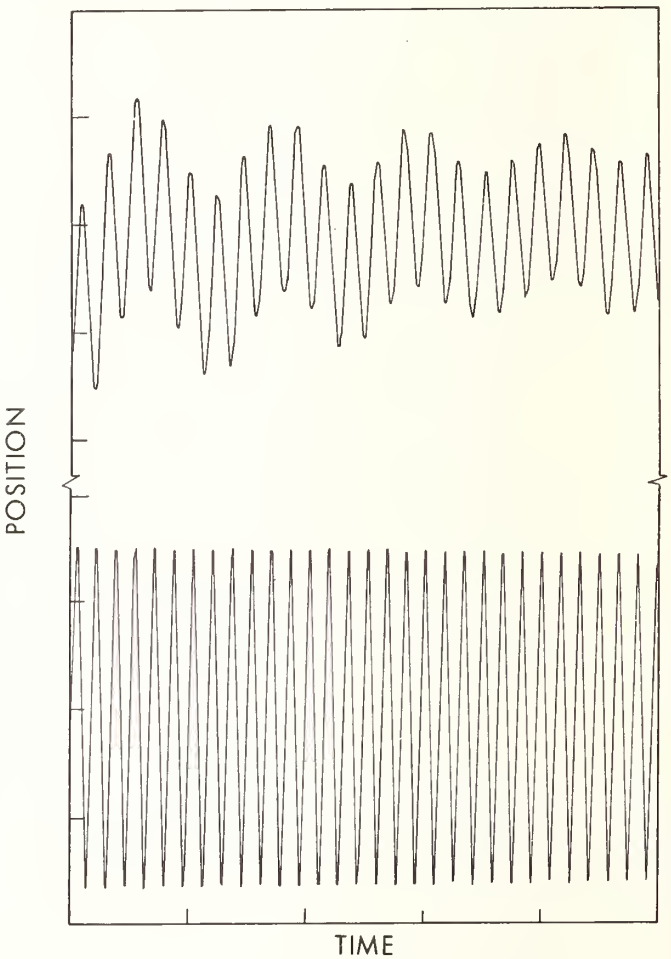


FIGURE 2. *Plot of angular position vs time without and with damping.*

3.5 Detector

The detector consists of a 170 mm focal length autocollimator with a 25 mm aperture in which the light through a 15 μm wide slit is collimated, reflected twice from the mirror attached to the small mass system, and refocussed on a 1024 element diode array. The light-sensitive elements are spaced 16 μm on centers, which gives a range of the autocollimator of 2×10^{-2} radians. The width of the refocussed image of the slit covers 4 to 5 elements of the array.

3.6 Computer

Periodically (every 20 seconds as determined by the internal clock of the computer and checked by a frequency counter), the computer turns the lamp of the autocollimator on for about 300 milliseconds, interrogates the array, and calculates the center of the image group to 1/8 of the spacing between elements. This, the angular position of the small mass system, is then stored.

3.7 Vacuum System

The small mass system, the fiber, and the damper are enclosed in an evacuated brass cylinder which is maintained at a pressure of a few times 10^{-6} Torr by an ion appendage pump. The windows in the chamber are coated with a conductive coat of tin oxide. A rotating vacuum seal through the top of the brass chamber allows positioning of the small mass system relative to the line joining the centers of the large masses.

3.8 Isolation Room

The apparatus is enclosed in an acoustically isolated, thermally controlled cube about 2.5 m on a side. This is mounted on a reinforced concrete slab of about 5000 kg. It is located in a large basement room of the Physics Building at the National Bureau of Standards, Gaithersburg, MD, where the temperature is held constant to within approximately 2 °C. The inner room temperature is controlled to within approximately 0.1 °C.

3.9 Metrology

The large masses are positioned by being placed on three small pads located in wells machined into a monolithic aluminum plate. The separation of the two large masses on the aluminum plate was measured in situ by the 3-D measuring machine of the Dimensional Metrology Group of the National Bureau of Standards, and determined to within 0.3 μm. The consistency of the separation, after removing and replacing the large masses, was found to be better than 0.05 μm rms.

The dimensions of the small mass system were determined initially by the Dimensional Metrology Group and finally by the Gauge Block Group at the National Bureau of Standards. The density of the tungsten in the small mass system was determined by the Length and Mass Measurements and Standards Division of NBS.

4. Data Collection, Reduction, and Results

The strategy adopted involved taking data only on weekends continuously from Friday afternoon until Monday morning, alternating between measurements with the large masses in place and with them removed. Usually, each of these conditions would extend from 6 to 12 hours. The frequencies for all the individual sets (6 to 12 hours) were determined using the algorithm of J. J. Snyder [11]. These frequencies are squared. The midpoint time of the run was assigned to each run. Each set of runs was separated into two groups (masses on, masses off) and a linear least squares fit was made of ω^2 vs time for each group with the constraint that the slopes of the fits be the same. The implication of this procedure is that any underlying drift (slope of the fits) affects the runs in the same way (Fig. 3), i.e., is independent of whether the masses are on or off.

As a check on the linearity of the system, runs were made using different starting amplitudes of the oscillations. The smaller oscillations had a higher percentage noise level (as may be expected). However, no difference in the frequencies using different amplitudes was detected beyond the normal scatter.

The calculation of G was made from the average of the $\Delta(\omega^2)$ value listed in Table 1.

The final assignment of the uncertainty in this determination of G is arrived at by the conventional method of the square root of the sum of the squares of the uncertainties listed in Table 2, and therefore represents a 1σ uncertainty.

The resulting value is $G = 6.6726 \pm 0.0005 \times 10^{-11}$ $m^3 \cdot s^{-2} \cdot kg^{-1}$.

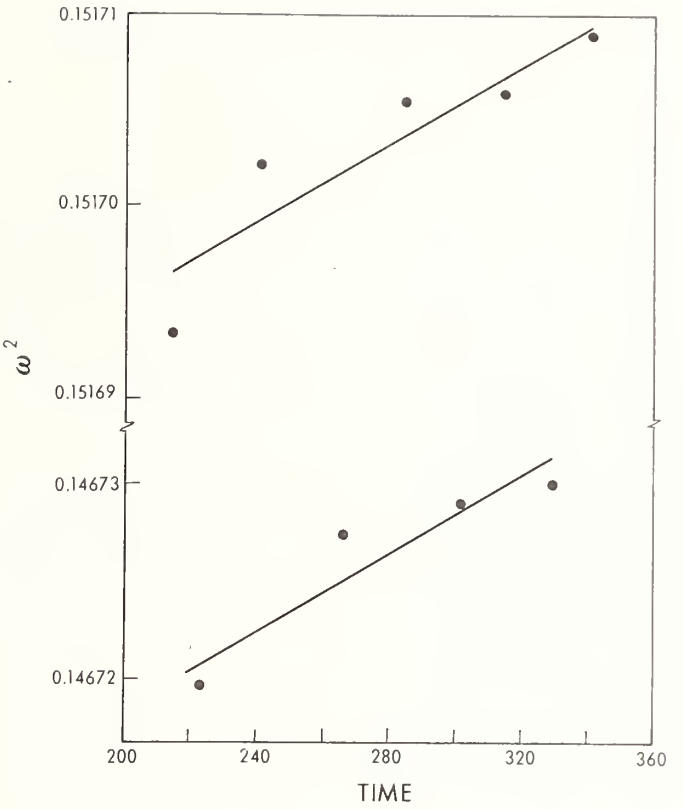


FIGURE 3. *Plot of the least squares fits of the ω^2 vs time constrained to equal slopes. (Note break in the scale for ω^2.) The unit of time is hours; of ω, radians per 20 seconds.*

TABLE 1. *Average value of $\Delta(\omega^2)$*

Date (1980)	$\Delta(\omega^2)$	Orientation
Aug 29–Sept 2	0.00497866	ne-sw
Sept 5–Sept 8	0.00497837	ne-sw
Sept 12–Sept 15	0.00497816	nw-se
Sept 19–Sept 22	0.00497729	nw-se
Sept 22–Sept 26	let up to air, autocollimator modified realigned	
Oct 3–Oct 4	0.00497773	ne-sw
Oct 8–Oct 10	0.00497788	ne-sw
Average:	0.00497801 ± 0.00000020 (40 ppm)	

5. Discussion

The measurement described here yields significantly smaller error bars than previous attempts using this method. A legitimate question is: why is this so? This experiment differs from Heyl's in that it has high vacuum, high Q, high density of the large masses, high data volume, and high damping of the pendulous modes. Some of these factors are interrelated.

TABLE 2. *Error budget*

Source of uncertainty	Uncertainty (ppm)
Position of the large masses	10
Mass of the large masses	1
Length of the small mass	22
Thickness of the small mass	36
Density of the small mass	6
Moment of inertia of mirror	23
$\Delta(\omega^2)$	40
RSS Total	64

Our choice of a quartz fiber which yielded the high Q came as a result of long trial and error. Experience with tungsten fibers indicates that it is difficult to achieve Q's greater than about 100. Also, the tungsten seems to have unpredictable changes in the ambient position of the small mass system and a drift which gets smaller in time, but never disappears entirely. Graphite, on the other hand, has low internal friction (the Q may be several hundred) but its torsion constant, even after being in vacuum for several months, changes by as much as 1% per day, and the ambient position seems to be temperature dependent with a coefficient of about 0.1 rad/°C for the fibers we used (6 μm diam., 40 cm long). The choice of a quartz fiber for its obvious advantages dictated high vacuum in order to exploit effectively the low internal friction of the quartz. This in turn required the installation of the damper.

In any G experiment of this type, the sensitivity is proportional to the density of the material from which the large masses are formed. Tungsten, which has a density of 19, was used in this experiment; earlier experimenters used materials with densities of less than 10.

Heyl did not have the advantage of the sophisticated computer and data acquisition equipment which was available to us. Under the circumstances, it is surprising that Heyl did so well or that we did not do better.

6. Future Improvements

The four largest components of the error budget are the uncertainty in the length and thickness of the discs of the small mass system, the uncertainty of the moment of inertia of the mirrors, and the uncertainty in the measurement of the frequency of the small mass system's oscillation. We are now constructing a new small mass system that will be more amenable to precise metrology than the present one. It is made from tungsten slices lapped flat and parallel to gauge block tolerances (1/10 fringe). These tungsten end pieces will be wrung onto a piece of pyrex, flat and parallel to the same tolerances. The dimensions of the pieces will be determined before assembly and then again after being wrung together in order to determine the quality and thickness of the wringing film. From this sandwich, a cylinder will be formed with approximately the same dimensions as the existing one. A small mass system made in this way will have 90% the sensitivity of the present one with the same mass, yet it should be much more rigid and maintain its dimensions through the metrology.

The mirror and stem of the small mass system will be made much lighter (20% as heavy), thus their contribution to the error will be proportionally less.

Finally, the largest contribution to the overall error is the uncertainty in the measurement of the frequency of the oscillation of the small mass system. This uncertainty will be the most difficult to reduce. In order to minimize this error, fibers which produce different periods will be tried. Experience has shown that the present small mass system, and consequently the new one, may be suspended on fibers of such small diameter so as to produce periods of up to twenty minutes. The frequency change due to the large masses in that case will be about 20%.

It is difficult to imagine doing a measurement of this type without the personnel and facilities of an institution such as the National Bureau of Standards. The following list is by no means a complete list of all the assistance given to us.

Bruce Borchardt measured the separation of the large masses and did the initial metrology on the small masses; Bob Crosson transposed data; Rich Davis and Randall Schoonover performed density measurements; R. D. Deslattes suggested the use of the diode array and participated in fruitful discussions; Jim Filliben and Stephen Leigh provided time series analysis, and general computational assistance; P. Thomas Olsen checked the magnetic susceptibility of the large masses and the material of the small masses; J. J. Snyder supplied the algorithm for computing the frequencies; Clyde Tucker performed metrology on the small mass system.

References

[1] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data **2**, 663 (1973).
[2] P. R. Heyl and P. Chrzanowski, J. Res. Natl. Bur. Stand. **29**, 1 (1942).
[3] C. V. Boys, Phil. Trans. R. Soc. London, Ser. A.: **1**, 1 (1895).
[4] G. G. Luther *et al.*, in *Atomic Masses and Fundamental Constants-5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum, New York, 1976), p. 592.
[5] R. D. Rose *et al.*, Phys. Rev. Lett. 23, 592 (1964).
[6] W. A. Koldewyn, A New Method for Measuring the Newtonian Gravitational Constant, thesis, Wesleyan University, 1976.
[7] C. Pontikis, C. R. Acad. Sci. Ser. B **274**, 437 (1972).
[8] O. V. Karagioz *et al.*, Izv. Akad. Nauk SSSR, Fiz. Zemli **1976** (5), 106 [Izv. Acad. Sci. USSR, Phys. Solid Earth **12**, 351 (1976)].
[9] W. R. Towler *et al.*, in *Precision Measurement and Fundamental Constants*, Ed. by D. N. Langenberg and B. N. Taylor, Natl. Bur. Stand. (U.S.), Spec. Publ. 343 (Aug. 1971), p. 485.
[10] J. H. Nash *et al.*, "High-Density Tungsten Spheres," Union Carbide Corporation Y-12 Plant Document Y-1654.
[11] J. J. Snyder, Appl. Opt. **19**, 1223 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Design of a Beam Balance for a Determination of *G*

C. C. Speake and A. J. F. Metherell

Cavendish Laboratory, Madingley Road, Cambridge CB3 OHE, England, U.K.

We are currently engaged in the design and construction of a beam balance which will be used for a precision determination of G. The balance is designed to give a value of G to about one part in 10^5. In place of knife edges the balance employs a cross-strip pivot for its fulcrum and flexure pivots are used to suspend the test masses from the ends of its arms. The balance is servo-controlled so that it maintains its position in the horizontal plane to an angle of about 10^{-9} rad. The attracting masses are in the form of rectangular slabs with rectangular holes at their centers. This shape ensures that the test masses suspended from the arms of the balance lie in regions of space where the gravitational field due to the slabs is highly uniform. A precise knowledge of the center-of-mass to center-of-mass distances is therefore unnecessary for a precision determination of G with this device.

Key words: beam balance; cross-strip pivot; Newtonian constant of gravitation; precision measurement.

1. Introduction

Apart from the very earliest measurements of G, which depended on finding the difference between the astronomical and geodetic latitudes on the north and south sides of a suitably surveyed mountain, all other measurements of the constant of gravitation have involved the use of either the torsion balance or the beam balance. The torsion balance has been used either in the form of a deflection instrument, as, for example, in the work of Cavendish [1], Boys [2], and Rose *et al.* [3], or in the form of a torsion pendulum, as exemplified by the experiments of Heyl [4] and Zahradnicek [5]. This versatility has meant that the torsion balance has played a dominant role in the measurement of G, whereas the use of the beam balance has been reported only three times in the literature [6], [7], [8], and these experiments were conducted over a period of less than twenty years towards the end of the 19th century. There are, however, no physical reasons for supposing that the beam balance is intrinsically less accurate than the torsion balance in detecting and measuring the small forces encountered in a determination of G. It is clear, given all the advantages of modern technology, that a great deal of scope exists to up-date and improve the beam balance method. We are currently involved in the construction of a beam balance designed for a precision measurement of G; the purpose of this paper is to describe the salient features of our method and to report on the progress so far made.

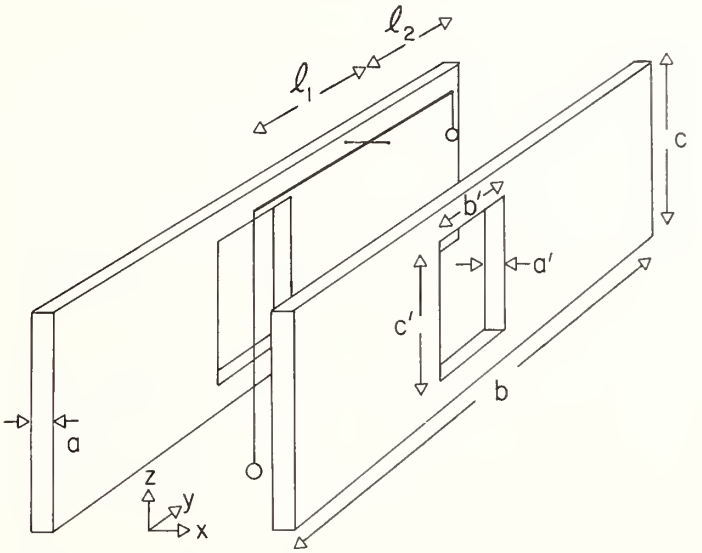


FIGURE 1. *Schematic arrangement of balance and attracting masses.*

2. The Principles of the Beam Balance Method

The balance and its associated masses are shown schematically in Fig. 1. The arm lengths ℓ_1 and ℓ_2 (Fig. 2) are nominally equal with $\ell_1 = \ell_2 \simeq 0.3$ m and the test masses m_1 and m_2 suspended from the ends of the arms are also nominally equal with $m_1 = m_2 \simeq 1$ kg. The attracting masses are in the form of a pair of rectangular parallelepiped slabs with linear dimensions $a \simeq 0.05$ m, $b \simeq 2.0$ m, $c \simeq 1.0$ m (Fig. 1) and separated by a distance $d \simeq 1.0$ m as measured from their geometric centers.

Each slab has a rectangular hole cut in it with dimensions $a' = a$, $b' = 0.175\,b$ and $c' = 0.7\,c$ thus ensuring that m_1 and m_2 sit in regions of space where the z components of the force possess extrema in the x, y, and z directions which are coincident [9]. The abandonment of the traditional spherical shape for these masses in favor of the more complicated mass distributions chosen here means that the center-of-mass to center-of-mass distances between the test masses and attracting masses are no longer quantities which have to be determined with great precision. The slabs produce gravitational fields in the regions of space near m_1 and m_2 which are highly uniform and hence these test masses do not have to be positioned with any great accuracy. Indeed our calculations show [9] that for a positioning error in m_1 or m_2 as large as 0.1 mm G can be determined to within a few parts in 10^7 provided that all other sources of error are negligible.

There is no need to remove the slabs from the vicinity of the balance in order to reduce the gravitational torque acting on it to zero, as is required for calibration purposes, since this can be achieved by the simple expedient of moving the slabs vertically through an appropriate distance. This arrangement therefore avoids problems associated with floor tilt since the mass distribution on the floor remains unaltered as the slabs are raised.

The effect of the rectangular slabs is to produce a force F_{G1} which acts upwards on m_1 and a force F_{G2} which acts downwards on m_2 and the resultant gravitational couple on the balance is cancelled by applying a voltage to an electrostatic transducer of the cylindrical capacitor type described by Durand [10]. This transducer produces the force F_{E1} shown in Fig. 2. The transducer voltage is used to servo-control the balance and maintain it in a horizontal position throughout the measurements. This arrangement facilitates error analysis because the voltage readings can be read out continuously during each experimental run.

When the gravitational couple is balanced by the electrostatic couple we have

$$F_{E1}\,\ell_1 = F_{G1}\,\ell_1 + F_{G2}\,\ell_2$$

$$\text{or } F_{E1} = F_{G1} + F_{G2}\frac{\ell_2}{\ell_1} = G\left\{\chi_1 m_1 + \chi_2 m_2 \frac{\ell_2}{\ell_1}\right\},$$

where G is the constant of gravitation and χ_1 and χ_2 are geometrical factors which depend on the positions and shapes of the attracting masses used in the experiment. The electrostatic force F_{E1} can be determined from the geometry of the transducer, its capacitance, and the applied voltage, leaving only the ratio ℓ_1/ℓ_2 to be determined. The factors χ_1 and χ_2 are not sensitive to the positions of the test masses m_1 and m_2 so that their positions do not form a part of the precision measurements required to determine G.

The ratio ℓ_1/ℓ_2 does however have to be determined to a high precision and the way this measurement is achieved is as follows: The rectangular slabs are raised into the position of zero torque (see above) and a second transducer is used to apply a force F_{E2} to the mass m_2 which is then balanced by the first transducer. The result is that the ratio F_{E1}/F_{E2} immediately gives ℓ_1/ℓ_2. Again servo-control is employed so that the ratio ℓ_1/ℓ_2 is monitored without altering the horizontal position of the balance.

The quantities which have to be measured with high precision are (a) the masses m_1 and m_2, (b) the masses of the rectangular slabs, (c) the dimensions of these slabs, (d) the dimensions and capacitances of the electrostatic transducers, and (e) the voltages applied to these transducers. To determine G to one part in 10^5 requires all these quantities to be determined to within a few parts in 10^6 and none of the above represents any special difficulties in this respect. The quantities which do not need to be known with any real precision are the positions of the test masses relative to the rectangular slabs. This should represent a real and marked improvement on all other determinations of G so far reported in the literature since all have required a precise knowledge of the center-of-mass to center-of-mass distances between the test masses and the attracting masses involved.

3. The Mechanical Design of the Balance

The overall length of the beam balance is restricted to about 0.6 m as a compact balance is a necessity when one is faced with the problem of controlling the temperature of its environment accurately and eliminating as far as possible undesirable thermal effects caused by temperature gradients. To this end the balance arm is constructed from carbon fiber reinforced plastic which has a very low thermal expansion coefficient ($-0.5 \times 10^{-6}\,\mathrm{K}^{-1}$ in the direction of its length). This material has the added advantages of low density, extremely high rigidity and meets the essential requirement of being nonmagnetic. The balance, together with its test masses, is to be operated inside a high-vacuum chamber to avoid problems arising from atmospheric convection currents and Brownian noise. In addition to insulating the balance as far as possible from thermal noise it may be necessary to float the apparatus because mechanical noise will manifest itself in motions of the horizontal reference plane defined by the fixed capacitor plates (Fig. 2). Noise at the pivot does not however couple to the rotation of the beam and hence sophisticated anti-vibration mountings are not required.

To avoid any problems arising from mechanical hysteresis, the arms of the balance will be allowed to rotate by only a very small amount (very much less than 10^{-9} rad.) This will be achieved by the use of servo-control to maintain the arms of the balance in the horizontal position.

The balance employs a cross-strip pivot (Fig. 2) instead of the conventional knife-edge for its fulcrum. The upper plate A (Fig. 2) of the pivot is held rigidly to the laboratory bench and the balance arms are attached to the lower plate B. The crossed-leaf springs of the pivot hold A and B together but allow plate B to rotate about the horizontal axis passing through the point P marked in Fig. 2.

Detailed studies of this device have been published by Haringx [11]. As far as we have been able to determine its mechanical properties make it ideally suited for our purposes and some of its more important properties are as follows: Firstly, its operation depends only on the elastic deformation of the cross-springs, thus completely eliminating the problem of sticking which ultimately limits the sensitivity of a knife-edge fulcrum. Secondly, provided the angular rotation of the beam is small (as in our case) the axis of rotation is particularly well defined [12].

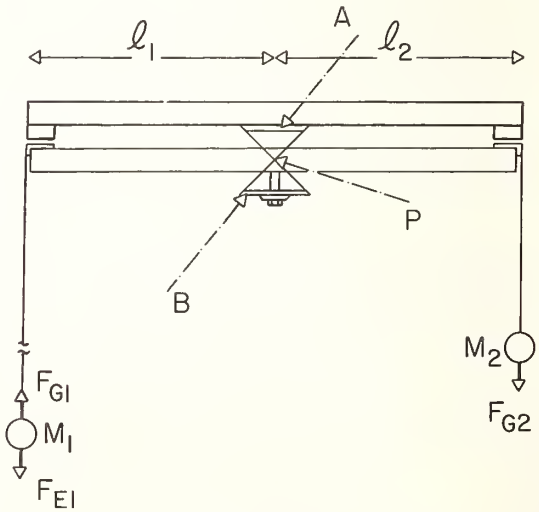


FIGURE 2. *The beam balance.*

Thirdly, when a tension is applied to the pivot, the restoring couple per unit angle of rotation decreases in proportion to the applied load until it reaches zero and the beam becomes unstable. By choosing appropriate values for the masses m_1 and m_2, the natural period of the device can therefore be made very large and the balance very sensitive.

The cross-strip pivot, which at present employs Be-Cu strips for its elastic members, does however suffer from the following problem: It is obviously desirable to perform all measurements at room temperature and in order to measure forces substantially smaller than those corresponding to thermal motion at room temperature it appears that the balance should have a long relaxation time [13]. Measurements made with angles of rotation of about 10^{-3} rad indicate that the damping due to internal friction in the Be-Cu strips is too large for our purposes. However for the very small oscillations the beam is designed to cope with ($\sim 10^{-9}$ rad), the damping parameter might be very much smaller [14] than the value obtained at an amplitude of 10^{-3} rad. We are currently extending our measurements to these very much smaller amplitudes and if the damping still proves to be too large our intention is to replace the Be-Cu strips with elastic members fabricated from fused silica.

The test masses are suspended from the rigid arms of the balance by means of flexure pivots (Fig. 2). The axis of rotation of this pivot is well defined provided that the tension of its elastic member is kept constant [15]. In our case m_1 and m_2 are fixed and the tension is therefore kept constant to an extremely high level. We are confident that with fixed loading and with proper temperature control of the balance's environment the arm lengths ℓ_1 and ℓ_2 will remain constant during any given experimental run to a very high degree of precision as is required for a measurement of G by this method.

4. Electronic Stabilization of the Balance

The electronic control of the balance is indicated schematically in Fig. 3. A differential capacitance transducer of the type described by Jones and Richards [16] is used to detect the motion of the beam. The off-set current produced by an imbalance between two small sensor capacitors flows from the center tap of the secondary windings of the bridge transformer (hand wound, ferrite pot core) through another transformer which acts as an input-impedance matcher for a low-noise amplifier. The beam is earthed. The oscillator produces approximately 10 V peak to peak at a frequency of 20 kHz. The control system exerts precisely zero force on the beam when it is exactly balanced. The polarity of the dc voltage output of the phase-sensitive detector depends on the sense of displacement of the beam relative to the horizontal plane. The micro-computer returns voltages to the electrostatic transducer which applies the force F_{E1} (Fig. 2) in such a way as to critically damp the motion and return the beam to its horizontal position. It is very likely that the electronic system is more than adequate for our requirements and we hope to restrict the motion of the beam to angles less than about 10^{-9} rad.

5. Summary

With the device described in this paper it is hoped to determine G to an accuracy of about 1 part in 10^5. The measurements will ultimately involve us in a comparison with the kilogram, the meter, and the volt, and our aim is to design a beam balance which will yield a value for G whose accuracy is limited to the accuracy with which these three comparisons can be made. Our design studies indicate that the problems associated with achieving this aim, although considerable, are not insurmountable.

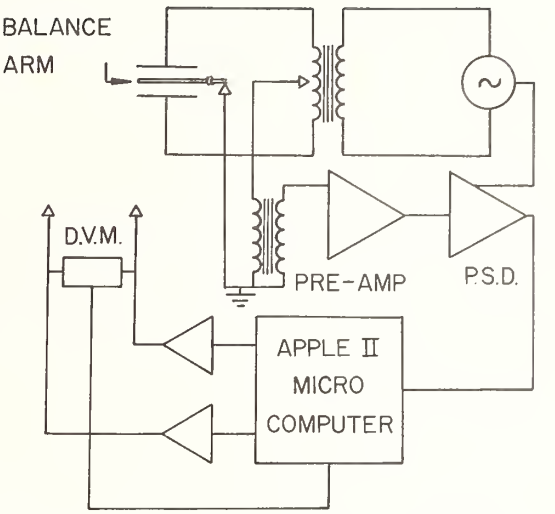


FIGURE 3. *The electronic control system.*

The authors are indebted to Professor A. H. Cook, FRS for many useful discussions and for his continued interest and encouragement.

References

[1] H. Cavendish, Phil. Trans. R. Soc. London 18, 469 (1798).
[2] C. V. Boys, Phil. Trans. R. Soc. London, Ser. A: 186, 1 (1895).
[3] R. D. Rose *et al.* Phys. Rev. Lett. 23, 655 (1969).
[4] P. Heyl, J. Res. Natl. Bur. Std. (U.S.) 5, 1243 (1930).
[5] V. J. Zahradnicek, Phys. Z., 34, 126 (1933).
[6] J. H. Poynting, Philos. Trans. R. Soc. Ser A: 182, 565 (1891).
[7] Ph. Von Jolly, Abh. Bayer. Akad. Wiss. (1898).
[8] F. Richarz and O. Krigar-Menzel, Anhang zu den Abhandlung (1898).
[9] A. J. F. Metherell *et al.*, these proceedings.
[10] E. Durand, *Electrostatique* (Masson et Cie, Paris, 1966), Vol. 2, p. 227.
[11] J. A. Haringx, Appl. Sci. Res. Sect. A: 1, 313 (1947).
[12] R. V. Jones, J. Sci. Instrum. 193, 39, (1962).
[13] V. B. Braginsky and A. B. Manukin, *Measurement of Weak Forces in Physics Experiments* (University of Chicago Press, 1977).
[14] R. V. Jones, private communication 1981.
[15] F. S. Eastman, Bulletin, University of Washington, Engineering Experiment Station Series Bulletin No. 86 4, (1933).
[16] R. V. Jones and J. C. S. Richards, J. Phys. E 6, 589, (1973).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Optimizing the Shape of the Attracting Mass in Precision Measurements of *G*

A. J. F. Metherell, C. C. Speake, and Y. T. Chen

Cavendish Laboratory, Madingley Road, Cambridge CB3 OHE, England, U.K.

and

J. E. Faller

Joint Institute for Laboratory Astrophysics, University of Colorado, Boulder, CO 80309

It is now recognized that the sphere is not necessarily the best shape to use for an attracting mass in a precision measurement of *G*. Two shapes which minimize the effect of errors in the center-of-mass to center-of-mass distance between the attracting mass and the test mass are (a) the right-circular hollow cylinder and (b) a pair of rectangular parallelepiped slabs with rectangular holes at their centers. The kinematic factors influencing the design of these mass distributions are considered and the optimum design parameters that should be employed with attracting masses of these shapes are determined.

Key words: attracting masses in gravitation; constant of gravitation, *G*; gravitation; gravitational attraction; Newtonian constant of gravitation; Newtonian gravitation; optimum mass geometries.

1. Introduction

Since the first precision measurement of G was reported by Cavendish [1] in 1798 a further thirty or so determinations have appeared in the literature. In nearly all these cases a spherical shape was chosen for the attracting mass and only during the last decade has it been recognized [2, 3, 4] that this shape is not necessarily the most suitable for a precision measurement of G. The major drawback of the sphere is this: If the center-of-mass distance between the attracting and attracted masses is uncertain by an amount Δr, then to the first order in small quantities, the error ΔF_r in the measured force F_r between these masses is given by $|\Delta F_r/F_r| = 2\Delta r/r$. The ultimate precision to which G can be determined with a spherical attracting mass is therefore limited by the accuracy Δr to which its center of mass can be located.

Recently it has been recognized [2] that the field along the axis of a hollow cylinder (Fig. 1a) possesses extrema (Fig. 1b) in the regions of space close to the end faces of the cylinder and that the quantity $|\Delta F_z/F_z|$ for $z \simeq z_m$ vanishes to the first order of small quantities. Clearly this property is not unique to the cylindrical shape; almost any solid with a hole cut through it will possess small regions of space in which the gravitational intensity is highly uniform. Once the general shape of the attracting mass has been decided on, it is obviously desirable to optimize its design to ensure that the regions of uniform intensity are as large as possible, which means that the extrema should be as flat as possible.

In this paper we report on an investigation aimed at finding the optimum design parameters for attracting masses of two quite different shapes; these being (a) the right-circular hollow cylinder and (b) a pair of parallel rectangular parallelepiped slabs with rectangular holes at their centers as illustrated by Fig. 3. The former has been used by Long [4] in a test of the inverse square law and the latter are to be used in a beam-balance method [5] for determining G. Although our investigations have included a study of the effects of density inhomogeneities and machining errors on the ultimate precision to which G can be determined when attracting masses of the above shapes are employed, in the interests of brevity the results of these studies will not be included in this paper but instead will be published elsewhere [6]. Here we will concern ourselves exclusively with the kinematic factors which determine the optimum design parameters; this means that throughout this paper it is assumed that the

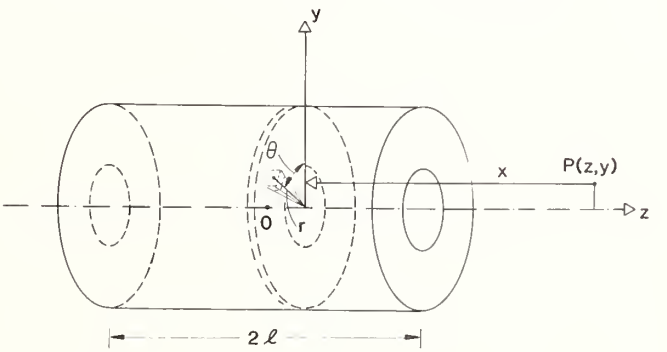


FIGURE 1a. *The right circular hollow cylinder.*

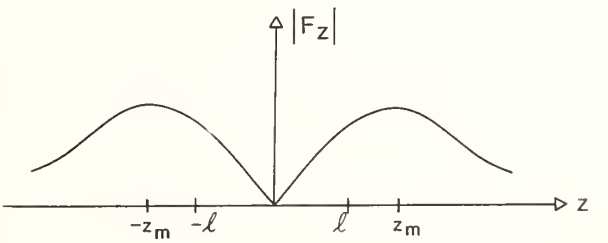


FIGURE 1b. *The variation of the axial component of the gravitational field with the axial coordinate z.*

materials from which the attracting masses are made are of constant density, and furthermore that these masses do not suffer from any machining errors.

2. The Right-Circular Hollow Cylinder

2.1 The Gravitational Field of a Hollow Cylinder

A closed analytical expression for the gravitational field of a hollow cylinder does not appear to have been published anywhere in the literature. Heyl [7] has published numerical values for the coefficients of the first eleven terms of a zonal harmonic expansion of the potential function of a solid cylinder. The resulting formula for the gravitational field is of open form and cumbersome to say the least. In principle the problem of finding the potential function for either a hollow or solid cylinder in closed form is simple enough; in practice however the calculation is lengthy and tedious and a full treatment of this problem is to be published elsewhere [8]. However, since the general solution to this problem is one of some importance we summarize the results of the calculation by quoting in the appendix at the end of this paper the relevant expressions for the field of a hollow cylinder.

2.2 The Paraxial Approximation

The expressions for the field intensities are very much simplified if the point (z,y) lies close to the axis of the cylinder. In this case the axial component of the intensity can be obtained by first noting that at the point (z,y)

$$F_z = -G\rho \int_{z-\ell}^{z+\ell} \int_{r_1}^{r_2} \int_0^{2\pi} \frac{xr}{(x^2 + r^2 + y^2 - 2ry\cos\theta)^{3/2}}\, d\theta\, dr\, dx$$

where ρ is the density of the cylinder and r, θ, and x are the coordinates indicated in Fig. 1a. On expanding the denominator of the integral and neglecting terms involving powers of y higher than the second we obtain after integration

$$F_z = 2\pi G\rho \left[A - B - C + D - \frac{y^2}{4}\left\{r_2^2\left[\frac{1}{A^3} - \frac{1}{B^3}\right] + r_1^2\left[\frac{1}{D^3} - \frac{1}{C^3}\right]\right\}\right]$$

where A, B, C, and D are the distances indicated in Fig. 2, Q being the axial point with the same value of z as the point P. This equation is a good approximation to F_z when y is small and is of course exact for all axial points ($y = 0$).

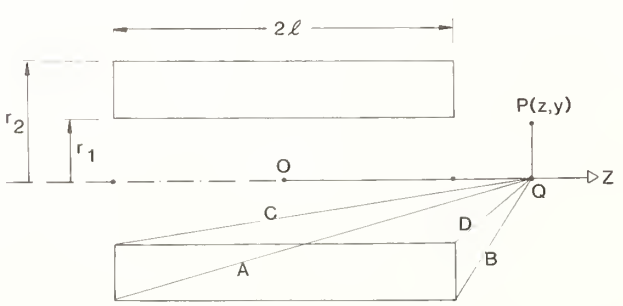


FIGURE 2. *A cross-section of a hollow cylinder showing the distances A, B, C, and D of the paraxial approximation.*

2.3 The Optimum Shape for the Cylinder

A detailed analysis [6] based on the paraxial approximation shows that in the region of the point $(z_m, 0)$, which lies on the axis of the cylinder and is the point where $|F_z|$ is a maximum in the z direction, the force $|F_z|$ is a minimum in the radial direction. The point z_m is therefore a stationary point in both the axial *and* the radial directions and hence the field is constant to the first order of small quantities for small displacements in any direction about z_m. Furthermore, the volume of space over which the field is effectively constant increases in size [6] if either the length of the cylinder is increased or if the inner and outer radii are allowed to approach one another. The optimum design for an attracting mass of this shape is therefore one for which the length of the cylinder is made as long as possible and the wall thickness as thin as possible, compatible of course with the requirement that sufficient mass of material be present in the cylinder to produce the required gravitational attraction. Finally it is worth noting as the length of the cylinder is increased the point z_m moves inwards towards the face of the cylinder until in the semi-infinite limit it lies in the plane of the face.

3. Rectangular Slabs

We are currently involved in the design and construction [5] of a beam balance which will be used for a precision measurement of G. The arrangement of masses used with this device is illustrated schematically in Fig. 3, where A_1 and A_2 are attracting masses in the form of a pair of parallel rectangular parallelepiped slabs with holes cut in them as shown in the figure. The gravitational couple of the slabs on the balance arms is zero whereas a couple is exerted on the test masses m (Fig. 3). The balance is servo-controlled and its position

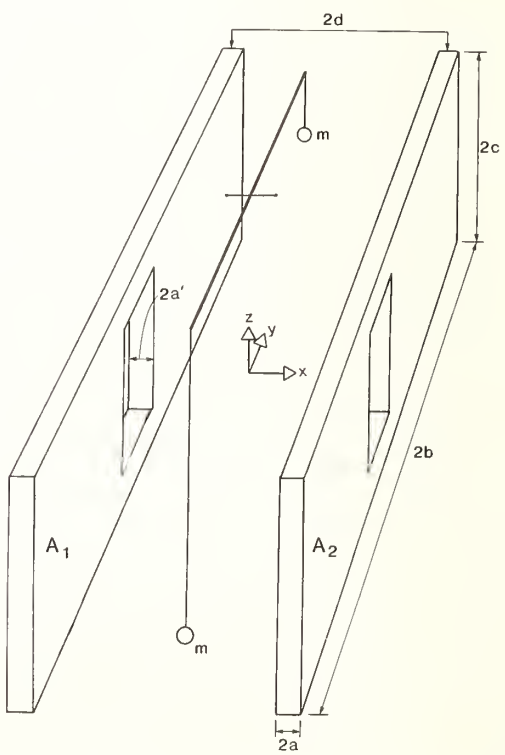


FIGURE 3a. *A schematic diagram of the beam balance and its associated masses.*

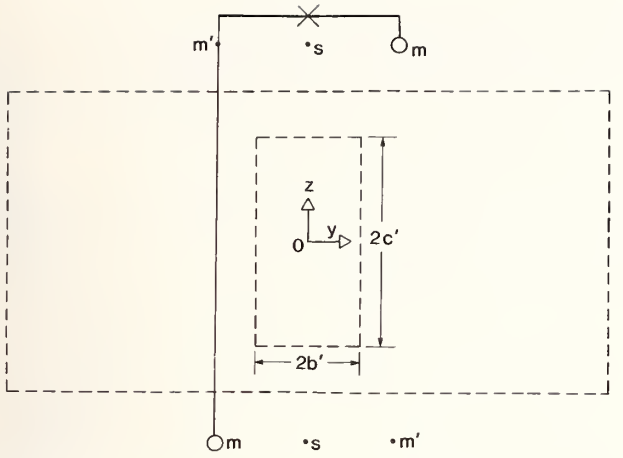


FIGURE 3b. *A cross-section in the plane $x = 0$ containing the balance.*

in the horizontal plane is maintained by applying an equal and opposite electrostatic couple (see Ref. [5] for further details).

3.1 The Gravitational Field of the Rectangular Slabs

The potential function of a homogeneous rectangular parallelepiped can be found in Ref. [9] and it is a simple matter to use the expression given there to determine the field at any point (x, y, z) in the vicinity of the slabs illustrated in Fig. 4. The component of force F_{x_n} in the Cartesian coordinate direction x_n is given by

$$F_{x_n} = (F_{x_n})_1 - (F_{x_n})_2 + (F_{x_n})_3 - (F_{x_n})_4 = \sum_{m=1}^{4} (-1)^{m+1}(F_{x_n})_m$$

where

$$(F_{x_n})_m = G\rho\left[\left\{\sum_{l=n-1}^{n}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}(-1)^{i+j+k}((x_l - (-1)^i a_l)\delta_{l+1,n} + (x_{l+1} - (-1)^j a_{l+1})\delta_{l,n}) \times \ln(\epsilon^l_{i,j,k} + \epsilon^l_{i,j,k+1} + (-1)^k 2a_{l+2})\right\} + \sum_{k=1}^{2}(x_n - (-1)^k a_n)\sum_{i=1}^{2}\sum_{j=1}^{2}(-1)^{i+j+k} \times \tan^{-1}\left[\frac{(x_{n+1} - (-1)^i a_{n+1})(x_{n+2} - (-1)^j a_{n+2})}{(x_n - (-1)^k a_n)\epsilon^n_{k,i,j}}\right]\right].$$

In this expression $\delta_{n,l}$ is the Kronecker delta and

$$\epsilon^p_{q,r,s} = [(x_p - (-1)^q a_p)^2 + (x_{p+1} - (-1)^r a_{p+1})^2 + (x_{p+2} - (-1)^s a_{p+2})^2]^{1/2},$$

it being understood that

$$x_4 \equiv x_1 \equiv x + d \text{ if } m = 1 \text{ or } 2, \text{ or } x - d \text{ if } m = 3 \text{ or } 4$$
$$x_5 \equiv x_2 \equiv y$$
$$x_6 \equiv x_3 \equiv z,$$

where x, y, z, and d are defined in Fig. 3a, and it being understood that

$$a_4 \equiv a_1 \equiv a \text{ if } m = 1 \text{ or } 3, \text{ or } a' \text{ if } m = 2 \text{ or } 4$$
$$a_5 \equiv a_2 \equiv b \text{ if } m = 1 \text{ or } 3, \text{ or } b' \text{ if } m = 2 \text{ or } 4$$
$$a_6 \equiv a_3 \equiv c \text{ if } m = 1 \text{ or } 3, \text{ or } c' \text{ if } m = 2 \text{ or } 4,$$

where a, b and c are also defined in Fig. 3a.

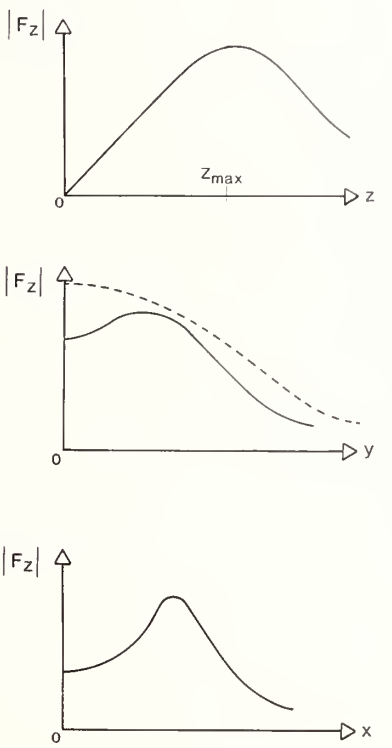


FIGURE 4. *The variation of $|F_z|$ with z, y, and x. The broken curve represents the variation obtained with solid slabs whereas the full curves represent those obtained with the slabs of Fig. 3a.*

3.2 The Optimum Shape for the Rectangular Slabs

Insensitivity to position of the test masses can be achieved provided that stationary points exist for the z component (F_z) of the gravitational field (the x and y components produce zero torque)[1] and provided that these points can be made coincident with the positions of the masses m (Fig. 3) which lie in the plane $x = 0$. The variation of $|F_z|$ as x, y, and z are varied is illustrated schematically in Fig. 4. If solid slabs (that is slabs without rectangular holes in them) are used then $|F_z|$ possesses two stationary points (S in Fig. 3b) in the plane $x = 0$ and these points have extrema in all three coordinate directions. For slabs with rectangular holes of suitable dimensions there are four stationary points (situated at m and m' in Fig. 3b) in the plane $x = 0$ and these also possess extrema in all three coordinate directions. Slabs which possess this property are the ones most suitable for the beam-balance experiment.

[1] A detailed analysis shows that when, as in our case, the test masses are pivoted, the torque due to the y components is not identically equal to zero but is nevertheless negligible; not only are the torques due to the y components acting in opposition and therefore subtractive but they also vanish to the first order of small quantities.

To determine the optimum shape of the rectangular slabs the following procedure was adopted: Firstly the dimensions of the slabs were chosen arbitrarily in the sense that a, b, c, and d (Fig. 3) were fixed by having regard only to the overall size of the apparatus. Secondly the dependence of $|F_z|$ on the three coordinate directions was investigated (by numerical methods) as a function of the dimensions a', b', c' of the rectangular hole until the optimum shape was determined. We have found that for slabs with dimension $a = 0.025$, $c = 0.5$, $d = 0.5$, relative to $b = 1.0$, the rectangular hole should have $a' = a$, $b = 0.175b$ and $c' = 0.35b$. With these dimensions the stationary points occur at $x = 0$, $y = \pm 0.2739b$, $z = \pm 0.6677b$, these being the positions at which the test masses should be placed if maximum insensitivity to positioning errors is to be achieved.

The work on the hollow cylinder described above was done in 1973/4 while one of us (AJFM) was a Visiting Fellow at the Joint Institute for Laboratory Astrophysics and he wishes to record his deep appreciation to the Fellows of JILA for allowing his participation in this work to be made possible.

Appendix: The Gravitational Field of a Hollow Cylinder

The coordinate system chosen for the problem is illustrated in Fig. 2, the radial coordinate being denoted by y and the axial coordinate by z. The inner and outer radii of the cylinder are r_1 and r_2 respectively, its density is ρ and its length is 2ℓ. The origin of the coordinate system is chosen so that the planes $z = \pm\ell$ are coincident with the end faces of the cylinder. We denote the binomial coefficients by ${}^nC_r$ and define the following terms; $C_i = (r_i^2 - y^2)^{1/2}$, where $i = 1$ or 2, unless $y \geq r_i$ in which case $C_i = 0$,

$$\Gamma_n = \frac{\left[\frac{1}{2}\right]!}{\left[\frac{1}{2} - n\right]! n!} \quad \text{and}$$

$$\gamma_{n,m} = \frac{(2(2n - m))!}{2^{2n} m!(2n - m)!(2(n - m))!}$$

Observing the convention we take $0! = 0^0 = 1$. The components of the field intensity in the axial and radial directions at the point (z,y) are then given by

$$F_z = F_z^{(a)} + F_z^{(b)} + F_z^{(c)} + F_z^{(d)}$$
$$F_y = F_y^{(a)} + F_y^{(b)} + F_y^{(c)} + F_y^{(d)},$$

where $F_z^{(a)} = F_y^{(a)} = 0$ *unless both* $z - \ell < C_1$ *and* $z + \ell > -C_1$ in which case

$$F_z^{(a)} = 2\pi G\rho \sum_{i=1}^{2} (-1)^i r_i \sum_{n=1}^{\infty} \frac{\Gamma_n}{r_i^{2n}} \sum_{m=0}^{n} (-1)^m \gamma_{n,m} \sum_{p=0}^{m} {}^mC_p\, y^{2p} [(z + \zeta)^{2(n-p)}]_{\zeta_1}^{\zeta_2},$$

$$F_y^{(a)} = 2\pi G\rho \sum_{i=1}^{2} (-1)^i r_i \sum_{n=1}^{\infty} \frac{\Gamma_n}{r_i^{2n}} \sum_{m=1}^{n} (-1)^m \gamma_{n,m} \sum_{p=1}^{m} 2p\, {}^mC_p\, y^{2p-1} \frac{[(z + \zeta)^{2(n-p)+1}]_{\zeta_1}^{\zeta_2}}{2(n - p) + 1},$$

the limits ζ_1 and ζ_2 being given by

$$\zeta_1 = -\ell \text{ if } z - \ell > -C_1 \text{ or}$$
$$\zeta_1 = -z - C_1 \text{ if } z - \ell \leq -C_1;$$
$$\zeta_2 = \ell \text{ if } z + \ell < C_1 \text{ or}$$
$$\zeta_2 = -z + C_1 \text{ if } z + \ell \geq C_1$$

and where $F_z^{(b)} = F_y^{(b)} = 0$ *unless either* $z + \ell > C_2$ or $z - \ell < -C_2$ in which case

$$F_z^{(b)} = 2\pi G\rho \sum_{i=1}^{2} (-1)^i r_i \sum_{n=1}^{\infty} \Gamma_n r_i^{2n-1} \sum_{m=0}^{n-1} (-1)^m y^{2m} [(z + \zeta)^2 + y^2)^{-(n+m-\frac{1}{2})}]_{\zeta_1}^{\zeta_2} \sum_{p=0}^{n-m-1} (-1)^p \gamma_{n-1,p}\, {}^{n-p-1}C_m$$

$$F_y^{(b)} = 2\pi G\rho \sum_{i=1}^{2} (-1)^i r_i \sum_{n=1}^{\infty} 2\Gamma_n r_i^{2n-1} \sum_{m=0}^{n-1} (-1)^m y^{2m} [m y^{-1} I_{(n+m-1)+1/2} - (n + m - \tfrac{1}{2}) y I_{(n+m)+1/2}]_{\zeta_1}^{\zeta_2} \times \sum_{p=0}^{n-m-1} (-1)^p \gamma_{n-1,p}\, {}^{n-p-1}C_m$$

where

$$I_{q+\frac{1}{2}} = \frac{\chi(\eta^2 - 1)}{(\eta - 1)} + \eta^q \ln(z + \zeta + ((z + \zeta)^2 + y^2)^{1/2})$$

in which

$$\chi = \frac{(z + \xi)}{y^2(2q - 1)((z + \zeta)^2 + y^2)^{q-1/2}}$$
$$\eta = \frac{2(q - 1)}{y^2(2q - 1)},$$

the limits being

$$\zeta_1 = -\ell \text{ if } z - \ell > C_2 \text{ or if } z - \ell < -C_2$$

or

$$\zeta_1 = -z - C_2 \text{ if } C_2 \geq z - \ell \geq -C_2$$

and

$$\zeta_2 = \ell \text{ if } z + \ell > C_2 \text{ or if } z + \ell < -C_2$$

or

$$\zeta_2 = -z + C_2 \text{ if } C_2 \geq z + \ell \geq -C_2$$

and where $F_z^{(c)} = F_y^{(c)} = 0$ *unless both* $z - \ell < C_2$ *and* $z + \ell > C_1$ in which case

$$F_z^{(c)} = F_z^{(a)}(i = 2 \text{ only}) + F_z^{(b)}(i = 1 \text{ only}) - 2\pi G\rho[z + \zeta]_{\zeta_1}^{\zeta_2}$$

and

$$F_y^{(c)} = F_y^{(a)}(i = 2 \text{ only}) + F_y^{(b)}(i = 1 \text{ only})$$

the limits being

$$\zeta_i = -\ell \text{ if } z - \ell > C_1 \text{ or } \zeta_1 = -z + C_1 \text{ if } z - \ell \leq C_1$$

and

$$\zeta_2 = \ell \text{ if } z + \ell < C_2 \text{ or } \zeta_2 = -z + C_2 \text{ if } z + \ell \geq C_2$$

and where $F_z^{(d)} = F_y^{(d)} = 0$ *unless both* $z - \ell < -C_1$ *and* $z + \ell > -C_2$ in which case $F_z^{(d)}$ and $F_y^{(d)}$ are given by the equations above for $F_z^{(c)}$ and $F_y^{(c)}$ except that the limits are

$$\zeta_1 = -\ell \text{ if } z - \ell > -C_2 \text{ or } \zeta_1 = -z - C_2 \text{ if } z - \ell \leq -C_2$$

and

$$\zeta_2 = \ell \text{ if } z + \ell < -C_1 \text{ or } \zeta_2 = -z - C_1 \text{ if } z + \ell \geq -C_1 .$$

References

[1] H. Cavendish, Phil. Trans. R. Soc. London 88, 469 (1798).
[2] M. Hulett, Senior Thesis, Wesleyan Univ., Conn. (1969).
[3] D. R. Long and D. Ogden, Phys. Rev. D 10, 1677 (1974).
[4] D. R. Long, Nature 260, 417 (1976).
[5] C. C. Speake and A. J. F. Metherell, these proceedings.
[6] A. J. F. Metherell *et al.* to be published in Metrologia.
[7] P. Heyl, J. Res. Natl. Bur. Stand. 5, 1243 (1930).
[8] A. J. F. Metherell *et al.* to be published in Phil. Trans. R. Soc. London.
[9] W. D. Macmillan, *The Theory of the Potential* (Dover Publications Inc., New York, 1958).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Vacuum Polarization and Recent Measurements of the Gravitational Constant as a Function of Mass Separation

Daniel R. Long

Department of Physics, Eastern Washington University, Cheney, WA 99004

We point out that the coulomb inverse square law has long been known to fail at small charge separations due to a vacuum polarization effect. We point out that gravitational inverse square law data should be analyzed for a vacuum polarization effect and present the QED result for the mass density shift of the vacuum due to an inducing mass. We point out that care must be used in analyzing the data because the point mass situation is very different from the case where one of the masses is quite extensive. We analyze the current data for the value of the vacuum polarization parameter λ. Much of that data has errors too large to give definite results although it tends to agree with my earlier result of $\lambda = 0.002$. Of the more precise results, Stacey and Tuck are in agreement with my value while Spero *et al.* disagree. It is pointed out that the vacuum polarization analysis of the Spero *et al.* result is ambiguous.

Key words: gravitation; gravitational inverse square law; inverse square law failure; non-Newtonian gravitation; vacuum polarization.

1. Introduction

It is surprising that the gravitational inverse square law was not experimentally investigated many years ago since the coulomb inverse square law has long been known to fail at small charge separations. The coulomb inverse square law fails by a very general process known as vacuum polarization. By this process, a source of force polarizes attracted charges out of the vacuum and builds up a charge density around itself. Just as a negative electron accumulates a cloud of positron density around it, a mass can attract a cloud of positive mass density around it. This mass density cloud around a mass causes the inverse square law to fail because an approaching mass sees less attracting mass as it penetrates the cloud.

The idea of a mass density in the vacuum near a mass seems strange; however, if the particles in such a cloud do not interact except gravitationally such a mass density shift of the vacuum would be very difficult to detect. Indeed, there could be a large non-interacting mass density of the entire vacuum and it would be undetectable since it would pull gravitationally equally in all directions. Hence, the mass density of the vacuum we are speaking of might be regarded as a small shift in an already large mass density of the vacuum.

2. Theory

The mass density shift of the vacuum $\delta\rho(\mathbf{r})$ has been calculated from quantum field theory by Uehling [1] and by Schwinger [2] and is given by

$$\delta\rho(\mathbf{r}) = \frac{-\lambda}{8\pi}\int_{|\mathbf{r}-\mathbf{r}'|\ll R^*} \frac{\left[1.410 + \log\frac{|\mathbf{r}-\mathbf{r}'|}{R^*}\right]}{|\mathbf{r}-\mathbf{r}'|}\nabla^2\rho(\mathbf{r}')\,dr' \,. \quad (1)$$

Here, $\rho(\mathbf{r}')$ is the external, zeroth order mass density distribution, R^* is the Compton wavelength of the fermions which couple to the graviton, and λ is the interaction strength.

When Eq. (1) is applied to a small, compact, mass it yields a mass density shift of the vacuum which drops off inversely as the cube of the distance away from the mass. When the force, F, between two such masses, M_1 and M_2, is calculated from Eq. (1), the result is

$$F(R) = \frac{G(R)M_1M_2}{R^2} \quad (2)$$

$$G(R) = G_0\left\{1 + \lambda\log\frac{R}{R^*}\right\}, \quad (3)$$

where R is the mass separation, G_0 the gravitational constant at $R = R^*$, and λ is the same fermion-graviton coupling constant as used in Eq. (1). It is also a result of quantum field theory that the induced mass density of the vacuum drops exponentially to zero for $R \gtrsim R^*$. Hence, the inverse square law holds perfectly at large mass separation, a result in accord with astronomical observations.

It is important to understand clearly how to apply Eq. (3) to gravitational data in order to test that data for vacuum polarization effects. By differentiating Eq. (3), we find that

$$\lambda = \frac{1}{G_0}\,\frac{dG(R)}{d(\log R)}\,, \quad (4)$$

and so it is attractive to assume that λ is simply the dimensionless logarithmic derivative given by

$$\lambda = \frac{d(\log G(R))}{d(\log R)}. \quad (5)$$

For the purpose of evaluating λ from $G(R)$ data taken at two different mass separations R_2 and R_1, $R_2 > R_1$, we can use the following expression as long as $G(R)$ only changes by a small percentage:

$$\lambda = \frac{G(R_2) - G(R_1)}{(1/2)[G(R_2) + G(R_1)]\log\frac{R_2}{R_1}}\,. \quad (6)$$

It must be greatly emphasized that *Eqs. (4), (5), and (6) only apply when the attracting masses are small and compact compared to the distance between them.* When one of the masses is very large and extensive, Eq. (1) must be used directly.

3. Comparison with Experiment

There now exist a number of sets of data involving compact masses from which values of λ can be inferred. Unfortunately most of these experiments quote errors too large to unambiguously support the claim that λ differs from the Newtonian value of zero. On the other hand, the general trend of the data suggests that $\lambda \simeq 0.002$ and hence that more careful experiments should be done.

In 1974, I published [3] an analysis of the historic measurements of G at various mass separations. A least squares fit to that data gives a historic value for λ

$$\lambda\,(\text{Historic}) = +2.11 \times 10^{-3} \pm 0.70 \times 10^{-3},$$

where the error is one standard deviation.

In 1976, I reported [4] a measurement which compared the value of the gravitational constant at a mass separation 29.9 cm with the value at a mass separation of 4.48 cm. The possible systematic errors in that experiment were carefully assessed. The resulting value for λ was

$$\lambda\,(\text{Long}) = +1.95 \times 10^{-3} \pm 0.4 \times 10^{-3},$$

where the quoted error contains both the random and systematic errors.

Also in 1976, a Russian group, O. V. Karagioz *et al.* [5], happened to measure G at a number of different mass separations. While they did not note the fact, their best and most complete data yielded a λ value of

$$\lambda\,(\text{Karagioz}) = +2.015 \times 10^{-3} \pm 0.61 \times 10^{-3},$$

where the given error is one standard deviation.

In 1979, Panov and Frontov reported an experiment to explicitly check out my result for λ. Their experiment measured the growing excitation of a torsion balance as the attracting masses were removed back and forth at the fundamental period of the torsion pendulum. In the part of the experiment where they compared the value of the gravitational constant at a mass separation of 295.8 cm with that at 42.06 cm, they found the value of λ

$$\lambda\,(\text{Panov, 1}) = +1.54 \times 10^{-3} \pm 3.07 \times 10^{-3}.$$

They also used a 600 kg mass to compare the value of G at a mass separation of 984.0 cm with that at 42.06 cm with the resulting value of λ

$$\lambda(\text{Panov, 2}) = -0.63 \times 10^{-3} \pm 4.12 \times 10^{-3}.$$

It is evident that the errors quoted for both of these λ values are too large to really check my result. However, considering the problems of floor tilt and the precision location of a 600 kg mass, it seem clear that their first value, $\lambda(\text{Panov}, 1)$ is to be preferred. That value is in good agreement with my result.

In 1980, Hirakawa *et al.* [7] used a 1400 kg aluminum gravitational antenna to measure the near wave amplitude due to a rotating bar mass at various mass separations. We can infer the following values of λ from their work.

$$\lambda(\text{Hirakawa, 1}) = +77.4 \times 10^{-3} \pm 172 \times 10^{-3}$$
$$\lambda(\text{Hirakawa, 2}) = +12.9 \times 10^{-3} \pm 93 \times 10^{-3}$$
$$\lambda(\text{Hirakawa, 3}) = +1.1 \times 10^{-3} \pm 75 \times 10^{-3}$$
$$\lambda(\text{Hirakawa, 4}) = +3.9 \times 10^{-3} \pm 49 \times 10^{-3}$$

As can be seen, their stated errors are far too large to really test my value of λ; however, their two most accurate data tend to agree with my result.

In 1979, Yu *et al.* [8] measured the force due to gravity near a large oil tank when it was full and when it was empty. In this way, they obtained a ratio of their value of the gravitational constant at a mass separation of some 30 m to the known value found by experiments at a mass separation of 13 cm. The fact that their attracting mass was so large compared to the mass separation makes this case a borderline case which probably should be treated using Eq. (1). We will, however, treat it as a point mass situation and calculate λ using Eq. (6). Their resulting values for λ are:

$$\lambda(\text{Yu, 1}) = +1.80 \times 10^{-3} \pm 21.8 \times 10^{-3}$$
$$\lambda(\text{Yu, 2}) = +3.54 \times 10^{-3} \pm 34.7 \times 10^{-3}$$
$$\lambda(\text{Yu, 3}) = +15.2 \times 10^{-3} \pm 41.3 \times 10^{-3}$$
$$\lambda(\text{Yu, 4}) = +52.1 \times 10^{-3} \pm 60.6 \times 10^{-3}.$$

Again we find that the stated errors are too large to test for my value of λ, but their best values tend to agree with my results.

The most unequivocal support for my value of λ comes from the geophysical work of Stacey and his collaborators. Stacey *et al.* [9] measured the force due to the earth's gravity at different points along a vertical mine shaft about 1 km deep. By precisely measuring the density of the surrounding rock they inferred a value for the gravitational constant at a range of about 0.5 km of

$$G\,(R = 0.5\,\text{km}) = 6.712 \times 10^{-8} \pm 0.025 \times 10^{-8}\ \frac{\text{dyne cm}^2}{\text{gm}^2}.$$

Furthermore, Stacey and Tuck [10] have examined all of the geophysical measurements of G available in the literature. All of that data also gave large values for G and the best value, that of McCulloh, was nearly identical to the one measured by Stacey *et al.*

A value of λ can be inferred from the Stacey and Tuck result by comparing that value of G with the authoritative result of Heyl and Chrzanowski [11] measured at a mass separation of 13 cm. However, in this case the attracting mass is immense and completely surrounds the gravimeter attracted mass. In this case Eq. (1) must be used to find λ. Using Eq. (1) we find that the vacuum polarization mass density shift of the vacuum within the 1 km thick rock slab is very small and *negative* (tending to reduce the observed G). The vastly dominant extra force arises from the attraction by the rock slab of the induced mass cloud coupled to the gravimeter mass. The compton wavelength of the fermions which couples to the graviton, R^*, enters into the calculation as a free parameter. For a value of $R^* = 349$ km the results of Stacey and Tuck are in complete agreement with my work:

$$\lambda(\text{Stacey}) = +2.00 \times 10^{-3} \pm 1.2 \times 10^{-3}.$$

It should be emphasized that the effects of the parameter R^* are small ($\sim 30\%$), and that the Stacey and Tuck result essentially agrees with my work irrespective of the precise value of R^*.

The paper by Spero *et al.* [12] presents the only result which, at first glance, appears to explicitly disagree with my value for λ. That group investigated the gravitational field inside of a long, uniform, cylinder which should vanish according to Newtonian gravitation. They found results which were consistent with Newtonian

gravitation to high accuracy.

This null experiment involved an attracting mass which is very extensive and even surrounds the attracted mass. Hence, we must go directly to Eq. (1) in order to deduce a value for λ. Unfortunately, it appears that Eq. (1) is too simple to follow such a complex topology. To illustrate this point we shall consider a null experiment using a spherical shell of mass rather than a cylinder. I have explicitly integrated Eq. (1) for the case of a solid sphere. Outside of the sphere we obtain an inverse cubed distance dependence of the induced, positive, mass density. Inside of the sphere we find the surprising result that the induced mass density of the vacuum is large and *negative*. Now if we cut a hole in the center of the sphere we really do not change the inducing gravitational fields very much so we might expect that Eq. (1) would give us a negative induced mass density at a point inside such a hole. Such a negative induced mass density would tend to explain the Spero *et al.* result.

Contrary to these physical expectations, Eq. (1) gives a positive induced mass density in the hole in the sphere. Physical intuition is offended even further by a detailed study of how Eq. (1) gets this positive mass density in the hole. It turns out that for points in the hole Eq. (1) acts just as if it were at a point outside of the sphere. This is very troubling because we would generally expect the polarization of the vacuum to depend on the nearby gravitational fields and the fields inside a hole in a sphere are very different from the field outside of a sphere. The fact that Eq. (1) "thinks" those two physical situations are identical tends to suggest that it is not a sufficiently good approximation to deal with the spherical shell case or the cylinder used by Spero *et al.*

At present, I am working to try to resolve this paradoxical behavior of Eq. (1) by expressing the induced mass density shift in terms of the nearby gravitational fields, but this work is not finished. For the present we shall use the point mass Eq. (6), to calculate a value of λ for the Spero *et al.* measurement and include a question mark with the number to indicate the above ambiguities. Thus,

$$\lambda(\text{Spero}) = +0.01? \times 10^{-3} \pm 0.97? \times 10^{-3}.$$

The foregoing discussion presents all of the data related to λ that I am aware of. These results are summarized in Table 1, and tend, I think to indicate a non-zero value of λ.

References

[1] A. E. Uehling, Phys. Rev. 48, 55 (1935). See also D. R. Long, Il Nuovo Cimento 55B, 252 (1980). Equation (1) has been modified for the gravitational case by the substitution $\alpha = -3\pi\lambda/4$.

[2] J. Schwinger, Phys. Rev. 75, 651 (1949).

[3] D. R. Long, Phys. Rev. D 9, 850 (1974).

[4] D. R. Long, Nature 260, 417 (1976).

[5] O. V. Karagioz, V. P. Isamaylov, N. I. Agafonov, E. G. Kocheryon, and Yu. A. Tarakanov, Izv. Adad. Nauk. SSR, Fiz. Zemli 12 (5), 351–354 (1976) [Izv. Acad. Sci. USSR, Phys. Solid Earth 12 (5), 106–111 (1976)].

[6] V. I. Panov and V. N. Frontov, Zh. Exsp. Teor. Fiz. 77, 1701 (1979) [Sov. Phys. JETP 50, 852 (1979)].

[7] H. Hirakawa, J. Tsubono, and K. Oide, Nature (London) 283, 184 (1980).

[8] H.-T. Yu, W.-T. Ni, C.-C. Hu, F.-H. Liu, C.-H. Yang, and W.-N. Liu, Phys. Rev. D 20, 1813 (1979).

[9] F. D. Stacey, G. J. Tuck, A. R. Maher, and D. Morris, Phys. Rev. D 23, 1683 (1981).

[10] F. D. Stacey and G. J. Tuck, Nature (London) 292, 230 (1981).

[11] P. Heyl and P. Chrzanowski, J. Res. Natl. Bur. Stand. 29, 1 (1942).

[12] R. Spero, J. K. Hoskins, R. Newman, J. Pellam, and J. Schultz, Phys. Rev. Lett. 44, 1645 (1980).

TABLE 1. *Experimental Values for* λ.

Source	$\lambda(10^{-3})$	R_2	R_1
λ(Historic)	+ 2.11 ± 0.70	80 cm	6.3 cm
λ(Long)	+ 1.95 ± 0.40	29.9 cm	4.48 cm
λ(Karagioz)	+ 2.02 ± 0.61	16.5 cm	5.6 cm
λ(Panov,1)	+ 1.54 ± 3.07	295.8 cm	42.06 cm
λ(Panov,2)	− 0.63 ± 4.12	984.0 cm	42.06 cm
λ(Hirakawa,1)	+77.4 ±172.	2.5 m	2.2 m
λ(Hirakawa,2)	+12.9 ± 93.	3.0 m	2.2 m
λ(Hirakawa,3)	+ 1.1 ± 75.	3.5 m	2.2 m
λ(Hirakawa,4)	+ 3.9 ± 49.	4.2 m	2.2 m
λ(Yu,1)	+ 1.80 ± 21.8	30.47 m	13.0 cm
λ(Yu,2)	+ 3.54 ± 34.7	32.35 m	13.0 cm
λ(Yu,3)	+15.2 ± 41.3	34.35 m	13.0 cm
λ(Yu,4)	+52.1 ± 60.6	36.30 cm	13.0 cm
λ(Stacey)	+ 2.00 ± 1.2	1 km	13.0 cm
λ(Spero)	+ 0.01? ± 0.07?	~30 cm	~4 cm

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Tests of the Gravitational Inverse Square Law Using Torsion Balances

J. K. Hoskins, R. Newman, J. Schultz, and R. Spero*

Department of Physics, University of California at Irvine, Irvine, CA 92717

We describe experiments at U.C. Irvine to test the inverse square distance dependence of the gravitational force at laboratory distances. One experiment has tested a distance range of approximately 2 to 5 cm, using a test mass suspended from a torsion balance to probe the gravitational field inside a mass tube. Results of this experiment support an inverse square law: Assuming a force law deviating from inverse square by a factor $[1 + \epsilon \ln r]$ with r expressed in cm, we find $\epsilon = (1 \pm 7) \times 10^{-5}$. A second experiment, now in progress, tests a distance range from 5 cm to 20 m. We discuss the feasibility of an experiment to test the inverse square law at distances less than 1 mm. Methods are discussed for reducing the effect of seismic noise on a torsion balance by damping pendulum oscillation modes of the balance.

Key words: gravitation; inverse square law; Newtonian gravity; torsion balance.

1. Introduction

The possibility of a deviation from inverse square behavior of the gravitational force on a laboratory distance scale has received considerable recent attention. Several authors have pointed out that a suprisingly large anomaly could be consistent with existing empirical evidence. An experiment indicating a failure of the inverse square law over a distance range 4.5–30 cm has in fact been reported by Long [1]. Several recent theoretical ideas suggest forces which could manifest themselves as a failure of the inverse square law of gravitation; of particular interest are the possibility of a long-range component of the strong interaction arising from two gluon exchange [2], and the suggestion arising from supergravity unification theories [3] of a possible laboratory scale deviation from inverse square behavior.

We are conducting experiments to test the inverse square law over distance ranges 2–5 cm and 5 cm–20 m, and have considered an experiment to probe a range < 1 mm. These experiments are based on the use of the torsion balance, an ancient instrument which has proven admirably suited for this application, and shows promise for vast further improvement in capability.

2. The 2–5 cm Experiment

This experiment has been completed, and results have been published [4]; we reproduce some of the details here. The method is essentially a null experiment; a torsion balance (Fig. 1) is used to measure the change in the force acting on a test mass suspended inside a long hollow cylinder as the cylinder is moved laterally so that the test mass hangs alternately near opposite walls. For an infinitely long perfect cylinder and exact inverse square force law the gravitational field due to the cylinder vanishes everywhere inside it, just as inside a spherical mass shell. For our finite cylinder of length $L = 60$ cm and inside diameter $D = 6$ cm there exists a small net "end effect" force on a test mass located near an inside wall, smaller than the nearly balanced opposing forces due to near and far wall by a factor $(D/L)^2 = 10^{-2}$. Thus,

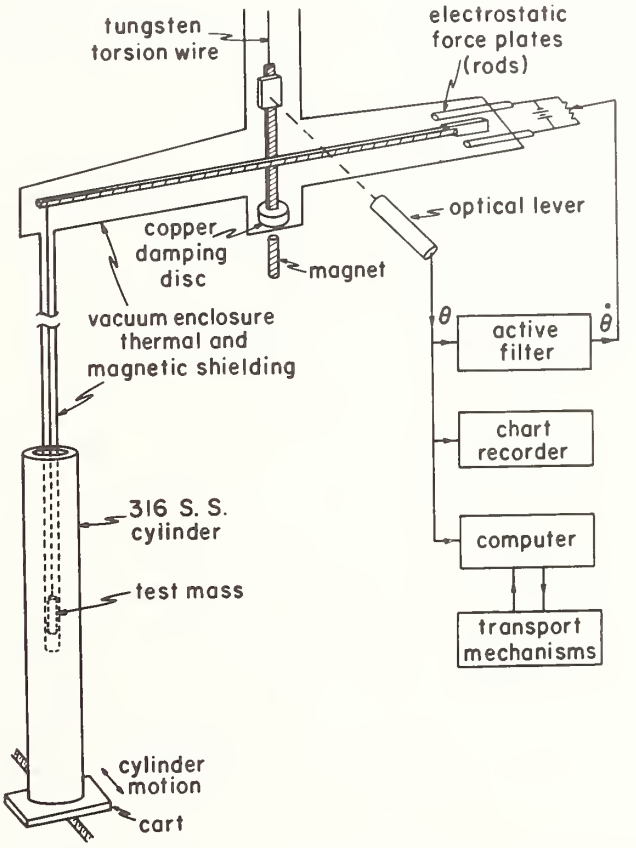


FIGURE 1. *Schematic of the experimental apparatus for the 2–5 cm inverse square law test.*

to compare the gravitational force at two distances corresponding to the distances from the near and far walls in our cylinder, to a level of 1 part/10^5, we need measure the end effect force to only 1 part/10^3. Furthermore, the end effect force is nearly a linear function of test mass radial position in the cylinder, so that we need measure only the relative motion of the cylinder with respect to the test mass, while the absolute position of the cylinder relative to the test mass need only be known to a few millimeters. These features of the finite cylinder

*Present address: Department of Physics, California Institute of Technology, Pasadena, CA 91125.

"null" experiment greatly simplify metrology tasks. The homogeneity and geometry of the cylinder itself must be known with a precision on the order of 1 part/10^5, but by averaging data taken at a set of equally spaced azimuthal orientations of the cylinder results are obtained equivalent to those for an azimuthally symmetric and homogeneous cylinder, leaving only axial cylinder mass variations to be precisely measured.

The cylinder, of mass 10.4 kg and wall thickness 1 cm, was made of high purity double vacuum melted 316 stainless steel. Its associated magnetic field was measured to be less than 5 nT. The test mass was a 20 gm 4.4 cm long rod made of high purity copper, hanging 83 cm below the end of a torsion balance boom of total length 60 cm, made of OFHC copper. All components were machined with care to avoid imbedding ferromagnetic fragments. The balance was suspended by a 32 cm long, 75 μm diameter tungsten wire in a vacuum of 2×10^{-7} Torr, maintained by an ion pump. The cylinder was external to the vacuum system, which was surrounded in the region of the test mass by magnetic and thermal shielding.

Calibration of the balance to the level required for the "null" test was accomplished by applying a known gravitational force, using a 133 g copper ring of radius R = 12.1 cm, located so that the test mass lay on the ring axis at the position $z = R/\sqrt{2}$ where the force due to the ring is a maximum and hence insensitive to the exact location of the test mass. The ring mass was chosen to produce a force on the test mass approximately equal to the end effect force it calibrates. At intervals the ring was moved by remote control to a corresponding position on the far side of the test mass, producing a calculable change in torque on the balance which served to calibrate it. In addition, the ring was used to enhance the null aspect of the experiment: whenever the cylinder was moved, the ring was simultaneously moved to its opposite position to largely cancel the change in force on the test mass due to the cylinder.

The experiment ran unattended during night hours, under the control of a computer. Data was taken with the test mass hanging 1.8 cm from the center of one wall. At intervals of 17 minutes the ring was moved, and the cylinder translated by 3.43 cm on a screw-driven cart using a remote magnetically shielded motor.

From a total of 132 measurements of the torque change $\Delta\Gamma$ on the balance associated with a roundtrip cylinder-ring cycle, at 8 cylinder azimuthal positions, we determined the experimental value $\Delta\Gamma = (4.49 \pm 0.03)$ μdyn-cm. (For comparison, the torque associated with only the portion of the cylinder to one side of the test mass is approximately 2000 μdyn cm.) The corresponding calculated value for $\Delta\Gamma$, based on a Newtonian force law, is $\Delta\Gamma = (4.47 \pm 0.14)$ μdyn cm, in excellent agreement with our measured value. The uncertainty in this calculated value is dominated by approximately equal contributions from: (a) uncertainty in the balance arm mass distribution, (b) limits on possible magnetic couplings, based on runs without shielding, and (c) uncertainty in corrections for nonuniformity in cylinder geometry and density. Deviations from perfect geometry were measured with 10^{-7} m resolution using a commercial differential gaging instrument. Variations in cylinder mass distribution were determined by computer-controlled axial scans at various azimuths of the transmission of γ-rays through the cylinder wall.

Analysing our data assuming the parameterization suggested by Long for a force law anomaly:

$$G = G(r) = G_0[1 + \epsilon \ln r]$$

we find: $\epsilon = (1 \pm 7) \times 10^{-5}$ for r measured in cm, compared to the value required to fit Long's data:

$$\epsilon_{\text{Long}} = (200 \pm 40) \times 10^{-5}$$

Another possible form for an anomaly involves a Yukawa potential, taking for the potential energy of masses M and m separated by r:

$$V(r) = \frac{G_\infty Mm}{r}(1 + \alpha e^{-r/\lambda}).$$

Assuming this form we have determined the range in (α, λ) parameter space which is excluded by our data. We find our result to be inconsistent with Long's for all values of λ. (See Fig. 3.)

Long has argued [5] that the effect he finds may not be observable in a null experiment such as ours. Non-null tests of the inverse square law may be desirable.

3. The 5 cm–20 m Experiment

This experiment, now in progress, is not of a null type. A torsion balance, consisting of a 60 cm long copper bar suspended at its midpoint by a tungsten fiber, acts as a gradiometer to determine the field gradient produced by a distant mass. The change in torque on the balance which results when 3 tons of water are pumped between two distant water tanks is compared with the torque change produced by moving a one gram mass from a position 5 cm to one side of the balance bar to a corresponding position on the other side. The balance operates in a vacuum enclosure in a thermally insulated steel-lined pit at a depth of 2 m, in an open field near the UCI campus. The balance arm has a mass of 500 gm and moment of inertia 157,000 g · cm^2. The suspension fiber has a diameter of 90 μm and torsion constant 49 dyn·cm/rad. The undamped period of the torsion oscillation is 350 seconds. Our hope is to achieve a sensitivity of 1 part/10^3 in the comparison of G(5 cm) and G(20 m), which would enable a determination of Long's parameter ϵ with an uncertainty of 2×10^{-4}. We consider in the following sections the contributions to the uncertainty in this comparison.

3.1 Metrology

Geometric and mass metrology for this experiment are, we believe, comfortably under control at the 1 part/10^3 level. Several features of the experiment simplify metrology tasks. The water tanks are suspended from single points, so that a precise determination of the positions of their centers of mass in the horizontal plane reduces simply to accurately surveying the suspension points. One of the tanks hangs from a precision load cell which determines the mass of water transferred between tanks to a precision of 5×10^{-4}. The arrangement of the 1 g "near mass" is such that the torque change on the balance, when the near mass is transferred from one side to the other, is sensitive only in second order to uncertainties in all three axes of the exact position of the balance arm relative to the near mass (see Fig. 2): (1) for vertical displacements the torque does not change to first order due to symmetry, (2) the near mass is located at the position along the length of the bar at which the torque it produces is a maximum; hence the torque is insensitive to small uncertainties in near mass position along the bar axis, (3) the near mass is transported between opposite sides of the balance arm by the rotating disk indicated in Fig. 2, in such a way that the distance between the two positions of the near mass is precisely determined by the

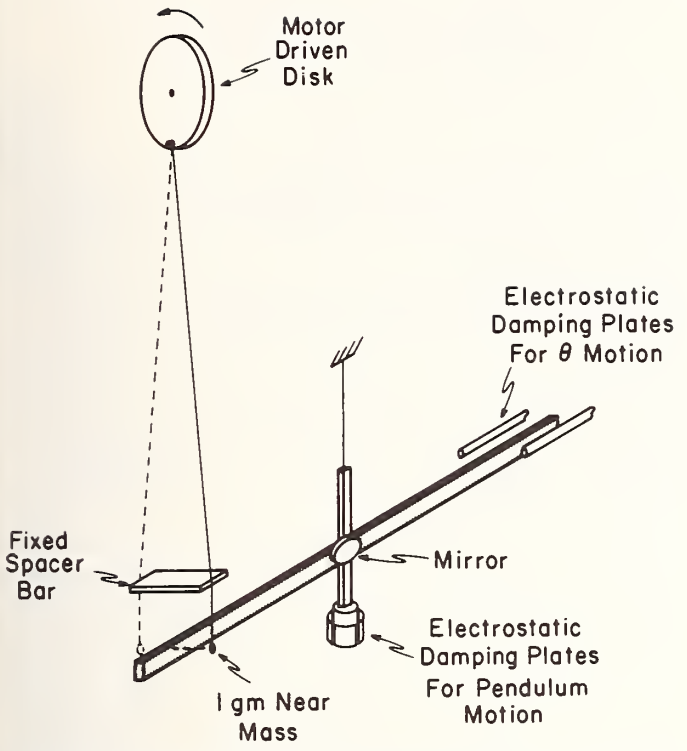


FIGURE 2. *Schematic of torsion balance for the 5 cm–20 m test, showing near mass transport mechanism and damping plate arrangement. The balance operates within a vacuum enclosure, to which the near mass is external.*

spacer bar indicated in the figure. The resulting torque change is sensitive to the exact position of the balance arm only in second order (if the bar is not quite centered between the two near mass positions the increased torque for one mass position is compensated to first order by a decreased torque in the other position). The result of these three considerations is that the balance arm position within its vacuum enclosure need be known only to about 1 mm in any direction for $1/10^3$ precision in this experiment. A window in the vacuum enclosure makes this determination readily feasible.

3.2 Nonthermal Noise

Noise in the torque signal is at present the main limitation in this experiment. Expressed in terms of the sensitivity of the balance as a gradiometer, the present effective noise level under quiet conditions, at our 0.75 cycle/hour operating frequency, is about $1E/\sqrt{Hz}$, where 1E = 1 Eötvös unit = 10^{-9} s^{-2}. Since the total change in field gradient when 3 tons of water are pumped between the two tanks is about $0.1\,E$, we can expect to compare G(20m) with G(5cm) with a reasonable integration time (10^6 s) only to about 1%. To reach our 1 part/10^3 goal will require a reduction of the present noise level.

A feature of the torsion balance used for this experiment is the use of electronic cooling not only of the torsion oscillation mode but also of pendulum oscillation modes. Pendulum motion of a torsion balance, excited by microseismic noise, couples by several subtle nonlinear mechanisms into torques. Because the mechanisms are nonlinear the resulting slowly varying part of the torsion noise signal is proportional to the square of the pendulum amplitude, and hence to the Q of the pendulum modes. Since the Q is on the order of 10^6 or more, very large reductions in the torsion signal noise may be expected from damping the pendulum oscillation modes. The torsion balance used for our 2–5 cm experiment (Fig. 1) used a magnet below a copper disk mounted on the torsion balance axis to produce eddy current damping of pendulum modes, reducing their Q by a factor of about 1000; this greatly reduced the observed noise in the torsion mode. The balance used in our 5 cm–20 m experiment (Fig. 2) uses two orthogonal pairs of fixed electrostatic plates, flanking a copper cylinder mounted on the balance axis, to sense pendulum motion of the balance and to apply appropriate damping forces. The electromechanical servo loop is similar to that applied by Block and Moore [6] to a gravimeter, and the analysis is similar. In our case two such servo loops are used, one for each orthogonal axis of pendulum motion. Also, in our case the pendulum motion of the balance in each axis has two normal modes, corresponding to swinging motion with top and bottom points of the rigid balance structure moving in phase and 180° out of phase. This complicates the analysis of the servo loop, but the resulting transfer function for the mechanical part of the system is simply expressed as a sum of two second order pole terms.

This damping system successfully reduces the Q of the pendulum modes of our balance from over 10^6 to on the order of 10^2, and greatly reduces the low frequency noise in the torsion signal. However the noise is still higher than desired, as indicated at the beginning of this section. What the dominant source of remaining noise may be is not clear at this point. It may be in part a result of high Q vertical spring oscillations in the balance, which are undamped by the present system, but were damped by the eddy current system of the 2–5 cm experiment. Gravitational noise from moving squirrels and rabbits in the vicinity of the balance may be a significant contribution. A number of other possible noise sources remain to be explored, and we still hope to reduce the noise to a level which will allow a 1 part/10^3 experiment.

3.3 Systematic Error

We have taken pains to minimize the danger of magnetic couplings, through the use of high purity materials, careful machining procedures, and magnetic shielding of the balance housing. Care was also taken to minimize thermal effects by shielding the balance housing with spaced concentric metal thermal shields. Our biggest concern is the possibility that ground tilt produced by the transfer of water between the distant tanks will shift the position of the hanging balance relative to the various surrounding electrostatic plates, with a consequent correlated change in torque on the balance. To determine if this is a problem we have taken two steps: (1) sensitive tiltmeters are mounted on the torsion balance base to determine the actual tilt generated when water is transported, and (2) the torsion balance base is mounted on three piezoelectric feet, which may be used to generate a large (3×10^{-6} rad) artificial tilt to determine the sensitivity of the torque signal to tilt. Preliminary measurements indicate that the tilt associated with water transport is less than 10^{-7} rad, assuring no problem for a 1% sensitivity experiment. A more precise measurement of the correlated tilts is possible, and will be necessary.

3.4 Experimenter Bias

Being concerned about the many subtle biases which can creep into an experiment of this type, such as the temptation to look with less intensity for sources of sys-

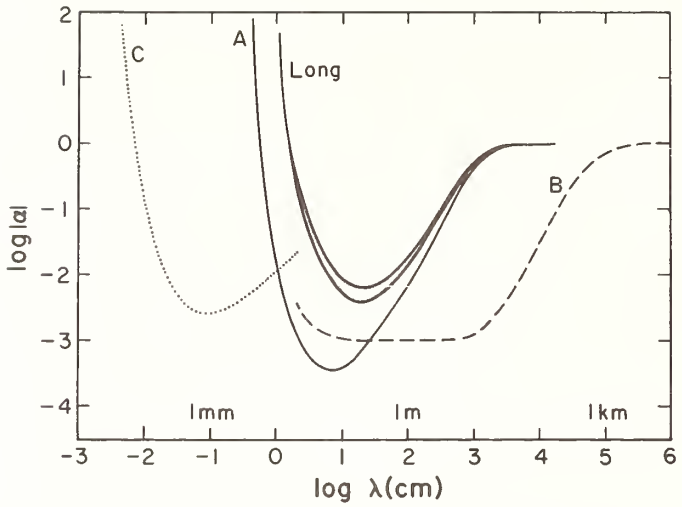


FIGURE 3. *Potential and present limits on the strength (α) and range (λ) of a Yukawa-type component of the gravitational force, for torsion balance experiments discussed in this paper. The curve marked "Long" is the locus of (α, λ) values consistent with D. Long's experimental results. Long's data require $\alpha < 0$. The other curves are: (A) the 2 σ limit on allowed (α, λ) values provided by our 2–5 cm experiment, assuming $\alpha < 0$, (B) the target sensitivity of our 5 cm–20 m experiment, and (C) the sensitivity which might be expected for an experiment with 0.05 cm mass separation.*

tematic error when the answer begins to look "right", we have protected ourselves by arranging that the value of the nominally 1 g "near mass" is known to us only to 10%, while the exact value is guarded by a colleague, unassociated with the experiment, who is instructed to reveal the exact value of the mass only when the book is closed on other aspects of the experiment.

4. Possible Inverse Square Law Tests for Mass Separations Less Than 1 mm

In view of the various theoretical motivations for testing the inverse square law [2, 3] it would be desirable to extend such tests to as small a mass separation as possible. In particular, the possible long range component of the strong interaction which could be associated with two gluon exchange between hadrons would have a van der Waals type distance dependence [2], increasing as a high power of the inverse mass separation. To search for such a force one should explore the minimum feasible mass separations. A good technique for doing this with a torsion balance might be to suspend a thin-walled metalized quartz tube of perhaps 2 cm diameter from the end of a torsion balance in such a way that it hangs within a similar stationary tube of slightly larger diameter which can be moved laterally in the fashion of our 2–5 cm experiment. In effect such an experiment measures the force as a function of distance between closely spaced mass plates. The possible advantages of using quartz tubes lie in the rigidity of a tube, high degree of homogeneity achievable with fused quartz, and the ability to repeat the experiment with a set of different azimuthal positions of the stationary tube relative to the suspended tube to reveal possible effects of mass non-uniformity. Suitable precision quartz tubes appear to be commercially available. A thin foil shield between the two tubes would be required to shield the electrostatic van der Waals forces, which become comparable to the gravitational force for mass separations of about 0.2 mm.

References

[1] D. R. Long, Nature (London) **260**, 417 (1976).
[2] G. Feinberg and J. Sucher, Phys. Rev. D **20**, 1717 (1979).
[3] J. Scherk, Phys. Lett. **88B**, 265 (1979).
[4] R. Spero *et al.*, Phys. Rev. Lett. **44**, 1645 (1980).
[5] D. R. Long, Nuovo Cimento B **55B**, 252 (1980), and these proceedings.
[6] B. Block and R. D. Moore, J. Geophys. Res. **71**, 4361 (1966).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Measurement of Gravitational Forces at Separations Around 10 Meters*

Hsou-Tao Yu and Wei-Tou Ni

Department of Physics, National Tsing Hua University, Hsinchu, China (Taiwan)

and

Chin-Cheng Hu, Fa-Hsiang Liu, Chia-Hsiang Yang, Wu-Nan Liu

Exploration and Exploitation Research Center, and Taiwan Petroleum Exploration Division, Chinese Petroleum Corporation, Miaoli, China (Taiwan)

We used a Worden gravimeter to measure the gravitational forces at fixed positions when a large oil tank was full and when it was empty. The results are reported here.

Key words: gravitational constant; inverse square law; oil tank experiment; tests of Newtonian gravity.

Since the last International Conference on Precision Measurement and Fundamental Constants, there has been considerable interest in the empirical foundation of the gravitational inverse square law [1, 2, 3]. In other words, the problem is whether the gravitational "constant" $G(r) \equiv F_{\text{grav}} r^2/Mm$ is independent of the separation of the two point masses m and M. Wagoner [4], Fujii [5], and O'Hanlon [6] have argued for the variation of $G(r)$ and proposed the α-μ model with $G(r) = G_c[1 + \alpha(1 + \mu r)e^{-\mu r}]$, where G_c, α and μ are constants.

Previous laboratory experiments had separations ranging from 5 cm to 1 m. Above 1 m there has been no direct measurement. In 1978, we used a Worden gravimeter to measure gravitational forces at fixed positions when a large oil tank was full and when it was empty. At a fixed point, the readings of the Worden gravimeter varied with time due to drift and tidal effect. To subtract these effects, we selected a base point as reference. For simplicity in analysis, readings were repeated between the base point and observation points within 1 hour. Since this time lapse is short compared with the period of tides, we can use linear interpolation methods to obtain $(g - g_{\text{base}})$ at equal time. The difference in g is due to the environment. We repeated the observations seven times to reduce errors. From these differences when the tank was full and when it was empty, we calculated $\Delta g = \bar{g} - g = g'_z$. The results are presented in Table 1 [7, 8].

According to the analysis of Mikkelsen and Newman [3], $G(r)$ is nearly constant in the ranges 5 cm $< r <$ 1 m and 10^4 km $< r < 3 \times 10^8$ km. Therefore in the α-μ model, we assume the range of variation μ^{-1} to be 1 m $< \mu^{-1} < 10^4$ km. Now using the data in Table I, the viable range of α and μ is shown as the shaded area in Fig. 1 [7, 8]. Mikkelsen and Newman [3] used models of the earth and sun to conclude that $0.6 < (1 + \alpha)^{-1} < 1.3$. Using these constraints and the possible values of α and μ in Fig. 1, we calculate the possible $G(r)$-range as the shaded area in Fig. 2 [7, 8]. This result, together with the accurate 2-5 cm experiment of Spero *et al.* [9], are incompatible with Long's result [2] in the α-μ model.

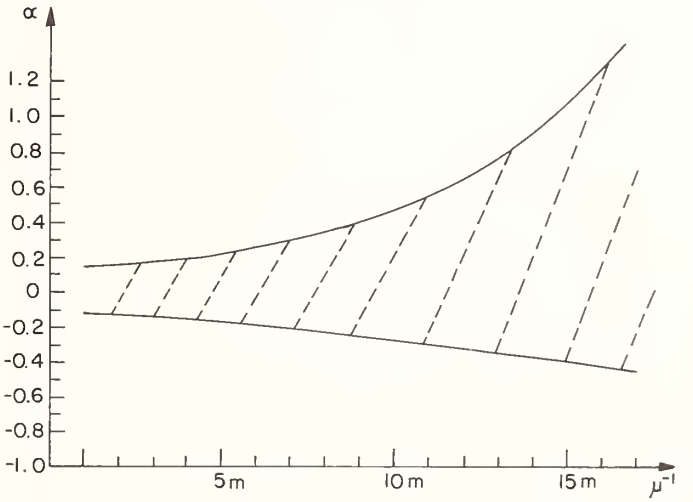


FIGURE 1. *Experimental constraints on the parameters α and μ in the α-μ model by Wagoner [4], Fuji [5], and O'Hanlon [6]. The viable range is shown as the shaded area.*

References

[1] D. R. Long, Phys. Rev. D 9, 850 (1974).
[2] D. R. Long, Nature (London) 260, 417 (1976).
[3] D. R. Mikkelsen and M. J. Newman, Phys. Rev. D 16, 919 (1977).
[4] R. V. Wagoner, Phys. Rev. D 1, 3209 (1970).
[5] Y. Fujii, Nature Phys. Sci. 234, 5 (1971); Ann. Phys. (N.Y.) 69, 494 (1972).
[6] J. O'Hanlon, Phys. Rev. Lett. 29, 137 (1972).
[7] H.-T. Yu, W.-T. Ni, C.-C. Hu, F.-H. Liu, C.-H. Yang, and W.-N. Liu, Chinese J. of Phys. 16, 201 (1978).
[8] H.-T. Yu, W.-T. Ni, C.-C. Hu, F.-H. Liu, C.-H. Yang, and W.-N. Liu, Phys. Rev. D 20, 1813 (1979).
[9] R. Spero, J. K. Hoskins, R. Newman, J. Pellam, and J. Schultz, Phys. Rev. Lett. 44, 1645 (1980).

*Supported in part by the National Science Council of China (Taiwan).

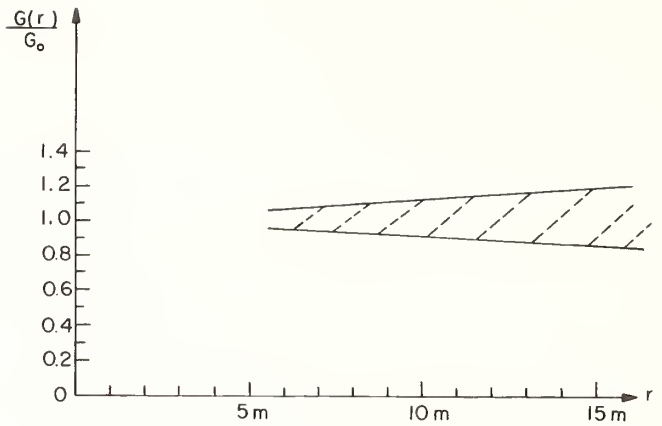


FIGURE 2. *Constraints on G(r) in the α-μ model. The viable range is shown as the shaded area.*

TABLE I. *Experimental Results*

Dates of measurement:	March 31, 1978		Full tank	Oil level: 13.473 m from the ground
	April 3, 1978		Empty tank	Oil level: 2.32 m from the ground
Tank diameter: 54.86 m Density of the oil: 0.885 gm/cm^3				
Distance of measured position from the rim of the tank	2 m	4 m	6 m	8 m
Mean experimental value $\lvert \Delta g_E \rvert$ (mGal)	0.122 ± 0.0155	0.085 ± 0.018	0.065 ± 0.017	0.060 ± 0.018
rms deviations in percent	13%	21.5%	26%	30%
Errors due to the uncertainty in the effective height of the mass-spring system of the Worden gravimeter	0.1%	0.3%	1%	3%
Newtonian prediction $\lvert \Delta g_N \rvert$	0.120	0.083	0.059	0.044
$\Delta g_E / \Delta g_N$	1.01 ± 0.13	1.02 ± 0.22	1.09 ± 0.29	1.35 ± 0.40

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Non-Newtonian Gravity: Geophysical Evidence

Frank D. Stacey and Gary J. Tuck

Physics Department, University of Queensland, Brisbane 4067, Australia

Six independent geophysical data sets have been used to estimate the value of the Newtonian gravitational constant G, yielding results that are consistently higher than the accepted laboratory-determined value G^* by 0.5% to 1.5%. Possibilities of unrecognized systematic errors preclude a definite conclusion, so that new large scale measurements of G are needed to clarify the matter. Two such experiments, using masses distributed on quite different scales, both much larger than any laboratory measurement of G, are under development. One determines the gravitational attraction by a 3.5 km layer of sea water by measurements of gravity in a bathyscaphe. The other makes use of frequent level changes of a hydroelectric pumped-storage lake, involving a 10 m layer of water.

Key words: geophysical measurements; gravitational constant; inverse square law; non-Newtonian gravity.

1. Introduction

Precise measurements of the Newtonian gravitational constant G have been made over a very limited range of separations of the attracting masses. Now that the validity of the inverse square law is being seriously questioned, measurements on scales quite different from those of conventional laboratory determinations are needed and geophysics has an important role to play. The development of geophysical experiments to determine G is partly a matter of retracing history, since the first methods of estimating G were geophysical [1, 2]. They were abandoned 100 years ago in the face of overwhelming evidence that laboratory measurements were much more precise. However, not only do we now have a new reason for reintroducing such large scale measurements, but geophysical techniques have advanced so dramatically since then that it may even be possible to produce a geophysical estimate that is more precise than the laboratory ones.

The feature of the current geophysical approach that gives it a particular advantage is that it uses very extensive plane layers of material as the attracting masses. In the case of water layers these can be very uniform and very precisely horizontal as well as effectively almost infinite in extent, so that their gravitational fields at all close points are independent of position and the forces exerted on test masses are independent of their shapes and homogeneities. All the geophysical estimates of G that we have reported recently [3] and both the experiments discussed here make use of this property.

Although we are suspicious of a deficiency in the inverse square law over a range extending to tens of meters at least, at the present stage our data are inadequate to suggest, on their own, the form of any required generalization. In making our calculations, we therefore assume validity of the inverse square law and formally calculate values of G in terms of it. A deficiency is then apparent as a disagreement with the conventional value G^* obtained from measurements on much smaller, laboratory scales. To proceed further we must recognize that we are measuring gravitational forces due to widely distributed masses and that we can determine directly not the law of gravity itself but some integral of it. Sufficiently precise data may nevertheless suffice to specify a form of generalization, but the present data are inadequate for this purpose.

2. Theoretical Basis of Geophysical Determinations of G

Although in principle any geological body of assumed known mass may be used to estimate G the only usable method in practice is that due, in its original form, to Airy [4] who measured gravity in a coal mine, in which extensive horizontal stratification is a reasonable assumption. Applied to the situation of a spherical, non-rotating Earth of radius r and mass M we can write the surface gravity

$$g(0) = GM/r^2 . \tag{1}$$

The product GM is inseparable in all astronomical and whole Earth geophysical observations, including rotational effects, precession and wobble and seismological observations including the free mode periods. To identify G independently of M it is necessary to make measurements at some depth z which is inside a known fraction ΔM of the Earth's mass so that

$$g(z) = G(M - \Delta M)/(r - z)^2 \tag{2}$$

and G is obtained by eliminating the unknown M from Eq. (1) and Eq. (2). In the approximation $z \ll r$, $g(z) - g(0) \ll g(0)$,

$$g(z) - g(0) = 2g(0)z/r - 4\pi G \int_0^z \rho\, dz . \tag{3}$$

A more complete equation that retains second order terms and includes the effects of the Earth's rotation and ellipticity is [5]

$$g(z) - g(0) = U(z) - 4\pi G X(z) , \tag{4}$$

where

$$U = 2g(0)z/r\,[1 + 3z/2r - (3/2)J_2(3\sin^2\phi_0 - 1)] + 3\omega^2 z(1 - \sin^2\phi_0) \tag{5}$$

$$X = (c/a)\,[1 + 2z/r + (1 - c^2/a^2)/2] \times \int_0^z \rho\, dz - (2/r)\int_0^z \rho\, z\, dz , \tag{6}$$

$J_2 = 1.08264 \times 10^{-3}$ is the mass ellipticity coefficient, $\omega = 7.2921 \times 10^{-5}$ rad s^{-1} is the rotation rate, ϕ_0 is geocentric latitude, and $c/a = 0.996647$ is the ratio of polar to equatorial axes. The refinements of Eqs. (4), (5), and (6) are necessary to our work.

It should be noted that, although we have presented the theory of the internal variation of gravity within the Earth in terms of a model Earth with complete spherical or elliptical layers, this is not in fact a requirement. All that is necessary is that the uniform layering should extend to a distance that is large compared with the depth range of the measurements. This is not an entirely trivial requirement for measurements over a depth range that may be several kilometers, but boundary corrections can readily be applied with sufficient accuracy.

3. Summary of Mine and Borehole Data

Equations (3) or (4) are the basis of a gravitational method of estimating the in situ densities of rock layers penetrated by mines and boreholes (by assuming the conventional value of G). In a few cases independent measurements of density have been reported, permitting the procedure to be reversed to calculate G. Including our own report of mine measurements undertaken specifically to determine G [5] we have found five data sets that appear reliable and give sufficient details for us to obtain values of G with standard deviations representing scatter of the data. The results are reported elsewhere [3] and summarized here (Table 1).

The values in Table 1 are consistently higher than the conventional laboratory estimate [10], $G^* = (6.672 \pm 0.004) \times 10^{-11}$ m^3 kg^{-1} s^{-2}, although only in the case of the McCulloh data is the discrepancy sufficiently large, relative to the formal standard deviation, to compel us to look carefully for possible systematic errors. We used 31 of McCulloh's gravity values [7] to obtain the estimate in Table 1 (4 values obtained in the top 50 m were omitted because weathering and fracturing of the rock in that range made the density estimates questionable, but including them makes little difference). Of these, the three subsets of 11, 10, 10 data points obtained over different ranges clearly indicate consistency of the data. The two possibilities for doubt are a systematic underestimate of sample densities and an anomalous regional gravity gradient due to some mass irregularity outside the range of observations. The supposition that the densities are all 1% low appears improbable, but is not totally impossible. We have attempted to allow for the possibility of a biasing mass anomaly by fitting a mass of arbitrary magnitude and depth to the data and found that including it further increased the estimate of G, although only marginally. Thus, if the high value of G is to be attributed to a systematic error, it must be due to a large, deep mass anomaly under Ohio, where McCulloh's measurements were made, causing an anomalously low free air gravity gradient over a broad area. Evidence for this is currently being sought.

4. Comparison of Sea-Floor and Sea-Surface Gravity Surveys

Before the development of methods of measuring gravity at the sea surface, marine surveys were made by a laborious procedure of lowering gravity meters to the sea floor. Subsequent surface measurements over the same area permit G to be estimated from the density and depth of the water. It is important to make the maximum use of such data, since they avoid one of the doubts that beset the mine and borehole data: there is no way that we can admit a 1% uncertainty in the density of sea water. Limited comparisons of surface and sea-floor gravity data have been made by the U.S. Geological Survey, with a reported hint of a systematic discrepancy [11], but from our examination of the data they appear inadequate to draw any conclusion.

By far the best data set that we have been able to discover and probably the only one in existence that is really useful for our purpose is held by the Exxon Company's Exploration Department. With their collaboration we have extracted from these records 703 pairs of sea-floor and corresponding sea-surface gravity values equally spaced over an approximately square area of 1100 km^2 in the Gulf of Mexico in water up to 700 m deep. The gravity differences show considerable scatter, which has at least two causes. One is noise in the sea-surface data, which are less accurate than the sea-floor data. The other is an effect related to topography of the sea floor. The survey data were corrected for sea-floor topography, but the correction may be imperfect, or else the data are influenced by sub-bottom mass anomalies that are correlated with topography. In this circumstance it is important that we have data uniformly spread over a substantial area, because the area mean of the gravity gradient averages the topographic effects. The mean vertical gradient of gravity in the sea over this area is $(2.2124 \pm 0.0020) \times 10^{-6}$ s^{-2}, which gives $G = (6.796 \pm 0.016) \times 10^{-11}$ m^3 kg^{-1} s^{-2}. As we point out [3], this estimate depends crucially upon the precision of the baseline ties of the two surveys. But the trend of the variation of gravity differences (Δg) with depth (z) is independent of these ties and if instead we seek a method of eliminating the topographic effect, we can use this trend. Although such an approach is necessarily rather qualitative, it does not appear reasonable to make assumptions that reduce the estimate of G below about 6.73×10^{-11} m^3 kg^{-1} s^{-2}, unless in

TABLE 1. *Estimates of G from mine and borehole data.*

Data source	Type of data	Depth range (m)	$(G \pm \sigma)(10^{-11}\,m^3\,kg^{-1}\,s^{-2})$
Whetton *et al.* [6]	Mine	96–587	6.795 ± 0.021
McCulloh [7]	Mine	57–685	6.7390 ± 0.0025
		57–208	6.724 ± 0.014
		223–388	6.726 ± 0.012
		418–685	6.747 ± 0.013
Hinze *et al.* [8]	Borehole	3712–3962	6.81 ± 0.07
Stacey *et al.* [5]	Mine	0–948	6.712 ± 0.037
Hussain *et al.* [9]	Mine	251–590	6.705 ± 0.060

this case also we happen to have hit upon an extensive low in the free air gradient of gravity.

5. Bathyscaphe Measurements of Gravity

In collaboration with staff of Texas A & M University we are planning to obtain a set of vertical gravity profiles in the deep sea, using the submersible Alvin, which has a maximum depth capability of 4000 m. That a free-floating bathyscaphe may be steady enough to permit such gravity measurements was demonstrated by Drake and Delauze [12], although their data were obtained in a sea-floor trench and the very anomalous gradient could be explained as a very local effect if the topography is much steeper than their map indicates. We are taking the additional precaution of using a stabilized platform for the gravity meter. Corrections for any slight vertical accelerations will be made from the doubly differentiated record of a very sensitive pressure gauge. The initial series of measurements is planned for the Sigsbee Abyssal Plain in the Gulf of Mexico, in an area selected as gravitationally and topographically featureless. In principle such measurements permit a very precise determination of G [13].

If the measurements are successful they will provide the largest scale determination of G. Measurements at the surface will not be possible as the gravity meter will not be adequately stabilized against wave action, but the available depth range below the effect of the waves is about 3500 m.

6. The Splityard Creek Gravity Experiment

Splityard Creek is a minor tributary of the Brisbane River in S.W. Queensland, just upstream of the now nearly complete Wivenhoe Dam across the river. A subsidiary dam across the creek will impound an additional volume of water at a higher level than the main lake and will be used as a hydroelectric pumped-storage reservoir. In the course of its operation the level will rise or fall by as much as 10 m in a few hours, depending upon the demand for electricity, remaining constant for a few hours before changing again. The situation is ideal for a fundamental experiment on gravity. A specially constructed electricity pylon near the middle of the lake will support an observing platform for our experiments (Fig. 1).

The principal experiment uses a vacuum balance to compare the weights of 10 kg masses, hanging (in evacuated tubes) at different levels in the lake. Considering two masses, one above high water and the other below low water, a measured rise in water level introduces an approximately infinite plane layer of uniform density ρ between the masses, pulling one up and the other down. For a layer of water of depth $h = 10$ m the gravity difference between the masses that is caused by the water layer is

$$\Delta g = 4\pi G \rho h = 8.38 \times 10^{-6} \text{ m s}^{-2} \qquad (7)$$

which is nearly 1 part in 10^6 of ambient gravity and so produces an unbalance of this magnitude in the comparison of the two weights. Assuming an initial balance of two masses m (10 kg) the balance may be restored by an additional mass Δm, where

$$\Delta m = m \, \Delta g / g \approx 8.5 \text{ mg} \qquad (8)$$

If the design sensitivity is achieved we will be able to measure an unbalance of this magnitude to 3 parts in 10^5. This sensitivity is based on some special design features

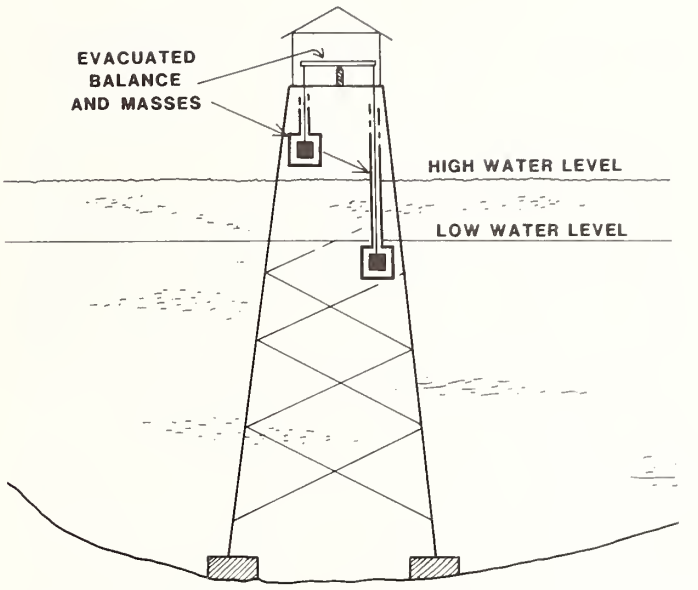


FIGURE 1. *Schematic view of the Splityard Creek gravity experiment. A balance compares the weights of masses suspended in evacuated tubes at different depths in a lake as the lake level is changed. The balance is supported by an observing platform on an electricity pylon which minimizes corrections for the absence of water in the volume occupied by the material of the pylon.*

that are indicated in Fig. 2.

Our balance design is based on the principles of capacitance micrometry that have had extensive use in our laboratory for more than 10 years [14]. Horizontal positioning of the balance arm is measured relative to a reference plane in which there are electrodes activated by a ratio transformer bridge operating at 3 kHz. This plane supports interconnected pools of mercury which are used to sense the adjustment of the plane to a perfectly horizontal position, the capacitances to the two pools being equal at perfect balance. The upper faces of the two electrodes operate a similar bridge with the balance arm, whose deparature from horizontal is therefore indirectly referred to the mercury levels. The balance arm itself is supported via its knife edge on an insulated support connected to a bridge preamplifier which is mounted on the reference plane. It is important to the bridge operation that, at balance of the bridge, electrostatic forces are also balanced and the balance arm is at earth potential, although as a precaution $1/\sqrt{2}$ switch is fitted to the bridge drive source. By dividing all voltages by $\sqrt{2}$, all electrostatic forces and any consequent unbalance will be

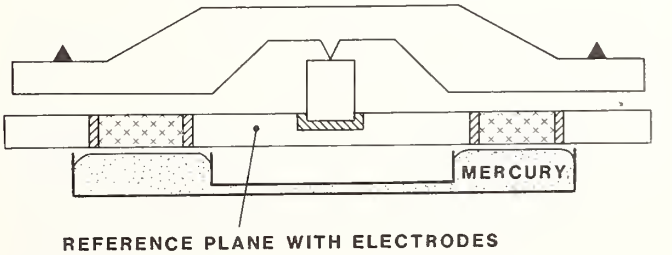


FIGURE 2. *Diagrammatic section of balance components for the Splityard Creek gravity experiment. The position of the balance arm is sensed by electrodes mounted in the reference plane on which it is mounted and this in turn is adjusted to horizontal by capacitance sensing to interconnected mercury pools.*

halved and the balance reading appropriate to zero applied voltage is readily calculated. The unbalance of the bridge output is used in a servo-circuit to apply restoring forces via additional electrodes in the reference plane (not shown). These are also used to apply critical damping. Thus the balance reading is of a restoring voltage that must be calibrated in terms of precisely known masses added to one side or the other.

Four suspended masses will be used, one above the high water level and one below on each side of the balance, the two masses down in the lake being at different levels. An electromechanical system for interchanging supports to the masses is built into the balance. In this way masses supported at different levels by the same balance arm can be repeatedly interchanged and precisely compared and dimensional stability of the balance arms over several hours is not required. The ends of the balance arm support the suspended masses not by conventional knife edges, but by diamond points, because this makes the exact point of support unambiguous. (The necessary parallelism of three knife edges cannot be achieved.)

Apart from the operation of the balance itself, the only quantities that must be measured accurately are the depth and density of the lake water. The shapes of the suspended masses and the geometries of the lake perimeter and supporting members of the instrument tower (where these represent absence of water) are important to the attainment of a very precise result but are not critical. Regional gravity inhomogeneities, tidal effects, depression of the lake floor by the water load and similar effects are either constant or affect all suspended masses so nearly equally that they can be ignored. The effect of observer mass is negligible at 8 m, which presents no difficulty as the balance is connected to its control console via an umbilical cord longer than this.

Details of corrections to be applied for the finite lake area (and irregular outline) are of some interest. The gravity due to a large sheet of density ρ and thickness dz at a point distant a from it, where a is much less than the distance r to any point of the perimeter is

$$dg = 2\pi G\,\rho\,dz\,(1 - a\,\langle r^{-1}\rangle + (a^3/2)\,\langle r^{-3}\rangle + \ldots) \quad (9)$$

where $\langle r^{-1}\rangle$ is the azimuthally averaged reciprocal shoreline distance. This binomial expansion shows that there is a first order correction and third and higher order corrections, but no second order terms. Since $a/r \approx 2 \times 10^{-3}$, third order terms are insignificant. Thus we can write the gravity difference between points at distances a above and $(z - a)$ below the layer

$$\Delta(dg) = d(\Delta g) = 4\pi G\,\rho\,dz\,(1 - (z/2)\,\langle r^{-1}\rangle), \quad (10)$$

where z is now identified as the vertical separation of two masses whose weights are compared. Now we can integrate over a finite layer whose outline varies over its total depth range, h, without complicating the form of the correction equation:

$$\Delta g = 4\pi G\,\rho\,h\,(1 - (z/2)\,\langle r^{-1}\rangle). \quad (11)$$

For a reason that follows, the lake perimeter is being carefully surveyed at all water levels of interest, but for the purpose of this correction we can estimate $\langle r^{-1}\rangle$ for the layer between any two levels from measurements on two pairs of masses at different vertical separations z. This is achieved by using the four masses referred to. The survey is a valuable check anyway, but is being done for another particular reason, which is to orient the balance so that $\langle r^{-1}\rangle$ is precisely the same at both ends of it. That this geometry has been achieved will be checked by turning the balance around between observing runs.

7. Conclusions

While the present geophysical evidence for a non-Newtonian gravitational effect lacks the observer control necessary to eliminate possibilities of systematic error, the fact that all of the available data give G at long range higher than G^* is circumstantial evidence that the effect is real. It is also of interest that the effect is of the same sign as that reported by Long [15] on the basis of experiments at laboratory ranges. Since we are contemplating an effect of order 1% for measurements on the scale of hundreds of meters, it cannot be long before sufficiently well controlled large scale experiments settle the matter conclusively.

The two experiments we have described are complementary. The bathyscaphe experiment is the largest scale G measurement that we have been able to devise and will be vital first in demonstrating whether a breakdown in Newton's law at kilometer range is positively demanded and, if so, then in deciding whether the upper range limit of the breakdown is within an observable range. If a breakdown is indicated then the greater accuracy of the lake experiment will be needed to indicate clearly the form of the breakdown. This experiment has all the advantages of a controlled laboratory experiment, with masses repeatedly introduced to and removed from an otherwise immobile apparatus. Further, it has several advantages, even over conventional G measurements, arising primarily from the geometrical simplicity and symmetry of the large mass (a layer of lake water). Corrections for attraction to the balance itself, extraneous cross attractions to masses on the wrong side of the balance, and errors arising from inhomogeneities of the attracting and attracted masses are all designed out of the experiment. Thus the lake experiment should provide a useful estimate of G quite apart from the search for non-Newtonian effects.

Our work is supported by the Australian Research Grants Committee and by Mount Isa Mines Ltd. Important ideas of Mr. G. Bell are incorporated in our balance design.

References

[1] J. H. Poynting, *The Mean Density of the Earth* (Griffin, London, 1894).
[2] K. E. Bullen, *The Earth's Density* (Chapman and Hall, London, 1975).
[3] F. D. Stacey and G. J. Tuck, Nature 292, 230 (1981).
[4] G. B. Airy, Philos. Trans. R. Soc. London 146, 297 and 343 (1856).
[5] F. D. Stacey, G. J. Tuck, S. C. Holding, A. R. Maher, and D. Morris, Phys. Rev. D 23, 1683 (1981).
[6] J. T. Whetton, J. O. Myers, and R. Smith, Geophysical Prospecting 5, 20 (1957).
[7] T. H. McCulloh, Geophysics 30, 1108 (1965).
[8] W. J. Hinze, J. W. Bradley, and A. R. Brown, J. Geophys. Res. 83, 5864 (1978).
[9] A. Hussain, G. Walach, and F. Weber, Geophysical Prospecting 29, 407 (1981).
[10] E. R. Cohen and B. N. Taylor, J. Phys. Chem. Ref. Data 2, 663 (1973).
[11] M. J. Yellin, U.S. Dept. of Commerce, ESSA Operational Data Report C & GSDR-2 (1968).
[12] C. L. Drake and H. Delauze, Ann. Inst. Oceanog. (Paris) 46, 7 (1965).
[13] F. D. Stacey, Geophys. Res. Lett. 5, 377 (1978).
[14] F. D. Stacey, J. M. W. Rynn, E. C. Little, and C. Croskell, J. Phys. E (J. Sci. Instrum.) 2, 945 (1969).
[15] D. R. Long, Nature 260, 417 (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Experimental Test of a Spatial Variation of the Newtonian Gravitational Constant at Large Distances*

H. A. Chan and H. J. Paik

Department of Physics and Astronomy, University of Maryland, College Park, MD 20742

The Poisson equation of Newtonian gravitational potential provides a source-independent null test of the Inverse Square Law. A convenient Laplacian detector consists of superconducting gravity gradiometers in three orthogonal directions. Matching and stability of the cryogenic detector are achieved by utilizing superconducting circuits. Since the Laplacian of the gravitational potential produced by an arbitrary source is zero outside the source in the Inverse Square Law, this experiment becomes a source-independent null test for the constancy of the gravitational constant. This characteristic allows a precision test of the Inverse Square Law at geological distances using natural objects like an ocean or the earth. We discuss experimental procedures and expected sensitivities of the null experiment for three different sources: a swinging pendulum, an ocean tide, and the earth itself. It appears that the empirical limits in the Inverse Square Law could be improved by three to six orders of magnitude in the range between 1 m and 10^7 km by this new null experiment.

Key words: gravitational constant; gravitational null experiment; inverse square law of gravitation; superconducting gravity gradiometer.

1. Introduction

The gravitational constant G is considered to be constant both with time and space in Newton's Universal Law of Gravitation and Einstein's General Theory of Relativity. It is possible, however, to make G a function of time [1] or a function of mass separation [2] within metric theories of gravity. A spatial variation of G has also been predicted on other theoretical grounds [3-6]. Most of these theories favor a functional dependence

$$G(R) = G[1 + \alpha(1 + \mu R)\, e^{-\mu R}], \tag{1}$$

which arises from a potential of the form

$$\phi(R) = \frac{GM}{R}\,(1 + \alpha e^{-\mu R}). \tag{2}$$

Therefore, a spatial dependence of G, if detected, could imply the existence of an additional short-range force which is superposed over the Newtonian long-range force rather than necessarily a failure of the Inverse Square Law.

Various authors have pointed out [7-9] that the existing data on absolute G and the product GM for the earth and other celestial bodies cannot rule out the possibility of α having a value as large as 1/3, if the range μ^{-1} falls somewhere between 10 m and 10 km. Efforts have been made recently to extend the Cavendish-type experiments up to a mass separation of $R \sim 10$ m [10-14]. Inference from orbits of artificial satellites on the constancy of G for $R \lesssim 10^3$ km is difficult on account of the rapid fall-off of any non-Newtonian signal as μ^{-1} is reduced below the satellite altitude. As a result, the intermediate distance range 10 m $\lesssim R \lesssim$ 10 km has been left largely untested to this date. It is highly desirable to be able to examine $G(R)$ on this range and improve the overall experimental limits in the constancy of G as a function of R. Such an experiment will test the scale invariance of the Universal Law of Gravitation and help settle issues raised by opposing theories of gravity.

*Supported in part by NASA.

In this paper, we discuss a series of experiments which could cover the desired geological scale and might improve the empirical limits on α at large distances by several orders of magnitude. These experiments are based on the principle of a *null* experiment proposed by Paik [15]. The "source-independent" nature of the experiments permits use of a large natural object like a mountain, an ocean tide, or the earth itself as a source to examine the Inverse Square Law at the characteristic distances of these objects. In the following sections, the principle of the gravitational null experiment is reviewed, experimental procedures for a laboratory, geological-scale, and earth-orbit experiment are discussed, and the expected resolutions of these experiments are presented. In addition, the principle and design of the gravitational null detector are briefly described.

2. Principle of a Null Experiment

The "source independent" null experiment invokes a well-known theorem in gravitostatics: Gauss's law for Newtonian gravity. In differential form, it is embodied in a Poisson equation

$$\nabla^2\phi(\mathbf{x}) = -\nabla \cdot \mathbf{g}(\mathbf{x}) = 4\pi G\,\rho(\mathbf{x}), \tag{3}$$

where $\mathbf{g} = -\nabla\phi$ is the gravitational field (force per unit mass). Since the unique solution of Eq. (3) is the $1/R$ potential, which is characteristic of an inverse square force law, it is clear that the Poisson equation for ϕ is equivalent to a constant G, $G \neq G(R)$, in the force equation

$$\mathbf{g}(\mathbf{x}) = -\int_V G(R)\,\frac{\rho(\mathbf{x}')}{R^2}\,d^3x', \tag{4}$$

where $R \equiv |\mathbf{x} - \mathbf{x}'|$. Notice that the Laplacian of the Newtonian potential, $\nabla^2\phi_N$, vanishes *identically* in vacuum, *independent* of mass distribution in the rest of the universe.

Hence, a *null* experiment for the constancy of G can be performed in the following way. First, construct a detector sensitive to $\nabla^2\phi$, which we call a gravitational "La-

placian detector." Second, using a large object of an arbitrary shape, modulate the source-detector separation periodically. Third, average the $\nabla^2\phi$ signal over many periods synchronously with the periodic motion. A statistically significant departure of $\nabla^2\phi$ from zero would then constitute a violation of the Inverse Square Law at the particular distance chosen for the experiment.

Unlike the Cavendish experiment, which basically determines G by comparing the integral in Eq. (4) with a measured force, the present experiment tests a differential equation which connects the field to a *local* quantity. As a result, the new experiment is insensitive to the global mass distribution of the source. This permits large geological objects of irregular shape to be used as a source. The source-detector separation is modulated periodically to overcome the 1/f noise in the Laplacian detector and to discriminate against gravity produced by stationary objects at other distances.

For a point source whose potential is given by Eq. (2), the field equation is modified by

$$\nabla^2\phi(R) = \frac{GM}{R^3}\,\alpha(\mu R)^2 e^{-\mu R}. \tag{5}$$

Of course, the right-hand side of this equation becomes a volume integral for an extended source:

$$\nabla^2\phi(R) = G\int_V \frac{\rho(\mathbf{x}')}{R^3}\,\alpha(\mu R)^2 e^{-\mu R}\,d^3x', \tag{6}$$

where $R \equiv |\mathbf{x} - \mathbf{x}'|$ as before. If the Inverse Square Law should fail at certain distances, the exact functional form of the gravitational potential could be determined by integrating the new field equation, Eq. (5).

3. Gravitational Laplacian Detector

The Laplacian of the gravitational potential is, by definition, a sum of three orthogonal gravity gradients (the trace of the gravity gradient tensor):

$$\nabla^2\phi = -\nabla\cdot\mathbf{g} = -\sum_{i=1}^{3}\frac{\partial g_i}{\partial x_i}, \tag{7}$$

where the x_i (i = 1,2,3) form any Cartesian coordinates. A three-axis in-line component gravity gradiometer[1] could, therefore, be employed as a Laplacian detector. A sensitive superconducting gravity gradiometer to be used for this experiment has been designed and is under construction [17,18]. Here we give a brief description of the instrument.

Figure 1 is a schematic circuit diagram of a single-axis in-line component gravity gradiometer. The shaded rectangles with arrows represent superconducting proof masses that are confined to move along the direction indicated by the arrows. A persistent current I_0 is stored with opposite polarity in each of the two loops formed by pairs of flat superconducting sensing coils. When each proof mass is displaced with respect to the coils in response to a gravitational force or a platform vibration, it causes a current to flow into and out of the persistent current loop due to the Meissner effect. The signals from the two accelerometers are directly added and subtracted by two SQUIDs (Superconducting Quantum Interference Devices) to obtain a common-mode force g_i and a force gradient $\Gamma_{ii} \equiv -\partial g_i/\partial x_i$. A Laplacian detector is composed of three such single-axis gradiometers assembled along three orthogonal directions.

[1] A similar null experiment in which a spherical mass shell is used as a Laplacian detector has been discussed by Paik [15] and proposed by Mills [16] independently.

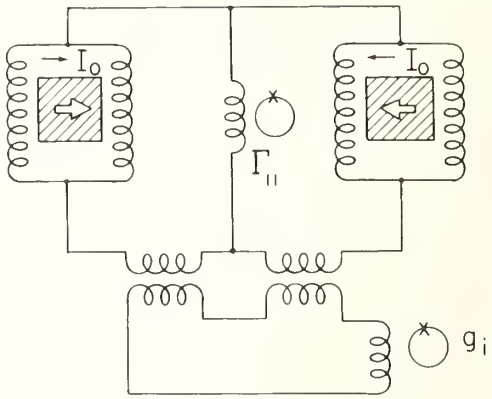


FIGURE 1. *A schematic circuit diagram of a single-axis superconducting gravity gradiometer. A Laplacian detector is composed of three single-axis gradiometers repeated in three orthogonal directions.*

In order to reject the common-mode forces to high precision and realize a very sensitive gravity gradiometer, the instrument must be fabricated with great care in a number of areas. The suspension springs for each proof mass must be soft along a well-defined axis and be linear. The sensitive axes of the two accelerometers in each gradiometer must be aligned. The sensitive axes of three gradiometers must then be aligned orthogonal to each other and the scale factors matched.

Figure 2 is a cut-away view of the superconducting proof mass and suspension. The suspension structure containing eight "folded cantilevers" is machined out of a single block of niobium (Nb). Plane surfaces are lapped parallel to each other before cantilevers are cut with an

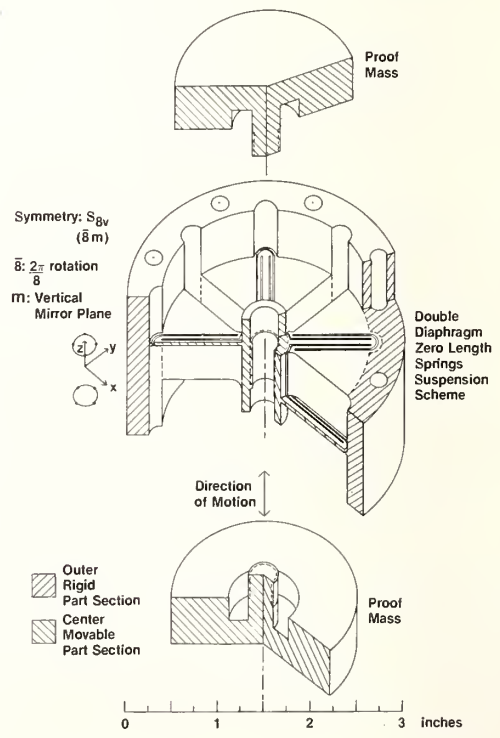


FIGURE 2. *A 55° cut-away view of the niobium proof mass and suspension of a superconducting accelerometer. Two halves of the proof mass are threaded into the suspension structure to be supported by eight "folded cantilevers."*

electric discharge machine. After the two halves of the Nb proof mass are threaded into the suspension structure, two "pancake"-shaped Nb coils each wound in a single layer are brought near the two outer surfaces of the proof mass to form an accelerometer. To make a Laplacian detector, six such accelerometers are mounted on six faces of a titanium cube whose deviations from orthogonality are controlled to a few arcseconds. Each pair of accelerometers located on opposing faces of the cube are coupled by the superconducting circuit shown in Fig. 1. The common-mode rejection error in each axis due to misalignment of the accelerometer axes becomes a second order effect when the cube is turned into three angular positions [19]. The uncompensated orthogonality error in the three axes gives rise to a coupling of 10^{-5} to zeroth order Newtonian gradients. The folded cantilever flexure employed here is expected to give a highly linear spring due to its pure bending motion, as well as low resonance frequencies of the proof masses.

In order to find $\nabla^2\phi$ by summing the outputs of three single-axis gradiometers, their scale-factors must be matched precisely. A support structure which enables common-mode balance and scale-factor match of a three-axis gravity gradiometer is shown in Fig. 3. An "umbrella" orientation, in which the three sensitive axes make an equal angle $\tan^{-1}\sqrt{2}$ with respect to the vertical, matches the gravity bias in three gradiometers. For common-mode balance, the gradiometer assembly is driven vertically at a desired frequency by means of a superconducting magnet. The ratio of two supercurrents in each of the three gradiometers is then adjusted until the differential output caused by the applied linear acceleration becomes sufficiently low. The scale-factor match between the three orthogonal components is accomplished by rotating the device in precise 120° steps around the vertical and comparing outputs of the three gradiometers for the same gravity gradient signal. A set of three positioning pins is located in the support structure for this purpose.

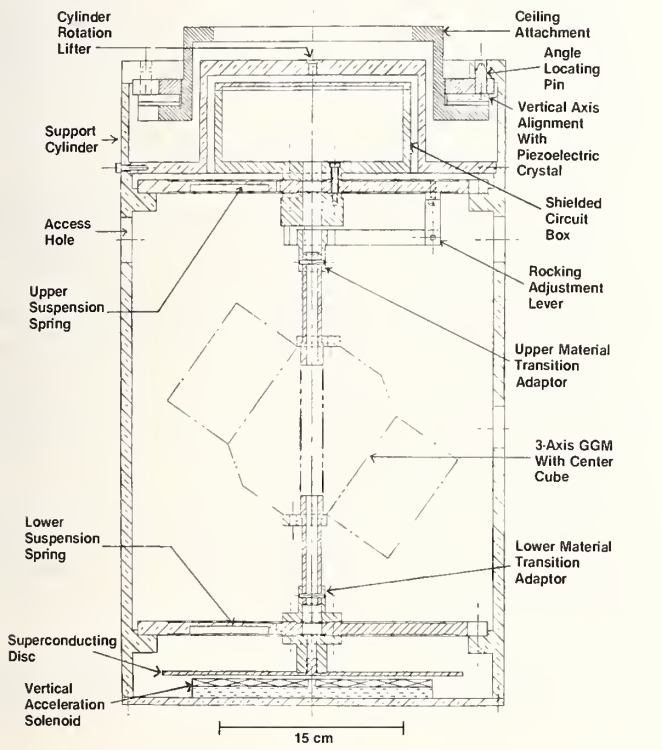


FIGURE 3. *A support structure which enables common-mode balance and scale-factor match of a three-axis gravity gradiometer in an "umbrella" suspension.*

A single-axis portion of the Laplacian detector described here has been constructed and is undergoing experimental test. The three-axis support structure is in the final stages of fabrication. An initial laboratory test of the Inverse Square Law is being prepared. The instrument-noise-limited sensitivity of the prototype Laplacian detector under construction is expected to be approximately 2×10^{-11} $s^{-2}Hz^{-1/2}$. For the earth-orbit experiment, the goal is to achieve a white noise level of 2×10^{-13} $s^{-2}Hz^{-1/2}$ in a larger model of the Laplacian detector. These sensitivities will be assumed to compute a minimum detectable α as a function of μ^{-1} in Section 5.

4. Experimental Procedures

4.1 Laboratory Experiment

In the laboratory scale $R \sim 10$ m, a dynamic gravitational field could be produced conveniently by a moving source while the detector is kept at rest. In our earlier paper [20], we considered an experiment in which a rotating dumbbell is used to produce a periodic *quadrupole* field as in the experiment of Hirakawa *et al.* [13]. Here we replace the rotating dumbbell by a swinging pendulum. The latter generates a periodic *monopole* field which falls off more slowly ($\sim R^{-3}$) than the quadrupole field ($\sim R^{-5}$) and allows a larger separation experiment.

Figure 4 shows an experimental configuration of the source and the detector. Typical dimensions are $l = 0.16$ m, $h \simeq 5$ m, $u \simeq 2$ m, and $R \simeq 10$ m. For the pendulum mass, we are constructing a spherical lead ball of 1600 kg. This will be used to test various instruments and procedures. Eventually the pendulum will carry a 10^4 kg stack of lead bricks for the Inverse Square Law experiment. If direct transmission of vibration through the building structure proves to be important, one could attempt to cancel the vibration using two identical pendula swinging in opposite directions. The detector could be located at the midpoint to double the gravity signal. This symmetric source configuration has an additional advantage of producing a uniform field in the middle thus making a distance error δR less important.

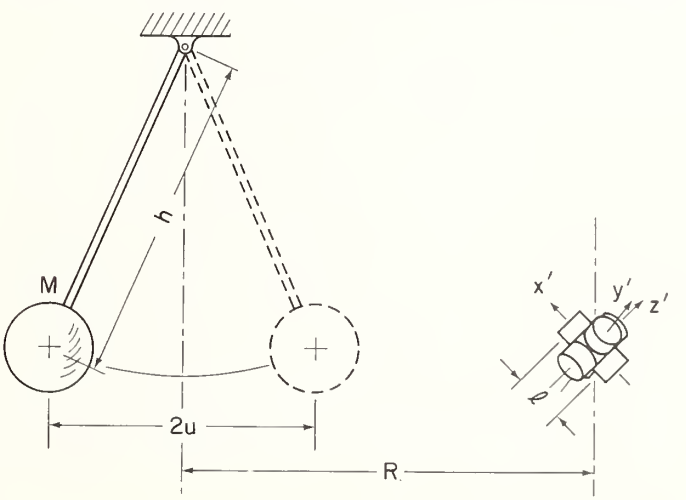


FIGURE 4. *Experimental configuration of the source and the Laplacian detector in a laboratory experiment.*

A completely source-independent null experiment requires a *true* Laplacian detector [21] which has a more sophisticated design. The three-axis gravity gradiometer that we are constructing at present is only an approximate Laplacian detector which still has a nonvanishing coupling to higher order spatial derivatives of ϕ. Hence, errors in the source geometry do not drop out completely. This residual coupling to the higher order Newtonian terms arises from the fact that a practical gradiometer measures a *difference* over a finite baseline l rather than taking a derivative at a point. For a potential given by Eq. (2), one can show [20] that the output of the three-axis gravity gradiometer is

$$\sum_{i=1}^{3} \delta g_i = \frac{GMl}{R^3} \left[\alpha(\mu R)^2 e^{-\mu R} - \frac{21}{8} \left(\frac{l}{R} \right)^2 \times \left[1 - \frac{5}{3} \frac{X^4 + Y^4 + Z^4}{R^4} \right] + O\left[\frac{l}{R} \right]^4 \right], \quad (8)$$

where $\mathbf{R} = (X, Y, Z)$ and $R = |\mathbf{R}|$. Thus, the finite baseline term varies as $(l/R)^2$ and becomes important for a relatively short distance experiment.

In the laboratory experiment, $(l/R)^2 \simeq 3 \times 10^{-4}$. Since the zeroth order Newtonian term is balanced out to 10^{-5}, the second order term needs to be compensated by computation to about 3% to bring down the errors to the same level. In order to model the Newtonian gravity correctly, we drive the pendulum to swing in a predetermined plane and read the actual pendulum position using a shaft encoder mounted at the pivot. The encoder output is then used to trigger a signal averager following the Laplacian detector. In this manner, errors associated with erratic pendulum motions are prevented from entering into the signal process. The signal is averaged over many days to improve the signal-to-noise ratio.

4.2 Geological-Scale Experiment

The source-independent nature of the Laplacian experiment permits precision tests of the gravitational force law at kilometer ranges using geological sources. Figure 5 illustrates experimental arrangements for two different sources: a mountain and an ocean tide. When a mountain is used as the source, the detector could be transported horizontally, modulating the horizontal distance to the mountain $d(t)$. An ocean tide gives the advantage of not having to move the detector at all though the signal frequency of 2.3×10^{-5} Hz is well inside the 1/f noise of the Laplacian detector. The inherent density homogeneity and the well-defined surface profile of water make the tide a very attractive source for a geological-scale gravity experiment [22].

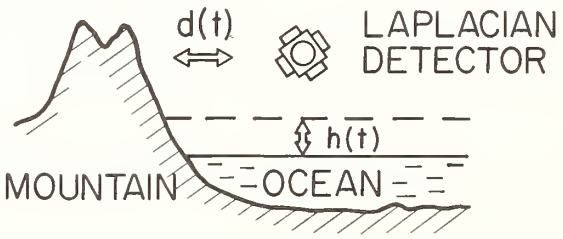


FIGURE 5. *Experimental arrangements for two different geological sources: a mountain and an ocean tide. For the mountain experiment, the detector is moved horizontally to modulate d(t). For the ocean tide experiment, the tide modulates a vertical distance h (t) while the detector is kept at rest.*

Moving a sensitive Laplacian detector without deteriorating its performance poses a major challenge in the mountain experiment. The problem could partially be overcome by modulating $d(t)$ in a square-wave fashion and averaging signals at the two end points while the detector is at rest. Although the Laplacian detector would still have to be moved gently so as to remain within the dynamic range of the SQUID detection circuits, dynamic error sources such as platform vibration and jitter could be eliminated by this procedure.

For an ocean-tide experiment, the cryostat containing the Laplacian detector could be fixed on a support tower above the water level or be lowered to the basin. The dynamic source of gravitation is the water level difference between the ebb and flow of the tide, $h \pm u$. This can be approximated as a plane sheet of mass with uniform density and thickness $2u$ for ranges smaller than the closest distance to the shore. It is well known that such a geometry produces a zero Newtonian force gradient so that our experiment becomes a *double null* experiment: a null source and a null detector.

Since the height modulation ($2u$) that is produced by an ocean tide is approximately 10 m, the resolution in α deteriorates at $\mu^{-1} > 10$ m. It is tempting to move the detector vertically on a flat plane[2] to a larger amplitude and improve the resolution at larger distances. In such an experiment, however, one encounters a large modulation of the vertical Newtonian gravity of the earth. Hence it is necessary to improve the common-mode rejection in the vertical direction.

4.3 Earth-Orbit Experiment

The three-axis superconducting gravity gradiometer described in this paper is under development for a gravity survey satellite of NASA [18]. After a successful test of the prototype model under construction, a larger, more sensitive model may be built and flown in a low altitude earth orbit to take a high resolution gravity map of the earth. Such a mission will give an excellent opportunity to test the Inverse Square Law at a distance range of the order of the earth's radius and improve the experimental limit of the Law by several orders of magnitude at $\mu^{-1} = 10^2$ to 10^3 km [23].

For the gravitational null experiment, the satellite could initially be launched in an elliptical orbit which modulates the satellite altitude from h_1 to h_2 at approximately 1.5×10^{-4} Hz as shown in Fig. 6. A full modula-

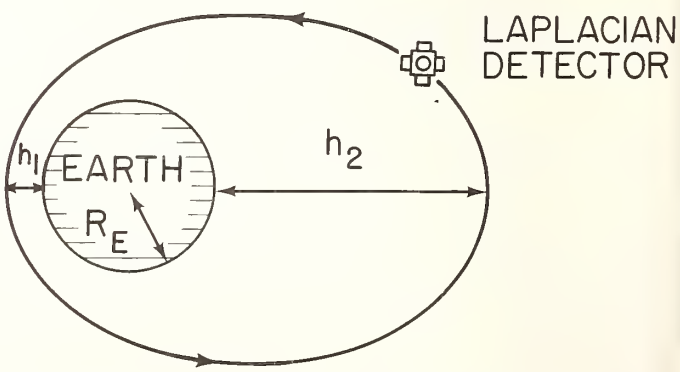


FIGURE 6. *The Laplacian detector in an elliptical earth orbit.*

[2]An elevator in a rocket launch tower or a tall urban tower could be used to transport the cryostat vertically.

tion of the non-Newtonian signal will be achieved if $h_1 \simeq 200$ km and $h_2 \gtrsim 6000$ km. There are several advantages of the earth-orbit experiment over the geological-scale experiment performed on the earth's surface. The quiet, zero-g nature of space provides an ideal environment for operation of a sensitive gravity gradiometer. The orbital motion of the satellite gives a continuous and full modulation of gravity signals at a reasonable frequency. Irregularities of the source are partially averaged out by the spin of the earth. It is also important that the satellite itself can be spun quietly in space to eliminate some of the important error terms.

The sensitivity of the flight model of the Laplacian detector is expected to be $10^{-6}\Gamma_E$ Hz$^{-1/2}$ at the signal frequency of 1.5×10^{-4} Hz, where $\Gamma_E = 3 \times 10^{-6}$ s^{-2} is the vertical bias gradient of the earth's gravity. If the signal is integrated for $\tau = 10^6$ s (12 days), the instrument noise level will go down to approximately $10^{-9}\ \Gamma_E$ implying a resolution of one part in 10^9 in the Inverse Square Law at the earth's radius $R_E = 6400$ km. An orthogonality error in the Laplacian detector, without compensation, would prevent a resolution better than one part in 10^5. Fortunately, there are ways in which one could eliminate this error to first order and thereby reduce the coupling to the Newtonian term to a level of 10^{-10}. One interesting approach is spinning the satellite around one of the sensitive axes of the Laplacian detector. Coupling to cross component gravity gradients arising from angular errors is then modulated at harmonics of the spin frequency and can be distinguished from the Laplacian signal which remains at dc. Errors in the Laplacian itself can be shown to be of the second order in misalignment angles for the particular choice of the spin axis. A more detailed analysis of the earth orbit experiment will be published separately [23].

5. Expected Resolution of the Null Experiment

In this section we discuss the expected resolution of various versions of the gravitational null experiment. The peak-to-peak amplitudes of non-Newtonian signals for various sources can be shown to be

$$(\nabla^2\phi)_{p-p} = \begin{cases} \dfrac{2GM}{R^3}\dfrac{\alpha(\mu R)^2 e^{-\mu R}}{1-(u/R)^2} \\ \times\left[\sinh(\mu u) + \dfrac{u}{R}\cosh(\mu u)\right], \quad \text{pendulum}, \\ 4\pi G\,\rho_w\,\alpha e^{-\mu h}\sinh(\mu u), \quad \text{ocean tide}, \\ 2\pi G\,\rho_E\,\alpha\left[(1+e^{-2\mu R_E}) - \dfrac{1}{\mu R_E}(1-e^{-2\mu R_E})\right] \\ \times\left[\dfrac{e^{-\mu h_1}}{1+h_1/R_E} - \dfrac{e^{-\mu h_2}}{1+h_2/R_E}\right], \quad \text{earth}, \end{cases} \tag{9}$$

where ρ_w and ρ_E are, respectively, the densities of ocean water and the earth, and the other parameters have been defined in earlier sections. We have assumed that the size of the pendulum mass is small compared to μ^{-1} or R and the earth is a sphere with a uniform density ρ_E.

The parameter values used for computation are $M = 10^4$ kg, $R = 10$ m and $u = 2$ m for the laboratory experiment; $h = 10$ m, $u = 5$ m and $\rho_w = 1.03 \times 10^3$ kg · m^{-3} for the geological-scale experiment; and $R_E = 6400$ km, $h_1 = 200$ km, $h_2 = 6400$ km, $\rho_E = 2.7 \times 10^3$ kg m^{-3} (surface density) for $\mu^{-1} \lesssim 10^3$ km and $\rho_E = 5.5 \times 10^3$ kg m^{-3} (mean density) for $\mu^{-1} \gtrsim 10^4$ km. The instrument sensitivities assumed for the three experiments are 2×10^{-11} s^{-2} Hz$^{-1/2}$ (at 0.3 Hz), 2×10^{-10} s^{-2} Hz$^{-1/2}$ (at 2.3×10^{-5} Hz) and 2×10^{-12} s^{-2} Hz$^{-1/2}$ (at 1.5×10^{-4} Hz), respectively, the latter two coming from the 1/f noise of the SQUID. In addition, an integration time of $\tau = 10^6$ s (12 days) has been assumed for all experiments.

Figure 7 shows the expected resolution $|\alpha_{min}|$ of the null experiment as well as limits set by other experiments as a function of μ^{-1}. The solid curves represent upper limits in α implied by previous experiments. Curves labeled "Panov," "Hirakawa," and "Newman" have been plotted from published results in Refs. [12-14]. The "lunar surface gravity" curve is the one obtained by Mikkelson and Newman [9]. The limit set by "LAGEOS-lunar ranging" has been obtained by comparing two recent data points in the measurements of the geocentric gravitational constant[3]: one determined from laser ranging on near-earth satellites [24] and the other from laser ranging of the moon [25]. The regions lying above the solid curves are forbidden by existing data. The shaded area labeled "Long" is the region to which α is limited by Long's experiment [10]. Long's positive result constitutes an exception in a general trend which favors the Inverse Square Law and is in direct contradiction with Newman's data [14]. The three dashed curves in Fig. 7 represent the resolutions expected from the three versions of the $\nabla^2\phi$ experiment considered in Section 4 and Eq. (9). Notice that a combination of the geological scale and the earth orbit experiment is capable of resolving α to better than 10^{-3} in ten decades over the range from $\mu^{-1} = 1$ m to $\mu^{-1} = 10^7$ km filling the gap between the laboratory and the astronomical scale. In the range of $\mu^{-1} \simeq 10^3$ km, the Laplacian experiment is expected to resolve the Inverse Square Law to better than one part in 10^9, matching the best limit obtained from solar system observations for much larger distances.

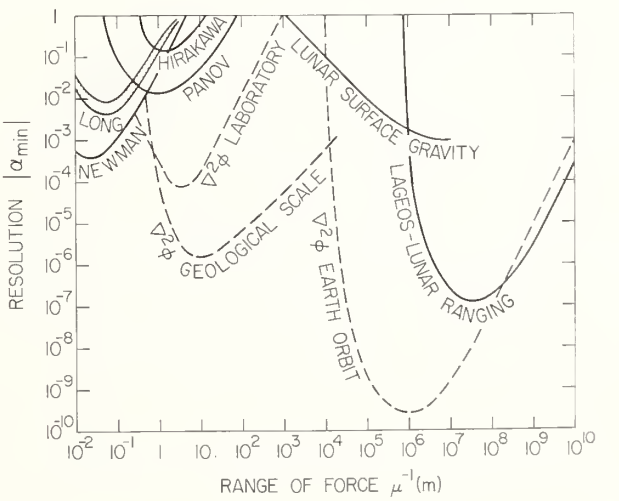


FIGURE 7. *Expected resolution in α of the Laplacian experiment (dashed curves) and limits set by previous experiments (solid curves) as a function of the range μ^{-1}. The null experiment is capable of resolving the Inverse Square Law to better than 10^{-3} in ten decades of range from 1 m to 10^{10} m.*

[3]The values of the geocentric gravitational constant in the two measurements are $GM_E = 398600.44 \pm 0.02$ km^3 s^{-2} at $R = 12300$ km (semimajor axis for LAGEOS) and $GM_E = 398600.461 \pm 0.026$ km^3 s^{-2} at $R = 384400$ km (mean earth-moon distance).

The proposed Laplacian experiment could remove embarrassingly large uncertainties that exist in our present knowledge of G as a function of mass separation. It is a *null* experiment which is capable of suppressing errors associated with the source. In the ocean and in earth orbit, the Newtonian terms could be eliminated completely by the additional null nature of the tide as source and by spinning the satellite quietly. As a result, excellent sensitivities are expected for the geological-scale and the earth-orbit experiment. The laboratory experiment is not completely free from source errors due to the proximity of the source and the detector. However, it will still be a useful first step in testing out the techniques of the gravitational null experiment.

In conclusion, a new concept has been developed to test one of the most fundamental hypotheses in the laws of physics: the *scale invariance* of the Inverse Square Law of Gravitation. The experimental examination of the spatial variation G constitutes a new test of General Relativity. An elegant null detector for the gravitational Inverse Square Law is being constructed using principles of superconductivity and an intricate mechanical design. We plan to carry out a series of null experiments in the coming years.

We gratefully acknowledge extensive contributions of Dr. Vol Moody and Mr. Frank Desrosier in the development of the superconducting gravity gradiometer and the construction of related experimental apparatus. We have also benefited from discussions with Dr. Peter Bender who has pointed out to us recent laser ranging data.

References

[1] C. Brans and R. H. Dicke, Phys. Rev. 124, 925 (1961).
[2] R. V. Wagoner, Phys. Rev. D 1, 3209 (1970).
[3] E. Pechlaner and R. Sexl, Commun. Math. Phys. 2, 227 (1966).
[4] J. O'Hanlon, Phys. Rev. Lett. 29, 137 (1972).
[5] Y. Fugii, Phys. Rev. D 9, 974 (1974).
[6] D. R. Long, Nuovo Cimento B 55, 252 (1980).
[7] Y. Fugii, Nature (London) Phys. Sci. 234, 4 (1971).
[8] D. R. Long, Phys. Rev. D 9, 850 (1974).
[9] D. R. Mikkelson and M. J. Newman, Phys. Rev. D 16, 919 (1977).
[10] D. R. Long, Nature (London) 260, 417 (1976).
[11] H. -T. Wu, W. -T. Ni, C. -C. Hu, F. -H. Liu, C. -H. Yang, and W. -N. Liu, Phys. Rev. D 19, 2320 (1979).
[12] V. I. Panov and V. N. Frontov, Sov. Phys. JETP 77, 1702 (1979).
[13] H. Hirakawa, K. Tsubono, and K. Oide, Nature (London) 283, 184 (1980).
[14] R. Spero, J. K. Hoskins, R. Newman, J. Pellam, and J. Schultz, Phys. Rev. Lett. 44, 1645 (1980).
[15] H. J. Paik, Phys. Rev. D 19, 2320 (1979).
[16] A. P. Mills, Jr., Genl. Relativ. Grav. 11, 1 (1979).
[17] H. J. Paik, E. R. Mapoles, and K. Y. Wang, in *Future Trends in Superconductive Electronics*, Ed. by B. S. Deaver *et al.* (Am. Inst. Phys. Conf. Proc. 44, New York, 1978), p. 166.
[18] H. J. Paik, J. Astronaut. Sci. 29, 1 (1981).
[19] H. A. Chan, unpublished analysis.
[20] H. J. Paik and H. A. Chan, in *Proceedings of the Second Marcel Grossman Conference on General Relativity*, Ed. by R. Ruffini (North Holland, Amsterdam, in press).
[21] H. A. Chan and H. J. Paik, in *Proceedings of the Ninth International Conference on General Relativity and Gravitation* (Jena, East Germany, 1980, in press).
[22] F. D. Stacey, Geophys. Res. Lett. 5, 377 (1978).
[23] H. J. Paik, paper in preparation.
[24] F. J. Lerch, R. E. Laubscher, S. M. Klosko, D. E. Smith, R. Kolenkiewicz, B. H. Putney, J. G. Marsh, and J. E. Brownd, Geophys. Res. Lett. 5, 1031 (1978).
[25] A. J. Ferrari, W. S. Sinclair, W. J. Sjogren, J. G. Williams, and C. F. Yoder, J. Geophys. Res. 85, 3939 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Measurement of G for Small Inter-Mass Spacings*

William C. Oelfke

Department of Physics, University of Central Florida, Orlando, FL 32816

An experimental project is currently under way at the University of Central Florida to measure the Newtonian gravitational constant G for inter-mass spacings r in the range 0.3 cm $> r >$ 3 cm. A Cavendish balance for measuring the gravitational attraction between disk-shaped masses has been constructed. This balance functions as a fully automated electrobalance with the incorporation of a microcomputer for the monitoring and recording of data and the manipulation of the balance. A complete description of the design and operation of this Cavendish balance is presented.

Key words: $G(r)$; inverse square law; measurement of G; short-range gravitational forces; torsion balance.

1. Introduction

A Cavendish balance has been constructed to measure the gravitational attraction between thin disks in order to provide a means of determining the gravitational constant G for inter-mass spacings in the range 0.3 cm $< r <$ 3 cm. It was pointed out in 1974 by Long [1] that the more recent values of G have been measured at inter-mass spacings of 7 to 10 cm and that although these measurements specify an accurate G they do not verify the inverse square law for gravity. In addition, historical data and some recent observations [2] seem to indicate that the gravitational forces do not strictly adhere to an inverse square law.

Theoretical arguments by Fujii [3] and O'Hanlon [4] followed by those of Scherk [5] tend to support these observations by suggesting that a massive spin 1 graviton may exist, along with the massless spin 2 graviton, that could give rise to an overall gravitational potential of

$$U(r) = -(GM/r)(1 - \tfrac{1}{3} e^{-mr}) .$$

The range, m^{-1}, of this force is unknown, however it may be assumed not to have a value around 10 cm because of the consistency of G measurements in this range. It may be ruled out, as well, in the range of orbital distances by astronomical observations. Experimental searches [6, 7] are therefore presently under way for short range forces consistent with m^{-1} in the ranges $m^{-1} < 10$ cm and 10 m $< m^{-1} <$ 1 km.

2. Experimental Design

This experiment is designed to measure gravitational forces between masses whose separations can be made as small as 0.3 cm. A flat disk geometry was chosen in order to provide a measurable gravitational force at these inter-mass spacings. In addition the Newtonian gravitational forces between such closely spaced disks has a much smaller gradient than the comparable force between spheres and as a result a Yukawa force component, if it existed, would become much more apparent. This is illustrated in Fig. 1 where a pure Newtonian gravitational force of attraction between the disks is compared with a hypothetical force with a Yukawa component of range $m^{-1} = 1$ cm.

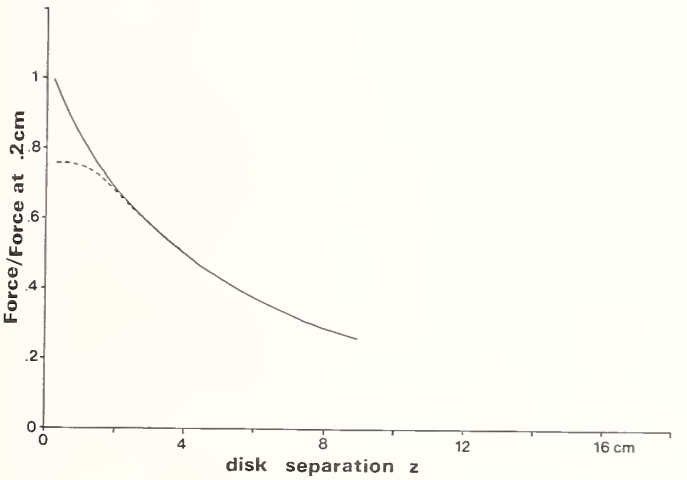


FIGURE 1. *The plot of the gravitational force of attraction between two thin disks as a function of disk separation z. The Newtonian force is plotted as a solid curve whereas the dashed curve shows the effect of adding the repulsive Yukawa type term for the case of* $m^{-1} = 1$ *cm.*

The gravitational balance is a symmetrical torsion balance of the Cavendish type with two equal suspended masses, each a disk 16 cm in diameter and 0.06 cm thick. Gravitational forces are exerted on these suspended disks by a symmetrical pair of disks, 20 cm in diameter and 0.3 cm thick. The balance, shown in Fig. 2, is 75 cm high with a torsion fiber length of 30 cm and a loaded torsion beam mass of 300 g. Each suspended disk has a mass of 100 g and each movable disk, a mass of 680 g. The suspended disks, torsion beam, and torsion fiber are shielded electrostatically and the entire balance is enclosed in a vacuum bell jar and evacuated to 10^{-5} Torr by an ion pump.

*Supported by a Joseph H. De Frees grant of the Research Corporation

The torsion balance is operated as an electrobalance so that any gravitational force tending to pull the torsion balance away from its equilibrium position is continuously balanced by electrostatic forces exerted by a pair of precision guard-ring capacitors located at the left end of the torsion beam as shown in Fig. 2. These precision capacitors can exert clockwise and counter-clockwise torques on the beam and provide a means for accurate measurement of gravitational torques. The voltages on these capacitors are controlled through digital-to-analog converters by a computer. Subroutines have been developed which provide for accurate measurement of the gravitational force while adjusting for torsion fiber drift and contact potential differences in the capacitors.

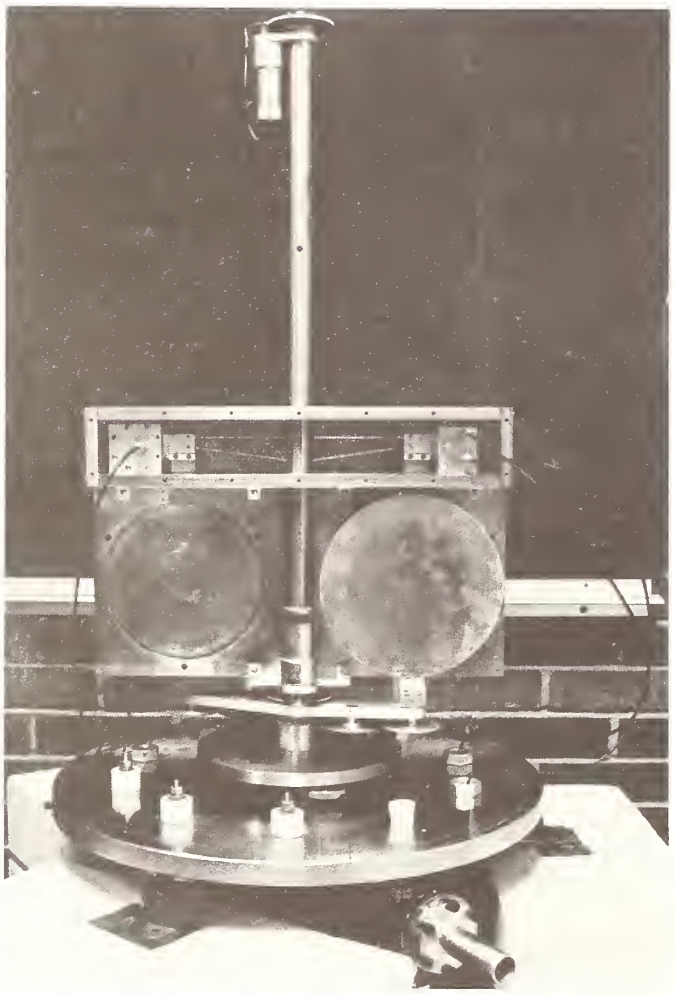


FIGURE 2. *The Cavendish balance for the measurement of short-range gravitational attractions.*

Beam position is sensed by a uhf resonant transducer system producing a voltage proportional to the angular displacement of the torsion beam from equilibrium. This transducer system, shown schematically in Fig. 3, is sensitive to angular beam displacements of 0.1 s of arc and provides a sufficiently large dynamic range to facilitate servo-locking the torsion beam to a stable null in an environment of floor noise and fiber drift common to all Cavendish balances.

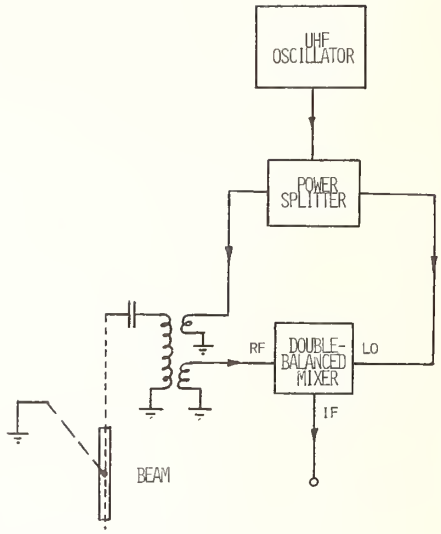


FIGURE 3. *The beam-position transducer system. Movement of the torsion beam changes the value of the electrical capacitance of the 500 MHz resonator producing a relative phase shift between the RF and the LO inputs to the double-balanced mixer.*

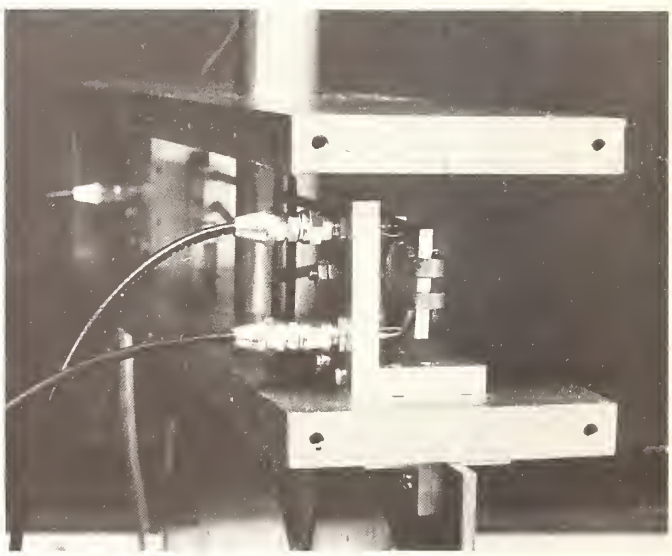


FIGURE 4. *The 500 MHz resonator showing the input and output coupling links, the inductance loop, and capacitors (partially hidden behind the ground plane).*

The uhf transducer shown in Fig. 4 is a unique 500 MHz resonator that consists of an inductance ring folded so that it is parallel to two semicircular, coplanar capacitance plates. A large capacitance of about 20 pF is established between these plates and the right end of the torsion beam which serves as a ground plane, sandwiched between the inductor and the capacitors. The electrical Q of this resonant circuit is 10^3 and when the resonator is operated as a transmission filter within its passband a small shift in the beam groundplane changes the resonant frequency of the transducer and produces a large phase shift in the output 500 MHz signal. The double-balanced mixer provides a means of detecting this phase shift and converting it to a voltage that can be read directly by an analog-to-digital converter.

Finally the manipulation of the large disks, the control of the electrobalance servo loop, correction of baseline drift, and the taking, processing and recording of data are all controlled by an LSI 11 microcomputer. The overall function of this system is represented in block diagram form in Fig. 5. A clock program establishes the sequence in which the movable disks are positioned relative to the suspended disks and the sequence in which

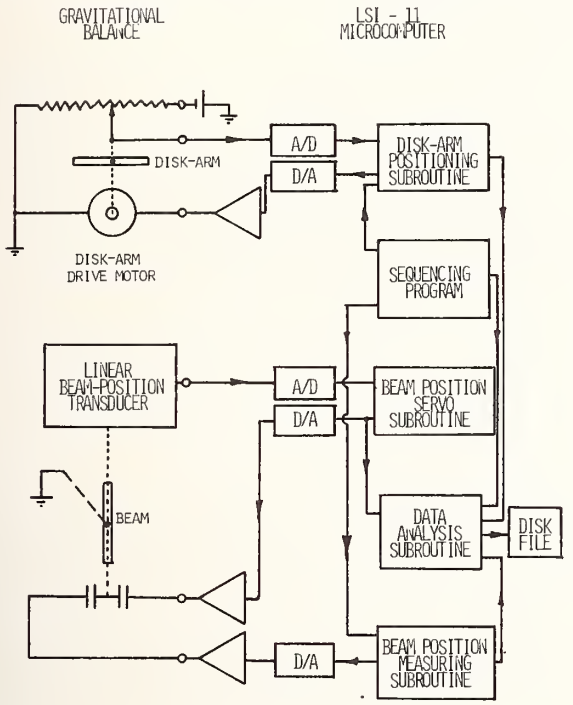


FIGURE 5. *A block diagram of the gravitational balance system and microcomputer control.*

the various measurements are taken. The movable disks are first positioned in the centered, symmetrical position and a series of measurements are made by the beam position measuring subroutine determining gap distances of the two guard-ring capacitors, their contact potentials, and force necessary to hold the beam at its electrical null position. This measurement sequence is repeated twice during each measurement cycle as a baseline check and to provide calibration of the gravitational force $F(z)$. The sequencing program next positions the movable disks in a clockwise position a predetermined distance z away from the suspended disks. The position z is read from a precision potentiometer and the gravitational force is read as a balancing voltage generated by the beam position servo subroutine. The disk arm is again centered for a baseline check and finally positioned a distance z away from the suspended disks and the force measurement is repeated. The measurement cycle is repeated a number of times for the same z and then a new series is initiated for a new z value. In this way a data set is generated from which $F(z)$ can be determined and $G(r)$ can be deduced.

The final assembly of this balance system was completed just before this conference. Work is now under way to begin the program of software testing and calibration and by the end of 1981 we will have begun making measurements of G.

References

[1] D. R. Long, Phys. Rev. D, 9, 850 (1974).
[2] F. D. Stacey and G. J. Tuck, these proceedings.
[3] Y. Fujii, Nature Phys. Sci. 234, 5 (1971).
[4] J. O'Hanlon, Phys. Rev. Lett. 43, 103 (1979).
[5] J. Scherk, Phys. Lett. 88B, 265 (1979).
[6] J. K. Hoskins, R. Newman, J. Schultz, and R. Spero, these proceedings.
[7] H. A. Chan and H. J. Paik, these proceedings.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Tests of Gravitation and Relativity*

Robert F. C. Vessot

Smithsonian Astrophysical Observatory, Cambridge, MA 02138

The weak but all pervasive force of gravity governs the overall behavior of the universe. The dramatic discoveries in astrophysics place gravitation in the forefront of interest and raise questions about the range of validity of Einstein's General Theory of Relativity and at what level it may prove to be a classical limit to a quantum theory. At present, after more than seventy years, the General Theory continues to be the most acceptable description of space-time despite continuing tests to challenge it by searching for contradictions to predictions based on this theory.

The kinds of tests being made can be roughly classified in terms of the contradictions to theory they seek to expose.

1. Tests to look for variations in the conventional "constants," for example, the constancy and isotropy of the velocity of light and the time dependence of the gravitational constant.

2. Tests to look for variations in the physics of the flat space-time of Lorentz inertial frames, which include studies of the second-order Doppler effect, length contraction, and mass variation.

3. Tests to confirm that gravitation is described by a metric theory in which a Riemannian geometry describes space-time such that everywhere, locally, the laws of special relativity apply.

4. Tests to evaluate the validity of theories proposed as alternatives to the General Theory. To assess these alternatives, theorists have devised a score-keeping system, which, in essence, is a theory of theories called the Parameterized-Post-Newtonian (PPN) metric theory. This provides a valuable means to test theories by clearly establishing their observable experimental consequences.

5. Tests from the anticipated detection of gravitational waves. The recent discoveries in astrophysics of superdense matter from gravitational collapse and the strong probability of the existence of black holes have kindled a high level of interest in the search for gravitational radiation which, when discovered, will open a totally new method of astronomical observation. Determining the characteristics of this radiation, i.e., its velocity of propagation and polarization qualities, will provide further critical tests of general relativity.

The chief purpose of this paper is to review the results of recent tests and assess the status of experimental gravitation, concentrating on the last three categories outlined above. The status of applicable present technology and developments for future measurements will be discussed.

Key words: experimental tests of gravitation; General Theory of Relativity tests; relativity.

1. Introduction

The physics of gravitation deals with the weakest of the four known particle interactions. Though by far the weakest force, the effects of gravity grow relentlessly with mass and pervade all space-time, and they literally encompass the whole universe. The dramatic recent discoveries in astronomy and astrophysics have heightened our interest in gravitation and raised questions as to the validity of the Einstein General Theory of Relativity when it is expected to apply to extraordinarily dense matter and extremely high gravitational fields. Questions have been raised about the range of validity of this theory and at what level we will find it to be a classical limit to a quantum theory.

After well over half a century, the General Theory continues to be the most acceptable description of space-time and to many questioning individuals appears to have almost become dogmatically accepted as a basis for theoretical research. At present, theoretical work has far outdistanced experimentation which, owing to the extremely small manifestations of the General Theory, makes testing relativistic gravity very difficult.

Historically, the early tests of light bending from eclipses of the sun, measurements of the perihelion of mercury, and the redshifting of spectral lines at the sun's surface gave little more than qualitative confirmation to the new theory. These tests, however, were crucial to its early acceptance and established Einstein's General Theory as one of the true masterpieces of human achievement. However, gravitational theories must, in the long run, be guided by experimental facts and it is only through the experimental approach that ultimate acceptance can be gained.

In other branches of physics, experiment and theory generally develop concurrently, each guiding the progress of the other. Gravitation is an exception because of the lack of instrumentation sensitive enough to provide clean, accurate, hard-core facts.

We have been limited by the levels of mass and velocity available to do truly definitive experiments in the laboratory, as distinguished from astrophysical observations. In the case of such observations, we clearly cannot use gravity theory as a tool for unraveling physical processes and, at the same time, expect the observation to provide a clear cut test of the gravity theory. This ambiguous position, at best, can lead us to a confirmation of the self-consistency between the physics of the observed effects and the theory of gravity we have employed. The theoretical development of gravitational physics contin-

*This work was supported by the National Aeronautics and Space Administration and the Smithsonian Institution.

ues despite the fact that no solid evidence has yet been found to invalidate the General Theory. The need for such evidence becomes more and more essential as startling astrophysical data accumulate, and calculation and speculation continue in our efforts to understand the new wonders that we observe.

The recent advances in technology give us considerable optimism in performing more crucial tests of gravitation. Atomic clocks, very long-baseline interferometry, interplanetary radar, laser and transponder ranging to space probes and planets, superconducting gyroscopes and other low-temperature techniques, provide means for making new instruments of previously unimaginable sensitivity. The spectacular advances in space technology have extended the dimensions of our laboratory to encompass the entire solar system—and beyond, and allow us to use massive bodies like the sun and planets as test objects for laboratory experimentation. As Misener, Thorne, and Wheeler [2] have said, "General Relativity is no longer a theorist's Paradise and an experimentalist's Hell."

It is clear that a rational testing program should, to a large extent, be guided by theory. However, in the case of gravitation it is more often governed by what is possible technologically and feasible economically. Also, we should not be blind to the possibilities that nature may hold some surprises, and opportunities to perform seemingly redundant tests by using disparate techniques should not be neglected.

There is little question that in the near future the best, clean, tests of gravitation will involve space-borne instrumentation, both to make use of the available massive bodies in our solar system and to escape the limitations placed on terrestrial instrumentation by vibration and earth tidal effects and the limits of signal propagation through the earth's troposphere and ionosphere. We can also expect that the use of space techniques in astronomy, over the whole electromagnetic spectrum from x rays to very low frequencies, will lead to discoveries posing new questions that will make even further demands on gravity theory.

At present, the most exciting prospect for the future is the anticipated detection of gravitational radiation. This would open up a totally new era of astronomy by providing data from processes completely different from the electromagnetic generators we have always observed.

1.1 Criteria for a Valid Theory of Gravitation

In seeking some orderly way to summarize gravity experiments, we should first look at the underlying criteria for a viable theory. Thorne and Will [3] give criteria for the viability of a gravitational theory as follows:

1. It must be complete—it must be capable of analyzing from first principles the outcome of every experiment of interest. The theory must incorporate and mesh with a complete set of electromagnetic and quantum mechanical laws which can be used to calculate the detailed behavior of bodies in a gravitational field. (Currently this does not extend to the extremes of quantum gravity.)
2. It must be self consistent. A theory must predict uniquely the outcome of every experiment.
3. It must be relativistic. If gravity is "turned off" the nongravitational laws of physics become those of special relativity.
4. It must have the correct Newtonian limit. In the limit of weak gravity fields and slow motion, we should see convergence to Newtonian physics.
5. It must embody the weak equivalence principle—otherwise known as the Universality of Free Fall, the experimental fact that all bodies fall in the same way, regardless of their composition.
6. It must embody the Universality of Gravitational Redshift.

These criteria provide guidance in classifying the types of tests we can perform. Criteria 1 and 2 will ultimately be met only by accumulation of experience of every possible kind. Criterion 3 requires that the nongravitational aspects of physics be governed by the Lorentz group as, for example, manifested by the everyday high-precision verification of four momentum conservation and other relativistic laws of kinematics in high-energy accelerators. These particle physics tests, however, are *very strictly* local and not global in their extent. Criterion 4 expresses the facts accumulated by centuries of solar system observation and the works of early astronomers and mathematicians culminating in the monumental *Principia* of Newton (1687) [4]. This system of physics is fundamental and gives the limiting description under conditions of weak gravitational fields and slow motion. Criteria 5 and 6 are statements that can be summarized as the Einstein Equivalence Principle, which he first enunciated in 1907 as a generalization of the observed proportionality between gravitational and inertial mass. This principle asserts that there is no way of distinguishing locally between a gravitational field and an oppositely directed acceleration. One of the consequences of the Einstein Equivalence Principle is the so-called gravitational redshift.

The Einstein Equivalence Principle (or strong equivalence principle as it is sometimes called) leads directly to the concept of a space-time that can be thought of as consisting of a manifold of freely falling Lorentz frames in each of which apply the laws of special relativity. This postulates that space-time is endowed with a metric, that the trajectories or world-lines of falling test bodies are geodesics of that metric, and that nongravitational laws of physics in any free-falling frame reduce to those of the Special Theory of Relativity.

We are thus led to a four-dimensional picture of space-time where, locally, the coordinates in four-space are rectilineal and orthogonal. Over larger excursions in space-time there occurs a bending of the coordinates that depends entirely on the distribution of matter and energy. This concept of how a spacetime continuum is shaped by the presence of mass and energy is Einstein's entirely original view of the structure of our universe.

Testing this concept of space-time presents a formidable challenge to metrology since it affects all aspects of space and time. In it we identify two "constants," the velocity of light, c, and the universal constant of gravitation, G. The nature of these two quantities throughout the space and time of our universe is fundamental to the study of cosmology and to the development of gravitation and relativity theory.

The following review of tests of gravitation and relativistic theories is intended to sketch an overview of the theory and describe the trends of the testing activity made possible by present technology. To those who wish to read further about the status of theory and experiment I recommend the recent (1979) book celebrating the centenary of Einstein's birth published by the Cambridge

University Press, edited by S.W. Hawking and W. Israel [5, 6], entitled *General Relativity. An Einstein Centenary Survey*; and C. M. Will's new book, *Theory and Experiment in Gravitational Physics*, by the same publisher.

1.2 Tests of Gravitation

The tests that have thus far been done can be classified in terms of the contradictions and inconsistencies they seek to expose.

1. Tests to look for variations in the conventional "constants," for example, the constancy and isotropy of the velocity of light and the time dependence of the gravitational constant.

2. Tests to look for variations in the physics of the flat space-time of Lorentz inertial frames, which include studies of the second-order Doppler effect, length contraction, and mass variation.

3. Tests to confirm that gravitation is described by a metric theory in which a four-dimensional Riemannian geometry describes space-time such that everywhere, locally, the laws of special relativity apply.

4. Tests to evaluate the validity of metric theories proposed as alternatives to the General Theory. To assess these alternatives, theorists have devised a score-keeping system, which, in essence, is a theory of theories called the Parameterized-Post-Newtonian (PPN) metric theory [1, 7]. This provides a valuable means to test theories by clearly establishing their observable experimental consequences.

5. Tests from the anticipated detection of gravitational waves. The recent discoveries of superdense matter from gravitational collapse and the strong probability of the existence of black holes have kindled a high level of interest in the search for gravitational radiation which, when discovered, will open a totally new method of astronomical observation. Determining the characteristics of this radiation, i.e., its velocity of propagation and polarization, will provide further critical tests of general relativity.

2. The Constancy of Constants

The question of the behavior of the gravitational constant, G, and the velocity of light is central to any theory of gravity and relativity. Since the focal point of this conference is on this type of question and because of the many papers given on these topics including the excellent review by H. de Boer that precedes this paper, I feel that I can justifiably give only a very rough overview of the behavior of the velocity of light and the universal gravitational constant.

Historically, behavior of the velocity of light has been the origin of our modern concepts of relativity. The Michaelson–Morely ether drift experiments and its successors provide the early experimental evidence for the Special Theory of Relativity. Central to this evidence is the question of the independence of c on source velocity, its constancy, and isotropy regardless of wavelength. Recent experimental evidence is strong that the velocity of light is independent of the relative velocity of the source and observer. If we take as a hypothesis that

$$c' = c + kv,$$

the criterion for independence is given by the limits that can be set on k.

Measurements of γ rays from π^0 mesons at energies greater than 6 GeV are reported by Alvager *et al.* [8] who find

$$k = (-3 \pm 13) \times 10^{-5}.$$

From astronomical observational data taken from a binary x-ray pulsar, Brecher [9] concludes that $k < 2 \times 10^{-9}$. Another postulate related to the velocity of light is that it is a limit to the velocity of massive particles. Tests have been made of the relative velocity of γ rays and high-energy electrons in the 10–20 GeV range. Here the prediction is that

$$\frac{c_\gamma - v_e}{c_\gamma} = \left[1 - \frac{v_e}{c_\gamma}\right] \sim 5 \times 10^{-10},$$

where c_γ is the velocity of γ rays and v_e is the velocity of the electrons.

Guiragossian *et al.* [10] confirm that $(1 - v_e/c) \sim \pm 2 \times 10^{-7}$. Brown *et al.* [11] report $(1 - v_e/c) \sim (1.2 \pm 2.7) \times 10^{-6}$ for 11-GeV electrons and visible light and also measure the relative velocity of light at visible wavelengths and of 7 GeV γ rays giving a result

$$\frac{c_{\text{vis}} - c_{7\,\text{GeV}}}{c} \sim (1.8 \pm 6) \times 10^{-6}.$$

A recent and very elegant measurement of the isotropy of the velocity of light has been reported by Brillet and Hall [12] who have used a sophisticated laser technique to search for an ether drift. Their work and its various interpretations will be discussed at this conference. Their conclusions are that the measurement of the length of a cavity resonator using electromagnetic signals is constant to within a few parts in 10^{15}. By making various hypotheses about the motion of this laboratory experiment in the universe we can arrive at very impressive limits on the validity of the Lorentz transform, which is fundamental to special relativity.

The question of the behavior of the gravitational constant has arisen from speculations relating to the expansion of the universe and models have been suggested that would necessitate a variation of G with time and with position in the cosmos and that there are possible violations of the inverse-square law of gravitational attraction.

Again, since there are many excellent papers addressing this question, I shall only report the most recent results from lunar occultation data and from planetary radar measurements.

From Lunar Occultation

$\left|\frac{\dot{G}}{G}\right| < 4 \times 10^{-11}/\text{yr}$ Morrison [13]

$\frac{\dot{G}}{G} = (-3.6 \pm 1.8) \times 10^{-11}/\text{yr}$ Van Flandern [14, 15]

From Planetary Radar Measurements

$\left|\frac{\dot{G}}{G}\right| < 4 \times 10^{-10}/\text{yr}$ Shapiro *et al.* [16]

$< 1.5 \times 10^{-10}/\text{yr}$ Reasenberg and Shapiro [17]

3. Tests of Special Relativity

The second category of tests is related to searches for violations of special relativity and comprises a number of subgroups that investigate the parameter $\gamma = (1 - v^2/c^2)^{-1/2}$.

3.1 $g-2$ Tests

These tests are based on the concept [18] that the factor $\gamma = (P/m_0)\, dp/dE$ applies differently from the cyclotron frequency ω_c, of an electron (lepton) in magnetic field B (where $\omega_c = eB/\gamma m_0 c$), and the $g-2$ precession frequency $\omega_s = (geB/2m_0c) + (1 - \gamma)\omega_c$.

The $g-2$ experiments determine the quantity $(\omega_s - \omega_c)(eB/m_0c)^{-1}$, which, in the conventional theory, equals $(g-2)/2 = a$. With the assumption that there is a different γ for electron dynamics and electron kinematics, the quantity a becomes $g/2 - \gamma/\bar{\gamma}$. The ratio $\gamma/\bar{\gamma}$ is written as an expansion $\gamma/\bar{\gamma} = 1 + C_1(\gamma - 1) + \ldots$, which makes $\gamma/\bar{\gamma} = 1$ in the limit $\gamma \to 1$ (low-velocity limit). The term C_1 is used to define a figure of merit with one or more values of γ, made at differing electron energies, to provide combinations like $C_1 = (a^{(2)} - a^{(1)})/(\gamma^{(1)} - \gamma^{(2)})$. For a given resonance linewidth, the larger the range of γ, the more sensitive the test.

A summary of the figure of merit parameter, C_1, made from various tests is shown in Table 1 from Cooper *et al.* [19].

The pioneering experiments of Ives and Stilwell [25] using fast-moving hydrogen atoms have been repeated by Mandelberg and Witten [26] and confirm the predictions to about 5%.

Experiments that observe the shift in a resonance owing to thermal motion of the source have been made by Pound and Rebka [27] as part of the necessary corrections of the resonance frequency of the ^{57}Fe Mossbauer radiation used to measure the gravitational redshift. Similar thermal motion effects occur in hydrogen masers [28]; however, these results assumed the validity of special relativity and were used as corrections to the frequency data and are not a convincing test of special relativity.

Far more sensitive tests have been reported in the last few years from muon decay measurements made with the CERN storage ring and reported by Bailey *et al.* [29]. This is the most recent of a long series of high-energy particle experiments to test the extension of the decay time involving many different particles.

TABLE 1. *Tests of g−2.*

Particle	$\gamma^{(1)}$	$\gamma^{(2)}$	C_1	References
μ^-, μ^+	12	29.2	$1.4 \pm 1.8 \times 10^{-8}$	Bailey *et al.* [20,21]
$\bar{e}$	1	1.2	$(-2.6 \pm 1.8) \times 10^{-8}$	Van Dyck *et al.* [22] Wesley and Rich [23]
$\bar{e}$	1	2.5×10^4	$(-1.0 \pm .8) \times 10^{-10}$	Cooper *et al.* [19]

3.2 Mass Increase Tests, $m = m_0\gamma$

Measurements of the mass increase for 385 MeV ($v/c \sim 0.7$) protons have been reported by Grove and Fox [24]. They measured the e/m ratio of 385-MeV protons in a synchrocyclotron and confirm the relation

$$\frac{e}{m} = \frac{e}{m_0}\left[1 - \frac{\omega_c^2 r^2}{c^2}\right]^{1/2}.$$

By putting the cyclotron frequency ω_c and the particle's radius of movement, r, in terms of velocity, v, we have

$$\frac{e}{m} = \frac{e}{m_0}\left[1 - \frac{v^2}{c^2}\right]^{1/2} = \frac{e}{m_0}\gamma^{-1}.$$

Their results are -0.0006 ± 0.001 for the quantity $(e/m_1 - e/m_2)(e/m_2)^{-1}$ for 385-MeV protons.

3.3 Tests of Second-Order Doppler Effects or Time Dilation, $\tau = \tau_0\gamma$

These tests generally involve the measurement of an electromagnetic signal, either in emission or absorption, of a particle that is in rapid motion with respect to the observer. The relation between the shift in the resonance frequency, or the extension of the decay time, and the relative velocity is experimentally determined and compared with predictions from special relativity.

Table 2 from J. Bailey *et al.* [29] summarizes these results using the parameter $(\tau_1/\gamma_1 - \tau_2/\gamma_2)(\tau_2/\gamma_2)^{-1} = F$ as a figure of merit. From this recent work we see that time dilation is confirmed at the 10^{-3} level.

A proposed high precision measurement of time dilation is reported in the work of Nachman, Rayman, and Hall [39] who will make time-dilation measurements using two-photon absorption resonance in a fast beam of metastable helium. They expect to obtain a 500-fold improvement over the CERN storage ring results discussed above. Their work will be reported elsewhere in these proceedings.

These tests clearly establish that the Special Theory of Relativity is on solid ground and it appears that this foundation of gravity theory is becoming more secure as further testing is done.

An interesting observation of the meshing of the Special Theory of Relativity with the Einstein Principle of Equivalence is discussed in the following section where the hydrogen maser space-probe relativity experiment is described. Owing to the 10,000-km altitude and very nearly free fall to the earth's surface, values of v/c of approximately 10^{-5} were encountered. In accounting for the total frequency variation during the flight, second-order Doppler effects at the 10^{-10} level are included in the relativistic prediction. As will be seen, the resulting overall confirmation of the predictions made from general relativity was at the 70×10^{-6} level, and can be interpreted as a test of time dilation over very nonlocal conditions.

TABLE 2. *Second-order Doppler effects or time dilation* $\tau = \tau_0\gamma$. *Particle-decay experiments, criterion* $F = (\tau_1/\gamma_1 - \tau_2/\gamma_2)/(\tau_2/\gamma_2)$

Particle	γ_1	γ_2	$F \times 10^2$	Reference
$\pi^{\pm}$	1.0	2.44	-2.5 ± 0.9	Nordberg *et al.* [30] Ayres *et al.* [31,32] Dunaitsev *et al.* [33]
$K^{\pm}$	1.0	3.38 4.17	0.9 ± 0.3	Ott and Pritchard [34] Lobkowicz [35]
K_s^0	1.63	15.2 20.2	0.5 ± 0.6 0.2 ± 0.7	Carithers *et al.* [36] Skjeggstad *et al.* [37] Geweniger *et al.* [38]
μ^+	1.0	29.3	0.02 ± 0.09	Bailey *et al.* [29]

4. Tests of the Metric Hypothesis

The Einstein Principle of Equivalence asserts that there is no way, locally, of distinguishing between a gravitational field and an oppositely directed acceleration. The traditional argument used to describe this principle for the behavior of light rays is shown in Fig. 1, where we consider the behavior of the frequency of a signal from a transmitter on the left-hand wall of the lab received at the right-hand wall a distance ℓ away. If we accelerate the lab to the right with a bulldozer, the signal received will be at a lower frequency because the right-hand wall received the signal at a time $\tau = \ell/c$ after it was transmitted during which time the receiver had a velocity $\Delta v = a\tau$ away from the transmitter. The Doppler shift, Δf, from this velocity is $\Delta f/f = \Delta v/c = a\ell/c$, where a is the acceleration of the lab. According to the Einstein Principle of Equivalence, if we tip the lab on end we should see a shift $\Delta f/f = g\ell/c$ where g is the local acceleration due to gravity. This shift is often referred to as the gravitational redshift. The question is, does a act in the same way as g?

A very elegant test of the equivalence principle in a lab (minus bulldozer) was done using the extremely narrow

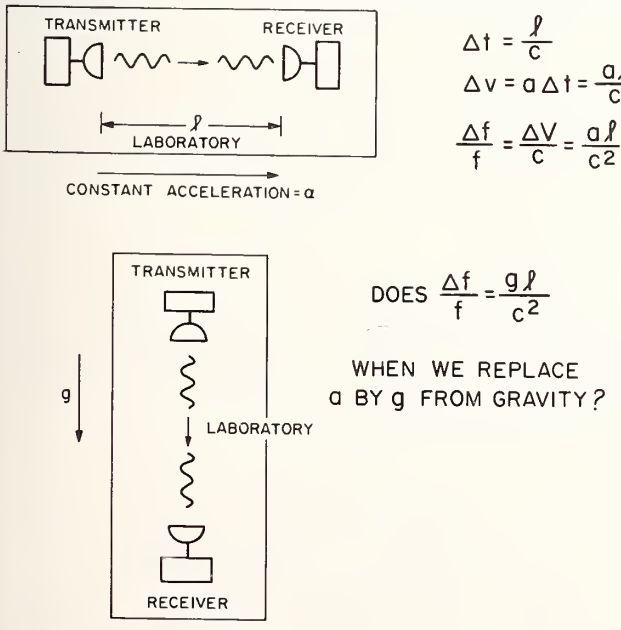


FIGURE 1. *Illustration of gravitational redshift.*

Mossbauer effect emission and absorption characteristics of ^{57}Fe gamma rays at a tower within Harvard's Jefferson Physics Laboratory by Pound, Rebka, and Snider [27, 40]. Their results have confirmed the hypothesis at the 1% level.

Another, but decidedly nonlocal, test of the gravitational redshift was performed [41] by launching a hydrogen maser oscillator (or clock) aloft in a rocket to an altitude of 10,000 km and comparing, with microwave signals, the frequency of the maser with other masers on earth. The first-order Doppler effects from the rapid changes in the propagation paths were canceled by a transponder system that measured the go-return (two-way) Doppler cycles, dividing the number of these cycles by two and subtracting them from the cycles in the one-way microwave link that connected the probe clock to earth clocks as shown in Fig. 2. Since the probe had a high relative velocity with respect to the earth station and since the earth station (located in Florida) was in motion owing to the earth's rotation, other effects of relativity must be considered. The predicted relationship is given in geocentric inertial frame as follows:

$$\frac{\Delta f}{f} = \frac{\phi_s - \phi_e}{c^2} - \frac{|\,v_e - v_s\,|^2}{2c^2} - \frac{r_{se}\cdot a_e}{c^2}.$$

Here $\phi_s - \phi_e$ is the Newtonian potential between the spacecraft and earth station, $\mathbf{r}_{se}$ is the vector joining the

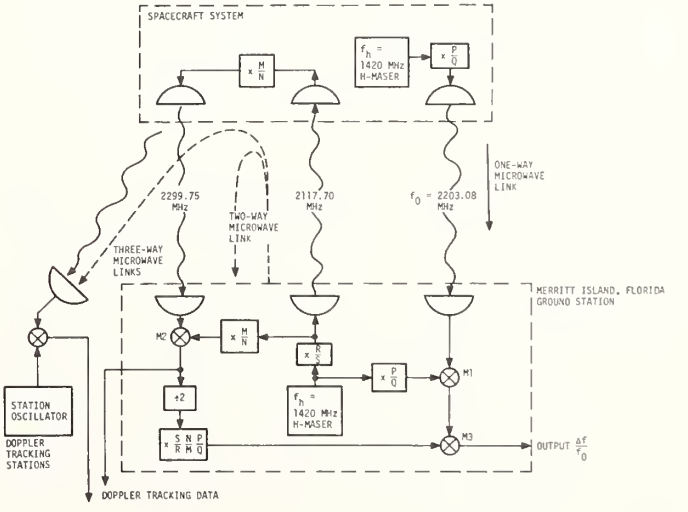


FIGURE 2. *Microwave system used in the 1976 NASA-SAO rocket-borne redshift experiment.*

spacecraft and earth station, $\mathbf{v}_s$ and $\mathbf{v}_e$ are the velocities of the spacecraft and earth stations, respectively, and $\mathbf{a}_e$ is the acceleration of the earth station owing to the earth's rotation.

The agreement of the measured data was within 70×10^{-6} of the total effect predicted in the above equation. There are various ways that this result can be interpreted. For example, if we affirm the correctness of the second term giving the second-order Doppler effect and assume that the third term correctly accounts for a Doppler shift resulting from the change of velocity of the earth station during the flight time, we can ascribe all the error to the equivalence principle and confirm it to approximately 140×10^{-6}. (Since the redshift and second-order Doppler terms are very nearly the same in magnitude, the fractional error in the total effect, now half as large, is doubled.)

Alternatively, if we place all the error in the second-order Doppler effect and assume the equivalence principle and the earth station term are correct, we have confirmation that the second-order frequency shift is correct within 140×10^{-6}. This conclusion is valid over space-time intervals that have considerable curvature, which is a different situation than obtained for the purely local tests described earlier.

Yet another interpretation of this data is possible if we ascribe all the error to a possible anisotropy of the velocity of light during the experiment. Since this experiment involved a substantial change of velocity and depended on the cancellation of the Doppler effect in the one-way path, by taking an average of the two-way Doppler and subtracting it from the one-way frequency, a residual Doppler signature would be present in the data if the velocity of light in the downlink and uplink signals was different.

Under the above assumptions, a limit of $\Delta c/c$ of $\pm 3 \times 10^{-9}$ can be set for the isotropy of the velocity of light.

The success of the Doppler-canceling techniques used in this experiment leads us naturally to think of other experiments in space with clocks, such as measuring the redshift to second order in a close-approach solar mission, and the use of a clock as an accurate distance-measuring tool. With the current stability of 6×10^{-16} for 1-h measuring intervals available from existing hydrogen masers [42], we can think of this in terms of a change of Doppler velocity $\Delta v/c = \Delta f/f$ at a level where Δv is about 2×10^{-5} cm s^{-1} when compared over hourly intervals. This sensitivity offers the possibility of detecting gravitational waves of very low frequency by looking for small anomalies in the motion of spacecraft at distances of many astronomical units from earth.

4.1 Universality of Free Fall as Historically Recorded

Tests leading to information that confirms the metric hypothesis begin in the early 15th century with the apocryphal experiment of Galileo who is said to have dropped a cannon ball and bocce ball from the Leaning Tower of Pisa. However, there are records that Galileo did perform dropping tests and also tests on an inclined plane to dilute gravity. The results are stated as follows by Galileo: "The variation in speed in air between balls of lead, gold, copper, porphyry, and other heavy materials is so slight that in a fall of 100 cubits (say 46 meters) a ball of gold would surely not outstrip one of copper by as much as four fingers. Having observed this, I came to the conclusion that in a medium totally void of resistance, all bodies would fall at the same speed" (see Ref. [2], p. 16). Later, Isaac Newton repeated these tests with pendulums made of different materials and, in *Principia*, states in Definition I, "The quantity of matter is the measure of the same arising from its density and bulk conjointly" ... and later ... "It is this quantity that I mean hereafter everywhere under the name of body or mass, and the same is known by the weight of each body, for it is proportional to the weight, as I have found by experiments on pendulums, very accurately made, which shall be shown hereafter" [4]. Here began a long series of experiments looking for a discrepancy of the proportionality of inertial mass to gravitational mass. Table 3 shows the trend of some of the tests of the Weak Equivalence Principle [43] given in terms of $\eta' \equiv 2(a_A - a_B)/(a_A + a_B)$, where a_A is the acceleration of substance A and a_B is the acceleration of substance B.

These tests of the universality of free fall, or Weak Equivalence Principle, are very important and a substantial effort has been made to connect the weak principle to the Einstein Equivalence Principle. This was begun by the late Leonard Schiff [50] and was continued by Lightman and Lee [51] in 1973 who derive proofs under specific mathematical frameworks for certain restricted conditions. This research has led to an understanding of the electrostatic and magnetostatic self-energies and their contributions to the gravitational and inertial mass of bodies. The resulting formalism points out possible deviations in the manner in which the electromagnetic laws and equations of motion mesh with gravitational fields if these laws do not follow the metric hypothesis. These deviations are described in terms of parameters Γ_1 and Λ_1, etc., for each order of the gravitational potential when expanded as a power series. These parameters describe the electrostatic and magnetostatic self-energy variation with potential, and are zero for metric theories. By making high precision resonators or oscillators depending on different physical processes, we can see how their frequencies are affected by gravitational potential by expressing their frequency dependence on gravity potential as prescribed by the nonmetric hypothesis, $\Delta f/f = (1 + \alpha)\Delta\phi/c^2$. Here, α is the measure of the nonmetric behavior or lack of proportionality in the redshift vs. gravitational potential relation for a particular clock:

$\alpha = -3\Gamma_0 + \Lambda_0$	for hydrogen hyperfine structure as in the hydrogen maser
$\alpha = -\frac{1}{2}(3\Gamma_0 + \Lambda_0)$	for an electromagnetic cavity resonator (superconducting cavity oscillator)
$\alpha = -2\Gamma_0$	for a high Q, dielectric crystal.

According to this hypothesis we can interpret the data from the various tests of the Weak Equivalence Principle shown in Table 3 and relate the data to tests that measure other parameters. Table 4 from Will [43] shows the limits of validity of the parameters.

Recent work by Will [43] has shown that the measurement of gravitational redshift is a valuable resource for gravity theory. The basis for this is the following conjecture: "Every nonmetric theory of gravity predicts a gravitational redshift which depends on the nature of the clock whose redshift is being measured." This can also be stated as, "Any complete, self-consistent and relativistic theory of gravity that embodies the universality of gravitational redshift is necessarily a metric theory"

TABLE 3. *Universality of free fall—Weak Equivalence Principle.*

Method and substances	Limit on η'	Reference
Pendulums of various substances	10^{-3}	Newton [4]
Pendulums of various substances	2×10^{-5}	Bessel [44]
Torsion balance using various substances	5×10^{-9}	Eötvös *et al.* [45]
Torsion balance using aluminum and gold	10^{-11}	Dicke [46]
Torsion balance using aluminum and platinum	10^{-12}	Braginsky and Panov [47]
Projected experiments in free fall in an earth-orbiting satellite are expected to reach limits of 10^{-18}		Everitt [48] Worden [49]

TABLE 4. *Lightman and Lee [51] connection between the weak and strong Principles of Equivalence [43].*

Parameter	Limit	Experiments
Γ_0	4×10^{-10}	Eötvös-Dicke-Braginsky (EDB) Electrostatic energy
Λ_0	6×10^{-6}	EDB magnetostatic energy
Γ_1	4×10^{-2}	EDB electrostatic energy
Λ_1	600	EDB magnetostatic energy
$3\Gamma_0 - \Lambda_0$	2×10^{-4}	Vessot *et al.* [52]
$\Gamma_0 + \Lambda_0$	10^{-5}	Hughes [53] Drever [54]
$\frac{3}{2}(\Gamma_0 - \Lambda_0)$	$< 2 \times 10^{-2}$	From H vs. SCSO test results, Turneaure *et al.* [55]

(from Will [56,57]). An alternative statement, with redshift replacing universality of free fall, is the Schiff Conjecture [58] (p. 343), "Any complete and self-consistent theory that obeys the weak equivalence principle must obey the Einstein equivalence principle." The conjecture in quotes is the classical analog of Schiff's original quantum mechanical conjecture, reformulated by Thorne, Lee, and Lightman [59].

This conjecture has led to experiments with clocks depending on various physical principles. In 1978 a series of experiments were conducted at Stanford University by Turneaure, Mattison, and Vessot, to compare the frequency of a superconducting cavity stabilized oscillator (SCSO) and a hydrogen maser (H). The object was to see if the variation in the gravitational potential of the sun as the earth rotated and moved in its eccentric orbit would manifest any change in the relative frequency of these oscillators as shown in Fig. 3. The gravitational potential variation is

$$u = \frac{\Delta\phi}{c^2}\Big|_{\text{sun}} = -3.2 \times 10^{-13} \cos 2\pi(t - t_0) + 2.8 \times 10^{-12}(t - t_0),$$

(where t is measured in solar days and t_0 is noon April 4). The Lightman and Lee parameters [51] for such a comparison give the ratio

$$f_H/f_{SCSO} = A\left[1 + u\,\frac{(\Gamma_0 - \Lambda_0)}{2}\right].$$

While the results of these tests are still to be published [55] it is now possible to state that the departure $(\alpha_H - \alpha_{SCSO}) < 2 \times 10^{-2}$.

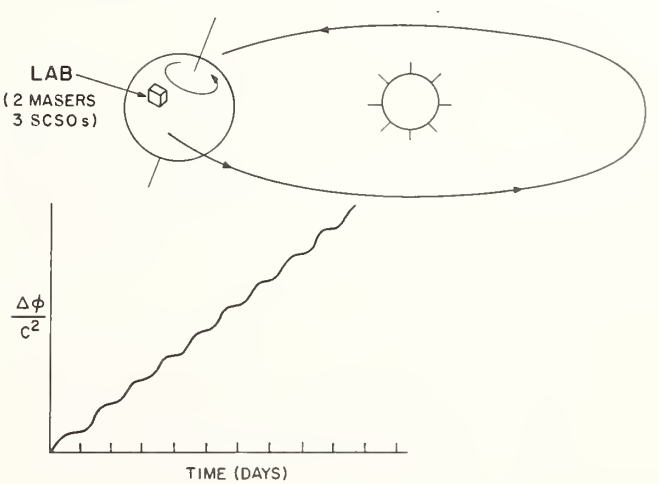


FIGURE 3. *Illustrating the variation of the sun's gravitational potential at Stanford, California, during the month of April.*

This question of how electrostatic and magnetostatic self-energy in particles behave as contributors to the "mass" has its gravitational analog called the "gravitational weak equivalence principle" and is also called the "super strong equivalence principle." The manner in which gravitational self-energy adds to the mass of bodies was studied by Nordtvedt [60, 61].

The gravitational self-energy for a body of mass m and characteristic size, a, is $Gm^2/a = E_g$ and the total energy is mc^2. The ratio of these, Gm/c^2a, is about 10^{-25} for laboratory-sized bodies and the Eötvös-Dicke-Braginsky tests cannot help distinguish the gravitational energy contributions to mass. However, for more massive bodies such as the earth, moon, and other solar system bodies, there could be a more measurable effect.

Nordtvedt defines a parameter, η, which is a measure of the possible departure from the equivalence of gravitational and inertial mass in the following expression

$$\frac{m_g}{m_i} = 1 + \eta \frac{G}{c^2} \frac{\int \frac{\rho(x)\rho(x')}{(x - x')} d^3x d^3x'}{\rho(x) d^3x} = 1 + \eta \frac{E_g}{mc^2},$$

where $\rho(x)$ is the mass density at position x, G is the gravitational constant, and c is the velocity of light.

He proposed tests of this effect using the earth-moon combination in the sun's gravity field. Under the conditions where the equivalence principle is violated owing to the possibility that the earth-moon gravitation self-energy divided by c^2 is not behaving like ordinary mass, the earth-moon separation would show anomalous motion.

This test has been performed using lunar laser-ranging techniques and the results are reported by Williams *et al.* [62], who found $\eta = 0.00 \pm 0.03$, and Shapiro *et al.* [63], who determine that $\eta = -0.001 \pm 0.015$. Since this test also involves other parameters that come from the assumption of a particular metric of space-time, more will be said in the following section.

Theoretical work continues in efforts to understand the relation between the Weak Equivalence Principle and the Einstein Equivalence Principle. Recent work by Ni [64] suggests Eötvös-type experiments testing the universality of free fall be done on polarized test bodies.

While the convincing experimental work and theoretical arguments have done much to bolster our confidence that whatever theory is found to be correct must be a metric theory, there is still room for error at some level. Theoretical study and experiments continue actively to pursue this question further.

5. Tests of Metric Theories

The most recent and complete classification of metric theories proposed as alternatives to the general theory of relativity is given by Will in *General Relativity, An Einstein Centenary Survey* [43] and in his forthcoming book *Theory and Experiment in Gravitational Physics* [6].

The role of the metric in such theories is quoted here from Will [43].

> When the postulates of metric theories of gravity are examined, one notices a crucial feature: no matter how complex the theory, no matter what additional gravitational or cosmological fields it deals with, matter and nongravitational fields respond only to the metric $g_{\mu\nu}$. The role of other fields which a given theory may contain can only be that of helping to generate the space time curvature associated with the metric. Matter may create these fields, and the fields plus the matter may generate the metric, but they cannot directly act back on the matter. *The matter responds only to the metric* (author's italics). From this point of view, the metric $g_{\mu\nu}$ becomes the primary theoretical entity, and all that distinguishes one metric thing from another is the particular way in which matter generates the metric.

A theory of theories has been developed by Thorne, Will, and Ni [65] that now comprises ten parameters that can be included in the metric to adapt it to the various proposed theories. This is known as the Parameterized Post-Newtonian Metric or PPN metric.

Table 5 (from Will, see Refs. [43, 6]) lists these parameters, their function, and their value in general relativity.

A complete discussion of the validity set by experiment on various different metric theories by the limits imposed on these metric parameters is outside the scope of this outline. Instead, I should like to concentrate on the experimental limits that can be assigned to the metric parameters associated with general relativity.

One of the earliest parameterizations of the General Theory was done by Eddington [66]. Later work by Robertson [67] and Schiff [68] described the solar system with a spherical nonrotating sun and treated the planets as idealized test bodies. This metric has the form

$$ds^2 = \left[1 + 2\alpha^* \frac{M}{r} + 2\beta \left(\frac{M}{r}\right)^2\right] dt^2 - \left[1 + 2\gamma \frac{M}{r}\right] (dx^2 + dy^2 + dz^2)$$

with $\alpha^* = 1$, $r = (x^2 + y^2 + z^2)^{1/2}$ is the radial distance, and $M = GM_{sun}/c^2$ and will be used in the following discussions as the "mass" of the sun.

This simple metric shows the behavior of the line element in space-time, ds, and includes terms in its temporal part accounting for gravitational redshift, $2M/r$, and the second-order redshift governed by the term β, which describes the nonlinearity in the superposition law for gravity. The use of the parameter α^* to describe the validity of the redshift at first order in c^{-2} is redundant; if α^* were other than exactly unity there would be a conflict with the idea of the metric itself and we would have a contradiction. In tests of planetary motion or particle motion any such departure from unity would be included in the definition of mass, M. However, in tests where the behavior of electromagnetic signals is observed (such as light bending and time delay from ray paths going near the sun), part of the observed deflection or delay is due to the temporal $2M/r$ term and part is due to the spatial $2\gamma M/r$ term where γ is the amount of space curvature from unit rest mass. Even though we may eventually have to face a logical inconsistency, I believe we can benefit by keeping α^* in the notation to help us see what is happening in the experimental results.

This metric provides a framework to test experimentally the warping of time and space using the sun as a massive test body by observing the behavior of light rays as they graze the sun and by observing the behavior of test bodies falling in the gravity field of the sun. The three early tests prescribed by Einstein, namely the advance of the perihelion of the orbit of the planet mercury, the deflection of starlight by the sun, and the gravitational redshift, are examples of this type of experiment.

TABLE 5. *The parameterized Post-Newtonian metric [43, 6].*

Parameter	What it measures relative to General Relativity	Value in General Relativity
γ	How much space curvature is produced by unit mass?	1
β	How much non-linearity is there in the superposition law of gravity?	1
ξ	Are there preferred location effects?	0
α_1, α_2, α_3	Are there preferred frame effects?	0
ζ_1, ζ_2, ζ_3, ζ_4	Is there violation of conservation of total momentum?	0

5.1 Solar System Tests of Falling Bodies

For the planet mercury in its eccentric orbit, the prediction is that the advance of its perihelion is given by

$$\delta\phi = \frac{6\pi M}{a(1-e^2)}\left\{\frac{2+\gamma-\beta}{3} + J_2\left[\frac{R_{\text{sun}}^2}{2Ma(1-e^2)}\right]\right\}.$$

Here, R_{sun} is the mean solar radius, a is the semimajor axis, e is the eccentricity of the orbit, M is the "mass" of the sun, and J_2 is the quadrupole moment of the solar gravitational potential.

The complication of the sun's angular momentum and consequent centrifugal flattening giving the term J_2 led to a very stimulating and fascinating alternative theory of gravitation where it was postulated that the measured value of the perihelion advance of about 43 arc seconds per century might, at least in part, be due to an appreciable value of J_2. This led to the Brans-Dicke [69] scalar-tensor theory and a resurgence of interest in gravitation that triggered a considerable effort to measure optically the roundness of the sun's disk in an effort to detect some measure of J_2 [70, 71]. The Brans-Dicke Theory introduces a variable parameter, ω, and gives the value $(1+\omega)/(2+\omega)$ for the parameter γ in Table 5. As $\omega \to \infty$ the value of $\gamma \to 1$, as predicted by general relativity. As will be seen, the value of ω is being pressed higher and higher and the validity of this theory is now highly questionable.

Measurements of the quantity $(2+2\gamma-\beta)/3$, the perihelion parameter, have been made by radar observations of mercury and the inner planets by Shapiro and his colleagues [63] who report a value of 1.003 ± 0.005 assuming the J_2 contribution is negligible.

The question of measuring the value of J_2 of the sun is of great interest to solar and stellar physicists. A space experiment to make an unambiguous measurement of J_2 is being studied by the U.S. National Aeronautics and Space Administration. This mission, now labeled "Starprobe," consists of a spacecraft that passes within four solar radii of the sun's center, which will compensate for all nongravitational forces such as solar light and particle flux pressure by using a drag compensating servo system. By accurately tracking the extremely eccentric orbit of this freely falling "test body," the value of J_2 and the other γ and β combinations will be separately identified through their time signatures in the tracking data. In addition, if the spacecraft were to contain a clock and a suitable microwave system, the direct measurement of β as the second-order behavior of the gravitational potential could be measured directly rather than derived from ballistics of falling particles as will be discussed later.

The Nordtvedt parameter η, discussed in the earlier section on the equivalence principle, when expressed in terms of the PPN metric coefficients [43] is

$$\eta = 4\beta - \gamma - 3 - \frac{10}{3}\xi - \alpha_1 + \frac{2}{3}\alpha_2 - \frac{2}{3}\zeta_1 - \frac{1}{3}\zeta_2.$$

Using the laser reflectors placed on the moon during the Apollo 11 mission a program of laser range measurements was begun in August 1969 at the McDonald Observatory in Texas [72] to measure the earth-moon motion as a pair of bodies in the field of the sun. Typically, distance measurements at the 30-cm level of accuracy have been made in the path from the laser to the lunar reflector. As stated earlier, the results are that $\eta = 0$ within about 2 parts in 10^2.

This offers a strong confirmation of the validity of metric theories and also provides information for separating the quantities β and γ when used with data from other experiments as will be seen later.

5.2 Tests of the Behavior of Electromagnetic Signals

The predicted deflection of light grazing the sun $\delta\theta$ is given by

$$\delta\theta = \frac{\alpha^* + \gamma}{2}\,\frac{4M}{d}\,\frac{1+\cos\theta}{2},$$

where M is the "mass" of the sun and θ is the angle between the earth-sun line and the incoming direction of the light. For a grazing ray, $d = R_{\text{sun}} = 1$ solar radius and $\theta \approx 0$.

Under these conditions the prediction is

$$d\theta \approx \frac{\alpha^* + \gamma}{2} \times 1.75 \text{ arcsec}.$$

For metric theories $(\alpha^* + \gamma)/2 = 1$.

In 1964, I. I. Shapiro [73] proposed another test involving the round-trip time delay of radio or radar signals passing very close to the sun. This "fourth" test of relativity predicts that the time is given by

$$\delta t = \frac{\alpha^* + \gamma}{2}\left[250 - 20\ln\left[\frac{d^2}{r}\right]\right] \mu\text{s},$$

where d is the distance of closest approach of the ray in solar radii, and r is the distance of the planet or satellite, in astronomical units.

5.3 Solar System Light-Deflection Measurements

Light-deflection measurements begin early in the post World War I era with the celebrated test by Eddington and his co-workers who, with 30% accuracy, confirmed Einstein's prediction. More recent optical measurements have been made but the accuracy has only been improved to

$$\frac{\alpha^* + \gamma}{2} \approx 0.95 \pm 0.11$$

from the solar eclipse of 30 June 1973 [74]. Many of the difficulties with such optical lunar eclipse measurements appear to have been overcome by recent photoelectric and astrometric techniques [75] and better results are expected.

5.4 Solar System Light-Bending Tests Using Radio Interferometry Techniques

Tests of light bending have been made using long baseline and Very Long Baseline Interferometry (VLBI) techniques observing microwave noise signals from quasi-stellar radio sources passing near the sun.

Figure 4 (from Will [43]) shows the way the data have been evolving in interferometric tests made in the last decade. These are plotted in terms of the quantity $(\alpha^* + \gamma)/2$, which, for the general theory is equal to 1.

The latest results for this quantity are given by Sramek [84] who puts $(\alpha^* + \gamma)/2 = 1.007 \pm 0.009$. According to Fomalant and Sramek [83,85] this technique is currently limited by uncertainties and irregularities in the tropospheric refractive index. While it is likely that a factor of two, or so, can be gained by modeling the lens-like behavior of the troposphere lying over the radio telescopes, the future improvements are more likely to occur using VLBI techniques with spaceborne radio telescopes.

5.5 Solar System Time-Delay Tests

Measurements of the time delay of signals passing close to the sun have given the most sensitive tests of $(\alpha^* + \gamma)/2$ thus far available. These measurements were first done by radar pulses reflected from planets at the far side of the sun (superior conjunction) and later from transponders on spacecraft in flight such as Mariner 6 and 7 and spacecraft orbiting the planet or on the surface of the planets as in the Mariner 9 Mars orbiter and Viking lander and orbiter. These results are shown in Fig. 5 (from Will [43]) with later additions.

The latest results are truly impressive. Reasenberg and Shapiro [91] give $(\alpha^* + \gamma)/2 = 1.000 \pm 0.0001$ as the result of using a 24-parameter model of the earth and mars motion and position. They have made over 100 least-squares solutions with several strategies to search for hidden biases. Their data are from the Viking Lander on mars, a space system that continues to operate and provide information. From further data, and better understanding of the propagation through the interplanetary plasma, Shapiro and Reasenberg express optimism that the final uncertainty could be at the 2×10^{-4} level. This would be just about the present limit of the accuracy of α^* made in the space-borne rocket experiment described in the section on Tests of the Metric Hypothesis and would bring the uncertainty in the equivalence principle into the conclusion for the time-delay tests.

By combining the lunar laser results cited earlier with the light deflection and retardation results and treating these as independent data sets, Shapiro, Counselman, and King [63] obtain the following value for the parameters β and γ excluding the other PPN parameters, that is, assuming general relativity is correct;

$$\beta = 1.003 \pm 0.005$$

and

$$\gamma = 1.008 \pm 0.008,$$

with a correlation of 0.6.

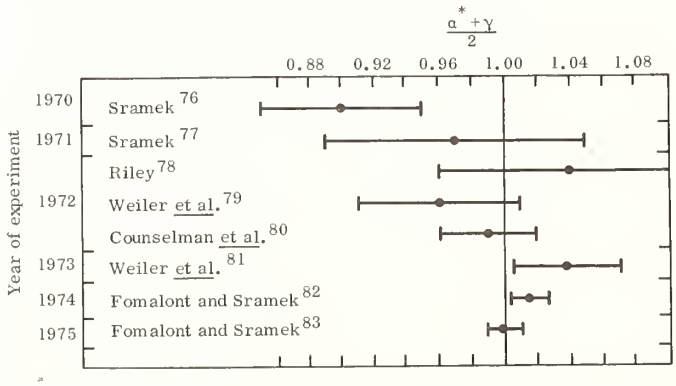


FIGURE 4. *Radio-wave-deflection experiments (from Will [43]). (The superscript number associated with each name is the appropriate reference number.)*

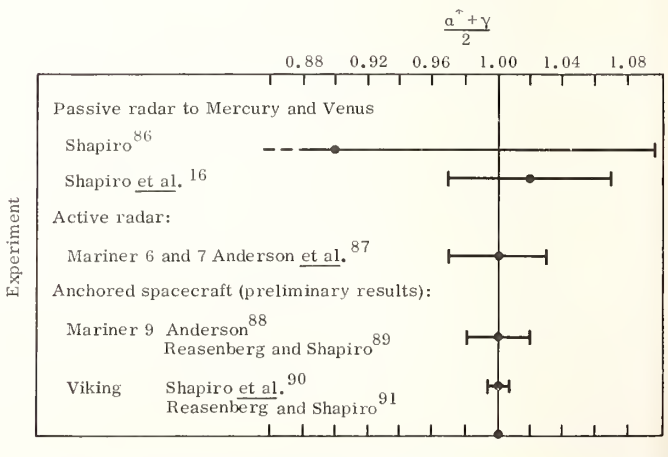


FIGURE 5. *Time-delay measurements. (The superscript number associated with each name is the appropriate reference number.)*

5.6 A Future Gyroscope Test in Earth Orbit

Thus far, all the tests discussed here have been related to the gravitational effects of static matter. There is now in preparation a test to measure the effects of moving matter, namely the rotating earth, using extraordinarily precise gyroscopes, operating at cryogenic temperatures, whose spin axes are sensed using superconducting techniques [48, 92, 93]. The spin axis direction will be measured relative to the direction of a distant star established by a telescope that is part of the space-borne system.

There are two relativistic effects that can be sensed by a gyro. For a spacecraft in a polar orbit and with the spin axis of the gyro in the plane of the orbit a relativistic effect known as the "geodetic precession," Ω_1, is predicted that depends on the space curvature parameter, γ, and is given by

$$\Omega_1 = (1/3)\ (1 + 2\gamma) \times 7 \text{ arcsec year}^{-1}.$$

The second, more interesting effect, is the Lense-Thirring or frame-dragging effect produced by the earth's rotation (see Ref. [2], §§ 19.2 and 33.4 for a discussion) and in this case, the gyroscope spin axis is aimed normal to the orbital plane as shown in Fig. 6.

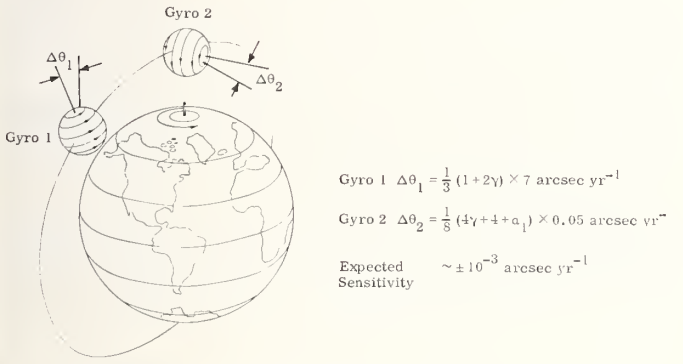


FIGURE 6. *Earth-orbiting gyro experiment [92].*

The PPN prediction for the Lense-Thirring effect in an earth-orbiting gyro is

$$\Omega_2 = (1/8)\ (4\gamma + 4) \times 0.05 \text{ arcsec year}^{-1}$$

and is proportional to the earth's angular momentum. The expected accuracy of this test is at the 10^{-3} arcsec year^{-1} level.

Another test proposed by Van Patten and Everitt [94] would involve two drag-compensated counter-rotating earth satellites in polar orbits. These, in fact, act as gyroscopes, each of whose axes are expected to rotate $\Omega_2 = (1/8)\ (4\gamma + 4) \times 0.18$ arcsec year^{-1} if placed in a circular orbit at 800-km altitude. Van Patten and Everitt estimate that Ω_2 would be measured to about 1% after two and a half years of data.

Apart from the very obvious importance of this type of test for confirming gravitational theory, there is the very interesting possibility of using the Lense-Thirring effect to measure the angular momentum of the sun by its effect in the close solar approach or "star probe" mission mentioned earlier. The measurement of both the J_2 of the sun and its angular momentum would be a boon to solar and stellar science.

A current assessment of the status of the measurement of relativistic frame dragging is given by Van Patten at this conference, and I should like to highlight the importance of this mission by reminding you that relativity and gravitation theory is an essential tool in making measurements of importance to many other branches of science.

After seeing these remarkably sophisticated experimental tests we are tempted to ask "How does general relativity theory shape up when compared to other possible theories?" The answer is that it has passed every test so far with no difficulty. The problem is that there is an ever growing list of theories that also pass these tests and that we must begin to look elsewhere to make more definitive tests. We must look for a laboratory with masses, densities, and velocities quite beyond any accessible to us in our solar system. This does not suggest that the use of our advancing technology in our terrestrial and solar system laboratory is of no avail, but that we should develop more sophisticated and sensitive detecting systems to search for other manifestations of gravitation and relativity in the form of gravitational radiation originating in other parts of our universe. We must realize that the region of space-time that mankind has traveled is extremely brief and encompasses a region that is infinitesimally small in the scale of distance, density, mass, and velocity of our universe.

6. The Search for Gravitational Radiation

Our present capability to devise laboratory experiments to test gravitation theory is limited by the masses available in our solar system. For experiments that seek to extend the range of validity of the theory we must look for naturally occurring astrophysical phenomena involving super dense matter and enormously higher masses. We are thus forced to change our posture to being observers, rather than cause-and-effect experimenters, and face the problem of unravelling from the data the characteristics of the generating phenomenon and the manifestation of the theory we are testing.

Gravitational radiation appears to be the key to understanding gravitation under these more extreme conditions and to pointing the way to post-Einstein theories of relativity. The instruments for detecting the expected extremely weak signals are at the cutting edge of today's technology and the sophistication, elegance, and astonishing sensitivity of these devices will surely contribute to other areas of science and technology.

An assessment of the possible astrophysical sources of gravitational radiation and the magnitude and characteristics of the signals that we might receive has been made in a recent book *Sources of Gravitational Radiation* edited by L. L. Smarr [95]. There are several ways in which radiation can be generated. A continuum background radiation akin to the 3 K electromagnetic radiation associated with the Big Bang origin of our universe is possible. It is highly probable that rotating binary star systems generate continuous radiation and wind-down as they lose energy. There is very strong evidence for this in the ongoing observations made by J. H. Taylor and his associates [96] from a binary pulsar system. However, detection of radiation from such systems is far below the sensitivity available in any existing gravitational wave detector. Higher level, low-frequency (10^{-3} Hz), pulsed gravitational radiation is believed to result from cataclysmic events such as supernova explosions and gravitational collapse or collision of black holes. Such signals that might arise from super massive black holes, such as are thought to exist as the cores of galaxies, are within the range of detectibility of highly sensitive Doppler detecting systems used in interplanetary tracking of spacecraft.

These closely reasoned speculations have provided a basis, or hunting license, to search for gravitational radiation. The largest amplitudes of radiation are expected from pulsed sources, a study of which is shown in Fig. 7 [97]. In this figure the amplitude, h, is a measure of the frequency difference observed between a pair of proper clocks, one at the crest, the other at the trough of a wave. this is also a measure of the fractional change in distance over that wave.

The frequency of the main pulse component is related to the size of the collapsing object. To the left, the region 10^4 to 10^2 Hz describes the results of supernova core collapses involving 10 to 100 solar masses in the Virgo Cluster of galaxies, which are at a distance of approximately 15 megaparsecs (Mpc) from earth. A burst rate greater than one per year is expected. The efficiency of the energy conversion, η, is shown as 1% (lower line) and 20% (upper line). As the frequency goes lower (to the right) the effect of more massive black hole formation involving $10^2 - 10^4$ solar masses at distances to 500 Mpc in globular clusters is shown.

At the lowest frequencies $\sim 10^{-3}$ Hz, the expected sources are supermassive ($10^6 - 10^7$ solar masses) black holes in galactic nuclei and quasars at distances associated with the early formation of galaxies. The rate of occurrence of these events is difficult to estimate, however. Smarr [97] estimates this to be at least one per year on the assumption that galaxies like our own are assumed to have a supermassive black hole at its core and that it is caused by a sudden collapse. Above the lines shown as "Black-Hole Upper Limits" is a line labeled fundamental upper limit. This is a constraint based on the conversion of all galaxies into gravitational radiation in a Hubble time, which characterizes the age of our universe. In order to exist at all we must clearly be well below this limit. The question is, at what rate of occurrence and at what level do gravitational signals exist? We should remember that proof of nonexistence of such waves at levels below the solid lines would be of enormous interest to theorists, as all gravitation theories predict some form of waves.

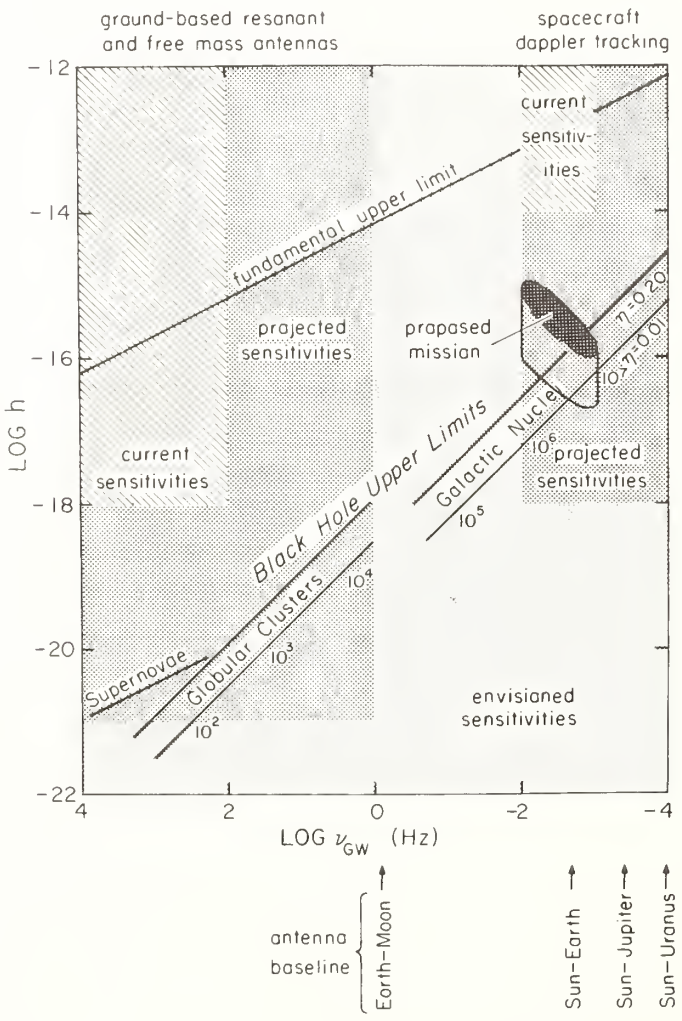


FIGURE 7. *Intensity vs. frequency of gravitational radiation and estimated sensitivity of detectors.*

The detection of gravitational waves was pioneered by Joseph Weber during the 1960s [98] using a large aluminum bar as a resonant detector to look for signals in the kilohertz range. Since that time similar detectors have been built at a dozen or more sites throughout Europe and North America and some of these have attained sensitivities of 1×10^{-16} in the parameter h. At Stanford University a second generation of these bars using cryogenic techniques [99, 100] has reached a sensitivity of 3×10^{-18} at temperatures below 2 K.

A generalization of the types of gravitational wave detectors is shown in Fig. 8 [101] from Douglass and Braginsky [10]. First is the ideally-free mass antenna where differences in distance measured between the two masses is used to detect the gravitational disturbance. Here we visualize a pair of widely separated stations, such as in a deep space probe and on earth, and the measurement of changes in separation arising from a very low frequency ($\sim 10^{-3}$ Hz) wave crossing the solar system by use of Doppler tracking techniques [97].

The tracking of distant spacecraft has already provided some upper limits to the magnitude of gravitational pulses [102]. The current levels of sensitivity are shown in the

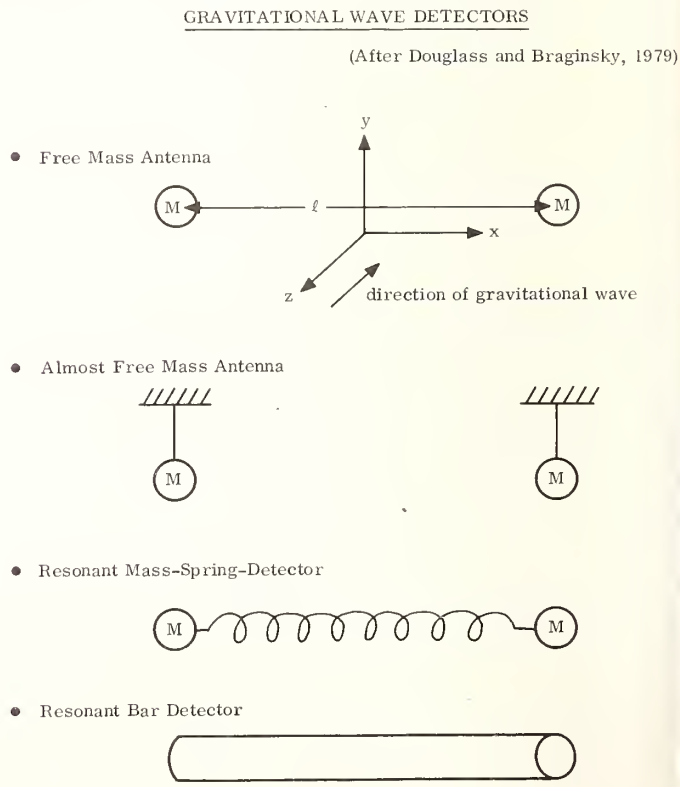


FIGURE 8. *Gravitational wave detectors [101].*

upper right corner of Fig. 7. Substantially improved sensitivity is possible and design changes are being implemented in deep space tracking systems to improve the sensitivity for gravitational wave detection.

Present atomic clock and Doppler tracking technology, if fully utilized, is at the stage where it would be possible to detect gravitational disturbances in the region described as "proposed mission." This involves a multilink [42] Doppler system with time correlated processing of Doppler data taken both at the spacecraft and at the earth station. Tracking from an earth orbiting station using cryogenically stabilized oscillators could improve the performance to the levels shown within the solid line below the "proposed mission" region. Work is in progress [103] to assess the possibilities of this system by computer simulation of Doppler data.

Another means of detection is the use of almost free masses. These are antennas made of several carefully suspended masses whose distances are measured using highly sophisticated laser techniques that extend the path distance by many reflections back and forth. Considerable effort is being made at several universities to develop this antenna and results are expected in the near future.

The resonant mass-spring and the closely related bar detector, discussed earlier, complete the present picture. Studies of these various forms lead us to project the sensitivity of the bar and almost free mass antennas to the 10^{-21} level at frequencies as low as 1 Hz. The projected sensitivity of free mass Doppler techniques could reach as low as 10^{-18} and, given the very precise use of highly stabilized lasers in position-stabilized spacecraft, we can envision reaching the 10^{-22} level.

The signals, once we get to see them, will provide a mixture of astrophysical information, depending on the manner in which they were generated, along with information relating to gravity theory. The latter is contained in the nature of the signals themselves, their velocity, polarization states, and the multipole characteristics, i.e., monopole, dipole, quadrupole, etc., of the radiation process.

The effort to detect gravitational waves is now a major, worldwide effort involving over 100 scientists. Progress has been very encouraging and the many inventions and new ideas from this work have provided valuable new technology. I believe we can look forward to some very exciting and beautiful results in the next decade.

7. Conclusions

The study of gravitation has been going on since the early days of recorded history. In the modern era, gravitational experiments and accurate astronomical measurements have guided the thinking of early scientists and philosophers and led to the monumentally important *Principia* of Isaac Newton and later to the immensely beautiful legacy left by Albert Einstein in his General Theory of Relativity. Questions of the validity of Einstein's theory still exist, however, and the field of theoretical gravitation is currently wide open in mathematical speculation and stands badly in need of experimental evidence. It is clear that Einstein's Special Theory is on solid ground. Ongoing high-energy physics experiments continue to press the speed of particles ever closer to the velocity of light and the behavior of particles and radiation under local conditions continues to be well described by the Special Theory.

The metric description of gravitation is based on the idea of a Riemannian geometry in which everywhere, locally, the flat space of special relativity is embedded. It is today the most valid theoretical description of physics that can accommodate gravitation. The original version of the metric theories, Einstein's General Theory, so far has beautifully withstood all experimental tests. However, we have only scratched the surface of this important topic; we have only been able to work in a very small range of gravitational potential ϕ, such that at best we have only explored $\Delta\phi/c^2$ for values below 1×10^{-6}. An important question is, what happens for $\Delta\phi/c^2$ for far larger values, even going to values, $\Delta\phi/c^2 \rightarrow 1$? The highly probable existence of black holes and the beguiling ideas of mathematical singularities imply physical conditions far beyond any we have ever tested and seriously challenge us to define some range of validity where the now-classical Einstein Theory could be considered as a limiting case of a Post-Einstein Theory. Whether it will be possible to see evidence for this in tests done in the next decades within our solar system seems doubtful. It is likely that any such evidence will have to come from observations of astrophysical phenomena and the role of the experimenter will become that of an observer equipped with the highest levels of available technology.

The future is very bright for this type of scientist-technologist-observer. I believe that we are at a threshold where new information will be forthcoming from large telescopes in space operating at all frequencies in the electromagnetic spectrum. Very high angular resolution will be possible with pairs of space telescopes using very long baseline interferometry techniques. The enormous increase in angular resolution and light gathering ability available from space will vastly extend our view of the universe. Also, I am convinced that the discovery of gravitational waves is not speculation, *it will happen* and will have an impact comparable to Galileo's use of the astronomical telescope in place of the naked eye. The likelihood that the results of these measurements and observations will pose far more new questions than answers to old ones is a powerful and constructive manifestation of our evolution and our search for knowledge. There are few, more worthwhile goals that our civilization can pursue.

I should like to thank Dr. B. N. Taylor for his recommendations and guidance in my attempt to organize this paper. The assistance and suggestions of Dr. Alan Lightman, in reviewing the manuscript, have been of great value; I am grateful for his patience and willingness to help. Also, I should like to thank Mrs. Anne Omundsen and Mrs. Gerda Schrauwen of the SAO Editorial Department for their excellent advice and cooperation in putting together this paper. The National Aeronautics and Space Administration (NASA) and the Smithsonian Institution have both provided support for this publication.

References

[1] C. M. Will, Astrophys. J. **169**, 125 (1971).
[2] C. W. Misener, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman and Company, San Francisco, 1973).
[3] K. S. Thorne and C. M. Will, Astrophys. J. **163**, 595 (1971).
[4] I. Newton, *Principia* (1687).
[5] S. W. Hawking and W. Israel, Eds., *General Relativity, An Einstein Centenary Survey* (Cambridge University Press, Cambridge, 1979).
[6] C. M. Will, *Theory and Experiment in Gravitational Physics* (Cambridge University Press, Cambridge, 1979).
[7] C. M. Will, in *Experimental Gravitation*, Ed. by B. Bertotti (Academic Press, New York, 1974), p. 1.

[8] T. Alvager *et al.*, Phys. Rev. Lett. 12, 260 (1964).
[9] K. Brecher, Phys. Rev. Lett. 39, 1051 (1977).
[10] Z. G. T. Guiragossian *et al.*, Phys. Rev. Lett. 34, 335 (1975).
[11] B. C. Brown *et al.*, Phys. Rev. Lett. 30, 763 (1973).
[12] A. Brillet and J. L. Hall, Phys. Rev. Lett. 42, 549 (1979).
[13] L. V. Morrison, Nature (London) 241, 519 (1973).
[14] T. C. Van Flandern, Bull. Am. Astron. Soc. 6(3), 431 (1975).
[15] T. C. Van Flandern, Mon. Not. R. Astron. Soc. 170, 333 (1975); Ann. N. Y. Acad. Sci. 262, 494 (1975).
[16] I. I. Shapiro *et al.*, Phys. Rev. Lett. 26, 1132 (1971).
[17] R. D. Reasenberg and I. I. Shapiro, in *On the Measurement of Cosmological Variations of the Gravitational Constant* (Proceedings of a Workshop, November 1975, Dept. of Physics, Florida State University, Tallahassee), Ed. by L. Halpern (University Press of Florida, Gainesville, 1978).
[18] D. Newman *et al.*, Phys. Rev. Lett. 40, 1355 (1978).
[19] P. S. Cooper *et al.*, Phys. Rev. Lett. 42, 1386 (1979).
[20] J. Bailey *et al.*, Nuovo Cimento A 9, 369 (1972).
[21] J. Bailey *et al.*, Phys. Lett. 68B, 191 (1977).
[22] R. S. Van Dyck, P. B. Schwenberg, and H. G. Dehmelt, Phys. Rev. Lett. 38, 310 (1977).
[23] J. C. Wesley and A. Rich, Phys. Rev. A 4, 1341 (1971).
[24] D. J. Grove and J. G. Fox, Phys. Rev. 90, 378 (1953) [Abstract UA7].
[25] H. E. Ives and S. P. Stilwell, J. Opt. Soc. Am. 28, 215 (1938).
[26] H. I. Mandelberg and L. Witten, J. Opt. Soc. Am. 52, 529 (1962).
[27] R. V. Pound and G. A. Rebka, Phys. Rev. Lett. 4, 272 (1960).
[28] R. F. C. Vessot and M. Levine, Metrologia 6, 116 (1970).
[29] J. Bailey *et al.*, Nature (London) 268, 301 (1977).
[30] M. E. Nordberg, F. Lobkowicz, and R. L. Barman, Phys. Lett. 24B, 594 (1967).
[31] D. S. Ayres *et al.*, Phys. Rev. 159, 1288 (1967).
[32] D. S. Ayres *et al.*, Phys. Rev. D 3, 105 (1971).
[33] A. F. Dunaitsev *et al.*, Sov. J. Nucl. Phys. 16, 292 (1973).
[34] R. J. Ott and T. W. Pritchard, Phys. Rev. D 3, 52 (1971).
[35] F. Lobkowicz *et al.*, Phys. Rev. 185, 1676 (1969).
[36] W. C. Carithers *et al.*, Phys. Rev. Lett. 34, 1244 (1975).
[37] O. Skjeggstad *et al.*, Nucl. Phys. B 48, 343 (1972).
[38] C. Geweniger *et al.*, Phys. Lett. 48B, 487 (1974).
[39] P. Nachman, M. Rayman, and J. L. Hall, these proceedings.
[40] R. V. Pound and J. L. Snider, Phys. Rev. 140, B788 (1965).
[41] R. F. C. Vessot and M. W. Levine, in *Experimental Gravitation*, Ed. by B. Bertotti (Accademia Nazionale, Lincei, Rome, 1977), p. 371.
[42] R. F. C. Vessot, Radio Sci. 14, 629 (1979).
[43] C. M. Will, in *General Relativity, An Einstein Centenary Survey*, Ed. by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979), p. 24.
[44] F. W. Bessel, Pogg. Ann. 25, 401 (1832).
[45] R. V. Eötvös, V. Pekar, and E. Fekete, Ann. Phys. (Leipzig) 68, 11 (1922).
[46] R. H. Dicke, in *Relativity Groups and Topology*, Ed. by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964), p. 165.
[47] V. B. Braginski and V. I. Panov, Sov. Phys. JETP 34, 463 (1972).
[48] C. F. W. Everitt, in *Experimental Gravitation*, Ed. by B. Bertotti (Academic Press, New York, 1974), p. 331.
[49] P. W. Worden, A cryogenic test of the Equivalence Principle, Ph. D. Thesis, Stanford University (1967).
[50] L. I. Schiff, Am. J. Phys. 28, 340 (1960); see page 343.
[51] A. P. Lightman and B. L. Lee, Phys. Rev. D 8, 7 (1973).
[52] R. F. C. Vessot *et al.*, Phys. Rev. Lett. 45, 2081 (1980).
[53] V. W. Hughes, H. G. Robinson, and V. Beltran-Lopez, Phys. Rev. Lett. 4, 342 (1960).
[54] R. W. P. Drever, Philos. Mag. 6, 683 (1961).
[55] J. Turneaure *et al.*, to be published (1981).
[56] C. M. Will, Phys. Rev. D 10, 2330 (1974).
[57] C. M. Will, Metrologia 13, 95 (1977).
[58] L. I. Schiff, Am. J. Phys. 28, 340 (1960).
[59] K. S. Thorne, B. L. Lee, and A. P. Lightman, Phys. Rev. D 7, 3563 (1973).
[60] K. Nordtvedt, Phys. Rev. 169, 1014 (1968a); Phys. Rev. 169, 1017 (1968b); Phys. Rev. 170, 1186 (1968c).
[61] K. Nordtvedt, Phys. Rev. 180, 1293 (1969).
[62] J. G. Williams *et al.*, Phys. Rev. Lett. 36, 585 (1976).
[63] I. I. Shapiro *et al.*, Phys. Rev. Lett. 36, 555 (1977).
[64] W.-T. Ni, Phys. Rev. Lett. 38, 301 (1977).
[65] K. S. Thorne, C. M. Will, and W.-T. Ni, in *Proceedings of the Conference on Experimental Tests of Gravitational Theories*, Ed. by R. W. Davies (NASA-JPL Tech Memo 33-499, 1971), p. 10.
[66] A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, Cambridge, 1922).
[67] H. P. Robertson, in *Space Age Astronomy*, Ed. by A. J. Deutsch and W. B. Klemperer (Academic Press, New York, 1962), p. 228.
[68] L. I. Schiff, in *Relativity Theory and Astrophysics. I. Relativity and Cosmology*, Ed. by J. Ehlers (American Mathematical Society, Providence, 1967), p. 105.
[69] C. Brans and R. H. Dicke, Phys. Rev. 124, 925 (1961).
[70] R. H. Dicke and H. M. Goldenberg, Astrophys. J. Suppl. Ser. 27, 131 (1974).
[71] H. A. Hill *et al.*, Phys. Rev. Lett. 33, 1497 (1974).
[72] P. L. Bender *et al.*, Science 182, 229 (1973).
[73] I. I. Shapiro, Phys. Rev. Lett. 13, 789 (1964).
[74] B. F. Jones, Astron. J. 81, 455 (1976).
[75] J. M. Hill, Mon. Not. R. Astron. Soc. 153, No. 3 (1971).
[76] R. A. Sramek, Astrophys. J. Pt. 2, Lett. 167, L55 (1971).
[77] R. A. Sramek, in *Experimental Gravitation*, Ed. by B. Bertotti (Academic Press, New York, 1974), p. 529.
[78] J. M. Riley, Mon. Not. R. Astron. Soc. 161, No. 3 (1973).
[79] K. W. Weiler *et al.*, Astron. Astrophys. 30, 241 (1974).
[80] C. C. Counselman *et al.*, Phys. Rev. Lett. 33, 1621 (1974).
[81] K. W. Weiler *et al.*, Phys. Rev. Lett. 35, 134 (1975).
[82] E. B. Fomalont and R. A. Sramek, Astrophys. J. 199, 749 (1975).
[83] E. B. Fomalont and R. A. Sramek, Phys. Rev. Lett. 36, 1475 (1976).
[84] R. A. Sramek, Astrophys. J. Pt. 2, Lett. 167, L55 (1971).
[85] E. B. Fomalont and R. A. Sramek, Comment Astrophys. 7, 19 (1977).
[86] I. I. Shapiro, Phys. Rev. Lett. 20, 1265 (1968).
[87] J. D. Anderson *et al.*, Astrophys. J. 200, 221 (1975).
[88] J. D. Anderson, in *Proceedings of the Third International Space Relativity Symposium* (27th Congress, International Astronomical Federation, 1977).
[89] R. D. Reasenberg and I. I. Shapiro, in *Proceedings of the International Meeting on Experimental Gravitation*, Ed. by B. Bertotti (Academia Nazionale dei Lincei, Rome, 1977), p. 143.
[90] I. I. Shapiro *et al.*, J. Geophys. Res. 82, 4329 (1977).
[91] R. D. Reasenberg and I. I. Shapiro, in *Proceedings of the Ninth Conference on Gravitation and General Relativity* (1980), in press.
[92] L. I. Schiff, Phys. Rev. Lett. 4, 215 (1960).
[93] J. A. Lipa *et al.*, in *Experimental Gravitation*, Ed. by B. Bertotti (Academic Press, New York, 1974), p. 361.
[94] R. A. Van Patten and C. F. W. Everitt, Phys. Rev. Lett. 36, 629 (1976).
[95] L. L. Smarr, Ed., *Sources of Gravitational Radiation, Proceedings of the Battelle Seattle Workshop* (Cambridge University Press, Cambridge, 1978).
[96] J. H. Taylor *et al.*, Astrophys. J. Pt. 2, Lett. 206, L 53 (1976).
[97] L. L. Smarr *et al.*, Gen. Rel. Grav. (in press).
[98] J. Weber, Phys. Rev. Lett. 18, 498 (1967).
[99] Stanford Gravity-Wave Detector Group, Preliminary evaluation of the performance of the Stanford 4800 kg low temperature gravitational wave detector, Stanford University Preprint, June 1980.
[100] V. B. Braginsky and K. S. Thorne, Present status of gravitational wave experiments, Preprint Series No. 609, California Institute of Technology, 1980.
[101] D. H. Douglass and V. B. Braginsky, in *General Relativity, An Einstein Centennial Survey*, Ed. by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979), p. 90.
[102] J. W. Armstrong, R. Woo, and F. B. Estabrook, Astrophys. J. 230, 570 (1979).
[103] T. Piran *et al.*, in preparation (1981).

Is the Gravitational Constant Changing?

T. C. Van Flandern

U. S. Naval Observatory, Washington, DC 20390

Forty-four years after Dirac's original conjecture, experiments have become accurate enough to test for a time variation of the Universal Gravitational Constant. Present results utilize the lunar orbit, and do give a non-zero result. Interpreting the observed quantities with the Canuto-Hsieh scalar covariant cosmology, $\dot{G}/G = (-6.4 \pm 2.2) \times 10^{-11}$ per year. The implications for relativity and cosmology are briefly discussed.

Key words: cosmology; gravitation; laws of physics; occultations; universe.

1. Introduction

Can a fundamental constant of physics change with time? No such variation has ever previously been detected, lending confidence to the widely-held belief that the fundamental constants really are constant. However, the 1937 conjecture of P.A.M. Dirac [1] cannot be ignored in this connection, not only because experimental tests have only recently been possible, but also because it provides a theoretical prediction that the Universal Gravitational Constant G alone among the fundamental constants of physics can vary.

Dirac's conjecture is a byproduct of the so-called Large Numbers Hypothesis. In essence, if the fundamental constants are all expressed as dimensionless ratios to rid them of man-made dimensions, for example by using units of length, mass, and time from the hydrogen atom, the constants tend to all lie within a few orders of magnitude of unity, with the notable exception of the gravitational constant. This last stands out from the others, its reciprocal being forty orders of magnitude greater. It seems that the age of the universe is also forty orders of magnitude greater than the light time for the Bohr radius of the hydrogen atom. This coincidence already mildly suggests a connection, unless nature is being somewhat perverse. Additionally, however, Dirac pointed out the formidable difficulties in explaining such large numbers in terms of purely mathematical and geometric concepts, as we hope to do someday with, e.g., the fine-structure constant. This reasoning also implies a time variation for these large numbers. In particular, if the coincidence with the age of the universe is not merely fortuitous, G would have to vary inversely with time to maintain its relation to the age of the universe.

2. The Experimental Evidence

Tests for a time variation of G have only recently become possible due to the availability of such devices as atomic clocks and retroreflectors on the Moon, and the capability of sending radar beams to the planets. The principle behind such experiments is to detect the acceleration in orbital motion of astronomical bodies which would occur if G changes. The precise relationship between variations in G and changes in a body's angular velocity (n) is theory dependent because G always appears as a factor of mass (M) in orbital motion, and different theories predict different behaviors for the product GM with time. In particular, no existing theory predicts simple adiabatic changes in G with no accompanying changes in M. In general, we have

$$\dot{G}/G = f(\dot{n}/n),$$

where a dot denotes a time derivative, and f is a constant supplied by the theory. For the Dirac theories, f must be either $+1$ or -1, depending on the form of mass change. For a so-called "primitive" theory in which mass does not change, $f = 1/2$. No existing theory known to this author has f numerically greater than 2.

Measuring changes in G therefore can be accomplished by measuring changes in the angular velocity of astronomical bodies. This process of necessity involves the use of an accurate clock. It will become apparent on reflection that no astronomical body can serve as a clock for this purpose, since the clock must not disguise the effect sought. This means that Ephemeris Time and Dynamical Time, both of which involve the orbital motion of bodies, cannot be used. On the other hand, Atomic Time is a fully suitable time scale. For this reason, we cannot utilize measurements of the orbital acceleration of, for example, the Moon over the last few centuries, when only dynamical clocks are available, except as a control (for example, to determine the angular acceleration from other causes, such as tidal friction). Only if the Moon's acceleration is different when measured with atomic and dynamical clocks can we impute the cause to a changing G.

3. Using the Moon

The Moon would provide an ideal celestial body for the necessary measurements if it were not for accelerations from other causes, such as tidal friction. However, for the reasons just detailed, it is possible to separate the tidal accelerations (using Dynamical Time) and the total accelerations (using Atomic Time) to determine whether there is any non-zero contribution from changes in G. Experiments which measure the tidal lunar acceleration [2] include classical telescopic observations, transits of Mercury across the Sun's disk, ancient eclipse observations, and ocean tide models. The mean of these individual determinations is

$$\dot{n}_{\text{tidal}} = (-28.8 \pm 1.5)''/\text{cy}^2$$

(cy = century).

Determinations of the total lunar acceleration (using

Atomic Time, available since 1955) have heretofore relied exclusively on timings of occultations of stars at the Moon's limb. Preliminary results using this method [3] have been affected by systematic errors due, in part, to the star catalogs and lunar ephemeris in use for the analysis. (See Ref. [2] for more discussion of current determinations of the lunar acceleration.) These problems have all been addressed and corrected insofar as possible; and the lapse of additional time has made the matter somewhat academic anyway, since the effects of acceleration eventually become larger than systematic errors of any given size. The lunar occultation results now give

$$\dot{n}_{\text{total}} = (-21.4 \pm 2.6)''/\text{cy}^2 .$$

Additionally, results for this same parameter have recently become available from lunar laser ranging to the retroreflectors left on the Moon by the astronauts. The mean errors of the laser experiment have also been rapidly decreasing, and are now for the first time better than those from the occultation results. It is therefore gratifying that the two experiments are in reasonable accord, with the laser ranging results giving

$$\dot{n}_{\text{total}} = (-23.6 \pm 1.5)''/\text{cy}^2$$

in the latest estimate [4]. Combining the occultation and laser results (and including a few additional estimates by other experimenters using the same laser data), we get

$$\dot{n}_{\text{total}} = (-23.2 \pm 1.2)''/\text{cy}^2 .$$

4. The Rate of Change of G

Comparing the tidal and total acceleration determinations, we note a discrepancy which has at least formal statistical significance. Taking the difference and dividing by the lunar angular velocity $n = 1.7333\times10^{11}''/\text{cy}$ gives for this discrepancy

$$\dot{n}/n = (+3.2 \pm 1.1) \times 10^{-11}/\text{yr} .$$

The most complete and internally consistent cosmological framework for variable G to date is that of Canuto and Hsieh [5], since it has addressed and resolved many supposed conflicts between variations in G and physical, geophysical, or astrophysical constraints. An important feature of this scalar-covariant theory is that it retains full consistency with the theory of general relativity, since G does not change with respect to Dynamical Time (only with respect to Atomic Time). Best agreement with most physical constraints is secured if the parameter f, relating $\dot{n}/n$ to $\dot{G}/G$, has the value -2. In this case, we arrive at the result

$$\dot{G}/G = (-6.4 \pm 2.2) \times 10^{-11}/\text{yr} ;$$

or equivalently, in the more usual choice of units in cosmology, $\dot{G}/G = (65 \pm 22)$ km/(s·Mpc) (Mpc = megaparsec). This value can now be compared directly with various determinations of the Hubble rate of expansion of the universe, with which it should be in agreement according to Dirac's Large Numbers Hypothesis. Recent determinations of the Hubble constant have ranged from 50 to somewhat over 100 in these units, which is clearly consistent with the hypothesis in view of the errors in both determinations.

The only other experiment with even preliminary results to date is planetary radar ranging [6]. These authors stress that the errors in their results at this stage make them consistent with a zero value for $\dot{G}/G$. The averaged determination from the individual results for each of the inner planets is

$$\dot{n}/n = (+12.4 \pm 6.6) \times 10^{-11}/\text{yr} ,$$

which is in even better agreement with the above results than it is with zero.

5. Cosmological Implications

The simplest and most literal meaning of the experimental measurements of a different angular acceleration of the Moon with respect to Atomic Time and Dynamical Time is that the number of atomic seconds in a dynamical interval (such as a revolution of the Earth about the Sun) is becoming fewer. Presumably, if the result has any generality to it, this means that atomic phenomena are slowing down with respect to dynamical phenomena. If the speed of light is constant (as assumed in the Large Numbers Hypothesis and the Canuto-Hsieh theory, but an assumption nonetheless), then we can make the same deduction about units of length: as time progresses, fewer and fewer hydrogen atoms will fit into the distance from the Sun to the Earth, or into any orbit whose dimensions are determined by the strength of gravitation.

The first deduction we can make about this result is that we cannot tell from existing data whether the changes are occurring on the atomic level or the dynamical level. Moreover experiments which purport to show that various atomic constants are invariant with respect to time do not rule out change occurring at the atomic level, since no properties of atoms, measured in atomic units, would be expected to vary. Indeed, we must take care to reinterpret all previous experimental results in the light of this new (although tentative) finding. Assumptions such as the constancy of the velocity of light, already made above, may be true only in one set of units (atomic or dynamical), but not the other. It is precisely such confusion which has led to some earlier erroneous deductions about implications of variations in G, which the scalar covariant theory of Canuto and Hsieh corrects [5].

The next important implication is that, even though these variations of scale which we call a decrease in G are related to the Hubble age of the universe, they are not by themselves a cause of or an explanation for the observed expansion of the universe unless the speed of light also varies, at least in dynamical units. The reason is that the redshifts of distant galaxies, which are used to measure the rate of expansion of the universe, can easily be expressed as a fraction of the velocity of light, c. For example, if the universe had constant linear dimensions in both dynamical and atomic units, the increase in redshift with distance (or equivalently, with lookback time) would imply an increase in c at past epochs, or that c was decreasing as time moves forward. If on the other hand c remains constant in dynamical units, then so do redshift measurements; and the Hubble rate of expansion of the universe in km/(s·Mpc) can be non-zero only if the dimension of megaparsecs (a dynamical unit of length) varies with atomic time. For this reason we deduce that, as long as c is taken constant, the Hubble age of the universe is its age in Dynamical Time.

If we let t measure time in dynamical units and τ measure time in atomic units, then let t_1 and τ_1 be the respective measures on both scales for the present epoch. We will wish to impose the constraint that $t_1 = \tau_1$ for convenience. Finally let t_o be the epoch of the "Big Bang" or start of the expansion of the universe on the dynam-

ical time scale. Then the Large Numbers Hypothesis and the experimental results apparently tell us that

$$d\tau/dt = 1/(t - t_o),$$

from which we can easily derive

$$(t - t_o)/(t_1 - t_o) = e^{(\tau - \tau_1)}.$$

As t goes to infinity, so does τ; but note that as t approaches t_o, τ approaches minus infinity. The singularity at the time of origin of expansion therefore occurs in atomic processes, but not necessarily in dynamical processes. More rigorously, it occurs in the ratio of atomic to dynamical time interval. This must certainly have profound implications for our current understanding of the origin and evolution of the universe, although it is not immediately clear what all of the ramifications may be. It seems intuitively to represent a curious wedding of the Big Bang and Steady State theories of cosmology which contradicts neither.

It may be worth mentioning one additional implication of varying G, insofar as it relates to Mach's Principle of Inertia and the question of the influence of the distant matter in the universe on local dynamical systems. Newton's law of universal gravitation may be written as

$$F = GMm/d^2,$$

where F is the force between two masses, M and m, at a distance d apart. We can also equate F to kma, where a is the resultant acceleration of the body of mass m in response to the force, and k is a constant of proportionality. It is customary to equate these two expressions for F, cancelling m in the process, and obtain an expression for the acceleration of m which is independent of m itself:

$$a = (G/k)M/d^2.$$

However the cancellation of m disguises an important assumption. In the law of gravitation, m is the gravitational mass of the body, whereas in $F = kma$, m is its inertial mass. The ratio of gravitational to inertial mass for a body has been shown to be constant to parts in 10^{13} over a range of different chemical compositions; but the time dependence of this ratio is not readily measurable. It may be seen that observed variations in G may equally well be ascribed to variations in k, or equivalently, to time variations in the ratio of gravitational to interial masses for all bodies. This may be thought of as a change in the local property of inertia, perhaps caused by the expansion of the distant matter in the universe, or perhaps due entirely to changes at the atomic level. In either case it permits the incorporation of Mach's Principle into cosmological theories in a new and interesting way.

6. Conclusion

The tentative conclusion from experimental evidence now available from the lunar occultation and lunar laser ranging experiments is that the universal gravitational constant G seems to be decreasing with time at a rate whose exact value is theory-dependent, but which is in qualitative agreement with the Hubble rate of expansion of the universe, as predicted by Dirac's Large Numbers Hypothesis. In accordance with the internally consistent scalar covariant theory of Canuto and Hsieh, which also retains full consistency with General Relativity, the observed rate would be $(-6.4 \pm 2.2) \times 10^{-11}$ per year. Very preliminary results from planetary radar ranging, which has the advantage that the Moon's orbit is not involved, in no way disagree with this measurement. The implications of this result for our understanding of the origin and ultimate fate of the universe are profound, although not yet fully elaborated.

References

[1] P. A. M. Dirac, Proc. R. Soc. London Ser. A: **165**, 199 (1938).
[2] T. C. Van Flandern, Astrophys. J. **248**, 813 (1981).
[3] T. C. Van Flandern, Mon. Not. R. Astron. Soc. **170**, 333 (1975).
[4] J. O. Dickey, J. G. Williams, and C. F. Yoder, EOS Trans. Am. Geophys. Union **61**, 939 (1980).
[5] V. Canuto and S. H. Hsieh, Astrophys. J. **224**, 303 (1978).
[6] R. D. Reasenberg and I. I. Shapiro, in *On the Measurement of Cosmological Variations of the Gravitational Constant*, Ed. by L. Halpern (Univ. of Florida Presses, Gainesville, FL, 1978), p. 71.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Experiments on Variation of the Gravitational Constant Using Precision Rotations

G. T. Gillies and R. C. Ritter

Department of Physics, University of Virginia, Charlottesville, VA 22901

The classical rotor for gravitational studies, the earth, has a varying decay time of $\sim 3 \times 10^9$ years and a rotational roughness $\sim 3 \times 10^{-8}$ per day. A more rigid, highly protected, laboratory rotor might be made more suitable for gravitational measurements such as tests for the temporal variation of Newton's G and for cosmological matter creation. In this paper past tests for variation of G with other parameters are reviewed and the first room-temperature tests of precision rotors for such "laboratory cosmology" experiments are discussed. The design of two such experiments at Virginia and the early results are presented for one protected rotor aimed at ultimately testing for matter creation at rates below $\dot{m}/m \simeq 10^{-10}$ year^{-1}. Rotor oscillations are observed in the rotating reference frame, having energy dissipation rates $\sim 10^{-14}$ W, thus giving some evidence that "corotation protection" against gas drag is effective.

Key words: gravitational constant variations; magnetic suspensions; matter creation; Newtonian gravitational constant G; precision rotations.

1. Introduction and Background

Much of our quantitative knowledge of the gravitational force has come from measurements of the Newtonian constant of gravitation, G. Because of gravity's unique weakness and long range, however, highly precise results are difficult to achieve in experiments of this type. Nevertheless, several rather conclusive tests have been performed in the last century, and the conclusions drawn from these efforts are usually summarized as follows: In weak fields, the gravitational force between two masses is almost perfectly described by Newton's Law, and the force is independent of the physical and chemical state of either the masses in question or the intervening medium (if any).

From other types of experimental observations and from certain types of astronomical data we can further conclude that in strong fields, Einstein's General Relativity more exactly explains gravitational and mechanical phenomenon. But is there more?

There have been about 200 measurements of G or equivalently, of the mean density of the earth. For a thorough review of them, see Refs. [1] and [2]. While most of these efforts were aimed solely at establishing and then improving the absolute value of G, some were studies of a more fundamental nature.

For instance, while most workers used solid spheres or cylinders as the fixed attracting masses in their balances, Cremieu [3] used 20 kg of liquid mercury instead. The good agreement between his result and the others rules out a dependence of G on the physical state of the masses. In a far more comprehensive experiment, Eötvös [4] demonstrated that the gravitational force was independent of both the physical state of the masses as well as their chemical structure down to the $10^{-9}\,G$ level. More recent experiments confirm this at even higher levels of accuracy (see Refs. [5] and [6]).

A different, but contemporary, series of experiments was devoted to a search for variations in G with respect to the direction of the test mass crystalline axis. Since so many properties of crystals depend on the direction of observation (e.g., refractive index, thermal conductivity, etc.), it seemed reasonable to search for similar types of gravitational anisotropies. Mackenzie [7] and Poynting and Gray [8] were the first to do this. Although their experiments differed in design, the outcome of each was a null result at the level of one part in 900 and one part in 16,000, respectively. Somewhat later, Stratton and then Heyl [9] made new sets of measurements. Heyl found that he could put a limit of $10^{-9}\,G$ on the gravitational anisotropy of crystals, a number he arrived at by carefully weighing samples from each of the five nonisometric crystal groups.

Along similar lines, Louis Bauer of the Carnegie Institution in Washington weighed both magnetized and unmagnetized steel samples in order to see if the magnetic energy density (or the lack of it) produced a variation in the gravitational attraction between the steel and the earth. Heyl points out [10] that Bauer found no such variation.

Shaw and Davy searched for a variation in G with respect to temperature. After a series of modifications to their Cavendish-type apparatus [11], they were able to eliminate some troublesome systematic effects and finally conclude that if $G = G_0(1 + \alpha T)$ where T is the temperature of the attracting masses in kelvins, then α is less than 2×10^{-6} kelvin.

Another classic series of experiments began in 1897 with Austin and Thwing's research on "gravitational permeability" [12]. They used a Boys-type balance which had been designed such that intervening masses made of various substances could be placed between the large and small masses of the balance. Measurements of G were then made with and without these intervening masses. Their data showed no variations in G as a result of screening of the gravitational force by the intervening masses, at least down to the $10^{-3}\,G$ level. This result, although null, created a great deal of interest, and during the following 30 years, no fewer than ten other experimental searches for gravitational screening were carried out. All of these reported null results, but the levels of sensitivity were steadily increasing. Then in 1920, Majorana [13] claimed to see a screening of the gravitational force at the $10^{-10}\,G$ level. Russell [14] disputed this, and

subsequent work by Majorana [15] with an altered Ruprecht balance confirmed that his initial results were in error and that gravitational screening was in fact undetectable even at the $10^{-11}\,G$ level. Reviews of these results are given by Majorana [16] and Heyl [17].

More recently, a number of workers have tested the exactness of the inverse square law by searching for a variation in G with distance. While most of these experiments have produced null results, some maintain that the question is still open [18]. For a review of this topic, see Ref. [19].

In spite of all this careful work, one crucial test has not yet been attempted in the laboratory, and that is a test of the temporal constancy of G; and relatedly, of the temporal constancy of masses.

Why is it crucial to search for a nonzero $\dot{G}/G$? The answer, in part, is that the existence of such an effect would be our first indication that General Relativity is incomplete in its ability to totally describe the gravitational force (since its field equations do not allow for a time varying G). From a Machian point of view one might even expect, naively, that the observed recession of the galaxies might give rise to a weakening of the gravitational constant locally. Einstein did, in fact, attempt to incorporate Mach's principle into his theory, but failed to do so. It would seem then that this would be a particularly vulnerable point in the theory, but hopefully one subject to experimental clarification.

Further, Dirac noticed that certain comparisons of the Eddington numbers (e^2/Gm_em_p, m_ec^3/e^2H_0, etc.) called for G to vary inversely as the age of the epoch, and for the total number of particles (say baryons) to vary directly as the square of the epoch [20]. He called this his large numbers hypothesis (LNH). He later used Milne's hypothesis to develop a two metric theory in which long range mechanical phenomena are described by the "Einstein Metric" and short range, atomic phenomena are described by the "Atomic Metric." In so doing he was able to link his LNH, in which G varies when measured in atomic time, to standard General Relativity in which G is constant. Recently, Canuto *et al.* [21] have succeeded in formalizing these concepts into a scale covariant theory of gravity from which the LNH emerges as a gauge condition. Dirac has recently changed his views on the existence of matter creation [22], but this only makes the overall question of its existence more intriguing.

In summary, Sir Fred Hoyle has posed a question which brings all these issues into focus. Our interpretation of Mach's principle and our knowledge of the fundamental constants is inseparable from an overall understanding of the nature of the universe. In turn, our understanding of the universe evolves simultaneously with our assessments of ever more precise experiments. He asks therefore [23], "Does the universe affect the laws we discover in the terrestrial laboratory, or is the universe no more than an exercise in the application of present laws...?" This question repeats a point made earlier, in that it links the expansion of the universe to the local constancy of G, and also, for that matter, to the constancy of masses.

Most theories calling for a nonzero value of $\dot{G}/G$ predict, in fact, that $\dot{G}/G \simeq H_0 \simeq 5 \times 10^{-11}\,\text{year}^{-1}$ (see for instance Ref. [20]). Rates of matter creation are predicted by the same theories to be $\dot{m}/m \simeq 10^{-10}\,\text{year}^{-1}$, where H_0 here and above is the Hubble constant.

Clearly, measurements of effects this small will be difficult, and it was reasonable to expect that the first tests would be astronomical. Reasonburg *et al.* [24] have used radar ranging to the inner planets to obtain a value of $\dot{G}/G = 5 \pm 10 \times 10^{-11}\,\text{year}^{-1}$. Van Flandern [19] has timed lunar occultations with an atomic clock and reports a value of $\dot{G}/G = (-6.2 \pm 2.0) \times 10^{-11}\,\text{year}^{-1}$. Finally, J. G. Williams [25] has used lunar laser ranging to establish limits of $|\dot{G}/G| \leq 3.0 \times 10^{-11}\,\text{year}^{-1}$. (For a review of other articles reporting values of $\dot{G}/G \neq 0$, see Ref. [19].) The relatively large uncertainties in these measurements and the fact that some of these results have changed algebraic sign over the last decade indicates that no concensus on a final value of $\dot{G}/G$, as determined astronomically, can be expected for some time.

Some workers are, however, also starting to look at rotating bodies—as opposed to revolving bodies—as potential sources of data for evaluating $\dot{G}/G$ and $\dot{m}/m$.

For instance, Paul Muller estimated the fractional decrease in the spin rate of the earth, $\dot{\omega}/\omega$, by analyzing ancient and modern eclipse records. After subtracting off components of $\dot{\omega}/\omega$ due to lunar tidal interactions and various geophysical effects he found a residual change in the earth spin rate that he claimed was consistent with the value of $\dot{G}/G$ predicted by several cosmologies [26].

At the same time, Beams and Ritter [27] began developing a laboratory experiment for $\dot{G}/G$ which was basically an altered version of Beams' rotation-based measurement of the absolute value of G [28]. The isolation from gravity gradients provided by rotating the system at constant speed will most likely be essential to any earth bound measurement of $\dot{G}/G$. This and other details of that experiment are discussed below.

Ritter has pointed out that a freely spinning rotor with a speed decay time constant greater than or equal to the age of the universe could be used to test matter creation cosmologies. The principles of such an experiment can be understood by appealing to the conservation of angular momentum: If matter is being created, the rotor moment of inertia must change, but then its angular speed must change too in such a way that the rotor angular momentum remains constant. But such a change in angular speed could only be seen if the rotor were sufficiently smooth and free of friction. Symbolically, we have a free rotor

$$\omega = \omega_0 e^{-t/\tau^*}\,, \tag{1}$$

where ω is the rotor angular speed as a function of time, t, and τ^* is the angular speed decay time constant. Since most theories predict

$$\dot{m}/m = 2H_0 \simeq 10^{-10}\,\text{year}^{-1}\,, \tag{2}$$

we must have a rotor with at most

$$\frac{\dot{\omega}}{\omega} = -\frac{1}{\tau^*} \leqslant 10^{-10}\,\text{year}^{-1}\,, \tag{3}$$

or

$$\tau^* \geqslant 10^{10}\,\text{years} = 3 \times 10^{18}\,\text{s}\,. \tag{4}$$

The earth as a rotor was an obvious candidate for study since it has presumably been spinning since its creation. Important obstacles to employing the earth for this purpose are its large and nonconstant rotational energy dissipation and fluctuations in the length of its day.

The rotation of the earth has been studied for many years, and it has been found to have a broad range of speed fluctuations leading to length of day variations (l.o.d.). The three types of variations which have been measured are [29, 30]: The secular drift or spindown, the

medium range (diurnal to annual) variations, and the short term-irregular variations. The spindown results from the unavoidable dissipation of the earth's rotational kinetic energy via tidal mechanisms. The medium range variations arise from the Chandler wobble and other nontidal geophysical mechanisms. Finally, the irregular variations appear to be uncorrelated with the others, yet associated in an overall way with fluctuations in the angular momentum of the atmosphere. The irregular variations can change the l.o.d. by as much as 10^{-4} s [32]. The spindown of the earth gives rise to a speed drift of $\dot{\omega}_e/\omega_e = 25 \times 10^{-11}/\text{year}^{-1}$ [29]. This means the earth angular speed decay time constant is $\tau_E^* = 4 \times 10^9$ years which, coupled with the suspected large variation of this rate, would be much too small to measure $\dot{m}/m$ for the earth at the 10^{-10} level. In reality, the l.o.d. variations which average out to 10^{-8} day^{-1} from all sources [31] might put a limit on the usefulness of this technique even if τ_E^* were much larger. Clearly we have to look elsewhere for a rotor which is both less dissipative and smoother.

Ritter *et al.* [33] proposed a laboratory experiment in which a rotor would be spun in a "protected" mode in order to achieve a sufficiently long speed decay time constant. The protection consists of a vacuum chamber which would corotate around the rotor, superconducting magnetic shields, and other necessary features. Beams had built a double magnetic suspension where one rotor was freely (magnetically) suspended from inside a second, outer rotor itself freely suspended. This system showed promise of virtually eliminating the otherwise limiting rotor friction caused by the residual atmosphere [34]. Reducing the friction increases the decay time constant. This can be seen by examining the differential equation for a free rotor whose moment of inertia is I:

$$I\dot{\omega} + \kappa\omega = 0 \ . \tag{5}$$

Here κ is the rotor damping coefficient. The solution is Eq. (1) where

$$\tau^* = \frac{I}{\kappa} \ . \tag{6}$$

Both Beams [35] and Fremery [36] had achieved decay times of $\sim 10^{10}$ s with small (1 mm) steel spheres spinning at high speeds (20 kHz) in vacuum. It became clear that if the rotor moment of inertia could be increased by 10^5 and the friction decreased by 10^3 or more (with the corotation scheme) that decay time constants of 10^{18} s could be achieved and the $\dot{m}/m$ predictions tested at that level.

The goals of the research reported here were to bring the room temperature version of the matter creation experiment into existence and to demonstrate the effectiveness of the corotation scheme in reducing the rotor energy dissipation. In the process of accomplishing these goals many of the techniques which will be necessary for successful completion of the $\dot{G}$ experiment have been developed and some of them carefully refined. This is a significant step forward in the creation of the more difficult $\dot{G}$ experiment.

2. Instrumentation

Since much of the instrumentation is common to both of these experiments, and since a $\dot{G}$ experiment has been thoroughly described before [27], only a thorough description of that used in the $\dot{m}$ experiment is given here. In principle, the inner and outer rotors of the $\dot{G}$ experiment are designed to maximize the gravitational attraction between them, while the inner and outer rotors of the $\dot{m}$ experiment are designed to minimize the attraction. In both experiments, all other forms of inter-rotor couplings including electrostatic, magnetic, mechanical and gaseous couplings (among others) must be minimized, unless they are to be used as part of a feedback system.

2.1 Instrumentation for the $\dot{m}$ Experiment

In its present form, the experiment has required the development and subsequent incorporation of five separate systems: A stable magnetic suspension for the inner rotor, a highly constant rotary drive for the outer rotor, a sensitive rotational period measurement system for both rotors, a vacuum system to pump out the outer rotor while it rotates, and a vibration isolation system which can support most of the above. For specific details of the construction, many of which are necessarily omitted below, consult Ref. [37].

The design of the magnetic suspension system was dictated by the properties of the rotor to be suspended. In this case a 250 g disk of glass ceramic (Zerodur) was chosen for its thermal and temporal stability. The disk is coaxially connected to a 16 cm by 1 cm diameter Zerodur rod at the other end of which the ferromagnetic support sphere is fastened. The support sphere has only a few percent of the total rotor moment of inertia to minimize eddy current losses. The 16 cm rod keeps the disk out of the strongest part of the support field thereby minimizing Faraday disk effects. A mirror used for rotational period measurement and counterweights are fixed to the disk. The total moment of inertia of this assembly is 1100 g-cm^2.

The position sensor for the suspension is a pulsed light beam which is parametrically sensed by a split cathode photodiode and the ensuing phase locked loop. A 200 W amplifier drives the support coil at the output end of the feedback loop, but a permanent magnet is used to provide most of the lifting force for the rotor in order to reduce ohmic losses in the coil. There is sufficient common mode rejection in this system to permit stable operation of the suspension while overhead room lights are on, and temperature drift of the photodetector is no longer problematic. We have kept rotors in suspension continuously for periods up to 0.1 year and could do so indefinitely if necessary. Detailed discussion of this and other magnetic suspensions developed in our laboratory is given in Refs. [38] and [39]. It is important to note that the pole piece used in the support coil does not rotate in the present system like it did in Ref. [37].

The need for an almost totally constant speed drive for the outer rotor is twofold: (1) as discussed above, in order to measure $\dot{m}/m$ to 10^{-10}/year, the decay time of the inner rotor must be $\simeq 10^{10}$ years. Any speed drift in the outer rotor at this level or above will cause the relative positions of the inner and outer rotors to shift. Such shifts would change the size of the restoring force acting on the inner rotor in such a way that the $\dot{m}$ signal might be masked. (2) Large short term speed fluctuations, if rectified by measurement system nonlinearities, might also introduce false $\dot{m}$ signals.

To overcome these problems, a new type of rotary drive system had to be developed which could ensure the long term accuracy ultimately needed, yet simultaneously deliver driving torque to the load without generating appreciable speed fluctuations. After several attempts to develop such a drive, J. W. Beams proposed the design presently in use called the magnetic averaging synchronous motor. It is a 50 pole synchronous motor whose armature is a thin strip of slightly retentive stainless steel.

This strip is epoxied around the periphery of a circular aluminum disk that is the motor rotor.

Temporary poles are established in the armature strip by flashing the drive coils at dc, and the motor is then spun by hand and allowed to coast until the free period is within ≈1% of the desired driven period. At that point, the drive circuitry is started, and within a few cycles, the motor locks in either on the fundamental or (if desired) one of the odd subharmonics of the electrical driving signal. Each phase of this drive signal partially rewrites the original induced poles, and eventually the armature pole strengths and their respective positions around the armature strip assume highly uniform values. This averaging effect smooths the drive. While the difficult subrotation measurements of the motor speed fluctuations have not yet been completed, it is known that on a period to period basis the fluctuations are driving-waveform dependent, but can be made as small as 2 ppm or less. Additional measurements have shown that these fluctuations average down by a factor of ten for every decade increase in averaging time, and it is very likely that this process will only be limited by drift in the frequency standard used as the time base for this motor (currently a cesium beam clock with $\Delta\nu/\nu \simeq 10^{-13}$).

The speed range of this motor is 0.1 to 10 rad/s, and the electrical drive signal is produced by a commercially available frequency synthesizer (Hewlett-Packard 3325 A) synchronized to the atomic clock.

The motor drives directly a smooth air bearing whose undriven spindown quality factor is ≈3000.

A 30 L/s ion pump is attached to the lower part of the motor bearing assembly. It is used to evacuate the 30 cm long by 10 cm diameter glass outer cylinder which is mounted on top of the bearing. They are connected by a 30 cm pumping sleeve which passes through the center of the motor bearing assembly, along its rotational axis. The pump is powered by two 12 V rechargeable batteries whose voltage is stepped up to ≈6000 V by a dc-dc converter. The batteries, converter, and associated metering are all attached to the pump and the entire system is free to rotate. A thin walled, cylindrical aluminum shroud is placed around the entire assembly and also corotates with it. The shroud streamlines the load driven by the motor and therefore reduces the size of the load fluctuations. This is important since the rotary drive system is both softly coupled and (presently) nonfedback, and large load fluctuations would result in unacceptable speed variations.

Operating pressure in the outer cylinder is typically limited to $\sim 10^{-6}$ mm Hg by the small diameter (≈3 cm) of the long pumping sleeve. While this limits the battery lifetime to about 4 days of corotation operation, the chamber pressure during this time is well below the molecular flow limit. This reduces the chances that collective gas motion, as might be expected in the viscous regime, will limit the inner cylinder spindown decay time constant.

The measurements of both the inner and outer cylinder periods were made with separate Jones recording optical levers [40] modified by Towler [41] and later by Gillies [37] for use as timers rather than angular displacement transducers. The levers themselves are accurate to at least 0.1 μs, and their output pulses are used to trigger digital counters and more recently a stand alone computer system used for making the actual timing measurements.

The magnetic suspension system, motor/bearing plus vacuum system load, and the Jones levers are all vibration isolated on a two stage passive system of our own design. An especially quiet room in the basement of the University of Virginia (U. Va.) Physics building was located and renovated for this work. The floor was sawed to separate the working surface from the rest of the building, and a 4000 kg stone was placed on Weber isolators in the center of that area. On this, the 250 kg surface plate which supports the experiment was set. It rests on pneumatically interconnected air bladders which isolate against floor noise to ≈3 Hz with a damping time of one second. Floor noise is unmeasurable (below 0.02 μm) at this station. An active isolation system sensed and driven by piezocrystals is currently under development and will be installed when necessary.

All control and sensing electronics are in an adjacent room, and the laboratory room containing the experiment is temperature controlled to within 1 °C by a dedicated air conditioner. Because a 1 Hz pendulum mode is easily excited in the rotor while in suspension, no one is allowed in the laboratory room while an experiment is in progress.

2.2 Instrumentation for the $\dot{G}$ Experiment

While this experiment is still in the design phase, one detailed proposal for construction has been formulated and presented in Refs. [19], [27], and [42] by workers at the U. Va.

In principle, the experiment is a rotating Cavendish balance. The angular position between the large and small masses is held constant by a servo system. Barring all other torquing mechanisms, a change in the gravitational interaction between the large and small masses would appear as a change in the restoring torque necessary to keep the system balanced.

The experiment, as mentioned above, is rotated to keep the effects of anomalous gravity gradients on the balance very small.

Such an experiment would have to be cooled to liquid helium temperatures in order to keep thermal noise effects small, and synchronous detection techniques would be necessary to remove the effects of $1/f$ noise.

The small mass system would most likely be magnetically suspended in order to maximize sensitivity, and the rotational angular speed of the whole system would have to be better than $\Delta\omega/\omega \simeq 10^{-12}$.

Subfringe laser interferometry or SQUID position sensing show promise for use in angle sensing, but the overall complexity of such an apparatus would make its construction both expensive and somewhat inelegant. Further experience at the National Bureau of Standards [43] with a rotating torsion balance has shown the practical difficulty of this combination, albeit with a fiber as the suspension method.

Two other schemes for measuring $\dot{G}$ were conceived at U. Va., but have not yet been studied in detail [44]. One would employ two superconducting spheres as masses on opposite ends of a "superconducting spring." This high Q system would be driven and observed to see if the strength of the drive signal varied presumably as a result of a change in G. A second system would synchronously detect the response of a weakly coupled sphere floating in a soft "ball trap" to a driving mass that rotated near the trapped sphere and gravitationally induced some motion in it. In order for either of these systems to be workable, it would be necessary to put them in drag-free artificial satellites.

3. Results and Discussion

A number of spindown and spinup tests were run in order to study the intercylinder gas coupling. Rotor speeds were typically 20 radians per second or less during spindown, and decay time constants of $\simeq 10^6$ s were observed as expected at the levels of vacuum used. The spinup tests were mostly done at atmospheric pressure, and the rotor was accelerated by the viscous coupling and asymptotically approached the angular speed of the driven outer cylinder. At relatively high angular speeds, 4 rad/s and above, interrotor mode locking was observed. Specifically, the inner rotor was observed to be smoothly spinning up from rest and then suddenly it would lock in on the angular frequency of rotation of the outer cylinder (within just a few cycles). While it is suspected that asymmetric flow in the interrotor atmosphere may have caused this, the question has not been important enough to pursue to a conclusion. Details of the experiments and samples of the data are available in Ref. [37].

The first corotation runs with the inner cylinder in vacuum were troubled by the rotation of the magnetic suspension support coil pole piece which was epoxied into the center of the upper vacuum plate covering the outer cylinder. The residual vertical and lateral misalignment in the pole piece put an unexpectedly large torque on the inner cylinder and caused it to almost exactly follow the motions of the outer cylinder. The strength of this "cranking effect" made it clear that it would be essential to have the pole piece position fixed in the laboratory frame.

The suspension was redesigned to permit this, and the data immediately reflected the improvement. For the first time, true corotation runs were conducted with an inner cylinder now free enough to be of real use in studying the now corotating atmosphere's effects on the total rotor damping.

The first runs showed that when the motor angular speed was within $\simeq 0.5\%$ of the rotor angular speed, the rotor would be caught in the potential formed by whatever interrotor coupling existed. As the rotors corotated, the inner one would oscillate about its preferred position with an amplitude of about 30 degrees or less and periods ranging from 10^3 s to 3×10^3 s. Ten period averages of the data showed that the oscillations were sometimes purely sinusoidal but often irregular and resembling a sinusoid with 3rd harmonic distortion. Attempts to investigate the dependence of the oscillation waveform on initial conditions of oscillation are in progress. The nature of the coupling itself is also being studied. Dipole moments of known amount are applied to the outer cylinder via a battery and cylindrical capacitor plates and the subsequent effects on the rotor motion are observed. A precision electrometer sensitive to 60 electrons/s of current flow is also being used to study the rotor charge distribution both during and after the startup of the ion pump. It is anticipated that coating of the rotor surfaces with a conducting film will be necessary in order to reduce electrostatic coupling effects.

Data covering a twenty-four hour period showing the rotor undergoing a decaying oscillation while rotating is displayed in Fig. 1. The most notable feature of this data is, in fact, the decay envelope bounding the data. It represents a power loss of $\simeq 9 \times 10^{-15}$ W, a remarkably small amount of energy given the fact that the rotor has 0.3 mJ of rotational kinetic energy. Given that gas drag is relative-velocity dependent, this result implies considerable effectiveness for the corotation protection method.

Figure 2 shows an equally interesting phenomenon.

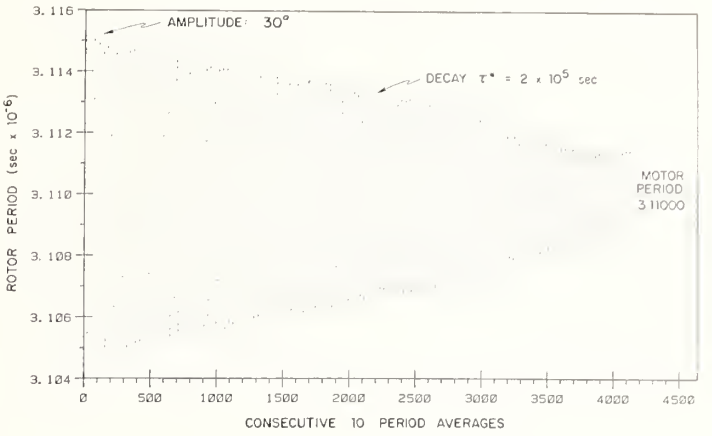


FIGURE 1. *Oscillations of inner rotor about synchronous position. The outer cylinder rotates at a constant rate with a period of 3.11000 s. The roughness of the decay envelope is due to the spacing of the points from the 10 period averaging.*

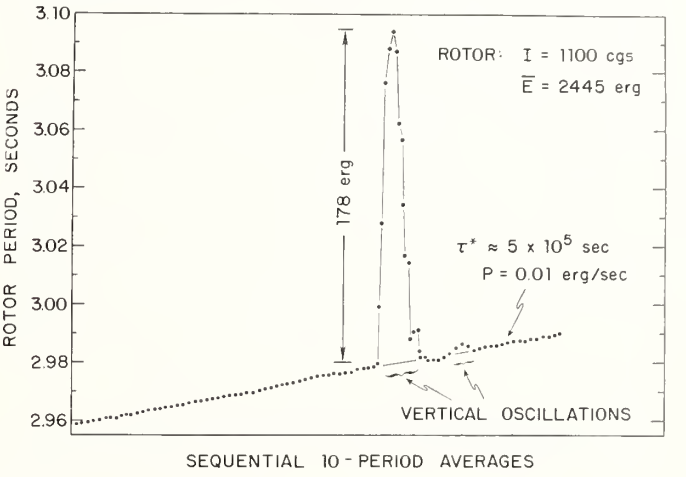


FIGURE 2. *Anomalous rotor slowing in non-corotation mode. The return to the extended decay baseline shows the conservative nature of the temporary mode.*

With the outer cylinder still, the inner rotor was allowed to undergo spindown. At a period length of $\simeq 2.980$ s, it began to oscillate vertically, as if the suspension became marginally unstable. The rotor suddenly lost rotational kinetic energy, but as the oscillation stopped, it seemed to regain all it had lost to the vibratory mode when its period reached $\simeq 2.985$ s; it returned to its original decay slope as if nothing happened. This mysterious activity might result from some type of rotational/vertical mode coupling, although it is difficult to understand the conservative nature of it in light of the electronic damping provided by the magnetic suspension feedback circuitry.

Inevitably, the rotor decay time when suspended by active magnetic suspensions will be eddy current limited. And, it is difficult to believe that vertical-rotational mode coupling can be made sufficiently low that an active suspension can be tolerated. For this reason it will probably be essential to do the final version of the *in* experiment with a passive, lossless superconducting bearing, such as that built by Harding at the Jet Propulsion Laboratory [45]. Such a system, in simplified form, is currently under construction by Brown at the State University of New York at Binghamton [46].

No other problem seems to be so critical to the final outcome of the experiment as that of bearing design. Speed fluctuations in the motor, while still only partially

characterized, might be corrected by suitable feedback, providing that the long term accuracy is not lost in the process. Higher levels of vacuum can be achieved by using an air bearing (or, perhaps, a magnetic bearing) with a wider throat—this is a hardware problem only. The nature of the couplings between the rotor and its environment will have to be studied individually, and judicious application of superconducting shielding and other protection mechanisms will be necessary. Ultimately, spherical rotors may be necessary, since they can be built freer of moment of inertia asymmetries than either cylindrical rotors or rotors of a hybrid geometry. Finally, if necessary, the experiment could be fedback and the inner rotor served to a null relative position with the outer rotor at the bottom of their potential well in order to obtain an unambiguous $\dot{m}$ signal.

4. Conclusions

The variation of G with many factors has been laboratory tested at various levels of precision in the past and no nonconstancy has been found. Now its variation with time and the related matter creation question are becoming of greater interest to gravitation and cosmology. Astronomical tests have become inconclusive and will probably always be challengeable under the "liberalization" of physical theory needed to encompass time varying G cosmologies. Laboratory tests are controllable and can assault these challenges.

Laboratory tests of matter creation are underway, and $\dot{G}$ tests are being planned. These rely on the prospect that laboratory rotors can be made much smoother and of longer decay time than the earth in its rotations. These efforts to overcome the classical barriers of bearing and gas-drag frictions use new methods, including magnetic suspension and corotation of the gas atmosphere.

Beams and Fremerey obtained rotational decay times of 10^{10} s for tiny steel spheres; and if the rotor moment of inertia could be increased by 10^5 while simultaneously decreasing the bearing and gas drag by 10^3, a cosmologically significant decay time of 10^{18} s would result.

While it is relatively easy, although not trivial, to scale up the rotor moment of inertia, it is a more difficult problem to reduce the bearing and gas friction. In fact, ferromagnetic suspension bearings seem, at this time, to have insurmountable limits which may dictate the use of the more nearly lossless superconducting suspension systems. Corotation of the gas, in principle, should eliminate it as a source of drag. Since this is accomplished by corotating an outer cylinder around the inner cylinder, however, new problems such as interrotor couplings are introduced. The electrical aspects of such couplings are only now beginning to be investigated, but other studies discussed have yielded interesting information. Measurements of the interrotor oscillation amplitude show energy losses in this mode of only 10^{-14} W. When studied in conjunction with the rotor secular decay rate, we can conclude that corotation can be effective in reducing the gas drag component of the overall rotor damping.

A host of other technology has been developed and much of it will be applicable to other areas of fundamental measurements as well.

We gratefully acknowledge support by the U.S. National Bureau of Standards under Precision Measurement Grant G8–9025, the National Science Foundation under grants PHY78–3208 and PHY80–07948, and the University of Virginia. We also thank George Jones for his work on the experiment and for taking the data presented here.

References

[1] A. S. MacKenzie, *Memoirs on the Laws of Gravitation* (American Book Company, New York, 1900).
[2] M. U. Sagitov, *The Gravitational Constant and the Mass of the Earth* (Nauka, Moscow, 1969).
[3] M. V. Cremieu, C. R. Acad. Sci. **149**, 700 (1909).
[4] R. V. Eötvös, D. Pekar, and E. Fekete, Ann. Phys. (Leipzig) 68, 11 (1922).
[5] P. G. Roll, R. Krotkov, and R. Dicke, Ann. Phys. (N.Y.) 26, 442 (1964).
[6] V. B. Braginsky and V. I. Panov, Sov. Phys. JETP **34**, 463 (1972).
[7] A. S. MacKenzie, Phys. Rev. **2**, 321 (1895).
[8] J. H. Poynting and P. H. Gray, Phil. Trans. R. Soc. London, Ser. A: **192**, 245 (1899).
[9] P. R. Heyl, Sci. Pap. Natl. Bur. Stand. **19**, 307 (1924).
[10] P. R. Heyl, Sci. Mon. 47, 114 (1938).
[11] P. E. Shaw and N. Davy, Proc. R. Soc. London **102**, 46 (1922).
[12] L. W. Austin and C. B. Thwing, Phys. Rev. 5, 294 (1897).
[13] Q. Majorana, Atti della Reale Accademia dei Lincei **28** (1919).
[14] H. Russel, Astrophys. J. 54, 334 (1921).
[15] Q. Majorana, C. R. Acad. Sci. 173, 478 (1921).
[16] Q. Majorana, Philos. Mag. 39, 488 (1920).
[17] P. R. Heyl, Science 55, 349 (1922).
[18] D. R. Long, Nuovo Cimento B **62**, 130 (1981).
[19] R. C. Ritter, in *The Proceedings of the Second Marcel Grossman Conference on General Relativity*, Ed. by R. Ruffini (North Holland, Amsterdam, in press).
[20] P. A. M. Dirac, Nature (London) **139**, 323 (1937).
[21] V. Canuto *et al.*, Phys. Rev. D **16**, 1643 (1977).
[22] P. A. M. Dirac, Proc. R. Soc. London, Ser. A: 365, 19 (1979).
[23] F. Hoyle, *From Stonehenge to Modern Cosmology* (W. H. Freeman, San Francisco, 1972).
[24] R. D. Reasonburg, in *Proceedings of the Second Marcel Grossman Conference on General Relativity*, Ed. by R. Ruffini (North Holland, Amsterdam, in press).
[25] J. G. Williams, W. S. Sinclair, and C. F. Yoder, Geophys. Res. Lett. 5, 943 (1978).
[26] P. M. Muller, in *Proceedings of the Workshop Meeting on the Measurement of Cosmological Variations of the Gravitational Constant*, Ed. by L. Halpern (University of Florida, Gainesville, 1978).
[27] R. C. Ritter and J. W. Beams, in *Proceedings of the Workshop Meeting on the Measurement of Cosmological Variations of the Gravitational Constant*, Ed. by L. Halpern (University of Florida, Gainesville, 1978).
[28] R. D. Rose *et al.*, Phys. Rev. Lett. 23, 655 (1969).
[29] R. H. Dicke, in *The Earth-Moon System*, Ed. by B. G. Marsden and A. G. W. Cameron (Plenum Press, New York, 1966).
[30] W. H. Munk and G. J. F. MacDonald, *The Rotation of the Earth*, (Cambridge University Press, London, 1975).
[31] G. P. Pil'nik, Sov. Astron. 20, 608 (1976).
[32] R. Hide *et al.*, Nature (London) **286**, 114 (1980).
[33] R. C. Ritter *et al.*, Nature (London) 271, 228 (1978).
[34] J. W. Beams, Rev. Sci. Instrum. 34, 1071 (1963).
[35] J. W. Beams *et al.*, Rev. Sci. Instrum. 33, 151 (1962).
[36] J. K. Fremerey, Phys. Rev. Lett. 30, 753 (1973).
[37] G. T. Gillies, Ph. D. Dissertation, U. Va. (University Microfilms, Ann Arbor, 1980).
[38] W. S. Cheung *et al.*, Precis. Eng. 2, 183 (1980).
[39] G. R. Jones, Jr., M. Sc. Thesis, University of Virginia, 1980, unpublished.
[40] R. V. Jones and J. C. S. Richards, J. Sci. Instrum. 36, 90 (1959).
[41] W. R. Towler, Thesis, University of Virginia, 1967, unpublished.
[42] R. C. Ritter *et al.*, in *Atomic Masses and Fundamental Constants 5*, Ed. by J. H. Sanders and A. H. Wapstra (Plenum Press, New York, 1976), p. 629.
[43] G. Luther, private communication, 1981.
[44] R. C. Ritter and G. T. Gillies, unpublished, 1980.
[45] J. T. Harding, Technical Report Number 32–897 (Jet Propulsion Laboratory, Pasadena, California, 1966).
[46] Y. J. Brown, private communication, 1981.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Interpreting Dirac's Large Numbers Hypothesis

W. Davidson

University of Otago, Dunedin, New Zealand

Dirac's large numbers hypothesis (LNH) and its implications are reviewed. Recent work shows that if the LNH is correct then not only does the Newtonian gravitational constant decrease on an atomic time scale ($G \propto t^{-1}$), as inferred by Dirac, but the electrical force between two charged particles becomes stronger relative to their mutual gravitational force regardless of which of Dirac's two space-time scales, atomic or gravitational, is used. This is made explicit by introducing a Coulomb constant γ analogous to G. The derivation of the two metrics directly from the LNH is demonstrated. Some observational implications of the LNH are listed.

Key words: Dirac cosmology; gravitational and electromagnetic constants; metric reciprocity; observations.

1. Introduction

As is well known, Dirac has suggested [1, 2] that the large dimensionless numbers of physics and astronomy are related by equations in which the coefficients are of order unity. This conjecture will be referred to as the "large numbers hypothesis" (LNH). Three such numbers are given special consideration by Dirac [1]. One is the ratio of the electric to the gravitational force between the electron and proton: $e^2/Gm_e\, m_p$. Another is the ratio of the Hubble time T of the expanding universe to a characteristic atomic time: $T/(e^2/m_e\, c^3)$. The third is: M/m_p, the number of baryons in the observable universe. Dirac finds that the first two numbers are approximately equal and the third is roughly the square of this.

McCrea and I have shown [3] that if the LNH is assumed to be true then not only does the Newtonian gravitational constant decrease on an atomic time scale, as inferred by Dirac [1], but the electrical force between two charged particles becomes stronger relative to their mutual gravitational force, regardless of which of Dirac's two space-time scales, atomic or gravitational, is used. For this purpose we introduced a Coulomb constant γ, to play the same role for the electrical force that G does for the gravitational force. Thus if we write appropriately

$$X = \frac{\gamma e^2}{G\, m_e\, m_p}, \tag{1}$$

$$Y = \frac{T}{\gamma e^2/m_e\, c^3}, \tag{2}$$

$$N = \frac{M}{m_p}, \tag{3}$$

then very roughly

$$X \sim Y \sim 10^{40}, \quad N \sim (10^{40})^2. \tag{4}$$

It will be appropriate to write $Y \sim t$, since Y is a measure of T in units of atomic time, and should vary with the age of the universe. In view of Eq. (4) the LNH then requires that

$$X \sim Y \sim t, \tag{5}$$

$$N \sim t^2. \tag{6}$$

It is then with remarkable self-consistency that we may derive from Eqs. (5) and (6) the two space-time metrics, atomic A and gravitational E, which have been suggested by Dirac as providing separate scales for atomic and gravitational phenomena [1]. This derivation direct from the LNH will be given in the next Section.

2. Derivation of Dirac's *A* and *E* Metrics from the LNH

It will be assumed, as in [1], that atomic masses and charges are constant regardless of the space-time scale used. In the present treatment this means that any time variations required by the LNH are to be borne by G and γ. We first seek a space-time metric A in which the result $Y \sim t$ of Eq. (5) stems from the assumption that

$$T_A \propto t \tag{7}$$

while γ is constant. Suppose the metric has the Robertson-Walker form

$$ds_A^2 = c^2 dt^2 - R_A^2(t)\,(dr^2 + r^2 d\Theta^2 + r^2 \sin^2\Theta\, d\phi^2) \tag{8}$$

in which, following Dirac's interpretation of the LNH [1], the spatial curvature is assumed to be zero. Then as in Ref. [3] the result Eq. (6) leads directly to

$$R_A(t) = \left[\frac{t}{t_o}\right]^{1/3} \tag{9}$$

where t_o is an arbitrary constant. The result $X \sim t$ of Eq. (5) then requires that in this metric G behaves as

$$G_A \propto t^{-1} \tag{10}$$

since it is inherently assumed that

$$\gamma_A = \text{constant}. \tag{11}$$

Suppose alternatively, that the relation $X \sim t$ of Eq. (5) is realized because γ varies while G is constant. We therefore seek a metric E having the properties that, on the E scale,

$$G_E = \text{constant}, \tag{12}$$

$$\gamma_E \propto t. \tag{13}$$

Let the metric be of the form

$$ds_E^2 = c^2 d\tau^2 - R_E^2(\tau)\,(dr^2 + r^2 d\Theta^2 + r^2 \sin^2\Theta\, d\phi^2)\,, \quad (14)$$

giving the matter the same comoving coordinates r, Θ, ϕ as in the A metric. Let r_{horiz} be the limiting r value of observed matter (seen by light leaving the matter at $\tau = 0$ and arriving at the observer at the present epoch τ_o). Then

$$r_{\text{horiz}} = c \int_0^{\tau_o} \frac{d\tau}{R_E(\tau)} = F(\tau_o)\,. \quad (15)$$

The number N of baryons observed up to the horizon at epoch τ_o is proportional to r_{horiz}^3. Hence by Eqs. (6) and (15) we obtain for general τ

$$F(\tau) \propto t^{2/3}$$

and so, on differentiating,

$$\frac{1}{R_E(\tau)} \propto t^{-1/3} \frac{dt}{d\tau}\,. \quad (16)$$

It should now be noted that the relation $Y \sim t$ of Eq. (5) combined with Eq. (13) means that

$$T_E \propto t^2\,. \quad (17)$$

But according to Eq. (14)

$$T_E = \frac{R_E(\tau)}{dR_E(\tau)/d\tau}\,. \quad (18)$$

Equations (16), (17) and (18) then give

$$R_E(\tau) = \left[\frac{\tau}{\tau_o}\right]^{2/3} \quad (19)$$

as well as $\tau \propto t^2$, and hence we write

$$\tau = \frac{t^2}{2t_o}\,. \quad (20)$$

We choose t_o in t time to correspond to τ_o in τ time. Hence

$$\tau_o = \frac{1}{2} t_o\,. \quad (21)$$

The A and E metrics are therefore

$$ds_A^2 = c^2 dt^2 - \left[\frac{t}{t_o}\right]^{2/3} (dr^2 + r^2 d\Theta^2 + r^2 \sin^2\Theta\, d\phi^2)\,, \quad (22)$$

$$ds_E^2 = c^2 d\tau^2 - \left[\frac{\tau}{\tau_o}\right]^{4/3} (dr^2 + r^2 d\Theta^2 + r^2 \sin^2\Theta\, d\phi^2) \quad (23)$$

which evidently stand in the conformal relation

$$ds_E^2 = \left[\frac{t}{t_o}\right]^2 ds_A^2\,. \quad (24)$$

In the A metric Eqs. (10) and (11) hold; in the E metric Eqs. (12) and (13) hold, and in particular $\gamma_E \propto \tau^{1/2}$ by Eq. (20). There is therefore a simple reciprocity between the metrics with regard to electromagnetic and gravitational phenomena. But evidently the electrical force between two charged particles becomes stronger relative to their mutual gravitational force, regardless of which of Dirac's two space-time scales is used [3].

If the LNH were the correct explanation of the large number coincidences, rather than say an anthropic principle [4], then according to the above interpretation electromagnetic and gravitational phenomena can be described in terms of conventional theory only if the appropriate space-time scales, A and E respectively, are used. It is therefore of the greatest importance to test such a radical proposition observationally.

3. Observational Interpretation of Dirac Cosmology

3.1 Observational Metric in Dirac Cosmology

The most refined methods of observation in astronomy and cosmology are undoubtedly those based on atomic and electromagnetic principles. For example, within the solar system the best measure of distance is now made by radar methods and atomic clocks. The distance to the nearer stars is determined by Euclidean geometry and the method of parallax. The distance to the nearer galaxies is the "luminosity distance" of their identifiable variable stars whose absolute magnitude is of known calibration; alternatively the brightest stars in galaxies of statistically known intrinsic brightness are used. The average absolute magnitude of, say, elliptical galaxies is then known and so the Hubble law between luminosity distance and redshift is established yielding the present value of Hubble's constant and its reciprocal T. All these criteria are based upon atomic or electromagnetic phenomena. Therefore the theoretical basis of observations in Dirac cosmology should presumably be the A metric Eq. (22). An exception to this would have to be anticipated where mass values have to be assigned using the gravitational properties of mass, and then the E metric would have to involved.

In particular, the Hubble time derived from the apparent magnitude-redshift data must be identified as T_A in Dirac cosmology. On the other hand the metric E given in Eq. (23) provides a theoretical Hubble time T_E. Since Einstein's gravitational field equations hold in this (Einstein-de Sitter) metric, it is T_E which must be inserted into the Einstein equation relating the present mass density to the present Hubble constant.

A summary of predictions, observable in principle, for Dirac cosmology is given below. Full details will be published elsewhere. Where appropriate, comparison is made with the predictions of general relativity (GR) for the case $\Omega_o = 1$ (when the present mass density is equal to the so-called critical density $3/8\,\pi G T^2$).

3.2 Predicted Features of Dirac Cosmology

(a) In Dirac cosmology the atomic age of the universe is only 1/3 of the observed Hubble time [1], compared with $2/3 \times T$ in GR ($\Omega_o = 1$). If the theory is correct then the estimated ages of globular star clusters [5, 6] are too high.

(b) In Dirac cosmology the deceleration parameter obtained by the apparent magnitude-redshift test should give $(q_o)_A = 2$; by the cosmic mass density method it should give $(q_o)_E = 0.5$. In GR ($\Omega_o = 1$) the prediction is $q_o = 0.5$ by both methods.

(c) The *present* cosmic mass density in Dirac cosmology is the same on scales A and E and is equal to $3/8\,\pi G_E T_E^2$ where $T_E = T_A/4$, T_A being the *observed* Hubble time. For a given observed value of the Hubble time the present mass density predicted for Dirac cosmology is therefore 16 times the critical density in GR. This may be a powerful method of ruling out Dirac cosmology if the so-called "missing mass" is not found. On the other hand the cluster evidence [7] for missing mass may support the Dirac theory.

(d) The epochs in atomic time for the cross-over from radiation dominance to matter dominance, recombination

of hydrogen and formation of galaxies are all much earlier and at higher material densities in Dirac cosmology compared with canonical GR. These circumstances together with the result that $G \propto t^{-1}$ on the A scale seem to assist the formation of galaxies in Dirac cosmology (see also Ref. [8]).

(e) The total mass density (after the cross-over) varies as t^{-1} on the A scale compared with τ^{-2} in canonical GR ($\Omega_o = 1$). The radiation density varies as $t^{-4/3}$ compared with $\tau^{-8/3}$ in GR.

(f) The redshift z in Dirac cosmology varies with atomic time such that $1 + z \propto t^{-1/3}$, in contrast with $1 + z \propto \tau^{-2/3}$ in canonical GR ($\Omega_o = 1$). Hence all cosmic events specified by redshift or temperature occur at a much earlier fractional age of the universe in Dirac cosmology compared with GR.

(g) Conventional Maxwellian electromagnetic theory can be applied consistently in A-scale Dirac cosmology, whereas on the E gauge electromagnetic quantities behave quite differently. In particular Planck's constant varies with epoch such that $h_E \propto \tau^{1/2}$, in addition to the variation of the Coulomb constant $\gamma(\propto \tau^{1/2})$. The latter relation means that the electrical space permittivity $\epsilon_o \propto \tau^{-1/2}$ and the magnetic permeability $\mu_o \propto \tau^{1/2}$. In addition the adiabatic relation for radiation density, viz. $\rho_r R^4 =$ constant, does not hold on the E scale. Therefore if the LNH is correct the results of combining Einstein's equations with the conventional mechanics of electromagnetic theory would be invalid. The canonical GR big-bang theory would therefore fail.

References

[1] P. A. M. Dirac, Proc. R. Soc. London Ser. A: 365, 19 (1979).
[2] P. A. M. Dirac, Proc. R. Soc. London Ser. A: 165, 199 (1938).
[3] W. Davidson and W. H. McCrea, Proc. R. Soc. London Ser. A: 374, 447 (1981).
[4] B. Carter, in *Confrontation of Cosmological Theories and Data: Proceedings*, Ed. by M. S. Longair (Kluwer, Boston, 1974), p. 291.
[5] M. Schwarzschild, Q. J. R. Astron. Soc. 11, 12 (1970).
[6] P. Demarque and R. D. McClure, in *The Evolution of Galaxies and Stellar Populations*, Ed. by B. M. Tinsley and R. B. Larson (Yale Univ. Obs., 1977), p. 199.
[7] G. O. Abell, in Ann. Rev. Astron. Astrophys. 3, Ed. by L. Goldberg (Annual Reviews Inc., Palo Alto, 1965), p. 1.
[8] R. H. Dicke, Rev. Mod. Phys. 34, 110 (1962).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

New Laboratory Test of the Equivalence Principle

P. T. Keyser, J. E. Faller,* and K. H. McLagan

Joint Institute for Laboratory Astrophysics, National Bureau of Standards and University of Colorado, Boulder, CO 80309

A test of the principle of equivalence using a large fluid (surrogate) fiber Eötvös apparatus is presently being undertaken at the Joint Institute for Laboratory Astrophysics. Preliminary measurements using a 0.25 m diameter fluid system were sufficiently encouraging that we have embarked on the construction of a five times larger (1.27 m diameter) system employing approximately 500 kg of lead and 500 kg of copper as the test masses. The first experimental results on the equivalence of gravitational and inertial mass with this new large apparatus are expected in 1983 or 1984.

Key words: equivalence principle; general relativity; gravitation.

1. Introduction

The ratio of the inertial mass of an object to its passive gravitational mass is postulated to be independent of the material in metric theories of gravitation such as general relativity [1]. In fact, in the last century, the equivalence of inertial mass (m) and gravitational mass (M) for two materials (A and B) as measured by the parameter η, where

$$\eta(\mathrm{A},\ \mathrm{B}) \equiv \frac{\left[\left(\frac{M}{m}\right)_{\mathrm{A}} - \left(\frac{M}{m}\right)_{\mathrm{B}}\right]}{\frac{1}{2}\left[\left(\frac{M}{m}\right)_{\mathrm{A}} + \left(\frac{M}{m}\right)_{\mathrm{B}}\right]}, \qquad (1)$$

was known to 2×10^{-5} [2], and was measured to be less than 3×10^{-9} by 1922 [3]. Two more recent measurements have lowered the upper limit on η to 3×10^{-11} and 0.9×10^{-12} respectively [4, 5]. Nevertheless, a fundamental motivation exists for checking this hypothesis to the highest possible accuracy: the principle of equivalence is basic to theories of gravity.

At the Joint Institute for Laboratory Astrophysics (JILA), we have developed a new type of torsion-pendulum apparatus for the purpose of improving the accuracy of the Eötvös experiment, which tests the equivalence of gravitational and inertial mass. All of the experiments mentioned above were done with a traditional torsion balance; but, despite a significant amount of effort, none of these experiments reached the sensitivity that would be expected if they were limited by the Brownian motion of the torsion balance. In all cases, other sources of noise dominated the fundamental thermal noise. The torsion fiber itself is a source of noise, while the classical torsion balance is extremely sensitive to seismic noise. Also, the sensitivity of a torsion balance is limited since the fiber both supports and torques the masses. Any attempt to increase the sensitivity of the balance by increasing the size of the masses also requires an increase in the diameter of the fiber, and this then places more severe requirements on the position detector since the torsional stiffness of a fiber increases as the fourth power of its diameter.

The approach we have taken at JILA to deal with these limitations of the traditional fiber involves the utilization of a new type of (surrogate) fiber, in which the functions of support and torque are separated. We do this by using a fluid (water at 3.98 °C, its maximum density point) to provide the support function of a fiber, and an electrode array to provide the centering and restoring functions. A 0.25 m diameter float using 3 kg test masses of copper and tungsten was successfully tested and used by Keiser and Faller [6]. While their result ($\eta = 0.2 \pm 4.0 \times 10^{-11}$) was an order of magnitude worse than what one would expect if the motion of the float were predominantly determined by Brownian motion at periods of 24 h, a power spectrum analysis showed that down to frequencies of 10 cycles/day the observed motion was in fact limited by the Brownian motion of the float.

2. Description and Analysis of the Apparatus

We are now constructing a 1.27 m diameter float, shown schematically in Fig. 1. Note there are seven balls in a hexagonal pattern on the lid of the float (see also Fig. 2), and three balls in a line on the underside of the lid of the float container. These latter three are electrically insulated from the container. The central balls function as centering electrodes—an ac potential of about 1 kV is applied across them to center the float. The outer balls function as torque electrodes. (The float can be zeroed in any of six positions, differing by 60°.) Varying

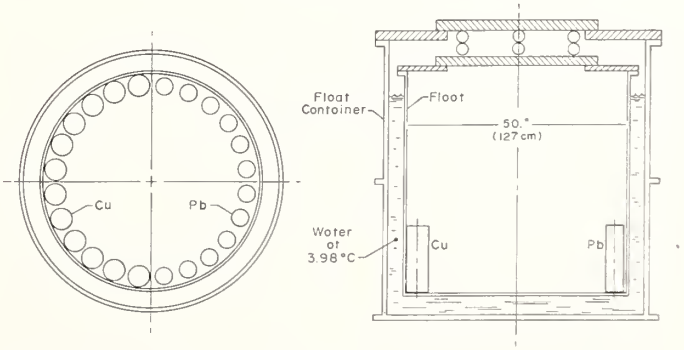


FIGURE 1. *Schematic of large float showing internal arrangement of masses.*

*Staff Member, Quantum Physics Division, National Bureau of Standards.

the ac potential applied to the outer balls serves to vary the torsion constant. As seen in Fig. 1, we are using 24 cylindrical masses; 12 of copper and 12 of lead. The total mass of each 12-piece test mass is about 490 kg. A breakdown of equivalence (i.e., a non-zero η) would result in a slight difference in the acceleration of the two sides of the float toward the sun, and this would evidence itself as a 24-hour-period variation in the position of the float.

The sensitivity of this type of apparatus can be expected to improve with size since most of the noise terms couple in through area (surface) effects, and therefore their relative importance decreases as one scales up the apparatus. Further, as one increases the size, D, of the apparatus, the fundamental thermal-noise-limited sensitivity improves, scaling as $D^{5/2}$. To see this, note that the noise torque in a bandwidth Δf due to Brownian motion is given by:

$$\tau_{\text{noise}} = (2k_b T\, b\, \Delta f)^{1/2} \tag{2}$$

where k_b is Boltzmann's constant, T is the absolute temperature, and b is the damping coefficient in the equation of motion for the float with an external torque, namely:

$$I\ddot{\Theta} + b\dot{\Theta} + K\Theta = \tau_{\text{external}} . \tag{3}$$

Here I is the moment of inertia of the float, and K is the torsion constant. The amplitude of the 24-hour-period signal torque is given by:

$$\tau_{\text{signal}} = \eta M\, a\, D \cos\delta , \tag{4}$$

where M is the mass of one of the two test masses, a is the acceleration toward the sun (0.62 cm/sec^2), D is the effective moment arm of the test masses, and δ is the declination of the sun. The coefficient of $\dot{\Theta}$ in Eq. (3) will depend on some geometric factor times the viscosity ν of the fluid. Straightforward dimensional analysis (remembering that the dimensions of ν are M/LT) shows that $b \propto \nu D^3$ (D being a characteristic linear size of the apparatus). Equating τ_{signal} and τ_{noise} (i.e., S/N = 1), noting that $M \propto D^3$, and solving for η, we see that

$$\eta_{\text{min}} \propto \frac{1}{D^{5/2}} \frac{(2k_b T\, \nu \Delta f)^{1/2}}{a \cos\delta} . \tag{5}$$

That is, the $k_b T$ limit on η improves as $D^{-5/2}$, or the sensitivity of a given apparatus improves as the five-halves power of its size. For our new apparatus, the calculated 1σ, one-day Brownian motion limit on the equivalence of gravitational and inertial mass is 7×10^{-14}.

3. Discussion

In both the Princeton [4] and the Russian experiments [5], seismic disturbances turned out to be a primary noise limit for their systems; and to this type of noise, the fluid approach seems to be remarkably insensitive. We have yet to detect any effect attributable to seismic disturbances on the position of fluid pendulums. One explanation for this can be found by noting that at least for low frequencies, time-varying accelerations do not—as they do with a normal fiber—stretch the system. Also, the fact that the zero of the system is established by a macroscopic electrode array rather than a microscopic fiber under stress probably accounts to some extent for the observed insensitivity of the zero to vibrations (and for that matter, to changes in temperature).

Of the various noise sources identified in the earlier experiment of Keiser and Faller [6], only the effect of the residual mass-quadrupole moments can be expected to be substantially worse in this new larger apparatus. This effect can be reduced by measuring the quadrupole moments of the float and systematically tuning them out by mass tuning (as was done on the small float); or one can remove the effect of residual mass-quadrupole moments on the data by directly measuring the changes in the gradients of the external gravitational field and then correcting the float's response for these changing gradients. In the case of the large float, we expect to employ both methods.

A number of practical problems have been encountered as a result of scaling up the smaller system to the large one, which is presently under construction, and which is probably near the practical size limit for scaling. As one example, the float as received from the fabricator was not as round on top as we had hoped. However, by using a hydraulic jack to deform the top just past the elastic limit, we were able to improve its as-received roundness of ± 0.035 cm to ± 0.010 cm, resulting in a fractional out-of-roundness of $\Delta R/R = 1.6 \times 10^{-4}$. Also, due to the vastly increased size of the apparatus, price and ease of fabrication became much more critical factors in choosing the materials (e.g., not only were gold and silver out of the question, but even tungsten was). It appears that lead and copper are the most practical choices. Initially we had planned to use Pb and Zn, however, the Zn castings were not as nonmagnetic as a sample had led us to expect, and they were returned to the manufacturer. We have subsequently switched to Cu. The paramagnetism problem associated with CuO is avoided by using OFHC (oxygen-free, high conductivity) copper, and by gold-plating the masses to prevent any oxidation with time.

At least a part of the non-$k_b T$ performance at long periods of the Keiser-Faller experiment has now been traced to their use of an external auto-collimator to monitor the float's angular position. This introduced not only the stability of the auto-collimator, but also the stability of the stand on which it was located as well as the mechanical integrity of the intermediate linkages between the auto-collimator stand and the container lid whose electrode array determined the "zero" of the fiber. A new sensing system (see Fig. 2) which we have devised for this experiment promises to have a much higher mechanical integrity in that the sensing optics are a part of the same container lid on which the zero-determining electrode array is located.

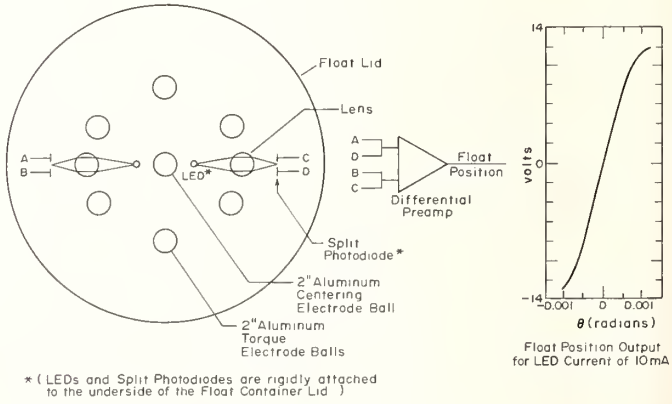


FIGURE 2. *Schematic of new angle-measuring system.*

In brief, this new sensing system consists of an infrared light-emitting diode (whose output light is focused into a line image by a cylindrical glass "lens" attached to the top of the float) and a split photodiode detector. As the float turns, the cylindrical lens moves with respect to

the LED-photodiode pair, causing the line image to move on the split photodiode. This results in a change in the output of the two halves of this device, giving rise, when amplified, to a rotation-dependent voltage as shown on the right half of Fig. 2. With two such detection systems—one on each side—and proper summing of their outputs, the detector is insensitive to sideways motion of the float without the sacrifice of any of its sensitivity to rotation.

4. Conclusion

This new large apparatus is now in the final stages of construction. We are awaiting delivery of the (encompassing) magnetic shield before installing this apparatus in the "Spectroscopy Lab" in the sub-basement of JILA to begin a prolonged (possibly a year-long) shake-down of the full apparatus. We plan to put a 0.2 to 0.5-cm layer of oil on the surface of the water, which we have experimentally determined lowers the evaporation rate by a factor of about 1000. This should help avoid local evaporative cooling of the surface of the water (which could lead to thermal gradients) and prevent the subsequent recondensation of water on the float (which could produce undesired quadrupole mass moments). We also plan to make the initial tests without cooling; and, by establishing a positive upward temperature gradient, we will try to achieve sufficient temperature stability to inhibit convection. Cooling, while straightforward and serving to reduce the linear expansion coefficient, α, water—and therefore the tendency to convect [7]—by more than an order of magnitude (at 20 °C, $\alpha \approx 7 \times 10^{-5}$/°C, while at 4 °C, $\alpha \approx 1.5 \times 10^{-7}$/°C), extracts a certain price in experimental inconvenience. This is particularly true in a system of this size, and we would rather not have to pay this price unless it proves absolutely necessary. However, even without cooling, the scale of this experiment is such that the first improved results on the equivalence principle using this large float should not be expected until 1983 or 1984.

This work is supported in part by the National Science Foundation, Grant No. PHY79–04928, and in part by the National Bureau of Standards as part of its research program on improved measurement techniques for applications to basic standards.

References

[1] R. J. Adler, M. J. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1975), p. 4.
[2] F. W. Bessel, Poggendorff's Annalen der Physik 25, 401 (1832).
[3] R. von Eötvös, D. Pekar, and E. Fekete, Ann. Phys. (Leipzig) 68, 11 (1922).
[4] P. G. Roll, R. Krotkov, and R. H. Dicke, Ann. Phys. (N.Y.) 26, 442 (1964).
[5] V. B. Braginsky and V. I. Panov, Sov. Phys. JETP 34, 463 (1972).
[6] G. M. Keiser and J. E. Faller, "Eötvös Experiment with a Fluid Fiber," in *Proceedings of the Second Marcel Grossman Meeting on Recent Developments on General Relativity*, Ed. by R. Ruffini (North-Holland, 1982), p. 969.
[7] C. Normand, Y. Pomeau, and M. G. Velarde, Rev. Mod. Phys. 49, 581 (1977).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

What Test Masses Are Best for an Eötvös Experiment?

D. F. Bartlett, J. Shepard, and C. D. Zafiratos

Department of Physics, University of Colorado, Boulder, CO 80309

No element lighter than aluminum has been used as a comparison mass in the recent precise tests of the equivalence of gravitational and inertial mass. We discuss how the substitution of solid hydrogen for aluminum as the light mass could increase the sensitivity of an Eötvös experiment by a factor between 10 and 100.

Key words: Eötvös experiment; equivalence principle; gravitational mass; inertial mass; test mass.

1. Introduction

In an Eötvös experiment, the inertial and gravitational masses of two test bodies of different materials are compared. The results of this comparison are characterized by the coefficient η:

$$\eta = M_A/m_A - M_B/m_B ,$$

where M is the gravitational mass and m the inertial mass of substances A and B. Since Galileo, the observed universality of the acceleration of differing masses in a gravitational field has been used to set increasingly stringent limits on the departure of η from 0. Since these tests form the basis of the equivalence principle, it is important to know how different A and B must be for the principle to be effectively tested.

Eötvös himself used magnalium and snakewood among a variety of test masses for the comparison of inertial and gravitational mass [1]. As experiments have become more precise, the choice of materials has narrowed [2, 3] until in the most recent published experiment only two materials were used [4]. Aluminum and platinum were chosen partly for their convenient physical properties. They also differ in nuclear characteristics such as ratio of neutrons to protons or of binding energy to nuclear mass. These differences in nuclear properties should make the experiment a sensitive test of the equivalence principle. However even aluminum is a heavy nucleus in that the distribution of its nucleons is well described by the standard two parameter Fermi function used to describe all nuclei heavier than carbon. Can one increase the sensitivity of an Eötvös experiment by choosing an element lighter than aluminum for the light test mass?

The answer is clearly yes if hydrogen is the choice. With no neutrons and no nuclear binding, hydrogen is clearly unique. If the strong or electromagnetic interactions contribute slightly differently to the inertial and gravitational mass of a nucleus, hydrogen is about 10 times as good a test mass as aluminum [5]. If the anomaly is in the weak interactions, the benefit is a factor of 100 [6].

More convenient choices for the light mass may include lithium and beryllium. In the absence of any model for a breakdown of the equivalence principle it is impossible to predict which complex nucleus would be the best choice. We will suggest that lithium is a better choice than aluminum, though not nearly as good as hydrogen.

2. Strong and Electromagnetic Interactions

Since most of an atom's mass is in the nucleus, it is most likely that a positive effect in an Eötvös experiment would come from differences in the nuclei of the two test masses. The most obvious criterion is the neutron to proton ratio. As is well known, the ratio N/Z remains close to one for all elements between helium and calcium after which the ratio increases gradually, reaching 1.53 at lead. As shown in Fig. 1, a reasonable approximation is

$$N/Z = 0, \quad \text{For } A = 1$$
$$N/Z = 1, \quad \text{For } 2 \leq A < 40$$
$$N/Z = 0.1 + 0.7 \log_{10} A, \quad \text{for } 40 < A \leq 208 ,$$

Where

$$A = N + Z .$$

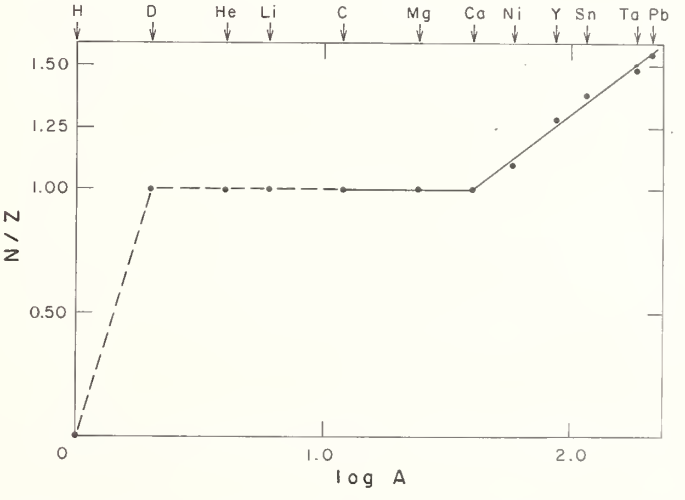


FIGURE 1. *Plot of N/Z vs $\log_{10}A$. Semilogarithmic plot emphasizes light nuclei.*

If neutrons and protons contribute differently to inertial and gravitational mass, hydrogen is clearly a unique choice for the light test mass. We may quantify this uniqueness by defining a figure of merit,

$$\mathrm{FM}_{np} = ((N - Z)/A)_{\text{heavy}} - ((N - Z)/A)_{\text{light}} .$$

For platinum as the heavy mass and aluminum as the light mass FM_{np} is 0.16. If hydrogen replaces aluminum as the light mass, FM_{np} increases to 1.2.

Alternatively, the test masses may differ in the amount of binding of the nucleons in the nucleus. A plot of binding energy per nucleon (BE/A) vs log A is given in Fig. 2 (A). The uniqueness of hydrogen is also evident from this curve. A convenient figure of merit is

$$FM_{BE} = ((A - M)/A)_{\text{heavy}} - ((A - M)/A)_{\text{light}},$$

where A is $N + Z$ and M is the atomic weight. For platinum as the heavy mass and either aluminum or hydrogen as the light mass, FM_{BE} is 0.4×10^{-3} and 8.5×10^{-3}, respectively.

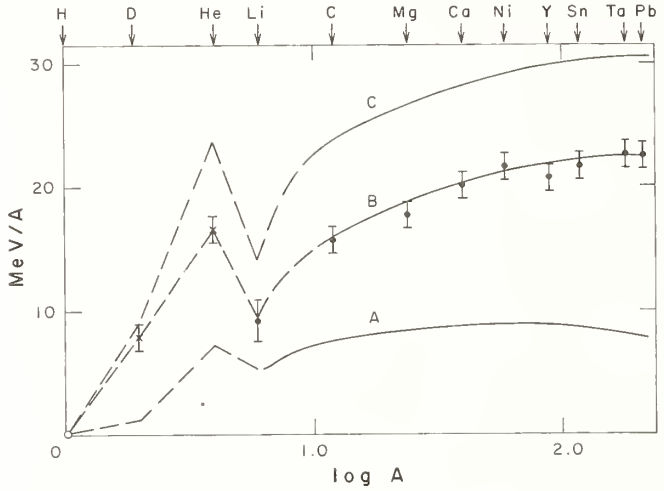


FIGURE 2. *Plot of (a) BE/A, (b) $\overline{T}$, and (c) $|\overline{V}| = BE/A + \overline{T}$ vs log A. The solid dots are from measurements of Ref. 8, the isotopes used being as naturally occurring except for ^{6}Li and ^{208}Pb; the crosses are from calculations of Ref. 7. In either case the errors represent fitting uncertainties.*

Of course these comparisons are not to be taken too literally—particularly since the mechanism for a violation of equivalence is not specified. However, the comparison does indicate that the use of hydrogen as a test mass should increase the sensitivity of the Eötvös experiment to either the strong or electromagnetic interactions by roughly a factor of 10.

Conceivably a positive effect in an Eötvös experiment could come from the *motion* of the nucleons in the nucleus. This could happen, for instance, if the usual relativistic increase in the mass of a moving proton or neutron were slightly different for inertial and gravitational masses. It is thus convenient to separate the binding energy into its kinetic and potential parts,

$$\frac{BE}{A} = |\overline{V}| - \overline{T}.$$

In an earlier paper, we used Hulthén wave functions to calculate the rms velocity and hence the kinetic energy of the nucleons in deuterium and helium [7]. Some time ago, Moniz and coworkers made careful measurements of the motion of nucleons in nine complex nuclei [8]. Using 500 MeV electrons as a probe, they determined that the distribution of nucleon velocities was well described as that of a Fermi gas having zero temperature and a Fermi momentum k_F ranging from 165 MeV/c for ^{6}Li to 265 MeV/c for ^{208}Pb. We have used the non-relativistic equations $\overline{k^2} = (3/5)k_F^2$ and $\overline{T} = \overline{k^2}/2m$, where m is the mass of nucleon, to determine $\overline{T}$ for the complex nuclei shown in Fig. 2 (B).

It is evident from this figure that although hydrogen is still the best choice its superiority is not as dramatic as in the previous discussion. Lithium is a close second. Using a figure of merit,

$$FM_{\overline{T}} = (\overline{T}/A)_{Pt} - (\overline{T}/A)_{\text{light}},$$

we find $FM_{\overline{T}} = 4 \times 10^{-3}$, 14×10^{-3}, and 24×10^{-3} for Al, ^{6}Li, and H, respectively.

Finally, Fig. 2 (C) showing the variation of mean potential $|\overline{V}|$ with A is obtained from the addition $|\overline{V}| = \overline{T} + BE$. In a separate calculation, we have determined $\overline{V}$ directly for nuclei heavier than carbon. We made numerical integrations using as inputs the standard two-parameter Fermi functions for optical model parameters [9] and for nuclear charge distributions [10]. Our results, which are not shown, confirm the trend of Fig. 2 (C), although with somewhat higher values for $\overline{V}$ that those obtained by addition.

3. Weak Interactions

The present accuracy of the Eötvös experiment is already sufficient to limit any anomaly in the interaction of gravity with the strong or electromagnetic interactions to about $\eta/FM_{BE} = 10^{-12}/0.0004 \simeq 10^{-9}$. Cognizant of this severe limit, Haugan and Will have recently looked at the sensitivity of the Eötvös experiment to the weak interactions [6]. They find that the contribution of the weak interactions to the observed inertial mass of the nucleus is $E_W/m_I = 2.2 \times 10^{-8}\ (NZ/A^2)F(N, Z)$, where $F(N,Z)$ is a slowly varying function about equal to one. After taking weak neutral currents into consideration, they find that

$$FM_W(\text{Al}) = (E_W/m_I)_{Pt} - (E_W/M_I)_{Al} = -0.8 \times 10^{-10}$$

for $\sin^2\theta_W = 0.2$, where θ_W is the Weinberg angle.

By contrast,

$$FM_W(\text{H}) = (E_W/m_I)_{Pt} = 0.7 \times 10^{-8},$$

about 100 times larger.

In Fig. 3 we have used the explicit form of $F(N,Z)$

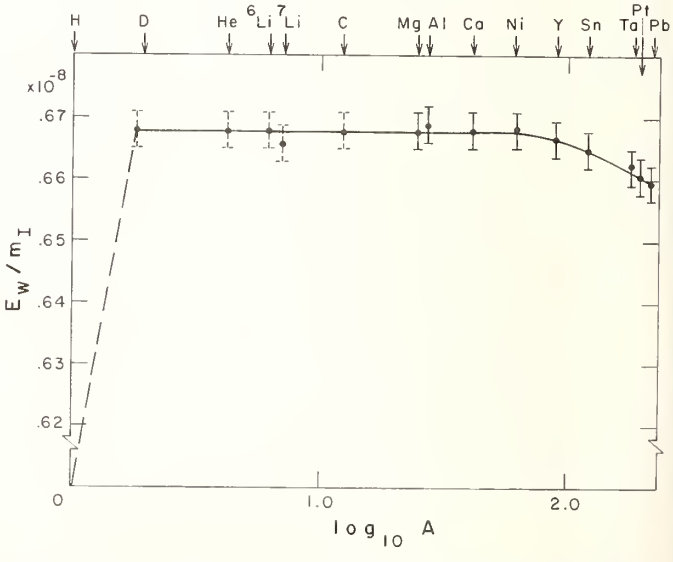


FIGURE 3. *Fraction of nuclear energy assignable to weak interactions (E_W/m_I) vs log A. All data from calculation of Ref. 6. The errors represent the unknown contribution from nuclear structure to E_W. For nuclei lighter than aluminum this contribution becomes hard to estimate.*

given in Eqs. (2) and (3) of Haugan and Will's article to plot E_W/m_I for various nuclei. The superiority of hydrogen is evident.

4. Conclusion

The virtues of hydrogen for an Eötvös experiment were first mentioned by Lee and Yang in 1955 [5]. At that time the theoretical advantages of hydrogen were outweighed by the obvious experimental difficulties involved in actually using hydrogen as a test mass. Within the last decade, however, cryogenic techniques have been increasingly used in precise gravitational measurements. Such techniques may well make solid hydrogen a feasible test mass. Meanwhile, lithium differs sufficiently in nuclear properties from heavier nuclides to make it an attractive second choice.

References

[1] R. v. Eötvös, D. Pekar, and E. Fekete, Ann. Phys. (Leibzig) 68, 11 (1922).
[2] P. G. Roll, R. Krotkov, and R. H. Dicke, Ann. Phys. (N.Y.) 26, 442 (1964).
[3] G. M. Keiser and J. E. Faller, Nuovo Cimento (to be published).
[4] V. B. Braginsky and V. I. Panov, Zh. Eksp. Teor. Fiz. 61, 873 (1971) [Engl. Trans.: Sov. Phys. JETP 34, 463 (1972)].
[5] T. D. Lee and C. N. Yang, Phys. Rev. 98, 1501 (1955).
[6] M. P. Haugan and C. M. Will, Phys. Rev. Lett. 37, 1 (1976).
[7] D. F. Bartlett, J. Shepard, C. Zafiratos, and B. F. L. Ward, Phys. Rev. D 20, 578 (1979).
[8] E. J. Moniz, I. Sick, R. R. Whitney, J. R. Ficenec, R. D. Kephart, and W. P. Trower, Phys. Rev. Lett. 26, 445 (1971).
[9] C. M. Percy and F. G. Percy, At. Data Nucl. Data Tables 17, 1 (1976).
[10] DeJager *et al.*, At. Data Nucl. Data Tables 14, 479 (1974).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Equivalence Principles and Precision Experiments*

Wei-Tou Ni

Department of Physics, National Tsing Hua University, Hsinchu, China (Taiwan)

We use a general framework—the χ-g framework—to study the empirical foundations of the Einstein Equivalence Principle (EEP), and to analyze the theoretical significance of various precision experiments and observations such as timing observations of pulsar signal propagation through galactic gravitational field, the Hughes-Drever experiments, the Eötvös-Dicke-Braginsky experiments, and the Vessot-Levine experiment. These experiments constrain various linear combinations of the components of χ to be metric to various degrees of accuracy, leaving only one component out of 21 not constrained. The theory with this component different from zero can be reformulated as a torsion theory. To test EEP and this theory, experiments on polarized test bodies are suggested.

Key words: equivalence principles; gravitation; pulsar signals; redshifts; symmetry violations; test-body experiments; torsion; unification of fundamental interactions.

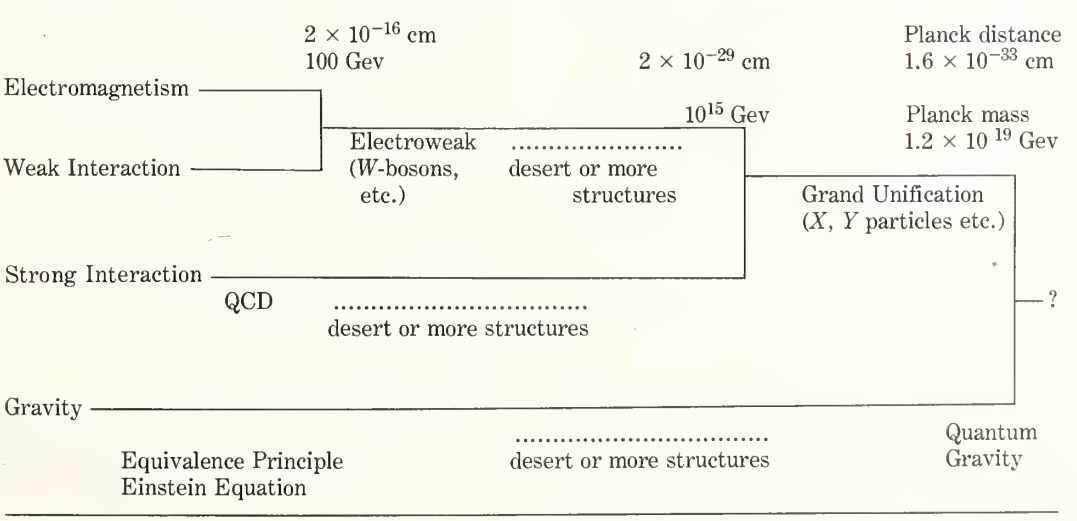
1. Introduction

In 1907, Einstein [1] proposed the complete physical equivalence of a homogeneous gravitational field to a uniformly accelerated reference system and derived clock and energy redshifts in a gravitational field from this equivalence. When applied to a spacetime region where inhomogeneities of the gravitational field can be neglected, this equivalence dictates the behavior of matter in a gravitational field. The postulate of this equivalence is called the Einstein Equivalence Principle (EEP). EEP is the cornerstone of the gravitational coupling of matter and non-gravitational fields in general relativity and metric theories of gravity.

The renaissance of general relativity research in the last two decades together with the fundamental discoveries and developments in particle physics leads to renewed interest in the microscopic origin of gravity. The discovery of parity violation [2, 3] and CP violation [4] in the weak interaction puts us into a symmetry broken world. The success of the Weinberg-Salam [5, 6] theory of unified electroweak interactions makes spontaneous symmetry breaking a promising way to generate "fundamental" constants and to unify interactions. The incorporation of spontaneous symmetry breaking in gravity to generate Newton's gravitational constant G_N has been considered by various authors [7]. Considering the role of spins and different gauge groups, various torsion theories and gauge-type theories of gravity have been proposed. Many of these theories violate EEP in one way or another at certain levels.

With our present knowledge, a unification scale can be drawn in Table 1. Electromagnetism and the weak interaction would be unified as the electroweak interaction at the W-boson energy (~100 Gev). According to the grand unification schemes [8], the electroweak and strong interac-

TABLE 1. *Energy and distance scale of possible unification of fundamental interactions. The distances in the scale are the reduced Compton wavelengths of the corresponding energies.*



*Supported in part by the National Science Council

tions would be unified at an energy of 10^{15} Gev. At present accelerator energies ($\sqrt{s} \sim 62$ Gev), Quantum Chromodynamics (QCD) looks promising in explaining the strong interaction. From this energy and the W-boson energy to the grand unification energy, there is a gap of 13-orders of magnitude. The important question is whether in this wide gap, there are more structures besides electroweak unification and QCD. In gravitation, the equivalence principle and Einstein equation are only verified empirically at lower energies and larger distances. Using dimensional arguments, quantum phenomena should be important in gravity at the Planck-mass energy $(\hbar c^5/G)^{1/2} = 1.223 \times 10^{19}$ Gev or Planck distance $(\hbar G/c^3)^{1/2} = 1.616 \times 10^{-33}$ cm. Above this energy, it is also possible that the grand unified interaction and quantum gravity can further be unified. Here the important question is whether the gravitational coupling of matter still obeys the Einstein equivalence principle. *To answer this question, ever more precise measurements in ever more diverse situations are desired.* This is a paramount challenge to the field of precision measurement.

As an illustration of why the naive point of view of no more structures might not be right, we quote the following historical example. Einstein's efforts of unifying electromagnetism and gravity were not successful. Now we understand why. Before electromagnetism and gravity can be unified, one has to take into account the weak interaction and, possibly, the strong interaction. Similar things could happen again and there would be more structures.

Both the fundamental role of EEP in general relativity and its possible violation at a certain level demand a close scrutiny of its empirical foundations. For this purpose, we use a general framework—the χ-g framework [9, 10]—to study the empirical foundations of EEP, to analyze the theoretical significance of various experiments and observations, and to propose new experiments.

2. The χ-g Framework and Three Theorems

Since here we are only concerned with the behavior of matter in a gravitational field, we treat the gravitational field as an external field. Most experiments concern electromagnetism. Therefore to be more specific, we consider electromagnetically interacting systems in this article. Generalizations to strong and weak interactions will be presented elsewhere. In previous work we have used a general framework—the χ-g framework [9, 10] to study Schiff's conjecture and theoretical relations of various equivalence principles. The χ-g framework is rather comprehensive [10] and we use it in sections 3–5 to study the empirical foundations of EEP.

This framework can be summarized in the following interaction Lagrangian density

$$\mathfrak{L}_I = -(\frac{1}{16\pi})\chi^{ijk\ell}F_{ij}F_{k\ell} - A_k j^k(-g)^{1/2} - \Sigma_I m_I \frac{ds_I}{dt}\delta(\mathbf{x} - \mathbf{x}_I), \quad (1)$$

where $\chi^{ijk\ell} = \chi^{k\ell ij} = -\chi^{k\ell ji}$ is a tensor density of the gravitational fields (e.g., g_{ij}, ϕ, etc.), and j^k, $F_{ij} \equiv A_{j,i} - A_{i,j}$ have the usual meaning. The gravitational constitutive tensor density $\chi^{ijk\ell}$ dictates the behavior of electromagnetism in a gravitational field and has 21 independent components in general. For a metric theory (when EEP holds), $\chi^{ijk\ell}$ is determined completely by the metric g^{ij} and equals $(-g)^{1/2}(\frac{1}{2}g^{ik}g^{j\ell} - \frac{1}{2}g^{i\ell}g^{kj})$.

In Refs. [9, 10], we have proved the following theorem.

Theorem I.—For a system whose Lagrangian density is given by Eq. (1), WEP I holds if and only if

$$\chi^{ijk\ell} = (-g)^{1/2}\left[\frac{1}{2}g^{ik}g^{j\ell} - \frac{1}{2}g^{i\ell}g^{kj} + \phi\epsilon^{ijk\ell}\right], \quad (2)$$

where ϕ is a scalar function of the gravitational field and $\epsilon^{ijk\ell} = (-g)^{1/2}e^{ijk\ell}$ with

$$e^{ijk\ell} = \begin{cases} 1, \text{ if } (ijk\ell) \text{ is an even permutation of (0123)} \\ -1, \text{ if } (ijk\ell) \text{ is an odd permutation of (0123)}, \\ 0, \text{ otherwise.} \end{cases} \quad (3)$$

By WEP I we mean Galileo's weak equivalence principle that macroscopic test bodies with the same initial conditions follow the same trajectories. If $\phi \neq 0$ in (2), the gravitational coupling to electromagnetism is not minimal and EEP is violated. Hence WEP I does not imply EEP. But WEP I does constrain the 21 degrees of freedom of χ to only one degree of freedom (ϕ). In the actual empirical situation, since Eötvös-Dicke-Braginsky experiments are performed on unpolarized bodies, they constrain only 2 degrees of freedom of χ (cf. §4). Only when these experiments are performed on polarized bodies with various electromagnetic energy configurations, can they constrain the other 18 degrees of freedom.

In the nonmetric theory with $\chi^{ijk\ell}$ ($\phi \neq 0$) given by Eq. (2) [11], there are anomalous torques on electromagnetic-energy-polarized bodies so that different test bodies will change their rotation state differently, like magnets in magnetic fields (cf. §5). Since the motion of a macroscopic test body is determined not only by its trajectory but also by its rotation state, the motion of polarized test bodies will not be the same. We, therefore, propose the following stronger weak equivalence principle (WEP II) to be tested by experiments, which states that in a gravitational field, the motion of a test body with a given initial motion state is independent of its internal structure and composition (universality of free-fall motion). Therefore, in this framework, the imposition of WEP II guarantees that EEP is valid. In this nonmetric theory the first-order gravitational redshift is metrical. These results can be stated as the following two theorems:

Theorem II—For the Lagrangian of Eq. (1), WEP II implies EEP.

Theorem III—WEP I implies that the first-order gravitational redshift is metrical.

3. Pulsar Signal Propagation in the Galactical Gravitational Field

If EEP is observed, photons with different polarizations as test particles shall follow identical trajectories in a gravitational field. In pulsar observations, the pulses and micropulses with different polarizations are correlated in their general structure and timing [12]. Due to precise timing and rich polarization data, pulsar signal propagation in the galactic gravitational field constitutes a high precision test of EEP.

Since our galactic Newtonian potential U is of the order of 10^{-6}, we use the weak field approximation in the χ-g framework. The vacuum Maxwell equation in this framework is

$$(\chi^{ijk\ell}A_{k,\ell})_{,j} = 0. \quad (4)$$

Neglecting $\chi^{ijk\ell}{}_{,p}$ in a slowly varying field, Eq. (4) be-

comes

$$\chi^{ijk\ell}\, A_{k,\ell j} = 0\,. \tag{5}$$

For weak field, we assume

$$\chi^{ijk\ell} = \chi^{(0)ijk\ell} + \chi^{(1)ijk\ell} \tag{6}$$

where

$$\chi^{(0)ijk\ell} = \frac{1}{2}\eta^{ik}\eta^{j\ell} - \frac{1}{2}\eta^{i\ell}\eta^{kj} \tag{7}$$

with η^{ij} the Minkowski metric and $|\chi^{(1)}\text{'s}| << 1$.

Using the eikonal approximation, we look for a plane-wave solution propagating in the z-direction. Imposing the radiation condition in the zeroth order and solving the dispersion relation for ω, we obtain

$$\omega_{\pm} = k\{1 + \frac{1}{4}[(K_1 + K_2) \pm \sqrt{(K_1 - K_2)^2 + 4K^2}]\} \tag{8}$$

where

$$\begin{aligned} K_1 &= \chi^{(1)1010} - 2\chi^{(1)1013} + \chi^{(1)1313}\,; \\ K_2 &= \chi^{(1)2020} - 2\chi^{(1)2023} + \chi^{(1)2323} \\ K &= \chi^{(1)1020} - \chi^{(1)1023} - \chi^{(1)1320} + \chi^{(1)1323}\,. \end{aligned} \tag{9}$$

Photons with two different polarizations propagate with different speeds $v_{\pm} = \omega_{\pm}/k$ and would split in 4-dimensional spacetime. The conditions for no splitting (no retardation) is $\omega_+ = \omega_-$, i.e.

$$K_1 = K_2, \quad K = 0\,. \tag{10}$$

Eq. (10) gives two constraints on the $\chi^{(1)}$'s.

The condition for no splitting (no retardation) in all directions gives ten constraints on the $\chi^{(1)}$'s. With these ten constraints, χ can be written in the following form

$$\chi^{ijk\ell} = (-H)^{1/2}\left[\frac{1}{2}H^{ik}H^{j\ell} - \frac{1}{2}H^{i\ell}H^{kj}\right]\psi + \phi\, e^{ijk\ell} \tag{11}$$

where $H = \det(H_{ij})$.

In actual observations, the pulses and micropulses with different polarizations are correlated in general structure and no retardation with respect to polarizations are observed. This means that conditions similar to Eq. (10) are satisfied to observational accuracy. For the Crab pulsar, micropulses with different polarizations are correlated in timing to within 10^{-4} sec. The distance of the Crab pulsar is 2200 pc; therefore to within 10^{-4} sec/(2200 × 3.26 light yr.) = 5×10^{-16} accuracy two conditions similar to Eq. (10) are satisfied. Over 300 pulsars in different directions are observed. Many of them have polarization data. Combining all of them, Eq. (11) is satisfied to an accuracy of $10^{-14} - 10^{-16}$. Since $U \sim 10^{-6}$, $\chi^{(1)}/U$ (or χ/U) agrees with that given by Eq. (11) to an accuracy of $10^{-8} - 10^{-10}$. Detailed analysis will reveal better results.

Thus, to high accuracy, photons are propagating in the metric field H^{ik} and two additional scalar fields ϕ and ψ. A change of H^{ik} to λH^{ik} does not affect $\chi^{ijk\ell}$ in Eq. (11). Thus we have eleven degrees of freedom in Eq. (11).

4. Hughes-Drever Experiments, Eötvös-Dicke-Braginsky Experiments, and Vessot-Levine Experiment

Since Eq. (11) is verified empirically to high accuracy from pulsar observations, we start from Eq. (11) in this section.

In Eq. (1), ds is the line element determined from the metric g_{ij}. From Eq. (11), the gravitational coupling to electromagnetism is determined by the metric H_{ij} and two scalar fields ϕ and ψ. If H_{ij} is not proportional to g_{ij}, then the hyperfine levels of the lithium atom will have additional shifts. But this is not observed to high accuracy in Hughes-Drever experiments [13, 14]. Therefore H_{ij} is proportional to g_{ij} to a certain accuracy. Since a change of H^{ik} to λH^{ik} does not affect $\chi^{ijk\ell}$ in Eq. (11), we can define $H_{11} = g_{11}$ to remove this scale freedom.

In Hughes-Drever experiments $\Delta m/m \leq 0.5 \times 10^{-22}$ or $\Delta m/m_{e.m.} \leq 0.3 \times 10^{-18}$ where $m_{e.m.}$ is the electromagnetic binding energy. Using Eq. (11) in Eq. (1), we have three kinds of contributions to $\Delta m/m_{e.m.}$. These three kinds are of the order of (i) $(H_{\mu\nu} - g_{\mu\nu})$, (ii) $(H_{0\mu} - g_{0\mu})\,v$, and (iii) $(H_{00} - g_{00})\,v^2$ respectively. Here the Greek indices μ, ν denote space indices. Considering the motion of laboratories from earth rotation, in the solar system and in our galaxy, we can set limits on various components of $(H_{ij} - g_{ij})$ from Hughes-Drever experiments as follows:

$$\begin{aligned} |H_{\mu\nu} - g_{\mu\nu}|/U &\leq 10^{-12} \\ |H_{0\mu} - g_{0\mu}|/U &\leq 10^{-7} - 10^{-8}\,, \\ |H_{00} - g_{00}|/U &\leq 10^{-4}\,. \end{aligned} \tag{12}$$

where U ($\sim 10^{-6}$) is the galactical gravitational potential.

Eötvös-Dicke-Braginsky [15, 16, 17] experiments are performed on unpolarized test bodies. In essence, these experiments show that unpolarized electric and magnetic energies follow the same trajectories as other forms of energy to a certain accuracy. The constraints on Eq. (11) are

$$|1 - \psi|/U < 10^{-9} \tag{13}$$

and

$$|H_{00} - g_{00}|/U < 10^{-5} \tag{14}$$

where U is the solar gravitational potential at the earth.

In 1976, Vessot and Levine [18] used an atomic hydrogen maser clock in a space probe to test and confirm the metric gravitational redshift to an accuracy of 1.4×10^{-4} [19]. The space probe attained an altitude of 10,000 km above the earth's surface. With Eq. (13), the constraint on Eq. (11) is

$$|H_{00} - g_{00}|/U \leq 1.4 \times 10^{-4}\,. \tag{15}$$

Thus, we see that for the constraint on $|H_{00} - g_{00}|/U$, Hughes-Drever experiments, Eötvös-Dicke-Braginsky experiments, and the Vessot-Levine experiment compete among each other in accuracy.

The empirical constraints from the last section and this section can be summarized in Table 2.

5. A Nonmetric Theory of Gravity, Torsion and Polarized Test-Body Experiments

From the results of the above two sections, we know that χ is constrained as in Eq. (2) to various degrees of accuracy. Eq. (2) together with Eq. (1) is the nonmetric theory we proposed earlier [9, 10, 11] as a counter example to Schiff's conjecture. In the following we discuss this theory.

The interaction Lagrangian is

$$\mathcal{L}_I = \mathcal{L}_I^{(M)} + \mathcal{L}_I^{(NM)} \tag{16}$$

TABLE 2. *Empirical foundations of the Einstein Equivalence Principle. With the constraints below,* $\chi^{ijk\ell} = (-g)^{1/2}\left[\frac{1}{2}g^{ik}g^{j\ell} - \frac{1}{2}g^{i\ell}g^{kj}\right] + \phi e^{ijk\ell}$ *to various degrees of accuracy, i.e., EEP is verified to various degrees of accuracy except for the freedom in* ϕ.

Experiments	Constraints	Accuracy
Pulsar Signal Propagation	$\chi^{ikl\ell} \to (-H)^{1/2}\left[\frac{1}{2}H^{ik}H^{j\ell} - \frac{1}{2}H^{i\ell}H^{kj}\right]\psi + \phi e^{ijk\ell}$	10^{-8} - 10^{-10}
Hughes-Drever Experiments	$H_{\mu\nu} \to g_{\mu\nu}$ $H_{0\mu} \to g_{0\mu}$ $H_{00} \to g_{00}$	10^{-12} 10^{-7} - 10^{-8} 10^{-4}
Eötvös-Dicke-Braginsky Experiments	$\psi \to 1$ $H_{00} \to g_{00}$	10^{-9} 10^{-5}
Vessot-Levine Redshift Experiment	$H_{00} \to g_{00}$	10^{-4}

where $\mathcal{L}_I^{(M)}$ is the usual metric Lagrangian and

$$\mathcal{L}_I^{(NM)} = (-\frac{1}{16\pi})\,\phi F_{ij}F_{k\ell}\,e^{ijk\ell}$$
$$= (-\frac{1}{4\pi})\,\phi_{,i}A_jA_{k,\ell}\,e^{ijk\ell} \quad \text{(mod. div.)}. \tag{17}$$

The Maxwell equations are

$$F^{ik}{}_{;k} + \epsilon^{ikm\ell}F_{km}\,\phi_{,\ell} = -4\pi j^i. \tag{18}$$

The Lorentz force law is the same as in metric theories of gravity. Gauge invariance and charge conservation are guaranteed. The Maxwell equations (18) are also conformally invariant.

This theory can be put into the form of a torsion theory. Define a metric compatible affine connection as

$$\Gamma^i{}_{jk} = \{\overset{i}{jk}\} + \Omega^i{}_{jk} \tag{19}$$

where $\{\overset{i}{jk}\}$ is the Christoffel symbol obtained from g_{ij} and

$$\Omega^i{}_{jk} = 2\phi_{,\ell}\,\epsilon^{\ell i}_{jk}. \tag{20}$$

The Maxwell equations (18) can be written as

$$\overline{F}^{ik}{}_{|k} = -4\pi j_i \tag{21}$$

where $|$ denotes covariant differentiation with respect to $\Gamma^i{}_{jk}$ (e.g., $V^i{}_{|j} = V^i{}_{,j} + \Gamma^i{}_{jm}V^m$ etc.) and

$$\overline{F}_{ik} \equiv A_{k|i} - A_{i|k} = A_{k,i} - A_{i,k} + 2\Omega^{\ell}{}_{ik}A_\ell. \tag{22}$$

The nonmetric part of the Lagrangian can be written in the form

$$\mathcal{L}_I^{(NM)} = 2A_jA_{k,\ell}\,\Omega^{jk\ell}\sqrt{-g}. \tag{23}$$

To complete this theory as a gravitational theory, we have to add a gravitational Lagrangian to it. For example, the gravitational Lagrangian $\mathcal{L}_G$ could be

$$\mathcal{L}_G = \frac{\sqrt{-g}}{16\pi}R(\Gamma^i{}_{jk}), \tag{24}$$

$$\mathcal{L}_G = \frac{\sqrt{-g}}{16\pi}\left[R(\Gamma^i{}_{jk}) + \eta\,\phi_{,i}\,\phi^{,i}\right], \tag{25}$$

or

$$\mathcal{L}_G = \frac{\sqrt{-g}}{16\pi}\left[\phi R(\{^i{}_{jk}\}) - \frac{1}{\phi}\omega(\phi)\,\phi_{,i}\,\phi^{,i}\right], \tag{26}$$

where η is a parameter and $\omega(\phi)$ is a function of ϕ.

Defining the electromagnetic stress-energy as

$$T^{(em)}_{ij} = \frac{1}{4\pi}(-F_{i\ell}F_j{}^{\ell} + \frac{1}{4}F_{\ell m}F^{\ell m}g_{ij}), \tag{27}$$

we have the following matter response equation:

$$T^k_{i;k} = -\frac{1}{4\pi}\epsilon^{jkm\ell}F_{ij}F_{km}\,\phi_{,\ell}. \tag{28}$$

From Theorem I in §2, we know that test bodies follow geodesics in this theory. Choose geodesic frames such that $g_{ij,k} = 0$ along the geodesic of the test body considered. Define the angular momentum tensor as

$$J_{ik} = \int (x_iT_{k0} - x_kT_{i0})dV. \tag{29}$$

The rate of change of the three angular momenta for the test body is

$$\frac{dJ_{\mu\nu}}{dt} = I_\mu\phi_{,\nu} - I_\nu\phi_{,\mu} \text{ or } \mathbf{J} = \mathbf{I}\times\nabla\phi \tag{30}$$

where

$$I_\mu = \frac{1}{16\pi}\int\epsilon^{njk\ell}F_{nj}F_{k\ell}\,x_\mu\,dV \tag{31}$$

is the gravitational polarization vector. Eq. (30) is similar to the one for a magnet in a magnetic field. Thus, the motion would be oscillatory about the $\nabla\phi$ axis. The order of magnitude of the torque would be $E_{e.m.}$ × fraction of polarization × d × $\nabla\phi$ where d is the dimension of the polarized electromagnetic field in the body. To constrain the remaining ϕ freedom of χ and to test this nonmetric theory, experiments on polarized test bodies with I_μ large are suggested.

The acceleration measurements on polarized test bodies would be tests of WEP I. The importance of these experiments as tests of EEP has been emphasized by Morgan

and Peres [20]. From Theorem I, WEP I constrains 20 degrees of freedom of χ. Eötvös-Dicke-Braginsky experiments constrain 2 degrees of freedom of χ. The acceleration measurements on polarized test bodies would constrain the other 18 degrees of freedom. This would constitute a check on the constraints obtained by other experiments in Table 2.

According to Theorem II, the acceleration and torque measurements of both polarized and unpolarized test bodies would be a complete set of tests of EEP.

Polarized test body experiments can also probe the role of spin in gravity. This has to be treated in an extended phenomenological framework.

6. Discussions

In this paper, we have used a general framework to study the empirical foundations of EEP and to analyze the theoretical significance of various precision experiments. We set empirical limits on χ's. This is useful in finding out the viable theories of gravity. For example, Hojman-Rosenbaum-Ryan-Shepley theory [21] violates the limits in Table 2, and, hence, is nonviable [22].

In the context of cosmology, the χ due to all cosmological matter should be used in the analysis of pulsar signal propagation and Hughes-Drever experiments. χ could also have a spontaneous-symmetry-breaking piece. This piece is also constrained by the analysis of §3 and §4. Detailed analysis will be presented elsewhere.

References

[1] A. Einstein, Jahrb. Radioakt. Elektron. 4, 411 (1907); corrections by Einstein in Jahrb. Radioakt. Elektron. 5, 98 (1908); English translations by H. M. Schwartz in Am. J. Phys., 45, 512, 811, 899 (1977).

[2] T. D. Lee and C. N. Yang, Phys. Rev. 104, 254 (1956).

[3] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, Phys. Rev., 105, 1413 (1957).

[4] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, Phys. Rev. Lett. 13, 138 (1964).

[5] S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967).

[6] A. Salam, in *Elementary Particle Theory*, Ed. by N. Svartholm (Wiley, New York, 1968), p. 367.

[7] Y. Fujii, Phys. Rev. D 9, 874 (1974); P. Minkowski, Phys. Lett., **71B**, 419 (1977); T. Matsuki, Prog. Theor. Phys. 59, 235 (1978); A. D. Linde, Pis'ma Zh. Eksp. Teor. Fiz., 30, 479 (1979) [JETP Lett., 30, 447 (1979)]; A. Zee, Phys. Rev. Lett., 42, 417 (1979); 44, 703 (1980).

[8] H. Georgi and S. L. Glashow; Phys. Rev. Lett. 33, 438 (1974). For a review see J. C. Pati's talk presented at the Einstein Centennial Symposium held in Jerusalem, March 1979. University of Maryland Preprint #80-006.

[9] W.-T. Ni, Bull. Am. Phys. Soc., 19, 655 (1974).

[10] W.-T. Ni, Phys. Rev. Lett. 38, 301 (1977).

[11] W.-T. Ni, A Nonmetric Theory of Gravity, preprint, Montana State University, Bozeman, Montana, USA (1973).

[12] See, e.g., F. G. Smith, *Pulsars* (Cambridge University Press, Cambridge, England, 1977).

[13] V. M. Hughes, H. G. Robinson, and V. Beltran-Lopez, Phys. Rev. Letters 4, 342 (1960); V. Beltran-Lopez, H. G. Robinson, and V. W. Hughes, Bull. Am. Phys. Soc. 6, 424 (1961).

[14] R. W. P. Drever, Philos. Mag. 6, 683 (1961).

[15] R. V. Eötvös, D. Pekar, and E. Fekete, Ann. Phys. (Leipzig) 68, 11 (1922); also R. V. Eötvös, Math. Naturwiss. Ber. Ungarn (Budapest) 8, 65 (1890).

[16] P. G. Roll, R. Krotkov, and R. H. Dicke, Ann. Phys. (N.Y.) 26, 442 (1964).

[17] V. B. Braginsky and V. I. Panov, Zh. Eksp. Teor. Fiz., 61, 873 (1971) [Sov. Phys. JETP 34, 463 (1972)].

[18] R. F. C. Vessot, and M. W. Levine, Gen. Relativ. Gravit. 10, 181 (1979).

[19] R. F. C. Vessot, *et al.*, Phys. Rev. Lett. 45, 2081 (1980).

[20] T. A. Morgan, and A. Peres, Phys. Rev. Lett., 9, 79 (1962); A. Peres, Phys. Rev. D 18, 2739 (1978).

[21] S. Hojman, M. Rosenbaum, M. P. Ryan, and L. C. Shepley, Phys. Rev. D 17, 3141 (1978).

[22] W.-T. Ni, Phys. Rev. D 19, 2260 (1979).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

An Inertial Clock to Test the Non-Metricity of Gravity

W. Stephen Cheung and Rogers C. Ritter

Department of Physics, University of Virginia, Charlottesville, VA 22901

A certain class of non-metric gravitation theories implies that clocks of different electromagnetic nature will interact with local gravity differently and result in slightly different gravitational redshift measurements. An ideal rotating body would have its moment of inertia, hence timekeeping, depend on the rotor dimensions, so that it would act as an "electrostatic clock." In the past, precision high speed rotations of small spheres have exhibited long decay times, $\sim 10^{10}$ s, and high Q. Simple extrapolation indicates that larger rotors should have much longer decay times. A double magnetic suspension of two rotors—an inner "proof" rotor and an outer "shroud" rotor—has been built and is under test as a method for reducing bearing and gas drags. This paper presents a description of a non-metricity test, analysis of the corotation scheme, and details of construction of the first, room-temperature inertial clock.

Key words: corotation; gravity; inertial clock; non-metricity; precision rotor.

1. Introduction

Except for the earth, precision macroscopic rotors have seldom been employed in experimental tests of gravitational theories. The Hughes-Drever experimental tests [1, 2] for the anisotropy of space, in fact, are one of the few microscopic rotor tests of the nature of gravity. One macroscopic rotor, the Stanford Relativity Gyro [3], is an example of a precision rotation but its gyro rather than pure rotor property is what is of interest. A second rotor used for gravitation, purported to test a gravitational radiation theory [5], was later found to be untenable [6].

As precision measurement devices macroscopic rotors suffer primarily from gas and bearing friction, in addition to the usual disturbances which constitute the ordinary hostility of our earthbound environment for sensitive measurements. If these problems can be solved sufficiently for a rotor to have a cosmological spin-down time, say 10^{18} sec, and if such a rotor can be interrogated appropriately, a useful new type of clock would result [7]. (Actually, the earth is such a clock but its spindown rate, $\sim 3 \times 10^{9}$ yr, is too great and too variable, $\sim 3 \times 10^{-8}$/day, for modern application in gravitation experiments.) Many applications might result from a clock whose stability came from a rotor of $Q \sim 10^{18}$; we discuss a particular one here.

A certain class of non-metric theories can be treated in the TH$\epsilon\mu$ formalism [8] to predict the non-uniqueness of the gravitational redshift as measured by ideal clocks of different electromagnetic nature [9]. The hydrogen maser and superconducting cavity stabilized oscillator (SCSO) have been compared for this purpose [10], but thus far with inconclusive results. The way in which the different electrical effects modify the gravitational redshift in the TH$\epsilon\mu$ formalism, independent of any particular gravitational theory, is given [11] below.

Here, the test might be made by comparing three clocks as they rotate around the earth's axis, moving a "vertical" distance h in the sun's potential from noon to midnight. In this case g would be the mean local field they experience from the sun, about 0.006 N/kg, so $gh/c^2 = 8 \times 10^{-13}$.

Two clocks could not in principle uniquely determine the parameters Γ_0 and Λ_0 from daily timekeeping differences in such an experiment. Furthermore, many ordinary local disturbances also have daily variations. Three different clocks, on the other hand, cannot only determine Γ_0 and Λ_0, but can give redundant information to test for these other effects. The fact that the sought-for variation is so small further amplifies the desirability for redundant measurement.

In this paper we discuss the conception, design, and construction of an inertial clock based on precision rotations with special attention to the unusual problems: the bearing friction and gas drag.

Group	Aspect of EM	Examples	Redshift in TH$\epsilon\mu$ Formalism
I	Electrostatic interactions	Principal lines of H Oscillations of high Q crystals Rotor clock	$Z = (1-2\Gamma_0)gh/c^2$
II	Hyperfine interactions	H-MAser	$Z = (1 - 3\Gamma_0 + \Lambda_0)gh/c^2$
III	Electrostatic + light propagation	SCSO	$Z = (1 - \frac{3}{2}\Gamma_0 - \frac{1}{2}\Lambda_0)gh/c^2$

2. The Rotational System

Beams [12], Fremery [4], and others have achieved long spindown times—10^{10} sec $\approx$ 300 yrs—using small (~3 mm dia) spheres ferromagnetically suspended. Magnetic bearing friction of a special kind and residual gas drag limited the decay time in these experiments.

The Keith Coriolis torque [13, 14] is an eddy-current drag on a ferromagnetic rotor, and has been tested by Fremery. In this the spin-spin coupling of a local rotor to the earth causes a continuous precession of the rotor in such a way that it experiences a hang-off angle (towards the north in the Northern Hemisphere) of a few microradians. This results in a compensatory eddy current in the rotor with a minute component causing drag. This is, surprisingly, almost independent of angular velocity. We have been unable, thus far, to make low-drag ferrite rotors which might be expected to alleviate this problem.

Superconducting rotors, on the other hand, should ideally be free of such an effect. (In addition, the shielding and low residual flux attainable makes a low temperature experiment the likely candidate for the ultimate inertial clock.) Holdeman [15] has estimated the drag torque for the superconducting Stanford Relativity Gyro experiment [3] at $\omega \approx 1200$ rad/s. He arrives at $\Delta\omega/\omega \approx 1.8 \times 10^{-8}$ as the worst case for a 6 month mission, which would be equivalent to a decay time $\tau^* \gtrsim 10^{17}$ sec.

Beams [12] has shown that in the free-molecule regime, gas drag on a rotor accurately obeys expected gas dynamic estimates. If an otherwise free rotor obeys the equation,

$$I\ddot{\theta} + \kappa\dot{\theta} = 0 \ , \tag{1}$$

where I is the moment of inertia and $\kappa = \kappa_g + \kappa_b$, the drag coefficient composed of gas and bearing components, we can write [7, 16] for a spherical rotor of radius a,

$$\kappa_g \approx \frac{2}{3} a^4 \hat{k} \sqrt{\pi \mu k T} \tag{2}$$

here $\hat{k}$ is the concentration of gas molecules of mass μ at temperature T.

For rotors ~ 10 cm diameter this would imply that a pressure $\lesssim 10^{-17}$ Torr would be needed to achieve spindown times $\tau^* > 10^{18}$ sec, a virtual impossibility.

Our solution has been to develop an old idea of Beams [17]. In this, the gas is moved along with the rotor by corotation of an outer cylinder. If collective gas effects are eliminated by operation in the free molecule regime, a huge reduction in gas drag could be expected, since the drag torque is proportional to the angular velocity of the rotor *relative* to its surrounding gas.

By this same reasoning (it can be shown formally) the outer cylinder exerts virtually zero torque on the inner through the gas medium as long as they rotate nearly together.

Thus the corotation method should virtually decouple the inner rotor from any gas drag. We have developed two versions of this: (1) Corotation Method I in which the outer cylinder is driven by a constant speed motor locked to a cesium-beam clock [18]; and (2) Corotation Method II in which the outer "shroud" cylinder is servoed to rotate with the inner "proof" rotor. It is this second method which is used as the basis for the inertial clock.

In order to give some feeling for the corotation properties, we analyze a freely rotating corotation system. In this system it is useful [7] to define an angle of departure of a fixed point on the inner rotor from uniform rotation at angular velocity ω_0 in the laboratory system:

$$\Delta = \theta - \omega_0 t \ . \tag{3}$$

With appropriate substitution into Eq. (1), plus the addition of a rotating restoring torque $c\Delta$ and a driving torque $T(t)$, this yields

$$I\ddot{\Delta} + \alpha_g\dot{\Delta} + \alpha_b\dot{\theta} + c\Delta = T(t) \ , \tag{4}$$

where $\kappa\dot{\theta}$ of Eq. (1) is replaced by $\alpha_g\dot{\Delta} + \alpha_b\dot{\theta}$ to account for its effect in the two different (laboratory and rotating) reference frames. (We have shown [19] that the support pole piece, hence the bearing, cannot, under ordinary circumstances, rotate as it develops a "cranking torque" which locks the rotor to the outer cylinder.)

For a free, undisturbed rotor, $c = T(t) = 0$, so that

$$I\ddot{\Delta} + (\alpha_g + \alpha_b)\dot{\Delta} = -\alpha_b\omega_0 \ . \tag{5}$$

If we define $\Omega = \dot{\Delta}$ and assume $\alpha_b\omega_0$ to be constant, the solution to Eq. (5) can be written

$$\Omega = \left[\frac{\alpha_b}{\alpha_g + \alpha_b}\right]\omega_0(e^{-t/\tau^*} - 1) \ , \tag{6}$$

where $\tau^* = I/(\alpha_g + \alpha_b)$ is the decay time due to all drag coefficients. From this we get the lag angle

$$-\Delta = \left[\frac{\alpha_b}{\alpha_g + \alpha_b}\right]\omega_0\tau^*(e^{-t/\tau^*} + t/\tau^* - 1) \ , \tag{7}$$

if $\Delta = 0$ at $t = 0$. For a measuring time $t \ll \tau^*$, this is approximately

$$-\Delta \approx \frac{1}{2}\left[\frac{\alpha_b}{I}\right]\omega_0 t^2 \ . \tag{8}$$

Thus, the free, passive, corotation system develops a phase lag $-\Delta$ which is quadratic in time but has negligible dependence on the gas friction. Analysis of the servo behavior which is important to the clock properties is beyond the intent of this paper; and these properties are presently under study.

3. Corotation and the Double Magnetic Suspension

In another experiment [18], which employs Corotation Method I, it is feasible to use an air bearing for the outer, constantly driven cylinder. In our experience, however, such a bearing is quite rough on the scale of disturbance considered here, and would constitute a considerable complication in the servo loop for the Method II experiment of the inertial clock. Consequently we have chosen to use the double magnetic suspension of Beams [17].

Beams had two types of double magnetic suspension. The inertial clock uses a variation of the second type, in which servo control of the position of the lower (inner) rotor is by means of a coil from below, thus affording greater independence of the two rotors. We have studied several topologies of the combination, including those in which the lower field is biassed to "push" or to "pull" the inner rotor. The latter, which Beams used, is more stable, but is more vulnerable to the cranking effect mentioned above. Ultimately a diamagnetic suspension, at least for the lower rotor, would seem to be most desirable in avoiding both the cranking effect and other inter-rotor vertical coupling. The reason for such attention to vertical coupling is the fact that our studies [16] have in-

dicated the immense difficulty of avoiding vertical-rotary mode coupling, which would, in effect, "shunt" the inner-shroud rotor servo loop.

In order to reduce drifts which had limited the earlier experiments of Beams and Fremery, the present rotors use low angular velocity, thus avoiding centrifugal stress. Typically 0.1 to 10 Hz are the rotation rates we use. Inter-rotor coupling via magnetic flux drive is also reduced.

4. The Hybrid Double Magnetic Suspension

The transducers in our current double suspension are different in nature, i.e., hybrid. A schematic description of this scheme is in Fig. 1. The vertical position of the upper rotor, R_1, is optically sensed while the lower rotor, R_2, is sensed by a sensing coil. Optical sensing is not employed in the lower suspension to avoid another possible type of coupling between the angular velocity of R_1 and the lower servo loop due to optical reflection and scattering from the attached shroud.

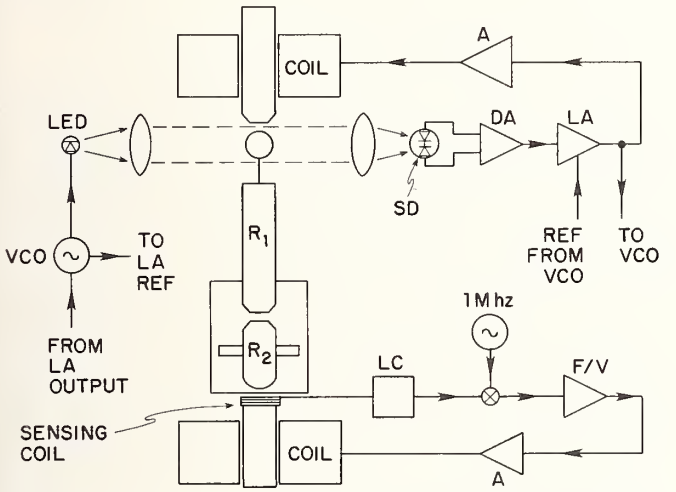


FIGURE 1. *Diagram of the hybrid double magnetic suspension. The upper rotor, R_1, consists of the cold rolled steel ball, the cylindrical magnet, and the shroud. The lower rotor, R_2, is typically made of ferrite and is suspended inside the shroud with the main lifting force provided by the permanent magnet. R_1 is optically sensed while R_2 is sensed by a coil. The labels are: SD = split-photodiode; DA = difference amplifier; LA = lock-in amplifier; LC = LC tank circuit; F/V = frequency to voltage converter; VCO = voltage controlled oscillator; A = typical suspension amplifier-differentiator.*

A parametric optical system is used [20]. In this the motional information of R_1 amplitude-modulates the carrier frequency, typically 10 kHz, at which the light source is pulsed. The light source is phase-locked with the subsequent signal demodulator through the voltage controlled oscillator (VCO). This parametric method of suspension has demonstrated significant noise reduction and disturbance immunity over the traditional optical sensor (dc light source). The 10 kHz light source is focused to give a parallel beam which passes evenly above and below the cold rolled steel ball (5/8" diameter) forming a symmetrical image on a split photodetector. With careful mechanical and electronics design, the electronic preamplifiers at the detector stage are balanced and hence further stability is achieved. Preliminary measurements show that the temporal drift, for example, of a rotor so suspended is less than 2×10^{-3} cm/°C. This drift is about a hundred times better than a similar drift measurement using a single photodetector and a dc light source. Other disturbances such as ambient light fluctuation and electromagnetic pickup have practically no effect on this balanced parametric method [20].

The coil sensor for the lower suspension operates on the principle of frequency modulation. The control motions of R_2 modulate the resonant frequency (as well as the amplitude) of the ~1 MHz tank circuit. These FM signals are mixed with a stable 1 MHz standard and the subsequent difference frequency is then converted into voltage. The amplitude changes are ignored by means of digital circuitry. This FM sensor is discussed further by Bernard *et al.* [21]. A ferrite rotor suspended by this means is shown in Fig. 2.

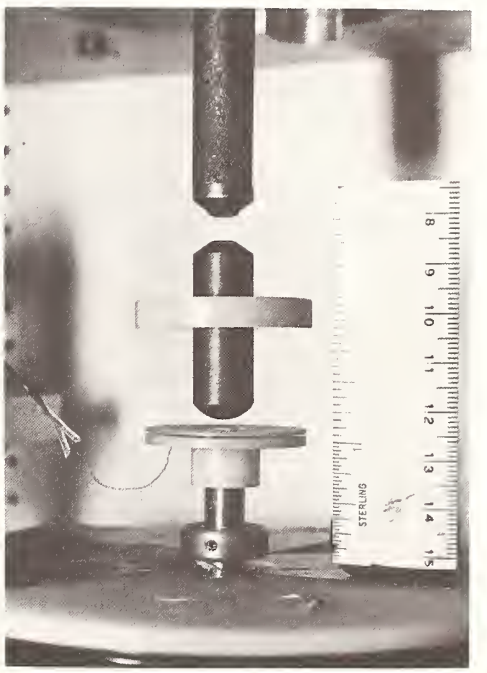


FIGURE 2. *A ferrite rotor in suspension. The magnet above provides the main lifting force. The vertical motion of the rotor is picked up by the sensing coil below it. The corrective magnetic fields come from the bigger coil below the rotor.*

5. Present Apparatus

The first version of the inertial clock experiment is at room temperature and is designed mainly to test a variety of basic corotation questions, including feedback. It is shown in moderate detail in Fig. 3. The aluminum assembly sits on a vibration isolation table.

The current-controlled pulsing LED and its detecting circuitry are housed separately in two aluminum tubings which are mechanically attached to the main aluminum chamber. The windows are coated with a thin, conducting coating of tin oxide and provide vacuum sealing.

The permanent magnet provides the main lifting force [22] for the rotor R_1. Thus the coil has only small, corrective currents and low heating. The shroud attached to R_1 has four baffle-like plates arranged so that gas molecules must follow multiple straight paths in order to get into or out of the interior of the shroud.

R_1 is driven to follow R_2 by means of two pairs of eddy-current drive coils. One pair has shoulders to act as "catchers" of R_1 if it falls. The second pair has hollow interiors to house a dc light source and a split photodetector, respectively. A zero-crossing signal results

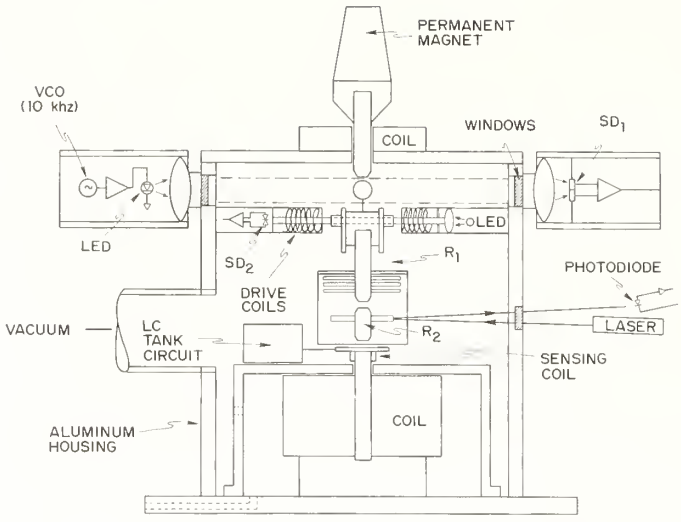


FIGURE 3. *Cross-sectional view of the present apparatus. R_1 and R_2 are doubly suspended in a hybrid fashion. Optical schemes of measuring the angular periods of the two rotors are also shown. SD_1 and SD_2 are split-photodiodes employed for the upper suspension and for measurement of the angular period of R_1, respectively.*

whenever the hole drilled in R_1 aligns with this optical path. The zero-crossing signals provide measurement of the angular period of R_1.

A simple scheme is adopted for the present apparatus to measure the angular period of R_2. A flat, shiny polished surface is machined on the teflon ringbelt attached to R_2. A laser beam is aimed at and reflected from this surface when it coasts by and is then detected by a photodetector. Modest performance of both R_1 and R_2 angle-sensors is sufficient for present experiments.

The vacuum level required here is not stringent as long as it is the molecular flow regime. At present, 10^{-5} Torr is acquired with a diffusion pump.

An overall picture of the present apparatus is shown in Fig. 4.

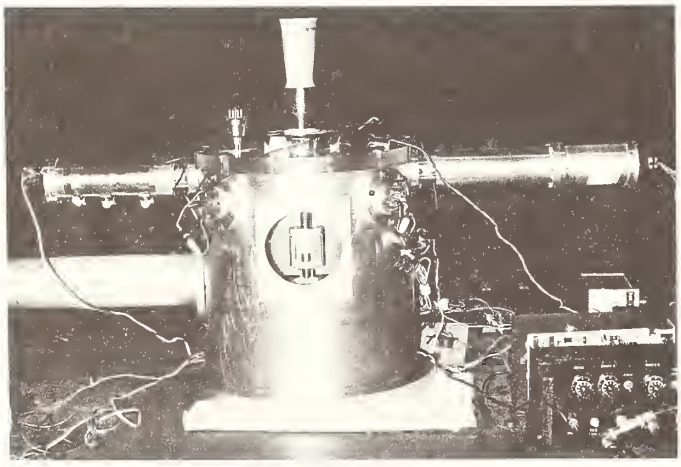


FIGURE 4. *Overall view of the present apparatus. Attached to the lower left of the aluminum housing is the vacuum port. The aluminum tubings to both sides of the housing house the pulsing light source and the detecting circuitry, respectively. Above the lid, one can also see the permanent magnet, the coil, and the vacuum gauges.*

6. Period Counting and the Angular Servomechanism

A challenging problem to an ultra high-Q rotor is the long experimental time needed to observe any significant period change.

It can be seen that if stability of 10^{-15} is finally achieved, and is to be used to measure diurnal variations at this level, the interrogation problem of a slow rotor is indeed significant.

An optical method for these relatively large period measurements suffers white phase noise, n_p, as well as quantization noise, n_q, of any electronic counting system. In ordinary averaging over time, t, n_p, and n_q go down as t^{-1}. An interesting algorithm suggested by Snyder [23], a sort of "statistical vernier," causes n_p and n_q to fall by $t^{-1.5}$. A microcomputer has been chosen to coordinate the period measurements and the angular control. The Snyder algorithm can be implemented in the microcomputer. When a difference between the periods of R_1 and R_2 is detected, the microcomputer gates open the eddy-current drive to drive R_1 to follow R_2 with the appropriate corrective phase.

7. Known Future Directions

Ferromagnetic suspension has exhibited [16] a tendency to couple vertical variations into rotary motion via parametric interactions. This might be anticipated, at some level, for any active suspension method. Hence, a particular point of diamagnetic, superconducting suspension would be to eliminate an active control. (It would, in addition, eliminate the Coriolis torque losses.)

An eddy-current drive, used thus far for convenience, introduces rotating magnetic flux in undesirable places, giving rise to stray inter-rotor coupling. The small torque needed to maintain (not start) the shroud rotor angular velocity against gas drag, is about 10^{-3} dyne-cm. This could be supplied by very modest laser power in a photon drive [7, 24]. The attendant intrinsic problems, discussed by Braginsky [25, 7], seem readily surmountable, although practical aspects have not been sufficiently tested.

Observation of the rotors will most likely require the Snyder algorithm. Even so, the subdivision of $\sim 10^5$ of these slow rotations in one day into $> 10^{15}$ parts is a formidable question, likely to take us into the technology of laser spectroscopy.

Hosts of disturbances of the standard varieties which plague many other precision measurements will likely cause the appearance of this experiment to change radically. The present apparatus will serve primarily to introduce us into the new problems of Method II Corotation. If our experience with Method I is an example, these will be many and interesting.

We gratefully acknowledge support by the U.S. National Bureau of Standards under grant G8-9025, the National Science Foundation under grants PHY78-3208 and PHY 80-07948, and the University of Virginia.

References

[1] V. W. Hughes, H. G. Robinson, and V. Beltran-Lopez, Phys. Rev. Lett. 4, 342 (1960).
[2] R. Drever, Philos. Mag. 6, 683 (1961).
[3] C. W. F. Everitt, in *Proc. First Marcel Grossman Conference on General Relativity*, Trieste, 1975, Ed. by R. Rufini (North Holland, Amsterdam, 1977), p. 545.

[4] J. K. Fremerey, Phys. Rev. Lett. 30, 753 (1973).
[5] J. C. Keith, J. Math. Phys. (N.Y.) 42, 248 (1973); J. C. Keith, Rev. Mex. Fiz. 12, 1 (1963); G. D. Birkoff, Proc. Nat. Acad. Sci. U.S.A. 30, No. 10, 324 (1944).
[6] C. M. Will, in *Experimental Gravitation*, Proc. Int'l School of Physics "Enrico Fermi," course LVI, Ed. by B. Bertotti (Academic Press, New York, 1974), p. 1.
[7] R. C. Ritter, in *Proc. Second Marcel Grossman Conference on General Relativity*, Trieste 1979, Ed. by R. Ruffini (North Holland, 1982), p. 1039.
[8] C. M. Will, Metrologia 13, 95 (1977).
[9] C. M. Will, Phys. Rev. D 10, 2330 (1974).
[10] S. R. Stein and J. P. Turneaure, Proc. IEEE 63, 1249 (1975); S. R. Stein and J. P. Turneaure, in *Future Trends in Superconductive Electronics*, Ed. by B. S. Deaver *et al.* (Am. Inst. Phys. Conf. Proc. No. 44, New York, 1978), p. 192.
[11] C. M. Will, private communication.
[12] J. W. Beams, D. M. Spitzer, and J. P. Wade, Rev. Sci. Instrum. 33, 151 (1962).
[13] J. C. Keith, J. Res. Natl. Bur. Stand. Sect. D: 67, 533 (1963).
[14] J. K. Fremerey, Rev. Sci. Instrum. 43, 1413 (1972).
[15] L. B. Holdeman and J. T. Holdeman, J. Appl. Phys. 47, 4936 (1976).
[16] G. R. Jones, Jr., "Decay Time Studies of Magnetically Suspended Precision Rotors," M. Sc. Thesis, University of Virginia, 1981.
[17] J. W. Beams, Rev. Sci. Instrum. 24, 1071 (1963).
[18] G. T. Gillies and R. C. Ritter, these proceedings; R. C. Ritter *et al.*, Nature (London) 271, 228 (1978).
[19] G. T. Gillies, "The Development of High Q, Precision Mechanical Rotors for Use in a Dynamic Measurement of Matter Creation," Ph.D. Dissertation, Univ. of Virginia (1980).
[20] W. S. Cheung *et al.*, Precis. Eng. 2, 183 (1980).
[21] B. E. Bernard *et al.*, Rev. Sci. Instrum., to be published.
[22] W. S. Cheung *et al.*, Am. J. Phys., to be published.
[23] J. J. Snyder, preprint and private communication.
[24] J. W. Beams, Phys. Rev. 72, 987 (1947).
[25] V. B. Braginsky and A. B. Manukin, *Measurement of Weak Forces in Physics Experiments*, Ed. by D. H. Douglass (Univ. of Chicago Press, Chicago, 1977).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Assessment of the Prospects for a Measurement of Relativistic Frame Dragging by 1990*

Richard A. Van Patten**

Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305

Relativistic frame dragging, i.e., the Schiff motional effect or the Lense-Thirring effect, has never been measured. Just two experiments proposed to date promise measurement accuracies of a few percent. The two are the gyro experiment and the twin satellite experiment. These two experiments, although very different, have a common element, a "gyroscope" in a rotating gravity field. With the gyro experiment, a small cryogenic gyro is placed in orbit and its spin axis history is compared with a star to 10^{-3} arc sec/yr accuracy. With the twin satellite experiment, two drag-free satellites are placed in reverse polar orbits and used as orbit-sized gyros. The sum of the nodal motion is checked to 30 cm against that predicted from earth oblateness using polar satellite-to-satellite Doppler data. This paper compares the errors and risks involved in carrying out each of these very demanding scientific space missions.

Key words: gyro experiment; relativistic frame dragging; twin satellite experiment.

1. Introduction for Gyro Experiment

In 1960, Leonard Schiff [1] predicted, using Einstein's general theory of relativity, that a gyroscope in orbit about the earth would experience a precession of its spin axis relative to the "fixed stars." Two relativistic precessions are predicted: a "geodetic" precession associated with the orbital motion of the gyro about the earth and a "motional" precession due to the earth's rotation. Fig. 1 shows the two effects for a gyro in a 520 km altitude polar orbit with its spin axis in the equatorial and orbital planes. The geodetic precession is north approximately 6.9 arc sec/yr and the motional precession is east approximately $0.05[(1 + \gamma)/2]$ arc sec/yr, where the parametrized post-Newtonian (PPN) parameter γ is near unity.

The Stanford Relativity Gyroscope Experiment, as presently envisioned [3], will use four superconducting electrostatically suspended spherical gyros carried in a drag-free earth satellite with spin axes nominally aligned with Rigel. Each spin axis orientation with respect to the line of sight to Rigel is measured using a SQUID magnetometer to compare the direction of the London magnetic moment with the readout of a telescope pointed nominally at Rigel, to ± 0.05 arc sec. The satellite is rolling about the line of sight to Rigel with a period of 10 minutes. By rolling about the line of sight the body fixed drift errors are eliminated leaving only inertially fixed quantities such as gyro drift and relativistic precessions. Quantities such as solar heating on the sunlit side of the rolling satellite and the earth's magnetic field have the potential of looking like inertially fixed drift errors and thus must be greatly attenuated (by 10^{-13} for the latter) in the vicinity of the gyros and telescope. A readout accuracy of 10^{-3} arc sec or 2 percent of the motional effect over one year is to be attained for any 70 hr estimating period. The uncertainty in the proper motion of Rigel turns out to be the dominant error source at about 4% of the motional effect [6].

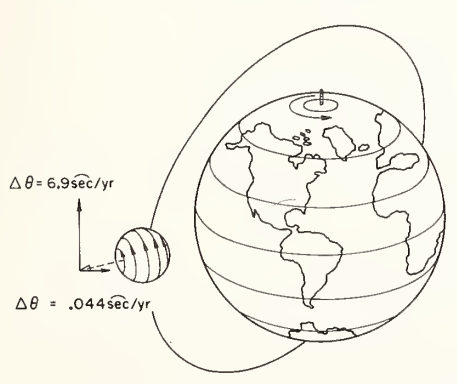


FIGURE 1. *Relativistic precessions: Geodetic and motional.*

2. Introduction for Twin Satellite Experiment

In 1974, the author and C. W. F. Everitt proposed a new test of the general theory of relativity [4, 5]. Using general relativity, Lense and Thirring predicted that the plane of a body orbiting a rotating mass is dragged a small amount in the direction of rotation of the mass. For a satellite orbit 800 km above the earth, this relativistic motion is an angular rate of $0.2[(1 + \gamma)/2]$ arc sec/yr or four times the Schiff motional effect for a gyro in polar orbit at the same altitude. This amounts to 31 m total eastward dragging for two orbits in 2.5 years.

At first it was thought that a single drag-free satellite in polar orbit might be sufficient for making the test. But the orbit plane position history relative to the poles would have to be known to centimeters to reduce the nonrelativistic orbit plane precession uncertainty to a level below that predicted by general relativity.

The two counter-orbiting (see Fig. 2) satellite experiment has significant advantages. Were it possible to

*Supported by NASA Contract NAS8–32355.

**Associate Investigator, "A Gyro Test of General Relativity in a Satellite".

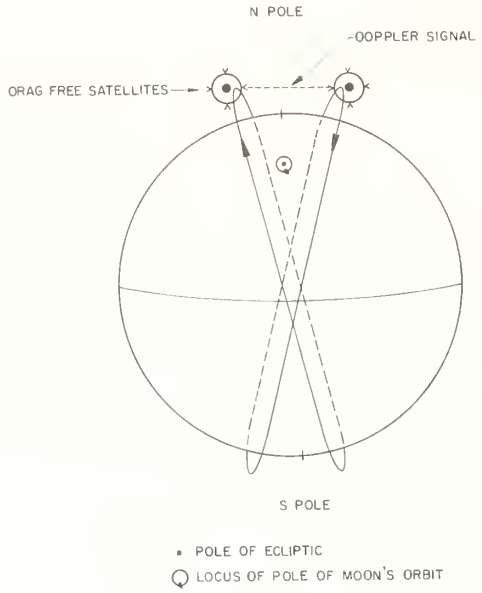


FIGURE 2. *Experimental configuration.*

place the two satellites in identical orbits (even if not polar) their planes would precess due to the non-relativistic effect of the earth's oblateness, in opposite directions at the same rate; i.e., the sum of the non-relativistic nodal precession rates would be zero. It is not possible to place the satellites precisely so, but now the non-zero sum of the non-relativistic nodal precession rates is a multiple (related to earth oblateness and known to 4 significant figures) of the mean plane separation which is very accurately deduced (to 0.2 cm accuracy) from satellite-to-satellite Doppler ranging measurements as they pass at or near the poles [7]. Subtracting the mean orbit plane precession due to earth oblateness (obtained by measuring mean plane separation) from the actual mean orbit plane precession (measured from ground tracking) a measurement of the Lense-Thirring effect is obtained to an accuracy of 1 to 3 percent for a 2.5 year experiment [8].

3. Errors in the Gyro Experiment

The three major error sources in the gyro experiment are:

(1) Precessions due to suspension system torques acting on a rotor which is not perfectly spherical or torques due to the center of mass being offset from the center of support.
(2) Noise and scale factor error associated with the SQUID magnetometer measurement of the gyro spin axis (the London moment vector) and noise in the measurement of roll angle using a side-looking star monitor (star blipper).
(3) Uncertainty in the proper motion of the guide star, Rigel.

Using a realistic measurement model, a Kalman filter covariance analysis was performed by R. Vassar and J. V. Breakwell [6] to analyze the effect of these errors on the determination from a single gyro of the geodetic and motional effect for a 1-year mission. Different *a priori* uncertainties in the proper motion of Rigel were assumed in three studies as follows:

(1) The first study assumed the proper motion of Rigel was known exactly.
(2) The second study using the latest astronometric data assumed the north-south proper motion and east-west proper motion are uncertain by 0.9 and 1.7 milliarc-sec/yr respectively [9].
(3) The third study assumed the error in Rigel's proper motion is a factor of 10 larger than the uncertainties computed from the accepted astronometric data and that an off-polar orbit could be used along with the likely eventuality that the measurements will show the geodetic term to be very nearly correct.

Table 1 summarizes the results for the three studies showing the expected error in the motional (frame dragging) term for the best orbit inclination under each assumption about proper motion [6].

TABLE 1. *Expected error in motional term.*

Proper Motion Uncertainty (milliarc sec/yr)	Orbit Inclination i	Month Launch	Expected Error in Frame Dragging (milliarc sec/yr)
No uncertainty	85.5°	Sept.	0.49 (1.1%)
NS = 0.9, EW = 1.7	85.5°	Sept.	1.79 (4.1%)
NS = 9.0*	70.5°	Oct.	2.67 (6.1%)

*Geodetic precession assumed equal to theoretical value.

Fig. 3 shows the expected error in the motional precession measurement with the accepted values of proper motion uncertainty for all inclinations from 67° to 90°.

Both Table 1 and Fig. 3 are based on a factor of 10 improvement in gyro readout noise relative to the existing laboratory readout using well established, planned design changes. This noise has a $1/f$ power spectral density. Thus a higher roll rate would decrease the effective noise. The roll rate is limited however by spacecraft mass properties in combination with the available pointing control authority to about 0.01 radians per second.

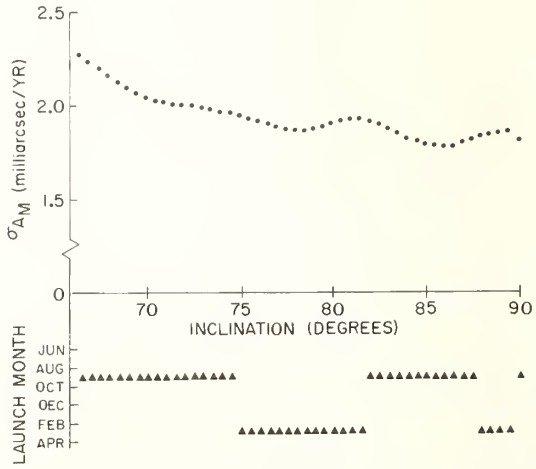


FIGURE 3. *Expected error in the motional precession with typical proper motion uncertainties.*

4. Errors in the Twin Satellite Experiment

Table 2 shows the error budget for the twin satellite experiment. As an aid to understanding Table 2 refer first to Fig. 4 and then to Fig. 5. Fig. 4a shows the form of the lateral polar plane separation (as deduced from the polar slant range Doppler data by Kalman filter). Fig. 4b shows the form of the sum of the nodal motions due to earth oblateness as calculated from the data of Fig. 4a and as measured from ground tracking. The non-relativistic no-

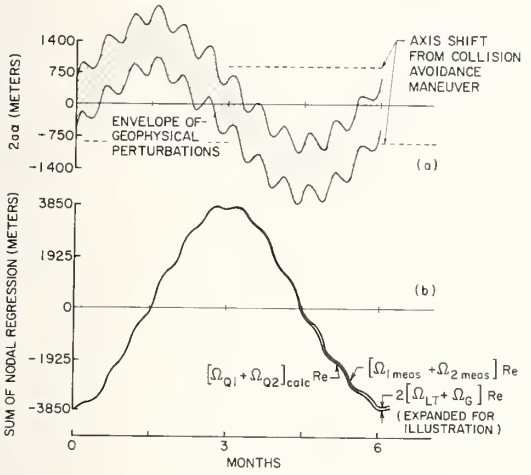


FIGURE 4a. *Form of lateral polar data.*

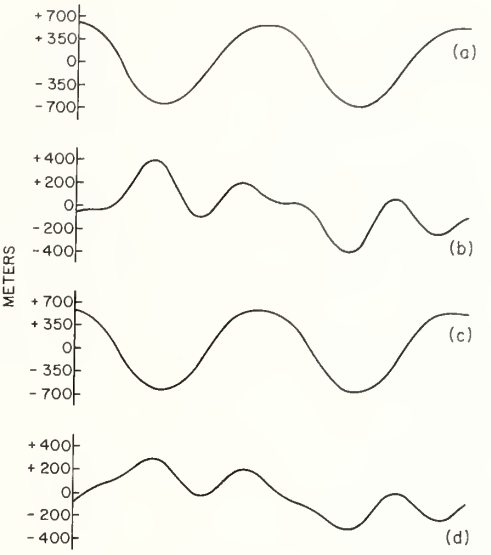


FIGURE 5. *Lateral and vertical polar perturbations over one sidereal day, as calculated from two different geophysical models. SAO 1973 Model (a) Lateral (b) Vertical, NOAA 1974 Model (c) Lateral (d) Vertical.*

TABLE 2. *Twin satellite experiment error budget.*

Error source	Equivalent Nodal Shift at Equator in 2.5 years (meters)
1. Error in measurement of mean orbit plane separation at the poles.	0.05
2. Error in determination of nodal angle history of satellite No. 1 (from ground tracking).	0.11
Same as above for satellite No. 2.	0.11
3. Error due to uncertainty of the Earth's quadrupole mass moment.	0.09
4. Error in nodal position caused by orbital adjustments performed at the equator.	0.08
5. Error in nodal position caused by fluctuations in self-gravity or mass attraction of satellite No. 1 on its proof mass.	0.04
Same as above for satellite No. 2.	0.04
Composite without time base error (root-sum-square)	0.21

Time Base Assumption	Optical UT 1(BIH)	Laser (LURE)	VLBI (POLARIS)
Equivalent time base error	0.5	0.01	0.05
Equivalent time base error allowing for correlation present with errors 2 above.	1.0	0.02	0.1
Composite with time base error	1.02	0.211	0.233
Combined relativistic effect	30.8	30.8	30.8
Probable experimental error (1 σ)	3.3%	0.69%	0.76%

dal plane motions are removed by subtracting these two results leaving only the relativistic plane motions. The total relativistic nodal plane motion over 2.5 years is made up of two components; 27.8 m due to the Lense-Thirring effect and 3.0 m geodetic precession due to the motion about the sun. The total relativistic nodal plane motion for both satellites is thus 30.8 m in 2.5 years. The envelope of geophysical perturbations shown in Fig. 4a, in addition to the solar (twice yearly) and lunar (twice monthly) gravity gradient perturbations, are shown more clearly in Fig. 5a, b, c and d. These meter per day vertical and lateral relative perturbations result from various combinations of earth gravity harmonics J_{lm}.

Error No. 1 in Table 2 has been determined from the results of a covariance analysis due to D. Schaechter and J. V. Breakwell [7], see Fig. 6, using a model with about 200 parameters for the polar vertical and lateral perturbations. The Doppler range error used for the analysis for a single polar encounter was 1.0 cm which is applicable for

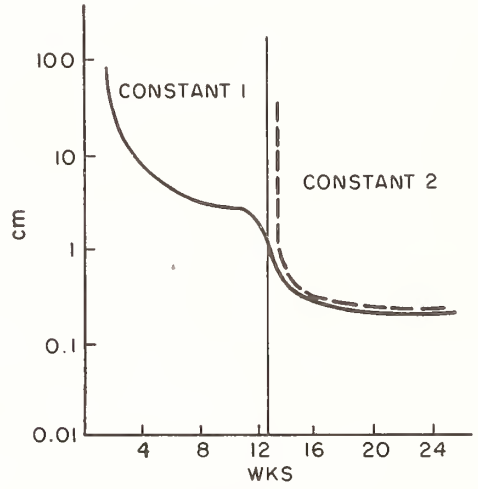


FIGURE 6. *Covariance evolution of polar mean plane separation parameters (note orbit adjustment at 13 weeks).*

distances up to 2 km assuming an oven stabilized crystal oscillator. Errors No. 2 through No. 5 are described in detail in Ref. [4]. The dominant error for this experiment depends on the time base assumption. The time base error relates directly to the uncertainty in earth rotation with respect to the inertial frame. This is needed to relate tracking data to the inertial frame. The equivalent time base errors shown in Table 2 reflect the accepted uncertainties available or attainable for each of the three sources indicated. The optical UT 1 time base is available while the laser and VLBI time bases may be available at the time the experiment could fly. Thus the expected accuracy for the twin satellite experiment is one to three percent depending on the time base assumption.

5. Risks Associated with the Gyro Experiment

The primary technical risk areas affecting our ability to successfully complete precision measurement space experiments such as the gyro are, of course, (1) accuracy performance and (2) reliability. The accuracy of the gyro experiment has received extensive attention in the research reports associated with the Stanford Gyro Relativity Project [6, 10, 11]. The reliability needed for obtaining an adequate probability of a successful flight is less well established and is also strongly coupled to accuracy since accuracy degradation is capable of destroying the scientific value of the mission even with a flight completely free of hard failures. A hard failure is any condition which prevents obtainment of additional valid science data from which recovery is not possible or feasible. The cryogenic gyro experiment is extremely complex and demanding in terms of being at or near the state of the art in a number of high technology areas. There are significant risks relating both to accuracy degradation and hard failures which will be listed. It is important to note that MTBF (mean time before failure) numbers relating to system reliability are not available at this time since flight hardware and redundancy are not yet fully defined. The objective of the discussion is to give the reader an idea of some of the numbers and conditions required for attaining the stated accuracy.

The quartz block assembly which contains 4 quartz gyros, 1 drag-free instrument and one quartz telescope along with associated plumbing, electrical, magnetic shielding and vacuum equipment operates inside a helium dewar at liquid helium temperatures. The gyro assemblies require a vacuum of 10^{-9} Torr to limit torques and spin down rates. The residual magnetic field in the vicinity of the gyros should be $\sim 10^{-8}$ to limit the trapped flux pinned to the superconducting niobium coatings on the gyro rotors at the time the coating is cooled below its critical temperature. Excessive trapped flux would saturate the SQUID magnetometer gyro readout. The earth's magnetic field must be attenuated by 10^{-13} at the gyro as the spacecraft rolls. This is accomplished by proper design of the lead bag supconducting shield system in conjunction with the gyro mounting cavities which also serve as superconducting shields. Magnetic properties of all parts inside the lead bag must be rigidly controlled, with magnetic inclusions eliminated to a level consistent with the 10^{-8} gauss design requirement. Thermal modulation of the SQUID magnetometer at the roll frequency due to solar heating on the exterior of the spacecraft as the 10-minute roll takes place should not exceed ten micro degrees K. This lateral heat flux requirement is accomplished with super insulation already required by the 1-year dewar hold time needed. Thermal heating through the telescope aperture from the sun and earth albedo must be similarly limited. A combination of gold coated windows, sun shade, and shutter arrangement is being studied as a solution.

The quartz gyro rotor must be spherical to 0.4 to 1.0 microinches (1 μ in. $\simeq$ 0.025 μm) with material homogeneity of 3 parts in 10^7 or better. The thin or thick film (4 or 100 μ in.) superconducting niobium coating on the rotor must be mass balanced well enough to be consistent with the above. The rotor coating should also be sufficiently robust to permit gyro rotor levitation at 1 g[1] for ground tests with an electrostatic suspension system. The rotor diameter may be decreased from 1.5 in. to say 0.75 in. to help attain the above. The gyro rotor will be spun up to 170 Hz with a gas spin-up system which requires spin-up channels to be located within 300 to 500 μ in. of the rotor surface. The flight suspension system must keep the rotor centered within ± 50 μ in. while supporting against acceleration forces of ± 1 g during the launch tests or $\pm 5 \times 10^{-8}$ g during flight. The support electrodes are located 1500 μ in. away from the

[1] In this context the symbol "g" refers to a unit of acceleration equal to that of gravity at the surface of the earth.

rotor. The suspension preload levels must be decreased to a low multiple of the gravity gradient acceleration from the earth which will be a twice per orbit sine function at about 5×10^{-8} g. The suspension system bandwidth must decrease from 4000 rad/s at 1 g to 4 rad/s during normal flight to reduce sensor noise acceleration to acceptable levels. During any emergency, such as micrometeor impacts or inadvertent handling or docking events, the suspension system will switch up as required in time to prevent the rotor from contacting the spin-up channels since any contact would cause severe damage to the gyro.

The 5 1/2-inch aperture telescope assembly is optically contacted together and must produce image linearity and smoothness (but not absolute null stability) of better than 1 milliarc sec over the central ± 0.05 arc sec in pitch and yaw. The attitude control system must point the telescope at Rigel within ± 0.05 arc sec including a ± 0.03 arc sec sinusoidal dither command signal with 1-minute period. This is required to measure and control the telescope photomultiplier readout scale factor to within ± 1 percent of that of the gyro. One hundred hours of smoothing time is required to reduce the scale factor mismatch by $1/e$ without permitting the SQUID noise to disturb the corrected scale factor by more than 1 percent rms. The above process permits the ± 50 milliarc sec attitude motions to be substracted from gyro to Rigel difference angle data to the level of 0.5 milliarc sec or better.

The drag-free instrument is located at the center of mass of the spacecraft to ± 1 mm and will be designed to maintain the proof mass drag free to $\pm 10^{-10}$ g. This should be compared with the level of 5×10^{-12} g attained by the TRIAD discos drag-free satellite launched in September 1972.

Both the drag-free and attitude control systems will use the helium boil-off gas in conjunction with proportional differential thrusters [11] for attainment of smooth, but adequately responsive, control. The attitude control system may employ inner pointing actuators for fine pointing of the quartz telescope. The spacecraft will be designed with nearly equal moments of inertia to minimize the gravity gradient torques, and minimum cross products of inertia to reduce pointing control authority requirements in the presence of the 10-minute period roll.

During initial spin-up with helium gas, steps must be taken to assure that the spin axes end up being within 10 to 30 arc sec of Rigel to assure functioning of the high resolution science data instrumentation systems.

The risk areas viewed as most important by the author are listed in order of perceived importance at this writing in Table 3.

6. Risks Associated with the Twin Satellite Experiment

This experiment has received much less attention than the gyro and therefore less time has been applied to both discovering and solving problems. However, during the three year period, 1975–1978, during which the experiment was actively studied a reasonably clear picture was formed of the major areas of risk associated with a flight program involving this mission.

To give the reader an understanding of the risks involved in conducting the twin satellite experiment it is necessary first to understand the nature of the drag-free orbits required. The orbit requirements for performing the mission are very strict, with a constraint on each of the six relative orbital elements [12].

(1) *Semi-major axes (a).* The semi-major axes must lie within 10 cm of each other, since otherwise the difference in orbital periods will cause the satellite encounter point to drift from the poles. There are several difficulties with these lower latitude encounter points. One is that a small component of the polar mean plane separation is lost. A second problem is that at lower latitudes the range measurement picks up small components of unmodeled fluctuations whose dominant effect is at the equator. Some work has been performed [13] in which the encounter point is purposely allowed to drift from the poles, so that additional geophysical information is obtained. However, the modeling of the earth gravity harmonic fluctuations at arbitrary latitudes (not just $\pm 90°$) is much more difficult. The normal altitude of both satellites is 800 km.

(2, 3) *Eccentricity* (ξ, η). A relative eccentricity between the two satellites produces relative altitude fluctuation at the poles whose magnitude is proportional to the relative eccentricity and whose period is equal to the perigee regression rate produced by the earth's oblateness. The separation of the mean orbit planes is determined using a Doppler ranging device which measures slant range at closest approach. The introduction of any relative vertical separation into this range measurement will somewhat degrade the information on mean plane separation. Also, the fluctuation model is greatly simplified by requiring the eccentricity to be nearly zero. Therefore, the two components of the eccentricity vector ($\xi = e \cos \omega$ and $\eta = e \sin \omega$) for each satellite should be trimmed to produce relative altitude fluctuations no greater than 300 m over the poles.

(4) *Inclination (i).* The orbital requirements for the inclinations of the two satellites are the most severe. This is because the uncertainties of the earth oblateness and relative inclination produce an uncertainty in the mean non-relativistic nodal precession rate, which directly affects the accuracy with which the Lense-Thirring nodal precession rate can be determined. As such, the average relative polar separation of the two orbit planes over the 2.5 year duration of the experiment must be less than 10 meters. This will be accomplished with the aid of small velocity vector corrections, applied every three months at the equator.

The mean position of the orbit planes must also be close to the pole of the earth. If it is not, the orbit planes will scissor slightly (i.e., the ascending nodes will not remain exactly 180° apart). Present ground tracking can determine in-track orbit position to about 10 m. Even with the scissor angle precisely known a component of the in-track uncertainty appears in the minimum range determination at the poles. To keep this range uncertainty small compared to the 1 cm Doppler ranging accuracy, it is required that the mean of the two satellite planes, averaged over the duration of the mission, be within 70 m of the earth's pole.

Both of these requirements on the inclinations must be viewed in light of the fact that periodic small changes in the satellite orbital inclination, applied every three months at the equator, will be required to prevent the possibility of a satellite collision [14] and in addition for keeping up with the moving pole of the earth due to the precession of the equinoxes.

(5) *Longitude of the node* (Ω). The relative longitude of the nodes, as mentioned above, must be kept close to 180°. So that the plane scissoring will not introduce large uncertainties in the polar slant range Doppler measure-

TABLE 3. *Risks associated with the gyro experiment.*

Description of Risk Area		Specific Defect or System		Type of Failure (accuracy or hard)
(1)	Ball to cavity contact or gyro performance	1a	Thin film ball coating damage due to 1g suspension	Hard (pre-launch)
		1b	Gyro contamination	[a]Accuracy/hard
		1c	Cryogenic deposits on gyro rotor	Accuracy/hard
		1d	Arcing due to 1a or or 1b	Hard (pre-launch)
		1e	Thick film ball coating thickness non-uniformity[b]	Accuracy
(2)	Pre-launch testing	2a	Inadequate ground testing due to avoidance of 1a	Accuracy if 1 to 3 gyros fail to function on orbit/hard if none function on orbit
(3)	Dewar system performance	3a	Loss of low (10^{-8} gauss) magnetic field	Accuracy/hard
		3b	Loss of high (10^{-9} Torr) vacuum	Accuracy/hard
		3c	Loss of superconductivity	Accuracy/hard
(4)	Electronics, assuming all critical functions are available initially on orbit.	4a	Loss of suspension	Accuracy
		4b	Loss of data readout system	Accuracy
		4c	Loss of control systems	Accuracy
		4d	Loss of flight computer(s)	Accuracy
		4e	Loss of communications	Accuracy
		4f	Loss of power system	Accuracy

[a]Note: Failure during the flight means reduced mission life which constitutes accuracy degradation.

[b]Note: The study of relative advantages of thick and thin films is incomplete and coupled with the rotor diameter study which is also in progress (the electrode force per unit area at 1g decreases directly as ball diameter decreases, but gyro drift error increases).

ment, the nodes must be within 0.7 km of each other for the duration of the mission.

(6) *In-track Encounter Point* (θ). The in-track encounter point requirement is that the encounter point is to occur within 2 km of the poles. This allows an enormously simplified fluctuation model to be used without introducing uncertainties comparable to the Doppler accuracy.

In addition to the usual method of using ground-tracking for orbit determination, this experiment can provide very accurate relative orbit information from the satellite-to-satellite Doppler data taken near the north and south poles. This is particularly useful in this experiment where high precision orbital information is required, and many of the orbit requirements are constraints on the relative orbital elements. Currently, ground tracking could supply the following accuracies on the relative state vector; δa, ± 10 m; $a\,\delta\xi$, ± 40 m; $a\,\delta\eta$, ± 40 m; $a\,\delta i$, ± 40 m; $a\,\delta\theta$, ± 40 m where δa, $\delta\xi$, $\delta\eta$, δi, and $\delta\theta$ are the differences between the two satellites' state vectors. An analysis due to D. B. Schaechter [15] has shown that the relative orbital elements can, in fact, be estimated very quickly and accurately from ground tracking augmented by the satellite-to-satellite Doppler measurement. An example of how the uncertainties evolve is shown in Table 4.

Reference [15] also presents a scheme for maneuvering the orbits into their required finely tuned configuration. The scheme makes the best use of the satellite-to-satellite tracking and also avoids any possible mishap through a satellite collision. In addition, Ref. [15] gives

TABLE 4. *Uncertainty in relative orbital elements after measurements.*

	Relative Orbital Elements				
No. of Satellite-to-Satellite Measurements	δa	$a\delta\xi$	$a\delta\eta$	$a\delta i$	$a\delta\Omega$
0	100	80	80	40	40
10	0.39	7	7	28	28
20	0.14	5	5	18	18
30	0.08	4	4	9	9
40	0.05	3	3	8	8

TABLE 5. *Risks associated with twin satellite experiment.*

Description of Risk Area		Specific Defect or System		Type of Failure (accuracy or hard)
(1)	Keeping ahead of the orbit tuning	1a	Excessive delay in orbit determination and tuning	Accuracy
(2)	Electronics, assuming all critical functions are available initially on orbit	2a	Loss of Doppler systems	[a]Accuracy
		2b	Loss of drag-free or attitude control	Accuracy
		2c	Loss of flight computer(s)	Accuracy
		2d	Loss of communications	Accuracy
		2e	Loss of power system	Accuracy
(3)	Collision of satellites	3a	Collision damage	Accuracy

[a]Note: Failure during the flight means reduced mission life which constitutes accuracy degradation.

closed form solutions for the station keeping requirements in terms of the relative orbital elements. Both two and three-impulse strategies are optimal, depending on the state of the orbits (state space regions).

From the above discussion it should be evident that orbit control will be a key factor in conducting the twin satellite experiment flight operations. The risk areas viewed as most important by the author are listed in order of perceived importance at this writing in Table 5.

7. Conclusion

A precise check of frame dragging would be an event of the highest scientific importance. It would extend the experimental verification of Einstein's general theory of relativity into a new area by providing the first measurements on the gravitational effects of moving matter, i.e., the "gravimagnetic" forces which are analogous in Einstein's theory to electromagnetic forces in Maxwell's theory. Agreement of the measured result with the prediction of Einstein's theory would allow application of that theory with much greater confidence to astrophysical calculations on rotating massive objects such as neutron stars or black holes. Disagreement with the theory would compell a radical rethinking of our concepts about gravitation.

Several ideas have been advanced for measuring frame dragging either by observations with the earth, the rotating sun, or rotating astrophysical objects. One difficulty with any observations involving the astrophysical bodies, including the sun, is the lack of accurate data or absence of any data on the angular momentum of the source. Although observations may be suggested which might reveal the existence of frame dragging they cannot give a quantitative check of the theory and their significance would remain uncertain. A more desirable approach is to make the measuements using the nearby rotating massive body whose angular momentum we do know (the earth) and to apply the information so obtained to deduce further information about astrophysical objects. Of experiments using the sun, the earth, or a laboratory body as the source, only the gyro and counter-orbiting polar satellite experiments appear to be feasible [16].

One interesting example of an astrophysical phenomenon which has been attributed to frame dragging is the existence of double lobed radio sources at the center of certain galaxies and quasars. The radio jets are observed to be collimated over ranges of hundreds of kiloparsecs. The currently most popular theory of this effect, due to R. Blandford and R. Shajek [17], is that there is a rotating charged supermassive black hole at the center of the quasar and that gravitational energy is coupled by frame dragging to the electromagnetic energy of the charged black hole to yield the radio signals. Another possible example of frame dragging is a theoretical explanation by

Fang and Ruffini [18] of the modulation on the large red and blue Doppler shift in the H_α lines of Object SS433. SS433 is observed to have periodic red and blue shifts modulated with a 164 day period. Fang and Ruffini have shown that a possible explanation of these observations is a rotating black hole with a compact object in a low orbit somewhat above the event horizon, $r < 10^4$ m. On this theory the red and blue shift result from the radial component of the orbital velocity and the 164 day modulation results from the Lense-Thirring precession of the orbiting compact object which is the source of the H_α radiation.

Because of the present situation with a NASA Phase B study in progress on the gyro experiment by the Marshall Space Flight Center and Stanford University following 17 years of laboratory research, it is the author's opinion that the gyro experiment must be successfully flown in the presently projected 1988–89 time frame if we are to have quantitative experimental frame dragging data by 1990. The twin satellite experiment also has the potential to yield accurate frame dragging data; however the prospects for its implementation during the 1980s are less favorable.

References

[1] L. I. Schiff, Proc. Nat. Acad. Sci., USA 46, 871 (1960); Phys. Rev. Lett. 4, 216 (1960).

[2] J. Lense and H. Thirring, Phys. Z. **XIX**, 156 (1918).

[3] C. W. F. Everitt, in *Experimental Gravitation*, Ed. by B. Bertotti (Academic Press, New York, 1974), pp. 331–360; J. A. Lipa, W. M. Fairbank and C. W. F. Everitt, *ibid.* p. 361.

[4] R. A. Van Patten and C. W. F. Everitt, Celestial Mechanics 13, 429 (1976).

[5] R. A. Van Patten and C. W. F. Everitt, Phys. Rev. Lett. 36, 629 (1976).

[6] R. Vassar, J. V. Breakwell, C. W. F. Everitt, and R. A. Van Patten, J. Spacecr. Rockets, in press.

[7] D. Schaechter, J. V. Breakwell, R. A. Van Patten, and C. W. F. Everitt, J. Spacecr. Rockets 17, 474 (1977).

[8] Richard A. Van Patten, in *Proceedings of the International School of General Relativity Effects in Physics and Astrophysics; Experiment and Theory (Third Course)* Ed. by J. Ehlers (Max-Planck-Institut für Physik und Astrophysik, Munich, 1977), p. 141.

[9] J. T. Anderson and C. W. F. Everitt, "Limits on the Measurement of Proper Motion and the Implications for the Relativity Gyroscope Experiment" (internal report, Stanford University Dept. of Physics, 1979).

[10] C. W. F. Everitt, Report to the National Aeronautics and Space Administration under grant No. 05–020–019 and contract No. NAS8–32–355 (June 1980).

[11] C. W. F. Everitt, Daniel B. DeBra, and Richard A. Van Patten, in *Advances in the Astronomical Sciences*, Ed. by E. J. Bauman (Univelt, San Diego, CA, 1981) Vol. 45, p. 265.

[12] D. B. Schaechter, Ph.D. Thesis, Stanford University (1977).

[13] J. V. Breakwell, C. W. F. Everitt, D. B. Schaechter, and R. A. Van Patten, Acta Astronautica, to be published (1982).

[14] David Schaechter, John V. Breakwell, Richard A. Van Patten, and Francis W. Everitt, J. Astronaut. Sci. 24, 137 (1976).

[15] D. B. Schaechter, J. V. Breakwell, R. A. Van Patten, and C. W. F. Everitt, in *Preprints of the 7th Triennial World Congress of the IFAC, Helsinki, Finland*, Ed. by A. Niemi (Permagon, Oxford 1978) Vol. 1, p. 1197.

[16] I. I. Shapiro, *et al.*, "Strategy for Space Research in Gravitational Physics in the 1980's," prepared by the Committee on Gravitational Physics, Space Science Board, National Academy of Sciences (27 July 1980).

[17] R. D. Blandford and R. L. Snajek, Mon. Not. R. Astron. Soc. 179, 435–456 (1977).

[18] Fang Li Zhi and Remo Ruffini, Phys. Lett. 86B, 193 (1979).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

The Status of the Velocity of Light in Special Relativity

Ernst Breitenberger

Department of Physics, Ohio University, Athens, OH 45701

The special theory of relativity is shown to be independent of the assumption that the velocity of light, c, is a universal constant. No more than an intuitively obvious monotonicity postulate of velocity addition is needed to prove the existence of a universal, kinematic limit velocity c_0, and the existence of a universal dispersion relation $\omega^2 = c_0^2(k^2 + \mu^2)$ for vacuum waves. The best empirical support for the signal postulate $c_0 = c$ still comes from uncorroborated observations of the Crab Nebula pulsar which admit a photon rest frequency of up to 7 MHz, far above the limit inferred from the consistency of quantum electrodynamics (QED). Existing theory-dependent arguments purporting to demonstrate the constancy of c are shown to be inadequate. Further experimental work on the variability of c with frequency and with source conditions is desirable.

Key words: photon rest mass; second postulate; special relativity; vacuum dispersion; velocity of light.

The special theory of relativity is traditionally developed using the premise that the velocity of light is a universal constant, c, which seems to rule out any logical argument involving the contrary, and worse, if believed to be unavoidable, would make any contrary experimental result appear as a refutation of relativity.

The subject can also be developed without reference to c, in a purely kinematic manner. This approach has many advantages (see Ref. [1]). It has a rich history dating back to 1910, but was never taken quite seriously, presumably because it remained strewn with redundant assumptions and inaccuracies of proof. However, a satisfactory kinematical derivation of the Lorentz transformation can be formulated [2]. In the following, I first outline the main steps of the argument, and then show some conclusions of interest to this Conference. In essence [1], the velocity of light becomes a legitimate object of investigation, both experimental and theoretical; it also emerges that not enough is known about its dispersion which seems intimately related to the validity of QED.

Note first that the traditional approach *requires* the existence of a universal signalling velocity c in order to make sure that it is operationally feasible to compare length and time *intervals* between two inertial systems K, K' which are in relative motion. Note next that we need not compare intervals because nowadays we can agree to use natural units furnished by atomic standards, such as ^{86}Kr and ^{133}Cs. Of course, we still need to communicate the *zeros* of coordinate and time scales, but for that purpose we need only *single* marker pulses whose velocity is not essential to the procedure; in fact, a particle burst could serve just as well as a light flash.

The admissible transformations (ATs) between K and K' are restricted by the relativity postulate to be linear. A chain of well-known, elementary arguments [2] shows that they have the form

$$\begin{aligned} x' &= \gamma(x - ut), \\ t' &= \epsilon x + \gamma t, \\ y' &= y, \quad z' = z, \end{aligned} \tag{1}$$

with $\gamma(u) > 0$ and $\gamma(-u) = \gamma(u)$, if the usual convention is made that K and K' coincide at $t = t' = 0$, and that O' moves in K with velocity u along the x-axis.

The traditional argument would next call upon the constancy of c and hence require the invariance of $x^2 + y^2 + z^2 - c^2t^2$ (from which the unknown coefficients γ and ϵ follow easily). This is quite unnecessary, though. First, the equivalence of K and K' requires that the inverse to the transformation of Eq. (1),

$$x = (x' + ut')/(\gamma + u\epsilon)$$

have the form $x = \gamma(-u)[x' - (-u)t']$, whence

$$\gamma = 1/(\gamma + u\epsilon) \tag{2}$$

must hold. For convenience, set now

$$\epsilon = -u\alpha\gamma \tag{3}$$

with a new function $\alpha(u)$, then from Eq. (2)

$$\gamma = (1 - \alpha u^2)^{-1/2} \tag{4}$$

follows; also, $\alpha(u)$ is seen to be even and therefore $\epsilon(u)$ is odd.

Second, the product of two ATs must be another AT; if we choose for convenience the second AT to lead from K' to a K'' which moves at u' against K' along the x'-axis and is oriented such that K' and K'' coincide at $t' = t'' = 0$, then it is necessary that

$$\begin{pmatrix} \gamma' & -u'\gamma' \\ -u'\alpha'\gamma' & \gamma' \end{pmatrix} \begin{pmatrix} \gamma & -u\gamma \\ -u\alpha\gamma & \gamma \end{pmatrix} = \begin{pmatrix} \gamma'' & -u''\gamma'' \\ -u''\alpha''\gamma'' & \gamma'' \end{pmatrix}, \tag{5}$$

where $\gamma' = \gamma(u')$ etc., for transformations of the type given in Eq. (1) as written in the notation of Eq. (3). Comparison of the matrix elements leads at once [2] to the triple conclusion

$$\alpha = \text{const}, \tag{6}$$

$$\gamma'' = \gamma\gamma'(1 + \alpha u u'), \tag{7}$$

$$u'' = (u + u')/(1 + \alpha u u'). \tag{8}$$

For the history of the results given in Eqs. (4) and (6), see Pauli [3].

Try here α negative and set $\alpha = -1/w^2$; then

$$\gamma = (1 + u^2/w^2)^{-1/2}, \tag{9}$$

$$u'' = (u + u')/(1 - uu'/w^2). \tag{10}$$

Evidently this AT has no limit velocity. Addition of two small positive velocities according to the law of Eq. (10) yields a resultant larger than their arithmetical sum, but if uu' is made large enough the resultant becomes negative. It is rather startling to recognize that such an aberrant velocity composition law is compatible with commonsense notions of kinematics although they have been shackled under the severe restriction of the relativity principle. In order to exclude this absurdity, we need a specific assumption which can be conveniently formulated as a *weak montonicity postulate of velocity addition:* the resultant of two parallel velocities never has the opposite direction.

On the other hand, if $\alpha > 0$ we obtain with $\alpha = 1/c_0^2$ the usual $\gamma = (1 - u^2/c_0^2)^{-1/2}$, and c_0 plays the role of a purely kinematic limit velocity. However, the concomitant Einstein addition law

$$u'' = (u + u')/(1 + uu'/c_0^2) \tag{11}$$

still admits complications which are best seen from its simple consequence [4]

$$u'' - u = u'(1 - u^2/c_0^2)/(1 + uu'/c_0^2).$$

If we again choose both u and u' positive, the difference $u'' - u$ will be negative if $u > c_0$ (and only then). Thus, adding on to the velocity will for a tachyon lead to a slowing down, as if cause and effect were interchanged [4], but the phenomenon is purely kinematic in origin because no dynamical (or "signalling") assumptions whatsoever have entered the argument. In order to exclude tachyons we may accordingly adopt a *strong monotonicity postulate of velocity addition:* the resultant of two parallel, positive velocities is never less than either. (The strong postulate evidently implies the weak one, which is the reason why the monotonicity formulation is here being put forward.)

Thus the price for the removal of c from the foundations of relativity is a kinematic axiom the need for which was never openly appreciated (to the detriment of our understanding of "causality" which ought to be shorn of such intrinsically kinematic elements). On the other side there are great benefits [1]. In particular, it becomes legitimate to ask whether or not $c_0 = c$ holds ("signal postulate" [1]).

By an elementary argument [5], for any wave (harmonic or not, but purely periodic) capable of propagating through vacuum, $(\mathbf{k}, i\omega/c_0) = \kappa$ must be a four-vector. It follows that

$$\kappa_j \kappa_j = k^2 - \omega^2/c_0^2 = \text{const} = -\mu^2. \tag{12}$$

The phase velocity is then

$$c_p = \omega/k = c_0(1 + \mu^2/k^2)^{1/2}, \tag{13}$$

and the group velocity $c_g = d\omega/dk$ is found to satisfy

$$c_p c_g = c_0^2. \tag{14}$$

The dispersion law of Eqs. (12) or (14) has long been known for de Broglie and for Klein-Gordon waves, but no one seems to have duly emphasized that it is universal; it is, in fact, a fundamental result of relativity with the same standing as, say, the addition law Eq. (11) or the time dilation formula. It is purely kinematic in nature (see the quoted argument [5]) and constitutes a restriction on all possible wave equations (whether tachyons are admitted or not). If the wave equation is linear and of the second order, the only possibility is the Klein-Gordon equation

$$\psi_{xx} - \psi_{tt}/c_0^2 = \mu^2\psi, \tag{15}$$

but there are many nonlinear and/or higher-order possibilities.

Once the kinematic origin of the constant μ is clearly understood, there is no harm in setting

$$\mu = 2\pi/\lambda_C \tag{16}$$

with λ_C a length, and in furthermore linking up with the established dynamical context by setting

$$\lambda_C = h/m_0 c_0 \tag{17}$$

as for a Compton wavelength. As a pragmatic middle way between logical accuracy and familiar context,

$$\nu_0 = c_0\mu/2\pi = c_0/\lambda_C \tag{18}$$

will here be called the "rest frequency" of the vacuum wave. If $\mu = 0$, then $\omega = c_0 k$ and there is no dispersion. If $\mu^2 > 0$, we have a real ν_0 to which a real rest mass may be thought to correspond. If $\mu^2 < 0$, then ν_0 is imaginary, while Eq. (13) shows that $c_p < c_0$ and hence Eq. (14) that $c_g > c_0$; this case may be thought to correspond to tachyons.

The signal postulate holds if and only if c is a constant [1]. Given the unique dispersion law of Eq. (12), if relativity is valid only two kinds of variability of c are possible: vacuum dispersion with one particular value of μ or ν_0; and vacuum dispersion with different values of μ which could only arise from different source processes, or as we say traditionally, "dependence of c on the state of motion of the source."

The best (but still uncorroborated) result on vacuum dispersion is due to Warner and Nather [6]; for discussion, see Ref. [1]. They deduced from observations of the Crab Nebula pulsar that between ≈ 350 nm and ≈ 550 nm the velocity of light is constant to better than 5 parts in 10^{17}; to be precise,

$$|\Delta c_g|/c_g < 4.8 \times 10^{-17}. \tag{19}$$

This astonishing figure certainly justifies the claim that for commercial purposes c in the visible equals c_0, yet it does not imply a satisfactorily low limit on ν_0.

From Eqs. (13) and (14) it follows that $c_g/c_0 = (1 + \mu^2/k^2)^{-1/2} \approx 1 - \mu^2\lambda^2/8\pi^2$, hence

$$|\Delta c_g|/c_0 \approx \mu^2\lambda\,|\Delta\lambda|/4\pi^2. \tag{20}$$

If in Eq. (20) we replace c_0 by c_g and combine it with the definitions of Eqs. (17) or (18), we have [1]

$$\nu_0 = m_0 c_0^2/h \approx c_0[(|\Delta c_g|/c_g)/(\lambda\,|\Delta\lambda|)]^{1/2}, \tag{21}$$

and if we take $c_0 \approx c$ and insert the experimental value, Eq. (19), we find

$$\nu_0 \lesssim 7 \text{ MHz} \tag{22}$$

(the value given in Ref. [1] is smaller by $\sqrt{10}$ because in Ref. [6] there is a misprint of one power of 10). This result is a disappointment. A ν_0 of some MHz would have been seen long ago in resonance spectroscopy, or as a discrepancy with QED predictions which are always based on $\nu_0 = 0$. For instance, the Lamb shift calculations in H agree with experiment (at present, roughly) within 10 kHz. A limit of $\nu_0 < 10$ kHz naturally suggests itself; wherefore we still (in 1981) have the irksome situation that QED provides comfort for standard Maxwellian electrodynamics [1], instead of the latter's being supported by a direct experimental result on c.

Evidently a photon rest frequency would show up better at long than at short wavelengths [1]. To have a yardstick, suppose that at $\lambda \approx 300$ m it were found that c remains constant to 1 ppm over a range of $\Delta\lambda = 1$ m; Eq. (21) then shows that $\nu_0 < 17$ kHz. That would be a result well worth the effort. It has never been clearly un-

derstood that measurements of the *variability* of c were just as important as knowledge of its numerical value [7] at some wavelength, for we never clearly recognized that a putative "photon rest mass" was an intrinsic feature of *kinematics* which should be checked out *before* relativity can safely be applied in electro*dynamics*. In addition to infrared or rf measurements, it may also be possible to exploit modern techniques in the visible; the feasibility of finding variations of c to 1 part in 10^{23} has already been envisaged [8]. A gain of 10^6 over the result of Eq. (19) could improve the limit of Eq. (22) by 10^3, depending on $\Delta\lambda$, and again, the effort would be worthwhile.

Claims have repeatedly been made on theoretical grounds to the effect that "the photon rest mass must be zero." They were almost invariably based on the assumption of harmonic waves and the Klein-Gordon equation, Eq. (15), which holds for the vector potential components in an electrodynamics due to Proca [9] (see Ref. [1] for brief comment). This generalization of Maxwell's theory is not very plausible to start with, for it is not gauge-invariant [10]. (Probably this is the reason for the occasional and otherwise unsupported statement that "gauge invariance excludes a non-zero photon rest mass.") Still, it is entirely reasonable to ask what finite values of the Proca parameter μ would be compatible with existing data. Through careful comparisons with known features of Maxwellian large-scale fields it has been shown [11] that the Proca μ must be effectively zero; in the present terminology, a Proca ν_0 could be at most a small fraction of 1 Hz. Thus Proca's electrodynamics is not a viable alternative to Maxwell's, and consequently cannot say anything about the photon rest frequency [1].

Indeed, there are (untold?) other alternatives to Maxwell's theory. Take the Born-Infeld one (even though it is not supported by evidence either, see Ref. [12]). It leads to a wave equation of the type

$$(1 - \psi_t^2)\psi_{xx} + 2\,\psi_x\,\psi_t\,\psi_{xt} - (1 + \psi_x^2)\psi_{tt} = 0 \qquad (23)$$

for the vector potential components [13]. This has undistorted wave solutions $f(x - t)$ and $g(x + t)$, and therefore is *free* of dispersion, yet its *static* solutions exhibit a typical $(a^4 + r^4)^{-1/2}$-dependence on distance [13], much as a Yukawa dependence with $e^{-\mu r}$-terms is typical for Proca fields [11]. Thus, a departure from the Maxwellian r^{-2}-fields is here compatible with a *zero* photon rest mass.

Consider also the possibility of linear *fourth*-order equations. The obvious $\Box^2\psi = 0$ is dispersion-free. For its static solutions with $(\nabla^2)^2\psi = 0$ there is (amongst others) the well-known elementary result: if ψ_1 and ψ_2 are harmonic, then $x\,\psi_1 + \psi_2$ is biharmonic (see any treatise on elasticity). Once again, fields quite different from the Maxwellian ones (and now with a departure from the Coulomb r^{-2}-character quite different from Yukawa's $e^{-\mu r}$-factor) are here compatible with a zero photon rest mass. These (counter) examples sharply illuminate the *kinematic* origin of a putative photon rest frequency $\nu_0 \neq 0$: electro*dynamical* speculations about the character of *static* fields, unsupported by other evidence, do not lead to unequivocal insights concerning ν_0, hence observations of fields *other* than those of vacuum waves cannot at present tell us anything definitive about ν_0.

A similar reservation applies to the interpretation of pulsar data subject to plasma dispersion in the interstellar medium [14]. It is not known how a non-zero ν_0 would couple with plasma dispersion; any hypothesis about that is dynamical and thus leads back to the previous situation inasmuch as it would require independent justification before it can be directed at the question, "what is ν_0?"

A somewhat more complex objection must be raised against the interpretation of a classic experiment often named nowadays after Cavendish, but previously named usually after Plimpton and Lawton [15] who obtained the first high-accuracy result. The experiment is said to validate Coulomb's law within extremely stringent tolerances, and hence to imply a stringent limit on ν_0. In this form, the claim is incorrect. First, the experiment does not test Coulomb's law

$$E(r) = 1/4\,\pi\,\epsilon_0\,r^2, \qquad (24)$$

but the law of Gauss

$$D(r) = 1/4\,\pi\,r^2, \qquad (25)$$

which leads to Eq. (24) only through the linearity hypothesis $D = \epsilon_0 E$. In fact, the experiment is an inverse Faraday pail experiment which tests the efficacy of an electrostatic shield by measuring the *charge electrostatically induced* on an electrometer placed inside when a known charge is put on the outside. Note that no measurement of a (ponderomotive) force such as Eq. (24) is involved, and that the induced charge would be quite the same if the shield (or pail) were filled with a dielectric.

There seems to be a universal lack of understanding that D is being measured, not E. In order to arrive at Coulomb's law, Eq. (24), one may, of course, combine the result on D with an independent estimate on the linearity of the vacuum; however, for the latter we can draw only upon QED, as far as we can see at the present time. Thus QED would again become involved in a fundamental, "macroscopic" problem. Furthermore, we again do not know how, nor if, a non-zero ν_0 would affect Coulomb's law over laboratory distances, because we do not know what sensible dynamical hypothesis to make, just as discussed above. (All this must on no account be taken as a slight on the beautiful work done recently [16] to improve the Plimpton-Lawton result; indeed it is pleasant to know how accurately the law of Gauss is valid.)

Lastly, consider possible variations of c arising from source processes. Ten years ago, the evidence in regard of the traditional "second postulate" was weak [1], but two results of considerable weight have been published since. Brecher [17] deduced from the observations of selected x-ray pulsars which are revolving quite fast in close binary systems that less than 2×10^{-9} of the (non-uniform!) orbital speed adds to the velocity of the x-rays; this is a very low limit on the traditional "ballistic" effects. Brown *et al.* [18] measured the velocity of 7 GeV bremsstrahlung bursts from SLAC to an accuracy of close to 1 part in 10^5 and found it to be consistent with the accepted value of c; this shows that fairly strong radiative reactions do not alter c. From the present, purely kinematic standpoint these observations are directed at the possibility that the μ in Eq. (12) is not a universal constant. Difficult as it is to imagine "several photon rest masses," there seems to be no first principle opposed to the idea. Thus it remains desirable to have further results under circumstances where nuclear forces or great changes of source acceleration play prominent roles, as pointed out in Ref. [1].

To leave nothing unsaid, it should be added that a non-null result on ν_0 or μ would raise ambiguities of interpretation: has a non-zero photon rest mass been demonstrated, or is the dispersion relation, Eq. (12), violated, with dire consequences for relativity? Thus in the kinematic approach the velocity of light indirectly becomes a test object for the validity of the basic result, Eq. (12), enhancing the desirability of further work.

References

[1] E. Breitenberger, Nuovo Cimento **1B**, 1 (1971).
[2] E. Breitenberger, to be submitted to Ann. Phys. (N.Y.).
[3] W. Pauli, *Relativitätstheorie* (Teubner, Leipzig 1921), translated as *Theory of Relativity* (Pergamon, New York 1958); end of Sec. 4 in either version.
[4] A. Einstein, Ann. Phys. (Leipzig) 23, 371 (1907).
[5] For example, J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York 1975), p. 508.
[6] B. Warner and R. E. Nather, Nature (London) 222, 158 (1969).
[7] See J. Terrien, Metrologia 10, 9 (1974); K. M. Baird, D. S. Smith, and B. G. Whitford, Optics Commun. 31, 367 (1979).
[8] A. M. Grassi Strini, G. Strini, and G. Tagliaferri, Phys. Rev. D 19, 2330 (1979).
[9] A. Proca, J. Phys. Radium 8, 23 (1936), and references given there.
[10] W. Pauli, Rev. Mod. Phys. 13, 203 (1941), Sec. II.2.
[11] General references: A. S. Goldhaber and M. M. Nieto, Rev. Mod. Phys. 43, 277 (1971); A. Barnes and J. D. Scargle, Phys. Rev. Lett. 35, 1117 (1975); A. D. Dolgov and Ya. B. Zeldovich, Rev. Mod. Phys. 53, 1 (1981), Sec. XII.
[12] For example, J. Rafelski, W. Greiner, and L. P. Fulcher, Nuovo Cimento **13B**, 135 (1973).
[13] M. Born and L. Infeld, Proc. R. Soc. London, Ser. A 144, 425 (1934).
[14] For example, J. R. Rawls, Phys. Rev. D 5, 487 (1972).
[15] S. J. Plimpton and W. E. Lawton, Phys. Rev. 50, 1066 (1936).
[16] E. R. Williams, J. E. Faller, and H. A. Hill, Phys. Rev. Lett. 26, 721 (1971).
[17] K. Brecher, Phys. Rev. Lett. 39, 1051 (1977); 39, 1236(E) (1977).
[18] B. C. Brown, G. E. Masek, T. Maung, E. S. Miller, H. Ruderman, and W. Vernon, Phys. Rev. Lett. 30, 763 (1973).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Relativistic Time Dilation: A Latter-Day Ives-Stillwell Experiment

Paul Nachman,* Marc D. Rayman, and J. L. Hall**

Joint Institute for Laboratory Astrophysics, National Bureau of Standards and University of Colorado, Boulder, CO 80309

Two-photon spectroscopy on a fast atomic beam offers the possibility of a dramatically improved test of the special relativistic time dilation. Our experimental arrangement will allow direct measurement of He atom resonance frequencies displaced from their rest-frame values by time dilation alone. The atomic beam velocity will be determined optically via the Doppler shift. The longitudinal interaction geometry will also permit us to test with increased precision for a possible vector anisotropy in the speed of light, a possibility raising renewed interest because of recent astrophysical discoveries. Further, we expect to refine experimental values for the energies of the Rydberg terms in He.

Key words: helium spectrum; preferred frames; special relativity; time dilation; two photon spectroscopy.

We are assembling a spectroscopic experiment for a precision test of the time dilation (second-order Doppler shift) predicted by special relativity. The "moving clocks" [1] to be used will be helium atoms in a fast beam ($v/c \lesssim 5 \times 10^{-3}$). The atoms will be in metastable states $2s^1S$ and $2s^3S$ and will be probed by two-photon transitions to higher triplet and singlet S and D energy levels.

The experiment, which is nearly ready for operation, is designed as shown in Fig. 1. Helium ions, produced in a discharge source, are accelerated to form an essentially monoenergetic beam (5 to 50 keV, with a $\lesssim$1 eV spread).

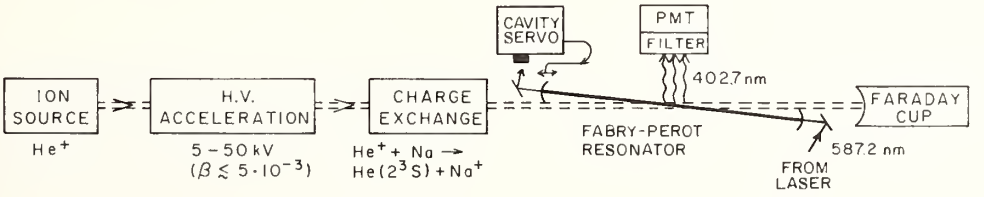


FIGURE 1. *Block diagram of fast-beam spectroscopy experiment.*

The ion beam (with several μA flux) is focused through and beyond a cell containing sodium vapor. Resonant charge exchange with the Na produces a beam that includes an appreciable fraction [2] of neutral He atoms in the metastable levels. In the ~milliradian forward cone, the velocity dispersion of the beam is not increased. This fast atom beam crossed, with small intersection angle, a standing wave light field from a stabilized cw ring dye laser. A Fabry-Perot resonator (build-up cavity) is used to increase the energy density of the light by approximately a factor 30 from that in the incident (~300 mW) laser beam. Figure 2 illustrates the proposed optical system.

The initial resonance used will be $5d^3D \leftarrow 2s^3S$. Light at three discrete wavelengths near 587.2 nm will drive this two-photon transition, producing three peaks in the 402.7 nm fluorescence $2p^3P \leftarrow 5d^3D$ [Fig. 3(a)] as the exciting wavelength is varied. The central resonance results from excitation of the $He(2s^3S)$ atoms by two counterpropagating photons and thus is shifted only by relativistic time dilation. The two other resonances, displaced by ~2 nm (~2 THz) from the central one [Fig. 3(b)], correspond to excitation by two copropagating photons, and thus are shifted primarily by the first-order Doppler effect. "Light shifts" of the resonances due to the ac Stark effect should be $\lesssim$10 kHz.

The angle of intersection between the atomic and optical beams is small (~20 milliradians) to provide long interaction times, hence minimizing laser intensity requirements and transit-broadening of the resonances. This geometry also permits direct determination of the atomic velocity by means of the first-order Doppler shifts, an improvement over earlier work in this laboratory wherein the particles' velocities were inferred from the accelerating potential of the beam [3].

The time dilation shift of the 587.2 nm resonance for He atoms with 50 keV kinetic energy is ~6 GHz. The line width of the resonance should be ~5 MHz, due to transit broadening. We expect to determine line center to 25 kHz (one half percent of the line width), thus achieving a measurement of the time dilation shift to a few parts per million. This accuracy level would represent ~300-fold improvement over that achieved in an experiment at much higher velocities, with muons at the CERN storage ring [4].

*NAS/NRC Postdoctoral Research Associate, National Bureau of Standards.
**Staff Member, Quantum Physics Division, National Bureau of Standards.

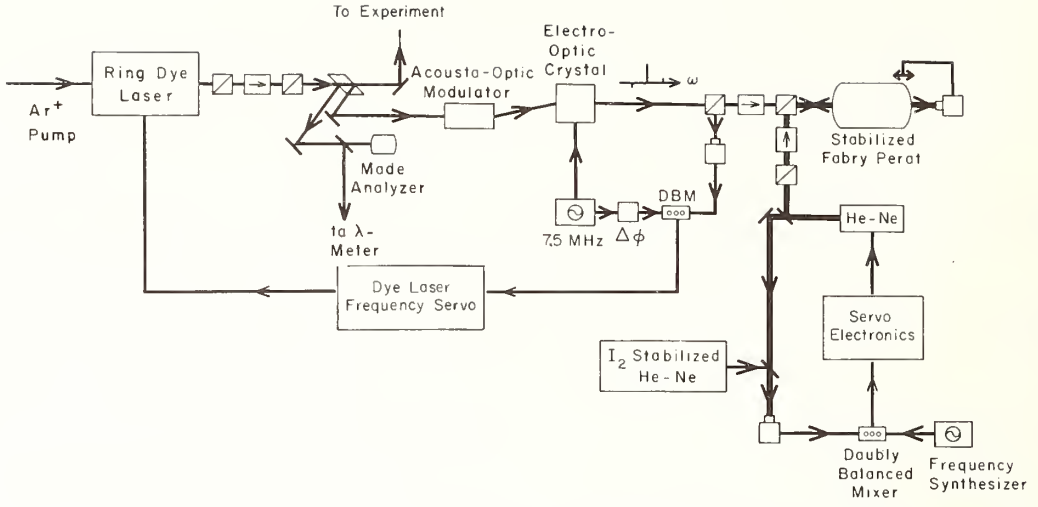


FIGURE 2. *Dye laser stabilization and tuning. A small portion of the output of a cw ring dye laser is sampled (after the optical isolator) to be used for diagnostics (mode analyzer and Lambdameter) and for laser frequency control. In the stabilization/fast tuning system this latter sample is frequency-shifted with an acousto-optic modulator (AOM) driven by a stable, but controllable, rf frequency source. The resulting beam is phase modulated with an electro-optic crystal to produce an FM sideband spectrum for locking the dye laser to the stabilized Fabry-Perot, using a reflection-mode Pound/Drever optical stabilizer. Changes in the AOM drive frequency are directly imposed on the laser by its frequency servo. For slower or larger scans, the cavity length is tuned via the frequency synthesizer which provides a programmable frequency offset between the HeNe local oscillator laser and the I_2-stabilized reference laser. The Fabry-Perot's transmission fringes are used to lock its length to the resonance condition with the HeNe local oscillator's variable wavelength, thus ultimately mapping the synthesizer's frequency change into a proportional frequency change of the dye laser.*

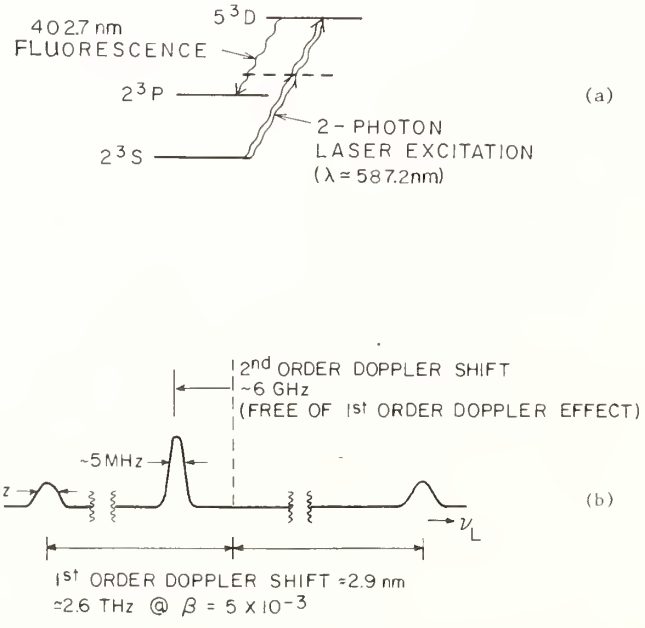


FIGURE 3. *(a) Energy levels and transitions for initial two-photon experiment. (b) Amplitude of 402.7 nm fluorescence [Fig. 2(a)] as function of exciting laser frequency. First-order-shifted peaks arise from absorption of two laser photons from same propagation direction. Central resonance is usual "Doppler-free" two-photon peak arising from one photon from each beam.*

Important future accuracy improvements can be obtained by using much smaller beam crossing angles in conjunction with two-photon transitions to longer-lived, higher excited states. Improved velocity definition can be accomplished with two-laser/two-photon techniques related to saturation spectroscopy. Line widths of 50 kHz seem realistic and would improve these expected results by two orders of magnitude if the wavelength/frequency metrology can be correspondingly improved. One possibility for this would be the use of CO_2 laser/dye laser sideband generation in a nonlinear crystal, followed by heterodyne detection with the second dye laser.

The experimental setup will allow us to make several other precision measurements. For example, recent years have seen renewed interest in the possibility of residual preferred frame effects. In particular, the observed anisotropies in the cosmic microwave background [5] and in the distribution of "local" mass (out to the Virgo Cluster of galaxies [6]) suggest that our galaxy has a peculiar velocity (compared to the Hubble expansion) with respect to the average mass of the universe. Using the present apparatus to monitor the relative positions (to ~10 kHz) of the first-order Doppler-shifted resonance and the central resonance as functions of sidereal time will permit us to test at the sub-10^{-8} level for a possible vector anisotropy in the speed of light. Our beam is oriented east-west to obtain the maximum sidereal modulation of such a hypothetical one-way effect.

Another interesting precision measurement to be made with our fast longitudinal beam is of the He electronic structure. Doppler-free two-photon spectroscopy of high-lying ("Rydberg") terms in He provides a way to refine substantially experimental knowledge of this structure. Results of such accurate measurements, in company with recent microwave-optical studies [7] of fine structure within and among Rydberg terms, would present a renewed challenge for theorists to improve our understanding of this basic three-body quantum-mechanical system.

Dye lasers tunable over the wavelength range 520–700 nm allow access by two-photon transitions from the 2s metastable levels to all n^1S and n^1D levels with $n \geq 6$ and to all n^3S and n^3D levels with $n \geq 4$, including the ionization limits of these series.

Analysis of the experimental results, using quantum defect ideas to model Rydberg series as they converge to the ionization limit, should allow us to determine absolute term energies with MHz accuracy (~0.00003 cm^{-1}), one to two orders of magnitude improvement over present knowledge. Similar two-photon measurements in a discharge afterglow have already been made by Giacobino [8]. Spectroscopy with a beam of He offers the advantage of minimizing pressure-broadening and pressure-induced shifts of the resonances compared to similar spectroscopy with a low-pressure cell.

References

[1] H. E. Ives and G. R. Stillwell, J. Opt. Soc. Am. 28, 215 (1938).

[2] E. H. Pedersen, Phys. Rev. Lett. 42, 440 (1979); and C. Reynaud, J. Pommier, V. N. Tuan, and M. Barat, Phys. Rev. Lett. 43, 579 (1979).

[3] J. J. Snyder and J. L. Hall, in *Laser Spectroscopy*, Ed. by S. Haroche, J. C. Pebay-Peyroula, T. W. Hänsch, and S. E. Harris (Springer-Verlag, Heidelberg, 1975), p. 6.

[4] J. Bailey, K. Borer, F. Combley, H. Drumm, F. Krienen, F. Lange, E. Picasso, W. von Ruden, F. J. M. Farley, J. H. Field, W. Flegel, and P. M. Hattersley, Nature (London) 268, 301 (1977).

[5] G. F. Smoot, M. V. Gorenstein, and R. A. Muller, Phys. Rev. Lett. 39, 898 (1977); S. P. Boughn, E. S. Cheng, and D. T. Wilkinson, Astrophys. J. (Letters) 243, L113 (1981).

[6] M. Aaronson, J. Mould, J. Huchra, W. T. Sullivan III, R. A. Schommer, and G. D. Bothun, Astrophys. J. 239, 12 (1980).

[7] J. W. Farley, K. B. MacAdam, and W. H. Wing, Phys. Rev. A 20, 1754 (1979).

[8] E. Giacobino, private communication (1981).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

An Experiment to Measure Relative Variations in the One-Way Velocity of Light

D. G. Torr and P. Kolen

Department of Physics, Utah State University, Logan, UT 84322

In this experiment two rubidium vapor frequency standards were placed approximately 500 m apart and the phase of their signals compared as a function of time. The diurnal rotation of the earth was used to introduce a change in the direction of propagation of the signal, thereby providing a test of the assumption of isotropy of propagation of electromagnetic radiation. The relative phase difference between clocks was also compared for negligible separation of the clocks. The drift rate changed detectably for the separated clocks, while the round-trip velocity remained constant to within 0.001% c. Typical variations observed in the one-way velocity imply a diurnal modulation of the order of $\pm$ 0.1% to 1.0% c. The relative precision of the measurements amounted to 1 part in 5×10^{13}.

Key words: fundamental constants; precision measurement; velocity of light.

1. Introduction

The assumption of the isotropic propagation of light in the special theory of relativity is traditional. The adoption of a convention involving anisotropic propagation does not necessarily violate the principle of relativity. For example, Winnie [1] formulated three synchrony free principles which include the factual core of the observations which support special relativity. The equations are expressed in ϵ-generalized Lorentz form, which requires no assumptions regarding the one-way velocity of light. Formulated in this generalized way, the special theory of relativity allows for the possibility of anisotropic propagation of electromagnetic radiation, which is not observable via round trip experiments, or any other similar symmetrical arrangement of the propagation path. However, we have not been able to find any experiments reported in the literature which are free of the symmetries which cancel the effects of possible anisotropies in space.

Sources of possible anisotropy in the propagation of light include anisotropies in the distribution of matter in the universe at large, or possibly the existence of an "absolute space" as advocated by Poincáre [2], in his strong theory of relativity, which has at its base the non-relativistic theory of Lorentz [3]. Giannoni [4] for example, has also formulated a spectrum of non-relativistic theories which satisfy the essence of Winnie's [1] synchrony free principles, and Torr and Kolen [5] have established that the special theory and the Poincáre [2] formulation of the Lorentz [3] theory lie at the extreme ends of a spectrum of theories, all of which satisfy the essence of Winnie's [1] principles as modified by Giannoni [4].

There is, therefore, significant theoretical justification for the measurement of the one-way velocity of light, provided that a suitable experiment can be devised for this purpose. In this paper we describe the results of an experiment in which the rotation of the earth is used to introduce changes in the direction of propagation of electromagnetic radiation with the purpose of detecting changes in phase produced by variations in the one-way velocity of light.

2. Historical Problems with the One-Way Experiment

Historically there has been a misconception regarding the feasibility of any experiment designed to measure the one-way velocity of light. Generally it is believed that the experiment is meaningless. The rationale behind this argument proceeds as follows. In order to make the measurement by timing the one-way flight of a light pulse it is necessary to synchronize two clocks separated by some distance d. To do this it is apparently necessary to make some assumptions about the propagation speed as a function of direction, since motion affects the frequencies of clocks, i.e., they cannot be locally synchronized and then separated without some unknown change in phase occurring. Thus a knowledge of the velocity of propagation is required to synchronize the clocks, which introduces a circular argument which in turn renders the concept of the experiment meaningless.

In the experiment reported in this paper we circumvent this problem by not requiring that the clocks be synchronized. We look instead for variations in the relative phases of signals generated by two clocks, which should only occur if the one-way velocity of light is direction dependent. It nevertheless transpires that even the absolute velocity of propagation can be deduced from this experiment as described below.

3. Concept of the Experiment

The experiment that we conducted utilized two Hewlett Packard model 5065A rubidium vapor frequency standards to time the flight of electromagnetic signals across a distance of ~500 m separating the clocks. The experimental arrangement that yielded the best results is schematically illustrated in Fig. 1. Each clock generates a ~1.5 V rms 5 MHz sinewave. The signal from clock A was used to trigger the start input of an HP model 5370A Universal Time Interval Counter. The signal from clock B was used to stop the counter.

The theory behind the experiment argues that if perfect clock stability is assumed, at some arbitrary time

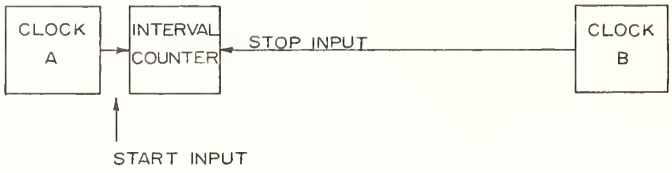


FIGURE 1. *Schematic illustration of the concept of the experiment.*

the interval counter will measure a signal given by

$$\Delta t_1 = \Delta t + \frac{d}{c^+} \tag{1}$$

where Δt represents some initial phase offset between the clocks, $d \simeq 500$ m, and c^+ is the velocity of propagation at that time and orientation of the experiment. The experiment relies on the rotation of the earth to interchange the positions of the two clocks, so that twelve hours later we measure

$$\Delta t_2 = \Delta t + \frac{d}{c^-} \tag{2}$$

where c^- potentially represents a different value for the velocity of propagation of the signal. Hence subtraction of the two measured intervals yields the quantity

$$\delta t = \Delta t_2 - \Delta t_1 = \frac{d}{c^-} - \frac{d}{c^+} \tag{3}$$

$$= d\frac{(c^+ - c^-)}{c^+ c^-} . \tag{4}$$

Naturally if the special theory of relativity is correct $c^+ = c^- = c$ and $\delta t = 0$.

To a first order approximation

$$\delta t = \frac{dv}{c^2} \simeq \Delta t \frac{v}{c} , \tag{5}$$

where $\Delta t = \frac{d}{c}$ (6)

and $v = \frac{c^+ - c^-}{2}$. (7)

Hence $v \simeq \frac{\delta t}{\Delta t} c$. (8)

Since a roundtrip measurement yields Δt, c^+ and c^- can be determined as a function of time.

In reality, however, the simplicity of this concept is marred by errors of measurement which we discuss in section 5.

4. Experimental Arrangement

Figure 2 schematically illustrates the details of the experimental arrangement that was used. The clocks were placed 500 m apart in an east-west orientation. The 5 MHz sine wave signal was propagated via a nitrogen filled coaxial cable maintained at constant pressure at ~2 psi above the ambient atmospheric pressure using a 2 stage regulator. Thermal control was achieved to within ± 1 K over a diurnal cycle by burying the cable at a depth of 5 feet below the surface, and the enclosures housing the equipment 10 feet below the surface. Electrical isolation was achieved by enclosing the equipment in a Faraday cage. Figure 2 also illustrates the dc power supply arrangement used to supply regulated power to within ± 10 mV. The power was introduced into the Faraday cages via electromagnetic interference filters.

The performance characteristics of the rubidium clocks were evaluated by running the clocks at a separation of 1 m. Limitations imposed in the measurement due to clock errors were determined, and requirements for a successful experiment established. These findings are discussed in section 6.

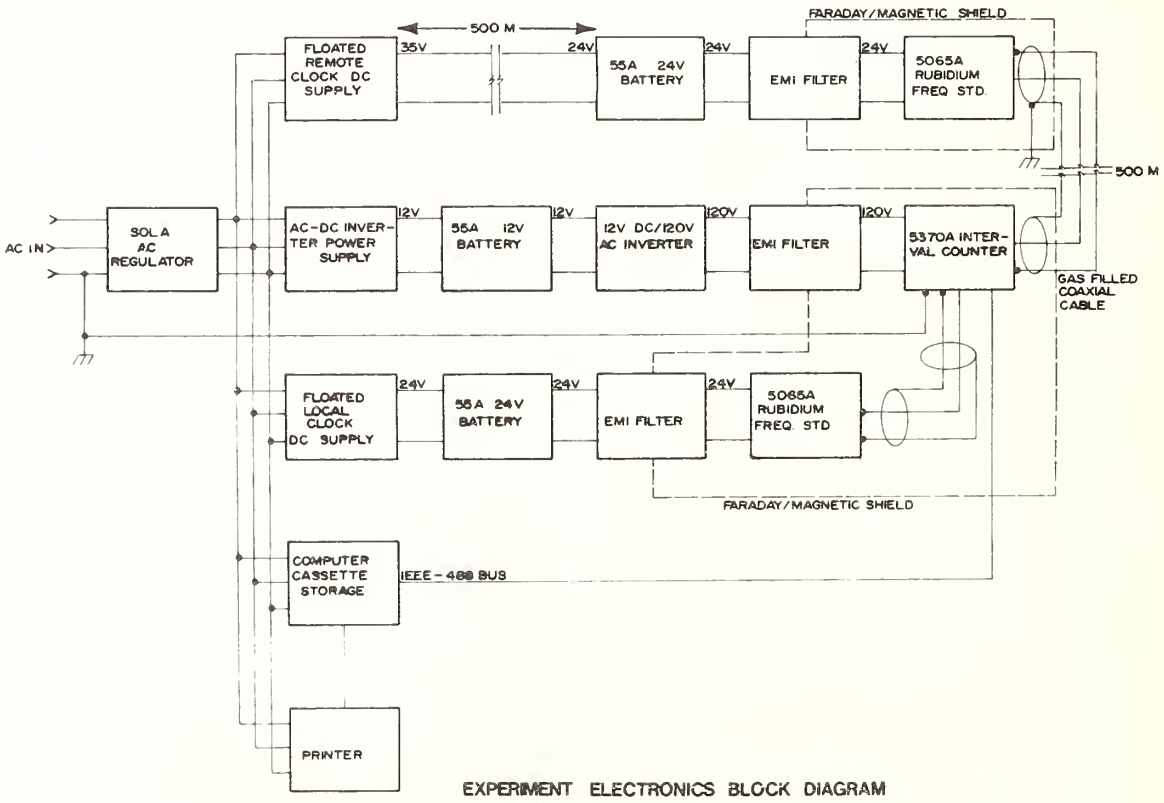


FIGURE 2. *Block diagram illustrating the details of the experimental arrangement.*

5. Evaluation of Errors of Measurement for Rubidium Vapor Frequency Standards

According to manufacturer-supplied information there are two primary sources of error which affect the behavior of rubidium clocks. These are:

a) *Settability*—This represents the practical limit of setting both clock oscillators to the identical frequency. This error accumulates with time.

b) *Long Term Frequency Drift*—This error is introduced by "long term" noise, and produces a non-linear drift with time.

By observing the behavior of the clocks at zero separation (i.e. $\simeq 1$ m) their relative drift rate could be measured and hence the clock errors functionally fitted and accurately defined. According to the manufacturer's specifications [6] the accumulated error of a rubidium clock can be parameterized as follows:

$$T(t) = (1/2)at^2 \pm bt \pm t_o \tag{9}$$

where

$T(t)$ = total accumulated error with respect to a perfect time base,

a = frequency drift rate $\simeq 1 \times 10^{-11}$ per month,

b = $(\frac{f_o}{f_r} - 1)$ = settability = $\pm 2 \times 10^{-12}$ seconds/second,

f_o = initial frequency of oscillator at $t = 0$,

f_r = reference frequency, and

t_o = initial time error at $t = 0$.

To calculate the difference in relative accumulated error between two rubidium clocks, Eq. (9) is modified to

$$\Delta T(t) = 1/2(a_1 - a_2)t^2 \pm b_r t \pm t_o \tag{10}$$

where

$\Delta T(t)$ = relative accumulated error,

$a_{1,2}$ = frequency drift constants of the respective clocks

$b_r = (b_1 - b_2) = \dfrac{f_{10} - f_{20}}{f_r}$, and

$f_{10,20}$ = initial frequencies of the respective clocks at $t = 0$.

It can be seen from Eq. (10) that if $a_1 \simeq a_2$, $\Delta T(t)$ can be approximated by

$$\Delta T(t) \simeq b_r t \pm t_o \tag{11}$$

for a short enough time such that the quadratic term can be neglected. This point is important for using rubidium clocks in this application which will be discussed in detail in the data reduction section.

In order to characterize the behavior of the clocks with repsect to each other, we monitored their performance for a period of ~ 8 weeks at zero separation. Figure 3 illustrates the typical relative drift rates observed. We have plotted the relative accumulated clock error $\Delta T(t)$ for 7 days. These results show a pattern that we have confirmed is typical. The clocks tend to drift linearly for periods of time which amount to several days. These linear drift periods are interspersed with non-linear drift periods were the slope of the drift curve changes rapidly, and then settles down to the linear pattern again.

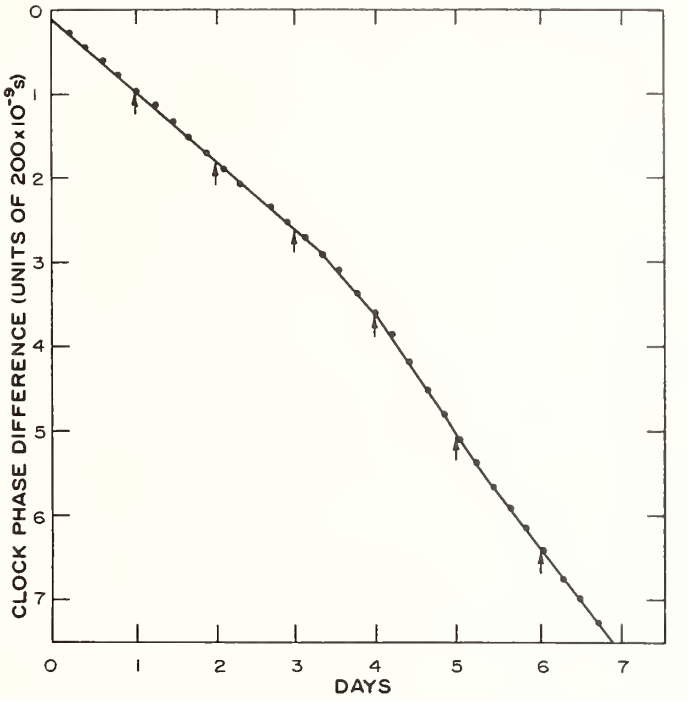


FIGURE 3. *Typical relative drift rate observed for the Rubidium Vapor Frequency Standards over a period of 7 days for zero separation.*

Figure 4 represents a typical linear segment over one day for zero clock separation. Figure 5 shows the residuals, i.e., with the linear trend removed.

While the mean departures from linearity over many days agree with the manufacturer's specifications, these departures occur "suddenly" providing large variations over a few hours which are clearly identifiable. Over the linear segments, the clock error is about a factor of 10 smaller than that specified by the manufacturer. Using the manufacturer's specifications we calculate a mean non-linear component of ~15 ns per day. Typically, however, the observed departure from linearity over 24 hours for a "linear segment" seldom exceeds 1.5 ns, which allows one to make a measurement of relative precision of ~1 part in 5×10^{13}.

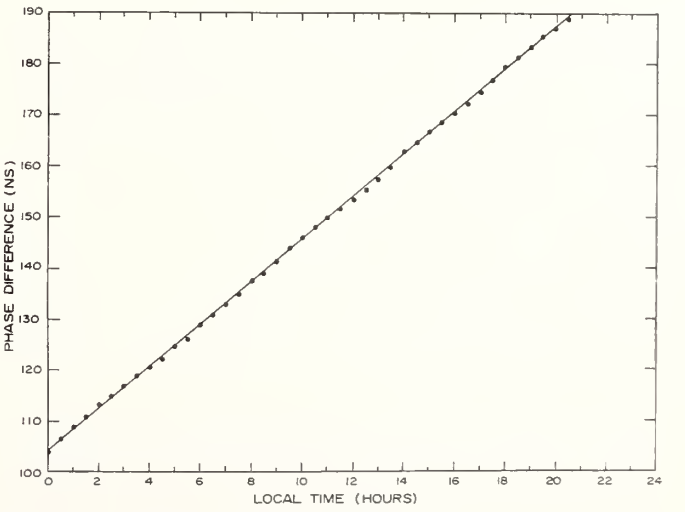


FIGURE 4. *A typical linear drift segment over a 24 hour period for zero separation of the clocks.*

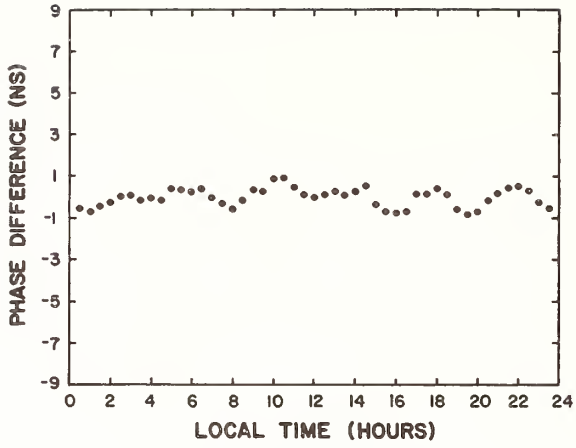


FIGURE 5. *Typical residual drift rate obtained when the linear trend is removed.*

6. Rubidium Results

Several months of data have been obtained for the clocks separated and for zero separation. The separated clocks exhibited a behavior pattern which is significantly different from that observed for zero separation. Very large (~ 10 ns) departures from linearity have been observed on several occasions for the separated clocks. Two examples are given in Figure 6. However, one cannot be absolutely sure that this behavior is not simply another form of the non-linear pattern identified in Fig. 3 even though the clocks have not exhibited this kind of behavior under zero separation conditions. Our results indicate that departures from linearity for the separated clocks remain below threshold (~ 1.5 ns) 30 percent of the time. Amplitudes between 1 and 3 ns are most common. However, not only does the amplitude of the signal vary from day to day, but the phase does also. Since there is no theory available which can account for these variations, we believe that it is essential to repeat the experiment with different clocks such as cesium beams which do not exhibit the non-linear variations peculiar to rubidium vapor frequency standards.

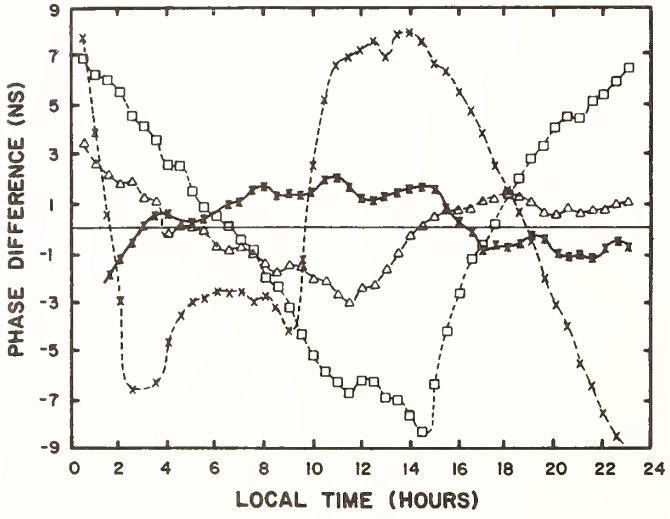


FIGURE 6. *The asterisks show a typical diurnal variation obtained for zero separation when the linear drift component is removed. Results obtained for a typical day for the separated clocks are shown as the triangles. The crosses and squares represent the maximum diurnal variation observed, excluding cases which exhibit sudden "quantum" changes in drift rate which are clearly due to clock errors.*

If these variations are real and the experiment is indicating the presence of a dynamical "absolute space," or the effects of anisotropic distribution of matter in the universe then the motion of the solar system in the galaxy should be embedded in the observed variations. This should be observable as a sine wave modulation of the signal over a 24 hour period. The component of the velocity of the solar system in the ecliptic plane is $\sim 10^5$ m/s^{-1} which translates into an amplitude variation of $\sim \pm 0.5$ ns. In the presence of variations which exceed ± 3 ns for 30 percent of the time, significantly more than a year of integration would be required to detect this signal at a signal-to-noise ratio of 3 to 1. We have therefore used a sample of data restricted to variation amplitudes <3 ns. The time required to obtain a given signal-to-noise ratio (SNR) is given by

$$SNR = \frac{\delta t \sqrt{N}}{\sigma_e} \tag{12}$$

or

$$\text{i.e. } N = \frac{(SNR)^2 \sigma_e^2}{\delta t^2}, \tag{13}$$

where δt = mean signal amplitude,
σ_e = standard deviation of the data, and
N = sidereal day count.
If we let $SNR = 3$,
$\delta_t = 0.5$ ns
$\sigma_e \simeq 3$ ns
$N = 324$ sidereal days at an occurrence rate of 30 percent, which considerably exceeds the number of days of data currently available. Hence the expected signal is not yet detectable with our present arrangement.

Figure 7 shows the result obtained integrating 23 days' data for which the maximum amplitude never exceeded 3 ns. From these results we can place an upper limit on δt of 0.5 ns, i.e.

$$v \lesssim 90 \text{ km/s}^{-1}$$

This fairly closely approximates the component in the ecliptic plane of the velocity of the solar system in the galaxy.

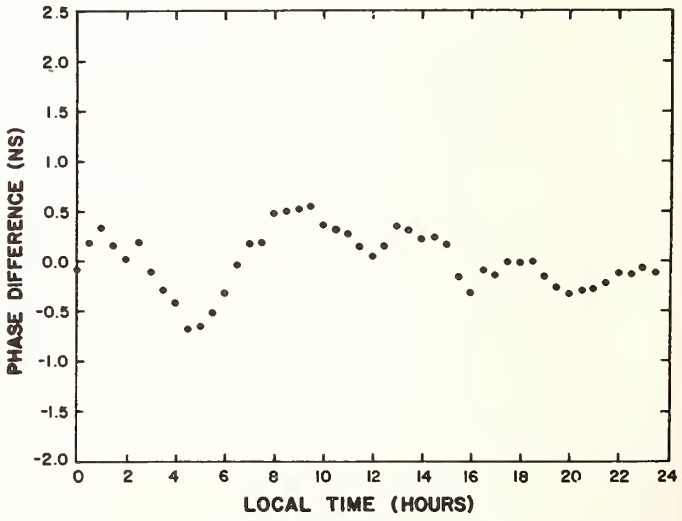


FIGURE 7. *The coherent sum of 23 days' data for the separated clocks for the period February to June, 1981. Summing was carried out using half hour bins.*

7. Conclusion

The main result that has emerged from this work is the demonstration of the viability of a measurement of the one-way velocity of light. Given perfect clocks the absolute one-way velocity could be accurately determined. With rubidium vapor frequency standards integration over 3 years would provide an unambiguous detection of the motion of the solar system in the galaxy, if such motion is indeed detectable by this technique.

The results we have obtained to date exhibit large variations in c (0.1% to 1% c) for the separated clocks, which are not observed (i.e., in the same form) for zero separation. More observational time will be required to unambiguously establish whether the observed variations are indeed due to clock errors or not.

We believe that the ambiguity could be greatly reduced by using cesium beam clocks with a high performance option which significantly improves the clock stability.

References

[1] J. A. Winnie, Phil. Sci. 37, 81 (1970); 37, 223 (1970).
[2] H. Poincáre, C. R. Acad. Sci. 140, 1504 (1905).
[3] H. A. Lorentz, Proc. K. Ned. Akad. Wet. 6, 809 (1904).
[4] C. Giannoni, Found. Phys., 9, 427 (1979).
[5] D. Torr and P. Kolen, Found. Phys. 12, 265 (1982).
[6] Hewlett Packard Company Report: Application Note 52-2, "Time Keeping and Frequency Calibration" (1976).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

High-Energy Gamma Rays Might Be Faster Than Visible Light

Kunio Fujiwara

Institute of Physics, College of General Education, University of Tokyo, Komaba, Tokyo 153, Japan

Our possible confusion between the wave and particle pictures arising from the linearity of the de Broglie relation might be the cause of the divergence in field theories. When we assume that every line segment in *three*-dimensional space has a quantum structure analogous to the atomic chain of atomic distance $2il_0$, the de Broglie relation is exponentially nonlinearized at sufficiently high momenta, and the symmetry between the two pictures is broken. It has been shown that, under such situations, the field theories automatically become finite *without changing their conventional form*. The present theory predicts that high-energy γ-rays would be faster than visible light, though the detectability of the effect depends on the value of the unknown constant l_0.

Key words: de Broglie relation; ether; light velocity; local field; quantization of space; regularization; special relativity.

1. Introduction

It is well known that, when the divergent integrals are renormalized, quantum electrodynamics (QED) can predict basic electromagnetic quantities with high accuracy. The agreement between the theory and the experiments is, in fact, excellent. What should we think is the physical meaning of this success of renormalized QED, then? In the author's opinion, the success implies that the conventional *form* of QED is, despite its divergence difficulties, already beautiful enough to be regarded as established. Any complication of the form would no longer be necessary upon the regularization of this theory. From this point of view, the author would show here that there exists at least one way to regularize QED as well as the other quantum field theories without any change in their conventional form.

We shall begin our discussion by reexamining some of our fundamental concepts. According to our understanding of the wave-particle duality, the wave and particle pictures are transformable into each other through the linear relations

$$\mathbf{p} = \hbar\mathbf{k} \quad \text{and} \quad E = \hbar\omega \,. \tag{1}$$

The equivalence between the wave and particle quantities assumed in these relations often implies a nearly complete confusion between the wave and particle pictures.

We know, on the other hand, two fundamental theories, i.e., the special theory of relativity and the quantum theory. In the former theory, relative motions between solid frames are considered; we should therefore understand that the former is implicitly based on the particle picture of matter, while the latter apparently emphasizes its wave properties.[1] A divergence problem arises, however, when we intend to unify these two theories into relativistic quantum field theories. Many physicists have treated this problem mainly from technical points of view, and it is usual to see that the result remarkably complicates the form of field theories.

The above divergence could, however, be a warning that our understanding of the wave-particle duality (cf. Eq. (1)) is inadequate at least at extremely high energies. In this connection it is worth noticing that we can assume, without conflicting with any past observations, that our *3-dimensional* space has an unknown quantum structure characterized by a very short length, l_0. If we really assume so, since only *waves* should be sensitive to such a structure, the equality between the wave and particle pictures would be broken at sufficiently high energies. Thus we would have

$$\mathbf{p} = \hbar\mathbf{k}\,\xi(k, l_0) \quad \text{and} \quad E = \hbar\omega \tag{2}$$

instead of Eq. (1). Here, $\xi(k, l_0)$ is the nonlinearization factor whose functional form will be proposed later. In spite of such a change in the de Broglie relation, however, the E-p relation for light, for instance, would remain in the usual form

$$E = cp \,, \tag{3}$$

because any particle properties are thought to be essentially insensitive to the quantum structure of space. Then substituting Eq. (2) into Eq. (3), we get a nonlinear ω-k relation for light

$$\omega = ck\,\xi(k, l_0) \,. \tag{4}$$

The set of relations Eqs. (3) and (4) suggests that a perfect 4-dimensional description, such as in the conventional theory of relativity, would be exact only in the particle picture, and that our possible confusion between the wave and particle pictures should be carefully solved whenever we want to discuss the high-energy behavior of field theories. In a recent paper [1] the author studied this problem by starting from the analogy of a linear chain, the simplest example of a quantized medium, and showed that the relativistic quantum field theories automatically become finite without changing their conventional form. In the following, the basic ideas will be briefly outlined with particular emphasis on the importance of the precise measurement of the velocity of high-energy γ-rays.

[1] The opposite appears to be true for light. In the theory of relativity, light is treated as wave signals, while the quantum theory emphasizes its particle properties.

2. The Outline of the Theory

An atomic linear chain can be regarded as a continuous string as long as the wavelength, $2\pi/k$, of the elastic wave propagating on the chain is much longer than the atomic distance a. The wave equation is then given by

$$(\partial^2/\partial x^2 - v^{-2}\partial^2/\partial t^2)u(x, t) = 0 , \quad (5)$$

which leads to a linear ω-k relation, $\omega = vk$, similar to that for light waves in vacuum. When the wavelength becomes comparable to or shorter than a, however, the effect of the "quantum structure" of this medium becomes important. In such cases, the ω-k relation is known [2] to be modified to

$$\omega = (2v/a)\sin(ak/2) . \quad (6)$$

The differential equation leading to this nonlinear ω-k relation will be obtained if we replace $\partial/\partial x$ in Eq. (5) by

$$(2i/a)\sin[-i(a/2)\partial/\partial x] . \quad (7)$$

Returning to our problem, we first assume that the above example is just suggesting how to quantize 3-dimensional space. Imagine that our 3-dimensional space is filled up with a special medium (we call it *ether*) in which any two points are interconnected by an "atomic chain" having an atomic distance $2il_0$ (l_0 real). The assumption of the *pure-imaginary* atomic distance appears to be essential for the successful regularization of field theories. The momentum operator in field equations would then be modified to

$$-i\hbar\nabla \sum_{n=0}^{\infty} [(-l_0^2\nabla^2)^n/(2n+1)!] , \quad (8)$$

when described by the observers fixed to the ether. (A similar, but 4-dimensional operator has been proposed by Danos, Greiner, and Rafelski [3].) Needless to say, for light and material waves whose wavelengths are long compared with l_0, these chains would behave as "continuous strings," and the operator, Eq. (8), would thereby reduce to $-i\hbar\nabla$. In such cases, the concept of ether as the medium of the waves would lose its importance, and we could be convinced of the absence of the ether as we have been so far.

When applied to a plane wave $\exp[i(\mathbf{k}\cdot\mathbf{x} - \omega t)]$, the operator, Eq. (8), will give the nonlinearized de Broglie relation

$$\mathbf{p} = \hbar\mathbf{k}[(\sinh kl_0)/kl_0] \equiv \hbar\mathbf{k}\,\xi(kl_0) \quad (9)$$

as preannounced in Eq. (2). The energy operator $i\hbar\partial/\partial t$ and the energy-frequency relation $E = \hbar\omega$ are, on the other hand, kept as usual because they are independent of 3-dimensional space quantization. The resulting ω-k and E-p relations for light in vacuum are shown in Figs. 1a and 1b, respectively; the former leads to the kl_0-dependent phase and group velocities

$$C_{ph} = c\,\xi(kl_0) \quad \text{and} \quad C_g = c\cosh kl_0 . \quad (10)$$

Our next problem is to reformulate the special theory of relativity. We start from the following two postulates:

(I) Every physical law (except that describing the quantum structure of space itself) is the same in all inertial frames.

(II) Light behaves isotropically, and the phase and group velocities of each individual light wave are invariant to all the inertial observers.

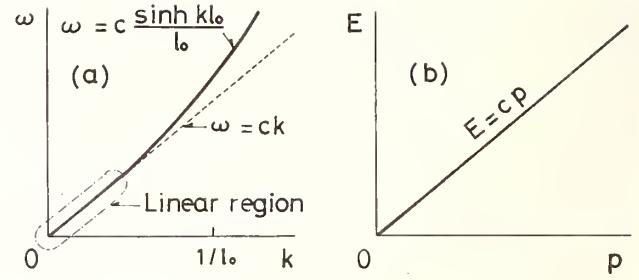


FIGURE 1. *The ω-k (solid curve in (a)) and E-p (b) relations for light in vacuum when the quantized ether is introduced into three-dimensional space. Notice the nonlinearity of the former relation. We may assume that we are now in the linear region shown in (a).*

Then, the first postulate requires that, in any inertial frame, the **p**-**k** relation should be of the form[2]

$$\mathbf{p} = \hbar\mathbf{k}[(\sinh kl)/kl] = \hbar\mathbf{k}\,\xi(kl) . \quad (11)$$

We have already assumed that $l = l_0$ in the frame fixed to the ether. In frames moving with respect to the ether, however, we can no longer regard l as a constant[3]. Thus, so far as l is concerned, we must admit the speciality of the "ether frame." This fact does not, however, put any essential restriction onto the covariant description of field theories, because l will appear in the theories only in the form kl, and we can reasonably assume that

$$kl = \text{invariant} \quad (12)$$

to all observers. Since light velocity is found to depend on kl, Eq. (12) accords also with the second postulate.

Next, to solve the possible confusion between the wave and particle pictures, we define a vector in the "reduced space" as

$$\mathbf{X} = \mathbf{x}/\xi(kl) , \quad (13)$$

where $\mathbf{x}$ is the vector in the usual space. Since we find

$$-i\hbar(\partial/\partial\mathbf{X})\phi_{\mathbf{k}} = \mathbf{p}\phi_{\mathbf{k}} \quad \text{and} \quad -i(\partial/\partial x)\phi_{\mathbf{k}} = \mathbf{k}\phi_{\mathbf{k}} , \quad (14)$$

where $\phi_{\mathbf{k}} = \exp[i(\mathbf{k}\cdot\mathbf{x} - \omega t)]$, we can understand that $\mathbf{X}$ is quantum-theoretically conjugate with the particle quantity, $\mathbf{p}$, while $\mathbf{x}$ is conjugate with the wave quantity, $\mathbf{k}$.

The Lorentz transformations are to be generalized in the way that they leave $\mathbf{x}^2 - c^2\xi^2(kl)t^2$ and kl invariant. Under such transformations (hereafter referred to as GLT), the following quantities *in the particle picture* form 4-vectors:

$$X_\mu \equiv (\mathbf{X}, ict), \quad D_\mu \equiv \partial/\partial X_\mu, \quad \text{and} \quad P_\mu \equiv (\mathbf{p}, iE/c) . \quad (15)$$

On the other hand, the quantities *in the wave picture* such as

$$x_\mu \equiv (\mathbf{x}, ict), \quad \partial_\mu \equiv \partial/\partial x_\mu, \quad \text{and} \quad k_\mu \equiv (\mathbf{k}, i\omega/c) \quad (16)$$

---

[2]Note that in particle scattering, momentum conservation would hold but wave-vector conservation would no longer hold.

[3]According to Eq. (44) in Ref. [1], we find

$$l = l_0(1 + B\cos\theta)/(1 - B^2)^{1/2}$$

for light waves, where B is the parameter analogous to $\beta = v/c$, and θ is the angle between the wave vector $\mathbf{k}$ and the direction of motion of the observer with respect to the ether. Note that l is independent of $k = |\mathbf{k}|$. This formula will hold also for material waves when $\hbar k\,\xi(kl) \gg mc$ is fulfilled in the ether frame. Needless to say, $kl = 0$ can be a good approximation if $\hbar k\,\xi(kl) \lesssim mc$.

can no longer form 4-vectors. Nevertheless we find that

$$k_\mu x_\mu = \text{inv.},\ d^4x = \text{inv., and } d^4k = \text{inv.}, \qquad (17)$$

which are easily proved if the invariance of $P_\mu X_\mu$, $\xi(kl)$, d^4X *etc.* is noticed. (cf. Appendix A in Ref. [1].) The invariance of the quantities shown in Eq. (17) will play an essential role in regularizing field theories without changing their form.

The covariance under the GLT is now to be the guiding principle in formulating field theories. The new field equations, for instance, are derived from the conventional ones by making the replacement $\partial_\mu \to D_\mu$. Nor will any formal change be required upon the reformulation of local field operators though, in return for that, a drastic change will be required in our concept of "localization" itself as mentioned below.

Consider, as a simplified example, a classical local field

$$f(x_\mu) \equiv (2\pi)^{-4} \int a(P_\mu) \exp(ik_\mu x_\mu) d^4k \qquad (18)$$

defined by an observer O in his x_μ space. Due to the replacement $a(k_\mu) \to a(P_\mu)$ and also to the invariance of $k_\mu x_\mu$ and d^4k, this expression is covariant under the GLT. Therefore, another observer O′ would also write it down in the form

$$f(x_\mu') = (2\pi)^{-4} \int a(P_\mu') \exp(ik_\mu' x_\mu') d^4k' \ . \qquad (19)$$

To that observer, however, the field of Eq. (19) can no longer be local because x_μ' is now dependent on the kl-value[4] of the Fourier component of the field. A typical example of x_μ' is given [1] by

$$x_1' = [x_1 - B\xi(kl)ct]/(1 - B^2)^{1/2}\ ,\ x_2' = x_2\ ,\ \text{etc.} \qquad (20)$$

It follows then that a local field defined by O at a point x_μ is accepted by O′ as a quasi-nonlocal field whose Fourier components are distributed on a curve in his x_μ' space (see Fig. 2), suggesting that "localization" might be a relative concept. Such a change in "localization" appears to have some resemblance to the well-known change in "simultaneity" which was required upon the establishment of the special theory of relativity. Equation (20) also suggests that the microcausality principle should be reformulated as

"The effect of the kl-component of a signal emitted at $x_\mu = 0$ cannot reach those points for which $x^2 - c^2\xi^2(kl)t^2 > 0$."

Our new field theories appear not to violate this requirement.

As expected from the replacement $\partial_\mu \to D_\mu$, the Fourier transforms of the propagators for free fields are to be written in terms of the particle quantity, P_μ. On the other hand, the wave quantity d^4k should be used as the volume elements of the Feynman integrals; this is obvious from the expression of the field operators as illustrated in Eqs. (18) and (19). If our confusion between the wave and particle pictures is solved in such a way, all the Feynman integrals would become finite because we find that $p = \hbar k\,\xi(kl) \sim \hbar(\exp kl)/2l$ for large k's.

3. On the Experimental Tests

The new theory is broad enough to include the conventional theory as its limiting case. The difference between the two will be appreciable only at sufficiently high energies, and precise measurements of the velocities of high-energy γ-rays or massive particles would be most effective for the test of the new theory. According to this theory, the group velocity of a light wave of photon energy E would be given by

$$C(E) = c\cosh[\sinh^{-1}(l_0E/\hbar c)] \simeq \begin{cases} c[1 + (l_0E/\hbar c)^2/2] & \text{for } E \ll \hbar c/l_0 \\ l_0E/\hbar & \text{for } E \gg \hbar c/l_0 \end{cases}, \qquad (21)$$

where l_0 is the possible third natural constant to be searched for. Equation (21) is also applicable to a particle of rest mass m if $p \gg mc$. In Fig. 3, the relative increase in light velocity $[C(E) - c]/c$ is plotted against E for various values of l_0. (The possible increase in light velocity with increasing photon energy has been discussed also by Pavlopoulos [4].) There are a few velocity

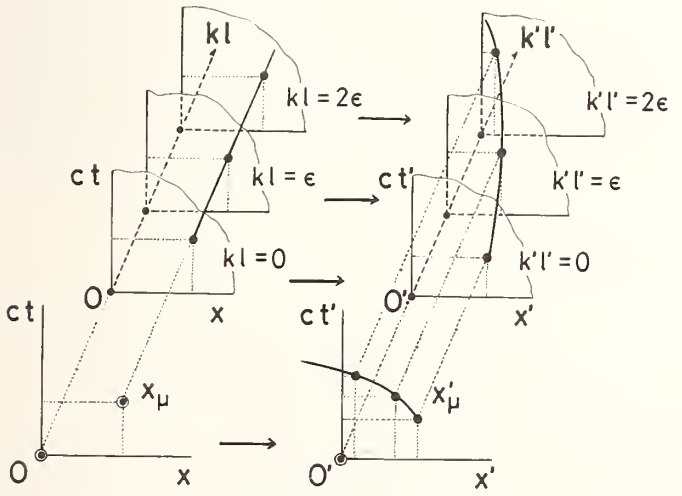


FIGURE 2. *An illustration of the transformation properties of the coordinate x_μ, which suggests that "localization" might be a relative concept. The new situation would be more easily understood if the kl-axis is introduced as shown in the upper half of the figure. Recall the invariance of kl.*

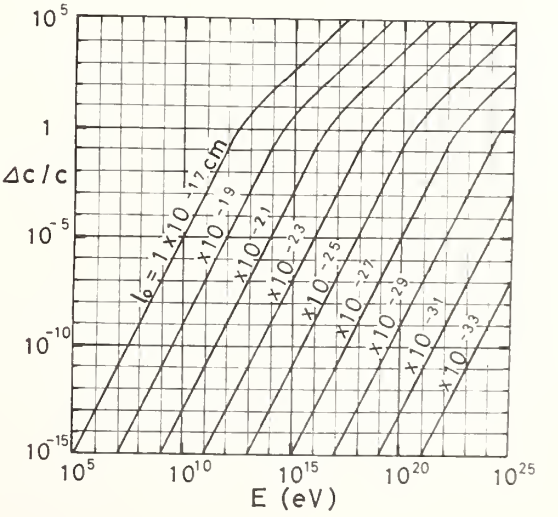


FIGURE 3. *Possible relative increase in light velocity $\Delta c/c$ plotted against the photon energy E for various values of l_0.*

[4]Note that the transformation of x_μ obeys $X_\mu^2 = \text{inv.}$ and $kl = \text{inv.}$

measurements at high energies [5, 6, 7]. Brown *et al.* [5] compared the velocities of 7 GeV photons and of 11 GeV electrons with that of visible light, and obtained the relative velocity differences, $[c\,(\mathrm{GeV}) - c\,(\mathrm{eV})]/c\,(\mathrm{eV}) = (1.8 \pm 6) \times 10^{-6}$ and $(c - v_e)/c = (-1.3 \pm 2.7) \times 10^{-6}$, respectively. Since they found no effect exceeding their smallest experimental uncertainty, 2.7×10^{-6}, we may conclude that l_0 is shorter than 5×10^{-18} cm.

References

[1] K. Fujiwara, Foundations of Phys. **10**, 309 (1980) and Contr. Paper No. 8 at 20th Int. Conf. High Energy Phys. (Madison, USA; July 1980).

[2] C. Kittel, *Introduction to Solid State Physics*, 2nd ed. (Wiley, New York, 1956), p. 104.

[3] M. Danos, W. Greiner, and J. Rafelski, Phys. Rev. D 6, 3476 (1972); and Z. Phys. 258, 147 (1973).

[4] T. G. Pavlopoulos, Phys. Rev. **159**, 1106 (1967); and Nuovo Cimento B 60, 93 (1969).

[5] B. C. Brown, G. E. Masek, T. Maung, E. S. Miller, H. Ruderman, and W. Vernon, Phys. Rev. Lett. **30**, 763 (1973).

[6] Z. G. T. Guiragossian, G. B. Rothbart, M. R. Yearian, R. A. Gearhart, and J. J. Murray, Phys. Rev. Lett. **34**, 335 (1975).

[7] T. Alväger, F. J. M. Farley, J. Kjellman, and I. Wallin, Phys. Lett. 12, 260 (1964).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Fiber Optic Ring as a Gravitational Wave Detector*

C. L. Mehta, D. Ranganathan, and G. Bose

Physics Department, Indian Institute of Technology, New Delhi–110016, India

We suggest the use of a Sagnac interferometer made of a fiber optic ring for detecting gravitational waves. Any distortion in the geometry of the ring on account of a gravitational wave is detectable as a phase shift between the counter propagating optical signals. In order to discriminate against angular velocity variations and any other local fluctuations a pair of orthogonal coils is to be used and the system rotated about an axis lying symmetrically between the two coils. It is estimated that a stabilized He-Ne laser used as the source and fiber coils of 1 m diameter with 100 turns rotating with an angular velocity of about 1000 rad/sec will provide adequate sensitivity.

Key words: fiber optic gyroscope; gravitational wave detector; Sagnac interferometer.

Soon after the advent of the laser, its use in rotation sensing devices was suggested and successfully demonstrated [1]. Optical gyroscopes using ring lasers have, more or less, been perfected and are soon expected to be standard devices for navigational purposes. Considerable interest has also been shown in fiber optic gyroscopes [2]. The basic idea involved in these rotation sensing devices is the Sagnac effect [3] first noted in 1913. Based on this effect, Haughan [4] *et al.* proposed a method of testing general relativity using a ring laser. In the present paper, we suggest, as a variant of this method, the use of a fiber optic ring for detecting gravitational waves. We first give a brief description of the Sagnac interferometer consisting of a rotating fiber optic ring. We then show how a pair of such fiber optic rings mounted suitably may be used to detect small changes in length caused by a gravitational wave.

Figure 1 illustrates the principle of a Sagnac interferometer. A laser beam is split into two beams by a beam divider and sent through a fiber optic coil of m turns and radius R in counter propagating directions. We also introduce a modulator in one of the beams as indicated in the diagram. Effectively, it changes the wave length of the clockwise propagating beam from λ to $\lambda - d\lambda$. The wave length of the counterclockwise propagating beam being λ. If the coil is rotating with an angular velocity Ω about its axis, one may readily see that the net phase shift produced between the two counterpropagating beams is to the first approximation given by

$$\Delta\phi = \frac{2\pi R}{\lambda^2} d\lambda + \frac{4\pi R^2 m\,\Omega}{\lambda c} . \qquad (1)$$

(It is being assumed that $d\lambda \ll \lambda$ and $\Omega R \ll c$.) If the modulation $d\lambda$ is so adjusted that there is no net phase shift ($\Delta\phi = 0$), then we find that

$$d\lambda = \frac{4\pi R^2}{2\pi R} \frac{m\,\lambda}{c} \Omega = \frac{4A}{P} \frac{m\,\lambda}{c} \Omega , \qquad (2)$$

where A is the area and P the perimeter of the coil cross-section. The measurement of $d\lambda$ thus enables us to measure Ω, i.e., we may use the interferometer as a gyroscope. Alternatively, knowing Ω we may determine the ratio of A, the area enclosed by the coil, to P, its perimeter.

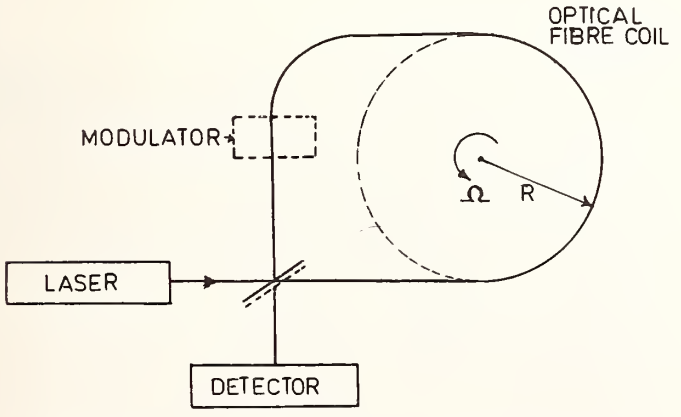


FIGURE 1. *Schematic diagram of a Sagnac interferometer.*

The main effect of a plane gravitational wave is to distort the geometry in the directions transverse to its wave vector. Figure 2 illustrates the direction of "tidal forces" due to a gravitational wave of e_+ polarization propagating in the z-direction [5]. Its effect on an object of circular cross-section is shown in Fig. 3. The change in length in the transverse direction is proportional to the strength of the gravitational wave. If h in suitable units denotes the norm of a typical element of the perturbation metric, then [5]

$$\frac{dR}{R} = h . \qquad (3)$$

Such a change is then detectable through a Sagnac interferometer as described below.

Two Sagnac interferometers are mounted orthogonal to each other on a shaft whose axis is symmetrically placed between the planes of the two coils as shown in Fig. 4. The input and the output coupling can be made by bringing the fiber ends out along the rotation axis. The shaft is rotated about its axis at an angular velocity Ω. Each coil is affected by the component of the angular velocity normal to its plane. In absence of any gravitational wave, two modulating signals for a null phase shift in the respective coil would be the same. Consider now a gravi-

*Supported in part by the Department of Electronics, Government of India, New Delhi.

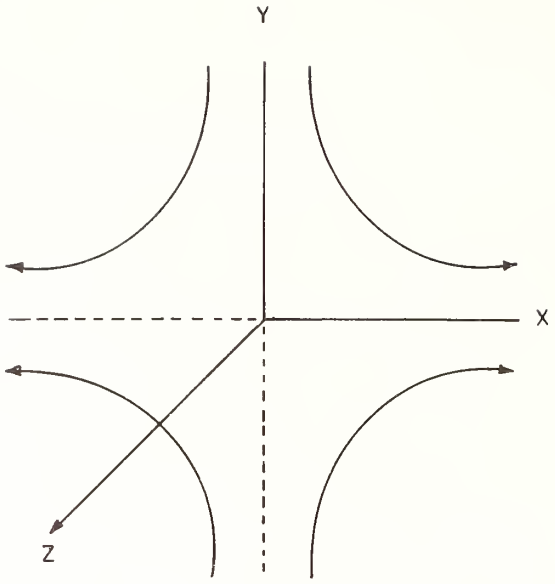


FIGURE 2. *Direction of "tidal forces" due to a gravitational wave of e_+ polarization propagating along z-axis.*

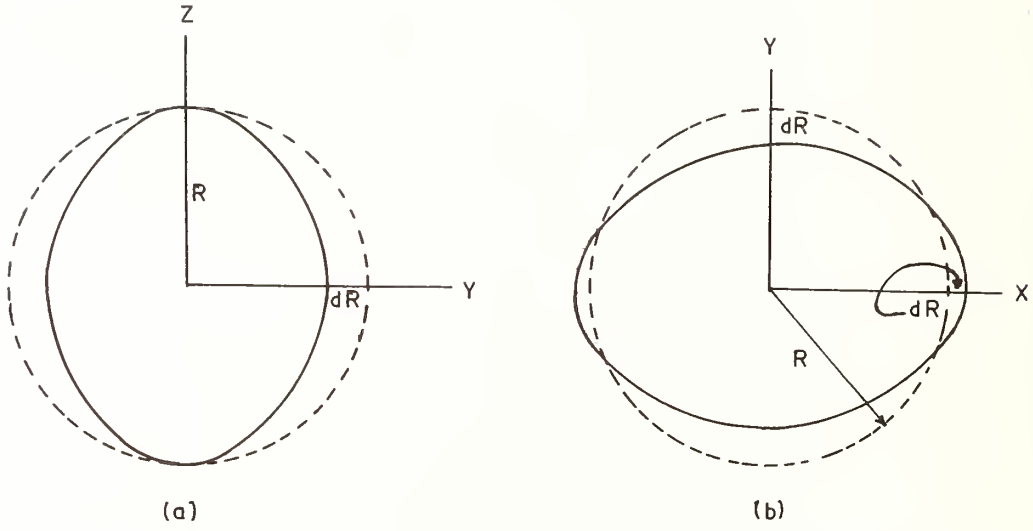


FIGURE 3. *Distortion of circular cross-section (a) in the yz-plane, (b) in the xy-plane, due to a gravitational wave of e_+ polarization propagating along z-axis.*

tational wave along oz, normal to Ω. When the line of intersection of the planes containing the two coils (line KK' as shown in Fig. 4) coincides with oz, coil 1 gets distorted into an ellipse of semi-major axis R and semi-minor axis $R - hR$, whereas coil 2 gets distorted into an ellipse of $R + hR$ and R as semi-major and semi-minor axes, respectively. In this position the difference of the modulation signals in the two coils is maximum. On the other hand, when KK' is perpendicular to oz, both coils are affected symmetrically and hence the difference of the modulation signals is zero. Assuming Ω^{-1} to be much less than the period of a typical gravitational wave, the output modulation difference will be an ac signal of twice the angular frequency of rotation of the shaft. Its amplitude is proportional to the strength of the gravitational wave. The system displays maximum sensitivity in the plane normal to the axis of rotation and zero sensitivity for waves travelling along the axis.

The advantage of mounting both coils on a common shaft is that any fluctuation in the angular velocity of the system affects both coils equally and thus do not show up in the output. Similarly, any variation on account of thermal changes which affect both coils equally can be ignored. In any case they will usually have a frequency different from the modulation difference signal, and hence could easily be eliminated. For further efficiency one may have to place the whole system in a temperature controlled environment.

In the above analysis, we considered a fiber optic ring of circular cross-section. However, it may readily be modified to include other geometrical cross-sections such as a triangular one etc. In fact, in place of a fiber optic ring one may use a ring laser gyroscope directly with minor modifications.

The probable estimate of the gravitational wave strength is [6]

$$\frac{dR}{R} = h \sim 10^{-12} - 10^{-16} \tag{4}$$

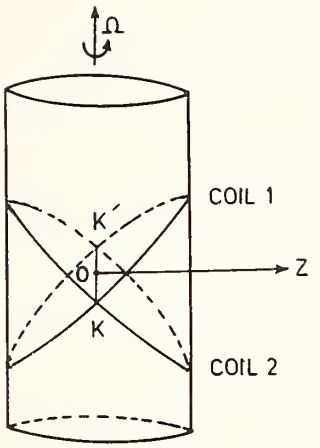


FIGURE 4. *Schematic diagram of the suggested gravitational wave detector.*

for waves with a period of $1 - 10^3$ sec. Let us use a pair of fiber optic coils of 0.5 m radius and 1000 turns each arranged as described above and the system is rotated at 10^3 rad/sec. If a He-Ne laser is used then the minimum detectable phase shift for such a system assuming shot noise limited detection [2, 7] is 10^{-8} fringe. If we ignore the modulation in Eq. (1), we find that

$$d(\Delta\phi) = \frac{8\pi R^2 m\,\Omega}{\lambda c}\,\frac{dR}{R}\,. \qquad (5)$$

Taking $d(\Delta\phi) \sim 10^{-8}$ and using typical values for the other parameters as given above, it is seen that the apparatus will be sensitive for measuring $dR/R \sim 10^{-13}$. Further improvement can be made by increasing the number of turns or the angular velocity of rotation. Also one may use a laser gyroscope which operates at the quantum limit [8]. It is estimated that a sensitivity adequate to measure $h \sim 10^{-18}$ could be obtained using a ring laser gyroscope. Fiber optic gyroscopes are less sensitive at present mainly on account of back scattering in the fiber.

References

[1] F. Aronowitz, in *Laser Applications*, Vol. 1, Ed. by M. Ross (Academic Press, New York, 1971) p. 134.
[2] V. Vali and R. W. Shorthill, Appl. Opt. 15, 1099 (1976).
[3] G. Sagnac, C. R. Acad. Sci., 157, 708 (1913).
[4] M. P. Haugan, M. O. Scully, and K. Just, Phys. Lett. 77A, 88 (1980).
[5] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, (Freeman, San Francisco, 1973) p. 1212.
[6] D. H. Douglass and V. B. Braginsky, in *General Relativity: An Einstein Centenary Survey*, Ed. by S. W. Hawking and W. Israel, (Cambridge University Press, Cambridge, 1979)
[7] S. Ezekiel and S. R. Balsamo, Appl. Phys. Lett. 30, 478 (1977).
[8] T. A. Dorschner, H. A. Haus, M. Holz, I. W. Smith, and H. Statz, IEEE, J. Quantum Electron. 16, 1376 (1980).

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

A Possible Laser Gravitational Wave Experiment in Space

J. E. Faller* and P. L. Bender*

Joint Institute for Laboratory Astrophysics, National Bureau of Standards and University of Colorado, Boulder, CO 80309

An investigation has been started of possible designs for a laser gravitational wave experiment with baseline lengths of roughly 10^6 km or longer. The objectives of the experiment are to search for narrow-band signals with periods of seconds to hours, for pulses of gravitational waves, and for broad-band background radiation. One of the main goals is to detect signals from known rotating binary stars, such as Am CVn, WZ Sge, and i Boo. The corresponding gravitational wave periods are 8.76, 40.5, and 193 minutes. The expected strain amplitudes are roughly 0.4, 0.5, and 6×10^{-21}, respectively, which correspond to equivalent accelerations of 50, 4, and 2×10^{-19} g[1] for a 10^6 km baseline. The main uncertainty in the expected signal strengths comes from the uncertainty in the distances. In view of the extremely tiny equivalent accelerations, care will be needed in designing the experiment in order to minimize spurious accelerations due to forces other than the gravitational attraction of the sun and planetary bodies.

The first approach considered [1] was to make laser heterodyne distance measurements over two non-parallel baselines which remain equal in length to 10^{-8} over the duration of a single experiment. For a laser stability in a 1 cycle/h bandwidth of 5×10^{-14} or better, periodic changes in the difference in length of the two baselines could be measured with roughly 10^{-22} sensitivity in one day. Three spacecraft equipped with laser transmitters and receivers are required. The laser beams would be sent both ways over each baseline, and the beat signals detected at each end. With 10 mW of transmitted laser power in the visible from each satellite, the required diameter of the optical systems is 50 cm.

One geometry considered was to have a central spacecraft at the L1 (or L2) libration point of the earth-sun system, which is about 10^6 km from the earth, and the other two spacecraft 1.5×10^8 km away at the L4 and L5 libration points. It is feasible to arrange the initial conditions so that the baselines from the central spacecraft to the other two stay equal to 10^{-8} over moderately long times. However, the number of signal counts detected per second would be quite low, and the propulsion capability necessary to reach the L4 and L5 libration points in one year is substantial.

A second geometry discussed involved putting two spacecraft about 1.4×10^6 km apart in coplanar one year period circular orbits around the sun, which define a reference plane, and the third spacecraft in a similar orbit which is inclined by about 0.3 degrees with respect to the reference plane [2]. With proper phasing of the orbital positions, the third spacecraft will be near the highest point above the reference plane when the other spacecraft are equally far ahead of and behind it. The distances from the third spacecraft to the other two would be 10^6 km, and would stay equal to 10^{-8} over roughly a day. With this separation the signal to noise would be high enough so that the lasers in the first two spacecraft could be phase locked to the signals arriving from the third spacecraft. However, thrusts would have to be applied to the third spacecraft about every day in order to reestablish the proper geometry.

Another approach which we would like to suggest is to let the two baselines differ in length by up to perhaps one part in 10^3, but measure the apparent changes in the length L of one baseline as well as the changes in the difference in length ΔL. The differences between the measured apparent changes in the baseline length L with periods of seconds to hours and those expected from celestial mechanics would be used to determine the laser frequency variations. This additional information makes it possible to correct for the laser frequency variations in determining whether changes in ΔL occur due to gravitational waves. The only restriction in interpreting the results is the implicit assumption that the change in L due to the gravitational waves is not orders of magnitude larger than the change in ΔL.

With the new approach, we are exploring a different geometrical arrangement with 10^6 km spacecraft separations which would allow the experiment to run for long periods of time without orbit corrections. The main spacecraft would be placed in a 1 year period circular orbit, which defines the reference plane. The other two spacecraft would be placed in orbits with eccentricities $e = 1/300$ and inclinations with respect to the reference plane of very nearly $\sqrt{3}e$. By proper phasing, the other spacecraft will appear to rotate about the main spacecraft with an annual period in a plane whose normal is tipped downward by 30 degrees with respect to the direction toward the sun.

The most difficult design problem for the proposed gravitational wave experiment probably will be shielding out or compensating for perturbing forces on the spacecraft due to fluctuations in the solar wind pressure and other effects. The magnitude of the fluctuating accelerations at the frequencies of interest will be roughly 10^{-13} g. A disturbance compensation system (DISCOS) which reduced perturbations from roughly 10^{-8} g to 5×10^{-12} g even at zero frequency was flown on the TRIAD satellite in 1972. Initial studies indicate that an improved DISCOS can be designed which would reduce spurious accelerations at the frequencies of interest to the necessary level in a 1 cycle/h bandwidth of less than 10^{-18} g. However, attention will have to be placed on avoiding the many sources of perturbations in such systems.

---

*Staff member, Quantum Physics Division, National Bureau of Standards.

[1] In this context, the gravitational acceleration at the earth's surface g, is used as a unit of acceleration.

References

[1] P. L. Bender, R. W. P. Drever, J. E. Faller, and R. Weiss, unpublished manuscript, 1977; referred to by R. Weiss in *Sources of Gravitational Radiation*, Ed. by L. L. Smarr (Cambridge Univ. Press, Cambridge, 1979), p. 26.

[2] R. Decker, J. L. Randall, P. L. Bender, and J. E. Faller, in *Active Optical Devices and Applications*, SPIE Vol. 220 (Society of Photo-Optical Instrumentation Engineers, Bellingham, Washington, 1980), p. 149.

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Author Index

	Page
Amin, S. R.	117
Anderson, J. T.	359
Arditi, M.	29
Baba, S.	529
Baird, K. M.	35
Bairnsfather, H.	469
Barker, P. H.	345
Bartlett, D. F.	643
Basov, N. G.	65
Becker, P.	317, 321
Becker, W.	99
Bell, G. A.	445
Bender, P. L.	689
Benjamin, D. J.	469
Bergquist, J. C.	93
Bertinetto, F.	69
Blatt, R.	99
Bliek, L.	553
Bollinger, J. J.	141
Borchert, G. L.	331
Bose, G.	685
Bouchareine, P.	81
Boulanger, J.-S.	21
Braginsky, V. B.	109
Braun, E.	535
Breeden, T.	153
Breitenberger, E.	667
Brillet, A.	73
Brodsky, S. J.	249
Burghardt, B.	49
Cabrera, B.	359
Cage, M. E.	539
Caldwell, C. D.	117
Camani, M.	223
Cérez, P.	73
Chan, H. A.	601
Chan, K. K.	127
Chen, Y. T.	581
Cheung, W. S.	653
Clothier, W. K.	469
Coarasa, M.-J.	453
Cohen, E. R.	391
Colclough, A. R.	263
Conti, R.	207
Cordiale, P.	69
Currey, M. F.	469
Daams, H.	21
Daniel, H.-U.	45
Davidson, W.	635
De Boer, H.	561
Dehmelt, H. G.	215, 219
Delker, L.	325
Deslattes, R. D.	303
DiMauro, L.	153
Dong, T.	153
Dress, W. B.	233
Dugan, G.	325
Dunn, A. F.	385
Dziuba, R. F.	539
Ebert, G.	519
Elnékavé, N.	465
Endo, T.	489,529
Erickson, G. W.	195
Fabre, C.	169
Faller, J. E.	405, 411, 581, 639, 689
Fang, Y.	419
Fau, A.	465
Felch, S. B.	359
Feldman, M.	153
Field, B. F.	539
Fontana, S.	69
Frieze, W.	149
Fujiwara, K.	681
Gabrielse, G.	219
Gähler, R.	257
Gallop, J. C.	365
Gao, J.	419
Gillies, G. T.	629
Girard, G.	453
Girvin, S. M.	557
Gossard, A. C.	549
Goy, P.	169
Gräff, G.	353
Gramlich, J. W.	357
Greene, G. L.	233, 515
Gross, M.	169
Gubin, M. A.	65
Guildner, L. A.	287
Guo, Y.	419
Gutmann, P.	535
Gygax, F. N.	223
Hall, J. L.	43, 671
Hammond, J. A.	423
Hänsch, T. W.	111
Hansen, P. G.	331
Hara, K.	479
Haroche, S.	169
Hartland, A.	543
Harvey, K. C.	163
Haugan, M. P.	181
Hein, G.	535, 553
Hellwig, H.	11
Hinds, E. A.	149
Hoeffgen, H.	49
Hoskins, J. K.	591
Howe, D. A.	25
Hu, C.	595
Huang, C.	419
Huang, D.	419
Hughes, V. W.	149, 237
Igarashi, T.	529
Iliff, R. L.	423
Inagaki, K.	529
Itano, W. M.	93
Ito, N.	53
Iwasaki, S.	53
Johnson, C. E.	229
Johnson, W. H.	335
Jonson, B.	331
Kalinowsky, H.	353
Kalus, J.	257
Kawaji, S.	529
Keyser, P. T.	639
Kibble, B. P.	461
Kinoshita, T.	203, 529
Klempt, E.	223
Knecht, J.	519
Kobayashi, Y.	441
Kochsiek, M.	427
Kolen, P.	675
Kötz, U.	159
Koyanagi, M.	489, 529
Kowalski, F. V.	181
Kowalski, J.	159

Page

Kponou, A. 149
Krüger, E. 369
Kubota, T. 479
Kugel, H. W. 145
Kurosawa, T. 53
Lahdenperä, K. 493
Lakkaraju, H. S. 103
Lavine, C. F. 539
Lepage, G. P. 185
Leventhal, M. 145
Levine, M. J. 201
Lewis, L. L. 25
Li, C. 77
Li, D. 419
Lichten, W. 117
Lindquist, W. B. 203
Littman, M. G. 173
Liu, F. 595
Liu, W. 595
Long, D. R. 587
Lovelock, M. J. 345
Lu, D. C. 325
Lundeen, S. R. 127, 131, 141
Luther, G. G. 573
Mampe, W. 233, 257
Man-Pichot, C. N. 73
Martin, J. E. 291
McGrory, J. B. 331
McLagan, K. H. 639
Mehl, J. B. 281
Mehta, C. L. 685
Meisel, G. 49
Melchert, F. 535
Metcalf, H. J. 153, 177
Metherell, A. J. F. 577, 581
Miller, P. D. 233
Moldover, M. R. 281
Morinaga, A. 53
Moriyama, J. 529
Müller, J. W. 375
Mungall, A. G. 21
Murakami, K. 529
Murnick, D. E. 145
Murphy, T. J. 357
Nachman, P. 671
Nakamura, A. 489, 529
Naylor, H. 345
Nelson, E. T. 145
Neumann, R. 159
Newman, D. 207
Newman, R. 591
Ni, W. 595, 647
Nikitin, V. V. 65
Nikulchin, A. V. 65
Nistler, W. 369
Niv, Y. 145
Noehte, S. 159
Obloh, H. 519
Oelfke, W. C. 607
Olsen, P. T. 475, 497
Ono, A. 299
Paik, H. J. 601
Panciera, R. 449
Pasin, W. 449
Patel, C. K. N. 145
Patterson, J. B. 445
Pendlebury, J. M. 233
Perrin, P. 233
Petley, B. W. 365
Petrovskiy, V. N. 65
Peuto, A. 449
Phillips, W. D. 173, 177, 475, 497
Pichanick, F. M. J. 149
Picotto, G. B. 69
Pipkin, F. M. 127, 131, 141
Ploog, K. 519

Page

Powell, L. J. 357
Protscenko, E. D. 65
Prowse, D. B. 437
Quinn, T. J. 291
Radcliffe, W. J. 365
Raimond, J. M. 169
Ramsey, N. F. 233
Ranganathan, D. 685
Rasetti, M. 449
Ravn, H. 331
Rayman, M. D. 671
Rebaglia, B. I. 69
Reilly, M. L. 287
Reinert, W. 49
Rendell, R. W. 557
Rich, A. 207
Rinker, R. L. 405, 411
Ritter, R. C. 629, 653
Robinson, H. G. 229
Roskies, R. Z. 201
Rowley, W. R. C. 57
Roxburgh, I. W. 1
Rüegg, W. 223
Sacconi, A. 449
Safinya, K. A. 127
Sakuma, A. 397
Sakurai, T. 53
Sands, R. W. 423
Sauder, W. C. 277
Schenck, A. 223
Schilling, H. 223
Schuessler, H. A. 103
Schult, O. W. B. 331
Schultz, J. 591
Schulze, R. 223
Schwinberg, P. B. 215, 349
Seppä, H. 493
Seyfried, P. 313
Shen, N. 77
Shepard, J. 643
Shiner, D. 123
Shiota, F. 479
Siegert, H. 317, 321
Sloggett, G. J. 469
Smith, K. 233
Smythe, R. M. 345
Sokolov, Y. L. 135
Speake, C. C. 577, 581
Spero, R. 591
Speth, J. 331
Stacey, F. D. 597
Steiner, M. 45
Suhr, H. 159
Sun, Y. 77
Sweetman, E. 207
Tanaka, K. 53
Tarbeyev, Y. V. 483
Torr, D. G. 675
Towler, W. R. 573
Traut, J. 353
Tsui, D. C. 549
Tuck, G. J. 597
Tyruikov, D. A. 65
Van Degrift, C. T. 457
Van Dyck, R. S., Jr. 215, 349
Van Flandern, T. C. 625
Van Patten, R. A. 659
Vessot, R. F. C. 611
Vitrant, G. 169
von Klitzing, K. 519, 535
Vowinkel, B. 49
Wagner, R. J. 539
Wallin, P. 493
Walls, F. L. 25, 93
Walther, H. 45
Wang, C. 77

Page

Wang, Z. 505
Warnecke, P. 535
Weirauch, W. 369
Werth, G. 99
Weyand, K. 509
White, R. E. 345
Wieman, C. 123
Williams, E. R. 475, 497
Wineland, D. J. 83, 93
Winkler, K. 159
Wöger, W. 383
Wolf, H. 223
Wood, O. R. II 145
Wu, C. S. 325
Wu, Y. 77
Xue, S. Q. 535
Yamanouchi, C. 529
Yang, C. 595
Yennie, D. R. 185
Yoshihiro, K. 529
Yu, H. 595
Zafiratos, C. D. 643
Zhang, G. 419
Zhang, X. 77
Zumberge, M. A. 405
zu Putlitz, G. 159

Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, Eds., Natl. Bur. Stand. (U.S.), Spec. Publ. 617 (1984).

Conference Registrants

Amin, S. R.
Arditi, M.
Aspden, H.
Baird, K. M.
Bankuti, L.
Barker, P. H.
Bartlett, D. F.
Becker, P.
Belecki, N. B.
Bell, G. A.
Bender, P. L.
Benjamin, S.
Bennett, S. J.
Bernard, B.
Bertinetto, F.
Bliek, L.
Bollinger, J. J.
Borchert, G. L.
Bouchareine, P.
Bourdet, G. L.
Braun, E.
Breeden, T.
Breitenberger, E.
Brillet, A.
Brodsky, S. J.
Brown, B. L.
Cabrera, B.
Cage, M. E.
Chandra, K.
Chen, Y. T.
Cheung, W. S.
Cohen, E. R.
Colclough, A. R.
Collé, R. P.
Conti, R. S.
Cooper, J. W.
Costain, C. C.
Daniel, H.-U.
Davis, R. S.
de Boer, H.
Dehmelt, H. G.
Deslattes, R. D.
DeVoe, R. G.
Dong, T.
Drever, R. W. P.
Dugan, G. F.
Dumas, J.
Dunn, A. F.
Dziuba, R. F.
Edwards, G. J.
Egan, P. O.
Eicke, W. G., Jr.
Eisenhart, C.
Erickson, G. W.
Evenson, K. M.
Faller, J. E.
Fau, A.
Faustov, R. N.
Feldman, M.
Field, B. F.
Frank, L. E.
Frueholz, R. P.
Fujiwara, K.
Furukawa, G. T.
Fystrom, D. O.
Gabrielse, G.
Gähler, R.
Garvey, R. M.
Garvin, D.
Giacomo, P.
Gillies, G. T.
Girard, G.
Gladisch, M.
Gonella, L.
Gossard, A. C.
Goy, P.
Grabe, M.
Gräff, G.
Greene, G. L.
Gubin, M. N.
Guien, M.
Guildner, L. A.
Guo, Y.
Hall, J. L.
Hammond, J. A.
Hänsch, T. W.
Hara, K.
Hartland, A.
Harvey, K. C.
Haugan, M. P.
Hazeltine, S. G.
Hellwig, H. W.
Henins, A.
Hill, R. M.
Hinds, E. A.
Holdeman, L. B.
Holt, H. K.
Holt, R. A.
Hoskins, J. K.
Hudson, R. P.
Hughes, V. W.
Itano, W. M.
Iuculano, G.
Jen, C. K.
Jimenez, J. J.
Johnson, C. E.
Johnson, W. H.
Jones, G. R., Jr.
Jones, O. C.
Jopson, R.
Kantor, F. W.
Kessler, E. G., Jr.
Kessler, K. G.
Keyser, P. T.
Kibble, B. P.
Kinoshita, T.
Kobayashi, M.
Kobayashi, Y.
Koch, W. F.
Kochsiek, M.
Kolen, P.
Koyanagi, M.
Kowalski, F. V.
Kowalski, J.
Kramer, B.
Krynetskiy, B. B.
Ku, H. H.
Kupper, W.
Layer, H. P.
Levine, M. J.
Lewis, L. L.
Lewis, R. R.
Leyh, C. H.
Li, M. C.

Lichten, W.
Lightbody, J. W., Jr.
Littman, M. G.
Lockyer, T. N.
Long, D. R.
Lough, T.
Lowke, J. J.
Lundeen, S. R.
Luther, G. G.
Mangum, B. W.
Marinenko, G.
Martin, J. E.
Martin, W. C.
Mathiesen, O.
Mattison, E.
Marx, E.
Mehta, C. L.
Meisel, G. O.
Metcalf, H. J.
Metherell, A. J. F.
Mills, A. P., Jr.
Mohr, P. J.
Moldover, M. R.
Morimura, M.
Muijlwijk, R.
Müller, J. W.
Nachman, P.
Nakamura, A.
Nelson, E. T.
Newman, D. E.
Newman, R.
Ni, W.-T.
Nistler, W.
Oelfke, W. C.
Olsen, P. T.
Ono, A.
Page, L. A.
Paik, H. J.
Pakay, P.
Petley, B. W.
Peuto, A. M.
Phillips, W. D.
Pipkin, F. M.
Powell, L. J.
Prestage, J. D.
Preston-Thomas, H.
Prowse, D. B.
Quinn, T. J.
Ramsey, N. F.
Rayman, M. D.
Reilly, M. L.
Rich, A.
Rinker, R. L.
Ritter, M. W.
Ritter, R. C.
Robinson, H. G.
Robiscoe, R. T.
Rosasco, G. J.
Roskies, R. Z.
Rowley, W. R. C.
Roxburgh, I. W.
Rueger, L. J.
Rumble, J. R., Jr.
Sacconi, A.
Sakuma, A.
Sauder, W. C.
Schuessler, H. A.
Schumer, D. B.
Schweitzer, W. G., Jr.
Schwitz, W.
Seppä, H.
Series, G. W.
Seyfreid, P.
Shen, N.
Shields, J. Q.
Shiner, D.
Shugart, H. A.
Siegert, H.
Silbar, M.
Sloggett, G. J.
Snyder, J. J.
Sokolov, Y. L.
Speake, C. C.
Stacey, F. D.
Steudel, A.
Stubbs, C. W.
Sweetman, E.
Tanaka, K.
Taylor, B. N.
Thomsen, J. S.
Torr, D. G.
Towler, W. R.
Twyford, L. C., Jr.
Urnov, A. M.
Van Baak, D. A.
Van Degrift, C. T.
Van Dierendonck, R. F.
Van Dyck, R. S., Jr.
Van Flandern, T. C.
Van Patten, R. A.
Verma, A. R.
Vessot, R. F. C.
Volk, C. H.
von Klitzing, K.
Wallin, P. T.
Walls, F. L.
Wang, Z.
Weber, A.
Weber, E. W.
Weirauch, W.
Werth, G.
Weyand, K.
White, J. A.
Wieman, C. E.
Williams, E. R.
Wineland, D. J.
Witt, T. J.
Wöger, W.
Woodle, K.
Yamanouchi, C.
Yennie, D. R.
Zumberge, M. A.

NBS-114A (REV. 2-8C)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR REPORT NO. NBS SP 617	2. Performing Organ. Report No.	3. Publication Date August 1984

4. TITLE AND SUBTITLE

Precision Measurement and Fundamental Constants II
Proceedings of the Second International Conference, held at the National Bureau of Standards, Gaithersburg, MD, June 8-12, 1981

5. AUTHOR(S)

B. N. Taylor and W. D. Phillips, Editors

6. PERFORMING ORGANIZATION (If joint or other than NBS, see instructions)

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

8. Type of Report & Period Covered
Final

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)

Same as in item 6 above.

10. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 84-601083

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)

This volume presents the Proceedings of the Second International Conference on Precision Measurement and Fundamental Constants, held at the National Bureau of Standards in Gaithersburg, MD, from June 8-12, 1981. Like its 1970 predecessor, the Conference provided an international forum for theoretical, experimental, and applied scientists actively engaged in research on precision measurements relating to the fundamental physical constants, and on the testing of related fundamental theory. More specifically, the purpose of the Conference was to assess the present state of the precision measurement-fundamental constants field, to examine basic limitations, and to explore the prospects for future significant advances. The principal subjects covered were: frequency, wavelength, spectroscopy, quantum electrodynamics, the gas constant, x-ray interferometry, nuclidic masses, uncertainty assignment, gravitational acceleration, mass, electrical quantities, gravity, and relativity. These proceedings contain the vast majority of both the invited review papers and the contributed current research papers presented at the Conference. The new results reported at the Conference were considered for inclusion in the 1983 least-squares adjustment of the constants carried out under the auspices of the Task Group on Fundamental Constants of the Committee on Data for Science and Technology (CODATA).

12. KEY WORDS (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)

data analysis; experimental tests of theory; fundamental constants; least-squares adjustments; precision measurements.

13. AVAILABILITY

☒ Unlimited
☐ For Official Distribution. Do Not Release to NTIS
☒ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
☐ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES
646

15. Price

USCOMM-DC 6043-P80

☆ U.S. GOVERNMENT PRINTING OFFICE : 1984 O - 450-414 : QL 3