NISTIR 90-4272

SELECTED PUBLICATIONS FOR THE EMAP WORKSHOP

H. H. Ku
Editor

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards and Technology
Center for Computing & Applied Mathematics
Statistical Engineering Division
Gaithersburg, MD 20899

U.S. DEPARTMENT OF COMMERCE
Robert A. Mosbacher, Secretary

Lee Mercer, Deputy Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
John W. Lyons, Director

NIST

NISTIR 90-4272

SELECTED PUBLICATIONS FOR THE EMAP WORKSHOP

H. H. Ku
Editor

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards and Technology
Center for Computing & Applied Mathematics
Statistical Engineering Division
Gaithersburg, MD 20899

March 1990

U.S. DEPARTMENT OF COMMERCE
Robert A. Mosbacher, Secretary
Lee Mercer, Deputy Under Secretary for Technology
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
John W. Lyons, Director

EDITOR'S NOTE

This volume is a collection of publications that has been found to be essential and useful in the conduct of the Electrical Measurement Assurance Program (EMAP) Workshop under the sponsorship of the National Institute of Standards and Technology (NIST) and the National Conference of Standards Laboratories (NCSL). Since 1981, the Electricity and the Statistical Engineering Divisions of NIST have collaborated to give seven EMAP Workshops at selected sites in the United States. These 5-day workshops, attended by about 30-40 participants, aim to disseminate NIST expertise in electrical measurements and calibration to professionals and senior technical personnel in industries, standards laboratories, and state and federal government agencies. Calibration designs and quality control techniques are emphasized for use in their own laboratories.

Publications included in this volume appeared originally as articles scattered in journals, NIST reports, technical notes, monographs and special publications. Some of these papers are now out of print. The present volume, with some publications edited specifically for the EMAP Workshop, will not only serve as a useful text and reference in class, but also for use by all professionals interested in electrical measurements as well.

The pagination for this publication has the following format: a unique alphabetic designation for each reprint is followed by the page number of the original article, e.g. A-14.

H. H. Ku

CONTENTS

BIBLIOGRAPHY

Precision Measurement and Calibration: Electricity
A.O. McCoubrey - Editor
NBS Special Publication 705, 1985, 853 pages

This Special Publication is a compilation of papers published mostly during the twelve year period, 1973-1985, by the Electricity Division Staff in the NBS Center for Basic Standards. This collection is focused upon the realization of SI units of electricity in the United States, the practical maintenance of such units, and the NBS calibration of industrial electrical measurement standards.

Precision Measurement and Calibration: Statistical Concepts and Procedures
H.H. Ku - Editor
NBS Special Publication 300, Vol. I, 1969, 436 pages

This volume deals with methodology in the generation, analysis, and interpretation of precision measurement data. The volume contains reprints of 40 papers on statistical concepts and procedures classified in six sections. Four works too long to be included here are represented by titles and abstracts. The interpretive foreword appearing at the beginning of each of the first six sections comments on the individual papers and thus characterizes the particular section. The index has been prepared to facilitate browsing.

Experimental Statistics
M.G. Natrella
NBS Handbook 91, 1966

This Handbook has been prepared as an aid to scientists and engineers engaged in research and development programs. It is intended for the user with an engineering background who has occasional need for statistical techniques, but does not have the time or inclination to become an expert on statistical theory and methods. This handbook covers the basic statistical concepts, the analysis of measurement data, the planning and analysis of experiments, and special topics which are important but non-standard statistical techniques.

ASTM Manual on Presentation of Data and Control Chart Analysis, STP 15D, American Society for Testing and Materials, 1976, 162 pages

A classical treatise on control charts.

Experimentation and Measurements
N.J. Youden
NBS Special Publication 672, 1984, 127 pages

An excellent introduction to the realistic consideration of errors of measurements. This book contains many ingenuous examples based on common everyday measurements, and illustrates how statistics can contribute to the design, analysis, and interpretation of experiments involving measurement data.

MEASUREMENT ASSURANCE

NBSIR 77-1240

MEASUREMENT ASSURANCE

J. M. Cameron

Institute for Basic Standards
National Bureau of Standards
Washington, D.C. 20234

April 1977

Final

Prepared for
Office of Measurement Services
Institute for Basic Standards
National Bureau of Standards
Washington, D.C. 20234

U.S. DEPARTMENT OF COMMERCE, Juanita M. Kreps, *Secretary*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Acting Director*

Measurement Assurance

Introduction

A single measurement can be the basis for actions taken to maintain our health, safety or the quality of our environment. It is important therefore that the errors of measurement be small enough so that the actions taken are only negligibly affected by these errors. We realize this necessity on a personal basis when we consider medical measurements, or our exposure to radioactivity. In any government regulatory action or measurement involved in legal actions it is also obvious that the shadow of doubt surrounding the measurements should be suitably small. But this is no less true for all other measurements in science and industry and even though legal action may not be involved, the validity of scientific inference, the effectiveness of process control, or the quality of production may depend on adequate measurements [2].

Allowable Limits of Measurement Error

How does one achieve this condition--that the measurements are "good enough" for their intended use? It would seem obvious that one has to start with the need--i.e., deciding upon what is "good enough". There are a number of cases where physiological restraints provide the definition such as in the allowable error in exposure to cobalt radiation in cancer treatment or in the amount of pollutant entering a lake. In nuclear materials control the allowable error is a function of the amount of material which would pose a hazard if diverted. In industrial production or commercial transactions, the error limit is determined by a balance between the cost of better measurement and the possible economic loss from poorer measurement.

By whatever path such requirements are arrived at, let us begin with the assumption that the allowable error should not be outside the interval (-a, +b) relative to the quantity being measured. Our problem is one of deciding whether the uncertainty of a single measurement is wholly contained in an interval of that size. We therefore need a means of assigning an uncertainty to a single isolated measurement and, in fact, we need a perspective (i.e., physical and mathematical model) in which to view measurement so as to give operational meaning to the term "uncertainty."

Reference Base to Which Measurements Must Be Related

It is instructive to contemplate the possible "cross-examination" of a measurement if it were to become an important element in a legal controversy. Two essential features emerge. First, that the contending parties would have to agree on what (actually realizable) measurement would be mutually acceptable. The logic of this seems unassailable--if one cannot state what measurement system would be

accepted as "correct," then one would have no defensible way of developing specifications or regulations involving such measurements. Second, the scientific cross-examination by which one establishes the "shadow of doubt" relative to this acceptable value gives one the uncertainty to be attached to the measurement.

The consensus or generally accepted value can be given a particularly simple meaning in dealing with measurements of such quantities as mass, volt, resistance, temperature, etc. One may require that uncertainties be expressed relative to the standards as maintained by a local laboratory or, when appropriate, to the national standards as maintained by NBS. In other cases, nationally accepted artifacts, standard reference materials or in some cases a particular measurement process may constitute a reference base. One basic quality should not be overlooked--all are operationally realizable. The confusion engendered by introducing the term "true value" as the correct but unknowable value is thus avoided.

Properties of Measurement Processes

In discussing uncertainty, we must account for two characteristics of measurement processes. First, repeated measurements of the same quantity by the same measurement process will disagree and, second, the limiting means of measurements by two different processes will disagree. These observations lead to a perspective from which to view measurement namely that the measurement be regarded as the "output" of a process analogous to an industrial production process. In defining the process, one must state the conditions under which a "repetition" of the measurement would be made, analogous to defining the conditions of manufacture in an industrial process.

The need for this specification of the process becomes clear if one envisions the "cross-examination" process. One would begin with such questions as

> Within what limits would an additional measurement by the same instrument agree when measuring some stable quantity?
>
> Would the agreement be poorer if the time interval between repetitions were increased?
>
> What if different instruments from the same manufacturer were used?
>
> If two or more types (or manufacturers) were used, how much disagreement would be expected?

To these can be added questions related to the conduct of the measurement.

What effect does geometry (orientation, etc.) have on the measurement?

What about environmental conditions--temperature, moisture, etc.?

Is the result dependent on the procedure used?

Do different operators show persistent differences in values?

Are there instrumental biases or differences due to reference standards or calibrations?

The questions serve to define the measurement process--the process whose "output" we seek to characterize.

The current understanding of a scientific or industrial process or of a measurement process is embodied in a physical model which explains the interactions of various factors, corrections for environmental or other effects, and the probability models necessary to account for the fact that repetitions of the same event give rise to nonidentical answers. For example, in noise level measurement one is involved with assumptions regarding frequency response, weighing networks, influence of procedures and geometry, and an accepted theory for making corrections for temperature and other environmental factors. In mass the properties of the comparator (balance) the environmental effects, and the procedure used all enter into the description of the method.

One thus begins with the specification of a measurement method--the detailed description of apparatus, procedures and conditions by which one will measure some quantity. Once the apparatus is assembled and checked out, one has a measurement process whose output can be studied to see if it conforms to the requirement for which it was created.

In industrial production one tries to produce identical items but usually a measurement process is set up to measure a variety of quantities and Ordinarily one does not measure the same quantity over and over. One thus has the problem of sampling the output of the measuring process so as to be able to make statements about the health of the process relative to the needs. The needed redundancy can sometimes be achieved by remeasuring some of the items, or by measuring a reference artifact periodically. It is essential that the repetitions be done under the same diversity of conditions as the regular measurements, and that the items being measured be typical of the regular workload.

As an example, a sequence of measurements was made using two sound level meters to measure a sound of nominally 90 dB re 20 μPa. The sound was generated by a loudspeaker fed broadband noise. On 16

different days measurements were made outdoors and over grass with the loudspeaker in the same orientation and location relative to a building 2 m behind the loudspeaker. The sound level meter was always the same distance (10 m) from the loudspeaker and on a line perpendicular to the face of the loudspeaker. Other than the grass, the person holding the sound level meter, and the building to the rear of the loudspeaker, there were no other reflecting surfaces or obstacles within 50 m. No measurements were made in the rain or in winds exceeding a few km/hr. The results from these 16 repetitions are shown in Figure 1. Typically, had duplicate measurements been made on the same day they would have given results as shown in Figure 2.

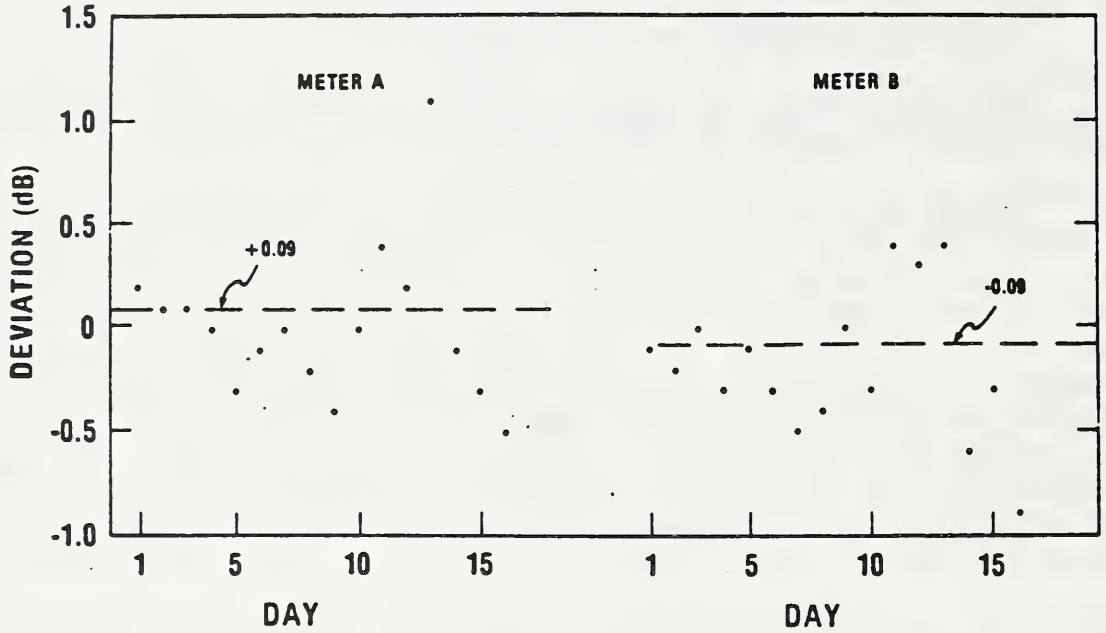


FIGURE 1: DAY-TO-DAY VARIATION IN METER READINGS.

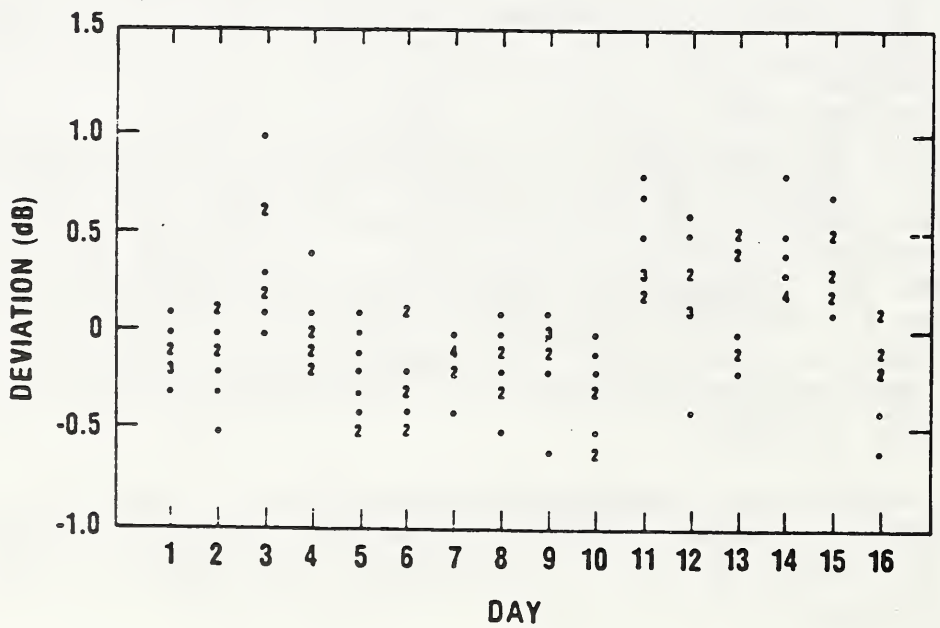


FIGURE 2: DAY-TO-DAY VARIATION IN METER READINGS WITH MULTIPLE VALUES PER DAY. (COINCIDENT POINTS INDICATED BY NUMBERS.)

One now faces the question of how to describe the variation that exists. Obviously there will be a different level of agreement expected between pairs on the same day, but this variation in no way predicts that encountered from day-to-day. The issue is not so much the statistical procedures to be used--these will follow after one defines the set of repetitions over which his conclusions must apply. For measuring the short term change in noise level, the difference between duplicates would apply; for any regulatory action, the day-to-day variation would have to be considered.

The crucial step in assessing the effects of random error is that of defining the set of repetitions over which the measurement is to apply. In the context of legal proceedings, one arrives at the degree of credibility of evidence by questions designed to find out how far the statement could be in error. In measurement, the uncertainty is arrived at by determining the amount of disagreement expected in the set of repetitions that would be appropriate in the context of the intended use of the measurement.

The Concept of a Repetition of a Measurement

Every measurement has a set of conditions in which it is presumed to be valid. At a very minimum, it is the set of repeated measurements with the same instrument-operator procedure-configuration. (This is the type of repetition one would envision in some process control oerations.) If the measurement is to be interchangeable with one made at another location, the repetition would involve different instrument-operator-procedure-environment configurations. (This type of repetition is involved in producing items to satisfy a specification and of manufacturing generally.) When the measurement is to be used for conformance to a health, safety, or environmental regulation even different methods may be involved in a "repetition."

To evaluate a measurement process some redundancy needs to be built into the system to determine the process parameters. This redundancy should be representative of the set of repetitions with which the uncertainty statement is to apply. In NBS' measurements of mass, a check standard is measured in parallel with the unknowns submitted for calibration. One thus generates a sequence of measurements of the same object covering an extended time period. From these results one can answer questions relating to the agreement expected in a recalibration and the operating characteristics of the measurement process. In this simple case the check standard is treated exactly the same way as the unknowns so that the properties of the process related to it are transferrable to the unknown.

The essential characteristic in establishing the validity of measurement is predictability that the variability remains at the same level and that the process has not drifted of shifted abruptly from its established values. One must build in redundancy in the form of a

control--the measurement of a reference quantity of known value--or by remeasuring some values by a reference method (or by an instrument with considerably smaller uncertainty). In cases where the phenomenon can be repeated, one can learn about random errors by remeasuring at a later time sufficiently far removed to guarantee independence.

In measuring an "unknown" one gets a single value, but one still is faced with the need to make a statement that allows for the scatter of the results. If we had a sufficiently long record of measurements, we could set limits within which we were fairly certain that the next measurement would lie. Such a statement should be based on a collection of independent determinations, each one similar in character to the new observation, that is to say, so that each observation of the collection and also the new observation can be considered as random drawings from the same probability distribution. These conditions will be satisfied if the collection of points is from a sufficiently broad set of environmental and operating conditions to allow all the random effects to which the process is subject to have a chance to exert their influence on the variability. Suitable collections of data can be obtained by incorporating an appropriate reference measurement into routine measurement procedures, provided they are representative of the same variability to which the "unknown" is subject. The statistical procedures for expressing the results will depend on the structure of the data but they cannot overcome deficiencies in the representativeness of the values being used.

The results from the reference item provide the basis for determining the parameters of the measurement process and the properties are transferable. One is saying, in effect, if we could have measured the "unknown" again and again, a sequence of values such as those for the reference item would have been obtained. Whether our single value is above or below the mean we cannot say, but we are fairly certain it would not differ by more than the bounds to the scatter of the values on the reference item.

The bound $\pm R$, to be used for the possible effect of random errors may be as simple as ± 3 (standard deviation) or may involve the combination of many components of variance. Once the set of repetitions over which one's conclusions must apply is defined, the structure of the random error bound can be determined.

Possible Offset of the Process

Once one has established that his measurement process is "in control" from the point of view of random variation, there remains the question of the possible offset of the process relative to other processes. It is not helpful to speak of the offset from a "true value" which exists only in the mathematical or physical model of the process. The usefulness of considering measurement in the context of legal proceedings helps clear away some of the classical confusion about

errors of measurement. In a legal or regulatory setting, one is forced to state what would be accepted as correct such as comparison (by a prescribed process) with national standards or with the results from a designated laboratory or consensus of many laboratories.

The idea of defining uncertainty as the extent to which a measurement is in doubt relative to a standard or process defined as correct finds expression in the recent Nuclear Regulatory Commission statement [12]:

> 70.57(a) "Traceability" means the ability to relate *individual measurement results* to national standards or nationally accepted measurement systems ... (italics added)

One could measure the offset of his process relative to the accepted process, and make suitable corrections to eliminate the offset. However, for most processes, one is content with setting bounds to the possible offset due to factors such as:

> Errors in the starting standards
>
> Departures from sought-after instrumentation (e.g., geometrical discrepancies)
>
> Errors in procedures, environment, etc.

and other effects which are persistent. From properly designed experiments one can arrive at a limit to the possible extent of errors from these sources in answer to the question, "If the process were set up ab initio, how large a difference in their limiting means would be reasonable?"

A bound to a number of factors can be determined as part of regular measurement. For example, the effect of elevation on sound level measurements could be evaluated by occasionally duplicating a measurement at a different height and taking an appropriate fraction of the observed difference as the limit to the possible offset due to any error in setting elevation. Figure 3 shows some results from sound level meters at two heights with the source at a constant height.

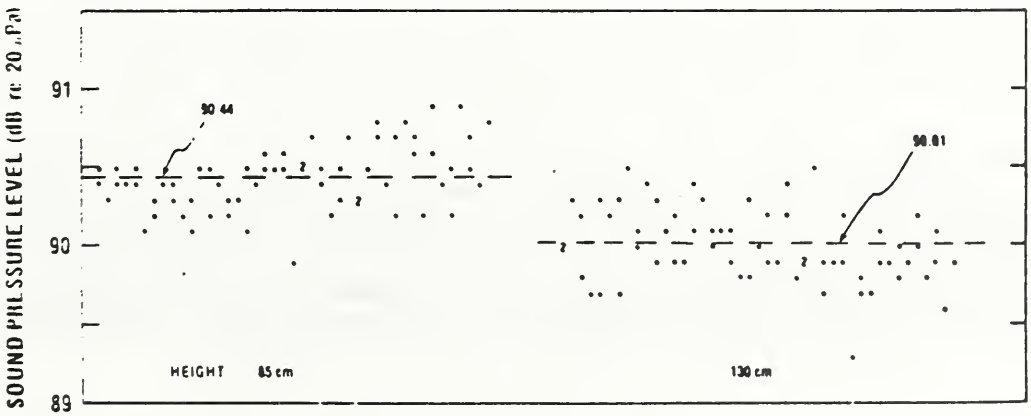


FIGURE 3: DIFFERENCE BETWEEN METER VALUES WITH CHANGE IN HEIGHT

Even if one has a functional relation, $y = f(h)$, expressing the dependence of the result, y, on height, h, one still has to carry out these measurements. The usual propagation of error approach involving partial derivatives, etc., implies that all instruments are equally dependent on the parameter under study, that there are no effects related to the factor except that contained in the formula. This can be verified for a particular instrument by actually measuring its response.

A similar comparison was made for a different orientation of the instrument with respect to this signal source and is shown in Figure 4. The effect of orientation is negligible and one would not be justified in adding an allowance for possible systematic error from this source based on a theoretical calculation.

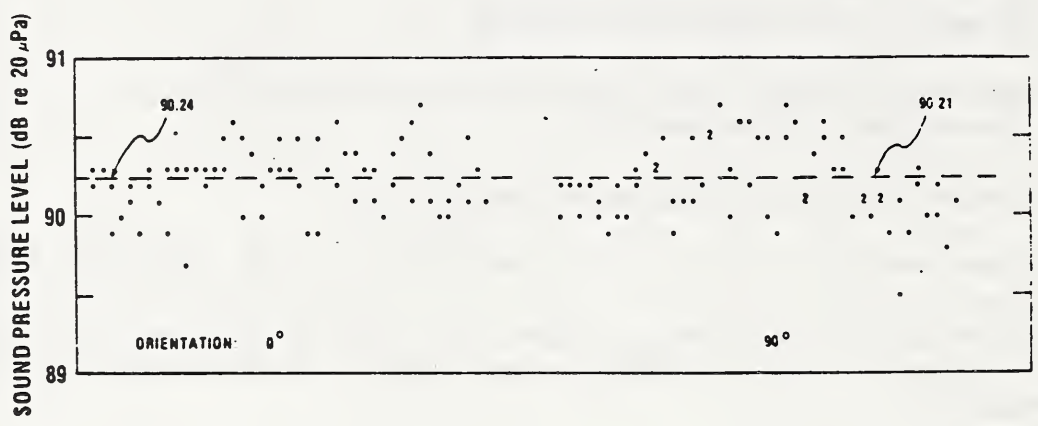


FIGURE 4: DIFFERENCE BETWEEN METER VALUES WITH A CHANGE IN ORIENTATION

From these measurements, one will have a set of bounds E_1, E_2, E_3, ... to the possible offset or systematic error from the various factors. The question as to how to combine these to a single bound to the possible offset depends on knowledge of the joint effects of two or more factors and on the physical model assumed for the process. For example, if the bounds E_i and E_j arise from independent random error

bounds, then it would be appropriate to combine them in quadrature, i.e., $\sqrt{E_i^2 + E_j^2}$. An error in the model e.g., assumed linearity even when nonlinearity exists) would act as an additive error. The properties of any combination rule can be evaluated and a selection made of the most appropriate. The result will be an overall value, E, for the possible offset for the limiting mean of the process from that of the nationally accepted process.

Uncertainty

What can one say about the uncertainty of a measurement made by a process that may be offset from the nationally accepted process by some amount $\pm E$, and is subject to random errors bounded by $\pm R$? How should these values be combined? To begin with, one could raise the question, "If the random error could be made negligible, what uncertainty would one attach to a value from the process?" Clearly the answer is $\pm E$. The next question, "If, in addition, a random error of size R is possible, what do we now say about the uncertainty?" The answer seems obvious--E and R are added to give an uncertainty of $\pm[E + R]$.

But what if E were itself the result of only random errors? The answer depends on what one calls a repetition. By the way E is defined, it is the bound for the systematic offset of the process and although it may be arrived at from consideration of random errors, the factor involved keeps the same (unknown) value throughout. Our ignorance does not make it a random variable.

Consider the case of a mass standard. NBS' certificate states that the uncertainty is based entirely on random variation, the effects from systematic errors being negligible. But unless one recalibrates, the error due to calibration remains fixed in all measurements by the user.

The uncertainty of a measurement--the width of its "shadow of doubt" in a legal proceeding--must therefore be the sum of the random error and systematic error limits.

Measurement Process Control

The essential feature for the validity of the uncertainty statement is that the process remain in a state of statistical control. Once an out-of-control condition occurs, one has lost predictability and the previous uncertainty statements are no longer valid.

To monitor the process some redundancy has to be built into the system. A variety of techniques can be used to give assurance of continued control. For example, one could periodically measure the same reference item or artifact or one could make duplicate measurements on some production items with enough delay to guarantee

independence. The American National Standards Institute Standard N15.18 for mass measurement [10] is an example where this approach is worked out in detail. But one has to verify more than just those parameters related to random variations. One needs to build in tests of the adequacy of the physical model by a variety of tests on the process (e.g., by repeating measurements under different conditions to verify the adequacy of the corrections for such changes) as well as periodic redetermination of the bounds for systematic error. One thus tests that the assumed model is still acceptable and that the parameters assigned to that model have not changed.

An excellent example of the efficacy of this approach is given by the recent announcement [6] of discrepancies of 1 mg in the assignment of mass to aluminum kilogram standards. The mass measurement system has long been shown to be nearly perfect for the usual standards. To check up on the performance of the system at densities nearer to that of most objects involved in practical measurement, an aluminum kilogram was sent to laboratories including several at high elevations. It turns out that the difference between the mass of a stainless steel and an aluminum kilogram is significantly different at different elevations. This unsuspected property of the real measurement system is now the subject of considerable study.

All measurements have some form of measurement assurance program associated with them although, as with quality control, we usually reserve the term for a formal program. In a formal program one treats the whole process--beginning with a study of the need, the development of a measuring process and a procedure for determining and monitoring its performance, and an evaluation of the effectiveness of the whole effort. One needs a criterion of success to be able to determine whether more of one's current measurement activity or perhaps some alternative would contribute most to the overall program, and this is not necessarily provided by the smallness of the uncertainty for a measurement.

For example, when the requirement is for matched sets (e.g., ball bearings) or mated assembly parts, then it is usually cheaper and more accurate to sort into finely divided classes and match for correctness of fit rather than perform direct measurement of each part.

When the measurement requirements are stated in terms of the needs of the system, (number of correctly matching parts, number of correctly measured dosimeters, etc.) one can measure success of the measurement effort in terms of closeness to meeting those goals. Measurement efficiency is thus judged in terms of the output of the organization rather than by the count of the number of significant digits. Also, one needs this measure of performance of the measurement effort to be able to identify those areas which need improvement.

Examples of Measurement Assurance Programs in NBS Measurements

Two easily described measurement assurance programs are those in mass and length. In routine calibration, a check standard is included with each set of weighings and process control is maintained by monitoring the value obtained for the check standard and of the random error from the least squares analysis [8, 9]. Control charts have been maintained since 1963. In the calibration of gage blocks, similar process control has been maintained since 1972 on both the interferometric process by which the assignment of length to the NBS master gage blocks is done and on the comparator process by which length values are transferred to customer gage blocks. [1, 7]

Similar programs are in effect in all divisions, but not all quantities involved in calibration have a formal program worthy of the name, measurement assurance.

Examples of Measurement Assurance Programs At Other Laboratories

Only two examples of measurement assurance programs at other laboratories have ever been reported. One at Autonetics [3] in length and one at Mounds Laboratory in mass. Once the mass measurement system for UF_6 is underway as part of the Safeguards program, NBS will be able to document the efficacy of the approach in practical measurement.

The NBS Measurement Assurance Programs Offered As A Part Of Our Calibration Service

Measurement Assurance Programs are listed as a calibration service in mass, volt, resistence, capacitance, voltage ratio, watthour meters, platinum resistance thermometry, and laser power. These are designed to measure the offset of measurement processes for the calibration of standards by other standards laboratories. These are applicable only to those laboratories who maintain and calibrate standards in the same manner as NBS. [See 11, 5, 13.]

These procedures enable a laboratory to determine the offset between its process of calibrating standards and that of NBS.

Need For Measurement Assurance Program For Practical Measurement

The UF_6 cylinder program for Safeguards [10] is an example of NBS' service in providing a direct method for measuring the offset of practical measurement processes from that accepted as correct, namely mass measurement by NBS. Investigation of the need and possible mechanisms or artifacts for monitoring the offset of practical measurements in quantities such as voltage, resistance, length, radioactivity is underway. (For examples of the application of these principles to sound level meters, see [5].)

In personnel dosimetry procedures are being worked out [14] to monitor the output of firms providing such services. In this case, a table of allowable limits of uncertainty are based on physiological considerations. Process parameters are to be determined by an initial study. Routine monitoring will be used to confirm that the process is "in control" at those levels, otherwise the parameters are redetermined *ab initio*. These "consistency" or "in control" criteria replace the usual one-time round robin approach. The amount of effort needed to establish this predictability is a function of the risk and costs of wrong decisions.

In industrial measurement we could ask

> If some critical measurements on the production line were repeated would the two measurements agree?
>
> How much bad material is passed, or good material rejected because of errors in measurement?

To those who have not properly answered these questions, dollar savings and improved product quality are possible without redesign or changes in production procedures.

Is our faith in instruments justified? Implicit faith in the correctness of instruments means that product variability (as determined by these instruments) is attributed to variability in components, raw materials or even poor design. One wonders how many times this has led to expensive changes in production procedures without apparent improvement because the variability actually arose in the measurements themselves.

How often has the installation and methods of use degraded the output of an instrument capable of much more accuracy than is required when handled properly? Without some surveillance of the actual measurements, one would never know.

One wonders how often a product is redesigned because measurement error has led to the decision that the product does not conform to specifications.

The result of this look at measurement is measurement assurance--the quality control of measurement. If adequate control exists, then one can look elsewhere for improvements in the product line. If it does not, then one has the possibility of savings without changing production procedures.

Some form of redundancy must be built into the process to answer these questions.

References

[1] John S. Beers, "A Gage Block Measurement Process Using Single Wavelength Interferometry," NBS Monograph 152, December 1975

[2] J. M. Cameron, "Measurement Assurance," Review of Standards and Specifications Section of Journal of Quality Technology, Vol. 8, No. 1, pp. 53-55, January 1976

[3] J. A. Hall, "Dimensional Measurement Assurance Programs," Autonetics Report No. X73-1146/031, presentation at 1973 Conference of the National Conference of Standards Laboratories at the National Bureau of Standards, Gaithersburg, Md., November 1973

[4] Edward B. Magrab, "Environmental Effects on Microphones and Type II Sound Level Meters," NBS Technical Note 931, October 1976

[5] N. Michael Oldham, "A Measurement Assurance Program for Electric Energy," NBS Technical Note 930, September 1976

[6] P. E. Pontius, "Mass Measurement: A Study of Anomalies," Science 190, pp. 379-380, October 1975

[7] P. E. Pontius, "Measurement Assurance Program--A Case Study: Length Measurements. Part 1. Long Gage Blocks (5 in. to 20 in.)," NBS Monograph 149, November 1975

[8] P. E. Pontius, "Measurement Philosophy of the Pilot Program for Mass Calibration," NBS Technical Note 288, May 1966

[9] P. E. Pontius, "Notes on the Fundamentals of Measurement and Measurement as a Production Process," NBSIR 74-545, September 1974

[10] "Mass Calibration Techniques for Nuclear Material Control," N15.18-1975, American National Standards Institute, New York, New York

[11] "Measurement Analysis Program in Mass," NBS Technical News Bulletin, September 1970

[12] "Measurement Control Program for Special Nuclear Materials Control and Accounting (10 CFR 70.57)," Federal Register, Vol. 40, No. 155, pp. 33651-33653, August 1975

[13] "New Calibration Service for the Volt," NBS Technical News Bulletin, pp. 46-47, February 1971

[14] "Proposed Standard--Criteria for Testing Personnel-Dosimetry Performance," Health-Physics Society, HPSSC WG/15, October 1976

Measurement Assurance Programs
Part II: Development and Implementation

Carroll Croarkin

Statistical Engineering Division
Center for Applied Mathematics
National Engineering Laboratory
National Bureau of Standards
Gaithersburg, MD 20899

Supersedes SP 676-II April 1984

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

Issued April 1985

Preface

The purpose of this document is to guide the reader through the logical development of a measurement assurance program as it is intended to

i) Tie a measurement process or reference standards to the defined unit of measurement for the quantity in question or to national standards; and

ii) Establish the uncertainty of values reported by the process through the maintenance of statistical control of the measurement process.

The discussion is approached in the context of the assumption that the tie to defined units or national standards is accomplished via a tie to NBS. Participation in a measurement assurance program can satisfy this tie where systematic error is evaluated via measurements made in the participating laboratory on an NBS transfer standard and where it can be shown that the measurement process is continuously in a state of statistical control. This in no way implies that measurement assurance cannot be attained without formal participation in an NBS sponsored program, but the presentation is made more concrete in this context.

The formulation of measurement assurance techniques for all measurement situations is not within the scope of this document. Obviously, such matter is best handled on a subject basis. The dearth of suitable documentation for specific measurement disciplines serves as a motivating factor in the development of this guide which describes statistical procedures and analyses that are generally pertinent to measurement assurance. It is hoped that the reader will be able to adapt the philosophy and techniques contained herein to his own particular measurement needs.

The material in this document is largely statistical in nature because of the measurement assurance approach to quantifying both the random and systematic errors that are generated by a measurement process. It should be recognized, however, that measurement assurance is not achieved by statistical techniques alone but by the totality of procedures such as correct measurement practice, adherence to recommended procedures, control of environmental factors and estimation of process parameters that relate the output of the measurement system to national standards.

This document is the second part of a general treatise, Measurement Assurance Programs, which is divided into Part I: General Introduction and Part II: Development and Implementation. Part I by Dr. Brian C. Belanger is intended as a statement of the goals of measurement assurance from a managerial perspective and advances the basic philosophy of quality in measurement. Part II, which was supported by the NBS Office of Measurement Services, extends the principles so stated to specific measurement situations, drawing extensively on programs that were developed by Mr. Joseph Cameron in consultation with NBS technical divisions. In addition to these examples, measurement control programs with verifiable uncertainty statements are outlined for measurement situations that the author has encountered in consultations with the measurement community outside NBS.

Table I and Table II in this manuscript were compiled using the NBS software package DATAPLOT developed by Dr. James Filliben.

TABLE OF CONTENTS

LIST OF FIGURES

Measurement Assurance Programs
Part II: Development and Implementation

Carroll Croarkin
Statistical Engineering Division
Center for Applied Mathematics
National Bureau of Standards
Gaithersburg, MD 20899

This document is a guide to the logical development of a measurement assurance program in which the tie between a measurement and its reference base is satisfied by measurements on a transfer standard. The uncertainty of values reported by the measurement process is defined; and the validation of this uncertainty for single measurements is developed. Measurement sequences for executing the transfer with NBS and procedures for maintaining statistical control are outlined for eight specific measurement situations with emphasis on characterizing parameters of the measurement process through use of a check standard.

Key Words: Calibration; check standard; measurement assurance; random error; statistical control; statistical methods; systematic error; uncertainty.

1. The Development of a Measurement Assurance Program

1.1 Historical Perspective

The development of measurement assurance at the National Bureau of Standards, over the more than eighty years that the nation's premier measurement laboratory has been in existence, has evolved hand in hand with the NBS central mission of providing quality measurement. We might date this evolution as starting with the early experiments on the velocity of light [1][1]. Since then the principles of measurement assurance have reached realizations of all SI units and numerous derived units of measurement, and even now are influencing innovations in measurement science related to electronics and engineering.

As the reader familiarizes himself with the concepts of measurement assurance, he will come to realize that quality in calibration is dependent upon the inclusion of a check standard in the calibration scheme. The first application of this principle at NBS came in the area of mechanical measurements where a prescribed series of observations known as a weighing design, so called because of the obvious connection to mass weighings, defines the relationship among reference standards, test items and check standards. The first weighing designs published by Hayford in 1893 [2] and Benoit in 1907 [3] had no provision for a check standard, and the creation of suitable designs had to await general progress in the area of experimental design which characterized statistical activity at NBS in the nineteen fifties.

[1] The numbers in brackets refer to references cited at the end of this document.

As early as 1926 an NBS publication by Pienkowsky [4] referred to a standard one gram weight whose mass as "determined in the calibrations just as though it were an unknown weight" was used as a gross check on the calibrations of the other unknown weights. It remained until the nineteen sixties for the concept of measurement as a process to be described by repetitions on a check standard such as the one gram weight described by Pienkowsky. At that time calibrations of mass and length standards were formalized into measurement assurance programs with demonstrable uncertainty of reported values and statistical control of individual calibrations. A compendium of weighing designs for mechanical and electrical quantities with allowance for a check standard in each calibration sequence was published in 1979 (Cameron et al [5]).

Although many experimenters, past and present, have contributed to the quality of measurement science at NBS, the formulation of measurement assurance is the special province of the Statistical Engineering Division. Three members of this group, C. Eisenhart, W. J. Youden and J. M. Cameron, were largely responsible for fruition of the check standard concept, and the advent of electronic computers aided in the rapid application of this concept to NBS calibration programs. In 1962 a paper by Eisenhart [6] laid the groundwork for defining a repetition for a measurement process and assessing the uncertainties associated with such a process. This paper still serves as the primary treatise on the subject. Concurrently, Youden was implementing "ruggedness" testing in physical measurements [7], and at the same time he was introducing experimental design into interlaboratory testing [8].

In 1967 the first documentation of a measurement assurance approach appeared in print as an NBS monograph. The tutorial by Pontius and Cameron [9], treated the entire spectrum of mass measurement as a production process and began the dissemination of measurement assurance outside the NBS community. In the years since then, measurement assurance, both within and outside NBS, has been applied to basic SI units such as length as formulated in reference [10] and complex measurement areas such as dimensional measurements for the integrated circuit industry as formulated in reference [11]. Recently the measurement assurance approach has found its way into an ANSI standard for nuclear material control [12] with the use of "artifact reference mass standards as references for uranium hexafluoride" cylinders reported by Pontius and Doher [13].

1.2 Introduction

The development of a measurement assurance program evolves logically from the specific interpretation that we will give to the term "measurement assurance". The reader is asked to lay aside interpretations given to this term from previous experiences and to concern himself with what it means to have demonstrable scientific assurance about the quality of a measurement. For calibration activities, quality of a measurement is defined by its uncertainty, and the validity of an uncertainty statement for an individual measurement is guaranteed via the measurement assurance program as it is intended to

i) Tie a single measurement to a reference base; and

ii) Establish the uncertainty of the measured value relative to this reference base.

Firstly, in the case of basic SI units, a single measurement of a characteristic embodied in an object or artifact must be related to the defined unit for that quantity; for example, until recently the length of a gage block was defined relative to the wavelength of radiation of krypton 86 as realized through interferometry [14]. Because derived units of measurement can only be indirectly related to basic units, the measurement assurance concept is extended to such quantities by requiring that they be related to a reference base such as artifact standards or a measurement system maintained by the National Bureau of Standards. Secondly, a measurement assurance program must provide a means of maintaining statistical control over the measurement system thereby guaranteeing the validity of the uncertainty for a single measured value relative to its reference base (Cameron [15]).

The definition of measurement assurance is completed by an examination of the properties of measurement. A single measurement is properly related to national standards only if there is agreement between it and a value that would be achieved for the same quantity at NBS--meaning a value that would be arrived at from a sufficiently long history of like measurements at NBS. In actuality it is not possible to estimate the disagreement between a single measurement in a given laboratory and the long-term NBS value. However, if the measurement system of the laboratory is stable or as we say operating in a state of statistical control, the single measurement can be regarded as a random draw from another long history of measurements which also tend to a long-term value. The purpose of calibration is to eliminate or reduce the disagreement, referred to as offset, between a laboratory's long-term value for a measurement and the corresponding NBS long-term value by corrections to the measurement system and/or reference standards.

Where offset cannot be eliminated or reduced by calibration, it is a systematic error accruing to the laboratory's measurement system. Even where there is an accounting for such disagreement, the fact that NBS has imperfect knowledge about the long-term value from its own measurement system, based as it is on a finite though large number of measurements, means that the limits of this knowledge contribute another systematic error to the measurement system of the laboratory. In some special cases systematic and random errors that arise as NBS attempts to tie its measurement system to defined units of measurement may also become part of the systematic error for the laboratory.

The uncertainty that surrounds any single measurement describes the extent to which that single number could disagree with its reference base. The uncertainty includes all systematic errors affecting the measurement system; it also includes limits to random error that define the degree to which the individual laboratory, just as NBS, may be in error in estimating the long-term value for the measurement. Where the calculation of a long-term value for a measurement and limits to random error cannot be done directly, which is the usual case for calibration measurements, the long-term value is referenced to a long-term value of measurements made on an artifact(s) called a check standard.

Measurement assurance is attained when the determination of all sources of systematic error is coupled with statistical control of the measurement process as achieved by adapting quality control techniques to measurements on the check standard. Statistical control consists of comparing current check standard measurements with the value expected for such measurements and making decisions about the condition of the process based on the outcome of this test. The establishment of suitable check standards and implementation of statistical control procedures are discussed in the next two chapters with implementation for specific cases being outlined in chapter 4.

The determination of systematic error is made by intercomparing the laboratory's reference standard(s) or measurement system with national standards or a measurement system maintained by the National Bureau of Standards. This intercomparison can be interfaced with NBS in one of three ways. Firstly, the reference standards can be submitted to the usual calibration exercise wherein values and associated uncertainties are assigned to the reference standards by NBS. The only sources of systematic error that are identifiable in this mode are directly related to the reference standards themselves and to the NBS calibration process. The name "measurement assurance program" is not formally attached to such efforts because the NBS involvement is limited and measurement control is left entirely to the participant, but the goal of measurement assurance is certainly realizable by this route.

Secondly, systematic error can be identified by internal calibration of instrumentation or reference standards through use of a standard reference material distributed by NBS. Thirdly, systematic error can be determined by a formal program in which an NBS calibrated artifact, called a transfer standard, is treated as an unknown in the participant's measurement process. The difference between the participant's assignment for the transfer standard and the NBS assignment determines the offset of the participant's process or reference standards from NBS.

The National Bureau of Standards provides measurement assurance related services that utilize the latter two courses, especially the use of transfer standards, in selected measurement areas [16]. A standard reference material and a transfer standard are comparable in the measurement assurance context. The latter is referred to more frequently in this publication because transfer standards are more publicized in connection with measurement assurance.

The development of a program which satisfies the goals of measurement assurance begins with the measurement problem which must be related to physical reality by a statement called a model. Models covering three aspects of metrology are discussed in this chapter. The first of these, the physical model, relates the realization of the quantity of interest by a measurement process to the fundamental definition for that quantity. Physical models change with changes in fundamental definitions.

For example, until 1960, the standard of length was "the distance between two scratch marks on the platinum-iridium meter bar at the Bureau International des Poids et Mesures" [17]. Models for realizing length related the intercomparison between the international meter bar and the national meter bar and subsequent intercomparison between the national meter bar and gage block standards. In 1960 length was redefined in terms of the wavelength of radiation of krypton-86. The defining wavelength of 86Kr was related to the wavelength of a stabilized laser light[a], thus establishing the relationship of interference fringe patterns observed with the laser interferometer to the length of gage blocks standards. Length has recently been redefined in terms of the velocity of light. This latest change will necessitate another model relating standards to "the length of the path traveled by light in a vacuum during a (given) time interval" [17].

The calibration model describes the relationship among reference standards, items to which the quantity is to be transferred such as secondary or laboratory reference standards, and the instrumentation that is used in the calibration process. For example, calibration of gage blocks by electromechanical intercomparison with gage block standards that have been measured interferometrically includes a correction for the temperature coefficient of blocks longer than 0.35 inches [19]. The calibration model for these intercomparisons assumes a constant instrumental offset that is canceled by the calibration experiment as discussed in section 1.4.

Statistical models further refine the relationship among calibration measurements in terms of the error structure. Section 1.5 describes the type of error structure that is assumed for measurement assurance programs taking as an example the eletromechanical comparison of gage blocks according to the scheme outlined in section 4.

Modeling, usually the responsibility of the national laboratory, is emphasized in this chapter partly to lay the foundation for the remainder of the text and partly so that the reader can form some idea of the degree of success that can be expected from a measurement assurance program. It is an implicit assumption that the validity of any intercomparison, either between transfer standards and reference standards or between reference standards and the workload, depends upon all items responding to test conditions in fundamentally the same way as described by the models.

---

[a] "Direct calibration of the laser wavelength against 86Kr is possible, but is relatively tedius and expensive. The procedure used is a heterodyne comparison of the stabilized He-Ne laser with an iodione stabilized laser" (Pontius [18]).

This logic leads us the next major phase in development--the test of the measurement prescription as a device for transferring a quantity of measurement from the national laboratory to a laboratory participating in a measurement assurance program. The final phase--the application of quality control techniques to the measurement process ensures a continuing tie to the national system of measurement. Several activities can take place during each of these phases. These are listed either in section 1.6 under the role of NBS or in section 1.7 under the role of the participant although it is clear that in practice there is some overlapping of these responsibilities.

In summary, measurement assurance implies that the determination of systematic error and the assignment of values to standards has been done correctly at every step in the measurement chain and, moreover, that this is guaranteed by a statistical control program that is capable of identifying problem measurements at every transfer point in the chain. Accomodation to these principles may require modification of the laboratory's calibration procedures. Where an NBS transfer standard is used to directly calibrate reference standards, the same measurement process and control procedures that are used for this intercomparison should be used for the regular workload. Where the transfer standard is used to calibrate a laboratory's primary standards, a statistical control program should be implemented for this intercomparison, along with similar control programs for the intercomparison of the primary standard with the reference standards and for the intercomparison of the reference standards with the workload. Obviously, the effort required to maintain such a system is greater than is required to maintain a current calibration on the reference standards. Measurement assurance places a substantial portion of the burden of proof on the participant, where it should rightfully be, because it is the quality of his measurements that is of ultimate interest.

1.3 Models for a Measurement System

A measurement system that relies on an artifact demands that the artifact play "two essential roles in the system; it must embody the quantity of interest, and it must produce a signal, (such as the deflection of a pointer on a scale or an electrical impulse) which is unambiguously related to the magnitude or intensity of the specified quantity" (Simpson [20]). The first step that must be undertaken in constructing a measurement system is to reduce the artifact to an idealized model which represents those properties believed to be pertinent to the intended measurement.

This model of the measurement process, based on the laws of physics, in the broadest sense embodies our understanding of the physical universe. It is usually a software model or statement that relates the signal produced by the artifact and all procedures used to produce the desired measured value, called the measurement algorithm, to the realization of the physical quantity of interest taking into account any factors such as environmental conditions that affect this realization.

The integrated circuit industry is a case study of a measurement problem not properly defined in terms of an artifact model. Inability throughout the industry to measure optically the widths of chromium lines to the accuracies needed for producing photomasks for integrated circuits can be traced to misconceptions about the nature of linewidth measurements--misconceptions that led to reliance on a line-scale calibration for making such measurements, in the hope that a correct line-scale for the optical system would guarantee accurate linewidth measurements.

Before attempting to produce a linewidth standard, NBS explored the nature of the systematic errors that are inherent in line-scale and linewidth measurements (Nyyssonen [21]). Line-scale defines the internal ruler of an instrument i.e. it is basically a left-edge to left-edge or a right-edge to right-edge measurement for which any bias in detecting edge location is assumed to cancel out. Linewidth, a more difficult determination, measures the width of a physical object, in this case a chromium line. It is a left-edge to right-edge measurement in which any bias in detecting edge location is assumed to be additive (Jerke [22]).

This theoretical modeling was corroborated by an interlaboratory study which demonstrated that an optical imaging system, although properly calibrated for line-scale, would not necessarily produce linewidth measurements with negligible systematic errors. The study also demonstrated that the same system when properly calibrated using a linewidth artifact would produce linewidth measurements with negligible systematic errors (Jerke et al [23]).

A model is never complete or perfect, and the difference between the model and reality leads to "a particular type of systematic error which exists if the measurement algorithm is flawless. Failure to recognize this fact can lead to major wastes of resources since no improvement in the measurement algorithm can reduce this error" (Simpson [24]).

Thus even though NBS semiconductor research has greatly enhanced linewidth measurement capability, the accuracy of linewidth measurement is still constrained by the difference between the real edge profile of a chromium line and a theoretical profile (Nyyssonen [25]) upon which the model depends. The discrepancy between the edges of chromium lines on production photomasks and the theoretical model is a limiting factor in attaining measurement agreement among photomasks makers, and it will not be reduced by finer tuning of the optical imaging systems or more accurate standards. This points out a problem that exists in going from the calibration laboratory with carefully fabricated artifacts to the production line and prompts us to include a caveat for the claims of measurement assurance programs. This type of systematic error is kept at an acceptable level only if the measured items are close in character to the standards and theoretical model on which their assignments depend. The only strategy which can reduce model ambiguity identically to zero uses objects called "prototypes" and, in effect, takes a particular object and defines it to be its own model. As pointed out by Simpson [26],

> This amounts to saying that this object is the perfect and complete realization of the class of objects to which it belongs, and hence the model ambiguity is, by definition, identically zero. The only SI unit still using this strategy is mass where the Paris[b] Kilogram is the kilogram of mass, and the only objects where mass can be unequivocally defined are one kilogram weights made of platinum.

The comparison of a non-platinum kilogram with the Paris kilogram would produce a systematic error unless the comparison was done in vacuum. High accuracy mass calibrations in air are corrected for air buoyancy -- a correction that depends on the material properties of the weight, temperature on the weight at the time of weighing and the local pressure and humidity. Any ambiguity between the model that drives this correction and the Paris kilogram in vacuum contributes a systematic error to the calibration process although admittedly this error is negligible.

1.4 Models for a Calibration Process

1.4.1 The Calibration Experiment

The exploration of the physical and mathematical models that relate a measurement to a quantity of interest leads to a measurement algorithm which defines a reference standard, instrumentation, environmental controls, measurement practices and procedures, and computational techniques for calibrating other artifacts or instruments with respect to the desired property.

> Calibration is a measurement process that assigns values to the response of an instrument or the property of an artifact relative to reference standards or measuring processes. This may involve determining the corrections to the scale (as with direct-reading instruments), determining the response curve of an instrument or artifact as a function of changes in a second variable (as with platinum resistance thermometers), or assigning values to reference objects (as with standards of mass, voltage, etc.) (Cameron [27]).

Calibration consists of comparing an "unknown" or test item which can be an artifact or instrument with reference standards according to the measurement algorithm. The calibration model, which addresses the relationship among measurements of test items and reference standards, must reflect the fact that the individual readings on the test items and reference standards are subject to systematic error that is a function of the measuring system and random error that may be a function of many uncontrollable factors.

---

[b]The international standard of mass resides at the Bureau International des Poids et Mesures in Sèvres, just outside Paris.

There are two common generic types of calibration models, additive models and multiplicative models. Reduction of systematic error by intercomparison with a reference standard involves estimating offset as either an additive factor Δ or a scale factor λ which in turn is used to assign a value to the test item relative to the known value of the reference standard. The choice of an additive or multiplicative model depends on the nature of the relationship among test items and reference standards and properties of the measuring system.

The calibration experiment is designed not only to assign values to test items that will account for systematic error between the requestor and the calibrator but also to estimate the magnitude of random errors in the calibration process. The nature of random error is discussed more fully in section 2.2, but suffice it to say for now that we are talking about small fluctuations that affect every measurement but are unmeasurable themselves for a given measurement. The statistical derivations in this manuscript assume that the random errors are independent and that they affect the measuring process symetrically i.e., that one is not predictable in size or direction from any other one and that the chances are equal of the resulting measurement being either too large or too small. It is also assumed that random errors for a given process conform to a law called a statistical distribution; quite commonly this is assumed to be the normal distribution, and the calibration experiment is designed to estimate a standard deviation which describes the exact shape of this distribution.

In the next three sections we list models that are in common usage in calibration work, and although the list is not exhaustive, it includes those models which form the basis for the calibration schemes in chapter 4. It is noted that the term "reading" or "measurement" in this context does not refer to a raw measurement, but rather to the raw measurement corrected for physical model specifications as discussed in the last section.

1.4.2 Models for Artifact Calibration[c]

In the simplest additive model for a calibration process, a test item x with a value X^*, as yet to be determined, and a reference standard R with a known or assigned value R^* are assumed to be related by:

$$X^* = \Delta + R^* \qquad (1.4.1)$$

where Δ is small but not negligible. The method for estimating the offset Δ between the two artifacts depends upon the response of the calibrating instrument.

If the calibrating instrument is without systematic error, the instrument response x for any item X will attain the value X^* except for the effect of random error; i.e., the instrument responds according to the model

$$x = X^* + \varepsilon$$

[c] The models for artifact calibration are also appropriate for single-point instrument calibration.

where ε represents the random error term. In this case there is no need to compare the test item with a reference standard because the capacity for making the transfer resides in the calibrating instrument. Such is assumed to be the case for direct reading instruments. Normally the calibrating instrument is not invested with such properties, and one calibration approach is to select a reference standard that is almost identical to the test item and compare the two using a comparator type of instrument for which additive instrumental offset is cancelled out in the calibration procedure. Given that the comparator produces a measurement x on the test item and a measurement r on the reference standard, the response is assumed to be of the form:

$$x = \psi + X^* + \varepsilon_x$$

and (1.4.2)

$$r = \psi + R^* + \varepsilon_r$$

where ψ is instrumental offset and the ε_x and ε_r are independent random errors. An estimate[¶] of Δ is gotten by the difference

$$\hat{\Delta} = x - r \ , \tag{1.4.3}$$

and the value of the test item is reported as

$$\hat{X}^* = \hat{\Delta} + R^* \ .$$

An inherent deficiency in relying on a single difference to estimate Δ is that it does not admit a way of assessing the size of the random errors. If the calibration procedure is repeated k times in such a way that the random errors from each repetition can be presumed to be independent, the model for k pairs of readings r_j, x_j $(j=1,\cdots,k)$ becomes

$$x_j = \psi + X^* + \varepsilon_{x_j}$$

$$r_j = \psi + R^* + \varepsilon_{r_j} \tag{1.4.4}$$

and the offset is estimated by

$$\hat{\Delta} = \frac{1}{k} \sum_{i=1}^{k} (x_i - r_i) \ . \tag{1.4.5}$$

Given the further assumption that all the random errors come from the same distribution, the magnitudes of the random errors can be quantified by a standard deviation (see Ku [28] for a clear and concise discussion of standard deviations).

Another less frequently assumed response for a calibrating instrument allows not only for instrumental offset ψ but also for a non-constant error that depends on the item being measured. This type of response is sometimes referred to as non-linear behavior, and in this case two reference standards with known values R_1^* and R_2^* are required to estimate X^*. Given measurements r_1 on the first standard and r_2 on the second standard, the instrument response for the three artifacts is described by:

[¶] The caret (^) over a symbol such as Δ denotes an estimate of the parameter from the data. It is dropped in future chapters where the intent is obvious.

$$x = \psi + \beta X^* + \varepsilon_x$$

$$r_1 = \psi + \beta R_1^* + \varepsilon_{r_1} \qquad (1.4.6)$$

$$\text{and} \quad r_2 = \psi + \beta R_2^* + \varepsilon_{r_2}$$

where the parameter β is non-trivial and different from one, and ε_x, ε_{r_1} and ε_{r_2} are independent random errors.

Then the measured differences $x-r_1$ and r_2-r_1 are used to construct an estimate of Δ, namely,

$$\hat{\Delta} = (R_2^* - R_1^*)\cdot(x - r_1)/(r_2 - r_1). \qquad (1.4.7)$$

The calibrated value of the test item is reported as

$$\hat{X}^* = \hat{\Delta} + R_1^*. \qquad (1.4.8)$$

Equivalently, Δ can be estimated by

$$\hat{\Delta} = (R_1^* - R_2^*)\cdot(x - r_2)/(r_1 - r_2)$$

in which case

$$\hat{X}^* = \Delta + R_2^*.$$

In order to achieve symmetry in the use of the reference standards, before and after readings, x_1 and x_2, can be taken on the test items with the readings in in order x_1, r_1, r_2, and x_2. Then Δ is estimated by

$$\hat{\Delta} = \frac{1}{2}(R_2^* - R_1^*)\cdot(x_1 - r_1 - r_2 + x_2)/(r_2 - r_1), \qquad (1.4.8a)$$

and the value for the test item is given by

$$\hat{X}^* = \hat{\Delta} + \frac{1}{2}(R_1^* + R_2^*).$$

In comparing the models in (1.4.2) and (1.4.6) one sees that the former model amounts to the slope β of the response curve of the instrument being identically one. If this slope is in fact close to one, which is certainly a reasonable assumption for most instruments, any departure from this assumption will contribute only a small systematic error in the assignment to the test item because of the small interval over which the measurements are taken. For this reason (1.4.2) is the commonly accepted model for calibration processes that use a comparator system of measurement.

The model in (1.4.6) amounts to a two-point calibration of the response function of the instrument; it is not dependent on a small calibration interval; and it is commonly used for direct-reading instruments. Notice that for either model a valid calibration for the test item does not depend on the response parameters of the instrument as long as they remain stable.

A multiplicative model for calibration assumes that the test item X and the reference standard R are related by

$$X^* = \gamma R^* \tag{1.4.9}$$

and that the measuring instrument has a response function of the form

$$x = \beta X^* + \varepsilon_x \tag{1.4.10}$$

$$r = \beta R^* + \varepsilon_r$$

where β and ε_x and ε_r are defined as before. The model leads to an estimate of γ; namely,

$$\hat{\gamma} = x/r \ . \tag{1.4.11}$$

The calibrated value of the test item is reported as

$$X^* = \hat{\gamma} R^* . \tag{1.4.12}$$

1.4.3 Models for Instrument Calibration

Models for instrument calibration relate the response of the instrument to a known stimulus called the independent variable. Where non-constant response of the instrument over a range of stimuli can be either theoretically or empirically related to the stimulus, the relationship is called a calibration curve.

The model for a calibration curve assumes that a response X is offset from a known stimulus W by an amount $\Delta(W)$ that depends on W and that the relationship holds over the entire calibration interval within a random error ε. A relationship of the form

$$X = \alpha + \beta W + \varepsilon \tag{1.4.13}$$

where α and β may be unknown parameters is called a linear calibration curve.

Once the parameters of the calibration curve are known or have been estimated by an experiment, future responses can be related back to their corresponding stimuli. In the general case this inversion is not easy nor is the attendant error analysis very tractable because the calibration curve is used in the reverse of the way that the data are fitted by least-squares.

The only case where the solution is straightforward is the linear case where a series of readings $X_j (j=1,\cdots,n)$ at designated points W_j^* $(j=1,\cdots,n)$ are used to obtain estimates $\hat{\alpha}$ and $\hat{\beta}$ of the parameters. The best estimate of offset for the linear case is

$$\hat{\Delta}(W) = \hat{\alpha} + \hat{\beta}(W). \tag{1.4.14}$$

Methods for estimating the parameters and quantifying the random error are discussed by Mandel [29].

1.5 Models for Error Analysis

The models in sections 1.4.2 and 1.4.3 admit random errors that come from a single error distribution whose standard deviation is of interest in quantifying the variability in the calibration process. We now expand this concept to models that include two distinct types of random errors; a random error term for short-term repetitions that is usually attributed to instrument variability and a random error term that allows for changes that are dependent on the conditions of the calibration and as such are assumed to remain constant for a given calibration. These two types of errors give rise to two distinct error distributions with associated standard deviations which can be estimated from the calibration data. The former is usually referred to as a "within" standard deviation and is designated by s_w.

The latter referred to as a "between" standard deviation, meaning between calibrations and designated by s_b, is attributed to changes in the calibration process from day-to-day. These include environmental changes that are not accounted for by modeling, changes in artifact alignment relative to the standard, and other fluctuations that are not reflected in the within standard deviation. For example, the model in (1.4.4) can be rewritten in terms of measured differences d_j $(j=1,\cdots,k)$ as

$$d_j = x_j - r_j = X^* - R^* + \varepsilon_j \qquad (1.5.1)$$

where the subscript j denotes short-term repetition and the ε_j are independent random errors that come from a distribution with standard deviation s_w. When this model is expanded to allow for day-to-day changes, the model becomes

$$d_j = (X^* + \delta_X) - (R^* + \delta_R) + \varepsilon_j \qquad (1.5.2)$$

where δ_X and δ_R are assumed to be independent random errors that come from a distribution with standard deviation s_b.

The quantities s_w and s_b, while of interest in their own right, are components of a "total" standard deviation that includes both "within" and "between" type variations in the measurement process. It is this total standard deviation, whose structure is discussed at length in this and later chapters, that is of primary interest in measurement assurance. The reader can verify that the proposed approach to error modeling is compatible with a components of variance model [30] by considering model (1.5.2) which leads to the estimate of offset given in (1.4.5). In terms of the error structure this offset is

$$\hat{\Delta} = (X^* - R^*) + (\delta_X - \delta_R) + \frac{1}{k}\sum_{j=1}^{k} \varepsilon_j \quad .$$

It can be shown‡ that a reported value based on a single (k=1) measured difference has standard deviation

$$s_r = (2s_b^2 + s_w^2)^{1/2} \quad .$$

‡The methodology for arriving at the standard deviation is not explained in this publication. See Ku [28], pages 312-314, for the computation of standard deviations when several independent errors are involved.

A reported value based on the average of k short-term differences has standard deviation

$$s_r = (2s_b^2 + s_w^2/k)^{1/2}.$$

Notice that the contribution of the component s_b to the standard deviation s_r is not reduced by taking multiple measurements that are closely spaced in time. This is the reason for discouraging short-term repetitions in measurement assurance and insisting that the definition of the total standard deviation encompass a broad range of operating conditions in the laboratory--implications which will be addressed in some detail in later chapters.

In this manuscript the total standard deviation s_c is defined to be the standard deviation of a "check standard" value as estimated from repeated calibration of the check standard. Where the error structure for the check standard value is the same as the error structure for the reported value of the test item, the standard deviation of the reported value which we call s_r, is exactly s_c. Otherwise, s_r must be adjusted accordingly. For example, suppose that a test item X with unknown value X^* is compared with two reference standards R_1 and R_2 with known values R_1^* and R_2^* by consecutive readings x_1, r_1, r_2, x_2 as described in section 4.2.

The error model for the measured differences

$$d_1 = x_1 - r_1$$

and

$$d_2 = x_2 - r_2$$

can be written as

$$\begin{aligned} d_1 &= (X^* + \delta_1) - (R_1^* + \delta_2) + \varepsilon_1 \\ d_2 &= (X^* + \delta_3) - (R_2^* + \delta_4) + \varepsilon_2 \end{aligned} \qquad (1.5.3)$$

where it is assumed that δ_1, δ_2, δ_3 and δ_4 have standard deviation s_b and ε_1 and ε_2 have standard deviation s_w.

The offset is estimated by

$$\hat{\Delta} = \frac{1}{2}(d_1 + d_2) \qquad (1.5.4)$$

and in terms of the error model

$$\hat{\Delta} = X^* - \frac{1}{2}(R_1^* + R_2^*) + \frac{1}{2}(\delta_1 - \delta_2 + \delta_3 - \delta_4 + \varepsilon_1 + \varepsilon_2). \qquad (1.5.5)$$

A check standard defined as the difference between R_1 and R_2 is computed for each calibration by

$$c = (d_2 - d_1). \qquad (1.5.6)$$

In terms of the errors the check standard measurement can be written

$$c = (R_1^* - R_2^*) + (-\delta_1 + \delta_2 + \delta_3 - \delta_4 - \varepsilon_1 + \varepsilon_2) \qquad (1.5.7)$$

The error model (1.5.5) for the reported value

$$X^* = \hat{\Delta} + \frac{1}{2}(R_1^* + R_2^*), \qquad (1.5.8)$$

and the error model (1.5.7) for the check standard measurment c are comprised of the same error terms and differ structurally by a factor of two.

Explicitly, the standard deviation of the reported value X^* is

$$s_r = \frac{1}{2}(4s_b^2 + 2s_w^2)^{1/2} \qquad (1.5.9)$$

and the standard deviation of c is

$$s_c = (4s_b^2 + 2s_w^2)^{1/2}. \qquad (1.5.10)$$

Therefore,

$$s_r = \frac{s_c}{2} \qquad (1.5.11)$$

In practice s_c is estimated by check standard measurements from many calibrations (see chapter 4), and this estimate is used in (1.5.11) to compute s_r.

Where the check standard value is a least-squares estimate from a design or a function of measurements on more than one artifact, the computation of the standard deviation of a reported value is more complicated. In such a case, one must first estimate s_w from a single calibration and compute s_b from an equation for s_c such as (1.5.10). Then the standard deviation of the reported value can be computed from an equation such as (1.5.9).

1.6 NBS Role in the Development of a Measurement Assurance Program

1.6.1 Study of Operations at Participating Laboratories

Before undertaking the development of a measurement assurance program for disseminating a unit of measurement, NBS technical staff familiarize themselves with operations at potential user laboratories so that the program can be structured around the equipment, facilities and personnel available to the laboratories. Suggestions for equipment modifications and additions are made at this time. The range of operating conditions in the participating laboratories is checked for consistency with the model, and in order to determine whether or not the accuracy goals of the measurement assurance program are attainable, NBS is advised of the individual laboratory's measurement requirements and capabilities.

1.6.2 Identification of Factors Capable of Perturbing the System

It is the responsibility of NBS to identify and isolate those factors capable of seriously disrupting the measurement system so that equipment and procedures can be designed to offset the impact of such factors (Youden [31]). This is particularly important if the measurement assurance program is intended for an industrial setting rather than a controlled laboratory setting.

An example of this type of testing, called "ruggedness" testing is found in the NBS flowmeter program for liquids (Mattingly et al [32]). The effects of three types of perturbation on turbine meters were studied experimentally, and it was found that the profile of the flow entering the meter has a significant effect on meter performance. This research led to the development of a flow conditioner which can be inserted in an upstream section of pipe to regulate the profile of the flow entering the meter. Because flow profiles vary from laboratory to laboratory depending on the source of the flow, such a flow conditioner is appended to the turbine meters that are circulated in the industry as NBS transfer standards.

1.6.3 Design of Interlaboratory Exchanges

The purpose of the interlaboratory study or round-robin test that is usually sponsored by NBS at the inception of a measurement assurance program is to determine the extent and size of offsets from NBS that are typical in the target industry. Secondary goals are the evaluation of the adequacy of proposed procedures for resolving the measurement problem, critique of the format and content of directions from NBS, and study of the ease of implementation on the part of participants. Frequently a preliminary interlaboratory test designed to identify significant problem areas is followed by a more comprehensive study which incorporates modifications to artifacts and protocols based on experience gained in the preliminary test.

1.6.4 Development of a Stable Transfer Standard or Standard Reference Material

Either a standard reference material or a transfer standard is developed for each measurement assurance program that is sponsored by NBS. The standard reference material (SRM) is a stable artifact produced either commercially or in-house that is calibrated, certified and sold by NBS in fairly large numbers.[d] Standard reference materials are well known for chemical applications. Recently NBS has certified two separate dimensional artifact standards as SRMs, one a linewidth standard for the integrated circuit industry [NBS SRM-474] and the other a magnification standard for scanning electron microscopes [NBS SRM-484]. An SRM has the unique property that it can be used not only for determining offset from NBS but also as an in-house standard for controlling the measurement process.

[d] A listing of SRM's is contained in the catalog of NBS Standard Reference Materials, NBS Special Publication 260, 1979-80 Edition, available from the Office of Standard Reference Materials, NBS, Gaithersburg, MD.

The transfer standard is a calibrated artifact or instrument standard that is used for disseminating the unit of measurement. It is loaned to the participant to be intercompared with the participant's standards or instrumentation under normal operating conditions in order to determine offset from NBS.

Artifacts that are stable with relation to a physical quantity, such as the mass of an object, do not usually pose any special problems when they are used as transfer standards because they can be shipped from one place to another without a change in the quantity of interest. Transfer standards that are not easily transported are packaged in environmentally controlled containers, but additional redundancy in the form of multiple standards and observations is always included in the measurement assurance program whenever the stability of the transfer standard is in question.

1.6.5 Dissemination of Measurement Technology and Documentation

The participant in a measurement assurance program is entitled to draw upon the expertise and experience that resides in the sponsoring NBS technical group. Technical assistance is disseminated by way of NBS publications, ASTM, standards, ANSI standards, laboratory visits, telephone conversations and NBS sponsored seminars. In conjunction with the advent of a new program a series of seminars is usually offered to the public to explain the philosophy, theory, measurement technology and statistical analyses which form the basis for a measurement assurance program in that discipline.

Documentation for standard reference materials is available through NBS Special Publication Series 260. As part of a long range plan to upgrade its calibration services, the National Bureau of Standards has instituted documentation requirements for all calibration services. Documentation includes theory, laboratory setup and practice, measurement technique, maintenance of standards, specification of measurement sequence, protocol for measurement control and determination of final uncertainty. When these publications become available, they will provide the bulk of the documentation that is needed for implementing a measurement assurance program that is related to an NBS calibration service. Insofar as a measurement assurance program as implemented by the participant may differ from the NBS calibration program in regard to the number of standards, specification of measurement sequence, corrections for environmental conditions, estimation of process parameters, and methods for determining offset and uncertainty, additional user oriented documentation may be made available.

1.6.6 Establishment of Measurement Protocol for Intercomparisons with NBS

Measurement assurance programs currently in existence fall into two categories. The first category contains those services which are highly structured for the participant, with regard to the number of laboratory standards to be employed in the transfer with NBS, the number of repetitions to be made in the exchange, and the protocol to be used for establishing an in-house measurement control program. At this time only the Gage Block Measurement Assurance Program (Croarkin et al [33]) and the Mass Measurement Assurance Program fall into this category.

All other programs allow the participant considerable leeway in regard to the items mentioned above in order to make the service compatible with the unique situation in each laboratory. The advantage of operating within the constraints of equipment and staff resources that are already allocated to the laboratory's normal workload is obvious, especially where accuracy requirements are not difficult to meet. However, there are drawbacks. The data analysis must be tailored to each participant, imposing an additional burden on NBS staff, and responsibility for instituting a rigorous measurement control program is left entirely to the participant.

1.6.7 Data Analyses and Determination of Offset

The determination of offset and associated uncertainty as realized by intercomparison of laboratory reference standards with NBS transfer standards is accomplished in one of two ways:

i) The transfer standard(s) is sent to the participant as a blind sample, and the data from the intercomparison are transmitted to NBS. Based upon the value assigned to the transfer standard by NBS and associated uncertainty from the NBS process, new values with associated uncertainties are assigned to the laboratory standards along with the uncertainty that is appropriate for an item measured by the participant's process.

ii) the transfer standard along with the its assigned value and associated uncertainty are transmitted to the participant, and the analyses and determination of offset become the responsibility of the participant.

Data analyses relating to the regular workload and measurement control procedures in a laboratory are best left to the individual participant. These analyses provide important insights into the pecularities of a measurement process, and, consequently, these analysis are best done internally. Even where much or all of the data analysis is undertaken by NBS, participants are encouraged to develop facility in this area in order to make themselves independent from NBS in the future. Some participants in measurement assurance programs have automated the analysis of calibration data, decisions relating to process control, updating of data files and final determination of uncertainty on minicomputers in their laboratories.

1.7 Participant's Role in a Measurement Assurance Program

1.7.1 Staff Preparation

The success of a properly conceived measurement assurance program depends upon the enthusiasm and dedication of the personnel who are making the measurements and resolving problems that arise in day-to-day operations. The measurement assurance approach is a long-term commitment in terms of evolving a measurement control technique that continually checks on the state of control of the process. Before undertaking such a program, there should be reasonable assurance of continuity of personnel assigned to the project, and steps should be taken to guarantee that new personnel are sufficiently prepared for taking on the assignment before the departure of experienced personnel.

The success of such a program also depends on a certain depth of understanding on the part of the staff. Here we are talking not so much about the intricacies of a particular analysis, but about a basic understanding of scientific methodology, the philosophy of measurement assurance, and the relationship between the control techniques and the validity of the values reported by the measurement process and their associated uncertainties. To this end, NBS offers seminars in which the attendees are instructed in these principles, but some prior staff preparation may be necessary in order to benefit fully from these expositions. Courses at local community colleges are recommended for exploring scientific principles and gaining facility with fundamental mathematical and statistical manipulations.

1.7.2 Selection of a Check Standard

The selection of a check standard must be considered in the preliminary planning for measurement assurance program. In short, its purpose is to provide a continuing thread that characterizes the operation of the measurement process over changing laboratory conditions and over time with regard to both the variability of the process and the long-term average of the process. It is a basic tenet of measurement assurance that the response of the process to the check standard be sufficiently similar to the response of the process to the test items that the performance of the process at all times can be adequately monitored by monitoring the response of the process to the check standard. The value of the check standard at any given time is a decision-making tool, and unexpected behavior on its part is grounds for discontinuing the process until statistical control is resumed.

Careful consideration should be given to the type of artifact that would be suitable for this purpose. It should certainly be of the same character as the items that constitute the workload in the laboratory. For some processes, such as processes dealing with basic units of measurement, the selection is obvious; check standard artifacts are similar to reference standards in design and quality. In general, an artifact that is less stable than the reference standards will not be useful as a check standard if its instability is large enough to mask the properties of the measurement process.

The check standard should be thought of not so much as an artifact but as a data base because it is the measurements that are of interest and not the artifact per se. The check standard data base consists of measurements, properly corrected for environmental factors, or some function of those measurements that have been made on the artifact check standard or on the reference standards. For example, a test item that is compared to two reference standards has its assignment based on the average of the values assigned to the two reference standards. The check standard can be defined to be the difference between the measurements on the reference standards thus eliminating the need for an extraneous measurement or other artifact. Where a calibration involves only one reference standard, an artifact that is similar in response to the test items can be designated as the artifact check standard. This need not be a calibrated artifact, and the properties of the measurement process are ascribed to it as long as it is measured in the same time frame as the other items in the calibration process. Several check standards used separately or in combination may be employed when the stability of the reference standards, such as a bank of standard cells, is cause for concern.

Where reference standards exist at several levels, such as mass standards or length standards, check standards are maintained and monitored at each level. Where the quantity of interest is propagated over several levels from one standard such as a one ohm resistor, which is used to propagate resistances between one and ten ohms, the same check standard artifact may be employed at the different levels, but the data bases for the different levels are regarded as separate check standards.

An SRM makes an ideal check standard if it is not contaminated or otherwise degraded by heavy usage. In any case the artifact or artifacts on which the check standard base is built must be readily available to the measurement process over a long period of time.

The proliferation of check standards involves no small amount of work in maintaining the data base, and serious thought should be given to placement of check standards in the measurement echelon. For a new program, one should start with check standards at a few critical points and gradually increase these as experience is gained with the program.

1.7.3 Initial Experiments to Estimate Process Parameters

The establishment of an initial data base for the laboratory's check standards is the first order of business in a new measurement assurance program. Before one attempts to quantify offset, it must be demonstrated that a measurement process does in fact exist; i.e., that measurements from the process satisfy the requirements for statistical control. This presupposes that the process precision is well known and that this can be documented. If, in fact, the documentation of the process has been lax, or if a substantially new process has been instituted for the measurement assurance program, then measurements taken over as long a time period as practical should be made on the check standard(s) in order to estimate the long-term average of the process and the standard deviation. Procedures for obtaining these initial estimates are discussed in subsequent chapters.

A laboratory planning a transfer with NBS should undertake these experiments well in advance of the arrival of the NBS transfer standard so that any problems encountered in the measuring system can be rectified. This provides a shake-down period for procedures, equipment and software involved in the measurement assurance program. Once the transfer standards are intercompared with the laboratory's reference standards, the resulting measurements involving the check standard are compared with the initial data base to decide if the process is in control at that time, and the transfer between the laboratory process and the NBS process is accomplished only if the process is judged in control. Therefore, participants are urged to make the initial experiments as representative of laboratory conditions as possible and to request help from the sponsoring NBS group if measurement problems or procedural ambiguities exist so that delays with the transfer can be avoided.

1.7.4 Calibration Procedures

Accomodation to measurement assurance principles can mandate a change in calibration procedures within the laboratory. Most often such change will amount to additional redundancy in the design and/or change in the order of measurements. The laboratory should settle upon one calibration design for the

transfer with NBS and the calibration workload. There is considerable advantage in doing this because the uncertainty determined from the transfer with NBS is only valid for that measurement process, and if the uncertainty is to have validity for the workload, the two measurement processes must be identical. There is a further advantage; the same statistical control program will suffice for both processes, and the check standard measurements from both sources can be combined into a single data base.

Another consideration is the manner in which systematic error is handled in the transfer experiment. Some measurement assurance programs are structured so that the determination of systematic error is made relative to the average of two or more reference standards as in section 4.2.4. For example, two reference gage blocks can be calibrated by intercomparison with two NBS transfer blocks by a design that assigns values relative to the average of the two reference blocks called the restraint. Systematic error is estimated as the difference between the restraint and the average computed for the two NBS blocks by the transfer experiment. The laboratory's restraint is then corrected for this offset. Meaningful values cannot be computed for the reference standards individually from the transfer experiment. Thus, the same design that is used for the transfer with NBS is employed in the calibration workload so that all assignments are made relative to the corrected restraint.

1.7.5 Process Control

The measurement assurance concept demands that a value be assigned to an artifact only when the measurement process is in control in order to guarantee the validity of the assignment and associated uncertainty statement. This means that statistical control is employed in the everyday workload of the laboratory as well as during the transfer with NBS. For highest accuracy work, comparable to calibrations at NBS, a check for control is made during every measuring sequence in which an artifact is calibrated by the system. Statistical control procedures based on check standard measurements along with the appropriate statistical tests are discussed in section 3.3.

The choice of a control procedure and its implementation are the responsibility of the participant. Those who are familiar with industrial quality control procedures and Shewhart type control charts should be able to adapt these methodologies to check standard measurements. A general discussion of control charts with examples is contained in chapter 5, and statistical control procedures for specific measurement situations are outlined in chapter 4.

1.7.6 Data Base Maintenance

A record of check standard measurements is kept separately from other laboratory records such as records of past calibrations. This permanent record should include all pertinent information relating to the measurement. For example, it normally includes an identification for the check standard, identification for the instrument, identification for the operator, day, month, year, identification for the type of statistical design used in the intercomparison, observed value of the check standard, environmental conditions that could affect the measurement such as temperature, pressure and relative humidity, standard deviation if applicable, and finally a flag denoting whether or not the check standard was in control on that occasion.

2. Characterization of Error

2.1 Introduction

It is the purpose of this chapter to introduce the reader to the concepts of random error, systematic error and uncertainty. It is the expressed purpose of measurement assurance to identify and quantify all sources of error in the measurement process, because in so doing, the worth of any value reported by the process can be stated in quantitative terms called an uncertainty. In a very real sense, a value assigned to an artifact only has meaning when there is an assessment of how well that number describes the property of interest (be it length, mass or whatever) in terms of its reference base. An uncertainty statement provides that assessment.

Error in measurement is categorized as either systematic, coming from a source that is constant and ever present in the measurement process, or random, coming from a source (or sources) that is continually fluctuating. Systematic error may be known or estimable for any given situation, but random error by its nature is never known for a given measurement situation. The point is that for a single measurement it may be possible to determine the size of the systematic error by intercomparison. On the other hand, the random error that is unique to a single measurement cannot be replicated because conditions of measurement cannot be repeated exactly. Therefore, it is common practice in metrology, as it is in process control [34], to quote limits to random error for all such experiments.

Classification of sources of error as either systematic or random is not always straightforward depending as it does on the way in which the potential source of error enters the measurement process, how it affects the output of that process, and the interpretation of the uncertainty. For example, the maximum observed difference between operators can define a systematic error for a system that is highly operator dependent and for which there are a restricted number of operators or, alternatively, a separate uncertainty statement can be issued for each operator's measurements. Measurement systems that routinely make use of many operators are better served by folding the effect of operator error into the total random error for that system.

At the National Bureau of Standards considerable attention is given to the classification of sources of error. For the participant in a measurement assurance program, systematic error is usually assumed to come from specific sources that are spelled out in this chapter, and remaining sources of error are assumed to be part of the random error of the participant's process and must be estimated as such.

2.2 Process Precision and Random Error

2.2.1 The Standard Deviation

A "measurement process" is said to exist for quantifying a physical attribute of an object, such as its length, only if the process is operating in a state-of-control (Eisenhart [35]). The fact is that, even for such a process, repeated measurements on the same object will not produce identical results. As long as the source of this disagreement is random in nature; i.e., its direction and magnitude not being predictable for any future measurement, the disagreement among measurements is referred to as the process imprecision. A measure of precision, such as the process standard deviation, quantifies this random error or scatter or, more aptly, describes the degree of agreement or closeness among successive measurements of the same object.

The term process precision as used in this publication is not limited to the characterization of the behavior of the particular measuring device per se, but it is intended to describe the total configuration of operator, environmental conditions, instrumentation and whatever other variables go into making any given measurement. As it is rarely possible to measure an item submitted for calibration over a representative set of environmental and working conditions in the laboratory, redundancy is obtained from measurements made on a check standard that is introduced into the measurement sequence on a routine basis. It is assumed that the check standard is similar in response to the test item and that the process precision can be estimated from the measurements made on the check standard.

The simplest measure of process precision is the range--the difference between the largest and smallest measurements in the group. The range is a satisfactory measure of precision when the number of measurements is small, say less than ten. It "does not enjoy the desirable property" (Ku [36]) of tending toward a limiting value as more measurements are taken; it can only increase and not decrease. Therefore, it is desirable to find a measure of precision which takes into account the information in all the measurements and which tends to a limiting value as the sample size increases if we are to use this measure to describe the process behavior as a stable phenomenon.

The standard deviation is such a measure. Small values for the standard deviation are indicative of good agreement and large values are indicative of poor agreement. Because it is necessary to distinguish different kinds of variability that contribute to the total process variability, it is likewise necessary to define different kinds of standard deviations. We routinely identify two levels of standard deviations in calibration work.

These two levels are described briefly in the first chapter where we are dealing with the models covering the error structure among measurements. Reiterating, the first type of standard deviation is a measure of the variability of the measurement process over a short period of time, usually the time necessary to complete one calibration using a particular sequence of measurements called a statistical design. This measure is called the "within standard deviation." Its usage as a check on the internal consistency of an individual calibration experiment is explained in chapter 3 and chapter 4 along with formulas and examples.

The second type of standard deviation that we are dealing with in measurement assurance, and by far the more important of the two, is the total standard deviation s_c. This latter measure includes both the "within" component of variability s_w and a "between" component of variability s_b, which the reader will recall explains incremental changes that can take place from calibration to calibration. The relationship among these quantites is assumed to be of the form

$$s_c = (s_w^2 + s_b^2)^{1/2} .$$

Therefore, the total standard deviation, including as it does both "within" and "between" components of variability, should accurately reflect both the short-term and long-term random errors that are affecting the measurement process.

The limits to random error quoted in the uncertainty statement are computed from the total standard deviation thus assuring that the conditions of a single calibration do not invalidate this measure of the quality of the reported value. As has been noted previously, the total standard deviation, not generally being available from the calibration data, is based on repeated check standard measurements that are structured to include all possible sources of random error. This is accomplished by monitoring the check standard over a long period of time and over the full range of environmental factors for which the uncertainty statement is assumed to be valid.

The total standard deviation depends on the physical model. The most familiar form

$$s_c = \left(\frac{1}{n-1} \sum_{i=1}^{n} (c_i - \bar{c})^2 \right)^{1/2} \qquad (2.2.1)$$

where the arithmetic mean is

$$\bar{c} = \frac{1}{n} \sum_{i=1}^{n} c_i \qquad (2.2.2)$$

assumes that check standard measurements $c_1, \cdots, c_n$ are independent of time and that the effect of other variables is negligible.

The term (n-1), called the degrees of freedom associated with s, is an indication of the amount of information in the standard deviation and is always reported along with the standard deviation.

2.2.2 Pooled Standard Deviation

If several standard deviations with small numbers of degrees of freedom are computed from the same process, they will vary considerably among themselves. It goes without saying that the standard deviation that is quoted in the uncertainty statement must have a sufficient amount of information to guarantee that it is a valid measure of process precision. The question is, "How much

redundancy is sufficient?" As a general rule, fifteen degrees of freedom is a minimum for the initial computation of the standard deviation. As the measurement assurance program progresses, the standard deviation is recomputed to take advantage of the increased data base, and assuming that the process is stable, this will assure a more reliable value of the standard deviation. A standard deviation based on as few as two data points can be combined with other similar estimates that have been obtained on separate occasions for the same process to obtain what is called a "pooled" standard deviation. If the individual standard deviations are $s_1,\cdots,s_k$ with degrees of freedom $\nu_1,\cdots,\nu_k$, respectively, the pooled standard deviation is

$$s_p = \left(\frac{\nu_1 s_1^2 + \cdots + \nu_k s_k^2}{\nu_1 + \cdots + \nu_k} \right)^{1/2} . \qquad (2.2.3)$$

The degrees of freedom associated with s_p is $\nu = \nu_1 + \cdots + \nu_k$.

2.2.3 Limits to Random Error

Limits to random error can be computed with a given probability if the distribution of random errors is known. Limits, so stated, depend upon assumptions concerning the average value and spread of the underlying distribution. For a calibration process it is assumed that random errors of measurement have an equal chance of being negative or positive such that their average value is zero. It is also assumed that the spread of the distribution is adequately estimated by the total process standard deviation.

Limits to random error for a single value from the measurement process are constructed so that the probability is $(1-\alpha)$, for α chosen suitably small, that if the measurement algorithm were to be repeated many times, the average outcome of these experiments would fall within $\pm\, s_c \cdot t_{\alpha/2}(\nu)$ of the reported value, where s_c is the total process standard deviation, ν is the number of degrees of freedom in s_c, and $t_{\alpha/2}(\nu)$ is the $\alpha/2$ percent point of Student's t distribution. (See Ku [37] for a further discussion of Student's t distribution.) Critical values for Student's t are given in Table I for $\alpha = 0.05$ and $\alpha = 0.01$ and degrees of freedom $\nu = 2(2)120$.

Frequently a precise probability interpretation for the limits to error is not needed and, in fact, will not be possible if it cannot be demonstrated that the underlying probability distribution for the data is exactly a normal distribution. In metrology the limits to random error are often taken to be three times the standard deviation. Other technical areas may use two standard deviations. The bounds, plus and minus three standard deviations, are statistically robust (with respect to the coverage of the distribution) in that if the experiment were to be repeated, the chance of reporting a value outside of these bounds would be extremely small. This, of course, assumes that the random errors affecting the experiment come from a distribution that is close in character to the normal distribution and that enough data have been collected to provide a reliable estimate of the standard deviation. The examples given in this chapter use three standard deviation limits.

2.3 Systematic Error

2.3.1 Conventional Calibration

Systematic error takes into account those sources of error, peculiar to the measurement system, that remain constant during the calibration process and explain a difference in results, say, between two different measuring systems trying to realize the same quantity through a large number of measurements. Some obvious examples are: uncertainties in values assumed for reference standards, uncertainties related to the geometry or alignment of instrumentation, differences between operators, differences between comparable systems, etc. The size of the possible discrepancy is estimated, either empirically or theoretically, but its direction is not always known.

In order to define systematic error for a calibration process, it is necessary to define the steps in a calibration echelon that relate the measured value of the quantity of interest back to its basic SI unit or to a national standard. NBS, except in the case of international comparisons, occupies the premier position in the U.S. calibration echelon. Thus the first transfer point in this calibration echelon involves the intercomparison of a laboratory reference standard with the national standard maintained by NBS which may be an artifact standard or an instrument. The second transfer point involves the intercomparison of the laboratory reference standard with an unknown which in turn can be a working standard from the same laboratory or an artifact standard from a lower level calibration laboratory or a finished product. The calibration chain is extended in this way until the final product has been calibrated by an intercomparison involving it and a standard which can be traced back to the National Bureau of Standards.

Systematic error is assessed at every transfer point and passed along to the next lower level in the calibration chain. Thus, the total systematic error for the measurement process that delivers the final product is an aggregate of systematic errors from all transfer points. Systematic error must be defined very specifically for each transfer point in terms of the long-term values for measurements from two systems, and it must also include an estimate of the amount by which the higher level system, such as NBS, may be in error in estimating its long-term value.

The purpose of each transfer point is to reduce or eliminate systematic errors at that level. If we look at an exchange between a laboratory and NBS, a potentially large source of systematic error comes from the values assigned to the laboratory's reference standards. Calibration of the reference standards at NBS can eliminate offset from this source, but the calibration itself is still a source of systematic error whose magnitude depends on how well NBS was able to conduct the calibration as measured by the uncertainty associated with the calibrated values.

The rationalization for assessing a systematic error from this source is that the values for the reference standards remain constant as they are used as a reference for assigning values to other artifacts or instruments. At least they remain constant until they are recalibrated at NBS, and the assignments resulting from their use are all affected in the same way, being either too low or too high, even though the direction and exact magnitude of this error are not known. Thus, uncertainties for values of reference standards are regarded as a systematic error in the laboratory's process (Youden [40]).

Systematic error associated with the uncertainty of a reference standard is assessed proportional to the nominal value of the test item and the nominal value of the reference standard. For example, if a one kilogram standard is used in a weighing design to calibrate a 500g weight, the systematic error from this source is one-half of the uncertainty associated with the assignment for the kilogram standard.

If the value for a test item is reported relative to the average of two reference standards R_1 and R_2, all artifacts being of the same nominal size, and if the assignments for R_1 and R_2 are independent, the systematic error from this source is assessed as

$$U = \frac{1}{2}\left(U_{R1}^2 + U_{R2}^2\right)^{1/2}$$

where U_{R1} and U_{R2} are the uncertainties for R_1 and R_2 respectively. Where the assignments to R_1 and R_2 are not done independently

$$U = (U_{R1} + U_{R2})/2.$$

2.3.2 Measurement Assurance Approach

A laboratory participating in a measurement assurance program measures a transfer standard(s) from NBS as if it were an unknown item using the reference standards and instrumentation that constitute the measurement system in that laboratory. The resulting value for the transfer standard, be it based on one measurement or on several repetitions in the laboratory, is compared with the value assigned the transfer standard by NBS. The relationship between the laboratory's assignment and the NBS assignment for the transfer standard defines an offset which is used to correct the values for the laboratory's reference standards.

This approach has an advantage over the usual calibration route as far as identifying systematic error in the laboratory. Either method suffices for identifying errors related to the values of the reference standards, but given that the reference standards are properly calibrated, the particular conditions of their usage in the laboratory may invite systematic errors that are unsuspected and unidentifiable. The dependence of optical systems on operator was mentioned in an earlier chapter, and systematic error caused by operator effect may be significant for other types of systems as well. Also, instrumentation can differ enough that the reference standards alone are not sufficient for eliminating systematic error. Of course, both of these sources of systematic error might be identifiable by proper experimentation, but it would be difficult to assess the magnitude of such errors without the

measurement assurance program. Other factors that are probably not identifiable within the laboratory itself are systematic errors related to lack of proper environmental control or incorrect measurement of temperature and humidity.

Two sources of systematic error are always present in a measurement assurance program. The uncertainty associated with the value of a transfer standard is one. Because another transfer point has been effectively added to the calibration chain, the limits to random error associated with the transfer measurements in the participating laboratory define another systematic error for the laboratory.

2.3.3 Calibration Curve

A more complex situation arises when the purpose of the program is to calibrate an instrument over a range for all continuous values. In this case transfer artifacts are provided at selected points covering the range of interest, and the intercomparisons are used to establish a functional relationship between the instrument and the NBS system. The assignment of values is based on this functional relationship. For example, systematic errors in linewidth measurements produced by an optical imaging system can be reduced relative to the NBS prototype optical system [38] from measurements made on an NBS dimensional artifact. (This artifact is a glass substrate with a series of chromium lines at spacings spanning the range of interest.)

Measurements made on individual lines on the artifact define a functional relationship between the two systems, and a least-squares technique is used to derive a best fitting curve to the measured values as a function of the NBS values. The empirical fit is called the calibration curve.

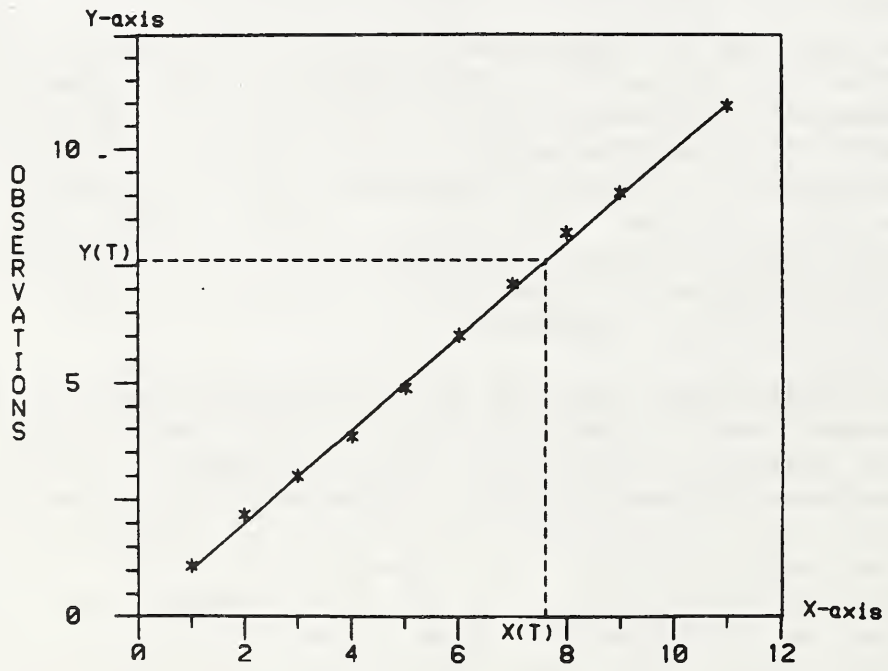


Figure 1
Schematic diagram of a linear calibration curve showing the relationship between an observed value Y(T) and its calibrated value X(T)

In figure 1, each optical measurement is plotted against the corresponding NBS value, and the calibration curve fitted to all the measurements is shown by the solid line. The offset between the user's system and the NBS system is reduced by relating any future measurement back to the NBS value. Schematically, for a future value Y(T) as shown on the Y-axis, a dotted line is drawn through Y(T) parallel to the X-axis. At the point where it intersects the calibration curve another dotted line is drawn parallel to the Y-axis, and its point of intersection on the X-axis, X(T), is the corresponding calibrated value relative to NBS.

Because the functional relationship is not known exactly but is estimated by a series of measurements, the calibration curve can be in error. A discussion of the effect of this error on the final uncertainty of a calibrated value is beyond the scope of this treatise. The reader is referred to Hockersmith and Ku [39] for a discussion relating to quadratic calibration curves and to Croarkin and Varner [40] for a discussion relating to linear calibration curves.

2.4 Uncertainty

2.4.1 Definition

The uncertainty statement assigns credible limits to the accuracy of the reported value stating to what extent that value may differ from its reference base. In practice it quantifies the magnitude of any possible discrepancy between the value actually obtained in the laboratory and the value which would be obtained at NBS for the same property of an object. An uncertainty provides both a measure of the worth of the values reported by the measurement laboratory and an estimate of the systematic error accruing to any organization that makes use of these values.

The uncertainty statement is composed of i) all sources of sytematic error that contribute to the offset from the reference base and ii) a limit to random error that quantifies the variability that is inherent in the measurement process as it transfers from a "known" or calibrated artifact or measurement system to an "unknown".

2.4.2 Combination of Random and Systematic Error

Once the systematic errors and the limits to random error have been estimated, they are combined into a single number which is called the uncertainty. Much controvery arises over the proper way to combine systematic and random errors in an uncertainty statement. Basic premises concerning measurement and its uncertainty as espoused by Youden [41], Eisenhart et al. [42] and others have long been adopted by NBS calibration services and are recommended for measurement assurance programs. A different philosophy that has recently been advanced by the Bureau International des Poids et Mesures is discussed in reference [43]. Basically the question revolves around whether systematic errors should be added linearly or combined in quadrature and around whether the systematic error and the limit to random error should be added linearly or combined in quadrature. For example, if there are several sources of systematic error $S_1,\cdots,S_k$, adding the systematic errors linearly assumes the worst possible combination of errors and gives a total systematic error of

total systematic error S where

$$S = S_1 + S_2 + \cdots + S_k \ . \qquad (2.4.1)$$

Combining the systematic errors in quadrature produces a total systematic error for those sources of

$$S = (S_1^2 + S_2^2 + \cdots + S_k^2)^{1/2} \ . \qquad (2.4.2)$$

Recommended practice for measurement assurance programs is to combine in quadrature systematic errors that are known to be independent as in (2.4.2), to add linearly systematic errors that may not be independent as in (2.4.1), and to combine systematic and random errors linearly.

2.4.3 Final Statement

Because there is no universal agreement on setting limits to random error, such as two or three standard deviation limits, and also because there is no universal agreement either at NBS or internationally as to how the systematic and random components should be combined, it is recommended that for maximum clarity the composition of the uncertainty statement be fully explained. The explanation should include a statement of the limits to random error, a list of sources of systematic error, and a description of the way in which they have been combined. An example of an uncertainty statement from an NBS calibration process is:

> The apparent mass correction for the nominal 10 gram weight is +0.583mg with an overall uncertainty of ±0.042mg, using three times the standard deviation of the reported value as a limit to the effect of random errors of measurement, the magnitude of systematic errors from all known sources being negligible.

The chain of uncertainty as propagated through a calibration echelon starts with the uncertainty assessed at NBS which consists of all sources of error, both systematic and random, associated with that process including the uncertainty of its reference standards relative to basic units of measurements. If the calibration echelon involves one or more standards laboratories, the total uncertainty as assessed at each echelon becomes a systematic error for the next lower echelon laboratory, and the uncertainties at each level are propagated in like manner. In the next section the propagation of uncertainties for a laboratory that uses an NBS calibrated artifact as a reference standard is compared with the propagation of uncertainties for a laboratory that calibrates its own measuring system through the use of an NBS transfer standard.

2.5 Uncertainty of Reported Values

2.5.1 Uncertainty via Conventional Calibration

The uncertainty associated with a value reported for a test item by a measurement process that is operating in a state of statistical control using

a reference standard calibrated by NBS is

$$U = 3s_r + U_{STD} . \qquad (2.5.1)$$

This assumes that the standard is not changed during transport and that environmental and procedural factors are not different from the conditions of calibration. The standard deviation of the reported value s_r depends on the total standard deviation s_c, the error structure for the reported value as discussed in section 1.5, and the number of measurements made on the test item. The quantity U_{STD} is the uncertainty associated with the reference standard as stated in the NBS calibration report.

Note that where the reported value is an average of p measurements, the usual standard deviation of an average, $s_r/\sqrt{p}$, sometimes called the standard error, will apply to the reported value only if the p repetitions were made over the same set of environmental conditions that were sampled in the calculation of the total standard deviation. In a calibration setting where repetitions are done within a day or two, the standard deviation of a reported value depends upon a between component of variability s_b and a within component s_w as explained in section 1.5.

2.5.2 Uncertainty via a Transfer Standard

Where a laboratory has calibrated its own reference standard using an NBS transfer standard, rather than using a reference standard calibrated at NBS, another echelon has effectively been added to the calibration chain. The uncertainty of that transfer must be assessed, and it contributes another systematic error to the process of subsequently assigning values to test items.

The uncertainty of a transfer involving a single transfer standard compared with a single laboratory standard is

$$U_{tr} = 3s_t + U_T \qquad (2.5.2)$$

and the uncertainty associated with a value reported for a test item is

$$U = 3s_r + 3s_t + U_T = 3s_r + U_{tr} \qquad (2.5.3)$$

where s_r is the standard deviation associated with the reported value of the test item as discussed in the last section; s_t is the standard deviation associated with the value assigned to the laboratory's reference standard via measurements made on the transfer standard; and U_T is the uncertainty assigned to the transfer standard by NBS.

Admittedly there can be some concern about qualifying a laboratory's systematic error by means of an NBS transfer standard because of the additional systematic error that this imposes on the uncertainty statement. This fact is inescapable, but the resulting uncertainty statement is, in fact, a realistic expression of the errors affecting the process whereas the usual calibration route does not provide a way of assessing systematic errors that may be affecting measurements, other than those directly involving the artifact standard.

The uncertainty, U_T, associated with a transfer standard will usually be smaller than U_{STD}, the uncertainty associated with a calibrated artifact. The calibration workload at NBS is at least one step removed from the NBS primary standard, and the size of U_T relative to U_{STD} can be reduced by eliminating this step in assignments to transfer standards. For example, transfer standards for voltage measurements are compared directly to an NBS primary reference bank that is in turn compared on a monthly basis to the Josephson effect, which provides a realization of the volt. The regular calibration workload is compared with a secondary bank of cells that is compared to the primary bank on a daily basis.

Transfer standards that are assigned values at NBS based on secondary standards are calibrated several times over a long time period in order to reduce the contribution of random error to the uncertainty of the assignment. For example, values for gage blocks that comprise the transfer set from NBS are averages of approximately nine electro-mechanical calibrations completed over a two year period. Furthermore, because s_t can be made small by sufficient repetition and careful excution of the transfer, the total uncertainty in (2.5.3) can be kept close to the uncertainty in (2.5.1) or at least small enough to meet the goals of the measurement assurance program. See figure 2 for a graphic comparison of uncertainties via measurement assurance and conventional calibration routes.

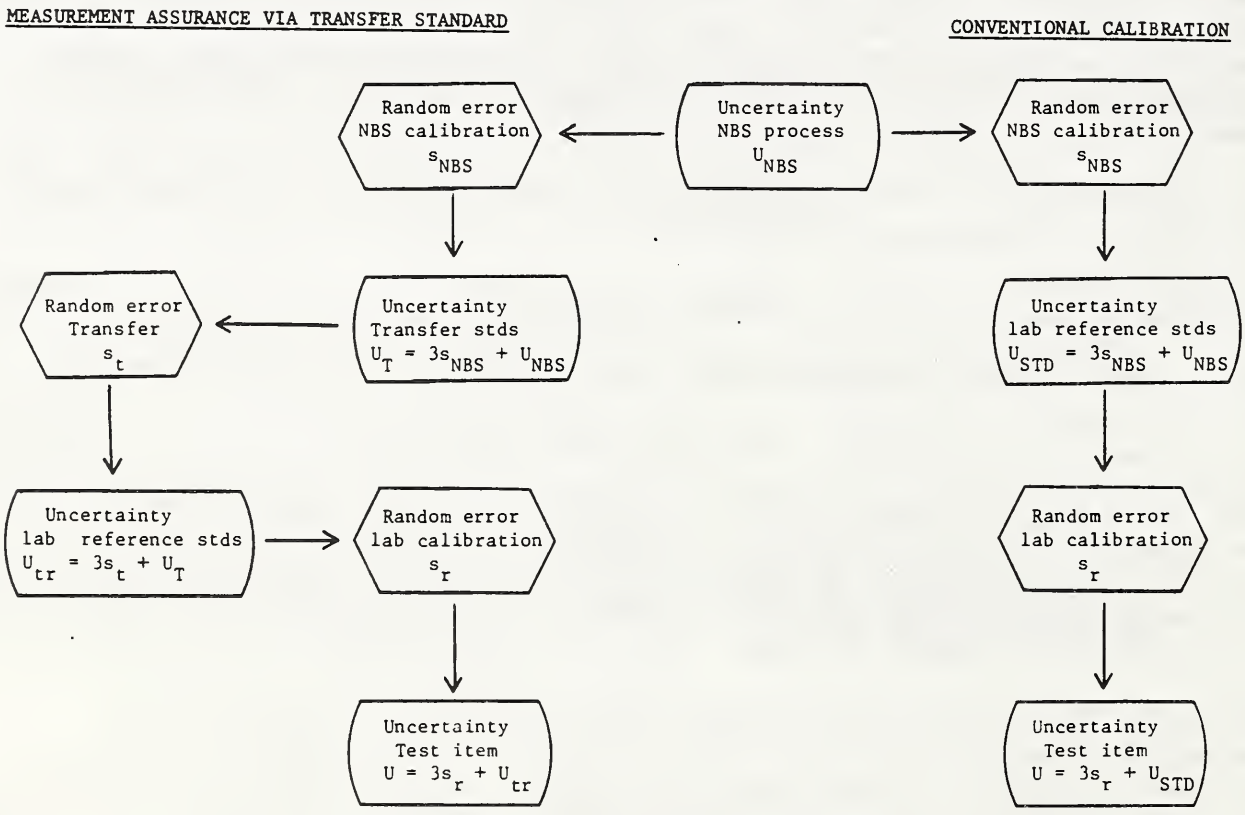


Figure 2
Diagram showing propagation of uncertainties from NBS process to final uncertainty for test item via measurement assurance route compared to the conventional calibration route

2.5.3 Example of an Uncertainty Statement

The principles of this chapter are illustrated by a preliminary experiment at NBS that eventually led to the development of a linewidth standard. Three sources of systematic error were identified in the NBS photometric process that related linewidth measurement to the fundamental definition of length through line-scale interferometry.

The uncertainty from the interferometric process, resulting from random errors associated with making the interferometric determinations and negligible systematic error, translated into a systematic error in the photometric process of 0.01μm. The maximum differences that were observed between the two operators and two instruments that were employed in the NBS system translated into systematic errors of 0.005μm and 0.020μm respectively.

Values assigned to linewidth artifacts were averaged from four photometric readings, and the standard deviation of each assignment was reported as s_r. The limits to random error were taken to be three times the standard deviation of the assignment. An error budget showing the various components contributing to the total uncertainty is shown below.

Components of Uncertainty

Limit to Random Error = $3s_r$		± 0.040μm
Systematic errors:		
a. Operator differences	± 0.005μm	
b. Instrument differences	± 0.020μm	
c. Uncertainty from interferometry	± 0.010μm	
Total systematic errors		± 0.035μm
Total Uncertainty[§]		± 0.075μm

Based on this analysis NBS assigned a total uncertainty of ± 0.075μm to artifacts that were calibrated by this system. If such an artifact were to be used by a laboratory for calibrating its optical imaging system, this uncertainty would become a systematic error for that process.

[§]It is suggested that uncertainties be stated to no more than two significant figures and that the last decimal place in the reported value of the measured item correspond in place value to the last decimal place in the uncertainty statement.

3. The Check Standard in a Measurement Assurance Program

3.1 Introduction

A check standard provides a means of characterizing the behavior of a measurement process by way of repeated measurements on the same artifact, combination of artifacts, or instrument over a substantial period of time and over fluctuating environmental conditions. It should be thought of as a data base of such measurements rather than as an artifact per se because it is the measurements, or some function of those measurements, corrected according to the model specifications, that actually describe process performance.

The structure of the check standard measurement depends on whether the calibration procedure is based on a single measurement or a calibration design. In some cases the check standard may be a function of readings on two reference standards, thus eliminating the need for an additional artifact. Check standard measurements of the following types form the basis for the measurement assurance programs in the next chapter.

1) Measurements made on a single artifact as close in time as possible to the measurements on the reference standard and the test item.

2) Differences between the observed values of two reference standards whose assigned values are the basis for assigning a value to a test item.

3) Computed value for single artifact from a statistical design involving k intercomparisons of reference standards, test items and artifact check standard.

4) Computed value of difference between two reference standards from a statistical design involving k intercomparisons of reference standards and test items.

5) Measurements made on a calibrated artifact by a direct reading instrument.

6) Calibrated value of a single artifact from a calibration process that uses a ratio technique.

3.2 Process Parameters Defined by the Check Standard

Measurement processes have two properties that are critical to a measurement assurance program. Measurements of a stable quantity tend to a long-term average which may not be the same average that would be achieved if a different laboratory produced the measurements. As discussed in detail in the last chapter, these measurements while tending to an average, will not be identical because of inability to reproduce conditions of measurement exactly, and this latter property is referred to as process variability or imprecision. Process parameters are quantities that describe the long-term value and the process precision from redundant measurements on a check standard.

The statistic for characterizing the long-term value is simply the arithmetic average of the check standard measurements and is referred to as the "accepted

value of the check standard." The check standard measurements supplant the ideal set of measurements that could be made on a test item if it were in the laboratory for a sufficiently long period of time. The average of those hypothetical measurements is, of course, the quantity that is of primary interest, but because such is not at our disposal, we define the process in terms of the accepted value of the check standard. This statistic defines a local base for the measurement process which is intimately related to any discrepancy between the reference base and the average of the measurements that could be made on a test item, and any change in the local base is reason to suspect that this systmatic error has changed.

The statistics for characterizing the process precision are: i) a total standard deviation computed from the same check standard measurements and ii) a within standard deviation computed from each calibration design or group of repetitions for cases where the calibration experiment reports a value based on more than a single measurement on a test item. Within standard deviations are pooled according to (2.2.3) into a single value called the "accepted within standard deviation" which reflects variations that typically take place in the measurement process during the course of a calibration.

If the check standard measurements are properly structured, the accepted total standard deviation reflects the totality of variability in the measurement process. The scatter of check standard measurements will be characteristic of measurements of a test item observed over a period of time in the calibration setting only if both types of measurements are affected by the same sources of error. Then the accepted total standard deviation computed from the check standard measurements can be used to compute the standard deviation for a value reported by the calibration process. Evidently, this computation depends on the type of measurements that are designated as check standard measurements and on the model for the calibration process. Specific examples are discussed in chapter 4.

Before embarking on a full-scale measurement assurance program, the participant conducts a series of experiments to establish a data base of check standard measurements. Accepted values for the process parameters are computed from this data base, and it is emphasized that these experiments should cover several weeks' time and should number at least fifteen to obtain reasonable estimates. The calibration schemes or designs for producing the check standard data must be identical to the procedures for calibrating test items in the workload and measuring transfer standards from NBS.

The importance of the initial check standard measurements dictates that they describe the system in its normal operating mode. Care should be exercised to guarantee that this is indeed the case, so that the standard deviation will be appropriate for an uncertainty statement constructed at any time in the future. This is done by varying the conditions of measurement to cover a representative range of laboratory conditions including operator and environmental variations. These measurements should be scrutinized for outliers because even one significant outlier in a small data set can seriously bias the estimates of the process parameters--perhaps causing an out-of-control condition when the transfer standard is being characterized in the laboratory and invalidating the transfer.

Methods for identifying outliers are highly dependent on underlying distributional assumptions. Several methods for detecting outliers are discussed in ASTM Standard E178[f], but for the foregoing reason, they may not be effective given a limited number of check standard measurements. A plot of the data points is usually satisfactory for detecting outliers. Each check standard measurement should be plotted against a time axis, thus creating a preliminary control chart, and measurements which are obviously aberrant should be deleted from the data set. On the other hand, the data should not be edited in order to achieve seemingly better precision because this will cause failures in the control mechanism at a later time. If a large number of points are suspected as outliers, say more than five percent, the check standard measurements do not constitute a strong data base, and the cause of large variations should be investigated and rectified before proceeding with the measurement assurance program.

3.3 The Check Standard in Process Control

Each check standard measurement is subjected to a statistical test for control, and the outcome of that test is used as a mechanism for accepting or rejecting the results of the measurement process. This presupposes that there is, in fact, a process that is in control, that sufficient data from the process exists to quantify this control, and that the behavior of future measurements is predictable from past behavior of the process. This test is exactly analogous to control chart methodology wherein values that fall inside control limits based on historical data are said to be in control, and values that fall outside the control limits are judged out-of-control.

The technique that is used for control is called a t-test wherein a test statistic is computed from the current check standard measurement, the accepted value of the check standard, and the total standard deviation. This test statistic, when large in absolute value compared to a critical value of Student's t distribution, is indicative of lack of control.

The critical value $t_{\alpha/2}(\nu)$ depends on ν, the number of degrees of freedom in the accepted total standard deviation, and on α, the significance level. The significance level α, the probability of mistakenly flagging a check standard measurement as out-of-control, should be chosen by the participant to be suitably small, say between 0.10 and 0.01, so that the number of remeasurements that must be made because of a chance failure is kept at an acceptable level.

Once the control procedure is installed in the laboratory, the assignments generated by the calibration process are accepted as valid within the stated uncertainty as long as the check standard measurements remain in control. Action is required whenever a check standard measurement is out-of-control. The immediate action is to discard the results of the calibration. Of course, at this point one is faced with a dilemna about what future actions should be taken in regard to the calibration process. Because of the probability of chance failure, exactly α, it is reasonable, while discarding the results of the calibration, to repeat the calibration sequence, hoping that check standard measurements will be in control.

[f] ASTM Standard E178 is available from the American Society for Testing Materials, 1916 Race Street, Philadelphia, Pennsylvania 19103.

In this happy event, one assumes that either something was amiss in the initial calibration, such as insufficient warm-up time for the instrument, or that one was the victim of chance failure. In either case it is permissible to accept the more recent result and proceed as usual. In the event of repeated successive failures or numerous failures over time, one must conclude that a major disruption in the calibration process is affecting the process offset, such as a change in a laboratory reference standard, and the calibration process should be shut down until the problem can be rectified and control reestablished. Each calibrration experiment is intended to reveal the offset of a test item or the client's process relative to NBS, and this offset will not be correctly estimated by the calibrating laboratory if the long-term average for its measurements is not constant relative to the reference base. Therefore, a failure of the check standard test implies that offset has not been eliminated or accounted for by the calibration experiment.

A consideration in choosing α is that the significance level for process control should be the same as the significance level for determining the limits of error in section 2.3. Smaller values of α, the probability of having to remeasure unnecessaarily, that is because of chance failure, correspond to larger associated limits of error. Thus the cost of remeasurement must be weighed against the impact of a larger uncertainty. Values of $\alpha = 0.05$ or $\alpha = 0.01$ are recommended.

An alternative to a critical value based on the t-distribution, as explained in section 2.3, is a factor such as three or two which can be used for computing limits to random error and testing for control. The factor three corresponds approximately to $\alpha = 0.003$ for the normal distribution and is well established in quality control applications. There are no hard and fast rules for picking either a significance level α or a factor such as three for the control procedure, but once it is chosen, it plays a large part in determining the frequency of remeasurement and the magnitude of the uncertainty for the process.

The measurement assurance procedures that are outlined in the next chapter are based upon a critical value of three in almost all cases. Those wishing a more stringent control procedure can substitute the appropriate value of $t_{\alpha/2}$ in the appropriate equations. In calibration work, the purpose of the control procedure is to flag those measurements which are clearly out-of-control, and a critical value of three is suitable for many situations. This approach is the current practice of many calibration services at NBS. Moreover, limits based on the factor three work well, covering a large percentage of the distribution of possible values of the test statistic, even where the test statistic is not strictly distributed as Student's t which is the case for some of the more complicated constructions in the next chapter.

If the measurement sequence allows for a within standard deviation, the ratio of this within standard deviation to the accepted within standard deviation is compared to a critical value based on Snedecor's F distribution (see Ku [44] for a discussion of the F test). A ratio that is large compared to the critical value is indicative of lack of control during the course of the measurement sequence.

The critical value $F_\alpha(\nu_1,\nu_2)$ depends on; ν_1, the number of degrees of freedom in the current within standard deviation; ν_2 the number of degrees of freedom in the accepted within standard deviation; and α, the significance level discussed in preceding paragraphs. Critical values of $F_\alpha(\nu_1,\nu_2)$ are tabulated in Table II for α=0.01, ν_1=1(1)10(2)30(10)120 and ν_2=10(1)20(2)30(5)120.†

The t-test and F test are invoked simultaneously--the failure of either test constituting grounds for discarding the measurement on the test item or transfer standard. The combination of these two tests is a powerful means of detecting shifts in the long-term average of the process as it defines systematic error.

The efficacy of the check standard as a device for guaranteeing that the process is functioning properly and that, therefore, the test items are assigned values with negligible offset relative to NBS, depends on the relationship among the measurements made on the test items, the measurements made on the reference standards and the measurements made on the check standards. The strongest case for measurement assurance exists when all assignments are statistically interrelated as in a statistical design. When the assignments are by nature statistically independent, it is essential that the measurements be temporally related by completing the measurement sequence in as short a time as possible.

There is really no guarantee that a predictable response on the part of the check standard assures a good measurement on the test item if it is possible for the process to change appreciably during the intervening time between the check standard measurement and the other measurements. However, a strong case for confidence in a measurement process exists for a process that is continuously in control. Furthermore, out-of-control findings for the check standard are almost unfailingly indicators of measurement problems because the control limits are specified so that the probability of a single value being out-of-control is extremely small.

The question of how often the process should be checked for control can only be answered in terms of the goals of the measurement program. A criterion based on economic considerations must balance the tradeoff between the cost of making additional measurements to ensure accuracy and the costs incurred when inaccurate measurements are allowed to occur. In order to achieve the highest level of measurement assurance, check standard measurements should be incorporated in every calibration sequence. When this is not possible or not necessary, a check for control should be incorporated in start-up procedures and repeated at intervals thereafter that depend on the level of system control that is desired and on past experiences with the control procedure.

A system that is always in control when checked can be presumed to remain in control between checks, and the time between check standard measurements can be lengthened. Conversely, the same presumption cannot be made for a system that is occasionally out-of-control, and the time between check standard measurements should be shortened if one is to determine how long the system can operate in-control.

†The notation 10(2)30, for example, indicates that the values go in steps of two from ten to thirty.

3.4 The Transfer with NBS

During the transfer between the participating laboratory and NBS, current check standard measurements that result from the transfer experiments are compared with the accepted value of the check standard by a t-test in order to ascertain whether or not there has been a significant change in the long-term average of the process. If the check standard measurements are continually out-of-control while the transfer standard is in the laboratory, the transfer measurements are invalid, and the transfer experiment should be discontinued until the initial check standard measurements are repeated and new accepted values are established. Isolated failures can be treated as they are treated in the calibration workload, and offending measurements that cannot be repeated are deleted from the transfer data.

Similarly, the within standard deviation computed from the transfer measurements is compared with the accepted within standard deviation by an F-test. If possible, sufficient repetitions spaced over a period of time are also included in the procedures for measuring the transfer standards so that the standard deviation for the transfer can be compared to the accepted total standard deviation.

After the completion of the transfer with NBS, the tests for control are continued for the calibration process. When an out-of-control condition is encountered in this mode, the measurement process is discontinued until control is restored which may amount to simply repeating the measurement sequence on the test item and check standard. When it is obvious that the process mean has shifted because of repeated out-of-control findings for the check standard, signifying that the offset from NBS has changed, it is time for another intercomparison with NBS. Theoretically one may be able to analyze the amount of change in the offset, but it seems judicious at this point to reestablish the values of the laboratory's reference standards.

3.5 Updating Process Parameters

After the control procedure has been in place for a year or more, sufficient data should be available so that the process parameters can be updated. The mechanics for doing this depend on the degree of automation that exists in the laboratory and on the computing capability at its disposal. In a sophisticated program one compares the accepted value for the check standard and the accepted total standard deviation with values computed from the check standard data that has been accumulated since the last update. If the two sets of data are essentially in agreement, updated process parameters are computed based on all check standard measurements. In cases where the process has changed significantly in regard to these parameters, the past historical data are discarded, and new process parameters are computed from the most recent data. For computer systems such as micro-computers with limited storage capacity, it may be feasible to retain only a fixed number of check standard measurements. Obviously the number should be sufficient for obtaining reliable estimates. The data file is continually updated by deleting the oldest measurement and adding the newest-thereby always keeping a fixed number of check standard measurements in the data file with which to compute the process parameters.

4. Implementation of Measurement Assurance for Specific Cases

This chapter contains the basic outlines for implementing measurement assurance programs for eight specific measurement situations where the sequence of measurements that constitute an intercomparison depends upon the number of reference standards, the number of test items and the number of redundant measurements to be employed in each intercomparison.

The essential elements that specify the measurement situation for each plan are as follows:

4.1 A comparator process in which one reference standard is compared to a test item and a check standard.

4.2 A comparator process in which a test item is compared to each of two reference standards, and control is maintained on the difference between readings on the two reference standards.

4.3 A comparator process in which three test items are compared to two reference standards in a statistical design, and control is maintained on the difference between the two standards.

4.4 A comparator process for mass calibrations illustrating the use of a 1, 1, 1 design and a 5, 3, 2, 1, 1, 1 design with provision for a check standard for each series.

4.5 A comparator process in which four test items are compared to four reference standards, without direct intercomparison between the test items or reference standards. Control is maintained on the difference between two reference standards.

4.6 Direct reading of the test item with the instrument as the standard. Control is maintained by repetitions on a calibrated artifact.

4.7 Simultaneous measurement of a group of test items relative to a bank of reference standards where a check standard is always included among the test items.

4.8 A ratio technique for one or more test items and one or two reference standards. Control is maintained on calibrated values of an artifact check standard.

Calibration as a process of intercomparing a test item with a reference standard and assigning a value to the test item based on the accepted value of the standard is frequently carried out by a comparator process. For high precision work, the comparator process makes use of an instrument or device which is capable of handling only very small differences between properties of similar objects such as a mechanical comparator for comparing gage blocks of the same nominal length or an electrical bridge for detecting very small differences between resistances. Where individual readings, in scale units, are taken on the unknown and the reference standards and converted to the appropriate units, a value can be assigned to the test item only through the

difference between the reading on the test item and the reading on the reference standard (See section 1.4.2). The calculated difference between the two readings is the "measurement of interest" and the number of such differences determines the redundancy in a measurement scheme.

Where the calibration experiment produces only a difference measurement, such as the difference in emf between two saturated cells as measured by a potentiometer, the term "reading on an unknown" or "reading on a standard" does not have a literal interpretation but refers to the logical intercomparison of the items. In either case, a value is assigned to an unknown relative to the known value of one or more reference standards. This known value is referred to as the restraint.

Where there are a small number of unknowns and reference standards, the calibration experiment may consist of all possible intercomparisons that can be made on the collection of items; this would amount to $k(k-1)/2$ difference measurements for k items being intercompared two at a time. A calibration design consists of a subset of all possible intercomparisons such that, given a restraint or assigned value for the reference standards, the series of intercomparisons can be solved for the unknowns. The method for finding a solution is least-squares, and the resulting values for the unknown items are least-square estimates.

Several factors dictate the choice of intercomparisons that constitute the design. Obviously, it is desirable to keep the number of intercomparisons small. Designs are usually structured so that precision in the assignments to the test items is the same for all items of the same nominal size and so that precision in this sense is optimized for a given number of intercomparisons. Other optimality criteria that are discussed in the statistical literature in references [45] and [46] may be of interest to the reader.

Calibration can also be carried out using a direct reading device or instrument in which case the device is regarded as the standard, and values, already in the appropriate units, are assigned directly to the test items. Such a device, for example an interferometer, can also be used in a comparator mode in which case the difference between a reading on the test item and a reading on the standard is regarded as the measurement of interest.

The eight measurement plans that are discussed in this section have been adapted to both mechanical and electrical measurements. Plan 4.1 is the simplest scheme for a comparator process and may be appropriate when accuracy requirements are moderate. It does not afford a high degree of protection because the linkage between the measurement on the test item and the measurement on the check standard is not as strong as it is for the other comparator schemes. Plan 4.2 affords a higher degree of protection against incorrect measurements by requiring redundant measurements on each test item. This plan is well suited to mechanical measurements and is currently utilized in the Gage Block Measurement Assurance Program. The program is illustrated with data from one participant in section 4.2.7.

Plans 4.3 and 4.5 involve calibration designs that are particularly appropriate for voltage and resistance measurements. The designs have a provision for estimating a so-called left-right effect which is an important

circuit parameter for voltage measurements. The discussion of plan 4.5, which is illustrated with data from the NBS Volt Transfer Program, explains the steps to be followed in process control using a check standard that is either stable or is drifting linearly with time.

Plan 4.4 describes a measurement assurance program for guaranteeing the accuracy of very precise weighings by means of two designs which are routinely used in the NBS mass calibration program. Weighing designs for different combinations of weights along with designs for mechanical and electrical measurements involving more standards and test items are described by Cameron et al [47]. Designs for eliminating temporal effects are described by Cameron and Hailes [48].

Surveillance testing as a means of ensuring the self-consistency of a weight set is described in detail in a recent publication by Jaeger and Davis [49]. The basic idea is to compare a given weight against a collection of other weights in the set whose nominal sum equals the first weight. The authors develop measurement assurance methods for monitoring the difference calculated from the comparison and resolving it with values assigned to the individual weights.

Plan 4.6 is probably the simplest and involves only direct readings on the test items. It is appropriate for large volume workloads that utilize an instrument standard such as interferometer, digital voltmeter, or electronic balance where there is a need to monitor or guarantee the accuracy of the instrument as a matter of course.

Plan 4.7 is appropriate for assigning values to test items or instruments relative to a bank of standards where the calibration consists of subjecting all items including the reference standards to the same stimuli, usually simultaneously. Control is maintained by a check standard which is included as a test item in each measurement sequence. Applications include watthour meter calibration where test meters and reference meters are connected to the same power source and very low pressure calibration where several pressure gages are confined in a vacuum chamber with a reference pressure gage.

By necessity, the analyses are outlined in a straightforward manner, and problems involving drifting reference standards or check standards must be considered separately. It is obviously impossible to anticipate the spectrum of complications that may arise in a given measurement area, and these analyses, offered as a simplistic approach to sometimes difficult problems, are intended to provide a starting point for measurement assurance.

Each measurement assurance program that is presented in this chapter relies upon a check standard concept as discussed at length in the last chapter, and the check standard measurements are crucial to the steps that constitute such a program; namely i) establishment of process parameters; ii) routine process control; iii) evaluation of systematic error by transfer with NBS; iv) determination of uncertainty for test items; v) update of process parameters.

The first four steps are outlined in detail for each program, and the fifth step relating to updating and maintaining the data base was discussed in generality in section 3.5.

4.1 Comparator Process for One Test Item, One Reference Standard, and One Check Standard

4.1.1 Measurement Sequence

This scheme is appropriate for a comparator process where the intercomparison of the test item X with the reference standard R is immediately followed by the intercomparison of an artifact check standard Y with the reference standard R in the sequence X, R, Y, R. The readings are denoted by x, r_1, y, r_2 respectively. This measurement sequence should be followed for all calibrations for which statistical control is to be achieved. The value of the check standard for one such sequence is defined from the reading on the artifact check standard and the duplicate readings on the reference standard as

$$c = y - \frac{1}{2}(r_1 + r_2) . \qquad (4.1.1)$$

All aspects of a measurement assurance program involving this design are explained and illustrated for gage blocks in reference [50].

4.1.2 Process Parameters

Initial values of the process parameters are obtained from n such measurement sequences, where $c_1,\cdots,c_n$ are the observed values of the check standard. The accepted value of the check standard is the mean of the check standard measurements; namely,

$$A_c = \frac{1}{n}\sum_{i=1}^{n} c_i . \qquad (4.1.2)$$

The accepted total standard deviation for the check standard is

$$s_c = \left(\frac{1}{n-1}\sum_{i=1}^{n} (c_i - A_c)^2 \right)^{1/2} \qquad (4.1.3)$$

with $\nu = n-1$ degrees of freedom.

The model assumed for the calibration process is the additive model (1.4.2). Under this model the error structure for the value of the test item and the error structure for the check standard measurement are identical. Thus s_c also estimates the standard deviation of the reported value of the test item which is shown in (4.1.6).

The control limits[h] that are appropriate for future check standard observations are given by

$$\text{Upper control limit} = A_c + 3s_c$$

$$\text{Lower control limit} = A_c - 3s_c \ .$$

4.1.3 Control Procedure

The control procedure applied to each calibration depends on a test statistic t_c that is computed from the observed value of the check standard c for that measurement sequence by

$$t_c = \frac{|c - A_c|}{s_c} \ . \qquad (4.1.4)$$

If

$$t_c < 3 \qquad (4.1.5)$$

the process is in control, and the value of the test item is reported as

$$X^* = x - \frac{1}{2}(r_1 + r_2) + R^* \qquad (4.1.6)$$

where R^* is the value assigned to the reference standard.

If

$$t_c \geq 3,$$

the calibration of the test item is invalid and must be repeated.

[h]The factor 3 is used in this and all subsequent computations in place of the appropriate percent point of the t distribution; $t_{\alpha/2}(\nu)$.

4.1.4 Transfer with NBS

The transfer with NBS is accomplished by p repetitions of the measurement sequence in which a transfer standard takes the place of the test item in each repetition. Process control as defined by (4.1.5) should be confirmed for each repetition. Any sequence that is out-of-control should be repeated until control is reestablished or else that repetition is deleted from the transfer. If the value assigned to the transfer standard by NBS is T^* with uncertainity U_T, the uncertainty of the transfer is

$$U_{tr} = \frac{3s_c}{\sqrt{p}} + U_T \ . \tag{4.1.7}$$

The offset Δ of the laboratory process from NBS is

$$\Delta = \frac{1}{p} \sum_{j=1}^{p} X_j^* - T^* \tag{4.1.8}$$

where $X_1^*, \cdots, X_p^*$ are values calculated according to (4.1.6) for the transfer standard for each of the p repetitions.

This offset is judged significant if

$$\frac{\sqrt{p}\,|\Delta|}{s_c} \geq 3 \ , \tag{4.1.9}$$

and in such case the assigned value of the reference standard becomes $R^* - \Delta$.

The assigned value of the reference standard is unchanged if

$$\frac{\sqrt{p}\,|\Delta|}{s_c} < 3 \ .$$

4.1.5 Total Uncertainty

The total uncertainty that is appropriate for a value assigned to a test item by <u>one</u> calibration sequence is

$$U = U_{tr} + 3s_c. \tag{4.1.10}$$

4.2 Comparator Process for One Test Item and Two Reference Standards

4.2.1 Measurement Sequence

This scheme involving duplicate measurements on the test item is appropriate for a comparator process where the assignment for the test item is made relative to the average of the values assigned to the two reference standards, called the restraint R^*. The intercomparison of the test item X with each of two reference standards, R_1 and R_2, in a trend eliminating design (Croarkin et al [51]) is accomplished by the sequence X, R_1, R_2, X, and the readings are denoted by x_1, r_1 r_2, x_2 respectively. The difference measurements are:

$$d_1 = x_1 - r_1$$

$$d_2 = x_2 - r_2$$

There is no artifact check standard for this design, and a check standard value is defined for each sequence as the calculated difference between the readings on the two reference standards as

$$c = d_2 - d_1 \tag{4.2.1}$$

The value c is structured so as to reflect the maximum variation that occurs in the measurement sequence between the first and the last readings on the test item and not just the variation that occurs between the readings on the two reference standards.

4.2.2 Process Parameters

Initial values of the process parameters are obtained from n such measurement sequences yielding check standard values $c_1, \cdots, c_n$. The accepted value of the check standard is given by the mean of the check standard values; namely,

$$A_c = \frac{1}{n} \sum_{i=1}^{n} c_i \ . \tag{4.2.2}$$

The total standard deviation of the check standard is defined by

$$s_c = \left(\frac{1}{n-1} \sum_{i=1}^{n} (c_i - A_c)^2 \right)^{1/2} \tag{4.2.3}$$

with $\nu = n-1$ degrees of freedom.

The control limits[i] that are appropriate for future observations on the check standard are given by

$$\text{Upper control limit} = A_c + 3s_c$$

$$\text{Lower control limit} = A_c - 3s_c \ .$$

[i] The factor 3 is used in this and all subsequent computations in place of the appropriate percent point of the t distribution; namely, $t_{\alpha/2}(\nu)$.

The model assumed for the process is the additive model (1.4.2). The error structures for the check standard measurement and the reported value of the test item are worked out in detail in section 1.5 where it is shown that the standard deviation for the reported value of the test item is $s_c/2$.

4.2.3 Control Procedure

The control procedure applied to each calibration depends on a statistic t_c that is computed from the observed value of the check standard c for that measurement sequence where

$$t_c = \frac{|c - A_c|}{s_c} . \qquad (4.2.4)$$

If

$$t_c < 3 \qquad (4.2.5)$$

the process is in control, and the value of the test item is reported as

$$X^* = \frac{1}{2}(d_1 + d_2) + R^* \qquad (4.2.6)$$

where the restraint $R^* = \frac{1}{2}(R_1^* + R_2^*)$, and R_1^* and R_2^* are the assigned values of the reference standards.

If

$$t_c \geq 3,$$

the calibration of the test item is invalid and must be repeated.

4.2.4 Transfer with NBS

The transfer with NBS can be accomplished with two tranfer standards T_1 and T_2. In this mode p_1 repetitions of the measurement sequence are made with T_1 taking the place of the test item and p_2 repetitions of the measurement sequence are made with T_2 taking the place of the test item. This produces a total of $p_1 + p_2$ repetitions for the transfer. Process control as defined by (4.2.5) should be confirmed for each repetition. Any sequence that is out-of-control should be repeated until control is reestablished or else that repetition is deleted from the transfer. If the values assigned to the transfer standards by NBS are T_1^* and T_2^* with uncertainties U_{T1} and U_{T2} respectively, the uncertainty of the transfer is

$$U_{tr} = \frac{3}{4}\left(\frac{p_1+p_2}{p_1 \cdot p_2}\right)^{1/2} s_c + U_T \qquad (4.2.7)$$

where

$$U_T = \frac{1}{2}\left(U_{T1}^{\ 2} + U_{T2}^{\ 2}\right)^{1/2} .$$

The offset Δ of the laboratory process from NBS is defined only in terms of the restraint; i.e., the average of the two reference standards. It is computed from the p_1 values assigned to the first transfer standard according to (4.2.6); namely, $X_1^*,\cdots,X_{p_1}^*$ and the p_2 values assigned to the second transfer standard according to (4.2.6); namely, $X_1^{**},\cdots,X_{p_2}^{**}$.

$$\Delta = \frac{1}{2p_1}\sum_{i=1}^{p_1} X_i^* + \frac{1}{2p_2}\sum_{i=1}^{p_2} X_i^{**} - \frac{1}{2}(T_1^* + T_2^*) \qquad (4.2.8)$$

The offset is judged significant if

$$\tilde{t} > 3, \qquad (4.2.9)$$

where

$$\tilde{t} = \frac{4\sqrt{p_1 \cdot p_2}\,|\Delta|}{\sqrt{p_1+p_2}\,s_c} \qquad (4.2.10)$$

and in such case the assigned value of the restraint is changed to $R^* - \Delta$. The restraint is unchanged if $\tilde{t} < 3$.

4.2.5 Uncertainty

The total uncertainty that is appropriate for a value assigned to a test item by (4.2.6) from one calibration sequence is

$$U = U_{tr} + \frac{3s_c}{2}. \qquad (4.2.11)$$

4.2.6 Example from the Gage Block Measurement Assurance Program

Two sets of eighty-one gage blocks from NBS were sent to industrial participants for the purpose of assigning values to their laboratory reference standards. Before the transfer blocks left NBS, each participant conducted a minimum of six experiments in which his two sets of reference standards were compared to a set of test blocks according to the measurement scheme in section 4.2.1. Because six measurements are not sufficient for estimating a standard deviation, the data were analyzed by groups, with about twenty blocks constituting a group.

In order to check a large data set for outliers, such as the data accumulated on the gage block check standards, it is sometimes possible to make use of the information in the individual standard deviations. Because the measurements are assumed to all come from the same process, a standard deviation that is large compared to the other standard deviations in the group suggests an outlier in the check standard measurements for that nominal size.

If there are k block sizes in a group, the test statistic is the ratio of a single standard deviation s_i to a quantity that has been pooled from the remaining standard deviations in that group; namely, s_j ($j=1,\cdots,k$; $j\neq i$). The test statistic is

$$F = \left(s_i/s_{p_i}\right)^2$$

where

$$s_{p_i} = \left(\frac{1}{k-1}\sum_{j\neq i} s_j^{\,2}\right)^{1/2}$$

and s_i has ν_1 degrees of freedom and each pooled standard deviation has ν_2 degrees of freedom. If all s_i have the same number of degrees of freedom ν, then $\nu_1 = \nu$ and $\nu_2 = (k-1)\cdot\nu$. If for α chosen suitable small,

$$F > F_\alpha(\nu_1,\nu_2)$$

where $F_\alpha(\nu_1,\nu_2)$ is the upper α percent point of the F distribution with ν_1 and ν_2 degrees of freedom, the standard deviation in question is considered significant, and the individual measurements for that check standard are inspected for an outlier--the outlier being either the largest or the smallest measurement.

Consider the standard deviations in exhibit 4.2.1 which were computed from check standard measurements for nine nominal sizes. The individual measurements are plotted in figure 3 as deviations from the mean for each nominal size as a function of nominal size. Test statistics computed for each nominal size show that the standard deviation for the 0.122000 inch check standard is significantly larger than the others, and figure 3 verifies that the smallest observation is not consistent with the other data for that size and is thus labeled an "outlier."

Exhibit 4.2.1 - Standard deviations from check standard measurements
Values in microinches

Nominal Length (Inches)	Std Devs	Degrees of Freedom	Pooled Std Devs	Degrees of Freedom	Test Statistic
	s_i	ν_1	s_{p_i}	ν_2	F
0.117000	0.445	5	0.723	40	0.38
0.118000	0.288	5	0.733	40	0.15
0.119000	0.952	5	0.659	40	2.09
0.120000	0.382	5	0.727	40	0.28
0.121000	0.616	5	0.707	40	0.76
0.122000	1.303	5	0.579	40	5.06†
0.123000	0.539	5	0.715	40	0.57
0.124000	0.674	5	0.700	40	0.93
0.125000	0.472	5	0.721	40	0.43

† $\left(s_i/s_{p_i}\right)^2 > F_{.01}(5,40)$ where $F_{.01}(5,40) = 3.51$ from Table II.

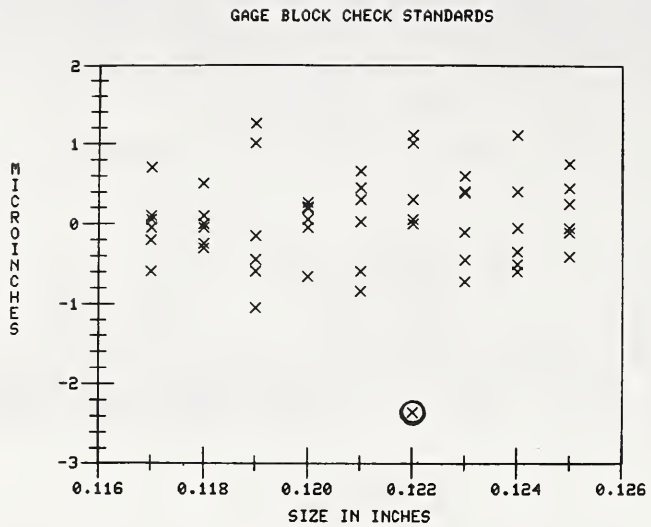


Figure 3
Deviations (microinches) from the mean versus nominal length (inches) for groups of six check standard measurements showing a single outlier

The initial data taken by the participants in the measurement assurance program were inspected for outliers by this method. All outliers were deleted from the data base before calculating the accepted values and standard deviations of the check standard measurements. A subset of the data for one participant is featured in exhibit 4.2.3 with the number of blocks being restricted to five for the purpose of the illustration. The exhibit shows the data from the initial experiments, with a check standard for each repetition computed according to (4.2.1) and initial values for the process parameters A_c and s_c computed using (4.2.2) and (4.2.3) respectively. After the initial data set was edited for outliers, the transfer blocks were sent to the participant. The values assigned to the transfer standards by NBS and the value for the participant's restraint are listed in exhibit 4.2.2.

Exhibit 4.2.2 - Participant's restraint and NBS values for transfer standards
Values in microinches

Nominal	Restraint	Transfer Stds		Uncertainties		Total§
	R^*	T_1^*	T_2^*	U_{T1}	U_{T2}	U_T
0.1006	1.30	-0.63	-0.56	2.17	2.06	2.12
0.1008	0.80	3.21	3.14	2.17	2.06	2.12
0.1010	2.65	2.33	2.52	2.17	2.06	2.12
0.1020	0.45	0.35	0.19	2.17	2.06	2.12
0.1030	-0.05	-2.09	-2.32	2.17	2.06	2.12

§ The systematic errors associated with the transfer standards are added linearly instead of in quadrature because the assignments T_1^* and T_2^* are not independent. Thus $U_T = (U_{T1} + U_{T2})/2$.

Exhibit 4.2.3 - Readings on unknown X and reference standards R_1 and R_2
Corrections to nominal size in microinches

Nominal (inches)	Reps	Readings				Check Standard	Mean	Total S.D.	D.F.
		x_1	r_1	r_2	x_2	c	A_c	s_c	ν
	1	53.9	52.7	46.8	53.9	5.9			
	2	54.8	49.9	45.0	54.5	4.6			
	3	56.0	50.0	44.1	56.1	6.0			
0.1006	4	56.3	50.0	44.1	56.3	5.9	5.80	0.616	5
	5	55.1	49.7	43.7	55.1	6.0			
	6	55.0	50.0	43.9	55.3	6.4			
	1	51.1	51.1	49.2	50.5	1.3			
	2	53.0	49.9	47.9	53.1	2.1			
	3	54.2	50.1	47.5	54.3	2.7			
0.1008	4	54.4	50.2	47.5	54.3	2.6	2.33	0.554	5
	5	53.2	49.9	47.2	53.2	2.7			
	6	53.3	50.0	47.3	53.2	2.6			
	1	52.0	50.1	49.1	52.5	1.5			
	2	54.8	48.8	47.7	54.7	1.0			
	3	55.5	50.0	48.4	55.5	1.6			
0.1010	4	55.4	50.0	48.3	55.4	1.7	1.70	0.593	5
	5	55.5	51.0	48.2	55.5	2.8			
	6	55.8	50.0	48.2	55.6	1.6			
	1	52.1	50.1	48.1	52.2	2.1			
	2	57.3	51.1	49.0	57.2	2.3			
	3	57.0	50.0	48.3	57.0	1.7			
0.1020	4	57.2	50.1	48.4	57.1	1.6	2.07	0.339	5
	5	55.3	50.0	47.6	55.3	2.4			
	6	55.1	49.9	47.6	55.1	2.3			
	1	53.9	49.2	48.5	54.1	0.9			
	2	58.8	50.0	49.0	58.8	1.0			
	3	59.4	50.0	49.1	59.5	1.0			
0.1030	4	59.4	50.0	49.1	59.4	0.9	0.73	0.361	5
	5	59.3	50.0	49.5	58.9	0.1			
	6	59.7	50.2	49.6	59.6	0.5			
							Pooled	0.507	25

Each transfer block was intercompared twice with the participants reference standards by the same scheme used to obtain the initial data, resulting in a total of $p_1 + p_2 = 4$ repetitions. The data for each repetition are shown in exhibit 4.2.4. The readings on the reference standards are designated by r_1 and r_2, and the duplicate readings on a transfer standard are designated by x_1 and x_2. The exhibit also lists the check standard that was computed for each repetition, the test statistic t_c, and the value reported for the NBS transfer standard according to (4.2.6).

Notice that on three occasions the check standard measurement failed the test for control defined by (4.2.5). Because the data were analyzed at NBS after the transfer standards left the participant's laboratory, it was not possible to repeat those sequences, and they were deleted from the transfer data thereby reducing the number of valid repetitions for those block sizes.

Exhibit 4.2.4 - Readings on transfer standards T_1 and T_2
Corrections to nominal size in microinches

Nominal (inches)	Stds	Reps	Readings				Check Std	Test Statistic	Transfer Std
			x_1	r_1	r_2	x_2	c	t_c	X*
	T_1	1	51.2	55.2	48.0	50.8	6.8	2.0	0.70
	T_1	2	51.3	55.2	48.9	51.2	6.2	0.8	0.50
0.1006	T_2	1	50.8	54.9	48.1	51.3	7.3	3.0§	----
	T_2	2	51.2	55.2	48.9	51.3	6.4	1.2	0.50
	T_1	1	56.5	55.3	52.4	56.3	2.7	0.7	3.35
	T_1	2	56.2	55.1	52.5	56.2	2.6	0.5	3.20
0.1008	T_2	1	56.1	55.1	52.3	56.4	3.1	1.5	3.35
	T_2	2	55.7	55.0	52.5	55.8	2.6	0.5	2.80
	T_1	1	54.0	54.9	53.2	54.0	1.7	0	2.60
	T_1	2	53.5	55.0	52.8	53.5	2.2	1.0	2.25
0.1010	T_2	1	53.9	54.9	53.3	53.9	1.6	0.2	2.45
	T_2	2	53.8	55.0	52.8	53.9	2.3	1.2	2.60
	T_1	1	54.9	54.3	52.1	54.7	2.0	0.1	2.05
	T_1	2	55.0	55.1	52.5	55.0	2.6	1.0	1.65
0.1020	T_2	1	54.6	54.3	52.1	54.6	2.2	0.3	1.85
	T_2	2	55.2	55.1	52.5	55.2	2.6	1.0	1.85
	T_1	1	52.9	53.9	52.9	52.8	0.9	0.3	-0.60
	T_1	2	53.9	54.9	52.4	53.9	2.5	3.5§	----
0.1030	T_2	1	52.4	53.9	52.9	52.5	1.1	0.7	-1.00
	T_2	2	53.4	54.9	52.4	53.4	2.5	3.5§	----

§Failed the test for control. The Gage Block Measurement Assuance Program uses a critical value of 3.

Offsets from NBS were computed for each block size by (4.2.8) and were tested for significance by (4.2.9). The participant was advised to change the value of the restraint for those block sizes which showed a significant offset from NBS. The uncertainty of the current transfer with NBS was computed. Results are reported in exhibit 4.2.5. The participant was further advised that the uncertainty appropriate for his process was U = 3.42 microinches as calculated by (4.2.11).

This uncertainty is valid for calibrations conducted according to the measurement scheme in Section 4.1.1 with the value of the restraint as stipulated as long as the process remains in control. Another transfer with NBS will be scheduled in two years to check on the state of the measurement assurance program, and it is anticipated that thereafter transfers with NBS will become increasingly rare. Specific blocks that shows signs of change can be recalibrated or replaced in the interim.

Exhibit 4.2.5 - Offsets from NBS and corrected restraints
Values in microinches

Nominal (inches)	Number of Repetitions	Offset	Test Statistic	Corrected Restraint	Uncertainty of Transfer
	$p_1 + p_2$	Δ	$\tilde{t}$	$R^* - \Delta$	U_{tr}
0.1006	3	1.14	7.3†	0.16	2.59
0.1008	4	0.00	0.0	0.80§	2.50
0.1010	4	0.05	0.4	2.65§	2.50
0.1020	4	1.58	12.5†	-1.13	2.50
0.1030	2	1.40	7.8†	-1.45	2.66

†The test statistic $\tilde{t} > 3$ indicating that the offset from NBS is significant and that the laboratory restraint should be decreased by the amount Δ.

§The restraint is unchanged because the offset is not significant.

4.3 Comparator Process for Three Test Items and Two Reference Standards

4.3.1 Measurement Sequence

In this scheme, which is particularly suitable for electrical measurements, the small difference between two items, such as the difference between the electromotive forces for two saturated cells, constitutes a measurement. The assignments of values to test items are done relative to two reference standards. The statistical design leads not only to equal precision in the assigned value for each test item, but it is also structured so that any position effect in the electrical connection, called left-right effect, is cancelled (Cameron & Eicke [52]). The theory of least-squares estimation which governs the solution of this type of design is explained by Cameron in reference [53].

The design is composed of a subset of all possible difference measurements that could be made on the two standards and three test items. The total number of measurements that could be made in order to achieve left-right balance on such a complement of standards and test items is twenty, and the design is parsimonious in that it requires a subset of ten of the possible measurements while still achieving equal precision for each assignment.

The reference standards are designated by R_1 and R_2, the test items by X, Y, and Z and the corresponding intercomparisons on each by r_1, r_2, x, y, z respectively. The order of measurements is given below:

$$\begin{aligned} d_1 &= r_1 - r_2 \\ d_2 &= r_2 - x \\ d_3 &= x - y \\ d_4 &= y - z \\ d_5 &= z - r_1 \\ d_6 &= y - r_1 \\ d_7 &= r_2 - y \\ d_8 &= z - r_2 \\ d_9 &= x - z \\ d_{10} &= r_1 - x \end{aligned} \qquad (4.3.1)$$

The left-right effect is estimated by

$$\hat{\zeta} = \frac{1}{10} \sum_{i=1}^{10} d_i \, . \qquad (4.3.2)$$

The differences of the reference standards from their average as estimated by least-squares are:

$$\hat{r}_1 = \frac{1}{10}(2d_1 - d_2 - d_5 - d_6 - d_7 + d_8 + d_{10})$$

$$\hat{r}_2 = \frac{1}{10}(-2d_1 + d_2 + d_5 + d_6 + d_7 - d_8 - d_{10})$$

and the corresponding differences for the test items are:

$$\begin{aligned}
\hat{x} &= \frac{1}{10}(-3d_2 + 2d_3 + d_5 + d_6 - d_7 + d_8 + 2d_9 - 3d_{10}) \\
\hat{y} &= \frac{1}{10}(-d_2 - 2d_3 + 2d_4 + d_5 + 3d_6 - 3d_7 + d_8 - d_{10}) \\
\hat{z} &= \frac{1}{10}(-d_2 - 2d_4 + 3d_5 + d_6 - d_7 + 3d_8 - 2d_9 - d_{10}).
\end{aligned} \tag{4.3.3}$$

The within standard deviation for each design is

$$s_w = \left(\frac{1}{\nu} \sum_{i=1}^{10} \xi_i^2 \right)^{1/2} \tag{4.3.4}$$

with degrees of freedom $\nu = 5$.

The individual deviations ξ_i from the least-squares fit are defined by:

$$\begin{aligned}
\xi_1 &= d_1 - \hat{r}_1 + \hat{r}_2 - \hat{\zeta} \\
\xi_2 &= d_2 - \hat{r}_2 + \hat{x} - \hat{\zeta} \\
\xi_3 &= d_3 - \hat{x} + \hat{y} - \hat{\zeta} \\
\xi_4 &= d_4 - \hat{y} + \hat{z} - \hat{\zeta} \\
\xi_5 &= d_5 - \hat{z} + \hat{r}_1 - \hat{\zeta} \\
\xi_6 &= d_6 - \hat{y} + \hat{r}_1 - \hat{\zeta} \\
\xi_7 &= d_7 - \hat{r}_2 + \hat{y} - \hat{\zeta} \\
\xi_8 &= d_8 - \hat{z} + \hat{r}_2 - \hat{\zeta} \\
\xi_9 &= d_9 - \hat{x} + \hat{z} - \hat{\zeta} \\
\xi_{10} &= d_{10} - \hat{r}_1 + \hat{x} - \hat{\zeta}.
\end{aligned} \tag{4.3.5}$$

This design can be used for measurement situations where there is no left-right effect to be estimated. In this case, the equations for the deviations ξ_i do not have the term ζ, and the degrees of freedom associated with s_w is $\nu = 6$. All other computations remain the same.

The value of the check standard for one such sequence is defined as the difference between the estimated values of the two reference standards for the sequence as

$$c = \frac{1}{5}(2d_1 - d_2 - d_5 - d_6 - d_7 + d_8 + d_{10}) \qquad (4.3.6)$$

4.3.2 Process Parameters

Initial values of the process parameters are obtained from n such designs, yielding check standard values $c_1, \cdots, c_n$ and within standard deviations $s_{w_1}, \cdots, s_{w_n}$. The accepted value of the check standard is defined as the mean of the check standard values; namely,

$$A_c = \frac{1}{n}\sum_{i=1}^{n} c_i \; . \qquad (4.3.7)$$

The accepted value of the within standard deviation, describing short-term phenomena that affect the measurements within the design, is the pooled value

$$s_p = \left(\frac{1}{n}\sum_{i=1}^{n} s_{w_i}^2\right)^{1/2} \qquad (4.3.8)$$

with degrees of freedom $\nu_1 = \nu \cdot n$.

The total standard deviation of the check standard is defined as

$$s_c = \left(\frac{1}{n-1}\sum_{i=1}^{n} (c_i - A_c)^2\right)^{1/2} \qquad (4.3.9)$$

with $\nu_2 = n-1$ degrees of freedom.

The model assumed for the process is the additive model (1.4.2). Under this model the error structure for the check standard measurement and the error structure for the reported value of an individual test item are such that the appropriate standard deviation for a value reported for a test item is

$$s_r = \frac{\sqrt{3}}{2} s_c \; .$$

The control limits[k] that are appropriate for future check standard values are given by

$$\text{Upper control limit} = A_c + 3s_c$$

$$\text{Lower control limit} = A_c - 3s_c \ .$$

4.3.3 Control Procedure

A test statistic t_c that depends on the observed value of the check standard c is computed for each design by

$$t_c = \frac{|c - A_c|}{s_c} \ . \qquad (4.3.10)$$

The control procedure depends upon this test statistic and the within standard deviation s_w for that design. A dual control procedure is applied as follows:

If $$t_c < 3 \qquad (4.3.11)$$

and if $$s_w < s_p \sqrt{F_\alpha(\nu, \nu_1)} \qquad (4.3.12)$$

for α chosen suitably small, the process is in control and values of the test items are reported as

$$\begin{aligned} X^* &= \hat{x} + R^* \\ Y^* &= \hat{y} + R^* \qquad (4.3.13) \\ Z^* &= \hat{z} + R^* \ . \end{aligned}$$

The restraint is defined as $R^* = \frac{1}{2}(R_1^* + R_2^*)$ where R_1^* and R_2^* are the assigned values of the reference standards.

If $$t_c > 3,$$

the calibration of the test items is invalid and must be repeated.

4.3.4 Transfer with NBS

Given three transfer standards T_1, T_2, and T_3, the transfer with NBS could be accomplished in one of several ways such as including only one transfer standard in each design. The most straightforward way is to let the transfer standards take the place of the test items X, Y, and Z in the design. The calibration design is repeated p times, and process control should be confirmed for each repetition as defined by (4.3.11) and (4.3.12).

[k]The factor 3 is used in this and all subsequent computations in place of the appropriate percent point of the t distribution; namely, $t_{\alpha/2}(\nu)$.

Any design that is out-of-control should be repeated until control is reestablished or else that design is deleted from the transfer. If the values assigned to the transfer standards by NBS are T_1^*, T_2^*, and T_3^* with uncertainties U_{T1}, U_{T2}, and U_{T3} respectively, the uncertainty of the transfer is

$$U_{tr} = \frac{3}{2\sqrt{15p}}\, s_c + \frac{1}{3}\left(U_{T1}^2 + U_{T2}^2 + U_{T3}^2\right)^{1/2}. \qquad (4.3.14)$$

A characteristic of the design that is not always recognized is that the offset Δ of the laboratory process from NBS is defined only in terms of the restraint and not in terms of individual reference standards. The reference standards should not be used separately and, if one standard is replaced, the value of the remaining standard and the replacement standard must be reestablished in relationship to NBS.

Given the p values assigned to each transfer standard by (4.3.13); namely,

$$X_1^*, \cdots, X_p^*$$

$$Y_1^*, \cdots, Y_p^*$$

$$Z_1^*, \cdots, Z_p^*,$$

the offset is computed as

$$\Delta = \frac{1}{3p} \sum_{i=1}^{p} (X_i^* + Y_i^* + Z_i^*) - \frac{1}{3}(T_1^* + T_2^* + T_3^*). \qquad (4.3.15)$$

The offset is judged significant if

$$\tilde{t} > 3\,, \qquad (4.3.16)$$

where

$$\tilde{t} = \frac{2\sqrt{15p}\ |\Delta|}{s_c} \qquad (4.3.17)$$

and in such case the assigned value of the restraint R^* is changed to $R^* - \Delta$. The restraint is unchanged if $\tilde{t} \leq 3$.

4.3.5 Uncertainty

The total uncertainty that is appropriate for a value assigned to a test item by (4.3.13) from one design is

$$U = \frac{3\sqrt{3}}{2}\, s_c + U_{tr}. \qquad (4.3.18)$$

4.4 Comparator Process for Mass Calibrations with One Check Standard for Each Series

4.4.1 Measurement Sequence

High precision mass determination is done by a sequence of intercomparisons that relate the mass of an object to the laboratory's kilogram reference standards which in turn are related to the Paris kilogram. An entire weight set may require several series of intercomparisons in order to assign values to all weights. The weights in each series are intercompared by a statistical design that prescribes the weighings. Each weighing involves a mass difference between two nominally equal weights or groups of weights. Values assigned thereby are least-squares estimates from the design. Provision for a check standard is included with the weights for each series. The reader is referred to Cameron et al. [5] for the statistical theory governing weighing designs; to Jaeger and Davis [54] for the physical theory; to Varner [55] for a description of the NBS software for mass determination; and to Appendix A in this publication for a description of the matrix manipulations needed for a solution to general weighing designs and the propagation of standard deviations and uncertainties through several series.

Normally the first series involves two kilogram reference standards, R_1 and R_2, a test kilogram X_{10}, and a summation Σ_1 of other weights totaling one kilogram nominally. The restraint is the average of the values assigned to R_1 and R_2, and the check standard is defined as the difference between R_1 and R_2 as estimated from the design.

The value assigned to the summation Σ_1 by the first series constitutes the restraint for the second series with the individual weights in the summation being calibrated separately in the second series. For example, if a 500 gram, a 300 gram, and a 200 gram weight make up the summation totaling one kilogram, those weights are assigned values in the second series of intercomparisons. Two series are needed to calibrate a weight set consisting of 1kg, 500g, 300g, 200g, and 100g weights, say. A summation of weights Σ_2 which becomes the restraint for third series is included in the second series if the weight set is to be extended to 50g, 30g, 20g, and 10g weights, and the calibration is extended to lesser weights in like manner.

The weighing designs for two such series are described generically as a 1, 1, 1, 1 design and a 5, 3, 2, 1, 1, 1 design representing the ratios of the weights in the series to each other. A design consists of a subset of all possible intercomparisons that can be made on the group of weights with several factors dictating this choice. A design is always constructed so that the standard deviation of reported values for weights of the same nominal size are equal. The number of intercomparisons is kept small, less than twenty, so that the weighings can be completed with thermal effects being minimized. Furthermore, the number of weights that one is willing to have on the pan at one time and the maximum load of the balance have some bearing on the choice of observations.

Two designs satisfying these criteria are shown below for calibrating the aforementioned weight set. These designs are used routinely in the NBS calibration program. Six observations designated by $d_1,\cdots,d_6$ suffice for the first series. A check standard for the first series is constructed by differencing the values of R_1 and R_2 that were estimated from the design. The second series has eleven observations designated by $d_1,\cdots,d_{11}$. Notice that a 100g weight designated as C is included in this design as a check standard. An observation for a single pan balance is defined as the mass difference between the weights marked by (+) and the weights marked by a (-) as defined by Jaeger and Davis [54].

Design for 1st Series

Obs	1kg	1kg	1kg	1kg
	R_1	R_2	X_{10}	Σ_1
d_1	+	-		
d_2	+		-	
d_3	+			-
d_4		+	-	
d_5		+		-
d_6			+	-

Design for 2nd Series

Obs	500g	300g	200g	100g	100g	100g
	X_5	X_3	X_2	X_1	Σ_2	C
d_1	+	-	-	+	-	
d_2	+	-	-		+	-
d_3	+	-	-	-		+
d_4	+	-	-			
d_5	+		-	-	-	-
d_6		+	-	+	-	-
d_7		+	-	-	+	-
d_8		+	-	-	-	+
d_9			+	-	-	
d_{10}			+	-		-
d_{11}			+		-	-

4.4.2 Process Parameters

The check standard for the first series is defined as

$$c_1 = (1/4)\{2d_1 + d_2 + d_3 - d_4 - d_5\} . \qquad (4.4.1)$$

The check standard for the second series is defined as

$$c_2 = (1/920)\{4d_1 - 111d_2 + 119d_3 + 4d_4 - 108d_5 - 102d_6 - 102d_7 + 128d_8 - 10d_9 - 125d_{10} - 125d_{11}\}. \qquad (4.4.2)$$

The within standard deviation for the first series is

$$s_{w_1} = \left(\frac{1}{4} \sum_{i=1}^{6} \xi_i^2 \right)^{1/2} \qquad (4.4.3)$$

with $\nu_1 = 4$ degrees of freedom.

The deviations ξ_i that are needed to compute s_{w_1} are defined by:

$$\xi_1 = d_1 - (1/4)\ [2d_1 - d_2 - d_3 + d_4 + d_5]$$

$$\xi_2 = d_2 - (1/4)\ [-d_1 + 2d_2 - d_3 - d_4 + d_6]$$

$$\xi_3 = d_3 - (1/4)\ [-d_1 - d_2 + 2d_3 - d_5 - d_6]$$

$$\xi_4 = d_4 - (1/4)\ [d_1 - d_2 + 2d_4 - d_5 + d_6]$$

$$\xi_5 = d_5 - (1/4)\ [d_1 - d_3 - d_4 + 2d_5 - d_6]$$

$$\xi_6 = d_6 - (1/4)\ [d_2 - d_3 + d_4 - d_5 + 2d_6] \ .$$

The within standard deviation for the second series is

$$s_{w_2} = \left(\frac{1}{6} \sum_{i=1}^{11} \xi_i^2 \right)^{1/2} \qquad (4.4.4)$$

with $\nu_2 = 6$ degrees of freedom.

The deviations needed to compute the within standard deviation s_{w_2} are defined as follows:

$$\xi_1 = d_1 - \hat{x}_5 + \hat{x}_3 + \hat{x}_2 - \hat{x}_1 + \hat{\Sigma}_2$$

$$\xi_2 = d_2 - \hat{x}_5 + \hat{x}_3 + \hat{x}_2 - \hat{\Sigma}_2 + c_2$$

$$\xi_3 = d_3 - \hat{x}_5 + \hat{x}_3 + \hat{x}_2 + \hat{x}_1 - c_2$$

$$\xi_4 = d_4 - \hat{x}_5 + \hat{x}_3 + \hat{x}_2 \qquad (4.4.5)$$

$$\xi_5 = d_5 - \hat{x}_5 + \hat{x}_2 + \hat{x}_1 + \hat{\Sigma}_2 + c_2$$

$$\xi_6 = d_6 - \hat{x}_3 + \hat{x}_2 - \hat{x}_1 + \hat{\Sigma}_2 + c_2$$

$$\xi_7 = d_7 - \hat{x}_3 + \hat{x}_2 + \hat{x}_1 - \hat{\Sigma}_2 + c_2$$

$$\xi_8 = d_8 - \hat{x}_3 + \hat{x}_2 + \hat{x}_1 + \hat{\Sigma}_2 - c_2$$

$$\xi_9 = d_9 - \hat{x}_2 + \hat{x}_1 + \hat{\Sigma}_2$$

$$\xi_{10} = d_{10} - \hat{x}_2 + \hat{x}_1 + c_2$$

$$\xi_{11} = d_{11} - \hat{x}_2 + \hat{\Sigma}_2 + c_2$$

where

$$\hat{x}_5 = (1/920)\{100(d_1 + d_2 + d_3 + d_4) + 60d_5 - 20(d_6 + d_7 + d_8 + d_9 + d_{10} + d_{11})\}$$

$$\hat{x}_3 = (1/920)\{-68(d_1 + d_2 + d_3 + d_4) - 4d_5 + 124(d_6 + d_7 + d_8) - 60(d_9 + d_{10} + d_{11})\}$$

$$\hat{x}_2 = (1/920)\{-32(d_1 + d_2 + d_3 + d_4) - 56d_5 - 104(d_6 + d_7 + d_8) + 80(d_9 + d_{10} + d_{11})\}$$

$$\hat{x}_1 = (1/920)\{119d_1 + 4d_2 - 111d_3 + 4d_4 - 108d_5 + 128d_6 - 102(d_7 + d_8) - 125(d_9 + d_{10}) - 10d_{11}\} \quad (4.4.6)$$

$$\hat{\Sigma}_2 = (1/920)\{-111d_1 + 119d_2 + 4(d_3 + d_4) - 108d_5 - 125d_6 + 128d_7 - 102d_8 - 125d_9 - 10d_{10} - 125d_{11}\}$$

Accepted values for the check standards, within standard deviations, and total standard deviations are obtained from n initial repetitions of the two series. Check standard values $c_{11},\cdots,c_{1n}$ and $c_{21},\cdots,c_{2n}$ from the respective series are averaged to obtain accepted values,

$$A_{c_1} = \frac{1}{n}\sum_{i=1}^{n} c_{1i}$$

and (4.4.7)

$$A_{c_2} = \frac{1}{n}\sum_{i=1}^{n} c_{2i} \quad .$$

Similarly, within standard deviations $s_{w_{11}},\cdots,s_{w_{1n}}$ from the first series and $s_{w_{21}},\cdots,s_{w_{2n}}$ from the second series are pooled to obtain accepted within standard deviations for the two series:

$$s_{p_1} = \left(\frac{1}{n}\sum_{i=1}^{n} s_{w_{1i}}^2\right)^{1/2}$$

and (4.4.8)

$$s_{p_2} = \left(\frac{1}{n}\sum_{i=1}^{n} s_{w_{2i}}^2\right)^{1/2} \quad .$$

The total standard deviations for the check standards for each series are respectively

$$s_{c_1} = \left(\frac{1}{n-1}\sum_{i=1}^{n} (c_{1i} - A_{c_1})^2\right)^{1/2}$$

and (4.4.9)

$$s_{c_2} = \left(\frac{1}{n-1}\sum_{i=1}^{n} (c_{2i} - A_{c_2})^2\right)^{1/2} \quad .$$

4.4.3 Control Procedure[ℓ]

Statistical control is maintained on the measurements by series. For the first series, test statistics computed from the current check standard value c_1, and the within standard deviation s_{w_1} are used to test for control. Let

$$t_{c_1} = \frac{|A_{c_1} - c_1|}{s_{c_1}} \quad . \tag{4.4.10}$$

If

$$t_{c_1} < 3 \tag{4.4.11a}$$

and if

$$s_{w_1} < s_{p_1} \sqrt{F_\alpha(4,4n)} \tag{4.4.11b}$$

for α chosen suitably small, the measurement process is in control, and the following values are assigned to the the test weight X_{10} and summation Σ_1:

$$X_{10}^* = -(1/8)\{3d_2 + d_3 + 3d_4 + d_5 - 2d_6\} + R^*$$

$$\Sigma_1^* = -(1/8)\{d_2 + 3d_3 + d_4 + 3d_5 + 2d_6\} + R^* \tag{4.4.12}$$

where $R^* = \frac{1}{2}(R_1^* + R_2^*)$ and R_1^* and R_2^* are the corrections to nominal size for the kilogram standards R_1 and R_2.

Statistical control for the second series depends upon the current check standard value c_2 and within standard deviation s_{w_2} for that series. Let

$$t_{c_2} = \frac{|A_{c_2} - c_2|}{s_{c_2}} \quad . \tag{4.4.13}$$

If

$$t_{c_2} < 3 \tag{4.4.14a}$$

and if

$$s_{w_2} < s_{p_2} \sqrt{F_\alpha(6,6n)} \tag{4.4.14b}$$

the measurement process is in control for that series.

Equations (4.4.10) and (4.4.13) are the simplest constructions for testing for offset using a t statistic. The technique for constructing these statistics follows the general method for t statistics; namely, the difference between

[ℓ]The factor 3 is used in this and all subsequent computations in place of the appropriate factor of the t distribution; namely, $t_{\alpha/2}(\nu)$.

the current value of the check standard and its accepted value divided by the standard deviation of the check standard. As such the construction is applicable to any design. In this case the statistic defined by (4.4.10) is precisely correct if the data base for check standard C_1 comes from identical designs with identical restraints, and similarly for the statistic deifined by (4.4.13). In practice a check standard, especially C_2, can be utilized in a variety of designs. This does not affect the interpretation of the accepted value of the check standard, but it does affect the interpretation of the total standard deviation. In such case the test statistics can be computed using the within standard deviations as follows:

$$t_{c_1} = \frac{\sqrt{2}|A_{c_1} - c_1|}{s_{w_1}} \qquad (4.4.10a)$$

$$t_{c_2} = \frac{|A_{c_2} - c_2|}{\left(\frac{29}{230} s_{w_2}^2 + \frac{1}{100} \cdot \frac{3}{8} s_{w_1}^2\right)^{1/2}} \qquad (4.4.13a)$$

These equations are compatible with the documenation in reference [55] where the between component of variance is assumed to be zero --an assumption that is true for the NBS mass calibration process. Notice that the construction of the relevant t statistic becomes increasingly complicated as one moves through the series of weighings depending as it does on the within standard deviations from all prior designs. See Appendix A for the general construction for any design.

Given that (4.4.14a) and (4.4.14b) are satisfied, values are reported for test items and summation for the next series as follows:

Weights	Reported Values
500g	$X_5^* = \hat{x}_5 + \frac{1}{2} \Sigma_1^*$
300g	$X_3^* = \hat{x}_3 + \frac{276}{920} \Sigma_1^*$
200g	$X_2^* = \hat{x}_2 + \frac{184}{920} \Sigma_1^*$
100g	$X_1^* = \hat{x}_1 + \frac{92}{920} \Sigma_1^*$
Σ100g	$\Sigma_2^* = \hat{\Sigma}_2 + \frac{92}{920} \Sigma_1^*$

(4.4.15)

where $\hat{x}_5$, $\hat{x}_3$, $\hat{x}_2$, $\hat{x}_1$ and $\hat{\Sigma}_2$ are defined in (4.4.6) and Σ_1^* is defined in (4.4.12). Whenever a series is out-of-control, the calibration results for the test weights in that series are invalid and must be repeated.

4.4.4 Transfer with NBS

For a mass measurement assurance program the laboratory's starting kilograms are calibrated at NBS and assigned values R_1^* and R_2^* and associated uncertainties U_{R1} and U_{R2}. The transfer is accomplished by relating all weighings to these standards as explained in section 4.4.1.

4.4.5 Uncertainty[m]

The uncertainty associated with the value assigned to any weight is a function of the design and the within standard deviations for that series and all prior series. It also includes as systematic error a proportional part of the uncertainty associated with the starting restraint. For example, the uncertainty for the value assigned to the one kilogram summation Σ_1^* which is the starting restraint for the second series is U_{1000} where

$$U_{1000} = 3\sqrt{k_1}\, s_{w_1} + \frac{1}{2}(U_{R1} + U_{R2}), \qquad k_1 = \frac{3}{8} \quad . \qquad (4.4.16)$$

The uncertainties for the 500g, 300g, 200g, and 100g test weights are respectively:

$$U_{500} = 3\left(k_2\, s_{w_2}^2 + \frac{3}{8}\, m_2^2 s_{w_1}^2\right)^{1/2} + \frac{m_2}{2}(U_{R1} + U_{R2}), \quad k_2 = \frac{50}{920}, \quad m_2 = \frac{1}{2}$$

$$U_{300} = 3\left(k_3\, s_{w_2}^2 + \frac{3}{8}\, m_3^2 s_{w_1}^2\right)^{1/2} + \frac{m_3}{2}(U_{R1} + U_{R2}), \quad k_3 = \frac{82}{920}, \quad m_3 = \frac{3}{10}$$

$$U_{200} = 3\left(k_4\, s_{w_2}^2 + \frac{3}{8}\, m_4^2 s_{w_1}^2\right)^{1/2} + \frac{m_4}{2}(U_{R1} + U_{R2}), \quad k_4 = \frac{64}{920}, \quad m_4 = \frac{1}{5}$$

$$U_{100} = 3\left(k_5\, s_{w_2}^2 + \frac{3}{8}\, m_5^2 s_{w_1}^2\right)^{1/2} + \frac{m_5}{2}(U_{R1} + U_{R2}), \quad k_5 = \frac{116}{920}, \quad m_5 = \frac{1}{10}$$

[m] Uncertainties are computed assuming the between component of variance is zero. See reference [55] for the general construction.

4.5 Comparator Process for Four Reference Standards and Four Test Items

4.5.1 Measurement Sequence

This design for four reference standards and four test items involves the intercomparison of items two at a time where each test item is intercompared with each standard one time, and there is no direct intercomparison among standards or test items. The design is routinely used for voltage measurements where the laboratory's reference standards R_1, R_2, R_3 and R_4 in one temperature controlled box are intercompared with test items W, X, Y and Z or transfer standards in another box, and there are no intercomparisons within a box.

Schematically, the intercomparisons are as shown below where a plus (+) or a minus (-) indicates relative position in the circuit.

Test \ Ref	R_1	R_2	R_3	R_4
W	+	-	+	-
X	-	+	-	+
Y	+	-	+	-
Z	-	+	-	+

Measurements on the laboratory standards R_1, R_2, R_3, and R_4 and the test items W, X, Y and Z are designated by r_1, r_2, r_3 and r_4 and w, x, y, and z respectively. The design consists of the following sequence of difference measurements:

$$\begin{aligned} d_1 &= r_1 - w \\ d_2 &= r_1 - y \\ d_3 &= r_3 - y \\ d_4 &= r_3 - w \\ d_5 &= r_2 - x \\ d_6 &= r_2 - z \\ d_7 &= r_4 - z \\ d_8 &= r_4 - x \\ d_9 &= x - r_1 \\ d_{10} &= z - r_1 \\ d_{11} &= z - r_3 \\ d_{12} &= x - r_3 \\ d_{13} &= w - r_2 \\ d_{14} &= y - r_2 \\ d_{15} &= y - r_4 \\ d_{16} &= w - r_4 \end{aligned} \qquad (4.5.1)$$

The design has several features that make it particularly suitable for intercomparing saturated standard cells. Let the observations d_i, ordered as in (4.5.1) so as to minimize the number of circuit connections, represent the differences in emf between two cells as measured by a potentiometer. The convention adhered to is, for example, that r_1-w represents the measured difference between R_1 and W with the cells reversed in the circuit relative to their positions for the difference w-r_1.

The design is balanced so as to cancel out any spurious emf that may be present in the circuit [56]. In the presence of such systematic error, called left-right effect, the measurements d_i are assumed to be related to the actual differences D_i in emf between two cells in the following way:

$$d_i = D_i + \zeta + \varepsilon_i \qquad i = 1,\cdots,16$$

where ζ is the left-right effect, and ε_i is random error. For a circuit with negligible left-right effect, one expects that the measurements would sum to zero except for the effect of random error. Any disparity between this expectation and the summation gives an estimate of the magnitude of left-right effect; namely,

$$\hat{\zeta} = \frac{1}{16} \sum_{i=1}^{16} d_i \,. \qquad (4.5.2)$$

A measuring process such as the one described in the foregoing paragraph can be characterized by:

i) a short-term or within standard deviation which describes variability during the time necessary to make the sixteen measurements for one design.

ii) accepted values for check standards which have been specifically chosen for this measurement situation.

iii) a total standard deviation for the process based on the check standard measurements.

The difference of each test item from the average of the reference group is computed by:

$$\begin{aligned} \hat{w} &= -\frac{1}{4}(d_1 + d_4 - d_{13} - d_{16}) \\ \hat{x} &= -\frac{1}{4}(d_5 + d_8 - d_9 - d_{12}) \\ \hat{y} &= -\frac{1}{4}(d_2 + d_3 - d_{14} - d_{15}) \\ \hat{z} &= -\frac{1}{4}(d_6 + d_7 - d_{10} - d_{11}) \end{aligned} \qquad (4.5.3)$$

The foregoing quantities in conjunction with the differences of the reference standards from their group average; namely,

$$\hat{r}_1 = \frac{1}{16}(3d_1+3d_2-d_3-d_4-d_5-d_6-d_7-d_8-3d_9-3d_{10}+d_{11}+d_{12}+d_{13}+d_{14}+d_{15}+d_{16})$$

$$\hat{r}_2 = \frac{1}{16}(-d_1-d_2-d_3-d_4+3d_5+3d_6-d_7-d_8+d_9+d_{10}+d_{11}+d_{12}-3d_{13}-3d_{14}+d_{15}+d_{16})$$

$$\hat{r}_3 = \frac{1}{16}(-d_1-d_2+3d_3+3d_4-d_5-d_6-d_7-d_8+d_9+d_{10}-3d_{11}-3d_{12}+d_{13}+d_{14}+d_{15}+d_{16}) \quad (4.5.4)$$

$$\hat{r}_4 = \frac{1}{16}(-d_1-d_2-d_3-d_4-d_5-d_6+3d_7+3d_8+d_9+d_{10}+d_{11}+d_{12}+d_{13}+d_{14}-3d_{15}-3d_{16})$$

and the estimated left-right effect $\hat{\zeta}$ are used to estimate a within standard deviation s_w for each design; namely,

$$s_w = \left(\frac{1}{8} \sum_{i=1}^{16} \xi_i^2 \right)^{1/2} \quad (4.5.5)$$

with $\nu=8$ degrees of freedom. The individual deviations ξ_i are given by:

$$\begin{aligned}
\xi_1 &= d_1 - \hat{r}_1 + \hat{w} - \hat{\zeta} \\
\xi_2 &= d_2 - \hat{r}_1 + \hat{y} - \hat{\zeta} \\
\xi_3 &= d_3 - \hat{r}_3 + \hat{y} - \hat{\zeta} \\
\xi_4 &= d_4 - \hat{r}_3 + \hat{w} - \hat{\zeta} \\
\xi_5 &= d_5 - \hat{r}_2 + \hat{x} - \hat{\zeta} \\
\xi_6 &= d_6 - \hat{r}_2 + \hat{z} - \hat{\zeta} \\
\xi_7 &= d_7 - \hat{r}_4 + \hat{z} - \hat{\zeta} \\
\xi_8 &= d_8 - \hat{r}_4 + \hat{x} - \hat{\zeta} \\
\xi_9 &= d_9 - \hat{x} + \hat{r}_1 - \hat{\zeta} \\
\xi_{10} &= d_{10} - \hat{z} + \hat{r}_1 - \hat{\zeta} \\
\xi_{11} &= d_{11} - \hat{z} + \hat{r}_3 - \hat{\zeta} \\
\xi_{12} &= d_{12} - \hat{x} + \hat{r}_3 - \hat{\zeta} \\
\xi_{13} &= d_{13} - \hat{w} + \hat{r}_2 - \hat{\zeta} \\
\xi_{14} &= d_{14} - \hat{y} + \hat{r}_2 - \hat{\zeta} \\
\xi_{15} &= d_{15} - \hat{y} + \hat{r}_4 - \hat{\zeta} \\
\xi_{16} &= d_{16} - \hat{w} + \hat{r}_4 - \hat{\zeta}
\end{aligned} \quad (4.5.6)$$

Check standards for electrical measurements are not easily defined because of the inherent nature of electrical quantities to drift over time. For this reason, three separate check standards are recommended for measurements on standard cells. The left-right effect reflects many of the sources of error in the measurement system and can be presumed to remain stable over time. For this reason it makes a suitable check standard for process control. Specifically, the value of the first check standard is defined for each design as ζ from (4.5.2).

There is also a need to check on the stability of the reference standards, changes or instabilities in which may not be reflected in the left-right effect. The least-squares estimates for the reference standards from the design (4.5.4) cannot be used to check on the stability of the standards themselves because these estimates are in effect a consequence of the design, subject to the restraint, and are not meaningful separately. For example, if the restraint is changed to exclude one of the reference standards, the least-squares estimates for the remaining reference standards as computed from the same observed differences (4.5.1) can change appreciably.

The information in a design does, however, allow a way of monitoring the change in one reference standard relative to another reference standard. A measured difference between two reference standards that is not subject to the restraint can be computed from each design, and two check standards, each one involving the difference between two reference standards, are recommended for monitoring the stability of the four reference standards.

Check standard C_1 is defined for the difference between R_1 and R_3, and check standard C_2 is defined for the difference between R_2 and R_4. Their respective values c_1 and c_2 are computed for each design as follows:

$$c_1 = \frac{1}{4}(d_1 + d_2 - d_3 - d_4 - d_9 - d_{10} + d_{11} + d_{12}) \qquad (4.5.7)$$

$$c_2 = \frac{1}{4}(d_5 + d_6 - d_7 - d_8 - d_{13} - d_{14} + d_{15} + d_{16}) .$$

Because it is anticipated that the change in one reference standard relative to another may not be stable over time, the method for analyzing check standards C_1 and C_2 is a modified process control technique that allows for linear drift.

4.5.2 Process Parameters for Stable and Drifting Check Standards

Initial values for the process parameters are established from n repetitions of the design in which the four reference standards are compared to any four test items. The resulting check standard measurements are $\zeta_1,\cdots,\zeta_n$; $c_{11},\cdots,c_{1n}$; and $c_{21},\cdots,c_{2n}$. For the left-right effect the n values are averaged to obtain the accepted value

$$A_\zeta = \frac{1}{n}\sum_{i=1}^{n}\hat{\zeta}_i \,. \qquad (4.5.8)$$

A total standard deviation for the left-right effect is also computed from the initial check standard measurements by

$$s_\zeta = \left(\frac{1}{(n-1)}\sum_{i=1}^{n}(\hat{\zeta}_i - A_\zeta)^2\right)^{1/2} \qquad (4.6.9)$$

with $\nu = (n-1)$ degrees of freedom.

The control limits[o] that are appropriate for future measurements on the left-right effect are:

$$\text{Upper Control Limit} = A_\zeta + 3s_\zeta$$

$$\text{Lower Control Limit} = A_\zeta - 3s_\zeta.$$

Similar calculations of accepted values and standard deviations are made for C_1 and C_2 where the check standard measurements $c_{11},\cdots,c_{1n}$ and $c_{21},\cdots,c_{2n}$ are stable over time. More often than not these quantities are not stable over time, and this fact must be taken into account in the analysis. If the check standard values show drift and if the drift is linear with time, check standard values $c_1,\cdots,c_n$ at times $t_1,\cdots,t_n$ can be characterized by

$$c_i = \alpha + \beta t_i \qquad i=1,\cdots,n$$

where the intercept α and the slope β are estimated by

$$\hat{\alpha} = \bar{c} - \hat{\beta}\,\bar{t}$$

and

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(t_i-\bar{t})(c_i-\bar{c})}{\sum_{i=1}^{n}(t_i-\bar{t})^2}$$

---

[o]The factor 3 is used in this and all subsequent computations in place of the appropriate percent point of the t distribution; namely, $t_{\alpha/2}(\nu)$.

with $$\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i \quad \text{and} \quad \bar{c} = \frac{1}{n}\sum_{i=1}^{n} c_i .$$

In the linear case the accepted total standard deviation for each check standard is

$$s_c = \left(\frac{1}{n-2} \sum_{i=1}^{n} (c_i - \hat{\alpha} - \hat{\beta} t_i)^2 \right)^{1/2} \qquad (4.5.10)$$

with $\nu = n-2$ degrees of freedom. See reference [59] for analyses relating to linear regression models.

The parameters of the linear fit and associated standard deviations should be computed for C_1 and C_2 separately resulting in estimates α_1, β_1, s_{c_1} with $\nu_1 = n-2$ degrees of freedom for check standard C_1 and α_2, β_2, s_{c_2} with $\nu_2 = n-2$ degrees of freedom for check standard C_2. The value that a check standard is expected to take on at any given time is thus dependent on the linear fit. Therefore, for a future time t', provided t' is not too far removed from t_n, the accepted values for the check standards are defined by

$$A'_{c_1} = \hat{\alpha}_1 + \hat{\beta}_1 t'$$

and (4.5.11)

$$A'_{c_2} = \hat{\alpha}_2 + \hat{\beta}_2 t' .$$

A total standard deviation for the measurements on C_1 and C_2 can be pooled from s_{c_1} and s_{c_2} by the formula

$$s_c = \left(\frac{1}{2}\left(s_{c_1}^2 + s_{c_2}^2 \right) \right)^{1/2} \qquad (4.5.12)$$

with $\nu = 2(n-2)$ degrees of freedom.

The control procedure assumes that t' is close to t_n because the standard deviation of a predicted value from a linear fit increases dramatically as the linear fit is extrapolated beyond the check standard data. Thus the chance of detecting a real shift in the process diminishes as the tests for control are continued into the future. This fact necessitates frequent updating of the parameters of the linear fit based on recent check standard values.

Furthermore, the control procedure and the assumption of a linear model are interdependent. Because there is no way of separating these two elements, an out-of-control signal can be caused by either lack of process control or a breakdown in the linearity of the check standard measurements. One must recognize this as a short-coming in the control procedure and arrange for other independent checks on the stability of the reference standards.

The control procedure also makes use of the accepted within standard deviation s_p which is not dependent upon model assumptionsfor the check standards . It is computed from the within standard deviations $s_{w_1}, \cdots, s_{w_n}$ for each design as follows:

$$s_p = \left(\frac{1}{n} \sum_{i=1}^{n} s_{w_i}^2 \right)^{1/2} \qquad (4.5.13)$$

with $\nu_3 = 8n$ degrees of freedom.

4.5.3 Process Control

Process control is maintained by monitoring the within standard deviation for each design and the performance of the three designated check standards. If check standard C_1 or C_2 repeatedly fails the test for control, it is likely that one of the two reference standards comprising the check standard has changed in value. In this case it will be necessary to replace one or both of the standards in question or reestablish their values relative to NBS.

Process control should be verified for the within standard deviation s_w as it is calculated for each design and for the current values of the check standards for that design; namely, ζ, c_1, and c_2. For the left-right effect ζ, the test statistic is:

$$t_\zeta = \frac{|\hat{\zeta} - A_\zeta|}{s_\zeta} . \qquad (4.5.14)$$

For check standards C_1 and C_2 that are drifting linearly over time the corresponding test statistics at time t' are:

$$t_{c_1} = \frac{|c_1 - A_{c_1}'|}{\tilde{s}} \qquad (4.5.15)$$

and

$$t_{c_2} = \frac{|c_2 - A_{c_2}'|}{\tilde{s}}$$

where

$$\tilde{s} = s_c \left(\frac{n+1}{n} + \frac{(t' - \bar{t})^2}{\sum_{i=1}^{n} (t_i - \bar{t})^2} \right)^{1/2} .$$

Then the following conditions can be imposed:

$$\text{If } t_\zeta \text{ and } t_{c_1} \text{ and } t_{c_2} \text{ are all } < 3 \qquad (4.5.16a)$$

and if

$$s_w < s_p\sqrt{F_\alpha(8, \nu_3)} \qquad (4.5.16b)$$

for α suitably small, the process is judged in control for that design.

The values of the test items are reported as

$$\begin{aligned} W^* &= \hat{w} + R^* \\ X^* &= \hat{x} + R^* \\ Y^* &= \hat{y} + R^* \\ Z^* &= \hat{z} + R^* \end{aligned} \qquad (4.5.17)$$

where the restraint $R^* = \frac{1}{4}(R_1^* + R_2^* + R_3^* + R_4^*)$, and R_1^*, R_2^*, R_3^*, and R_4^* are the values assigned to the laboratory's reference standards.

If the results of the control procedures along with other experimental evidence indicate instability or other anomalous behavior on the part of one of the reference standards, the entire experiment need not necessarily be discarded. It is possible to delete the reference standard in question from the restraint and obtain new values for the test items if the values of the remaining reference standards are known individually. For example, if one is involved in a transfer with NBS, and if reference standard R_1 shows signs of serious malfunction after several days of intercomparisons between the reference standards and the transfer standards, the values for the transfer standards can be recomputed for each design as follows:

$$\hat{w} = \frac{1}{48}\{-9d_1+3d_2-d_3-13d_4-d_5-d_6-d_7-d_8-3d_9-3d_{10}+d_{11}+d_{12}+13d_{13}+d_{14}+d_{15}+13d_{16}\}$$

$$\hat{x} = \frac{1}{48}\{3d_1+3d_2-d_3-d_4-13d_5-d_6-d_7-13d_8+9d_9-3d_{10}+d_{11}+13d_{12}+d_{13}+d_{14}+d_{15}+d_{16}\} \qquad (4.5.18)$$

$$\hat{y} = \frac{1}{48}\{3d_1-9d_2-13d_3-d_4-d_5-d_6-d_7-d_8-3d_9-3d_{10}+d_{11}+d_{12}+d_{13}+13d_{14}+13d_{15}+d_{16}\}$$

$$\hat{z} = \frac{1}{48}\{3d_1+3d_2-d_3-d_4-d_5-13d_6-13d_7-d_8-3d_9+9d_{10}+13d_{11}+d_{12}+d_{13}+d_{14}+d_{15}+d_{16}\}$$

and W^*, X^*, Y^* and Z^* are computed according to (4.5.17) with the restraint R^* changed to:

$$R^* = \frac{1}{3}(R_2^* + R_3^* + R_4^*).$$

The differences of reference standards R_2, R_3 and R_4 from their average value are recomputed to be:

$$\hat{r}_2 = \frac{1}{12}\{-d_3-d_4+2d_5+2d_6-d_7-d_8+d_{11}+d_{12}-2d_{13}-2d_{14}+d_{15}+d_{16}\}$$

$$\hat{r}_3 = \frac{1}{12}\{2d_3+2d_4-d_5-d_6-d_7-d_8-2d_{11}-2d_{12}+d_{13}+d_{14}+d_{15}+d_{16}\} \quad (4.5.19)$$

$$\hat{r}_4 = \frac{1}{12}\{-d_3-d_4-d_5-d_6+2d_7+2d_8+d_{11}+d_{12}+d_{13}+d_{14}-2d_{15}-2d_{16}\}$$

The within standard deviation for each design (see equations (4.5.5) and (4.5.6)) can be computed using either the original quantities in (4.5.3) and (4.5.4) or the adjusted quantities in (4.5.18) and (4.5.19) with identical results.

4.5.4 Transfer with NBS

Transfer with NBS is accomplished by means of p repetitions of the design in which four transfer standards T_1, T_2, T_3, and T_4 replace the four test items. If one of the tests for control defined by (4.5.16a) and (4.5.16b) is not satisfied, the design should be repeated or else that repetition should be deleted from the transfer.

Given p repetitions of the design in which T_1 replaces W, T_2 replaces X, T_3 replaces Y and T_4 replaces Z, the p values assigned to each transfer standard by the participant's process are computed from (4.5.17); namely,

$$W_1^*,\cdots,W_p^*$$

$$X_1^*,\cdots,X_p^*$$

$$Y_1^*,\cdots,Y_p^*$$

$$Z_1^*,\cdots,Z_p^*.$$

NBS assigns values to electrical transfer standards that take into account their individual and collective behavior both before, during, and after their sojurn in the participant's laboratory. A transfer standard that displays unstable behavior during one of these periods may be excluded from the analysis. Normally the averages for the four transfer standards from the "before and after" NBS determinations are fit by least-squares to a linear function of time; then average values T_j^* are predicted for the times $t_j (j=1,\cdots,p)$ that the transfer standards were in the participant's laboratory by the equation

$$T_j^* = \hat{\alpha}_o + \hat{\beta}_o t_j \qquad j=1,\cdots,p$$

where $\hat{\alpha}_o$ and $\hat{\beta}_o$ are estimated from NBS measurements.

This makes it possible to compute daily offsets $\Delta_j (j=1,\cdots,p)$ for the reference group where

$$\Delta_j = \frac{1}{4}(W_j^* + X_j^* + Y_j^* + Z_j^*) - T_j^* \qquad j=1,\cdots,p \quad (4.5.20)$$

and assuming the reference group is stable, an average offset for the reference group is computed by

$$\bar{\Delta} = \frac{1}{p}\sum_{j=1}^{p} \Delta_j . \quad (4.5.21)$$

The offset is judged significant if

$$\tilde{t} > 3$$

where

$$\tilde{t} = \frac{4\sqrt{p}\,|\bar{\Delta}|}{(4s_c^2 - s_p^2)^{1/2}} .$$

In such case the value of the laboratory restraint is changed to $R^* - \Delta$.

Otherwise, the restraint is unchanged.

4.5.5 Uncertainty

The uncertainty of the transfer is

$$U_{tr} = \frac{3(4s_c^2 - s_p^2)^{1/2}}{4\sqrt{p}} + U_T \quad (4.5.22)$$

where U_T is the uncertainty assigned to the transfer standards by NBS.

The uncertainty that is appropriate for the laboratory's process as it assigns a value to a test item based on a single design is

$$U = \frac{3}{4}(10s_c^2 - s_p^2)^{1/2} + U_{tr} . \quad (4.5.23)$$

4.5.6 Example

An example is presented from the Volt Transfer Program where an environmentally controlled box of four standard cells was sent to an industrial participant to be intercompared with the participant's box of four standard cells. After the NBS cells had been in the participant's laboratory for two weeks, thereby giving them a chance to recover from the trip, the laboratory's reference cells were intercompared with the NBS cells each day for 16 days using the design described in sec 4.5.1. The data corrected for the temperature in each box are shown in exhibit 4.5.1.

Exhibit 4.5.1 - Intercomparison of laboratory standard cells with NBS cells
Value in microvolts

Obs \ Day	1	2	3	4	5	6	7	8
d_1	86.70	86.92	86.86	86.98	86.97	86.99	87.07	87.17
d_2	87.28	87.02	86.77	86.69	86.57	86.60	86.63	86.68
d_3	89.29	89.01	88.75	88.64	88.52	88.52	88.54	88.57
d_4	88.84	88.98	88.87	88.94	88.97	88.94	88.98	89.10
d_5	88.06	87.97	87.84	87.89	88.33	88.17	88.03	88.11
d_6	87.43	87.04	86.94	86.94	86.96	86.92	86.92	87.07
d_7	88.96	88.58	88.46	88.47	88.45	88.47	88.50	88.55
d_8	89.63	89.52	89.39	89.43	89.85	89.70	89.63	89.60
d_9	-86.47	-86.60	-86.32	-86.34	-86.71	-86.71	-86.61	-86.71
d_{10}	-85.81	-85.70	-85.40	-85.41	-85.39	-85.49	-85.54	-85.66
d_{11}	-87.82	-87.67	-87.34	-87.37	-87.34	-87.40	-87.42	-87.54
d_{12}	-88.49	-88.62	-88.25	-88.32	-88.72	-88.64	-88.54	-88.59
d_{13}	-88.80	-89.11	-88.84	-88.92	-88.90	-88.88	-88.92	-89.03
d_{14}	-89.15	-89.16	-88.73	-88.64	-88.48	-88.47	-88.48	-88.55
d_{15}	-90.90	-90.69	-90.24	-90.13	-89.93	-89.98	-90.02	-90.08
d_{16}	-90.38	-90.64	-90.33	-90.40	-90.35	-90.41	-90.49	-90.60

Obs \ Day	9	10	11	12	13	14	15	16
d_1	87.25	87.28	87.32	87.45	87.46	87.50	87.53	87.59
d_2	86.72	86.80	86.81	86.87	86.90	86.91	86.92	86.97
d_3	88.60	88.52	88.52	88.62	88.59	88.59	88.59	88.62
d_4	89.13	89.07	89.09	89.16	89.18	89.17	89.21	89.24
d_5	88.09	88.00	87.78	88.12	88.05	88.06	88.04	88.05
d_6	87.07	86.99	86.89	87.17	87.15	87.09	87.07	87.07
d_7	88.58	88.56	88.62	88.69	88.82	88.68	88.69	88.70
d_8	89.60	89.55	89.60	89.68	89.79	89.67	89.65	89.68
d_9	-86.66	-86.66	-86.74	-86.78	-86.78	-86.84	-86.92	-86.89
d_{10}	-85.63	-85.67	-85.79	-85.79	-85.80	-85.88	-85.96	-85.93
d_{11}	-87.52	-87.48	-87.55	-87.58	-87.58	-87.60	-87.59	-87.60
d_{12}	-88.57	-88.47	-88.51	-88.53	-88.57	-88.54	-88.57	-88.57
d_{13}	-89.04	-89.00	-89.01	-89.10	-89.06	-89.10	-89.10	-89.12
d_{14}	-88.53	-88.44	-88.47	-88.55	-88.47	-88.51	-88.55	-88.53
d_{15}	-90.07	-90.00	-90.05	-90.12	-90.14	-90.10	-90.12	-90.14
d_{16}	-90.60	-90.54	-90.58	-90.68	-90.69	-90.69	-90.75	-90.77

Exhibit 4.5.2 - Estimates for transfer standards and reference standards
Values in microvolts

Day	NBS Standard Cells T_1	T_2	T_3	T_4	Laboratory Reference Cells R_1	R_2	R_3	R_4	L-R Effect
	$\hat{w}$	$\hat{x}$	$\hat{y}$	$\hat{z}$	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_3$	$\hat{r}_4$	$\hat{\zeta}$
1	-88.68	-88.16	-89.15	-87.50	-1.811	-0.016	0.234	1.592	-0.102
2	-88.91	-88.17	-88.96	-87.24	-1.767	-0.007	0.243	1.531	-0.197
3	-88.72	-87.94	-88.62	-87.03	-1.746	+0.004	0.219	1.522	-0.098
4	-88.80	-87.99	-88.50	-87.04	-1.739	+0.003	0.223	1.513	-0.097
5	-88.80	-88.40	-88.38	-87.04	-1.743	+0.015	0.235	1.492	-0.075
6	-88.81	-88.31	-88.40	-87.07	-1.696	-0.033	0.232	1.497	-0.104
7	-88.87	-88.20	-88.42	-87.10	-1.683	-0.058	0.225	1.515	-0.108
8	-88.97	-88.25	-88.46	-87.20	-1.671	-0.036	0.224	1.482	-0.119
9	-89.00	-88.23	-88.48	-87.20	-1.664	-0.047	0.226	1.486	-0.098
10	-88.98	-88.17	-88.44	-87.18	-1.587	-0.082	0.196	1.473	-0.093
11	-89.00	-88.16	-88.46	-87.22	-1.543	-0.171	0.209	1.504	-0.129
12	-89.10	-88.28	-88.54	-87.31	-1.583	-0.071	0.167	1.487	-0.086
13	-89.10	-88.30	-88.53	-87.34	-1.579	-0.132	0.166	1.546	-0.072
14	-89.12	-88.28	-88.53	-87.32	-1.526	-0.118	0.167	1.477	-0.099
15	-89.15	-88.30	-88.55	-87.30	-1.496	-0.139	0.161	1.474	-0.116
16	-89.18	-88.30	-88.57	-87.33	-1.497	-0.149	0.166	1.481	-0.102

Figures 4-7 show the individual behavior of the transfer standards, and figure figure 8 shows the behavior of the transfer group on the average. One might conclude based on these graphs that the cells were not sufficiently stabilized at the beginning of the experiment and that the first two measurements in the participant's laboratory should be deleted from the transfer data.

The differences for the transfer cells and the reference cells from their group means (See equations (4.5.3) and (4.5.4)) are listed in exhibit 4.5.2. The behavior of the reference cells during the transfer with NBS is of interest because the final assignment of offset depends on the assumption that the reference cells are stable. As was noted earlier in this section, the quantities listed in exhibit 4.5.2 do not describe the behavior of the individual reference cells because these quantities are constrained so that their sum is equal to zero.

The only way to observe the individual cells during the transfer is to reverse the way in which the assignments are currently made; i.e., to analyze the data from the intercomparisons using the reference cells as unknowns and the value of the transfer group from NBS as the restraint. This will give individual values for each reference cell and can be done after the fact if the transfer group proves sufficiently stable. The rationalization for computing an offset using the reference cells as the restraint is that one would expect the reference cells, if they are of the same quality as the transfer cells, to be more stable considering they have not recently been in transit.

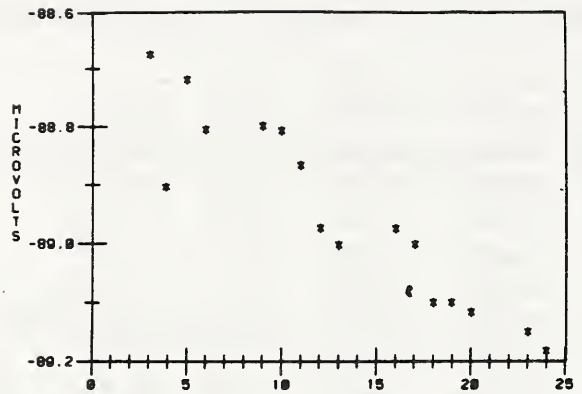


Figure 4
Values (μV) assigned to transfer standard T_1 versus time (days)

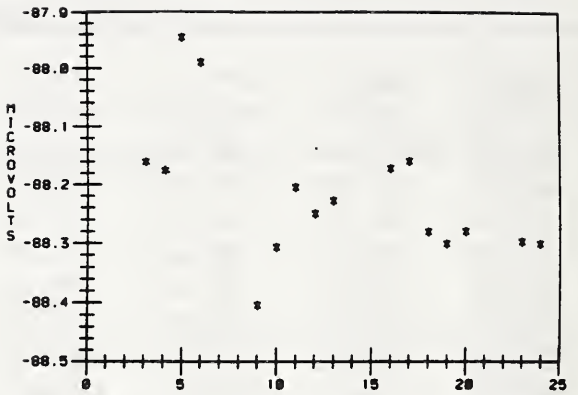


Figure 5
Values (μV) assigned to transfer standard T_2 versus time (days)

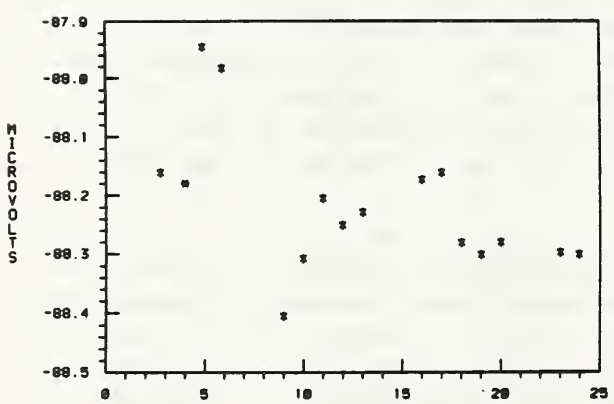


Figure 6
Values (μV) assigned to transfer standard T_3 versus time (days)

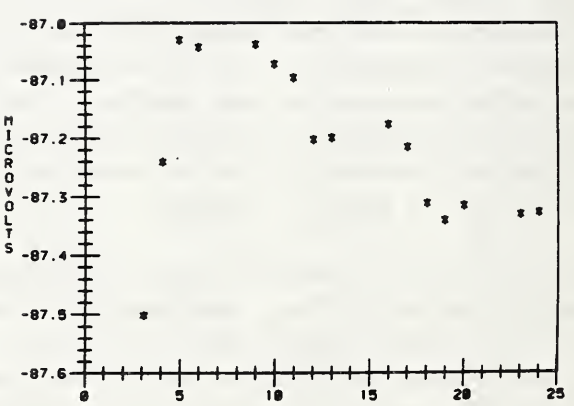


Figure 7
Values (μV) assigned to transfer standard T_4 versus time (days)

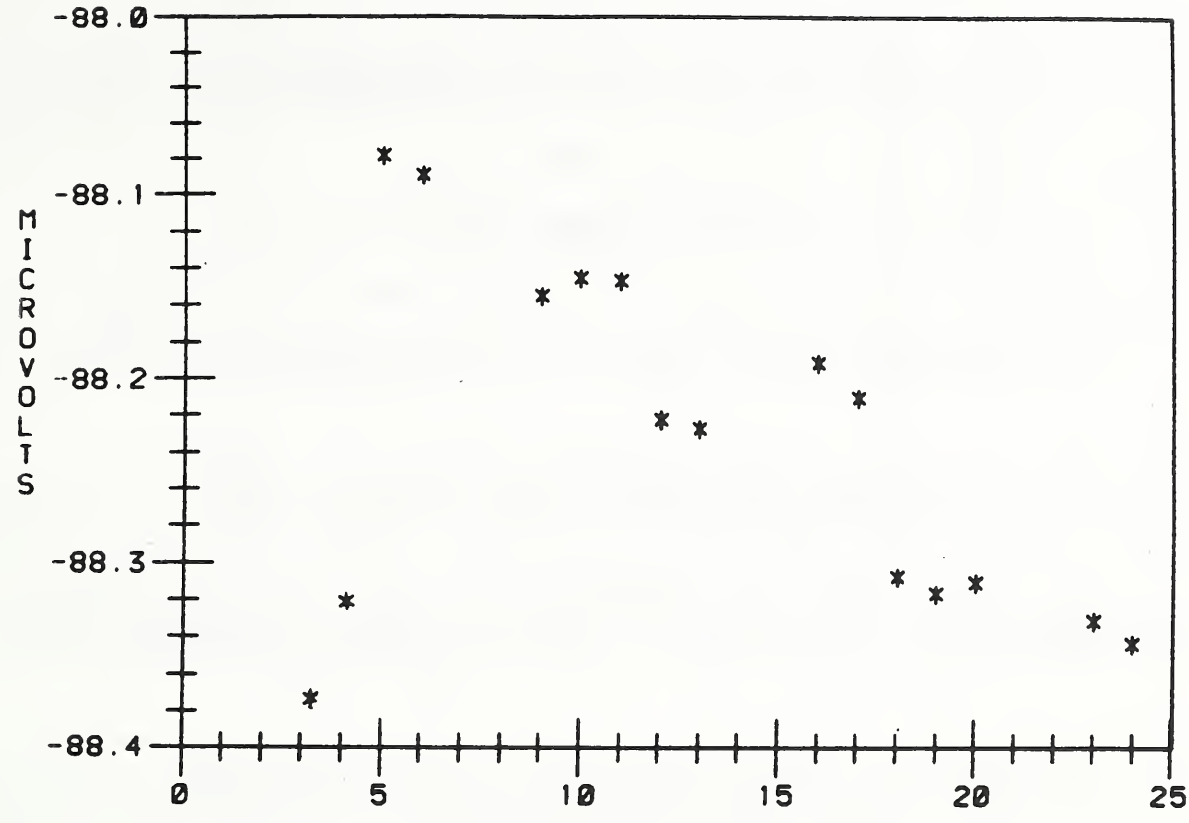


Figure 8
Average values (μV) assigned to four transfer standards versus time (days)

Each day's intercomparisons are analyzed for internal consistency via an F test on the within standard deviation for that day. The stability of the three designated check standards is also tested each day. Results of those designs which show evidence of lack of statistical control or anomalous behavior on the part of one of the check standards are excluded from the transfer experiment. Because we do not have prior history on this measurement process, we rely on hypothetical data to demonstrate to the reader the analysis that should be done for each design.

The left-right effects (4.5.2) are plotted in figure 9. Their respective test statistics (4.5.8) are listed in exhibit 4.5.3. Upper and lower control limits in figure 9 are indicated by dashed lines, and points that fall outside these control limits are equivalent to the corresponding test statistics being significant. These computations assume that prior data on the left-right effect established a standard deviation for the left-right effect of $s_\zeta = 0.02\mu V$ with $\nu_1 = 50$ degrees of freedom and that the accepted value of the left-right effect was established as $A_\zeta = 0.100\mu V$ from the same data.

Check standards C_1 and C_2 as constructed in (4.5.7) are observed differences between two reference cells and do not depend on the restraint or the design. Tracked over a period of time they show the way in which two cells are changing in respect to each other. Their values are listed in exhibit 4.5.3 and plotted as a function of time in figures 10-11.

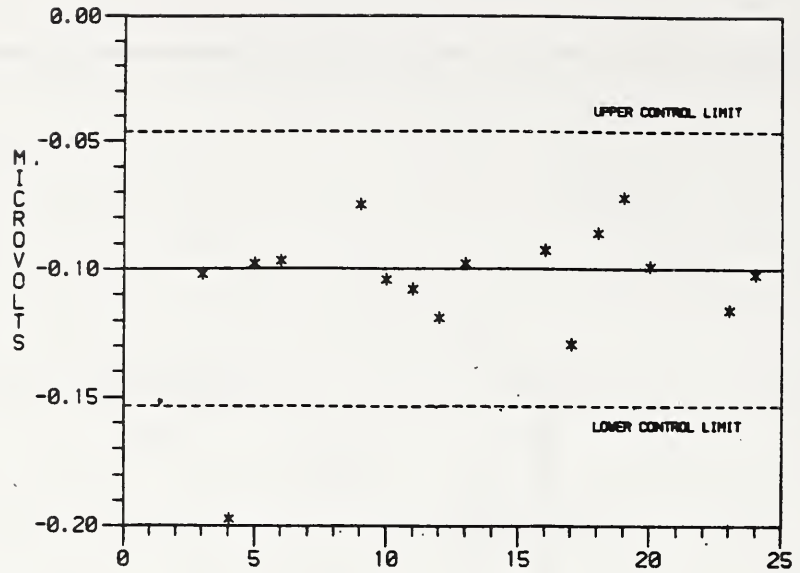


Figure 9
Left-right effect (μV) plotted against time (days) with dashed lines indicating upper and lower control limits at the 1% significance level

Exhibit 4.5.3 - check standards and test statistics[†]
Values in microvolts

Run	Date	Left-right Effect	Test Stat	Check Std	Test Stat	Check Std	Test Stat
#	t'	$\hat{\zeta}$	t_ζ	c_1	t_{c_1}	c_2	t_{c_2}
1	3	-0.102	0.1	-2.0450	0.22	-1.6075[¶]	2.90[¶]
2	4	-0.197[§]	4.8[§]	-2.0100	0.29	-1.5375	0.51
3	5	-0.098	0.1	-1.9650	1.10	-1.5175	0.29
4	6	-0.097	0.2	-1.9625	0.58	-1.5100	0.68
5	9	-0.075	1.2	-1.9775	1.66	-1.4775	2.16
6	10	-0.104	0.2	-1.9275	0.69	-1.5300	0.69
7	11	-0.108	0.4	-1.9075	0.66	-1.5725	0.46
8	12	-0.119	1.0	-1.8950	0.85	-1.5175	1.37
9	13	-0.098	0.1	-1.8900	1.27	-1.5300	1.14
10	16	-0.093	0.4	-1.7825	0.25	-1.5550	0.84
11	17	-0.129	1.4	-1.7525	0.58	-1.6675	2.35
12	18	-0.086	0.7	-1.7500	0.09	-1.5575	1.05
13	19	-0.072	1.4	-1.7450	0.32	-1.6775	2.31
14	20	-0.099	0.0	-1.6925	0.65	-1.5950	0.25
15	23	-0.116	0.8	-1.6575	0.01	-1.6125	0.18
16	24	-0.102	0.1	-1.6625	0.66	-1.6300	0.17

[†]We choose to illustrate the control procedure at the 1% significance level.
[§]Failed test for control at 1% level of significance based on a critical value $t_{.005}(50) = 2.678$ from Table I.
[¶]Failed test for control at 1% level of significance based on a critical value $t_{.005}(100) = 2.626$ from Table I.

For this analysis, it was assumed that data from fifty-one initial designs established a linear relationship with time for each check standard as follows:

$$c_1 = -2.095 + 0.0190t$$
$$c_2 = -1.501 - 0.00513t \qquad (4.5.25)$$

and that standard deviations, s_{c_1} for C_1 and s_{c_2} for C_2, were pooled to form a process standard deviation $s_c = 0.030\mu V$ with $\nu = 100$ degrees of freedom.

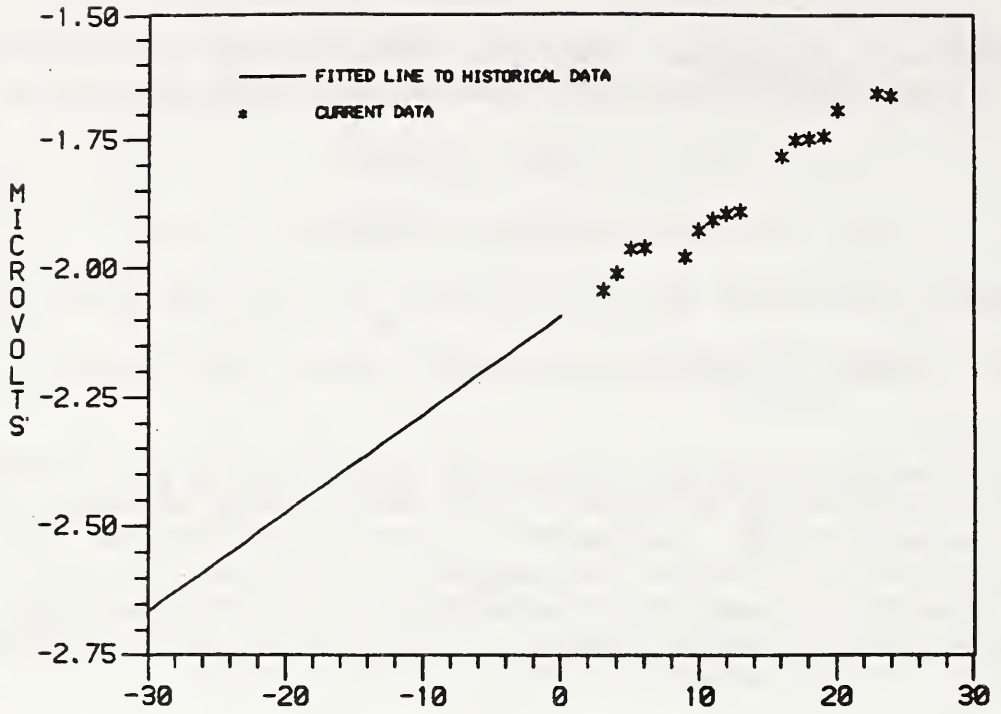
Based on the foregoing assumption, predicted values (4.5.11) for the check standards were computed for each time t' that the transfer standards were measured in the participant's laboratory. Given this information, the check standard measurements on each day were tested for agreement with the extrapolated line by the test statistics listed in exhibit 4.5.3. The test statistics for C_1 and C_2 that are shown in exhibit 4.5.3 were computed from (4.5.15) with n = 31 and values of $t_i(i=1,\cdots,31) = -30(1)0$.

The same analysis is shown graphically in figures 10-11. The upper portion of figure 10 shows the linear fit to the historical data as a solid line, and the values of check standard C_1 for the transfer experience are shown as discrete points, (*) with the convention that the transfer experiment starts at time t = 0.

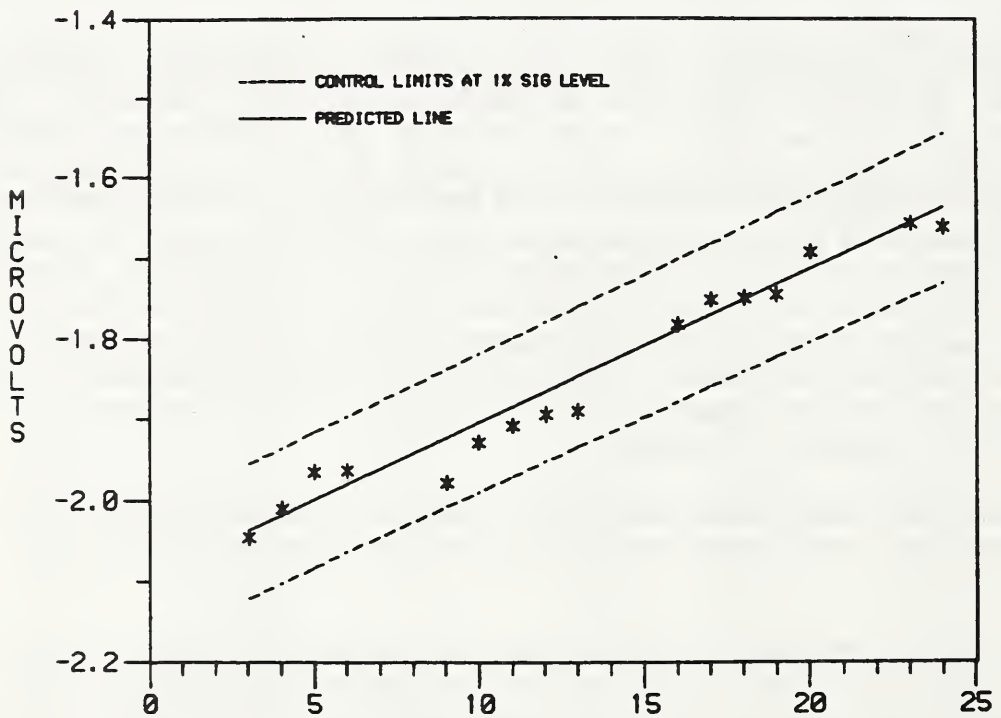
The lower portion of figure 10 shows the analysis of the check standard measurements. The solid line is an extrapolation of the linear fit from the upper portion of the same figure to the time of the transfer experiment. The dashed lines are upper and lower control limits that show the range within which the check standard measurements are expected to deviate from the extrapolated line. A point being outside these control limits is exactly analogous to the corresponding test statistic being significant in exhibit 4.5.3. Although it is not readily apparent from the graph, the control limits become wider as the check standard measurements are further removed in time from their historical data base. Thus, there is a smaller chance of detecting anomalous behavior as the experiments are continued into the future if the data base is not updated frequently.

Figure 11 shows the same analysis for the values of check standard C_2 from the transfer experiment with check standard C_2 out-of-control on the first day.

The within standard deviations are listed in exhibit 4.5.4 and plotted in figure 12. An F test based on an accepted standard deviation $s_p = 0.02\mu V$ with $\nu_3 = 408$ degrees of freedom indicates that there are measurement problems on the first and eleventh days. It is interesting to note that check standard C_2 is low on the eleventh day although it is not actually out-of-control and that the left-right effect is very close to being out-of-control on that same day. Given the responses of the check standards and the transfer standards and the information garnered from the control procedure, it would seem reasonable to delete three measurements from the transfer data; namely, the first, second and eleventh days' measurements.

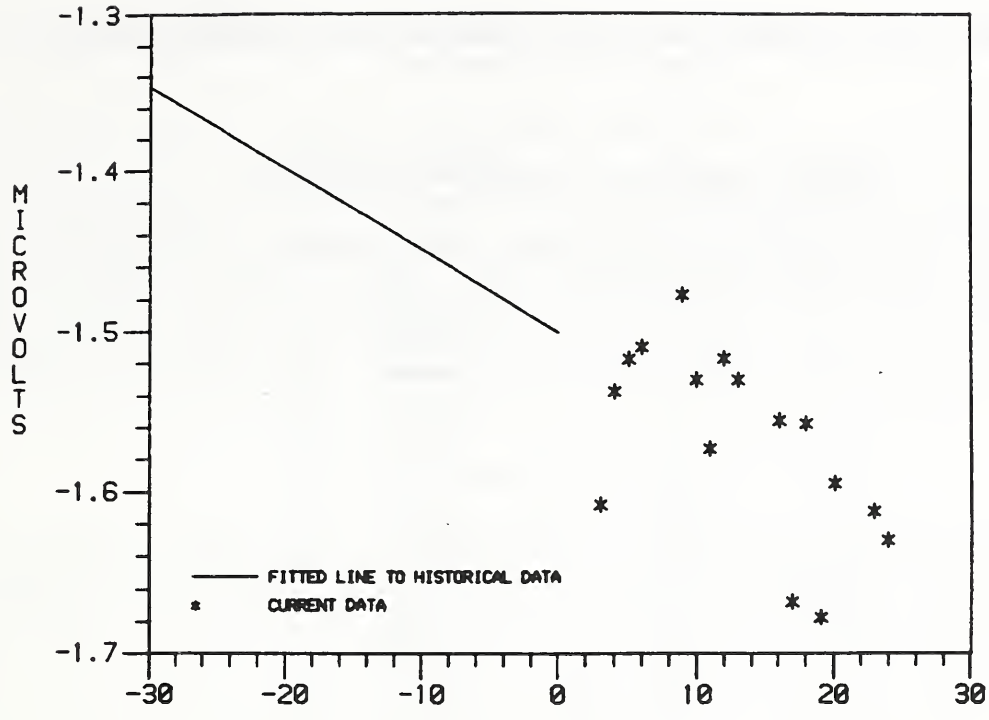


(a)

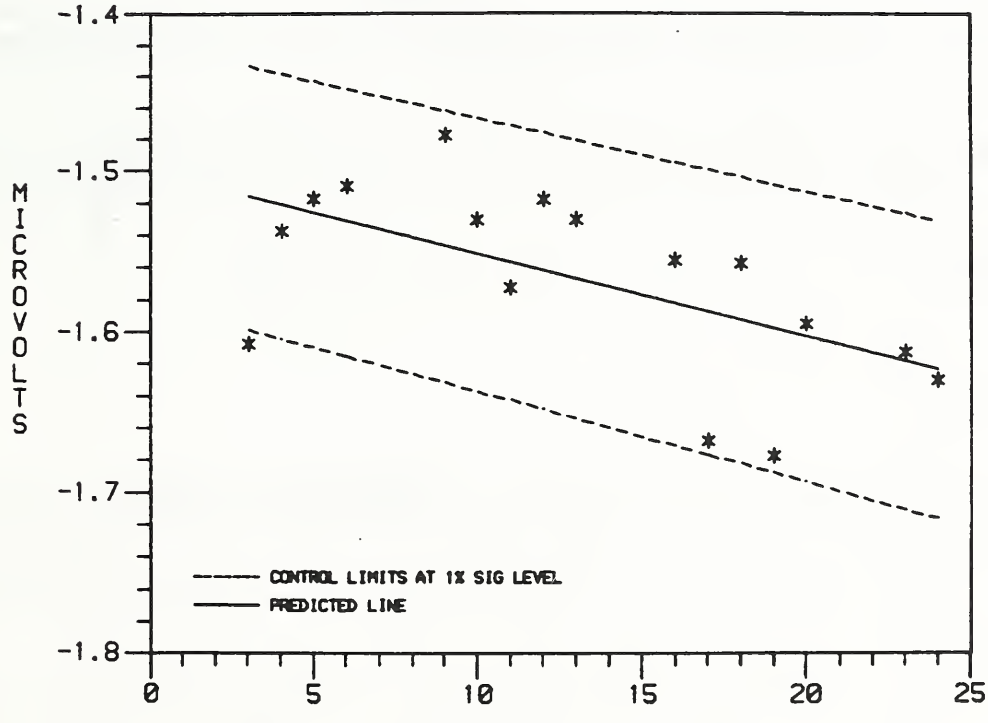


(b)

Figure 10
Check standard C_1 (μV) plotted against time (days)
with a solid line indicating a predicted linear fit and dashed lines
indicating upper and lower control limits at the 1% level of significance



(a)



(b)

Figure 11
Check standard C_2 (μV) plotted against time (days)
with a solid line indicating a predicted linear fit and dashed lines
indicating upper and lower control limits at the 1% level of significance.

Exhibit 4.5.4 - Within sandard deviations and test statistics
Values in microvolts

Run #	Date t	Within SD s_w	DF ν_3
1	3	0.054§	8
2	4	0.018	8
3	5	0.019	8
4	6	0.015	8
5	9	0.022	8
6	10	0.011	8
7	11	0.016	8
8	12	0.021	8
9	13	0.013	8
10	16	0.021	8
11	17	0.054§	8
12	18	0.018	8
13	19	0.031	8
14	20	0.013	8
15	23	0.018	8
16	24	0.011	8

§Failure to satisfy the inequality $s_w < s_p\sqrt{F_{.01}(8,\infty)}$ at the 1% significance level based on $s_p = 0.02\mu V$ and a critical value $F_{.01}(8,\infty) = 2.5$ from Table II.

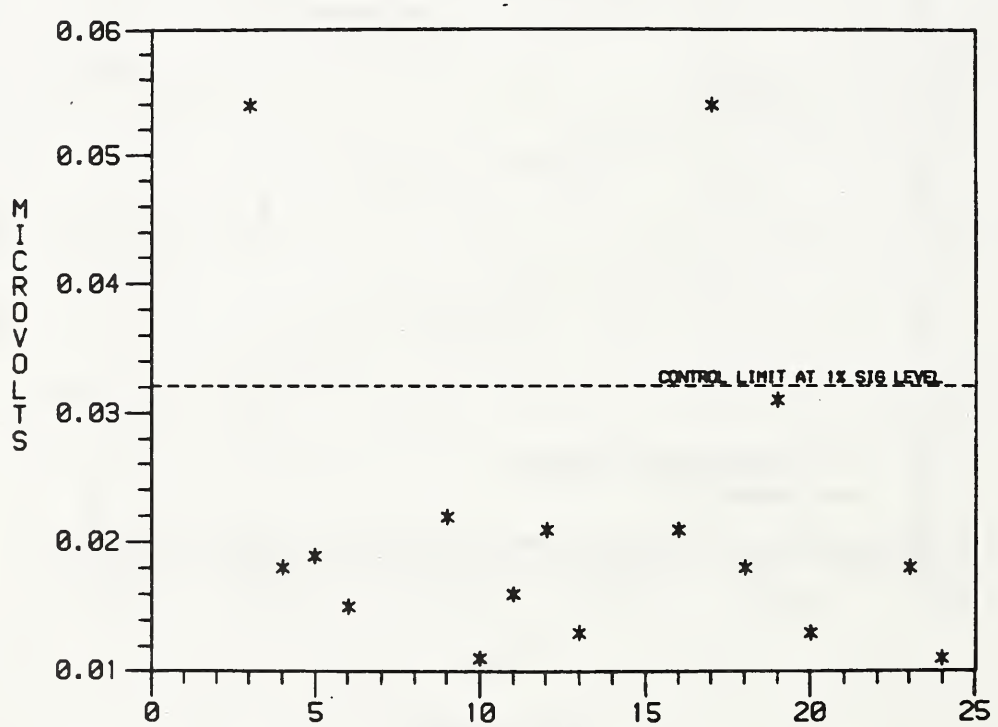


Figure 12
Within standard deviations (μV) plotted against time (days)
with dashed line indicating control limit at 1% level of significance.

4.6 Direct Reading of the Test Item with an Instrument Standard

4.6.1 Measurement Sequence

In this mode of operation a value is directly assigned to a test item X by a calibrated instrument. Observations on a stable artifact that takes on the role of the check standard C are used to establish a base line for the instrument and to maintain and control its variability in what amounts to a surveillance type test. An observation on the test item is denoted by x, and an observation on the check standard is denoted by c.

4.6.2 Process Parameters

Initial values of the process parameters are obtained from n independent measurements on the check standard $c_1, \cdots, c_n$. The accepted value of the check standard is defined by the mean of the check standard measurements; namely,

$$A_c = \frac{1}{n} \sum_{i=1}^{n} c_i \, . \tag{4.6.1}$$

The total standard deviation of the instrument is

$$s_c = \left(\frac{1}{n-1} \sum_{i=1}^{n} (c_i - A_c)^2 \right)^{1/2} \tag{4.6.2}$$

with $\nu = n-1$ degress of freedom. The control limits[n] that are appropriate for future observations on the check standard are given by

$$\text{Upper control limit} = A_c + 3s_c$$

$$\text{Lower control limit} = A_c - 3s_c \, .$$

4.6.3 Control Procedure

The primary purpose of the control procedure is to monitor instrumental drift, and observations on the check standard should be taken frequently enough to ensure that such drift is being contained. A test statistic t_c computed from the most recent check standard measurement c is given by

$$t_c = \frac{|c - A_c|}{s_c} \, .$$

The process is in control at the time of the check standard measurement c if

$$t_c < 3 \, . \tag{4.6.3}$$

[n]The factor 3 is used in this and all subsequent computations in place of the appropriate percent point of the t distribution; namely, $t_{\alpha/2}(\nu)$.

If

$$t_c > 3 ,$$

the process is not in control at the time of the check standard measurement, and measurements should be discontinued until the problem with the instrument is rectified.

4.6.4 Transfer with NBS

Determination of systematic error can be made by making p measurements $r_1,\cdots,r_p$ on a calibrated artifact or transfer standard which has an assigned value T^* and associated uncertainty U_T. Instrumental offset ψ defined by

$$\psi = \frac{1}{p}\sum_{i=1}^{p} r_i - T^* \qquad (4.6.4)$$

is not significant if

$$\frac{\sqrt{p}\,|\psi|}{s_c} < 3 . \qquad (4.6.5)$$

It is extremely important to recognize that this approach makes two important assumptions that must be verified experimentally; namely, that the instrument has a constant offset from the NBS process over the regime of interest as in (1.4.2) and that the precision of the instrument is constant over this same regime. The question of constant offset is considered first. A single point is not sufficient for the determination, and the system must be checked using several calibrated artifacts that span the regime of interest. Assume that m transfer standards are sufficient to verify the points of interest and that the transfer standards have assigned values $T_1^*,\cdots,T_m^*$ and associated uncertainties $U_{T1},\cdots,U_{Tm}$. Assume also that m offsets $\psi_1,\cdots,\psi_m$ computed according to (4.6.4) have been determined from measurements made on the transfer standards.

If all ψ_j $(j=1,\cdots,m)$ are insignificant as judged by (4.6.5), no adjustment to the instrument is needed. If the offsets are of varying magnitudes, and if it can be shown that these offsets are functionally related to the assigned values of the transfer standards, it may be possible to calibrate the instrument using a calibration curve based on the offsets (see section 2.3.3). Finally, if the offsets are significant and of the same magnitude, either the instrument is adjusted for the average offset

$$\bar{\psi} = \frac{\sum_{j=1}^{m} p_j \cdot \psi_j}{\sum_{j=1}^{m} p_j} \qquad (4.6.6)$$

where p_j $(j=1,\cdots,m)$ represents the number of measurements on the j^{th} transfer standard or a reading x on a test item x is reported as

$$X^* = x - \bar{\psi} .$$

The uncertainty of the transfer is

$$U_{tr} = \frac{3s_c}{\sqrt{p_1+\cdots+p_m}} + \frac{1}{m}\left(U_{T1}^2 + \cdots + U_{Tm}^2\right)^{1/2} . \qquad (4.6.7)$$

4.6.5 Uncertainty

The total uncertainty that is appropriate for one measurement made on a test item using the calibrated instrument is

$$U = U_{tr} + 3s_c . \qquad (4.6.8)$$

4.6.6 Process Precision.

The question concerning whether or not the precision of an instrument remains constant over a given regime can be addressed by comparing standard deviations from several levels in the regime. A familiar example is an electronic balance that is used over a large range of loads where the precision of the instrument may be load dependent. This assumption can be checked either with calibrated or uncalibrated artifacts.

Standard deviations with their associated degrees of freedom should be tabulated by load and inspected for consistency. It is possible to quote one uncertainty over the entire regime only if the precision is constant over all load levels; i.e., if these standard deviations are all of the same magnitude.

A visual inspection of the values may be sufficient for determining whether or not the standard deviations are of roughly the same magnitude in which case the standard deviations should be pooled using (2.2.3) and the uncertainty computed by replacing s_c in equations (4.6.7) and (4.6.8) with the pooled standard deviation.

If there is some question about the propriety of combining all the standard deviations, the largest standard deviation can be checked for agreement with the others using a test developed by Cochran [57]. A description of the test statistic and tables for deciding whether or not the largest standard deviation in a group is significantly different from the group are tabulated by Eisenhart [58].

If it is logical to assume that the precision of the instrument will vary with the magnitude of the quantity of interest, then a series of check standards should be established, one at each level of interest, with the estimate of process precision (4.6.2), the test for statistical control (4.6.3), and computation of uncertainty (4.6.8) being made at each level independently, thus begging the question of constant variability.

4.7 Simultaneous Measurement of a Group of Test Items and a Group of Reference Standards

4.7.1 Measurement Sequence

This scheme is appropriate for assigning values to individual test items or instruments relative to the average of a bank or group of reference standards, called the restraint R*, when all items including the standards are simultaneously subjected to the same stimuli such as a power source or a vacuum chamber. Assume there are m reference standards $R_1,\cdots,R_m$, and ℓ test items $X_1,\cdots,X_\ell$. One position in the configuration of test items should be reserved for a check standard Y, an artifact similar to the test items, where a reading on Y is always recorded along with the other readings.

Assume that a measurement sequence produces readings $r_1,\cdots,r_m$ on the standards, $x_1,\cdots,x_\ell$ on the test items and y on the check standards. The value that is recorded as the check standard measurement for one sequence is

$$c = y - \frac{1}{m}\sum_{i=1}^{m} r_i \ . \qquad (4.7.1)$$

In other words the measured difference between the artifact check standard and the average of the reference standards is the check standard measurement. In the remainder of this section, the term check standard refers to this recorded difference rather than the measured value y.

4.7.2 Process Parameters

Initial values of the process parameters are obtained from n such measurement sequences where $c_1,\cdots,c_n$ are the check standard measurements.

The accepted value of the check standard is the mean of these values; namely,

$$A_c = \frac{1}{n}\sum_{i=1}^{n} c_i \ . \qquad (4.7.2)$$

The total standard deviation of the check standard is

$$s_c = \left(\frac{1}{n-1}\sum_{i=1}^{n} (c_i - A_c)^2 \right)^{1/2} \ . \qquad (4.7.3)$$

Control limits[q] that are appropriate for future check standard observations are given by

$$\text{Upper Control Limit} = A_c + 3s_c$$

$$\text{Lower Control Limit} = A_c - 3s_c \ .$$

[q]The factor 3 is used in this and all subsequent computations in place of the appropriate percent point of the t distribution; namely, $t_{\alpha/2}(\nu)$.

The control procedure applied to each calibration depends on a test statistic t_c computed from the value of the check standard c for that measurement sequence by

$$t_c = \frac{|c - A_c|}{s_c} . \qquad (4.7.4)$$

If

$$t_c < 3 \qquad (4.7.5)$$

the process in control, and the value of a test item is reported as

$$X_j^* = x_j - \frac{1}{m} \sum_{i=1}^{m} r_i + R^* \qquad j=1,\cdots,\ell \qquad (4.7.6)$$

where $R^* = \frac{1}{m} (R_1^* + \cdots + R_m^*)$ and $R_1^*,\cdots,R_m^*$ are the values assigned to the reference standards. If

$$t_c > 3$$

the calibration of the test items is invalid and must be repeated.

4.7.3 Transfer with NBS

The transfer with NBS is accomplished by p repetitions of the measurement sequence during which a group of ℓ transfer standards $T_1,\cdots,T_\ell$ replaces the group of test items. Process control as defined by (4.7.5) should be confirmed for each repetition. Any sequence that is out-of-control should be repeated until control is restored or else that repetition is deleted from the transfer. The values assigned the transfer standards are $T_1^*,\cdots,T_\ell^*$ with uncertainties $U_{T1},\cdots,U_{T\ell}$.

The offset Δ_i $(i=1,\cdots,p)$ of the laboratory process from NBS for the ith repetition is based on the values assigned to the ℓ transfer standards by (4.7.6); namely, $X_1^*,\cdots,X_\ell^*$ where

$$\Delta_i = \frac{1}{\ell} \sum_{j=1}^{\ell} (X_j^* - T_j^*) \qquad i=1,\cdots,p$$

and the average offset computed for the p repetitions is

$$\overline{\Delta} = \frac{1}{p} \sum_{i=1}^{p} \Delta_i . \qquad (4.7.7)$$

The uncertainty of the transfer is

$$U_{tr} = \frac{3s_c}{\sqrt{p\ell}} + \frac{1}{\ell} \left(U_{T1}^2 + \cdots + U_{T\ell}^2 \right)^{1/2} . \qquad (4.7.8)$$

The offset is judged significant if

$$\frac{\sqrt{p\ell}\,|\Delta|}{s_c} > 3 \qquad (4.7.8)$$

and in such case the assigned value of the restraint is changed to $R^* - \Delta$. The restraint is unchanged if

$$\frac{\sqrt{p\ell}\,|\Delta|}{s_c} < 3.$$

4.7.5 Uncertainty

The total uncertainty that is appropriate for a value assigned to a test item by (4.7.6) from one calibration is

$$U = U_{tr} + 3s_c. \qquad (4.7.9)$$

4.8 Ratio Technique for One or More Test Items and One or Two Reference Standards

4.8.1 Measurement Scheme

In this section we describe calibration of a test item X by an instrument such as a scanning electron microscope which has only short-term stability. Consider the case where the test item X and the reference standard R are related by (1.4.9) and the instrument response is of the form (1.4.10). One reference standard R is sufficient to provide a calibrated value X^* for the test item given a single reading x on the test item and a single reading r on the reference standard. The calibrated value is

$$X^* = x \cdot R^*/r \qquad (4.8.1)$$

where R^* is the value assigned to the reference standard.

Where the test item and reference standard are related by (1.4.1) and the instrument response is of the form (1.4.6), two reference standards R_1 and R_2 are needed to calibrate a test item X (Cameron [60]). The artifacts should be measured in the sequence R_1, X, R_2 with the corresponding measurements denoted by r_1, x, r_2. The calibrated value for the test item is

$$X^* = R_1^* + \frac{(R_2^* - R_1^*)\cdot(x - r_1)}{(r_2 - r_1)} \qquad (4.8.2.)$$

where R_1^* and R_2^* are the values assigned to R_1 and R_2 respectively.

If before and after readings are taken on the test item in the sequence X, R_1, R_2, X with the measurements denoted by x_1, r_1, r_2, x_2 respectively, then the calibrated value for the test item is

$$X^* = \frac{1}{2}\left\{ (R_1^* + R_2^*) + \frac{(R_2^* - R_1^*)\cdot(x_1 - r_1 - r_2 + x_2)}{(r_2 - r_1)} \right\} . \quad (4.8.3)$$

More than one unknown can be calibrated from the same pair of readings on R_1 and R_2 only if the sequence of measurements can be arranged so that no test item is too far removed from R_1 and R_2 in the measurement scheme. For example, for test items X, Y, and Z, the sequence X, R_1, Y, R_2, Z minimizes the separation between unknowns and standards, and the calibrated value for each unknown is calculated according to (4.8.2).

In practice, it may be necessary to have several artifact standards that cover the operating range of the instrument. In addition to artifact standards for every level, it is necessary to have one artifact check standard Y for every level. A measurement y on the check standard should be included in the calibration program on a regular basis, and if feasible, with every calibration scheme. The check standard value that is used for controlling the process and for estimating random error is computed in exactly the same way as X^*. For example, for the measurement sequence described by (4.8.2), the check standard value from one calibration is

$$c = R_1^* + \frac{(R_2^* - R_1^*)\cdot(y - r_1)}{(r_2 - r_1)} . \quad (4.8.4)$$

4.8.2 Process Parameters

Initial values of the process parameters are obtained from n such calibration sequences yielding check standard values $c_1, \cdots, c_n$. The accepted value of the check standard is defined as the mean of the check standard values; namely,

$$A_c = \frac{1}{n}\sum_{i=1}^{n} c_i . \quad (4.8.5)$$

The total standard deviation of the check standard is defined by

$$s_c = \left(\frac{1}{n-1} \sum_{i=1}^{n} (c_i - A_c)^2 \right)^{1/2} \quad (4.8.6)$$

with $\nu = n-1$ degrees of freedom.

In this case s_c is the standard deviation of a calibrated value X^* and will reflect not only the imprecision in the measurements x, r_1, and r_2 but also any changes in the response curve for the instrument that are not accounted for by the ratioing device.

The control limits[r] that are appropriate for future check standard values are:

$$\text{Upper control limit} = A_c + 3s_c$$

$$\text{Lower control limit} = A_c - 3s_c \ .$$

4.8.3 Control Procedure

A control procedure is applied to each calibration sequence which includes a check standard measurement. The control procedure is based on a test statistic t_c computed from the check standard value c for that sequence; namely,

$$t_c = \frac{|c - A_c|}{s_c} \ .$$

If

$$t_c < 3 \qquad (4.8.7)$$

the process is in control, and the value of a test item X is reported as X^*.

If

$$t_c \geq 3,$$

the process is out-of-control, and the calibration of the test item is invalid and must be repeated.

4.8.4 Transfer with NBS

The tie to NBS is via the reference standards which are either standard reference materials from NBS or secondary calibrated artifacts.

4.8.5 Uncertainty

The uncertainty for an artifact calibrated according to (4.8.1) is

$$U = 3s_c + U_R \qquad (4.8.8)$$

where U_R is the uncertainty for R^*. The uncertainty for an artifact calibrated according to (4.8.2) or (4.8.3) is

$$U = 3s_c + \frac{1}{2}\left(U_{R1}^2 + U_{R2}^2\right)^{1/2} \qquad (4.8.9)$$

where U_{R1} and U_{R2} are the uncertainties for R_1^* and R_2^* respectively.

---

[r]The factor 3 is used in this and all subsquent computations in place of the appropriate percent point of the t distribution; namely, $t_{\alpha/2}(\nu)$.

5. Control Charts

5.1 Introduction

The industrial application of control charts involves a production process that yields product that is assumed to be homogeneous with respect to a particular property that is measurable. The control chart is devised to detect any variation in the production process that is not random in nature and which, therefore, can be assigned a cause. Guaranteeing that all variation in the production process is random in nature guarantees that the process is operating in an optimal fashion, and if, given these circumstances, the product is not within specifications, major adjustments to the process are required in order to substantially affect its output.

Once a base line and control limits have been defined for the process, based on prior data from the same process, the control chart is set up with a solid horizontal line representing the base line and dashed lines above and below the base line representing the control limits. Samples drawn at random from the production process are measured for the property of interest, and the resulting values are plotted on the control chart as a function of time. Values that fall within the control limits are referred to as being "in statistical control" and values that fall outside the control limits are referred to as being "out of control". Values outside the control limits are a sufficient indication that the "process should be investigated and corrected" (Bicking & Gryna [61]).

The Shewhart control chart discussed above is appropriate for individual measurements or averages of "natural" groups. This type of control chart, used in conjunction with a control chart for standard deviations, is a powerful means of detecting changes in the measurement process. Other types of control procedures include a cusum chart (Duncan[62]) which is particularly useful for detecting gradual drifts in a continuous process as compared with abrupt shifts. Methods for detecting changes in both the base line of the process and in the variability of the process on a single control chart are discussed by Reynolds and Ghosh in reference [63].

Statistical control as originated by Shewhart [64] assumes that repeated measurements of a reproducible property are available and that these measurements constitute a random sample of all such possible measurements from a known distribution such as the normal distribution. The term random sample implies two important properties of the measurements; namely, that they are independent and that they all come from the same distribution. The average value and standard deviation calculated from a random sample in conjunction with known properties of the distribution are used to calculate limits within which a certain percentage of all measurements should fall. In other words, a series of initial measurements are made to characterize the distribution of all possible measurements, and future measurements are checked for conformity with this distribution.

Notice that one is not concerned with whether or not the product is within certain specification limits, but rather with whether or not the production process is behaving properly. The control procedure for a measurement process is similar in many respects to industrial control. In the measurement

assurance context the measurement algorithm including instrumentation, reference standards and operator interactions is the process that is to be controlled, and its direct product is measurement per se. The measurements are assumed to be valid if the measurement algorithm is operating in a state of control; i.e., if the variations in that process are due to random causes which can be quantified, thus assuring that a value reported by the process will have negligible offset from national standards within predictable limits. This will be the case if the control chart shows that the base line for the process is not changing.

Statistical control in the measurement assurance context can conversely be predicated on the assumption that the measurement process is stable and that lack of control indicates a change in the artifact being measured. There are circumstances where this type of control is needed--that is, when it is necessary to know whether or not an artifact has changed with respect to the property being measured. For example, a transfer standard that is being circulated to several laboratories must be checked periodically at NBS. Similarly, intercomparisons between working standards and primary standards can be subjected to a control procedure to ensure that the working standards have not changed appreciably. In these instances, lack of control will result in either replacing the artifact in question or in reassigning its accepted value.

Calibration control is perhaps dissimilar to industrial control in that although artifacts submitted for measurement are of the same general type, their properties must be quantified individually. Thus, there is an inherent problem in controlling the values assigned to individual artifacts or instruments because the measurement is rarely repeated, let alone repeated sufficiently often to characterize the distribution of possible values. Without a historical data base there is no way of determining whether or not the current calibration is in control or is, in fact, a proper assignment for the item. For this reason a check standard is introduced into the measurement sequence in such a way that it can be assumed that the measurement algorithm acts on the check standard in much the same way as it acts on the item being calibrated. The redundant measurements on the check standard are the basis for both characterizing the distribution of measurements and deciding if the measurement process is in control on a given occasion.

The control limits are chosen so that the probability is 100α percent that future measurements will fall outside the control limits strictly by chance. Therefore, α if always chosen small, say $\alpha = .01$ or $\alpha = .05$ so that very few measurements will be discarded unnecessarily. Smaller values of α correspond to wider control limits which result in the measurement almost always being accepted unless there is a serious shift in the process. The converse is also true--larger values of α correspond to narrower control limits which result in tighter control of the measurement process with more frequent remeasurement. Obviously, the success that can be expected in detecting changes in the process which is referred to as the power of the control procedure is linked to the choice of α.

The reader may have already noted that the procedure for determining control or lack thereof is exactly analogous to a statistical t-test for deciding whether or not a single observation comes from a process with known mean and unknown standard deviation.

5.2 Control Charts for Single Measurements

The measurements for initiating the control chart must be collected over a sufficiently wide range of operating conditions to ensure a correct characterization of the distribution and over a sufficently long period of time to ensure independence. Grant and Leavenworth state that ideally twenty-five measurements should be spread over several months time [65]. As few as ten or fifteen measurements can suffice if this data base is updated when more measurements are available. The measurements are plotted as a function of time without imposing a base line or control limits on the plot in order to track the measurement process and verify that it produces stable measurements whose variability is random in nature. Such a plot also allows one to check specification limits, but specification limits do not constitute statistical control because they do not have a probabilistic interpretation.

When one is satisfied that the initial measurements are adequate for representing the distribution and that process variability is tolerable, a base line and control limits are computed from this data base.

For single measurements the base line is taken to be the average of initial measurements $x_1, \cdots, x_n$; namely

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{5.2.1}$$

and the control limits are taken to be

$$\begin{aligned} &\bar{x} + s \cdot t_{\alpha/2}(\nu) \\ &\bar{x} - s \cdot t_{\alpha/2}(\nu) \end{aligned} \tag{5.2.2}$$

where s, the total standard deviation computed from the initial measurements is

$$s = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{1/2} \tag{5.2.3}$$

with $\nu = n-1$ degrees of freedom. The number $t_{\alpha/2}(\nu)$ is the $\alpha/2$ percentage point of Student's t distribution with ν degrees of freedom.

Once the average value and the control limits have been established, future measurements are tested for control. One concludes that measurements that fall within the control limits come from the hypothesized distribution, and that, therefore, the measurement process is acting in an acceptable and predictable manner. The converse is also true. Measurements that fall outside the control limits infer a significant change in the process. Where such a change is noted, one must determine whether the change is permanent or transitory.

In a measurement assurance context, every violation of the control limits requires a remedial action. It may be sufficient to simply repeat the offending measurement in order to reestablish control, but all measurements since the last successful test for control are discarded once an out-of-control condition occurs.

As an example, consider how repeated measurements on a calibrated weight can be used to demonstrate that an electronic balance is, indeed, weighing accurately at all times. Accuracy in this context means that values delivered by the balance are in agreement with national standards (prototype kilogram) as maintained by NBS within the stated uncertainty. Parobeck et al [66] describe a measurement assurance program for large volume weighings on electronic balances where redundancy and control are achieved by repeating weighings of selected test items on different days.

A program to control a weighing process is begun by making n initial measurements on the calibrated weight, being sure to allow enough time between successive measurements to cover a range of operating conditions in the laboratory, and using these initial measurements as a historical base for computing the average $\bar{x}$ and the standard deviation s of the balance.

Given a calibrated value A with uncertainty U_A for the weight, the balance is accurate within the uncertainty $U_A \pm s \cdot t_{\alpha/2}(n-1)$ if

$$A - \frac{s \cdot t_{\alpha/2}(n-1)}{\sqrt{n}} - U_A < \bar{x} < A + \frac{s \cdot t_{\alpha/2}(n-1)}{\sqrt{n}} + U_A.$$

Notice that this test takes into account both the limits to random error for the measurement process, $\pm s \cdot t_{\alpha/2}(n-1)/\sqrt{n}$, and the uncertainty associated with the calibrated value of the weight, U_A.

Once the accuracy has been verified, the control phase of the program is pursued by remeasuring the weight from time to time. The resulting values are plotted on a control chart having base line and control limits as defined in equations (5.2.1) and (5.2.3), and it is presumed that the balance continues to be accurate as long as

$$\bar{x} - s \cdot t_{\alpha/2}(n-1) < y_i < \bar{x} + s \cdot t_{\alpha/2}(n-1)$$

for all future measurements y_i.

There is always a question, in this type of application, of how often one should check for control. It seems obvious, particularly if one is dealing with electronic instrumentation, that there should always be a check for control as part of any start-up procedures. After that, the frequency is dictated by the past performance of the system and by the amount of inconvenience and expense that is generated when an out-of-control condition is encountered--keeping in mind that when the balance is found to be out-of-control, it is necessary to recall all the measurements that were made on that balance since the previous successful check for control.

5.3 Control Charts for Averages or Predicted Values

Thus far, the discussion has centered on control charts for individual measurements, and it is easily extended to include control charts for averages that are completely analagous to the control charts for individual measurements. When the reported value of a measurement sequence, be it an average or a predicted value from a least-squares analysis, is computed from k intercomparisons that were made over a relatively short period of time, the "measurement of interest" is the corresponding average or predicted value of the check standard. This quantity is treated analogously to a single measurement with base line and control limits for the control chart determined from n such initial quantities. That is, given check standard values $x_1,\cdots,x_n$ each of which is an average or predicted value from k intercomparisons, the grand mean $\bar{x}$ computed from (5.2.1) represents the base line of the process and control limits as in (5.2.2) can be calculated using the total standard deviation s from (5.2.3). In this case the quantity s is the standard deviation of an average or predicted value and not the standard deviation of a single measurement from the process.

5.4 Control Charts for Within Standard Deviations

For a measurement scheme involving k intercomparisons, it is possible to generate a control chart for what is called the "within" or short-term variability of the process.

Assume that each check standard value x_i (i=1,$\cdots$,n) is the result of k intercomparisons; namely, $x_{i1},\cdots,x_{ik}$ where the quantity x_i is the average of these intercomparisons,

$$x_i = \frac{1}{k} \sum_{j=1}^{k} x_{ij} \quad .$$

The within standard deviations are estimated by

$$s_{w_i} = \left(\frac{1}{k-1} \sum_{j=1}^{k} (x_{ij} - x_i)^2 \right)^{1/2} \qquad (5.4.1)$$

with degrees of freedom $\nu_i = k-1$. Where the intercomparisons form a statistical design, the quantity x_i and the within standard deviation are computed from a least-squares analysis.

The base line and limits for controlling short-term process variability make use of the same intercomparisons that were used to establish the control chart for averages. The base line is the pooled within standard deviation

$$s_p = \left(\frac{\nu_1 s_{w_1}^2 + \cdots + \nu_n s_{w_n}^2}{\nu_1 + \cdots + \nu_n} \right)^{1/2} \qquad (5.4.2)$$

The degrees of freedom $\nu = \nu_1 + \cdots + \nu_n$ allow for a different number of degrees of freedom in each estimate of the within standard deviation in (5.4.1). If all measurement schemes contain the same number of intercomparisons, say k, then $\nu = n(k-1)$.

Because a standard deviation is a positive quantity, it is only necessary to test against an upper limit in order to test the short-term variability. Thus for any future measurement sequence involving k intercomparisons, the within standard deviation s_w is computed as in (5.4.1) and is said to be in-control if

$$s_w < s_p \sqrt{F_\alpha(k-1,\nu)} \qquad (5.4.3)$$

where $F_\alpha(k-1,\nu)$ is the upper α percent point of the F distribution with k-1 degrees of freedom in the numerator and ν degrees of freedom in the denominator.

The control chart for averages used in conjunction with the control chart for within standard deviations is a powerful means of detecting changes in the process. The two control procedures are evoked simultaneously, and if an out-of-control condition is encountered for either test, the process is assumed to be out-of-control and the measurement sequence is repeated.

5.5 Alternative Control Limits

The reader may be familiar with control charts with control limits computed as the product of the total standard deviation and a fixed multiplicative factor, such as two or three, instead of the appropriate percentage point of the F or t-distribution. Control charts for within standard deviations should always be based on the F distribution because the critical values of the F distribution change rapidly with changes in degrees of freedom.

The consideration of whether a control chart for averages should be based on the percentage points of Student's t distribution or on a fixed multiplicative factor, such as three or two, is really a matter of choice depending on the level of control that one is hoping to achieve and on the type of measurements that are in question. The use of Student's t distribution is the most rigorous test if the measurements truly represent a random sample from a normal distribution. It allows a strict probability interpretation of the control procedure.

It cannot always be shown, and indeed is not always the case, that measurements come from an idealized distribution such as the normal distribution. If one looked at a large number of measurements on the same item, they might come from a distribution that is slightly skewed; i.e., for example, extreme large values may be more likely than extreme small values.

> The problem of deciding whether to use limits based on the normal distribution, those based on some other distribution, or those which involve no assumption about the form of the distribution is one which, though of a kind common in applied statistics, has no satisfactory solution. Limits based on the normal distribution are substantially shorter for a fixed sample size than those based on no assumption about the distribution, but they may be irrelevant if the distribution is too far from normal. (Bowker [67]).

For this reason it is customary in the United States to use plus or minus three standard deviations as the control limits (Duncan [68]). The factor three guarantees that a large proportion of the distribution is covered for measurements coming from any distribution that is close to the normal distribution in character. These limits are robust, and should be used when the intent is to identify measurements that are clearly out-of-control. Because these limits are so wide, an out-of-control finding is almost certainly an indication of a serious malfunction in the measurement process. If a somewhat tighter control is desired, two standard deviation limits can be considered. Very few values will fall between the two and three standard deviation limits, and the price of remeasuring for those few may be worth the added degree of control.

5.6 Control Charts for Drifting Check Standards

Another consideration concerns the problem of drifting check standards and whether or not they can be used for control purposes. The assumption is made in most measurement control programs that the check standard is stable and that any change that is noted by the control procedure is caused by changes in the measurement process itself. Obviously if the check standard is not completely stable, the ability to detect a change in the process is confounded with any possible drift in the check standard.

Unfortunately the situation in reality is that artifacts may not be completely stable, and this instability will be detected when it is large compared to the process precision. Changes in check standards over time can be expected. Of the forty or more check standards that are in continual use in the NBS mass calibration program, only about half of those standards are completely stable or do not show any drift over time. The question is, "Can a drifting check standard be used for control purposes?" Sometimes it can, but a drifting check standard causes complications in the analysis when, depending on the rate of change, the control limits pick up this change.

There are a few ad hoc procedures that can be used in lieu of a rigorous approach to this problem. Probably the simplest approach is to determine the time interval over which the check standard is stable by studying historical data and to enforce the control procedure over this interval. When this time interval has elapsed or when numerous values have been flagged as being out-of-control, the base line and control limits can be adjusted based on more recent measurements on the check standard.

It the check standard is changing steadily, as is the case for many artifact standards at NBS, it is sometimes possible to model the rate of drift and to predict from this model a value for the check standard at a future time that is not too far removed from the present. This involves fitting a regression equation to the measurements as a function of time by the method of least-squares and computing the values of the check standard for future times. Then the control procedure is time dependent; the base value is the predicted value from the regression equation at that time, and the control limits which depend on the standard deviation of this predicted value become wider with time. This approach has been used at NBS for check standards with linear drift rate as a function of time. It can work reasonably well as long as the drift remains linear, but the cause of a breakdown in the linearity assumption cannot be easily identified because it is never really possible to separate the change in the artifact from the change in the process. In such a situation it is imperative that the process be checked frequently for offset by comparison to a national standard or to other stable laboratory standards.

5.7 Synopsis and Examples of Control Charts

Four important ideas that are pertinent to calibration programs should emerge from the disucssion thus far. First, when dealing with statistical control of the properties of an artifact or statistical control of a measurement process, the control parameters are not imposed upon the process externally but are characteristic of the measurement process itself as described by historical data.

Secondly, if the check standard measurement is outside the established control limits, the calibration sequence is presumed to be out-of-control, and the calibrations of the test items are considered invalid. When such a condition is initially encountered, the instrumentation can be checked and the measurement sequence repeated--testing again for control. Any intervening results should be discarded. If control cannot be restored, a significant change has occurred in the process, and this change must be investigated. If a process is repeatedly out-of-control, the base line and control limits should be reestablished based on more recent data.

If the check standard measurement is in-control, this is taken as evidence that the process is behaving as expected in relation to the item submitted for calibration, and its assignment is assumed to be correct. Lack of control is certainly grounds for rejecting the calibration of the test item, but the complimentary argument is not as strong. The relationship between the measurement on the test item and the measurement on the check standard must be interrelated or executed very close together in time in order to be satisfied that the assignment ot the check standard has, indeed, been done properly.

Thirdly, the process precision is very well characterized by a total standard deviation calculated from measurements on the check standard. In some cases, such measurements provide the only way of obtaining a realistic estimate of this source of uncertainty. Fourthly, even though the tests for control can be automated, it is not only advantageous to visually examine the control charts in order to detect anomalies or slight shifts in the process and possible drifting of the check standard over time, but it is essential for understanding the long term behavior of the measurement process.

In order to demonstrate the value of such critical examinations, four examples that have been encountered in NBS measurement assurance programs are discussed.

The National Bureau of Standards maintains control charts on about forty check standards that are used in the mass calibration program. The control chart shown in figure 13 depicts values of the one kilogram check standard as it has been estimated from the measurement sequence used in the calibration workload for one kilogram weights. The three standard deviation limits shown by the dashed lines are the control limits that are used for this program, and if one compares these limits with the two standard deviation limits shown by the dotted lines, it is apparent that very few points fall between the two sets of limits. It can also be noted that the two standard deviation control limits are almost identical with control limits based on student's t distribution at significance level $\alpha = 0.01$ when the number of points is large as in this case.

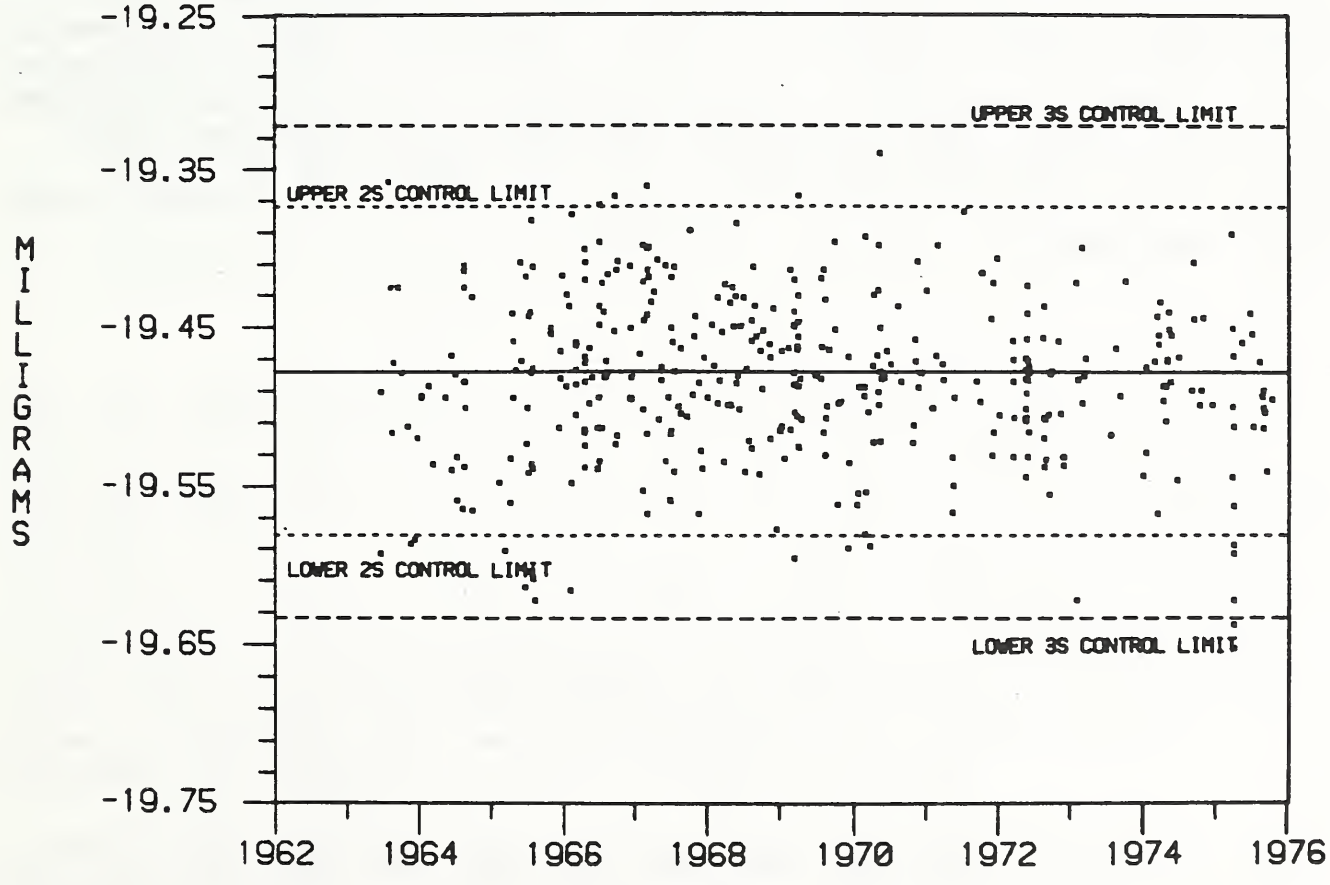


Figure 13
Check standard #41 (mg) as measured on NBS balance #4
plotted against time (years)

At this point the reader should be sufficiently sensitized to this approach to be aware of one shortcoming in this control chart. The chart implies that the process, which is demonstrably in-control, has never been out-of-control. A few points should fall outside of the control limits merely by chance, and as it happens other out-of-control situations have occurred in this program over the years. In fact, the control procedure would serve no useful purpose in the calibration program if there were no out-of-control situations to be detected. Actually this graph represents only the successful tests for control that were made with the one kilogram check standard because the calibration results and the check standard values were automatically discarded whenever the control limits were violated. The software for the NBS mass calibration program has been changed so that all values of the check standard are retained, and each value is flagged as to whether or not it was in control on that occasion. One should know when and how often control limits have been violated, and control charts should contain all findings.

The short-term or within variability of the same process is charted in figure 14 which shows within standard deviations for calibration sequences involving all weights calibrated on NBS balance #4. A calibration sequence typically requires between three and fifteen measurements, and the within standard deviation that is calculated from each sequence reflects the inherent variability of the balance and the effect of any environmental changes that occur during the time needed to make the requisite measurements. The base line for this control procedure, shown by the solid line, is the pooled within standard deviation in (5.4.2). Because the number of degrees of freedom varies with the design, it is not possible to establish a single upper control limit for this process; the control limit for each point is calculated separately, and the control procedure is automated using the control limit based on the F distribution as shown in (5.4.3). Once a year the within standard deviations are plotted to see if any degradation has occurred in the balance over the year.

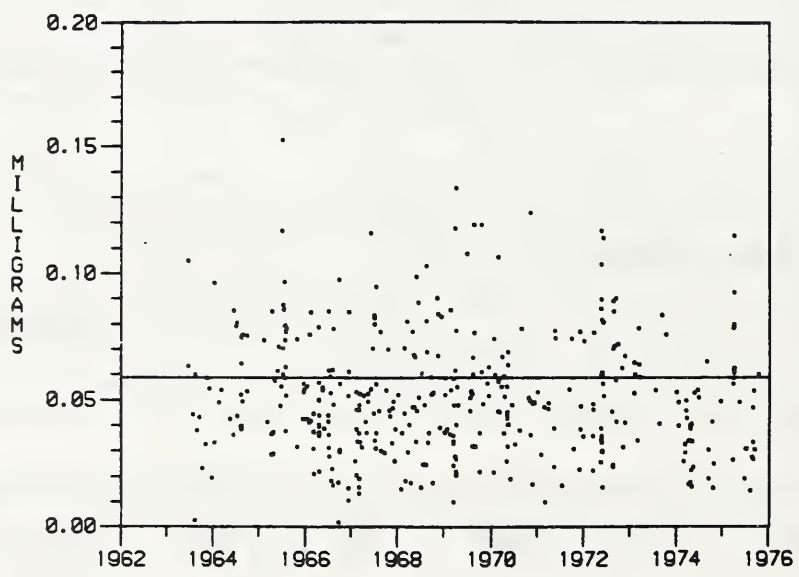


Figure 14
Within standard deviations (mg) for NBS balance #4
plotted against time (years)

The examples cited in figures 13 and 14 are for a process, as was said before, that has been in existence for a long time and that is demonstrably in-control. It may be instructive to examine a few processes, or at least the data from those processes, that have not been carefully monitored and that are not necessarily in-control.

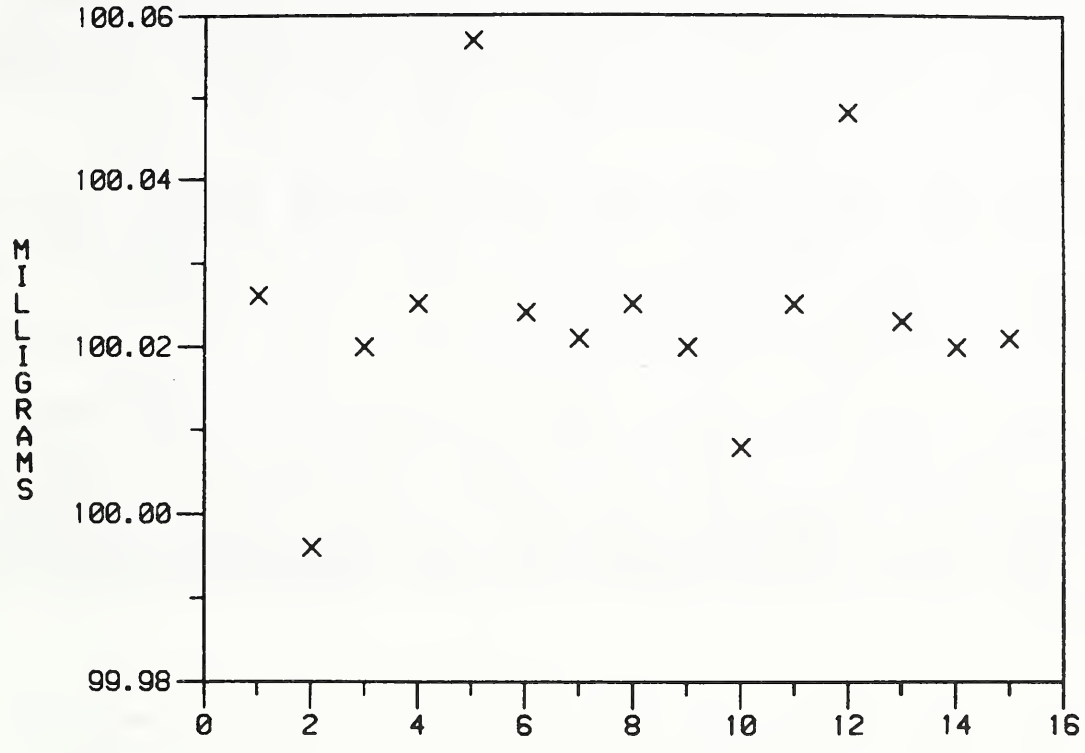


Figure 15
Measurements (mg) on a 100g weight plotted against time (months)

Take, for example, the data in figure 15 which represent repeated weighings made over a fifteen month period on a calibrated weight. Notice that the majority of the values are clustering close together but that there are a relatively large number of extremely discordant values. It is not sensible in this case to ask, "What base line and control limits are appropriate for this process?" In fact, at this point in time, a measurement process does not exist because it is not possible to predict a future value of the process, or in other words, the data as plotted in figure 15 do not represent a random sample from a single error distribution. In this case, a critical deficiency in the measurement process was tracked down; namely, that the elapsed time between two weighings being made on the balance in succession was not sufficient for the balance to come to proper equilibrium.

A control procedure involving a power instrument standard is shown in figure 16. The graph shows assignments made to the power standard as it was intercompared with its primary power source over a two-week period. The sixteen resulting measurements define the base line and control limits for the process.

The results of sixteen additional measurements taken a year later are shown in figure 17, and although they are clearly out-of-control with respect to the initial measurements, they are consistent among themselves raising a question as to whether the power standard itself is changing radically, whether the initial measurements were, in fact, out-of-control and should be discounted, or whether the process is not properly characterized by either set of measurements. Really only one thing is clear at this point -- that the assignment cannot be made with any degree of confidence and that the power standard should not be the basis for a calibration program until the process of assigning a value to the power standard is adequately characterized.

This was accomplished by repeating the intercomparison at three month intervals taking only two or three measurements each time instead of sixteen. The results are shown in figure 18. A large component of variance that did not show up in the initial two-week interval affects the measurement process, and the standard deviation computed from the short-term measurements under-estimates the process variability as it exists over, say, a year's time.

This example demonstrates an extremely important principle of measurement assurance; namely that in general there is little value in closely spaced repetitions. These should be kept to a minimum, and measurements should be taken frequently over a long period of time in order to correctly characterize a process. This practice should be continued until the process parameters are well established and only then should the intervals between intercomparisons be lengthened.

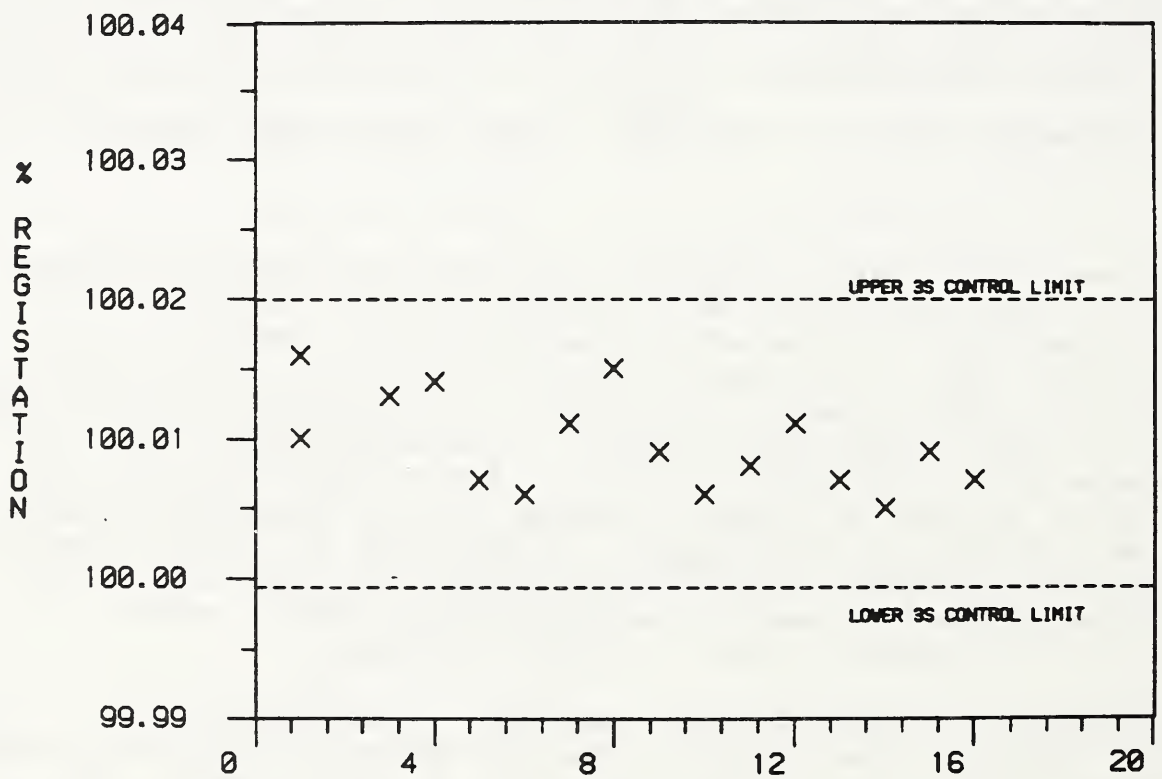


Figure 16
Measurements (% reg) on a power standard plotted against run sequence showing upper and lower three standard deviation limits

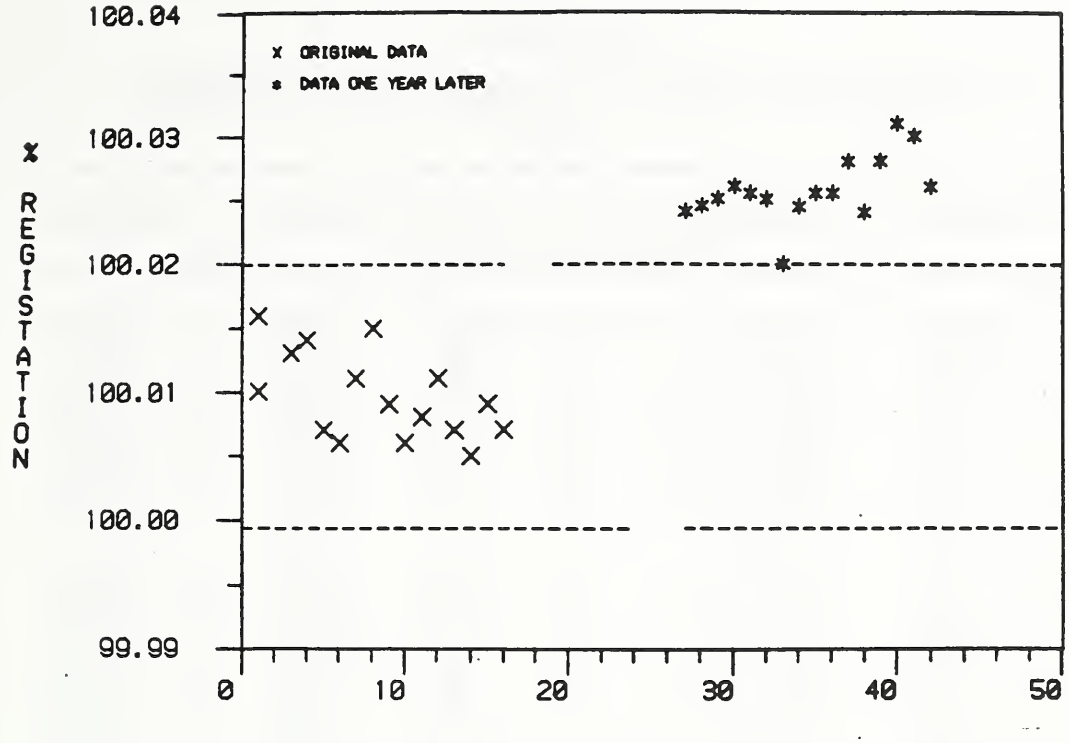


Figure 17
Original measurements (% reg) on power standard and measurements on the same standard a year later with original control limits

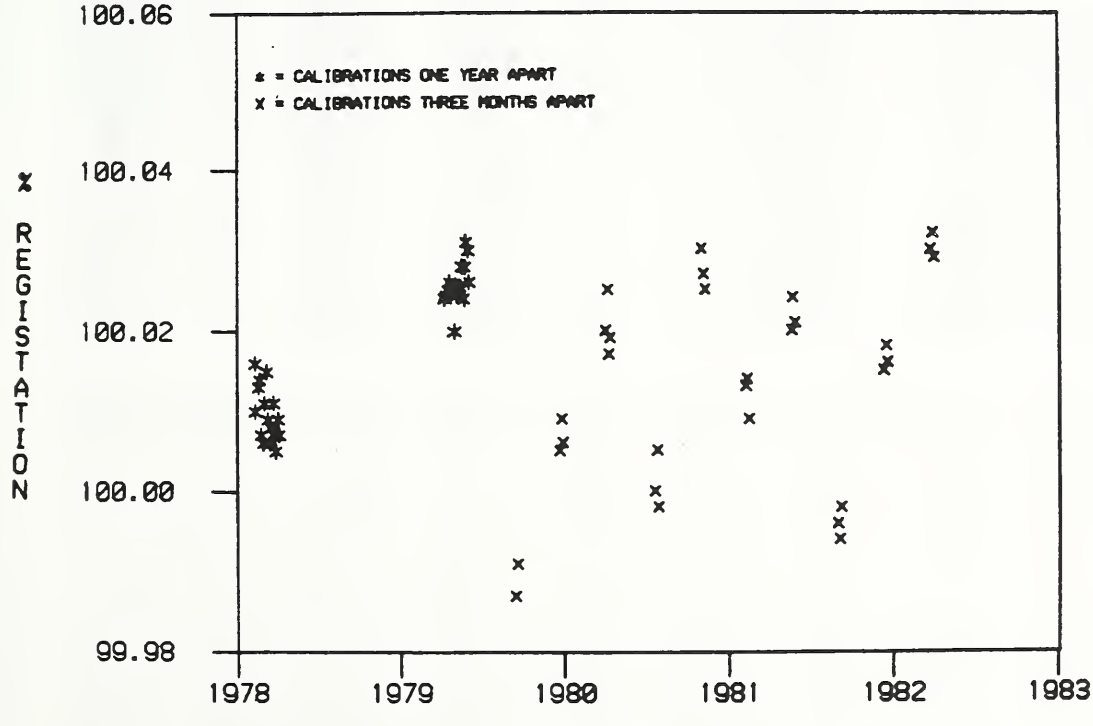


Figure 18
Measurements (% reg) on the power standard at three month intervals over three years

Table I
Critical Values $t_{\alpha/2}(\nu)$ of Student's t Distribution

ν	α=0.05	α=0.01	ν	α=0.05	α=0.01
2	4.303	9.925	62	1.999	2.657
4	2.776	4.604	64	1.998	2.655
6	2.447	3.707	66	1.997	2.652
8	2.306	3.355	68	1.995	2.650
10	2.228	3.169	70	1.994	2.648
12	2.179	3.055	72	1.993	2.646
14	2.145	2.977	74	1.993	2.644
16	2.120	2.921	76	1.992	2.642
18	2.101	2.878	78	1.991	2.640
20	2.086	2.845	80	1.990	2.639
22	2.074	2.819	82	1.989	2.637
24	2.064	2.797	84	1.989	2.636
26	2.056	2.779	86	1.988	2.634
28	2.048	2.763	88	1.987	2.633
30	2.042	2.750	90	1.987	2.632
32	2.037	2.738	92	1.986	2.630
34	2.032	2.728	94	1.985	2.629
36	2.028	2.719	96	1.984	2.628
38	2.024	2.712	98	1.983	2.627
40	2.021	2.704	100	1.983	2.626
42	2.018	2.698	102	1.983	2.625
44	2.015	2.692	104	1.982	2.624
46	2.013	2.687	106	1.982	2.623
48	2.011	2.682	108	1.981	2.622
50	2.009	2.678	110	1.981	2.621
52	2.007	2.674	112	1.981	2.620
54	2.005	2.670	114	1.981	2.620
56	2.003	2.667	116	1.981	2.619
58	2.002	2.663	118	1.980	2.618
60	2.000	2.660	120	1.980	2.617
			∞	1.960	2.576

ν = number of degrees of freedom in the total standard deviation.

Table II

Critical values $F_\alpha(\nu_1, \nu_2)$ of the F Distribution
α=0.01

DF ν_2	Degrees of freedom ν_1									
	1	2	3	4	5	6	7	8	9	10
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
45	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70
55	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
65	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59
75	6.99	4.90	4.05	3.58	3.27	3.05	2.89	2.76	2.65	2.57
80	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55
85	6.94	4.86	4.02	3.55	3.24	3.02	2.86	2.73	2.62	2.54
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52
95	6.91	4.84	3.99	3.52	3.22	3.00	2.83	2.70	2.60	2.51
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50
105	6.88	4.81	3.97	3.50	3.20	2.98	2.81	2.69	2.58	2.49
110	6.87	4.80	3.96	3.49	3.19	2.97	2.81	2.68	2.57	2.49
115	6.86	4.79	3.96	3.49	3.18	2.96	2.80	2.67	2.57	2.48
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Table II continued

Critical Values $F_\alpha(\nu_1, \nu_2)$ of the F Distribution
$\alpha = 0.01$

DF ν_2	Degrees of freedom ν_1									
	12	14	16	18	20	22	24	26	28	30
10	4.71	4.60	4.52	4.46	4.41	4.36	4.33	4.30	4.27	4.25
11	4.40	4.29	4.21	4.15	4.10	4.06	4.02	3.99	3.96	3.94
12	4.16	4.05	3.97	3.91	3.86	3.82	3.78	3.75	3.72	3.70
13	3.96	3.86	3.78	3.72	3.66	3.62	3.59	3.56	3.53	3.51
14	3.80	3.70	3.62	3.56	3.51	3.46	3.43	3.40	3.37	3.35
15	3.67	3.56	3.49	3.42	3.37	3.33	3.29	3.26	3.24	3.21
16	3.55	3.45	3.37	3.31	3.26	3.22	3.18	3.15	3.12	3.10
17	3.46	3.35	3.27	3.21	3.16	3.12	3.08	3.05	3.03	3.00
18	3.37	3.27	3.19	3.13	3.08	3.03	3.00	2.97	2.94	2.92
19	3.30	3.19	3.12	3.05	3.00	2.96	2.92	2.89	2.87	2.84
20	3.23	3.13	3.05	2.99	2.94	2.90	2.86	2.83	2.80	2.78
22	3.12	3.02	2.94	2.88	2.83	2.78	2.75	2.72	2.69	2.67
24	3.03	2.93	2.85	2.79	2.74	2.70	2.66	2.63	2.60	2.58
26	2.96	2.86	2.78	2.72	2.66	2.62	2.58	2.55	2.53	2.50
28	2.90	2.79	2.72	2.65	2.60	2.56	2.52	2.49	2.46	2.44
30	2.84	2.74	2.66	2.60	2.55	2.51	2.47	2.44	2.41	2.39
35	2.74	2.64	2.56	2.50	2.44	2.40	2.36	2.33	2.30	2.28
40	2.66	2.56	2.48	2.42	2.37	2.33	2.29	2.26	2.23	2.20
45	2.61	2.51	2.43	2.36	2.31	2.27	2.23	2.20	2.17	2.14
50	2.56	2.46	2.38	2.32	2.27	2.22	2.18	2.15	2.12	2.10
55	2.53	2.42	2.34	2.28	2.23	2.18	2.15	2.11	2.08	2.06
60	2.50	2.39	2.31	2.25	2.20	2.15	2.12	2.08	2.05	2.03
65	2.47	2.37	2.29	2.23	2.17	2.13	2.09	2.06	2.03	2.00
70	2.45	2.35	2.27	2.20	2.15	2.11	2.07	2.03	2.01	1.98
75	2.43	2.33	2.25	2.18	2.13	2.09	2.05	2.02	1.99	1.96
80	2.42	2.31	2.23	2.17	2.12	2.07	2.03	2.00	1.97	1.94
85	2.40	2.30	2.22	2.15	2.10	2.06	2.02	1.98	1.95	1.93
90	2.39	2.29	2.21	2.14	2.09	2.04	2.00	1.97	1.94	1.92
95	2.38	2.28	2.20	2.13	2.08	2.03	1.99	1.96	1.93	1.90
100	2.37	2.27	2.19	2.12	2.07	2.02	1.98	1.95	1.92	1.89
105	2.36	2.26	2.18	2.11	2.06	2.01	1.97	1.94	1.91	1.88
110	2.35	2.25	2.17	2.10	2.05	2.00	1.96	1.93	1.90	1.88
115	2.34	2.24	2.16	2.10	2.04	2.00	1.96	1.92	1.89	1.87
120	2.34	2.23	2.15	2.09	2.03	1.99	1.95	1.92	1.89	1.86
∞	2.19	2.09	2.00	1.94	1.88	1.84	1.79	1.76	1.73	1.70

Table II continued

Critical Values $F_\alpha(\nu_1, \nu_2)$ of the F Distribution
α=0.01

DF ν_2	Degrees of freedom ν_1									
	40	50	60	70	80	90	100	110	120	∞
10	4.17	4.12	4.08	4.06	4.04	4.03	4.01	4.00	4.00	3.91
11	3.86	3.81	3.78	3.75	3.73	3.72	3.71	3.70	3.69	3.60
12	3.62	3.57	3.54	3.51	3.49	3.48	3.47	3.46	3.45	3.36
13	3.43	3.38	3.34	3.32	3.30	3.28	3.27	3.26	3.25	3.17
14	3.27	3.22	3.18	3.16	3.14	3.12	3.11	3.10	3.09	3.01
15	3.13	3.08	3.05	3.02	3.00	2.99	2.98	2.97	2.96	2.87
16	3.02	2.97	2.93	2.91	2.89	2.87	2.86	2.85	2.84	2.76
17	2.92	2.87	2.83	2.81	2.79	2.78	2.76	2.75	2.75	2.65
18	2.84	2.78	2.75	2.72	2.70	2.69	2.68	2.67	2.66	2.57
19	2.76	2.71	2.67	2.65	2.63	2.61	2.60	2.59	2.58	2.49
20	2.69	2.64	2.61	2.58	2.56	2.55	2.54	2.53	2.52	2.42
22	2.58	2.53	2.50	2.47	2.45	2.43	2.42	2.41	2.40	2.31
24	2.49	2.44	2.40	2.38	2.36	2.34	2.33	2.32	2.31	2.21
26	2.42	2.36	2.33	2.30	2.28	2.26	2.25	2.24	2.23	2.13
28	2.35	2.30	2.26	2.24	2.22	2.20	2.19	2.18	2.17	2.07
30	2.30	2.24	2.21	2.18	2.16	2.14	2.13	2.12	2.11	2.01
35	2.19	2.14	2.10	2.07	2.05	2.03	2.02	2.01	2.00	1.89
40	2.11	2.06	2.02	1.99	1.97	1.95	1.94	1.93	1.92	1.81
45	2.05	2.00	1.96	1.93	1.91	1.89	1.88	1.86	1.85	1.74
50	2.01	1.95	1.91	1.88	1.86	1.84	1.82	1.81	1.80	1.69
55	1.97	1.91	1.87	1.84	1.82	1.80	1.78	1.77	1.76	1.64
60	1.94	1.88	1.84	1.81	1.78	1.76	1.75	1.74	1.73	1.60
65	1.91	1.85	1.81	1.78	1.75	1.74	1.72	1.71	1.70	1.57
70	1.89	1.83	1.78	1.75	1.73	1.71	1.70	1.68	1.67	1.54
75	1.87	1.81	1.76	1.73	1.71	1.69	1.67	1.66	1.65	1.52
80	1.85	1.79	1.75	1.71	1.69	1.67	1.65	1.64	1.63	1.50
85	1.83	1.77	1.73	1.70	1.67	1.65	1.64	1.62	1.61	1.48
90	1.82	1.76	1.72	1.68	1.66	1.64	1.62	1.61	1.60	1.46
95	1.81	1.75	1.70	1.67	1.65	1.63	1.61	1.60	1.58	1.45
100	1.80	1.74	1.69	1.66	1.63	1.61	1.60	1.58	1.57	1.43
105	1.79	1.73	1.68	1.65	1.62	1.60	1.59	1.57	1.56	1.42
110	1.78	1.72	1.67	1.64	1.61	1.59	1.58	1.56	1.55	1.41
115	1.77	1.71	1.66	1.63	1.60	1.58	1.57	1.55	1.54	1.40
120	1.76	1.70	1.66	1.62	1.60	1.58	1.56	1.54	1.53	1.39
∞	1.60	1.53	1.48	1.44	1.41	1.38	1.36	1.35	1.33	

REFERENCES

[1] Dorsey, N. E. The velocity of light. Trans. Am. Phil. Soc. XXXIV.; 1944, pp.1-110.

[2] Hayford, J. A. On the least square adjustment of weighings, U.S. Coast and Geodetic Survey Appendix 10, Report for 1892; 1893.

[3] Benoit, M. J. R. L'Etalonnage des Series de Poids Travaux et Memoirs du Bureau International des Poids et Mesures 13(1); 1907.

[4] Pienkowsky, A. T. Short tests for sets of laboratory weights. Scientific Papers of the Bureau of Standards, 21(527); 1926.

[5] Cameron, J. M.; Croarkin, M. C.; Raybold, R. C. Designs for the Calibration of Standards of Mass. Nat. Bur. Stand. (U.S.) Tech Note 952; 1977.

[6] §Eisenhart, C. Realistic evaluation of the precision and accuracy of instrument calibration systems. J. Res. Nat. Bur. Stand. (U.S.) 67C(2); 1962. pp. 161-187.

[7] §Youden, W. J. Experimental design and ASTM committees. Materials Research and Standards, 1(11); 1961. pp. 862-867.

[8] §Youden, W. J. Physical measurements and experiment design. Colloques Internationaux du Centre National de la Recherche Scientifique No. 110, le Plan d'Experiences; 1961. pp. 115-128.

[9] §Pontius, P. E. and Cameron, J. M. Realistic Uncertainties and the Mass Measurement Process. Nat. Bur. Stand. (U.S.) Monogr. 103; 1967.

[10] Croarkin, M. C.; Beers, J. S.; Tucker, C. Measurement Assurance for Gage Blocks. Nat. Bur. Stand. (U.S.) Monogr. 163; 1979.

[11] Croarkin, M. C.; Varner, R. N. Measurement Assurance for Dimensional Measurements on Integrated Circuit Photomasks. Nat. Bur. Stand. (U.S.) Tech. Note 1164; 1982.

[12] American National Standard ANSI N15.18-1975. Mass calibration techniques for nuclear materials control. Available from ANSI, Inc., 1430 Broadway, New York, NY 10018.

[13] Pontius, P. E. and Doher, L. W. The joint ANSI-INMM 8.1-Nuclear Regulatory Commission study of uranium hexafluoride cylinder material accountability bulk measurements. Proc. 18th Ann. Mtg. INMM, VI(III); 1977. p. 480.

§Reprinted in Nat. Bur. Stand. (U.S.) Spec. Publ. 300, Vol I. Precision Methods and Calibration: Statistical Concepts and Procedures. H. H. Ku, editor. 1969.

[14] Beers, J. S. A Gage Block Measurement Process Using Single Wavelength Interferometry. Nat. Bur. Stand. (U.S.) Monogr. 152; 1975.

[15] Cameron, J. M. Measurement Assurance. Nat. Bur. Stand. (U.S.) NBSIR 77-1240; 1977.

[16] Kieffer, L. J., ed. Calibration and Related Measurement Services of the National Bureau of Standards. Nat. Bur. Stand. (U.S.) Spec. Publ. 250; 1982.

[17] Pipkin, F. R., Ritter, R. C. Precision measurements and fundamental constants. Science, 219 (4587); 1983. p. 917.

[18] Pontius, P. E. The Measurement Assurance Program - A Case Study: Length Measurements Part I. Long Gage Blocks (5 in to 20 in). Nat. Bur. Stand. (U.S.) Monogr. 149; 1975.

[19] Beers, J. S., Tucker, C. D. Intercomparison Procedures for Gage Blocks Using Electromechanical Comparators. Nat. Bur. Stand. (U.S.) NBSIR 76-979; 1976. p. 9.

[20] Simpson, J. A. Foundations of metrology. J. Res. Nat. Bur. Stand. (U.S.). 86(3), 1981. p. 282.

[21] Nyyssonen, D. Linewidth measurement with an optical microscope: The effect of operating conditions on the image profile. Appl. Opt. 16(8); Ausust 1977. pp. 2223-2230.

[22] Jerke, J. M., ed. Semiconductor Measurement Technology: Accurate Linewidth Measurements on Integrated Circuit Photomasks. Nat. Bur. Stand. (U.S.) Spec. Publ. 400-43; 1980. pp. 7-15.

[23] Jerke, J. M.; Croarkin, M. C.; Varner, R. N. Semiconductor Measurement Technology: Interlaboratory Study on Linewidth Measurements for Antireflective Chromium Photomasks. Nat. Bur. Stand. (U.S.) Spec. Publ. 400-74; 1982.

[24] See reference 20, p. 283.

[25] See reference 22, p. 50.

[26] See reference 20, p. 283.

[27] Cameron, J. M. Encyclopedia of Statistical Sciences, Vol 1. S. Kotz and N. L. Johnson, ed. New York: John Wiley & Sons, Inc.; 1982. pp. 341-347.

[28] Ku, H. H. Statistical concepts of a measurement process. Precision Methods and Calibration: Statistical Concepts and Procedures. Nat. Bur. Stand. (U.S.) Spec. Publ. 300, Vol I. H. H. Ku, ed. 1969. pp 296-20 to 330-54.

[29] Mandel, J. The Statistical Analysis of Experimental Data. New York: Interscience Publ; 1964. pp. 278-279.

[30] Snedecor, G. W. and Cochran, W. G. Statistical Methods, Sixth ed. Ames, Iowa: The Iowa State University Press; 1976. pp. 279-280.

[31] See reference [7]. pp. 862-863.

[32] Mattingly, G. H.; Pontius, P. E.; Allion, H. H.; Moore, E. F. A laboratory study of turbine meter uncertainty. Proc. Symp. on Flow in Open Channels and Closed Circuits; Nat. Bur. Stand. (U.S.) Spec. Publ. 484; 1977.

[33] See reference [10].

[34] Duncan, A.J. Quality Control and Industrial Statistics, Fourth ed. Homewood: Richard D. Irwin, Inc; 1974. p. 381.

[35] See reference [8], p. 21-22.

[36] See reference [28], p. 299.

[37] See reference [28], p. 305-308.

[38] See reference [23].

[39] Hockersmith, T. E.; Ku, H. H. Uncertainties associated with proving ring calibration. Precision Methods and Calibration: Statistical Concepts and Procedures. Nat. Bur. Stand. (U.S.) Spec. Publ. 300, Vol. 1. H. H. Ku, ed.; 1969. pp. 257-1 to 264-8.

[40] See reference [11], p. 30-33.

[41] Youden, W. J. Uncertainties in calibration. IRE Transactions on Instrumentation, I-11, (3,4); 1962. p. 137.

[42] Eisenhart, C., Ku, H. H. and Colle, R. Expression of the Uncertainties of Final Measurement Results; Reprints. Nat. Bur. Stand. (U.S.) Spec. Pub. 644; 1983.

[43] Giacomo, P. News from BIPM. Metrologia, 17; 1981. pp. 73-74.

[44] See reference [28]. pp. 322-323.

[45] Raghavarao, D. Construction and Combinatorial Problems in Design of Experiments. New York: John Wiley & Sons, Inc; 1971. p. 315.

[46] Mood, A. M. On Hotelling's Weighing Problem. Annals of Mathematical Statistics, 17; 1946. pp.432-446.

[47] See reference [5].

[48] Cameron, J. M.; Hailes, G. E. Designs for the Calibration of Small Groups of Standards in the Presence of Drift. Nat. Bur. Stand. (U.S.) Tech Note 844; 1974. p. 1.

[49] Jaegar, K. B. and Davis, R. S. A Primer for Mass Metrology. Nat. Bur. Stand. (U.S.) Special Publication: Industrial Measurement Series 700-1; 1984.

[50] See reference [10], pp. 13-25.

[51] See reference [10], pp. 27-39.

[52] Cameron, J. M.; Eicke, W. G. Designs for Surveillance of the Volt Maintained by a Small Group of Saturated Standard Cells. Nat. Bur. Stand. (U.S.) Tech Note 430; 1967.

[53] Cameron, J. M. The Use of the Method of Least Squares in Calibration. Nat. Bur. Stand. (U.S.) NBSIR 74-587; 1974.

[54] See reference [49].

[55] Varner, R. N. Mass Calibration Computer Software. Nat. Bur. Stand. (U.S.) Tech. Note 1127; 1980.

[56] See reference [52]. pp. 1-2.

[57] Cochran, W. J. The distribution of the largest of a set of estimated variances as a fraction of their total. Annals of Eugenics, 11; 1941. pp. 47-52.

[58] Eisenhart, C. Significance of the largest of a set of sample estimates of variance. Chapter 15 of Selected Techniques of Statistical Analysis. Eisenhart, C., Hastay, M. W., Wallis, W. A., editors. New York: McGraw Hill Book Co., Inc.; 1947. pp. 383-394.

[59] Draper, N. R.; Smith, H. Applied Regression Analysis. New York: John Wiley & Sons, Inc.; 1966. p. 1-32.

[60] See reference [22], p. 350.

[61] Bicking, C. A.; Gryna, F. M. Jr. Process Control by Statistical Methods. Section 23 of Quality Control Handbook, Third Edition, J. M. Juran, ed. New York: McGraw-Hill Book Co.; 1974. p. 23-2, 23-3.

[62] See reference [34]. pp. 464-484.

[63] Reynolds, J. R. Jr.; Ghosh, B. K. Designing control charts for means and variances. 1081 ASQC Quality Congress Transactions: San Francisco; 1981.

[64] Shewhart, W. A. Statistical Method from the Viewpoint of Quality Control. The Graduate School, U.S. Department of Agriculture, Washington, DC; 1939.

[65] Grant, E. L.; Leavenworth, R. S. Statistical Quality Control, 4th Edition. New York: McGraw Hill Book Co.; 1976. p. 129.

[66] Parobeck, P.; Tomb, T.; Ku, H. H.; Cameron, J. M. Measurement assurance program for weighings of respirable coal mine dust samples. J. Quality Tech, 13(3); 1981. pp. 157-165.

[67] Bowker, A. H. Tolerance limits for normal distributions. Chapter 2 of reference [58]. p. 99.

[68] See reference [63]. p. 381.

APPENDIX A

The purpose of this appendix is to define the matrix manipulations‡ that produce the least-squares solution to a weighing design along with the propagation of associated standard deviations and uncertainties.† The theory is explained by Cameron et al. in reference [5]. It is assumed that a series of weighing designs is required in order to calibrate an entire weight set and that assignments to individual weights depend upon a starting restraint with known value that is invoked in the first design. The starting restraint is usually the known sum of two reference kilograms. It is also assumed that the designs are interconnected in such a way that a value assigned to an individual weight or sum of weights from one design constitutes the restraint for the next design in the series.

Each design in the series involves n intercomparisons among p weights where the p weights include the reference standards composing the restraint, the test weights, and check standard.

The model for the measurement process assumes that these observations are related to the values of the weights by

$$D = AX^* + \varepsilon \tag{A.1}$$

where D is the (nx1) vector of observations; A is an (nxp) design matrix of zeroes and ones such that a plus or minus one in the ijth position indicates that the jth weight is measured by the ith observation, and a zero indicates the converse; X^* is the (px1) vector of unknown values for the p weights; and ε is the (nx1) vector of random errors.

Define

$$D' = (d_1 \cdots d_n) \tag{A.2}$$

$$A = \begin{pmatrix} a_{11} \cdots a_{1p} \\ \vdots \quad\quad \vdots \\ \vdots \quad\quad \vdots \\ \vdots \quad\quad \vdots \\ a_{n1} \cdots a_{np} \end{pmatrix} \tag{A.3}$$

$$(X^*)' = (X_1^* \cdots X_p^*) \tag{A.4}$$

and

$$\varepsilon' = (\varepsilon_1 \cdots \varepsilon_n) \tag{A.5}$$

‡The matrix notation that is used in this appendix denotes the transpose of the matrix M by M' and the inverse of the matrix M by M^{-1}.

†Assuming that there is no significant between component of variance in the measurement process.

In order to define various linear combinations of the weights, we will also define several vectors of size (pxl) which have the general form

$$\ell' = (\ell_1 \cdots \ell_p)$$

where each element ℓ_i (i=1,...,p) is either zero, plus one or minus one.

The least-squares estimate for (A.4) depends upon the inverse of the normal equations A'A. The usual case for calibration experiments is that A'A has rank p-1. Where A'A has rank less than p, the inverse does not exist and a solution can be obtained only by imposing a restraint upon the system of equations. Therefore, we let R^* be a scalar with known value called the restraint; and ℓ_R be a (pxl) vector of zeroes and ones such that a one in the jth position indicates that the jth weight is in the restraint, and a zero indicates the converse. For example,

$$\ell_R' = (1\ 1\ 0 \cdots 0)$$

indicates that the restraint is over the first two weights.

One approach to finding the least-squares estimate for X^* is via an augmented matrix B where

$$B = \begin{pmatrix} A'A & \ell_R & A'D \\ \ell_R' & 0 & R^* \\ 0 & 0 & -1 \end{pmatrix} \qquad (A.6)$$

is a (p+2)x(p+2) matrix whose inverse

$$B^{-1} = \begin{pmatrix} Q & h & \hat{X}^* \\ h' & 0 & . \\ . \ . \ . & & . \end{pmatrix} \qquad (A.7)$$

can be partitioned as shown above. The (pxp) matrix Q in the upper left hand corner of B^{-1} contains information relating to the variances of the estimates, and the (pxl) matrix X^* in the upper right hand corner of B^{-1} contains the least-squares estimates for the p weights. The other quantities in B^{-1} are not of interest for this application. Notice that once the inverse of B has been computed, the estimates are immediately available without further matrix multiplications.¶

The individual deviations of the observations from their fitted values are given by the (pxl) vector ξ where

$$\xi' = (D - AX^*)', \qquad (A.8)$$

¶ The caret (^) indicating a least-squares estimate from the data is dropped in future references to X^*.

and the within standard deviation for the design is

$$s_w = \left(\frac{\xi' \xi}{n-p+1} \right)^{1/2} \tag{A.9}$$

with n-p+1 degrees of freedom.

The restraint for the next design in the series can be written in the form

$$\Sigma^* = \ell_\Sigma' X^* \tag{A.10}$$

where ℓ_Σ is a (px1) vector of zeroes and ones where a one in the ith position indicates that the ith weight is to be included in the restraint for the next design, and a zero indicates the converse. The standard deviation for the outgoing restraint is given by

$$s_\Sigma = \left(\ell_\Sigma' Q \ell_\Sigma \, s_w^2 + \left(\frac{\ell_\Sigma' W}{\ell_R' W} \right)^2 s_R^2 \right)^{1/2} \tag{A.11}$$

where s_R is the standard deviation of the incoming restraint R^* as computed from the previous design, and

$$W' = (W_1 \; . \; . \; . \; W_p)$$

where W is a (px1) vector of nominal values for the p weights. If the current design is the first design in the series, then s_R is zero.

Notice that the computation of the standard deviation associated with the check standard as defined in (A.14) and the computation of the standard deviation associated with the values of the test weights as defined in (A.16) are also dependent on s_R. Thus, the standard deviations for each series are dependent on all prior series as they are propagated starting with the first series.

The current value for the check standard from the design can be written in the form

$$c = \ell_c' X^* \tag{A.13}$$

where ℓ_c is a (px1) vector of zeroes and ones such that a plus or minus one in the ith position indicates that the ith weight is in the check standard, and a zero indicates the converse.

The standard deviation of the check standard value is given by

$$s_c = \left(\ell_c' Q \ell_c \, s_w^2 + \left(\frac{\ell_c' W}{\ell_R' W} \right)^2 s_R^2 \right)^{1/2} . \tag{A.14}$$

Then given that the accepted value for the check standard is known from previous experiments to be A_c, a test for control is made by computing the test statistic

$$t_c = \frac{|A_c - c|}{s_c} \tag{A.15}$$

and comparing it to a critical value.

Finally, we are interested in the uncertainty of the value assigned to a single weight or to a collection of weights. For each summation or difference of weights that is of interest, we define a (pxl) vector ℓ_S of zeroes, plus ones and minus ones such that a one in the ith position indicates that the ith weight is involved in the summation or difference, and a zero indicates the converse. The reported value for the summation S is S^* where

$$S^* = \ell_S{}' X^*.$$

The standard deviation for the summation, designated by s_S is

$$s_S = \left(\ell_S{}' Q \ell_S \; s_w{}^2 + \left(\frac{\ell_S{}' W}{\ell_R{}' W} \right)^2 s_R{}^2 \right)^{1/2} \tag{A.16}$$

and the uncertainty associated with the summation is

$$U = 3s_S + \frac{\ell_S{}' W}{\ell_{SR}{}' W} U_{SR} \tag{A.17}$$

where U_{SR} is the uncertainty assigned to the starting restraint in the series, and similarly ℓ_{SR} is the (pxl) vector of zeroes, plus ones and minus ones such that a plus or minus one in the ith position indicates that the ith weight is in the starting restraint.

Notice that if we are talking about a single weight whose value is $X_j{}^*$, then the quantity

$$\ell_S{}' Q \ell_S = q_{jj}$$

where q_{jj} is the jth diagonal element in Q.

For the next design in the series, let the restraint be $R^* = \Sigma^*$ with standard deviation $s_R = s_\Sigma$ and proceed with the calculation starting with equation (A.1).

Expression of the Uncertainties of Final Measurement Results: Reprints

Churchill Eisenhart

Harry H. Ku

R. Collé

National Bureau of Standards
Washington, DC 20234

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

Issued January 1983

Foreword

The reporting of final measurement results, and the uncertainties associated with the measurement processes used to obtain these results, has always been and continues to be a source of difficulty. The three articles reprinted in this publication are collected here as a convenient reference source for experimenters who must face the difficult task of deciding how to express measurement uncertainties. The philosophical basis, general guidelines, and specific recommendations for expressing uncertainties contained within these articles have evolved at NBS over a period of many years.

The first article originally appeared in *Science* in 1968. This article develops the underlying basis and general guidelines on the forms of expression needed for uncertainty statements, and presents specific recommendations for four distinct cases: (i) when both systematic error and imprecision are negligible; (ii) when systematic error is not negligible, and imprecision is negligible; (iii) when neither systematic error nor imprecision is negligible; and (iv) when systematic error is negligible, and imprecision is not negligible.

The second article, written as a companion to the first, originally appeared in a 1968 issue of *M&D: Measurements and Data.* It gives a condensed summary of the recommendations presented in the first article, and provides tabular guides to commonly used statements of imprecision, systematic error, and uncertainty.

The third article is a postscript to the two preceding articles, and was prepared in 1980 for an internal NBS communications manual. It reinforces the major thrust and content of the earlier articles, but includes more recent thought particularly in regard to overall uncertainty statements.

The first two articles have since been reprinted in several NBS publications including Special Publication 300, Volume 1, *Precision Measurement and Calibration: Statistical Concepts and Procedures* (Harry H. Ku, ed., 1969). The 1980 NBS communications manual incorporated the second and third articles, but did not reprint the first article. Furthermore, this manual is not accessible outside NBS. This special publication, therefore, collects all three articles, for the first time, in one convenient source which is available to the many scientists and engineers throughout the entire measurement community.

Contents

Expression of the Uncertainties of Final Results

Clear statements of the uncertainties of reported values are needed for their critical evaluation.

Churchill Eisenhart

Measurement of some property of a thing in practice always takes the form of a sequence of steps or operations that yield as an end result a number that serves to represent the amount or quantity of some particular property of a thing—a number that indicates how much of this property the thing has, for someone to use for a specific purpose. The end result may be the outcome of a single reading of an instrument, with or without corrections for departures from prescribed conditions. More often it is some kind of average, for example, the arithmetic mean of a number of independent determinations of the same magnitude, or the final result of a least squares "reduction" of measurements of a number of different magnitudes that bear known relations with one another in accordance with a definite experimental plan. In general, the purpose for which the answer is needed determines the precision or accuracy required and ordinarily also the method of measurement employed.

Although the accuracy required of a reported value depends primarily on the *intended* use, or uses, of the value, one should not ignore the requirements of other uses to which it is likely to be put. A reported value whose accuracy is entirely unknown is worthless.

Strictly speaking, the actual *error* of a reported value, that is the magnitude and sign of its deviation from the truth (*1*), is usually unknowable. Limits to this error, however, can usually be inferred—with some risk of being incorrect—from the precision of the measurement process by which the reported value was obtained, and from reasonable limits to the possible bias of the measurement process. The *bias*, or *systematic error*, of a measurement process is the magnitude and direction of its tendency to measure something other than what was intended; its *precision* refers to the typical closeness together of successive independent measurements of a single magnitude generated by repeated applications of the process under specified conditions; and its *accuracy* is determined by the closeness to the true value characteristic of such measurements.

Precision and accuracy are inherent characteristics of the measurement process employed and not of the particular end result obtained. From experience with a particular measurement process and knowledge of its sensitivity to uncontrolled factors, one can often place reasonable bounds on its likely systematic error (bias). It is also necessary to know how well the particular value in hand is likely to agree with other values that the same measurement process might have provided in this instance, or might yield on remeasurement of the same magnitude on another occasion. Such information is provided by the estimated *standard error* (*2*) of the reported value, which measures (or is an index of) the characteristic disagreement of repeated determinations of the same quantity by the same method, and thus serves to indicate the precision (strictly, the imprecision) of the reported value (*3*).

Four Distinct Forms of Expression Needed

The uncertainty of a reported value is indicated by stating credible limits to its likely inaccuracy. No single form of expression for these limits is universally satisfactory. In fact, different forms of expression are recommended, which will depend on the relative magnitudes of the imprecision and likely bias, and their relative importance in relation to the intended use of the reported value, as well as to other possible uses to which it may be put (*4*).

Four distinct cases need to be recognized: (i) both systematic error and imprecision negligible, in relation to the requirements of the intended and likely uses of the result; (ii) systematic error not negligible, imprecision negligible; (iii) neither systematic error nor imprecision negligible; and (iv) systematic error negligible, imprecision not negligible.

Specific recommendations with respect to each of these cases are made below. General guidelines upon which these specific recommendations are based are discussed in the following paragraphs.

Perils of Shorthand Expressions

Final results and their respective uncertainties should be reported in sentence form whenever possible. The shorthand form "$a \pm b$" should be avoided in abstracts and summaries; and never used without explicit explanation of its connotation. If no explanation is given, many persons will take $\pm b$ to signify bounds to the inaccuracy of a. Others may assume that b is the "standard error," or the "probable error," of a, and hence the uncertainty of a is at least $\pm 3b$, or $\pm 4b$, respectively. Still others may take b to be an indication merely of the imprecision of the individual measurements, that is, to be the "standard deviation," or the "average deviation," or the "probable error" of a single observation. Each of these interpretations reflects a practice of which instances can be found in current scientific literature. As a step in the direction of reducing this current confusion, it is recommended that the use of "$a \pm b$" in presenting results be limited to that sanctioned for the case of tabular results in the fourth recommendation of the section below headed "Systematic error not negligible, imprecision negligible."

The author is a senior research fellow and former chief of the Statistical Engineering Laboratory at the National Bureau of Standards, Washington, D.C. 20234. The recommendations presented in this paper have evolved at the Bureau over a period of many years and are made public here for general information, and to educe comments and suggestions.

Imprecision and Systematic Error Require Separate Treatment

Since imprecision and systematic error are distinctly different components of inaccuracy, and are subject to different treatments and interpretations in usage, two numerics respectively expressing the imprecision and bounds to the systematic error of the reported result should be used whenever both of these errors are factors requiring consideration. Such instances are discussed in the section below for the case of "Neither systematic error nor imprecision negligible."

In quoting a reported value and its associated uncertainty from the literature, the interpretation of the uncertainty quoted should be stated if given by the author. If the interpretation is not known, a remark to this effect is in order. This practice may induce authors to use more explicit formulations of their statements of uncertainty.

Standard Deviation and Standard Error

The terms *standard deviation* and *standard error* should be reserved to denote the canonical values for the measurement process, based on considerable recent experience with the measurement process or processes involved. When there is insufficient recent experience, an estimate of the standard error (standard deviation) must of necessity be computed by recognized statistical procedures from the same measurements as the reported value itself. To avoid possible misunderstanding, in such cases, the term "computed (or estimated) standard error" ("computed standard deviation") should be used. A formula for calculating this computed standard error is given in the section below for the case of "Neither systematic error nor imprecision negligible."

Uncertainties of Accepted Values of Fundamental Constants or Primary Standards

If the uncertainty in the accepted value of a national primary standard or of some fundamental constant of nature (for example, in the volt as maintained at the National Bureau of Standards, or in the acceleration of gravity g on the Potsdam basis) is an important source of systematic error affecting the measurement process, no allowance for possible systematic error from this source should be included ordinarily in evaluating overall bounds to the systematic error of the measurement process. Since the error concerned, whatever it is, affects all results obtained by the method of measurement involved, to include an allowance for this error would be to make everybody's results appear unduly inaccurate relative to each other. In such instances one should state: (i) that measurements obtained by the process concerned are expressed in terms of the volt (or the kilogram, or other unit) "as maintained at the National Bureau of Standards," or (ii) that the indicated bounds to the systematic error of the process are exclusive of the uncertainty of the stated value adopted for some particular constant or quantity. An example of the latter form of statement is:

. . . neglecting the uncertainty of the value 6.6256×10^{-34} joule seconds adopted for Planck's constant.

Systematic Error and Imprecision Both Negligible

In this case the reported result should be given, after rounding, to the number of significant figures consistent with the accuracy requirements of the situation, together with an explicit statement of its accuracy. An example is:

. . . the wavelengths of the principal visible lines of mercury-198 have been measured relative to the 6057.802106 Å (angstrom units) line of krypton-98, and their values in vacuum are

5792.2685 Å
5771.1984 Å
5462.2706 Å
4359.5625 Å
4047.7146 Å

correct to eight significant figures.

It needs to be emphasized that if no statement of accuracy or precision accompanies a reported number, then, in accordance with the usual conventions governing rounding, this number will ordinarily be interpreted as being accurate within ±½ unit in the last significant figure given; that is, it will be understood that its inaccuracy before rounding was less than ± 5 units in the next place. The statement "correct to eight significant figures" is included explicitly in the foregoing example, rather than left to be understood in order to forestall any concern that an explicit statement of lesser accuracy was inadvertently omitted.

Systematic Error Not Negligible, Imprecision Negligible

When the imprecision of a result is negligible, but the inherent systematic error of the measurement process concerned is not negligible, then the following rules are recommended:

1) Qualification of a reported result should be limited to a single quasi-absolute type of statement that places bounds on its inaccuracy.

2) These bounds should be stated to no more than two significant figures.

3) The reported result itself should be given (that is, rounded) to the last place affected by the stated bounds (unless it is desired to indicate and preserve such relative accuracy or precision of a higher order that it may possess for certain particular uses).

4) Accuracy statements should be given in sentence form in all cases, except when a number of results of different accuracies are presented, for example, in tabular arrangement. If it is necessary or desirable to indicate the respective accuracies of a number of results, the results should be given in the form $a \pm b$ (or $a \ {}^{+b}_{-c}$, if necessary) with an appropriate explanatory remark (as a footnote to the table, or incorporated in the accompanying text) to the effect that the $\pm b$, or ${}^{+b}_{-c}$, signify bounds to the systematic errors to which the a's may be subject.

5) The fact that the imprecision is negligible should be stated explicitly.

The particular form of the quasi-absolute type of statement employed in a given instance will depend ordinarily on personal taste, experience, current and past practice in the field of activity concerned, and so forth. Some examples of good practice are:

. . . is (are) not in error by more than 1 part in (x).
. . . is (are) accurate within ± (x units) [or ± (x) percent].
. . . is (are) believed accurate within (. . . .).

Positive wording, as in the first two of these quasi-absolute statements, is appropriate only when the stated bounds to the possible inaccuracy of the reported value are themselves reliably established. However, when the indicated bounds are somewhat conjectural, it is desirable to signify this fact (and put the reader on guard) by inclusion of some modifying expression such as "believed," "considered," "estimated to be," "thought to be," and

so forth, as exemplified by the third of the foregoing examples.

The term *uncertainty* may sometimes be used effectively to achieve a conciseness of expression otherwise difficult or impossible to attain. Thus, one might make a statement such as:

The uncertainties in the above values are not more than ± 0.5°C in the range 0°C to 1100°C, and then increase to ± 2°C at 1450°C,

or

The uncertainty in this value does not exceed . . . excluding (or, including) the uncertainty of . . . in the value . . . adopted for the (reference standard involved).

A statement giving numerical limits of uncertainty as in the above should be followed by a brief discussion telling how the limits were derived.

Finally, the following forms of quasi-absolute statements are considered poor practice, and are to be avoided:

The accuracy of . . . is 5 percent.
The accuracy of . . . is ± 2 percent.

These are presumably intended to mean that the result concerned is not inaccurate, that is, not in error, by more than 5 percent or 2 percent, respectively, but they explicitly state the opposite.

Neither Systematic Error Nor Imprecision Negligible

When neither the imprecision nor the systematic error of a result are negligible, then the following rules are recommended:

1) A reported result should be qualified by a quasi-absolute type of statement that places bounds on its systematic error, and a separate statement of its standard error or its probable error, or of an upper bound thereto, whenever a reliable determination of such value or bound is available. Otherwise a computed value of the standard error, or, probable error, so designated, should be given together with a statement of the number of degrees of freedom on which it is based.

2) The bounds to its systematic error and the measure of its imprecision should be stated to no more than two significant figures.

3) The reported result itself should be stated at most to the last place affected by the finer of the two qualifying statements (unless it is desired to indicate and preserve such relative accuracy or precision of a higher order that it may possess for certain particular uses).

4) The qualification of a reported result with respect to its imprecision and systematic error should be given in sentence form, except when results of different precision or with different bounds to their systematic errors are presented in tabular arrangement. If it is necessary or desirable to indicate their respective imprecisions or bounds to their respective systematic errors, such information may be given in a parallel column or columns, with appropriate identification.

Here, and in the next section, the term *standard error* is to be understood as signifying the standard deviation of the reported value itself, not as signifying the standard deviation of the single determination (unless, of course, the reported value is simply the result of a single determination).

The above recommendations should not be construed to exclude the presentation of a quasi-abolute type of statement placing bounds on the inaccuracy, that is, on the overall uncertainty, of a reported value, provided that separate statements of its imprecision and its possible systematic error are included also. To be in good taste, the bounds indicating the overall uncertainty should not be numerically less than the corresponding bounds placed on the systematic error outwardly increased by at least three times the standard error. The fourth of the following examples of good practice is an instance at point:

The standard errors of these values do not exceed 0.000004 inch, and their systematic errors are not in excess of 0.00002 inch.

The standard errors of these values are less than (x units), and their systematic errors are thought to be less than ± (y units). No additional uncertainty is assigned for the conversion to the chemical scale since the adopted conversion factor is taken as 1.000275 exactly.

. . . with a standard error of (x units), and a systematic error of not more than ± (y units).

. . . with an overall uncertainty of ± 3 percent based on a standard error of 0.5 percent and an allowance of ± 1.5 percent for systematic error.

When a reliably established value for the relevant standard error is available, and the dispersion of the present measurements is in keeping with this experience, then this canonical value of the standard error should be used (*5*). If such experience indicates that the standard error is subject to fluctuations greater than the intrinsic variation of such a measure, then an appropriate upper bound should be given, for example, as in the first two of the above examples, or by changing "a standard error . . ." in the third and fourth examples to "an upper bound to the standard error . . ."

When there is insufficient recent experience with the measurement processes involved, an estimate of the standard error must of necessity be computed by recognized statistical procedures from the same measurements as the reported value itself. It is essential that such computations be carried out according to an agreed-upon standard procedure, and the results thereof presented in sufficient detail to enable the reader to form his own judgment, and make his own allowances for their inherent uncertainties. To avoid possible misunderstanding, in such cases, first, the term *computed standard error* should be used; second, the estimate of the standard error employed should be that obtained from

$$\text{estimate of standard error} = \left(\frac{\text{sum of squared residuals}}{n\nu}\right)^{1/2}$$

where n is the (effective) number of completely independent determinations of which a is the arithmetic mean (or other appropriate least-squares adjusted value) and ν is the number of degrees of freedom involved in the sum of squared residuals (that is, the number of residuals minus the number of fitted constants or other independent constraints on the residuals); and third, the number of degrees of freedom should be explicitly stated. If the reported value a is the arithmetic mean, then:

$$\text{estimate of standard error} = (s^2/n)^{1/2}$$

where

$$s^2 = \sum_{i=1}^{n} (x_i - a)^2/(n-1)$$

and n is the number of completely independent determinations of which a is the arithmetic mean. For example:

. . . which is the arithmetic mean of (n) independent determinations and has a standard error of . . .

. . . with an overall uncertainty of ± 5.2 km/sec based on a standard error of 1.5 km/sec and estimated bounds of ± 0.7 km/sec on the systematic error. (The figure 5.2 is equal to 0.7 plus 3 times 1.5.)

or, if based on a computed standard error,

The computed probable error (or, standard error) of these values is (x units),

based on (ν) degrees of freedom, and the systematic error is estimated to be less than $\pm$ (y units).

. . . with an overall uncertainty of $\pm$ 7 km/sec derived from bounds of $\pm$ 0.7 km/sec on the systematic error and a computed standard error of 1.5 km/sec based on 9 degrees of freedom. [The number 7 is approximately equal to $0.7 + (4.3 \times 1.5)$, where 4.3 is the value of Student's t for 9 degrees of freedom exceeded in absolute value with 0.002 probability. As $\nu \to \infty$, $t_{.002}(\nu) \to 3.090$.]

When the reported value is the result of a complex measurement process and is obtained as a function of several quantities whose standard errors have been computed, these several quantities and their standard errors should usually be reported, together with a description of the method of computation by which the standard errors were combined to provide an overall estimate of imprecision for the reported value.

Systematic Error Negligible, Imprecision Not Negligible

When the systematic error of a result is negligible but its imprecision is not, the following rules are recommended:

1) Qualification of a reported value should be limited to a statement of its standard error or of an upper bound thereto, whenever a reliable determination of such value or bound is available. Otherwise a computed value of the standard error, so designated, should be given together with a statement of the number of degrees of freedom on which it is based.

2) The standard error or upper bound thereto, should be stated to not more than two significant figures.

3) The reported result itself should be stated at most to the last place affected by the stated value or bound to its imprecision (unless it is desired to indicate and preserve such relative precision of a higher order that it may possess for certain particular uses).

4) The qualification of a reported result with respect to its imprecision should be given in sentence form, except when results of different precision are presented in tabular arrangement and it is necessary or desirable to indicate their respective imprecisions in which event such information may be given in a parallel column or columns, with appropriate identification.

5) The fact that the systematic error is negligible should be stated explicitly.

The above recommendations should not be construed to exclude the presentation of a quasi-absolute type of statement placing bounds on its possible inaccuracy, provided that a separate statement of its imprecision is included also. To be in good taste, such bounds to its inaccuracy should be numerically equal to at least three times the stated standard error. The fourth of the following examples of good practice is an instance at point.

The standard errors of these values are less than (x units).

. . . with a standard error of (x units).

. . . with a computed standard error of (x units) based on (ν) degrees of freedom.

. . . with an overall uncertainty of $\pm$ 4.5 km/sec derived from a standard error of 1.5 km/sec. (The figure 4.5 is equal to 3×1.5.)

or, if based on a computed standard error,

. . . with an overall uncertainty of $\pm$ 6.5 km/sec derived from a computed standard error of 1.5 km/sec (based on 9 degrees of freedom). (The number 6.5 is equal to 4.3×1.5, where 4.3 is the value of Student's t for 9 degrees of freedom exceeded in absolute value with 0.002 probability. As $\nu \to \infty$, $t_{.002}(\nu) \to 3.090$.)

The remarks with regard to a computed standard error in the preceding section apply with equal force to the last two examples above.

Conclusion

The foregoing recommendations call for fuller and sharper detail than is general in common pactice. They should be regarded as minimum standards of good practice. Of course, many instances require fuller treatment than that recommended here.

Thus, in the case of determinations of the "fundamental physical constants" and other basic properties of nature, the author or authors should give a detailed account of the various components of imprecision and systematic error, and list their respective individual magnitudes in tabular form, so that (i) the state of the art will be more clearly revealed, (ii) each individual user of the final result may decide for himself which of the indicated components of imprecision or systematic error are, or are not, relevant to his use of the final result, and (iii)—most important—the final result itself or its uncertainty can be modified appropriately in the light of later advances. This is, and has long been, the practice followed in the best reports of fundamental studies, but current efforts to prepare critically evaluated standard reference data have revealed that far too great a fraction of the data in the scientific literature "cannot be critically evaluated because the minimum of essential information is not present" (*6*).

References and Notes

1. The true value defined conceptually by an exemplar measurement process, or the target value intended in a practical measurement process.
2. The standard error is the standard deviation of the probability distribution of estimates (that is, reported values) of the quantity that is being measured. See M. G. Kendall and W. R. Buckland, *A Dictionary of Statistical Terms* (Hafner, New York, 1957).
3. For a comprehensive discussion on precision and accuracy, and a selected bibliography of 80 references, see C. Eisenhart, "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," *J. Res. Nat. Bur. Std.* **67C**, No. 2, 161–187 (1963). (Reprints are available upon request.)
4. The essential elements of the present recommendations first appeared in a 1955 National Bureau of Standards task group report prepared principally by Malcolm W. Jensen (Office of Weights and Measures), Leroy W. Tilton (Optics and Metrology Division), and Churchill Eisenhart (Applied Mathematics Division), which was based for the most part on detailed recommendations developed some years earlier by Dr. Tilton for the internal guidance of the Optics and Metrology Division. In September 1961, new introductory material was added to the recommendations of the 1955 task group; a few minor changes were made in the illustrative examples, and the resulting revised version was circulated as a working paper of the Subcommittee on Accuracy Statements of the NBS Testing and Calibration Committee. This 1961 version was incorporated without essential change as chapter 23, "Expression of the Uncertainties of Final Results," of NBS Handbook 91, *Experimental Statistics* (U.S. Government Printing Office, Washington, 1963), reprinted with corrections in 1966. (This handbook brought together in a single volume the material on experimental statistics prepared at the National Bureau of Standards for the U.S. Army *Ordnance Engineering Design Handbook*, and printed in 1962 for limited distribution as U.S. Army Ordnance Corps Pamphlets ORDP 20-110 through 20-114. Subsequently, when these five pamphlets became parts of the *AMC Engineering Design Handbook*, they were designated Army Materiel Command Pamphlets AMCP 706-110 through 706-114.)

   In the present version, the content of chapter 23 has been rearranged and, in order to be more appropriate to calibration work, more explicit consideration has been given to the case where the value of the standard deviation σ of the measurement process involved has been well established by recent past experience. A terse summary of the principal recommendations of the present paper in the form of a text figure (Fig. 1) is contained in H. H. Ku, "Expressions of Imprecision, Systematic Error, and Uncertainty Associated with a Reported Value," to be published in *Measurements and Data*. The earlier versions were addressed primarily to the case of isolated experiments or tests, where the relevant value of σ is usually unknown in advance, and the statistical uncertainty of the final results must therefore be expressed entirely in terms of quantities derived from the data of the experiment itself.
5. The control chart is an invaluable tool in providing justification for the use of a canonical value of the standard error. See, for example, ASTM Manual on *Quality Control of Materials* (American Society for Testing and Materials, Philadephia, 1951).
6. L. M. Branscomb, "The misinformation explosion: Is the literature worth reviewing?," a talk presented to the Philosophical Society of Washington, 17 November 1967, and to be published in *Scientific Research*.

EXPRESSIONS OF IMPRECISION, SYSTEMATIC ERROR, AND UNCERTAINTY ASSOCIATED WITH A REPORTED VALUE

HARRY H. KU, National Bureau of Standards

The work of a calibration laboratory may be thought of as a sequence of operations that result in the collection, storage, and transmittal of information. In making a statement of uncertainty of the result of calibration, the calibration laboratory transmits information to its clients on the particular item calibrated.

It is logical, then, to require the transmitted information to be meaningful and unambiguous, and to contain all the relevant information in the possession of the laboratory. *The information content of the statement of uncertainty determines, to a large extent, the worth of the calibrated value.*

A common deficiency in many statements of uncertainty is that they do not convey all the information a calibration laboratory has to offer, information acquired through much ingenuity and hard work. This deficiency usually originates in two ways:

1. Loss of information through oversimplification, and

2. loss of information through the inability of the laboratory to take into account information accumulated from its past experience.

With the increasingly stringent demands for improved precision and accuracy of calibration work, calibration laboratories as a whole just cannot afford such luxury.

Traceability to the national standards, accuracy ratios, and class tolerance requirements are simplified concepts that aim to achieve different degrees of accuracy requirements. These concepts and the resulting statements are useful on certain occasions, but fail whenever the demand is exacting. The general practice of obliterating all the identifiable components of uncertainty, by combining them into an overall uncertainty, just for the sake of simplicity, is another case in point. After all, if the calibration laboratory reports all the pertinent information in separate components, the user can always combine them or use them individually, as he sees fit. On the other hand, if the user is given only one number, he can never disentangle this number into its various components. Since the information buried under these oversimplified statements is available, and may well be useful to sophisticated customers, such practices result in substantial waste of effort and resources.

In calibrating an item by repeating the same calibration procedure, the calibration laboratory gains increments of information about its calibration *system*. These increments of information are quantified and accumulated for the benefit of the calibration laboratory. If the precision of the calibration process remains unchanged, the statistical measure of dispersion (s) – i.e., the standard deviations computed from these sets of data – can be pooled together, weighted by their respective degrees of freedom. When many such increments of information are combined, an accepted or canonical value of standard deviation (σ) is established. This established (canonical) value of standard deviation characterizes the precision of the calibration process, and is treasured information in any calibration laboratory.

Hence, the canonical value of standard deviation is the quantification of information accumulated from past experiences of the calibration laboratory, and is an essential element of the statement of uncertainty. The standard deviation (s) computed from the current calibration is used to check the precision of current work, and to add to the pool of information on the process, but certainly does not represent all the information available in the possession of an established calibration laboratory. Only by passing its accumulated information to the users is the calibration laboratory performing a complete service.

STATEMENT OF UNCERTAINTY

In the preparation of a statement of uncertainty, it is helpful to bear in mind that:

1. The derivation of a statement of uncertainty has as its foundation the work done in the laboratory, and is based on information accumulated from past experience, and

2. In general, information is lost through oversimplification, and demands for improved precision and accuracy cannot be met with simplified statements of uncertainty.

Unless a statement of uncertainty is well formulated and supported, it is difficult to say what is meant by the statement, a difficulty frequently encountered. Since the evaluation of uncertainty is part and parcel of the high standard of work of a calibration laboratory, the statement of uncertainty deserves all the attention required to make the statement both realistic and useful. To this end, Tables 1, 2 and 3 give terms and expressions compiled as a ready reference for those who are searching for some appropriate format or wording, to carry out the thoughts expressed. They summarize the recommended practices on expression of uncertainties as given in Chapter 23 of NBS Handbook 91. A revised version of this chapter with the title "Expression of Uncertainties of Final Results" by Churchill Eisenhart may be found in NBS Special Publication 300-1. Figure 1 gives a condensed summary of this material. Tables 1, 2, and 3 give details of forms of imprecision, systematic error, and uncertainty statements.

TABLE 1 – IMPRECISION STATEMENTS

Value reported	Index or Measure of Error	Remarks
Precision of a measurement (calibration) process	(a). Standard deviation (σ) of a single determination (observation)	σ (or s with the associated degrees of freedom[1]) is of main interest as an index of precision of the measurement process. If the average of n such measurements is also reported, see (b) below.
Arithmetic mean ($\bar{x}_n$) of n numbers	(b). Standard error ($\sigma/\sqrt{n}$) of the reported value	$\bar{x}_n$ is of main interest; the number n is also essential information; σ assumed known.[1]
	(c). 2 sigma limits (d). 3 sigma limits	Commonly used bounds of imprecisions; usually used when σ known, or when n large.
	(e). Confidence interval (indicate one- or two-sided)	Data points assumed to be normally distributed; report confidence coefficient (level) $100(1-\alpha)\%$. [2]
	(f). Half-width of confidence interval (or confidence limits)	Same as (e) above; for symmetrical two-sided intervals; an index to bounds of imprecision. [2]
	(g). Probable error of the reported value	Probable error $= .6745\,\frac{\sigma}{\sqrt{n}}$ for normally distributed data points when σ known. Use of $\sigma/\sqrt{n}$ preferred. Incorrect if σ not known.
	(h). Mean deviation, or average deviation, of a measurement from the mean calculated from the sample	Limiting mean of mean deviation $= \sqrt{\frac{2}{\pi}}\,\sqrt{\frac{n-1}{n}}\,\sigma$ for normally distributed data points when σ known. Use of σ usually preferred.
	(i). Any of the above expressed in percent, or ppm of $\bar{x}_n$.	State what is being expressed in percent, eg., $(\sigma/\sqrt{n})(100/\bar{x}_n)$, $\bar{x}_n$ being a fairly constant value.
m means each computed from n measurements	(j). (b), (c), (d) and (f) above	If the measurements are of equal precision and σ unknown, use $s_p^2 = \frac{1}{m}\sum_{i=1}^{m} s_i^2$ as estimate of σ^2. The no. of degrees of freedom associated with s_p is $m(n-1)$.
	(k). Sample coefficient of variation $(v = \frac{s}{\bar{x}_n})$ or relative percent $(v \times 100)$	Appropriate when the m means cover a wide range and where the v's computed for the m sets are about the same magnitude. Give range of v's for the m sets. The means must be positive and bounded away from zero.
Weighted mean $\bar{\bar{x}} = \frac{w_1\bar{x}_1 + w_2\bar{x}_2}{w_1 + w_2}$	(l). Standard error ($\sigma_{\bar{\bar{x}}}^2$) of the weighted mean	If $w_1 = 1/\sigma_{\bar{x}_1}^2$ and $w_2 = 1/\sigma_{\bar{x}_2}^2$, then $\sigma_{\bar{\bar{x}}}^2 = \frac{1}{w_1+w_2}$. Not recommended when the σ's are not known and are estimated by s computed from small number of measurements.
An equation (theoretical or empirical) fitted to data points by the method of least squares	(m). Standard deviation computed from the deviations (residuals) of data points from the fitted curve	Report n, the number of data points, and k, the number of constants fitted, $s^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2/(n-k)$, where $\hat{y}_i$ is the value on the fitted curve for the particular x_i.[3] Value of s usually given in computer print-out.
Constants (coefficients) in the equation fitted to the data points by the method of least squares	(n). Standard errors of the coefficients based on the standard deviation computed under (m)	Standard errors usually given in computer print-out. Report n and k as above.[3]

TABLE 1 – IMPRECISION STATEMENTS – (Continued)

Value reported	Index or Measure of Error	Remarks
A predicted point on the curve $\hat{y}$ for a particular x_o	(o). Standard error ($s_{\hat{y}}$) of the predicted point	For the straight line case, the computer print-out gives the variance-covariance matrix $\begin{pmatrix} s_{11} & \\ s_{12} & s_{22} \end{pmatrix}$. $s_{\hat{y}}^2 = s_{11} + 2s_{12}x_o + s_{22}x_o^2$.[3] Report n and k.
A predicted observed value for a particular x_o	(p). Standard error of the predicted value of y	For the straight line case, $s_y^2 = s_{\hat{y}}^2 + s^2$ where $s_{\hat{y}}^2$ and s^2 are that given in (o) and (m) respectively.[3] Report n and k
Value of function of the arithmetic means of several measured variables	(q). Standard error calculated by the use of propagation of error formulas	Appropriate when errors of measurements are small compared to the values of variables measured. Use standard error of the means of the variables in the formulas.[4] Report number of measurements from which these standard errors are computed.
Percentage or proportion (r/n), r and n being counts	(r). Confidence limits of the true proportion P	Procedures for obtaining exact and approximate confidence limits are discussed in Chapter 7, NBS Handbook 91. State one-sided or two-sided.

TABLE 2 – SYSTEMATIC ERROR[5] (BIAS) STATEMENTS

Value reported	Index or Measure of Error	Remarks
Numerical value resulting from a measurement process	Reasonable bounds ascribed to the value originating from: (i). systematic error reliably established	Detailed discussions of systematic errors are always helpful. Positive wording is appropriate: ". . . is not in error by more than . . ." ". . . is accurate within ± . . ."
	(ii). systematic error estimated from experience or by judgment	Use modifier such as "believed", "estimated", "considered", to signify the conjectural nature of the statement.
	(iii). combination of a number of elemental systematic errors	State explicitly the method of combination such as "the simple sum of the bounds" or "the square root of the sum of squares".
	(iv). uncertainty in some fundamental constant	Give reference to the value of constant used.
	(v). uncertainty in calibrated values	Ascertain the meaning of the systematic and random components of the uncertainty from the calibration laboratory so that decisions on the uses of these components can be made from the correct interpretations.
	(vi). bias in the method of computation	Correct if feasible, or give the magnitude.

TABLE 3 – UNCERTAINTY STATEMENTS

Value reported	Index or Measure of error	Remarks
Numerical value resulting from a measurement process	Bounds to inaccuracy: (1). Systematic error and imprecision both negligible	Explicit expression of correctness to the last significant figure, interpreted as being accurate within $\pm 1/2$ units in the last significant figure given.
	(2). Imprecision negligible. Bounds on inaccuracy given to no more than two significant figures.	Sentence form preferred such as given under remark for (i) and (ii). Footnote needed if bounds are given in tabular form.
	(3). Systematic error negligible. Index of precision (b), (g), (h), (i), (k), or (n) stated to no more than two significant figures	State explicitly the index used and give essential information associated with the index. Qualify index calculated by the word "computed". Avoid using expressions of the form $a \pm b$ unless the meaning of b is explained fully immediately following or in footnote.
	(3'). Systematic error negligible. Bounds to imprecision (c), (d), (e), or (f) stated to no more than two significant figures.	Same as under (3).
	(4). Neither systematic error nor imprecision negligible. Two numerics indicating bounds to systematic error and index of imprecision respectively	(2) and (3) above separately stated.
	(4'). Bounds to systematic error and imprecision combined, indicating the likely inaccuracy of the value	(2) and (3') above where the two components either have been previously described, or explained immediately following (or in footnote).
	(5). Quoted from literature	State reference and give author's interpretation of the uncertainty; add remark if meaning unknown or ambiguous.

[1] If σ is not known, use the computed standard deviation s based on k measurements as an estimate of σ, where $s^2 = \frac{1}{k-1} \sum_{i=1}^{k} (x_i - \bar{x}_k)^2$. The number (k–1) is the degrees of freedom associated with s.

[2] For interpretation see Chapter 1, NBS Handbook 91, **Experimental Statistics**, by M. G. Natrella, 1963.

[3] For details see Chapter 5 (straight line), and Chapter 6 (multivariate and polynomial), NBS Handbook 91.

[4] For details see "Notes on the use of propagation of error formulas", by Harry H. Ku, NBS Journal of Research, Vol. 70C, No. 4, October-December, 1966.

[5] See "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems" by Churchill Eisenhart, NBS Journal of Research, Vol. 67C, No. 2, April-June, 1963, and "Systematic Errors in Physical Constants" by W. J. Youden, Physics Today 14, 1961.

FIGURE 1 – SUMMARY OF RECOMMENDATIONS ON EXPRESSIONS OF THE UNCERTAINTIES OF FINAL RESULTS

SYSTEMATIC ERROR AND IMPRECISION BOTH NEGLIGIBLE (CASE 1)

In this case, the reported result should be given correct to the number of significant figures consistent with the accuracy requirements of the situation, together with an explicit statement of its accuracy or correctness.

SYSTEMATIC ERROR NOT NEGLIGIBLE, IMPRECISION NEGLIGIBLE (CASE 2)

(a) Qualification of a reported result should be limited to a single quasi-absolute type of statement that places bounds on its inaccuracy;

(b) These bounds should be stated to no more than two significant figures;

(c) The reported result itself should be given (i.e., rounded) to the last place affected by the stated bounds, unless it is desired to indicate and preserve such relative accuracy or precision of a higher order that the result may possess for certain particular uses;

(d) Accuracy statements should be given in sentence form in all cases, except when a number of results of different accuracies are presented, e.g., in tabular arrangement. If it is necessary or desirable to indicate the respective accuracies of a number of results, the results should be given in the form $a \pm b$ (or $a \pm \frac{b}{c}$, if necessary) with an appropriate explanatory remark (as a footnote to the table, or incorporated in the accompanying test) to the effect that the $\pm b$, or $\pm \frac{b}{c}$, signify bounds to the errors which the a's may be subject.

(e) The fact that the imprecision is negligible should be stated explicity.

NEITHER SYSTEMATIC ERROR NOR IMPRECISION NEGLIGIBLE (CASE 3)

(a) A reported result should be qualified by: (1) a quasi-absolute type of statement that places bounds on its systematic error; and, (2) a separate statement of its standard error or of an upper bound thereto, whenever a reliable determination of such value or bound is available — otherwise, a computed value of the standard error so designated should be given, together with a statement of a number of degrees of freedom on which it is based;

(b) The bounds to its systematic error and the measure of its imprecision should be stated to no more than two significant figures;

(c) The reported result itself should be stated, at most, to the last place affected by the finer of the two qualifying statements, unless it is desired to indicate and preserve such relative accuracy or precision of a higher order that the result may possess for certain particular uses;

(d) The qualification of a reported result, with respect to its imprecision and systematic error, should be given in sentence form, except when results of different precision or with different bounds to their systematic errors are presented in tabular arrangement. If it is necessary or desirable to indicate their respective imprecisions or bounds to their respective systematic errors, such information may be given in a parallel column or columns, with appropriate identification.

SYSTEMATIC ERROR NEGLIGIBLE, IMPRECISION NOT NEGLIGIBLE (CASE 4)

(a) Qualification of a reported value should be limited to a statement of its standard error or of an upper bound thereto, whenever a reliable determination of such value or bound is available. Otherwise, a computed value of the standard error so designated should be given, together with a statement of the number of degrees of freedom on which it is based;

(b) The standard error, or upper bound thereto, should be stated to not more than two significant figures;

(c) The reported result itself should be stated, at most, to the last place affected by the stated value or bound to its imprecision, unless it is desired to indicate and preserve such relative precision of a higher order that the result may possess for certain particular uses;

(d) The qualification of a reported result with respect to its imprecision should be given in sentence form, except when results of different precision are presented in tabular arrangement and it is necessary or desirable to indicate their respective imprecisions, in which event such information may be given in a parallel column or columns, with appropriate identification.

(e) The fact that the systematic error is negligible should be stated explicitly.

POSTSCRIPT

Over the intervening years since the publication of Eisenhart's and Ku's articles, it has become apparent that a few additional comments may be useful. It is equally apparent that a complete revision is neither necessary nor desirable inasmuch as the major thrust and content of the articles remain as valid and as appropriate as when first written. For this reason, these comments are made as a postscript.

Uncertainty Assessments Must Be Complete

The uncertainty of a reported value is meant to be a credible estimate of the likely limits to its actual *error,* i.e., the magnitude and sign of its deviation from the truth. As such, uncertainty statements must be based on as nearly complete an assessment as possible. This assessment process must consider every conceivable source of inaccuracy in the result.

A measurement process generally consists of a very complicated sequence of many individual unit operations or steps. Virtually every step in this sequence introduces a conceivable source of inaccuracy whose magnitude must be assessed. These sources include:

- Inherent stochastic variability of the measurement process;
- Uncertainties in standards and calibrated apparatus;
- Effects of environmental factors, such as variations in temperature, humidity, atmospheric pressure, and power supply voltage;
- Time-dependent instabilities due to gradual and subtle changes in standards or apparatus;
- Inability to realize physical model because of instrument limitations;
- Methodology procedural errors, such as incorrect logic, or misunderstanding what one is or should be doing;
- Uncertainties arising from interferences, impurities, inhomogeneity, inadequate resolution, incomplete discrimination, etc.;
- Metrologist errors, such as misreading of an instrument;
- Malfunctioning or damaged apparatus;
- Laboratory practice including handling techniques, cleanliness, etc.; and
- Computational uncertainties as well as errors in transcription of data, and other calculational or arithmetical mistakes.

This list should not be interpreted as exhaustive, but rather as illustrative of the most common generic sources of inaccuracy that may be present.

The various sources of inaccuracy are generally classified into sources of *imprecision* (random components) and sources of *bias* (fixed offsets). To which category a particular source should be properly assigned is often difficult and troublesome. In part, this is because many experimental procedures or individual steps in the overall measurement process embody both systematic and stochastic (random) elements. (For an alternative discussion that questions the need for a clear cut distinction between random and systematic components of uncertainty, see [7].) One practical approach is to classify the sources of inaccuracy according to how the uncertainty is estimated. In this way, sources of imprecision are considered to be those components which *can be* and *are* estimated by a statistical analysis of replicate determinations. For completeness, the *systematic uncertainty components* can be considered to be the residual set of conceivable sources of inaccuracy that are biased and not subject to random variability, and those that may be due to random causes but *cannot be* or *are not* assessed by statistical methods. The systematic category includes sources of inaccuracy other than biases in order to obtain a complete accounting of all sources of inaccuracy in the measurement process. Hence, it is meaningful to report a random uncertainty contribution, only if one has a computed statistic for the magnitude of its imprecision or random variation. Many sources of inaccuracy may exist consisting of several components from both the random and systematic categories and can be assessed only after consideration of the more fundamental processes involved. The uncertainty in the calibration of an instrument with a standard reference material, for example, would have not only components from the uncertainty in the standard itself, but also uncertainty components arising from the use of the standard in performing the calibration.

Assessment of Imprecision (Random Uncertainties)

Although the treatment and expressions of reporting the imprecision of measurement results were adequately covered in the original article, a number of points are of sufficient importance to deserve reemphasis.

The only way to assess realistically the overall imprecision is to make direct—or preferably, when possible, indirect—replicate determinations [1] and calculate an appropriate statistic such as the standard error of the mean. It is extremely important to be definite on what constitutes a "replicate determination" because the extent to which conditions are allowed to vary freely over successive "repetitions" of the measurement process determines the scope of the statistical inferences that may be drawn from measurements obtained [2, sec. 4.1]. When measurements of a particular quantity made on a single occasion exhibit closer mutual agreement than measurements made on different occasions so that differences between occasions are indicated, the value of the computed standard error of the mean of all the measurements obtained by lumping all of the measurements together will underestimate the actual standard error of the mean. A more realistic value is given by taking the arithmetic means of the measurements obtained on the respective occasions as the *replicate determinations* and calculating the standard error of their mean in the usual way [3, sec. 3.5].

In many situations, it may not be possible or feasible because of time and cost constraints to perform a sufficient number of completely independent determinations of the measurement result. For results derived from several component quantities, the individual imprecision estimates must be propagated to obtain the imprecision of the final result. It must be emphasized, however, that

these estimates of imprecision should not be based exclusively on the information derived from just the present measurements. Presently derived information should be added to the information accumulated in the past on the imprecision of the measurement process. In this way, more realistic and reliable canonical values of the imprecision statistics may be established over time. Ideally, every major step or component of the measurement process should be independently assessed. This would include not only the variability inherent in the particular measurement of concern, but also the imprecision arising from corrections, calibration factors, and any other quantities that make up the final result.

Assessment of Systematic Uncertainties

Although a general guideline for the approach to the assessment of systematic uncertainties can be formulated, there are, unfortunately, no rules to objectively assign a magnitude to them. For the most part, it is a subjective process. Their magnitudes should preferably be based on experimental verification, but may have to rely on the judgment and experience of the metrologist. In general, each systematic uncertainty contribution is considered as a quasi-absolute upper bound, overall or maximum limit on its inaccuracy. Its magnitude is typically estimated in terms of an interval from plus to minus δ about the mean of the measurement result. By what method then should the magnitude of these maximum limits be assigned? It may be based on comparison to a standard, on experiments designed for the purpose [4], or on verification with two or more independent and reliable measurement methods. Additionally, the limits may be based on judgment, based on experience, based on intuition, or based on other measurements and data. Or the limits may include combinations of some or all of the above factors. Whenever possible, they should be empirically derived or verified. The reliability of the estimate of the systematic uncertainty will largely depend on the resourcefulness and ingenuity of the metrologist.

The Need for an Overall Uncertainty Statement

Without deprecating the perils of shorthand expressions, there is often a need for an overall uncertainty statement which combines the imprecision and systematic uncertainty components. Arguments that it is incorrect from a theoretical point of view to combine the individual components in any fashion are not always practical. First, an approach which retains all details is not amenable for large compilations of results from numerous sources. And second, this approach shifts the burden of evaluating the uncertainties to users. Many users need a single uncertainty value resulting from the combination of all sources of inaccuracy. These users believe, and rightly so, that this overall estimate of inaccuracy can be most appropriately made by the person responsible for the measurement result. It must be emphasized, however, that there is no one clearly superior appropriate method for reporting an overall uncertainty, and that the choice of method is somewhat arbitrary. Several methods are commonly employed [5,6].

One method is to add linearly all components of the systematic uncertainty and linearly add the total to the imprecision estimate. Since the individual systematic uncertainties (δ_j) are considered to be maximum limits, it logically should be added to an imprecision estimate at a similar confidence level. That is, for example, the overall uncertainty u may be given by

$$u = [t_\nu(\alpha)]s + \sum_{j=1}^{q} \delta_j$$

where s is the computed standard error based on ν degrees of freedom, $t_\nu(\alpha)$ is the Student-t value corresponding to a two-tail significance level of $\alpha = 0.05$, 0.01, or 0.001 (depending on the practice in the measurement field concerned), and δ_j is the magnitude of the estimated systematic uncertainty for each of the identified q systematic uncertainty components. This approach probably overestimates the inaccuracy, but can be considered as an estimate of the maximum possible limits. For example, if someone estimated that five contributions of about equal magnitude made up the total systematic error, that person would have to be very unlucky if all five were plus, or all five were minus. Yet, if there was one dominant contributor, it might be a very valid approximation.

Two other approaches have also been widely used. These methods add in quadrature all of the systematic uncertainty components, and either add the resulting quantity *linearly* to the standard error estimate,

$$s + \sqrt{\sum_{j=1}^{q} \delta_j^2} \quad ,$$

or add it *in quadrature* to the standard error estimate,

$$\sqrt{s^2 + \sum_{j=1}^{q} \delta_j^2}$$

These are frequently considered (erroneously) to correspond to a confidence level with $P = 68\%$.

In another method, often termed the PTB approach [6], the component systematic uncertainties are assumed to be independent and distributed such that all values within the estimated limits are equiprobable (rectangular or uniform distribution) [8]. With these assumptions, the rectangular systematic uncertainty distributions can be convoluted to obtain a combined probability distribution for which the variance may be computed. This may then be combined in quadrature with that for the random uncertainty. In its simplest form, the uncertainty components are combined to form an overall uncertainty by

$$u = k\sqrt{s^2 + (1/3)\sum_{j=1}^{q} \delta_j^2} \quad ,$$

where k is customarily taken as 2 or 3. The above simple form is not appropriate when one of the component δ_j's is much larger than the others; in such a case it will be more informative to keep that component separate from the others and add it linearly.

A Concluding Thought

If there is one fundamental proposition for the expression of uncertainties, it is

> The information content of the statement of uncertainty determines, to a large extent, the worth of the final result.

This information content can be maximized by following a few simple principles:

BE EXPLICIT

PROVIDE DETAILS

DON'T OVERSIMPLIFY

When an overall uncertainty is reported, one should explicitly state how the separate components were combined. In addition, for results of primary importance, a detailed discussion and complete specification of all of the separate uncertainty components is still required. In this way, some users will benefit from having the metrologist's estimate of the overall uncertainty, while more sophisticated users will still have access to all of the information necessary for them to evaluate, combine, or use the uncertainties as they see fit.

REFERENCES AND NOTES

[1] Youden, W. J. Statistical aspects of analytical determinations. The Analyst 77: 874–878; 1952, Dec: Reprinted in Journal of Quality Technology 4: 45–49; 1972, Jan: Youden, W. J., Connor, W. S., Making one measurement do the work of two; Chemical and Engineering Progress 49: 549–552; 1953, Oct. Reprinted in Journal of Quality Technology 4: 25–28; 1972, Jan.

[2] Eisenhart, Churchill. Realistic evaluation of the precision and accuracy of instrument calibration systems. J. Res. Nat. Bur. Stand. (U.S.) 67C (2): 161–187; 1963; Reprinted as paper 1.2 in NBS Special Publication 300-1.

[3] Eisenhart, Churchill. Contribution to panel discussion of adjustments of the fundamental physical constants. Langenberg, D. N., Taylor, B. N., eds. Precision Measurement and Fundamental Constants, Nat. Bur. Stand. (U.S.) Spec. Publ. 343: 509–518; 1971.

[4] Youden's burette experiment, Journal of Quality Technology 4: 20-23 1972, Jan. Youden, W. J., Systematic errors in physical constants. Physics Today, 14, 32–34, 36, 38, 40, 43; 1961, Sept. Reprinted as paper 1.4 in NBS Special Publication 300-1. Youden, W. J., Enduring values, Technometrics 14: 1–11; 1972, Feb.

[5] Campion, P. J.; Burns, J. E.; Williams, A. A code of practice for the detailed statement of accuracy. National Physical Laboratory. London: Her Majesty's Stationery Office. 1953. II-57.

[6] Wagner, Siegfried R. On the quantitative characterization of the uncertainty of experimental results in metrology: PTB-Mitteilungen 89: 83–89; 1979, Feb.

[7] Müller, Jörg G. Some second thoughts on error statements. Nuclear Instruments and Methods 163, 241–251; 1979.

[8] A numerical comparison of uncertainty limits resulting from these assumptions with those implied by several alternative distributional assumptions is provided by table 1 on page 184 of [2], and discussed on the same and following page.

Churchill Eisenhart
Ronald Collé

July 1980

VOLTAGE METROLOGY

NBS Technical Note 1239

Solid-State Voltage Standard Performance and Design Guidelines

Bruce F. Field

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

September 1987

Sponsored in part by:
DC and Low Frequency Subgroup of the Calibration
Coordination Group, Department of Defense

U.S. Department of Commerce
Clarence J. Brown, Acting Secretary

National Bureau of Standards
Ernest Ambler, Director

Table of Contents

List of Figures

List of Tables

NBS MEASUREMENT SERVICES: SOLID-STATE DC VOLTAGE STANDARD CALIBRATIONS

Bruce F. Field

Abstract: This document describes the procedures used at NBS to calibrate solid-state dc voltage standards in terms of the U.S. Legal Volt. The process involves calibrating client standards at NBS approximately 10 times over a two week period. The operational procedures and apparatus used to compare the client standards to the U.S. Legal Volt are discussed in detail.

1. Introduction

DC voltage standards based on solid-state devices are now available with stabilities approaching 1 ppm/year. Such voltage standards can be used to maintain and disseminate (transfer) a laboratory unit of voltage with accuracies approaching those obtainable with saturated standard cells. Presently-available voltage standards contain at least one (and sometimes more) Zener diode reference devices which develop a voltage that is scaled using internal resistive dividers and low-noise amplifiers to produce a range of output voltages from 1 to 10 V. These standards typically have their best accuracy at some voltage other than the 1.01-V level of standard cells. NBS has developed a calibration service based on a specialized measuring system that automatically compares the output of any arbitrary voltage source, within the voltage range of one to ten volts, to a 1.018-V reference standard.

Solid-state voltage standards that are accepted for calibration must be complete instruments, must be powered by the ac line or internal batteries, and must continuously produce one or more stable voltages. Such standards, when sent to NBS, are allowed to stabilize, and then are directly compared to NBS standards which are in turn calibrated in terms of the U.S. Legal Volt. Typically ten comparisons are made (once each working day) over a two week period. At the conclusion of the test a report is issued with a statement of the values of the voltage outputs of the standard and the estimated uncertainties for the values while the standard was at NBS. No additional uncertainty is included for transportation effects, long term drifts of the outputs, or measurement errors in the client's laboratory.

The calibration of solid-state standards is closely tied to, and dependent on, the calibration of working groups of standard cells. The reader is thus referred to reference 1, NBS SP250-24, *NBS Measurement Services: Standard Cell Calibrations*, for a complete explanation of the apparatus and procedures used for the calibration of standard cells.

2. Description of Service

Types of Standards Accepted for Calibration

NBS offers a calibration service for solid-state voltage standards that are complete functioning standards which continuously produce the output voltages to be measured. We do not accept bare reference devices or modules that require a stabilized power supply for operation. Also not accepted are instruments with multiple outputs that must be manually switched to one set of output terminals. Each output voltage to be measured must be continuously available at a set of terminals dedicated to only that output. A potentiometric (non-loading) method is used for the measurement of the standards so there is no requirement that the standard be capable of delivering current to the measuring system. Standards having output resistances as high as 1 kΩ can be accurately measured.

Shipment of Standards

Solid-state standards may be carried to NBS by hand, or shipped to NBS by common carrier; arrangements can be made with a local delivery service for pickup and delivery to local airports. Standards may be shipped under battery power to maintain temperature control and/or to power the reference device, or they may be shipped with all power turned off. If they are shipped with the power turned off we have observed that a short (2-3 days) settling time is occasionally necessary before high-accuracy measurements can be started.

Calibration Procedures

Most solid-state standards operate with the reference device in a temperature-controlled oven. The temperature of the oven is recorded when the standard is first put on test (if the standard contains a temperature measuring device), but unless specially requested, the temperature of the oven is not measured or recorded during the remainder of the test.

Routine calibrations of solid-state standards typically take 2 to 3 weeks to complete. The voltages are read daily for a period of ten working days. After ten readings the data are reviewed by NBS personnel and if the stability and day-to-day scatter of the outputs are within acceptable limits, a calibration report is issued. If the measured voltages fluctuated or drifted unduly during the measurement period, the report will reflect these circumstances. In some cases additional measurements may be taken to confirm the existence of a problem. If the standard exhibits an initial drift but later becomes stable (indicating a transportation problem), additional measurements will be taken until at least ten stable readings have been obtained. Only in rare circumstances would more than twenty measurements be taken; if the standard has not stabilized by that time it is likely that it is de-

fective. Only the stable readings are used for the final values in the calibration report.

3. Calibration Apparatus and Measurement Approach

System Description

The system used at NBS to calibrate solid-state voltage standards is specially designed; it employs a potentiometric method to measure the standard which allows voltage measurements to be made without loading the standard. (This system is also described in reference 2). The system is basically a scaling device for comparing arbitrary voltage standards to 1.018-V standard cells. It will measure any dc voltage in the range from 0 to 10 V, and has input channels for up to 54 unknown standards.

The system is also capable of making high-accuracy standard cell comparisons. Standard cells may be connected to the input channels of the scanner in place of (or in addition to) solid-state standards and these cells may be intercompared with one another.

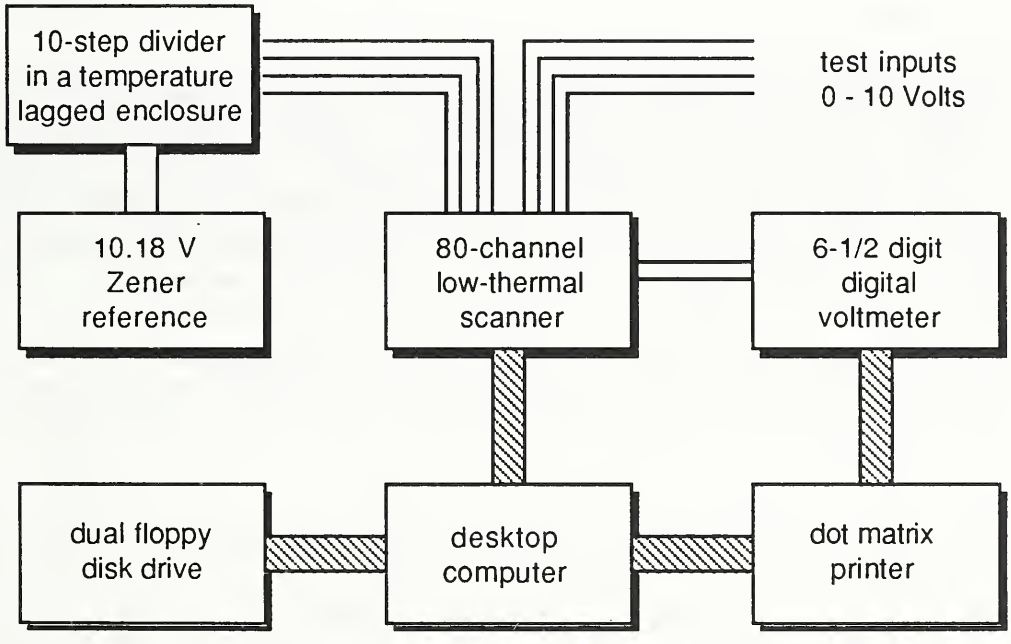


Fig. 1. Block diagram of the dc voltage measuring system.

Figure 1 is a simplified block diagram of the measurement system. It includes a modified 10.18-V Zener reference, a 10-step resistive divider consisting of 10 nominally equal 250-Ω resistors, a low-thermal-emf crossbar scanner, a 6-1/2 digit digital voltmeter (DVM), and a desktop computer. The scanner, voltmeter, disk drive, and printer are connected to the computer with an IEEE-488 bus system. Not shown in the figure is an addition-

al connection to a 16 Mbyte hard disk which is common to several measuring systems in the laboratory and which is used for program and data storage. The floppy disk system is used for backup data storage in addition to the main storage on the hard disk.

Also not shown in Fig. 1 is the group of four saturated standard cells used as a reference for the system (Working Group C). All solid-state voltage measurements are expressed as a numerical ratio times the mean emf of these four cells. These cells in turn are calibrated daily in terms of the U.S. Legal Volt using standard NBS procedures [1].

Zener Reference

The Zener reference is a commercial Zener standard that has been modified to produce 10.1817 V rather than the 10 V normally supplied. With 16 AWG wire connecting the standard and the resistive divider the voltage at the divider is very close to 10.181 V, approximately ten times the voltage of a saturated standard cell at 30 °C. Interactions between the Zener reference and the other parts of the measuring system (notably the DVM) are negligible even when the standard is plugged into the ac power lines. The Zener reference is operated on ac power for all the measurements.

Resistive Divider

The resistive divider driven by the Zener reference consists of 10 nominally equal, stable 250-Ω wire-wound resistors. The 250 Ω choice is a compromise between reducing the resistance to minimize the effects of leakage resistances and increasing the load current that must be supplied by the Zener reference. The resistors are selected for matched temperature coefficients; all the coefficients are within ±1 ppm/°C of one another. The resistors are mounted in a thick-walled aluminum box for temperature lagging. During a measurement sequence the individual resistors typically change value less than 0.03 ppm with respect to the total divider resistance.

Scanner

The scanner used to select the voltage inputs is a modified crossbar switch identical to the VTP scanner described in reference 1. It has 80 2-wire input channels and uses latching relays to reduce the thermal emfs within the switches. The scanner is wired so that any two of the 80 channels may be selected by the computer and connected together in series opposition with the difference voltage presented to the DVM. By selecting the two channels in reverse order the voltage difference with opposite polarity is presented to the DVM. The unresolved thermal emfs of the scanner have been periodically measured and are less than 0.015 μV per switch. Twenty of the channels are used for the resistive divider, four channels are used for the reference standard cells, and two channels are reserved for crossbar thermal tests, leaving 54 channels available for unknown standards.

Digital Voltmeter

A 6-1/2 digit DVM with 0.1 μV resolution on the 0.1-V range is used as the system voltmeter. The voltmeter is programmed and read by the computer. All voltages are measured by the DVM in both the forward and reverse direction to cancel offset voltages in the DVM. The 0.1-V range is used when measuring microvolt level voltages associated with standard cell differences, and an integration time of 9 s for each polarity yields a random measurement uncertainty of 0.03 μV (1σ) for a single measurement. For all measurements of solid-state standards, except at the 1.018-V level, the voltmeter is calibrated and used on the 1-V range.

Working Group C

Working Group C consists of four saturated standard cells housed in a commercial temperature-regulated enclosure. The enclosure contains an internal thermistor and Wheatstone bridge for measurement of the cell temperatures with a resolution of 0.001 °C. The enclosure is powered from a regulated 12 V dc power supply to minimize any effects of a changing ac line voltage.

Measurement Process

The general measurement procedure is to 1) self-calibrate the system using Working Group C, 2) compare the unknown solid-state standards to the calibrated system, and 3) self-calibrate the system again to make sure no significant drifts have occurred. The exact procedure is controlled by a measurement procedure file (RUNFILE) which is executed by the computer. The measurement procedure file is created and modified daily by the operator as necessary.

Figure 2 shows some critical parts of the measuring system in more detail. The modified 10.18-V Zener reference is permanently connected to the resistive divider; this combination (the Transfer Reference) is considered to be a stable voltage source with ten series-connected 1.018 V outputs. The Transfer Reference is calibrated by comparing the voltage drop across each resistor to each cell in Working Group C, measuring the microvolt-level difference voltages with the digital voltmeter. All four Working Group C cells are compared to the ten Transfer Reference voltages for a total of 40 measurements which results in a redundant set of measurements. [1] By referencing each output tap to the bottom tap, the Transfer Reference can be used as a source of ten calibrated voltages from 0 V to 10.18 V in steps of 1.018 V. Then (as shown in Fig. 2) the output voltage developed across N of the divider resistors is compared to the voltage of the unknown standard-under-test using the DVM. N is chosen to minimize the magnitude of the DVM reading. All the solid-state standards to be measured are compared in turn to the divider tap appropriate for that standard. Then the Transfer Reference voltages are compared a second time to the standard cell voltages after all the standards-under-test measurements have been completed. The

second "calibration" of the Transfer Reference reduces the error caused by any drift in the Reference during the measurements. The low-thermal scanner switch is used to connect the standard cells and solid-state standards to the measurement system.

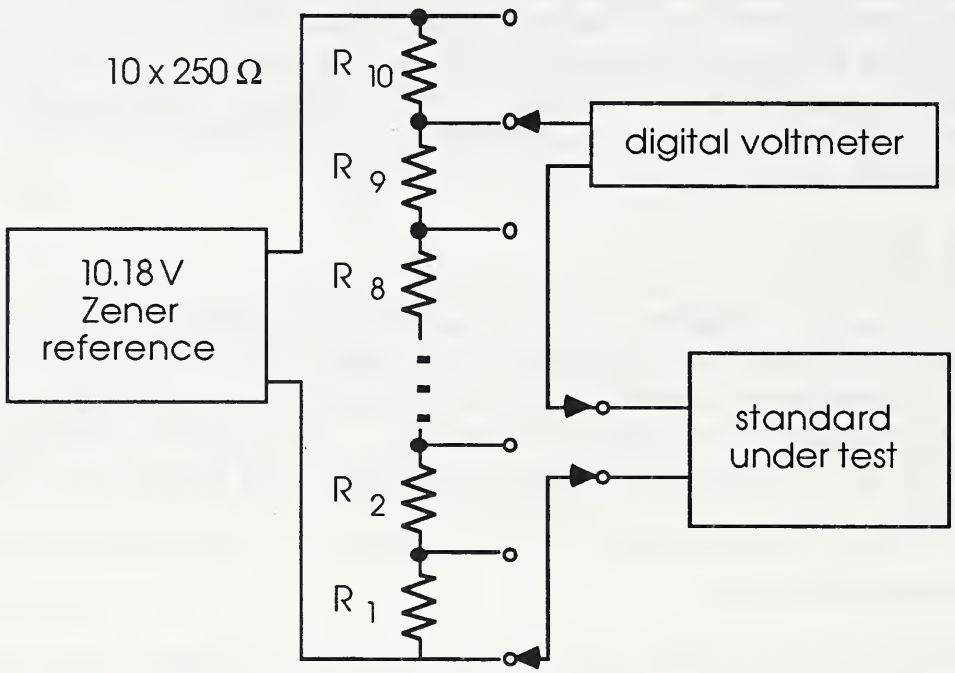


Fig. 2. Simplified wiring diagram of the measuring system set up to compare a standard-under-test to the calibrated resistive divider (Transfer Reference). Not shown is the low-thermal crossbar selector switch.

Measurement Theory

With this system the digital voltmeter is used to read only a fraction of the voltage of the standard-under-test which reduces the contribution of the DVM uncertainty to the overall measurement uncertainty. When comparing the standard-under-test to the Transfer Reference (with voltage taps at 1.018 V increments) the maximum reading required of the DVM is 0.509 V. Thus,

$$U = \frac{V_Z - (N)1.018}{V_Z} U_{DVM}$$

where U_{DVM} is the DVM uncertainty expressed as a percentage of reading, V_Z is the voltage of the solid-state standard, N is the number of resistors, and U is the final DVM uncertainty as a percentage of V_Z. For a worst case of V_Z = 5.6 V, U = 0.09 U_{DVM}, and for V_Z = 10 V, U = 0.02 U_{DVM}. Therefore, if the DVM can be calibrated to an accuracy of 1 ppm the worst case error contribution is 0.09 ppm.

The linearity of the DVM is checked as necessary, but the gain of the one volt range of the DVM is calibrated every time measurements are made by measuring the 1.018 V developed across two of the resistors in the Transfer Reference; typically resistors R_1 and R_2 are used (see Fig. 2.). This is done to evaluate the effect of measuring voltages at ground and above ground. Gain calibrations from several days can be combined to detect a systematic bias in the measurements with respect to the particular resistor used. Such measurement errors can be caused by excessive DVM leakage or bias currents and will also affect the accuracy of the comparisons between the Transfer Reference and the standard cell reference group.

Once the Transfer Reference has been calibrated in terms of the reference standard cells the linearity of the DVM on the 10 V range can be checked by directly measuring the ten voltages, 1.018 to 10.18 V, developed across the resistors. Although not a thorough calibration of the DVM linearity, this test can be used to detect significant changes in the linearity errors.

The DVM is used on the 0.1-V range to improve the resolution of comparisons of 1.018-V solid-state standards to the Transfer Reference, and of the comparisons of the Transfer Reference to the reference standard cells. The gain error of the 0.1-V range is not calibrated daily by the system but, the largest expected measurement difference for these comparisons is only 200 μV. Thus for an overall measurement error of 0.005 ppm (of 1.018 V) the gain need only be known to 250 ppm. The gain of the 0.1-V range is periodically checked using external standards to ensure that it is within this limit.

The final value for each solid-state standard under test is computed by correcting the measured difference between the standard and the Transfer Reference for the gain of the DVM, and adding this to the value of the appropriate tap of the Transfer Reference. The calibrated values of the Transfer Reference are determined using a least-squares analysis of the redundant measurements that compared the 10 outputs of the Reference to the reference standard cell group. The mean of the "before" and "after" calibrations is used.

4. Operational Procedures

Client Standards Handling

Standards shipped to NBS under power are hand carried from the Building 301 Receiving Room as soon as Electricity Division personnel are notified. The temperature is checked by qualified laboratory staff (if the standard has a temperature measuring device) and the standard is brought to the Volt Facility Laboratory. All solid-state standards to be calibrated are connected to the measuring system scanner at the time of their arrival and are left undisturbed until the test is completed. When the standard is received it is

assigned a unique ID number, usually a subset of the serial number, and this is entered into a log book along with the date and the condition of the standard. Measurements are begun the day after the arrival of the unit. After the test is completed, the client is notified and the standard is repacked in its original shipping material and returned to the client.

Daily Procedures

Calibration of Working Group C

Working Group C is used as the reference for all solid-state measurements in the Volt Facility Laboratory. Working Group C is compared daily to two groups of primary cells (Primary Groups A and B) using the automated standard cell comparison system VTP [1]. These measurements are made early in the morning before measurements of solid-state standards are begun and the data is immediately reduced to determine the values of the cell emfs of Working Group C. These values are used for the solid-state standard measurements done that day.

Calibration of Solid-State Standards

The operator must set up a measurement procedure file (RUNFILE) that has a list of all the standards to be calibrated. Program EDITA, a general purpose text editor, is usually used to modify the previous day's file. The exact format of the file is described in Appendix B, but in general the file contains the measurement sequence for calibrating the Transfer Reference, a list of identification numbers of the standards-under-test, their scanner circuit numbers, and the approximate voltage of each of the standards. The operator manually measures the temperature of Working Group C before the measurements are begun and enters this into the file along with the cell emf values (from the comparisons with the primary groups). When the file is complete the measurement program (READZEN) is started and measurements proceed automatically with the fully corrected values of the solid-state standards printed out and saved on disk at the conclusion of the measurements.

READZEN first checks the measurement system apparatus to make sure everything is functionally working. This includes sending a clear-all-channels command to the scanner and checking to see if it responds, sending an initialization command to the DVM and checking it's response, checking that the manual front/rear input switch on the DVM is set properly, checking to see if the hard disk/file server is connected and turned on (if it is to be used), checking that the printer is turned on and loaded with paper, and finally, checking that a properly initialized data disk and a program disk containing a RUNFILE are in their respective disk drives. Multiple RUNFILEs may be included on the program disk for different sets of measurements; the operator selects the desired one at the time the program READZEN is run.

Measurements are started by setting the DVM to the 0.1-V range and shorting the DVM input to discharge the input capacitance. Then the ten 1.018-V outputs of the Transfer Reference are compared to the Working Group C cell emfs using the redundant measurement design discussed in section 3. Forty measurements are required and the data is saved on disk; this process takes about 1/2 hour. Measurements of the unknown solid-state standards are preceded by gain calibrations of the DVM 1-V range. DVM gain calibrations consist of directly measuring (in both forward and reverse polarity) two of the Transfer Reference voltages (1.018 V) using the 1-V range of the DVM as discussed above.

The solid-state standard voltages are measured by comparing each one in turn to the tap on the Transfer Reference that is closest in voltage to the standard's voltage and measuring the difference with the DVM set on the 1-V range (in both forward and reverse polarity). Once all the unknown standards have been measured the measurement sequence repeats starting with the DVM calibrations. This measurement sequence (including the DVM calibrations) is repeated three times and then a final (fourth) DVM calibration is done. At the conclusion of these measurements, the Transfer Reference is recalibrated against Working Group C. To summarize, the measurement sequence is – calibrate the Transfer Reference using Group C, calibrate the DVM, measure the unknown standards, calibrate the DVM, measure the unknown standards, calibrate the DVM, measure the unknown standards, calibrate the DVM, and finally calibrate the Transfer Reference a second time.

Analysis of the Daily Data

At the conclusion of all the measurements for the day, the READZEN program automatically calculates the emfs of the standards-under-test and produces a two page printout summarizing the calibration of the Transfer Reference and listing the emfs of the solid-state standards. The emfs of the standards (as well as intermediate results) are also saved on two disk files, one on a floppy disk, and the other on the hard disk. These data are combined with other daily data sets for assigning the final emf values to the standard.

Figures 3a and 3b are copies of the two pages of a typical daily printout produced by READZEN. At the top of page 1, below the date and the name of the RUNFILE that was used, is a summary of the measurements made, and their starting time. In the example shown the Transfer Reference (denoted Transfer Zener in Fig. 3a) was compared to Working Group C (2800) at 8:03 and 10:07 am. The unknown solid-state standards were compared to the Transfer Reference three times: 8:35, 9:05, and 9:35 am. This part of the log is printed as the measurements are being taken so the operator can track the operation of the program.

```
ZENER MEASUREMENT LOG FOR:  9 SEP 1987    [1.0]            WEEK: 8737

RUNFILE:  RUNFILEA:,700,1

-------------------------------------------------------------------------

08:03   DESIGN: (2800  - Zener )    Zener  = 7970.159      SD =  .022
08:35   Zener measurement completed -- Set 1
09:05   Zener measurement completed -- Set 2
09:35   Zener measurement completed -- Set 3
10:07   DESIGN: (2800  - Zener )    Zener  = 7970.155      SD =  .017

-------------------------------------------------------------------------
 REFERENCE CELL STATISTICS

           .022       .017       Standard Deviations (uV)
           .008       .006       Left-Right Components (uV)
        30.0135    30.0135       Reference Cell Box Temps (C)

 Reference          Value at    Difference   Value at         Mean
   Cells     Res    08:03:56       (uV)      10:07:02        Values

 2801        1     8166.558    <  .003>     8166.561       8166.560
 2802        1     8164.409    <-0.000>     8164.409       8164.409
 2803        1     8165.602    <  .007>     8165.609       8165.606
 2804        1     8166.147    <-0.010>     8166.137       8166.142
 Mean              8165.679    < 0.000>     8165.679       8165.679

 Transfer Zener

 Z1                7978.379    < 0.000>     7978.380       7978.379
 Z2                7976.446    < -.021>     7976.425       7976.436
 Z3                7932.627    < -.006>     7932.621       7932.624
 Z4                7956.967    < -.004>     7956.963       7956.965
 Z5                8052.081    < -.011>     8052.071       8052.076
 Z6                7954.107    < -.015>     7954.092       7954.099
 Z7                7956.854    < -.006>     7956.848       7956.851
 Z8                7966.002    < -.006>     7965.996       7965.999
 Z9                7958.452    <  .019>     7958.471       7958.461
 Z10               7969.671    <  .015>     7969.686       7969.678
 Mean              7969.671    < -.004>     7970.155       7970.157

 DVM CALIBRATION SUMMARY

 Number of DVM calibrations =     8
       Mean DVM error (ppm) =   2.0
    Std. Dev. of mean error =    .1

-------------------------------------------------------------------------
```

Fig. 3a. Page 1 of a typical daily printout from the measurement program READZEN.

ZENER CALIBRATION SUMMARY FOR: 9 Sep 1987

Zener ID	Voltage (volts)	Std. Dev. (ppm)
188/1	9.99999654	.145
188/2	1.01813796	.220
C5950/1	9.99999937	.006
C6015/1	10.00000245	.002
5027/1	10.00002831	.004
5027/2	1.01799749	.011
5027/3	.99999292	0.000
5010N/1	10.00003443	0.000
5010N/2	1.01800020	.008
5010N/3	1.00000492	0.000
5011/1	10.00002566	.031
5011/2	1.01799755	.028
5011/3	1.00000773	.036
5018/1	9.99999464	.004
5018/2	1.01799740	.007
5018/3	.99997592	0.000
BFF1/1	1.01899358	.020
49105/1	9.99987041	.026
49105/2	1.01798278	.010
49105/3	.99998092	0.000
49105/4	6.99989104	.052
49111/1	9.99990381	.027
49111/2	1.01798826	.017
49111/32	.99998952	.042
49111/4	6.99992458	.047
5035/1	10.00000727	.014
5035/2	1.01799416	.013
5035/3	.99999792	0.000

ALL CALCULATIONS COMPLETED AND SRM UPDATED

Fig. 3b. Page 2 of a typical daily printout from the measurement program READZEN.

The comparisons of the Transfer Reference to 2800 (Working Group C) are summarized under the heading REFERENCE CELL STATISTICS. The standard deviations are the standard deviations of a single observation from the redundant measurement design [3]. The emfs of the reference cells in 2800 and the 1.018-V outputs of the Transfer Reference (labelled Z1 to Z10) are expressed in microvolts reduced by 1.01 volts. The mean values from the two calibrations are used as the values of the Transfer Reference for calculating the values of the solid-state standards under test. The emf differences between two calibrations are calculated and printed in the Difference column to aid the operator in detecting abnormal changes.

At the bottom of Fig. 3a a short summary of the DVM calibrations appears. In this example the DVM was compared to two different resistors before and after each set of Zener measurements ([3 sets +1]x2) for a total of 8 calibrations. The mean DVM error on the 1-V range was found to be (2.0 ± 0.1) ppm.

Fig. 3b, page 2 of the printout, lists the standards under test, their average measured voltage, and the standard deviation of the measured voltage as determined from the repeated sets of measurements (3 in this case). The repeated measurements are performed over a relatively short period of time (1 to 2 hours) to reduce the variability of the final averaged values. The complete set of measurements described above is considered to be a single measurement of each standard consisting of the one averaged value for that standard.

5. Assignment of the Final Value

Client standards are measured in terms of NBS working cells once a day (using the above procedure) for at least ten working days. The measured values are plotted and the data are reviewed by laboratory staff to determine if the outputs exhibit sufficient stability for a report to be issued. If so, the mean of the all the daily measurements is used as the final assigned value for the output of the standard. In some rare cases standards have shown an initial drift for the first few days after arriving at NBS. If this happens additional measurements are taken so that at least ten stable measurements are obtained, and the early unstable readings are not used in determining the final value. Even if the standard behaves properly, the exact number of measurements used for any calibration is not fixed; it depends on when the standard was received, when the client requires that it be returned, etc., but at least ten measurements are always used and usually not more than fifteen.

Figure 4 shows the measured values of three outputs of a typical Zener standard. The outputs are all derived from the same Zener reference device; the 1.018-V and 1-V outputs are obtained from internal resistance dividers connected to the 10-V output.

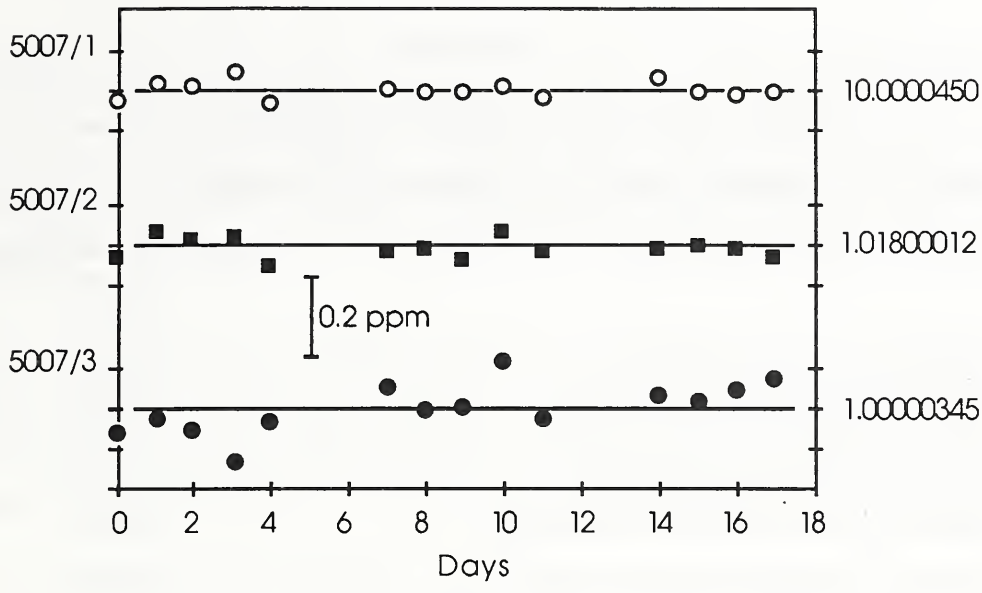


Fig. 4. The three voltage outputs of a typical solid-state standard sent to NBS for calibration. The outputs are plotted in ppm from their mean voltage. The solid line is drawn at the mean value which is also labelled (in volts) on the right axis.

Uncertainty

The measuring system was designed principally to measure solid-state standards in the range 5 - 10 V. Table I lists the sources of uncertainty in the measuring system extrapolated to the worst case unknown voltage in the 5 - 10 V range. The system may be used over the 1 - 5 V range but with somewhat reduced accuracy when the unknown voltage is not near a cardinal value (1.018, 2.036, 3.054,... V).

Explanation of Table Items

By far the most critical component of the system is the digital voltmeter. The DVM gain is measured during the course of the Zener measurements and the gain error of the 1-V range is calculated and applied as a correction to the DVM readings. An allowance is included in the Table for the inaccuracy of this gain measurement.

The DVM linearity was initially checked on the 10-V range using a calibrated, manual 7-dial Kelvin-Varley divider, and measurements on similar DVMs indicate little or no change in the linearity with time. The linearity error is typically a maximum of 0.7 ppm or less at half-scale input and no correction to the data is made for it. Provision has been made for the system to calibrate automatically the linearity of the DVM at 10 points on the 10-V range by measuring the 10 voltages 1.018, 2.036, ..., 10.18 V although this is not done routinely.

TABLE I

UNCERTAINTIES IN THE FINAL ASSIGNED VALUE

Source of Uncertainty	1 std. dev. estimate (ppm)
DVM gain uncertainty	0.020
DVM linearity uncertainty	0.062
DVM leakage/bias currents	0.023
Standard cell leakage currents	0.006
Scanner switch thermal emfs	0.013
Random uncertainty in calibrating the Transfer Reference	0.007
Random uncertainty in calibrating the client standard (within-day)	0.013
RSS subtotal	0.072
Uncertainty in value of Working Group C	0.040
Change in Working Group C during the day	0.047
Uncertainty of the value assigned to the client standard	
RSS Total	**0.095**

The DVM leakage and bias currents and standard cell leakage currents were measured directly with an electrometer and their effects on the measurement calculated.

The crossbar scanner switch thermal emfs were measured with the DVM by shorting the input leads and measuring the residual voltages.

The random uncertainties in calibrating the Transfer Reference and the client standards are estimated from the pooled standard deviations of the measurement designs. Although these uncertainties tend to be reduced by averaging ten measurements of the client standard we prefer to include them for completeness.

Two sets of DVM readings are taken for every voltage measurement (both for standard cells and solid-state standards), one with the input voltage source connected normally and a second with the voltage source polarity reversed by the crossbar scanner. Subtracting the second reading from the first and dividing by two yields a measurement free from any error due to a non-zero DVM offset, thus no allowance has been included in the Table for DVM offset errors.

Ac effects observed in an earlier system were found to be negligible in this system [2]. The ac effects on the Zener reference of the Transfer Reference were estimated by monitoring changes in its voltage using a passive measurement circuit referenced to a standard cell while ac was introduced into the system by connecting the DVM to the Transfer Reference taps.

The RSS subtotal represents the ratio error in comparing an arbitrary voltage source to a group of 1.018-V standard cells. For establishing an uncertainty for the client standards with respect to the U.S. Legal Volt, the uncertainty of Working Group C must be added. The uncertainty of the value of Working Group C computed from earlier measurements is developed in reference 1 and will not be repeated here. An additional uncertainty is included, however, for temperature instabilities and other related emf variations in the present standard cell enclosure used for Working Group C that cause the cell emfs to change during the day.

As far as is known, all systematic errors listed in Table I are independent of each other. Therefore, they are combined in root-sum-square (RSS) fashion along with the random errors. The final reported uncertainty for a client standard ($U_{reported}$) is three times the root-sum-square of the above measurement uncertainty and the standard deviation of the mean of the ten daily measurements (s_m), i.e.,

$$U_{reported} = 3\left[\sqrt{(0.095^2 + s_m^2)}\right].$$

6. Quality Control Procedures

Measurement System Verification

As discussed in Section 5 all major sources of error in the measuring system are evaluated during the measurements and corrections are applied to the data as necessary. The printout from READZEN is reviewed daily by laboratory personnel to verify that all the measurements are in statistical control.

Other potential sources of error include uncompensated thermal emfs in the crossbar switches and in the leads to the standards and circulating ground currents within the measuring instrument and between the measuring instrument and the client standard. Several quality control procedures are periodically performed to estimate or eliminate these uncertainties.

Thermal emfs in the leads from the standards to the crossbar switch are evaluated regularly. The positive and negative leads are shorted together at the end where they would normally connect to the standard and a redundant measurement design is done between the two sets of shorted leads using the DVM as the detector. These thermal emf measurements are performed as standards leave the laboratory and leads become free. Experience has shown that the thermal emfs are usually small and fairly constant (approximately 15 nV or less).

Leakage resistance and leakage current measurements are made periodically between the different parts of the measuring system using an electrometer to ensure that they are at acceptable levels.

A 10-V standard that is not moved is measured every day as part of the workload and serves as a check standard to monitor the long term stability of the measurement system. The 10-V output of the standard is plotted versus time in Fig. 5.

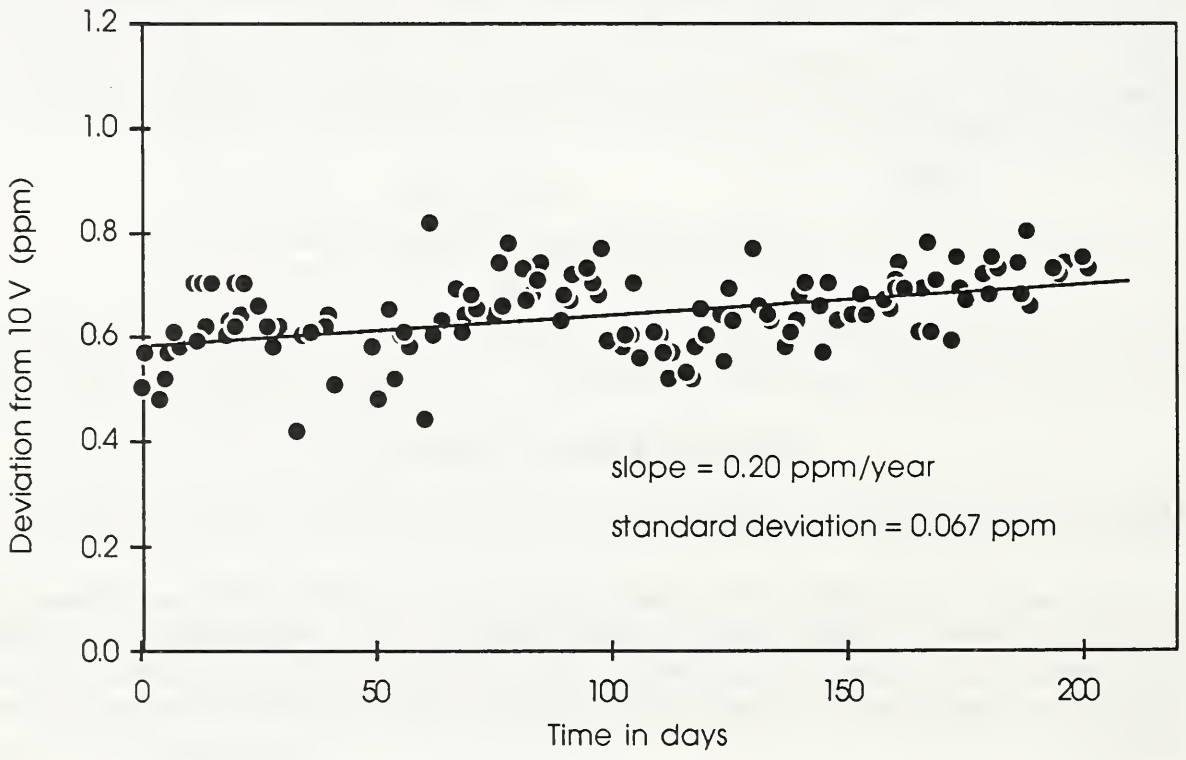


Fig. 5. Ten volt output of a check standard as measured by the system.

Client Standard Evaluation

Tests for ac effects are done for each new type of standard received for test. Some models of standards produce a dc shift of the output voltage when an ac source is connected to the output terminals. Experiments have shown that for some standards the DVM can produce a significant shift [2]. These experiments also showed that the effect was model dependent, not standard dependent; all items of a particular model tested either had the problem or they did not. Any new, unfamiliar model received is tested for ac effects by applying a small ac signal to the output while observing the dc voltage.

7. Bibliography

1. Field, B. F., "NBS Measurement Services: Standard Cell Calibrations," *Natl. Bur. Stand. (U.S.), Special Publ. 250-24,* 52 pages (1987).

2. Field, B. F., "A Sub-PPM automated 1–10 Volt DC Measuring System," *IEEE Trans. Instrum. Meas.,* Vol. IM-34, No. 2, pp. 327-330 (1985).

3. Eicke, W. G. and Cameron, J. M., "Designs for Intercomparing Small Groups of Saturated Standard Cells," *Natl. Bur. Stand. (U.S.), Tech. Note 430*, 19 pages (1967).

4. Eicke, W. G., "Comments on Zener diodes as voltage standards," *Comité International des Poids et Mesures, Comité Consultatif d'Electricité,* 10th Session, p. 27 (1963).

5. Eicke, W. G., "Making precision voltage measurement on Zener diodes," *IEEE Trans.*, Paper CP 63-416 (1963).

6. Eicke, W. G., "The Operating Characteristics of Zener Reference Diodes and their Measurement," *ISA Transactions*, Vol. 3, No. 2, pp. 93-99 (1964).

7. Huntley, L., "The Fluke direct voltage maintenance program," *Proc. 1984 Meas. Sci. Conf.*, pp. 147-151 (1984).

8. Koep, K. J., "The use of a solid state dc voltage transport standard to transfer the dc volt in a regional standard cell MAP program," *Proc. 1984 Meas. Sci. Conf.*, pp. 142-146 (1984).

9. Murray, W. and Dencher, P., "High-accuracy voltage transfer using a modified commercial solid-state reference," *J. Phys. E.: Sci. Instrum.*, Vol. 17, No. 5, pp. 354-356 (1984).

10. Spreadbury, P. J. and Everhart, T. E., "Ultra-stable portable voltage sources," *IEE Conf. Publ.*, Vol. 174, pp. 117-120.

Appendix A - Report of Calibration

UNITED STATES DEPARTMENT OF COMMERCE
National Bureau of Standards
Gaithersburg, Maryland 20899

REPORT OF CALIBRATION

Solid-State Voltage Standard
XXXXX Model YYYY DC Reference Standard
Serial No. 1111111

Submitted by
XXXXXXXX
XXXXXXXXXXXX
XXXXXXXXXXXXXX

This voltage standard was calibrated during the period June 1, 1987 to June 18, 1987, by comparing its output voltages with those from a calibrated ten step divider driven by a stable ten volt source. The difference voltages were measured using a calibrated digital voltmeter. This measuring system was standardized by saturated standard cells calibrated in terms of the U.S. Legal Volt (volts-NBS) using standard NBS procedures. The solid-state voltage standard was received with the IN CAL light on and with a thermistor resistance of 3.91 kilohms. The unit was plugged into the ac power line for all measurements.

The values given in the table below are the mean values of the results of the individual calibrations summarized in the Appendix. The uncertainties given are three standard deviation estimates equal to three times the root-sum-square of the estimated 0.095 ppm one standard deviation uncertainty in the measurement system (primarily systematic or type B), and the one standard deviation random (type A) uncertainty of the mean of the measurements on the individual voltage standard output.*

Nominal Output	Measured Voltage (volts-NBS)	Uncertainty (ppm)	Uncertainty (microvolts)
10 volts	10.0000450	0.29	2.86
1.018 volts	1.0180001	0.29	0.29

* The nomenclature 'type A' and 'type B' and the procedure used to combine them to obtain the total uncertainty has been recommended by the International Bureau of Weights and Measures (BIPM) Working Group on the Statement of Uncertainties, see Metrologia, vol. 17, No. 2, p. 73 (1981).

Test No. 521/XXXXXXXXX

Solid-State Voltage Standard
XXXXX Model YYYY DC Reference Standard
Serial No. 1111111

These uncertainty figures contain no allowance for the effects of transport of the standard back to the user. Any valid uncertainty statement applying to the above values when the standard is used other than in this laboratory must contain an additional uncertainty component. Also not included in the above uncertainties is an allowance for long term drift of the output voltages of this standard. This must be determined from historical data on a case by case basis.

For the Director
National Measurement Laboratory

Norman B. Belecki, Physicist
Electricity Division
Center for Basic Standards

Test No.: 521/XXXXXXXXX
Reference: XXXXXXXXXXX
Date: September 8, 1987

Solid-State Voltage Standard
XXXXX Model YYYY DC Reference Standard
Serial No. 1111111

APPENDIX
(values in volts-NBS)

Date	Time	10 volts	1.018 volts
06/01/87	10:19	10.0000448	1.0180001
06/02/87	10:58	10.0000452	1.0180002
06/03/87	09:52	10.0000451	1.0180001
06/04/87	12:16	10.0000454	1.0180001
06/05/87	10:07	10.0000447	1.0180001
06/08/87	10:27	10.0000450	1.0180001
06/09/87	09:12	10.0000450	1.0180001
06/10/87	09:34	10.0000450	1.0180001
06/11/87	09:33	10.0000451	1.0180002
06/12/87	11:20	10.0000448	1.0180001
06/15/87	11:03	10.0000453	1.0180001
06/16/87	11:03	10.0000449	1.0180001
06/17/87	09:39	10.0000448	1.0180001
06/18/87	09:38	10.0000450	1.0180001
Mean (volts-NBS)		10.0000450	1.0180001
Standard Deviation of the Mean (ppm)		.006	.007
Number of readings		14	14

Test No.: 521/XXXXXXXXX
Reference: XXXXXXXXXXX
Date: September 8, 1987

Appendix B - Measurement Software

The READZEN program is a generalized measurement program that runs under the direction of an ASCII text file (RUNFILE) and is used to make all the measurements required for solid-state standard calibrations. Program EDITA is a general purpose editor program for creating and modifying RUNFILEs. The following sections describe how to operate the programs and the purpose of some of the important program variables. It is assumed that the reader is familiar with the BASIC language environment (Rocky Mountain BASIC) used with the Hewlett/Packard model 200/300 desktop computers.*

Program READZEN

READZEN is the main measurement and calculation program. It has the capability of doing standard cell intercomparisons, least-squares cell data reduction, Zener-Transfer Reference comparisons, and Zener data reduction. The operation of the program is controlled by an ASCII text file named RUNFILExxx that is stored on the program disk. The xxx indicates a 1 to 3 character suffix of numbers or letters. The suffix is used to differentiate between several Runfiles that may be stored on the same disk but were created for different measurement procedures, i.e., normal workload measurements, special tests, etc. At runtime the operator is prompted to select one.

At the beginning of the program code of READZEN are comments on the purpose of the program, array variable dimension statements, and a number of variables that define the various defaults of the program. These variables control things like the specification of the program disk and data disk storage units, the 488 bus addresses of the measuring system instruments, and other internal variables used to control calculations and data storage. These variables are described in detail below. Program operations that may be optionally changed by the user are generally controlled by these variables. No modification to the program code itself should be necessary.

Running the program

When the program is started an informational screen is shown giving the name, version number, and purpose of the program, and all the scanner channels are opened. If the scanner is not connected or turned on a message is printed and the program is stopped. Similarly, the printer is checked and the disk drives are checked for the program and data disks. If

---

* Specific model names are used for clarification only and their use does not constitute an endorsement by NBS as the best available equipment. Other instruments may be as suitable, or more suitable, for the applications described.

all is in order, the Runfile names on the program disk are displayed along with an ID number, a 20 character identifying line, and the creation date of the file. The operator must enter the ID number of one of the Runfiles. Once this has been done the measurements proceed automatically with no further operator control. At the conclusion of the measurements the results are printed on the printer and saved on disk.

Program options

Variables to control the data storage and bus addresses of the measuring instruments are contained in lines 1270 to 1641 of READZEN. A description of some useful variables follows with their normal values shown in parentheses.

Observer$	is an identifying name for the system. (Observer$="Micky")
Zsets	is a counter for the number of times the Zener references are compared to the transfer reference. Under special circumstances more or fewer measurements may be desired. (Zsets=3)
Dvmcorr	is a "gain" value that is used to multiply all DVM readings taken on the 0.1 V scale. If the 0.1 V scale is known to have a significant gain error this variable can be set to compensate for it; e.g., if the gain error is -100 ppm, the DVM reading is lower than the actual value, set Dvmcorr = 1.00010. Generally, the DVM is physically adjusted if the gain is significantly in error. (Dvmcorr=1)
Dk$	is the data disk specifier. (Dk$=":,700,0")
Progdk$	is the program disk specifier. (Progdk$=":,700,1")
Srm$	is the suffix for the SRM. (Srm$=":REMOTE")
Dir$	is a prefix directory name for the SRM. This name will preface the data file name, and Srm$ will be appended to the end to form a complete filename. (Dir$="ZENDATA/")

Zener_volt$ is the partial filename for result files. Calculated standard cell results preface Zener_volt$ with a "D" and Zener reference results preface it with an "R". For both, the 2 digit year and a 2 digit week number within the year are appended to the filename. Zener_volt$ should be 4 characters or less, otherwise the filename will be too long.
(Zener_volt$="ZEN")

Hcopy controls the printing of measurement data, intermediate results, and final results to the printer. If Hcopy=0 (normal default) one line is printed on the printer each time a measurement set (cells or Zeners) is completed. The printer output becomes a log of the measurements; the actual measurements are not printed. The results of the cell calculations and the final Zener results are also printed. If Hcopy = 1 then log printing is suppressed but cell and Zener results are printed. If Hcopy = 2 all the measurement data (cell-pair differences and Zener differences) are printed to the printer as well as the results. If Hcopy = –1 nothing is printed to the printer. (Hcopy = 0)

Srm controls whether data is saved on an SRM in addition to the floppy disks. It should be set to either Yes or No. If Yes, check to make sure that Dir$ is correctly set to the subdirectory desired. (Srm=Yes)

Saveflag controls saving of data on floppy disk and the SRM. For Saveflag set equal to 1, all measurement data files and result files are saved on disk. To save space on disk Saveflag may be set equal to 0 so only the cell and Zener result files are saved. If Saveflag = –1 only the final Zener result files are saved. (Saveflag = 1)

File use

READZEN automatically creates as needed a number of data files. These are all ASCII files with data stored as strings. Each line in the file is stored as one string and may contain an arbitrary number of items with the items

separated by at least one space. All files are read using a free-field format; only the order of the data is important, not the position on any one line. File names are generally composed of a unique prefix identifying the type of file and a suffix based on the date. The date suffix is either of the form YYMMDD or YYWW where YY is the year, MM is the month, DD is the day, and WW is the number of the week in the year. WW starts with 1 for the first week of the year and automatically changes on Sunday for the next week. The exact style of the date suffix depends on whether the file contains one day's data or one week's. The exact format of the data stored in the files is discussed below.

The creation and use of the files is as follows: Intercomparisons between the reference standard cells and the Transfer Reference produce 40 voltage difference readings (4 cells by 10 resistors). These difference readings are stored in an "S" file, i.e., SYYMMDD. All intercomparisons performed in one day are stored one after the other in the file. READZEN performs a least-squares calculation on the "S" data and saves the calculated cell values and "resistor-voltages" from the Transfer Reference in a "D" file, i.e., DZEN_YYWW. All standard cell and Transfer Reference voltages for the entire week are stored in this file. The raw difference readings between the Zener standards under test and the Transfer Reference are stored in "Z" files, i.e., ZYYMMDD. After calculation the final Zener values are stored in "R" files, i.e., RZEN_YYWW.

The files are initially created with a specified size and once created cannot be increased beyond this size. They have generally been made large enough to handle one or two sets of measurements per day. If the file size is consistently too small the initial size specification may be increased. "S" files are created in line 4780 of READZEN and are specified as 68 records of 256 characters/record. "D" files are created in line 11840 and are specified with a length of 20 records of 256 characters/record. "Z" files are created in line 8100 with a length of 20 records of 256 characters/record. "R" files are created in line 12980 with a length of 20 records of 256 characters/record.

Program EDITA

EDITA is a general purpose ASCII text file editor for viewing, printing, and modifying the input and output files generated by READZEN. Prior to running READZEN a RUNFILE is created using EDITA which details the measurements to be made for that day. The specific RUNFILE format is described in the next section on file formats.

Running the program.

When the program is started it asks for a file storage device. This is the floppy disk or hard disk on which a previously created file is stored or a new file is to be stored. Enter a standard H/P disk designator without any quotes, e.g., :,700,0. The program then prompts for a file name; enter the

name of an existing file or a name for the new file to be created. If the program cannot find the file on the specified disk, it will ask if a new one should be created. Enter Y or N as appropriate.

Most operations are controlled by the function keys at the top of the keyboard. They are:

f1	PREV SCREEN	Scroll up to display the previous screen of data.
f2	NEXT SCREEN	Scroll down to display the next screen of data.
f3	FIND CHRS	Find a string of characters within the text.
f4	2ND KEY LEVEL	Change the key functions to the alternate (2nd) set.
f5	APPEND ON (OFF)	With append on additional lines may be added to the file, with append off the original length is preserved.
f6	GO TO LINE	Go to the specified line number.
f7	MOVE/DEL COPY	Move, delete, or copy a block of lines.
f8	END PROGRAM	Stop the program without saving the file.

2nd Key Level

f1	STORE TEXT	Save the file in the previously specified name.
f2	MERGE FILES	Append another file to the end of the existing one in memory.
f3	NEW STORAGE	Specify a new disk designator
f4	1ST KEY LEVEL	Return to the original key functions
f5	NEW PRINTER	Specify the printer select code (usually 701). Note: this must be done before printing the first time.

f6	PRINT TEXT	Print the file on the printer select code from f5.
f7	CAT	Print a catalog of specified disk.
f8	END PROGRAM	Stop the program without saving the file.

In addition all the normal H/P program editing keys may be used – insert line, delete line, insert character, delete character, and all the arrow keys.

File Formats

All files are ASCII files with data stored as strings. Each line in the file comprises one string and may contain an arbitrary number of items. All files are read using a free-field format; only the order of the data is important, not the position on any one line. Each item is separated from its neighbor by one or more spaces (usually one). For convenience we have adopted a convention of grouping several items on a line together to make it easier for the operator to "read" the file. For each file an example of actual data is given. Line numbers are shown for identification purposes only, they are not included in the file. The exact number of lines is unimportant; data may be stored all one line, or one item per line if desired.

Runfiles

Runfiles contain the measurement design data and commands to control operation of the READZEN program. The exact format will vary depending on the measurements required but will follow the general pattern of making cell measurements between the reference cell group and the Transfer Reference (DESIGN), measuring the unknown Zeners (ZMEAS), repeating the cell measurements (REDESIGN), and calculating the final results (ZCALC).

> Note: The ten resistor voltages of the transfer reference are considered to be equivalent to cells for the purposes of the measurement programs and files. The ten voltages are assigned a group name of 'Zener' and individual IDs of Z1, Z2, Z3,...,Z10.

Sample data: RUNFILEA

```
1     2.1138921114E11 Zener_workload
2     ! 4x10 design, 1600 vs transfer Zener
3     DESIGN
4     2 1600 4 Zener 10
5     4010 40 8165.455 0.034
```

```
6     1 1 1 1 0 0 0 0 0 0 0 0 0 0
7     1601 1602 1603 1604 Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z9 Z10
8     29 30 31 32  19 20 21 22 23 24 25 26 27 28
9     1 5 1 7 1 9 1 11 1 13 2 6 2 8 2 10 2 12 2 14
10    3 5 3 7 3 9 3 11 3 13 4 6 4 8 4 10 4 12 4 14
11    13 4 11 4 9 4 7 4 5 4 14 3 12 3 10 3 8 3 6 3
12    13 2 11 2 9 2 7 2 5 2 14 1 12 1 10 1 8 1 6 1
13    30 30.011 30 30
14    ENDDATA
15    !
16    ZMEAS
17    5018-1 10 8
18    5018-2 1.018 7
19    5018-3 1 12
20    ENDDATA
21    !
22    REDESIGN
23    ZCALC
```

Description:

1 — time and date in seconds (standard H/P time as returned by TIMEDATE), identifying label (up to 30 characters with no spaces) displayed by READZEN to identify different Runfiles

2 — any line preceded by an exclamation point (!) is ignored by the program; this is a comment for the operator

3 — command to the program to expect DESIGN data for intercomparing cell groups to follow

4 — number of cell groups, group 1 name, number of cells in group 1, group 2 name, number of cells in group 2

5 — a meaningless design number - not used but something must be there; total number of observations, mean of reference group, accepted standard deviation (the mean of the reference group is expressed in microvolts reduced by 1.01 volts, i.e., 8165.455 ≡ 1.018165455 volts)

6 — restraint - one number for each cell in each group, 1 identifies the cell as included in the mean of the reference group, 0 the cell is not included

7 — cell IDs - group 1 first, group 2 next

8 — cell circuit number corresponding to above cell IDs

9 - 12 — measurement design - cells in line 7 are assigned position numbers 1 through N (14 here). Pairs of numbers indicate difference measurements to be made, i.e., 1 - 5 indicates to measure cell "1" (1601) minus cell "5" (Z1). The number of pairs of numbers must agree with the number of observations specified on line 5.

13	nominal temperature of group 1, actual temperature of group 1, nominal temperature of group 2, actual temperature of group 2
14	ENDDATA command to indicate end of DESIGN data set
15	comment
16	ZMEAS command to start accepting Zener measurement data for unknown Zeners to be measured
17 - 19	Zener ID, nominal voltage (volts), scanner channel number
20	ENDDATA command to indicate end of ZMEAS data set
21	comment
22	REDESIGN command, repeat previous cell intercomparison specified by the DESIGN data
23	ZCALC command, calculate standard cell measurement designs and final Zener voltages. ZCALC requires that a DESIGN measurement on the transfer reference be done before and after the ZMEAS measurements.

"S" files

"S" files contain the raw standard cell intercomparison data; only one set is shown here. A file may contain additional sets taken on the same day. The similarity between lines 4 through 12 of the Runfile with lines 2 through 10 of the "S" file is no coincidence!

Sample data: S860730

```
1     07 30 86 13:54:46 Micky
2     2 1600 4 Zener 10
3     40 8165.455 0.034
4     1 1 1 1 0 0 0 0 0 0 0 0 0 0
5     1601 1602 1603 1604 Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z9 Z10
6     1 5 1 7 1 9 1 11 1 13 2 6 2 8 2 10 2 12 2 14
7     3 5 3 7 3 9 3 11 3 13 4 6 4 8 4 10 4 12 4 14
8     13 4 11 4 9 4 7 4 5 4 14 3 12 3 10 3 8 3 6 3
9     13 2 11 2 9 2 7 2 5 2 14 1 12 1 10 1 8 1 6 1
10    30 30.011 30 30
11    157.742 199.267 87.808 173.733 173.267 149.658
12    169.544 170.258 159.317 157.133 154.942 196.600
13    84.825 170.825 170.467 151.842 171.758 172.492
14    161.475 159.417 -169.067 -169.608 -83.544 -195.217
15    -153.542 -160.475 -162.833 -173.717 -173.000 -153.217
16    -166.683 -167.267 -81.258 -192.875 -151.333 -163.433
17    -165.633 -176.592 -176.000 -156.083
```

Description:

1	month, day, year, time, observer (system) name
2	number of cell groups, group 1 name, number of cells in group 1, group 2 name, number of cells in group 2
3	total number of observations, mean of reference group, accepted standard deviation
4	restraint - one number for each cell in each group; 1 identifies the cell as included in the mean of the reference group, 0 the cell is not included
5	cell IDs - group 1 first, group 2 next
6 - 9	measurement design - cells in line 5 are assigned position numbers 1 through N (14 here) pairs of numbers indicate difference measurements to be made, i.e., 1 - 5 indicates to measure cell "1" (1601) minus cell "5" (Z1). The number of pairs of numbers must agree with the number specified on line 3.
10	nominal temperature of group 1, actual temperature of group 1, nominal temperature of group 2, actual temperature of group 2
11 - 17	actual measured cell differences in microvolts

"D" files

"D" files contain the results of the standard cell intercomparison data; only one set is shown here. A file may contain additional sets of data taken within the same week.

Sample data: DZEN_8630

```
1     860730  14  40  26  0.08247  0.0133  NBS-Z  12:15:11  2  -0.596
2     1600  4  30  30.011  -0.6226
3     1601  1602  1603  1604
4     Zener  10  30  30  0
5     Z1  Z2  Z3  Z4  Z5  Z6  Z7  Z8  Z9  Z10
6     1  1601  8103.3617  1  1602  8096.9666  1  1603  8100.4951  1
        1604  8099.1765
7     0  Z1  7945.575  0  Z2  7947.2625  0  Z3  7904.0645  0  Z4
        7927.444
8     0  Z5  8015.7105  0  Z6  7926.848  0  Z7  7929.7768  0  Z8
        7937.7688
9     0  Z9  7930.194  0  Z10  7939.9563
```

Description:

1	YYMMDD, total number of cells, number of observations, degrees of freedom, standard deviation, left-

right component, measurement system ID, time, number of cell groups, average DVM offset

2 group 1 name, number of cells in group 1, nominal temperature, actual temperature, temperature correction

3 group 1 cell IDs

4 group 2 name, number of cells in group 2, nominal temperature, actual temperature, temperature correction

5 group 2 cell IDs

6 - 9 restraint flag, cell ID, cell emf in μV reduced by 1.01 V. The restraint flag=1 if the cell is included in the reference group mean, =0 if not. The restraint, ID, and emf triplet is repeated for each cell measured.

"Z" files

"Z" files contain the raw Zener measurements taken between the unknown Zener references and the transfer Zener; only one set is shown here. A file may contain additional sets taken on the same day.

Sample data: Z860730

```
1     860730 15:2847 3 2 3
2     DVM 1 1.0181223E+6
3     DVM 2 1.0181238E+6
4     5018-1 10 -179729.8
5     5018-2 1 21.9
6     5018-3 1 -17988.95
7     DVM 1 1.01812225E+6
8     DVM 2 1.01812395E+6
9     5018-1 10 -179730.15
10    5018-2 1 22.05
11    5018-3 1 -17988.8
12    DVM 1 1.0181223E+6
13    DVM 2 1.0181236E+6
14    5018-1 10 -179729.9
15    5018-2 1 21.7
16    5018-3 1 -17988.95
17    DVM 1 1.01812225E+6
18    DVM 2 1.0181237E+6
```

Description:

1 YYMMDD, time, number of repeated sets of Zener measurements, number of DVM calibrations per Zener measurement, number of Zeners measured

2 - 3	string "DVM" to indicate a direct DVM reading of one of the resistors of the resistive divider, the number of the resistor measured (1-10), the reading in μV
4 - 6	Zener ID, tap number of resistive divider to which Zener was compared (0-10), measured difference in μV
7 - 18	repeated measurements, stored the same as lines 2 - 6

"R" files

"R" files contain the final calculated results of the Zener data; only one set is shown here. A file may contain additional sets of data taken within the same week.

Sample data: RZEN_8632

```
1     860730  12:53:36  3  3
2     4  8131.1000  1601  1602  1603  1604
3     5018-1  10.0000047045  0.00333333176584
4     5018-2  1.01799810104  0.142727802049
5     5018-3  0.999989513271  0
```

Description:

1	YYMMDD, time, number of Zeners calibrated, number of repeated measurements (Zsets)
2	number of reference cells, mean emf of reference cells, reference cell IDs
3 - 5	Zener ID, calculated voltage (volts), standard deviation of the mean (ppm) calculated from the repeated measurement sets

<u>System Software Requirements</u>

The measurement software will run satisfactorily on all H/P Basic systems 3.0 and higher. The following BIN files are required for proper operation.

CLOCK	CS80
KBD	MAT
ERR	HPIB

For use with an SRM the following additional BIN files are needed.

SRM	DCOMM

NBS MEASUREMENT SERVICES: STANDARD CELL CALIBRATIONS

Bruce F. Field

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

U.S. DEPARTMENT OF COMMERCE, Clarence J. Brown, Acting Secretary
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director
Issued October 1987

Table of Contents (cont.)

Table of Contents

List of Figures

List of Tables

NBS MEASUREMENT SERVICES: STANDARD CELL CALIBRATIONS

Bruce F. Field

Abstract - This document describes the procedures used at NBS to calibrate standard cells in terms of the U.S. Legal Volt. The two calibration services that are offered by the Electricity Division are discussed; Regular Calibration of client standards at NBS; and the Volt Transfer Program which is a process to determine the difference between the U.S. Legal Volt and the volt as maintained by a client laboratory. The operational procedures used to compare standard cells and to maintain the U.S. Legal Volt via the ac Josephson effect are discussed in detail.

1. Introduction

To maintain a unit of electromotive force (emf), standards laboratories typically use one or more groups of electrochemical cells called standard cells. Standard cells are physical representations of the unit of emf and are used as standards against which the emf of other cells and systems are compared. At present, saturated cadmium-sulfate-type (Weston) cells are used when high accuracy, 1 ppm or less, is required in maintaining a unit traceable to NBS.

Standard cells are electrochemical systems composed of two dissimilar electrodes immersed in an electrolytic solution. They are not intended to supply electric current and, therefore, are of different design from those electrochemical systems which are intended for such purpose. The stability of the emf of the cell depends on the chemical equilibrium within the cell. Although the emf of a cell is highly reproducible and generally exhibits a fairly constant emf, it must be periodically recalibrated to eliminate the effects of long term drift. The frequency with which recalibrations are required is a function of the accuracy requirement, the number of cells used to maintain the volt, and the stability of the individual cells. This must be determined by the client laboratory using the cells.

Calibration of cells in terms of the U.S. Legal (or NBS) Volt is accomplished via two services offered by the Electricity Division of the NBS.

Regular Calibration

Clients may send cells to NBS for direct comparison to NBS standards which are in turn calibrated in terms of the NBS Volt. At the conclusion of the test a report is issued with a statement of the values of the emfs of the cells and estimated uncertainties for the emf values, while the cells were at NBS. No additional

uncertainty is included for transportation effects, long term drifts of the emfs of the cells, or measurement errors in the client's laboratory.

Volt Transfer Program

The Volt Transfer Program (VTP) is a measurement assurance program designed to determine the difference in the unit of voltage maintained by a client laboratory using standard cells and the NBS Volt as maintained by NBS. In this program, a group of NBS-owned cells in a thermoregulated enclosure is measured at NBS, shipped to the client laboratory, compared to the client cell group, and returned to NBS for final measurements. The measurements to be made at the client laboratory are specified by NBS and the data is reduced and reviewed by NBS. A report is issued containing the difference of the client unit of voltage from the NBS Volt and specifying an uncertainty of that difference observed during the time of the test. In this experiment the uncertainty in the transportation of the standard cells is estimated and included in the final uncertainty.

2. Description of Service

Regular Calibration

NBS offers a voltage calibration service for saturated and unsaturated standard cells of the Weston type. All cells are compared to NBS standards which are calibrated in terms of the U.S. Legal Volt defined via the ac Josephson effect. Cells accepted for test include shippable or unshippable saturated cells that are either designed for immersion in an oil bath or that are housed in a thermoregulated enclosure. NBS maintains two oil baths, one at 28 °C and one at 30 °C, for testing immersible-type cells.

Routine calibrations of standard cells involve the following considerations:

1) Unsaturated cells require approximately 3 weeks for a complete calibration. The emfs of such cells are read daily for a period of 10 days. If the measured emfs fluctuate unduly or are unusually low, or if the cells show abnormal indications, the report of calibration will reflect these circumstances. Unsaturated cells are not likely to be injured by normal transportation (mail or express) if they are packed carefully. Because of the possible hazard from freezing, shipment during very cold weather should be avoided.

2) Saturated standard cells of the unshippable type should always be transported by messenger because such cells should never be tipped from an upright position by more than 45 degrees in any direction. Unshippable saturated cells contained in portable, temperature-regulated enclosures

should also be transported by messenger, and with the enclosure activated or under power if possible.

3) Saturated standard cells of the shippable type housed in portable thermoregulated enclosures should be packed carefully and shipped under power if possible. Liquid-in-glass thermometers normally mounted in such devices should be removed and provided with additional rigid packing for protection against breakage. Enclosures having a nominal cell temperature of 28 °C or lower should not be transported during the summer due to the danger of over heating. To prevent overheating, enclosures should not be energized by the ac power line while they are packed in shipping containers. When an enclosure is shipped under temperature control, arrangements should be made by the client to have it promptly delivered to and picked up from the airport (if shipping by air freight) to avoid exceeding the capacity of the batteries powering the enclosure.

4) Saturated standard cells which arrive having been maintained continuously at their nominal temperature of use will, workload permitting, undergo test immediately upon receipt for a period not to exceed 4 weeks, unless other arrangements are made. If such cells perform abnormally with respect to the typical performance of like cells in similar environments, the owner will be notified. Arrangements for further testing may be made at that time if desired. Cells will be returned as soon as possible after calibration.

5) Saturated cells arriving at a temperature other than their nominal temperature of use will be brought to their use temperature as soon as possible after receipt. Starting one month after they are initially brought to temperature, daily readings will be taken to observe the stability of the cells. When the cells stabilize, 10 daily readings will be taken and used to assign values to them. This process will not exceed 90 days unless special arrangements are made.

6) Prior to each daily reading of the cell emfs the temperature of the enclosure (or the oil bath) will be read using the temperature measuring device included in the enclosure, or special arrangements can be made to have the temperature monitored by a calibrated NBS platinum resistance thermometer. A calibrated platinum resistance thermometer is always used to monitor the temperature of the oil bath with the temperature reported relative to the International Practical Temperature Scale (IPTS-68).

The Volt Transfer Program experiment consists of four phases.

1) Preliminary Laboratory Study. During this phase NBS personnel will discuss with client laboratory personnel various aspects of the experiment, suggest experiments to evaluate the measurement process, and establish a tentative shipment and measurement schedule. If necessary NBS personnel may visit the laboratory. NBS may restrict participation in the Volt Transfer Program to those clients who use quality-control procedures to monitor their representation of the volt. Guidelines for setting up control chart procedures can be found in VTP Control Chart Requirements [40].

2) Shipment of an NBS Transport Standard. An NBS transport standard, a group of 4 cells (usually) in a thermoregulated enclosure with a battery pack, is shipped via air freight to the client laboratory. The success of the experiment depends on keeping the transport standard under power continuously and expediting its transport. Detailed instructions for transport are given to the client.

3) Measurement of the Transport Standard in the Client Laboratory. Measurements comparing the cell emfs of the transport standard to the emfs of the client laboratory cells are done in accordance with NBS specified procedures. Usually measurements are made daily over a two to three week period. Data is sent to NBS immediately after each measurement for reduction and review. The conclusion of this phase is determined by NBS after review of the client data; in some cases additional measurements may be required.

4) Analysis of the Data and Issuance of a Final Report. Upon return to NBS the transport standard will be remeasured to determine if it changed during the experiment. After sufficient measurements have been made the data is analyzed to determine the difference between the units of voltage at NBS and the client laboratory ($V_{LAB} - V_{NBS}$) and its uncertainty. If the difference exceeds the estimated uncertainty, new emf values for the client cells will be recommended to reduce the difference to zero.

3. Calibration Apparatus and Measurement Approach

Voltage Difference Measurements

All cells at NBS are calibrated by comparing them directly to NBS owned standard cells which are in turn calibrated in terms of the U.S. Legal Volt. All comparisons use redundant measurement designs and a series opposition method, as described in NBS Technical Note 430 [16]. Figure 1 shows a simplified schematic of the measuring circuit.

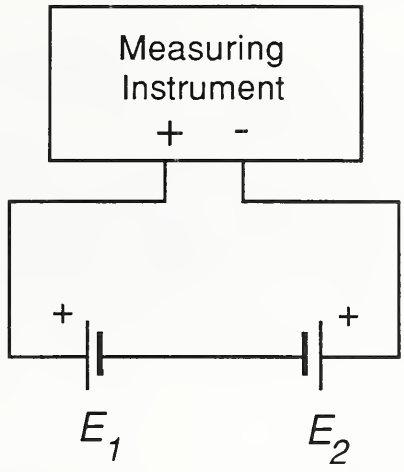


Fig. 1. Simplified measuring circuit.

Two cells are connected in series opposition (one unknown, one reference) and the small voltage difference is read using a digital voltmeter (DVM). In an ideal situation the difference in emf as measured by the voltmeter is:

$$E = E_1 - E_2$$

where E_1 and E_2 are the emfs of the two cells being compared. However, in the real situation there may be spurious emfs in the circuit. In general these can be classified into two categories:

1. Those emfs that remain constant, or relatively so, in relation to the interval over which a complete set of measurements is made.

2. Those emfs that vary rapidly (referenced to the interval over which a complete set of measurements is made).

If the emfs are of the second type they will have the effect of decreasing the precision of the process. On the other hand, if they are of the first type they will have the effect of introducing a systematic error into the measurement result such that

$$E = E_1 - E_2 + P$$

where P is the constant, and as yet unknown, emf. It is possible to estimate P by taking a second measurement E' as shown in Fig. 2, where

$$E' = E_2 - E_1 + P\,.$$

Taking the difference between the two expressions gives

$$E - E' = 2(E_1 - E_2),$$

thus yielding an estimate of $E_1 - E_2$ free of P. The pair of measurements are said to be "left-right" balanced. That is, if there is a positional effect it is balanced out of the final result. This technique is analogous to that used to eliminate the inequality of balance arms in precision weighing on a two-pan equal-arm balance. In order to designate the cell positions from the operational point of view, they are frequently designated as unknown and reference (or left and right) relative to the input terminals of the measuring instrument.

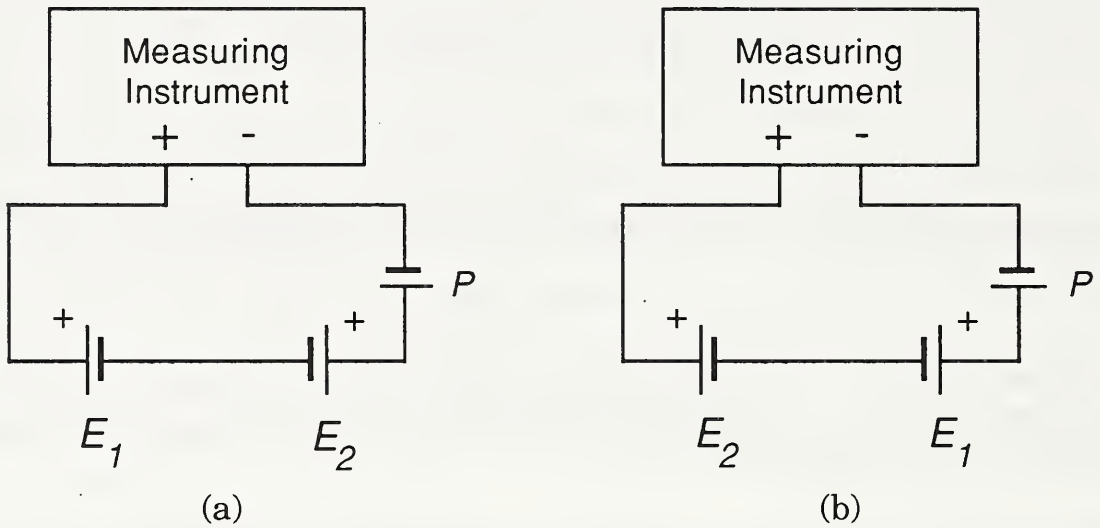


Fig. 2. Two measurement configurations to eliminate positional errors (P).

With classical potentiometers the left-right effect P does not vary significantly over the time of the measurements and it is sufficient to require that the total design be left-right balanced, i.e., each cell appears on the left side and right side of the measuring system an equal number of times irrespective of their order of appearance in the design. Digital voltmeters, however, have been found to introduce a small time-varying dc offset that must be eliminated by immediately reading every cell pair twice, once in the normal mode and once with the polarity of the input reversed. The algebraic difference of these two measurements divided by two is considered to be a single measure of the cell difference with the DVM offset eliminated.

Measurement Apparatus

Two voltage measuring systems are currently in use at NBS for cell comparisons. Except for the switching systems, they are identical. One system (System VTP) is used for comparison of NBS primary standards with NBS working standards and for comparison of the working standards to VTP transport standards. The second system (System REG) is used for comparison of NBS working standards to client cells sent in for regular calibration. Figure 3 is a block diagram of the measurement apparatus of the systems.

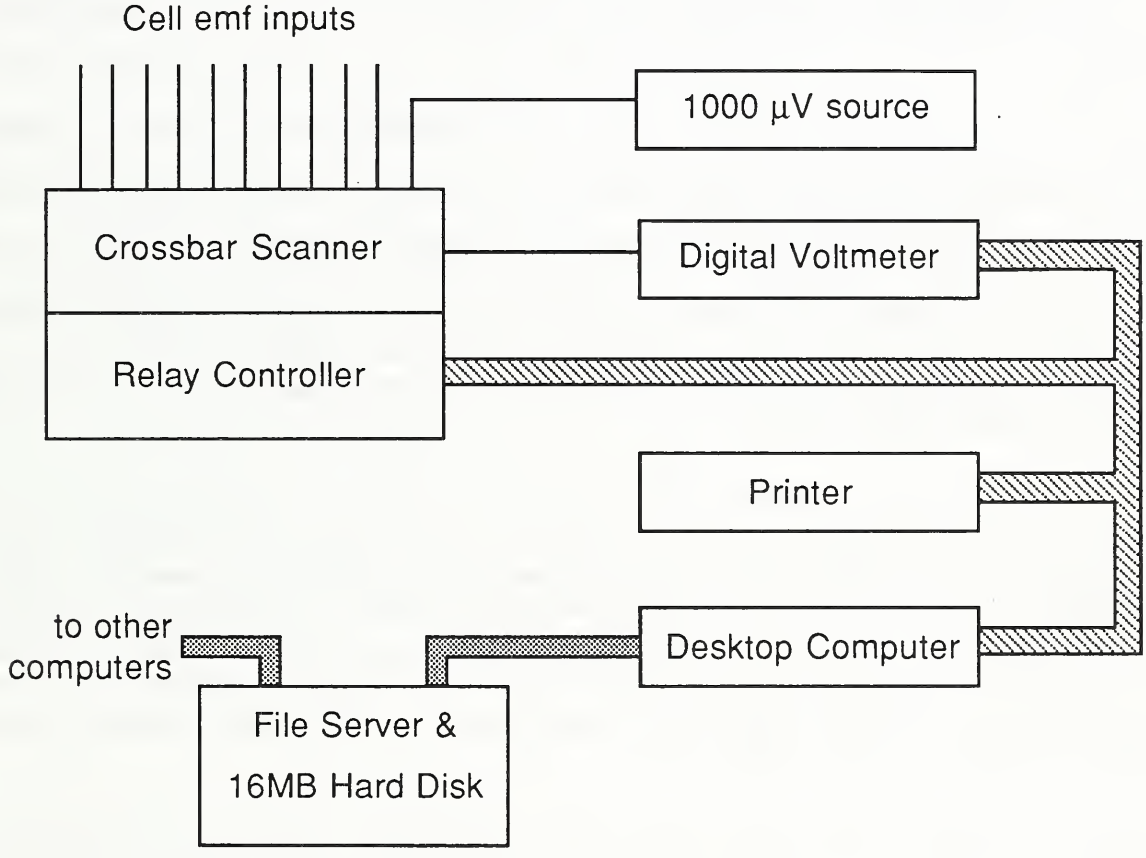


Fig. 3. Block diagram of the two measurement systems used for comparison of pairs of standard cells.

Using a measurement design, the computer commands the crossbar scanner to connect two cells in series opposition to the DVM; the DVM takes a number of readings of the voltage difference (usually three); the crossbar scanner reconnects the cells with reversed polarity; and the DVM takes three more readings. The computer records the readings, and when the experimental design is complete, it calculates the estimates of the cell differences using a least-squares method and records the cell emfs on the hard disk and on a floppy disk for backup. The hard disk is shared among several desktop computers to permit the data from both measurement systems to be stored in one location,

and to permit access to the data from other machines even while measurements are being made. A summary of the basic statistics is printed after each measurement set.

The digital voltmeters used to make the difference measurements are 6-1/2 digit high-resolution voltmeters. Two performance parameters of the voltmeters are critical to this application: low input bias current and small random measurement error. The design of the saturated standard cell is such that the emf stability depends on electrochemical equilibrium within the cell. Small charging or discharging currents that may be present at the input of the DVM will produce changes in the cell emf that may be significant. The voltmeters used at NBS have a measured input bias current of 12 pA which contributes a negligible error to the measurement. On the lowest voltage range of the voltmeter the standard deviation of a single measurement is 0.034 μV with an integration time of 18 seconds. The integration time was chosen specifically to obtain a measurement standard deviation of less than 0.04 μV. Since only small voltage differences are being measured (the worst case is about 600 μV when comparing cells at 37 °C to cells at 30 °C), only a modest DVM reading accuracy of 17 ppm (parts-per-million) is required for 0.01 μV accuracy in the voltage differences. (Unless otherwise specified, all uncertainties in this document are meant to correspond to a one standard deviation estimate.) To ensure that the voltmeter is within the required accuracy, it is checked every day by reading a calibrated 1000 μV source.

A commercial Zener standard that produces 1000 μV is connected to one input of the scanner switch, and a piece of copper wire (to provide a low-thermal short) is connected to a second input. Connecting the first input minus the second input presents the 1000 μV source to the DVM. Actually, two Zener standards and two copper wires are used, one set for each measuring system. Each 1000 μV source is read daily with the DVM to perform a one point calibration on the gain accuracy of the DVM. The sources are periodically calibrated using a resistive divider and a standard cell.

Pairs of cells are connected to the digital voltmeters by one of three low-thermal switching systems. The first system consists of a 10x10 matrix crossbar switch with latching solenoids. This switch is used to connect up to 80 VTP transport cells to measuring system VTP. A second modular switching system is also used to connect NBS reference cells to measuring system VTP. This system consists of up to five switch modules each capable of connecting one or two of ten cells to a common 4-wire signal bus [27]. The output of this bus is connected to the DVM of measuring system VTP. The third system, consisting of two 10x10 matrix crossbar switches with latching solenoids and a 2x5 selector switch, is used to connect up to 300 client cells to measuring system REG.

All the switching systems are based on a low-thermal crossbar switch; however, each system is constructed somewhat differently with capabilities for switching different numbers of cells. Common to all three systems is the capability to randomly select any two cells under computer control and connect them in

series opposition to the DVM in "normal" or "reversed" polarity. The matrix crossbar switches are actuated by momentarily energizing the switch solenoids, and then permanent magnets within the switches latch the contacts closed until a second release solenoid is energized. This momentary action is designed to minimize the heat input to the switch which in turn would produce thermal gradients and generate large thermal emfs in the switch. The uncompensated thermal emfs in these switching systems have been measured and found to be less than 20 nV. The solenoids in the modular switch system (used for intercomparison of NBS reference cells) are continuously energized during contact closure, but the switch has been modified to locate the solenoids outside the chamber containing the switch contacts to minimize the heat input to the contacts. The uncompensated thermal emfs have been determined to be less than 6 nV.

Redundant Measurement Designs

Redundant measurement designs are used to compare cell emfs because they provide diagnostic information about potential errors or problems with the measuring system. Two important parameters are estimated by the designs: the within-day standard deviation and the left-right (positional) effect in the measuring system. The within-day standard deviation permits laboratory personnel to estimate the quality of the measurements and, if necessary, to eliminate anomalous observations from the measurement design and re-compute the cell emfs with little loss in accuracy.

Two specific designs are generally used, one for comparison of two groups of four cells, and one for comparison of one group of six cells to one group of four cells. These designs are described in detail in Appendix A. On occasion additional designs are required for groups with varying numbers of cells; these are constructed similarly to the ones in the Appendix. The cell emfs are estimated from the cell difference measurements by solving the overdetermined set of equations using the least-squares method and including the constraint that the mean emf of a given group of cells (i.e., the reference group) is known. This results in assigning individual emf values to all the cells in terms of the mean emf of the reference group. This calculation is performed by the computer program READBOX which is discussed in Appendix D.

Temperature Measurements

All saturated standard cells exhibit an emf change with temperature and must be maintained at a constant temperature. Cells sent to NBS for calibration are either housed in portable thermoregulated enclosures, or are immersed in NBS-provided constant-temperature oil baths. Thermoregulated enclosures generally contain a temperature sensing element which can be used to monitor the temperature variations within the cell enclosure. Typical devices are mercury-in-glass thermometers, thermistor bridges, and platinum resistance thermometers. NBS follows the manufacturer's recommended procedure for

monitoring the temperature using these devices. The temperature scale embodied in the temperature device is taken as correct, primarily because accurate knowledge of the temperature is unnecessary. Cell emfs are corrected only for small *changes* in temperature, referenced to a nominal temperature as established by the temperature device.

For cells housed in the NBS oil baths, calibrated platinum resistance thermometers and an ac resistance thermometer bridge are used to measure the temperature. The temperature of the oil bath is stable and uniform to at least 0.001 °C. The estimated uncertainty of the temperature measurement with respect to the International Practical Temperature Scale (IPTS-68) is 0.005 °C (3 standard deviation estimate), and includes uncertainty for the power dissipation in the thermometer, errors in the determination of the triple point of water, and errors in the drift of the calibration constants between calibrations.

Cell emfs are corrected for small day-to-day temperature changes by monitoring the temperature as described above and applying an emf correction based on the International (or Wolff) Temperature Formula [13]:

$$E_T = E_{20} - 0.00004060(T-20) - 0.000000950(T-20)^2 + 0.000000010(T-20)^3$$

in volts where E_T is the emf at temperature T and E_{20} is the emf at 20 °C. This equation is used to correct the cell emfs to any arbitrary nominal temperature by computing the correction for the actual temperature with respect to 20 °C and subtracting the correction for the nominal temperature with respect to 20 °C. The formula is an approximation and is not exact for all cells although it provides reasonable accuracy (better than 0.1 μV) if the total correction is less than 1 μV.

4. Maintenance of the Volt

The ac Josephson Effect

The U.S. Legal (or NBS) Volt is presently defined in terms of the atomic constants h (the Planck constant) and e (the elementary charge) via the ac Josephson effect [26]. Critical to this definition is the role played by a Josephson junction which may be regarded as a frequency-to-voltage converter, where the frequency-to-voltage ratio is precisely equal to the combination of physical constants 2e/h. The current value of 2e/h used for maintaining the NBS Volt is 483593.420 GHz/V_{NBS}. The U.S. Legal Unit of voltage is known to be smaller than the SI unit by about (9 ±1) ppm (3σ) [41]. (The dominant system of units used throughout the world to express the results of physical measurements is *Le Systèm International d'Unités* or International System of Units, abbreviated SI.)

When two weakly coupled superconductors are irradiated with microwave energy, the assembly (a Josephson junction) can be used to produce precise voltages described by the following equation:

$$E = nf/(2e/h)$$

where n is an integer and f is the frequency of the irradiating microwave energy. A variety of experimental tests (for material dependence, temperature dependence, etc.) and theoretical investigations of the Josephson relation have been made which indicate that for ordinary Josephson devices (particularly tunnel junctions) with conventional current-voltage lead configurations the ratio is exact to at least a few parts in 10^8 [23]. In practice, frequency measurements of the microwave energizing signal are based on the NBS unit of time interval, the atomic second.

The Josephson Array

Arrays consisting of from 1500 to 2076 Josephson junctions have been used to produce a total voltage of up to 1.2 V [42,43]. These arrays do not require individual control of the bias currents as is usual, but avoid the multiple bias problem by using constant-voltage steps which cross the zero current axis of the junction I-V curve. This arrangement allows a large array of junctions to share a common current bias at or near zero. The arrays are fabricated using niobium and lead alloys and are stable at room temperature.

Microwave Apparatus

The microwave radiation applied to the Josephson array is supplied by a 60 mW Gunn diode oscillator at 94 GHz which is frequency-stabilized by a frequency locking counter containing a quartz-crystal oscillator. The short term frequency stability of the microwave radiation (15 min) is about 1 part in 10^9. The frequency is measured by the frequency counter with a resolution of 1 part in 10^{10}. The accuracy of the counter time base is regularly checked against the U.S. frequency standard by comparing the counter time base to a 100 kHz high-stability oscillator which is simultaneously compared to the signal from WWVB using a VLF comparator.

DC Measurement Apparatus

Figure 4 shows some of the dc measurement apparatus used to compare the array voltage at 1.018 V to a 1.018 V Zener reference standard. (Not shown is a low-thermal-emf crossbar selector switch used to connect one of three Zener references to the measurement system.) To measure the voltage of the Zener reference, the array is adjusted to produce a voltage nearly equal to the Zener reference by (a) adjustment of the bias current and microwave power to select a voltage step, and (b) adjustment of the microwave frequency to fine-tune the step voltage. The step is observed on the oscilloscope to check for any abnormalities.

The difference between the array and Zener reference is then measured by averaging several readings (E_1) of the digital voltmeter. Low-thermal-emf reversing switch A is reversed and several more readings (E_2) are taken to eliminate offsets in the digital voltmeter. The bias current to the array is reversed and adjusted to produce -1.018 V and reversing switch B is switched. Two more sets of readings (E_3 and E_4 respectively) are taken with switch A in its normal and reversed positions. This action is required to eliminate thermal emfs in the leads from the array to the reversing switches. The Zener voltage is calculated as $(E_1 - E_2 - E_3 + E_4)/4$ + nf/(2e/h) where n is the integer step number, f is the microwave frequency, and 2e/h is as defined above. This measurement sequence is repeated five times for each Zener reference, with a typical standard deviation of 0.009 μV for the five measurements, and takes about 12 minutes.

A final check is made on the thermal emfs in the leads from the Zener reference to reversing switch B by replacing the Zener reference with a short and adjusting the array to operate on the zero-voltage-step with the microwave power set to zero. The same measurement sequence is run as for the 1.018 volt measurement. The residual thermal emfs thus determined are subtracted from the 1.018 volt measurements. (The Zener references are connected to one of the modular crossbar switches described in section 3. The output of this switch is considered to be the output of the Zener references as all measurements of these references are made through the switch. Any stable thermal emfs within the switch, or between the switch and the Zener reference, are cancelled as they add equally to both the 2e/h measurements and the cell comparisons.)

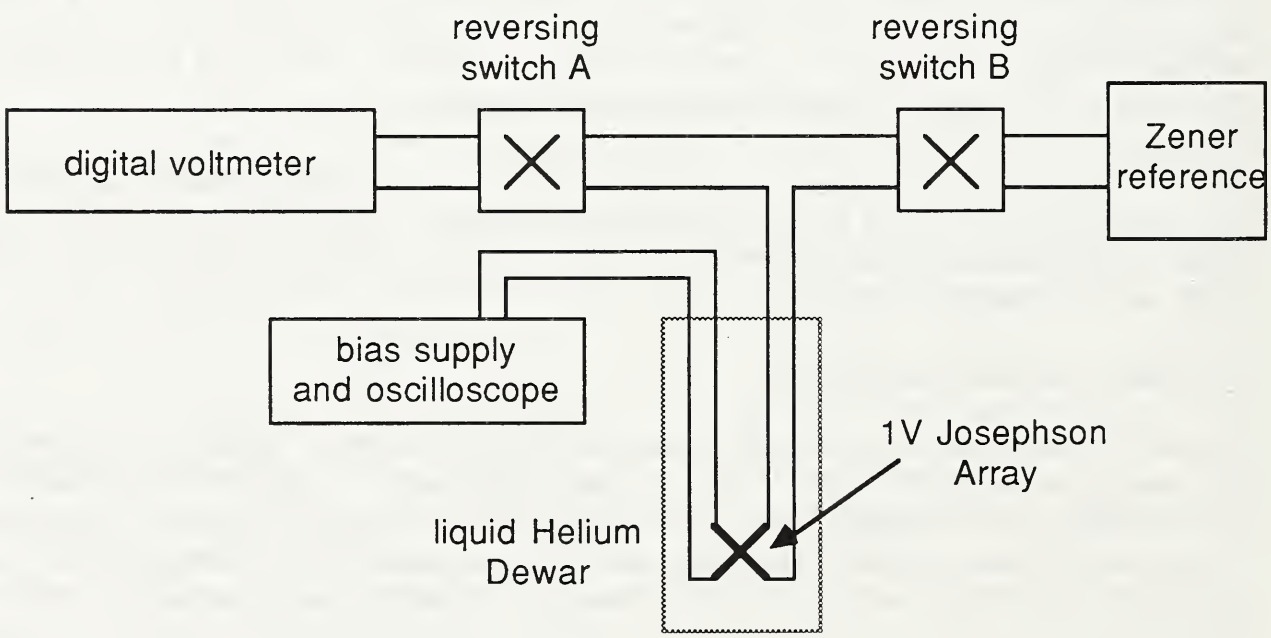


Fig. 4. Simplified diagram of the measurement system used for calibration of three Zener references.

Standard Cell Comparisons

Two standard cell groups (consisting of 10 cells total and considered to be the NBS primary groups) are calibrated in terms of 2e/h at weekly intervals. The calibrations consist of making cell comparison measurements (using redundant measurement designs) between the primary groups of cells and a group of three Zener reference standards each producing 1.018 V. The pattern of group comparisons currently in use is shown in Fig. 5. A first set of cell-Zener comparisons are made and then the three Zener standards are each calibrated in turn by the Josephson Array. After the Array calibrations the cell comparisons are repeated to determine if the Zener standards changed during the measurements.

Only Zener standards are compared to the Array because the Array voltage occasionally and unpredictably jumps abruptly to a slightly different voltage step. If the Array were being compared to a cell while this shift occurred, a small charging or discharging current would be introduced into the cell, slightly changing it's emf. Zener standards, however, are not significantly affected by this type of current pulse.

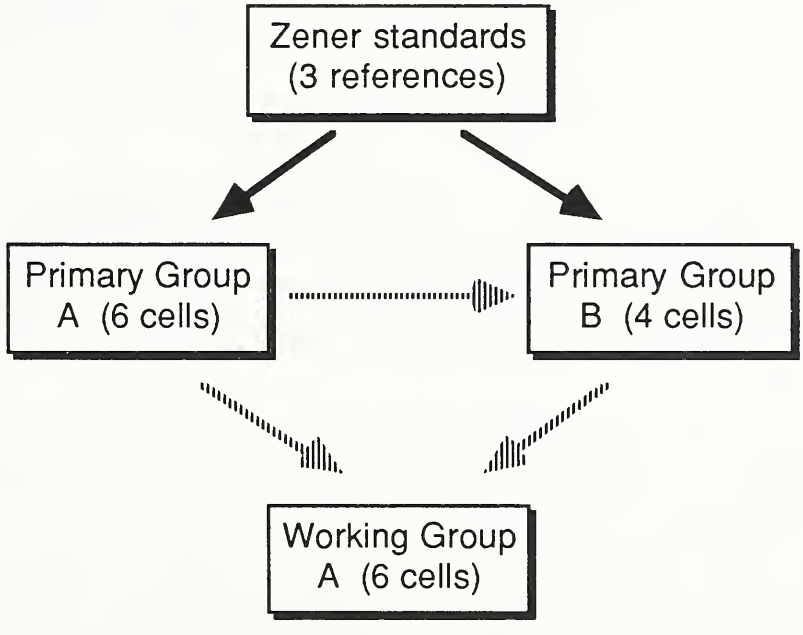


Fig. 5. Measurement sequence used to assign values to the primary groups based on 2e/h measurements.

TABLE I

UNCERTAINTIES IN THE ASSIGNMENT OF THE MEAN EMFS OF THE PRIMARY GROUPS

Source of uncertainty	1 std. dev. estimate (ppm)
Microwave frequency	0.005
Assignment of Zener reference group using Array	0.007
Random uncertainty of comparison of Zener reference group to the mean of Primary Group A	0.009
Change in the Zener reference group during the 2e/h measurement	0.019
Uncompensated thermal emfs in the cell switches	0.005
RSS total	**0.023**

Uncertainty

Table I summarizes the sources of uncertainty in assigning a value to the mean emf of Primary Group A at the time of a 2e/h measurement. The microwave frequency is measured by a frequency counter which is calibrated in terms of WWVB. An uncertainty is included for measurement uncertainty and drift of the oscillator frequency.

For each daily 2e/h measurement each Zener reference is compared to the Array five times. The pooled standard deviation of the mean calculated from individual comparisons is 0.0042 μV (60 degrees of freedom). This value, divided by the square root of three (0.0024 μV, 0.002 ppm), is used as the limit of the random component of uncertainty in comparing the mean of the three Zener standards to the Josephson Array. The thermal emfs in the leads to the Zener reference are measured and subtracted from the Zener values. The uncertainty of determining the thermal emfs is 0.0042 μV (0.004 ppm). All other known sources of systematic error are negligible.

The random component of uncertainty in comparing the Zener standards to the primary group is determined from the 0.034 μV pooled standard deviation of a

single cell-Zener comparison. Because of the redundant measurements of the measurement design, the uncertainty of the difference of the mean of the Zener reference group to the mean of the primary group is 0.012 μV. Two measurement designs are made for Primary Group A (before and after the 2e/h measurements); thus, the uncertainty of it's mean emf is reduced by the square root of two for a total uncertainty of 0.009 μV (0.009 ppm).

The difference of the mean Zener group emf minus the mean primary group emf typically shows a change of (0.008±0.011) μV in the before and after comparisons described above. These appear to be caused by small random shifts in the Zener references emfs. Taking a conservative approach, we assign an uncertainty of 0.019 ppm (0.008+0.011) for changes in the mean Zener group emf during the measurements.

Modular low-thermal switches (described in Section 3) are used to connect the various cells to the measuring system. The thermal emfs in these switches that do not cancel on cell reversal have been measured to be 0.005 μV (0.005 ppm).

The Root-Sum-Square (RSS) total is an estimate of the uncertainty in assigning a value to the mean emf of Primary Group A based on one 2e/h measurement. This is an estimate of how well the present system would agree with another totally independent 2e/h system. Based on data from actual 2e/h measurements (see Fig. 6, for example) the reproducibility with which the mean emf of a single group of standard cells can be determined using the present Josephson apparatus is about 0.020 ppm.

5. Operational Procedures

Calibration of the Laboratory Primary Cells

Standard cell emfs drift with time so the use of a simple time invariant model for the cell emf can lead to unacceptably large step changes in the disseminated volt each time the volt is reassigned from Josephson Array comparisons (2e/h measurements). In addition, it is desirable to "average" several 2e/h measurements to reduce the random error of the measurements. Thus, we use a model for the primary cell group emfs that predicts a linear drift with time and new model coefficients calculated after every fourth 2e/h measurement.

2e/h measurements are made at approximately weekly intervals. For every fourth measurement (a cardinal measurement) additional cell comparisons between the primary and working cell groups are made. These data are used to reassign the emfs of all the primary cell groups. In general, least-squares lines are fit to the last five or so cardinal measurements and these lines are used to predict weekly values for the means of the cell emf groups for the coming month. The exact number of measurements chosen for the fitted lines depends on (a) how well the cell emfs fit a linear model, and (b) the random scatter in the 2e/h measurements of the cell emfs. The judgement and experience of the

laboratory staff are used to determine the models. Figure 6 shows the typical behavior of the mean emf of six cells in Primary Group A plotted against time. The line plotted in Fig. 6 is the least-squares fitted line to the data; the residual standard deviation of the fitted points about the line is 0.020 μV.

The three 2e/h measurements made on "off-weeks" each month are used to check the prediction of the assignment of the cells. The collection of cell groups that constitute the primary groups changes as cells need replacement or enclosures need repair. In general, the primary groups consist of two or three groups of four to six cells each, with about 10 to 12 cells total. Generally, these standard cell emfs drift with time; drift rates of the different cells range from +0.1 to -0.7 ppm/year.

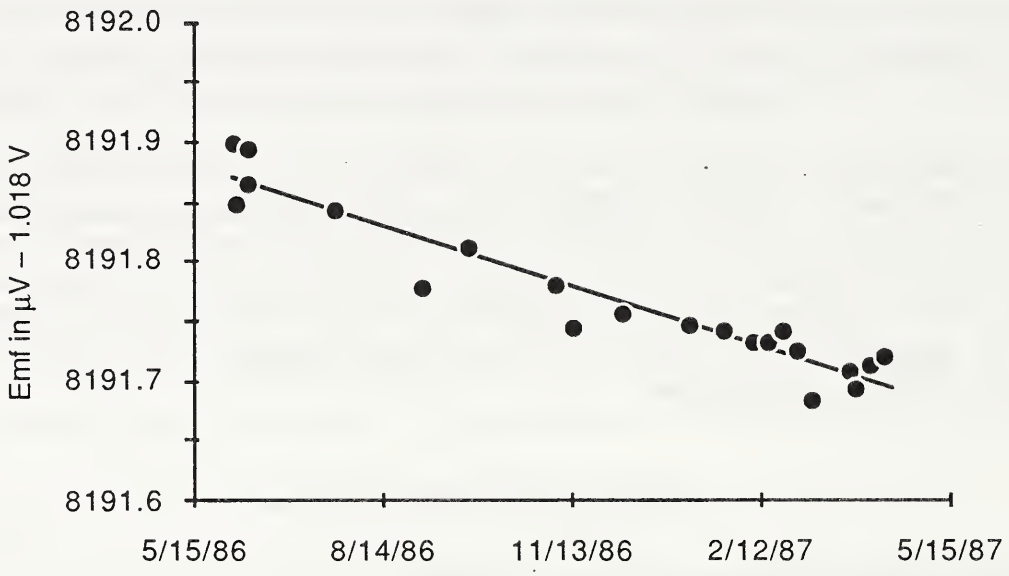


Fig. 6. Mean emf of Primary Group A as determined from 2e/h measurements.

The primary cell groups also serve as a check standard for the 2e/h measurements. If an individual 2e/h measurement assigns values to the primary cells that are inconsistent with the predicted values, all the measurement systems are investigated to determine the source of the inconsistency. If the problem cannot be resolved, the 2e/h measurement is usually repeated. In the exceptional circumstance where later 2e/h measurements confirm the deviation from the predicted model (usually more than 0.07 ppm) appropriate past workload data are corrected to reflect the change.

Note: Prior to February 1987 two series-connected Josephson junctions producing 10 mV were used as the basis for the U.S. Legal Volt. The small voltage produced by the junctions was "stepped-up" to 1.019 V by special

potentiometers and compared to unsaturated standard cells. These unsaturated cells were compared to the primary cells using a manual potentiometer. Because of the difficulty of these measurements they were made only monthly and they are described in detail in reference 26. Some data (particularly figures 8 and 13) are based on these older measurements. The use of the Josephson Array has reduced the total uncertainty of Table I from 0.036 ppm for the old system to 0.023 ppm for the Array.

Calibration of Working Cell Groups

To minimize the possibility of disturbing the primary cells, the client cells are not compared directly to the primary cells. Instead, as shown in Fig. 7, the primary cells are used to calibrate working groups of cells once each day, immediately before calibration of the client cells or the NBS transport standards for the Volt Transfer Program. The working cell emfs are determined from the predicted values of the primary cells, and are used for that day. The working cells are then compared to client cells and the NBS transport standards. This is feasible because of the small within-day uncertainty of 0.004 to 0.007 ppm which is introduced by the intercomparison of standard cells. Measurements are made using redundant designs and the VTP automated measuring system.

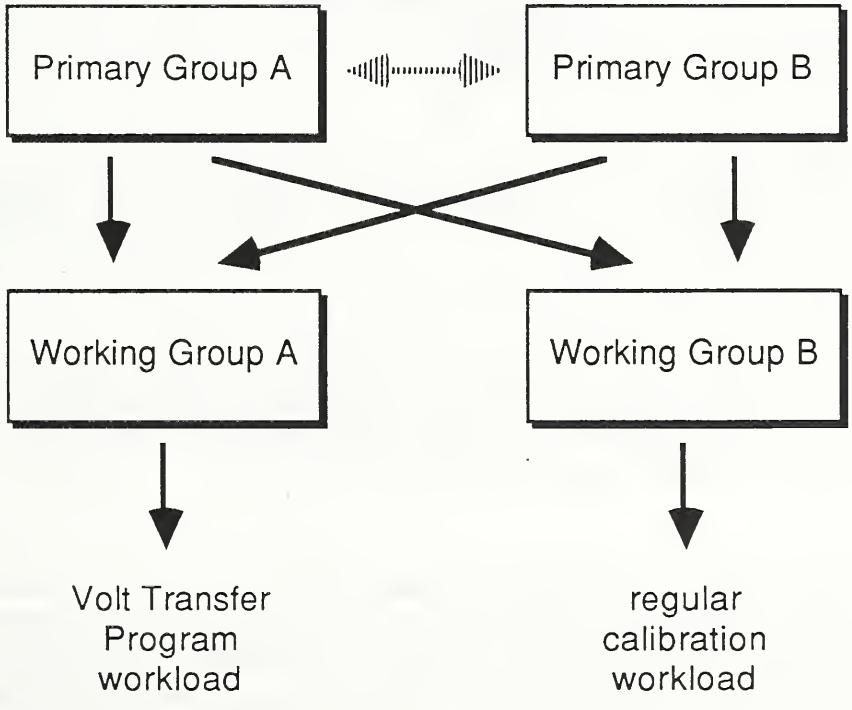


Fig. 7. Measurement sequence used to compare the Regular Calibration and VTP workload to NBS working groups.

This procedure is modified as necessary when cells show erratic behavior or temperature-regulated enclosures fail. At present the primary cells consist of two groups, in different thermoregulated enclosures, that are independently compared to each of the working groups. This results in two assigned values to each working group. In an ideal situation the two values should agree; the difference is a measure of the error of the predicted values for the primary cells. Figure 8 shows the typical behavior of this difference. If the assigned values disagree by more than 0.08 μV (as they occasionally do in Fig. 8), direct comparisons are made between the two primary groups to try and resolve the disagreement. In many cases the disagreement can be traced to abnormal behavior of the temperature controller of one of the enclosures, or a rapidly changing cell emf.

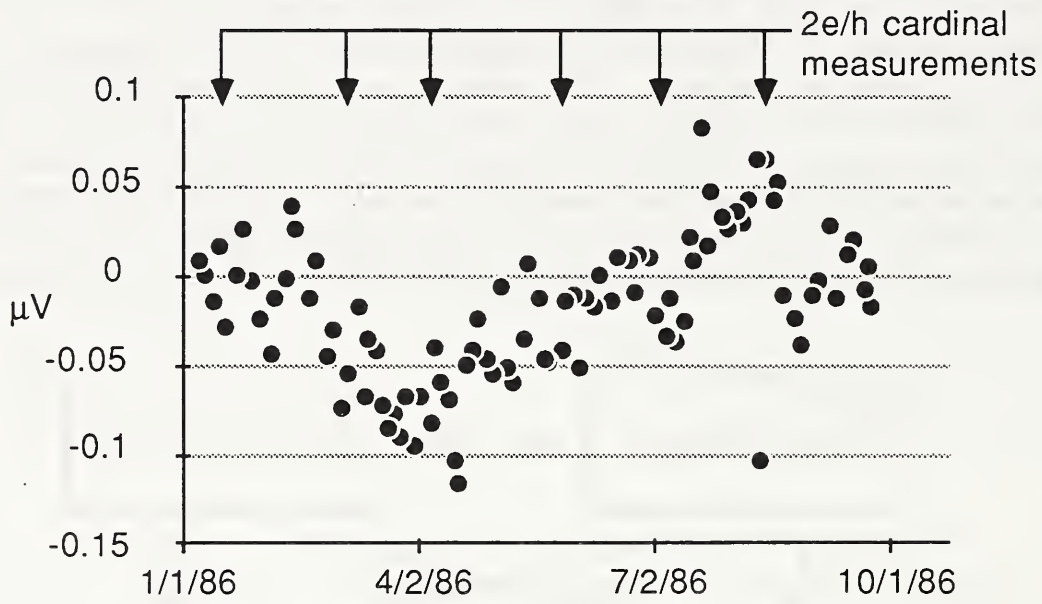


Fig. 8. The difference between the two daily values assigned to Working group A from the two Primary groups. The arrows indicate dates on which the cell emfs were reassigned based on 2e/h measurements.

If the lines fit to the emfs of the primary groups are poor predictors of the emfs, it would be expected that the difference would be nearly zero immediately after a 2e/h measurement and gradually diverge thereafter. The data of Fig. 8 do not indicate this. The data do show correlation with changes in room temperature, even after correcting for the indicated cell temperatures [29]. This appears to be related to less than optimum temperature control of Primary Group B, which is a commercial enclosure. Plans are underway to replace this enclosure with an NBS-designed enclosure similar to the one for Primary Group A [29].

TABLE II

UNCERTAINTIES IN THE ASSIGNMENT OF THE MEAN EMF OF THE WORKING GROUP

Source of uncertainty	1 std. dev. estimate (ppm)
Assignment of primary group (from Table I)	0.023
Day-to-day fluctuation of primary groups	0.028
DVM scale error	0.005
Random uncertainty in comparison of primary group to working group	0.004
Uncompensated thermal emfs in cell switches	0.015
Uncertainty of assignment to working group RSS total	**0.040**

Table II summarizes the uncertainties in assigning a daily value to the mean emf of the working groups. Although several 2e/h measurements are combined to predict values for the primary groups, the uncertainty of the primary group assignment is not reduced from the value quoted in Table I because part of the uncertainty may be systematic to all the 2e/h measurements. An uncertainty is included for the day-to-day fluctuation of the primary groups and the additional uncertainty due to predicting the values of the primary groups up to one month ahead. This is estimated by taking the standard deviation of the difference of the two assigned values to a working group, calculated from the daily comparisons with the primary groups (see Fig. 8).

The DVM scale error is estimated based on readings of the calibrated 1000 μV source which show a scale error of 70 ppm and a maximum cell difference of 64 μV. The random uncertainty of the cell comparisons is estimated from the redundant measurement designs as before with a pooled standard deviation of a single measurement of 0.034 μV.

Low-thermal switches connected to measuring system VTP (described in Section 3) are used to connect the various cells to the measuring system. The

thermal emfs in these switches that do not cancel on cell reversal have been measured to be 0.015 μV (0.015 ppm).

The RSS total uncertainty in assigning a value to the mean emf of either working group is estimated to be 0.040 ppm.

Regular Calibration

Calibration of Client Saturated Standard Cells

Cells received for calibration may be subjected to a stabilization period before measurements are begun. The length of the stabilization period depends on whether the cells were shipped to NBS under constant temperature control and may be as long as 4 weeks. If space is available on the measuring system, and cells have been shipped under temperature control, they are generally connected to the measuring system immediately. However, later review of the data may result in these early measurements being discarded.

Three types of saturated standard cells are calibrated: cells in oil at 28 °C or 30 °C and groups of cells in temperature-controlled enclosures. The first two types are placed in oil baths whose temperature is stable and uniform to at least 0.001 °C and whose temperatures are determined using NBS-owned platinum resistance thermometers. Temperature-controlled standard cell enclosures are tested under the following ambient conditions:

Temperature	(23 ± 1) °C
Relative Humidity	50% or less

The operating temperature of the cells in temperature-regulated enclosures is determined using the temperature measuring device supplied with the enclosure. If the enclosure has a temperature indicating bridge, an NBS owned null detector is used to make the readings to within the resolution of the bridge, usually 0.001 °C. An NBS or customer owned platinum resistance thermometer may be used if requested. Temperature measurements are made each day before the cell emf measurements are started.

Assignment of the Final Value

Client cells are compared to NBS working cells for a minimum of one measurement design per day for ten working days. The cell emfs are plotted and the data is reviewed by two staff members who decide if the cell emfs exhibit sufficient stability for a report to be issued. If so, the means of the ten measurements are used as the final assigned values of emf of the cells. If the cells show excessive drift or other unusual behavior, additional measurements may be taken.

TABLE III

UNCERTAINTIES IN THE ASSIGNMENT OF THE MEAN EMF OF A CLIENT CELL

Source of uncertainty	1 std. dev. estimate (ppm)
Uncertainty of assignment to working group	0.040
Change in working group during the day	0.040
DVM scale error	0.010
Random uncertainty of comparison of working group to Client cell	0.007
Uncompensated thermal emfs in the cell switches	0.030
Uncertainty of assignment to Client cell (not including temperature measurement errors; see text)	
RSS total	**0.065**

Table III summarizes the sources of uncertainty in calibrating a client cell in terms of the U.S. Legal Volt.

The DVM scale error is calculated as described for Table I, except cell differences as large as 120 μV may be measured. The random measurement uncertainty of the working group - client cell comparison is estimated from the redundant measurement design as before.

No allowance for temperature has been included in the above uncertainties. Fluctuations in cell emfs due to inexact measurement and correction of temperature changes of the primary and working groups are included in the day-to-day fluctuation of the primary groups, and the within-day fluctuation of the working groups. The uncertainty of the cell emfs due to the imprecision of the temperature monitoring device is estimated by calculating the change in cell emf (according to the International Temperature formula) for a change of one least count of the temperature monitoring device. In the case of cells in enclosures containing internal thermistor bridges, where a least count of 0.001 °C is possible, an additional uncertainty of approximately 0.05 ppm is

included. Enclosures monitored by mercury-in-glass thermometers are assigned an uncertainty equivalent to half the smallest graduation marked on the thermometer, approximately 0.5 ppm for a resolution of 0.01 °C. For cells calibrated in NBS oil baths an emf uncertainty equivalent to a temperature uncertainty of 0.005 °C is used to account for possible calibration errors in the platinum resistance thermometers used at NBS.

The day-to-day random component of the uncertainty of the client cell emf is determined by calculating the standard deviation of the measured emfs of the cell from the ten daily measurement designs and comparing it, at the 99% confidence interval (CI) using an F-test, with the pooled standard deviation of a large population of measurements of similar standards (0.135 - 0.188 μV, depending on the type of enclosure). If the statistic is determined to belong to that population, then the population standard deviation of the mean is used as the estimate of the limit of the random component. If not, the computed standard deviation of the mean is used as the estimate.

The final reported uncertainty is the direct sum of this random component, the total assignment uncertainty from Table III, and the temperature uncertainty, multiplied by a factor of three.

The calibration service has recently been extensively modified, with completely new automated measuring systems and new quality control procedures. The values quoted in the tables above are estimates of the uncertainties of the new system. Until a complete evaluation is completed, the following (previously determined) uncertainties are being used. The client cell assignment uncertainty is taken as 0.075 ppm, the pooled standard deviation for random error is 0.135 μV, and the temperature uncertainty is the same as described.

The reported uncertainty contains no allowance for long term drift of the cells under test. Long term behavior must be determined by the client by analysis of the history of each individual standard. In addition, no allowances are made for the possible effects of transporting the standard between laboratories or the possible existence of a gross temperature dependence on ambient (room) temperature of a standard cell enclosure.

Volt Transfer Program

The Volt Transfer Program involves sending an NBS owned transport voltage standard to the client laboratory. The transport cells are compared to NBS working standards at NBS before and after shipment to the client laboratory, and to the client reference group of standard cells while at the client laboratory. The process is described in more detail below.

Measurement of the Transport Standard at NBS

The transport standard consists of a commercial thermoregulated enclosure usually containing 4 saturated shippable standard cells, operating at an internal temperature of 30 °C or 32 °C. The temperature is measured using the internal temperature bridge and an external null detector to provide 0.001°C resolution. Corrections to the cell emfs are applied according to the discussion given in Section 3. Transport standard cells are compared daily to Working Group A using redundant measurement designs and the VTP measurement system (Fig. 7). Before shipment to a client laboratory, the cell emfs are plotted and examined for stability. In any case, at least 15 daily measurements are required after the enclosure's return to NBS and before it is shipped to another laboratory.

Upon receipt of a purchase order NBS schedules the shipment of a transport standard to the client laboratory. The transport standard is shipped with an external power supply containing a battery to power the standard during shipping, and in a special container designed to provide physical shock protection and temperature lagging. The battery supply can maintain temperature control of the standard for about 24 hours during normal ambient shipping conditions. Since the temperature control system cannot supply cooling to the enclosure, shipment to warmer locations is avoided during extremely hot weather to prevent the enclosure from overheating. Shipment is normally via air freight with special delivery service from NBS to the airport just in time for the scheduled flight. The delivery service may also be able to arrange, with an affiliated service, for pickup and delivery of the enclosure to the client laboratory. If not, client laboratory personnel are notified of the time of arrival of the standard and are expected to provide transportation from the airport to their laboratory within the 24 hour lifetime of the batteries.

Comparison of the NBS Transport Standard to the Client Standard

Each laboratory participating in the Volt Transfer Program must identify a group of saturated standard cells that are considered to be the "laboratory reference group" and constitute the "laboratory volt". The transport standard cells are compared to the cells of the laboratory reference using redundant measurement designs similar to the ones used at NBS. Each laboratory uses its own measuring equipment in the manner it would normally use to calibrate saturated cells, except for the requirement that an NBS-specified measurement design be used. NBS provides data sheets to record the measurements which are returned to NBS where the cell comparison observations are reduced to determine the cell emfs in terms of the client laboratory volt. Laboratories are required to make daily design measurements and the data is reviewed by NBS, as it is taken, to determine when sufficient data has been obtained to permit shipment of the standard back to NBS. A minimum of 10 measurement designs over 10 days is required for the transfer; however, the number is more typically in the range from twelve to twenty. In addition, the within-day standard deviation and left-right components from the measurement designs are checked

to see if they are statistically in control with respect to the expected values determined from measurements at NBS and other laboratories.

Upon its return to NBS, the standard is again compared to Working Group A as before and, when the standard has stabilized and sufficient data has been taken, a final report is issued. Typical measurement data is shown in Fig. 9 where the mean emf of a transport standard over the period of one transfer is plotted. The emf in the figure is expressed in microvolts with 1.01 volts subtracted from the total cell emf.

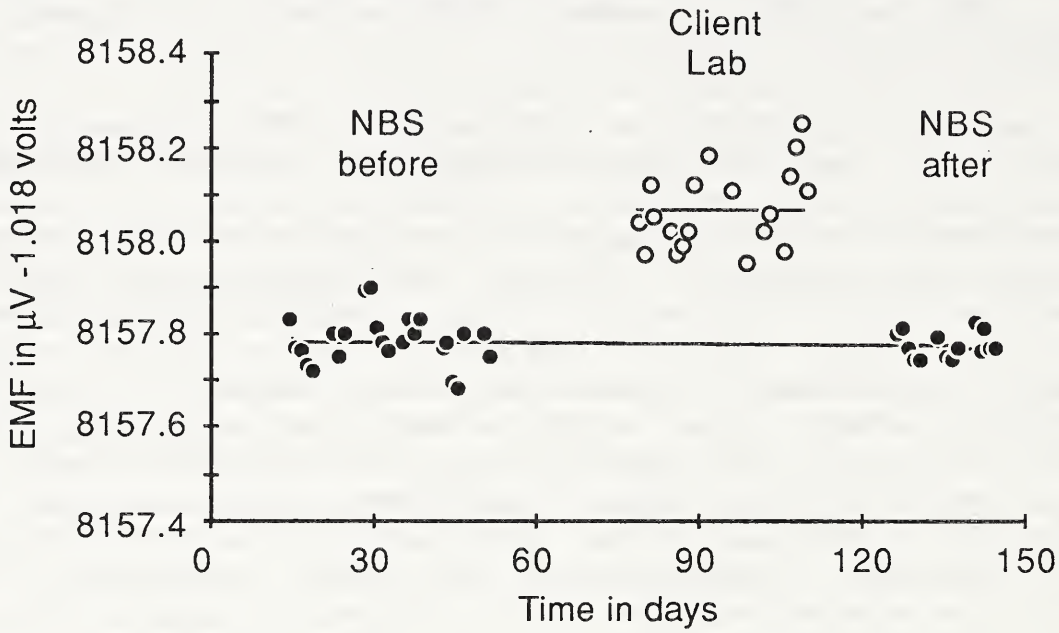


Fig. 9. Mean emf of a typical transport standard as measured at NBS and a client laboratory.

Assignment of Final Value

Emf assignments at NBS and the client laboratory for each cell of the transport standard are plotted as in Fig. 9 above. Data showing cell emf changes due to shipment or other effects are removed from the analysis; often the first measurement after shipment is anomalous. The data are also reviewed for abnormal temperature readings and to determine if the data reasonably conform to a linear model. As the before and after NBS data do not usually exactly agree, a least-squares line is fitted to the NBS data for each transport cell, and values are predicted for each of the times the cell was compared to the client laboratory cells. V_{LAB}-V_{NBS} is determined using each cell in the transport enclosure by subtracting the NBS predicted values of the transport cell from the respective client laboratory emf assignments, and taking the mean of these differences.

Although the difference in the units as determined from each of the four cells in the transport standard should agree, on occasion some cells show abnormal behavior, usually excessive drift, or excessively long recovery from physical shock or electrical disturbance. The judgement and experience of the laboratory staff are used to determine when to exclude a cell from the analysis or, on rare occasions, use only the NBS data taken before or after the transfer.

TABLE IV

UNCERTAINTIES IN THE DETERMINATION OF $V_{LAB} - V_{NBS}$

Source of uncertainty	1 std. dev. estimate (ppm)
Uncertainty of NBS assignment to transport group	0.061
Correlated temperature effects of the transport	0.050
Temperature resolution	0.050
Random component due to individual cell assignments and changes during transport	0.093
Uncertainty of $V_{LAB} - V_{NBS}$ determination using a four cell transport enclosure	
RSS total	**0.132**

Table IV summarizes the sources of uncertainty in the determination of V_{LAB}-V_{NBS}.

The uncertainty in assigning a value to the transport standard while at NBS is obtained through an analysis identical to Table III, except the switches used for the VTP transport standards have somewhat lower uncompensated thermal emfs (0.020 μV as opposed to the 0.030 μV shown in Table III).

The transport standards used for the Volt Transfer Program all contain thermistor bridges with 0.001 °C resolution; an uncertainty for the cell emf equivalent to this temperature change is included. In addition, the cell temperature within the enclosures changes slightly with changes in room

temperature. The thermistor bridge does not properly reflect the change of the temperature of the cell in this case, probably due to a temperature sensitive component of the bridge circuitry that is at room temperature. An uncertainty estimate of 0.05 ppm is included to account for a difference between the client laboratory ambient temperature and the NBS laboratory ambient temperature.

Each of the four cells in the transport standard is used to determine a value for the difference in laboratory units, $V_{LAB} - V_{NBS}$. The standard deviation of these four values includes an uncertainty in predicting the cell emfs while at the client laboratory, an uncertainty in comparing the client laboratory reference to the transport standard, and an uncertainty for random changes in the cell emfs due to shipment. A pooled standard deviation of 0.19 μV has been computed from 50 transfers and is used as the population standard deviation. The calculated standard deviation for each new transfer is compared to the population standard deviation at the 99% confidence interval using an F-test. If the statistic is determined to belong to that population, then the population standard deviation of the mean is used as the uncertainty estimate; if not, the actual standard deviation of the mean is used. Thus the uncertainty estimate for a four cell transport standard, based on the population standard deviation, is (0.19 μV)/2 or 0.095 μV (0.093 ppm).

If the difference $V_{LAB} - V_{NBS}$ exceeds 0.20 ppm (1.5 times the uncertainty based on the population standard deviation), the report will recommend adjusting the assigned values of the client reference cell emfs to reduce the difference to zero. Adjusted values for the cell emfs are calculated based the measurements made in the client laboratory comparing the client reference group to the NBS transport standard. Each comparison results in a determination of the difference of each client reference cell emf from the mean emf of the client reference group. The average difference for each cell, over all the comparisons, is added to the newly determined mean emf of the reference group to calculate the new values for each of the client reference cells.

6. Quality Control Procedures

Measurement System Verification

Potential sources of error in the cell measurement systems include scaling or gain errors in the digital voltmeters, uncompensated thermal emfs in the crossbar switches and cell leads, leakage currents to ground from the measurement apparatus, and circulating ground currents. Several quality control procedures are periodically performed to estimate or eliminate these uncertainties.

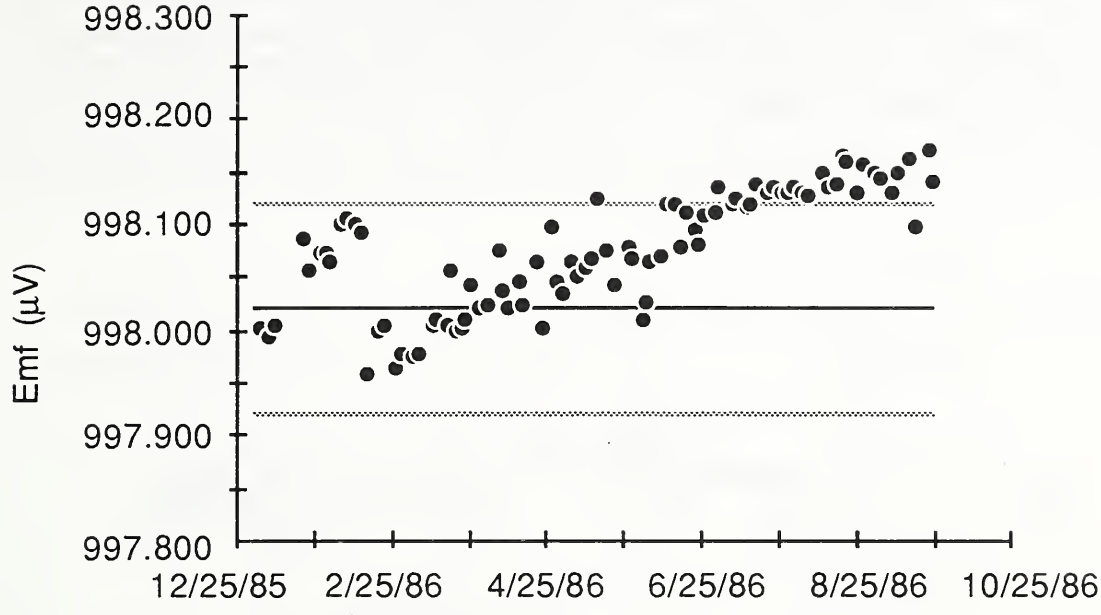


Fig. 10. Daily DVM readings of the 1000 μV Zener source.

Before each daily set of cell comparisons the measurement system reads the output of a 1000 μV Zener source to monitor the gain error of the 0.1 volt range of the digital voltmeter. All voltmeter measurements are taken with the applied voltage in the normal polarity and again with the polarity reversed by the crossbar switch to eliminate any uncertainty due to zero offset of the voltmeter and thermal emfs in the leads to the voltmeter. The daily measured values of the 1000 μV source are plotted on a control chart and compared to predetermined limits. Figure 10 is a plot of the voltage of the Zener source that is connected to measuring system REG. The control limits correspond to a worst case cell comparison uncertainty of approximately 0.01 ppm. If the limits are exceeded, the reason for the out-of-tolerance condition is investigated by laboratory personnel. Usually the problem is found to be that the voltage of the 1000 μV Zener source has drifted (as it has in Fig. 10), and the source is recalibrated.

After each measurement design, the computer controlling the measurement system reduces the data using a least-squares technique and computes the standard deviation for a single measurement for that design. This standard deviation is compared to a control limit based on the population standard deviation to determine if it is in statistical control. If it is out of control the measurement design may be repeated. Measuring system VTP will repeat the measurement design once if the standard deviation is out of control. If the standard deviation of the second design is also out of control the system does not repeat it again. Measuring system REG does not repeat the design under any

circumstances. The standard deviation is a function of the individual cell group being measured, and control limits for newly arrived client cell groups are not well known. Each day, a pooled standard deviation from all the measurement designs that are in control is calculated and plotted on a control chart for each of the measurement systems (REG and VTP). Figure 11 is the control chart for measurement system REG.

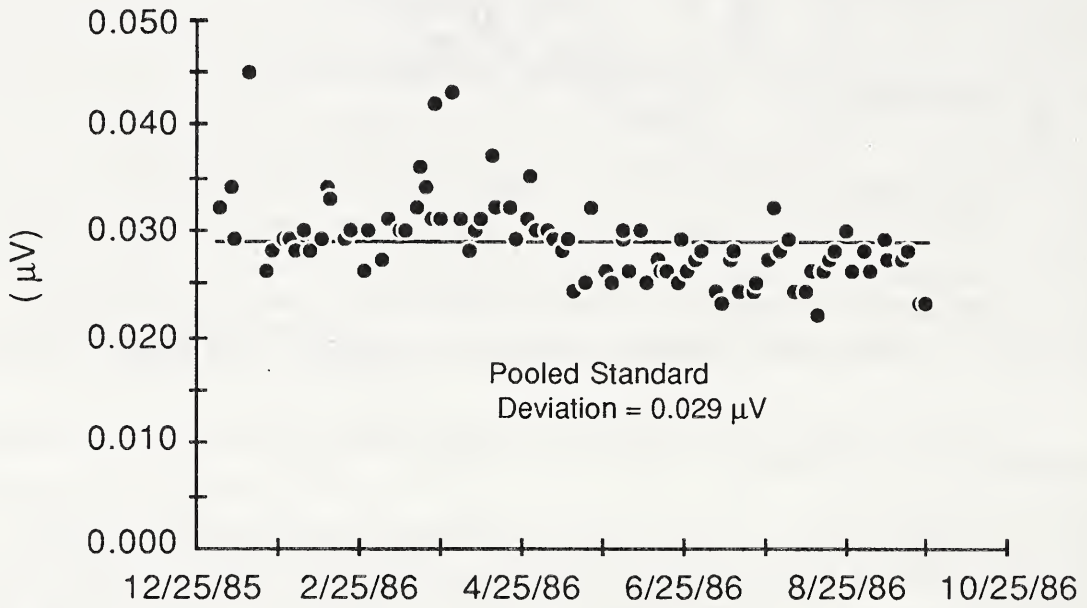


Fig. 11. Within-day standard deviations calculated from the redundant measurement designs.

Thermal emfs in the leads from the cell enclosures to the crossbar switch are evaluated regularly. The positive and negative leads are shorted together at the end where they would normally connect to the cells and a redundant measurement design is made between two sets of shorted leads. These thermal emf measurements are performed frequently, usually every two months or so as cell enclosures leave the laboratory and leads become free. Experience has shown that the thermal emfs are usually small and fairly constant (approximately 15 nV and 30 nV for measuring systems VTP and REG respectively).

Both measurement systems print out a daily log of the measurements performed by each system. Figure 12 is a typical example produced by the VTP measurement system. This log contains the measured value of the 1000 μV source, and a list of the groups of cells that were measured. The time of the measurement, the reference group used, the measured temperature of the group, the change in temperature correction, the change in the group mean emf, the standard deviation of the redundant design, and the measured DVM

offset voltage are printed for each cell group measured. The change in group temperature correction and mean emf are calculated by subtracting the values obtained for the previous day's readings from the current day's readings. If any of the values exceed predefined limits the measurements are flagged on the log and lab personnel investigate the problem. The limits are set at 0.1 μV for both the change in temperature correction and mean emf, and are based on laboratory personnel experience. The exact values are not overly important since they serve only to alert the operator to a potential problem.

VTP MEASUREMENT LOG FOR: 29 May 1986 [1.4] WEEK: 8622

DVM check Zener measurement = 1000.014 (+/- 0.005) uV

Time	STD-ID	UNK-ID	UNK Temp (C)	Change T-corr (uV)	Change Mean (uV)	SD (uV)	DVM Offset (uV)	
10:20	18	2800	30.0070	-.113	.084	.020	-.251	T
10:31	18	19	29.9790	-.034	.028	.009	-.274	
10:42	18	2000	31.9965	-.030	-.000	.021	-.299	
10:53	18	1400	30.0080	-.038	.002	.029	-.271	
11:05	18	2100	32.0055	-.059	.007	.027	-.300	
		MEAN of daily runs:		-.053	.024		-.279	
		Standard deviation:		.036	.035	.022	.021	

Some runs were marked for exceeding predefined limits.
The letters in the right hand column mean:

T - the temperature correction changed by more than .1 uV
M - the mean of the unknown group changed by more than .1 uV
S - the std. dev. exceeded 1.4 time the accepted std. dev.
D - the magnitude of the dvm offset exceeded 2 uV

ALL CALCULATIONS COMPLETED AND SRM UPDATED

Fig. 12. Typical daily log printout from the VTP measuring system.

As discussed earlier, two working groups of cells are calibrated daily in terms of the primary groups. Cell emf values are predicted for these working groups at the same time predictions are made for the primary cell groups, i.e., once every 4 weeks when 2e/h measurements are made. Occasional problems with temperature regulated enclosures, oil baths, and unstable cells have made long term predictions of either the primary or working groups meaningless. However values for up to one month ahead are predicted for the primary and

working groups based on four or more months of data. Each day the values of the working groups are calculated based on the daily intercomparisons of the primary groups against the working groups and the predicted values of the primary groups. The measurements or measurement systems are scrutinized whenever there is significant (0.08 ppm) disagreement between the directly predicted values of the working groups and the values assigned through the primary groups. Figure 13 is a plot of the difference between the daily values assigned to Working Group A and the predicted values.

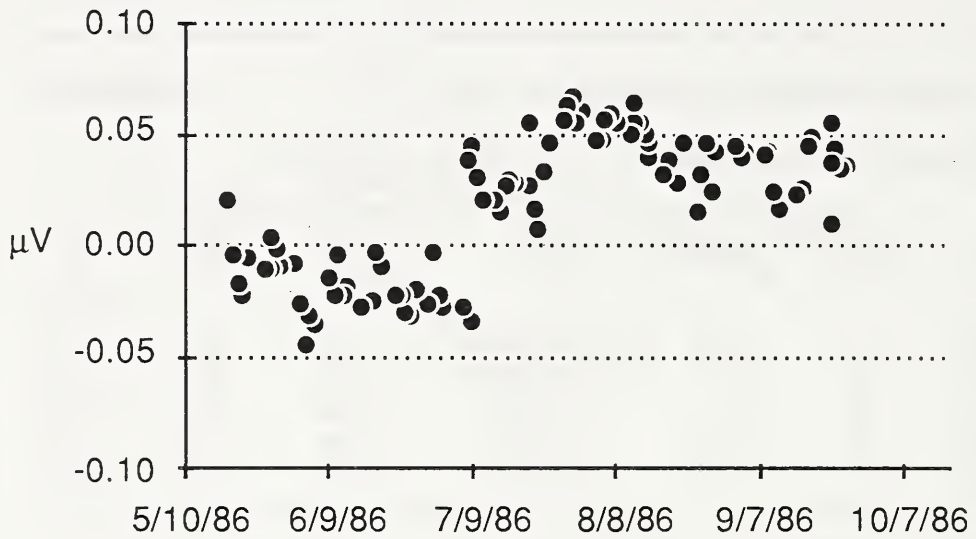


Fig. 13. The difference between the daily value assigned to the mean emf of Working Group A from the two Primary groups, and from the predicted value of the Working Group.

An enclosure that can be used exclusively as a long term check standard is not available because of a shortage of good quality enclosures and cells. Instead, the average of the changes in emfs of the workload from the previous day's measurements is calculated daily and noted by laboratory personnel.

Occasional additional tests are performed on the automated systems. Selected cell enclosures are calibrated with both systems to ensure that their results agree, and the automated systems are compared to a calibrated millivolt potentiometer . Measurements of the insulation/leakage resistance of the measuring systems are done occasionally and recorded in laboratory notebooks. Closure experiments are done monthly (at least) to detect systematic errors due to leakage resistance. Closure experiments consist of redundantly comparing three cell enclosures with the pattern, A-B, B-C, and C-A. Summing the three mean emf differences should yield a value of zero; the disagreement of the actual value is an indication of measurement error. Using twenty-four such experiments performed on different enclosures between July 17, 1986 and

October 1, 1986, the mean closure error was -0.0009 μV with a standard deviation of the mean of 0.0017 μV.

Measurement Assurance for Client Standards

Standards shipped to NBS in temperature regulated enclosures under power are hand carried from the Building 301 Receiving Room as soon as Electricity Division personnel are notified. The temperature is checked by qualified laboratory technicians and the enclosure is brought to the Volt Facility Laboratory. Laboratory personnel within the Volt Facility connect and disconnect the standards to the measuring systems to minimize the possibility of incorrect connections.

After stabilization and measurement of the standard, the data are plotted for each standard cell emf in the enclosure. These data are reviewed by laboratory personnel and anomalous data usually related to stabilization are removed. The remaining data are used to generate a calibration report. The final calibration report and intermediate emf plots are also reviewed by the leader of the Electricity Division Dissemination Services Group.

7. Bibliography

1. ASTM Manual on Quality Control of Materials, Special Technical Publication 15-C, Part 3 (Jan., 1951).

2. Harris, F. K., Electrical Measurements, J. Wiley & Sons, pp. 168-185 (1952).

3. Field, B. F., Eicke, W. G, and Kleimann, R. E., unpublished computer programs for reducing standard cell data.

4. Linnik, J. W., *Method of Least Squares and Principles of the Theory of Observations*, Translated from the Russian by Regina C. Elandt, Pergamon Press, New York (1961).

5. Brodd, R. J. and Eicke, W. G., "Effect of vibration and shock on unsaturated standard cells," *J. Res. Natl. Bur. Stand.* (U.S.), Vol. 66C (Eng. and Instr.), No. 2, pp. 85-97 (Apr.-June 1962).

6. Finch, D. I., "Temperature: Its Measurement and Control in Science and Industry," *General Principles of Thermoelectric Thermometry*, Vol. 3, Part 2, Reinhold Publishing Corporation, New York (1962).

7. Eicke, W. G., "Comments on Zener diodes as voltage standards," *Comité International des Poids et Mesures, Comité Consultatif d'Electricité,* 10th Session, p. 27 (1963).

8. Eicke, W. G., "Making precision voltage measurement on Zener diodes," *IEEE Trans.*, Paper CP 63-416 (1963).

9. Natrella, M. G., Experimental Statistics, NBS Handbook 91, pp. 18-1, 18-4 (1963).

10. Dunfee, B. L., "Method for calibrating a standard volt box," *J. Res. Natl. Bur. Stand. (U.S.)*, Vol. 67C (Engr. and Instr.), No. 1, pp. 1-13 (Jan.-Mar. 1963).

11. Ramaley, D., "Practical method for calibration of potentiometers," *Natl. Bur. Stand. (U.S.) Tech. Note 172*, 44 pages (1963).

12. Eicke, W. G, "The Operating Characteristics of Zener Reference Diodes and their Measurement," *ISA Transactions*, Vol. 3, No. 2, pp. 93-99 (1964).

13. Hamer, W. J., "Standard cells, their construction, maintenance, and characteristics," *Natl. Bur. Stand. (U.S.) Monogr. 84*, 38 pages (1965).

14. Pontius, P. E., "Measurement Philosophy of the Pilot Program for Mass Measurement," *Natl. Bur. Stand. (U.S.) Tech. Note 288* (1966).

15. Hamer, W. J., "The Volt Standard Moves To Gaithersburg, Maryland," *J. Wash. Acad. of Sci.*, Vol. 56, pp. 101-108 (1966).

16. Eicke, W. G. and Cameron, J. M., "Designs for Intercomparing Small Groups of Saturated Standard Cells," *Natl. Bur. Stand. (U.S.) Tech. Note 430*, 19 pages (1967).

17. Volt Transfer Program Instructions, Electricity Division, Natl. Bur. Stand. (U.S.) (1968).

18. Belecki, N. B., "Transfer of the unit of voltage," *Proc. ISA Ann. Conf.*, Vol. 5, p. 608 (1968).

19. Belecki, N. B., "A Guarded Shielded Standard Cell Measuring System," *ISA Preprint 706-70* (1970).

20. Dziuba, R. F. and Dunfee, B. L., "A resistive voltage-ratio standard and measuring circuit," *IEEE Trans. Instrum. Meas.*, Vol. IM-19, No. 4, pp 266-277 (1970).

21. "NBS offers new calibration service for the volt," *Natl. Bur. Stand., NBS Special Publ. 250*, Users Bull., No. 1 (1971).

22. Eicke, W. G., "The Volt, Yesterday, Today and Tomorrow," *ISA Reprint 71-703* (1971).

23. Finnegan, T. F., Denenstein, A., and Langenberg, D. N., "ac Josephson Effect Determination of e/h: A Standard of Electrochemical Potential Based on Macroscopic Quantum Phase Coherence in Superconductors," *Phys. Rev. B*, Vol. 4, No. 5, pp. 1487-1522 (1971).

24. Eicke, W. G., and Taylor, B. N., "Summary of international comparison of as-maintained units of voltage and values of 2e/h," *IEEE Trans. Instrum. Meas.*, Vol. IM-21, No. 4., pp. 316-319 (1972).

25. Popenoe, C. H., and Campbell, M. S., "MIDAS: Modular interactive data acquisition system - description and specification," *Natl. Bur. Stand. (U.S.) Tech. Note 790* (1973).

26. Field, B. F., Finnegan, T. F., and Toots, J., "Volt maintenance at NBS via 2e/h: A new definition of the NBS volt," *Metrologia*, Vol. 8, pp. 155-166 (1973).

27. Braudaway, D. W. and Kleimann, R. E., "A high-resolution prototype system for automatic measurement of standard cell voltages," *IEEE Trans. Instrum. Meas.*, Vol. IM-23, No. 4, pp. 282-286 (1974).

28. Eicke, W. G., and Auxier, L. M., "Regional maintenance of the volt using NBS volt transfer techniques," *IEEE Trans. Instrum. Meas.*, Vol. IM-23, No. 4, pp. 290-294 (1974).

29. Cutkosky, R. D. and Field, B. F., "Standard cell enclosure with 20-uK stability," *IEEE Trans. Instrum. Meas.*, Vol. IM-23, No. 4, pp. 295-298 (1974).

30. Eicke, W. G. and Kleimann, R. E., "An automated standard cell comparator controlled by a desk calculator," *NBS Preliminary Report* (1975).

31. Field, B. F. and Hesterman, V. W., "Josephson junction based voltage standard," *Proc. 5th Cal Poly Meas. Sci. Conf.* (1975).

32. Field, B. F. and Hesterman, V. W., "Laboratory voltage standard based on 2e/h," *IEEE Trans. Instrum. Meas.*, Vol. IM-25, No. 4, pp. 509-511 (1976).

33. Eicke, W. G., Leedy, T., et al., "Measurement assurance programs in a field environment," *Natl. Conf. of Stand. Lab. Newsletter*, Vol. 23, No. 1, pp. 51-55 (1983).

34. "Supplementary Instructions for Group Volt Transfer Programs," Electricity Division, Natl. Bur. Stand. (U.S.) (1984).

35. Huntley, L., "The Fluke direct voltage maintenance program," *Proc. 1984 Meas. Sci. Conf.*, pp. 147-151 (1984).

36. Koep, K. J., "The use of a solid state dc voltage transport standard to transfer the dc volt in a regional standard cell MAP program," *Proc. 1984 Meas. Sci. Conf.*, pp. 142-146 (1984).

37. Murray, W. and Dencher, P., "High-accuracy voltage transfer using a modified commercial solid-state reference," *J. Phys. E.: Sci. Instrum.*, Vol. 17, No. 5, pp. 354-356 (1984).

38. Spreadbury, P. J. and Everhart, T. E., "Ultra-stable portable voltage sources," *IEE Conf. Publ.*, Vol. 174, pp. 117-120.

39. Field, B. F., "A sub-ppm automated one-to-ten volt measuring system," *IEEE Trans. Instrum. Meas.*, Vol. IM-34, No. 2, pp. 327-330 (1985).

40. VTP Control Chart Requirements, Electricity Division, Natl. Bur. Stand. (U.S.) (1984).

41. Taylor, B. N., "Possible Changes in the U.S. Legal Units of Voltage and Resistance," *J. Res. Natl. Bur. Stand.* (U.S.), Vol. 91, No. 5, pp. 299-305 (Sep.-Oct. 1986).

42. Hamilton, C. A., Kautz, R. L., and Lloyd, F. L., "A practical Josephson voltage standard at one volt," *IEEE Elec. Dev. Lett.*, EDL-6, pp. 623-625.

43. Hamilton, C. A., Kautz, R. L., Lloyd, F. L., Steiner, R. L., and Field, B. F., "The NBS Josephson Array Voltage Standard," *IEEE Trans. Instrum. Meas.*, Vol. IM-36, No. 2, pp. 258-261 (1987).

Appendix A - Redundant Measurement Designs

Redundant measurement designs are used for all standard cell comparisons as they are efficient in providing the maximum amount of information per measurement and provide information about possible systematic effects in the measuring system. The designs are used to provide estimates of:

1) the cell emfs with respect to a group mean emf,
2) the within-day standard deviation of a single observation,
3) the left-right (positional) effect,
4) the standard deviations of the cell emfs,
5) and the deviations of individual observations from the predicted values.

In general, when comparing two or more cell groups, a "full" design is used where all possible pair differences that involve two cells from different groups are measured. The 4x6 design and 4x4 design below illustrate this process. For intercomparison of cells within a single group the appropriate design from NBS Technical Note 430 is used [16].

On occasion an anomalous reading may be discovered during the data reduction of the cell comparison data. Since some redundancy is provided by the design it is possible to remove one or two observations and recalculate with little loss in accuracy in determining the cell emfs with respect to the group mean.

For the examples below, the cells in the groups are designated R1 - R4 (R for reference) and X1 - X4 or X1 - X6 (X for unknown) depending on the design. The voltage difference measured is "left cell" - "right cell".

4x4 Design - 16 observations

Observation number	Left cell	Right cell	Observation number	Left cell	Right cell
1	R1	X1	9	X3	R4
2	R1	X3	10	X1	R4
3	R2	X2	11	X4	R3
4	R2	X4	12	X2	R3
5	R3	X1	13	X3	R2
6	R3	X3	14	X1	R2
7	R4	X2	15	X4	R1
8	R4	X4	16	X2	R1

4x6 Design - 24 observations

Observation number	Left cell	Right cell	Observation number	Left cell	Right cell
1	R1	X1	13	X5	R4
2	R1	X3	14	X3	R4
3	R1	X5	15	X1	R4
4	R2	X2	16	X6	R3
5	R2	X4	17	X4	R3
6	R2	X6	18	X2	R3
7	R3	X1	19	X5	R2
8	R3	X3	20	X3	R2
9	R3	X5	21	X1	R2
10	R4	X2	22	X6	R1
11	R4	X4	23	X4	R1
12	R4	X6	24	X2	R1

U. S. Department of Commerce
NATIONAL BUREAU OF STANDARDS
National Measurement Laboratory
Gaithersburg, MD 20899

REPORT OF CALIBRATION
DC Voltage Standard

Description of Standard:
Standard Cell Enclosure
Model XXXXXXXXXXXXXXXX
Containing 4 Saturated Standard Cells

Submitted By:
XXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXX

This standard cell enclosure was received 21 Mar 1986, not under power with an internal temperature far below its normal operating temperature.

The data from which the values in the table below are computed are the result of ten daily measurements of the differences between the EMF's of the cells under test and those of NBS working standards during the period between 5 May 1986 and 16 May 1986.

POSITION NUMBER	EMF (volts)	UNCERTAINTY (microvolts)	EMF (volts)	UNCERTAINTY (microvolts)
1	1.0181270	0.27	1.0181269	0.27
2	1.0181271	0.27	1.0181270	0.27
3	1.0181272	0.27	1.0181271	0.27
4	1.0181268	0.27	1.0181267	0.27

The electromotive forces on the left above were corrected to nominal temperature (30.0 degrees Celsius) using the International Temperature formula proposed by F.A. Wolff. The electromotive forces on the right are at the mean operating temperature during the test (30.0019 degrees C) as determined by use of a temperature deviation measuring device mounted in the enclosure.

The above uncertainties include components for random fluctuations in the cell under test and in NBS equipment and standards, for a systematic error of 0.076 ppm in the measurements of NBS working standards in terms of the Josephson effect, and for the systematic error in transfer due to the finite resolution of the apparatus used to determine the temperature of the cells under test. In the case of standard cells tested in NBS oil baths, the latter uncertainty is replaced by the EMF equivalent of 0.005 degrees Celsius which is the uncertainty of the temperature measurements in this laboratory. The random error component is computed from the standard deviation of the mean EMF and is at the three sigma level.

These uncertainty figures contain no allowance for the effects of transportation upon this standard. The minimum uncertainty due to such effects under very carefully controlled transport conditions has been seen to be 0.42 ppm (3 sigma). Any valid uncertainty statement applying to the above values when the standard has been moved from the NBS Volt Facility must contain such a component of error. If data from which to estimate the transport error are not available, one part-per-million is not an unreasonable value. Also not included in the above uncertainties is an allowance for long term drift of the values of the outputs of this standard. This must be determined from historical data on a case by case basis.

A summary and analysis of the data upon which the above values are based is appended. A complete explanation of the uncertainty statements given above as well as additional information regarding NBS calibration of such voltage standards is also included.

For the Director
National Measurement Laboratory

Norman B. Belecki, Physicist
Center for Basic Standards
Electricity Division

Test No. XXXXXXXX
Date: 21 May 1986

U. S. Department of Commerce
NATIONAL BUREAU OF STANDARDS
National Measurement Laboratory
Gaithersburg, MD 20899

REPORT OF CALIBRATION
DC Voltage Standard

Description of Standard:
Unsaturated Standard Cell
Model No. XXXXXXXXXXXX
Serial No. XXXXXXXXXXXX

Submitted By:
XXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXX

The electromotive force of this cell at 22.4 degrees Celsius was, at the time of test, 1.01923 volts. This value, correct to 0.005 percent, is the mean of a series of measurements concluded 24 Apr 1986. The stated uncertainty (0.005%) includes allowance of plus or minus 50 microvolts for variability in the EMF of the cell during test.

This is an unsaturated cell of the cadmium sulfate type, suitable for work requiring no greater accuracy than 0.005 percent. Such cells have a temperature coefficient that is negligible within the ordinary range of room temperature. Rapid changes in temperature may, however, produce temporary alterations of several hundredths of one percent in the electromotive force.

Precautions in using standard cells: (1) the cell should not be exposed to temperatures below 4 degrees Celsius, (2) abrupt changes in temperature should be avoided, (3) all parts of the cell should be at the same temperature, (4) current in excess of 0.0001 ampere should never pass through the cell, (5) unsaturated cells should be recalibrated at intervals of a year or two because the electromotive force of an unsaturated cell usually decreases with time.

For the Director
National Measurement Laboratory

Norman B. Belecki, Physicist
Center for Basic Standards
Electricity Division

Test No. XXXXXXXX
Date: 1 May 1986

U. S. Department of Commerce
N A T I O N A L B U R E A U O F S T A N D A R D S
National Measurement Laboratory
Gaithersburg, MD 20899

R E P O R T O F T E S T

Determination of $V_{TSCC} - V_{NBS}$

Laboratory: (LAB)
XXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXX
Reference No. XXXXXXXXXX

Reference Cells:
5403 5404 5402
5406

The difference between the unit of electromotive force maintained by the above laboratory and the U.S. Legal Volt, as maintained by the National Bureau of Standards (NBS), was experimentally determined during the period between November 11, 1985 and February 28, 1986. The results, given in the equation below, are based on the previously-assigned mean emf of the reference cells listed above.

$$V_{LAB} - V_{NBS} = 0.15 \times 10^{-6} V_{NBS}$$

The above number is based upon the experiment described in Appendix A and the data summarized in Appendix B of this report.

The uncertainty of this difference, $V_{LAB} - V_{NBS}$, is 0.15 microvolts at the 99% confidence level. It was determined by combining in quadrature random error components resulting from variations among the cells in the transport standard due to the transfer, day-to-day fluctuations in the results of NBS measurements of the transport standard, and day-to-day variations in the difference between the results of the client's measurements and the corresponding NBS-predicted values of the emf's of the cells in the transport standard. As far as is presently know, systematic effects are negligible. The pertinent data and their reduction are given in Appendix B, along with plots of all data taken both at NBS and in the client laboratory.

Test No. XXXXXXXX

This uncertainty, the difference cited above, and the results of any adjustments made to the values of the reference cells whose numbers are given above are valid only under conditions approximating those under which this experiment was performed. In particular, the operating temperature of the reference cells must be within 0.05 degree C of that measured during the course of the experiment. This condition must be met in order that errors in the algorithm used to compute temperature corrections not affect the experimental uncertainty.

On July 1, 1972, NBS changed from using saturated standard cells to maintain the U.S. Legal Volt to the use of the ac Josephson effect. Based upon data taken to date, it is estimated that the U.S. Legal Volt is constant, with respect to time to at least 0.1 ppm. This figure is not included in the above uncertainty. As further data are obtained, the uncertainty for this type of experiment will be refined to take into account the improved method for maintaining the volt.

If the measured difference between the unit of voltage as disseminated by this laboratory and the U.S. Legal Unit of voltage is greater in magnitude than the characteristic standard deviation of a measurement of this type (0.20 ppm), it is recommended that the assigned mean value of the laboratory's reference group be adjusted so that $V_{LAB} - V_{NBS} = 0$. Appendix C has be included in this report to facilitate recommended adjustments.

For the Director
National Measurement Laboratory

Norman B. Belecki, Physicist
Center for Basic Standards
Electricity Division

Test No. XXXXXXXX
Date: May 14, 1986

Appendix D - SOFTWARE

All software for controlling the measurement systems, reducing the cell data, and producing test reports, was written in-house specifically for the intended application. At present the software is in transition; a minicomputer used to do data reduction is in the process of being replaced by a number of dedicated desktop computers. Software has been completed for these computers to make standard cell comparisons and calculate the cell emfs from the cell difference observations. These data are stored on a shared hard disk. Existing software for the minicomputer to maintain a data base of cell emfs of NBS and client standards, produce control charts, and produce test reports is currently being converted to run on a new, more modern, minicomputer. This is only an interim step since these programs need to be substantially rewritten to provide better quality control charts and easier access to the data base.

The following programs are complete and operate on the desktop computers.

READBOX - This program intercompares standard cell voltages using a redundant measurement design, reduces the data to determine the cell emfs, and saves the data to disk. The actual voltage differences are measured with measuring system VTP, using the digital voltmeter, with the cells connected to the voltmeter by a crossbar switch. The data collected by the program is stored on a floppy disk in drive 0 and optionally on the Shared Resource Management (SRM) hard disk. The files on the floppy disk are designed to hold approximately one week's worth of data. A printed log is generated for each day's measurements listing the boxes compared, the standard deviation of the measurements, the change from the previous day's measurements, and the DVM offset. At the end of all the measurements, the above quantities are summarized.

READCAL - This program is a slight modification of READBOX to run on measuring system REG. The crossbar switch used with system REG is a 300 position switch rather than the 80 position switch used with system VTP, and is addressed slightly differently by the computer.

STDCEL - This program is contained within READBOX (and READCAL) and is used to reduce standard cell difference observations to standard cell emfs. This program is used when individual cell difference observations, within a measurement design, are out of control. It permits the operator to remove selectively those suspect observations from the design and store the corrected cell emfs on floppy and hard disks.

EDIT - This program is a general purpose text (ASCII) file editing program and is used to create or modify an existing "runfile". The "runfile" contains a list of standards and their temperatures that are to be read by READBOX. This file is updated daily using EDIT to contain the current enclosure temperatures, cell identification numbers, and crossbar circuit numbers.

This program can also be used to view or modify any of the data files containing cell difference observations or cell emfs.

The following programs operate on the old minicomputer and are in the process of being converted to run on the new minicomputer.

VFORMAT - This program takes data in the format produced by READBOX and converts it to "Analysis" format for the program described below.

ANALYSIS - This program (and a number of variants created for special purposes) produces control charts of cell emfs and cell temperatures on a line printer. The program also performs a linear fit to the cell emf and can be used to predict the emf of cells at a given time. This is used for calculating V_{LAB} - V_{NBS} from the Volt Transfer Program transport cells.

REPORT - This program in conjunction with ANALYSIS above is used to produce a final VTP report. This program performs text formatting, combining the test report text with the values obtained from ANALYSIS.

BFORMAT - This program converts the READBOX data into a format compatible with EMF below. This is necessary as EMF was originally written for operation with an entirely different measurement system. When completely converted to the new minicomputer, BFORMAT will be bound with EMF to produce one program.

EMF - This program compiles standard cell emf data for the regular calibration service. The data are printed out, statistical tests are applied to determine if the data are sufficiently stable and if the temperature correction is reasonable, and a calibration uncertainty is calculated.

CELL REPORT - This program in conjunction with EMF produces the final calibration report for the regular calibration service.

NBS MEASUREMENT SERVICES:
Solid-State DC Voltage Standard Calibrations

Bruce F. Field

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

January 1988

U.S. DEPARTMENT OF COMMERCE, C. William Verity, Secretary
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Solid-State Voltage Standard Performance and Design Guidelines

BRUCE F. FIELD

I. INTRODUCTION

OVER the past six years the Electricity Division has examined and analyzed the performance of nearly all high-quality commercially-available solid-state (Zener) voltage standards. Based on our observations to date we offer the following set of design guidelines to define what we believe is required in a modern solid-state voltage standard to supplement or replace current standards using saturated cadmium-sulfate standard cells. This document is not a complete specification for a voltage standard but the ideas contained herein should be considered when defining the requirements for a voltage standard. It should also not be inferred that any standards that meet the requirements of this document are necessarily endorsed by NBS as the best or only suitable standards available.

For our purpose here we define a voltage standard as a complete instrument in one box that is based on a solid-state reference, is powered by the ac line or internal batteries, and continuously produces one or more stable voltages. This note presents guidelines that describe two types of solid-state standards with outputs at the 10 V and 1.018 V levels. The first type is a laboratory standard intended for maintenance of a local unit of voltage, while the second is a transport standard designed for comparing two laboratory units of voltage at the 10 V and 1.018 V levels. The laboratory standard is intended to be used as part of a group of like standards to maintain a unit of voltage at the 10 V level to an accuracy of 0.3 ppm (1σ) after corrections have been applied for drift of the standard, and the transportable standard is to be used to transfer a unit of voltage between laboratories to an accuracy of 0.1 ppm (1σ). (All uncertainties in this note are expressed as one standard deviation estimates.)

The guidelines have been divided into two categories, one describing the operational performance of a standard and the second describing important circuit design considerations. The performance guidelines identify the important characteristics of standards such as voltage output stability, output noise, battery life, weight, etc. In the discussion of the performance guidelines we generally do not recommend a particular design for the circuitry of the standard, we only consider the end performance. However, there are several qualities we consider important in the design of the electrical circuitry and these are discussed as design guidelines. Table I is a list of all the guidelines in approximate order of importance.

For each performance guideline a specific goal has been developed to serve as a guide for writing a detailed solid-state voltage standard specification and also as a guide to anyone evaluating such a standard. Certain goals have been made intentionally stringent because either they are additive in nature or they are easily achievable with present technology. We believe that most of the goals described here are attainable using present technology. Where appropriate, differing goals between the laboratory standard and transport standard are noted in the discussion of the guideline. A summary of all the performance goals is given in Table II at the end of the paper.

TABLE I

List of Guidelines

Performance

P1. Long-term drift (stability) of the voltage outputs.
P2. Sensitivity of the voltage outputs to power interruptions.
P3. Noise on the voltage outputs.
P4. Temperature coefficient of the voltage outputs.
P5. Regulation of the voltage outputs with respect to the supply voltage.
P6. Load regulation of the voltage outputs.
P7. Change of the voltage outputs with ac imposed on the output terminals.
P8. Operating time under battery power.
P9. Recovery time of the voltage outputs after transport.
P10. Electrical isolation of the voltage outputs.
P11. Protection of the voltage outputs.
P12. Battery recharge time.
P13. Adjustment range of the voltage outputs.
P14. Terminal posts for the voltage outputs.
P15. Environmental operating conditions.
P16. Physical shock during shipment.
P17. Weight.
P18. Panel indicators.
P19. Battery life.
P20. Provision for an extra battery.
P21. Compliance with electrical safety standards.

Design

D1. Multiple independent references.
D2. Independence of multiple outputs.
D3. Quality of the 1.018 V output.
D4. Electrical isolation.

II. PERFORMANCE GUIDELINES

P1. *Long-term drift (stability) of the voltage outputs.*

Goal P1: The long term drift of each reference should be less than 2 ppm/year at 10 V with day-to-day variations less than 0.1 ppm.

A standard with a stable low-drift output voltage is essential when the standard is to be used to maintain a local laboratory unit of voltage. Although we have observed that the drifts of most standards are generally linear and predictable, a standard with a large drift may require that it be periodically adjusted or that corrections be applied to the data. Presently-available standards are capable of stabilities of ±4 ppm/year or better at the 10 V level. Figure 1 shows the stability performance of the 10 V output of a typical commercial standard. This particular standard has a drift of +0.95 ppm/year with a residual standard deviation of the fitted line of 0.07 ppm.

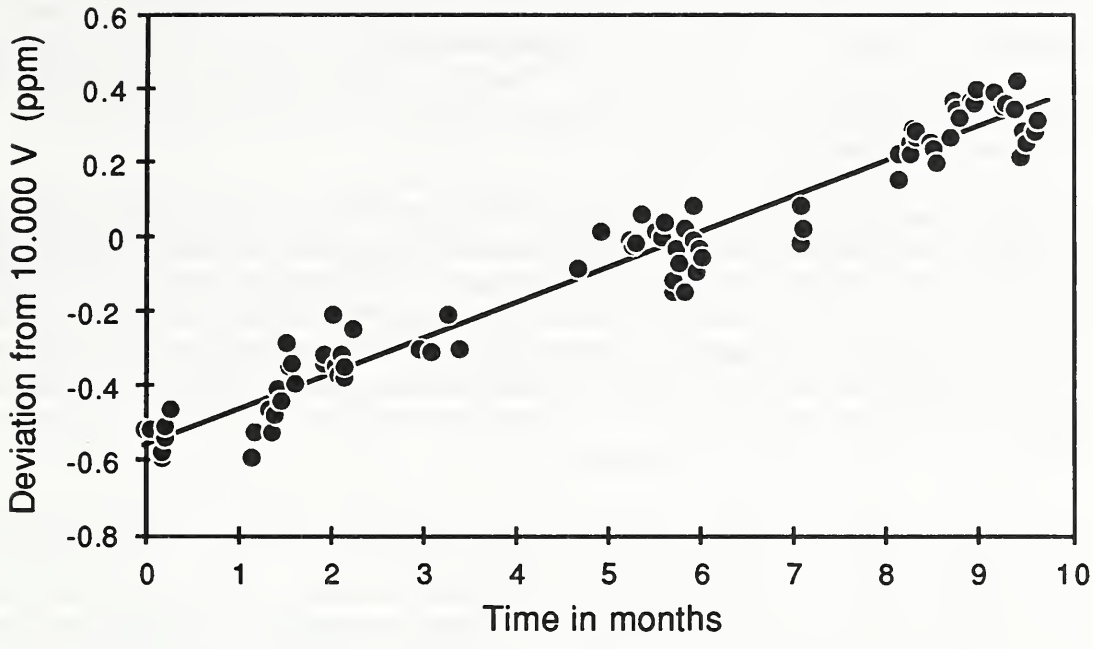


Fig. 1. Long term drift of a typical 10 V standard.

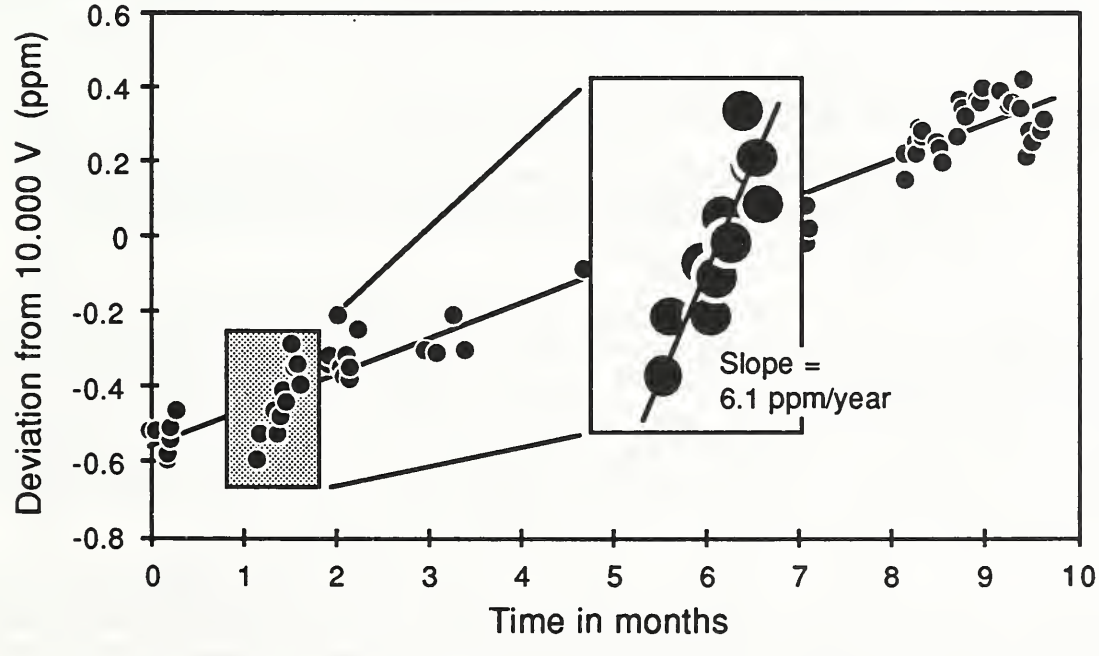


Fig. 2. Detail of structure in the voltage output of a typical 10 V standard.

As can be seen in Fig. 1 there is additional structure in the output voltage that produces day-to-day variations of up to several tenths of a part-per-million from the general drift line. This is typical of most standards tested but the cause is not yet understood. For best accuracy in deter-

mining the stability of the standard it should be monitored for at least six months to predict an annual drift rate. Figure 2 illustrates what can happen if insufficient data are used. Using data taken over a three week period between months 1 and 2, a slope of +6.1 ppm/year is calculated, but this is in error by a factor of six! (In fact for this example no one month period of data comes close to predicting the annual drift rate.) This day-to-day variation in the output makes it difficult to accurately predict an annual drift rate with much less than six months of data. (This problem also exists with standard cells!)

Most standards tested exhibited fairly linear drift rates. For each standard used as part of a volt maintenance procedure the drift rate should be accurately determined by long-term measurements and periodic corrections applied to the value of the standards to correct for the expected drift. If instead each standard is assumed constant between calibrations an additional uncertainty must be added due to its drift. Data to date show that the drift rate of most standards is considerably larger than the uncertainties of the drift corrections, thus the uncertainty of maintaining a volt at the 10 V level can be reduced substantially, usually from about 2 ppm to better than 0.5 ppm, by correcting for the expected drift.

In all presently-available commercial standards the 1.018 V (and 1 V) outputs are derived from the 10 V output using internal resistive dividers. The 1.018 V (and 1 V) outputs of most standards have been found to be significantly less stable and have more day-to-day variation than the 10 V outputs. Drift rates are typically 2 or 3 times worse than the drift rate of the 10 V output. Two standards have been observed for which the 1.018 V output drifted at a rate of greater than 1 ppm/week while the 10 V output showed random variations of 0.2 ppm with no detectable drift. For presently-available standards we do not recommend that the 1.018 V outputs be used as a general replacement for standard cells. One exception is the use of the 1.018 V output as a transfer standard where it is carefully calibrated and used within a short period of time (<1 day).

P2. *Sensitivity of the voltage outputs to power interruptions.*

> **Goal P2:** Voltage output shifts resulting from power interruptions or abrupt ambient temperature changes of 20 °C or less should be less than 0.1 ppm.

Although standards based on Zener diodes generally perform best if the diode is continuously powered and kept at a constant temperature, it is likely, especially during shipment, that the standard will occasionally lose power. In addition to the interruption of current to the diode, the temperature-controlled oven (if there is one) may cool to ambient temperature. The power loss may be due to lengthy shipping times or an extended ac power outage in the laboratory. If this happens it is necessary that after restoration of ac power the standard return to exactly the same voltage it had before the power outage occurred.

We have conducted power interruption tests on a number of commercial standards and have found that the magnitude of the observed voltage shift is vaguely dependent on the individual standard rather than the type of standard being tested; some standards consistently showed small changes while others exhibited changes as large as 2 ppm [1,2].

Figure 3 shows NBS measurements of the 10 V outputs of two temperature-controlled standards. During the two gaps the standards were shipped to (and returned from) another laboratory with the oven turned off during shipment. We estimate the units were off power for approximately 8 hours during each shipment. The first one or two points of SN 10 starting at 3 months and possibly the first six points at 4.6 months may be inconsistent with the remaining points and likely indicate a change and recovery of the standard. Fitting straight lines to the two

sets of data (excluding the two points at 3 months) yield residual standard deviations of 0.100 ppm and 0.052 ppm for SN 10 and SN 11, respectively. Examining the deviations of the individual points from the fitted lines, we conclude that except for the initial recovery of SN 10 there is no indication that the standards were significantly affected (<0.1 ppm) by the shipping process.

A second test on the same two temperature-controlled standards was conducted by carefully calibrating their 10 V output in terms of the U.S. Legal Volt for a five-day period, abruptly removing the power and allowing the ovens to cool to room temperature for a two-day period (typically Saturday and Sunday), and then restoring the power Monday morning and repeating the process six times. The first six points for each of the standards in Fig. 4 represent the mean of the five (approximately) measurements. After the sixth week, the standards were shipped to other laboratories with the power turned off during shipment. The last four points of Fig. 4 are the calibrations while at NBS. Each point represents the mean of from 11 to 64 daily measurements. Least-square lines were fitted to the data where each point was weighted inversely proportional to the number of daily measurements. The residual standard deviations based on an average of ten daily measurements are 0.049 and 0.045 ppm for SN's 10 and 11, respectively.

One nontemperature-controlled standard was tested by cooling the standard from room temperature (23 °C) to approximately 4 °C and holding it there for about 10 hours with the power removed. Measurements were begun one day after resumption of power to the standard and its return to room temperature. Figure 5 summarizes the results of the test. The 10 V output of the standard showed a consistent increase in value after each outage but the magnitude of the shift was unpredictable.

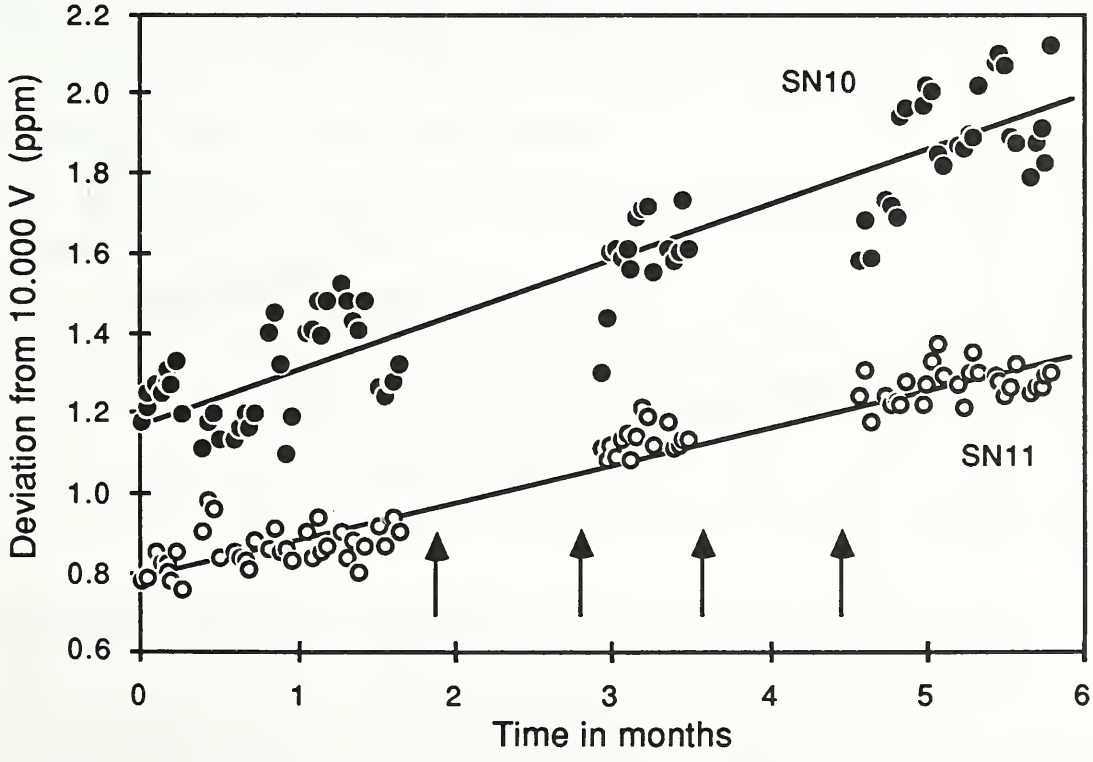


Fig. 3. NBS measurements of the ten volt outputs of two temperature-controlled standards. The standards were shipped via air freight to another laboratory and returned to NBS at the times indicated by the arrows. During all four shipments the power was turned off.

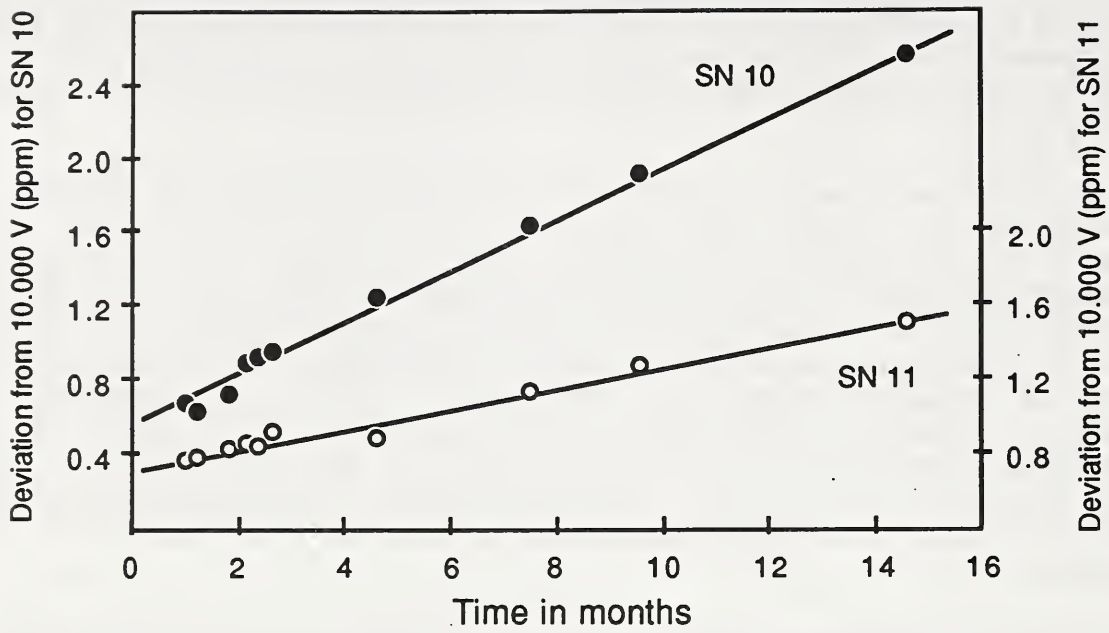


Fig. 4. NBS measurements of the ten volt outputs of two temperature-controlled standards. Each point represents the mean of a number of measurements. For the first six points the power was turned off between each point to simulate the shipping environment. The standards were shipped air freight to several laboratories between the latter four points.

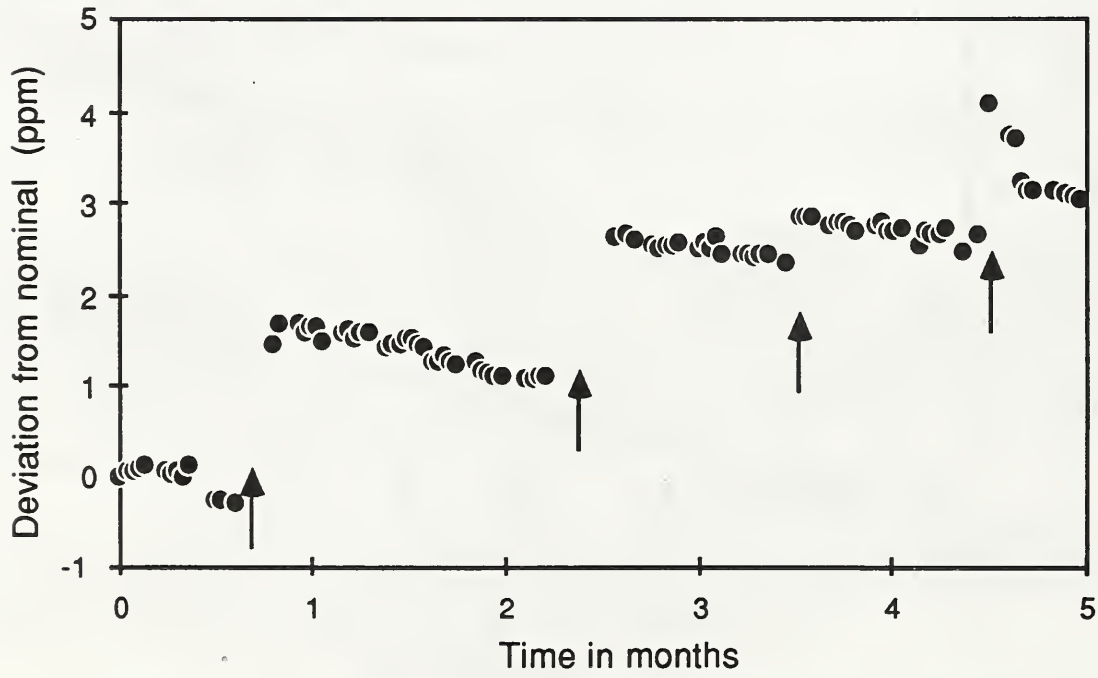


Fig. 5. Measurements of ten volt outputs of a nontemperature-controlled standard. At the times indicated by the arrows, power was removed from the unit and it was cooled to 4 °C for approximately 10 hours. Measurements were resumed after the unit was returned to room temperature (23 °C).

Almost all standards tested (temperature-controlled or not) showed non-reproducible shifts when subjected to power interruptions and abrupt temperature changes. Although it was not generally possible to predict exactly the direction and magnitude of the shift, several standards consistently showed small random shifts, <0.1 ppm. Such a value could be used for these standards as a reliable estimate of the uncertainty caused by power interruptions .

P3. *Noise on the voltage outputs.*

> **Goal P3:** Noise output of any voltage output should be <0.1 ppm rms in a bandwidth of 0.01 - 10 Hz. Day-to-day variations (where sufficient measurements are averaged to negligibly reduce short-term noise) should be less than 0.1 ppm (1σ).

Measurement errors caused by self-generated noise on the voltage standard outputs in the frequency range 0.01 - 10 Hz (short-term noise) can be reduced by having the measuring system integrate the signal over a suitable period. Noise produced at higher frequencies is (or should be) rejected by the measuring system. Noise in the frequency range 0.00001 - 0.01 Hz (day-to-day scatter) in some cases may be reduced by averaging measurements of the standard over several days, but for many tests it must be included as part of the uncertainty of the standard. Available standards typically limit the noise at the output terminals to <0.1 ppm (<1 μV rms on the 10 V range, and <0.1 μV rms on the 1.018 V range) in a bandwidth of 0.01 - 10 Hz which is consistent with the day-to-day scatter observed for most standards [4-6]. The short term noise should be smaller than the day-to-day scatter of the standard so as not to contribute significantly to the latter. Special tests may be required to ensure that all parts of the measuring system are insensitive to noise produced in any other part of the measuring system.

P4. *Temperature coefficient of the voltage outputs.*

> **Goal P4:** The temperature coefficient of any voltage output should be less than 0.01 ppm/°C.

Standards intended for use in a laboratory environment (±2 °C) should have temperature coefficients of the output voltages of 0.01 ppm/°C or less to preclude the necessity of applying temperature corrections. This can be readily achieved with temperature-controlled standards. Figure 6 shows the temperature dependence of the 10 V output of a typical temperature-controlled standard. An additional allowance will usually have to be included for the 1.018 V output because of the temperature coefficient of the resistive divider. Including the divider in the oven will minimize the temperature coefficient and eliminate any temperature hysteresis effect of the resistors.

Figure 7 shows the temperature dependence of the 10 V output of a typical nontemperature-controlled standard with respect to the ambient temperature. The standard is designed to to have a zero-temperature-coefficient at normal room temperature but does not meet goal P4. Nontemperature-controlled standards may also have compensation circuits to monitor the ambient temperature and apply an electrical correction to the output voltage. This can be done with reasonable success over a limited temperature range. But, nontemperature-controlled standards may be affected by large abrupt changes in ambient temperature causing their output to permanently change. Any nontemperature-controlled standard intended for transport should be checked for this property.

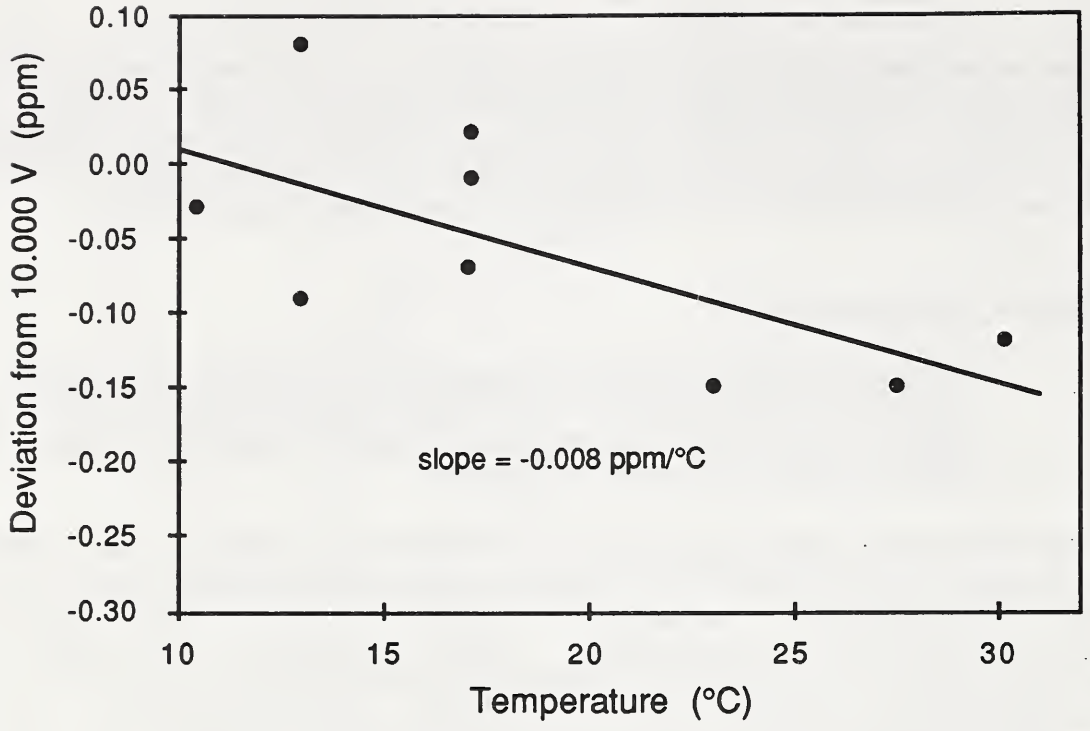


Fig. 6. Deviation of the ten volt output of a temperature-controlled standard when subjected to changes in ambient temperature.

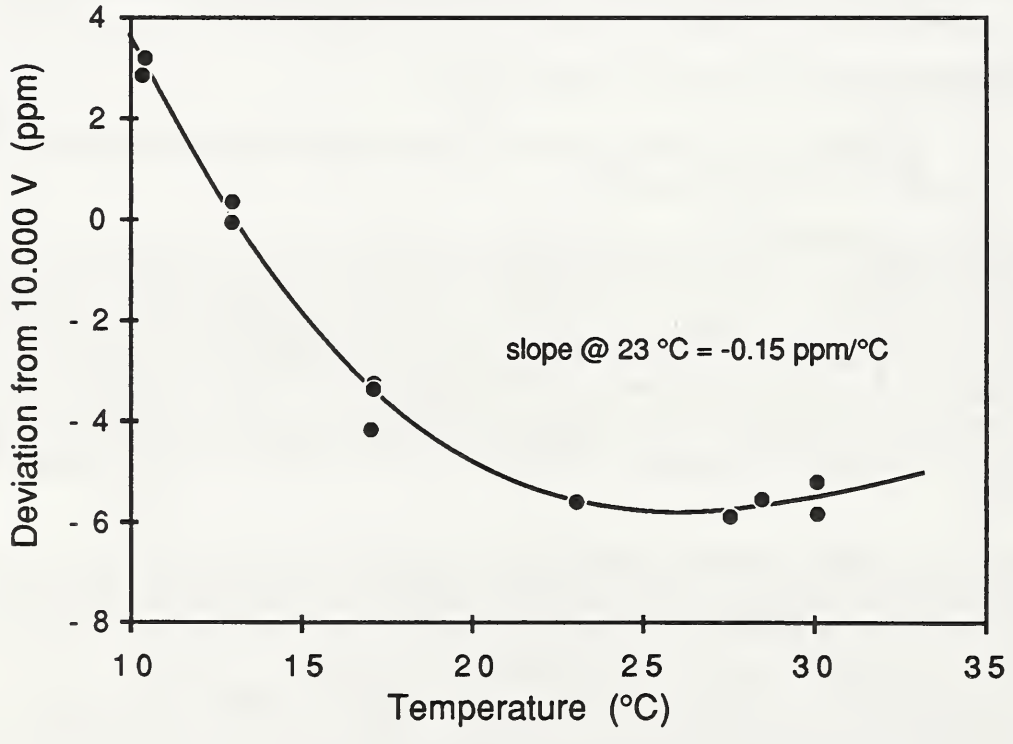


Fig. 7. Deviation of the ten volt output of a nontemperature-controlled standard when subjected to changes in ambient temperature.

P5. *Regulation of the voltage outputs with respect to the supply voltage.*

> **Goal P5:** The maximum change in any output voltage should be 0.01 ppm or less over the supply voltage range (ac and battery) of the standard.

Specifications for the maximum change in the output voltages of presently-available standards are typically 0.05 ppm or less for a momentary or prolonged change in the ac mains voltage anywhere within the operating range specified for the standard. In some cases a small settling time is also specified. If the standard is to be operational at full accuracy under battery power then the supply regulation specification must also apply to battery operation. A light or other indicator should be included to indicate when the battery voltage is sufficient for the standard to be within specifications. For highest-accuracy standards, supply-regulation-dependence should be 0.01 ppm or less over the operating range.

Figure 8 demonstrates a typical change in a nontemperature-controlled standard when it is unplugged from the ac mains at time 0 and allowed to run from its internal batteries. In this case we believe the initial drift during the first hour is due to cooling of the power transformer within the standard as the output voltage is not correlated with the supply voltage. A similar but opposite change is observed when the standard is reconnected to ac power.

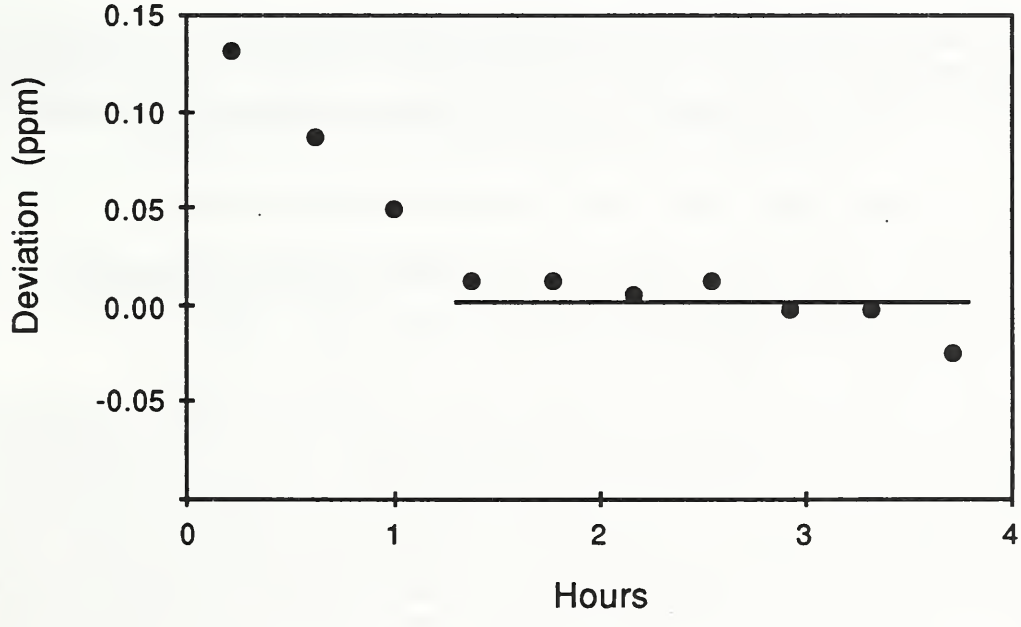


Fig. 8. Change of the ten volt output of a nontemperature-controlled standard when switched from ac mains to battery operation at time 0.

P6. *Load regulation of the voltage outputs.*

> **Goal P6:** The output resistance of the 10 V range should be 0.001 Ω or less with a 2 mA current capability for a laboratory standard and less than 1 kΩ for a transport standard. Output resistance of the 1.018 V range should be 1 kΩ or less. The output resistances of all ranges should be specified by the manufacturer so that the user may apply a loading correction if desired.

All presently-available commercial standards use a buffer amplifier to provide a low resistance output at the 10 V level which is capable of supplying or sinking 2 to 10 mA. The 10 V output may be used in a limited manner to accurately drive a Kelvin-Varley divider for calibration purposes.

Available standards specify output resistances from 0.005 to ≤0.5 Ω. Connecting a 100 kΩ Kelvin-Varley divider to the 10 V tap of a standard with a 0.005 Ω output resistance will change the output 0.05 ppm, while a standard with an output resistance of 0.5 Ω will change 5 ppm. We have observed that even standards with output resistances as high as 0.5 Ω provide a stable, albeit different, output voltage when driving a divider. But, caution must be used if the standard is calibrated without the divider and then used with the divider to calibrate other instruments. In this situation it is preferable to leave the divider permanently attached and calibrate the standard through the Kelvin-Varley, i.e., set the divider to 0.999999X and use the output of the divider. Small errors from voltage drops in the input leads to the divider are also cancelled using this method.

The 1.018 V outputs of standards are generally derived from the 10 V outputs by internal resistive dividers with typical output resistances from 800 to 1000 Ω. Thus no loading is permitted on the 1.018 V output – all measurements should be done using a potentiometric method.

P7. *Change of the voltage outputs with ac imposed on the output terminals.*

> **Goal P7:** All voltage outputs should exhibit a change of less than 0.01 ppm when a DVM (8 mV noise pk-pk, 1 kHz - 5 MHz BW) is connected to that output.

Diodes and other non-linear elements in the circuitry can rectify ac noise introduced at the output terminals from external sources such as digital voltmeters [3]. This can produce a substantial dc shift in the output voltage of the standard when the noise source is connected to the standard. These shifts have been observed using the monitoring system shown in Fig. 9. A digital voltmeter was used as the measuring instrument shown in the figure and was alternately connected and disconnected to the standard under test while null detector (D) was monitored. The null detector must be known to be insensitive to ac for this test; a mechanical galvanometer is recommended. Additionally a filter may be added at the output of the divider to reduce ac coupling to the detector and the standard cell. A number of standards were tested and showed changes in the range of <0.01 ppm to 30 ppm. In each case the outputs immediately returned to their original values when the digital voltmeter was disconnected from the circuit. The voltmeter used for this test was a common 6-1/2 digit model that produced approximately 8 mV of noise peak-peak in the 1 kHz to 5 MHz frequency band.

This problem can introduce a nearly undetectable systematic error in a calibration process if the user is unaware of it. Suppose the standard is calibrated in the calibration laboratory against standard cells using passive apparatus with presumably little ac noise; the "correct" value is thus obtained. If later the standard is used on the production line to calibrate a digital voltmeter, the

standard's output shifts because of ac noise produced by the voltmeter and the voltmeter reading is in error. We found that the dc changes produced by individual instruments (e.g. voltmeters) are extremely reproducible from day-to-day and thus reproducible measurements cannot be taken as a sign that there is no problem.

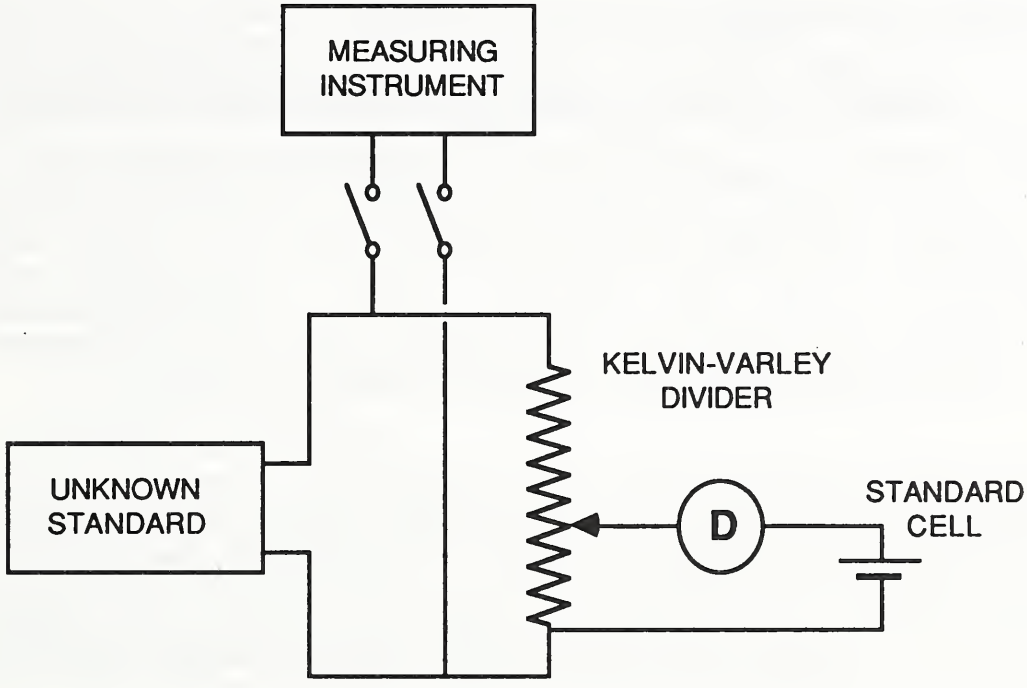


Fig. 9. Test circuit for measuring the sensitivity of an unknown solid-state standard to ac generated by the measuring instrument.

P8. *Operating time under battery power.*

> **Goal P8:** The battery should supply power for operation of a transport standard for 72 hours at a 20 °C ambient temperature.

Laboratory standards may require battery operation for maintenance of the standard during laboratory ac power outages to prevent unpredictable shifts in the outputs, or for special tests that require the standard to be completely isolated from the ac mains and/or ground. The duration of power outages is unpredictable of course, but an 8 to 24 hour battery operating time would seem reasonable. Special tests involving the standards may impose other battery operating conditions and require a more lengthy battery operating time.

Transport standards that are to be shipped under battery power via air freight will need considerably longer battery operating times. Our experience with transporting three standards within the United States via several guaranteed 24-hour/overnight delivery services leads us to conclude that 72 hour battery operation is desirable [1]. Shipping standards by air within a 24 hour time frame requires careful coordination between the laboratories, often with laboratory personnel delivering the standard to, and picking up the standard from, the airport. We consider a battery operating time of 24 hours to be unacceptable for a transport standard, although in some cases it is possible to extend the operating time by including additional batteries in the shipping container. This is less desirable as two massive objects in the same container are more likely to cause damage to each other than one alone.

An alternative to this approach is to design the standard to be shipped with the power turned off. Thus there is no limit imposed on shipping time by the battery operating time. (This may also save substantially on the weight of the standard; see guideline P17.)

P9. *Recovery time of the voltage outputs after transport.*

> **Goal P9:** All voltage outputs should recover to within 0.02 ppm of their final value in less than 2 hours after restoration of ac power.

If the standard is shipped under battery power with the oven operating there is no reason to expect a significant change in the output after restoration of ac power and hence no recovery time is expected. Most manufacturers recommend shipping their standards under power for highest accuracy transfers. If the manufacturer suggests that transfers can be made not under power a recovery or settling time to the final-expected-value should be specified to indicate when the standard will be ready for use. For example: "After shipping, the output voltage will be within 0.5 ppm of it's final value 8 hours after restoration of ac power."

We have made a number of transfers with two temperature-controlled standards that were shipped not-under-power (see guideline P2). One of the units never showed any significant recovery effects; the unit was received in the laboratory one day, and measurements begun the next day agreed with succeeding measurements within the normal day-to-day scatter. But, the first two or three day's measurements on the second unit often, but not always, were in slight disagreement with succeeding measurements. Figure 3 shows an example of this recovery behavior where both standards were shipped together in the same shipping container. SN 10 at 3 and 4.6 months apparently shows recovery effects, while SN 11 appears to be unaffected.

A special situation may exist where the unit is manually switched to a "transit" mode and during this time the standard is not intended to provide it's specified accuracy. In this mode the batteries may be used only to maintain a constant diode current while the oven control is turned off to conserve battery power or, the diode current may not be closely regulated. In this case a recovery or settling time should be specified, or a front panel light provided, to indicate when the standard is at full specified accuracy after being switched back to "operate".

P10. *Electrical isolation of the voltage outputs.*

> **Goal P10:** Greater than 10^{11} Ω from any output to any other output, to ground, or to the ac mains.

The output(s) of the standard must be well isolated from the ac mains and ground, and when multiple references are provided they should be isolated from each other. Many experiments and calibration procedures rely on the standard producing an output that may be "floated" off ground. The typical user is most likely familiar with standard cells which usually have excellent isolation, between cells, to ground, and to the ac mains, and is unlikely to consider that Zener standards should behave any differently. If the multiple outputs of the standard cannot be connected in series to produce a larger voltage, the user should be specifically warned of this. Unlike the cell, the Zener standard is connected to the ac mains and operates with much higher internal voltages; they typically have voltages as large as 24 V at some portions of the circuitry. Thus a 1.018 V output could possibly be driven to as much as 24 V above ground by leakage resistances from the circuitry to ground.

P11. *Protection of the voltage outputs.*

> **Goal P11:** There should be no lasting effects from shorting or applying up to 1000 V (current limited to 25 mA) across any output or between any output and ground.

As a minimum, the standard should be unaffected by indefinitely shorting any of the outputs; the output should return to its original value soon after the short is removed. If the time required to return to the original value is greater than a few seconds, a settling time should be specified. There should also be protection against inadvertent application of 1000 V to any of the outputs. Such a situation could happen during calibration of a dc calibrator with 1000 V capability.

P12. *Battery recharge time.*

> **Goal P12:** The battery recharge time should be 24 hours or less for fully discharged batteries and the charging circuit should not overcharge the batteries.

The time required to recharge the internal battery (if any) is generally not a problem. If the standard is being shipped to a laboratory for calibration, several days at a minimum will be required for the calibration and this usually far exceeds the battery recharge time. The optimum recharge time will depend on the charging method and the particular battery being used. A recharge time from 14 to 24 hours for fully discharged batteries is reasonable. The charging circuit should not overcharge that batteries if left permanently connected.

P13. *Adjustment range of the voltage outputs.*

> **Goal P13:** No adjustable elements should be included for regulation of the final output voltages.

For best stability we recommend that there be no adjustable elements in the output circuitry. The output should be trimmed initially at the factory using fixed-valued components, and not adjusted afterward. The standard should be used the same way standard cells currently are; each standard is assigned a calibrated value which is not necessarily the nominal value. We also recognize, however, that some applications require standards that produce an exact nominal value. In this case a separate adjustable output, with an adjustment range only large enough to compensate for the expected drift of the standard, may be added to the standard. The adjustment device should have a continuous resolution of 0.1 ppm or better.

Because of the physical shock encountered by transport standards during shipping (we have observed over 120 g's), we recommend that adjustable elements never be included in standards designed for transport.

P14. *Terminal posts for the voltage outputs.*

> **Goal P14:** The standard should have separate low-thermal-emf terminals for each reference output, arranged for easy interconnection.

Low-thermal-emf (e.g., copper) binding posts should be used for all voltage outputs. Separate common terminals should be provided for each voltage output. We judge separate common terminals to be more reliable as only one wire or lug is connected to the terminal. Where multiple references or output voltages use a single common terminal, there is a greater likelihood that one or more of the wires on the terminal will make a poor contact. Switched outputs (i.e., multiple references switched to a single output) should not be used under any circumstances because of the probability of poor switch performance and the inability to use the standard with an automated switching system.

In a standard with multiple references, intercomparisons between the references can be done easily, even with multiple commons, if the common terminals are arranged in line with one another to permit a single copper shorting wire to be placed across all of them. The measuring instrument can then be connected between pairs of positive terminals to complete the measurement circuit.

P15. *Environmental operating conditions.*

> **Goal P15:** The standard should operate at full accuracy under normal temperature, pressure, and humidity excursions encountered in the laboratory.

All accuracy specifications should apply when the standard is at laboratory conditions. If degradation of the specifications is necessary for use under less optimum conditions, e.g., on a production line, the revised accuracy specifications should also be stated.

Laboratory conditions	
Temperature range:	(23 ± 2) °C
Humidity:	10 to 60 %RH
Altitude:	-300 to 2000 m

P16. *Physical shock during shipment.*

> **Goal P16:** A transport standard and its shipping container should be designed to accept g-forces as high as 120 g's in any direction without damage. A shipping container should be recommended or supplied.

Standards designed for transport will likely require protection with a shipping container containing additional packing material. We have shipped a number of standards in foam lined shipping containers and have instrumented them with ball-and-spring type shock indicators. The combined weight of the standard and shipping container was approximately 27 to 36 kg and the shock indicators were securely fastened to the standards. These indicators are rated for a particular g-force and the internal springs and balls fly apart if the enclosure is subjected to the rated or greater force. During almost all shipments forces of at least 60 g's were encountered, and during one shipment a force of greater than 120 g's was recorded.

During this time we noted a pattern of serious damage when standards were shipped in a particular shipping container with approximately 5 cm of foam insulation surrounding the standard. In each case the standard was of a type with a removeable battery pack, and a printed circuit board that mates with a connector inside the instrument shifted during shipment, shorting out the battery and charring the printed circuit board.

P17. *Weight.*

> **Goal P17:** A transport standard should weigh less than 9 kg (20 lb.), 13.6 kg (30 lb.) including it's shipping container.

Since laboratory standards are (or should be) rarely moved, weight is not an important consideration. Practically, the standard should be movable by one person, and most presently-available standards meet this criteria having a weight of 18 kg or less.

Weight is more of a problem with transport standards as heavy items generally suffer rougher handling during shipment. This is one area where present standards are seriously lacking. It is necessary to produce a multiple reference transport standard that weighs less than 9 kg. Anything weighing much more than this is unwieldy and distinctly less useful as a transport standard. Remember, the competition is a 4-cell standard cell enclosure weighing 11 kg (13.6 kg with the shipping container). As batteries usually account for a substantial portion of the weight of a standard, a considerable savings in weight can be achieved if they can be reduced or eliminated. Thus, the weight of a transport standard can be reduced by designing it to be shipped with the power turned off.

P18. *Panel indicators.*

> **Goal P18:** A standard should have suitable front-panel indicators to clearly verify that the unit is operating properly.

The standard should have suitable front-panel indicators to verify that the unit is operating properly. These include (1) an oven temperature indicator or monitoring device, (2) a battery charge light to indicate whether the battery is charging and when it has reached full charge, (3) a power failure indicator to monitor any power interruptions to the reference or oven, (4) an ac power light to show when the unit is operating from the ac mains, and (5) an indicator to show when the battery is within its operating voltage limits.

P19. *Battery Life.*

> **Goal P19:** Batteries should supply at least 50% of rated capacity for 2 years.

As batteries age, their capacity decreases, decreasing the operating time of the standard while on battery power. One of the most annoying problems we have had is verifying the capacity of a set of batteries installed in a standard. The most frequent cause of unsuccessful transfers is unknown battery capacity that is a fraction of the specified capacity. The manufacturer should recommend a test procedure for verifying the capacity of the batteries. Alternatively, a regular replacement schedule could be recommended.

It would be extremely desirable to include some kind of test circuit in the standard to detect marginal or failing batteries. Another approach might be to mount the batteries in a chassis or box that can be removed without opening the instrument, to provide for the easy interchange of suspect batteries with good batteries. The suspect batteries could then be tested outside the standard using a procedure recommended by the manufacturer.

P20. *Provision for an extra battery.*

> **Goal P20:** A connector should be provided on the standard to permit the use of additional external batteries to extend the operating time of the standard.

Transport standards should provide a connector on the standard for connecting an external battery to extend the operating time for lengthy shipments. The extra batteries should be charged by the internal charger of the standard. A desirable feature would be to provide for operation on 12 V dc so that during shipment by automobile the electrical system of the car can be used to power the standard.

P21. *Compliance with electrical safety standards*

> **Goal P21:** The standard should comply with all applicable U.S. and international safety standards, such as UL 1244, IEC 348, and VDE 0411-1973.

III. DESIGN GUIDELINES

D1. *Multiple independent references.*

It is absolutely necessary to use multiple standards, or a standard containing multiple references, to evaluate the errors associated with transporting a standard from one location to another. When only one reference/standard is used there is no way to assess the uncertainty of a particular transfer other than by using data from similar previous experiments. When using multiple references/standards, changes in the relative differences between the references as measured at both locations can be used as a statistical check or assessment of that part of the uncertainty involved with the transport of the standard.

Multiple independent references contained within a single standard (not multiple outputs from the same reference) are a convenient way to provide redundancy in establishing or transporting a unit of voltage. Just as standard cell enclosures are never designed for only one cell, Zener standards should contain more than one reference device. More is almost always better, however a reasonable number of reference outputs is on the order of four to six, with each reference providing a 10 V and 1.018 V output. If there are too few references there is not enough redundancy, if there are too many references then too many measurements are required. The alternative of using multiple standards instead of a single standard with multiple references is not recommended. This approach is expensive (i.e., more standards, more shipping weight), inconvenient, and more likely to produce damage to the standards because of rougher handling during shipment.

As noted above (P14.), each reference within a standard should have its own separate terminals brought out to the front panel. This permits intercomparison of the individual references and allows the user to identify noisy references or references that are drifting excessively with respect to the rest of the group. Algorithms can be developed and applied for statistical removal of abnormal references from the group to improve the overall stability of the group mean.

D2. *Independence of multiple outputs.*

The statistical procedures and evaluation of uncertainties referred to in the last section generally require that the multiple outputs (references) of the standard be independent from one another with respect to all environmental conditions. If independence is not achieved then the procedures will underestimate the uncertainty.

Independence among multiple references can be achieved (but not guaranteed) by having separate power supplies, separate pre-regulators, separate voltage dividers, and separate ovens, for each diode reference, i.e., build several completely separate standards and house them in one cabinet. If it can be shown that one or more of these items contribute very little to the overall performance of the standard then it may be possible to have one common element for all the references, e.g., a common power supply or oven. We recommend that wherever practical the designer should avoid using circuit elements common to all the references.

We have evaluated several multiple-reference standards and have observed that the day-to-day fluctuations in the output voltages are correlated with one another indicating a dependence between the references. This may be caused by the power supply or the oven (or both) which are common to all the references. The manufacturer claims that the diode references are specifically chosen with different temperature coefficients to minimize correlation between the outputs.

D3. *Quality of the 1.018 V output.*

Zener standards are currently being used as replacements for standard cells and will continue to be used as such for some time. The quality of the 1.018 V output in most standards is very poor compared to the 10 V output. Improved dividers must be developed for this application. Bulk-metal-film dividers may be considered for use in a high-quality 1.018 V standard.

Another divider technology of interest is the time division divider (TDD) used in most high-quality dc calibrators. This technique involves switching the output between two references, usually zero volts and some other fixed voltage, and filtering the output to produce a voltage that is equal to the duty cycle times the fixed voltage. Linearities of 0.1 ppm or better have been claimed by manufacturers. This application requires only a simple version of the TDD as only stability is required – it will operate at a fixed duty cycle. An added advantage, if the duty cycle is adjustable, is that any required voltage adjustment may be made digitally.

D4. *Isolation.*

We recommend that each reference in a multiple-reference standard be fully and independently guarded, starting with a shield on the secondary of the power transformer and continuing to the front panel binding posts. If a single transformer is used for multiple references it should have multiple secondaries, each with it's own shield. Ideally, each reference should have an individual battery contained within it's guard, although this presents other problems with testing, recharging, and replacing the batteries. As an alternative, high-isolation switches (relays) could be used to disconnect one set of batteries from all the reference circuits when the batteries are not needed.

IV. CONCLUSIONS

Presently-available Zener voltage standards are reasonable and useful tools for maintenance of a unit of voltage at the 10 V level to an accuracy of 1 ppm. It has been demonstrated that when selected standards are carefully used as a transport standard a 10 V unit of voltage may be transferred between two laboratories to an accuracy of 0.08 ppm. However, present day standards are lacking in many areas and the preceding guidelines and goals are intended to address their shortcomings. The goals are generally realistic, being well within the grasp of current technology; major breakthroughs in technology are not required.

We also recognize that the quality of any individual standard depends heavily upon the quality of the particular Zener reference contained within it. This problem can be traced back to the poorly understood diode manufacturing process. Additional research needs to be done on the relationship between Zener diode performance characteristics, especially stability, and manufacturing process parameters before significantly improved Zener standards can be developed.

TABLE II

Summary of Preformance Goals

Guideline number

P1. *Long-term drift (stability) of the voltage outputs.*
The long term drift of each reference should be less than 2 ppm/year at 10 V with day-to-day variations less than 0.1 ppm.

P2. *Sensitivity of the voltage outputs to power interruptions.*
Voltage output shifts resulting from power interruptions or abrupt ambient temperature changes of 20 °C or less should be less than 0.1 ppm.

P3. *Noise on the voltage outputs.*
Noise output of any voltage output should be <0.1 ppm rms in a bandwidth of 0.01 - 10 Hz. Day-to-day variations should be less than 0.1 ppm (1σ).

P4. *Temperature coefficient of the voltage outputs.*
The temperature coefficient of any voltage output should be less than 0.01 ppm/°C.

P5. *Regulation of the voltage outputs with respect to the supply voltage.*
The maximum change in any output voltage should be 0.01 ppm or less over the supply voltage range (ac and battery) of the standard.

P6. *Load regulation of the voltage outputs.*
The output resistance of the 10 V range should be 0.001 Ω or less with a 2 mA current capability for a laboratory standard and less than 1 kΩ for a transport standard. Output resistance of the 1.018 V range should be 1 kΩ or less. The output resistances of all ranges should be specified by the manufacturer so that the user may apply a loading correction if desired.

P7. *Change of the voltage outputs with ac imposed on the output terminals.*
All voltage outputs should exhibit a change of less than 0.01 ppm when a DVM (8 mV noise pk-pk, 1 kHz - 5 MHz BW) is connected to that output.

P8. *Operating time under battery power.*
The battery should supply power for operation of a transport standard for 72 hours at a 20 °C ambient temperature.

P9. *Recovery time of the voltage outputs after transport.*
All voltage outputs should recover to within 0.02 ppm of their final value in less than 2 hours after restoration of ac power.

P10. *Electrical isolation of the voltage outputs.*
Greater than 10^{11} Ω from any output to any other output, to ground, or to the ac mains.

P11. *Protection of the voltage outputs.*
There should be no lasting effects from shorting or applying up to 1000 V (current limited to 25 mA) across any output or between any output and ground.

P12. *Battery recharge time.*
The battery recharge time should be 24 hours or less for fully discharged batteries and the charging circuit should not overcharge the batteries.

P13. *Adjustment range of the voltage outputs.*
No adjustable elements should be included for regulation of the final output voltages.

P14. *Terminal posts for the voltage outputs.*
The standard should have separate low-thermal-emf terminals for each reference output, arranged for easy interconnection.

P15. *Environmental operating conditions.*
The standard should operate at full accuracy under normal temperature, pressure, and humidity excursions encountered in the laboratory.

P16. *Physical shock during shipment.*
A transport standard and its shipping container should be designed to accept g-forces as high as 120 g's in any direction without damage. A shipping container should be recommended or supplied.

P17. *Weight.*
A transport standard should weigh less than 9 kg (20 lb.), 13.6 kg (30 lb.) including it's shipping container.

P18. *Panel indicators.*
A standard should have suitable front panel indicators to clearly verify that the unit is operating properly.

P19. *Battery life.*
Batteries should supply at least 50% of rated capacity for 2 years.

P20. *Provision for an extra battery.*
A connector should be provided on the standard to permit the use of additional external batteries to extend the operating time of the standard.

P21. *Compliance with electrical safety standards.*
The standard should comply with all applicable U.S. and international safety standards, such as UL 1244, IEC 348, and VDE 0411-1973.

REFERENCES

[1] L. S. R. Becker, B. F. Field, and T. E. Kiess, "10-V Round-Robin Test Conducted on a Solid-State DC Voltage Standard," *IEEE Trans. Instrum. Meas., vol. IM-35, pp. 383-386, Dec. 1986.*

[2] "Transportable Ten Volt DC Standard," *NBS-CCG Quarterly Progress Report - 1 July 1985 - 30 September 1985.*

[3] W. G. Eicke, "Operating Characteristics of Zener Reference Diodes and Their Measurements," *ISA Trans., vol. 3, no. 2, pp. 93-99, Apr. 1964.*

[4] *732A DC Reference Standard Instruction Manual,* John Fluke Mfg. Co., Inc., 1985.

[5] *Trancell Precision Solid State Voltage Reference Standards Specifications,* Koep Precision Standards, Inc.

[6] *2734A Direct Voltage Reference Bank, Operation and Maintenance Manual,* Valhalla Scientific, 1987.

Josephson Array Voltage Calibration System: Operational Use and Verification

Richard L. Steiner
Bruce F. Field

Reprinted from
IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT
Vol. 38, No. 2, April 1989

Josephson Array Voltage Calibration System: Operational Use and Verification

RICHARD L. STEINER AND BRUCE F. FIELD, MEMBER, IEEE

Abstract—**A new Josephson array system now maintains the U.S. Legal Volt. This system is almost fully automated, operates with a typical precision of 0.009 μV, and readily allows U.S. Legal Volt measurements weekly, or more frequently if desired. This system was compared to the previous volt maintenance system, and agreement was made to within 0.03 ppm. This verification is limited by uncertainties in the resistive divider instruments of the previous system.**

I. Introduction

IMPLEMENTATION of a new Josephson array measurement system to maintain the U.S. Legal Volt began February 10, 1987. With a Josephson array operating near 1.018 V, groups of primary standard cells are now calibrated more easily and quickly than was previously possible with an earlier measurement system based on a 10 mV, two junction device [1]. Measurements of the primary cell groups are made more frequently and at least three times greater accuracy than the older system. To be assured of this accuracy, the array system was tested extensively against the original Josephson calibration system with its two independent 100:1 resistive scaling instruments. This assured that neither the array chip circuitry nor the new measurement instrumentation and procedures introduced any unknown errors.

The development of the Josephson array and its use in a measurement system have already been described in some detail [2]–[7]. In summary, the NBS device, simply called an array, consists of 1500 or more Josephson junctions connected in series as part of a microstripline distribution network. A 94-GHz millimeter wave Gunn oscillator drives each junction into zero current, quantized voltage states. (Since this is the frequency normally used, we will be referring to the step voltage as approximately 200 μV.) The frequency f and the assigned value of the quotient of fundamental constants e and h, determine the voltage between each step. The total number of steps n of all junctions results in the voltage generated across the device

$$V_n = nf/(2e/h).$$

Selecting different step numbers, as low as one step, provides many possible voltages. A specific, precisely known voltage is then generated by adjusting the millimeter wave frequency. Negative voltages are generated as easily by reverse biasing the array (choosing the negative step number).

Manuscript received June 10, 1988. This work was supported in part by the Calibration Coordination Group of the U.S. Department of Defense.

The authors are with the National Bureau of Standards, Electricity Division, Gaithersburg, MD 20899.

IEEE Log Number 8825811.

II. Array Calibration System

The array calibration system was planned to be flexible in providing a range of output voltages, to be at least as accurate as a previous system, and to be automated for fast, uncomplicated operation. Most of the system components are commercially available instruments. The system consists of a millimeter wave section for supplying a known frequency to the array device, an electronic biasing section for selecting a particular step voltage, and a voltage comparison section for measuring the value of an unknown reference. A computer coordinates all data acquisition, instrument control, and data reduction. Fig. 1 shows a block diagram of the entire array calibration system.

A. Millimeter Wave Hardware

The source for this system is a low noise Gunn diode oscillator supplying 45 mW over the range 93–95 GHz. It has several advantages over a klystron: it requires only an inexpensive and safe 15-V power supply, it does not require water cooling, and has a long operational lifetime.

A frequency controller containing a synthesizer, phase-lock circuitry, power supply, and computer interface, is dedicated to phase-locking the Gunn source over the whole frequency range. This controller was designed to stabilize the center frequency to within 20 Hz and reduce phase noise to a very low value of a few kilohertz. Furthermore, the frequency can be changed in 1-kHz increments, corresponding to 10-nV changes in the 1-V array output. The controller also directly supplies voltage for biasing the Gunn source. It allows electronic frequency changes of up to several hundred megahertz without causing extraneous shifts in frequency and power, which could produce voltage step jumps in the array. The time base is an external 10-MHz signal, which is derived from a LORAN C receiver for NBS frequency traceability with an error of less than 5 parts in 10^{11}.

The millimeter waves are transmitted through WR-10 (75–110 GHz) waveguide to a vane attenuator to regulate the power. A narrow-band low-loss isolator follows and helps the source lock stably at any attenuation level. To

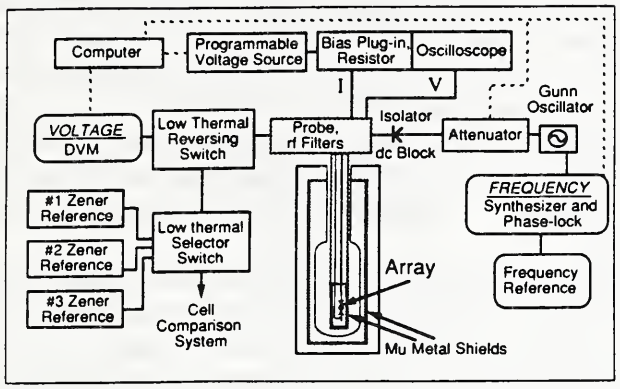


Fig. 1. Schematic of the voltage measurement system based on a Josephson array.

decouple the electrical ground of the millimeter wave section from the probe and dc measurement section, a piece of Mylar tape and nylon screws interrupt the dc circuit of the waveguide just before it enters the probe. A short waveguide transition section is placed here to match the probe WR-12 (60–90 GHz) waveguide.

B. Electronic Bias Circuitry

The main function of the electronic biasing section is to select a step at a particular voltage and to provide an oscilloscope display of the *I–V* characteristic of the array for step observation. Presently a digitally-programmable voltage source with 1-ppm resolution and a low noise, low impedance output provides a bias voltage to the array. A custom built oscilloscope plug-in module buffers the bias source, adds a series resistance to aid in step selection, and provides the current channel for the *I–V* display. The voltage channel for the display is supplied from an isolation amplifier plug-in connected to the array with a pair of leads that are separate from the main voltage sense leads. Because small currents flow through the array, care must be taken to insure that the connections at the array are superconducting, otherwise an offset error will be added to the DVM voltage. The bias plug-in unit itself can serve as a general purpose, manual bias supply. It provides a versatile bias control for diagnostic testing of array chips; and in exploring the array's behavior it has proven quite indispensable.

C. DC Voltage Measurement Apparatus

The voltage measurement section of the system (see Fig. 1) consists not only of a digital voltmeter (DVM) and low-thermal switches, but three local Zener voltage references as well. The switches connect the array in series opposition to each Zener reference, allowing the DVM to record only the small difference voltage. Though presented last, this section is arguably the most significant, since the components used here limit the accuracy of the system.

The most critical component is the DVM. Low noise emission from the input terminals is the first requirement for its selection. Since any electrical noise interferes with the functioning of the array, it is hard to quantify an intolerable noise level; increased noise leads to increased step instability until ultimately the array fails to generate steps. Five DVM models were tested in the array system, three of which were studied with a spectrum analyzer for noise emission. Differences were seen, but only by connecting each model into the array circuit and looking at step stability could we determine whether it was acceptable or not. A three-stage 40-kHz low-pass *LC* filter is currently built into the probe, but two DVM models were still too noisy. Different units of the same model responded similarly. Increased filtering is an undesirable solution to this problem; the time constant becomes excessively long, the leakage resistance decreases, and the series resistance increases, all potentially increasing the systematic errors. Other considerations for a DVM are low offset-current high input-impedance, and good isolation from ground, since the array system is susceptible to error from these sources as is any other precision voltage measurement apparatus.

An important benefit of using a DVM is the flexibility to make voltage measurements over a wide voltage range. Analog nullmeters were tried as an alternative to a DVM, but since the array steps are about 200 μV at 95 GHz, spontaneous step jumps overload the more sensitive scales of a nullmeter, causing delays for meter recovery or even worse, zero offset changes which invalidate previous data. DVM's, even on the most sensitive range required for 10-nV resolution, can easily handle voltages of several hundred microvolts. Also, having a DVM capable of measuring at the 1-V level through the same input makes it possible to determine the step number without prior knowledge of any reference voltage. Finally, a programmable DVM frees the computer to vary voltage bias parameters while the DVM simultaneously records the variations as a check for constant voltage. Because of this the automatic step selection and step viability checking is fast and reliable.

Two sets of low-thermal switches connect the references to be calibrated to the array. The low thermal selector switch (Fig. 1) is a specially modified telephone relay matrix in a temperature lagged enclosure and is operated with a manual controller to select each Zener reference as needed. (This switch is also operated under computer control to connect the Zener reference group to a comparison system in the standard cell facility for calibration of the primary standard cell group.) The low thermal reversing switch completes the measurement circuit, connecting the array, DVM, and Zener reference in appropriate series combinations to eliminate spurious thermal EMF's.

A group of three Zener references comprise an intermediate voltage standard for transferring the array calibration to the primary group of standard cells. The references are adequately stable during the short times required to transfer to the cells, and during the step selection procedure they can be abused with large voltages and transient currents that cells could not possibly endure. Each reference is individually calibrated under battery

power, otherwise they suffer from poor ground isolation and significant ac pickup.

III. Operation

A. Procedure

The procedure for operating the array is constantly evolving as programmable instruments are added, therefore, only a brief description is given here. There are three basic stages: 1) initial voltage measurement and frequency calculation, 2) adjustment of the array output to the proper voltage, and 3) dc voltage measurement. The low thermal switching is still manually controlled. Step 2) is occasionally done manually when the initial millimeter wave power needs special adjustment or the array is particularly unstable.

Stage 1) is simple. The DVM measures the Zener reference voltage directly. The Gunn oscillator is manually tuned to an approximate frequency where the array operates efficiently, and then the computer calculates the step number and sets the frequency to closely match the array to the reference. To determine the correct step, the DVM must accurately resolve about 100 μV, one-half the difference between two steps. This routine is flexible as it will work as well at any voltage as at 1 V and does not require the operator to input the approximate value of the Zener reference.

Next, the array is connected in series opposition with the Zener reference using low-thermal switches, and then the computer gradually decreases the millimeter wave power while adjusting the external voltage bias supply to select a stable step which nearly equals the reference voltage. This second stage of operation has proven to be the most difficult to automate. Generating array steps requires a rather specific amount of millimeter wave power which depends upon the frequency, step number, and array design; but this power cannot be predetermined. A narrow range of power is needed to maximize the current amplitude and stability of steps near 1 V, about 20 mW at the probe top for this system. Once a proper power level is found, further power changes are generally unnecessary.

The step selection routine starts with a large bias voltage applied to the array, roughly 3 mV times the total number of junctions per array, to force all the junctions into the same polarity. To maximize the probability of selecting the center of a step, a dc bias (approximately the desired step voltage) plus an ac oscillation are applied to the array, with the ac oscillation gradually reduced to zero. Next, a dither voltage, much smaller than the step current, is applied to check for constant voltage and simultaneously determine the step voltage. This step finding sequence is repeated with small dc bias corrections until the correct step is selected. A high resolution (1 μV) voltage supply is thus mandatory for the selection of a particular step. If the computer is unable to find a step, the operator resorts to observing the step pattern on an oscilloscope and manually adjusts the power and voltage bias. The present step selection routine fails perhaps five percent of the time.

Once the desired step is found, all bias oscillations are stopped to prevent unnecessary step jumps. The frequency is fine tuned until the voltage difference between the array and the reference is less than 0.2 μV and the frequency is not changed again during the measurements of that reference. Voltages from meter or thermal offsets remain small so additional frequency tuning is unnecessary.

Each voltage difference measurement is an average of six DVM readings where each reading integrates for about 6 s. Every data point is in turn an average of two measurements, one normal DVM measurement and a second with the input to the DVM reversed. This immediately eliminates meter offsets which fluctuate more rapidly than the thermal EMF's in the rest of the circuit. Because the thermal EMF's are stable over the short 15 min of the full measurement cycle, five data points of a single array polarity are recorded before the array is reverse biased, and five more points in the reversed polarity complete the measurement. Alternatively, reversing after each data point, a desirable procedure if the thermal offsets or Zener reference voltages change unpredictably during the measurement, adds 10–20 extra minutes for the additional step searches.

B. System Characteristics

Unless the array produces only discrete steps of constant voltage, array stability is generally reduced and it is also more difficult to automatically verify a true Josephson voltage state. A certain amount of millimeter wave power is required to produce steps at the desired voltage, but too much power decreases the current amplitude of the steps or worse, generates "array noise," i.e., fractional step values, voltage states between steps, and nonconstant voltage steps. This power attenuation is complicated by the unique response of each array. Definite resonance coupling is evident where arrays operate perfectly with less than 15 mW, while at some frequencies more than 40 mW is required. Sometimes the same array generates noisy, unstable steps outside of the frequency ranges where the steps are ideal.

The voltage biasing circuit provides an external voltage across the array, but it cannot simply force the array onto a desired step, because the array itself is an active voltage source. Constant voltage steps must all cross the zero current axis, and each has an upper and lower current limit determined by the critical current of the junctions and the millimeter wave power level. Minutely increasing or decreasing the array current beyond a critical value, the edge of the smallest step, forces the array to jump from that step to another where the array voltage is closer to the bias and the current is reduced to within these critical limits. While the array maintains a constant voltage, there can exist a voltage difference between the adjustable bias source and the array, and some current flows within this circuit through the array. The array current is fixed proportionally to the difference voltage by a resistance in series with the low-impedance voltage source (see Fig. 2).

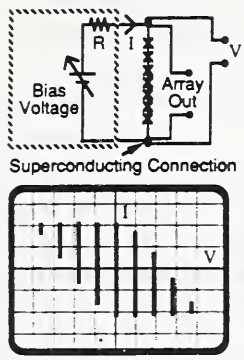


Fig. 2. Voltage biasing circuit. The display on the right shows a typical I–V plot at any array voltage. The top and bottom of the parallelogram envelope represent the critical current limits of the array steps. The diagonal sides represent the resistive load line for a constant bias voltage.

TABLE I
ESTIMATED ONE STANDARD DEVIATION SOURCES OF UNCERTAINTY IN THE ARRAY SYSTEM OPERATING AT 1.018 V

Sources of Uncertainty	Uncertainty (ppm)
(1) Random DVM measurement uncertainty	0.007
(2) Switch thermal emfs	0.005
(3) Leakage resistance	0.003
(4) Time base and frequency counting	0.0001
RSS subtotal	0.009
(5) Typical Zener reference noise and drift for the system	0.019
RSS Total	0.021

Sources (1) and (2) are actually constant in microvolt values and must be considered as limiting values for array voltages much less than 1 volt. Source (3) represents the systematic uncertainty from measurements of leakage resistance in the probe rf filters.

A high-valued resistor allows the array voltage to be many hundreds of steps away from the applied bias voltage, but a low-valued resistor effectively limits the voltage difference so the array voltage is always close to the bias voltage, thus simplifying step selection.

Fig. 2 shows the circuit configuration used for step selection, and the simulated oscilloscope screen plots the array current against the array voltage to show how the steps cross zero current. Several different steps are displayed by sweeping the bias voltage. The diagonal edge of the step envelope reveals the circuit load line which defines the array current values that are possible for a constant bias voltage. Based upon testing of this system, 5–10 Ω sufficiently limits the number of states (about 5–10) in which the array can exist at a single bias voltage. The voltage-to-current relation could be reduced to only one step per bias voltage setting by further decreasing the resistance. Unfortunately, the resistance necessary for this, about 3 Ω, induces array instabilities that are not completely understood.

IV. Sources of Uncertainty

The uncertainties in Table I are in parts-per-million at the 1.018-V level. We are mainly concerned with calibrating voltages at the level of standard cells, so we refer to these values for the rest of this paper. The total uncertainty of the array system as one part of the volt maintenance procedure is about 0.009 ppm. Note, however, that the extreme flexibility of the array allows measurements from 200 μV to ultimately 10 V, therefore, the uncertainty increases considerably for measurements smaller than 1 V and decreases for higher voltages. The largest contribution comes from the random uncertainties, which are mostly due to the DVM count error. Another fixed value contribution is caused by switch thermal EMF's. Although thermal EMF's drift minimally in the short measurement time, individual switch contact closures produce small random thermal EMF's that are difficult to eliminate.

The total uncertainty more appropriately includes the influence of the Zener references. The main purpose of this system is to maintain the unit of voltage, and the Zener reference group is an integral part of the transfer process. The volt is transferred from the Zener group to ten standard cells used as the primary reference group for daily workload calibrations. The pooled standard deviation of the mean value for the reference group is typically 0.004 2 ppm, where all the measurements are taken within one hour. However, the Zener references change randomly from the time of the first standard cell comparisons to the second set of comparisons about 3 h later. We estimate an uncertainty of 0.015 ppm should be included for these changes. Thus as a conservative estimate for the complete transfer, we assign 0.019 ppm for a one standard deviation estimate [8].

Fig. 3 shows the mean EMF of our primary cell group over time as measured by the array. The prediction of the primary group voltage is effectively limited by the day-to-day changes and variable seasonal effects (e.g., humidity changes) in the voltages which are significantly greater than the overall measurement uncertainty estimated for the array system. This implies that the limiting uncertainty in using the array to calibrate standard cells is the variations of the standard cells.

V. Array System Verification

The problem of making an accurate voltage measurement is strongly illustrated in this section, which reviews our efforts to verify the accuracy of the array system. Before adoption as the source of the U.S. Legal Volt, it was necessary to insure that the new array system had no unknown errors. Previous tests for differences between arrays resulted in no measurable effect to 5 parts in 10^{13}, and later to 2 parts in 10^{17} [9], [10]. However, this tested the universality of the Josephson frequency-to-voltage conversion ratio and applied only to the array chip itself. We conducted two different types of tests, an indirect comparison of the array to a conventional two-junction Josephson device at 10 mV, and a check of the ratio of the array output at 1 V and 10 mV using two room temperature voltage scaling instruments.

A brief description of the scaling instruments involved is necessary to understand these tests. Both scaling instruments contain resistance networks and power supplies designed to produce two voltages, 1.018 V and 10.18 mV where the 100:1 ratio can be self-calibrated. One of the

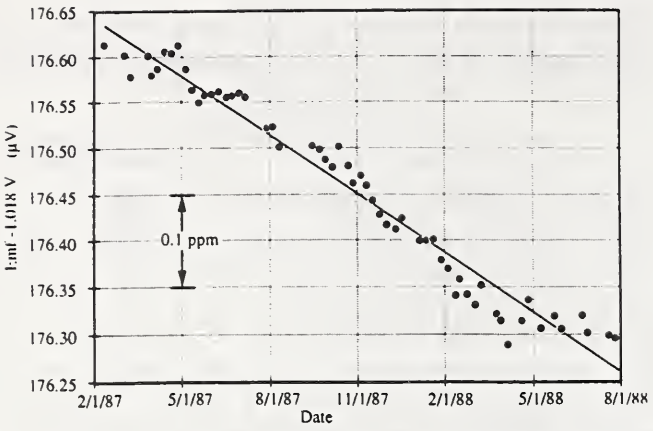


Fig. 3. History of the primary cell group since the adoption of the array based-voltage measurement system. The solid line is the linear fit to all of the points.

instruments, a series-parallel comparator (SPC), was used for maintaining the U.S. Legal Volt from 1972 to 1987 [11]. It contains two resistance networks, each of ten equal resistors with one network connected in series while the other is connected in parallel for a 100:1 voltage ratio. The second divider instrument, the Cascade-Interchange Comparator (CIC), had originally been used to check the scaling accuracy of the SPC [12]. The CIC basically consists of two strings of resistors to provide a 100:1 ratio but they can also be compared to each other using a series of Wheatstone bridges for a complete self-calibration. Thorough checks for possible systematic error sources in both instruments were performed, the estimated totals being 2.8 and 3.1 parts in 10^8 for the SPC and CIC, respectively. For the CIC, the operating procedure requires at least six h to complete.

Initially we calibrated a Zener reference with both the array system at 1.018 V and the established volt maintenance system based upon a two-junction Josephson device. To our surprise, we observed a 0.129-ppm difference. We later substituted the array chip (10.18 mV) and its millimeter wave source for the equivalent section of the older apparatus. This produced a difference of 0.165. This indirect comparison implied that an array at the 10-mV level agreed with a conventionally-biased Josephson device to -0.036 ± 0.05 ppm.

It was still necessary to resolve the 1-V level differences to verify the scaling of the array and/or the scaling instruments. For these tests a Zener reference was calibrated with the array at 1.018 V and also with the SPC using the array at 10.18 mV. All aspects of the SPC instrument were checked before the array comparisons: individual resistor matching, low-thermal solder connections, leakage resistance, temperature stability, tetrahedral junction errors, and switch contact integrity. Eventually a thermal offset voltage was located in one of the switch contacts of the SPC, which was producing the unanticipated error. We measured this offset at various times as 0.14–0.18 μV and incorporated it as a correction to the SPC comparisons. Complete comparisons were performed three separate times over more than a year, each measurement taking more than 3 h. The difference between the two systems (Array-SPC) in scaling to the 1-V level was -0.009 ± 0.031 ppm, as shown in Fig. 4. Error bars represent the pooled random and estimated systematic uncertainties (1σ). The CIC checks and array comparisons were similar to those involving the SPC but taking even longer, up to 6 h. The CIC/Array comparison (CIC-Array) resulted in a difference of 0.003 ± 0.069 ppm. The uncertainty assigned to this value is high because only one test was conducted due to its greater difficulty.

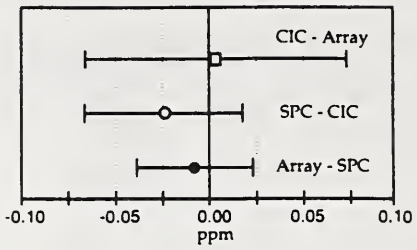


Fig. 4. Average differences obtained from the comparisons between the array, SPC, and CIC systems. The error bars represent the combined random and estimated systematic uncertainties of each system.

Additional experiments directly checked the 100:1 ratios of the SPC and CIC by comparing the high and low voltage sides of each instrument. Fig. 4 summarizes the average result from three SPC-CIC comparisons, -0.023 ± 0.043 ppm. The assigned uncertainty includes the systematic uncertainties of both the CIC and the SPC. The uncertainties in the SPC and CIC obviously limit all these verification tests of the array system. However, the overall agreement among these comparisons is clearly quite satisfactory and we can safely conclude that the array system uncertainty, relative to the previous U.S. Legal Volt system, is within 0.031 ppm at the 1-V level.

VI. Conclusions

The array system has unquestionably improved the accuracy and speed with which our volt calibrations are made. As expected we found no measurable errors in the voltage produced by the array but, it is quite possible to assemble a measurement system that does have systematic errors arising from ground loops, leakage resistance, and switch and cable thermal EMF's. Since it is necessary to test each new array system, development of a transportable array system for on-site verification of new array systems is desirable.

After experimenting with nine array devices of different configurations, we have seen that some arrays work well and are easy to use, but problems can arise in even the best arrays. At least three of these arrays have deteriorated to the point where steps are no longer stable enough to measure. Some arrays can suddenly form steps either with a resistive slope or with voltages which are not at clearly defined integer values. Strong RF interference can also induce array instability even through the filtering circuitry, a problem requiring array system operation within an RF shielded room. It is not easy to operate a measurement system, even manually, when these different behaviors are intermittently present, and accuracy cannot be

taken for granted. Furthermore, because the coupling resonances of some devices do not overlap with the narrow frequency range typical of Gunn-type millimeter wave sources, array devices are not necessarily interchangeable in an existing array system.

Fully automating the system requires replacement of the two sets of manually operated low-thermal switches. We need to build an automated switch network as commercial switches typically have internal thermal EMF's much greater than our measurement uncertainty. Further enhancements will rely on characterizing array properties to lead to improvements in step selection and an understanding of step anomalies, so that a completely automated measurement procedure is reliable as well as accurate.

Acknowledgment

The authors wish to especially thank C. Hamilton for his work in designing the array, bias plug-in, and cyrogenic probe, and F. Lloyd for fabricating the array devices. They also thank B. N. Taylor for his continued support and encouragement.

References

[1] B. F. Field, T. N. Finnegan, and J. Toots, "Volt maintenance at NBS via $2e/h$: A new definition of the NBS Volt," *Metrologia*, vol. 9, pp. 155-166, Oct. 1973.

[2] M. T. Levinsen, R. Y. Chiao, M. J. Feldman, and B. A. Tucker, "An inverse ac Josephson effect voltage standard," *Appl. Phys. Lett.*, vol. 32, pp. 776-778, Dec. 1977.

[3] R. L. Kautz and G. Costabile, "A Josephson voltage standard using a series array of 100 junctions," *IEEE Trans. Magn.*, vol. MAG-17, pp. 780-783, Jan. 1981.

[4] R. L. Kautz, "On a proposed Josephson-effect voltage standard at zero current bias," *Appl. Phys. Lett.*, vol. 36, pp. 386-388, Mar. 1980.

[5] J. Niemeyer, J. H. Hinken, and R. L. Kautz, "Microwave induced constant voltage steps at one volt from a series array of Josephson junctions," *Appl. Phys. Lett.*, vol. 45, pp. 478-480, Aug. 1984.

[6] C. A. Hamilton, R. L. Kautz, F. L. Lloyd, R. L. Steiner, and B. F. Field, "The NBS Josephson array voltage standard," *IEEE Trans. Instrum. Meas.*, vol. IM-36, pp. 258-261, June 1987.

[7] R. L. Kautz, C. A. Hamilton, and F. L. Lloyd, "Series-array Josephson voltage standards," *IEEE Trans. Magn.*, vol. MAG-23, pp. 883-890, Mar. 1987.

[8] B. F. Field, "NBS Measurement services: Standard cell calibrations," Natl. Bur. Stand. Special Publication 250-24, Oct. 1987.

[9] J. Niemeyer, L. Grimm, C. A. Hamilton, and R. L. Steiner, "High precision measurement of a possible resistive slope of Josephson array voltage steps," *IEEE Electron Device Lett.*, vol. EDL-7, pp. 44-46, Jan. 1986.

[10] R. L. Kautz and F. L. Lloyd, "Precision of series-array Josephson voltage standards," *Appl. Phys. Lett.*, vol. 51, pp. 2043-2045, Dec. 1987.

[11] T. F. Finnegan and A. Denenstein, "High accuracy potentiometers for use with ten millivolt Josephson devices: I. Double series-parallel exchange comparator," *Rev.Sci. Instrum.*, vol. 44, pp. 944-953, Aug. 1973.

[12] ——, "High accuracy potentiometers for use with ten millivolt Josephson devices: II. Cascade-interchange comparator," *Rev. Sci. Instrum.*, vol. 45, pp. 735-741, June 1974.

CALIBRATION DESIGN AND STATISTICAL ANALYSIS

UNITED STATES DEPARTMENT OF COMMERCE
Alexander B. Trowbridge, Secretary
NATIONAL BUREAU OF STANDARDS • A. V. Astin, Director

TECHNICAL NOTE 430

ISSUED OCTOBER 9, 1967

Designs for Surveillance of the Volt Maintained by a Small Group of Saturated Standard Cells

W. G. Eicke
Electricity Division
Institute for Basic Standards
National Bureau of Standards
Washington, D.C. 20234

J. M. Cameron
Applied Mathematics Division
Institute for Basic Standards
National Bureau of Standards
Washington, D.C. 20234

For sale by the Superintendent of Documents, Government Printing Office
Washington, D.C., 20402 - Price 20 cents

FOREWORD

When a local standard such as that for electromotive force is maintained by a group of standards, procedures must be established to provide evidence that the group has maintained its original value. One also needs methods for the transfer of the value to test items that provide efficient use of measurement effort while monitoring the measurement process and providing information for updating the values of process parameters. Solutions to the more general problem of transferring the value from laboratory to laboratory and of maintaining agreement among laboratories depend on the existence of control within the laboratories.

This note is one of a number of contemplated reports having the general aim of providing methods for the surveillance of measurement processes with emphasis on the amount and kind of information needed for the estimation and control of the uncertainty in measurement.

August 1967

M. B. Wallenstein
Acting Director,
Institute for Basic
Standards

Designs for Surveillance of the Volt Maintained by a Small Group of Saturated Standard Cells

W. G. Eicke
Electricity Division
Electrochemistry Section

and

J. M. Cameron
Applied Mathematics Division
Statistical Engineering Laboratory

This technical note describes a procedure for maintaining surveillance over a small group of saturated standard cells. The measurement process is briefly discussed and the principle of left-right balance as a means of eliminating certain systematic errors is developed. Specific designs and their analysis for intercomparing 3, 4, 5 and 6 cells in a single temperature controlled environment are given. Procedures for setting up control charts on the appropriate parameters are given, and a technique is described for detecting certain types of systematic errors.

Key words: Control charts, experiment design, saturated standard cells, standard cells calibration, statistics, voltage standard.

I. INTRODUCTION

At the local level the primary standard of electromotive force is maintained by a group of saturated standard cells, the same type of cell used to maintain the National unit of electromotive force. Many laboratories use groups containing from 3 to 6 cells mounted in either a temperature controlled air or oil bath. The cells are in general calibrated by the National Bureau of Standards at periodic intervals and the mean emf of the group is assumed to remain constant between calibrations. Since such calibrations are done infrequently (at intervals of one year or more) some technique must be employed to maintain surveillance over the local unit between calibrations.

Starting with assigned values for each of the cells of the group as a given set of reference points, one can check on the relative stability of the cells by measuring differences among them. One could measure all possible differences and have equal precision in the knowledge of the values of all cells or one could pick a favorite cell and compare all others with this one (but this leads to high precision in the knowledge of the selected cell and relatively low precision in all others). For small groups it is quite practical to measure all possible differences, but as the group size increases the number of measurements would increase rapidly with N, where N is the group size. As N increases compromise schemes that lead to equal precision in the knowledge of each cell can be used. One such design is given in [1] for a group of 20 cells in which only 40 differences are measured (instead of 190 if all $N(N-1)/2$ were measured).

This note discusses methods for maintaining surveillance of groups containing three, four, five, or six cells in a single temperature controlled enclosure. The procedures suggested are designed to yield information on:

1) the stability of the differences in emf among the group,

2) the components of variability and dependence of the measurement process precision on environmental influences or procedural changes,

3) possible systematic errors and estimation of the accuracy of the process.

Furthermore, they tend to maximize the yield of useful information per measurement.

II. The Measurement Technique

The opposition method [3] is usually employed in the intercomparison of saturated standard cells. In this method the small difference between two cells connected in series opposition is measured using a suitable instrument. The instrument is usually a potentiometer designed for the measurement of very small emfs [2].

In the ideal situation the difference in emf as measured by the potentiometer is:

$$\Delta E = V_1 - V_2 \tag{1}$$

where V_1 and V_2 are the emfs of the two cells being compared.

However in the real situation there may be spurious emfs in the circuit. In general these can be classified into two categories;

1. Those emfs that remain constant, or relatively so, in relation to the interval over which a complete set of measurements is made.
2. Those emfs that vary rapidly (referenced to the interval over which a complete set of measurements is made).

If the emfs are of the second type they will have the effect of decreasing the precision of the process. On the other hand if they are of the first type they will have the effect of introducing a systematic error, thereby making the measurement

$$\Delta E = V_1 - V_2 + P \tag{2}$$

where P is the constant emf. It is possible to estimate P by taking a second measurement

$$\Delta E' = V_2 - V_1 + P \tag{3}$$

and summing the two

$$2P = \Delta E + \Delta E' \tag{4}$$

The difference between eqs. (2) and (3) gives

$$\Delta E - \Delta E' = 2(V_1 - V_2) \tag{5}$$

an estimate of $V_1 - V_2$ free of P. The pair of measurements (eqs. (2) and (3)) are said to be "left-right" balanced. That is if there is a positional effect it is balanced out of the final result. This technique is analogous to that used to eliminate the inequality of balance arms in precision weighing on a two pan equal arm balance. In order to designate the cell positions from the operational point of view they are frequently designated as unknown and reference: Relative to the input terminals of the measuring instrument they are as shown in fig. 1. In the next section the principle of "left-right" balance will be extended to groups containing three or more cells.

III. Designs for Groups of 3, 4, 5, or 6 Cells

Experimental designs of groups of 3, 4, 5, and 6 cells are given in Appendix A. The designs presented have been selected to be (1) efficient from the standpoint of the operators making the measurement (2) statistically efficient, in the sense of minimum standard deviation for the estimated cell values, and (3) relatively easy to analyze using conventional desk calculators. All of the analyses presented are the least square solutions for the associated design assuming that the sum of the differences from the mean of all is zero. For groups of 3, 4, 5, total "left-right" balance has been achieved and the estimate of the "left-right" effect is

$$\hat{P} = 1/n \sum_{i=1}^{n} y_i \tag{6}$$

where n is the number of measurements and y the observed difference in emf between two cells.* For the case of six cells left-right balance is achieved for only the first 12 measurements.

*In terms of the notation of fig. 1 y is the observed ΔE.

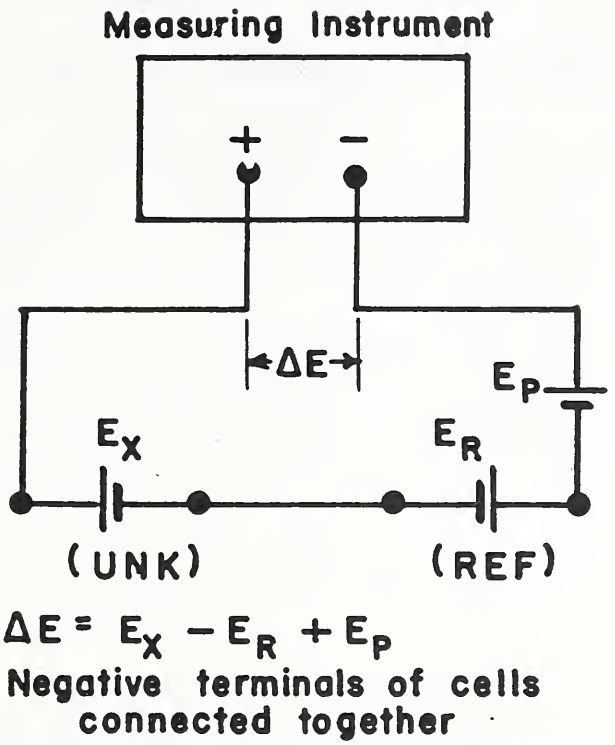


Figure 1

Two ways in which two cells can be connected in series-opposition.

For each size group the design and its analysis are given as a complete entity, with the left half giving the general procedure and the right half a numerical example. The suggested order for making the measurements requires moving one set of leads at a time thereby minimizing the possibility of connecting the wrong cells.

The definitions of the symbols used in the tables are as follows:

Symbol	Definition
V_i	The emf of the ith cell
M	The group mean
M*	The mean of
v_i	The difference V_i-M
y_i	The ith measured difference
$\hat{P}$	The calculated circuit residual
Q_i	A sum of those y's that involve the ith cell
$\hat{v}_i$	The calculated ith v
$\hat{y}_i$	The predicted y_i calculated from $\hat{v}$ and $\hat{P}$
d_i	Deviations, $(y_i - \hat{y}_i)$
s	The standard deviation of a single observation

For all of the designs given in Appendix A the assumption is that the mean of the whole group is known and serves as the restraint in the least square solution. In the next section, procedures for changing the restraint will be given. The analysis produces the following basic information which can be used to monitor the process:

1. The emf of each cell (or the difference from the group mean)
2. The residual emf, $\hat{P}$
3. The standard deviation of a single observation
4. The deviation of each observation from the predicted value

The frequency with which these intercomparisons should be run may vary considerably depending on the particular installation. Once it is established that the process is in a state of control then one intercomparison each week should be sufficient.

IV. Change of Restraint

In the previous section it was assumed that the mean for the whole group was known, such as would be the case if the group had been assigned values by the National Bureau of Standards. Because only differences in emf are measured, this average value is the restraint on the values which provide the "ground zero" to which the cell values are related.

When one or more of the cells show a change so large as to be inconsistent with its assigned value, it becomes necessary to remove these cells from the defining group. Evidence of such changes would be discovered from control charts on either the cell values or control charts on differences between cells (see the next section on control charts).

To illustrate, let us assume that the assigned values for the 5th and 6th cells of the example in Table A-4 had been 1.0182536 and 1.0182501 instead of the values (1.0182416, 1.0182381) given in

section 1 of the table. This is a change of +12.0μV in each cell so that the new emf values for the cells would be as shown in the following table.

Table 1-A

Cell	Assigned values	(Table A-4 Sec. 4) $\hat{v}_i$	Emf of cell $M + \hat{V}_i$	Difference from assigned
1	1.0182605	10.47	1.01826445	3.95
2	2655	15.62	1.01826960	4.10
3	2466	- 3.40	1.01825058	3.98
4	2476	- 2.29	1.01825169	4.09
5	2536*	- 8.37	1.01824561	-7.99
6	2501*	-12.04	1.01824194	-8.16

Average = 1.01825398 = M

*These values differ by 12μV from the data of Table A-4.

The last two cells are obviously inconsistent with their assigned values, so that one would want to remove these from the restraint and establish "ground zero" with the first four cells.

To do this one calculates

(1) $\bar{V}_A$: the average of the assigned value of the cells to be retained in the restraint as shown in column 2 of the table below;

(2) $\bar{\hat{v}}_i$: the average of the cell estimates in the restraint as shown in column 3 below; and

(3) adds $\bar{V}_A - \bar{\hat{v}}$ to each $\hat{v}_i$ to give the cell values, $\hat{V}_i$, as shown in column 4.

Table 1-B

Cell	Assigned values	(Table A-4 Sec. 4) $\hat{v}_i$	Emf of cell $\bar{V}_A - v + v_i = \hat{V}_i$
1	1.0182605	10.47	1.01826042
2	2655	15.62	26557
3	2466	- 3.40	24655
4	2476	- 2.29	24766
Average = 1.01825505 = V_A		$\bar{v}$ = 5.10	
5	1.0182536	- 8.37	24158
6	2501	-12.04	23792

The cell values are now expressed in terms of the average of the "good" cells as the reference point. The misbehaving cells would ordinarily continue to be measured in the hope that they would stabilize at some new value.

V. Control Charts

Control charts [5] [6] on process parameters such as the cell values and standard deviations of a single observation provide an effective means of determining whether or not the process is in a state of statistical control. Control charts for each cell (or difference between cells), process precision (standard deviation of an observation), and the residual emf P should be maintained. These charts provide the verification of that part of the uncertainty statement that deals with bounds for the effect of random error. Such statements say in effect "If this measurement process is used a large number of times, the values obtained for a single quantity will vary within the stated limits." The charts permit one to demonstrate the validity of such statements on current data.

For each run one will have values for each of the cells, the standard deviation, and the residual emf. To check on the state of control of the measurement process and on the stability of the cells, one would study the sequence of values for these parameters. Control charts on the cells can be established on the emf of the cells, the difference between successive cells (e.g. cell 1 - cell 2, cell 2 - cell 3, cell 3 - cell 4, etc.), or both. The former has the difficulty that it is not sensitive to a change in the emf of a single cell. However, by following the differences between successive cells (i.e. 1st minus 2nd, 2nd minus 3rd and so on) one has an easily interpreted set of results even though the successive differences are not independent. A single "bad" cell will show up as out of control on two successive differences, whereas the remaining differences are unaffected.

In order to establish control limits one has to know the precision of the measurement process (see discussion on measurement processes in ref. [7]). However, under the assumption that the standard deviation, σ, of the process is known one can, for a given design, write down the standard deviations of the individual cell emfs, the difference between two cells, and for the residual P. One can use three times the appropriate standard deviation as control limits. For the designs given the values for setting limits are shown in Table 2.

Unfortunately, when starting such surveillance the process precision, σ, is usually not known and must be estimated from the available data. In this case one would pool a number, m, of individual standard deviations using the formula

$$s_p^2 = 1/m \sum_{i=1}^{m} s_i^2 \qquad (7)$$

for a particular design. This might entail making several runs a week for the first month or so to obtain starting control limits. After about 100 degrees of freedom have been accumulated a new value of s_p should be calculated and the control limits revised. For such a large number of degrees of freedom s_p approaches σ very closely. Using s_p and the control limits from the factors in Table 2, a control chart on the standard deviation of the process (i.e. on s as computed in Appendix A) should be constructed.

It is also desirable to maintain a control chart on the residual emf using the limits given in Table 2. Initially the accepted value of P would be taken as zero. However, if after repeated measurements the value of P is other than zero and constant, the central value and control limits should be adjusted accordingly.

The start of each type of control chart is shown in Fig. 2. For the charts on the cells the central values for both cells and difference between cells should be based on the assigned values. The chart can either be kept on a run number or a time basis. The latter has the advantage that one can estimate rate of drift if any cell shows a trend.

Table 2

Factors for Setting the 3σ Control Limits for the Designs in Appendix A

No. of cells Excluded from group mean	Number of Cells in Group: 3	4	5	6
		Cell Values		
0	1.0000	0.9186	1.2000	1.1260
1*	0.8660 1.500	0.8660 1.2247	1.1619 1.5000	1.1071 1.3512
2*	0. 1.7320	0.750 1.2990	1.0954 1.5492	1.0794 1.3839
3*	--- ---	0 1.5000	.9487 1.6432	1.0000 1.4392
		Residual (P)		
	1.2247	0.8660	0.9487	0.8018
		Successive differences $(\hat{v}_i - \hat{v}_{i+1})$		
	1.7321	1.5000	1.8974	(1)
		Standard deviations		
Upper limit	1.945	1.585	1.737	1.552
Central line	0.888	0.950	0.933	0.963

*The upper figures are for those cells included in the mean and the lower figures are for those cells excluded from the mean.

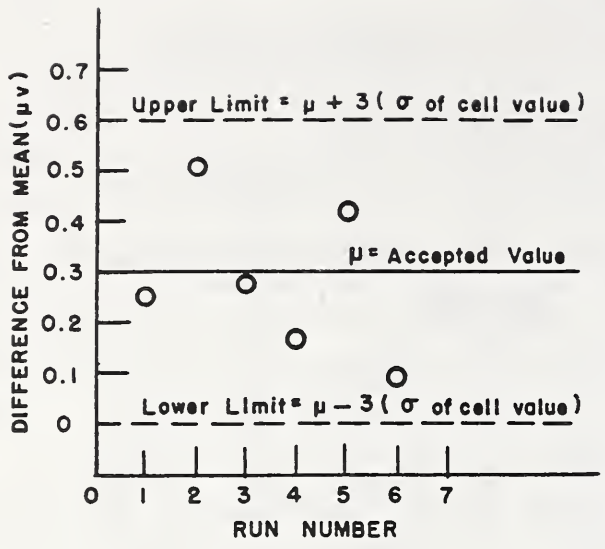
(1) For differences (1-2)(2-3)(4-5)(5-6) the limit is 1.7321.
For differences (4-5)(6-1) the limit is 1.7525.

To compute control limit multiply σ or pooled s_p by the appropriate factor and add or subtract as required.

If a cell should "go bad" and be removed from the mean, but still kept in the group, then the limits should be altered accordingly (see Table 2). It is important to bear in mind that the control charts on the cells only indicate change in the emf of cells relative to each other. If the whole group is changing it will not show up in any of the charts and can only be ascertained by comparison with other cells whose values are known. This situation does occur because small groups of cells are usually from the same manufacturer and lot, and therefore have similar aging characteristics.

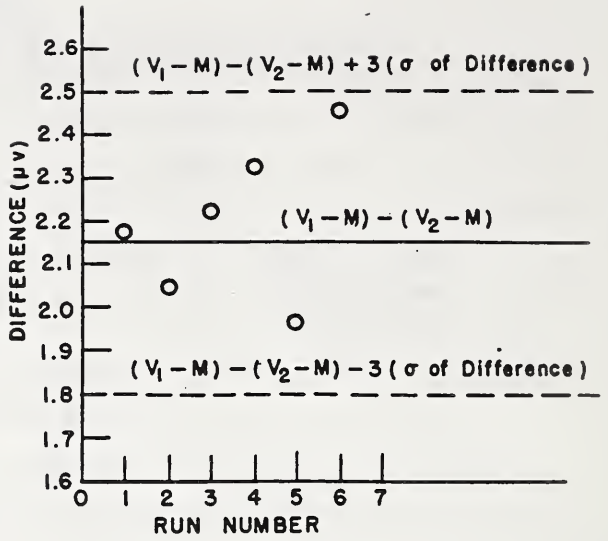
VI. Systematic Errors

Ideally a measurement process should be free of systematic error, however, this is not often the case. In fact, the residual P is a systematic error. Its effect on the values of the cells is readily removed and its magnitude estimated using the suggested designs. Other systematic errors are not so easily detected. Indications of their presence in some cases can be obtained by analysis of deviations from two or more successive designs run on the same group of cells.



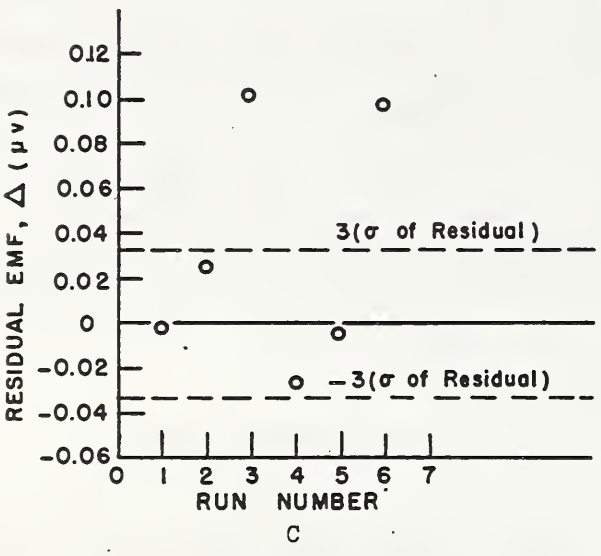
A

Control Chart for value of a cell.



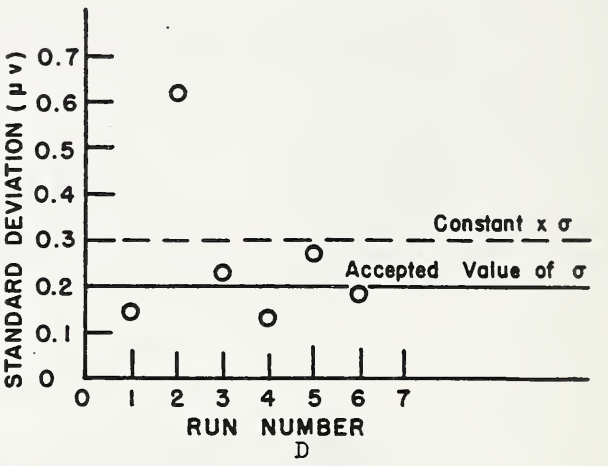
B

Control Chart for the difference between sucessive cells.



C

Control Chart for residual.



D

Control Chart for standard deviation of a single observation.

Figure 2

Typical control charts for maintaining surveillance over a group of standard cells.

Detection is based on the assumption that the deviations for a particular observation (cell 1 - cell 2, cell 1 - cell 3, etc.) are independent in successive runs.

If the magnitude and sign of corresponding deviations from successive runs tend to agree then one would suspect the presence of a systematic error. Such an analysis can be done on 5 and 6 cell groups graphically by plotting the deviations of one run as a function of the second run. If the deviations tend to fall on a straight line having slope 1 and passing through the origin, then one would suspect a systematic error. If there is none then the points would be distributed randomly. Figure 3 shows an example with no systematic error present. Figure 4 was created from the data of Figure 3 by adding 0.3 μV to the absolute value of each observation to simulate an offset error such as failing to correct for the instrument zero. The presence of such a systematic error will (1) cause the deviations to string along the line, (2) inflate the standard deviation, and (3), introduce a bias into each calculated $\hat{v}_i$. The magnitude of the latter will depend on the particular set of observations. Instead of the model for a single observation being

$$E(y_{ij}) = X_i - X_j + P$$

as in the case of Figure 3, and as assumed in the appendix, it is now

$$E(y_{ij}) = X_i - X_j + P + \frac{y_{ij}}{|y_{ij}|} C$$

where C is the zero offset.

For sets with less than 5 cells one would examine the deviations of successive runs for patterns. If the deviations for a given observation have the same sign and approximately the same magnitude, one would suspect a possible systematic error. Studies are being conducted to develop relatively simple tests that may be used to detect the presence of many types of systematic errors.

The cause or causes of systematic errors will depend on a particular measuring system. Some possible causes are:

1. Failure to make applicable corrections
2. Zero offset
3. Operator reading bias
4. Operator setting bias
5. Leakage currents

This list is by no means complete, but merely suggests some possible causes.

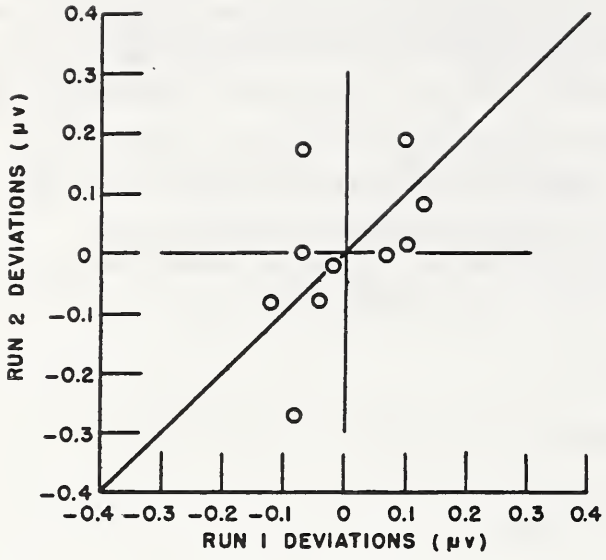


Figure 3

Youden plot for two runs without systematic error.

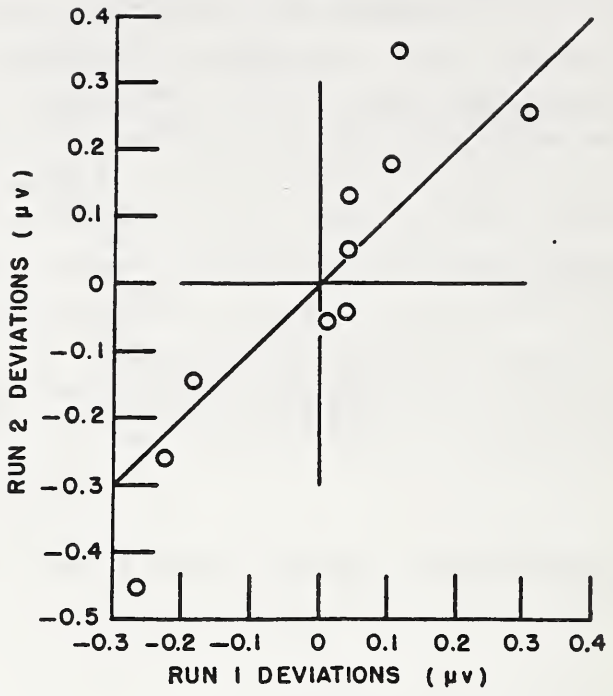


Figure 4

Youden plot for two runs with systematic error.

REFERENCES

[1] Hamer, W. J., "The Volt Standard" Moves To Gaithersburg, Maryland, J. Wash. Acad. of Sci. 56 (1966), pp. 101-108.

[2] Harris, F. K., Electrical Measurements, J. Wiley & Sons (1952), pp. 168-185.

[3] Hamer, W. J., Standard Cells--Their Construction, Maintenance, and Characteristics, NBS Monograph 84 (1965), pp. 8-10.

[4] Hamer, W. J., ibid. pp. 5-8.

[5] ASTM Manual on Quality Control of Materials, Special Technical Publication 15-C, Part 3, January 1951.

[6] Natrella, M. G., Experimental Statistics, NBS Handbook 91 (1963), pp. 18-1, 18-4.

[7] Pontius, P. E., Measurement Philosophy of the Pilot Program for Mass Measurement, NBS Technical Note 288, May 1966.

APPENDIX A

Designs for Groups of 3, 4, 5 and 6 Saturated Standard Cells

APPENDIX A

Designs for Groups of 3, 4, 5 and 6 Saturated Standard Cells

TABLE A-1

THE INTERCOMPARISON OF A GROUP OF THREE SATURATED STANDARD CELLS

	Example
1. Given: The emfs of three saturated standard cells (V_1, V_2 and V_3) are assigned by calibrating them in terms of a known standard of electromotive force. The mean of the group is $M = 1/3 \sum_{i=1}^{3} V_i$ and the difference from the mean of each cell is $v_i = (V_i - M)$	1. From an NBS Report of Calibration: $V_1 = 1.0182571$ $v_1 = -1.2$ μV $V_2 = 1.0182535$ $v_2 = -4.8$ $V_3 = 1.0182643$ $v_3 = +6.0$ Mean = 1.0182583 sum = 0
2. Assuming that there is a small constant emf P associated with the measuring process then the expected value of a single observation is $E(y_i) = V_j - V_k + P$ $j \neq k$; for j and k = 1,2,3. For all possible values of j and k the following schedule of measurements is convenient and requires changing the connections to only one cell at a time.	2. Observations:

Measurement	Cell in UNK position*	Cell in REF position*	
y_1	1	2	$y_1 = +4.8$ μV
y_2	1	3	$y_2 = -6.6$
y_3	2	3	$y_3 = -10.6$
y_4	2	1	$y_4 = -3.4$
y_5	3	1	$y_5 = +7.4$
y_6	3	2	$y_6 = +10.4$

*See Fig. 1 for definition of positions.

3. Estimation of P: $\hat{P} = 1/6 \sum_{i=1}^{6} y_i$	3. $\hat{P} = 1/6$ (+4.8 -6.6 -10.6 -3.4 +7.4 +10.4) $\hat{P} = .333$ μV

TABLE A-1 Continued

4. Estimation of v_i: $\hat{v}_1 = \hat{V}_1 - M = 1/6\ (y_1 + y_2 - y_4 - y_5)$ $\hat{v}_2 = \hat{V}_2 - M = 1/6\ (-y_1 + y_3 + y_4 - y_6)$ $\hat{v}_3 = \hat{V}_3 - M = 1/6\ (-y_2 - y_3 + y_5 + y_6)$ Arithmetic Check = $\Sigma\ \hat{v}_i = 0$ (within round-off)	4. $\hat{v}_1 = 1/6\ (4.8 - 6.6 + 3.4 - 7.4) = -.967\ \mu V$ $\hat{v}_2 = 1/6\ (-4.8 - 10.6 - 3.4 - 10.4) = -4.867$ $\hat{v}_3 = 1/6\ (+6.6 + 10.6 + 7.4 + 10.4) = +5.833$ Check = $-.967 - 4.867 + 5.833 = -.001\ \mu V$
5. Calculation of $\hat{y}$'s, the predicted y's: $\hat{y}_1 = \hat{v}_1 - \hat{v}_2 + \hat{P}$ $\hat{y}_2 = \hat{v}_1 - \hat{v}_3 + \hat{P}$ $\hat{y}_3 = \hat{v}_2 - \hat{v}_3 + \hat{P}$ $\hat{y}_4 = \hat{v}_2 - \hat{v}_1 + \hat{P}$ $\hat{y}_5 = \hat{v}_3 - \hat{v}_1 + \hat{P}$ $\hat{y}_6 = \hat{v}_3 - \hat{v}_2 + \hat{P}$ Arithmetic Check = $\Sigma\ \hat{y}_i = 6\hat{P}$	5. $\hat{y}_1 = -.967 + 4.867 + .333 = +4.233\ \mu V$ $\hat{y}_2 = -.967 - 5.833 + .333 = -6.467$ $\hat{y}_3 = -4.867 - 5.833 + .333 = -10.367$ $\hat{y}_4 = -4.867 + .967 + .333 = -3.567$ $\hat{y}_5 = +5.833 + .967 + .333 = +7.133$ $\hat{y}_6 = +5.833 + 4.867 + .333 = +11.033$ Check = $\Sigma\ \hat{y}_i = 1.998$
6. Calculation of the deviations ($d_i = y_i - \hat{y}_i$): $d_1 = y_1 - \hat{y}_1$ $d_2 = y_2 - \hat{y}_2$ $d_3 = y_3 - \hat{y}_3$ $d_4 = y_4 - \hat{y}_4$ $d_5 = y_5 - \hat{y}_5$ $d_6 = y_6 - \hat{y}_6$ Check = $\sum_{i=1}^{6} d_i = 0$ (within round-off)	6. $d_1 = 4.8 - 4.233 = .567\ \mu V$ $d_2 = -6.6 + 6.467 = -.133$ $d_3 = -10.6 + 10.367 = -.233$ $d_4 = -3.4 + 3.567 = +.167$ $d_5 = 7.4 - 7.133 = +.267$ $d_6 = 10.4 - 11.033 = -.633$ $\Sigma d = .002$
7. The standard deviation of a single observation (s) is $s = \sqrt{\sum_{i=1}^{6} d_i^2}$ where 3 represents the number of degrees of freedom in this error estimate.	7. $s = \sqrt{\frac{.8933}{3}} \cong .55\ \mu V$
8. Emf values of the cells The emf's of the cells are calculated by restoring the mean value to give $\hat{V}_i = \hat{v}_i + M$	8. Mean (from Section 1): 1.01825830 $\hat{V}_1 = 1.01825830 - 0.00000097 = 1.01825733$ $\hat{V}_2 = 1.01825830 - 0.00000487 = 1.01825343$ $\hat{V}_3 = 1.01825830 + 0.00000583 = 1.01826413$

TABLE A-2

THE INTERCOMPARISON OF A GROUP OF 4 CELLS

	Example
1. Given: The emfs of four saturated standard cells (V_1, V_2, V_3 and V_4) are assigned by calibrating them in terms of a known standard of electromotive force. The mean of the group is $M = 1/4 \sum_{i=1}^{4} V_i$ and the difference from the mean of each cell is $v_i = (V_i - M)$	1. From an NBS Report of Calibration: $V_1 = 1.0182459$ $v_1 = -4.1\ \mu V$ $V_2 = 1.0182488$ $v_2 = -1.2$ $V_3 = 1.0182526$ $v_3 = +2.6$ $V_4 = 1.0182527$ $v_4 = +2.7$ Mean = 1.0182500 sum = 0
2. Assuming that there is a small constant emf P associated with the measuring process, the expected value of a single observation is $E(y_i) = V_j - V_k + P \quad j \neq k$; for j and k = 1,2,3,4. For all possible values of j and k the following schedule of measurements is convenient, and in most cases requires changing the connections to only one cell at a time.	2. Observations:

Measurement	Cell in UNK position	Cell in REF position	
y_1	1	2	$y_1 = -3.1\ \mu V$
y_2	1	3	$y_2 = -6.9$
y_3	2	3	$y_3 = -3.8$
y_4	2	4	$y_4 = -4.0$
y_5	3	4	$y_5 = -.4$
y_6	3	1	$y_6 = 6.3$
y_7	3	2	$y_7 = 3.3$
y_8	4	2	$y_8 = 3.4$
y_9	4	1	$y_9 = 6.4$
y_{10}	4	3	$y_{10} = -.2$
y_{11}	2	1	$y_{11} = 2.7$
y_{12}	1	4	$y_{12} = -7.0$

3. Estimation of P: $\hat{P} = 1/12 \sum_{i=1}^{12} y_i$	3. $\hat{P} = 1/12\ (-3.1 -6.9 -3.8 -4.0 -.4 +6.3 +3.3 +3.4 +6.4 -.2 +2.7 -7.0)$ $\hat{P} = -.275\ \mu V$

TABLE A-2 Continued

4. Estimation of v_i:

$$\hat{v}_1 = \hat{V}_1 - M = 1/8\,(y_1 + y_2 - y_6 - y_9 - y_{11} + y_{12})$$

$$\hat{v}_2 = \hat{V}_2 - M = 1/8\,(-y_1 + y_3 + y_4 - y_7 - y_8 + y_{11})$$

$$\hat{v}_3 = \hat{V}_3 - M = 1/8\,(-y_2 - y_3 + y_5 + y_6 + y_7 - y_{10})$$

$$\hat{v}_4 = \hat{V}_4 - M = 1/8\,(-y_4 - y_5 + y_8 + y_9 + y_{10} - y_{12})$$

Arithmetic Check = $\sum_1^5 \hat{v}_i = 0$ (within round-off)

4.

$\hat{v}_1 = 1/8\,(-3.1 - 6.9 - 6.3 - 6.4 - 2.7 - 7.0) = -4.05\ \mu V$

$\hat{v}_2 = 1/8\,(3.1 - 3.8 - 4.0 - 3.3 - 3.4 + 2.7) = -1.088$

$\hat{v}_3 = 1/8\,(6.9 + 3.8 - .4 + 6.3 + 3.3 + .2) = 2.512$

$\hat{v}_4 = 1/8\,(4.0 + .4 + 3.4 + 6.4 - .2 + 7.0) = 2.625$

Check = $-4.050 - 1.087 + 2.512 + 2.625 = 0.001$

5. Calculation of $\hat{y}$'s, the predicted y's:

$$\hat{y}_1 = \hat{v}_1 - \hat{v}_2 + \hat{P}$$
$$\hat{y}_2 = \hat{v}_1 - \hat{v}_3 + \hat{P}$$
$$\hat{y}_3 = \hat{v}_2 - \hat{v}_3 + \hat{P}$$
$$\hat{y}_4 = \hat{v}_2 - \hat{v}_4 + \hat{P}$$
$$\hat{y}_5 = \hat{v}_3 - \hat{v}_4 + \hat{P}$$
$$\hat{y}_6 = \hat{v}_3 - \hat{v}_1 + \hat{P}$$
$$\hat{y}_7 = \hat{v}_3 - \hat{v}_2 + \hat{P}$$
$$\hat{y}_8 = \hat{v}_4 - \hat{v}_2 + \hat{P}$$
$$\hat{y}_9 = \hat{v}_4 - \hat{v}_1 + \hat{P}$$
$$\hat{y}_{10} = \hat{v}_4 - \hat{v}_3 + \hat{P}$$
$$\hat{y}_{11} = \hat{v}_2 - \hat{v}_1 + \hat{P}$$
$$\hat{y}_{12} = \hat{v}_1 - \hat{v}_4 + \hat{P}$$

Arithmetic Check = $\sum \hat{y}_i = 12\hat{P}$

5.

$\hat{y}_1 = -4.050 + 1.088 - .275 = -3.237\ \mu V$
$\hat{y}_2 = -4.050 - 2.512 - .275 = -6.837$
$\hat{y}_3 = -1.088 - 2.512 - .275 = -3.875$
$\hat{y}_4 = -1.088 - 2.625 - .275 = -3.988$
$\hat{y}_5 = 2.512 - 2.625 - .275 = -0.388$
$\hat{y}_6 = 2.512 + 4.050 - .275 = 6.287$
$\hat{y}_7 = 2.512 + 1.088 - .275 = 3.325$
$\hat{y}_8 = 2.625 + 1.088 - .275 = 3.438$
$\hat{y}_9 = 2.625 + 4.050 - .275 = 6.400$
$\hat{y}_{10} = 2.625 - 2.512 - .275 = -0.162$
$\hat{y}_{11} = -1.088 + 4.050 - .275 = 2.687$
$\hat{y}_{12} = -4.050 - 2.625 - .275 = -6.950$

sum = $(-3.237 - 6.837 - 3.875 - 3.988 - 0.388 + 6.287 + 3.325 + 3.438 + 6.400 - 0.162 + 2.687 - 6.950) = -3.300$

6. Calculation of the deviations ($d_i = y_i - \hat{y}_i$):

$$d_1 = y_1 - \hat{y}_1$$
$$d_2 = y_2 - \hat{y}_2$$
$$d_3 = y_3 - \hat{y}_3$$
$$d_4 = y_4 - \hat{y}_4$$
$$d_5 = y_5 - \hat{y}_5$$
$$d_6 = y_6 - \hat{y}_6$$
$$d_7 = y_7 - \hat{y}_7$$
$$d_8 = y_8 - \hat{y}_8$$
$$d_9 = y_9 - \hat{y}_9$$
$$d_{10} = y_{10} - \hat{y}_{10}$$
$$d_{11} = y_{11} - \hat{y}_{11}$$
$$d_{12} = y_{12} - \hat{y}_{12}$$

Arithmetic Check = $\sum d_i = 0$ (within round-off)

6.

$d_1 = -3.1 + 3.237 = .137\ \mu V$
$d_2 = -6.9 + 6.837 = -.063$
$d_3 = -3.8 + 3.875 = .075$
$d_4 = -4.0 + 3.988 = -.012$
$d_5 = -0.4 + .388 = -.012$
$d_6 = 6.3 - 6.287 = .013$
$d_7 = 3.3 - 3.325 = -.025$
$d_8 = 3.4 - 3.438 = -.037$
$d_9 = 6.4 - 6.400 = 0$
$d_{10} = -.2 + .162 = -.038$
$d_{11} = 2.7 - 2.687 = .013$
$d_{12} = -7.0 + 6.950 = -.050$

$(.137 - .063 + .075 - .012 - .012 + .013 - .025 - .038 + 0 - .038 + .013 - .050) = 0$

7. The standard deviation of a single observation (s) is

$$s = \sqrt{\frac{\sum_{i=1}^{12} (d_i)^2}{8}}$$

where 8 represents the number of degrees of freedom in this error estimate.

7.

$$s = \sqrt{\frac{.035}{8}} = .066\ \mu V$$

8. Emf values of the cells

The emf's of the cells are calculated by restoring the mean value to give

$$\hat{V}_i = \hat{v}_i + M$$

8. Mean (from Section 1): 1.01825000

$\hat{V}_1 = 1.01825000 - 0.00000405 = 1.01824595$
$\hat{V}_2 = 1.01825000 - 0.00000109 = 1.01824891$
$\hat{V}_3 = 1.01825000 + 0.00000251 = 1.01825251$
$\hat{V}_4 = 1.01825000 + 0.00000262 = 1.01825262$

TABLE A-3

THE INTERCOMPARISON OF A GROUP OF 5 CELLS

	Example
1. Given: The emfs of five saturated standard cells (V_1, V_2, V_3, V_4 and V_5) are assigned by calibrating them in terms of a known standard of electromotive force. The mean of the group is $M = 1/5 \sum_{i=1}^{4} V_i$ and the difference from the mean of each cell is $v_i = (V_i - M)$	1. From an NBS Report of Calibration: V_1 = 1.0182538 v_1 = .8 μV V_2 = 1.0182531 v_2 = .1 V_3 = 1.0182518 v_3 = -1.2 V_4 = 1.0182532 v_4 = .2 V_5 = 1.0182531 v_5 = .1 Mean = 1.0182530 sum = 0
2. Assuming that there is a small constant emf P associated with the measuring process then the expected value of a single observation is $E(y_i) = V_j - V_k + P \quad j \neq k$; for j and k= 1,2...5 For certain values of j and k a set of 10 measurements having "left-right" balance that is convenient and requires changing the connections to only one cell at a time is	2. Observations.

Measurement	Cell in UNK position	Cell in REF position	
1	1	2	y_1 = .5 μV
2	1	3	y_2 = 1.6
3	2	3	y_3 = .9
4	2	4	y_4 = -.4
5	3	4	y_5 = -1.5
6	3	5	y_6 = -1.3
7	4	5	y_7 = 0
8	4	1	y_8 = -.8
9	5	1	y_9 = -1.0
10	5	2	y_{10} = -.2

3. Estimation of P: $\hat{P} = 1/10 \sum_{i=1}^{10} y_i$	3. $\hat{P}$ = 1/10 (.5 +1.6 +.9 -.4 -1.5 -1.3 +0. -.8 -1.0 -.2) $\hat{P}$ = -.22 μV
4. Estimation of v_i: $\hat{v}_1 = \hat{V}_1 - M = 1/5 (y_1 + y_2 - y_8 - y_9)$ $\hat{v}_2 = \hat{V}_2 - M = 1/5 (-y_1 + y_3 + y_4 - y_{10})$ $\hat{v}_3 = \hat{V}_3 - M = 1/5 (-y_2 - y_3 + y_5 + y_6)$ $\hat{v}_4 = \hat{V}_4 - M = 1/5 (-y_4 - y_5 + y_7 + y_8)$ $\hat{v}_5 = \hat{V}_5 - M = 1/5 (-y_6 - y_7 + y_9 + y_{10})$ Arithmetic Check = $\sum_{i=1}^{5} V = 0$ (within round-off)	4. $\hat{v}_1$ = 1/5 (.5 +1.6 +.8 +1.0) = .78 μV $\hat{v}_2$ = 1/5 (-.5 +.9 -.4 +.2) = .04 $\hat{v}_3$ = 1/5 (-1.6 -.9 -1.5 -1.3) = -1.06 $\hat{v}_4$ = 1/5 (+.4 +1.5 +0. -.8) = .22 $\hat{v}_5$ = 1/5 (1.3 +0. -1.0 -.2) = 0.02 Check: $\sum_{i=1}^{5} \hat{v}_i$ = .78 +.04 -1.06 +.22 +.02= 0

TABLE A-3 Continued

5. Calculation of $\hat{y}$'s, the predicted y's:

$$\hat{y}_1 = \hat{v}_1 - \hat{v}_2 + \hat{P}$$
$$\hat{y}_2 = \hat{v}_1 - \hat{v}_3 + \hat{P}$$
$$\hat{y}_3 = \hat{v}_2 - \hat{v}_3 + \hat{P}$$
$$\hat{y}_4 = \hat{v}_2 - \hat{v}_4 + \hat{P}$$
$$\hat{y}_5 = \hat{v}_3 - \hat{v}_4 + \hat{P}$$
$$\hat{y}_6 = \hat{v}_3 - \hat{v}_5 + \hat{P}$$
$$\hat{y}_7 = \hat{v}_4 - \hat{v}_5 + \hat{P}$$
$$\hat{y}_8 = \hat{v}_4 - \hat{v}_1 + \hat{P}$$
$$\hat{y}_9 = \hat{v}_5 - \hat{v}_1 + \hat{P}$$
$$\hat{y}_{10} = \hat{v}_5 - \hat{v}_2 + \hat{P}$$

Arithmetic Check: $\Sigma \hat{y}_i = 10\hat{P}$

5.

$\hat{y}_1 = .78 - .04 - .22 = .52$ μV
$\hat{y}_2 = .78 + 1.06 - .22 = 1.62$
$\hat{y}_3 = .04 + 1.06 - .22 = .88$
$\hat{y}_4 = .04 - .22 - .22 = -.4$
$\hat{y}_5 = -1.06 - .22 - .22 = -1.50$
$\hat{y}_6 = -1.06 - .02 - .22 = -1.30$
$\hat{y}_7 = .22 - .02 - .22 = -.02$
$\hat{y}_8 = .22 - .78 - .22 = -.78$
$\hat{y}_9 = .02 - .78 - .22 = -.98$
$\hat{y}_{10} = .02 - .04 - .22 = -.24$

Check: $\Sigma \hat{y} = -2.20 = 10(-.22) = -2.20$

6. Calculation of the deviations ($d_i = y_i - \hat{y}_i$):

$$d_1 = y_1 - \hat{y}_1$$
$$d_2 = y_2 - \hat{y}_2$$
$$d_3 = y_3 - \hat{y}_3$$
$$d_4 = y_4 - \hat{y}_4$$
$$d_5 = y_5 - \hat{y}_5$$
$$d_6 = y_6 - \hat{y}_6$$
$$d_7 = y_7 - \hat{y}_7$$
$$d_8 = y_8 - \hat{y}_8$$
$$d_9 = y_9 - \hat{y}_9$$
$$d_{10} = y_{10} - \hat{y}_{10}$$

Check = $\sum_{i=1}^{10} d_i = 0$ (within round-off)

6.

$d_1 = .5 - .52 = -.02$ μV
$d_2 = 1.6 - 1.62 = -.02$
$d_3 = .9 - .88 = .02$
$d_4 = -.4 + .4 = 0$
$d_5 = -1.5 + 1.50 = 0$
$d_6 = -1.3 + 1.30 = 0$
$d_7 = 0. + .02 = .02$
$d_8 = -.8 + .78 = -.02$
$d_9 = -1.0 + .98 = -.02$
$d_{10} = -.2 + .24 = .04$

Check = $(-.02 - .02 + .02 + .3(0) + .02 - .02 - .02 + .04) = 0$

7. The standard deviation of a single observation (s) is

$$s = \sqrt{\frac{\sum_{i-1}^{10} d_i}{5}}$$

where 5 represents the number of degrees of freedom in this error estimate.

7.

$$s = \sqrt{\frac{.004}{5}} = .028 \text{ μV}$$

8. Emf values of the cells

The emf's of the cells are calculated by restoring the mean value to give

$$\hat{V}_i = \hat{v}_i + M$$

8. Mean (from Section 1): 1.01825300

$\hat{V}_1 = 1.01825300 + 0.00000078 = 1.01825378$
$\hat{V}_2 = 1.01825300 + 0.00000004 = 1.01825304$
$\hat{V}_3 = 1.01825300 - 0.00000106 = 1.01825194$
$\hat{V}_4 = 1.01825300 + 0.00000022 = 1.01825322$
$\hat{V}_5 = 1.01825300 + 0.00000002 = 1.01825302$

TABLE A-4

THE INTERCOMPARISON OF A GROUP OF 6 CELLS

1. Given: The emfs of six saturated standard cells (V_1, V_2, V_3, V_4, V_5 and V_6) are assigned by calibrating them in terms of a known standard of electromotive force. The mean of the group is

$$M = 1/6 \sum_{i=1}^{6} V_i$$

and the difference from the mean of each cell is

$$v_i = (V_i - M)$$

Example

1. From an NBS Report of Calibration:

V_1 = 1.0182605	v_1 = 10.52
V_2 = 1.0182655	v_2 = 15.52
V_3 = 1.0182466	v_3 = -3.38
V_4 = 1.0182476	v_4 = -2.38
V_5 = 1.0182416	v_5 = -8.38
V_6 = 1.0182381	v_6 = -11.88

Mean = 1.01824998 sum = 0

2. Assuming that there is a small constant emf P associated with the measuring process, the expected value of a single observation is

$E(y_i) = V_j - V_k + P \quad j \neq k$; for j and k = 1,2...6

For certain values of j and k a set of 15 measurements, 12 of which are "left-right" balanced, can be made.

Measurement	Cell in UNK position	Cell in REF position
1	1	2
2	1	3
3	2	3
4	2	4
5	3	4
6	3	5 —
7	4	5 —
8	4	6 —
9	5 —	6 —
10	5 —	1
11	6 —	1
12	6 —	2
13	1	4
14	2	5 —
15	3	6 —

2. Observations:

y_1 = -5.4 μV
y_2 = 13.7
y_3 = 18.8
y_4 = 17.7
y_5 = -1.3
y_6 = 4.8
y_7 = 5.9
y_8 = 9.5
y_9 = 3.5
y_{10} = -19.1
y_{11} = -22.7
y_{12} = -27.9
y_{13} = 12.5
y_{14} = 23.7
y_{15} = 8.4

3. Create a set of sums Q_i, S and T:

$Q_1 = (y_1 + y_2 - y_{10} - y_{11} + y_{13})$
$Q_2 = (-y_1 + y_3 + y_4 - y_{12} + y_{14})$
$Q_3 = (-y_2 - y_3 + y_5 + y_6 + y_{15})$
$Q_4 = (-y_4 - y_5 + y_7 + y_8 - y_{13})$
$Q_5 = (-y_6 - y_7 + y_9 + y_{10} - y_{14})$
$Q_6 = (-y_8 - y_9 + y_{11} + y_{12} - y_{15})$
$S = Q_1 + Q_2 + Q_3$

T = sum of all measurements

Arithmetic Check = $\sum_{i=1}^{6} Q_i = 0$

3.

$Q_1 = (-5.4 + 13.7 + 19.1 + 22.7 + 12.5) = +62.6$
$Q_2 = (5.4 + 18.8 + 17.7 + 27.9 + 23.7) = +93.5$
$Q_3 = (-13.7 - 18.8 - 1.3 + 4.8 + 8.4) = -20.6$
$Q_4 = (-17.7 + 1.3 + 5.9 + 9.5 - 12.5) = -13.5$
$Q_5 = (-4.8 - 5.9 + 3.5 - 19.1 - 23.7) = -50.0$
$Q_6 = (-9.5 - 3.5 - 22.7 - 27.9 - 8.4) = -72.0$
$S = (62.6 + 93.5 - 20.6) = 135.5$

T = (-5.4 +13.7 +18.8 +17.7 -1.3 +4.8 +5.9 +9.5 +3.5 -19.1 -22.7 -27.9 +12.5 +23.7 +8.4) = 42.1

Check = (62.6 +93.5 -20.6 -13.5 -50.0 -72.0) = 0

4. Calculate $\hat{P}$:

$$\hat{P} = \frac{3T - S}{42}$$

4.

$$\hat{P} = \frac{126.3 - 135.5}{42} = -.219 \ \mu V$$

TABLE A-4 Continued

5. Calculate $\hat{v}$'s:

$$\hat{v}_1 = 1/6\ (Q_1 - \hat{P})$$
$$\hat{v}_2 = 1/6\ (Q_2 - \hat{P})$$
$$\hat{v}_3 = 1/6\ (Q_3 - \hat{P})$$
$$\hat{v}_4 = 1/6\ (Q_4 + \hat{P})$$
$$\hat{v}_5 = 1/6\ (Q_5 + \hat{P})$$
$$\hat{v}_6 = 1/6\ (Q_6 + \hat{P})$$

Check = $\sum_{i=1}^{6} \hat{v}_i = 0$

5.

$\hat{v}_1 = 1/6\ (62.6 + .219) = 10.470\ \mu V$
$\hat{v}_2 = 1/6\ (93.5 + .219) = 15.620$
$\hat{v}_3 = 1/6\ (-20.6 + .219) = -3.397$
$\hat{v}_4 = 1/6\ (-13.5 - .219) = -2.286$
$\hat{v}_5 = 1/6\ (-50.0 - .219) = -8.370$
$\hat{v}_6 = 1/6\ (-72.0 - .219) = -12.036$

Check = 1/6 (62.819 +93.719 -20.819 -13.781 -50.219 -72.219) = 0

6. Calculation of $\hat{y}$'s, the predicted y's:

$$\hat{y}_1 = \hat{v}_1 - \hat{v}_2 + \hat{P}$$
$$\hat{y}_2 = \hat{v}_1 - \hat{v}_3 + \hat{P}$$
$$\hat{y}_3 = \hat{v}_2 - \hat{v}_3 + \hat{P}$$
$$\hat{y}_4 = \hat{v}_2 - \hat{v}_4 + \hat{P}$$
$$\hat{y}_5 = \hat{v}_3 - \hat{v}_4 + \hat{P}$$
$$\hat{y}_6 = \hat{v}_3 - \hat{v}_5 + \hat{P}$$
$$\hat{y}_7 = \hat{v}_4 - \hat{v}_5 + \hat{P}$$
$$\hat{y}_8 = \hat{v}_4 - \hat{v}_6 + \hat{P}$$
$$\hat{y}_9 = \hat{v}_5 - \hat{v}_6 + \hat{P}$$
$$\hat{y}_{10} = \hat{v}_5 - \hat{v}_1 + \hat{P}$$
$$\hat{y}_{11} = \hat{v}_6 - \hat{v}_1 + \hat{P}$$
$$\hat{y}_{12} = \hat{v}_6 - \hat{v}_2 + \hat{P}$$
$$\hat{y}_{13} = \hat{v}_1 - \hat{v}_4 + \hat{P}$$
$$\hat{y}_{14} = \hat{v}_2 - \hat{v}_5 + \hat{P}$$
$$\hat{y}_{15} = \hat{v}_3 - \hat{v}_6 + \hat{P}$$

6.

$\hat{y}_1 = 10.470 - 15.620 - .219 = -5.369$
$\hat{y}_2 = 10.470 + 3.397 - .219 = 13.648$
$\hat{y}_3 = 15.620 + 3.397 - .219 = 18.798$
$\hat{y}_4 = 15.620 + 2.286 - .219 - 17.687$
$\hat{y}_5 = -3.397 + 2.286 - .219 = -1.330$
$\hat{y}_6 = -3.397 + 8.370 - .219 = 4.754$
$\hat{y}_7 = -2.286 + 8.370 - .219 = 5.865$
$\hat{y}_8 = -2.286 + 12.036 - .219 = 9.531$
$\hat{y}_9 = -8.370 + 12.036 - .219 = 3.447$
$\hat{y}_{10} = -8.370 - 10.470 - .219 = -19.059$
$\hat{y}_{11} = -12.036 - 10.470 - .219 = -22.725$
$\hat{y}_{12} = -12.036 - 15.620 - .219 = -27.875$
$\hat{y}_{13} = 10.470 + 2.286 - .219 = 12.537$
$\hat{y}_{14} = 15.620 + 8.370 - .219 = 23.771$
$\hat{y}_{15} = -3.397 + 12.036 - .219 = 8.420$

7. Calculation of the deviations ($d_i = y_i - \hat{y}_i$):

$$d_1 = y_1 - \hat{y}_1$$
$$d_2 = y_2 - \hat{y}_2$$
$$d_3 = y_3 - \hat{y}_3$$
$$d_4 = y_4 - \hat{y}_4$$
$$d_5 = y_5 - \hat{y}_5$$
$$d_6 = y_6 - \hat{y}_6$$
$$d_7 = y_7 - \hat{y}_7$$
$$d_8 = y_8 - \hat{y}_8$$
$$d_9 = y_9 - \hat{y}_9$$
$$d_{10} = y_{10} - \hat{y}_{10}$$
$$d_{11} = y_{11} - \hat{y}_{11}$$
$$d_{12} = y_{12} - \hat{y}_{12}$$
$$d_{13} = y_{13} - \hat{y}_{13}$$
$$d_{14} = y_{14} - \hat{y}_{14}$$
$$d_{15} = y_{15} - \hat{y}_{15}$$

7.

$d_1 = -5.4 + 5.369 = -.031\ \mu V$
$d_2 = 13.7 - 13.648 = .052$
$d_3 = 18.8 - 18.798 = .002$
$d_4 = 17.7 - 17.687 = .013$
$d_5 = -1.3 + 1.330 = .030$
$d_6 = 4.8 - 4.754 = .046$
$d_7 = 5.9 - 5.865 = .035$
$d_8 = 9.5 - 9.531 = -.031$
$d_9 = 3.5 - 3.447 = .053$
$d_{10} = -19.1 + 19.059 = -.041$
$d_{11} = -22.7 + 22.725 = .025$
$d_{12} = -27.9 + 27.875 = -.025$
$d_{13} = 12.5 - 12.537 = -.037$
$d_{14} = 23.7 - 23.771 = -.071$
$d_{15} = 8.4 - 8.420 = -.020$

8. The standard deviation of a single observation (s) is

$$s = \sqrt{\frac{\Sigma(d_i)^2}{9}}$$

where 9 represents the degrees of freedom for error.

8.

$$s = \sqrt{\frac{.021590}{9}} = .0490\ \mu V$$

9. Emf values of the cells

The emf's of the cells are calculated by restoring the mean value to give

$$\hat{V}_i = \hat{v}_i + M$$

9. Mean (from Section 1): 1.01824998

$\hat{V}_1 = 1.01824998 + 0.00001047 = 1.01826045$
$\hat{V}_2 = 1.01824998 + 0.00001562 = 1.01826560$
$\hat{V}_3 = 1.01824998 - 0.00000340 = 1.01824658$
$\hat{V}_4 = 1.01824998 - 0.00000229 = 1.01824769$
$\hat{V}_5 = 1.01824998 - 0.00000837 = 1.01824161$
$\hat{V}_6 = 1.01824998 - 0.00001204 = 1.01823794$

Designs for the Calibration of Small Groups of Standards in the Presence of Drift

Joseph M. Cameron and
Geraldine E. Hailes

Institute for Basic Standards
National Bureau of Standards
Washington, D.C. 20234

U.S. DEPARTMENT OF COMMERCE, Frederick B. Dent, *Secretary*
NATIONAL BUREAU OF STANDARDS, Richard W. Roberts, *Director*

Issued August 1974

DESIGNS FOR THE CALIBRATION OF SMALL GROUPS OF STANDARDS IN THE PRESENCE OF DRIFT

by

Joseph M. Cameron and Geraldine E. Hailes

The process of calibrating a small number of "unknown" standards relative to one or two reference standards involved determining differences among the group of objects. Drift, due most often to temperature effects, or a "left-right" polarity effect can bias both the values assigned to the objects and the estimate of the effect of random errors. This note presents schedules of measurements of differences that eliminate the bias from these sources in the assigned value and variances and at the same time gives estimates of the magnitude of these extraneous components. The use of these designs in measurement process control is discussed and a computer program in BASIC is presented.

Key Words: Calibration; calibration design; experiment design; instrumental drift; measurement process; statistical analysis; trend elimination

1. Introduction

In very few processes can the effect of time be ignored. Instability in the object being measured, inability to maintain constant conditions or procedures, and variations in the detector or comparator all contribute to changes with time. A number of approaches have been suggested for reducing or eliminating the effects of these temporal effects on the validity of one's measurements. One way is to make measurements far enough apart in time (usually with some formal randomization procedure to guarantee statistical independence of the measurements) that the cumulative effects from the various sources appear in the random error component. At the other extreme, one can go to great lengths to eliminate these time dependent effects by achieving better environmental control, better instruments, better procedures, etc. If the measurements are to be transferred, as with instrument calibrations, then the first procedure leads to error bounds in which the random error limits include a between-time component whereas the latter procedure suppresses such a component. Neither of these represent the conditions of use adequately.

A compromise consists of arranging the experiment under its normal conditions so that it is as nearly as possible free of time dependent effects. The classic example of this is afforded by the calibration of thermometers in a bath with a gradually rising temperature using the schedule whose structure is as follows for a standard S, and 4 unknowns, T_1, T_2, T_3, T_4.

$$S\ T_1\ T_2\ T_3\ T_4\ S\ T_4\ T_3\ T_2\ T_1\ S$$

If the measurements are evenly spread in time, then the average of the bath temperature for all thermometers are the same (see [3] for a discussion of this practice). A similar procedure has been followed in weighing where in the substitution method one measures in scale units

$$A,\ B,\ B+S,\ A+S$$

to obtain the difference A-B and the deflection corresponding to the sensitivity weight, S.

The calibration of a small number of "unknown" objects relative to one or two reference standards involves determining differences among the group of objects. Instrumental drift, due most often to temperature effects, or a "left-right" polarity effect can bias both the values assigned to the objects and the estimate of the effect of random errors. This note presents schedules of measurements of differences that eliminate the bias from these sources and at the same time gives estimates of the magnitude of these extraneous components. The use of these designs in measurement process control is discussed and a computer program in BASIC is presented in this report.

2. Measurement as a Process

A single isolated measurement, like a single event in history, is difficult to interpret unless it can be regarded as a part of a continuing process. When the measurement is looked upon as the output of a process--a production process whose output is the measured values--then one can attribute to the single measurement the properties of the process from which it arose (for a discussion of this approach, see Eisenhart [2]). Just as with any production process, the operating characteristics are determined by building some redundancy into the system. Redundancy is needed to assure oneself that he has indeed measured the sought after quantity, uncontaminated by extraneous factors related to the operator, instrument, environment, or other items.

Among the characteristics of the process are those associated with the ability to repeat a measurement both in the short term and in the long term. Repetitions made within a few hours, such as with designs having more observations than unknowns, usually exhibit less variation than those made at long time intervals. This additional long-term component of variance can be measured from the agreement among repeated measurements on the same quantity. In addition to these properties related to variability, one needs to incorporate checks on the systematic errors which may possibly affect the process, and, if possible, measurements that provide information as to the adequacy of the assumptions in the underlying physical model.

In calibration it is often convenient to measure a check standard along with the calibration of one or more unknowns. One thus has a value for monitoring the process that is on an equal footing with the unknowns. By tracking its long-run performance, one can determine not only the presence of components of variance, but also by recording ancillary information on environmental and other factors one can develop information for assessing the adequacy of the assumed physical model and for setting bounds to the effect from known sources of possible systematic error. This "check standard" need not be the value of a single item but may take the form of a difference between two such items or some linear combination of several.

The effect of some sources of systematic error can be eliminated by "balancing out" the effect by repeating the measurement of a difference, $(x - y)$ in the reverse order, $(y - x)$. Time dependent effects can be balanced out by using the techniques of this report. For others, it is sometimes possible to alter the conditions to levels of a factor beyond that known to have been in effect at the time of the measurement and to use the changes produced in the output at these extremes as a bounds to the effect of the factor.

In all cases one has to continually monitor the process output just as one does with an industrial production process if he is to have assurance that the calibrations are correct.

3. Substitution Weighing

Consider first the simple situation of scale deflections produced on a balance by adding weights A and B and a sensitivity weight S. One could use either of the following sequences.

Sequence 1	Sequence 2
A	A
A+S	B
B	B+S
B+S	A+S

In high precision work one invariably finds a change in balance response with time so that the value for the difference (A-B) will obviously be contaminated by whatever time effects exist for Sequence 1. If Sequence 2 is used, it may be represented as follows.

Quantity	Effect of Drift	Scale Divisions
A	-3Δ	x_1
B	$-\Delta$	x_2
B+S	Δ	x_3
A+S	3Δ	x_4

and the quantity

$$\frac{1}{2}(x_1 - x_2 - x_3 + x_4)$$

can be seen to give an unbiased value for (A-B) because the drift effect (a 2Δ change in scale reading between each observation) cancels out. The least squares values for A-B, S, and Δ in scale divisions are

$$\widehat{(A-B)} = \frac{1}{2}(x_1 - x_2 - x_3 + x_4)$$

$$\hat{S} = \frac{1}{2}(x_1 - 3x_2 + 3x_3 - x_4)$$

$$\hat{\Delta} = \frac{1}{4}(-x_1 + x_2 - x_3 + x_4)$$

4. Thermometry

At NBS, the calibration of liquid in glass thermometers is usually carried out in a controlled bath which is continually heated so as to give a slight temperature increase with time. The temperature of the bath is measured by resistance thermometry at the start, middle, and end of the run with the test thermometers being run once each in the first interval and once again in reverse order in the second. The time sequence for the resistance measurements R_1, R_2, R_3 and the two series of test thermometer values denoted by $T_1\ T_2 \cdots T_k$ are as follows:

$$R_1\ T_1\ T_2 \cdots T_k\ R_2\ T'_k \cdots T'_2\ T'_1\ R_3$$

If equal time intervals are maintained between readings of the test thermometers, then one would expect an increase, ΔT in temperature with each interval except perhaps the middle one in which the resistance thermometer reading, R_2, is made. The analysis of this form of data is given in Appendix A.

5. Polarimeter Data

In determining the optical rotation of a quartz control plate used as reference standards in polarimeters, one measures the voltage response of a synchronous detector as the angle is varied. However, the response, y, of the system has a nearly linear drift with the angle so that one can represent this drift effect relative to the centroid of the data as being either $\cdots -3\Delta, -2\Delta, -\Delta, 0, \Delta, 2\Delta, 3\Delta, \cdots$ with Δ being the increment to the response added in each time interval. [For even n it is convenient to use $\cdots -3\Delta, -\Delta, \Delta, 3\Delta \cdots$ or 2Δ increment per time interval.]

In the polarimeter experiment the response is a linear function of angle so that the observation becomes

$$y_i = \alpha + \beta x_i + (i - \frac{n+1}{2})\Delta + \text{random error}$$

where the x_i are evenly spaced deviations from the nominal angle, e.g., x = 0", 10", 20", 30", $\cdots$ If the usual estimate of α and β are to remain unbiased and unchanged in precision, then one must have

$$\Sigma x_i\ (i - \frac{n+1}{2}) = 0$$

so that the estimates are orthogonal to the drift in the detector. The following orderings have this property:

Measurement Number	n = 4		n = 5	
	Quantity to be Measured	Setting for Polarimeter	Quantity to be Measured	Setting for Polarimeter
1	$\alpha + 2\beta$	20"	$\alpha + \beta$	10"
2	α	0"	$\alpha + 4\beta$	40"
3	$\alpha + 3\beta$	30"	$\alpha + 2\beta$	20"
4	$\alpha + \beta$	10"	α	0"
5	---	---	$\alpha + 3\beta$	30"

6. Calibration Designs

The term calibration design has been applied [1] to experiments where only differences between nominally equal objects or groups of objects can be measured. Perhaps the simplest such experiment consists in measuring the differences between the two objects of the $n(n-1)/2$ distinct pairing that can be formed from n objects. Ordinarily the order in which these measurements are made is of no consequence. However, when the response of the comparator is time dependent, attention to the order is important if one wishes to minimize the effect of these changes. When this effect can be adequately represented by a linear drift, it is possible to balance out the effect by proper ordering of the observations. As with the polarimeter data, this drift can be represented by the series $\cdots -3\Delta, -2\Delta, -\Delta, 0, \Delta, 2\Delta, 3\Delta, \cdots$ if $n(n-1)/2$ is odd or by $\cdots -5\Delta, -3\Delta, -\Delta, \Delta, 3\Delta, 5\Delta \ldots$ if $n(n-1)/2$ is even.

For $n = 4$, $n(n-1)/2 = 6$, and it turns out that it is not possible to balance out the drift effect with 6 measurements. However, with 8 measurements the balance can be achieved by the following order, denoting the four objects by A, B, C, D.

Observation	Observation is a Measurement of				
	A	B	C	D	Δ
y_1	+	-	0	0	-7
y_2	-	0	0	+	-5
y_3	0	0	+	-	-3
y_4	0	+	-	0	-1
y_5	0	+	0	-	1
y_6	-	0	0	+	3
y_7	+	0	-	0	5
y_8	0	-	+	0	7

The notation used here, the plus and minus signs, indicate the items entering into the difference measurement. Thus, y_2 is a measurement of the difference (D-A).

To see how the drift effect is balanced out, consider item C which occurs in the third, fourth, seventh, and eighth observations. In the third and eighth observations the item occurs positively and the corresponding drift effects are -3Δ and 7Δ respectively. For the fourth and seventh observations, item C occurs negatively while the corresponding

drift effects are $-\Delta$ and 5Δ. The overall effect can be represented by the sum of cross products of the columns for C and Δ, namely

$$[1](-3\Delta) + [-1](-\Delta) + [-1]\ (5\Delta) + [1]\ (7\Delta) = 0$$

using square brackets to denote the coefficient attached to the direction of the difference and parenthesis for the drift effect. For A, one has

$$[1](-7\Delta) + [-1](-5\Delta) + [-1](3\Delta) + [1](5\Delta) = 0$$

In general, if the cross products sum out to zero, then the drift effect is said to be completely "balanced out" or "orthogonal" to the items being measured.

7. Restraints

In calibration designs only differences between items are measured so that unless one or more of them are standards for which values are known, one cannot assign values for the remaining "unknown" items. Algebraically, one has a system of equations that is not of full rank and needs the value for one item or the sum of several items as the restraint to lead to a unique solution.

In the design of Section 6, for example, if one has a single standard and three unknowns, the standard can be assigned to any one of the letters. (The same would be true of three standards and one unknown.) If there are two standards and two unknowns, the choice of which pair of letters to assign for the standards is important in terms of minimizing the uncertainty in the unknown.

It turns out that the pairing of A with D or of B with C is slightly less efficient (see Appendix B) than the other pairings A with B or C with D. This results from the fact that the observation on the differences (A-D) and (B-C) are repeated and it is usually better (to achieve smaller variance for the test items) to measure differences between standards and unknowns than between pairs of standards.

8. Use of Calibration Design in Gage Block Calibration

The calibration design of Section 6 is used in gage block calibration at the National Bureau of Standards and the analysis and interpretation of the design for this application is representative of the principles involves in the use of the design in other applications.

At NBS two master sets of gage blocks are maintained for transferring length calibration to users gage blocks, these are designated A and B and their sum is designated by K. These are combined with two sets of unknowns, designated C and D. The difference (A-B) is used as the check standard.

If we denote the values determined for A B C and D by $\hat{A}$ $\hat{B}$ $\hat{C}$ $\hat{D}$ in accordance with the statisticians' practice of distinguishing the value from the experiment from the sought-after or long-run value, we may then write

$$\hat{A} = \frac{1}{24}(5y_1 - 2y_2 - y_3 - 2y_4 - 3y_5 - 2y_6 + 3y_7 + 2y_8) + \frac{K}{2}$$

$$\hat{B} = \frac{1}{24}(-5y_1 + 2y_2 + y_3 + 2y_4 + 3y_5 + 2y_6 - 3y_7 - 2y_8) + \frac{K}{2}$$

$$\hat{A-B} = \frac{1}{24}(10y_1 - 4y_2 - 2y_3 - 4y_4 - 6y_5 - 4y_6 + 6y_7 + 4y_8)$$

$$\hat{C} = \frac{1}{24}(-y_1 + 2y_2 + 5y_3 - 6y_4 - y_5 + 2y_6 - 7y_7 + 6y_8) + \frac{K}{2}$$

$$\hat{D} = \frac{1}{24}(y_1 + 6y_2 - 5y_3 - 2y_4 - 7y_5 + 6y_6 - y_7 + 2y_8) + \frac{K}{2}$$

where $\hat{A} + \hat{B}$ necessarily sum to K.

These values have the following standard deviations in terms of the long run precision as represented by the process standard deviation σ.

$$\text{s.d. } (\hat{A}) = \text{s.d. } (\hat{B}) = \sigma \sqrt{\frac{5}{48}}$$

$$\text{s.d. } (\hat{A-B}) = \sigma \sqrt{\frac{5}{12}}$$

$$\text{s.d. } (\hat{C}) = \text{s.d. } (\hat{D}) = \sigma \sqrt{\frac{13}{48}}$$

One also obtains values $\hat{\Delta}$ for Δ where

$$\hat{\Delta} = \frac{1}{168}(-7y_1 - 5y_2 - 3y_3 - y_4 + y_5 + 3y_6 + 5y_7 + 7y_8)$$

$$\text{s.d. } (\hat{\Delta}) = \sigma \sqrt{\frac{1}{168}}$$

Because this is an overdetermined system (more observations than unknowns) the deviation between observed and computed value is, in general, different from zero and reflects the random errors of measurement. These deviations, d_1 d_2 $\cdot$ $\cdot$ $\cdot$ d_8 are as follows:

$$d_1 = \frac{1}{168}(49y_1 - 7y_2 - 7y_3 + 21y_4 + 49y_5 + 49y_6 - 7y_7 + 21y_8)$$

$$d_2 = \frac{1}{168}(-7y_1 + 87y_2 + 13y_3 - 5y_4 + 33y_5 - 41y_6 + 53y_7 + 35y_8)$$

$$d_3 = \frac{1}{168}(-7y_1 + 13y_2 + 89y_3 + 25y_4 - 39y_5 + 37y_6 + 57y_7 - 7y_8)$$

$$d_4 = \frac{1}{168}(21y_1 - 5y_2 + 25y_3 + 111y_4 - 27y_5 + 3y_6 - 23y_7 + 63y_8)$$

$$d_5 = \frac{1}{168}(49y_1 + 33y_2 - 39y_3 - 27y_4 + 97y_5 + 25y_6 + 9y_7 + 21y_8)$$

$$d_6 = \frac{1}{168}(49y_1 - 41y_2 + 37y_3 + 3y_4 + 25y_5 + 103y_6 + 13y_7 - 21y_8)$$

$$d_7 = \frac{1}{168}(-7y_1 + 53y_2 + 57y_3 - 23y_4 + 9y_5 + 13y_6 + 73y_7 - 7y_8)$$

$$d_8 = \frac{1}{168}(21y_1 + 35y_2 - 7y_3 + 63y_4 + 21y_5 - 21y_6 - 7y_7 + 63y_8)$$

These deviations provide the information needed to obtain a value, s, which is the experiment's value for the process standard deviation, σ.

$$s = \sqrt{\frac{\Sigma(\text{dev})^2}{4}} \qquad \text{degrees of freedom} = 4$$

The number of degrees of freedom results from taking the number of observations less the number of unknowns then adding one (for the restraint) to give 8 - 5 + 1 = 4.

9. Example

Routine calibration of gage blocks is carried out with two NBS master blocks (designated S. and S..) and two test blocks (designated X and Y). The blocks are placed close together on a metal platen for a sufficiently long time to insure temperature equilibrium. A mechanical intercomparator is used to determine the difference between the blocks by first determining a reading for the block indicated by "+" then following with the block indicated by "-". The difference between these two readings is the observation, y (all values are in micro-inch). For a set of 0.101 in. blocks, the following data was obtained.

DATA FROM NBS CALIBRATION OF FOUR 0.101 INCH GAGE BLOCKS

i	Schedule	Difference Measured	First Reading	Second Reading	Difference y(i)	Deviation
1	+ - 0 0	S.-S..	52.0	52.5	-0.5	0.029
2	- 0 0 +	Y-S.	45.2	52.1	-6.9	-0.046
3	0 0 + -	X-Y	50.0	45.1	4.9	0.113
4	0 + - 0	S..-X	53.1	50.0	3.1	0.571
5	0 + 0 -	S..-Y	52.3	45.2	7.1	-0.238
6	- 0 0 +	Y-S.	45.1	52.0	-6.9	-0.079
7	+ 0 - 0	S.-X	52.0	50.1	1.9	-0.154
8	0 - + 0	X - S..	50.1	52.3	-2.2	0.304

S.+S.. = 6.4 used as restraint

S.-S.. = -0.133 used as check standard

σ = .32 accepted standard deviation

The values for the blocks and the drift effect, Δ, are

$$\hat{S}. = \frac{1}{24}[5(-0.5) - 2(-6.9) - (4.9) - 2(3.1) - 3(7.1) - 2(-6.9) + 3(1.9) + 2(-2.2)] + \frac{(6.4)}{2}$$

$$= \frac{1}{24}(-6.0) + 3.2 = 2.9500$$

$$\hat{S}.. = \frac{1}{24}[-5(-0.5) + 2(-6.9) + (4.9) + 2(3.1) + 3(7.1) + 2(-6.9) - 3(1.9) - 2(-2.2)] + \frac{(6.4)}{2}$$

$$= \frac{1}{24}(6.0) + 3.2 = 3.4500$$

$$S_{\cdot}\hat{-}S_{\cdot\cdot} = -0.5$$

$$\hat{X} = \frac{1}{24}[-(-0.5) + 2(-6.9) + 5(4.9) - 6(3.1) - (7.1) + 2(-6.9) - 7(1.9) + 6(-2.2)] + \frac{(6.4)}{2}$$

$$= \frac{1}{24}(-54.8) + 3.2 = 0.9167$$

$$\hat{Y} = \frac{1}{24}[(-0.5) + 6(-6.9) - 5(4.9) - 2(3.1) - 7(7.1) + 6(-6.9) - (1.9) + 2(-2.2)] + \frac{(6.4)}{2}$$

$$= \frac{1}{24}(-170.0) + 3.2 = -3.8833$$

$$\hat{\Delta} = \frac{1}{168}[-7(-0.5) - 5(-6.9) - 3(-4.9) - 1(3.1) + (7.1) + 3(-6.9) + 5(1.9) + 7(-2.27)$$

$$= \frac{1}{168}(.7) = 0.0042$$

The accepted standard deviation for the process is 0.32 μ-in so that one can compare the observed standard deviation, s,

$$s = \sqrt{\Sigma dev^2/4} = \sqrt{\frac{.5208}{4}} = 0.361$$

to the accepted value by computing

$$F = \left(\frac{s}{\sigma}\right)^2 = \frac{0.1302}{0.1024} = 1.27$$

Had the ratio $(s/\sigma)^2$ exceeded 3.32 (the critical value for the 1% probability level of the F distribution), then the measurements would be regarded as being "out of control" and would be repeated. The other check on process performance is provided by the check standard for which the difference between $(S_{\cdot}\hat{-}S_{\cdot\cdot})$ and its accepted value should be less than 3 times the standard deviation of $(S_{\cdot}\hat{-}S_{\cdot\cdot})$. See Section 10 for a discussion of this test.

The drift term, $\hat{\Delta}$, has a standard deviation of $\sigma/\sqrt{168}$ or 0.025. The statistical significance of $\hat{\Delta}$ can be judged by forming the ratio $\frac{\hat{\Delta}}{\sigma/\sqrt{168}}$.

If this ratio exceeds 3, then $\hat{\Delta}$ would be regarded as significant. However, because the design has eliminated the effect of drift on the

values of the blocks, one would not be concerned about a "significant" $\hat{\Delta}$ unless it was greatly in excess of previously encountered values.

The deviations are computed as shown in Section 8, for example, for the deviation corresponding to y_8 is given by

$$(\text{dev})_8 = \frac{1}{168}[21(-0.5) + 35(-6.9) - 7(4.9) + 63(3.1) + 21(7.1) - 21(-6.9) -$$

$$7(1.9) + 63\ (-2.2)]$$

$$= \frac{1}{168}[51.1] = 0.304$$

10. Process Control

As mentioned in Section 2, continued monitoring of the measurement process is required to assure that predictions based on the accepted values for process parameters are still valid. For gage block calibration, the process is monitored for precision by comparison of the observed standard deviation to the accepted value, σ_w, by means of the F-test. In the case of the design of Section 6, the square of the ratio of the two standard deviations is compared to the critical value, $F(4, \infty, \alpha)$, which is the α probability point of the F distribution for degrees of freedom 4 and ∞. [For calibrations at NBS, α is chosen as .01 to give $F(4, \infty, .01) = 3.32$].

The check for systematic error is given by comparison of the observed value of the difference, S. - S.., between the two standards with its accepted value. The uncertainty of this difference is given by $\sigma_T = \sqrt{(5/12)\sigma_w^2 + 2\sigma_B^2}$ where σ_w is the "within run" standard deviation and σ_B is the component of variance arising from variations from run-to-run. The value of σ_T is obtained directly from the sequence of values of S. - S.. arising in regular calibrations. The check standard test is therefore,

$$t = \frac{\left|\text{observed } (S. - S..) - \text{accepted } (S. - S..)\right|}{\sigma_T} < 3$$

i.e., t is compared to the critical value 3.0 which would correspond to the .003 probability level for the normal distribution.

If both the "precision" (F-test) and "accuracy" (t-test) criteria are satisfied, the process is regarded as being "in control" and values for the unknowns, X and Y, and their associated uncertainties are regarded as valid. Failure on either criterion is an "out-of-control" signal and the measurements are repeated.

When the between run component, σ_B, is present, the standard deviation associated with the values for the unknowns are given by *

$$\sigma(S.) = \sigma(S..) = \sqrt{\frac{5}{48}\sigma_w^2 + \frac{1}{2}\sigma_B^2} = \frac{1}{2}\sigma_T$$

$$\sigma(X) = \sigma(Y) = \sqrt{\frac{13}{48}\sigma_w^2 + \frac{3}{2}\sigma_B^2} = \sqrt{\frac{3}{4}\sigma_T^2 - \frac{2}{48}\sigma_w^2}$$

The value for the drift serves as an indicator of possible trouble if it changes markedly from its usual range of values. However, because any linear drift is balanced out, a change in the value does not of itself vitiate the results.

*See M.C. Croarkin, "An Extended Error Model for Comparison Calibration" for an explanation.

If the uncertainty attached to the restraint value is not negligible, this will lead to a possible systematic error in all measurements based on this restraint. Therefore, as a bound to this error one should, for the design of section 6, add to the uncertainty from random error an allowance of one-half the uncertainty in the sum ($S_i + S_j$). This is shown in the computer example.

11. Computer Program

Appendix C lists a computer program in BASIC for carrying out the calculation for the gage block example. The program can be used with any design provided one has the arrays of coefficients for the determination of the values of the unknowns and the deviations corresponding to the two arrays given in Section 8 for the gage block example.

The program calls for input of:

a) Administrative data--designation of blocks, operator, date, etc.

b) Process parameters--standard deviations, value for check standard, etc.

c) Comparator readings

The computer programs provide in the output:

a) Deviations, s.d.

b) Values for unknowns, drift, and associated undertainties.

c) Statistical tests as to whether process can be regarded as "in control": on standard deviation and on value of check standard.

12. Other Designs for Elimination of Drift With Order of Measurement

The number of observations over which a linear drift could be expected to be valid varies with the type of measurement, but experience indicates that it is unusual if it is as large as 20. For all distinct pairings of n items $n(n-1)/2$ exceeds 20 for $n \geq 7$. The table below gives designs for n = 5, 6, 7 which are balanced for linear drift.

n = 5 10 Observations		n = 6 18 Observations		n = 7 21 Observations	
+ - 0 0 0	1-2	+ - 0 0 0 0	1-2	+ - 0 0 0 0 0	1-2
0 + - 0 0	2-3	0 + - 0 0 0	2-3	0 + - 0 0 0 0	2-3
0 0 + - 0	3-4	0 0 + - 0 0	3-4	0 0 + - 0 0 0	3-4
0 0 0 + -	4-5	0 0 0 + 0 -	4-6	0 0 0 + - 0 0	4-5
- 0 0 0 +	5-1	0 0 0 0 - +	6-5	0 0 0 0 + - 0	5-6
		- 0 0 0 + 0	5-1	0 0 0 0 0 + -	6-7
- 0 0 + 0	4-1			- 0 0 0 0 0 +	7-1
0 + 0 - 0	2-4	0 0 + 0 0 -	3-6		
0 - 0 0 +	5-2	- 0 0 + 0 0	4-1	0 + 0 - 0 0 0	2-4
0 0 + 0 -	3-5	0 + 0 0 - 0	2-5	0 0 + 0 - 0 0	3-5
+ 0 - 0 0	1-3	0 0 - 0 0 +	6-3	0 0 0 + 0 - 0	4-6
		+ 0 0 - 0 0	1-4	0 0 0 0 + 0 -	5-7
		0 - 0 0 + 0	5-2	- 0 0 0 0 + 0	6-1
				0 - 0 0 0 0 +	7-2
		0 0 0 - + 0	5-4	+ 0 - 0 0 0 0	1-3
		0 0 - 0 0 +	6-3		
		0 - 0 + 0 0	4-2	0 0 - 0 0 0 +	7-3
		- 0 + 0 0 0	3-1	0 - 0 0 0 + 0	6-2
		0 + 0 0 - 0	2-5	- 0 0 0 + 0 0	5-1
		+ 0 0 0 0 -	1-6	0 0 0 + 0 0 -	4-7
				0 0 + 0 0 - 0	3-6
				0 + 0 0 - 0 0	2-5
				+ 0 0 - 0 0 0	1-4

An alternate form of displaying the design is shown for n = 5 and is used for the other designs.

Designs that involve a subset of all possible pairings are given below:

n = 6 12 Observ.	n = 7 14 Observ.	n = 8 16 Observ.	n = 9 18 Observ.
1-2	1-2	1-2	1-2
5-1	2-3	2-3	2-3
2-3	3-4	3-4	6-5
4-6	4-5	4-5	3-1
3-4	5-6	5-6	5-4
6-5	6-7	6-7	8-9
	7-1	7-8	4-7
2-4		8-1	9-6
4-5	3-1		7-8
6-2	5-3	4-1	
3-1	7-5	7-4	7-1
1-6	2-7	2-7	4-6
5-3	4-2	5-2	9-7
	6-4	8-5	1-4
	1-6	3-8	3-9
		6-3	5-8
		1-6	6-3
			2-5
			8-2

REFERENCES

1. Bose, R. C. & Cameron, J. M., The Bridge Tournament Problem and Calibration Designs for Comparing Pairs of Objects, NBS J. of Res. 69B (1965), 323-332.

2. Eisenhart, C., Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems, NBS J. of Res. 67C (1963), 161-187.

3. Swindells, J. F., Calibration of Liquid-in-Glass Thermometers, NBS Monograph 90, GPO, 1965.

4. Zelen, M., Linear Estimation and Related Topics, Chapter 17 of Survey of Numerical Analysis edited by J. Todd, McGraw Hill, New York City (1962), 558-584.

APPENDIX A

Thermometer Calibration

Liquid in glass thermometers are calibrated at NBS in a controlled bath in which the temperature is increasing in a nearly linear fashion with time. The temperature of the bath is measured by platinum resistance thermometry at the beginning, middle, and end of a run with the test thermometers being read once in the first interval and again in reverse order in the second interval. The time sequence for the resistance measurements, R_1, R_2, R_3 and the two series of thermometer values denoted by T'_i and T_i are as follows:

$$R_1\ T'_1\ T'_2 \cdots T'_k\ R_2\ T_k \cdots T_2\ T_1\ R_3$$

with uniform time intervals between the thermometer readings. Figure shows a schematic of the situation with the increment to the bath temperature being Δ for each time period except for the middle reading involving resistance thermometry where a step of α in temperature is assumed.

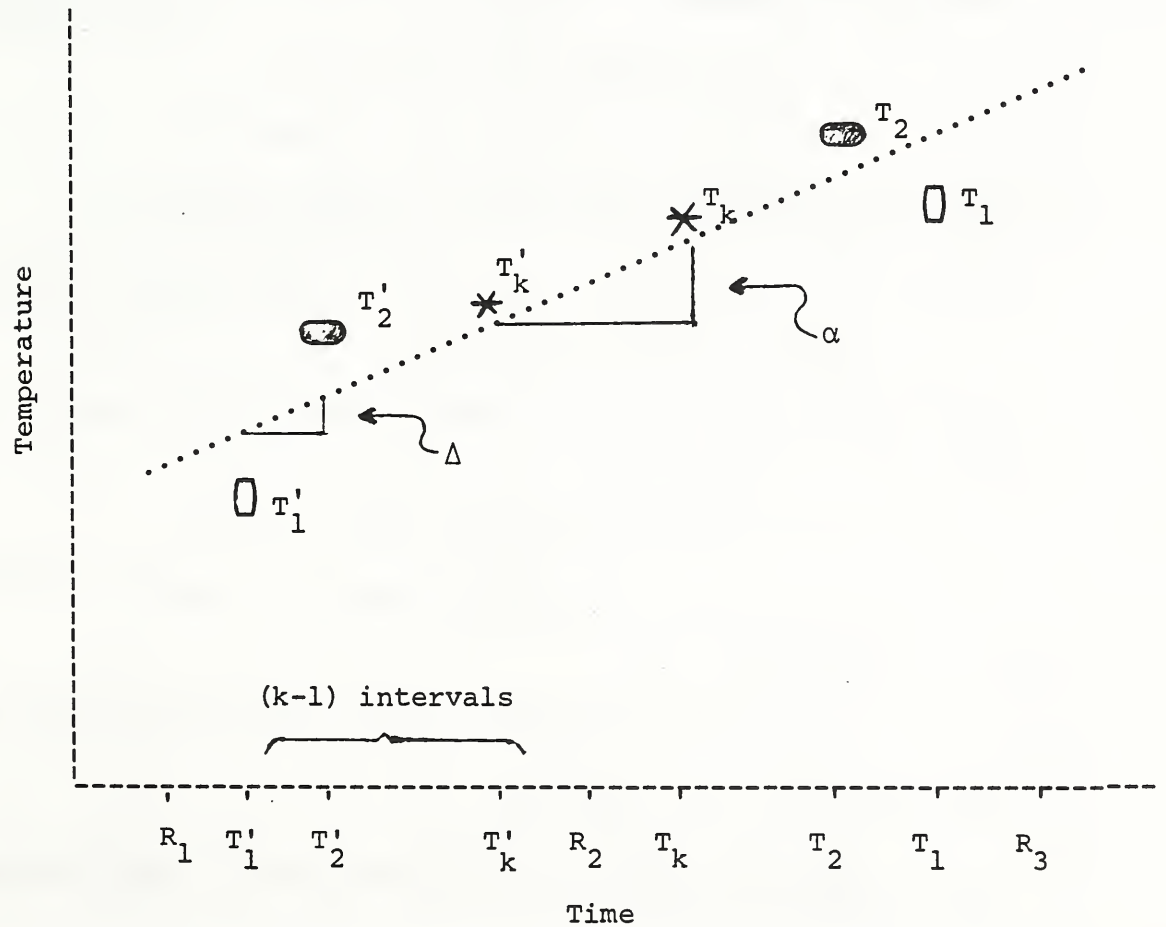


Figure Thermometer reading at fixed time intervals in a bath with linear drift.

The average $(T_i' + T_i)/2$ will be the indication of the i-th thermometer at the temperature implied by $(R_1 + R_2 + R_3)/3$. The differences, $d_i = T_i - T_i'$ will be a measure of $\alpha + 2(k - i)\Delta$ so that the observational equations may be written

$$E(d) = E\begin{bmatrix} T_1 - T_1' \\ T_2 - T_2' \\ \cdot \\ \cdot \\ T_{k-1} - T_{k-1}' \\ T_k - T_k' \end{bmatrix} = \begin{bmatrix} \alpha + 2(k-1)\Delta \\ \alpha + 2(k-2)\Delta \\ \cdot \\ \cdot \\ \alpha + 2\Delta \\ \alpha \end{bmatrix} = \begin{bmatrix} 1 & 2(k-1) \\ 1 & 2(k-2) \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \Delta \end{bmatrix} = X \begin{bmatrix} \alpha \\ \Delta \end{bmatrix}$$

where X stands for the indicated matrix, and where E( ) stands for the "expected value of," i.e., the limiting value if the effects of random error were eliminated.

The least squares estimates of α and Δ are given by the solution to the normal equations

$$(X'X)\begin{bmatrix} \alpha \\ \Delta \end{bmatrix} = X'd = \begin{bmatrix} \Sigma d \\ 2\Sigma d\ (k-i) \end{bmatrix}$$

where the inverse of the matrix of normal equations is

$$(X'X)^{-1} = \begin{bmatrix} k & k(k-1) \\ k(k-1) & 2k(k-1)(2k-1)/3 \end{bmatrix}^{-1} = \frac{1}{k(k^2-1)} \begin{bmatrix} 2(k-1)(2k-1) & -3(k-1) \\ -3(k-1) & 3 \end{bmatrix}$$

The estimates of α, Δ and σ^2, the variance of the observations are given by

$$\hat{\alpha} = \frac{2}{k(k+1)}\,[3\Sigma id - (k+1)\Sigma d]$$

$$\hat{\Delta} = \frac{3}{k(k^2-1)}\,[(k+1)\Sigma d - 2\Sigma id]$$

$$\hat{\sigma}^2 = \frac{1}{k-2}\,[\Sigma d^2 - \hat{\alpha}\Sigma d - 2\hat{\Delta}\Sigma(k-i)d] = \frac{1}{k-2}\,\Sigma(dev)^2$$

where $dev_i = d_i - \hat{\alpha} - 2(k-i)\hat{\Delta}$.

The standard deviation of the value for the test thermometer is $\sigma/\sqrt{2}$ and for α and Δ,

$$\text{s.d. } (\alpha) = \sigma\sqrt{2(2k-1)/k(k+1)}$$

$$\text{s.d. } (\Delta) = \sigma\sqrt{3/k(k^2-1)}$$

Control on the measurement process is maintained by two forms of redundancy--one to check on the process average and the other to check on process variability. The first of these is provided by incorporating an NBS standard thermometer among the k thermometers and requiring that its value be within random error of its accepted value. The variability check is given by comparing $\hat{\sigma}$ with the long run value established for the process. When these conditions are satisfied, then one can regard the process as being in a state of control.

A typical set of data for this type of calibration is given in the following table. For simplicity the resistance measurements have been suppressed and the temperature reported directly.

Calibration of Thermometers
Data From NBS Calibration of 22 August 1972
Provided by J. Wise, NBS, Thermometry Section

Thermometer	Observation	Averages	PRT - OBS = Correction at 40°
Reference (PRT)	39.9378		
T_1'	39.983	T_1 39.9870	-0.0436
T_2'	39.913	T_2 39.9150	0.0284
T_3'	39.966	T_3 39.9675	-0.0241
T_4' (check standard)	39.840	T_4 39.8410	0.1024*
Reference (PRT)	39.9422	PRT 39.9434	
T_4	39.842		
T_3	39.969		
T_2	39.917		
T_1	39.991	*accepted value is 0.1000	
Reference PRT	39.9501		

i	$d = T_i' - T_i$	α	Δ		Predicted d	dev.
1	-0.008	1	6	$\Sigma d = -0.077$	-0.0071	-0.0009
2	-0.004	1	4	$\Sigma id = -0.033$	-0.0052	0.0012
3	-0.003	1	2	$2/k(k+1) = 1/10$	-0.0033	0.0003
4	-0.002	1	0	$3/k(k^2-1) = 1/20$	-0.0014	-0.0006

$$\Sigma dev^2 = 0.00000270$$

$$\hat{\alpha} = \frac{1}{10}\,[3(-0.033) - 5(-0.017)] = -0.00140$$

$$\hat{\Delta} = \frac{1}{20}\,[5(-0.017) - 2(-0.033)] = -0.00095$$

$$\hat{\sigma} = \sqrt{\Sigma dev^2/2} = \sqrt{0.00000135} = 0.00115$$

s.d. $(\hat{\alpha}) = \sigma\sqrt{\frac{7}{10}}$ s.d. (average T) $= \sigma/\sqrt{2}$

s.d. $(\hat{\Delta}) = \sigma\sqrt{\frac{1}{20}}$

APPENDIX B

Least Squares Analysis of Calibration Designs

In this appendix the least squares analysis is presented in matrix form for those wishing to prepare a general analysis. Each formal statement will be illustrated by its application to the calibration design of Section 6.

It is assumed that the expected value of the observations represented in vector form as $y' = (y_1\ y_2\ \ldots\ y_n)$ have expected values $E(y) = X\beta$ where β is the vector of parameters and X is the design matrix. It is also assumed that the errors of measurement are uncorrelated and have equal variance, i.e., $V(y) = \sigma^2 I$.

For the design of section 5,

$$X = \begin{pmatrix} 1 & -1 & 0 & 0 & -7 \\ -1 & 0 & 0 & 1 & -5 \\ 0 & 0 & 1 & -1 & -3 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 1 \\ -1 & 0 & 0 & 1 & 3 \\ 1 & 0 & -1 & 0 & 5 \\ 0 & -1 & 1 & 0 & 7 \end{pmatrix} \qquad \beta = \begin{pmatrix} A \\ B \\ C \\ D \\ \Delta \end{pmatrix}$$

The matrix of normal equations is given by $(X'X)\beta = X'y$ which for calibration designs is not of full rank.

$$X'X\beta = \begin{pmatrix} 4 & -1 & -1 & -2 & 0 \\ -1 & 4 & -2 & -1 & 0 \\ -1 & -2 & 4 & -1 & 0 \\ -2 & -1 & -1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 168 \end{pmatrix} \qquad \beta = X'y = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 \\ 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 \\ -7 & -5 & -3 & -1 & 1 & 3 & 5 & 7 \end{pmatrix} y$$

In order to solve this system, a restraint in the form $h'\beta = K$ is imposed leading to the augmented equations (see Zelen [4]),

$$\begin{pmatrix} X'X & h \\ h' & 0 \end{pmatrix} \begin{pmatrix} B \\ \lambda \end{pmatrix} = \begin{pmatrix} X'y \\ K \end{pmatrix}$$

For the design as used in the calibration of gage blocks the restraint is that A + B = K, giving h' = (1 1 0 0 0) and the augmented equations are

$$\begin{pmatrix} 4 & -1 & -1 & -2 & 0 & 1 \\ -1 & 4 & -2 & -1 & 0 & 1 \\ -1 & -2 & 4 & -1 & 0 & 0 \\ -2 & -1 & -1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 168 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \beta = \begin{pmatrix} X'y \\ K \end{pmatrix}$$

The solution for the parameter values $\hat{\beta}$ are

$$\begin{pmatrix} \hat{\beta} \\ \lambda \end{pmatrix} = \begin{pmatrix} X'X & h \\ h' & 0 \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ K \end{pmatrix} = \begin{pmatrix} C & g \\ g' & 0 \end{pmatrix} \begin{pmatrix} X'y \\ K \end{pmatrix}$$

where C is the indicated KxK matrix arising in the inversion process n. For the example

$$\begin{pmatrix} \hat{\beta} \\ \lambda \end{pmatrix} = \frac{1}{336} \begin{pmatrix} 35 & -35 & -7 & 7 & 0 & 168 \\ -35 & 35 & 7 & -7 & 0 & 168 \\ -7 & 7 & 91 & 21 & 0 & 168 \\ 7 & -7 & 21 & 91 & 0 & 168 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 168 & 168 & 168 & 168 & 0 & 0 \end{pmatrix} \begin{pmatrix} X'y \\ K \end{pmatrix} = \frac{1}{168} \begin{pmatrix} 35 & -14 & -7 & -14 & -21 & -14 & 21 & 14 & 84 \\ -35 & 14 & 7 & 14 & 21 & 14 & -21 & -14 & 84 \\ -7 & 14 & 35 & -42 & -7 & 14 & -49 & 42 & 84 \\ 7 & 42 & -35 & -14 & -49 & 42 & -7 & 14 & 84 \\ -7 & -5 & -3 & -1 & 1 & 3 & 5 & 7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ K \end{pmatrix}$$

for which $C = \frac{1}{336} \begin{pmatrix} 35 & -35 & -7 & 7 & 0 \\ -35 & 35 & 7 & -7 & 0 \\ -7 & 7 & 91 & 21 & 0 \\ 7 & -7 & 21 & 91 & 0 \\ 2 & 0 & 0 & 0 & 2 \end{pmatrix}$

The variances of the parameters are given by $C_{ii}\sigma^2$ and of linear functions, $\ell'\beta$, the variance is $\ell'C\ell\sigma^2$. For the example

$$V(\hat{A}) = V(\hat{B}) = C_{11}\sigma^2 = C_{22}\sigma^2 = 35\sigma^2/336 = 5\sigma^2/48$$

$$V(\hat{C}) = V(\hat{D}) = C_{33}\sigma^2 = C_{44}\sigma^2 = 91\sigma^2/336 = 13\sigma^2/48$$

$$V(\hat{A-B}) = (C_{11} + C_{22} - 2C_{12})\sigma^2 = 140\sigma^2/336 = 5\sigma^2/12$$

$$V(\hat{A+B}) = 0$$

$$V(\hat{C+D}) = (C_{33} + C_{44} + 2C_{34})\sigma^2 = 224\sigma^2/336 = 2\sigma^2/3$$

$$V(\hat{C-D}) = (C_{33} + C_{44} - 2C_{34})\sigma^2 = 140\sigma^2/336 = 5\sigma^2/24$$

$$V(\hat{\Delta}) = C_{55}\sigma^2 = 2\sigma^2/336 = \sigma^2/168$$

FOOTNOTE 1

If one had assigned the two standards to positions B and C instead of to A and B as was done, then one would be repeating the measurement of the difference (B-C), and of the difference (A-D). These differences are internal to the pair of standards over the pair of unknowns and one might suspect that they add little to the transfer from standard to test item. This is confirmed by examination of the inverse of the matrix of normal equations,

$$\begin{bmatrix} 4 & -1 & -1 & -2 & 0 & 0 \\ -1 & 4 & -2 & -1 & 0 & 1 \\ -1 & -2 & 4 & -1 & 0 & 1 \\ -2 & -1 & -1 & -14 & 0 & 0 \\ 0 & 0 & 0 & 0 & 168 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}^{-1} = \frac{1}{168} \begin{bmatrix} 56 & 0 & 0 & 28 & 0 & 84 \\ 0 & 14 & -14 & 0 & 0 & 84 \\ 0 & -14 & 14 & 0 & 0 & 84 \\ 28 & 0 & 0 & 56 & 0 & 84 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 84 & 84 & 84 & 84 & 0 & 0 \end{bmatrix}$$

The variances for the standards are

$$V(\hat{B}) = V(\hat{C}) = 14\sigma^2/168 = \sigma^2/12$$

$$V(\hat{B-C}) = 56\sigma^2/168 = \sigma^2/3$$

which is smaller than for the restraint A + B = K.

However, for the test items the variances are

$$V(\hat{A}) = V(\hat{D}) = 56\sigma^2/168 = \sigma^2/3$$

which is larger than that with the restraint A + B = K for which the corresponding variance is $13\sigma^2/48$.

The estimate for the test item, A, is

$$\hat{A} = \frac{1}{24}\left[8y_1 - 4y_2 - 4y_3 - 4y_5 - 4y_6 + 8y_7\right] + \frac{K}{2}$$

which does not involve y_4 and y_8 which are measurements of the difference between the two standards, i.e., of B-C. Thus, there is a gain in efficiency in the calibration of the test block by using positions A and B for the standards, the efficiency factor being $(\sigma^2/3)/(13\sigma^2/48) = 16/13$.

FOOTNOTE 2

If there were but a single standard, A, the inverse of the matrix of normal equations would be

$$\begin{pmatrix} 4 & -1 & -1 & -2 & 0 & 1 \\ -1 & 4 & -2 & -1 & 0 & 0 \\ -1 & -2 & 4 & -1 & 0 & 0 \\ -2 & -1 & -1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 168 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^{-1} = \frac{1}{168} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 168 \\ 0 & 70 & 42 & 28 & 0 & 168 \\ 0 & 42 & 70 & 28 & 0 & 168 \\ 0 & 28 & 28 & 56 & 0 & 168 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 168 & 168 & 168 & 168 & 0 & 0 \end{pmatrix}$$

The variances of the test items are

$$V(\hat{B}) = V(\hat{C}) = 70\sigma^2/168 = 5\sigma^2/12$$

$$V(\hat{D}) = 56\sigma^2/168 = \sigma^2/3$$

FOOTNOTE 3

If the sum of all four were taken as the restraint, the inverse of the matrix of normal equations would be

$$\begin{pmatrix} 4 & -1 & -1 & -2 & 0 & 1 \\ -1 & 4 & -2 & -1 & 0 & 1 \\ -1 & -2 & 4 & -1 & 0 & 1 \\ -2 & -1 & -1 & 4 & 0 & 1 \\ 0 & 0 & 0 & 0 & 168 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}^{-1} = \frac{1}{336} \begin{pmatrix} 49 & -21 & -21 & -7 & 0 & 84 \\ -21 & 49 & -7 & -21 & 0 & 84 \\ -21 & -7 & 49 & -21 & 0 & 84 \\ -7 & -21 & -21 & 49 & 0 & 84 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 84 & 84 & 84 & 84 & 0 & 0 \end{pmatrix}$$

The variances of all four test items are the same

$$V(\hat{A}) = V(\hat{B}) = V(\hat{C}) = V(\hat{D}) = 49\sigma^2/336 = 7\sigma^2/48$$

```
 5 ************G A G E    B L O C K    C A L I B R A T I O N***********
10 *    THIS PROGRAM COMPUTES THE VALUES OF THE UNKNOWN GAGE BLOCKS,
15 *    AND PERFORMS TWO STATISTICAL TESTS, THE F-TEST AND THE T-TEST,
20 *    TO DETERMINE IF THE PROCESS IS IN CONTROL.
25 ***************************************************************
30 *                              INPUT
35 *
40 *    THIS PROGRAM CALLS FOR A USER CREATED DATA FILE WHICH CONSISTS
45 *    OF THE FOLLOWING:
50 *    (1) (CS(I),I=1,3) - DATE,OBSERVER,INSTRUMENT
55 *    (2) K   - VALUE OF RESTRAINT(MICROINCHES)
60 *    (3) NO  - NOMINAL SIZE OF TEST BLOCKS(INCHES)
65 *    (4) S5  - ACCEPTED VALUE OF CHECK STANDARD(MICROINCHES)
70 *    (5) B5  - ACCEPTED S.D. OF THE INSTRUMENT(MICROINCHES)
72 *    (6) B6  - ACCEPTED TOTAL S.D.
75 *    (7) M2  - UNCERTAINTY IN THE RESTRAINT
80 *    (8) X(I),Y(I) - OBSERVED READINGS (EIGHT PAIRS)
85 ***************************************************************
90 *
95 *
100 **************************************************************
105 *
110 *    DATA VALUES WHICH ARE DETERMINED BY THE CALIBRATION DESIGN,
115 *    AND WHICH ARE STORED WITHIN THIS PROGRAM IN DATA STATEMENTS,
120 *    ARE AS FOLLOWS:
125 *    (1) B(I,J) - LEAST SQUARES COEFF TO COMPUTE THE UNKNOWNS
130 *    (2) M(I,J) - LEAST SQUARES COEFF TO COMPUTE THE DEVIATION
135 *    (3) E(I)   - VARIANCE FACTOR
138 *    (4) F(I)   - DRIFT VECTOR
140 *    (5) C1     - MATRIX DIVISOR
142 *    (6) GS(I)  - BLOCK DESIGNATION
145 *
150 *    OTHER VARIABLES:
155 *    (1) N  - NO. OF BLOCKS IN THE CALIBRATION (N = 4)
160 *    (2) G1 - NO. OF OBSERVATIONS (G1 = 8)
165 *    (3) F4 - F RATIO (CRITICAL VALUE FOR P = .01); (F4 = 3.32)
170 *    (4) A4 - NO. OF DEGREES OF FREEDOM (A4 = 4)
175 **************************************************************
180 *
185 *
190 **************************************************************
195 *                              OUTPUT
200 *
205 *    (1)   DATE,OBSERVER,INSTRUMENT
210 *    (2)   COMPARATOR READINGS
215 *    (3)   OBSERVED DIFFERENCES
220 *    (4)   DEVIATIONS
225 *    (5)   VALUES OF THE UNKNOWNS
230 *    (6)   OBSERVED STANDARD DEVIATION
235 *    (7)   STATISTICAL TESTS
240 *    (8)   UNCERTAINTY STATEMENT
245 **************************************************************
```

```
250 DIM C$(3),B(4,9),M(8,8),A(9),X(8),Y(8),G$(4)
255 DATA 35,-14,-7,-14,-21,-14,21,14,84
260 DATA -35,14,7,14,21,14,-21,-14,84
265 DATA -7,14,35,-42,-7,14,-49,42,84
270 DATA 7,42,-35,-14,-49,42,-7,14,84
275 DATA 49,-7,-7,21,49,49,-7,21
280 DATA -7,87,13,-5,33,-41,53,35
285 DATA -7,13,89,25,-39,37,57,-7
290 DATA 21,-5,25,111,-27,3,-23,63
295 DATA 49,33,-39,-27,97,25,9,21
300 DATA 49,-41,37,3,25,103,13,-21
305 DATA -7,53,57,-23,9,13,73,-7
310 DATA  21,35,-7,63,21,-21,-7,63
312 DATA -7,-5,-3,-1,1,3,5,7
315 DATA .31250,.31250,.145833,.145833
320 DATA 168
322 DATA S.,S..,X,Y
330 N=4
335 G1=8
340 F4=3,32
345 A4=4
350 R1=G1+1
355 * READ COEFFICIENTS USED TO COMPUTE VALUES OF THE BLOCKS
360 FOR I = 1,N
365 FOR J1 = 1,R1
370 READ B(I,J1)
375 NEXT J1
380 NEXT I
385 * READ COEFFICIENTS USED TO COMPUTE THE DEVIATIONS
390 FOR I = 1,G1
395 FOR J1= 1,G1
400 READ M(I,J1)
405 NEXT J1
410 NEXT I
411 * READ DRIFT VECTOR
412 FOR I = 1,G1
413 READ F(I)
414 NEXT I
415 * READ VARIANCE VECTOR
420 FOR I = 1,N
425 READ E(I)
430 NEXT I
435 * READ MATRIX DIVISOR
440 READ C1
441 * READ BLOCK DESIGNATIONS
442 FOR I = 1,N
443 READ G$(I)
444 NEXT I
```

```
445 * DEFINE USER DATA FILE - DATA1
450 FILES DATA1
455 * READ ADMINISTRATIVE DATA AND PROCESS PARAMETERS
460 READ#1,C$(1),C$(2),C$(3)
465 READ#1,K,N0,S5,B5,B6,M2
470 * READ COMPARATOR READINGS AND COMPUTE THEIR DIFFERENCES
471 * ALSO, COMPUTE DRIFT =D1, AND S.D.(DRIFT) = S1
473 D1=0
475 FOR I = 1,G1
480 READ#1,X(I),Y(I)
485 A(I)=X(I)-Y(I)
487 D1=D1+A(I)*F(I)
490 NEXT I
492 D1=D1/C1
493 S1=B5*(1./C1)↑.5
495 * SET A(9)= RESTRAINT
500 A(9)=K
505 * COMPUTE VALUES = V(I), S.D. = Z(I), AND UNCERTAINTY = C(I
510 FOR I = 1,N
515 Y1=0
520 FOR J1=1,R1
525 Y1=Y1+B(I,J1)*A(J1)
530 NEXT J1
535 V(I)=Y1/C1
540 Z(I)=(B6↑2-B5↑2*E(I))↑.5
545 C(I)=3*Z(I)+.5*M2
550 NEXT I
555 * COMPUTE CHECK STANDARD
560 C5=V(1)-V(2)
565 * COMPUTE  THE DEVIATIONS AND THE OBSERVED S.D.
570 S0=0
575 FOR I = 1,G1
580 D0=0
585 FOR J1=1,G1
590 D0=D0+M(I,J1)*A(J1)
595 NEXT J1
600 D(I)=D0/C1
605 S0=S0+D(I)↑2
610 NEXT I
615 S=(S0/A4)↑.5
620 * PERFORM STATISTICAL TESTS
625 F=(S/B5)↑2
630 T=(C5-S5)/B6
```

```
640 PRINT,906
645 PRINT DATE  ,C$(1)
650 PRINT OBS.  ,C$(2)
655 PRINT INSTR.,C$(3)
660 PRINT,905
665 PRINT OBSERVATIONS
670 FOR I = 1,G1
675 PRINT X(I),Y(I)
680 NEXT I
685 *PRINT OBSERVED DIFFERENCES AND DEVIATIONS
690 PRINT,695
695 FMT //,X18,A(I),X8, DEV(I)
700 FOR I = 1,G1
705 PRINT,710,A(I),D(I)
710 FMT X14,F9.3,X4,F9.3
715 NEXT I
720 * PRINT VALUES OF THE BLOCKS
722 PRINT,723
723 FMT //,X53,UNCERTAINTY
725 PRINT,730
730 FMT X19,NOM.,X8,CORR.,X10,S. D.,X4,3(S.D.)*.5(S.E.)
735 FOR I =1,N
740 PRINT,745,G$(I),NO,V(I),Z(I),C(I)
745 FMT F12.6,F11.2,X7,F8.5,F15.5
750 NEXT I
755 * PRINT STATISTICAL INFORMATION
760 PRINT,765
765 PRINT OBS. S.D,ACC. S.D.,F TEST,F RATIO,D.F.
770 PRINT,775,S,B5,F,F4,A4
775 FMT F7.4,X5,F9.5,X3,F8.3,F12.2,I10
780 IF F>F4 THEN 790
785 GO TO 800
790 PRINT,795
795 PRINT  ************S. D. IS NOT IN CONTROL**********************
800 PRINT,805
805 PRINT OBS. CHECK,ACC. CHECK,T TEST
810 PRINT,815,C5,S5,T
815 FMT F10.5,F11.5,F12.5
820 IF ABS(T) > 3 THEN 830
825 GO TO 840
830 PRINT,835
835 PRINT ************CHECK STANDARD IS NOT IN CONTROL***************
840 PRINT,905
845* PRINT DRIFT AND S.D.(DRIFT)
885 PRINT,890,D1,S1
890 FMT  DRIFT        = ,F10.4/ S.D.(DRIFT) = F10.4
895 PRINT,900,K
900 FMT ///, RESTRAINT (S.*S..)                    = ,F8.2
901 PRINT,903,M2
902 PRINT,906
903 FMT  SYSTEMATIC EPROR(S.E.) IN RESTRAINT = ,F8.2
905 FMT //
906 FMT //////
910 STOP
915 END
```

INPUT = USER DATA FILE

```
10 MAY 28 1974,HOWELL,FEDERAL 1
20 6.4,.101,-.133,.32,.49,.20
30 52.0,52.5
40 45.2,52.1
50 50.0,45.1
60 53.1,50.0
70 52.3,45.2
80 45.1,52.0
90 52.0,50.1
100 50.1,52.3
```

OUTPUT

```
DATE          MAY 28 1974
OBS.          HOWELL
INSTR.        FEDERAL 1

OBSERVATIONS
     52            52.5
     45.2          52.1
     50            45.1
     53.1          50
     52.3          45.2
     45.1          52
     52            50.1
     50.1          52.3
```

	A(I)	DEV(I)
	-.500	.029
	-6.900	-.046
	4.900	.113
	3.100	.571
	7.100	-.237
	-6.900	-.079
	1.900	-.154
	-2.200	.304

	NOM.	CORR.	S. D.	UNCERTAINTY 3(S.D.)+.5(S.E.)
S.	.101000	2.95	.45618	1.46854
S..	.101000	3.45	.45618	1.46854
X	.101000	.92	.47452	1.52355
Y	.101000	-3.88	.47452	1.52355

OBS. S.D	ACC. S.D.	F TEST	F RATIO	D.F.
.3607	.32000	1.271	3.32	4

OBS. CHECK	ACC. CHECK	T TEST
-.50000	-.13300	-.74898

```
DRIFT          =       .0042
S.D.(DRIFT)    =       .0247

RESTRAINT (S.+S..)                      =      6.40
SYSTEMATIC ERROR(S.E.) IN RESTRAINT     =       .20
```

Designs for the Calibration of Standards of Mass

J. M. Cameron, M. C. Croarkin,
and R. C. Raybold

Institute for Basic Standards
National Bureau of Standards
Washington, D.C. 20234

U.S. DEPARTMENT OF COMMERCE, Juanita M. Kreps, Secretary
Dr. Sidney Harman, Under Secretary
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Acting Director

Issued June 1977 - Condensed version for this publication

DESIGNS FOR THE CALIBRATION OF STANDARDS OF MASS

by

J. M. Cameron, M. C. Croarkin, R. C. Raybold
Office of Measurement Services
Institute for Basic Standards
National Bureau of Standards

This report presents a collection of designs for the intercomparison of sets of weights for use in precision calibration of standards of mass. These include a number of previously unpublished designs which have an additional weight in each set to serve as the check standard for monitoring the performance of the weighing process. Also included are the classical designs of Benoit and Hayford. The complete least squares analysis is presented in integer form (i.e., with a common division) for the most widely used designs; and for the others, the standard deviations are given for various weight combinations when used as an ascending or as a descending series. Designs for sets of nominally equal objects, the 2 2 . . . 1 1 . . . series, the binary sequences, the 5 2 2 1 1 series, and the 5 3 2 1 1 and some miscellaneous series are given.

Key Words: Design of experiments, least squares, mass calibration, statistical design, weighing design

INTRODUCTION

Calibration of a set of weights consists of assigning values for the unknown weights in terms of the known mass of one or more standards. For high precision work, this involves the use of the balance as a comparator which measures the difference between two objects (or two groups of objects) which must have nominally the same mass because of the small "on-scale" range of the comparator. In deriving units which are subdivisions of the basic unit or multiples thereof, a variety of different weighing sets have been used because of convenience or other practical considerations. A typical set is the 5 3 2 1 series which bridges the range from 10 to 1. In this paper, designs are presented

for sets of weights of the same nominal size, for the most common subdivisions currently in use, and for a miscellaneous group included for completeness. In most cases, the designs provide for a check standard, treated as an additional unknown weight, to be used for monitoring the performance of the measuring process.

Precision weighing is usually done by some form of transposition weighing on a two-pan balance and by substitution methods on a one-pan balance. Matters relating to weighing procedures are discussed in [8, 9]. For the purposes of this report, it will be assumed that a well behaved comparator is available and that measurements of differences in the mass of two objects or groups of objects are corrected for air buoyancy effects and other environmental or procedural factors. It is further assumed that the measurements are uncorrelated in the statistical sense and all are of equal precision. (These latter two assumptions are non-trivial and special care has to be taken to insure their validity so that the random error component of the uncertainty is properly evaluated.)

NEED FOR A CHECK STANDARD

In a calibration laboratory, it is necessary to have checks on the measurement process to provide assurance that the process measures what it was intended to measure and that it does so with a nearly constant precision [10]. A direct check on the limiting mean of the measurement process is provided if a known weight is calibrated regularly as if it were an unknown test weight. If the value obtained for the weight differs from its accepted value by an amount larger than can be accounted for by the imprecision of measurement, then the process

would be regarded as being out of control. One is saying that if he cannot calibrate his own weight correctly, he can have little confidence in the values for the calibration of unknown weights derived from the same data.

There is another equally important reason for routine calibration of the same weight--the results on it provide the true measure of the variability of the process. In the course of a year the weighings would have been done under diverse weighing conditions and, hence, the sequence of values would reflect the actual variability of the process--variability which may not be reflected in the internal agreement of one series of measurements.

If an unknown test weight was repeatedly measured, one would expect variability similar to that shown by the check standard. If one has a single measurement on an unknown, it would be like a random selection from the sequence. From the sequence of values on the check standard, one can establish limits to the variability of the process and, because of the equivalence in the method of measurements for both the standard and the unknown, one can legitimately transfer the properties of this sequence to the unknowns.

To establish that the measurement process is in control requires also that the measurements be internally consistent within the limits of random error. If more weighings are made than there are unknowns, then there will be a "closure" error because the values for the observations calculated from "best" values for the weights will differ from that actually observed. The standard deviation computed from these deviations can be tested against the long-run value of this process parameter.

LEAST SQUARES ANALYSIS

We begin then with a set of n observations, $y_1, y_2, \ldots y_n$ involving k objects whose values, $\beta_1, \beta_2, \ldots \beta_k$ are to be determined. The set of observations can be represented by the equations for their expected values, $E(y_i)$,

$$E(y_1) = x_{11}\beta_1 + x_{12}\beta_2 \ldots x_{1k}\beta_k \tag{1}$$
$$E(y_2) = x_{21}\beta_1 + x_{22}\beta_2 \ldots x_{2k}\beta_k$$
$$\vdots$$
$$E(y_n) = x_{n1}\beta_1 + x_{n2}\beta_2 \ldots x_{nk}\beta_k$$

or in matrix form $E(y) = X\beta$ where the element, x_{ij}, of the X matrix is 0 if the weight is absent, and 1 or -1 depending on the direction of the comparison. In this note we shall adopt the convention of using just the signs so that, for example, all possible comparisons (ignoring direction) of 4 nominally equal objects will have the representation

β_1	β_2	β_3	β_4
+	-		
+		-	
+			-
	+	-	
	+		-
		+	-

$$X = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

In the least squares analysis one forms the normal equations

$$X'X\hat{\beta} = X'y$$

where the entries in X'X are merely the sums of squares and sums of cross products of the columns of X. In the above case, one gets

$$\begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix} \hat{\beta} = \begin{bmatrix} y_1 + y_2 + y_3 \\ -y_1 + y_4 + y_5 \\ -y_2 - y_4 + y_6 \\ -y_3 - y_5 - y_6 \end{bmatrix}$$

where $\hat{\beta}$ is the column vector with elements $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$, the caret being used to denote the fact that the values are functions of the observations, and not the sought-after values, β.

It can easily be verified in this case that the system of equations is not of full rank (e.g., the column totals are zero) and this is a property of all designs where only differences are measured. In mass calibration, one has one or more standards whose value can be taken as known and these provide the restraint on the system needed to give a unique set of answers. Usually these involve a starting kilogram or a unique summation such as 5 + 3 + 2 which has been determined in a previous series or is the initial unit value for an ascending series such as the 1, 2, 3, 5 series. One can write the restraint* in the form

$$r_1\beta_1 + r_2\beta_2 \ldots + r_k\beta_k = m \qquad (2)$$

and use the method of Lagrangian multipliers (with multipliers 2λ) to minimize the function

*In all cases treated here a single restraint is sufficient. See Zelen [12] and Goldman and Zelen [6] for a discussion of the general case.

$$\Phi = \Sigma(\text{deviations})^2 + 2\lambda(r_1\beta_1 + \ldots r_k\beta_k - m) \qquad (3)$$

The normal equations now contain an additional "unknown," namely λ and written out in full are as follows:

$$\begin{aligned}
\Sigma x_1^2\hat{\beta}_1 &+ \Sigma x_1x_2\hat{\beta}_2 \ldots \Sigma x_1x_k\hat{\beta}_k + r_1\lambda = \Sigma x_1y \qquad (4)\\
\Sigma x_2x_1\hat{\beta}_1 &+ \Sigma x_2^2\hat{\beta}_2 \ldots \Sigma x_2x_k\hat{\beta}_k + r_2\lambda = \Sigma x_2y\\
&\vdots\\
\Sigma x_kx_1\hat{\beta}_1 &+ \Sigma x_kx_2\hat{\beta}_2 \ldots \Sigma x_k^2\hat{\beta}_k + r_k\lambda = \Sigma x_ky\\
r_1\hat{\beta}_1 &+ r_2\hat{\beta}_2 \ldots r_k\hat{\beta}_k = m
\end{aligned}$$

where

$$\Sigma x_ix_j = \sum_{k=1}^{n} x_{ik}x_{jk}$$

$$\Sigma x_iy = \sum_{k=1}^{n} x_{ik}y_k$$

or in matrix notation

$$\begin{bmatrix} X'X & r \\ r' & o \end{bmatrix}\begin{bmatrix} \hat{\beta} \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ m \end{bmatrix} \qquad (5)$$

The solution may be written out formally as follows:

$$\begin{bmatrix} \hat{\beta} \\ \lambda \end{bmatrix} = \begin{bmatrix} C & h \\ h' & o \end{bmatrix}\begin{bmatrix} X'y \\ m \end{bmatrix} = \begin{bmatrix} CX' & h \\ h'X' & o \end{bmatrix}\begin{bmatrix} y \\ m \end{bmatrix} \qquad (6)$$

where $r' = (r_1r_2 \ldots r_k)$.

To facilitate computation it is convenient to have the values, $\hat{\beta}$, written out as linear functions of the y's and m, i.e., $\hat{\beta} = [CX', h]\begin{bmatrix} y \\ m \end{bmatrix}$. This leads to a set of multipliers of the observations of the form

$$\hat{\beta}_1 = g_{11}y_1 + g_{12}y_2 \cdots g_{1n}y_n + h_1 m \tag{7}$$

$$\vdots$$

$$\hat{\beta}_k = g_{k1}y_1 + g_{k2}y_2 \cdots g_{kn}y_n + h_k m$$

These multipliers, g_{ij} and h_i, are given in Appendix B in transposed form for some of the designs. The matrix C is important because the variances and covariances of the estimates are given by

$$\text{Variance } (\hat{\beta}_i) = C_{ii}\sigma^2, \quad \text{Covariance } (\hat{\beta}_i, \hat{\beta}_j) = C_{ij}\sigma^2 \tag{8}$$

The quantity, σ^2, is the variance (square of the long run value of the standard deviation) associated with the process. In a set of n observations on k items and r = 1 restraints one has n - k + r = n - k + 1 degrees of freedom for a standard deviation, s, formed by

$$s^2 = \frac{1}{n - k + 1} \Sigma(\text{deviations})_i^2 \tag{9}$$

$$(\text{deviation})_i = y_i - (x_{i1}\hat{\beta}_1 + x_{i2}\hat{\beta}_2 \cdots x_{ik}\hat{\beta}_k)$$

One can write these deviations as a function of the observations by noting that the predicted values are just $X\hat{\beta}$ and the deviations are thus

$$\text{dev} = y - X\hat{\beta} = y - X[CX', h]\begin{bmatrix} y \\ m \end{bmatrix} = y - [XCX', 0]\begin{bmatrix} y \\ m \end{bmatrix} \tag{10}$$

$$= [I - XCX']y$$

which can be written as

$$\text{dev}_1 = d_{11}y_1 + d_{12}y_2 \cdots d_{1n}y_n \tag{11}$$

$$\vdots$$

$$\text{dev}_n = d_{n1}y_1 + d_{n2}y_2 \cdots d_{nn}y_n$$

The array of coefficients, d_{ij}, is given in Appendix B for some of the designs

Weights are often used in combination and one needs to know the standard deviation for the various sums. For a sum of two items, $\hat{\beta}_i$ and $\hat{\beta}_j$, one has

$$\text{Var}(\hat{\beta}_i + \hat{\beta}_j) = \text{Var}(\hat{\beta}_i) + \text{Var}(\hat{\beta}_j) + 2\text{Cov}(\hat{\beta}_i, \hat{\beta}_j)$$

and for a linear combination

$$L = \ell_1\hat{\beta}_1 + \ell_2\hat{\beta}_2 \ldots \ell_k\hat{\beta}_k \qquad (12)$$

$$\text{Variance } (L) = \ell' C \ell \sigma^2$$

where $\ell' = (\ell_1, \ell_2 \ldots \ell_k)$, C comes from the inverse of the matrix of normal equations [see equation (6)]. In Appendix A each design has a list of the factors D_i for computing the standard deviations for all usual weight combinations, L_i where $\text{Variance}(L_i) = D_i^2\sigma^2$.

DESIGNS FOR WEIGHING

The criteria for good weighing designs depend to some extent on the use intended for the resulting values. For example, if the weights are to be used independently of each other, then one would want the standard deviation [$\sigma\sqrt{C_{ii}}$ from formula (8)] for the value for each unknown weight to be the minimum possible. If the weights are to be used in combination, then one wants the variance of all appropriate linear functions to be as small as possible.

Further, the desirability of a design depends somewhat on the restraint being used. In some cases, one's judgment of a design changes depending on whether one starts with a summation as known (e.g., 5 + 3 + 2) and works down, or with a unit as known and works up (e.g., by use of a 1, 2, 3, 5 series). For a given number of measurements

only a finite set of possible designs exist for a series and only occasionally is one of these designs uniformly and undeniably "best".

The designs are grouped into categories in Appendix A. In Appendix B the complete analysis is given for one design. For the others, the complete analysis is on file with the authors, and the factors for computing the standard deviations for different weight combinations are given in Appendix A.

A. DESIGNS FOR NOMINALLY EQUAL GROUPS

A.1 All distinct intercomparisons. If k weights are to be intercompared by measuring the difference between weights of the $k(k-1)/2$ distinct pairings, then a general analysis can be written out as a function of the number of weights that are regarded as known and used as the restraint. The inverse of the normal equations with the sum of the first m of the k weights taken as known is as follows:

$$\begin{bmatrix} kI-J & -J & \pm \\ -J & kI-J & 0 \\ \pm' & 0 & 0 \end{bmatrix}^{-1} = \frac{1}{mk} \begin{bmatrix} mI-J & 0 & k\pm \\ 0 & mI+J & k\pm \\ k\pm' & k\pm' & 0 \end{bmatrix} \qquad (13)$$

where $\pm' = (1, 1, \ldots 1)$ and J is a matrix of all ones and the matrices on the diagonal are of the dimension mxm, (k-m)x(k-m), and 1x1.

Thus the standard deviation of the value for weights within the restraint is $\sigma\sqrt{\frac{m-1}{mk}}$ and for the unknowns, $\sigma\sqrt{\frac{m+1}{mk}}$. The standard deviation of a sum of h unknowns is $\sigma\sqrt{\frac{h(h+m)}{mk}}$.

The $\hat{\beta}_i$ are given by

$$\hat{\beta}_i = \frac{T_i}{k} - \frac{\sum_1^m T_i}{mk} + \frac{K}{m} \qquad \text{for weights within restraint}$$

$$\hat{\beta}_i = \frac{T_i}{k} + \frac{\sum_{m+1}^k T_i}{mk} + \frac{K}{m} \qquad \text{for unknown weights} \tag{14}$$

where T_i is the sum of the y values invlving β_i in the positive sense minus the sum of y values involving β_i in the negative sense (e.g., if $E(y) = \beta_i - \beta_j$, then y would be added to T_i but subtracted from T_j), and K is the value of the restraint ($\sum_1^m \beta = K$).

The standard deviation s is given by

$$s^2 = \frac{2}{(k-1)(k-2)} \{\Sigma dev^2\}$$

$$= \frac{2}{(k-1)(k-2)} \{\Sigma y_i^2 - \frac{\Sigma T_i^2}{k}\} \tag{15}$$

Designs for which a linear drift with time is balanced out are also included (see [4] for details of the analysis).

A.2 Subsets of all distinct intercomparisons. For large k (say $k \geq 6$) the number of possible pairings becomes large enough that the time involved in completing the measurements leads to a degradation of the precision as environmental changes, operator, fatigue, etc., become important. For that reason, subsets of the k(k-1)/2 pairings are used to form the design.

In some of the designs the sum of all weights is taken as the restraint. This is appropriate when the design is used to monitor within group behavior. In others, there is an implied grouping into two classes the sum for one of which is taken as the restraint.

A.3 Designs involving grouping of weights. When differences between groups of two or more weights are measured, a reduction in the variance of the values can be achieved in comparison with an equal number of differences between single weights. However, for large k the problems of identifying and handling the groupings may outweigh the possible gain in efficiency. Bose and Cameron [2, 3] have tabulated all designs up to k = 13 and give methods of construction for $k \leq 50$, for the special case of designs balanced so that all weights appear equally often with each other on the same pan and a similar property holds for their occurrence in opposite pans. Partially balanced designs have been developed by Suryanarayana and Chakravarti [5, 11].

TREND ELIMINATION

When responses are time dependent due to temperature and atomospheric changes, proper ordering can make the values for the weights independent of any drift effect. The designs A.1.3, A.2.5, A.2.2, and A.2.4 have this property. If one wishes to use the trend eliminating property of this design (they are valid as given, of course), then one can account for a drift effect of the form . . . -3Δ, -2Δ, $-\Delta$, 0, Δ, 2Δ, 3Δ, . . . if n is odd or by . . . -5Δ, -3Δ, $-\Delta$, $-\Delta$, 3Δ, 5Δ, . . . if n is even. This will not change the computations for the weights or their variances. However, the degrees of freedom are reduced by one and the deviations will be different.

USES OF THE DESIGNS

Description of Designs

Each design lists k, the number of weights; n, the number of measurements; and d.f., the degrees of freedom associated with the standard deviation.

The identification or nominal size of each weight β_i is given next to the heading "Observations." Y(1), Y(2), . . . denotes the measurements where + indicates the weight is present positively, and - indicates the weight is present negatively.

For example, Design A.1.1 involves three equal weights. If used as a starting design for a 1 kg to 1 mg set, the three weights would be kilograms, and the first observation Y(1) would be the difference between the first and second kilograms.

Two restraints are listed although others are possible. Usually "Restraint A" is appropriate for descending series involving, for example, 5 + 3 + 2 which would be calibrated as a "Ten" weight in the higher series involving 50, 30, 20, 10. For ascending series, "Restraint B" is usually a single unit weight.

In the case of three equal weights, Restraint A takes the sum of two kilograms as known; Restraint B takes the single third weight as the reference standard.

"Factors for Computing Standard Deviations" give the multipliers needed to calculate the standard deviations of linear combinations of the weights. "Wt" identifies the total weight combination or load L_i where

$$L_i = \sum_{i=1}^{k} \ell_i \beta_i, \quad \ell_i = 0 \text{ or } 1$$

The next two columns under Restraints A and B list the factors D_i where D_i is calculated so that

$$\text{Variance}(L_i) = D_i^2\sigma^2$$

The remaining columns, under the nominal sizes of the weights, show the actual weights involved in the combinations, i.e. a + indicates $\ell_i = 1$ and a blank indicates $\ell_i = 0$.

DESIGN A.1.1 THREE EQUAL WEIGHTS K= 3 N= 3 D.F.= 1

OBSERVATIONS	1	1	1
Y(1)	+	−	
Y(2)	+		−
Y(3)		+	−
RESTRAINT A	+	+	
RESTRAINT B			+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1
1	.7071	.0000			+
1	.4082	.8165		+	
1	.4082	.8165	+		
2	.0000	1.4142	+	+	
3	.7071	1.4142	+	+	+

DESIGN A.1.2 FOUR EQUAL WEIGHTS K= 4 N= 6 D.F.= 3

OBSERVATIONS	1	1	1	1
Y(1)	+	−		
Y(2)	+		−	
Y(3)	+			−
Y(4)		+	−	
Y(5)		+		−
Y(6)			+	−
RESTRAINT A	+	+		
RESTRAINT B				+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1
1	.6124	.0000				+
1	.6124	.7071			+	
1	.3536	.7071		+		
1	.3536	.7071	+			
2	.0000	1.2247	+	+		
3	.6124	1.7321	+	+	+	
4	1.0000	1.7321	+	+	+	+

DESIGN A.1.3 FOUR EQUAL WEIGHTS TREND ELIMINATION* (CAMERON-HAILES) K= 4 N= 8 D.F.= 5

OBSERVATIONS	1	1	1	1
Y(1)	+			−
Y(2)	−		+	
Y(3)		+	−	
Y(4)		−		+
Y(5)			−	+
Y(6)	−		+	
Y(7)	+	−		
Y(8)		+		−
RESTRAINT A	+	+		
RESTRAINT B				+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1
1	.5204	.0000				+
1	.5204	.6455			+	
1	.3227	.5774		+		
1	.3227	.6455	+			
2	.0000	1.0408	+	+		
3	.5204	1.5275	+	+	+	
4	.8165	1.5275	+	+	+	+

*See page 13.

DESIGN A.1.4 FIVE EQUAL WEIGHTS K= 5 N=10 D.F.= 6

OBSERVATIONS	1	1	1	1	1
Y(1)	+	−			
Y(2)	+		−		
Y(3)	+			−	
Y(4)	+				−
Y(5)		+	−		
Y(6)		+		−	
Y(7)		+			−
Y(8)			+	−	
Y(9)			+		−
Y(10)				+	−
RESTRAINT A	+	+			
RESTRAINT B					+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1
1	.5477	.0000					+
1	.5477	.6325				+	
1	.5477	.6325			+		
1	.3162	.6325		+			
1	.3162	.6325	+				
2	.0000	1.0954	+	+			
3	.5477	1.5492	+	+	+		
4	.8944	2.0000	+	+	+	+	
5	1.2247	2.0000	+	+	+	+	+

DESIGN A.1.5 FIVE EQUAL WEIGHTS K= 5
TREND ELIMINATION * N=10
(CAMERON-HAILES) D.F.= 6

OBSERVATIONS	1	1	1	1	1
Y(1)	+	-			
Y(2)		+	-		
Y(3)			+	-	
Y(4)				+	-
Y(5)	-				+
Y(6)	-			+	
Y(7)		+		-	
Y(8)		-			+
Y(9)			+		-
Y(10)	+		-		
RESTRAINT A	+	+			
RESTRAINT B					+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1
1	.5477	.0000					+
1	.5477	.6325				+	
1	.5477	.6325			+		
1	.3162	.6325		+			
1	.3162	.6325	+				
2	.0000	1.0954	+	+			
3	.5477	1.5492	+	+	+		
4	.8944	2.0000	+	+	+	+	
5	1.2247	2.0000	+	+	+	+	+

*See page 13.

DESIGN A.2.1 SIX EQUAL WEIGHTS K= 6
N= 8
D.F.= 3

OBSERVATIONS	1	1	1	1	1	1
Y(1)	+		-			
Y(2)	+			-		
Y(3)	+				-	
Y(4)	+					-
Y(5)		+	-			
Y(6)		+		-		
Y(7)		+			-	
Y(8)		+				-
RESTRAINT A	+	+				
RESTRAINT B	+	+	+	+	+	+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1
1	.7071	.6236						+
1	.7071	.6236					+	
1	.7071	.6236				+		
1	.7071	.6236			+			
1	.3536	.4249		+				
1	.3536	.4249	+					
2	.0000	.4714	+	+				
3	.7071	.7071	+	+	+			
4	1.0000	.7454	+	+	+	+		
5	1.2247	.6236	+	+	+	+	+	
6	1.4142	.0000	+	+	+	+	+	+

DESIGN A.1.6 SIX EQUAL WEIGHTS K= 6
N=15
D.F.=10

OBSERVATIONS	1	1	1	1	1	1
Y(1)	+	-				
Y(2)	+		-			
Y(3)	+			-		
Y(4)	+				-	
Y(5)	+					-
Y(6)		+	-			
Y(7)		+		-		
Y(8)		+			-	
Y(9)		+				-
Y(10)			+	-		
Y(11)			+		-	
Y(12)			+			-
Y(13)				+	-	
Y(14)				+		-
Y(15)					+	-
RESTRAINT A	+	+				
RESTRAINT B						+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1
1	.5000	.0000						+
1	.5000	.5774					+	
1	.5000	.5774				+		
1	.5000	.5774			+			
1	.2887	.5774		+				
1	.2887	.5774	+					
2	.0000	1.0000	+	+				
3	.5000	1.4142	+	+	+			
4	.8165	1.8257	+	+	+	+		
5	1.1180	2.2361	+	+	+	+	+	
6	1.4142	2.2361	+	+	+	+	+	+

DESIGN A.2.2 SIX EQUAL WEIGHTS K= 6
TREND ELIMINATION * N=12
D.F.= 7

OBSERVATIONS	1	1	1	1	1	1
Y(1)	+		-			
Y(2)	-			+		
Y(3)			+		-	
Y(4)		+				-
Y(5)		-			+	
Y(6)				-		+
Y(7)		-	+			
Y(8)		+		-		
Y(9)			-			+
Y(10)	-				+	
Y(11)	+					-
Y(12)				+	-	
RESTRAINT A	+	+				
RESTRAINT B						+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1
1	.5401	.0000						+
1	.5401	.7071					+	
1	.5401	.6455				+		
1	.5401	.6455			+			
1	.3536	.6455		+				
1	.3536	.6455	+					
2	.0000	1.0801	+	+				
3	.5401	1.5546	+	+	+			
4	.8165	2.0000	+	+	+	+		
5	1.1365	2.5495	+	+	+	+	+	
6	1.4142	2.5495	+	+	+	+	+	+

*See page 13.

DESIGN A.2.3 SEVEN EQUAL WEIGHTS K= 7
N=10
D.F.= 4

OBSERVATIONS	1	1	1	1	1	1	1
Y(1)	+		−				
Y(2)	+			−			
Y(3)	+				−		
Y(4)	+					−	
Y(5)	+						−
Y(6)		+	−				
Y(7)		+		−			
Y(8)		+			−		
Y(9)		+				−	
Y(10)		+					−
RESTRAINT A	+	+					
RESTRAINT B	+	+	+	+	+	+	+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1	1
1	.7071	.6389							+
1	.7071	.6389						+	
1	.7071	.6389					+		
1	.7071	.6389				+			
1	.7071	.6389			+				
1	.3162	.3886		+					
1	.3162	.3886	+						
2	.0000	.4518	+	+					
3	.7071	.7284	+	+	+				
4	1.0000	.8207	+	+	+	+			
5	1.2247	.7954	+	+	+	+	+		
6	1.4142	.6389	+	+	+	+	+	+	
7	1.5811	.0000	+	+	+	+	+	+	+

DESIGN A.2.4 SEVEN EQUAL WEIGHTS K= 7
TREND ELIMINATION * N=14
D.F.= 8

OBSERVATIONS	1	1	1	1	1	1	1
Y(1)	+	−					
Y(2)		+	−				
Y(3)			+	−			
Y(4)				+	−		
Y(5)					+	−	
Y(6)						+	−
Y(7)	−						+
Y(8)	−		+				
Y(9)			−		+		
Y(10)					−		+
Y(11)		+					−
Y(12)		−		+			
Y(13)				−		+	
Y(14)	+					−	
RESTRAINT A	+	+					
RESTRAINT B	+	+	+	+	+	+	+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1	1
1	.5694	.4447							+
1	.6157	.4447						+	
1	.6504	.4447					+		
1	.6157	.4447				+			
1	.5694	.4447			+				
1	.3231	.4447		+					
1	.3231	.4447	+						
2	.0000	.6112	+	+					
3	.5694	.7110	+	+	+				
4	.9945	.7110	+	+	+	+			
5	1.4507	.6112	+	+	+	+	+		
6	1.8337	.4447	+	+	+	+	+	+	
7	2.1394	.0000	+	+	+	+	+	+	+

*See page 13.

DESIGN A.2.5 EIGHT EQUAL WEIGHTS K= 8
N= 8
D.F.= 1

OBSERVATIONS	1	1	1	1	1	1	1	1
Y(1)	+				−			
Y(2)	+					−		
Y(3)		+					−	
Y(4)		+						−
Y(5)			+					−
Y(6)			+		−			
Y(7)				+		−		
Y(8)				+			−	
RESTRAINT A	+	+	+	+				
RESTRAINT B	+	+	+	+	+	+	+	+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1	1
1	.8660	.8101							+
1	.8660	.8101						+	
1	.8660	.8101					+		
1	.8660	.8101				+			
1	.7906	.8101					+		
1	.7906	.8101			+				
1	.7906	.8101		+					
1	.7906	.8101	+						
2	.7071	.7906	+	+					
3	.7906	.9520	+	+	+				
4	.0000	.7071	+	+	+	+			
5	.8660	.9520	+	+	+	+	+		
6	1.2247	1.0607	+	+	+	+	+	+	
7	1.3229	.8101	+	+	+	+	+	+	+
8	1.4142	.0000	+	+	+	+	+	+	+

DESIGN A.2.6 EIGHT EQUAL WEIGHTS K= 8
N=12
D.F.= 5

OBSERVATIONS	1	1	1	1	1	1	1	1
Y(1)	+		−					
Y(2)	+			−				
Y(3)	+				−			
Y(4)	+					−		
Y(5)	+						−	
Y(6)	+							−
Y(7)		+	−					
Y(8)		+		−				
Y(9)		+			−			
Y(10)		+				−		
Y(11)		+					−	
Y(12)		+						−
RESTRAINT A	+	+						
RESTRAINT B	+	+	+	+	+	+	+	+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1	1
1	.7071	.6495							+
1	.7071	.6495						+	
1	.7071	.6495					+		
1	.7071	.6495				+			
1	.7071	.6495					+		
1	.7071	.6495			+				
1	.2887	.3608		+					
1	.2887	.3608	+						
2	.0000	.4330	+	+					
3	.7071	.7395	+	+	+				
4	1.0000	.8660	+	+	+	+			
5	1.2247	.8927	+	+	+	+	+		
6	1.4142	.8292	+	+	+	+	+	+	
7	1.5811	.6495	+	+	+	+	+	+	+
8	1.7321	.0000	+	+	+	+	+	+	+

DESIGN A.2.7 EIGHT EQUAL WEIGHTS K= 8
N=15
D.F.= 8

OBSERVATIONS	1	1	1	1	1	1	1	1
Y(1)	+			−				
Y(2)	+				−			
Y(3)	+					−		
Y(4)	+						−	
Y(5)	+							−
Y(6)		+		−				
Y(7)		+			−			
Y(8)		+				−		
Y(9)		+					−	
Y(10)		+						−
Y(11)			+	−				
Y(12)			+		−			
Y(13)			+			−		
Y(14)			+				−	
Y(15)			+					−
RESTRAINT A	+	+	+					
RESTRAINT B	+	+	+	+	+	+	+	+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	RESTRAINTS B	1	1	1	1	1	1	1	1
1	.5774	.5254								+
1	.5774	.5254							+	
1	.5774	.5254						+		
1	.5774	.5254					+			
1	.5774	.5254				+				
1	.3651	.3992			+					
1	.3651	.3992		+						
1	.3651	.3992	+							
2	.3651	.4873	+	+						
3	.0000	.4841	+	+	+					
4	.5774	.6455	+	+	+	+				
5	.8165	.6960	+	+	+	+	+			
6	1.0000	.6614	+	+	+	+	+	+		
7	1.1547	.5254	+	+	+	+	+	+	+	
8	1.2910	.0000	+	+	+	+	+	+	+	+

DESIGN A.2.6 EIGHT EQUAL WEIGHTS K= 8
TREND ELIMINATION N=16
D.F.= 9

OBSERVATIONS	1	1	1	1	1	1	1	1
Y(1)	+						−	
Y(2)						−	+	
Y(3)			−			+		
Y(4)		−	+					
Y(5)		+						−
Y(6)					−			+
Y(7)				−	+			
Y(8)	−			+				
Y(9)	−		+					
Y(10)			−		+			
Y(11)					−		+	
Y(12)		+					−	
Y(13)		−		+				
Y(14)				−		+		
Y(15)						−		+
Y(16)	+							−
RESTRAINT A	+	+						
RESTRAINT B	+	+	+	+	+	+	+	+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	RESTRAINTS B	1	1	1	1	1	1	1	1
1	.5590	.4507								+
1	.5590	.4507							+	
1	.6124	.4507						+		
1	.6124	.4507					+			
1	.5590	.4507				+				
1	.5590	.4507			+					
1	.3536	.4507		+						
1	.3536	.4507	+							
2	.0000	.5590	+	+						
3	.5590	.6731	+	+	+					
4	.8660	.7071	+	+	+	+				
5	1.2748	.6731	+	+	+	+	+			
6	1.6583	.5590	+	+	+	+	+	+		
7	1.9526	.4507	+	+	+	+	+	+	+	
8	2.2361	.0000	+	+	+	+	+	+	+	+

DESIGN A.2.9 NINE EQUAL WEIGHTS K= 9 N=12 D.F.= 4

OBSERVATIONS	1	1	1	1	1	1	1	1	1
Y(1)	+			−					
Y(2)	+						−		
Y(3)	+				−				
Y(4)	+							−	
Y(5)		+			−				
Y(6)		+						−	
Y(7)		+				−			
Y(8)		+							−
Y(9)			+			−			
Y(10)			+						−
Y(11)			+	−					
Y(12)			+				−		
RESTRAINT A	+	+	+						
RESTRAINT B	+	+	+	+	+	+	+	+	+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	RESTRAINTS B	1	1	1	1	1	1	1	1	1
1	.7454	.6939									+
1	.7454	.6939								+	
1	.7454	.6939							+		
1	.7454	.6939						+			
1	.7454	.6939					+				
1	.7454	.6939				+					
1	.4714	.5092			+						
1	.4714	.5092		+							
1	.4714	.5092	+								
2	.4714	.6086	+	+							
3	.0000	.5774	+	+	+						
4	.7454	.8389	+	+	+	+					
5	1.0274	.9329	+	+	+	+	+				
6	1.2247	.9129	+	+	+	+	+	+			
7	1.4337	.8714	+	+	+	+	+	+	+		
8	1.5966	.6939	+	+	+	+	+	+	+	+	
9	1.7321	.0000	+	+	+	+	+	+	+	+	+

DESIGN A.2.10 TEN EQUAL WEIGHTS K=10 N=10 D.F.= 1

OBSERVATIONS	1	1	1	1	1	1	1	1	1	1
Y(1)	+					−				
Y(2)	+							−		
Y(3)		+					−			
Y(4)		+							−	
Y(5)			+					−		
Y(6)			+							−
Y(7)				+					−	
Y(8)				+		−				
Y(9)					+					−
Y(10)					+		−			
RESTRAINT A	+	+	+	+	+					
RESTRAINT B	+	+	+	+	+	+	+	+	+	+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	RESTRAINTS B	1	1	1	1	1	1	1	1	1	1
1	.9487	.9083										+
1	.9487	.9083									+	
1	.9487	.9083								+		
1	.9487	.9083							+			
1	.9487	.9083						+				
1	.8944	.9083					+					
1	.8944	.9083				+						
1	.8944	.9083			+							
1	.8944	.9083		+								
1	.8944	.9083	+									
2	.8944	.9487	+	+								
3	.8944	1.0124	+	+	+							
4	.8944	1.0954	+	+	+	+						
5	.0000	.7906	+	+	+	+	+					
6	.9487	1.0954	+	+	+	+	+	+				
7	1.0954	1.0124	+	+	+	+	+	+	+			
8	1.3038	.9487	+	+	+	+	+	+	+	+		
9	1.5492	.9083	+	+	+	+	+	+	+	+	+	
10	1.5811	.0000	+	+	+	+	+	+	+	+	+	+

DESIGN A.2.11 TEN EQUAL WEIGHTS K=10 N=12 D.F.= 3

CESERVATIONS	1	1	1	1	1	1	1	1	1	1
Y(1)	+				−					
Y(2)	+					−				
Y(3)	+						−			
Y(4)		+			−					
Y(5)		+						−		
Y(6)		+							−	
Y(7)			+			−				
Y(8)			+					−		
Y(9)			+							−
Y(10)				+			−			
Y(11)				+					−	
Y(12)				+						−
RESTRAINT A	+	+	+	+						
RESTRAINT B	+	+	+	+	+	+	+	+	+	+

FACTCRS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1	1	1	1	1
1	.7906	.7450										+
1	.7906	.7450									+	
1	.7906	.7450								+		
1	.7906	.7450							+			
1	.7906	.7450						+				
1	.7906	.7450					+					
1	.6124	.6364				+						
1	.6124	.6364			+							
1	.6124	.6364		+								
1	.6124	.6364	+									
2	.7071	.7874	+	+								
3	.6124	.8031	+	+	+							
4	.0000	.6928	+	+	+	+						
5	.7906	.9354	+	+	+	+	+					
6	1.1180	1.0630	+	+	+	+	+	+				
7	1.3693	1.1158	+	+	+	+	+	+	+			
8	1.5000	.9849	+	+	+	+	+	+	+	+		
9	1.6202	.7450	+	+	+	+	+	+	+	+	+	
10	1.7321	.0000	+	+	+	+	+	+	+	+	+	+

DESIGN A.2.12 TEN EQUAL WEIGHTS K=10 N=15 D.F.= 6

OBSERVATIONS	1	1	1	1	1	1	1	1	1	1
Y(1)	+					−				
Y(2)	+						−			
Y(3)	+								−	
Y(4)		+					−			
Y(5)		+						−		
Y(6)		+								−
Y(7)			+					−		
Y(8)			+						−	
Y(9)			+			−				
Y(10)				+					−	
Y(11)				+						−
Y(12)				+			−			
Y(13)					+					−
Y(14)					+	−				
Y(15)					+			−		
RESTRAINT A	+	+	+	+	+					
RESTRAINT B	+	+	+	+	+	+	+	+	+	+

FACTORS FCR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1	1	1	1	1
1	.6276	.5865										+
1	.6276	.5865									+	
1	.6276	.5865								+		
1	.6276	.5865							+			
1	.6276	.5865						+				
1	.5721	.5865					+					
1	.5721	.5865				+						
1	.5721	.5865			+							
1	.5721	.5865		+								
1	.5721	.5865	+									
2	.6606	.7092	+	+								
3	.6606	.7657	+	+	+							
4	.5721	.7707	+	+	+	+						
5	.0000	.6455	+	+	+	+	+					
6	.6276	.7707	+	+	+	+	+	+				
7	.8385	.7657	+	+	+	+	+	+	+			
8	1.0180	.7092	+	+	+	+	+	+	+	+		
9	1.1807	.5865	+	+	+	+	+	+	+	+	+	
10	1.2910	.0000	+	+	+	+	+	+	+	+	+	+

DESIGN A.3.1 SIX EQUAL WEIGHTS K= 6
N=10
(BOSE-CAMERON) D.F.= 5

OBSERVATIONS	1	1	1	1	1	1
Y(1)	+	-	-	+	-	+
Y(2)	-	+	-	-	+	+
Y(3)	+	-	+	-	-	+
Y(4)	-	+	-	+	-	+
Y(5)	-	-	+	-	+	+
Y(6)	+	-	-	+	+	-
Y(7)	+	+	-	-	+	-
Y(8)	+	+	+	-	-	-
Y(9)	-	+	+	+	-	-
Y(10)	-	-	+	+	+	-
RESTRAINT A	+	+				
RESTRAINT B						+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1
1	.3536	.0000						+
1	.3536	.4082					+	
1	.3536	.4082				+		
1	.3536	.4082			+			
1	.2041	.4082		+				
1	.2041	.4082	+					
2	.0000	.7071	+	+				
3	.3536	1.0000	+	+	+			
4	.5774	1.2910	+	+	+	+		
5	.7906	1.5811	+	+	+	+	+	
6	1.0000	1.5811	+	+	+	+	+	+

DESIGN A.3.2 SIX EQUAL WEIGHTS K= 6
N=15
(BOSE-CAMERON) D.F.=10

OBSERVATIONS	1	1	1	1	1	1
Y(1)	+	-		+		-
Y(2)		+	-		+	-
Y(3)	+		+	-		-
Y(4)		+		+	-	-
Y(5)	-		+		+	-
Y(6)	+	-	-			+
Y(7)		+	-	-		+
Y(8)			+	-	-	+
Y(9)	-			+	-	+
Y(10)	-	-			+	+
Y(11)	-	+	+	-		
Y(12)		-	+	+	-	
Y(13)	-		-	+	+	
Y(14)	+	-		-	+	
Y(15)	+	+	-		-	
RESTRAINT A	+	+				
RESTRAINT B						+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1
1	.3536	.0000						+
1	.3536	.4082					+	
1	.3536	.4082				+		
1	.3536	.4082			+			
1	.2041	.4082		+				
1	.2041	.4082	+					
2	.0000	.7071	+	+				
3	.3536	1.0000	+	+	+			
4	.5774	1.2910	+	+	+	+		
5	.7906	1.5811	+	+	+	+	+	
6	1.0000	1.5811	+	+	+	+	+	+

DESIGN A.3.3 SEVEN EQUAL WEIGHTS K= 7
N= 7
(BOSE-CAMERON) D.F.= 1

OBSERVATIONS	1	1	1	1	1	1	1
Y(1)	+	+	-	+	-	-	
Y(2)		+	+	-	+	-	-
Y(3)	-		+	+	-	+	-
Y(4)	-	-		+	+	-	+
Y(5)	+	-	-		+	+	-
Y(6)	-	+	-	-		+	+
Y(7)	+	-	+	-	-		+
RESTRAINT A	+	+					
RESTRAINT B							+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1	1
1	.4629	.0000							+
1	.4629	.5345						+	
1	.4629	.5345					+		
1	.4629	.5345				+			
1	.4629	.5345			+				
1	.2673	.5345		+					
1	.2673	.5345	+						
2	.0000	.9258	+	+					
3	.4629	1.3093	+	+	+				
4	.7559	1.6903	+	+	+	+			
5	1.0351	2.0702	+	+	+	+	+		
6	1.3093	2.4495	+	+	+	+	+	+	
7	1.5811	2.4495	+	+	+	+	+	+	+

DESIGN A.3.4 EIGHT EQUAL WEIGHTS K= 8
N= 7
(BOSE-CAMERON) D.F.= 0

OBSERVATIONS	1	1	1	1	1	1	1	1
Y(1)	+	+	+	+	-	-	-	-
Y(2)	+	+	-	-	+	+	-	-
Y(3)	+	-	+	-	+	-	+	-
Y(4)	+	+	-	-	-	-	+	+
Y(5)	+	-	+	-	-	+	-	+
Y(6)	+	-	-	+	+	-	-	+
Y(7)	+	-	-	+	-	+	+	-
RESTRAINT A	+	+						
RESTRAINT B								+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1	1
1	.4330	.0000							+
1	.4330	.5000						+	
1	.4330	.5000					+		
1	.4330	.5000				+			
1	.4330	.5000			+				
1	.2500	.5000		+					
1	.2500	.5000	+						
2	.0000	.8660	+	+					
3	.4330	1.2247	+	+	+				
4	.7071	1.5811	+	+	+	+			
5	.9682	1.9365	+	+	+	+	+		
6	1.2247	2.2913	+	+	+	+	+	+	
7	1.4790	2.6458	+	+	+	+	+	+	+
8	1.7321	2.6458	+	+	+	+	+	+	+

DESIGN A.3.5 EIGHT EQUAL WEIGHTS K= 8
N=14
(BOSE-CAMERON) D.F.= 7

OBSERVATIONS	1	1	1	1	1	1	1	1
Y(1)	+	-		-			+	
Y(2)	+	+	-		-			
Y(3)		+	+	-		-		
Y(4)			+	+	-		-	
Y(5)	-			+	+	-		
Y(6)		-			+	+	-	
Y(7)	-		-			+	+	
Y(8)	+	-		+				-
Y(9)		+	-		+			-
Y(10)			+	-		+		-
Y(11)				+	-		+	-
Y(12)	+				+	-		-
Y(13)		+				+	-	-
Y(14)	-		+				+	-
RESTRAINT A	+	+						
RESTRAINT B								+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	B	1	1	1	1	1	1	1	1
1	.4330	.0000								+
1	.4330	.5000							+	
1	.4330	.5000						+		
1	.4330	.5000					+			
1	.4330	.5000				+				
1	.4330	.5000			+					
1	.2500	.5000		+						
1	.2500	.5000	+							
2	.0000	.8660	+	+						
3	.4330	1.2247	+	+	+					
4	.7071	1.5811	+	+	+	+				
5	.9682	1.9365	+	+	+	+	+			
6	1.2247	2.2913	+	+	+	+	+	+		
7	1.4790	2.6458	+	+	+	+	+	+	+	
8	1.7321	2.6458	+	+	+	+	+	+	+	+

DESIGN A.3.6 NINE EQUAL WEIGHTS K= 9
N= 9
(BOSE-CAMERON) D.F.= 1

OBSERVATIONS	1	1	1	1	1	1	1	1	1
Y(1)	+	−	+	−	+	−	+	−	
Y(2)		+	−	+	+	−	−	−	+
Y(3)	+		−	+	−	−	+	+	−
Y(4)	−	−		+	−	+	+	−	+
Y(5)	+	+	+		−	+	−	−	−
Y(6)	+	−	−	−		+	−	+	+
Y(7)	−	−	+	+	+		−	+	−
Y(8)	−	+	+	−	−	−		+	+
Y(9)	−	+	−	−	+	+	+		−
RESTRAINT A	+	+							
RESTRAINT B									+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	RESTRAINTS B	1	1	1	1	1	1	1	1	1
1	.4082	.0000									+
1	.4082	.4714								+	
1	.4082	.4714							+		
1	.4082	.4714						+			
1	.4082	.4714					+				
1	.4082	.4714				+					
1	.4082	.4714			+						
1	.2357	.4714		+							
1	.2357	.4714	+								
2	.0000	.8165	+	+							
3	.4082	1.1547	+	+	+						
4	.6667	1.4907	+	+	+	+					
5	.9129	1.8257	+	+	+	+	+				
6	1.1547	2.1602	+	+	+	+	+	+			
7	1.3944	2.4944	+	+	+	+	+	+	+		
8	1.6330	2.8284	+	+	+	+	+	+	+	+	
9	1.8708	2.8284	+	+	+	+	+	+	+	+	+

DESIGN A.3.7 NINE EQUAL WEIGHTS K= 9
N=12
(BOSE-CAMERON) D.F.= 4

OBSERVATIONS	1	1	1	1	1	1	1	1	1
Y(1)	+	+	+	−	−	−			
Y(2)	+	+	+				−	−	−
Y(3)				+	+	+	−	−	−
Y(4)	+	−		+	−		+	−	
Y(5)	+		−	+		−	+		−
Y(6)		+	−		+	−		+	−
Y(7)	+	−			+	−	−		+
Y(8)	+		−	−	+			−	+
Y(9)		+	−	−		+	+	−	
Y(10)	+	−		−		+		+	−
Y(11)	+		−		−	+	−	+	
Y(12)		+	−	+	−		−		+
RESTRAINT A	+	+							
RESTRAINT B									+

FACTORS FOR COMPUTING ST DEVS

WT	RESTRAINTS A	RESTRAINTS B	1	1	1	1	1	1	1	1	1
1	.4082	.0000									+
1	.4082	.4714								+	
1	.4082	.4714							+		
1	.4082	.4714						+			
1	.4082	.4714					+				
1	.4082	.4714				+					
1	.4082	.4714			+						
1	.2357	.4714		+							
1	.2357	.4714	+								
2	.0000	.8165	+	+							
3	.4082	1.1547	+	+	+						
4	.6667	1.4907	+	+	+	+					
5	.9129	1.8257	+	+	+	+	+				
6	1.1547	2.1602	+	+	+	+	+	+			
7	1.3944	2.4944	+	+	+	+	+	+	+		
8	1.6330	2.8284	+	+	+	+	+	+	+	+	
9	1.8708	2.8284	+	+	+	+	+	+	+	+	+

TABLE 1: MULTIPLIERS OF THE OBSERVATIONS FOR DETERMINING PARAMETER VALUES AND DEVIATIONS FOR DESIGN A.1.2

DESIGN A.1.2 FOUR EQUAL WEIGHTS K= 4
N= 6
D.F.= 3

OBSERVATIONS	1	1	1	1
Y(1)	+	−		
Y(2)	+		−	
Y(3)	+			−
Y(4)		+	−	
Y(5)		+		−
Y(6)			+	−
RESTRAINT	+	+		

PARAMETER VALUES
DIVISOR = 8

OBSERVATIONS	1	1	1	1
Y(1)	2	-2	0	0
Y(2)	1	-1	-3	-1
Y(3)	1	-1	-1	-3
Y(4)	-1	1	-3	-1
Y(5)	-1	1	-1	-3
Y(6)	0	0	2	-2
M(1)	4	4	4	4

DEVIATIONS
DIVISOR = 4

	1	2	3	4	5	6
Y(1)	2	-1	-1	1	1	0
Y(2)	-1	2	-1	-1	0	1
Y(3)	-1	-1	2	0	-1	-1
Y(4)	1	-1	0	2	-1	1
Y(5)	1	0	-1	-1	2	-1
Y(6)	0	1	-1	1	-1	2

MATRIX OF NORMAL EQUATIONS

3	-1	-1	-1	1
-1	3	-1	-1	1
-1	-1	3	-1	0
-1	-1	-1	3	0
1	1	0	0	0

INVERSE DIVISOR = 8

1	-1	0	0	4
-1	1	0	0	4
0	0	3	1	4
0	0	1	3	4
4	4	4	4	0

TABLE 2: MULTIPLIERS OF THE OBSERVATIONS FOR DETERMINING PARAMETER VALUES AND DEVIATIONS FOR DESIGN A.1.2

DESIGN A.1.2 FOUR EQUAL WEIGHTS K= 4
N= 6
D.F.= 3

OBSERVATIONS	1	1	1	1
Y(1)	+	−		
Y(2)	+		−	
Y(3)	+			−
Y(4)		+	−	
Y(5)		+		−
Y(6)			+	−
RESTRAINT				+

PARAMETER VALUES
DIVISOR = 4

OBSERVATIONS	1	1	1	1
Y(1)	1	-1	0	0
Y(2)	1	0	-1	0
Y(3)	2	1	1	0
Y(4)	0	1	-1	0
Y(5)	1	2	1	0
Y(6)	1	1	2	0
M(1)	4	4	4	4

DEVIATIONS
DIVISOR = 4

	1	2	3	4	5	6
Y(1)	2	-1	-1	1	1	0
Y(2)	-1	2	-1	-1	0	1
Y(3)	-1	-1	2	0	-1	-1
Y(4)	1	-1	0	2	-1	1
Y(5)	1	0	-1	-1	2	-1
Y(6)	0	1	-1	1	-1	2

MATRIX OF NORMAL EQUATIONS

3	-1	-1	-1	0
-1	3	-1	-1	0
-1	-1	3	-1	0
-1	-1	-1	3	1
0	0	0	1	0

INVERSE DIVISOR = 4

2	1	1	0	4
1	2	1	0	4
1	1	2	0	4
0	0	0	0	4
4	4	4	4	0

REFERENCES

[1] Benoit, M. J. R., L'Etalonnage des Series de Poids, Travaux et Memoires du Bureau International des Poids et Mesures, No. 13, 1-47 (1907)

[2] Bose, R. C. and Cameron, J. M., The Bridge Tournament Problem and Calibration Designs for Comparing Pairs of Objects, National Bureau of Standards Journal of Research 69B, 323-332 (1965)

[3] Bose, R. C. and Cameron, J. M., Calibration Designs Based on Solutions to the Tournament Problem, National Bureau of Standards Journal of Research 71B, 149-160 (1967)

[4] Cameron, J. M. and Hailes, G. E., Designs for the Calibration of Small Groups of Standards in the Presence of Drift, NBS Technical Note 844 (August 1974)

[5] Chakravarti, I. M. and Suryanarayana, K. V., Partial Difference Sets and Partially Balanced Weighing Designs for Calibration and Tournaments, Journal of Combinatorial Theory 13(3), 426-431 (1972)

[6] Goldman, A. J. and Zelen, M., Weak Generalized Inverses and Minimum Variance Linear Unbiased Estimation, NBS Journal of Research 68B, No. 4, 151-172 (1964)

[7] Hayford, John A., On the Least Squares Adjustment of Weighings, U.S. Coast & Geodetic Survey Report for 1892, Appendix 10, USCGS (1893)

[8] National Bureau of Standards Circular 3, Design and Test of Standards of Mass, 1918. This circular is out of print but

its contents have been incorporated in NBS Handbook 77, Precision Measurement and Calibration, Volume III, Optics, Metrology and Radiation, pp. 588-706. This handbook is available for reference or circulation in most Government Depository Libraries throughout the United States.

[9] Pontius, P. E., Mass and Mass Values, NBS Monograph 133, 39 pages (1974)

[10] Pontius, P. E., Measurement Philosophy of the Pilot Program for Mass Calibration, NBS Technical Note 288, 39 pages (1966)

[11] Suryanarayana, K. V., Contributions to Partially Balanced Weighing Designs, Ph.D. dissertation submitted to the Department of Statistics, University of North Carolina at Chapel Hill, Institute of Statistics Mimeo Series No. 621 (1969)

[12] Zelen, M., Linear Estimation and Related Topics, Survey of Numerical Analysis, Chapter 17, edited by J. Todd, McGraw Hill, New York City, 558-584 (1962)

Metrologia 26, 107–113 (1989)

metrologia

An Extended Error Model for Comparison Calibration

C. Croarkin

Statistical Engineering Division, National Institute of Standards and Technology*, Gaithersburg, MD 20899, USA

Received: September 10, 1988 and in revised form November 18, 1988

Abstract

The usual error model for calibration experiments is extended to situations where there are both short-term and long-term random errors of measurement. Such error models are useful where short-term errors are related to instrumentation, and long-term errors are related to operating procedures, environmental factors or changes in the artifacts themselves. The concept of a check standard is advanced for estimating variability and maintaining statistical control of the measurement process.

Introduction

Comparison calibration relates a characteristic of an artifact or instrument to the defined unit for the quantity of interest. A reference standard, whose value has been independently established, is the basis for assigning a value to the unknown artifact. For calibrations at the highest accuracy levels, very precise comparators with linear responses over a small on-scale range are used to quantify small differences between artifacts of the same nominal value. We describe an error model and analysis where two unknowns are compared with two reference standards according to a specific design.

Calibration Model

In the simplest case, an unknown X with value X^*, yet to be determined, is assumed to be related to a reference standard R with known value R^* by

$$X^* = \Lambda + R^*$$

where Λ is small but not necessarily negligible.

Given a measurement x on the unknown and a measurement r on the reference standard, the responses are assumed to be of the form

$$x = \eta + X^* + e_x$$

and (1)

$$r = \eta + R^* + e_r$$

where η is instrumental offset and e_x and e_r are independent random errors which come from a distribution with mean zero and standard deviation σ.

The value of Λ is estimated [1] by the difference $\boldsymbol{\Lambda}$ where

$$\boldsymbol{\Lambda} = x - r \tag{2}$$

and the value assigned to the unknown artifact is based on the known value of the reference standard, R^*, called the restraint, according to

$$\boldsymbol{X}^* = \boldsymbol{\Lambda} + R^* . \tag{3}$$

The standard deviation of this estimate, $\sigma_{\boldsymbol{X}}$, depends on the error structure for $\boldsymbol{X}^*$ which is of the form

$$\boldsymbol{X}^* = X^* + e_x - e_r \tag{4}$$

so that

$$\sigma_{\boldsymbol{X}} = \sqrt{2}\,\sigma . \tag{5}$$

Calibration Designs

A more complicated case involves the calibration of several unknowns, such as a weight set of various denominations or a group of voltage cells in a temperature-controlled enclosure, relative to a single reference standard or group of standards. Any difference measurements which compare unknowns and reference standards with one another and each other are candidates for the calibration procedure.

* Formerly, the U.S. National Bureau of Standards

[1] Boldface type is used to denote a least-squares estimate from the data such as $\boldsymbol{\Lambda}$

A calibration design is a subset of all candidate measurements which admits a least-squares solution for the unknowns. The design is constructed to be parsimonious so as, on one hand, to minimize the number of measurements and, on the other hand, to give estimates with reasonably high precision. We recognize that precision depends on the number of measurements, and Grabe [1] has shown how precision depends on the construction of the design. As we show in this paper, precision can also be limited by other factors.

In the earliest references to designs by Hayford and Benoit [2, 3], the term "weighing design" is used to describe a sequence of measurements for calibrating a weight set. In papers published in the 1960s and 1970s, Bose and Cameron [4, 5] and Chakravarti and Suryanarayana [6] extend the theory and application of designs; Cameron and Eicke [7] solve a problem peculiar to electrical circuits; and Cameron and Hailes [8] discuss the situation where there is drift in the measurement process. Recent publications [9–12] show that designs now enjoy general acceptance in the calibration laboratory and are routinely used for the calibration of mechanical and electrical units of measurement, as well as for mass measurements.

Expanded Calibration Model

Throughout this development, the one constant assumption has been that random errors of measurement are independent and come from a single error distribution (such as the normal distribution). With more precise measurement systems, we are now able to identify situations where these assumptions are called into question and a more realistic model is needed. We find that random errors of measurement for a single design, which takes at most a few hours' time, are not of the same magnitude as errors which afflict the measurement process over the course of several designs or days [2]. Thus, we are forced to admit two error distributions, one that arises in the short-term and one that arises in the long term.

It is convenient to think in terms of short-term instrumental variations and long-term artifact changes caused by environmental conditions and the like. The latter are assumed to vary randomly from design to design and to be constant for a single design. The model in (1) is expanded to include both types of errors so that

$$\begin{aligned} x &= \eta + \{X^* + \delta_X\} + e_x \\ r &= \eta + \{R^* + \delta_R\} + e_r \end{aligned} \tag{6}$$

where e_x and e_r are short-term errors of (1), and δ_X and δ_R, which represent long-term changes associated with X and R, come from a distribution with mean zero and standard deviation σ_b.

The error structure of the estimate, X^*, given by

$$X^* = X^* + \delta_X - \delta_R + e_x - e_r \tag{7}$$

now contains both types of error terms, and the standard deviation σ_X becomes

$$\sigma_X = (2\sigma_b^2 + 2\sigma^2)^{1/2} .$$

Application to Designs

Standard deviations associated with solutions to a design depend upon the error structures of the model. We illustrate with an example where two unknown artifacts X_1 and X_2 with unknown values X_1^* and X_2^* are calibrated relative to two reference standards R_1 and R_2 with values R_1^* and R_2^*. All items have the same nominal value. A design consisting of the six comparisons $d_1, \ldots, d_6$ that can be made among the four items, two at a time, can be represented as:

Obs	R_1	R_2	X_1	X_2
d_1	1	−1		
d_2	1		−1	
d_3	1			−1
d_4		1	−1	
d_5		1		−1
d_6			1	−1

The model that follows from this design is:

$$\begin{aligned}
d_1 &= \{R_1^* + \delta_{R_1}\} - \{R_2^* + \delta_{R_2}\} + \varepsilon_1 \\
d_2 &= \{R_1^* + \delta_{R_1}\} - \{X_1^* + \delta_{X_1}\} + \varepsilon_2 \\
d_3 &= \{R_1^* + \delta_{R_1}\} - \{X_2^* + \delta_{X_2}\} + \varepsilon_3 \\
d_4 &= \{R_2^* + \delta_{R_2}\} - \{X_1^* + \delta_{X_1}\} + \varepsilon_4 \\
d_5 &= \{R_2^* + \delta_{R_2}\} - \{X_2^* + \delta_{X_2}\} + \varepsilon_5 \\
d_6 &= \{X_1^* + \delta_{X_1}\} - \{X_2^* + \delta_{X_2}\} + \varepsilon_6
\end{aligned} \tag{8}$$

[2] The statistical term for this phenomenon is components of error with the errors sometimes referred to as within-time and between-time random errors

The terms $\varepsilon_1, \ldots, \varepsilon_6$ represent random errors of measurement and the terms $\delta_{R_1}, \delta_{R_2}, \delta_{X_1}$, and δ_{X_2} represent random changes in the artifacts. It is assumed that the ε terms come from a distribution with mean zero and standard deviation σ_w and that the δ terms come from a distribution with mean zero and standard deviation σ_b. All random errors are assumed to be mutually independent.

The solution to the design depends on the restraint. If the restraint is taken to be the average of the reference standards or

$$R^* = \tfrac{1}{2}(R_1^* + R_2^*),$$

then least-squares estimates (see, for example, Cameron et al. [13]) are as follows:

$$\begin{aligned}
R_1^* &= \tfrac{1}{8}(\ \ 2d_1 + d_2\ \ + d_3\ \ - d_4\ \ - d_5) + R^* \\
R_2^* &= \tfrac{1}{8}(-2d_1 - d_2\ \ - d_3\ \ + d_4\ \ + d_5) + R^* \\
X_1^* &= \tfrac{1}{8}(\ \ \ \ - 3d_2 - d_3\ \ - 3d_4 - d_5\ \ + 2d_6) + R^* \\
X_2^* &= \tfrac{1}{8}(\ \ \ \ - d_2\ \ - 3d_3 - d_4\ \ - 3d_5 - 2d_6) + R^*
\end{aligned} \tag{9}$$

We rewrite the solutions in terms of model (8) and collect error terms to obtain

$$\begin{aligned}
R_1^* &= R_1^* + \tfrac{1}{8}(\ \ 4\delta_{R_1} - 4\delta_{R_2} + 2\varepsilon_1\ \ + \varepsilon_2\ \ + \varepsilon_3\ \ - \varepsilon_4\ \ - \varepsilon_5) \\
R_2^* &= R_2^* + \tfrac{1}{8}(-4\delta_{R_1} + 4\delta_{R_2} - 2\varepsilon_1\ \ - \varepsilon_2\ \ - \varepsilon_3\ \ + \varepsilon_4\ \ + \varepsilon_5) \\
X_1^* &= X_1^* + \tfrac{1}{8}(-4\delta_{R_1} - 4\delta_{R_2} + 8\delta_{X_1} - 3\varepsilon_2 - \varepsilon_3\ \ - 3\varepsilon_4 - \varepsilon_5\ \ + 2\varepsilon_6) \\
X_2^* &= X_2^* + \tfrac{1}{8}(-4\delta_{R_1} - 4\delta_{R_2} + 8\delta_{X_2} - \varepsilon_2\ \ - 3\varepsilon_3 - \varepsilon_4\ \ - 3\varepsilon_5 - 2\varepsilon_6)
\end{aligned} \tag{10}$$

Associated standard deviations are found from (10) as follows [3]:

$$\sigma_{R_1} = \sigma_{R_2} = (\tfrac{1}{2}\sigma_b^2 + \tfrac{1}{8}\sigma_w^2)^{1/2}$$

and (11)

$$\sigma_{X_1} = \sigma_{X_2} = (\tfrac{3}{2}\sigma_b^2 + \tfrac{3}{8}\sigma_w^2)^{1/2}.$$

The structure of (11) indicates how precision depends on the relationship between the components of error. For all four estimates, the contribution to the total variance from σ_b^2 is four times larger than the contribution from σ_w^2; thus, the size of σ_b relative to σ_w determines to what extent precision is affected by the number of design points.

Check Standard

The quantity σ_b can only be estimated from many designs involving the same artifact. Because calibrations are usually performed on a one-time basis, the prerequisite data for this analysis does not usually exist on the unknown itself. Thus, we designate a check standard for this purpose, and values of the check standard from many designs provide the basis for estimating σ_b.

For designs involving two reference standards, we create a check standard based on the difference between the two reference standards. For the design of (8), this difference

$$C = R_1^* - R_2^* \tag{12}$$

which is independent of the restraint, has an error structure of the form

$$\begin{aligned}
C = R_1^* - R_2^* + \tfrac{1}{8}(&8\delta_{R_1} - 8\delta_{R_2} + 4\varepsilon_1 + 2\varepsilon_2 + 2\varepsilon_3 \\
&- 2\varepsilon_4 - 2\varepsilon_5).
\end{aligned} \tag{13}$$

with associated standard deviation

$$\sigma_C = (2\sigma_b^2 + \tfrac{1}{2}\sigma_w^2)^{1/2}. \tag{14}$$

Hence

$$\sigma_b = (\tfrac{1}{2}\sigma_C^2 - \tfrac{1}{4}\sigma_w^2)^{1/2} \tag{15}$$

and (11) can be reduced to

$$\sigma_{R_1} = \sigma_{R_2} = \tfrac{1}{2}\sigma_C$$

and (16)

$$\sigma_{X_1} = \sigma_{X_2} = \frac{\sqrt{3}}{2}\sigma_C.$$

Estimates of Standard Deviations from the Data

Given n designs with check standard values $C_1, \ldots, C_n$, the quantity σ_C is estimated with $(n-1)$ degrees of freedom by

$$s_C = \left(\frac{1}{n-1}\sum_{i=1}^{n}(C_i - \bar{C})^2\right)^{1/2} \tag{17}$$

where $\bar{C}$ is the average of the check standard values [4].

The standard deviation, σ_w, is estimated from a single design with $(m-k+1)$ degrees of freedom where m is the number of comparisons in the design; k is the number of artifacts; and the additional degree of freedom comes from the known value of the restraint. For the design given by (8), the standard deviation σ_w is estimated with three degrees of freedom by

$$s_w = \left(\frac{1}{3}\sum_{i=1}^{6}(d_i - \bar{d}_i)^2\right)^{1/2} \tag{18}$$

[3] These equations are valid where R^* is known without random error; see the section headed, "A Matrix Approach", for the case where R^* is subject to random error

[4] This method of estimating the standard deviation assumes that the check standard is not drifting over time

where d_i is the predicted value for each difference measurement from the design; i.e.,

$$
\begin{aligned}
d_1 &= R_1^* - R_2^* \\
d_2 &= R_1^* \qquad - X_1^* \\
d_3 &= R_1^* \qquad\qquad - X_2^* \\
d_4 &= \quad R_2^* - X_1^* \\
d_5 &= \quad R_2^* \qquad - X_2^* \\
d_6 &= \qquad\quad X_1^* - X_2^*
\end{aligned}
$$

We can improve the estimate of σ_w by pooling the standard deviations $s_{w_1}, \ldots, s_{w_n}$ from the n designs. The pooled value s_p, which has 3 n degrees of freedom, is computed as

$$s_p = \left(\frac{1}{n} \sum_{i=1}^{n} s_{w_i}^2 \right)^{1/2}. \tag{19}$$

For the purpose of making statements of precision or uncertainty the population standard deviations σ_w, σ_b and σ_C are replaced by their respective estimates in the appropriate equations.

Process Control

Two aspects of statistical process control are relevant in the calibration process. Short-term control for measurements constituting a single design depends on σ_w, and long-term control for calibrations over time depends on σ_b via check standard measurements. The latter depends upon reliable estimates from historical data for the mean, $\bar{C}$, and the standard deviation, s_C. For any new calibration, the check standard value, C, is tested for agreement with past data by a t statistic where

$$t = \frac{|C - \bar{C}|}{s_C}.$$

The process is judged to be in control if

$$t \leq t_{\alpha/2}(\nu)$$

where $t_{\alpha/2}(\nu)$ is the upper $\alpha/2$ percentage point of Student's t distribution [14] with ν degrees of freedom. Otherwise, the calibration is discarded.

Short-term control for each design is exercised by comparing the standard deviation from the design, s_w, with a pooled value s_p from historical data. An F statistic is computed as

$$F = s_w^2 / s_p^2.$$

Short-term precision is regarded as being in control if

$$F \leq F_\alpha(\nu_1, \nu_2)$$

where $F_\alpha(\nu_1, \nu_2)$ is the upper α percentage point of Snedecor's F distribution [15] with ν_1 degrees of freedom in s_w and ν_2 degrees of freedom in s_p. Failure to meet this condition is taken as an indication that precision has deteriorated, and the current calibration results are discarded.

Case Study From Mass Calibration

The National Institute of Standards and Technology (NIST) maintains about thirty check standards for mass calibrations. These check standards, which cover a variety of designs, load levels, and balances, constitute the data base for constructing uncertainties associated with mass calibrations and for implementing statistical control of the calibration process.

The data base, which covers the last twenty years of calibration history at NIST, is reviewed on an annual basis to update uncertainty statements and to expose any trends or anomalies in the process. Standard deviations from the designs, s_w, are pooled by balance. Standard deviations for each check standard, s_C, are estimated by (17).

Analysis confirms that the long-term component of error, s_b, is negligible for the mass-calibration process except at the critical kilogram level. The majority of mass calibrations at NIST start at the kilogram level using the design of (8) with the restraint as the average of two reference kilograms and a check standard C as defined by (12). Standard deviations for this process are shown in the table below.

Standard Deviations at the Kilogram Level

Source	Notation	Eq.	Std. dev.
Kg balance	s_p	(19)	0.0316 mg
Check standard	s_C	(17)	0.0277 mg
Long-term change	s_b	(15)	0.0116 mg
Unknowns	s_{X_1}, s_{X_2}	(16)	0.0240 mg

Weights other than kilograms are related to the NIST unit of mass via a hierarchy of designs where the restraint for each design is taken from the solution to the previous design. For example, at the kilogram level, the unknown X_2 is a group of weights totaling a kilogram; the group constitutes the starting restraint for the next design in the series. Thus, any random error that influences the value assigned to X_2 is propagated to all other weights.

Application to Other Designs

The standard deviation associated with a measurement must be defined on a design-by-design basis. A

matrix approach is outlined in the next section; also see Croarkin [16, 17] for specific formulations for a design involving two reference standards and three unknowns and a design involving four reference standards and four unknowns.

The problem of definition can sometimes be avoided by judicious choice of a check standard. If one chooses a check standard with the same error structure as the artifacts being calibrated, then the standard deviation for the check standard also applies to the calibrated artifacts. For example, if we make all ten comparisons among five artifacts of the same nominal value, where one artifact is a designated check standard, then the check standard will have the same error structure as the unknowns.

A Matrix Approach

A matrix approach is outlined for estimating components of variance for any measurement design where there are both short-term random errors of measurement and long-term random changes in the artifacts. We also allow for the situation where the restraint has been estimated from a previous experiment, and the random errors associated with that measurement process are taken into account.

Given m difference measurements among k artifacts, where some artifacts are regarded as reference standards and some are regarded as test items or unknowns, the model for the measurement process

$$D = A[X^* + \delta] + \varepsilon \qquad (20)$$

is shown in terms of matrix elements. The elements and their respective dimensions are defined as follows:

- $\underset{(m \times 1)}{D}$ a matrix of difference measurements
- $\underset{(m \times k)}{A}$ a matrix of zeroes and ones such that a plus or minus one in the j^{th} position indicates that the j^{th} artifact is involved in the i^{th} comparison and a zero indicates the converse
- $\underset{(k+1)}{X^*}$ a matrix of unknown values for the k artifacts
- $\underset{(k \times 1)}{\delta}$ a matrix of random errors with zero mean and standard deviation σ_b
- $\underset{(m \times 1)}{\varepsilon}$ a matrix of random errors with zero mean and standard deviation σ_w

Because the matrix A has rank $(k-1)$, a solution for an unknown X^*, as shown by Zelen [18], is achieved by imposing upon the system a restraint, or known value for a linear combination of the artifacts. Let the scalar R^* be the restraint, and let $\mathscr{L}_R$ be a vector of zeroes and ones such that a one in the j^{th} position indicates that the j^{th} artifact is in the restraint and a zero indicates the converse.

For example, the vector [5]

$$\underset{(1 \times k)}{\mathscr{L}'_R} = (1 \quad 1 \quad 0 \ldots 0)$$

indicates that the *restraint* R^* is the *summation* for the first two artifacts.

Then a solution can be found from an augmented matrix B where

$$\underset{(k+2)\times(k+2)}{B} = \begin{pmatrix} \underset{(k \times k)}{A'A} & \underset{(k \times 1)}{\mathscr{L}_R} & \underset{(k \times 1)}{A'D} \\ \underset{(1 \times k)}{\mathscr{L}'_R} & \underset{(1 \times 1)}{0} & \underset{(1 \times 1)}{R^*} \\ \underset{(1 \times k)}{0} & \underset{(1 \times 1)}{0} & \underset{(1 \times 1)}{-I} \end{pmatrix}$$

has an inverse of the form

$$\underset{(k+2)\times(k+2)}{B^{-1}} = \begin{pmatrix} \underset{(k \times k)}{Q} & \underset{(k \times 1)}{h} & \underset{(k \times 1)}{X^*} \\ \underset{(1 \times k)}{h'} & \underset{(1 \times 1)}{0} & \underset{(1 \times 1)}{\bullet} \\ \underset{(1 \times k)}{\bullet} & \underset{(1 \times 1)}{\bullet} & \underset{(1 \times 1)}{\bullet} \end{pmatrix}$$

and Q is the covariance matrix; X^* is the vector of estimates for the unknowns; and other entries ($\bullet$) are irrelevant for this application.

The deviations from the fit are given by the vector ζ where

$$\underset{(1 \times m)}{\zeta'} = [D - A X^*]'$$

and the standard deviation for the design σ_w is estimated by

$$s_w = \left(\frac{\zeta'\zeta}{m-k+1}\right)^{1/2}$$

with $m-k+1$ degrees of freedom.

It is now assumed that a check standard C is tracked for many applications of the same design over time. The estimated value of C for any particular design is given by

$$C = \mathscr{L}'_C[X^*]$$

where, for example,

$$\underset{(1 \times k)}{\mathscr{L}'_C} = (1 \quad -1 \quad 0 \ldots 0)$$

indicates that the check standard is the computed difference between the first and second artifacts.

[5] The mark (′) indicates the transpose of a matrix

The standard deviation σ_b can be estimated from the relationship

$$\sigma_b^2 = \frac{\sigma_C^2 - \mathscr{L}_C'[Q]\,\mathscr{L}_C\,\sigma_w^2}{\mathscr{L}_C'[Q\,A'\,A]\,\mathscr{L}_C}$$

where σ_C and σ_w should be estimated from the data of several designs [6].

Now consider a single unknown X_J whose estimated value is

$$X_J^* = [\mathscr{L}_{X_J}]'[X^*] \tag{21}$$

where, for example,

$$\underset{(1\times k)}{\mathscr{L}'_{X_J}} = (0 \quad 1 \quad 0 \ldots 0)$$

signifies that X_J refers to the second artifact in the design. Then the appropriate standard deviation for X_J^* is given by

$$\sigma_{X_J} = ([\mathscr{L}_{X_J}]'[Q\,A'\,A][\mathscr{L}_{X_J}]\,\sigma_b^2 + [\mathscr{L}_{X_J}]'[Q][\mathscr{L}_{X_J}]\,\sigma_w^2)^{1/2} \tag{22}$$

and the standard deviation associated with any linear combination of the unknowns is computed in a similar fashion. At this stage we assume that R^* is known without random error. Eq. (25) is appropriate if this assumption is not valid.

Mass calibration is a special case because values are assigned to sets of weights covering several denominations of mass. All values are related to a starting restraint, such as a kilogram reference standard, by a series of interrelated designs. The first series includes as an unknown, a single weight or a summation of weights, which becomes the restraint for the following series and so on throughout the entire weight set. Thus, we must account for imprecision associated with restraints after the first series.

Let $\mathscr{L}_\Sigma$ be a (k × 1) vector that defines the unknown whose value will be used as the restraint in the next series; this out-going restraint has value

$$X_\Sigma = [\mathscr{L}_\Sigma]'[X^*]. \tag{23}$$

The standard deviation associated with this restraint is computed as

$$\sigma_{X_\Sigma} = ([\mathscr{L}_\Sigma]'[Q\,A'\,A][\mathscr{L}_\Sigma]\,\sigma_b^2 + [\mathscr{L}_\Sigma]'[Q][\mathscr{L}_\Sigma]\,\sigma_w^2)^{1/2}. \tag{24}$$

[6] See the discussion under "Check Standard" and Eqs. (16) and (18)

To account for weights of various denominations, let W be a vector of nominal values for the k weights of the second series so that

$$\underset{(1\times k)}{W'} = (W_1, \ldots, W_k).$$

Now we redefine the design matrix A and the restraint vector $\mathscr{L}_R$ for the next series and let

$$R^* = X_\Sigma$$

and

$$\sigma_R = \sigma_{X_\Sigma}.$$

The matrix B and its inverse B^{-1} follow accordingly. The $\mathscr{L}_{X_J}$ vectors are also redefined for the weights in the series so that estimates can be computed according to (21). Then the appropriate standard deviation for the j^{th} weight, X_J, is given by (25) which follows:

$$\sigma_{X_J} = \left[[\mathscr{L}_{X_J}]'[Q\,A'\,A][\mathscr{L}_{X_J}]\,\sigma_b^2 + [\mathscr{L}_{X_J}]'[Q][\mathscr{L}_{X_J}]\,\sigma_w^2 + \left(\frac{[\mathscr{L}_{X_J}]'W}{[\mathscr{L}_R]'W}\right)^2 \sigma_R^2\right]^{1/2} \tag{25}$$

Standard deviations for the check standard for this series and other combinations of weights are computed similarly. It is noted that the process standard deviations, σ_w and σ_b, depend on the balance and the denominations of weights calibrated in the series; thus, they should be estimated separately for each series.

The process is extended to the next series by redefining the vector $\mathscr{L}_\Sigma$ so that it identifies the out-going restraint whose value is given by (23). Then the standard deviation for this restraint is given by (26) which follows:

$$\sigma_{X_\Sigma} = \left[[\mathscr{L}_\Sigma]'[Q\,A'\,A][\mathscr{L}_\Sigma]\,\sigma_b^2 + [\mathscr{L}_\Sigma]'[Q][\mathscr{L}_\Sigma]\,\sigma_w^2 + \left(\frac{[\mathscr{L}_\Sigma]'W}{[\mathscr{L}_R]'W}\right)^2 \sigma_R^2\right]^{1/2} \tag{26}$$

The standard deviations given by (22) and (24) are appropriate for values estimated in the initial series of weighings where the starting restraint is a known value. For values assigned by subsequent series of weighings, the imprecision of the estimated restraint contributes a component to the total standard deviation. Thus, (25) and (26) are appropriate.

Concluding Remarks

The proposed error model is especially enlightening where short-term errors are related to instrumentation. Then long-term errors are the result of operating procedures or environmental changes which affect the artifacts over time but are reasonably constant in the short-term so as not to affect the standard deviation from the design. Thus, there is motivation for isolating the long-term component in order to ascer-

tain whether precision can be improved given current instrumentation.

Other models may prove more useful or descriptive for other situations. For example, for mass calibrations which deal with weights of the same nominal mass, it is reasonable to assume that random changes in the weights can be characterized by a single error distribution. However, for weights which are not of the same nominal mass, we would allow for errors proportional to mass or, perhaps, to surface area.

Finally, the analysis of the design for four artifacts demonstrates that improved precision cannot always be attained by increasing the number of measurements in the design. The relative magnitudes of σ_w and σ_b and their contribution to the total variance must be understood before one can improve precision.

Acknowledgments. The author wishes to thank Dr. Richard Davis of the National Institute of Standards and Technology, who is responsible for maintaining the NIST unit of mass, for his interest and advice on the subject at hand.

References

1. M. Grabe: Metrologia **14**, 143–146 (1978)
2. J. A. Hayford: *On the least-square adjustment of weighings*, U.S. Coast and Geodetic Survey Report for 1892, Appendix 10
3. J.-R. Benoît: *L'étalonnage des séries de poids*, Trav. Mém. BIPM vol. **13**, 1907
4. R. C. Bose, J. M. Cameron: J. Res. Natl. Bur. Stand., Sect. **B: 79**, 323–332 (1965)
5. R. C. Bose, J. M. Cameron: J. Res. Natl. Bur. Stand., Sect. **B: 71**, 149–160 (1967)
6. I. M. Chakravarti, K. V. Suryanarayana: J. Combin. Theory **13**, 426–431 (1964)
7. J. M. Cameron, W. G. Eicke: *Designs for the Surveillance of the Volt Maintained by a Small Group of Saturated Standard Cells*, Natl. Bur. Stand. (U.S.) Tech. Note 430 (1967)
8. J. M. Cameron, G. E. Hailes: *Designs for the Calibration of Small Groups of Standards in the Presence of Drift*, Natl. Bur. Stand. (U.S.) Tech. Note 844 (1974)
9. C. P. Reeve: *The Calibration of Indexing Tables by Subdivision*, Natl. Bur. Stand. (U.S.) report NBSIR 75-750 (1975)
10. J. M. Cameron, M. C. Croarkin, R. C. Raybold: *Designs for the Calibration of Standards of Mass*, Natl. Bur. Stand (U.S.) Tech. Note 952 (1977)
11. C. P. Reeve: *The Calibration of a Roundness Standard*, Natl. Bur. Stand. (U.S.) report NBSIR 79-1758 (1979)
12. C. P. Reeve: *The Calibration of Angle Blocks by Intercomparison*, Natl. Bur. Stand. (U.S.) report NBSIR 80-1967 (1980)
13. J. M. Cameron, M. C. Croarkin, R. C. Raybold: Ref. 10 (above), pp. 10–11
14. E. S. Pearson, H. O. Hartley (eds.): *Biometrika Tables for Statisticians, Vol. I* (Cambridge University Press, Cambridge 1956), p. 138
15. Ibid., pp. 157–163
16. M. C. Croarkin: *Measurement Assurance Programs, Part II: Development and Implementation*, Natl. Bur. Stand. (U.S.) Spec. Publ. 676-II (1985), pp. 54–58
17. Ibid., pp. 66–75
18. M. Zelen: *Linear Estimation and Related Topics*, in: *Survey of Numerical Analysis*, J. Todd (ed.) (McGraw Hill, New York 1962), pp. 563–565

NBSIR 74-587

THE USE OF THE METHOD OF LEAST SQUARES IN CALIBRATION

J. M. Cameron

Institute for Basic Standards
National Bureau of Standards
Washington, D. C. 20234

September 1974

U. S. DEPARTMENT OF COMMERCE, Frederick B. Dent, Secretary
NATIONAL BUREAU OF STANDARDS, Richard W. Roberts, Director

THE USE OF THE METHOD OF LEAST SQUARES IN CALIBRATION

by

J. M. Cameron

1. Introduction

When more than one measurement is made on the same quantity, we are accustomed to taking an average and we have the feeling that the result is "better" than any single value that might be chosen from the set. Exactly why the average should be better needs some justification and the fundamental step toward a general approach to the problem of measurement was taken by Thomas Simpson in 1755. In showing the advantage of taking an average of values arising from a number of probability distributions, "he took the bold step of regarding errors, not as individual unrelated happenings, but as properties of the measurement process itself . . . He thus opened the way to a mathematical theory of measurement based on the mathematical theory of probability" [3, page 29].

The taking of an average is a special case of the method of least squares for which the original justification by Lengendre in 1805 did not involve any probability considerations but was advanced as a convenient method for the combination of observations. It was Gauss who recognized that one could not arrive at a "best" value unless the probability distribution of the measurement errors were known. In 1798 he showed the optimality of the least squares values when the underlying distribution is normal and in 1821 showed that the method of least squares leads to values of the parameters which have minimum variance among all possible unbiased linear functions* of the observations regardless of the underlying distribution. It is this property that gives the method of least squares its position of dominance among methods of combination of observations.

In this paper the statistical concepts needed for the method of least squares will be stated as a prelude to the usual modern version of the Gauss theorem. The formation of the observational equations and the derivation of the normal equations are illustrated for several situations arising in calibration. The role of restraints in the solution of systems which are not of full rank is discussed. The results are presented in a form designed to facilitate computation.

*An example of a nonlinear function with smaller variance than the average (the "best" linear estimator) is given by the midrange for the rectangular distribution. The midrange (average of the largest and smallest observation) has variance $1/[2(N+1)(N+2)]$ when based on n measurements, whereas the average has variance $1/12N$. Thus if $N \geq 3$, the midrange is to be preferred.

2. The Physical and Statistical Model of an Experiment

In physics, one is familiar with the construction and interpretation of the physical model of an experiment. One has a substantial body of theory on which to base such a model and one need only consider the determination of length by interferometric measurements to remind oneself of the various elements involved: a defined unit, the apparatus, the procedure, the corrections for environmental factors, etc. One realization of the experiment leads to values for the quantities of interest.

But one realizes that a repetition of the experiment will lead to different values--differences for which the physical model does not provide corrections. One is thus confronted with the need for a statistical model to account for the variations encountered in a sequence of measurements. In building the statistical model, one is first faced with the issue of what is meant by a repetition of the experiment--many readings within a few minutes or *ab initio* determinations a week apart.

The objective is to describe the output of the physical process not only in terms of the physical quantities involved but also in terms of the random variation and systematic influences due to environmental, procedural, or instrumental factors in the experiment.

3. Equation of Expected Values of the Observation

If one measured the same quantity again and again to obtain the sequence

$$y_1, y_2, \cdots y_n \cdots$$

then if the process that generates these numbers is "in control," the long run average or *limiting mean*, μ, will exist. By "in control" one means that the values of y behave as random variables from a probability distribution (for a discussion of this topic, see Eisenhart [1]). This limiting mean, μ, is usually called the *expected value* of y designated by the operator E( ) so that the statement becomes in symbols $E(y) = \mu$. Because y is regarded as a random variable one can represent it as

$$y = \mu + \varepsilon$$

where ε is the random component that follows some probability distribution with a limiting mean of zero, i.e., $E(\varepsilon) = 0$.

The quantity μ may involve one or more parameters. Consider the measurement of the difference in length of all distinct pairings of

four gage blocks, A, B, C, D. Denote the 6 measurements by $y_1, y_2, \ldots y_6$, then one may write

$$E(y_1) = A\text{-}B$$
$$E(y_2) = A\text{-}C$$
$$E(y_3) = A\text{-}D$$
$$E(y_4) = B\text{-}C$$
$$E(y_5) = B\text{-}D$$
$$E(y_6) = C\text{-}D$$

Other representations are useful.

Observation	Expected Value: E(y)	Matrix Form: $X\beta$
y_1	A - B	$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix}$
y_2	A - C	
y_3	A - D	
y_4	B - C	
y_5	B - D	
y_6	C - D	

Consider a sequence of measurements of the same quantity in the presence of a linear drift of Δ per observation. The expected values are thus:

Observation	Matrix Form: $X\beta$
$E(y_1) = \mu$	$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \\ 1 & (n-1) \end{bmatrix} \begin{bmatrix} \mu \\ \Delta \end{bmatrix}$
$E(y_2) = \mu + \Delta$	
$E(y_3) = \mu + 2\Delta$	
$E(y_n) = \mu + (n-1)\Delta$	

There is an alternative representation that measures the drift from the central point of the experiment so that the drift is represented by . . . -3Δ, -2Δ, $-\Delta$, 0, Δ, 2Δ, 3Δ . . . for an odd number of observations and by . . . $\frac{-5\Delta}{2}$, $\frac{-3\Delta}{2}$, $\frac{-\Delta}{2}$, $\frac{\Delta}{2}$, $\frac{3\Delta}{2}$, $\frac{5\Delta}{2}$ . . . for an even number of observations.

If, as for example with some gage blocks, the value changes approximately linearly with time; then one can represent the observation as follows:

Expected Value E(y)	Matrix Form: $X\beta$
$E(y_1) = \alpha + \beta x_1$	$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$
$E(y_2) = \alpha + \beta x_2$	
.	
.	
.	
$E(y_n) = \alpha + \beta x_n$	

The sequence of measurements for the intercomparison of 4 gage blocks is as follows:

Observation	Expected Value: E(y)	Matrix Form: $X\beta$
y_1	S. - S.. - $7\Delta/2$	$\begin{bmatrix} 1 & -1 & 0 & 0 & -7 \\ -1 & 0 & 0 & 1 & -5 \\ 0 & 0 & 1 & -1 & -3 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 1 \\ -1 & 0 & 0 & 1 & 3 \\ 1 & 0 & -1 & 0 & 5 \\ 0 & -1 & 1 & 0 & 7 \end{bmatrix} \begin{bmatrix} S. \\ S.. \\ X \\ Y \\ \Delta/2 \end{bmatrix}$
y_2	Y - S. - $5\Delta/2$	
y_3	X - Y - $3\Delta/2$	
y_4	S..- X - $\Delta/2$	
y_5	S..- Y + $\Delta/2$	
y_6	Y - S. + $3\Delta/2$	
y_7	S. - X + $5\Delta/2$	
y_8	X - S.. + $7\Delta/2$	

(Note that for simplicity, $\Delta/2$ is regarded as the parameter.)
For a detailed analysis of this and related experimental arrangements, see J. M. Cameron and G. E. Hailes [1]. The notation is that used in [1] where S. and S.. refer to reference standards and X and Y are the objects being calibrated.

If, as often occurs in the intercomparison of electrical standards, the comparator has a left-right polarity effect, this can be represented as an additive effect, α, as shown below for the intercomparison of 5 standards.

Observation	Expected Value: E(y)	Matrix Form: Xβ						
y_1	$A - B + \alpha$	$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 1 \end{bmatrix}$						$\begin{bmatrix} A \end{bmatrix}$
y_2	$B - C + \alpha$	$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 1 \end{bmatrix}$						$\begin{bmatrix} B \end{bmatrix}$
y_3	$C - D + \alpha$	$\begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}$						$\begin{bmatrix} C \end{bmatrix}$
y_4	$D - E + \alpha$	$\begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 1 \end{bmatrix}$						$\begin{bmatrix} D \end{bmatrix}$
y_5	$-A + E + \alpha$	$\begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$						$\begin{bmatrix} E \end{bmatrix}$
y_6	$-A + D + \alpha$	$\begin{bmatrix} -1 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}$						$\begin{bmatrix} \alpha \end{bmatrix}$
y_7	$B - D + \alpha$	$\begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 1 \end{bmatrix}$						
y_8	$-B + E + \alpha$	$\begin{bmatrix} 0 & -1 & 0 & 0 & 1 & 1 \end{bmatrix}$						
y_9	$C - E + \alpha$	$\begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 1 \end{bmatrix}$						
y_{10}	$A - C + \alpha$	$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}$						

4. Statistical Independence

The sequence of differences from a zero measurement, y_0,

$$A: \quad y_1-y_0,\ y_2-y_0,\ y_3-y_0, \ldots, y_n-y_0, \ldots$$

are clearly dependent because an error in y_0 will be common to all. Similarly, the successive differences

$$B: \quad y_2-y_1,\ y_3-y_2, \ldots, y_n-y_{n-1}, \ldots$$

will be correlated in pairs because an error in y_n affects both the (n-1)st and n-th difference.

If it is assumed in both cases that each y_i has the form $\mu_i = \mu_i + \varepsilon_i$ where $E(\varepsilon_i) = 0$, Var $(\varepsilon_i) = \sigma^2$ and cov $(\varepsilon_i, \varepsilon_j) = 0$, then the variance of the differences for sequence A is, as one would expect,

$$V(y_i - y_0) = 2\sigma^2$$

and the covariance of two differences is

$$\text{cov}\ (y_i - y_0,\ y_j - y_0) = E[(\varepsilon_i - \varepsilon_0)(\varepsilon_j - \varepsilon_0)] = E(\varepsilon_0^2) = \sigma^2$$

because terms of the form $E(\varepsilon_i, \varepsilon_j) = 0$

For sequence B the variance is also $V(y_i - y_{i-1}) = 2\sigma^2$ and the covariance terms are

$$\text{cov}(y_i - y_{i-1},\ y_j - y_{j-1}) = E[(\varepsilon_i - \varepsilon_{i-1})(\varepsilon_j - \varepsilon_{j-1})] = \begin{cases} 0 & \text{if } |i-j| \geq 2 \\ -\sigma^2 & \text{if } |i-j| = 1 \end{cases}$$

These variance-covariance relationships can be represented in matrix form:

Sequence A: $V = \begin{bmatrix} 2 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ \vdots & & & & \\ 1 & 1 & 1 & \cdots & 2 \end{bmatrix} \sigma^2$ Sequence B: $V = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & \cdots & 2 \end{bmatrix} \sigma^2$

All are familiar with the phenomenon of much closer agreement among measurements taken immediately after each other when compared to a sequence of values taken days or weeks apart. The simplest statistical model for this case is that each day has its own limiting mean, $\mu_i = \mu + \delta_i$, where $E(\delta_i) = 0$, $\text{Var}(\delta_i) = \sigma_B^2$, $\text{Cov}(\delta_i, \delta_j) = 0$, and the successive values on each day have the form

$$y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \delta_i + \varepsilon_{ij}$$

where $E(\varepsilon_{ij}) = 0$, $\text{Var}(\varepsilon_{ij}) = \sigma_W^2$, $\text{Cov}(\varepsilon_{ij}, \varepsilon_{k\ell}) = 0$, and $\text{Cov}(\varepsilon_{ij}, \delta_k) = 0$.

These three examples serve to illustrate the point that the physical conduct of the experiment is the essential element in dictating the appropriate statistical analysis. In all three cases the correlation among the variables vitiates the usual formula: standard deviation of the mean = $(1/\sqrt{n})$ standard deviation. (See Appendix, Section 1(b).)

It is in the physical conduct of the experiment that one has to build in the independence of the measurements. For Sequence A one could remeasure the zero setting each time or in Sequence B, make an independent duplicate measurement. Ordinarily this is too much of an expense to pay to achieve uncorrelated variables just for a simpler analysis.

Statistical independence is to be desired in the sense that if the successive measurements are highly correlated, then many measurements are only slightly better than a single one. The really important issue is that the proper statistical model be used so that the results are valid.

5. Normal Equations For the Method of Least Squares (independent random variables)

When there are more observations than parameters, the "best" (in the sense of minimum variance) linear unbiased estimates for the parameters are given by the so-called least squares estimators. For example, assume one has the problem of deriving values for A, B, C, and D from the following measurements.

Measurements	Expected Value: E(y)	Matrix Form: $X\beta$
y_1	A	$\begin{bmatrix}1&0&0&0\end{bmatrix}\begin{bmatrix}A\\B\\C\\D\end{bmatrix}$
y_2	B	$\begin{bmatrix}0&1&0&0\end{bmatrix}$
y_3	C	$\begin{bmatrix}0&0&1&0\end{bmatrix}$
y_4	D	$\begin{bmatrix}0&0&0&1\end{bmatrix}$
y_5	A + B	$\begin{bmatrix}1&1&0&0\end{bmatrix}$
y_6	B + C	$\begin{bmatrix}0&1&1&0\end{bmatrix}$
y_7	C + D	$\begin{bmatrix}0&0&1&1\end{bmatrix}$
y_8	D + A	$\begin{bmatrix}1&0&0&1\end{bmatrix}$

An obvious estimator, $\tilde{A}$, is the average of the three values,

	Expected Value
y_1	A
y_5-y_2	(A+B)-B
y_8-y_4	(A+D)-D

so that, assuming independent measurements with variance, σ^2,

$$\tilde{A} = \frac{1}{3}(y_1 + y_5 - y_2 + y_8 - y_4)$$

$$\text{Var}(\tilde{A}) = \frac{5}{9}\sigma^2$$

The least squares estimator is obtained by forming the normal equations (see Appendix, Section 2).

$$3A + B + \quad + D = y_1 + y_5 + y_8$$

$$A + 3B + C \quad = y_2 + y_6 + y_5$$

$$B + 3C + D = y_3 + y_7 + y_6$$

$$A \quad + C + 3D = y_4 + y_8 + y_7$$

The solution gives the following estimators for the parameters.

$$\hat{A} = (7y_1 - 3y_2 + 2y_3 - 3y_4 + 4y_5 - y_6 - y_7 + 4y_8)/15$$

$$\hat{B} = (-3y_1 + 7y_2 - 3y_3 + 2y_4 + 4y_5 + 4y_6 - y_7 - y_8)/15$$

$$\hat{C} = (2y_1 - 3y_2 + 7y_3 - 3y_4 - y_5 + 4y_6 + 4y_7 - y_8)/15$$

$$\hat{D} = (-3y_1 + 2y_2 - 3y_3 + 7y_4 - y_5 - y_6 + 4y_7 + 4y_8)/15$$

Using formula (1.11) of Appendix, gives

$$\text{Var}(\hat{A}) = 105\sigma^2/225 = 21\sigma^2/45 = 7\sigma^2/15$$

which can be compared to the variance of $\tilde{A}$ which was $25\sigma^2/45$. The Gauss theorem on least squares guarantees that no other linear unbiased estimator will have smaller variance.

In matrix form one has

$$(X'X)\hat{\beta} = \begin{bmatrix} 3 & 1 & 0 & 1 \\ 1 & 3 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 1 & 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_8 \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{15} \begin{bmatrix} 7 & -3 & 2 & -3 \\ -3 & 7 & -3 & 2 \\ 2 & -3 & 7 & -3 \\ -3 & 2 & -3 & 7 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} y$$

$$\hat{\beta} = \frac{1}{15} \begin{bmatrix} 7 & -3 & 2 & -3 & 4 & -1 & -1 & 4 \\ -3 & 7 & -3 & 2 & 4 & 4 & -1 & -1 \\ 2 & -3 & 7 & -3 & -1 & 4 & 4 & -1 \\ -3 & 2 & -3 & 7 & -1 & -1 & 4 & 4 \end{bmatrix} y$$

When only differences among a group of objects (such as gage blocks, voltage cells, etc.) are measured the normal equation will not be of full rank so that a unique solution will not exist. For the design involving differences between all distinct pairings of objects the normal equations are, for the case of 4 objects discussed in Section 3,

$$3A - B - C - D = y_1 + y_2 + y_3 = q_1$$

$$-A + 3B - C - D = -y_1 + y_4 + y_5 = q_2$$

$$-A - B + 3C - D = -y_2 - y_4 + y_6 = q_3$$

$$-A - B - C + 3D = -y_3 - y_5 - y_6 = q_4$$

Or in matrix form:

$$X'X\hat{\beta} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \beta = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix} \beta = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{bmatrix} y$$

which can be seen not to be of full rank because the sum of the four equations is zero.

One needs a baseline to which the differences can be referred--a restraint to bring the system of equations up to full rank. If one of the objects were designated as the standard, or if a number (or all) of them were regarded as a reference group whose value was known, values for the items could be obtained.

If the restraint $A = K_o$ is invoked, the normal equations become (using the methods of Appendix, Section 3)

$$\begin{cases} 3A - B - C - D + \lambda = q_1 \\ -A + 3B - C - D = q_2 \\ -A - B + 3C - D = q_3 \\ -A - B - C + 3D = q_4 \\ A = K_o \end{cases} \qquad \begin{bmatrix} 3 & -1 & -1 & -1 & 1 \\ -1 & 3 & -1 & -1 & 0 \\ -1 & -1 & 3 & -1 & 0 \\ -1 & -1 & -1 & 3 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'y \\ K_o \end{bmatrix}$$

The solution is given by

$$\hat{A} = K$$

$$\hat{B} = K + (-2y_1 - y_2 - y_3 + y_4 + y_5)/4$$

$$\hat{C} = K + (-y_1 - 2y_2 - y_3 - y_4 + y_6)/4$$

$$\hat{D} = K + (-y_1 - y_2 - 2y_3 - y_5 - y_6)/4$$

$$\lambda = 0$$

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 0 & 0 & 0 & 0 & 4 \\ 0 & 2 & 1 & 1 & 4 \\ 0 & 1 & 2 & 1 & 4 \\ 0 & 1 & 1 & 2 & 4 \\ 4 & 4 & 4 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ K_o \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 4 \\ -2 & -1 & -1 & 1 & 1 & 0 & 4 \\ -1 & -2 & -1 & -1 & 0 & 1 & 4 \\ -1 & -1 & -2 & 0 & -1 & -1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ K_o \end{bmatrix}$$

The variances of the values are $V(\hat{A}) = 0$; $V(\hat{B}) = V(\hat{C}) = V(\hat{D}) = \sigma^2/2$.

If the restraint $A + B + C + D = K_1$ is invoked, the normal equations become

$$\begin{aligned} 3A - B - C - D + \lambda &= q_1 \\ -A + 3B - C - D + \lambda &= q_2 \\ -A - B + 3C - D + \lambda &= q_3 \\ -A - B - C + 3D + \lambda &= q_4 \\ A + B + C + D &= K_1 \end{aligned} \qquad \begin{bmatrix} 3 & -1 & -1 & -1 & 1 \\ -1 & 3 & -1 & -1 & 1 \\ -1 & -1 & 3 & -1 & 1 \\ -1 & -1 & -1 & 3 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{B} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'y \\ K_1 \end{bmatrix}$$

and the solution is given by

$$\begin{aligned} \hat{A} &= (y_1+y_2+y_3+K_1)/4 \\ \hat{B} &= (-y_1+y_4+y_5+K_1)/4 \\ \hat{C} &= (-y_2-y_4+y_6+K_1)/4 \\ \hat{D} &= (-y_3-y_5-y_6+K_1)/4 \\ \lambda &= 0 \end{aligned}$$

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 3 & -1 & -1 & -1 & 4 \\ -1 & 3 & -1 & -1 & 4 \\ -1 & -1 & 3 & -1 & 4 \\ -1 & -1 & -1 & 3 & 4 \\ 4 & 4 & 4 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ K_1 \end{bmatrix}$$

$$= \frac{1}{16} \begin{bmatrix} 4 & 4 & 4 & 0 & 0 & 0 & 4 \\ -4 & 0 & 0 & 4 & 4 & 0 & 4 \\ 0 & -4 & 0 & -4 & 0 & 4 & 4 \\ 0 & 0 & -4 & 0 & -4 & -4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ K_1 \end{bmatrix}$$

The variances of the values are $V(\hat{A}) = V(\hat{B}) = V(\hat{C}) = V(\hat{D}) = 3\sigma^2/16$.

Although it is a simple matter to change the reference point for the parameters (i.e., change the restraint) after one solution has been found, the corresponding change of variances for the parameter values should not be ignored. These variances are given by the diagonal terms of the inverse of the matrix of normal equation, the inverse being indicated by double brackets in these examples. The difference in variance for $\hat{B}$ in the last example, arises from the fact that in the first case one is concerned only with the difference between A (the standard) and B, whereas in the second case it is the difference between B and the average of the others that is involved.

For completeness, the matrices of normal equations and their inverses for the examples of Section 3 are shown below.

Linear Drift

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \cdot & \\ \cdot & \\ 1 & (n-1) \end{bmatrix} \qquad X'X = \begin{bmatrix} n & n(n-1)/2 \\ n(n-1)/2 & n(n-1)(2n-1)/6 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{12}{n^2(n^2-1)} \begin{bmatrix} n(n-1)(2n-1)/6 & -n(n-1)/2 \\ -n(n-1)/2 & n \end{bmatrix}$$

y a linear function of x

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \\ \cdot & \\ 1 & x_n \end{bmatrix} \qquad X'X = \begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix} \qquad (X'X)^{-1} = \frac{1}{n\Sigma x^2 - (\Sigma x)^2} \begin{bmatrix} \Sigma x^2 & -\Sigma x \\ -\Sigma x & n \end{bmatrix}$$

Gage block design

$$X = \begin{bmatrix} 1 & -1 & 0 & 0 & -7 \\ -1 & 0 & 0 & 1 & -5 \\ 0 & 0 & 1 & -1 & -3 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 1 \\ -1 & 0 & 0 & 1 & 3 \\ 1 & 0 & -1 & 0 & 5 \\ 0 & -1 & 1 & 0 & 7 \end{bmatrix}$$

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix} = \begin{bmatrix} 4 & -1 & -1 & -2 & 0 & 1 \\ -1 & 4 & -2 & -1 & 0 & 1 \\ -1 & -2 & 4 & -1 & 0 & 0 \\ -2 & -1 & -1 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 168 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix}^{-1} = \frac{1}{336} \begin{bmatrix} 35 & -35 & -7 & 7 & 0 & 168 \\ -35 & 35 & 7 & -7 & 0 & 168 \\ -7 & 7 & 91 & 21 & 0 & 168 \\ 7 & -7 & 21 & 91 & 0 & 168 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 168 & 168 & 168 & 168 & 0 & 0 \end{bmatrix}$$

Intercomparison of 5 standards (Sum of all used as restraint)

$$X = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ -1 & 0 & 0 & 0 & 1 & 1 \\ -1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & -1 & 1 \\ 1 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix} = \begin{bmatrix} 4 & -1 & -1 & -1 & -1 & 0 & 1 \\ -1 & 4 & -1 & -1 & -1 & 0 & 1 \\ -1 & -1 & 4 & -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & 4 & -1 & 0 & 1 \\ -1 & -1 & -1 & -1 & 4 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 10 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix}^{-1} = \frac{1}{25}\begin{bmatrix} 4 & -1 & -1 & -1 & 1 & 0 & 5 \\ -1 & 4 & -1 & -1 & -1 & 0 & 5 \\ -1 & -1 & 4 & -1 & -1 & 0 & 5 \\ -1 & -1 & -1 & 4 & -1 & 0 & 5 \\ -1 & -1 & -1 & -1 & 4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 5/2 & 0 \\ 5 & 5 & 5 & 5 & 5 & 0 & 0 \end{bmatrix}.$$

6. Standard Deviation

By substituting the computed values for the parameters into the equations of expected values for the observation, one has a *predicted value* to compare to the actual observation. The difference, d, between the observed and predicted value is called the *deviation* and is used to determine an estimate, s, of the standard deviation, σ, of the process

$$s = \sqrt{\frac{\Sigma d_i^2}{n-k+m}}$$

where n is the number of measurements, k is the number of parameters and m is the number of restraints.

Ordinarily one has available a sequence of values of the standard deviation say $s_1, s_2, s_3, \ldots, s_n$ based on $\nu_1, \nu_2, \nu_3, \ldots, \nu_n$ degrees of freedom. One forms the estimate of σ by combining these in quadrature

$$\hat{\sigma} = \sqrt{\frac{\nu_1 s_1^2 + \nu_2 s_2^2 + \ldots + \nu_n s_n^2}{\nu_1 + \nu_2 + \ldots + \nu_n}}$$

with degrees of freedom $N = \Sigma\nu$. In assigning a standard deviation to the parameters or linear combinations of them, the value $\hat{\sigma}$ is used rather than the value of s from a single experiment.

The variance of the sums of two parameter values is given by adding the corresponding diagonal terms (variances) in the inverse of the matrix of normal equations and the appropriate off diagonal terms (covariances) and multiplying by σ^2. For the case of the intercomparison of 5 standards given at the end of Section 5:

$$\text{s.d. } (\hat{A}+\hat{B}) = \sqrt{\sigma_A^2 + \sigma_B^2 + 2\sigma_{AB}} = \frac{\sigma}{\sqrt{25}}\sqrt{[4+4+2(-1)]} = \frac{\sigma\sqrt{6}}{5}$$

For the variance of the difference, the covariance terms enter negatively so that for the same example

$$\text{s.d.}(\hat{A}-\hat{B}) = \sqrt{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}} = \frac{\sigma}{\sqrt{25}}\sqrt{[4+4-2(-1)]} = \frac{\sigma\sqrt{10}}{5}$$

For other linear combinations, formula 1.10-M of the Appendix would be used.

For the linear function example, the predicted value of y for x_o is $\hat{y}_o = \hat{\alpha} + \hat{\beta}x_o$ which has a variance of

$$\begin{bmatrix} 1 & x_o \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix} \begin{bmatrix} 1 \\ x_o \end{bmatrix} \sigma^2 = (C_{11} + x_o^2 C_{22} + 2x_o C_{12})\sigma^2$$

where the terms C_{11}, C_{12}, C_{22} are the elements of $(X'X)^{-1}$ given in Section 5 for the case of y as a linear function of x.

7. Correlated Measurements

In the previous section it was assumed that the observations were uncorrelated, i.e., that $V(y_i) = \sigma^2$, $\text{cov}(y_i, y_j) = 0$ or in matrix form $V = \text{Var}(y) = \sigma^2 I$ where I is the identity matrix. Section 4 of the Appendix discusses the general case where one knows the matrix, V, of variances and covariances for the observations.

Quite often a transformation of variables can be achieved to obtain variables that are uncorrelated. A simple example is provided by the case of cummulative errors, i.e., in the case where

$$y_1 = \mu_1 + \varepsilon_1$$

$$y_2 = \mu_2 + \varepsilon_1 + \varepsilon_2$$

$$y_3 = \mu_3 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3$$

The variance covariance matrix of the y's assuming $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon) = \sigma^2$, $\text{cov}(\varepsilon_i\varepsilon_j) = 0$ is given by

$$V = \sigma^2 \begin{bmatrix} 1 & 1 & 1 & . & . & 1 \\ 1 & 2 & 2 & & & 2 \\ 1 & 2 & 3 & & & 3 \\ . & & & & & \\ . & & & & & \\ 1 & 2 & 3 & . & . & n \end{bmatrix}$$

If one transforms to variables x_i where

$$x_1 = y_1 \qquad = \mu_1 + \varepsilon_1$$

$$x_2 = y_2 - y_1 \quad = \mu_2 - \mu_1 + \varepsilon_2$$

$$x_3 = y_3 - y_2 \quad = \mu_3 - \mu_2 + \varepsilon_3$$

$$\cdot$$

$$\cdot$$

$$x_n = y_n - y_{n-1} = \mu_n - \mu_{n-1} + \varepsilon_n$$

The expected values and variances become

$$E(X) = \begin{bmatrix} \mu_1 \\ \mu_2 - \mu_1 \\ . \\ . \\ \mu_n - \mu_{n-1} \end{bmatrix} \qquad V(X) = \begin{bmatrix} \sigma^2 & 0 & . & . & 0 \\ 0 & \sigma^2 & & & 0 \\ . & & & & \\ . & & & & \\ 0 & 0 & & & \sigma^2 \end{bmatrix}$$

In matrix form $X = Ty$ where $T = \begin{bmatrix} 1 & 0 & 0 & . & . & 0 & 0 \\ -1 & 1 & 0 & & & 0 & 0 \\ 0 & -1 & 1 & & & 0 & 0 \\ . & & & & & & \\ . & & & & & & \\ 0 & 0 & 0 & & & -1 & 1 \end{bmatrix}$

and if one computes Var(Ty) = TVT', one gets

$$\text{Var}(Ty) = \begin{bmatrix} 1 & 0 & 0 & . & . \\ -1 & 1 & 0 & & \\ 0 & -1 & 1 & & \\ . & & & & \\ . & & & & \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & . & . & . & 1 \\ 1 & 2 & 2 & & & & 2 \\ 1 & 2 & 3 & & & & 3 \\ . & & & & & & \\ . & & & & & & \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & . & . \\ 0 & 1 & -1 & & \\ 0 & 0 & 1 & & \\ . & & & & \\ . & & & & \end{bmatrix} \sigma^2 = \sigma^2 I$$

REFERENCES

(1) Cameron, J. M., and Hailes, G. E., "Designs for the Calibration of Small Groups of Standards in the Presence of Instrumental Drift," NBS Technical Note 844, Aug. 1974.

(2) Eisenhart, C., "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," NBS J. Research, 67C (1963), 161-187.

(3) Eisenhart, C., "The Meaning of 'Least' in Least Squares," J. Wash. Acad. Sci., 54 (1964), 24-33.

(4) Goldman, A. J., and Zelen, M., "Weak Generalized Inverses and Minimum Variances Linear Unbiased Estimation," NBS J. Research 68B (1964) 151-172.

(5) Zelen, M., "Linear Estimation and Related Topics," Chapter 17 of Survey of Numerical Analysis, edited by J. Todd, McGraw Hill, New York, 1957, 558-584.

APPENDIX: FORMULAS FROM STATISTICS

1. Background and Notation

(a) Expected Value

The expected value, μ, of a random variable, y, will be written

$$E(y) = \mu$$

The mean μ may represent a linear function of some basic parameters $\beta_1, \beta_2, \ldots \beta_k$ with known coefficients $x_1, x_2, \ldots, x_k$

$$E(y) = \mu = x_1\beta_1 + x_2\beta_2 + \ldots + x_k\beta_k$$

The expected value of n observed values $y_1, y_2, \ldots, y_n$ can then be written

$$E(y_1) = x_{11}\beta_1 + x_{12}\beta_2 + \ldots + x_{1k}\beta_k \qquad (1.1)$$

$$E(y_2) = x_{21}\beta_1 + x_{22}\beta_2 + \ldots + x_{2k}\beta_k$$

$$\vdots$$

$$E(y_n) = x_{n1}\beta_1 + x_{n2}\beta_2 + \ldots + x_{nk}\beta_k$$

This may be written in matrix notation as

$$\begin{bmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_n) \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \cdots x_{1k} \\ x_{21} & x_{22} \cdots x_{2k} \\ & \\ x_{n1} & x_{n2} \cdots x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \\ \beta_k \end{bmatrix} \qquad (1.1\text{-M})$$

or as $E(y) = X\beta$

where the vectors y and β and the matrix, X, are easily identified.

(b) Variance, Covariance

The variance, σ_i^2, of a random variable, y_i, is defined as

$$\sigma_i^2 = E\{(y_i - \mu_i)^2\} = E(y_i^2) - 2\mu_i E(y_i) + \mu_i^2 = E(y_i^2) - \mu_i^2 \qquad (1.2)$$

and the covariance σ_{ij} of the variables y_i and y_j by

$$\sigma_{ij} = E\{(y_i - \mu_i)(y_j - \mu_j)\} = E(y_i y_j) - \mu_i \mu_j \qquad (1.3)$$

The variance of cy where c is some constant is

$$\text{Var}(cy) = E\{(cy - c\mu)^2\} = c^2\sigma^2 \qquad (1.4)$$

The variance of a sum of two variables

$$\text{Var}(y_1 + y_2) = E\{[y_1 + y_2 - (\mu_1 + \mu_2)]^2\} = E\{[(y_1 - \mu_1) + (y_2 - \mu_2)]^2\}$$

$$= E(y_1 - \mu_1)^2 + E(y_2 - \mu_2)^2 + 2E\{(y_1 - \mu_1)(y_2 - \mu_2)\}$$

$$= \sigma_1^2 + \sigma_2^2 + 2\sigma_{12} \qquad (1.5)$$

which we may write as

$$\sigma_1^2 + \sigma_2^2 + 2\sigma_{12} = \begin{bmatrix}1 & 1\end{bmatrix} \begin{bmatrix}\sigma_1^2 + \sigma_{12} \\ \sigma_{12} + \sigma_2^2\end{bmatrix} = \begin{bmatrix}1 & 1\end{bmatrix} \begin{bmatrix}\sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2\end{bmatrix} \begin{bmatrix}1 \\ 1\end{bmatrix} \qquad (1.5\text{-M})$$

For independent random variables $\sigma_{ij} = 0$ and

$$\text{Var}(\Sigma y_i) = \Sigma\sigma_i^2 \qquad (1.6)$$

EXAMPLE:

$$\text{Var}(ay_1 + by_2 + cy_3) = E\{[(ay_1 - a\mu_1) + (by_2 - b\mu_2) + (cy_3 - c\mu_3)]^2\}$$

$$= a^2\sigma_1^2 + b^2\sigma_2^2 + c^2\sigma_3^2 + 2ab\sigma_{12} + 2ac\sigma_{13} + 2bc\sigma_{23} \qquad (1.7)$$

which may be written as

$$\begin{bmatrix}a & b & c\end{bmatrix} \begin{bmatrix}\sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2\end{bmatrix} \begin{bmatrix}a \\ b \\ c\end{bmatrix} \qquad (1.7\text{-M})$$

(c) Linear Function of Random Variables

A linear function

$$L = a_1y_1 + a_2y_2 + \ldots + a_ny_n \tag{1.8}$$

has expected value

$$E(L) = a_1E(y_1) + a_2E(y_2) + \ldots + a_nE(y_n) \tag{1.9}$$

or in matrix notation

$$E(L) = (a_1\ a_2\ \ldots\ a_n)\begin{bmatrix} E(y_1) \\ E(y_2) \\ E(y_n) \end{bmatrix} = a'\mu \tag{1.9-M}$$

The variance is given, by analogy with (1.7) by

$$V(L) = [a_1\ a_2\ \ldots\ a_n]\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2n} \\ \cdot & & & \\ \cdot & & & \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_n^2 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_n \end{bmatrix} \tag{1.10-M}$$

which reduces to the usual formula

$$V(\Sigma a_iy_i) = \Sigma a_i^2\sigma_i^2 \tag{1.11}$$

if $\sigma_{ij} = 0$.

For two linear functions L_1 and L_2 the covariance term is given by

$$E\{[a_1(y_1-\mu_1) + \ldots\ldots + a_n(y_n-\mu_n)][b_1(y_1-\mu_1) + \ldots\ldots + b_n(y_n-\mu_n)]\}$$

$$= a_1b_1E(y_1-\mu_1)^2 + a_2b_2E(y_2-\mu_2)^2 + \ldots + a_nb_nE(y_n-\mu_n)^2$$

$$+ (a_1b_2 + a_2b_1)E(y_1-\mu_1)(y_2-\mu_2) + (a_1b_3 + a_3b_1)E(y_1-\mu_1)(y_3-\mu_3)$$

$$+ (a_2b_3+a_3b_2)E(y_2-\mu_2)(y_3-\mu_3) + \ldots$$

This reduces to the usual formulas:

$$\begin{aligned} &\text{If } \sigma_{ij} = 0 \qquad \text{then} \quad \text{Cov}\,(L_1, L_2) = \Sigma a_i b_i \sigma_i^2 \\ &\text{If } \sigma_i = \sigma \qquad \text{then} \quad \text{Cov}\,(L_1, L_2) = \sigma^2 \Sigma a_i b_i \end{aligned} \tag{1.12}$$

For the case of $L_1 = a_1y_1 + a_2y_2 + a_3y_3$ and $L_2 = b_1y_1 + b_2y_2 + b_3y_3$, the covariance can be written:

$$(a_1\ a_2\ a_3)\begin{bmatrix} b_1\sigma_1^2 + b_2\sigma_{12} + b_3\sigma_{13} \\ b_2\sigma_2^2 + b_1\sigma_{12} + b_3\sigma_{23} \\ b_3\sigma_3^2 + b_1\sigma_{13} + b_2\sigma_{23} \end{bmatrix} = (a_1\ a_2\ a_3)\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \tag{1.12-M}$$

giving the general formula for the variance and covariance of two linear functions

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix}\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ a_n & b_n \end{bmatrix} \tag{1.13-M}$$

or in general for p such function, i.e., for a pxn matrix A

$$\text{Var}(AY) = AVA' \tag{1.14-M}$$

(d) Quadratic Forms in Random Variables

We have from (1.2)

$$E(y^2) = \sigma^2 + \mu^2 \tag{1.15}$$

We wish to extend this to include the case of a more general quadratic expression in the y's, consider for example

$$E[(ay_1 + by_2)^2] = Ea^2y_1^2 + Eb^2y_2 + 2abE(y_1y_2)$$

$$= a^2\sigma_1^2 + a^2\mu_1^2 + b^2\sigma_2^2 + b^2\mu_2^2 + 2ab\mu_1\mu_2 + 2ab\sigma_{12}$$

which may be displayed as a matrix product as follows:

$$[y_1\ y_2]\begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = [\mu_1\ \mu_2]\begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + [a\ \ b]\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix}$$

This example illustrates the general formula:

$$E[(a_1y_1 + \ldots + a_ny_n)^2] = E\left\{[y_1\ y_2 \ldots y_n]\begin{bmatrix} a_1^2 & a_1a_2 & \ldots & a_1a_n \\ a_2a_1 & a_2^2 & \ldots & a_2a_n \\ \cdot & & & \\ \cdot & & & \\ a_na_1 & a_na_2 & \ldots & a_n^2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix}\right\}$$

$$= [\mu_1 \ldots \mu_n]\begin{bmatrix} a_1^2 & \ldots & a_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{1n} & \ldots & a_n^2 \end{bmatrix}\begin{bmatrix} \mu_1 \\ \cdot \\ \cdot \\ \mu_n \end{bmatrix} + [a_1 \ldots a_n]\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \cdot & & & \\ \cdot & & & \\ \sigma_{1n} & \sigma_{2n} & \cdot & \sigma_n^2 \end{bmatrix}\begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ a_n \end{bmatrix}$$

or

$$E\{y'Ay\} = \mu'A\mu + a'Va \qquad (1.16\text{-M})$$

where $A = \begin{bmatrix} a_1^2 & a_1a_2 & \ldots \\ a_1a_2 & a_2^2 & \ldots \\ \cdot & & \\ \cdot & & \end{bmatrix}$ and $V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots \\ \sigma_{12} & \sigma_2^2 & \ldots \\ \cdot & & \\ \cdot & & \end{bmatrix}$

The last term can be replaced by the trace of AV so that we have

$$E(Y'AY) = \mu'A\mu + \text{Trace}(AV) \qquad (1.17\text{-M})$$

For an excellent treatment of these statistical topics one should consult Zelen [5].

2. The Gauss Theorem on Least Squares (Independent, Equal Variance, Full Rank)

Let the n observations $y_1, y_2, \ldots, y_n$ have expected values

$$\begin{aligned} E(y_1) &= x_{11}\beta_1 + x_{12}\beta_2 \ldots x_{1k}\beta_k \\ E(y_2) &= x_{21}\beta_1 + x_{22}\beta_2 \ldots x_{2k}\beta_k \\ &\cdot \\ &\cdot \\ &\cdot \\ E(y_n) &= x_{n1}\beta_1 + x_{n2}\beta_2 \ldots x_{nk}\beta_k \end{aligned} \qquad (2.1)$$

and be statistically independent with common variance, σ^2. These two conditions can be expressed in matrix form as follows:

$$E(y) = \begin{bmatrix} E\,y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ x_{n1} & x_{n2} & & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} = X\beta \qquad (2.1\text{-M})$$

$$V(y) = \begin{bmatrix} \sigma^2 & o & \ldots & o \\ o & \sigma^2 & \ldots & o \\ o & o & \ldots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

The Gauss theorem states that the minimum variance unbiased linear estimator of any linear function, L, of the parameters, $\beta_1\ \beta_2 \ldots \beta_k$, say

$$L = a_1\beta_1 + a_2\beta_2 + \ldots a_k\beta_k$$

is given by substituting the values of β_i which minimize

$$Q = \Sigma[y_i - (x_{i1}\beta_1 + \ldots + x_{ik}\beta_k)]^2 \qquad (2.2)$$

considered as a function of the β_i. These values, $\hat{\beta}_1, \hat{\beta}_2 \ldots \hat{\beta}_k$ are the solutions to the k equations, called the *normal equations*.

$$\begin{aligned}
&\Sigma x_{i1}^2\hat{\beta}_1 + \Sigma x_{i1}x_{i2}\hat{\beta}_2 + \ldots + \Sigma x_{i1}x_{ik}\hat{\beta}_k = \Sigma x_{i1}y_i \\
&\Sigma x_{i2}x_{i1}\hat{\beta}_1 + \Sigma x_{i2}^2\hat{\beta}_2 + \ldots + \Sigma x_{i2}x_{ik}\hat{\beta}_k = \Sigma x_{i2}y_i \\
&\cdot \\
&\cdot \\
&\cdot \\
&\Sigma x_{ik}x_{i1}\hat{\beta}_1 + \Sigma x_{ik}x_{i2}\hat{\beta}_2 + \ldots + \Sigma x_{ik}^2\hat{\beta}_k = \Sigma x_{ik}y_i
\end{aligned} \qquad (2.3)$$

or in matrix form

$$(X'X)\hat{\beta} = X'y \qquad (2.3\text{-M})$$

The solution to these equations can be written as

$$\hat{\beta} = (X'X)^{-1}X'y \qquad (2.4\text{-M})$$

because X was assumed to be of rank k. The matrix $(X'X)^{-1}$ is the *inverse of the matrix of normal equations* and plays an important role in least squares analysis. Let its elements be denoted by c_{ij} so that

$$(X'X)^{-1} = \begin{bmatrix} c_{11} & c_{12} \ldots c_{1k} \\ c_{21} & c_{22} \ldots c_{2k} \\ \cdot & \\ \cdot & \\ \cdot & \\ c_{k1} & c_{k2} \ldots c_{kk} \end{bmatrix} \qquad (2.5\text{-M})$$

The standard deviation, σ, is estimated from the deviations d_i, where

$$d_i = y_i - (x_{i1}\hat{\beta}_1 + x_{i2}\hat{\beta}_2 + \ldots + x_{ik}\hat{\beta}_k) \qquad (2.6)$$

by the quantity, s,

$$s = \sqrt{\frac{\Sigma d^2}{n-k}} \tag{2.7}$$

and is said to have n-k *degrees of freedom.*

The standard deviation of the values for the coefficients $\hat{\beta}_i$ are given by

$$s.d.(\hat{\beta}_i) = \sigma\sqrt{c_{ii}} \tag{2.8}$$

and for a linear function $L = a_1\hat{\beta}_1 + a_2\hat{\beta}_2 \ldots a_k\hat{\beta}_k$ is [see equation (1.10-M)]

$$s.d.(L) = \sigma\left\{[a_1 \ldots a_k]\begin{bmatrix} c_{11} & \ldots & c_{1n} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ c_{n1} & \ldots & c_{nn} \end{bmatrix}\begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_k \end{bmatrix}\right\}^{1/2} \tag{2.9-M}$$

3. The Gauss Theorem on Least Squares (Independent, Equal Variance, With Restraints)

If the parameters, β_i, are required to satisfy the m linear equations

$$\left\{\begin{array}{l} \psi_1 = b_{11}\beta_1 + b_{12}\beta_2 \ldots b_{1k}\beta_k = K_1 \\ \quad . \\ \quad . \\ \psi_m = b_{m1}\beta_1 + b_{m2}\beta_2 \ldots b_{mk}\beta_k = K_m \end{array}\right. \tag{3.1}$$

or in matrix form

$$B'\beta = K \tag{3.1-M}$$

then using the method of Lagrangian multipliers, it turns out that the minimum variance unbiased linear estimators are given by minimizing

$$F = Q + 2\lambda_1 (\psi_1 - K_1) + 2\lambda_2(\psi_2 - K_2) + \ldots + 2\lambda_m(\psi_m - K_m) \tag{3.2}$$

considered as a function of the β's and λ's. ($2\lambda_i$ is chosen rather than just λ_i so that in setting $\partial F/\partial\beta_i = 0$, a common factor of 2 can be divided out.)

This leads to the normal equations

$$\left\{\begin{array}{l} \Sigma x_1^2\beta_1 + \ldots + \Sigma x_1 x_k\beta_k + b_{11}\lambda_1 + \ldots + b_{m1}\lambda_m = \Sigma x_1 y \\ \quad . \\ \quad . \\ \quad . \\ \Sigma x_k x_1\beta_1 + \ldots + \Sigma x_k^2\beta_k + b_{1k}\lambda_1 + \ldots + b_{mk}\lambda_m = \Sigma x_k y \\ b_{11}\beta_1 + \ldots + \beta_{1k}\beta_k = K_1 \\ \quad . \\ \quad . \\ \quad . \\ b_{m1}\beta_1 + \ldots + \beta_{mk}\beta_k = K_m \end{array}\right. \tag{3.3}$$

or in matrix form

$$\begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ K \end{bmatrix} \qquad (3.3\text{-M})$$

and the solution is given by

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'X & B \\ B' & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ K \end{bmatrix} \qquad (3.4\text{-M})$$

If X'X was already of full rank, then B must be of rank m for the inverse to exist. If X'X is of rank (k-m) and B' consists of m rows, then the indicated inverse will exist if B is orthogonal to X'X, i.e. that $(X'X)B = 0$, and B is of rank m. Also if B is a combination of such an orthogonal set of restraints, denoted by H, and the vectors of X'X, then the inverse exists if the mxm matrix B'H is of rank m, i.e., the determinant $|B'H| \neq 0$.

EXAMPLE: If the differences A-B, B-C, C-D, D-E, E-A are measured, then the 5 measurements y_1, y_2, y_3, y_4, y_5 (assumed independent with equal variance) can be represented as

$$E(y) = \begin{bmatrix} A-B \\ B-C \\ C-D \\ D-E \\ -A+E \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = X\beta$$

$$X'X = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{bmatrix} \qquad \text{rank of X'X is 4}$$

The restraint $A+B+C+D+E = [1\ 1\ 1\ 1\ 1] \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = H' \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = K$

is orthogonal to X'X because H'(X'X) = (1 1 1 1 1) (X'X) = [0 0 0 0 0]. If the given restraint were A + B = K_0, then B' = (1 1 0 0 0) and $|B'H| = 2 \neq 0$ so that the restraint is sufficient to produce a solution.

The standard deviation estimate is changed from that given in formula (2.7) to become

$$s = \sqrt{\frac{\Sigma d^2}{n-k+m}} \qquad \text{degrees of freedom} = (n-k+m) \tag{3.5}$$

where m is the number of restraints.

Formulas (2.8) and (2.9) still apply for the standard deviation of the parameter values and of linear combinations of them.

4. The Gauss Theorem on Least Squares (General Case)

If the observed values $y_1\ y_2\ \ldots\ y_n$ have variances σ_i^2 and covariances σ_{ij} so that

$$\text{Var}\,(y) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & & \sigma_{2n} \\ \cdot & & & \\ \sigma_{1n} & \sigma_{2n} & & \sigma_n^2 \end{bmatrix} = V \tag{4.1-M}$$

and the parameters are subject to the m restraints

$$\begin{cases} b_{11}\beta_1 + \ldots + b_{1k}\beta_k = K_1 \\ \cdot \\ b_{m1}\beta_1 + \ldots + b_{mk}\beta_k = K_m \end{cases} \tag{4.2}$$

or in matrix form

$$B'\beta = K \tag{4.2-M}$$

Then the least squares estimators for β are given by

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'V^{-1}X & B \\ B' & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'V^{-1}y \\ K \end{bmatrix}$$

where as before $\hat{\lambda}' = [\hat{\lambda}_1\ \hat{\lambda}_2\ \ldots\ \hat{\lambda}_m]$ is a vector of Lagrangian multipliers entering into the minimization process.

For a discussion of this general case, the reader is referred to the Goldman-Zelen article [4].

Precision Calibration of Phase Meters

RAYMOND S. TURGEL, SENIOR MEMBER, IEEE, AND DOMINIC F. VECCHIA

Abstract—**Using the calibration of a phase meter with a nominally linear response as an example, a statistical approach is discussed for predicting worst-case offsets of the meter response characteristic from the value of the reference standard. A linear calibration curve is used to model the meter response, and statistical tests are described which test the appropriateness of the model and whether the calculated calibration curve differs significantly from the ideal. Various levels of corrections to be applied can then be determined on the basis of these tests, and limits to offsets are calculated for each of the levels. By extending this approach, it is possible to predict limits of uncertainty when using the calibrated meter to make measurements.**

I. Introduction

THIS paper discusses a statistical treatment of calibration data which leads to the prediction of measurement uncertainties after appropriate corrections are applied to the readings of the calibrated instrument. The method is illustrated using a phase meter as an example.

In any measurement, the "true" value of the measurand is hidden by random effects and systematic offsets inherent in the measuring instrument and the measurement process. The purpose of a calibration is to try to eliminate the systematic offset by determining suitable corrections which, when applied to the instrument reading, bring the measurement result into closer agreement with the reference standard. Since a degree of uncertainty is inherently associated with the process of calibration itself, the corrections for the systematic offset cannot be established precisely. However, it is possible to estimate limits to the uncertainty of the measurement result after the corrections have been applied. The statistical approaches that lead to these estimates are discussed in this paper.

II. Calibration Curve

The calibration of a measuring instrument can be represented mathematically by a "calibration curve" (Fig. 1) which relates the readings of the instrument under test to the corresponding values of the calibration standard. Since random fluctuations tend to mask the limiting mean of the instrument response at any particular test point, the corrections calculated based on predicted values derived from the calibration curve will, in general, give more reliable results than those obtained from the test data directly. A necessary condition is that the calibration curve models the instrument response correctly. Therefore, it is important to test whether the *a priori* assumption that the model fits the instrument response is justified.

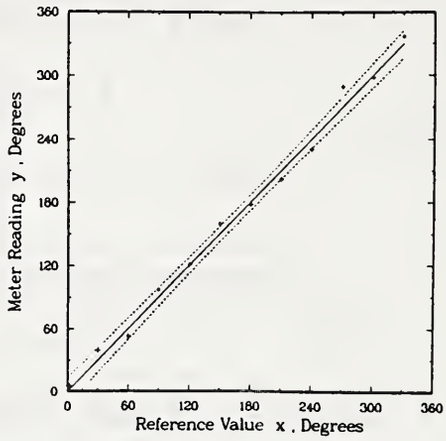


Fig. 1. Calibration curve. Dotted lines indicate 95 percent confidence bands.

Once the correctness of the model is established, the computed calibration curve can be compared with an ideal curve that represents an instrument which agrees perfectly with the standard. To do this, the authors examine whether there are statistically significant differences between the parameters of the computed calibration curve and corresponding parameters of an ideal calibration curve [1]. The outcome of such tests helps to decide what level of corrections, if any, will be necessary.

A. Illustrative Example

To illustrate the application of the above concepts, a simple instrument having a linear response is used as an example. However, the validity of the method is by no means restricted to linear systems and can be extended to more complex relationships. The formulas are derived for the statistical analysis of a nominally linear relationship between the phase angles indicated by a phase meter and the phase angle supplied by a calibrating standard signal source—a phase angle standard.

In this linear case, the calibration curve is a straight line which can be characterized by a slope and an intercept. The corrections to be applied to the phase meter readings can be derived from the linear equation. The extent to which corrections need to be applied must be regarded as a function not only of the calibration data but also of the accuracy specifications of the instrument. In general, there is no point in applying corrections if the uncorrected meter readings are already within the specified accuracy lim-

Manuscript received April 28, 1987; revised July 7, 1987.

R. S. Turgel is with the Electrosystems Division, National Bureau of Standards, Gaithersburg, MD 20899.

D. F. Vecchia is with the Statistical Engineering Division, National Bureau of Standards, Boulder, CO.

IEEE Log Number 8716863.

its, even though applying corrections will always reduce the predicted offset of the measurement result that is computed from the statistical parameters. If the intercept of a linear calibration curve differs appreciably from zero, but the slope is not significantly different from its ideal value, a simple additive constant will bring the measurement result to within the specification limits. There is then no need to calculate individual corrections for every data point, although doing so may result in smaller numerical values for the predicted offset. In the following sections formulas are developed to evaluate the limits of the predicted offset for three levels of applied corrections. A comparison of the numerical values of these limits with the instrument specifications will guide the decision on selecting the appropriate level of corrections.

III. Experimental Procedure Used for Calibrating a Phase Meter

A. Choice of Calibration Procedure

The example of a phase meter calibration is particularly suitable because the straightforward experimental procedure provides a good illustration of a generalized calibration method that could apply equally well to other types of instruments. The output reading of the meter is in the same units and of the same magnitude as the phase angle provided by the calibration source, and no intermediate steps or conversion factors are involved.

B. Circuit Configuration

A phase angle standard [2]–[4] which generates two sinusoidal signals adjustable in phase and independently adjustable in amplitude is used as a calibration source. The standard is designed so that the selected phase angle is known precisely and, therefore, can be used as the reference to which the readings on the phase meter are compared. For convenience, the phase angle standard can be operated via the IEEE-488 bus, allowing the test points to be selected under software control. Signals from the output of the calibration standard are applied directly to the input terminals of the instrument under test.

C. Test Point Selection

For the purpose of the calibration, a "range" is defined by the frequency and the amplitudes of the two test signals. In each range, measurements are made at several phase angles chosen to cover the desired span, usually from 0° to 360°. The exact number of test points is not important, as long as it is large enough to provide the appropriate accuracy for the calculation of the calibration curve. Experience has shown that for a phase meter with a 0.01° resolution, twelve points spread over the 360° span are a satisfactory compromise between the effort involved in making measurements and the accuracy obtained. The results of the measurements in each range are treated as a statistically independent population, and separate accuracy parameter values are computed for each range.

D. Measurement Procedure

For the statistical treatment, it is important to make replicate measurements, usually three or four, at each phase angle tested. This replication provides a measure of the variability of the readings due to the phase meter. Any variation in the output of the phase standard is generally at least an order of magnitude smaller, and is disregarded for the present discussion. The sequence of measurements at the selected phase-angle test points is randomized to minimize time dependent trends and thereby reduce a possible bias in the measurements.

The computer program determines the randomized sequence of phase angles to be tested, and the output of the phase standard is set accordingly. Readings from the phase meter are then recorded and stored in the computer.

IV. Model of the Response Characteristic

A. Estimated Calibration Curve

Using a least squares fit to the data collected, a calibration curve is derived for the response characteristic of the meter under test for each range. For our example, assuming a linear response (phase reading versus phase standard), the model of the calibration data is a straight line of the form

$$y = a + bx + e$$

where

a and b	are the intercept and slope of the straight line,
x	phase angle given by the standard,
y	reading on the phase meter
e	term for the random effect.

If the subscript i ($i = 1, \cdots, k$) denotes the index of the test point, and the subscript j ($j = 1, \cdots, n$) the number of the replicate reading, then the estimated values (denoted by a caret) for the coefficient a and b of the calibration curve $\hat{y} = \hat{a} + \hat{b}x$ can be expressed as [1]

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$\hat{b} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y})(x_i - \bar{x})}{n\sum_{i=1}^{k}(x_i - \bar{x})^2}$$

where the average $\bar{y}$ and the average $\bar{x}$ are

$$\bar{y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}}{nk} \quad \text{and} \quad \bar{x} = \frac{\sum_{i=1}^{k} x_i}{k}.$$

B. Adequacy of the Model

To test whether the calibration data fit the linear model, the fitted value for each phase angle is compared to the average of the repeat measurements at corresponding phase angles by an "F-Test." This test provides a criterion to decide if the calibration data fit the linear model [5]. The calibration data are not consistent with the linear

model if

$$F' = \frac{S_L^2}{S_e^2} > F_{0.01}\{k-2,\ k(n-1)\}$$

where

$$S_L^2 = \frac{n}{k-2}\sum_{i=1}^{k}(\hat{y}_i - \bar{y}_i)^2$$

$$S_e^2 = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2$$

(see footnote 1)

and

$$\hat{y}_i = \hat{a} + \hat{b}x_i, \qquad i = 1, \cdots, k$$

$$\bar{y}_i = \frac{\sum_{j=1}^{n} y_{ij}}{n}, \qquad i = 1, \cdots, k$$

and $F_{0.01}\{k-2,\ k(n-1)\}$ is the upper one-percent point of the F distribution with $k-2$ degrees of freedom in the numerator and $k(n-1)$ degrees of freedom in the denominator. Use of a small significance level ($\alpha = 0.01$) in the test implies that we are only interested in detecting very substantial departures from linearity in the phase meter characteristic.[2]

Similarly, if the linear model is appropriate, levels of significance can be calculated for the coefficients of the calibration curve. These are based on the statistics t_1 and t_2 which test whether $a = 0$ and $b = 1$, respectively. The test statistics are

$$t_1 = \frac{\hat{a}}{S(\hat{a})} \qquad t_2 = \frac{\hat{b}-1}{S(\hat{b})}$$

where $S(\hat{a})$ and $S(\hat{b})$ are the estimated standard deviations of the coefficients.

Using the tables of the Student's t distribution for ($nk-2$) degrees of freedom, the attained levels of statistical significance associated with t_1 and t_2 can be computed. A significance level near zero (<0.05) for t_1 indicates that the intercept is probably different from its "ideal" value of zero, and a significance value near zero for t_2 indicates that the slope is probably different from its "ideal" value of one.

V. Calculation of Phase Meter Uncertainty

When making a phase measurement, the reading obtained from the phase meter differs from the corresponding value of the standard by a systematic offset plus a random effect. As mentioned, the systematic offset can be reduced by an appropriate correction, while the random effect can only be reduced by averaging several readings. For the correction of the systematic component, we consider three cases:

Level 1. No correction applied:

$$\hat{x} = y$$

Level 2. A constant correction applied:

$$\hat{x} = y + C; \quad C = \bar{x} - \bar{y}$$

Level 3. Full correction applied using the calibration curve:

$$\hat{x} = (y - \hat{a})/\hat{b}.$$

For each case, we can estimate the limits to the uncertainty in the phase meter reading. We denote the systematic offset of an uncorrected reading Δ_x at a phase angle x by

$$\Delta_x = E(y \mid x) - x = a + (b-1)x$$

where $E(y \mid x)$ is the expected phase meter reading, and the other symbols are defined as before.

It is evident that the offset is a function of the phase angle as well as the parameters of the calibration curve. In most cases, however, we would like to know the limits of the offset over the entire span of phase angles. For the straight line calibration curve, the equation for the upper and lower limits for the systematic offset at the point x_t, which can be derived [6] from the confidence bands of the calibration curve, shown as dotted lines in Fig. 1, are given by

$$u(\Delta_{x_t}) = A + (B-1)x_t + R(x_t) \qquad (1a)$$

and

$$l(\Delta_{x_t}) = A + (B-1)x_t - R(x_t) \qquad (1b)$$

where

$$R(x_t) = s\sqrt{2F_{0.05}(2,\ nk-2)} \cdot \left[\frac{1}{nk} + \frac{(x_t - \bar{x})^2}{n\sum_{i=1}^{k}(x_i - \bar{x})^2}\right]^{1/2}.$$

In the equations x_t denotes the phase angle given by the standard, s is the standard deviation of fit of the straight line, the coefficients A and B are assigned appropriate values for level 1 and level 2 as shown below, and F is the value of the F distribution for the upper 5-percent point with two degrees of freedom in the numerator and $nk-2$ degrees of freedom in the denominator. Fig. 1 is a plot of the characteristic curve for a phase meter and shows the upper and lower limits of likely systematic offsets. Note that the largest values occur at 0° and at 360°.

A. Estimated Limits for the Systematic Offset

For the three levels of applied corrections, the upper and lower limits of the systematic offset can be calculated

[1]This variance accounts for the random effect, e, in the equation for the straight-line response characteristic.

[2]A special condition arises when the variability about the average is of the same order as the resolution of the meter, and consequently the readings at each test point are truncated to the same numerical value, or a value differing by only one significant digit. In this case the distribution of the deviations will be far from normal, and values of the F-test using tables based on a normal distribution cannot be applied.

and the magnitude of the largest offset for each level $\hat{\Delta}_i$ can be found. These values can then be compared to performance specification limits for the instruments in order to determine what level of corrections need to be applied. It should be noted, however, that the $\hat{\Delta}_i$'s account only for the uncertainty due to the calibration process. If it is desired to include the additional uncertainty that arises from the user's measurement with the meter, then the standard deviation for the user's measurement process must be taken into account as shown below.

The limits of systematic offset can be estimated as follows:

Level 1, no corrections applied:
Using (1a) and (1b), setting $A = \hat{a}$, $B = \hat{b}$

$$\hat{\Delta}_1 = \text{maximum of } \left\{ \left| l(\Delta_0) \right|, \ \left| u(\Delta_0) \right|, \ \left| l(\Delta_{360}) \right|, \ \left| u(\Delta_{360}) \right| \right\}.$$

Level 2, constant correction C applied:
Using (1a) and (1b), setting $A = \hat{a} + C$, $B = \hat{b}$

$$\hat{\Delta}_2 = \text{maximum of } \left\{ \left| l(\Delta_0) \right|, \ \left| u(\Delta_0) \right|, \ \left| l(\Delta_{360}) \right|, \ \left| u(\Delta_{360}) \right| \right\}.$$

Level 3, full correction applied:
Using $u(\Delta_{x_i}) = +R(x_i)/\hat{b}$ and $l(\Delta_{x_i}) = -R(x_i)/\hat{b}$

$$\hat{\Delta}_3 = \text{maximum of } \left\{ \left| l(\Delta_0) \right|, \ \left| u(\Delta_0) \right|, \ \left| l(\Delta_{360}) \right|, \ \left| u(\Delta_{360}) \right| \right\}.$$

B. Estimated Limits for Phase Meter Reading Uncertainty

To obtain an overall estimate of the uncertainty of a phase meter reading, the variability of replicated readings must be included as well as the systematic offset (relative to the standard) given above. The estimate of the standard deviation for the user's measurement process, s_p, must be calculated from the data obtained under the test conditions in the user's laboratory. This standard deviation may well be different than that calculated from the calibration data.

The value for the standard deviation s_p may now be added to the $\hat{\Delta}$ limits of the systematic offsets for the three levels of corrections applied to provide a bound $\hat{E}$ to the uncertainty of the meter reading relative to the value supplied by the standard.

Level 1, no corrections applied:

$$\hat{E}_1 = \hat{\Delta}_1 + s_p \cdot t_{\alpha/2}(\nu_p).$$

Level 2, a constant correction A applied:

$$\hat{E}_2 = \hat{\Delta}_2 + s_p \cdot t_{\alpha/2}(\nu_p).$$

Level 3, full correction applied:

$$\hat{E}_3 = \hat{\Delta}_3 + s_p \cdot t_{\alpha/2}(\nu_p).$$

where,

s_p standard deviation of repeat measurements
ν_p degrees of freedom associated with s_p
$t_{\alpha/2(\nu p)}$ $1 - \alpha/2$ percentile of the Student's t distribution with ν_p degrees of freedom.

The standard deviation s_p should have at least 15 degrees of freedom. Additional repeat measurement can be combined for a pooled value of the standard deviation by computing the square root of the weighted average (weight = ν_i) of the variances of each set of repeat measurements. A sample calculation is shown in the Appendix.

VI. Conclusion

A statistical procedure has been described for the calibration of a phase meter with a nominally linear response. The systematic offset of the meter reading relative to the values provided by the calibration standard can be predicted from a calibration curve. Three levels of correction are considered which will bring the meter readings to within the specified accuracy. The level is selected depending on how closely the actual calibration agrees with an ideal calibration curve. The overall uncertainty of the phase meter reading can be estimated by applying the appropriate level of corrections as well as a term for the random effects of the measurement process.

Appendix

A. Sample Calculation

The predicted values of the phase meter reading are obtained by fitting the calibration data for the set of current, voltage, and frequency conditions to a linear equation which models the average response of the phase meter. In the equations shown below $\hat{y}$ is the predicted value of the phase meter response for a phase angle value x given by the standard:

Test Conditions:	Input A: 100 V Input B: 100 V Frequency: 60 Hz	
Predicted value	$\hat{y} = -0.019402 + 0.999749\ x$	
Standard errors	(0.006952) (0.000036)	
Significance levels[3]	(0.009) (0.000)	
Residual Standard Deviation	$s = 0.02218$	
Lack of fit	$F' = 0.466$	
Significance level[4]	(0.895)	$2s = 0.04436$
Level of Correction	*Correction Equation*	*Limit to Offset*[5]
No correction	$\hat{x} = y$	0.091
Constant correction	$\hat{x} = y + 0.021944$	0.070
Complete calibration curve	$\hat{x} = 1.000251 \cdot y + 0.019407$	0.020

Assuming arbitrarily for this example that with 20 degrees of freedom the user's standard deviation is 20 percent

[3]Significance levels are derived using the statistical t-tests to decide if the intercept and slope of the linear model are different from zero and one, respectively. A level near zero (less than or equal to 0.05) indicates that the associated parameter is probably different from the ideal value.

[4]The significance level of F' is associated with an objective statistical test for the adequacy of a linear model relating the phase meter under test and the NBS assigned values. Levels near zero indicate that the assumption of a straight line relationship may be incorrect.

[5]Phase meter offset relative to the reference standard.

larger than that calculated from the calibration data, then $s_p = 0.027$, and for a the confidence factor $\alpha = 0.05$, the estimated uncertainty of the phase meter readings becomes:

Level of Correction	*Correct Equation*	*Limit to Uncertainty*[6]
No correction	$\hat{x} = y$	0.147
Constant correction	$\hat{x} = y + 0.021944$	0.126
Complete calibration curve	$\hat{x} = 1.000251 \cdot y + 0.019407$	0.076

[6]Uncertainty of phase meter reading relative to the reference standard for a given (user's) standard deviation.

References

[1] C. Croarkin and R. N. Varner, "Measurement assurance for dimensional measurement on integrated-circuit photomasks," NBS Tech. Note 1164, Aug. 1982.

[2] R. S. Turgel, "A precision phase angle calibration standard for frequencies up to 50 kHz," *IEEE Trans. Instrum. Meas.*, vol. IM-34, pp. 509-516, Dec. 1985.

[3] R. S. Turgel, N. M. Oldham, G. N. Stenbakken, and T. H. Kibalo, "NBS phase angle calibration standard," NBS Tech. Note 1144, July 1981.

[4] R. S. Turgel, "NBS 50 kHz phase angle calibration standard," NBS Tech. Note 1220, Apr. 1986.

[5] N. R. Draper and H. Smith, *Applied Regression Analysis*. New York: Wiley, 2nd edition, 1981, p. 37.

[6] ——, *Applied Regression Analysis*. 2nd edition. New York: Wiley, 1981, p. 31.

PRACTICAL REPRESENTATION OF ELECTRICAL UNITS, EFFECTIVE JANUARY 1990

Volume 94, Number 2, March-April 1989
Journal of Research of the National Institute of Standards and Technology

Special Report on Electrical Standards

New Internationally Adopted Reference Standards of Voltage and Resistance

Volume 94 Number 2 March-April 1989

B. N. Taylor

National Institute of Standards and Technology,
Gaithersburg, MD 20899

This report provides the background for and summarizes the main results of the 18th meeting of the Consultative Committee on Electricity (CCE) of the International Committee of Weights and Measures (CIPM) held in September 1988. Also included are the most important implications of these results. The principal recommendations originating from the meeting, which were subsequently adopted by the CIPM, establish new international reference standards of voltage and resistance based on the Josephson effect and the quantum Hall effect, respectively. The new standards, which are to come into effect starting January 1, 1990, will result in improved uniformity of electrical measurements worldwide and their consistency with the International System of Units or SI. To implement the CIPM recommendations in the U.S. requires that, on January 1, 1990, the value of the U.S. representation of the volt be increased by about 9.26 parts per million (ppm) and the value of the U.S. representation of the ohm be increased by about 1.69 ppm. The resulting increases in the U.S. representations of the ampere and watt will be about 7.57 ppm and 16.84 ppm, respectively. The CCE also recommended a particular method, affirmed by the CIPM, of reporting calibration results obtained with the new reference standards that is to be used by all national standards laboratories.

Key words: CCE; CIPM; Consultative Committee on Electricity; International Committee of Weights and Measures; International System of Units; Josephson effect; Josephson frequency-to-voltage quotient; ohm; quantum Hall effect; quantized Hall resistance; SI; volt.

Accepted: December 7, 1988

1. Background

The 18th meeting of the Consultative Committee on Electricity (CCE) of the International Committee of Weights and Measures (CIPM) was held September 27 and 28, 1988, at the International Bureau of Weights and Measures (BIPM), which is located in Sèvres (a suburb of Paris), France. NIST Director E. Ambler, a member of the CIPM and President of the CCE, chaired the meeting and the author attended as NIST representative. Some 30 individuals from 15 countries participated.

As discussed in this journal in the author's 1987 report on the 17th meeting of the CCE held at the BIPM in September 1986 [1], the CCE is one of eight CIPM Consultative Committees which together cover most of the areas of basic metrology. These Committees give advice to the CIPM on matters referred to them. They may, for example, form "Working Groups" to study special subjects and make specific proposals to the CIPM concerning changes in laboratory reference standards and in the definitions of units. As organizational entities of the Treaty of the Meter, one of the responsibilities of the Consultative Committees is to ensure the propagation and improvement of the International System of Units or SI, the unit system used throughout the world. The SI serves as a basis for

the promotion of long-term, worldwide uniformity of measurements which is of considerable importance to science, commerce, and industry.

However, scientific, commercial, and industrial requirements for the long-term repeatability and worldwide consistency of voltage and resistance measurements often exceed the accuracy with which the SI units for such measurements, the volt[1] and the ohm, can be readily realized. To meet these severe demands, it is necessary to establish representations[1] of the volt and ohm that have a long-term reproducibility and constancy superior to the present direct realizations of the SI units themselves.

Indeed, as discussed by the author in reference [1], in 1972 the CCE suggested that the national standards laboratories adopt 483 594 GHz/V exactly as a conventional value of the Josephson frequency-to-voltage quotient for use in maintaining an accurate and reproducible representation of the volt by means of the Josephson effect. While most national laboratories did adopt this value, three decided to use different values. Moreover, it has become apparent that the CCE's 1972 value of this quotient is about 8 parts per million (ppm) smaller than the SI value, implying that representations of the volt based on the 1972 value are actually about 8 ppm smaller than the volt.

It has also become apparent that because most national standards laboratories base their representation of the ohm on the mean resistance of a particular group of wire-wound resistors, the various national representations of the ohm differ significantly from each other and the ohm, and some are drifting excessively. Although the Thompson-Lampard calculable capacitor can be used to realize the ohm with an uncertainty[2] of less than 0.1 ppm, it is a difficult experiment to perform routinely. Hence, the 1980 discovery of the quantum Hall effect (QHE) by K. von Klitzing [6] was enthusiastically welcomed by electrical metrologists because it promised to provide a method for basing a representation of the ohm on invariant fundamental constants in direct analogy with the Josephson effect. The QHE clearly had the potential of eliminating in a relatively straightforward way the problems of nonuniformity of national representations of the ohm, their variation in time, and their inconsistency with the SI.

To address the problems associated with current national representations of the volt and ohm as discussed above, the CCE at its 17th meeting established through Declaration E1 (1986),[3] "Concerning the Josephson effect for maintaining the representation of the volt," the CCE Working Group on the Josephson Effect. The CCE charged the Working Group to propose a new value of the Josephson frequency-to-voltage quotient consistent with the SI value based upon all relevant data that became available by June 15, 1988. Similarly, recognizing the rapid advances made in understanding the QHE since its comparatively recent discovery, the CCE established through Declaration E2 (1986),[3] "Concerning the quantum Hall effect for maintaining a representation of the ohm," the Working Group on the Quantum Hall Effect. The CCE charged the Working Group to (i) propose to the CCE, based upon all relevant data that became available by June 15, 1988, a value of the quantized Hall resistance consistent with the SI value for use in maintaining an accurate and stable national representation of the ohm by means of the QHE; and (ii) develop detailed guidelines for the proper use of the QHE to realize reliably such a representation.[4]

Further, the CCE stated its intention to hold its 18th meeting in September 1988 with a view to recommending that both the proposed new value of the Josephson frequency-to-voltage quotient and the proposed value of the quantized Hall resistance come into effect on January 1, 1990. These values would be used by all those national standards

[1] The volt is the SI unit of electromotive force (emf) and electric potential difference. Occasionally it may be referred to in the literature as the absolute volt. As-maintained volt, representation of the volt, laboratory representation of the volt, "national unit of voltage", "laboratory unit of voltage", "practical realization of the volt", and other similar terms are commonly used to indicate a "practical unit" for expressing measurement results. However, to avoid possible misunderstanding, it is best not to use the word *unit* in this context. The only unit of emf in the SI is, of course, the volt. In keeping with references [2] and [3], from which this report has drawn heavily, we use the expression *representation of the volt* and variations thereof. The expression *reference standard of voltage* is also used occasionally in a similar or related sense. The situation for the ohm and resistance is strictly analogous.

[2] Throughout, all uncertainties are meant to correspond to one standard deviation estimates in keeping with CIPM Recommendation 1 (CI-1986) [4,5].

[3] The complete declaration is given in reference [1], but see also references [5] and [7].

[4] The members of the CCE Working Group on the Josephson Effect were R. Kaarls, Van Swinden Laboratorium (VSL), The Netherlands; B. P. Kibble, National Physical Laboratory (NPL), U.K.; B. N. Taylor, (NIST); and T. J. Witt, Coordinator (BIPM). The members of the CCE Working Group on the Quantum Hall Effect were F. Delahaye (BIPM); T. Endo, Electrotechnical Laboratory (ETL), Japan; O. C. Jones (NPL); V. Kose, Physikalisch-Technische Bundesanstalt (PTB), F. R. G.; B. N. Taylor, Coordinator (NIST); and B. M. Wood, National Research Council of Canada (NRCC), Canada.

laboratories (and others) that base their representation of the volt on the Josephson effect, and that choose to base their representation of the ohm on the QHE. These proposals of the CCE were subsequently approved by the CIPM [8] and by the General Conference of Weights and Measures (CGPM) [9] under whose authority the CIPM functions.

In response to the CCE's directives, each Working Group prepared a report which focused on the review and analysis of the values of the Josephson frequency-to-voltage quotient or quantized Hall resistance in SI units that were available by June 15, 1988; and the derivation of a recommended value for the purpose of establishing an accurate and internationally uniform representation of the volt and of the ohm based on the Josephson effect and on the quantum Hall effect, respectively. Submitted to the CCE in August 1988, the reports include useful background information as well as a discussion as to how the new representations might be used in practice to express calibration results. In keeping with the CCE's charge, the QHE Working Group also prepared a companion report entitled "Technical Guidelines for the Reliable Measurement of the Quantized Hall Resistance." Because unbiased quantized Hall resistance determinations are required for an accurate and reproducible representation of the ohm based on the QHE, these guidelines are of exceptional importance.[5]

2. CCE 18th Meeting Discussion and Principal Decisions

As an aid to the reader, this section of the report also includes some tutorial information.

2.1 Josephson Effect

2.1.1 Definition of Josephson Constant When a Josephson junction is irradiated with microwave radiation of frequency f, its current vs voltage curve exhibits steps at highly precise quantized Josephson voltages U_J. The voltage of the nth step $U_J(n)$, n an integer, is related to the frequency of the radiation by

$$U_J(n) = nf/K_J, \tag{1}$$

where K_J is commonly termed the Josephson frequency-to-voltage quotient [11]. The Working Group on the Josephson Effect (WGJE) proposed that this quotient be referred to as the Josephson constant and, since no symbol had yet been adopted for it, that it be denoted by K_J. It follows from eq (1) that the Josephson constant is equal to the frequency-to-voltage quotient of the $n = 1$ step.

The theory of the Josephson effect predicts, and the experimentally observed universality of eq (1) is consistent with the prediction, that K_J is equal to the invariant quotient of fundamental constants $2e/h$, where e is the elementary charge and h is the Planck constant [11]. For the purpose of including data from measurements of fundamental constants in the derivation of their recommended value of K_J, the WGJE assumed that $2e/h = K_J$. However, K_J is not intended to represent the combination of fundamental constants $2e/h$.

2.1.2 Josephson Effect Reference Standard of Voltage The CCE reviewed the report from the WGJE and discussed at some length the draft recommendation E1 (1988), "Representation of the volt by means of the Josephson effect," prepared jointly by the WGJE and the Working Group on the Quantum Hall Effect. The CCE then agreed:

(i) to use the term "Josephson constant" with symbol K_J to denote the Josephson frequency-to-voltage quotient;

(ii) to accept the WGJE's recommended value of K_J, namely, $K_J = (483\ 597.9 \pm 0.2)$ GHz/V, where the 0.2 GHz/V assigned one-standard-deviation uncertainty corresponds to a relative uncertainty of 0.4 ppm;

(iii) to use this recommended value to define a conventional value of K_J and to denote it by the symbol K_{J-90}, so that $K_{J-90} \overset{\text{def}}{=} 483\ 597.9$ GHz/V exactly. (The subscript 90 derives from the fact that this new conventional value of the Josephson constant is to come into effect starting January 1, 1990, a date reaffirmed by the CCE.) The CCE also noted

(iv) that since K_{J-90} exceeds the CCE's 1972 conventional value of the Josephson constant by 3.9 GHz/V or about 8.065 ppm, the new representation of the volt will exceed that based on the 1972 value by about 8.065 ppm; and further agreed

(v) that because the purpose of the new volt representation is to improve the worldwide uniformity of voltage measurements and their consistency with the SI, laboratories which do not base their national representation of the volt on the Joseph-

[5] The complete reports of the Josephson and Quantum Hall Effect Working Groups including the "Technical Guidelines" (Rapports BIPM 88/77, 88/8, and 88/9) will appear in the proceedings of the CCE's 18th meeting [2]. Additionally, a combined, somewhat condensed version of the two reports may be found in reference [3] and the "Technical Guidelines" in reference [10].

son effect should, on January 1, 1990, adjust the value of their national volt representation so that it is consistent with the new representation. Further, this consistency should be maintained by having a transportable voltage standard periodically calibrated by a laboratory that does base its representation of the volt on the Josephson effect;

(vi) that even if future, more accurate measurements of K_J indicate that the recommended value differs from the SI value by some small amount, the conventional value K_{J-90} should not be altered. Rather, the CCE could simply note the difference between a representation of the volt based on K_{J-90} and the volt; and

(vii) that because an accurate representation of the volt is important to science, commerce, and industry, laboratories should continue their efforts to realize the volt with greater accuracy, either directly or indirectly via measurements of fundamental constants. This could lead to a significant reduction in the uncertainty assigned to the new volt representation.

Having concurred on these points, the CCE edited the draft recommendation E1 (1988) to bring it to final form. The following week it was submitted to the CIPM for approval at its 77th meeting held on October 4–6, 1988, at the BIPM. After some minor editorial changes, the CIPM adopted it as its own recommendation [12]. The following is the English language version (the French language version is the official one and is given in references [2] and [12]):

Representation of the Volt by Means of the Josephson Effect

Recommendation 1 (CI-1988)

The Comité International des Poids et Mesures,

acting in accordance with instructions given in Resolution 6 of the 18th Conférence Générale des Poids et Mesures concerning the forthcoming adjustment of the representations of the volt and the ohm,

considering

—that a detailed study of the results of the most recent determinations leads to a value of 483 597.9 GHz/V for the Josephson constant, K_J, that is to say, for the quotient of frequency divided by the potential difference corresponding to the $n=1$ step in the Josephson effect,

—that the Josephson effect together with this value of K_J can be used to establish a reference standard of electromotive force having a one-standard-deviation uncertainty with respect to the volt estimated to be 4 parts in 10^7, and a reproducibility which is significantly better,

recommends

—that 483 597.9 GHz/V exactly be adopted as a conventional value, denoted by K_{J-90}, for the Josephson constant, K_J,

—that this new value be used from 1st January 1990, and not before, to replace the values currently in use,

—that this new value be used from this same date by all laboratories which base their measurements of electromotive force on the Josephson effect, and

—that from this same date all other laboratories adjust the value of their laboratory reference standards to agree with the new adopted value,

is of the opinion

—that no change in this recommended value of the Josephson constant will be necessary in the foreseeable future, and

draws the attention of laboratories to the fact that the new value is greater by 3.9 GHz/V, or about 8 parts in 10^6, than the value given in 1972 by the Comité Consultatif d'Électricité in its Declaration E-72.

2.2 Quantum Hall Effect

2.2.1 Definition of the von Klitzing Constant The QHE is characteristic of certain high mobility semiconductor devices of standard Hall-bar geometry when in a large applied magnetic field and cooled to a temperature of about one kelvin. For a fixed current I through a QHE device there are regions in the curve of Hall voltage vs gate voltage, or of Hall voltage vs magnetic field depending upon the device, where the Hall voltage U_H remains constant as the gate voltage or magnetic field is varied. These regions of constant Hall voltage are termed Hall plateaus. Under the proper experimental conditions, the Hall resistance of the ith plateau $R_H(i)$, defined as the quotient of the Hall voltage of the ith plateau to the current I, is given by

$$R_H(i)=U_H(i)/I=R_K/i, \qquad (2)$$

where i is an integer [13]. Because $R_H(i)$ is often referred to as the quantized Hall resistance regardless of plateau number, the Working Group on the Quantum Hall Effect (WGQHE) proposed that to avoid confusion, the symbol R_K be used as the Hall voltage-to-current quotient or resistance of the $i=1$ plateau and that it be termed the von Klitzing constant after the discoverer of the QHE. It thus follows from eq (2) that $R_K=R_H(1)$.

The theory of the QHE predicts, and the experimentally observed universality of eq (2) is consistent with the prediction, that R_K is equal to the invariant quotient of fundamental constants h/e^2 [13]. For the purpose of including data from measurements of fundamental constants in the derivation of their recommended value of R_K, the WGQHE assumed that $h/e^2 = R_K$. However, in analogy with K_J, R_K is not intended to represent the combination of fundamental constants h/e^2.

2.2.2 Quantum Hall Effect Reference Standard of Resistance The CCE reviewed the report of the WGQHE and discussed the draft recommendation E2 (1988), "Representation of the ohm by means of the quantum Hall effect," prepared jointly by the two Working Groups. Because of the similarities between the QHE and the Josephson effect, the review and discussion proceeded expeditiously. Indeed, the second half of point (iii) as given here in section 2.1.2 on the Josephson effect and all of points (v), (vi), and (vii) were viewed by the CCE as applying to the quantum Hall effect as well. Also in analogy with the Josephson effect, the CCE agreed:

(i) to use the term "von Klitzing constant" with symbol R_K to denote the Hall voltage to current quotient or resistance of the $i = 1$ plateau;

(ii) to accept the WGQHE's recommended value of R_K, namely, $R_K = (25\ 812.807 \pm 0.005)\ \Omega$, where the 0.005 Ω assigned one-standard-deviation uncertainty corresponds to a relative uncertainty of 0.2 ppm; and

(iii) to use this recommended value to define a conventional value of R_K and to denote it by the symbol R_{K-90}, so that $R_{K-90} \stackrel{\text{def}}{=} 25\ 812.807\ \Omega$ exactly.

The same procedure was followed for draft recommendation E2 (1988) as for E1 (1988) regarding the Josephson effect. The final CIPM English language version is as follows:

Representation of the Ohm by Means of the Quantum Hall Effect

Recommendation 2 (CI-1988)

The Comité International des Poids et Mesures,

acting in accordance with instructions given in Resolution 6 of the 18th Conférence Générale des Poids et Mesures concerning the forthcoming adjustment of the representations of the volt and the ohm,

considering

—that most existing laboratory reference standards of resistance change significantly with time,

—that a laboratory reference standard of resistance based on the quantum Hall effect would be stable and reproducible,

—that a detailed study of the results of the most recent determinations leads to a value of 25 812.807 Ω for the von Klitzing constant, R_K, that is to say, for the quotient of the Hall potential difference divided by current corresponding to the plateau $i = 1$ in the quantum Hall effect,

—that the quantum Hall effect, together with this value of R_K, can be used to establish a reference standard of resistance having a one-standard-deviation uncertainty with respect to the ohm estimated to be 2 parts in 10^7, and a reproducibility which is significantly better,

recommends

—that 25 812.807 Ω exactly be adopted as a conventional value, denoted by R_{K-90}, for the von Klitzing constant, R_K,

—that this value be used from 1st January 1990, and not before, by all laboratories which base their measurements of resistance on the quantum Hall effect,

—that from this same date all other laboratories adjust the value of their laboratory reference standards to agree with R_{K-90},

—that in the use of the quantum Hall effect to establish a laboratory reference standard of resistance, laboratories follow the most recent edition of the "Technical Guidelines for Reliable Measurements of the Quantized Hall Resistance" drawn up by the Comité Consultatif d'Électricité and published by the Bureau International des Poids et Mesures,

and *is of the opinion*

—that no change in this recommended value of the von Klitzing constant will be necessary in the foreseeable future.

2.3 Practical Implementation of Recommendations

As implied by the discussion of section 1, the results of voltage and resistance measurements expressed in terms of representations of the volt and ohm based on the Josephson and quantum Hall effects, respectively, will have a higher precision than the same measurement results expressed in terms of the volt and ohm themselves. Indeed, this is one of the principal reasons for establishing such

representations.[6] The question arises, however, as to how such measurement results should be reported in practice. The Working Groups recognized that the potential for significant confusion internationally could best be eliminated by having each national standards laboratory adopt the same approach. To this end, in their reports the Working Groups identified and considered the advantages and disadvantages of three different approaches to the reporting problem, two of which are both rigorous and correct [2]. In the first, new "practical units" "V_{90}" and "Ω_{90}" are defined; in the second, new, so-called "conventional physical quantities" for electromotive force (and electric potential difference) and resistance, "E_{90}" and "R_{90}," are defined.

The CCE discussed at length the three approaches identified by the Working Groups and concluded that there was an alternative solution, similar to the Working Groups' third approach, that is also rigorous but avoids

(i) defining new practical units of emf and resistance that are likely to differ from the volt and ohm by small amounts and which would be parallel to and thus in competition with the volt and ohm. (Defining such units automatically leads to practical electrical units for current, power, capacitance, etc., thereby giving the appearance that a complete new system of electrical units has been established outside of the SI.) The CCE's alternative solution also avoids

(ii) defining new conventional physical quantities for emf and resistance which are likely to differ from traditional or true emf and resistance by small amounts. (Defining such quantities automatically leads to conventional physical quantities for current, power, capacitance, etc.; and to the peculiar situation of, for example, the same standard cell having both a conventional emf and a true emf.) Further, the alternative solution avoids

(iii) the use of subscripts or other distinguishing symbols of any sort on either unit symbols or quantity symbols. (With the elimination of such subscripts and symbols, for example, those denoting particular laboratories or dates, the national standards laboratories can avoid giving the impression to the users of their calibration services that there is more than one representation of the volt and of the ohm in general use, that there may be significant differences among national realizations of the new volt and ohm representations, and that either the national realizations or the new representations differ significantly from the SI.)

The CCE's solution, which was affirmed by the CIPM at its 77th meeting [12] and which all national standards laboratories are requested to follow, is indicated in the following variation of the example given by the CCE [2] (the treatment of resistance measurements is strictly analogous):

The emf E of an unknown standard cell calibrated in terms of a representation of the volt based on the Josephson effect and the conventional value of the Josephson constant K_{J-90}, may be rigorously expressed in terms of the (SI) volt V as (to be specific):

$$E = (1.018\ 123\ 45)\ \text{V} \pm \epsilon, \tag{3}$$

where ϵ represents the total uncertainty, in volts, and is composed of the following two components: ΔE, the combined uncertainty associated with the calibration itself and with the realization of the Josephson effect volt representation at the particular standards laboratory performing the calibration; and ΔA, the uncertainty with which the ratio K_{J-90}/K_J is known (i.e., it is assumed that $K_{J-90}/K_J = 1 \pm \Delta A$). According to Recommendation 1 (CI-1988), ΔA is 4 parts in 10^7 or 0.4 ppm (assigned one standard deviation).

Since, by international agreement, ΔA is common to all laboratories, the two uncertainties ΔE and ΔA need not be formally combined to obtain the total uncertainty ϵ but may be separately indicated. Hence, the measured emf E may be expressed as

$$E = (1.018\ 123\ 45)\ \text{V} \pm \Delta E \tag{4}$$

for all practical purposes of precision electrical metrology and trade, with ΔA appearing separately on the calibration certificate when the precision of the calibration warrants it. If, for example, $\Delta E/E$ is significantly greater than 0.4 ppm, ΔA may be omitted with negligible effect.

An example of the wording that might be used on a NIST Report of Calibration for a standard cell enclosure for the case where ΔA may not be omitted and which is a variation of the wording given in an example developed by the CCE [2], is as follows:

[6] As noted by the CCE [2], the Josephson and quantum Hall effects and the values K_{J-90} and R_{K-90} cannot be used to define the volt and ohm. To do so would require a change in the status of the permeability of vacuum μ_0 from an exactly defined constant, thereby abrogating the definition of the ampere. It would also give rise to electrical units which would be incompatible with the definition of the kilogram and units derived from it.

Sample Hypothetical NIST Calibration Report

This standard cell enclosure was received (date) under power at its normal operating temperature.

The values given in the table below are based on the results of daily measurements of the differences between the emfs of the cells in this standard and those of NIST working standards calibrated in terms of the Josephson effect using the new conventional value of the Josephson constant internationally adopted for use starting January 1, 1990 (see Note A). The measurements were made in the period from (date) to (date).

Cell number	emf (volts, V)	Uncertainty (microvolts, μV)
1	1.018 119 85	0.27
2	1.018 133 77	0.27
3	1.018 126 42	0.27
4	1.018 141 53	0.27

(Information relating to the measurements and their uncertainties to be given here.)

Note A

The value of the Josephson constant used in this calibration, namely, K_{J-90}=483 597.9 GHz/V exactly, is that adopted by international agreement for implementation starting on January 1, 1990, by all national standards laboratories that base their national representation of the volt (i.e., their national "practical unit" of voltage) on the Josephson effect. Since all such laboratories now use the same conventional value of the Josephson constant while prior to this date several different values were in use, the significant differences which previously existed among the values of some national representations of the volt no longer exist. Moreover, the national standards laboratories of those countries that do not use the Josephson effect for this purpose are requested to maintain their own national representation of the volt so as to be consistent with the above conventional value of the Josephson constant, for example, through periodic comparisons with a laboratory that does use the Josephson effect. An ideal representation of the volt based on the Josephson effect and K_{J-90} is expected to be consistent with the volt as defined in the International System of Units (SI) to within an assigned relative one-standard-deviation uncertainty of 0.4 ppm (0.41 μV for an emf of 1.018 V). Because this uncertainty is the same for all national standards laboratories, it has not been formally included in the uncertainties given in the table. However, its existence must be taken into account when the utmost consistency between electrical and nonelectrical measurements of the same physical quantity is required.

2.4 Future Work on Electrical Units

The ideas agreed upon by the CCE as given in point (vii) in Sect. 2.1.2 on the Josephson effect, and which apply equally as well to the quantum Hall effect, led the CCE to adopt the following formal recommendation which was also approved by the CIPM at its 77th meeting [12].

Realization of the Electrical SI Units

Recommendation E3 (1988)

The Comité Consultatif d'Électricité

recognizing

—the importance to science, commerce and industry of accuracy in electrical measurements,

—the fact that this accuracy depends on the accuracy of the reference standards of the electrical units,

—the very close ties that now exist between electrical metrology and fundamental physical constants,

—the possibility of obtaining more accurate reference standards of the electrical units either directly from the realizations of their definitions or indirectly from measurements of fundamental constants, and

—the continuing need to compare among themselves independent realizations of the units and independent measurements of fundamental constants to verify their accuracy,

recommends

—that laboratories continue their work on the electrical units by undertaking direct realizations of these units and measurements of the fundamental constants, and

—that laboratories pursue the improvement of the means for the international comparison of national standards of electromotive force and electrical resistance.

3. Conclusion

The apparatus currently being used by the national standards laboratories is such that the total experimental uncertainty associated with a particular national representation of the volt based on the Josephson effect generally lies in the range 0.01 to 0.2 ppm. As a consequence, with the worldwide adoption starting January 1, 1990, of the new conventional value of the Josephson constant K_{J-90}, all national representations of the volt should be equivalent to within a few tenths of a ppm. Similarly, the total experimental uncertainty associated with the measurement of quantized Hall resistances also generally lies in the range 0.01 to 0.2 ppm. Hence, with the worldwide adoption starting on January 1, 1990, of a new representation of the ohm based on the QHE and the conventional value of the von Klitzing constant R_{K-90}, all national representations of the ohm should also be equivalent to within a few tenths of a ppm. Moreover, these new national volt and ohm representations should be consistent with the volt and the ohm to better than 0.5 ppm.

In the U.S., the value of the present national representation of the volt maintained by NIST will need to be increased on January 1, 1990, by about 9.26 ppm to bring it into agreement with the new representation of the volt. This is sufficiently large that literally thousands of electrical standards, measuring instruments, and electronic systems throughout the Nation will have to be adjusted or recalibrated in order to conform with the new representation. Most other countries will be required to make a similar change in the value of their present representation of the volt as can be seen from figure 1. On the same date, the value of the U.S. representation of the ohm maintained by NIST will need to be increased by about 1.69 ppm to bring it into agreement with the new representation of the ohm based on the quantum Hall effect. This too is an amount which is of significance to many existing standards, instruments, and systems.

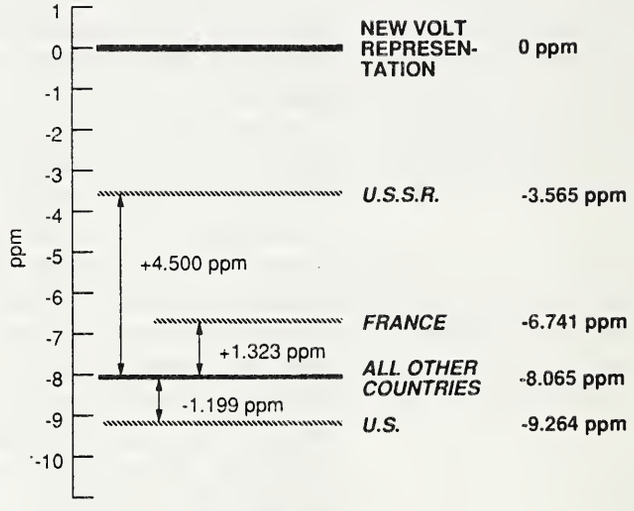


Figure 1. Graphical comparison of the value of the present representation of the volt of various countries as based on the Josephson effect, with the new representation of the volt based on the Josephson effect and the CIPM conventional value of the Josephson constant K_{J-90} which is to come into effect starting on January 1, 1990. The value of the volt representation indicated by "All Other Countries" is based on the conventional value of the Josephson constant stated by the CCE in 1972, namely, 483 594 GHz/V. The countries that currently use this value include Australia, Canada, Finland, F.R.G., G.D.R., Italy, Japan, The Netherlands, and the U.K. The BIPM uses this value as well, but NIST uses 483 593.420 GHz/V. Thus, as the figure shows, on January 1, 1990, the value of the present U.S. volt representation will need to be increased by 9.264 ppm to bring it into conformity with the new representation.

The change required in the value of the national representation of the ohm of other countries varies between a decrease of a few tenths of a ppm to an increase in excess of 3 ppm.

Since $A = V/\Omega$ where A is the ampere as defined in the SI; and $W = V^2/\Omega$ where W is the watt as defined in the SI, the 9.264 ppm and 1.69 ppm increase in the U.S. representation of the volt and of the ohm, respectively, imply that on January 1, 1990, (i) the U.S. representation of the ampere will increase by about 7.57 ppm and (ii) the U.S. electrical representation of the watt will increase by about 16.84 ppm. Because an ideal volt representation based on the Josephson effect and K_{J-90} is expected to be consistent with the volt to within an assigned relative one-standard-deviation uncertainty of 0.4 ppm; and an ideal ohm representation based on the QHE and R_{K-90} is expected to be consistent with the ohm to within an assigned one-standard-deviation uncertainty of 0.2 ppm, ampere and watt representations derived from such ideal volt and ohm representations via the above equations are expected to be consistent with the ampere and watt to within a one-standard-deviation uncertainty of 0.45 ppm and 0.83 ppm, respectively.

The CCE strongly believes, and the author fully concurs, that the significant improvement in the international uniformity of electrical measurements and their consistency with the SI which will result from implementing the new representations of the volt and ohm will be of major benefit to science, commerce, and industry throughout the world; and that the costs associated with implementing the new representations will be far outweighed by these benefits.

About the author: Barry N. Taylor, a physicist, is head of the Fundamental Constants Data Center in the NIST National Measurement Laboratory and Chief Editor of the Journal of Research of the National Institute of Standards and Technology.

4. References

[1] Taylor, B. N., J. Res. Natl. Bur. Stand. **92**, 55 (1987).
[2] BIPM Com. Cons. Électricité **18** (1988), to be published.
[3] Taylor, B. N., and Witt, T. J., Metrologia **26**, No. 1 (1989), to be published.
[4] BIPM Proc.-Verb. Com. Int. Poids et Mesures **54**, 35 (1986). See also Kaarls, R., BIPM Proc. Verb. Int. Poids et Mesures **49**, A1, 26 (1981); and Giacomo, P., Metrologia **18**, 41 (1982).
[5] Giacomo, P., Metrologia **24**, 45 (1987).
[6] v. Klitzing, K., Dorda, G., and Pepper, M., Phys. Rev. Lett. **45**, 494 (1980).
[7] BIPM Com. Cons. Électricité **17**, E6, E88; E10, E92 (1986).
[8] Reference [4], p. 10.
[9] BIPM Comptes Rendus 18e Conf. Gén. Poids et Mesures, p. 100 (1987). See also Giacomo, P., Metrologia **25**, 113 (1988).
[10] Delahaye, F., Metrologia **26**, No. 1 (1989), to be published.
[11] Clarke, J., Am. J. Phys. **38**, 1071 (1970).
[12] BIPM Proc.-Verb. Com. Int. Poids et Mesures **56** (1988), to be published.
[13] The Quantum Hall Effect, eds. R. E. Prange and S. M. Girvin, Springer-Verlag, NY (1987). This book provides a comprehensive review of the quantum Hall effect.

Metrologia 26, 47–62 (1989)

metrologia

New International Electrical Reference Standards Based on the Josephson and Quantum Hall Effects

B. N. Taylor [a] and **T. J. Witt** [b]

[a] National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
[b] Bureau International des Poids et Mesures, Pavillon de Breteuil, F-92312, Sèvres Cedex, France

Received: November 17, 1988

Abstract

We give here the background and basis for the new international electrical reference standards of voltage and resistance that are to come into effect worldwide starting on 1st January 1990. Founded on the Josephson and quantum Hall effects, respectively, these new reference standards will improve significantly the international uniformity of electrical measurements and their consistency with the SI.

1. Introduction

This paper is a combined, somewhat condensed version of the August 1988 reports to the Comité Consultatif d'Électricité (CCE) from the Working Group on the Josephson Effect and from the Working Group on the Quantum Hall Effect (QHE). The authors served as coordinators of the two Working Groups.[1] In keeping with the CCE's charge to the Groups, the principal focus of the reports is the review and analysis of the values of the Josephson frequency-to-voltage quotient and the quantized Hall resistance in units of the Système International d'Unités (SI) that were available by 15 June 1988; and the derivation of a recommended value of each for the purpose of establishing accurate and internationally uniform reference standards of voltage and resistance based on the Josephson and quantum Hall effects, respectively. The reports also include helpful background information as well as a discussion of three approaches to how the new reference standards might be used in practice. The latter topic is omitted from this paper since an alternative approach was eventually adopted by the CCE at its 18th meeting held at the Bureau International des Poids et Mesures (BIPM), 27–28 September 1988.

We apprise the reader that the complete reports of the Josephson and Quantum Hall Effect Working Groups, published by the BIPM as Rapport BIPM-88/7 and Rapport BIPM-88/8, respectively, will appear in the proceedings of the CCE's 18th meeting, BIPM Com. Cons. Électricité **18,** (1988). The QHE Working Group's companion report entitled *Technical Guidelines for Reliable Measurements of the Quantized Hall Resistance*, Rapport BIPM-88/9, will also appear in the same publication and is reprinted in this issue of Metrologia under the authorship of QHE Working Group member F. Delahaye. Because unbiased quantized Hall resistance measurements are required for an accurate and reproducible representation of the ohm based on the QHE, these guidelines are of exceptional importance.

A report on the 18th meeting of the CCE by BIPM Director T. J. Quinn is given in this issue of Metrologia in "News from the BIPM". His report includes the formal recommendations of the CCE as edited and adopted by the Comité International des Poids et Mesures (CIPM) at its meeting on 4–6 October 1988 and the alternative approach mentioned above.

1.1. Background

The CCE is one of eight Consultative Committees to the CIPM. These committees, which together cover most of the areas of basic metrology, give advice to the CIPM on matters referred to them. For example, they may establish temporary or permanent "Working Groups" to study special subjects, coordinate the international work carried out in their respective fields, advise the CIPM about the work of the BIPM in these

[1] The members of the CCE Working Group on the Josephson Effect were R. Kaarls (VSL), B. P. Kibble (NPL), B. N. Taylor (NIST), and T. J. Witt, Coordinator (BIPM). The members of the CCE Working Group on the Quantum Hall Effect were F. Delahaye (BIPM), T. Endo (ETL), O. C. Jones (NPL), V. Kose (PTB), B. N. Taylor, Coordinator (NIST), and B. M. Wood (NRC). (See the Appendix for the laboratory abbreviations used throughout this paper.)

fields, and propose appropriate actions to the CIPM including recommendations concerning changes in the definitions and representations of units. The CIPM may endorse, modify, or reject these recommendations, submitting as appropriate those which will have a very broad impact to the Conférence Générale des Poids et Mesures (CGPM) for final approval. (See Ref. [1] for further details.)

As an organ of the Convention du Mètre, one of the responsibilities of the CCE is to ensure the propagation and improvement of the SI, the unit system used throughout the world. In particular, the SI serves as a basis for the promotion of long-term, world-wide uniformity of electrical measurements which is of considerable technical and economic importance to commerce and industry.

Consequently, at its 13th meeting held in 1972, the CCE suggested that the national standards laboratories adopt 483 594.0 GHz/V as the conventional value of the Josephson frequency-to-voltage quotient for use in realizing and maintaining accurate and stable national representations of the volt[2] by means of the Josephson effect [2]. While most national laboratories did in fact adopt this value, three did not. The US, France, and the USSR adopted values of the quotient which are, respectively, $(1-1.20\times10^{-6})$, $(1+1.32\times10^{-6})$, and $(1+4.50\times10^{-6})$ times the value stated by the CCE in 1972 [3]. As a consequence, the national representations of the volt of these countries differ by -1.20 μV, 1.32 μV, and 4.50 μV, respectively, from the national representations of those countries which use the 1972 value. Moreover, it has recently become evident that the 1972 value is about $(1-8\times10^{-6})$ times the SI value, implying that the national representations of the volt of those countries that have adopted it are about 8 μV smaller than the SI unit [4]. For the US, France, and the USSR, the differences from SI are about -9.2 μV, -6.7 μV, and -3.5 μV, respectively.

The CCE has also become increasingly concerned that, because most national standards laboratories base their representation of the ohm on the mean resistance of a particular group of precision wire-wound standard resistors, and because these artifact standards age, the various national representations of the ohm differ significantly from each other and from the ohm, and some are drifting excessively. Indeed, current evidence indicates that most national representations of the ohm are from a few tenths μΩ larger to a few μΩ smaller than the ohm and that their drift rates lie in the range -0.07 to $+0.07$ μΩ/year [5].

Although in principle a so-called Thompson-Lampard calculable capacitor can be used to realize the ohm with an uncertainty of less than 0.1 μΩ, it is in practice a difficult experiment to carry out routinely; only one laboratory in the world has had such an apparatus in continuous operation since the method was first developed in the early 1960s [6]. Consequently, electrical metrologists enthusiastically welcomed von Klitzing's 1980 discovery of the QHE [7] since it promised to provide a method for basing a representation of the ohm on fundamental constants in much the same manner as the Josephson effect has provided a method for basing a representation of the volt on fundamental constants. The QHE clearly had the potential of virtually eliminating in a relatively simple way the problems of non-uniformity of national representations of the ohm, their variation with time, and their inconsistency with the SI.

1.2. Working Groups on the Josephson and Quantum Hall Effects

To address the issue of non-uniformity of national representations of the volt and their inconsistency with the SI, the CCE at its 17th meeting held in September 1986 established through Declaration E1 (1986), "Concerning the Josephson effect for maintaining the representation of the volt", the CCE Working Group on the Josephson Effect [4]. The CCE charged the Working Group to propose a new value of the Josephson frequency-to-voltage quotient consistent with the SI value based upon all relevant data that become available by 15 June 1988. Similarly, recognizing the rapid advances made in understanding the QHE since its comparatively recent discovery, the CCE established through Declaration E2 (1986), "Concerning the quantum Hall effect for maintaining a representation of the ohm", the Working Group on the Quantum Hall Effect [8]. The CCE charged the Working Group to (i) propose to the CCE, based upon all relevant data that become available by 15 June 1988, a value of the quantized Hall resistance consistent with the SI Value for use in realizing and maintaining accurate and stable national representations of the ohm by means of the QHE; and (ii) develop detailed guidelines for the proper use of the QHE to realize such representations reliably.

Further, the CCE stated its intention of holding its 18th meeting in September 1988 with a view to re-

[2] The volt is the SI unit of electromotive force (emf) or electric potential difference. Occasionally it may be referred to in the literature as the absolute volt. As-maintained volt, representation of the volt, laboratory representation of the volt, "national unit of voltage", "laboratory unit of voltage", practical realization of the volt, or other similar terms are commonly used to indicate a "practical unit" for expressing measurement results. However, to avoid possible confusion, it is preferable not to use the word *unit* in this context. (The only unit of emf in the SI is the volt.) This paper uses the expression *representation of the volt* and variations thereof. The situation for the ohm and resistance is strictly analogous

commending that both the proposed new value of the Josephson frequency-to-voltage quotient and the proposed value of the quantized Hall resistance come into effect on 1st January 1990. These values would be used by all those national standards laboratories (and others) that base their representation of the volt on the Josephson effect, and that choose to base their representation of the ohm on the QHE. These proposals of the CCE were subsequently approved by the CIPM at its 75th meeting in 1986 [9] and by the 18th CGPM in 1987 [10].

1.3. Permanence of the New Representations of the Volt and Ohm

In its discussions leading to Declarations E1 and E2 (1986), the CCE agreed that while world-wide uniformity of electrical measurements can only be assured through the SI, in the particular areas of voltage and resistance, scientific, commerical, and industrial requirements for long-term reproducibility now exceed the accuracy with which the SI units can be readily realized. To meet these very exacting demands, the CCE believes it is necessary that representations of the volt and the ohm be established that have a superior long-term reproducibility and constancy than the present direct realizations of the SI units themselves.

Although the recommended values for the Josephson frequency-to-voltage quotient and quantized Hall resistance upon which the new representations of the volt and ohm are to be based are believed to be consistent with the SI values within their assigned uncertainties, it is recognized that future, more accurate measurements will probably show that the new recommended values differ from the SI values by some small amount. In keeping with the point of view of the CCE, it is envisaged that should such a situation occur, the CCE could simply note the differences between the volt and ohm and their new representations. This would be useful for those workers (mostly in the fields of realizing the electrical units and determining the fundamental physical constants) for whom the small differences may be significant. Since any such differences are expected to be sufficiently small that practical electrical measurements will be unaffected, it is strongly believed that the new recommended values will not need to be altered in the foreseeable future.

However, this last statement must not be interpreted to mean that improved realizations of the volt and ohm are now unnecessary. Because accurate representations of the volt and ohm are important to science, commerce, and industry, it is important for laboratories to continue their efforts to realize the volt and ohm with greater accuracy, either directly or indirectly through measurements of relevant fundamental constants. This could result in a significant reduction of the uncertainties assigned to the new representations.

1.4. Laboratories that do not Use the Josephson and Quantum Hall Effects

The purpose of the new volt and ohm representations is to improve the world-wide uniformity of national representations of the volt and ohm and their consistency with the SI. The question thus arises as to the procedure to be followed by those laboratories which do not base their representations of the volt and ohm on the Josephson and quantum Hall effects. In keeping with the viewpoint expressed by the CCE during its discussions in connection with Declarations E1 and E2 (1986), it is recommended that on 1st January 1990, such laboratories adjust the values of their representations of the volt and ohm so that they are consistent with the new representations. Furthermore, this consistency should be maintained by having transportable voltage and resistance standards periodically calibrated by a laboratory that does base its representations of the volt and ohm on the Josephson and quantum Hall effects, for example BIPM.

2. The Josephson and Quantum Hall Effects

2.1. Josephson Frequency-to-Voltage Quotient

As is now well known, the Josephson effects (ac and dc) are characteristic of weakly coupled superconductors when cooled below their transition temperatures [11]. An example is two thin films of superconducting lead separated by an approximately 1 nm-thick thermally-grown oxide layer.

When, under the proper experimental conditions, such a "Josephson device" is irradiated with electromagnetic radiation of frequency ν, its current-voltage curve exhibits current steps at highly precise quantized Josephson voltages U_J. The voltage of the nth step $U_J(n)$, n an integer, is related to the frequency of the radiation by

$$U_J(n) = n\nu/K_J\,, \tag{1}$$

where K_J is commonly termed the *Josephson frequency-to-voltage quotient* and which will be referred to as the *Josephson constant*. (Since no symbol has yet been adopted for this quotient, the use of K_J is proposed. It follows from (1) that the Josephson constant is equal to the frequency-to-voltage quotient of the $n = 1$ step.)

A significant amount of experimental evidence supports the view that the Josephson constant K_J is a universal quantity. For example, K_J has been shown to be independent of experimental variables such as type of superconductor, temperature, and irradiation fre-

quency and power, to very high precision [12–22]. Indeed, in one experiment it was shown that K_J was the same for two Josephson devices made of different superconducting materials to within a relative difference of 2×10^{-16} [18]. A Josephson device may thus be viewed as a nearly perfect frequency-to-voltage transducer.

The theory of the Josephson effects predicts, and the experimentally observed universality of the Josephson frequency-voltage relation [Eq. (1)] is consistent with the prediction, that K_J is equal to the invariant quotient of fundamental constants $2e/h$, where e is the elementary charge and h is the Planck constant [11, 23–25]. It is thus assumed for the purpose of including data from measurements of fundamental constants in the derivation of the recommended value of K_J that $2e/h = K_J$. (The same assumption was made by the CODATA Task Group on Fundamental Constants in obtaining their 1986 set of recommended values of the constants [26].)

2.2. Hall Voltage-to-Current Quotient or Quantized Hall Resistance

As is also now well known, the quantum Hall effects (integral and fractional) are characteristic of a two-dimensional electron gas (2DEG). A 2DEG may be realized in a high-mobility semiconductor device such as a silicon MOSFET (metal-oxide-semiconductor field-effect transistor) or $GaAs-Al_xGa_{1-x}As$ heterostructure, of standard Hall-bar geometry, when the applied magnetic flux density is of the order of 10 T and the device is cooled to a temperature of a few kelvins [27]. Under these conditions, the 2DEG is completely quantized and for a fixed current I through the device there are regions in the curve of Hall voltage vs. gate voltage, or of Hall voltage vs. magnetic flux density, where the Hall voltage U_H remains constant as the gate voltage or magnetic flux density is varied. These regions of constant Hall voltage are termed Hall plateaus.

In the limit of zero dissipation in the direction of current flow, the Hall resistance of the ith plateau $R_H(i)$, defined as the quotient of the Hall voltage of the ith plateau to the current I, is quantized:

$$R_H(i) = U_H(i)/I = R_K/i\,, \qquad (2)$$

where i is an integer and R_K is the von *Klitzing constant.*[3] (It follows from (2) that R_K is equal to the resistance of the $i=1$ plateau, $R_H(1)$. Since $R_H(i)$ is often referred to as the quantized Hall resistance independent of plateau number i, to avoid confusion it is proposed that R_K be used as the symbol for the Hall voltage-to-current quotient or resistance of the $i=1$ plateau, and to refer to it as the von Klitzing constant after the discoverer of the quantum Hall effect.)

A significant amount of experimental evidence supports the view that the von Klitzing constant R_K is a universal quantity, provided that the particular quantum Hall effect device used meets certain criteria. While the universality of R_K has not yet been demonstrated to a level of precision approaching that of the Josephson frequency-to-voltage quotient or Josephson constant K_J, studies of the influence of experimental variables such as current, temperature, device type, device material, and plateau number have shown that if certain precautions are taken and tests performed, then R_K may be reproduced with a relative precision approaching one part in 10^8 or possibly even several parts in 10^9 [28–34]. Carrying out quantized Hall resistance measurements according to the QHE Working Group's *Technical Guidelines for Reliable Measurements of the Quantized Hall Resistance* should allow this level of precision to be reached. Throughout the remainder of this paper it is assumed that these guidelines are implemented.

The current theory of the quantum Hall effect predicts, and the experimentally observed universality of the fundamental quantized Hall resistance relation [Eq. (2)] is consistent with the prediction, that R_K is equal to the invariant quotient of fundamental constants h/e^2 [27, 35–37]. Although the accuracy of this equality and (2) are still under active theoretical and experimental investigation, it is assumed for the purpose of including data from measurements of fundamental constants in the derivation of the recommended value of R_K, that $h/e^2 = R_K$. (The same assumption was made by the CODATA Task Group on Fundamental Constants in obtaining their 1986 recommended values of the constants [26].)

In particular, the fine-structure constant $\alpha \simeq 1/137$ and h/e^2 are related by defined quantities: $h/e^2 = \mu_0 c/2\alpha$, where $\mu_0 = 4\pi \times 10^{-7}$ N/A^2 exactly is the permeability of vacuum and $c = 299\,792\,458$ m/s exactly is the speed of light in vacuum. As a consequence of the above assumption, a measurement of α having a particular relative uncertainty will yield a value of R_K in ohms with the same relative uncertainty.

3. Derivation of Recommended Values of the Josephson and von Klitzing Constants, K_J and R_K

3.1. Approach

Because the recommended values of K_J and R_K are for use in realizing practical representations of the volt

[3] This discussion is restricted to the integral quantum Hall effect for which i is an integer. The fractional quantum Hall effect, for which i is the ratio of two integers, has not yet been studied sufficiently to warrant its use as a basis for a representation of the ohm

Table 1. Summary of values of the Josephson constant K_J.[a] For ease of comparison, the values are given in two forms: in GHz/V (column 2); and in parts in 10^6 relative to the value for the Josephson constant stated by the CCE in 1972, namely, 483 594 GHz/V (column 3)

Item No.	K_J (GHz/V)	$[(K_J/483\ 594\ \text{GHz/V})-1]\times 10^6$	Remarks and references[b]
1.	483 597.91 ±0.13	8.09 ±0.27	CSIRO/NML Hg electrometer [40, 41]
2.	483 598.77 ±0.17	9.86 ±0.35	U. Zagreb capacitor volt balance, realization of farad via calculable capacitor and voltage calibrations in terms of K_J from other laboratories [42]
3.	483 597.903 ±0.037	8.070 ±0.077	NPL realization of watt via moving-coil balance, realization of ohm via calculable capacitor [43]
4.	483 597.84 ±0.32	7.94 ±0.67	NIST realization of watt via moving-coil balance, realization of ohm via calculable capacitor [44–46]
5.	483 597.94 ±0.33	8.15 ±0.67	NIST γ'_p (high) from F, γ'_p (low), realization of ohm via calculable capacitor [47]
6.	483 597.88 ±0.48	8.02 ±0.99	NIM γ'_p (high), γ'_p (low), realization of ohm via calculable capacitor [48–51]
7.	483 597.54 ±0.25	7.33 ±0.52	NPL γ'_p (high), NIST γ'_p (low), NBS and CSIRO/NML realizations of ohm via calculable capacitor [52–55]
8.	483 597.40 ±0.29	7.03 ±0.60	2 e/h from NIST N_A, $\alpha(a_e)$ [56, 57]
9.	483 597.70 ±0.32	7.65 ±0.66	2 e/h from PTB N_A using NIST silicon reference sample of known molar mass, $\alpha(a_e)$ [58, 57]
10.	483 595.90 ±0.39	3.92 ±0.80	2 e/h from ASMW γ'_p via γ'_p (low) and γ'_p (high), $\alpha(a_e)$ [59, 57]

[a] To minimize rounding errors, calculations were carried out with values generally having one or more digits in addition to those shown

[b] See Appendix for laboratory abbreviations

and ohm by means of the Josephson and quantum Hall effects, the following guiding principle is adopted for their derivation: The values should be so chosen that they are unlikely to require significant change in the foreseeable future. This means that the number of digits given for the recommended values should be the minimum possible and that the uncertainties should be conservatively assigned. This principle also implies that it is unnecessary to carry out a complete least-squares adjustment of the fundamental physical constants to derive the recommended values; a straightforward treatment of the individual measurements of K_J and R_K currently available should suffice.

3.2. *Summary of Values of K_J*

Table 1 summarizes the measurements of K_J to be considered while Fig. 1 compares them graphically. (To aid in the comparison, the most precise value and its uncertainty are indicated by dashed and full lines, respectively, as well as by the usual point and error bars.) Values are included only if they were available by 15 June 1988 as stated by the CCE in its Declaration E1 (1986) and for which some form of documentation was available. Although it will be assumed that $K_J = 2\ e/h$ as discussed in Sect. 2.1, only the last three entries of Table 1 (Items 8, 9, 10) require this assumption. Such values are usually termed *indirect*, while those which do not require this assumption (items 1 through 7) are termed *direct*.

In general, an earlier result from a particular experiment has been excluded when it has been replaced by a more recent and presumably more reliable result from the same or a closely related experiment. Also excluded are measurements having a relative uncertainty[4] larger than about 1×10^{-6} because they cannot contribute in a significant way to the derivation of the recommended value.

The values given in Table 1 require further explanation.

Item 1. The relative uncertainty of the CSIRO/NML result, obtained using an elevated mercury column or so-called mercury electrometer, has been reduced from 0.31×10^{-6} to 0.27×10^{-6} based on further measurements relating to the density of the reference mercury

[4] Throughout this paper, uncertainties are treated following the suggestions of the BIPM Working Group on the Statement of Uncertainties as embodied in Recommendation INC-1 (1980) and affirmed by the CIPM in Recommendation 1 (CI-1981) [38]. In particular, all uncertainties are one-standard-deviation estimates in keeping with CIPM Recommendation 1 (CI-1986) [39]

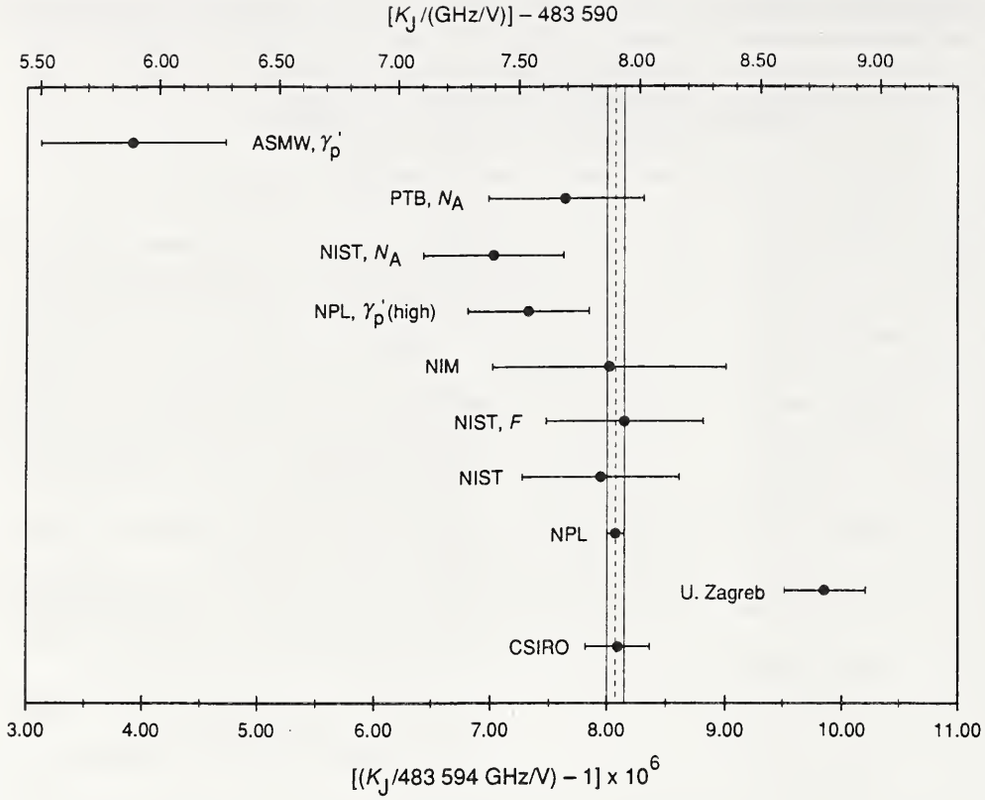


Fig. 1. Comparison of the values of K_J and their standard deviation uncertainties as given in Table 1. The vertical dashed and solid lines indicate the value and standard deviation uncertainty of the most precise result

used in the NML experiment and to the stability of the density of mercury during long-term storage [40, 41].

Item 2. The result from the U. Zagreb given in the table is from their latest and most precise measurements [42]. It was obtained using volt balance ETF-84 during late 1987 and the first half of 1988. However, it differs significantly from the results obtained from 1981 to 1985 using volt balances ETF-80 and ETF-82. Possible sources of systematic error in the present balance and associated equipment are being vigorously investigated. This experiment requires knowledge of the value of a reference capacitor in farads, but because it enters into the calculation of K_J to the one-half power, its contribution to the uncertainty is reduced by a factor of two.

Item 3. To obtain a value of K_J from a watt-realization experiment of the moving-coil type developed at the NPL requires knowledge of a reference resistance in ohms. This resistance can either be an artifact-based resistance standard or a quantized Hall resistance. (The contribution of the uncertainty of this reference resistance to K_J is reduced by a factor of two since it enters to the one-half power.) The NPL value of K_J [43] given in the table is based on an NPL realization of the ohm using a calculable capacitor. If it were based on the recommended value of the von Klitzing constant (see Sect. 3.7), it would be 4.2 parts in 10^8 larger, somewhat over one half the standard deviation of the NPL value. Because this is a comparatively small shift and the relative uncertainty assigned to the recommended value of the von Klitzing constant is 2×10^{-7}, the NPL result is taken as given.

Item 4. The experiment to realize the watt by the moving-coil method at the NIST is similar to that at the NPL but it has not yet reached the same level of precision because a much weaker magnetic field is currently being used. The NIST result [44–46] is based on a realization of the ohm at the NIST via a calculable capacitor. Using instead the recommended value of the von Klitzing constant would increase the NIST result by less than one part in 10^8.

Item 5. A value of K_J can be obtained from so-called low and high field measurements of the gyromagnetic ratio of the proton, γ_p' (low) and γ_p' (high), and a realization of the ohm [26] (here and throughout this paper the prime indicates a spherical, pure H_2O nuclear-magnetic-resonance or NMR sample at 25 °C). For this result, γ_p' (high) was derived from a NIST measurement of the faraday constant F and the accepted values of well-known constants [47]. The experiments to realize the ohm via the NIST calculable capacitor and to measure F and γ_p' (low) were carried out at the NIST during the period 1973 to 1978.

Item 6. This result from the NIM was obtained in the same way as Item 5 except γ_p' (high) was measured directly using NMR and a force balance [48–51]. The experiments to realize the ohm and to measure γ_p' (low)

and γ'_p (high) upon which it is based were carried out from October 1987 to May 1988 and supersede those of the 1970s.

Item 7. Like data Items 5 and 6, data Item 7 is based on a realization of the ohm and measurements of γ'_p (low) and γ'_p (high). It was derived from the 1974 NPL measurement of γ'_p (high) [52], the 1978 NIST measurement of γ'_p (low) [53], the 1973 NIST calculable capacitor realization of the ohm [54], NML calculable capacitor realizations of the ohm carried out over the period 1964 to 1987 [6, 55], and the results of the international comparisons of national representations of the ohm organized by the BIPM over the same period. Because the same value of γ'_p (low) was used to obtain Item 5, Items 5 and 7 are not completely independent; their correlation coefficient is 0.03. This correlation is taken into account in the calculations carried out in the next section as appropriate. It is relatively small because the uncertainties of the two values of γ'_p (high) upon which Items 5 and 7 are based are about six and five times larger, respectively, than the uncertainty of the NIST 1978 γ'_p (low) value.

Item 8. A value of $K_J = 2\,e/h$ can be obtained from a measurement of the Avogadro constant N_A via the relation

$$2\,e/h = [16\,R_\infty\,(m_p/m_e)\,N_A/(\mu_0\,c^2\,M_p\,\alpha)]^{1/2}\,, \qquad (3)$$

where R_∞ is the Rydberg constant for infinite mass, m_p/m_e is the proton-to-electron mass ratio, and M_p is the molar mass of the proton. Item 8 was derived from this equation using (i) $N_A = 6.022\,129\,7(72) \times 10^{23}\ \mathrm{mol}^{-1}$ based on the most recent NIST silicon-lattice-spacing measurements [56] and the NIST value for the molar volume of silicon used in the 1986 CODATA adjustment of the fundamental constants [26] but updated to account for a new mass adjustment [60] (the change is inconsequential); (ii) the CODATA value for m_p/m_e [26]; (iii) $R_\infty = 10\,973\,731.573(4)\ \mathrm{m}^{-1}$ [61], a more recent and accurate value than that of CODATA; (iv) $M_p = 1.007\,276\,468(7) \times 10^{-3}$ kg/mol based on the new value for the nuclidic mass of hydrogen from the new mass adjustment [60]; and (v), the most recent value of the fine-structure constant from the electron-magnetic-moment anomaly a_e [57], $1/\alpha\,(a_e) = 137.035\,991\,4(11)$. While the NIST silicon lattice spacing result is not final, the value is unlikely to change by an amount of any significane in comparison with the 1.15×10^{-6} relative uncertainty of the silicon molar volume.

Item 9. This result for $K_J = 2\,e/h$ was derived from (3) using the value $N_A = 6.022\,137\,3(79) \times 10^{23}\ \mathrm{mol}^{-1}$ as obtained from the PTB measurements of the lattice spacing and molar volume of silicon [58]. Data Items 8 and 9 are not entirely independent because they are based on the molar mass of the same silicon reference material. The 0.42×10^{-6} relative uncertainty of the molar mass of this material leads to a correlation coefficient between the two values of 0.11. This correlation is taken into account as appropriate in the calculations carried out in the next section.

Item 10. This result for $K_J = 2\,e/h$ was derived from the ASMW low- and high-field measurements of the proton gyromagnetic ratio completed in 1985 [59]. It is based on the relations

$$\gamma'_p = [\{\gamma'_p\,(\mathrm{low})\}\,\{\gamma'_p\,(\mathrm{high})\}]^{1/2}\,\mathrm{s}^{-1}\,\mathrm{T}^{-1} \qquad (4)$$

$$2\,e/h = 4\,R_\infty\,\gamma'_p\,[c\,\alpha^2\,(\mu'_p/\mu_B)]^{-1}\,, \qquad (5)$$

where { } indicates numerical value only and it is assumed that γ'_p (low) and γ'_p (high) are measured in terms of the same laboratory representations of the volt and ohm; μ'_p/μ_B is the magnetic moment of the proton in units of the Bohr magneton. Using the CODATA value of μ'_p/μ_B [26], $\gamma'_p = 2.675\,142\,7(21) \times 10^8\ \mathrm{s}^{-1}\,\mathrm{T}^{-1}$ from the ASMW measurements, and the values for the other constants indicated previously, yields the result in the table.

An alternative approach would have been to re-express the two ASMW measurements in terms of the BIPM representations of the volt and ohm and to use the NML calculable capacitor realization of the ohm to obtain a value of K_J rather than of $2\,e/h$ in a manner similar to that used to obtain Items 5, 6, and 7. However, the use of (4) and (5) minimizes the considerable problems associated with the transfer to the BIPM representations without introducing any significant additional uncertainty since the constants entering (5) are well known in comparison with the 0.80×10^{-6} relative uncertainty of the ASMW value of γ'_p.

3.3. *Analysis of Values of* K_J

The simple mean and standard deviation of the mean of the ten measurements given in Table 1 are

$$K_J = (483\,597.68 \pm 0.23)\ \mathrm{GHz/V} \qquad (6a)$$

$$= K_{J-72}\,[1 + (7.61 \pm 0.47) \times 10^{-6}]\,, \qquad (6b)$$

where for convenience the value of the Josephson constant stated by the CCE in 1972 is denoted by the symbol K_{J-72}; that is, $K_{J-72} = 483\,594$ GHz/V exactly.

The simple mean and its standard deviation have little significance in the present case because of the large differences in precision among the measurements. The more appropriate weighted mean, taking as the weight of each measurement the reciprocal of the square of its assigned one-standard-deviation un-

certainty, $w_i = 1/s_i^2$, yields [5]

$$K_J = (483\,597.907 \pm 0.086)\ \text{GHz/V} \quad (7a)$$

$$= K_{J-72}\,[1 + (8.08 \pm 0.18) \times 10^{-6}], \quad (7b)$$

where the uncertainty has been calculated on the basis of external consistency. [6] That is, the usual standard deviation of the weighted mean calculated on the basis of internal consistency, $s_I = \left[\sum_{i=1}^{N} w_i\right]^{-1/2}$, has been multiplied by the scale factor or Birge ratio $R_B = [\chi^2/\nu]^{1/2}$, where χ^2 is the statistic "chi square" and ν is the number of degrees of freedom ($\nu = N - 1 = 9$ in the present case). This was done because the data are in disagreement; $R_B = 2.55$ and $\chi^2 = 58.7$ compared with its expected value of $9 = \nu$. The probability that such a large value of χ^2 has occurred by chance is essentially zero [i.e., $P(58.7|9) \simeq 0$].

The problem, of course, is that the U. Zagreb result, Item 2, and the ASMW result, Item 10, strongly disagree with each other as well as with most of the remaining data. This is readily apparent from an examination of Table 1 and Fig. 1. Indeed, Item 2 accounts for about 44% and Item 10 for about 46% of the above value of χ^2, respectively. If these two clearly discrepant items are deleted, one finds for the weighted mean

$$K_J = (483\,597.887 \pm 0.035)\ \text{GHz/V} \quad (8a)$$

$$= K_{J-72}\,[1 + (8.039 \pm 0.071) \times 10^{-6}], \quad (8b)$$

where the uncertainty is now calculated on the basis of internal consistency (as will be the case for the remainder of this section). The eight values are in excellent agreement: $\chi^2 = 5.22$ for $\nu = 7$, $R_B = 0.86$, and $P(5.22|7) \simeq 0.63$. (We assume as usual that $P > 0.05$ indicates an acceptable level of agreement.)

It is clear that the NPL result, Item 3, will dominate any weighted mean in which it is included because its assigned uncertainty is significantly smaller than that of any other value. If it is deleted along with the discrepant Items 2 and 10, the weighted mean of the remaining seven items is

$$K_J = (483\,597.794 \pm 0.092)\ \text{GHz/V} \quad (9a)$$

$$= K_{J-72}\,1 + [(7.84 \pm 0.19) \times 10^{-6}], \quad (9b)$$

with $\chi^2 = 4.03$ for $\nu = 6$, $R_B = 0.82$, and $P(4.03|6) = 0.67$. Again, these values are in excellent agreement among themselves. Moreover, their weighted mean is consistent with the highly precise NPL result, Item 3. The relative difference is $(0.23 \pm 0.21) \times 10^{-6}$.

If the next most precise value, the CSIRO/NML result (Item 1), is deleted along with the two discrepant Items 2 and 10 and the NPL result (Item 3), one finds

$$K_J = (483\,597.67 \pm 0.13)\ \text{GHz/V} \quad (10a)$$

$$= K_{J-72}\,[1 + (7.60 \pm 0.27) \times 10^{-6}], \quad (10b)$$

with $\chi^2 = 2.38$ for $\nu = 5$, $R_B = 0.69$, and $P(2.38|5) = 0.79$. The relative difference between this value and the NPL value is $(0.47 \pm 0.28) \times 10^{-6}$, which is acceptable agreement.

Because Item 10 is discrepant and the two remaining indirect values, Items 8 and 9, are of low precision relative to the two most precise direct values, Items 1 and 3, little can be learned from a detailed comparison of the means of the direct and indirect values. However, one notes that the relative difference between the weighted mean of the six consistent direct measurements, Items 1, 3, 4, 5, 6, and 7, and that of Items 8 and 9 is $(0.75 \pm 0.47) \times 10^{-6}$. The agreement is acceptable.

3.4. *Selection of Recommended Value of K_J*

It is clear from the above analysis that taking as the recommended value $K_J = 483\,597.9$ GHz/V is highly consistent with any reasonable treatment of the data [e.g. Eq. (8)] and the adopted guiding principle discussed in Sect. 3.1. The question remains as to the uncertainty to be assigned to this value which will also be consistent with the principle.

Considering (i) that the total spread of the individual measurements upon which the highly precise NPL result is based is about 0.35 GHz/V, or 0.73 parts in 10^6; (ii) that the difference between (8) and (10) is 0.21 GHz/V, or 0.44 parts in 10^6; and (iii) the existence of the discrepant Items 2 and 10, it is believed that adopting 0.2 GHz/V as the uncertainty, which corresponds to a relative uncertainty of 4×10^{-7}, is consistent with both the guiding principle and the data. In conformity with footnote 4, this uncertainty is taken as one standard deviation. Thus the recommended value and assigned uncertainty are

$$K_J = 483\,597.9\ \text{GHz/V} \quad (11a)$$

$$\text{Assigned standard deviation: } 0.2\ \text{GHz/V} \quad (11b)$$

$$\text{Corresponding relative standard deviation: } 4 \times 10^{-7}. \quad (11c)$$

Figure 2 graphically compares this value with the data of Table 1. Equation (11) is consistent with the 1986 CODATA value $K_J = (483\,597.67 \pm 0.14)$ GHz/V [26].

---

[5] This is the appropriate equation if all of the covariances or correlation coefficients between the measurements are zero. In fact, a generalized variance matrix was used because of the correlations between some of the data that were indicated in Sect. 3.2

[6] Throughout, the procedures and terminology employed in least-squares adjustments of the fundamental physical constants are used; see Ref. [26] and the references cited therein for details

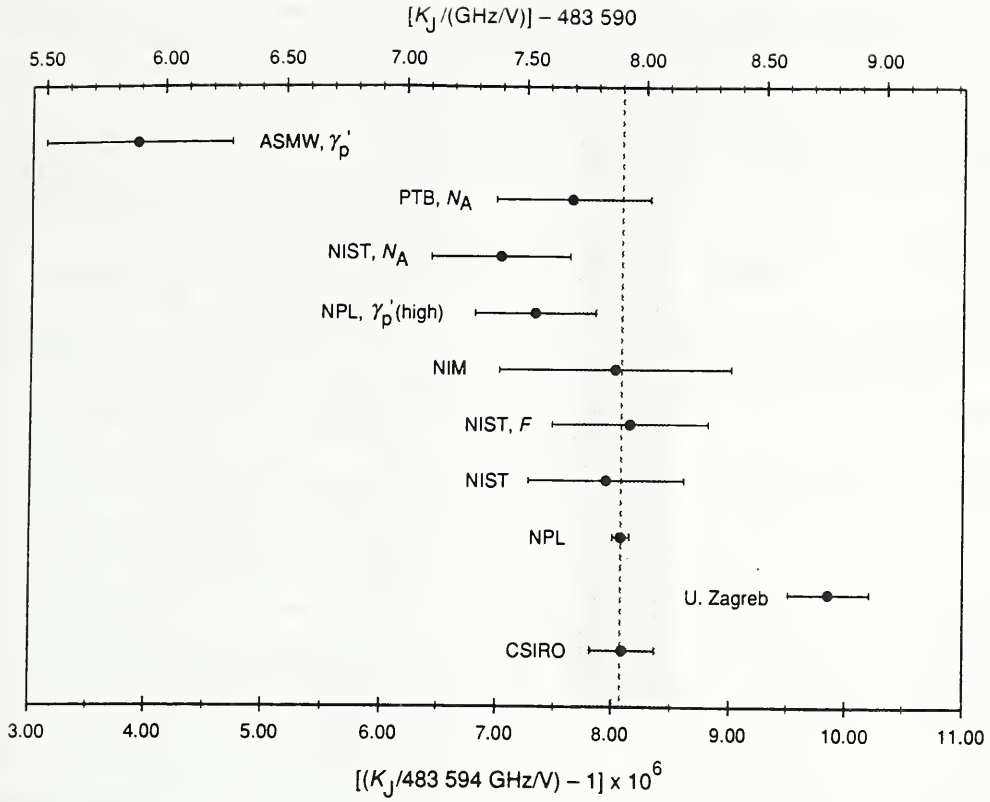


Fig. 2. Comparison of the recommended value of K_J (vertical dashed line) and its assigned standard deviation uncertainty (delimited by the shading) with the values of K_J and their standard deviation uncertainties as given in Table 1

3.5. Summary of Values of R_K

Table 2 summarizes the measurements of R_K to be considered while Fig. 3 compares them graphically. (To aid in the comparison, the most precise value and its uncertainty are again indicated by dashed and full lines, respectively, as well as by the usual point and error bars.) Values are included only if they were available by 15 June 1988 as stated by the CCE in its Declaration E2 (1986) and for which some form of documentation was available. Although it will be assumed that $R_K = h/e^2 = \mu_0 c \alpha^{-1}/2$ as discussed in Sect. 2.2, only the last four entries of Table 2 (Items 8 through 11) require this assumption. These values are termed *indirect*, while those which do not require this assumption (Items 1 through 7) are termed *direct*. Again, an earlier result from a particular experiment has been excluded when it has been replaced by a more recent and presumably more reliable result from the same or a closely related experiment.

The values given in Table 2 require further explanation.

Item 1. This result was obtained by the CSIRO/NML from its own quantized Hall resistance (QHR) measurements (i.e., determinations of $R_H(i)$ in terms of laboratory reference resistors) and realizations of the ohm via the CSIRO/NML calculable capacitor [62, 63]. The values labeled (a) and (b) were obtained from QHR measurements carried out at the BIPM [55, 64] and the G. U. [65], respectively, and CSIRO/NML ohm realizations transferred directly to these laboratories by means of one ohm artifact resistance standards. Clearly, Item 1 agrees well with 1a but not with 1 b.[7] The G. U. result is based on measurements using a silicon MOSFET sample. A more recent but still preliminary G. U. result using a GaAs heterostructure is about 9 parts in 10^8 smaller [65], thereby bringing into question Item 1b. (Item 6 was also obtained from silicon MOSFET measurements. All other values were obtained using heterostructures.) Further, Item 1 is based on QHR and ohm-realization measurements, both of which were carried out in the CSIRO/NML. In the opinion of some members of the QHE Working Group, it is preferable to use values of R_K based solely on data obtained in the same laboratory because of the difficulties associated with accurately transferring representations of the ohm between laboratories. For these reasons, only Item 1 is included in the calculations.

Items 2 to 7. The values of R_K from the NPL (Item 2) [66], the LCIE (Item 3) [67], the ETL (Item 4) (68], the NIST (Item 5) [69, 45, 46], and the NIM (Item 7) [72, 73], are all based on QHR and ohm-realization

[7] The agreement of Items 1 and 1a implies agreement of the CSIRO/NML QHR measurements with those of BIPM. The QHR measurements of a number of other laboratories also agree well with those of BIPM and hence with those of CSIRO/NML; see Ref. [55], and also Ref. [77]

Table 2. Summary of values of the von Klitzing constant R_K.[a] For ease of comparison, the values are given in two forms: in Ω (column 2); and in parts in 10^6 relative to the convenient reference resistance 25 812.8 Ω (column 3)

Item No.	R_K (Ω)	$[(R_K/25\,812.8\,\Omega)-1]\times 10^6$	Remarks and references[b]
1.	25 812.809 4 ±0.001 7	0.363 ±0.066	CSIRO/NML quantized Hall resistance (QHR) measurements and realization of ohm via calculable capacitor [62, 63]
(a)	25 812.808 6 ±0.001 7	0.333 ±0.065	BIPM QHR, CSIRO/NML realization of ohm via calculable capacitor [55, 64]
(b)	25 812.813 4 ±0.002 1	0.520 ±0.080	G.U. QHR, CSIRO/NML realization of ohm via calculable capacitor [65]
2.	25 812.809 2 ±0.001 4	0.356 ±0.054	NPL QHR and realization of ohm via calculable capacitor [66]
3.	25 812.801 8 ±0.005 7	0.070 ±0.220	LCIE QHR and realization of ohm via calculable capacitor [67]
4.	25 812.806 4 ±0.006 7	0.247 ±0.260	ETL QHR and realization of ohm via calculable capacitor [68]
5.	25 812.807 23 ±0.000 61	0.280 ±0.024	NIST QHR and realization of ohm via calculable capacitor [69, 45, 46]
6.	25 812.806 5 ±0.008 3	0.250 ±0.320	IMS QHR, IMM realization of ohm via calculable capacitor [70, 71]
7.	25 812.805 5 ±0.015 6	0.214 ±0.606	NIM QHR and realization of ohm via calculable capacitor [72, 73]
8.	25 812.805 99 ±0.000 21	0.232 1 ±0.008 0	α^{-1} from electron magnetic moment anomaly a_e [74, 57]
9.	25 812.806 2 ±0.004 2	0.241 ±0.163	α^{-1} from muonium groundstate hyperfine splitting ν (Muhfs) [26]
10.	25 812.804 60 ±0.000 95	0.178 ±0.037	α^{-1} from NIST γ'_p (low), QHR, and Josephson $2e/h$ [75, 69, 76, 46]
11.	25 812.803 3 ±0.001 5	0.127 ±0.056	α^{-1} from NIST γ'_p (low), realization of ohm via calculable capacitor, and Josephson $2e/h$ [75, 45, 76, 46]

[a] To minimize rounding errors, calculations were carried out with values generally having one or more digits in addition to those shown

[b] See Appendix for laboratory abbreviations

measurements carried out in the same laboratory. For Item 6, the QHR measurements were carried out at the IMS in Moscow, and the ohm-realization measurements at the IMM in Leningrad [70, 71]; an artifact resistance standard was used to transfer the measurements between the two laboratories. The variation in the uncertainty assigned to the seven direct values, Items 1 through 7, is mainly due to the design and construction details of the calculable capacitor and associated impedance bridges used in the ohm-realization experiments.

Item 8. This indirect result is based on the value of the inverse fine-structure constant α^{-1} obtained from the experimental measurement of the electron-magnetic-moment anomaly a_e at the University of Washington (relative uncertainty of 4×10^{-9}) [74]; and the theoretical expression for a_e given by T. Kinoshita, Cornell University (relative uncertainty of 7×10^{-9} arising from numerical integrations) [57]. Although Kinoshita's calculations are not final, he has assigned the uncertainty conservatively and the value of R_K is not expected to change significantly. This is the most precise result currently available, assuming the correctness of the relation $R_K = \mu_0 c \alpha^{-1}/2$.

Item 9. This result is based on the value of α^{-1} obtained from the ground-state hyperfine splitting of muonium ($\mu^+ e^-$ atom) following the CODATA 1986 least-squares adjustment of the fundamental constants [26]. However, the more recent and accurate value of R_∞ given previously has been used [61] to evaluate the theoretical expression for the splitting as taken by CODATA from the work of Sapirstein et al. [78]. Further, this expression has been updated to include the additional terms calculated by Eides et al. [79] and

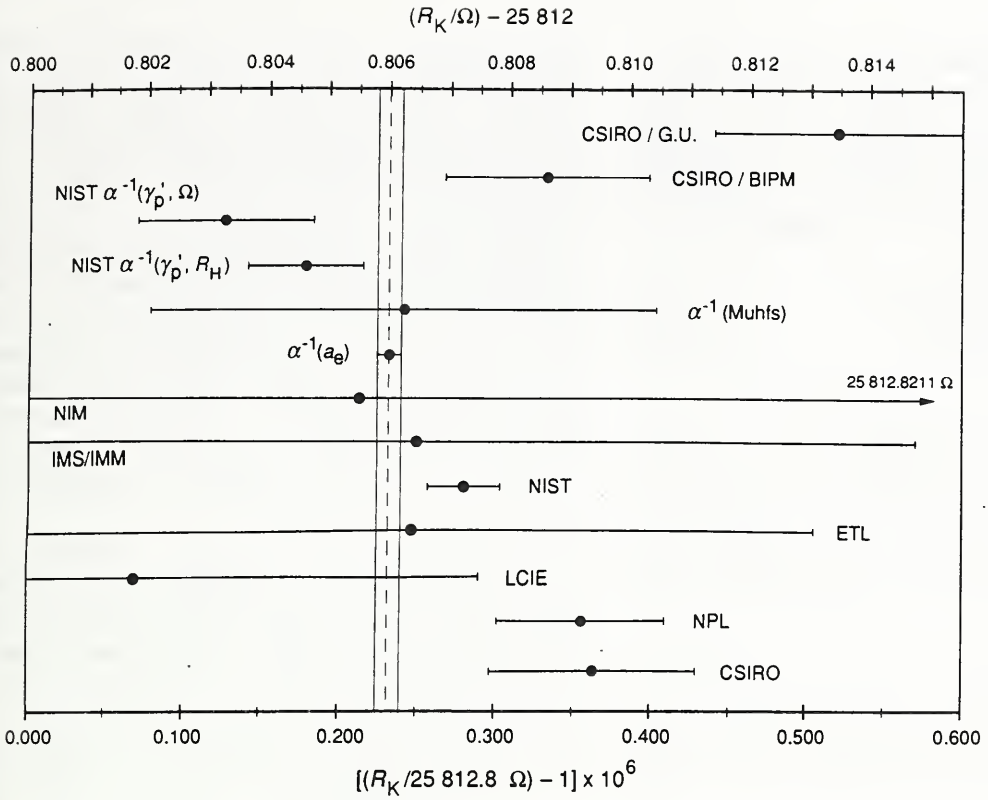


Fig. 3. Comparison of the values of R_K and their standard deviation uncertainties as given in Table 2. The vertical dashed and solid lines indicate the value and standard deviation uncertainty of the most precise result

Starshenka and Faustov [80], and the exact analytic expressions obtained by Karshenboim et al. [81] and by Eides et al. [82] for the corresponding numerically evaluated terms given by Sapirstein et al.

Item 10. The relationship [26]

$$\alpha^{-1} = [(\mu_p'/\mu_B)(2e/h)\ R_K/(2\mu_0 R_\infty \gamma_p')]^{1/3}, \qquad (12)$$

where as before μ_p'/μ_B is the magnetic moment of the proton in units of the Bohr magneton and γ_p' is the gyromagnetic ratio of the proton, has the unique property that it remains valid if $2e/h$ is measured by the Josephson effect and is expressed in terms of V_{LAB}; R_K is expressed in terms of Ω_{LAB}; and γ_p' is measured by the low-field method and is expressed in terms of $T_{LAB} \propto A_{LAB} = V_{LAB}/\Omega_{LAB}$, where the quantities A_{LAB}, T_{LAB}, V_{LAB}, and Ω_{LAB} are the laboratory representations of the tesla, ampere, volt, and ohm, respectively. Item 10 was obtained by the NIST from this equation and a new NIST determination of γ_p' (low) [75], maintenance of V_{NIST} using Josephson arrays [76], measurements of R_K in terms of Ω_{NIST} [69], the 1986 CODATA value of μ_p'/μ_B [26], and the value of R_∞ given previously. Because Items 5 and 10 are based on the same QHR measurements, they are not totally independent; their correlation coefficient is 0.04.

Item 11. If the relation $R_K = \mu_0 c \alpha^{-1}/2$ is used to eliminate R_K from (12), the resulting expression is

$$\alpha^{-1} = [c(\mu_p'/\mu_B)(2e/h)/(4R_\infty \gamma_p')]^{1/2}. \qquad (13)$$

If as before $2e/h$ is measured by the Josephson effect and is expressed in V_{LAB}, and γ_p' is measured by the low-field method and is expressed in $T_{LAB} \propto A_{LAB} = V_{LAB}/\Omega_{LAB}$, the quantity Ω_{LAB} is introduced in the denominator. That is, γ_p' is replaced by $\Omega_{LAB}\gamma_p'$ (low), where Ω_{LAB} is to be expressed in ohms. Item 11 was obtained by the NIST from this equation using the result of its recent experiment to realize the ohm via the NIST calculable capacitor [45]. Because the same ohm realization result was used by the NIST to obtain Item 5, the two are not independent; their correlation coefficient is -0.17. Similarly, the correlation coefficient of Items 10 and 11 is 0.98, mainly because both are based on the same value of γ_p' (low). The NIST Items 5, 10, and 11 are related in such a way that, assuming their correlations are properly considered, the weighted mean of any two of them (and all three of them) gives the same result. This is taken into account as appropriate in the calculations carried out in the following section.

3.6. *Analysis of Values of* R_K

The simple mean and standard deviation of the mean of the ten measurements 1 through 9 plus 11 are

$$R_K = (25\,812.806\,26 \pm 0.000\,80)\,\Omega \qquad (14a)$$

$$= R_0[1 + (0.242 \pm 0.031) \times 10^{-6}], \qquad (14b)$$

where the convenient reference resistance 25 812.8 Ω is denoted by the symbol R_0. (Including Item 10 instead of Item 11 yields a similar result.)

However, the simple mean and its standard deviation have little significance in the present case because of the large differences in precision of the measurements. The more appropriate weighted mean, taking as the weight of each measurement the reciprocal of the square of its assigned one standard deviation uncertainty, $w_i = 1/s_i^2$, yields[5]

$$R_K = (25\,812.806\,15 \pm 0.000\,25)\,\Omega \quad (15a)$$
$$= R_0\,[1 + (0.238\,4 \pm 0.009\,6)], \quad (15b)$$

where the uncertainty has been calculated on the basis of external consistency (see Sect. 3.3). The reason is that the data are only marginally in agreement; $R_B = 1.31$ and $\chi^2 = 15.6$ compared with its expected value of $9 = \nu$. The probability that this value of χ^2 has occurred by chance is about 8%, i.e., $P(15.6|9) \simeq 0.08$. (We assume as usual that $P > 0.05$ indicates an acceptable level of agreement.)

It is clear that the value of R_K from $\alpha^{-1}(a_e)$, Item 8, will dominate any weighted mean in which it is included because its assigned uncertainty is significantly smaller than that of any other value. If it is deleted, one obtains

$$R_K = (25\,812.806\,97 \pm 0.000\,56)\,\Omega \quad (16a)$$
$$= R_0\,[1 + (0.270 \pm 0.022) \times 10^{-6}], \quad (16b)$$

where the uncertainty is calculated on the basis of external consistency; $\chi^2 = 11.8$ for $\nu = 8$, $R_B = 1.22$, and $P(11.8|8) \simeq 0.16$. The agreement is reasonable.

It is of interest to calculate a value of R_K based solely on the seven direct measurements, Items 1 through 7. The result is

$$R_K = (25\,812.807\,65 \pm 0.000\,52)\,\Omega \quad (17a)$$
$$= R_0\,[1 + (0.296 \pm 0.020) \times 10^{-6}], \quad (17b)$$

where the uncertainty has been calculated on the basis of internal consistency; $\chi^2 = 3.85$ for $\nu = 6$, $R_B = 0.80$, and $P(3.85|6) \simeq 0.70$. The agreement is excellent.

Because Items 1, 2, and 5 are significantly more precise than Items 3, 4, 6, and 7, they essentially determine the weighted mean of the seven direct measurements. Indeed, the weighted mean of just these three more precise values is

$$R_K = (25\,812.807\,72 \pm 0.000\,61)\,\Omega \quad (18a)$$
$$= R_0\,[1 + 0.299 \pm 0.024) \times 10^{-6}], \quad (18b)$$

where the uncertainty has been calculated on the basis of external consistency; $\chi^2 = 2.70$ for $\nu = 2$, $R_B = 1.16$, and $P(2.70|2) \simeq 0.26$. The agreement is quite reasonable.

An indirect value based on the weighted mean of Items 8, 9, and 11 may be obtained for comparison:

$$R_K = (25\,812.805\,94 \pm 0.000\,27)\,\Omega \quad (19a)$$
$$= R_0\,[1 + (0.230 \pm 0.010) \times 10^{-6}], \quad (19b)$$

where the uncertainty has been calculated on the basis of external consistency; $\chi^2 = 3.40$ for $\nu = 2$, $R_B = 1.30$, and $P(3.40|2) \simeq 0.18$. The agreement is reasonable. The difference between Equations. (18) and (19) is $(0.001\,78 \pm 0.000\,67)\,\Omega$, which corresponds to a relative difference of $(0.069 \pm 0.026) \times 10^{-6}$. Compared in this way, the direct and indirect values are not in particularly good agreement. (Using Item 10 in place of Item 11 yields a similar result.) Indeed, the uncertainty of Item 8 is sufficiently small compared with the uncertainties of the other indirect values that it essentially determines the indirect value. The difference between (18) and Item 8 is $(0.001\,73 \pm 0.000\,65)\,\Omega$, which corresponds to a fractional difference of $(0.067 \pm 0.025) \times 10^{-6}$.

Finally, the result of the above comparison of (18) and (19) leads one to calculate the weighted mean of just items 1, 2, 5, 8, and 10. These five more precise values of R_K have relative uncertainties of less than 7×10^{-8}, which is less than one half that of the next most precise value. They yield

$$R_K = (25\,812.806\,16 \pm 0.000\,37)\,\Omega \quad (20a)$$
$$= R_0\,[1 + (0.239 \pm 0.014) \times 10^{-6}], \quad (20b)$$

where the uncertainty has been calculated on the basis of external consistency; $\chi^2 = 15.0$ for $\nu = 4$, $R_B = 1.93$, and $P(15.0|4) \simeq 0.005$. As could be anticipated from the above comparison, the agreement is poor. (Including Item 11 in place of Item 10 yields the identical result because of the relationship between Items 5, 10, and 11 discussed previously.)

3.7. *Selection of Recommended Value of R_K*

Based on the weighted mean of all the data as given in (15), and the guiding principle discussed in Sect. 3.1, an obvious choice for the recommended value is 25 812.806 Ω. That this is identical to the value of R_K obtained from $\alpha^{-1}(a_e)$, Item 8, is in large part due to the latter's small uncertainty in comparison with the uncertainties of the other values. On the other hand, based on the weighted mean of just the direct measurements rather than of all the measurements, an obvious choice is 25 812.808 Ω as given in either (17) or (18).

In the opinion of the QHE Working Group, the recommended value of R_K should reflect the results of the direct measurements and should not be dominated by a single indirect value which has not been verified by independent experiments and calculations. On this basis, the recommended value is taken as the simple mean of the above two values, namely, 25 812.807 Ω. It is noteworthy that this recommended value is only 3.9 parts in 10^8 larger than the value 25 812.806 Ω and only 7.7 parts in 10^8 smaller than the value 25 812.809 Ω. Based on the available data, the former

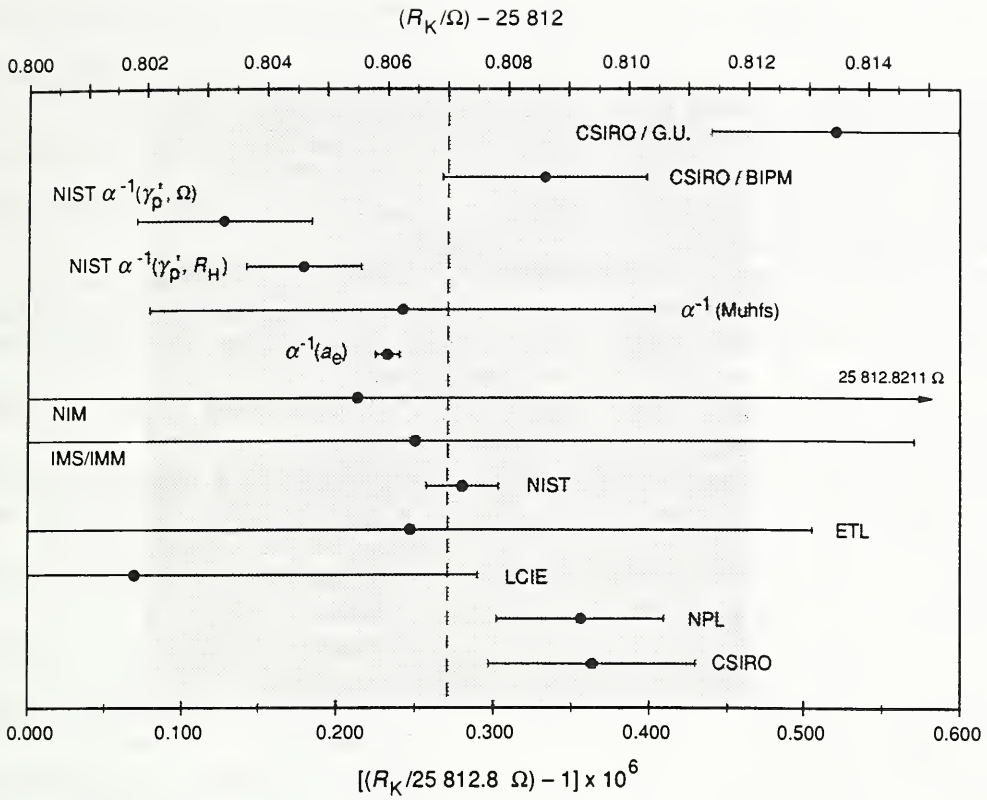


Fig. 4. Comparison of the recommended value of R_K (vertical dashed line) and its assigned standard deviation uncertainty (delimited by the shading) with the values of R_K and their standard deviation uncertainties as given in Table 2

is the smallest value for R_K one might reasonably choose while the latter, being the simple mean of the three most precise direct measurements (Items 1, 2, and 5) is the largest value one might conceivably choose.[8]

The question remains as to the uncertainty to be assigned to this recommended value which will also be consistent with the adopted guiding principle. Considering that the total spread of the values of R_K given in Table 2 (including item 1b) is about 0.01 Ω, and that the relation $R_K = \mu_0 c \alpha^{-1}/2$ as well as (2) are under active theoretical and experimental investigation, it is believed that adopting 0.005 Ω as the uncertainty, which corresponds to a relative uncertainty of 2×10^{-7}, is consistent with both the guiding principle and the current situation. Again in conformity with footnote 4, this is taken as one standard deviation. The recommended value and assigned uncertainty are thus

$$R_K = 25\,812.807\,\Omega \quad (21\,a)$$

Assigned standard deviation: 0.005 Ω (21 b)

Corresponding relative standard deviation:

$$2 \times 10^{-7}. \quad (21\,c)$$

[8] This value results if one takes the extreme position that the recommended value of R_K should be based solely on direct measurements and that the differences in the uncertainties assigned Items 1, 2, and 5 arise from the differences in the approaches used to estimate them rather than from any real differences in the accuracies of the experiments themselves

Figure 4 graphically compares this value with the data of Table 2. Equation (21) is consistent with the 1986 CODATA value $R_K = (25\,812.805\,6 \pm 0.001\,2)\,\Omega$ [26].

4. Conclusion

For the purpose of basing a representation of the volt on the Josephson effect, (11a) is used to define the following conventional value for the Josephson constant:

$$K_{J-90} \overset{\text{def}}{=} 483\,597.9\ \text{GHz/V} \quad (22)$$

exactly, where the subscript 90 derives from the fact that the new representation of the volt is to come into effect starting on 1st January 1990. It follows from (11c) that the assigned one-standard-deviation uncertainty of the new volt representation as based on K_{J-90} is 0.4 μV, corresponding to a relative uncertainty of 4×10^{-7}. Because K_{J-90} is approximately $(1 + 8.065 \times 10^{-6})$ times the value 483 594 GHz/V stated by the CCE in 1972, the new representation of the volt will exceed a representation of the volt based on the 1972 value by about 8.065 μV. It is expected that the new conventional value of the Josephson constant will not need to be altered in the foreseeable future.

For the purpose of basing a representation of the ohm on the quantum Hall effect, (21 a) is used to define the following conventional value for the von Klitzing

constant:

$$R_{K-90} \overset{\text{def}}{=} 25\,812.807\,\Omega \qquad (23)$$

exactly, where again the subscript 90 derives from the fact that the new representation of the ohm is to come into effect starting on 1st January 1990. It follows from (21 c) that the assigned one-standard deviation uncertainty of the new ohm representation as based on R_{K-90} is 0.2 μΩ, corresponding to a relative uncertainty of 2×10^{-7}. It is expected that this conventional value of the von Klitzing constant will not need to be altered in the foreseeable future.

Based on the apparatus in current use at the national standards laboratories, the experimental uncertainty associated with the maintenance of a laboratory representation of the volt by means of the Josephson effect and that with the measurement of quantized Hall resistances each lies in the range 0.1 to 2 parts in 10^7 [3, 5]. Consequently, with the world-wide adoption starting on 1st January 1990 of the conventional value of the Josephson constant K_{J-90}, and of the QHE as a reference standard of resistance using the conventional value of the von Klitzing constant R_{K-90}, all national representations of the volt and of the ohm should be equivalent to within a few parts in 10^7, as presaged in Sect. 1.4. Furthermore, these representations should be consistent with the volt and with the ohm, respectively, to better than 5 parts in 10^7. The resulting significant improvement in the international uniformity of electrical measurements and their consistency with the SI will no doubt be of major benefit to science, commerce, and industry throughout the world.

Note Added in Proof. The University of Zagreb group under Professor V. Bego has identified several unsuspected systematic errors in their volt balance and associated equipment (see *Item 2*, Sect. 3.2 and Table 1). When these are taken into account, the U. Zagreb value of K_J agrees with the recommended value. However, a final result from the experiment will not be available until additional data are obtained and the analysis of all known possible sources of error is completed (V. Bego, private communication, January 1989).

Appendix

The following are the laboratory abbreviations used throughout this paper.

ASMW	Amt für Standardisierung, Meßwesen und Warenprüfung, Berlin, GDR
BIPM	Bureau International des Poids et Mesures, Sèvres
CSIRO/ NML	Commonwealth Scientific and Industrial Research Organization, Division of Applied Physics, National Measurement Laboratory, Lindfield, Australia
ETL	Electrotechnical Laboratory, Tsukuba, Japan
G. U.	Department of Physics, Gakushuin University, Tokyo, Japan
IMM	Mendeleev Institute of Metrology, Leningrad, USSR
IMS	Institute of Metrological Service, Moscow, USSR
LCIE	Laboratoire Central des Industries Électriques, Fontenay-aux-Roses, France
NIM	National Institute of Metrology, Beijing, PRC
NPL	National Physical Laboratory, Teddington, UK
NIST	National Institute of Standards and Technology (formerly National Bureau of Standards), Gaithersburg, USA
NRC	National Research Council of Canada, Ottawa, Canada
PTB	Physikalisch-Technische Bundesanstalt, Braunschweig, FRG
VSL	Van Swinden Laboratorium, Delft, The Netherlands
U. Zagreb	Faculty of Electrical Engineering, University of Zagreb, Yugoslavia

References

1. *Le BIPM et la Convention du Mètre* (Bureau International des Poids et Mesures, Sèvres, 1987)
2. BIPM Com. Cons. Électricité **13,** E13–E14 (1972). For the English-language version see J. Terrien: Metrologia **9,** 40–43 (1973)
3. BIPM Com. Cons. Électricité **17,** E112–E113 (1986)
4. Reference [3], pp. E6–E7; E88–E89. Also, P. Giacomo: Metrologia **24,** 45–51 (1987)
5. Reference [3], pp. E123, E127
6. G. W. Small: IEEE Trans. Instrum. Meas. **IM-36,** 190–195 (1987)
7. K. v. Klitzing, G. Dorda, M. Pepper: Phys. Rev. Lett. **45,** 494–497 (1980)
8. Reference [3], pp. E10–E12; E92–E94. Also, P. Giacomo: Metrologia **24,** 45–51 (1987)
9. BIPM Proc.-Verb. Com. Int. Poids et Mesures **54,** 8–10 (1986)
10. BIPM Comptes Rendus 18e Conf. Gén. Poids et Mesures, p. 100, (1987). For the English language version see P. Giacomo: Metrologia **25,** 113–119 (1988)
11. J. Clarke: Am. J. Phys. **38,** 1071–1095 (1970)
12. J. Clarke: Phys. Rev. Lett. **21,** 1566–1569 (1968)
13. W. H. Parker, D. N. Langenberg, A. Denenstein, B. N. Taylor: Phys. Rev. **177,** 639–664 (1969)
14. T. F. Finnegan, A. Denenstein, D. N. Langenberg, J. C. McMenamin, D. E. Novoseller, L. Cheng: Phys. Rev. Lett. **23,** 229–233 (1969)
15. T. D. Bracken, W. O. Hamilton: Phys. Rev. B **6,** 2603–2609 (1972)
16. J. C. MacFarlane: Appl. Phys. Lett. **22,** 549–550 (1973)
17. J. C. Gallop: *A Test of the Material Independence of the AC Josephson Effect* (NPL Report Qu 25, Division of Quantum Metrology, National Physical Laboratory, Teddington, UK 1973), 10 pp

18. J.-S. Tsai, A. K. Jain, J. E. Lukens: Phys. Rev. Lett. **51,** 316–319 (1983). Erratum: ibid., 1109
19. J. Niemeyer, L. Grimm, W. Meier, J. H. Hinken, E. Vollmer: Appl. Phys. Lett. **47,** 1222–1223 (1985)
20. J. Niemeyer, L. Grimm, C. A. Hamilton, R. L. Steiner: IEEE Electron Device Lett. **EDL-7,** 44–46 (1986)
21. A. K. Jain, J. E. Lukens, J.-S. Tsai: Phys. Rev. Lett. **58,** 1165–1168 (1987)
22. R. L. Kautz, F. L. Lloyd: Appl. Phys. Lett. **51,** 2043–2045 (1987)
23. D. N. Langenberg, J. R. Schrieffer: Phys. Rev. B **3,** 1776–1778 (1971)
24. J. B. Hartle, D. J. Scalapino, R. L. Sugar: Phys. Rev. B **3,** 1778–1781 (1971)
25. T. A. Fulton: Phys. Rev. B **7,** 981–982 (1973)
26. E. R. Cohen, B. N. Taylor: Rev. Mod. Phys. **59,** 1121–1148 (1987)
27. R. E. Prange, S. M. Girvin (eds.): *The Quantum Hall Effect* (Springer-Verlag, New York 1987). This book provides an excellent summary of the theoretical and experimental situation through 1985
28. M. E. Cage, R. F. Dziuba, B. F. Field, E. R. Williams, A. C. Gossard, D. C. Tsui, R. J. Wagner: Phys. Rev. Lett. **51,** 1374–1377 (1983)
29. M. E. Cage, B. F. Field, R. F. Dziuba, S. M. Girvin, A. C. Gossard, D. C. Tsui: Phys. Rev. **B30,** 2286–2288 (1984)
30. L. Bliek, E. Braun, F. Melchert, P. Warnecke, W. Schlapp, G. Weimann, K. Ploog, G. Ebert, G. E. Dorda: IEEE Trans. Instrum. Meas. **IM-34,** 304–305 (1985)
31. A. Hartland, D. J. Davies, D. R. Wood: IEEE Trans. Instrum. Meas. **IM-34,** 309–314 (1985)
32. K. Yoshihiro, J. Kinoshita, K. Inagaki, C. Yamanouchi, T. Endo, Y. Murayama, M. Koyanagi, A. Yagi, J. Wakabayashi, S. Kawaji: Phys. Rev. B **33,** 6874–6896 (1986)
33. F. Delahaye, D. Dominguez, F. Alexandre, J. P. Andre, J. P. Hirtz, M. Razeghi: Metrologia **22,** 103–110 (1986)
34. F. Delahaye, D. Dominguez: IEEE Trans. Instrum. Meas. **IM-36,** 226–229 (1987)
35. H. Aoki, T. Ando: Phys. Rev. Lett. **57,** 3093–3096 (1986)
36. H. Aoki: Rep. Prog. Phys. **50,** 655–730 (1987)
37. D. R. Yennie: Rev. Mod. Phys. **59,** Pt. 1, 781–824 (1987)
38. R. Kaarls: BIPM Proc.-Verb. Com. Int. Poids et Mesures **49,** A1–A12; 26 (1981). For the English-language version see P. Giacomo: Metrologia **17,** 69–74 (1981) and **18,** 41–44 (1982)
39. BIPM Proc.-Verb. Com. Int. Poids et Mesures **54,** 35 (1986). For the English language version see P. Giacomo: Metrologia **24,** 45–51 (1987)
40. G. J. Sloggett, M. F. Currey, D. J. Benjamin: Document CCE/88-1a and Document CCE/88-1b (Addendum) submitted to the 18th meeting of the CCE, 1988
41. W. K. Clothier, G. J. Sloggett, H. Bairnsfather, M. F. Currey, D. J. Benjamin: Metrologia **26** (1989), to be published
42. V. Bego, J. Butorac, K. Poljancic, I. Kunst, B. Ferkovic, G. Gasljevic, D. Vujevic, M. Brezinscak: *Determination of the Volt and 2 e/h at the Faculty of Electrical Engineering in Zagreb* (Report of the Electrotechnical Faculty, University of Zagreb 1988), 153 pp
43. B. P. Kibble, I. A. Robinson, J. H. Belliss: *A Realization of the SI Watt by the NPL Moving-Coil Balance* (NPL Report DES 88, Division of Electrical Science, National Physical Laboratory, Teddington, UK 1988), 51 pp; and Document CCE/88-16 submitted to the 18th meeting of the CCE, 1988
44. P. T. Olsen, R. E. Elmquist, W. D. Phillips, E. R. Williams, G. R. Jones, Jr., V. E. Bower: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published; and Document CCE/88-37 submitted to the 18th meeting of the CCE, 1988
45. J. Q. Shields, R. F. Dziuba, H. P. Layer: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published; and Document CCE/88-31 submitted to the 18th meeting of the CCE, 1988
46. M. E. Cage, R. F. Dziuba, R. E. Elmquist, B. F. Field, G. R. Jones, Jr., P. T. Olsen, W. D. Phillips, J. Q. Shields, R. L. Steiner, B. N. Taylor, E. R. Williams: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published; and Document CCE/88-32 submitted to the 18th meeting of the CCE, 1988
47. B. N. Taylor: Document CCE/88-12 submitted to the 18th meeting of the CCE, 1988. See also J. Res. Natl. Bur. Stand. **90,** 91–94 (1985)
48. H. Liu, R. Liu, P. Shen, T. Jin, Z. Lu, X. Du, B. Yu: Document CCE/88-11 submitted to the 18th meeting of the CCE, 1988
49. Y. Ruan, X. Wang, T. Yin, Z. Zhang: Document CCE/88-7 submitted to the 18th meeting of the CCE, 1988
50. N. Shen et al.: Document CCE/88-23 submitted to the 18th meeting of the CCE, 1988
51. G. Zhou, G. Zhao, M. Fa, L. Wang, Y. Zhang: Document CCE/88-24 submitted to the 18th meeting of the CCE, 1988
52. B. P. Kibble, G. J. Hunt: Metrologia **15,** 5–30 (1979)
53. E. R. Williams, P. T. Olsen: Phys. Rev. Lett. **42,** 1575–1579 (1979)
54. R. D. Cutkosky: IEEE Trans. Instrum. Meas. **IM-23,** 305–309 (1974)
55. T. J. Witt, F. Delahaye, D. Bournaud: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published; and Document CCE/88-36 (Rapport BIPM-88/6) submitted to the 18th meeting of the CCE, 1988
56. R. D. Deslattes: Private communication, 1987; and B. Kramer (ed.), *The Art of Precise Measurement* (Physik Verlag, Weinheim 1988), pp. 193–208
57. T. Kinoshita: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published
58. P. Seyfried, R. Balhorn, M. Kochsiek, A. F. Kozdon, H.-J. Rademacher, H. Wagenbreth, A. M. Peuto, A. Sacconi: IEEE Trans. Instrum. Meas. **IM-36,** 161–165 (1987)
59. J. Forkert, W. Schlesok: Metrologische Abhandlungen **6,** No. 3–4, 165–175 (1986) (in German)
60. A. H. Wapstra, G. Audi, R. Hoekstra: *Atomic Masses from (Mainly) Experimental Data*, privately distributed report; and also At. Data Nucl. Data Tables **39,** 281–393 (1988)
61. P. Zhao, W. Lichten, H. Layer, J. Bergquist: Phys. Rev. Lett. **58,** 1293–1295 (1987). Erratum: ibid., p. 2506
62. G. W. Small, B. W. Ricketts, P. C. Coogan: Document CCE/88-18 submitted to the 18th meeting of the CCE, 1988
63. G. W. Small, B. W. Ricketts, P. C. Coogan: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published
64. F. Delahaye, A. Satrapinsky, T. J. Witt: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published; and Document CCE/88-25 submitted to the 18th meeting of the CCE, 1988
65. S. Kawaji, N. Nagashima, N. Kikuchi, J. Wakabayashi, B. W. Ricketts, K. Yoshihiro, J. Kinoshita, K. Inagaki, C. Yamanouchi: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published
66. A. Hartland, R. G. Jones, D. J. Legg: Document CCE/88-9 submitted to the 18th meeting of the CCE, 1988
67. F. Delahaye, A. Fau, D. Dominguez, M. Bellon: IEEE Trans. Instrum. Meas. **IM-36,** 205–207 (1987)
68. K. Shida, T. Wada, H. Nishinaka, K. Segawa: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published
69. M. E. Cage, R. F. Dziuba, C. T. Van Degrift, D. Yu: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published; and Document CCE/88-39 submitted to the 18th meeting of the CCE, 1988
70. V. A. Kuznetzov, V. Yu. Kaminskii, S. N. Lebedev, V. M. Pudalov, V. V. Sazhin, S. G. Semenchinsky, A. K. Yanysh, Yu. V. Tarbeyev, V. P. Shigorin: Document CCE/88-3 submitted to the 18th meeting of the CCE, 1988

71. I. Ya. Krasnopolin, V. M. Pudalov, S. G. Semenschinsky: Document CCE/88-5 submitted to the 18th meeting of the CCE, 1988
72. Z. Zhang, D. Wang, T. Yin, Z. Hu, J. Zheng, F. Liu, J. Zhou, L. Cai: Document CCE/88-8 submitted to the 18th meeting of the CCE, 1988
73. Y. Ruan, X. Wang, T. Yin, Z. Zhang: Document CCE/88-7 submitted to the 18th meeting of the CCE, 1988
74. R. S. Van Dyck, Jr., P. B. Schwinberg, H. G. Dehmelt: Phys. Rev. Lett. **59,** 26–29 (1987)
75. E. R. Williams, G. R. Jones, Jr., S. Ye, R. Liu, H. Sasaki, P. T. Olsen, W. D. Phillips, H. P. Layer: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published; and Document CCE/88-38 submitted to the 18th meeting of the CCE, 1988
76. R. L. Steiner, B. F. Field: IEEE Trans. Instrum. Meas. **IM-38** (1989), to be published
77. E. Braun, P. Gutmann, P. Klinzmann, H. Leontiew, P. Warnecke: Document CCE/88-41 submitted to the 18th meeting of the CCE, 1988
78. J. R. Sapirstein, E. A. Terray, D. R. Yennie: Phys. Rev. D **29,** 2290–2314 (1984)
79. M. I. Eides, V. A. Shelyuto: Phys. Lett. **146B,** 241–243 (1984)
80. V. V. Starshenko, R. N. Faustov: Kratkie Soobshcheniy OIYaI 7, 39–44 (July 1985) (in Russian). [Joint Institute for Nuclear Research (JINR), USSR Rapid Communication.]
81. S. G. Karshenboim, V. A. Shelyuto, M. I. Eides: Zh. Eksp. Teor. Fiz. **92,** 1188–1200 (1987) [English transl.: Sov. Phys. JETP **65,** 664–670 (1987)]
82. M. I. Eides, S. G. Karshenboim, V. A. Shelyuto: Phys. Lett. **202B,** 572–574 (1988)

A Measurement of the NBS Electrical Watt in SI Units

P. Thomas Olsen
Randolph E. Elmquist
William D. Phillips
Edwin R. Williams
George R. Jones, Jr.
Vincent E. Bower

Reprinted from
IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT
Vol. 38, No. 2, April 1989

A Measurement of the NBS Electrical Watt in SI Units

P. THOMAS OLSEN, MEMBER, IEEE, RANDOLPH E. ELMQUIST, WILLIAM D. PHILLIPS, EDWIN R. WILLIAMS, MEMBER, IEEE, GEORGE R. JONES, JR., AND VINCENT E. BOWER

Abstract—**We have measured the NBS electric watt in SI units to be: $W_{NBS}/W = K_W = 1 - (16.69 \pm 1.33)$ ppm.The uncertainty of 1.33 ppm has the significance of a standard deviation and includes our best estimate of random and known or suspected systematic uncertainties. The mean time of the measurement is May 15, 1988. Combined with the recent measurement of the NBS ohm in SI units: $\Omega_{NBS}/\Omega = K_\Omega = 1 - (1.593 \pm 0.022)$ ppm, this leads to a Josephson frequency/voltage quotient of $E_J = E_0[1 + (7.94 \pm 0.67)$ ppm] where $E_0 = 483\,594$ GHz/V.**

I. Introduction

IN THE International System of Units (SI), the electrical units are defined in such a way that the electrical unit of power, the volt-ampere, is identical to the mechanical unit of power, the newton-meter/second, and each of these units is a watt. In the laboratory representation of the electrical units, or "laboratory system of units," where the volt is defined by the Josephson effect and the ohm by the quantum Hall effect (or by reference to an artifact resistance standard), the units of electrical and mechanical power are not necessarily equivalent. Our experiment compares the NBS laboratory electrical watt to the mechanical, SI watt. It is, in effect, an electrical realization of the SI watt. In spirit, this experiment is very much like a realization of the ampere, and its relationship to more traditional ampere realizations or absolute ampere experiments, has been described earlier [1]. The measurement is based on an idea first proposed by Kibble [2].

II. Theory

Consider the circuit of Fig. 1(a). Two coils carry currents I_1 and I_2. The vertical, z-component of the force between them, F_z, is given by the derivative of the mutual inductance: $F_z = I_1 I_2 (dM_{12}/dz)$. This vertical force can be compared to a gravitational force mg using a balance. Now consider the same two coils in Fig. 1(b), where coil 2 is open-circuit and a voltmeter measures the EMF generated across it as it is moved along z with respect to coil 1. The generated EMF is given by: $\mathcal{E} = I_1 (dM_{12}/dt)$. Combining the expressions for F_z and $\mathcal{E}$, we have

$$F_z v = I_2 \mathcal{E} \tag{1}$$

Manuscript received June 10, 1988.

The authors are with the National Bureau of Standards, Electricity Division, Gaithersburg, MD 20899.

IEEE Log Number 8824967.

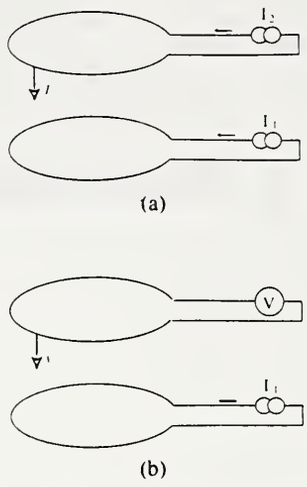


Fig. 1. Conceptual diagram of the experiment showing (a) the force measurement and (b) the EMF measurement.

where $v = dz/dt$. This simply expresses the equivalence of mechanical and electrical power, all quantities being expressed in SI units.

To express the electrical quantities in our laboratory units we define $K_A = A_{NBS}/A$, the ratio of the NBS and SI units of current. K_V, K_Ω, and K_W are defined in a similar way. Taking I_{NBS} to mean the current measured in NBS units, and similarly for $\mathcal{E}_{NBS}$, we have $I_2 = I_{NBS} K_A$, and $\mathcal{E} = K_V \mathcal{E}_{NBS}$. Substituting these expressions in (1) we have

$$K_W = K_A K_V = \frac{F_z v}{I_{NBS} \mathcal{E}_{NBS}}. \tag{2}$$

If K_Ω is separately measured, as in an absolute ohm experiment, we can obtain K_V and K_A from K_W:

$$K_A^2 = K_W/K_\Omega,\ K_V^2 = K_W K_\Omega. \tag{3}$$

Since $K_V = E_{0,NBS}/(2e/h)$ where $E_{0,NBS} = 483\,593.420$ GHz/V, a determination of K_V determines $2e/h$ in SI units.

Because it is difficult in practice to measure accurately an instantaneous velocity or voltage, we do not use (2) directly to determine K_W. Instead, we use an integral form of (2), obtained by integrating the expression for the generated EMF over time and the expression for the force over distance. Experimentally, we realize these integrals by: a) measuring the EMF as a function of time while moving coil 2, and b) measuring the current required in

coil 2 to maintain a constant force between the coils for various static positions of coil 2. It can be shown that K_W is given by

$$K_W = \frac{mg \int_{z1}^{z2} [I_0/I(z)]dz}{I_{0NBS} \int_{t1}^{t2} \mathcal{E}_{NBS}(t)\, dt}. \qquad (4)$$

Here $\mathcal{E}_{NBS}(t)$ is the generated voltage when the moveable coil is moved along some path in the field of the fixed coil; mg is the fixed gravitational force against which the magnetic force is balanced. $I(z)$ is the current in the moveable coil needed to create that magnetic force at a position z. The current I_0 is $I(z_0)$, where z_0 is some arbitrary reference point where $I(z)$ is measured. The quantity $\int[I_0/I(z)]dz$ is called the force integral and is a quantity which depends only on the geometry of the coils and on the path of integration, but not on the current in the coils. Equation (4) requires that the current in the fixed coil (i.e., the field coil) and the geometry of all the coils remain constant between the time that $\int\mathcal{E}_{NBS}(t)\, dt$ and I_0 are measured. Furthermore, the path of the moveable coil in the time integral and in the distance integral (i.e., the force integral) must be identical. Finally, the integral $\int F_z dz$ (to which $\int mg[I_0/I(z)]\, dz$ is equivalent) must account for all of the mechanical work done in moving the coil from z_1 to z_2; that is, either the displacements dx and dy must be zero along the chosen path or the force components F_x and F_y must be zero.

III. Apparatus

The apparatus used to make the measurements required in (4) has been described in some detail previously [1]. Here we will review the apparatus only very briefly. The apparatus is shown schematically in Fig. 2. Coil C is the moveable, suspended coil which is equivalent to coil 2 in Fig. 1. The fixed field coils B, above and below the suspended coil, are equivalent to coil 1 in Fig. 1. In the experiment reported here, these B coils are wound with copper wire and cooled by being immersed in a circulating oil bath. We also have available a superconducting version of coils B which produces a field two orders of magnitude larger than the present coils, and which will be used in future measurements [3].

The suspended coil C is attached to the "balance wheel" A through the spider E. The spider also supports a pan for the standard masses. Wheel A is an aluminum disc with a knife edge at its center. A band consisting of 40 fine wires hangs from the wheel on both sides. As the wheel rotates, it acts like a pulley, raising or lowering the suspended coil. The advantage of this arrangement, compared to hanging the coil from one arm of a conventional beam balance, is that the translation of the coil is almost perfectly vertical, helping to satisfy the requirement for validity of (4) that there be no horizontal displacements.

Coils G along with permanent magnets F provide the means by which the balance wheel is made to rotate. A servo feedback to these drive coils controls the motion of the suspended coil for the generation of the EMF $\mathcal{E}$. The position of the suspended coil is measured with a laser interferometer. Servo feedback to the suspended coil maintains this position while we measure the current in the suspended coil required to balance the force mg.

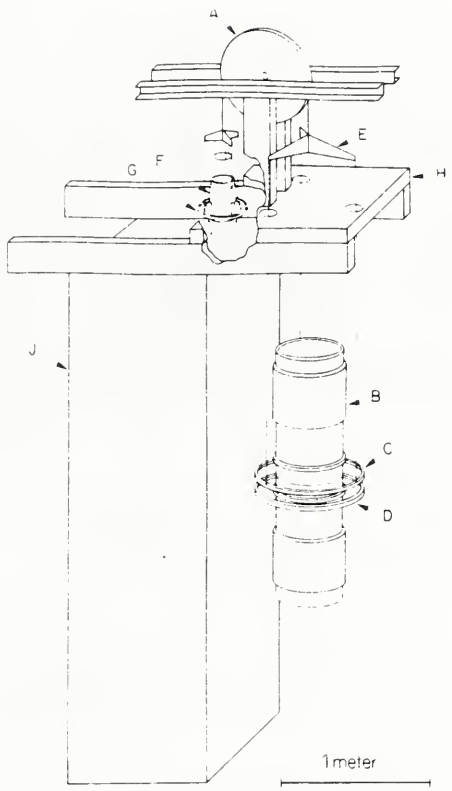


Fig. 2. The apparatus of the NBS absolute watt measurement.

For more details concerning the apparatus, the reader is referred to [1]. Among the significant changes since that wiring is that the suspended coil now is wound in three separate concentric sections with a total of 2355 turns of copper wire. The suspended coil has a total resistance of 480 Ω. Reversing a current of 50 mA through this coil changes the magnetic force by about the equivalent force produced by a mass of 105 g. Gold or brass 105-g mass standards are used to measure the force. We also have used a 70-g mass standard, reversing 33 mA through the coil. The 50- or 33-mA current is passed through a 20- or 30-Ω resistor which allows comparison of the voltage against our working 1-V voltage standard. The 20- or 30-Ω resistor is also used to generate 20 or 30 mV as a reference for the generated EMF when the coil moves. Use of the same 20- or 30-Ω resistor for these two functions eliminates the need for precise calibration of these resistors. Furthermore, the 20-Ω resistor is made of two 10-Ω resistors and alternate use of these 10-Ω resistors to generate the 10-mV reference eliminates the need for their calibration. These changes, along with complete automation of the measurement, have allowed us to achieve a substantial improvement in precision over that reported previously.

IV. Procedures

The measurement of K_W is divided into three phases, corresponding to three key measurable quantities in (4). In the first phase, the suspended coil moves under servo control so as to generate a nearly constant EMF. Data accumulated during this "voltage" phase is used to compute the time integral of the voltage. In the second, or "force" phase, we measure the current I_0 which balances the force mg at a specific position of the suspended coil. These two measurement phases are repeated many times in succession. At a different time, often separated from the first two phases by several days, we perform the third phase or "force integral" measurement. This is accomplished by comparing the current needed to balance mg at many z-positions of the suspended coil.

All of the measurements use a 1-mA current source to supply working voltage references. This current source, which can be reversed with a ramped reversing switch, supplies 10-, 20-, 30-Ω, and 1-kΩ resistors, producing 10-, 20-, 30-mV, and 1-V references. These working voltage references, produced from stable resistors, are used throughout the measurement. The current source is calibrated by comparing the 1-V reference to a Zener voltage standard which is periodically calibrated against the NBS volt and by calibrating the 1-kΩ resistor in terms of Ω_{NBS}. Note that the 10-, 20-, and, 30-Ω resistors need not be precisely calibrated (see discussion at the end of the previous section).

In the voltage phase of the measurement the EMF generated by the moving coil is servoed to be nominally equal to the 10-, 20-, or, 30-mV reference. At 20-mV, this produces a velocity of about 2 mm/s. When the moving coil position, as measured by the laser interferometer, reaches selected positions along the vertical path, the computer is triggered to perform measurements of the error voltage (difference between the reference voltage and the generated voltage) and of the time from a 1-MHz clock. The triggerings occur at rates up to 400 Hz for the 20-mV reference level and at somewhat lower rates for the other levels, limited by the data-taking speed of the computer.

With a perfect servo, the error voltage would be zero, as the velocity is adjusted to produce a constant EMF. In reality, there is a typical error on the order of a microvolt. In principle, even with such an error voltage, a correct measurement of the voltage and the times should lead to the correct time integral of the voltage in (4). Unfortunately, because of the time constant of the linear amplifier used to detect the error voltage, we do not have the true instantaneous generated EMF at the time data points are acquired. To correct for this in the calculation of the time integral, we have digitally filtered the measured time points with the same time constant measured for the linear amplifier.

The total path length over which we collect data is 7.6 cm. Up to 15 000 points are recorded during a traversal. After each such traversal the current source is reversed, reversing the reference voltage. Under servo control, this reverses the direction of travel of the suspended coil, and another measurement is taken going in the opposite direction. Some computation of the time integral of the generated EMF, including digital filtering of the time data, is performed during the period when the coil turns around for another traversal. The compressed data is stored for later analysis. The integral is taken over a path whose length is 5.9 cm (370 000 interferometer fringes of $\lambda/4$ for λ = 633 nm). The time interval for the voltage integral is about 30 s when the generated EMF is 20 mV. A set of 10 traversals in each direction is measured, which, with the time for turning around, requires between 18 and 36 min, depending on the chosen reference voltage. To eliminate the effect of drift and of zero offsets, we interpolate the voltage integral obtained for two up (down) traversals to the time of a down (up) traversal and compute the up/down difference in voltage integrals. Typical scatter in such differences over a set of traversals is a bit less than 1 ppm.

During the force phase the suspended coil position is servoed to a fixed location by feeding back the interferometer measurement as a current to the coil. This current is measured with the standard mass alternately on or off the pan connected to the spider from which the coil is hung. A force measurement typically consists of eight reversals where the mass is lowered onto or raised from the suspended pan. For each positioning of the mass, the coil current is measured by passing it through a 20- or 30-Ω resistor and comparing the voltage drop to the 1-V reference. The balance is arranged so that when the mass is lifted or replaced the current in the coil must reverse in order to maintain the servo position. Current is measured for about 2 min in each mass position. Including the time for raising and lowering the mass and stabilizing the servo, the elapsed time for force measurements is about 25 min. To eliminate the effect of drift and zero offsets we interpolate between two measurements with the mass off (on) to the time of a measurement with the mass on (off) to obtain the on/off current difference. The scatter in such differences, made over 25 min, is considerably less than 1 ppm for most measurements.

On a typical day of taking data, we obtain 10–16 pairs of voltage and force measurements, at the rate of about one pair per hour, including the time required to automatically reconfigure the apparatus when changing from voltage to force and vice versa.

The final phase of the measurement, the determination of the "force profile," is very similar to the force measurements described above. Here, we simply make repeated force measurements, but at various positions of the suspended coil. To eliminate the effect of drifts, we continually measure the reversal current at the reference position used in the force phase. Thus we obtain the ratio of the reversal current at various positions, $I(z)$, to the current at the reference position, I_0. The measurements are repeated at different positions continuously for several

days. The ratios obtained are fit to a polynomial and analytically integrated over the same interval used for the voltage integral.

In order to minimize the influence of choosing a particular set of endpoints for the integration interval, we calculated the voltage integral for 20–50 overlapping regions with different sets of endpoints. The endpoints of the fixed-length intervals used were changed by about a third of the total distance used in the analysis. From each voltage integral, and the appropriate force integrals for each interval, we obtain a value of K_W. To reduce the noise and to eliminate effects which contribute at the endpoints of the integral, such as variations in the velocity which might be due to servo errors, we averaged these with equal weight. (This process tends to weight measurements of voltage and time taken near the ends of the analyzed region less than those in the central region, since the integrated region always included one or two centimeters of the region nearest the center.)

V. Results and Discussion

Fig. 3 shows a histogram of 191 values of K_W obtained over a period of approximately one month. Each value contributing to the histogram is obtained as follows: from the values of two successive voltage integrals, typically taken an hour apart, we interpolate a voltage integral at the mean time of the intervening weighing measurement and calculate a K_W from the actual weighing and the interpolated voltage integral. Similarly, we interpolate between two successive weighings to the time of a voltage integral measurement and obtain a K_W from those values. The values in the histogram are averages of two successive values of K_W, one from two voltage integrals and a weighing, and one from two weighings and a voltage integral. The next two values of K_W so obtained are used to produce another point on the histogram. Each point, therefore, contains information from two weighings and two voltage integrals, corresponding to approximately two hours of elapsed time. While some of the same information is used in adjacent points, they are nearly independent. The standard deviation of all the values of K_W obtained is 1.1 ppm.

The interpolation process described above is supposed to account for variations in the current or geometry of the fixed field coils between voltage and force measurements. Fig. 4 shows the voltage and force measurements themselves during the course of a typical day. The fact that they follow the same pattern of variation is good evidence that the experiment is in control and operating as expected.

Fig. 5 shows the results of a typical measurement of the force profile, along with the polynomial fit to the measured points. A similar profile can be obtained by integrating the voltage data over appropriately small time intervals. Because of the short time intervals, such a profile is noisier than the directly measured force profile, but it has the same shape, as expected.

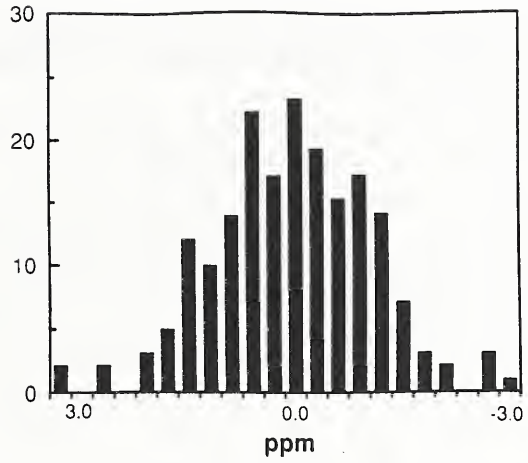


Fig. 3. A histogram showing the distribution of watt measurements taken in May 1988. The number of two-hour measurements obtained are represented by the lengths of the solid bars; the value of the watt for the measurements is given in ppm from the mean.

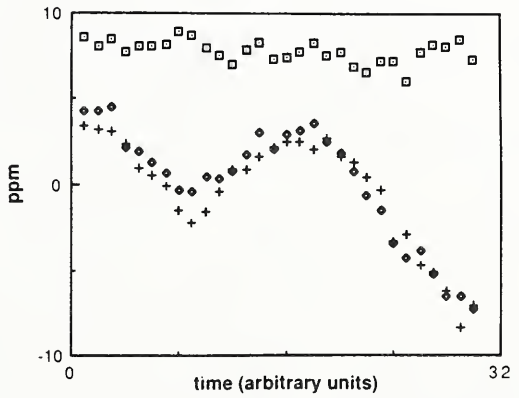


Fig. 4. Relative values of the watt (squares), force (diamonds), and EMF/velocity (crosses) in a measurement sequence lasting about 24 h.

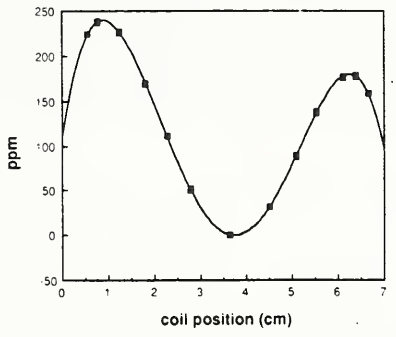


Fig. 5. A set of force measurements at thirteen vertical positions of the moveable coil and a sixth-order fit to the force profile.

For the final analysis, we have divided the data by days and averaged the various days together with equal weighting. The standard deviation of 13 days of data is 0.80 ppm. The fact that the standard deviation of 13 days is not much reduced from the standard deviation of 191 individual points suggests that day-to-day scatter is not purely statistical, but hides some systematic error possibly related to conditions which change from one day to the next. Because of this, we felt it was prudent to average the days together without weighting them either according to the time spent taking data on a given day or according

to the variance of a given day's data. For a discussion of the statistical treatment of data with differences between groups of measurements, see [4].

The mean of 13 days of data taken over a period of approximately the month of May 1988 was $K_W = 1 + \delta_W$, where $\delta_W = -16.69$ ppm. If we had taken the mean of all the results in the histogram (equivalent to weighting each day according to the amount of data taken on that day) we would get a δ_W 0.12-ppm lower. The voltage generated in the voltage-velocity part of the measurement was nominally 10, 20, or 30 mV. The velocity in turn also varied as the voltage. If we consider the data obtained at each of these voltages separately, we obtain $\delta_W = -16.96$, -16.51, and -16.74 ppm for 10, 20, and 30 mV, respectively. The standard deviation as estimated from these three values is 0.23 ppm.

Data was actually taken on 17 days during May, but four of these sets were discarded. Those four were the ones in which the voltage for the voltage-velocity part was sampled at rates of ~100 and 50 Hz. We kept only that 10-mV data sampled at ~200 Hz. The reason for discarding that data is as follows: both the velocity and the voltage measured during the voltage-velocity phase are modulated at about 30 Hz with an amplitude as high as 2 percent of the velocity and 0.1 percent of the voltage. This modulation is thought to be the result of mechanical vibrations of the building which couple into the balance. Computer simulations of the acquisition of data from such a modulated source indicate that when the sampling rate is too low, a significant error (at the parts per million level) can result from the inability to accurately integrate the modulated voltage signal. For 20-mV data, the percent modulation is smaller, and simulations show the effect to be much smaller. We find experimentally that the value of K_W obtained at 10 mV depends on the sampling rate, while at 20 mV it does not. We, therefore, eliminate the questionable 10-mV data taken at the low sampling rate. Had we included that data, we would have $\delta_W = -16.94$ ppm. Taking all the 10-mV data by itself gives $\delta_W = -17.41$, while if we eliminate all the 10-mV data, the 20- and 30-mV data give $\delta_W = -16.60$ ppm. Considering the results of the simulations and the observed variation with sampling rate, we feel justified in discarding the slow 10-mV data. However, we do consider the effect of the sampling rate in assigning the final uncertainty.

The voltages used to compute the voltage integral are measured with a low noise dc linear amplifier and sampled using an A/D converter. The amplifier has a frequency response which effectively filters the voltage appearing at its input. Roughly speaking, this filtering corresponds to a time constant of about 50 ms. As discussed above, to correct this, we digitally apply the measured filter function of the null detector to the time data. The result quoted is for data so filtered. Comparing the filtered data with unfiltered data where the raw times are used to calculate the voltage integral, we find that for unfiltered data $\delta_u = -15.88$ ppm, a 0.75-ppm difference from the filtered data. The difference is greater for higher velocity data. Computer simulations of the data indicate that filtering is the proper way to handle the data, so we have used only filtered data.

TABLE I
ESTIMATED ONE STANDARD DEVIATION UNCERTAINTIES

Statistical and undiscovered systematics		0.80 ppm
Analysis		0.60
Different force profiles		0.50
Polynomial representation of force profile		0.42
Gain of detector		0.36
Effect of vibrations		0.30
Effect of different velocities		0.20
Verticality		0.10
Standards		0.30
TOTAL (Root of summed squares)		1.33 ppm
Standards		
Volt transfer	0.15	
Current calibration	0.10	
Laser wavelength	0.18	
Resistor calibration	0.09	
Mass calibration	0.11	
Gravity	0.05	

Table I shows our one standard deviation uncertainty estimates for K_W.

1) Statistical and Undiscovered Systematics: The entry of 0.80 ppm is the day-to-day standard deviation of the 13 days used in the determination. While we might reduce this by $\sqrt{13}$ to 0.22 ppm, the suggestion that the data are not statistical and that there may be some undiscovered systematics (such as a remaining error due to insufficient sampling rate, for example) leads us to use the full day-to-day standard deviation.

2) Analysis: The entry of 0.6 ppm refers to uncertainties arising from shifts found to arise in K_W when different data analyses are used, including filtering of time data and averaging of voltage data before integrating.

3) Different Force Profiles: This represents uncertainties which were estimated by deliberately changing the configuration of the fixed coils or by using different parts of the total path of the moveable coil to determine K_W. Shifts in K_W from such changes lead to an uncertainty of 0.5 ppm.

4) Polynomial Representation of the Force Profile: The force profile is measured at a large number of points and fit with a polynomial of order 4–6. The statistical uncertainty of the fit and the changes in the integral with the order of the polynomial fit contribute 0.42 ppm.

5) Gain of Detector: The gain of the dc linear amplifier used in voltage measurements and in current measurements during weighing is important because the mea-

surements are always somewhat off null. Uncertainties in this calibration contribute 0.36 ppm.

6) Effect of Vibrations: Mechanical vibrations are responsible for the observed 30-Hz modulation, which, when added to simulated data, leads to errors. We can also deliberately apply vibrations which are seen to shift K_W. The size of these shifts for large applied vibrations leads us to assign 0.3-ppm uncertainty associated with the actual vibrations. The equipment necessary to cool the large electromagnet is one source which produces mechanical and acoustic vibrations. This noise source is a source of uncertainty in the measurement which will be eliminated with the superconducting magnet.

7) Effect of Different Velocities: Choice of different generated voltages in the voltage phase leads to different velocities. Because of the finite time response of the detector and the servo system, this might lead to errors. We assign 0.2 ppm-uncertainty based on the scatter between results obtained at different velocities.

8) Verticality: The interferometric measurements of position of the moveable coil must be made along a vertical line to correspond to the direction in which forces can be measured by the balance. The error in setting the interferometer path to vertical contributes 0.1 ppm.

9) Standards: The total effect of uncertainties in the various calibrations and standardizations is 0.3 ppm, which reflects not only the basic calibration uncertainty but also uncertainties in corrections applied to the calibrations. Volt transfer includes the accuracy of the calibration of our Zener reference against the NBS volt and the drift between calibrations. "Current calibrations" refers to errors in establishing our working standard of voltage against which the weighing currents and the generated voltages are measured. Laser wavelength uncertainties include errors in the refractive index of air due to uncertainty in temperature, pressure, humidity, and CO_2 content of the local atmosphere. Resistor calibration includes uncertainty in the drift of the resistor compared to the NBS ohm. Mass calibration includes uncertainties in the buoyancy corrections due to measurement of the parameters of the local atmosphere. Gravity uncertainty includes uncertainty in the absolute gravimeter measurement at a reference site, transfer to the experimental site, and tidal variations in the gravitational acceleration.

We have considered other sources such as leakage resistances, frequency standard errors, and the influence of nonvertical forces; and have concluded that all such identified errors are significantly less than 0.1 ppm and so do not need to be considered in our uncertainty estimate. We should point out that the evaluation of our most significant errors is limited by the day-to-day fluctuations in measurement. We expect that the accumulation of a significantly larger set of data will allow most errors to be reduced.

Furthermore, use of the superconducting field coils will increase the fixed magnetic field by two orders of magnitude. This will allow the force and the generated voltage

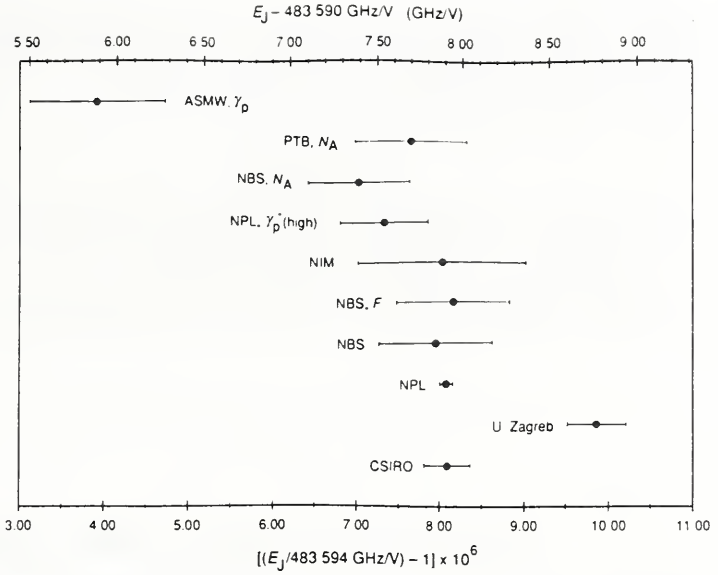


Fig. 6. Comparison of the values of E_J and their one-standard-deviation uncertainties (values as compiled by B. N. Taylor).

to be increased substantially, reducing most of the largest errors by about an order of magnitude.

The final result for our present measurement of the NBS watt in SI units is

$$\frac{W_{\text{NBS}}}{W} = K_W = 1 - (16.69 \pm 1.33) \times 10^{-6}.$$

(May 15, 1988)

The uncertainty of 1.33 ppm has the significance of a standard deviation and includes our best estimate of random and known or suspected systematic uncertainty. The mean time of the measurement is May 15, 1988, and refers to the NBS volt and ohm as of that date. Combined with the measurement of May 17, 1988 of the NBS ohm in SI units which implies $\Omega_{\text{NBS}}/\Omega = K_\Omega = 1 - (1.593 \pm 0.022)$ ppm, this leads to a Josephson frequency/voltage quotient of

$$E_J = E_0\left[1 + (7.94 \pm 0.67)\ \text{ppm}\right]$$

where $E_0 = 483\ 594$ GHz/V.

A comparison of the current experimental values of E_J are given in Fig. 6. Direct force measurements are contributions from CSIRO (Australia), U. Zagreb (Yugoslavia), NPL (U.K.), and this paper NBS (U.S.). Values of derived E_J are from the Faraday: NBS F (U.S.); from high and low field gamma-p: NIM (PRC), NPL high NBS low, ASMW (DDR); and the derivation resulting from Avogadro's number: NBS N_A (U.S.) and PTB, N_A (FRG).

Acknowledgment

The authors would like to thank the following colleagues for their valuable contributions to this measurement: R. F. Dziuba for resistor calibrations; R. S. Davis for mass calibrations and density measurements; H. Layer for laser wavelength calibrations; R. Steiner and Y. Sheng for voltage calibrations; J. A. Hammond and M. Zum-

berge for separate measurements of the acceleration of gravity near the site of the experiment; and B. N. Taylor for his continued support and encouragement.

References

[1] P. T. Olsen, V. E. Bower, W. D. Phillips, E. R. Williams, and G. R. Jones, Jr., "The NBS absolute ampere experiment," *IEEE Trans. Instrum. Meas.*, vol. IM-34, pp. 175–181, June 1985.

[2] B. P. Kibble, "A measurement of the gyromagnetic ratio of the proton by the strong field method," *Atomic Masses and Fundamental Constants*. J. H. Sanders and A. H. Wapstra, Eds. New York: Plenum, 1976.

[3] W. Y. Chen, J. R. Purcell, P. T. Olsen, W. D. Phillips, and E. R. Williams, "Design and construction of a superconducting magnet system for the absolute ampere experiment," in *Advances in Cryogenic Engineering*, R. W. Fast, Ed. New York: Plenum, vol. 27, pp. 97–104, 1982.

[4] H. H. Ku, "Statistical concepts in metrology," in *Precision Measurement and Calibration*, NBS Special Publ. 300, vol. 1, pp. 296–328, Harry H. Ku, Ed., 1969, see p. 317. Reprinted from the *Handbook of Industrial Metrology*. Amer. Soc. of Tool and Manufacturing Engineers. Englewood Cliffs, NJ: Prentice-Hall, pp. 20–52, 1967.

T-IM/38/2/1/26209

New Realization of the Ohm and Farad Using the NBS Calculable Capacitor

John Q. Shields
Ronald F. Dziuba
Howard P. Layer

Reprinted from
IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT
Vol. 38, No. 2, April 1989

New Realization of the Ohm and Farad Using the NBS Calculable Capacitor

JOHN Q. SHIELDS, RONALD F. DZIUBA, MEMBER, IEEE, AND HOWARD P. LAYER

Abstract—**Results of a new realization of the ohm and farad using the NBS calculable capacitor and associated apparatus are reported. The results show that both the NBS representation of the ohm and the NBS representation of the farad are changing with time, Ω_{NBS} at the rate of −0.054 ppm/year and F_{NBS} at the rate of 0.010 ppm/year. The realization of the ohm is of particular significance at this time because of its role in assigning an SI value to the quantized Hall resistance. The estimated uncertainty of the ohm realization is 0.022 ppm (1σ) while the estimated uncertainty of the farad realization is 0.014 ppm (1σ).**

I. Introduction

THE NBS REPRESENTATION of the ohm Ω_{NBS} is based on the mean resistance of five Thomas-type wire-wound resistors maintained in a 25°C oil bath at NBS Gaithersburg, MD. Similarly, the NBS representation of the farad F_{NBS} is based on the mean capacitance of four fused-silica capacitors in a comparable oil bath. By means of the United States calibration hierarchy, measurements of resistance, capacitance, and inductance made throughout the country are generally traceable to these representations.

Realization of the ohm and farad is necessary for two distinct reasons: first, to determine Ω_{NBS} and F_{NBS} in SI units, thereby ensuring that measurements based on these electrical quantities are consistent with the SI, the unit system used throughout the world; second, to determine in SI units a number of fundamental physical constants of importance to both physics and electrical metrology. These include the fine-structure constant α, the quantized Hall resistance $R_H = h/e^2$, and the Josephson frequency to voltage quotient $E_J = 2e/h$ (h is the Planck constant, e is the elementary charge). Indeed, it is likely that starting January 1, 1990, representations of the ohm worldwide will be based on a conventional value of R_H and representations of the volt will be based on a new conventional value of E_J, both consistent with the SI. These values are to be derived from the data available by June 15, 1988 [1].

In order to contribute to the pool of data, NBS, like other national standards laboratories, is carrying out experiments to determine R_H and E_J. Realizing the ohm by means of the NBS calculable capacitor is an important part of the NBS effort. This paper describes our measurements and gives our latest results.

Manuscript received June 10, 1988.

J. Q. Shields and R. F. Dziuba are with the Electricity Division, National Bureau of Standards, Gaithersburg, MD 20899.

H. P. Layer is with the Length and Mass Division, National Bureau of Standards, Gaithersburg, MD 20899.

IEEE Log Number 8826209.

II. AC Measurements

The measurement sequence used in the 1974 ohm and farad determinations [2] has been retained in the present NBS measurements. A 0.5-pF calculable cross-capacitor is used to measure a transportable 10-pF reference capacitor which is carried to the laboratory containing the NBS bank of 10-pF fused silica reference capacitors. A 10:1 bridge is used in two stages to measure two 1000-pF capacitors which are in turn used as two arms of a special frequency-dependent bridge for measuring two 100-kΩ resistors. A 100:1 bridge is used to compare each of the two 100-kΩ resistors with a 1000-Ω transportable resistor, R311, which is carried to the laboratory containing the NBS bank of 1-Ω resistors where the dc stepdown is made. The ac–dc difference of R311 is determined by means of a special 1000-Ω coaxial resistor of negligible ac–dc difference. All ac measurements are carried out at $\omega = 10^4$ rad/s $\simeq$ 1592 Hz.

The calculable capacitor [2], the ac bridges [3] and standards [4], the ac–dc resistance standard [5], and the equipment used to measure transformer ratios [6], [7] and voltage dependencies [8] remain basically the same as in the 1974 measurements. The calculable capacitor was partially disassembled in order to realign the electrical and optical axes, clean the optical flats, install larger diameter PTFE rings on the guard tubes, check for microphonic coupling errors [9], and measure the distributed inductances and capacitances used to calculate frequency corrections.

Residual gas in the calculable capacitor has been reduced to a negligible level by the installation of a turbomolecular pump. All of the ac voltage sources and some of the preamplifiers and phase-sensitive detectors are new. An automated data acquisition system is now used in the comparison of the calculable capacitor with the 10-pF reference capacitor, resulting in a standard deviation of 0.003 ppm for the random scatter in one complete measurement requiring about one hour.

III. DC Measurements

Relative to the 1974 measurements the most significant reduction in uncertainty is in the dc stepdown. To relate the 1000-Ω transportable resistor R311 to the reference

bank of 1-Ω resistors, measurements at 1, 100, and 1000 Ω are necessary using three different measurement systems. NBS-built Hamon-type resistance transfer standards [10] are used to provide accurate 1 : 100 and 1 : 10 resistance ratios to extend the 1-Ω reference bank to the 100-Ω level, and then to the 1000-Ω level. All of the resistance standards used in the scaling process to measure resistor R311 are measured *in situ*, immersed in specially designed circular oil baths with temperature maintained at (25.000 ± 0.002)°C.

An automated direct-current comparator system [11] is used to compare the five 1-Ω resistors in the reference bank to Hamon transfer standard H10A containing ten 10-Ω coils configured in the parallel mode. The resistance coils for Hamon H10A are individually sealed in brass cans filled with silicone heat-sink compound and terminated with mercury-wetted contacts. The connectors for the parallel configuration consist of (a) low-resistance amalgamated copper shorting bars for the current terminations, and (b) one of two fixtures having fan resistances of either 1 or 10 Ω for the potential terminations. A second Hamon standard of similar design but different construction is used as a check standard for scaling from 1 to 100 Ω.

The measurements at 100 Ω are made using an automated direct-current comparator system similar in design, construction, and operation to the 1-Ω system. Hamon H10A in its series mode is compared to a guarded Hamon standard H1kA containing ten 1000-Ω elements in its parallel configuration. The card-type resistance elements for H1kA are sealed in a thick-walled aluminum box filled with silicone fluid. The resistor terminations are British Post Office (BPO) connectors. A second Hamon standard containing ten 100-Ω resistors is used as a check standard for scaling from 100 to 1000 Ω.

Resistor comparisons at the 1000-Ω level are made using a guarded, resistance-ratio bridge. The adjustable part of the bridge is a modified direct-reading, double-ratio set. With this bridge the individual resistance sections of Hamon H1kA are compared to the 1000-Ω transportable resistor R311. Alternatively, Hamon H1kA can be compared to resistor R311 by first measuring the series–parallel combination of its first nine coils, and then measuring its tenth coil individually.

A set of three measured values is obtained for resistor R311 over a period of three consecutive days. Each day's measurement procedure consists of scaling from 1 to 1000 Ω and then back down to 1 Ω. The total uncertainty in assigning a value to resistor R311 in terms of the NBS 1-Ω reference bank is 0.010 ppm, the root-sum-square of the individual uncertainties listed in Table III. (Throughout this paper, all uncertainties are one standard deviation estimates.)

IV. Interferometer

The optical apparatus used to illuminate the capacitor interferometer has undergone complete renovation since the last series of measurements made in the early 1970's.

TABLE I
Uncertainties in the Measurement of F_{NBS} in SI Units

	one standard deviation estimate (ppm)
Type A (random)	0.003
Geometrical imperfections in the calculable capacitor	0.007
Laser/interferometer misalignment	0.003
Frequency (loading) corrections	0.004
Microphonic coupling	0.005
Voltage dependence	0.005
Drift between calibrations/failure to close	0.006
Transformer ratio measurement	0.002
Bridge linearity and phase adjustment	0.003
Detector uncertainties	0.002
Coaxial choke effectiveness	0.001
Temperature corrections for 10 pF capacitors	0.002
RSS	0.014

The Fabry–Perot (flat plate) interferometer is illuminated by a mode-stabilized He–Ne local oscillator whose wavelength is stable to within a few hundred kilohertz over the course of a measurement. Because this laser is not stabilized to a well-defined wavelength and may drift several megahertz per day, its frequency is continuously compared to an iodine stabilized laser whose wavelength is accurate to about three parts in 10^{10} and is stable to several parts in 10^{13} [12]. The lasers are compared by measuring the frequency of the heterodyne signal using a high-speed photo detector, electronic amplification, and a digital frequency counter. The frequency comparison contributes no significant error to the laser calibration.

The interferometer cavity is isolated from the working laser by using an acousto-optic modulator which shifts the frequency of the incident beam by 90 MHz and any light reflected back from the interferometer system by an additional 90 MHz. Because of the narrow bandwidth of the laser cavity, the return light, shifted from the incident light by 180 MHz, is no longer resonant in the laser cavity and is not coupled to the laser modes. No frequency pulling of the working laser wavelength due to back reflections from the interferometer or the beam steering optics could be detected. The uncertainty due to imperfect laser/interferometer alignment (Table I) was determined empirically by making comparative capacitance measurements after repeatedly misaligning the cavity and the beam steering optics and then realigning the system.

The laser beam diameter is expanded so that it fills the interferometer mirrors, and the entire transmitted beam is used to illuminate the photodetector so there is no beam truncation error. An integrating servo system locks the interferometer to the local oscillator with an accuracy of 1×10^{-3} of the interferometer transmission width and contributes negligible error to the determination of capacitance.

TABLE II
UNCERTAINTIES IN THE MEASUREMENTS RELATING THE NBS BANK OF 10-pF CAPACITORS TO THE TRANSPORTABLE 1000-Ω RESISTOR

	one standard deviation estimate (ppm)
Type A (random)	0.666
Transformer ratio measurements	0.665
Voltage dependence measurements	0.665
Bridge linearity and phase adjustment	0.666
AC/DC difference of transportable 1000 Ω resistor	0.665
Quadrature bridge harmonics	0.663
Drift between calibrations/failure to close	0.666
Coaxial choke effectiveness	0.667
Auxiliary adjustments on four terminal pair bridges	0.662
Detector uncertainties	0.663
RSS	0.674

TABLE III
UNCERTAINTIES IN THE STEPDOWN FROM THE TRANSPORTABLE 1000 Ω RESISTOR TO THE NBS BANK OF 1-Ω RESISTORS

	one standard deviation estimate (ppm)
Type A (random)	0.666
Temperature dependence of standards	0.663
Pressure dependence of standards	0.669
Loading effects of standards	0.663
Leakage effects of Hamon standards	0.663
Series parallel connections of Hamon standards	0.666
Linearity of detector systems	0.663
Drift between calibrations/failure to close	0.663
RSS	0.676

V. RESULTS

The results of the measurements are

$$\Omega_{NBS} = [1 - (1.594 \pm 0.022) \times 10^{-6}]\,\Omega \qquad (1)$$

$$F_{NBS} = [1 + (0.143 \pm 0.014) \times 10^{-6}]\,F \qquad (2)$$

where the mean date of each is May 17, 1988 and the uncertainties are one standard deviation estimates representing the root-sum-square of the appropriate uncertainties listed in Tables I–III.

Using the results of the last NBS ohm realization [2] which had a mean date of December 2, 1973, we find that Ω_{NBS} has changed by (-0.775 ± 0.035) ppm and the average rate of change of Ω_{NBS} is $(-0.053\,6 \pm 0.002\,4)$ ppm/year. This is in good agreement with the result $(-0.052\,9 \pm 0.004\,0)$ ppm/year obtained from measurements of the quantized Hall resistance [13] in terms of Ω_{NBS} which have been made on a regular basis since August 1983. Of more importance, the quantized Hall resistance measurements [13] can be used with (1) to calculate an SI value for the quantized Hall resistance. This calculation and other fundamental constant calculations using (1) are described in another paper [14]. Suffice it to say here that on the mean date May 17, 1988

$$R_H = 25\,812.8\,[1 + (1.874 \pm 0.012) \times 10^{-6}]\,\Omega_{NBS}$$

which, in combination with (1), yields

$$R_H = 25\,812.8\,[1 + (0.280 \pm 0.024) \times 10^{-6}]\,\Omega$$

$$= (25\,812.807\,23 \pm 0.000\,61)\,\Omega$$

where the correlation between the Ω_{NBS} and R_H measurements due to the fact that some uncertainties are common to both has been taken into account.

Noting that F_{NBS} was adjusted to coincide with the SI value following the last NBS farad realization [2], we find that F_{NBS} has changed by (0.143 ± 0.016) ppm and the average rate of change of F_{NBS} is $(0.009\,9 \pm 0.001\,1)$ ppm/year.

ACKNOWLEDGMENT

The authors wish to thank G. R. Jones for his assistance in bringing the calculable capacitor into operation, G. W. Small for his help in refurbishing the calculable capacitor and for many valuable discussions; L. H. Lee for his continuous help throughout the project; and B. N. Taylor for his continued support and encouragement.

REFERENCES

[1] P. Giacomo, "News from the BIPM," *Metrologia*, vol. 24, pp. 45–51, 1987.

[2] R. D. Cutkosky, "New NBS measurements of the absolute farad and ohm," *IEEE Trans. Instrum. Meas.*, vol. IM-23, pp. 305–309, Dec. 1974.

[3] R. D. Cutkosky, "Techniques for comparing four terminal pair admittance standards," *J. Res. Nat. Bur. Stand.*, vol. 74C, pp. 63–78, July–Dec. 1970.

[4] R. D. Cutkosky and L. H. Lee, "Improved ten picofarad fused silica dielectric capacitor," *J. Res. Nat. Bur. Stand.*, vol. 69C, pp. 173–179, July–Sept. 1965.

[5] R. J. Haddad, "A resistor calculable from dc to $\omega = 10^5$ rad/s," M.S. thesis, School Eng. Appl. Sci., George Washington Univ., Apr. 1969.

[6] R. D. Cutkosky and J. Q. Shields, "The precision measurement of transformer ratios," *IRE Trans. Instrum.*, vol. I-9, pp. 243–250, Sept. 1960.

[7] J. Q. Shields, "Measurement of four pair admittances with two pair bridges," *IEEE Trans. Instrum. Meas.*, vol. IM-23, pp. 348–352, Dec. 1974.

[8] J. Q. Shields, "Voltage dependence of precision air capacitors," *J. Res. Nat. Bur. Stand.*, vol. 69C, pp. 265–274, Oct.–Dec. 1965.

[9] J. Q. Shields, "Absolute measurement of loss angle using a toroidal cross capacitor," *IEEE Trans. Instrum. Meas.*, vol. IM-27, pp. 464–466, Dec. 1978.

[10] R. V. Hamon, "A 100 Ω build up resistor for the calibration of standard resistors," *J. Sci. Instrum.*, vol. 31, pp. 450–453, Dec. 1954.

[11] K. P. Baker and R. F. Dziuba, "Automated NBS 1 Ω measurement system," *IEEE Trans. Instrum. Meas.*, vol. IM-32, pp. 154–158, Mar. 1983.

[12] H. P. Layer, "A portable iodine stabilized helium–neon laser," *IEEE Trans. Instrum. Meas.*, vol. IM-29, pp. 358–361, Dec. 1980.

[13] M. E. Cage, R. F. Dziuba, C. T. Van Degrift, and D. Y. Yu, "Determination of the time dependence of Ω_{NBS} using the quantized Hall resistance," pp. 263–265, this issue.

[14] M. E. Cage *et al.*, "NBS determination of the fine structure constant and of the quantized Hall resistance and Josephson frequency to voltage quotient in SI units," pp. 284–289, this issue.

T-IM/36/2/2/14038

Monitoring the U.S. Legal Unit of Resistance via the Quantum Hall Effect

M.E. Cage
R.F. Dziuba
B.F. Field
T.E. Kiess
C.T. Van Degrift

Reprinted from
IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT
Vol. IM-36, No. 2, June 1987

Monitoring the U.S. Legal Unit of Resistance via the Quantum Hall Effect

MARVIN E. CAGE, RONALD F. DZIUBA, MEMBER, IEEE, BRUCE F. FIELD, MEMBER, IEEE, THOMAS E. KIESS, AND CRAIG T. VAN DEGRIFT

Abstract—**The quantum Hall effect is being used to monitor the resistances of the five 1-Ω Thomas-type resistors which define the U.S. legal unit of resistance, the ohm maintained at the National Bureau of Standards (Ω_{NBS}). Typically, the total one-standard-deviation (1σ) accuracy for the transfer between three different GaAs quantum Hall devices and the five 1-Ω resistors is ±0.05 ppm. Measurements to date provide the first direct evidence that the value of Ω_{NBS} is decreasing by about (0.05 ± 0.02) ppm per year.**

I. Introduction

THE quantum Hall resistance R_H of a two-dimensional electron gas is, under certain special conditions, quantized in units of h/e^2 [1]:

$$R_H(i) = \frac{V_H(i)}{I} = \frac{h}{e^2 i} \approx \frac{25\ 812.80}{i}\ \Omega \qquad (1)$$

where V_H is the Hall voltage across the sample, h the Planck constant, e the elementary charge, and i is an integer quantum number. Equation (1) is written in absolute or International System (SI) units. It can be expressed in as-maintained laboratory units by replacing R_H and Ω by the quantized Hall resistance and ohm at the National Bureau of Standards ($(R_H)_{NBS}$ and Ω_{NBS}, respectively), where Ω_{NBS} is the United States legal unit of resistance, and is defined in terms of the mean resistance of five 1-Ω Thomas-type resistors maintained at NBS. One measures the value of R_H in laboratory units, and then expresses it in SI units once the ratio (Ω_{NBS}/Ω) has been determined. This ratio can be obtained in two ways: either from the calculable capacitor experiment [2], [3], or by combining the low-field gyromagnetic ratio of the proton, γ_p', and $2e/h$ via the Josephson effect [4]–[6]. Both approaches are currently being pursued at the NBS. (The value of Ω_{NBS} can, however, be expressed in terms of the SI resistance unit as realized at the National Measurement Laboratory (NML), Australia because one quantum Hall device and three 1-Ω resistors have been used as transfer standards between the two laboratories [7], [8]. Those measurements imply that the Ω_{NBS} was (1.341 ± 0.062) ppm smaller than the realization of the International System ohm (SI Ω) at NML in October, 1985 [8], and that h/e^2 is (0.40 ± 0.08) ppm larger than the 25 812.80 Ω nominal value. These results will be described in more detail elsewhere.)

Manuscript received June 23, 1986. This work was supported in part by the United States Office of Naval Research, the Calibration Coordination Group of the Department of Defense, and the Army Research Office.

M. E. Cage, R. F. Dziuba, B. F. Field, and T. E. Kiess are with the Electricity Division, National Bureau of Standards, Gaithersburg, MD 20899.

C. T. Van Degrift is with the Temperature and Pressure Division, National Bureau of Standards, Gaithersburg, MD 20899.

IEEE Log Number 8714038.

II. Experimental Details

Three high-quality quantum Hall effect devices are being used to monitor Ω_{NBS}. Each is a GaAs–$Al_xGa_{1-x}As$ heterostructure ($x = 0.29$) grown using molecular beam epitaxy by A. C. Gossard at AT&T Bell Laboratories, and then prepared into Hall bar geometries and mounted by D. C. Tsui at Princeton University. The devices are ~4.6 mm long and ~0.4 mm wide, and have three sets of Hall potential probes, with two sets symmetrically displaced ±1.0 mm along the channel from the center set. Two different sets of Hall probe pairs are used for each of the three devices, the center set, and an off-center set. The devices designated GaAs(7) and GaAs(8) have ~100 000 $cm^2/(V \cdot s)$ zero magnetic field mobilities at 4.2 K, while the GaAs(9) device has a mobility of ~75 000 $cm^2/(V \cdot s)$. (GaAs(9) was the device used in the NML transfer.) The epitaxially grown film thicknesses and doping density profiles of these devices are optimized for the $i = 4$ quantum Hall step, where $R_H(4) \approx 6{,}453.20\ \Omega$. The centers of this step occur at ~5.6 – 6.0 T for the three devices; the corresponding electron densities are ~$5.4 - 5.8 \times 10^{11}\ cm^{-2}$.

A set of wire-wound reference resistors have been constructed to have values R_R within a few parts-per-million of the value of $R_H(4)$. They are hermetically sealed in silicone fluid-filled containers and placed in specially constructed, temperature-regulated air bath enclosures. The air temperature is controlled to within ±0.002°C of a nominal temperature of ~28°C.

III. Quantum Hall Effect Measurements

Two different measurement systems are used to compare the quantum Hall voltages V_H with the voltage drops V_R across the wire-wound reference resistors: a manually-operated potentiometric comparator [9] and an automated and guarded resistance bridge [10]. Figs. 1 and 2 show simplified schematic diagrams of these two systems. The potentiometric system has a ±0.011-ppm random, or type

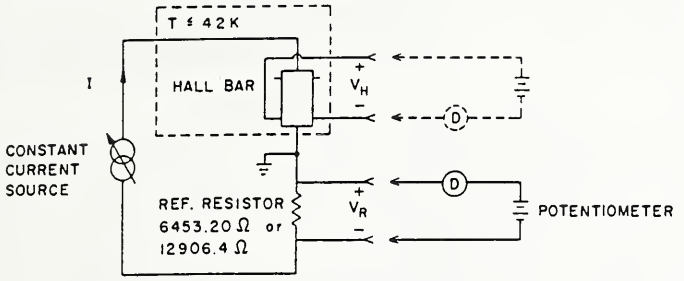


Fig. 1. A simplified schematic of the manually operated potentiometric comparator measurement system, where *D* is an electronic detector.

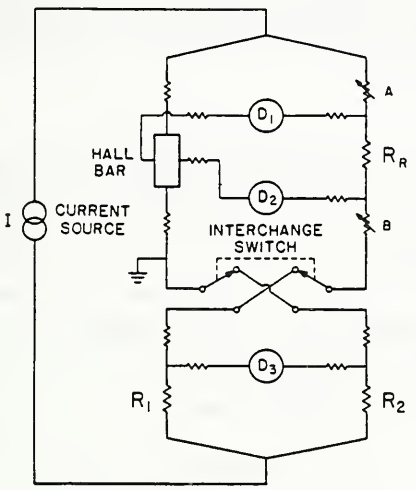


Fig. 2. A simplified schematic of the automated quantum Hall resistance bridge, which uses three electronic detectors.

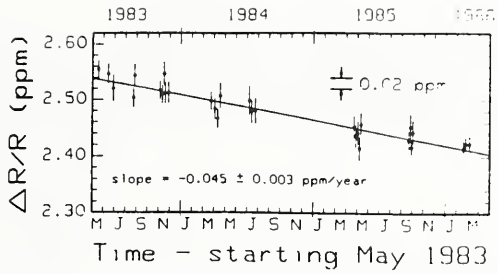


Fig. 3. Relative comparisons as a function of time of the resistance of the $i = 4$ steps of three different quantum Hall devices with that of a nominal 6,453.20-Ω wire-wound reference resistor ($\Delta R/R = (V_H - V_R)/V_R$. The value of this particular resistor is increasing by (0.045 ± 0.003) ppm per year.)

A, uncertainty after a 1-h measurement period for a device current of 25.5 μA; the random uncertainty of the resistance bridge is typically ±0.006 ppm for a comparable measurement period at 25.5 μA.

Both measurement systems have been used to compare the values of R_H with those of R_R for the two Hall probe sets on the three GaAs devices for both magnetic field directions. To be useful as a resistance standard, the Hall steps must be flat within the experimental resolution. All twelve quantum Hall steps are flat to within ±0.01 ppm over a magnetic field range that is ~2 percent of the central field values when the devices are cooled to ~1.2 K. Fig. 3 of [9] shows a digital mapping of one of these $i = 4$ steps for GaAs(7).

IV. Quantum Hall Effect Systematic Measurement Uncertainties

In addition to the random measurement uncertainties, there are systematic corrections with associated systematic, or type *B*, uncertainties. One such correction is due to a measurement system offset error in which the value of the device under test depends on whether it is measured in the R_H position or the R_R position of the measurement circuit. The correction is determined by replacing the quantum Hall device with a 6,453.20-Ω reference resistor and then using either of the measurement systems to intercompare the resistor pairs. This offset error has been observed on both of the NBS measurement systems, as well as on the NML automated potentiometric comparator system which uses a different detector [7]. The source of the error is not understood, but it is probably associated with the electronic detectors. It does not seem to be due to dc leakage currents because the leakage resistances are $>10^{12}$ Ω for both NBS measurement systems. It is also independent of the detector input current. The position-dependent measurement offset error is sometimes as large as (0.025 ± 0.016) ppm for the potentiometric system and (0.019 ± 0.011) ppm for the resistance bridge. The resistor interchange procedure to determine this offset correction is done each day that R_H is measured.

There is an uncertainty in calibrating the gains and linearities of the electronic detector–digital voltmeter pairs. Both the detector–digital voltmeter pairs used at NBS and at NML [7] *appear* to have gains which vary by a few tenths of a percent over the input voltage range. This nonlinearity is due to a 1-μV dead band of the digital voltmeters at zero volts. This problem can be avoided by either using digital voltmeters which have no dead band or by increasing the output voltages of the detectors. The voltmeters used in calibrating the detector gains must, of course, be the same ones that are used in the quantum Hall resistance measurements. There still remains, however, the problem of stability; the gains of the detector–digital voltmeter pairs vary by ~0.1 percent during a day if the room temperature is controlled to ~±1°C. This instability typically contributes a ±0.003-ppm uncertainty to the measurements for the potentiometric comparator system, and a ±0.015-ppm uncertainty for the resistance bridge.

There is also a correction for the temperature dependence [11] of R_H for each Hall probe set of every quantum Hall device for both magnetic field directions. The corrections to the values of R_H for some devices are found to vary linearly with the minimum values of the voltage drop along the device, $V_x^{\min}$. These corrections can be quite significant. Reference [11, Figs. 3, 4] demonstrates that these linear relationships hold over at least four orders of magnitude change in $V_x^{\min}$ for GaAs(7) and GaAs(8).

Every quantum Hall device is unique; the effects reported in [11] are not always observed, nor are they necessarily the *only* temperature-dependent effects. For example, GaAs(9) has a nonlinear dependence on $V_x^{\min}$ similar to that for one Hall probe set of GaAs(7) [11]. This nonlinearity is probably due to the asymmetry of the

TABLE I
ESTIMATED 1σ (68 PERCENT CONFIDENCE LEVEL) UNCERTAINTIES FOR THE QUANTUM HALL RESISTANCE MEASUREMENTS, $R_H \rightarrow R_K$

Sources of Uncertainty	Uncertainties (ppm)	
	Potentiometric System	Bridge System
Random Measurement Uncertainty	0.011	0.006
Measurement Offset Error	0.016	0.011
Detector Gains and Linearity	0.003	0.015
Temperature Dependence Corrections	$\lesssim$ 0.002	$\lesssim$ 0.002
Current Dependence Corrections	$<$ 0.001	$<$ 0.001
ROOT-SUM-SQUARE TOTAL (ppm)	0.020	0.020

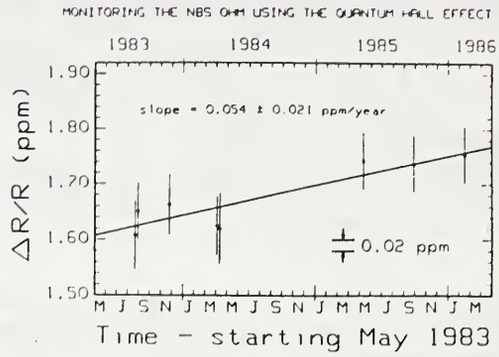


Fig. 4. Monitoring as a function of time the value of $R_H(4)$ expressed as a difference in ppm from a reference value of 6,453.20 Ω_{NBS}. (These data indicate that the U.S. legal ohm, Ω_{NBS}, is decreasing by $\sim (0.05 \pm 0.02)$ ppm per year.)

TABLE II
ESTIMATED 1σ (68-PERCENT CONFIDENCE LEVEL) UNCERTAINTIES FOR THE STEP-DOWNS TO THE U.S. LEGAL OHM, $R_H \rightarrow \Omega_{NBS}$

Sources of Uncertainty	Uncertainties (ppm)
QHE Resistance Measurement Uncertainty	0.020
Resistance Scaling Uncertainty	0.044
Self-Heating of Reference Resistors	0.020
ROOT-SUM-SQUARE TOTAL (ppm)	0.052

Hall step with respect to V_x^{min}; thus the value of R_H in this case includes the effect of structure on the side of the step. In another example, the value of R_H is too small at higher temperatures for one Hall probe set of GaAs(9), but then becomes consistently *too large* by $\sim$0.13 ppm over the temperature range 3.4 $-$ 2.5 K before dropping to the "correct" value at 1.2 K. One could thus infer a temperature-independent (but incorrect) value of R_H over the 2.5–3.4-K temperature range. All three GaAs devices have temperature-dependent effects which are completely repeatable over many cool-downs from room temperature. To date the largest correction to R_H, necessary to extrapolate from the 1.2–0-K values, has been (0.026 ± 0.002) ppm.

No current dependence nor current breakdown phenomena [12] were observed for the three GaAs devices for $I \leq 25.5\ \mu$A, so no correction for finite current is required. Table I summarizes the assigned uncertainties; the total root sum square (rss) uncertainty for each measurement system is typically $\pm$0.020 ppm.

V. QUANTUM HALL EFFECT RESULTS

Fig. 3 shows comparisons of $i = 4$ quantum Hall resistances of the three GaAs devices with that of a nominal 6,453.20-Ω reference resistor during a 34-month time period starting in May 1983. These data are independent of the Hall device, the Hall probe set, the magnetic field direction, and the measurement system once the appropriate offset, gain, and temperature-dependent corrections are made. A weighted linear least squares fit, which takes into account the total uncertainty of each measurement, shows that the resistance of this particular reference resistor is increasing at a rate of (0.045 ± 0.003) ppm per year. This unusually small and linear drift rate enables us to continuously monitor the reliability of the two measurement systems.

VI. STEP-DOWNS TO THE NBS OHM

To monitor the NBS ohm, the nominal 6,453.20-Ω reference resistors must be calibrated in terms of the set of five 1-Ω resistors which define Ω_{NBS}. This is done in two stages: the first uses a 6,453.20 to 100-Ω series/parallel Hamon network configuration [13] consisting of eight 800-Ω resistors plus a series-connected 53.2-Ω resistor; the second uses a 100 to 1-Ω Hamon network consisting of ten 10-Ω resistors. Transfers from 6,453.20 to 1 Ω_{NBS} are currently estimated to have an uncertainty of $\pm$0.044 ppm [9].

The current used in the step-downs is 1.25 mA for the 6,453.20-Ω reference resistors, whereas it is 25.5 μA in the quantum Hall effect resistance comparisons. The reference resistors are maintained in constant-temperature air baths, which enhance the self-heating effect in the reference resistors. The self-heating increases the temperature of the silicone fluid at higher currents. This typically produces a $(+0.02 \pm 0.02)$-ppm correction to the value of the 6,453.20-Ω reference resistors in the step-down procedure.

Measurements involving the entire sequence (quantum Hall resistance comparisons with nominal 6,453.20-Ω reference resistors and then step-downs to Ω_{NBS} have been made over a 31-month interval commencing in August 1983. Fig. 4 shows the results of these measurements to date. The total 1σ rss uncertainty is typically $\pm$0.052 ppm for each datum, as indicated in Table II.

The data of Fig. 4 show the first direct evidence that Ω_{NBS} is decreasing with time. A weighted linear least squares fit yields a drift rate of (0.054 ± 0.021) ppm per year, but data must be accumulated over a longer time

span in order to reduce the uncertainty and to verify that the drift is indeed linear.

Acknowledgment

The authors thank A. C. Gossard of AT&T Bell Laboratories, Murray Hill, NJ, for growing the GaAs devices; D. C. Tsui of Princeton University, Princeton, NJ, for preparing and selecting the devices; G. Marullo Reedtz, a Guest Scientist at the National Bureau of Standards from the Istituto Elettrotecnico Nazionale, Turin, Italy for assisting in the reference resistor self-heating measurements; and B. N. Taylor at the National Bureau of Standards for his continued support and encouragement.

References

[1] K. von Klitzing, G. Dorda, and M. Pepper, "New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance," *Phys. Rev. Lett.*, vol. 45, pp. 494–497, Aug. 1980.

[2] A. M. Thompson, "An absolute determination of resistance based on a calculable standard of capacitance," *Metrologia*, vol. 4, pp. 1–7, Jan. 1968.

[3] R. D. Cutkosky, "New NBS measurements of the absolute farad and ohm," *IEEE Trans. Instrum Meas.*, vol. IM-23, pp. 305–309, Dec. 1974.

[4] D. C. Tsui, A. C. Gossard, B. F. Field, M. E. Cage, and R. F. Dziuba, "Determination of the fine-structure constant using GaAs/$Al_xGa_{1-x}As$ heterostructures," *Phys. Rev. Lett.*, vol. 48, pp. 3–6, Jan. 1982.

[5] B. N. Taylor, "Impact of quantized Hall resistance on SI electrical units and fundamental constants," *Metrologia*, vol. 21, pp. 37–39, Jan. 1985.

[6] B. N. Taylor, "New results from previously reported NBS fundamental constant determinations," *J. Res. Nat. Bur. Stand.*, vol. 90, pp. 91–94, Mar.–Apr. 1985.

[7] B. W. Ricketts and M. E. Cage, "Quantized Hall resistance measurements at the National Measurement Laboratory, Australia," pp. 245–248, this issue.

[8] G. W. Small, "Revised value and estimate of uncertainties in NML's realization of the ohm," presented at Com. Consult. d'Electricite, Com. Int. Poids Meas., Trav. 17e Sess., Sept. 1986.

[9] M. E. Cage, R. F. Dziuba, and B. F. Field, "A test of the quantum Hall effect as a resistance standard," *IEEE Trans. Instrum. Meas.*, vol. IM-34, pp. 301–303, June 1985.

[10] B. F. Field, "A high-accuracy automated resistance bridge for measuring quantum Hall devices," *IEEE Trans. Instrum. Meas.*, vol. IM-34, pp. 320–322, June 1985.

[11] M. E. Cage, B. F. Field, R. F. Dziuba, S. M. Girvin, D. C. Tsui, and A. C. Gossard, "Temperature dependence of the quantum Hall resistance," *Phys. Rev. B.*, vol. 30, pp. 2286–2288, Aug. 1984.

[12] M. E. Cage *et al.*, "Dissipation and dynamic nonlinear behavior in the quantum Hall regime," *Phys. Rev. Lett.*, vol. 51, pp. 1374–1377, Oct. 1983.

[13] B. V. Hamon, "A 1–100 ohm build-up resistor for the calibration of standard resistors," *J. Sci. Instrum.*, vol. 31, pp. 450–453, Dec. 1954.

NBS Determination of the Fine-Structure Constant and of the Quantized Hall Resistance and Josephson Frequency-to-Voltage Quotient in SI Units

M. E. Cage
Ronald F. Dziuba
Randolph E. Elmquist
Bruce F. Field
George R. Jones, Jr.
P. Thomas Olsen
William D. Phillips
J. Q. Shields
R. L. Steiner
Barry N. Taylor
Edwin R. Williams

Reprinted from
IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT
Vol. 38, No. 2, April 1989

NBS Determination of the Fine-Structure Constant, and of the Quantized Hall Resistance and Josephson Frequency-to-Voltage Quotient in SI Units

M. E. CAGE, RONALD F. DZIUBA, MEMBER, IEEE, RANDOLPH E. ELMQUIST, BRUCE F. FIELD, MEMBER, IEEE, GEORGE R. JONES, JR., P. THOMAS OLSEN, MEMBER, IEEE, WILLIAM D. PHILLIPS, J. Q. SHIELDS, R. L. STEINER, BARRY N. TAYLOR, FELLOW, IEEE, AND EDWIN R. WILLIAMS, MEMBER, IEEE

Abstract—**Results from NBS experiments to realize the ohm and the watt, to determine the proton gyromagnetic ratio by the low field method, to determine the time dependence of the NBS representation of the ohm using the quantum Hall effect, and to maintain the NBS representation of the volt using the Josephson effect, are appropriately combined to obtain an accurate value of the fine-structure constant and of the quantized Hall resistance in SI units, and values in SI units of the Josephson frequency-to-voltage quotient, Planck constant, and elementary charge.**

I. Introduction

TO COMPARE critically predictions of quantum electrodynamics (QED) with experimental results often requires an accurate value of the fine-structure constant α which is independent of QED. For example, a highly accurate, QED-independent value of α is necessary for obtaining a theoretical value of the electron magnetic moment anomaly a_e from QED for comparison with the experimentally determined value of a_e [1].

Under the assumption that the quantized Hall resistance R_H as measured using a quantum Hall effect (QHE) device is equal to h/e^2 (h is the Planck constant and e is the elementary charge), there are at present two virtually independent ways of obtaining a reliable, QED-independent value of α from high precision experiments in the field of electrical metrology [2]. The first is to use the QHE and a direct realization of the ohm by means of a calculable capacitor to determine $R_H = h/e^2$ in SI units, that is, in ohms:

$$\alpha = \mu_0 c/2R_H \qquad (1)$$

where $\mu_0 = 4\pi \times 10^{-7}$ N/A^2 exactly, is the permeability of vacuum and $c = 299\,972\,458$ m/s exactly, is the speed of light in vacuum.

Manuscript received June 10, 1988.

The authors are with the National Bureau of Standards, Electricity Division, Gaithersburg, MD 20899.

IEEE Log number 8824970.

The second involves the use of the relation

$$\alpha^3 = 2\mu_0 R_\infty \gamma_p'/(\mu_p'/\mu_B)\, R_H E_J \qquad (2)$$

where R_∞ is the infinite mass Rydberg constant, γ_p' is the proton gyromagnetic ratio (the prime indicates a spherical, pure H_2O proton nuclear magnetic resonance or NMR sample at 25°C), μ_p'/μ_B is the magnetic moment of the proton in units of the Bohr magneton, $R_H = h/e^2$ is again the quantized Hall resistance, and E_J is the Josephson frequency-to-voltage quotient and is assumed equal to $2e/h$.

Although (1) and (2) both involve R_H, they do so in very different ways. Equation (1) requires R_H in SI units, whereas (2) has the important property that it is independent of the electrical units used, that is, it holds if γ_p', R_H, and E_J are measured in SI units, or if γ_p' is determined by the weak or low-field method [3] and it, R_H, and E_J are measured in terms of the EMF of an arbitrary battery and the resistance of an arbitrary resistor. Because of this property, method two does not require the direct realization of an SI electrical unit. Since the two methods involve radically different experiments but can yield α values of comparable uncertainty, each can serve as a critical check on the other.

Equation (1) is also noteworthy. Since μ_0 and c are exact quantities, it implies that if α is known with a certain fractional uncertainty, R_H will be known in SI units (i.e., ohms) with the same fractional uncertainty. This is significant because the Comité Consultatif d'Électricité (CCE) of the Comité International des Poids et Mesures (CIPM) has decided to recommend that starting on January 1, 1990, practical or laboratory representations of the ohm be based on the QHE using an adopted value of R_H in agreement with the SI value. This value is to be selected by the CCE at its September 1988 meeting and is to be used by all those laboratories which choose to base their ohm representation on the QHE [4].

Similarly, the CCE has decided to recommend that starting on January 1, 1990, all those laboratories that base

their practical or laboratory representation of the volt on the Josephson effect use a new value of the Josephson frequency-to-voltage quotient $E_J = 2e/h$ in agreement with the SI value. This value, which is to replace the one suggested by the CCE in 1972 and used by most countries, is also to be selected by the CCE at its September 1988 meeting [4].

II. NBS Measurements

We have been carrying out a number of experiments at NBS that are now yielding results and bear on the issue of a reliable value of α, and of R_H and E_J in SI units. These are:

1) maintenance of the NBS representation of the volt V_{NBS} using Josephson 1-V arrays [5];
2) determination of the NBS representation of the ohm $\underline{\Omega}_{NBS}$ (which is based on the mean resistance of five Thomas-type wire-wound resistors) in ohms using the NBS calculable capacitor [6];
3) determination of the time dependence of the NBS representation of the ohm $\underline{\Omega}_{NBS}$ through comparisons with the quantized Hall resistance R_H [7];
4) determination of the NBS electrical representation of the watt $W_{NBS} = V_{NBS}^2/\underline{\Omega}_{NBS}$ in watts by equating electrical and mechanical power using a moving-coil force balance [8];
5) determination of γ_p' by the low field method, γ_p'(low), in terms of the NBS representation of the ampere $A_{NBS} = V_{NBS}/\underline{\Omega}_{NBS}$ using a specially constructed 2.1-m long, single-layer, precision solenoid [9].

Two remarks about these determinations are in order. First, to minimize the contribution of uncertainties associated with laboratory electrical standards to the final uncertainties of the experiments, a great deal of effort is devoted to the problem of standards calibration and monitoring. The goal is to ensure that $\underline{\Omega}_{NBS}$ and the reference resistors used in the γ_p' (low) and W_{NBS} experiments are well known in terms of R_H and that the Zener voltage references used in these two experiments are well known in terms of E_J. Because of the need for close ties between these artifact standards and the invariant reference resistance and voltage provided by the quantum Hall and Josephson effects, and because the QHE and Josephson effect apparatus are located in a building 1.5 km from that in which the γ_p' (low) and W_{NBS} apparatus are located, this aspect of the experiments is not at all trivial. It requires the coordination of a number of different measurement systems and the collaboration and cooperation of many individuals. We are fortunate at NBS to have all of these efforts in one group and believe that the complexity and diversity of these five experiments precludes them from being undertaken at a single institution other than a national standards laboratory.

Second, the values of R_H we shall actually use to evaluate (1) and (2) are based on the same measurements of R_H in terms of $\underline{\Omega}_{NBS}$. Consequently, the two resulting values of α are not completely independent and the small correlation between their uncertainties should be considered when comparing them or when combining them to obtain a single "best" value of α. Similarly, in principle the value of E_J obtained from the determination of W_{NBS} is correlated with both values of α through $\underline{\Omega}_{NBS}$ or R_H, and in the case of the value of α obtained from (2), through V_{NBS} as well. These correlations (and others) are all taken into account in the calculations of Section III as appropriate.

III. Data, Calculations, and Results

A. Data

Since July 1, 1972 [10], the NBS representation of the volt V_{NBS} has been based on the Josephson effect via the relation

$$V_{NBS} = \nu_{J\text{-}NBS}/E_J \tag{3a}$$

with the following exact adopted value for the Josephson frequency:

$$\nu_{J\text{-}NBS} = 483\ 593.420\ \text{GHz}. \tag{3b}$$

Beginning February 10, 1987 [5], Josephson arrays have been used to implement this definition of V_{NBS} with an inherent one standard deviation uncertainty of 0.0089 ppm, where ppm = part-per-million or 1 part in 10^6. (Throughout, all uncertainties are meant to correspond to one standard deviation estimates.)

Based on a realization of the ohm using the NBS calculable capacitor having a mean date of May 17, 1988, it was found that [6]

$$\underline{\Omega}_{NBS} = [1 - (1.594 \pm 0.022)\ \text{ppm}]\ \Omega. \tag{4}$$

Since August 1983, $\underline{\Omega}_{NBS}$ has been monitored in terms of the quantized Hall resistance R_H via the quantum Hall effect. If we write in analogy with (3a)

$$\underline{\Omega}_{NBS} = R_H/r_{K\text{-}NBS} \tag{5}$$

where the dimensionless quantity $r_{K\text{-}NBS}$ is equal to the numerical value of R_H expressed in $\underline{\Omega}_{NBS}$, then the result of the monitoring may be written as [7]

$$\begin{aligned} r_{K\text{-}NBS} = 25\ 812.8\ [1 &+ (1.842\ 0 \pm 0.011\ 6)\ \text{ppm} \\ &+ (0.052\ 9 \pm 0.004\ 0) \\ &\times (t - 0.778\ 5)\ \text{ppm/yr}] \end{aligned} \tag{6}$$

where t is measured in years from January 1, 1987 and the intercept and slope uncertainties are uncorrelated. (The quantity $r_{K\text{-}NBS}$, where the subscript K stands for von Klitzing, the discoverer of the QHE, is time dependent because $\underline{\Omega}_{NBS}$ is based on the mean resistance of five time-varying artifact resistance standards. This is in contrast to V_{NBS} which is based directly on the Josephson effect rather than on a group of electrochemical cells.)

Measurements of $W_{NBS} = V_{NBS}^2/\underline{\Omega}_{NBS}$ using the NBS moving-coil force balance having a mean date of May 15,

1988 yielded [8]:

$$W_{NBS} = [1 - (16.69 \pm 1.33)\ \text{ppm}]\ \text{W}. \quad (7)$$

Measurements of $\gamma'_p(\text{low})$ in terms of $T_{NBS} \propto A_{NBS} = V_{NBS}/\Omega_{NBS}$ with a mean date of April 3, 1988 gave the result [9]:

$$\gamma'_p(\text{low}) = 2.675\ 133\ 76(29) \times 10^8\ \text{s}^{-1}\ T_{NBS}^{-1}\ (0.11\ \text{ppm}). \quad (8)$$

(Here T_{NBS}, the NBS representation of the tesla, is directly proportional to A_{NBS} since it is based on a precision solenoid of known dimensions carrying a current known in terms of A_{NBS}.)

B. Calculations

For numerical convenience, here and throughout the remainder of this paper we consider the inverse fine-structure constant $\alpha^{-1} \simeq 137 = 1/\alpha$ rather than α. For ease of comparison, we present results in the form $\alpha^{-1} = \sigma^0_{-1}[1 + (\Delta \pm \epsilon)\ \text{ppm}]$ which is equivalent to $R_H = R_0[1 + (\Delta \pm \epsilon)\ \text{ppm}]$, where $\alpha_0^{-1} = 2R_0/\mu_0 c = 137.035\ 959\ 5 \cdots$ exactly and $R_0 = 25\ 812.8\ \Omega$ exactly. Similarly, we give $E_J = 2e/h$ in the form $E_J = E_0[1 + (\Delta' \pm \epsilon')\ \text{ppm}]$ with $E_0 = 483\ 594\ \text{GHz/V}$ exactly, for ease of comparison with the value suggested by the CCE in 1972 which numerically is $483\ 594 \times 10^9$ [11].

With the aid of (5), (1) may be written as

$$\alpha^{-1} = 2R_H/\mu_0 c = (2/\mu_0 c)\ r_{K\text{-}NBS}\Omega_{NBS} \quad (9)$$

where it is assumed that the values for $r_{K\text{-}NBS}$ and Ω_{NBS} correspond to the same time t. Using the value of $r_{K\text{-}NBS}$ given by (6) on the mean date of the NBS ohm realization, and the value of Ω_{NBS} given in (4), we obtain from (9)

$$\alpha^{-1} = \alpha_0^{-1}[1 + (0.280 \pm 0.024)\ \text{ppm}] \quad (10a)$$

$$= 137.035\ 997\ 9(32) \quad (10b)$$

$$R_H = 25\ 812.807\ 23(61)\ \Omega. \quad (10c)$$

In a similar fashion, with the aid of (3a) and (5), (2) may be written as

$$\alpha^{-1} = \left\{(\mu'_p/\mu_B)\ r_{K\text{-}NBS}\nu_{J\text{-}NBS}/2\mu_0 R_\infty \gamma'_p(\text{low})\right\}^{1/3} \quad (11)$$

where as usual { } indicates numerical value only. Using the value of $r_{K\text{-}NBS}$ given by (6) on the mean date of the NBS $\gamma'_p(\text{low})$ measurements, the value of $\nu_{J\text{-}NBS}$ given in (3b), $\gamma'_p(\text{low})$ as given in (8), the 1986 CODATA recommended value for μ'_p/μ_B (fractional uncertainty = 0.011 ppm) [3], and $R_\infty = 10\ 973\ 731.573(4)\ \text{m}^{-1}$ (fractional uncertainty = 0.0004 ppm) [12], a more up-to-date and accurate value than that of CODATA, (11) yields

$$\alpha^{-1} = \alpha_0^{-1}[1 + (0.178 \pm 0.037)\ \text{ppm}] \quad (12a)$$

$$= 137.035\ 984\ 0(51) \quad (12b)$$

$$R_H = 25\ 812.804\ 60(95)\ \Omega. \quad (12c)$$

This result does not agree as well as one would like with that obtained from the realization of the ohm as given in (10); the two differ by (0.102 ± 0.043) ppm or 2.37 combined standard deviations. All relevant calculations and corrections are being reviewed in an attempt to understand the significance of this difference.

The values of α^{-1} given in (10) and (12) involve the assumption that R_H is a measure of h/e^2 and thus of α^{-1} ((12) also requires the assumption $2e/h = E_J$). A value of α^{-1} independent of this assumption (but still requiring the assumption $2e/h = E_J$) can be obtained from our data via the following relation derived by using (1) to eliminate R_H from (2):

$$(\alpha^{-1})^2 = c(\mu'_p/\mu_B)\ E_J/4R_\infty\gamma'_p \quad (13a)$$

which may be written as

$$\alpha^{-1} = \left\{c(\mu'_p/\mu_B)\ \nu_{J\text{-}NBS}/4R_\infty\Omega_{NBS}\gamma'_p(\text{low})\right\}^{1/2}. \quad (13b)$$

For historical reasons, the value of α^{-1} obtained from (13b) is often referred to as the "Josephson junction value of alpha." Using the result

$$d\Omega_{NBS}/dt = (-0.052\ 9 \pm 0.004\ 0)\ \mu\Omega/\text{yr} \quad (14)$$

implied by (6) to correct the value of Ω_{NBS} given in (4) to the mean date of the $\gamma'_p(\text{low})$ measurements, (3b) for $\nu_{J\text{-}NBS}$, (8) for $\gamma'_p(\text{low})$, the 1986 CODATA value for μ'_p/μ_B, and the above value of R_∞, we find[1]

$$\alpha^{-1} = \alpha_0^{-1}[1 + (0.127 \pm 0.056)\ \text{ppm}] \quad (15a)$$

$$= 137.035\ 977\ 0(77) \quad (15b)$$

$$R_H = 25\ 812.803\ 29(145). \quad (15c)$$

Equation (15) also differs from (10) by 2.37 combined standard deviations, as do (12) and (15) with each other. The reason is that the three values are highly interdependent; any two of (9), (11), and (13b) determine the third. (The correlation coefficient of (10) and (12) is 0.040; of (10) and (15), −0.170 and of (12) and (15), 0.978.)

To obtain E_J from the measurement of W_{NBS} in watts, we first use the relation $W_{NBS} = V^2_{NBS}/\Omega_{NBS}$ and (3a), (5), and (1) to write

$$E_J^2 = \nu^2_{J\text{-}NBS}/\Omega_{NBS}W_{NBS} \quad (16a)$$

$$= \nu^2_{J\text{-}NBS}r_{K\text{-}NBS}/R_H W_{NBS} \quad (16b)$$

$$= 2\nu^2_{J\text{-}NBS}r_{K\text{-}NBS}/\mu_0 c\alpha^{-1}W_{NBS}. \quad (16c)$$

Using (14) to correct the result for Ω_{NBS} given in (4) to the mean date of the W_{NBS} measurements, (3b) for $\nu_{J\text{-}NBS}$, and (7) for W_{NBS}, we find from (16a)

$$E_J = E_0[1 + (7.94 \pm 0.67)\ \text{ppm}] \quad (17a)$$

$$= 483\ 597.84(32)\ \text{GHz/V}. \quad (17b)$$

[1]Taking $d\Omega_{NBS}/dt$ from quantized Hall resistance measurements does not negate the independence of this value of α^{-1} from the assumption $R_H = h/e^2$. The only assumption required is that R_H is time-independent.

Alternatively, we may use (16b) with $r_{K\text{-}NBS}$ from (6) and R_H from (12c). The result is

$$E_J = E_0[1 + (7.99 \pm 0.67)\ \text{ppm}] \quad (18a)$$

$$= 483\ 597.87(32)\ \text{GHz/V}. \quad (18b)$$

If we repeat this calculation using R_H from (15c), we obtain

$$E_J = E_0[1 + (8.02 \pm 0.67)\ \text{ppm}] \quad (19a)$$

$$= 483\ 597.88(32)\ \text{GHz/V}. \quad (19b)$$

The differences among these three values of E_J see ((17)–(19)) are small compared to the uncertainties of the three values, which are completely dominated by the 1.33-ppm uncertainty of W_{NBS}. Hence, the differences among the values of R_H discussed above are not critical to deriving a value of E_J from W_{NBS}.

C. Final Results

To obtain a single best value of α^{-1} (or equivalently, R_H) and of E_J from the NBS data, and subsequently values of h and e from these best values, we use the method of least squares as employed in fundamental constants adjustments [3]. Taking α^{-1} and E_J as the adjustable constants or unknowns, the observational equations for the three measured quantities $\underline{\Omega}_{NBS}$, $\gamma'_p(\text{low})$, and W_{NBS} are

$$\alpha^{-1} = (2/\mu_0 c)\, r_{K\text{-}NBS}\underline{\Omega}_{NBS}$$

$$\alpha^{-3} = \left\{(\mu'_p/\mu_B)\, r_{K\text{-}NBS}\nu_{J\text{-}NBS}/2\mu_0 R_\infty \gamma'_p(\text{low})\right\} \quad (20b)$$

$$\alpha^{-1}E_J^2 = 2\nu_{J\text{-}NBS}^2 r_{K\text{-}NBS}/\mu_0 c W_{NBS} \quad (20c)$$

which follow from (9), (11), and (16c). Then since $R_H = \mu_0 c\alpha^{-1}/2 = h/e^2$ and $E_J = 2e/h^2$

$$h = 4/R_H E_J^2 \quad (21a)$$

$$e = 2/R_H E_J. \quad (21b)$$

It is apparent from (20) and (6) that $r_{K\text{-}NBS}(t = 0)$ and $dr_{K\text{-}NBS}/dt$ (i.e., the intercept and slope of (6)) could be taken as adjustable constants in addition to α^{-1} and E_J. We choose not to for the following reason: because of their comparative simplicity and the extended period over which they have been carried out (i.e., about five years), we believe that at present, our quantized Hall resistance measurements are inherently more reliable than our other measurements (i.e., $\underline{\Omega}_{NBS}$ and $\gamma'_p(\text{low})$), more so perhaps than is indicated by their *a priori* assigned uncertainties. Consequently, we do not believe $r_{K\text{-}NBS}(t = 0)$ or $dr_{K\text{-}NBS}/dt$ should be subject to adjustment. Moreover, because of the structure of (20a)–(20c), taking these two quantities as adjustable constants yields identical values for α^{-1} and E_J. In addition, the adjusted value of $r_{K\text{-}NBS}(t = 0)$ would be only 0.0034 ppm smaller than the input value from (6) (0.3 standard deviations of the latter); and the adjusted value of $dr_{K\text{-}NBS}/dt$ would be only 0.0004 ppm/yr smaller than the input value from (6) (0.1 standard deviations of the latter).

To solve (20a)–(20c), we use: 1) (6) to calculate $r_{K\text{-}NBS}$ at the mean dates of the $\underline{\Omega}_{NBS}$, $\gamma'_p(\text{low})$, and W_{NBS} measurements as before, 2) the values given above for these quantities as well as for $\nu_{J\text{-}NBS}$, μ'_p/μ_B, and R_∞, and 3) the correlated error least squares approach [13]. This latter method is employed to take into account the approximate 0.012-ppm uncertainty in $r_{K\text{-}NBS}$ common to all three equations; some comparatively small systematic uncertainties associated with resistance measurements common to the R_H, $\underline{\Omega}_{NBS}$, $\gamma'_p(\text{low})$, and W_{NBS} experiments; and for completeness the 0.0089-ppm uncertainty in implementing the definition of V_{NBS} common to (20b) and (20c). (All of these were also taken into account in the calculations of the previous section.)

The difference between the values of α^{-1} from $\underline{\Omega}_{NBS}$ and $\gamma'_p(\text{low})$ indicated above (see (10) and (12)) is apparent in the least-squares treatment; the statistic "chi square" (χ^2) for the adjustment is 5.60 compared with its expected value of $f = 1$ (f is the number of degrees of freedom); the Birge ratio is $R_B = [\chi^2/f]^{1/2} = 2.37$ compared with its expected value of 1. The probability that this value of χ^2 has occured by chance is about 1.8 percent. The resulting adjusted value of α^{-1} is $\alpha_0^{-1}[1 + (0.252 \pm 0.020)\ \text{ppm}]$.

The above value of χ^2 arises entirely from the difference between the two values of α^{-1}; W_{NBS} does not contribute to χ^2 or to the determination of α^{-1} since $\alpha^{-1}E_J^2 \propto 1/W_{NBS}$ (see (20c)) and (20a) and (20b) do not involve E_J. Thus a one variable (i.e., α^{-1}) least squares adjustment involving just $\underline{\Omega}_{NBS}$ and $\gamma'_p(\text{low})$ ((20a) and (20b)) yields the same adjusted value for α^{-1} and also has $\chi^2 = 5.60$ for $f = 1$ with $R_B = 2.37$. If $R_B = 2.37$ is used as a multiplicative scale factor for the uncertainties of the one variable adjustment data so that χ^2 has its expected value of 1,[3] and if the two variable (i.e., α^{-1} and E_J) adjustment is then repeated with these expanded uncertainties for (20a) and (20b) with the result that χ^2 for this adjustment also has its expected value of 1, we find

$$\alpha^{-1} = \alpha_0^{-1}[1 + (0.252 \pm 0.048)\ \text{ppm}] \quad (22a)$$

$$= 137.035\ 994\ 0(65) \quad (22b)$$

$$R_H = 25\ 812.806\ 50(123)\ \Omega \quad (22c)$$

$$E_J = E_0[1 + (7.96 \pm 0.67)\ \text{ppm}] \quad (23a)$$

$$= 483\ 597.85(32)\ \text{GHz/V} \quad (23b)$$

$$h = 6.626\ 070\ 4(88) \times 10^{-34}\ \text{J} \cdot \text{s}\ (1.33\ \text{ppm}) \quad (24)$$

[2] In principle, to calculate properly the uncertainty of h and e, the covariance of α^{-1} and E_J, which is one of the outputs of the adjustment, must be taken into account.

[3] The adjusted value of α^{-1} is unchanged but its uncertainty is increased to 0.048 ppm = 2.37 × 0.020 ppm.

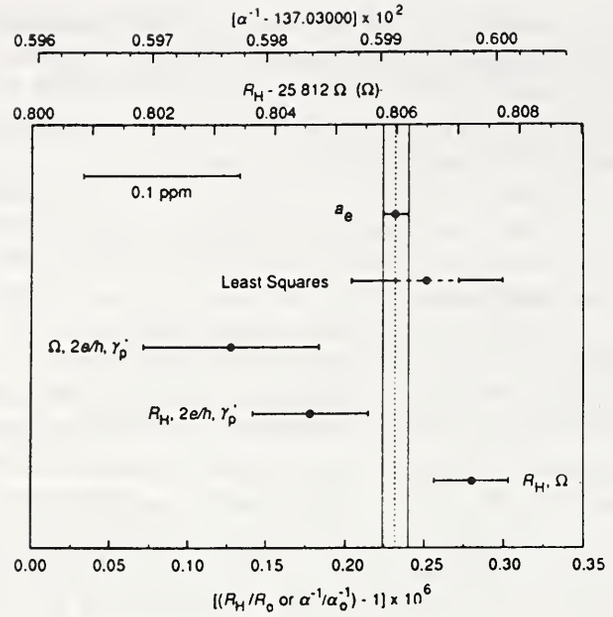


Fig. 1. Comparison of values of R_H and α^{-1} discussed in the text. (R_0 = 25 812.8 Ω, $\alpha_0^{-1} = 2R_0/\mu_0 c$ = 137.035 959 5 · · · .) Starting from the bottom of the figure, the equation number of the value as given in the text is: (10), (12), (15), (22), (26). The least squares value, (22), is shown with and without its uncertainty increased by the scale factor R_B = 2.37; see text. (As a comparison aid, the most accurate value (26) and its uncertainty are indicated by dashed and full lines, respectively, as well as by the usual point and error bars.)

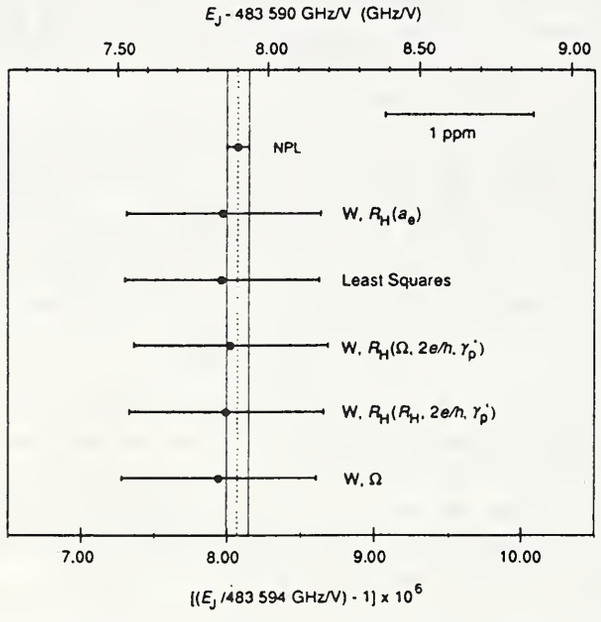


Fig. 2. Comparison of values of E_J discussed in the text. Starting from the bottom of the figure, the equation number of the value as given in the text is: (17)–(19), (23), (27), (28). (As a comparison aid, the most accurate value (28) and its uncertainty are indicated by dashed and full lines, respectively, as well as by the usual point and error bars.)

$$e = 1.602\ 176\ 70(107) \times 10^{-19}\ \mathrm{C}\ (0.67\ \mathrm{ppm}). \quad (25)$$

In Figs. 1 and 2 we graphically compare the principal values of α^{-1} and E_J obtained throughout this paper from our data. Also included in Fig. 1 is the highly accurate but still preliminary QED value of α^{-1} obtained by Kinoshita from the electron magnetic moment anomaly a_e [14]:

$$\alpha^{-1}(a_e) = \alpha_0^{-1}[1 + (0.232\ 1 \pm 0.008\ 0)\ \mathrm{ppm}] \quad (26a)$$

$$= 137.035\ 991\ 4(11) \quad (26b)$$

$$R_H = 25\ 812.805\ 99(21). \quad (26c)$$

If this value of R_H is used to calculate a value of E_J from (16b), the result is

$$E_J = E_0[1 + (7.97 \pm 0.67)\ \mathrm{ppm}] \quad (27a)$$

$$= 483\ 597.85(32)\ \mathrm{GHz/V}. \quad (27b)$$

This value is also included in Fig. 2 as is the most precise result reported to date, that obtained by Kibble *et al.* [15] at the National Physical Laboratory (NPL), U.K., from their version of the moving-coil watt realization experiment:

$$E_J = E_0[1 + (8.070 \pm 0.077)\ \mathrm{ppm}] \quad (28a)$$

$$= 483\ 597.903(37)\ \mathrm{GHz/V}. \quad (28b)$$

It should be borne in mind while examining Figs. 1 and 2, especially Fig. 2, that the NBS values are highly interdependent. Indeed, all the values of E_J except that of NPL are essentially the same value since the uncertainty of W_{NBS} is so overwhelmingly dominant.

IV. Conclusion

We have presented a detailed analysis of the currently available NBS results from ongoing experiments to realize the ohm, to realize the watt, to determine the proton gyromagnetic ratio by the low field method, to determine the time dependence of the NBS representation of the ohm using the quantum Hall effect, and to maintain the NBS representation of the volt using Josephson 1-V arrays. Either of our values for α^{-1} and thus R_H in SI units as derived from (1) and (2) (see (10) and (12)) is more accurate than any QED-independent value presently available. This is even true of our least-squares adjusted value with expanded uncertainty, (22). All three values agree with the highly accurate QED value of α^{-1} from the electron magnetic moment anomaly, (26), since the difference from $\alpha^{-1}(a_e)$ for each is less than two combined standard deviations. However, the two values derived from (1) and (2) differ from each other by somewhat more than two combined standard deviations. Every effort will be made in the near future to understand if this difference is significant.

Our values for E_J, h, and e have rather larger uncertainties than their 1986 CODATA recommended counterparts, although they are quite consistent with them. However, it is expected that when the room-temperature magnetic field-generating solenoid of the NBS watt realization experiment is replaced by a specially constructed superconducting solenoid, the uncertainties of the NBS values will be reduced by between a factor of 15 and 50. This would make them the most accurate values available.

References

[1] T. Kinoshita, "The anomalous magnetic moment of the electron and the quantum electrodynamics determination of the fine-structure constant," *IEEE Trans. Instrum. Meas.*, vol. IM-36, pp. 201-204, June 1987.

[2] B. N. Taylor, "Impact of quantized Hall resistance on SI electrical units and fundamental constants," *Metrologia*, vol. 21, pp. 37-39, 1985.

[3] E. R. Cohen and B. N. Taylor, "The 1986 adjustment of the fundamental physical constants," *Rev. Mod. Phys.*, vol. 59, no. 4, pp. 1121-1148, Oct. 1987.

[4] P. Giacomo, "News from the BIPM," *Metrologia*, vol. 24, pp. 45-51, 1987.

[5] R. L. Steiner and B. F. Field, "Josephson array voltage calibration system: Operational use and verification," *IEEE Trans. Instrum. Meas.*, pp. 296-301, this issue.

[6] J. Q. Shields, R. F. Dziuba, and H. P. Layer, "New realization of the ohm and farad using the NBS calculable capacitor," *IEEE Trans. Instrum. Meas.*, pp. 249-251, this issue.

[7] M. E. Cage, R. F. Dziuba, C. T. Van Degrift, and D. Y. Yu, "Determination of the time dependence of Ω_{NBS} using the quantized Hall resistance," *IEEE Trans. Instrum. Meas.*, pp. 263-269, this issue.

[8] P. T. Olsen, R. E. Elmquist, W. D. Phillips, E. R. Williams, G. R. Jones, Jr., and V. E. Bower, "A measurement of the NBS electrical watt in SI units," *IEEE Trans. Instrum. Meas.*, pp. 238-244, this issue.

[9] E. R. Williams, G. R. Jones, Jr., Ye Sheng, Liu Ruimin, H. Sasaki, P. T. Olsen, W. D. Phillips, and H. P. Layer, "A low field determination of the proton gyromagnetic ratio in H_2O," *IEEE Trans. Instrum. Meas.*, pp. 233-237, this issue.

[10] B. F. Field, T. F. Finnegan, and J. Toots, "Volt maintenance at NBS via $2e/h$: A new definition of the NBS volt," *Metrologia*, vol. 9, pp. 155-166, 1973.

[11] J. Terrien, "News from the Bureau International des Poids et Mesures," *Metrologia*, vol. 9, pp. 40-43, 1973.

[12] P. Zhao, W. L. Lichten, H. Layer, and J. Bergquist, "New value for the Rydberg constant from the hydrogen Balmer-β transition," *Phys. Rev. Lett.*, vol. 58, no. 13, pp. 1293-1295, Mar. 1987; Erratum, no. 23, p. 2506, June 1987.

[13] M. G. Natrella, *Experimental Statistics*, National Bureau Standard, Handbook 91 USGPO, Washington, DC, 1963, pp. 6-22-6-26.

[14] T. Kinoshita, "Accuracy of the fine-structure constant," *IEEE Trans. Instrum. Meas.*, pp. 172-175, this issue.

[15] B. P. Kibble, I. A. Robinson, and J. H. Belliss, "A realization of the SI watt by the NPL moving-coil balance," *NPL Rep. DES 88*, National Physical Lab., Teddington, U.K., May 1988, 54 p.

T-IM/38/2/1/26007

A Low Field Determination of the Proton Gyromagnetic Ratio in Water

Edwin R. Williams
George R. Jones, Jr.
Sheng Ye
Ruimin Liu
Hitoshi Sasaki
P. Thomas Olsen
William D. Phillips
Howard P. Layer

Reprinted from
IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT
Vol. 38, No. 2, April 1989

A Low Field Determination of the Proton Gyromagnetic Ratio in Water

EDWIN R. WILLIAMS, MEMBER, IEEE, GEORGE R. JONES, JR., SHENG YE, RUIMIN LIU, HITOSHI SASAKI, P. THOMAS OLSEN, MEMBER, IEEE, WILLIAM D. PHILLIPS, AND HOWARD P. LAYER

Abstract—We measure the proton gyromagnetic ratio in H_2O by the low field method, $\gamma_p'(\text{low})$. The result $\gamma_p'(\text{low}) = 2.67\ 513\ 376\ 10^8\ s^{-1}$ T_{NBS}^{-1} (0.11 ppm), leads to a value of the fine structure constant of $\alpha^{-1} = 137.0\ 359\ 840$ (0.037 ppm) and a value for the quantized Hall resistance in SI units of $R_H = 25\ 812.80460\ \Omega$ (0.037 ppm). To achieve this result, we measured the dimensions of a 2.1-m solenoid with an accuracy of 0.04 μm, and then measured the NMR frequency of a water sample in the field of the solenoid.

I. Introduction

AFTER COMPLETING a measurement of $\gamma_p'(\text{low})$ in 1979 [1] we began building an entirely new apparatus to further improve our measurements and consequently test quantum electrodynamic (QED) theory more stringently. We now report the first results of this latest effort.

The low field method of measuring the proton gyromagnetic ratio in H_2O, $\gamma_p'(\text{low})$, involves two experiments. (The prime indicates that a spherical sample of pure H_2O at a temperature of 25°C is used.) First, we measure the dimensions of a precision single-layer solenoid by an inductive technique in which the position of the current in the wire is located [1], [2]. In the second part, we measure the proton precession frequency ω_p' by standard NMR techniques. $\gamma_p'(\text{low})$ is then obtained from: $\gamma_p'(\text{low}) = \omega_p'/\xi I$, where ξ is the coil constant equal to the magnetic flux density for unit current calculated from the measured dimensions and I is the current in the solenoid. The motivation for improving the accuracy of $\gamma_p'(\text{low})$ comes from its important contribution to our knowledge of the values of the fundamental constants, particularly the fine structure constant α, since

$$\alpha^{-1} = \left\{ \frac{(\mu_p'/\mu_B)\,[R_H]_{LAB}\,[2e/h]_{LAB}}{2\mu_0 R_\infty [\gamma_p']_{LAB}} \right\}^{1/3} \qquad (1)$$

Manuscript received June 10, 1988. This work was supported in part by the U.S. Department of Energy.

E. R. Williams, G. R. Jones, Jr., P. T. Olsen, and W. D. Phillips are with the Electricity Division, National Bureau of Standards, Gaithersburg, MD 20899.

Y. Sheng is with the National Bureau of Standards, Gaithersburg, MD 20899, on leave from the Huazhong University of Science and Technology, Wuhan, China.

L. Ruimin is with the National Bureau of Standards, Gaithersburg, MD 20899, on leave from the National Institute of Metrology, Beijing, China.

H. Sasaki is with the National Bureau of Standards, Gaithersburg, MD 20899, on leave from the Electrotechnical Laboratory, Japan.

H. P. Layer is with the Length and Mass Division, National Bureau of Standards, Gaithersburg, MD 20899.

IEEE Log Number 8826007.

and, therefore, the quantized Hall resistance R_H, since

$$R_H = \left\{ \frac{\mu_0^2 c^3 (\mu_p'/\mu_B)\,[R_H]_{LAB}\,[2e/h]_{LAB}}{16 R_\infty [\gamma_p']_{LAB}} \right\}^{1/3} \qquad (2)$$

because of the current interest in adopting a value of R_H as a representation of the ohm. In these equations μ_0 is the permeability of free space, c is the speed of light in vacuum, μ_p'/μ_B is the magnetic moment of the proton in units of the Bohr magneton, and R_∞ is the Rydberg constant for infinite mass. These quantities are known to a few parts in 10^8 or better. Note that the three electrical constants R_H, $2e/h$, and γ_p' must be measured in the same laboratory (LAB) units, and that there is a cube root dependence on the measured quantities. Moreover, (1) and (2) do not depend on any direct measurement of SI electrical units. Therefore, our more accurate value of γ_p' will help test the quantum Hall theory and the ac Josephson effect theory, as well as help to test QED.

The principal uncertainty in our previous determination was caused by the measurement of the solenoid diameter. For this new experiment we constructed a 2.1-m long by 0.295-m diameter solenoid. This longer solenoid allows us to employ a compensation technique which eliminates the need for an accurate measurement of the mean solenoid diameter.

II. Reducing the Sensitivity to Diameter

The magnetic field of an ideal infinite helical solenoid has two favorable properties: the field inside is uniform, and the magnetic field is independent of the diameter, as it depends only on the number of turns per unit length (sometimes called the pitch). For solenoids of finite length, we have developed a compensation technique which retains these properties of field uniformity and reduced sensitivity to coil diameter when measuring the field at the center of the solenoid. This is accomplished by having five current sources, each of which puts current into segments of the solenoid as shown in Fig. 1. Using a computer program, we found many useful configurations and chose the one that gives a field uniform to better than 0.2 ppm within a spherical volume 8 cm in diameter, with an insensitivity to solenoid diameter similar to that of a 1.5-km long solenoid. In effect, the extra current in the end windings compensates for the finite length of the

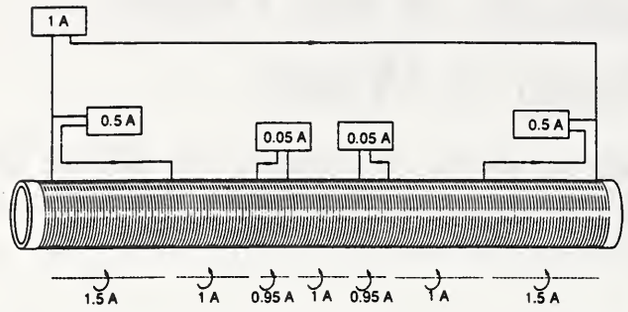


Fig. 1. Compensation scheme for obtaining an "infinite" solenoid for NMR measurements. By using five current sources connected as indicated, a magnetic field is produced that is both uniform to 2 parts in 10^8 over a 6-cm diameter sphere and essentially independent of the average diameter of the solenoid. Leakage between the five current sources must be kept very low.

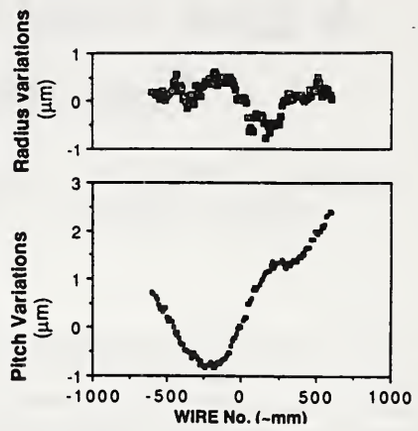


Fig. 2. Radius variations (top) and axial position variations (bottom) in micrometers versus wire number (~ 1 mm/turn). The radius variations are with respect to the center wires. For the axial variations a uniform pitch of 1.058 588 mm/turn has been subtracted from the laser interferometer readings at each wire number.

solenoid. The solenoid has to be long enough as otherwise it is difficult to find a useful solution using practical current levels. Therefore, we built this longer solenoid for our present experiment. This single layer solenoid is wound on a fused silica form, and the grooves in the form have been hand lapped in order to achieve high uniformity in the radius and pitch. The wire is gold-plated copper and was drawn through a die directly onto the silica form to ensure wire uniformity and roundness. The solenoid pitch is 1.058 mm/turn and the wire diameter is 0.8 mm. Fig. 2 shows how uniformly it was constructed. With this solenoid and our compensation technique, we greatly reduce the need to measure the diameter. However, solenoid radius variations and pitch variations are still the critical dimensions that must be measured in order to calculate the magnetic flux density.

III. Dimensional Measurements

The magnetic techniques used to measure the solenoid radius variations and the axial position (pitch) variations have been described in earlier papers [1]–[3], so we will just summarize the method here. An ac current with a special wave form [2], [3] is injected into the turns being measured (we measure ten at a time), and a voltage is induced into five coils wound on another fused silica form (see Fig. 3). Coils A, A', B, B' have about 600 turns each, while coil C has 300 turns, and the entire probe system can be moved along the solenoid axis. Coils A and A' are wired in opposition and form a linear differential transformer that has a sensitivity to axial displacements of 0.01 μm. As we move the current injector from one set of ten turns to the next, a laser interferometer measures the displacement of the pick-up coil assembly between successive null readings. This displacement is a measure of the distance between the centers of current of the measured turns. At the same time the other three coils are used to measure the radius variations. The voltage induced in coil C is inversely proportional to the radius of the activated turns, while the voltages B and B' are directly proportional. The number of turns in coil C is adjusted to cancel the sum of coils B and B'. Therefore, the total small voltage difference is very sensitive to radius changes, yet insensitive to axial position. This three-coil radius-to-voltage transducer has a sensitivity of 0.01 μm. We calibrate it by having extra turns at the end of the solenoid that have 50 μm greater diameter than the main solenoid. A 10-percent error in this calibration produces a 0.016-ppm uncertainty in γ_p'. When measuring both the axial position and radius variations, it is necessary to measure the effect of horizontal and vertical displacements of the probe coils as well as changes in the probe's horizontal and vertical angle. The corrections that must be made for each of these departures from the axis have a quadratic dependence on the radius variations, so we position the probe at the peak of these curves and define this to be the center. Corrections are applied for motions off this center, but in all cases they are less than 0.3 μm and their uncertainties are at least ten times less. A bubble level is used to measure the departure from straightness as the current injector travels along the solenoid. The laser used was calibrated on a daily basis against an iodine stabilized laser [4].

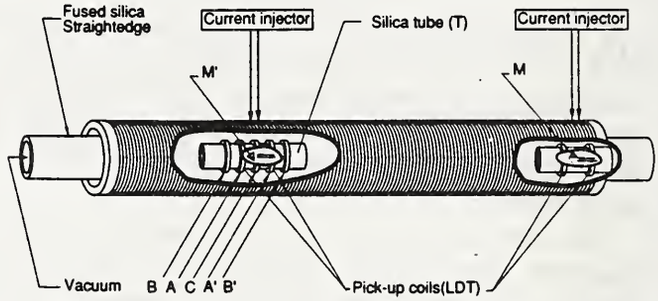


Fig. 3. System used to measure the dimensions of the solenoid by determining the axial position and radius variations of the wires. The five coils A, A', B, B', and C are attached to a fused silica tube and can be pushed or pulled along the axis of the solenoid. Coils A and A' locate the axial position 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 fused silica straight edge is used to guide the five-coil probe and forms part of the vacuum chamber.

Using these magnetic induction techniques we measured the solenoid radius variations and axial position variations as a function of turn number. Our technique measures the average position of ten turns, so the 2100

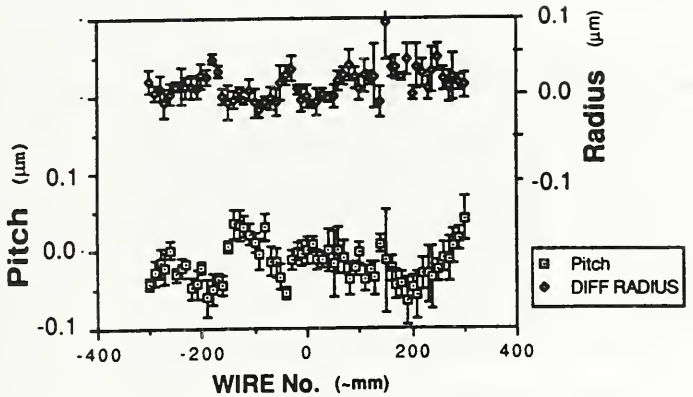


Fig. 4. Difference between radius variations (top) and axial position variations (bottom) from measurements before and after the NMR measurements (should be zero in an ideal case). The error bars represent the standard deviation achieved in one data set.

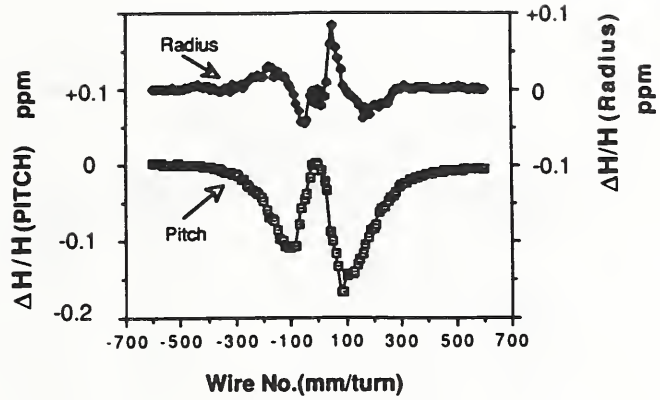


Fig. 5. The correction to the magnetic field due to the radius and axial position data from one of five such measurement sets. Corrections are in parts per million of the field and must be summed.

turns are measured as 210 positions. A pitch of 1.058 588 mm/turn has been subtracted from the axial measurements, and the result is plotted in Fig. 2. Fig. 4 shows the difference between one measurement set carried out on May 2–4, 1988, just after the NMR measurements and a set carried out on March 27–28, just before the NMR measurements. Only the critical region between ±300 turns was measured on the later occasion. The error bars represent the standard deviation of the four measurements taken for each wire in the March set. From each of these measurement sets we calculate a correction to the magnetic flux density as shown in Fig. 5. Five such data sets were taken, three before the NMR measurements and two after. The standard deviation of these five is 0.019 ppm ($\sigma_m = 0.009$) for the axial variations correction and 0.009 ppm ($\sigma_m = 0.004$) for the radial variations correction. From the dimensional measurements we can also calculate the gradient in the magnetic field along the axis. The dashed curve in Fig. 6 is the gradient in the magnetic field so calculated. We will compare these gradients to those measured with NMR.

IV. NMR Results

We have resolved a major systematic error that prevented us from reporting a preliminary result for γ_p' at the CPEM'86 meeting. We discovered and eliminated a time dependent leakage current that was caused by high voltage breakdown when we reversed the current through our solenoid. This leakage current passed directly between the solenoid terminals, so that it was difficult to detect even though it was large ($\sim 10^{-5}$ A).

We measure the NMR using a 3.5-cm diameter water sample. The 52-kHz frequency is detected by a tuned pickup coil having 800 turns wound in sections in a spherical arrangement. Helmholtz RF drive coils have 5 turns per section and are perpendicular to the detector coil. Both detector and drive coils are perpendicular to the solenoid field. A frequency synthesizer is used to sweep through the 0.015-Hz wide resonance. A lock-in amplifier detects both the in-phase and quadrature signals. The sensitivity of the result to the accuracy of setting the phase is 0.022 ppm/deg, and the phase can be adjusted to ±1.5 deg.

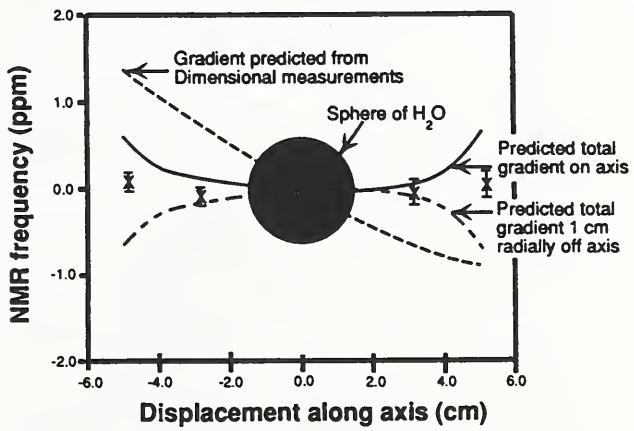


Fig. 6. The 3.5-cm sphere of water is used to measure the gradient of the magnetic flux density. The NMR data are the points plotted as x's. They are a measure of the field averaged over the entire volume. The dashed sloped curve is the contribution to the gradient caused by the solenoid imperfections and is calculated from the dimensional measurements of the solenoid. The other two curves are the gradients calculated taking account of all compensation currents and imperfections along the axis of the solenoid or 1 cm radially displaced from it.

We used this NMR detector to measure the flux density gradient along the axis, but first we calculated the linear gradient produced by unbalancing the two 0.05-A current sources in Fig. 1. The solid curve in Fig. 5 is a sum of all predicted gradients along the axis while the double dashed curve is that predicted 1-cm off axis. The 3.5-cm sample averages the field over its volume, so the gradient measured via NMR (the x's with error bars) agrees with the calculated gradient.

In testing for systematic effects we found that different detector coils produced different results. One coil was loaned to us by Vigoureux of NPL and another by Weyand of PTB, and we also used two NBS coils. To measure the correction for the susceptibility of the detector coil we designed another coil which was wound directly on the water sample like the PTB coil, but was small enough to fit inside the other coils. We then measured the NMR frequency with and without the coil under test. Table I shows that after correcting for the susceptibility of the detector coil, all coils give similar results. The susceptibility measured for the NBS no. 1 coil is in agreement with that measured for the 1979 experiment by another method.

TABLE I
SUSCEPTIBILITY MEASUREMENTS

Coil name	Susceptibility Correction to B (ppm)	Difference from NBS No. 2 (ppm)	NMR difference (ppm)
NBS No. 2	-0.033 ± 0.045	0	0
NBS No. 1	-0.068 ± 0.072	-0.035 ± 0.08	+0.080 ± 0.08
NPL	+0.273 ± 0.05	+0.306 ± 0.06	+0.322 ± 0.12
PTB			-0.472 ± 0.1

We used the NBS no. 2 coil for the final data taking because its spherical shape gives a more uniform filling factor. We measured the critical 1-Ω resistor against the NBS reference group just before and just after the NMR measurement, and the standard deviation of the measurements was 0.009 ppm. A 1.5-km long cable connecting the γ_p' and the Josephson voltage standard laboratories was used each day in between NMR measurements to calibrate our Zener voltage reference against $2e/h$. One key feature of this cable is a coaxial guard system that has current flowing in the guard coaxial shield [5]. This guard reduces leakage by between 10–100 times, but even with the guard current off, the voltage transfer changed by only 0.01 ppm. Fig. 7 is a plot of the daily average of the NMR frequencies with error bars indicating the standard deviation of the day's average. The standard deviation of these daily averages is 0.052 ppm and represents one of the largest sources of error. We do not use the standard deviation of the mean because this is a daily average and we have not carried out systematic tests to higher accuracy. Although the long-term day-to-day scatter of the Zener diode reference does not affect the NMR observations, the short-term hourly scatter may be a significant contribution. We would like a better voltage reference. The temperature coefficient of the solenoid, 0.495 ppm/°C measured by NMR data and 0.496 ppm/°C by dimensional measurements, is consistent with the nominal value for the expansion of fused silica.

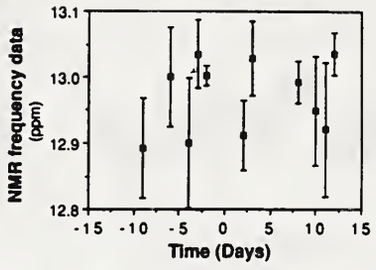


Fig. 7. The NMR readings, averaged over each day. The error bars are the standard deviation of that day's measurements. The unweighted mean is 12.97 ± 0.052 ppm (σ_m = 0.016 ppm).

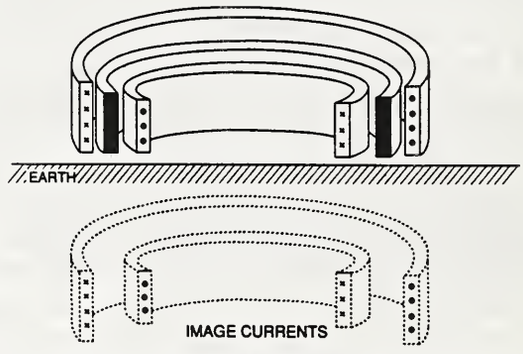


Fig. 8. The coil assembly used to determine the susceptibility of the Earth. Current in the inner and outer coils produces only a small flux in the center detector coil, which detects flux from currents in the Earth.

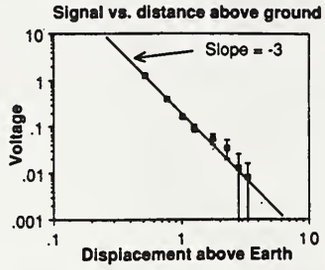


Fig. 9. A log-log plot of the voltage induced in the susceptibility detector coil versus height above the ground. The solid line has a slope of exactly 3 thus showing a cubic dependence on distance.

V. CORRECTIONS FOR IRON

By moving an iron object closer to the solenoid during the NMR measurements and knowing that its systematic falls off as the 6th power of the distance, we can correct for the objects required near the experiment. The largest such correction (-0.0018 ppm) is required for the pump that circulates a fluorocarbon cooling fluid to a shower that controls the solenoid temperature. The iron in the ground around the nonmagnetic facility must also be accounted for. Fig. 8 shows a set of coils that were constructed to measure both the earth's iron content and the solenoid susceptibility. Current from the special waveform used in the dimensional measurements is passed through the inner and outer coils in opposite directions. Their turns ratio has been chosen such that a very small voltage is induced in the middle coil, but the two coils have as large a net dipole moment as possible. The image currents produced by this dipole in the earth can be detected as we move the three coil assembly closer or further from the earth. Fig. 9 is a log–log plot of the detected voltage as a function of distance above the earth just outside the room housing the solenoid. The line has the slope of three predicted by the cubic dependence in the distance from an infinite permeable sheet. By performing such measurements from 0.5 to 1.2 m underneath the solenoid we obtain a correction to γ_p' of (-0.12 ± 0.03 ppm) for the solenoid position which is 2.7 m above the ground.

Our solenoid is manufactured from fused sand, so iron impurities are possible although the manufacturer promised they would be small. In the center of the coil assembly used to measure the earth's susceptibility there is a sufficiently large flux density, and the hole is large enough to insert our 29-cm diameter solenoid. We, therefore, calibrated the assembly with a material of known susceptibility such as ethyl alcohol and measured the change off voltage induced in the detector coil as we placed a 10-cm section taken from the end of the solenoid inside the as-

sembly. We also measured a section of the fused silica vacuum chamber. Fickett of NBS/Boulder also calibrated small core sections taken from another section. The measured susceptibilities all agree with the accepted susceptibility of pure fused silica. The correction calculated using this value of susceptibility is negligible because the five current configuration also makes this correction approach the zero correction of an infinite cylinder.

VI. Results and Conclusions

Table II summarizes most of the corrections that must be applied to calculate γ_p' and the corresponding estimates of uncertainty if appropriate. The dimensional measurements, the NMR measurements, and the various calibrations all contribute about 0.05 ppm each. Thus no one item is presently limiting the accuracy of the results. Our value for γ_p' is

$$\gamma_p'(\text{low}) = 2.67\ 513\ 376 \times 10^8\ (0.11\ \text{ppm})\ \text{s}^{-1}\ T_{\text{NBS}}^{-1}.$$

In another paper [6] we compute the following quantities that are derived from this work and compare them to other measurements:

$$\text{from (1): } \alpha^{-1} = 137.0\ 359\ 840(51)\ (0.037\ \text{ppm})$$

$$\text{from (2): } R_{\text{H}} = 25\ 812.80\ 460(95)\ (0.037\ \text{ppm}).$$

This value of α^{-1} agrees with the QED value, the difference being $(-0.054 \pm 0.038$ ppm), but differs by somewhat more than two combined standard deviations from the NBS absolute ohm realization, the difference being $(-0.102 \pm 0.043$ ppm). In Fig. 10 we have plotted our value of γ_p' along with the values that were considered in the 1986 adjustment of the fundamental physical constants [7]. The agreement with the precise QED value is satisfying, but the difference between our value and the NBS ohm value, which also has a relatively small uncertainty, is disconcerting. We plan to continue our measurements to test further for any errors. This result provides one of the most accurate routes for measuring R_{H} in SI units and will help ensure that the value adopted for R_{H} is near its SI value.

For convenient future reference, we express our value of γ_p' in terms of representations of the volt and ohm based on the following adopted values of the Josephson frequency to voltage quotient and the quantized Hall resistance:

$$R_{\text{H}} \equiv 25\ 812.807\ \Omega$$

$$2e/h \equiv 483\ 597.9\ \text{GHz/V}.$$

These values are those likely to be adopted by the Comité Consultatif d'Électricité, CCE, for basing practical representations of the volt and ohm on the Josephson and quantum Hall effects. The result is

$$\gamma_p'(\text{low}) = 2.67\ 515\ 427 \times 10^8\ \text{s}^{-1}\ T^{-1}\ (0.11\ \text{ppm}).$$

Acknowledgment

The authors would like to thank R. Steiner, F. Fickett, B. Petley, and R. Dziuba who helped calibrate our Zener voltage references, measure sample susceptibilities, measure NMR systematic errors, and calibrate resistors, respectively. They also thank B. N. Taylor for his continued support and encouragement.

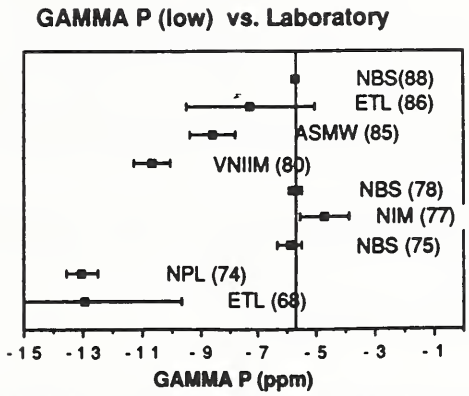


Fig. 10. The values of γ_p'(low) from various laboratories compared with our value.

TABLE II
One Standard Deviation Uncertainty Estimates

Item	Correction to γ_p' (ppm)	Uncertainty (ppm)
NMR measurements	-	0.052
Other NMR related corrections	-0.0051	0.034
Detector susceptibility	0.033	0.045
Solenoid susceptibility	0	0.02
Earth's susceptibility	-0.12	0.03
Pitch measurements	-	0.050
Radius variations	0.166	0.019
Return leads etc.	-0.131	0.02
Resistor calibrations	35.057	0.012
One ohm power coefficient	-0.228	0.05

$\gamma_p' = 2.67513376\ 10^8\ \text{Hz}/T_{\text{NBS}}\ \pm 0.11$ ppm

$\alpha^{-1} = 137.0359840 \pm 0.037$ppm

References

[1] E. R. Williams and P. T. Olsen, "New measurement of the proton gyromagnetic ratio and a derived value of the fine-structure-constant accurate to a part in 10^7," *Phys. Rev. Lett.*, vol. 42, pp. 1575-1579, June 1979.

[2] ——, "A noncontacting magnetic pickup probe for measuring the pitch of a precision solenoid," *IEEE Trans. Instrum. Meas.*, vol. IM-21, pp. 376-379, Nov. 1972.

[3] E. R. Williams, G. R. Jones, Jun-Shou Song, W. D. Phillips, and P. T. Olsen, "Report on the new NBS determination of the proton gyromagnetic ratio," *IEEE Trans. Instrum. Meas.*, vol. IM-34, pp. 163-167, June 1985.

[4] H. P. Layer, "A portable iodine stabilized helium-neon laser," *IEEE Trans. Instrum. Meas.*, vol. IM-29, pp. 358-361, Dec. 1980.

[5] E. R. Williams, P. T. Olsen, and B. F. Field, "Standard cell calibration via current transfer," *IEEE Trans. Instrum. Meas.*, vol. IM-23, pp. 299-301, June 1974.

[6] M. E. Cage, R. F. Dziuba, R. E. Elmquist, G. R. Jones, Jr., P. T. Olsen, W. D. Phillips, J. Q. Shields, R. L. Steiner, B. N. Taylor, and E. R. Williams, "NBS determination of the fine-structure constant, and of the quantized Hall resistance and Josephson frequency to voltage quotient in SI units," pp. 284-289, this issue.

[7] E. R. Cohen and B. N. Taylor, "The 1986 adjustment of the fundamental physical constants," *Rev. Mod. Phys.*, vol. 59, pp. 1121-1148, Oct. 1987.

NBS-114A (REV. 2-8C)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR REPORT NO. NISTIR 90-4272	2. Performing Organ. Report No.	3. Publication Date MARCH 1990

4. TITLE AND SUBTITLE

Selected Publications for the EMAP Seminar

5. AUTHOR(S)

H.H. Ku, editor

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

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)

10. SUPPLEMENTARY NOTES

☐ 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 is a collection of publications of NBS/NIST that has been found to be essential and useful in the conduct of the Electricity Measurement Assurance Program (EMAP) Workshop under the sponsorship of the National Institute of Standards and Technology and the National Conference of Standards Laboratories. Publications included in this volume appeared originally as articles scattered in journals, NIST reports, technical notes, monographs and special publications. Some of these papers are now out of print. The present volume also contains papers which are relevant to the 1990 changes in the electrical units. This publication, combining many of the relevant publications edited for the Workshop, will not only serve as a useful text and reference in class, but also for use by all professionals interested in electrical measurements as well.

12. KEY WORDS (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)
calibration designs; electrical units; measurement assurance; statistical analysis; statistical control

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
456

15. Price
A20