JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 88, No. 5, September-October 1983

CONTENTS

Library of Congress Catalog Card Number: 63-37059

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, DC 20402
Single copy price $5.50 Domestic; $6.90 Foreign.
Subscription price: $18.00 a year; $22.50 foreign mailing.
UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON: 1983

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 88, No. 5, September-October 1983

History of Atomic Clocks

Norman F. Ramsey
Mount Holyoke College, South Hadley, MA 01075
and Harvard University, Cambridge, MA 02138

Accepted: June 21, 1983

The history of atomic and molecular standards of time and frequency is traced from the earliest work on molecular and atomic beam resonance techniques to more recent developments that promise improved standards in the future. The various devices currently used as standards are discussed in detail from an historical prospective. The latter part of the article is devoted to a discussion of prospective developments which hold promise for major improvements in accuracy, stability and reproducibility.

Key words: accuracy; atomic beam; clocks; frequency; frequency stabilization; laser; magnetic resonance; maser; microwave absorption; molecular beam; optical pumping; reproducibility; stability; superconducting cavities; time; trapped-ion devices.

Introduction

In discussing the history of atomic and molecular standards of time and frequency, two alternative approaches are available. One is to treat all devices in parallel on a year-by-year basis. The other is to discuss each alternative device in succession. It is clear that the latter approach is the more suitable and will be followed here, but frequent cross-references will be made to other devices. In following this procedure, it is clear that the first technique discussed should be the molecular and atomic beam magnetic resonance method; historically, this was the first method and it remains one of the most effective time standards.

The Molecular Beam Resonance Method

The earliest molecular beam experiment was that of Dunoyer [1][1] to show that Na atoms travel in straight lines in an evacuated tube. The development of molecular beams as a valuable research technique was largely due to the work of Otto Stern [2] and his collaborators in Hamburg in the 1920's and early 1930's, and to contributions from I. I. Rabi and his associates at Columbia in the 1930's.

The early molecular beam experiments did not use oscillatory fields and were of limited accuracy. The molecular beam magnetic resonance method arose from a succession of ideas, the earliest of which can be

About the Author, Paper: Higgins Professor of Physics at Harvard University, Norman F. Ramsey is one of America's foremost scientists in the precision measurement of the basic properties of matter. This invited paper is an outgrowth of the Fourteenth Annual Precise Time and Time-Interval meeting held at the NASA Goddard Space Flight Center on Dec. 2, 1982, at which a panel of distinguished scientists, including I. I. Rabi, H. G. Dehmelt, R. H. Dicke, R. V. Pound and C. H. Townes, as well as Professor Ramsey, discussed "An Historical Context for the Scientific Principles of Atomic Clocks." The same day marked the opening of a new exhibit on "Atomic Clocks" at the Smithsonian National Museum of American History. This paper was written by Professor Ramsey in response to an invitation of the committee that coordinated these events; it is based on an article by the author in the report, *Science, Technology, and the Modern Navy,* issued in 1976 by the U.S. Office of Naval Research in Arlington, VA as report ONR-37. An earlier review by the author appeared in *IEEE Transactions on Instrumentation and Measurement* **IM-21**, 90 (1972), and a more recent review by H. Hellwig, K. M. Evenson and D. J. Wineland appeared in *Physics Today* **31**, 23 (December 1978).

[1] Numbers in brackets indicate literature references at the end of this paper.

traced back to 1927, although that idea was rather remote from the principle of magnetic resonance. Physicist Sir Charles Darwin [3]—the grandson of the great evolutionist—discussed theoretically the nonadiabatic transitions that make it possible for an atom's angular momentum components along the direction of a magnetic field to be integral multiples of $h/2\pi$ both before and after the direction of the field is changed an arbitrary amount. Inspired by Darwin's theoretical discussion, Phipps and Stern [4] in 1931 performed the first experiments on paramagnetic atoms passing through weak magnetic fields whose directions varied rapidly in space. Guttinger [5] and Majorana [6] developed further the theory of such experiments. Frisch and Segre [7] continued atomic beam experiments with adiabatic and nonadiabatic transitions of paramagnetic atoms and found, in agreement with Guttinger's and Majorana's theories, that transitions took place when the rate of change of the direction of the field was larger than or comparable to the Larmor frequency,

$$\omega_0 = \gamma_I H_0, \tag{1}$$

which is the classical frequency of precession of a classical magnetized top with the same ratio γ_I of magnetic moment to angular momentum. Transitions did not take place when the rate of change of the direction of H was small compared to the Larmor frequency. However, some of the results of Frisch and Segre were not consistent with theoretical expectations. Rabi [8] pointed out that these discrepancies arose from the effects of the nuclear magnetic moments since some of the transitions were performed in such weak fields that strong or intermediate coupling between the nuclei and the electrons prevailed. The transitions in such circumstances were quite different from those for which the effects of the nuclear spins could be neglected. Rabi showed that the results of Frisch and Segre were consistent with expectations if the effects of the nuclei were included. Rabi also pointed out that such nonadiabatic transitions could be used to identify the states and hence to determine the signs of the nuclear magnetic moments. Motz and Rose [9], Rabi [10], and Schwinger [11] in 1937 calculated the transition probability for molecules that passed through a region in which the direction of the field varied rapidly.

In all of the above experiments, however, the direction of the field varied in space, the only time variation arising as the atoms in the atomic beam passed through the region. Since the atoms possessed a Maxwellian velocity distribution, the atomic velocities varied and the apparent frequencies of the changing field were different for different velocities. Furthermore, the change in field direction ordinarily went through only a portion of a full cycle. For both of these reasons, no sharp resonance effects could be expected. Rabi in his brilliant 1937 theoretical paper entitled "Space Quantization in a Gyrating Magnetic Field" [10] assumed for simplicity that the field was oscillatory in time even though the primary application was to a field varying with the position along the beam rather oscillating with time. As a consequence, all the formulae in that paper are applicable to the resonance case with oscillatory fields and the paper, without alteration, provides the fundamental theory for present molecular beam magnetic resonance experiments as well as for other experiments with magnetic resonance.

While writing his theoretical paper on the gyrating field, Rabi discussed with some of his colleagues the possibility of using oscillatory rather than space varying magnetic fields, but Rabi's laboratory had a full program of important experiments which did not require oscillatory fields, and no experiments utilizing oscillatory fields were started during the six months following the submission of Rabi's theoretical paper on the gyrating field. In September of 1937, C. J. Gorter visited Rabi's laboratory [12] and described his brilliantly conceived but experimentally unsuccessful efforts to observe nuclear magnetic resonance in lithium fluoride, as described in Gorter's publications of the previous year [13]. The research efforts in Rabi's laboratory were soon directed primarily toward the construction of molecular beam magnetic resonance apparatus with oscillator driven magnetic fields. Two successful molecular beam devices using this method were soon constructed by Rabi [14,15], Zacharias [14,15], Millman [14], Kusch [14], Kellogg [15] and Ramsey [15]. A schematic view of these [14] is shown in figure 1. In these experiments the atoms and molecules were deflected by a first inhomogeneous magnetic field and refocused by a second one. When a resonance transition was induced in the region between the two inhomogeneous fields, the occurrence of the transition could easily be recognized by the reduction of intensity asociated with the accompanying failure of refocussing. For transitions induced by the radio-frequency field, the apparent frequency was almost the same for all molecules independent of molecular velocity. As a result, when eq (1) was satisfied sharp resonances were obtained

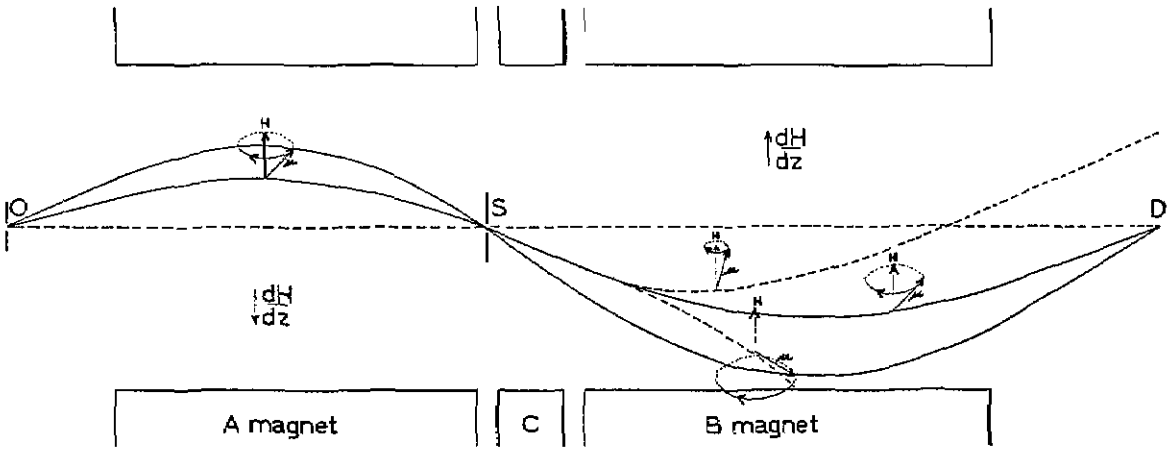


Figure 1–Schematic diagram [12] showing the principle of the first molecular beam resonance apparatus. The two solid curves indicate two paths of molecules having different orientations that are not changed during passage through the apparatus. The two dashed curves in the region of the B magnet indicate two paths of molecules whose orientation has been changed in the C region so the refocusing is lost due to the change in the component of the magnetic moment along the direction of the magnetic field.

Rabi et al. [15] soon extended the method to the molecule H_2, for which the resonance frequencies depended not only on eq (1) but also on internal interactions within the molecule. The transitions in this case occurred whenever the oscillatory field was at a Bohr frequency for an allowed transition

$$h\nu = E_1 - E_2 \qquad (2)$$

For the first time these authors began speaking of their results as "radio-frequency spectroscopy."

Molecular Beam Magnetic Resonance Experiments

By 1939 the new molecular beam magnetic resonance method had demonstrated its usefulness sufficiently well that it appeared to Rabi and his associates to be of possible value for the definition of standard magnetic fields and for use as a time and frequency standard; these potentialities were discussed informally with some scientists at the National Bureau of Standards in 1939. Rabi [16] in his 1945 Richtmeyer Lecture publicly discussed the possibility of atomic clocks and expanded on the implications in a subsequently reported *New York Times* interview.

In most respects the molecular beam technique in 1939 was more suitable as a standard of magnetic field than of frequency or time since the observed resonances then were largely dependent on the externally applied magnetic field. From the point of view of atomic clocks, it was consequently a great step forward when in 1940 Kusch, Millman, and Rabi [17,18] first extended the method to paramagnetic atoms and in particular to $\Delta F = \pm 1$ transitions of atoms where the relative orientation of the nuclear and electronic magnetic moments was changed, in which case the resonance frequencies were determined dominantly by fixed internal properties of the atom rather than by interactions with an externally applied magnetic field. The first resonance measurements of the Cs hyperfine separation, which has been so extensively used in frequency control, was reported [18] in 1940.

In 1941 the research with the atomic beam magnetic resonance method was mostly interrupted by World War II and did not resume until 1946. In 1949 Kusch and Taub [19] pointed out the possibility of observing the hyperfine resonances at magnetic fields such that the resonance frequency was an extremum, in which case the frequency to first order was independent of the strength of the magnetic field.

In 1949 Ramsey [20,21] invented the separated oscillatory field method for a molecular beam resonance experiment on molecular hydrogen. In this new method the oscillatory field, instead of being distributed uniformly throughout the transition region, was concentrated in two coherently driven oscillatory fields in short regions at the beginning and end of the transition region. The theoretical shape of a resonance curve with this apparatus is shown in figure 2. Ramsey pointed out that this method has the following advantages: (1) the resonances are 40% narrower than even the most favorable Rabi resonances with the

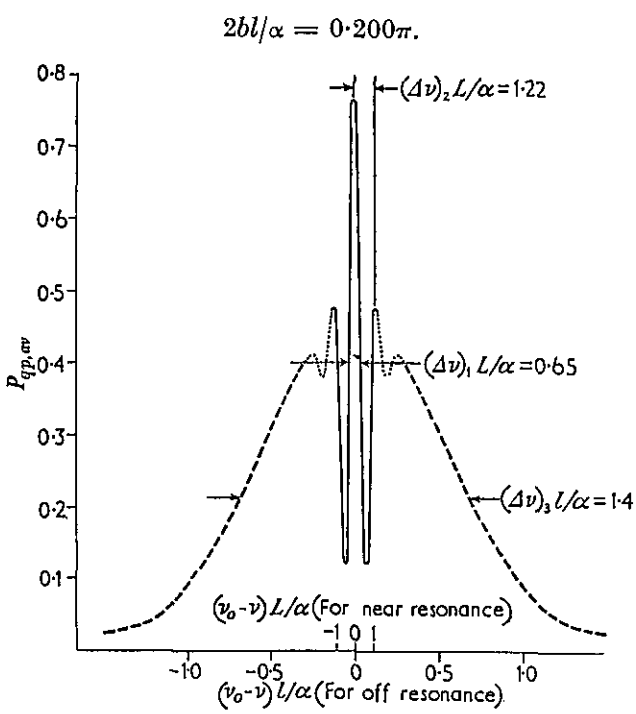


Figure 2–Theoretical shape for separated oscillatory field resonance pattern [20].

same length of apparatus; (2) the resonances are not broadened by field inhomogeneities; (3) the length of the transition region can be much longer than the wavelength of the radiation, provided that the two oscillatory regions are short, whereas there are difficulties with the Rabi method due to phase shifts when the length of the oscillatory region is comparable to the wavelength; (4) the first-order Doppler shift can mostly be eliminated when sufficiently short oscillatory field regions are used; (5) the sensitivity of resonance measurements can be increased by the deliberate use of appropriate relative phase shifts between the two oscillatory fields; (6) with short-lived states the resonance width can be narrowed below that expected from the lifetime of the state and the Heisenberg uncertainty principle if the separation of the oscillatory fields is sufficiently great that only molecules living longer than average in the excited state can reach the second oscillatory field before decaying [22]. The apparatus for which the separated oscillatory field was first proposed is shown in figure 3 [23].

Atomic Beam Frequency Standards

With the above developments, it was apparent to most molecular beam researchers that atomic beam methods could be highly effective for standards of time and frequency even though this view was not shared by advocates of alternative frequency standards, such as crystals. In 1952, Sherwood, Lyons, McCracken and Kusch [24,25] reported briefly on atomic beam resonance research supported by the National Bureau of Standards and directed primarily toward the development of an atomic beam clock. A schematic diagram of a proposed atomic beam clock at that time is given in figure 4. The financial support for such work soon dwindled due to advances in the then new field of microwave spectroscopy and to the view then held at the National Bureau of Standards that a molecular clock based on the microwave absorption by ammonia at its inversion frequency would be simpler and more promising.

A few years later, Zacharias [26,27] stimulated renewed interest in an atomic beam cesium clock. His initial concern was for an entirely new type of cesium beam in which ultrahigh precision would be obtained by the use of extremely slow molecules moving upwards in a vertical apparatus at such low velocities that they would fall back down by the action of gravity. Although this fountain experiment eventually failed due to the unexpected deficiency of the required ultraslow molecules emerging from the source, it stimulated Zacharias to develop and to urge others to develop well-engineered atomic beam frequency standards using normal atomic velocities. The unsuccessful fountain experiment of Zacharias illustrates the value to science even of some unsuccessful experiments; this unsuccessful effort directly and indirectly stimulated three quite different but important developments: (1) the use of conventional but well-engineered atomic beams for frequency control; (2) the development by Kleppner, Ramsey, and others [28–30] of the stored-atom technique, which eventually led to the hydrogen maser; and (3) high-precision resonance experiments with ultraslow neutrons [31]. The first report on an atomic beam frequency standard at a Frequency Control Symposium was that of Zacharias in 1955. Zacharias claimed a short-time stability of 1 part in 10^9 for his atomic cesium frequency standard.

In 1955 Essen and Parry [32] of the British National Physical Laboratory successfully operated the first practical laboratory atomic cesium beam apparatus that was extensively used as an actual frequency standard. Their construction and effective use of this device provided a major impetus to the subsequent development of atomic beam cesium frequency standards.

In 1956 the first commercial model of an atomic beam frequency standard appeared on the market.

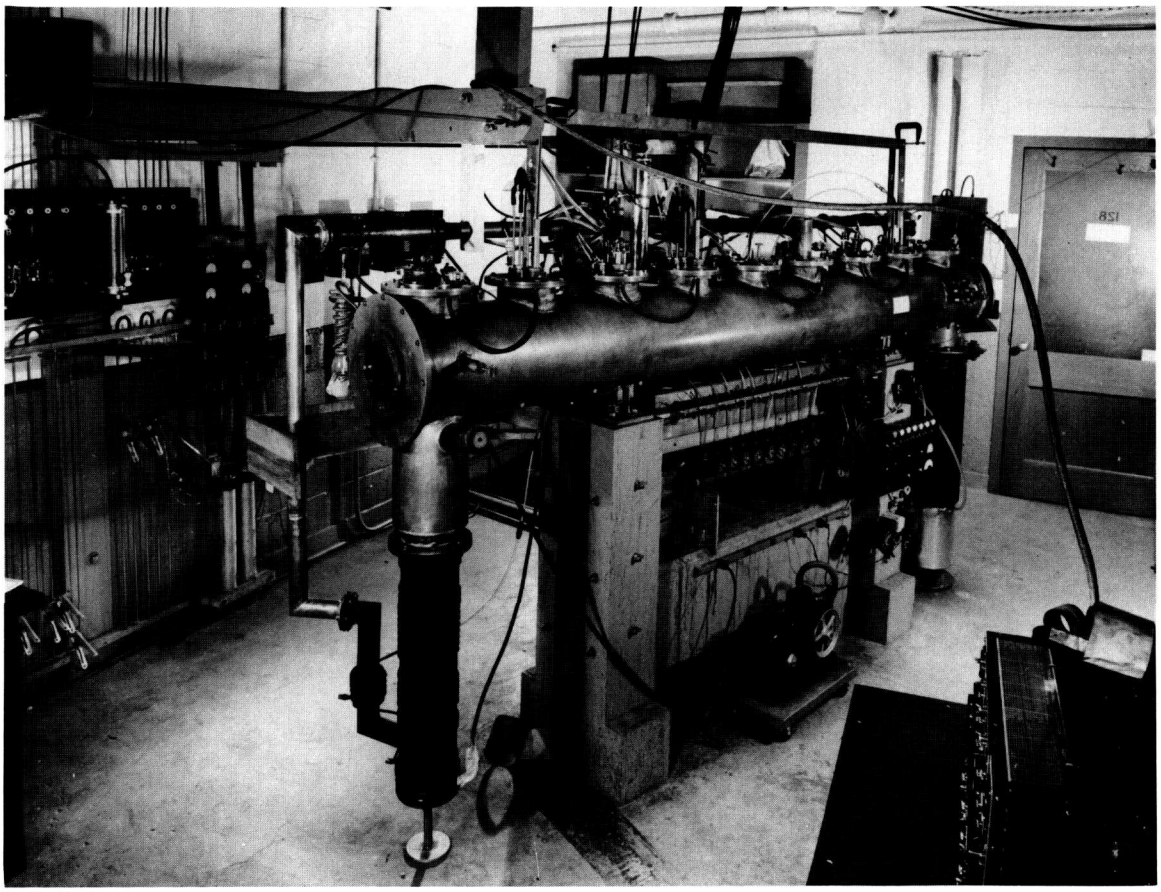


Figure 3–Apparatus for which separated oscillatory field was first proposed [18].

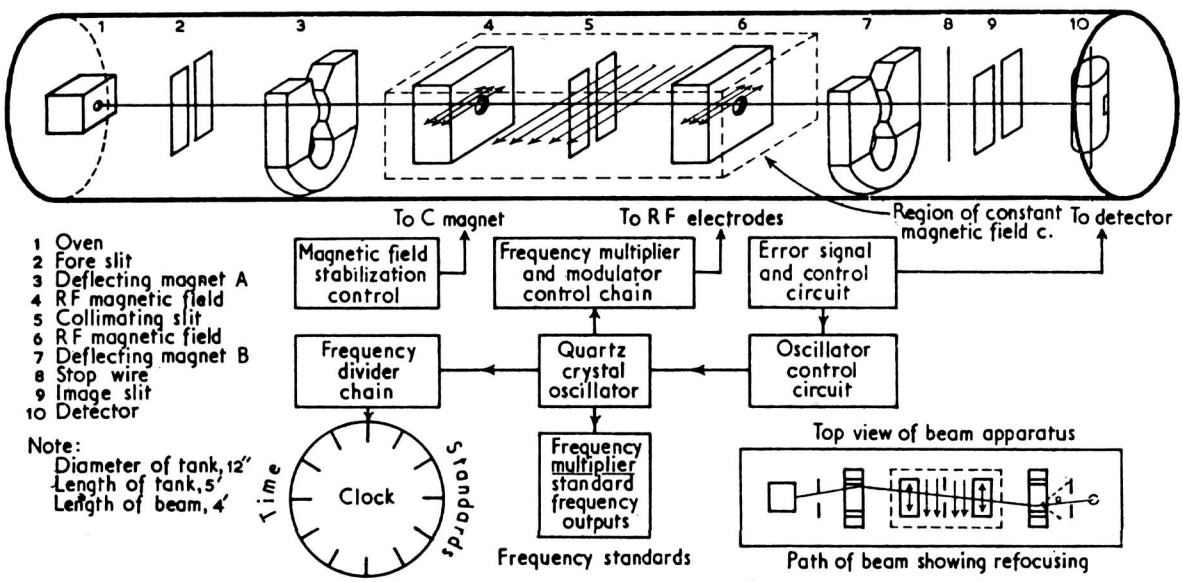


Figure 4–Schematic diagram of a proposed atomic beam clock [25].

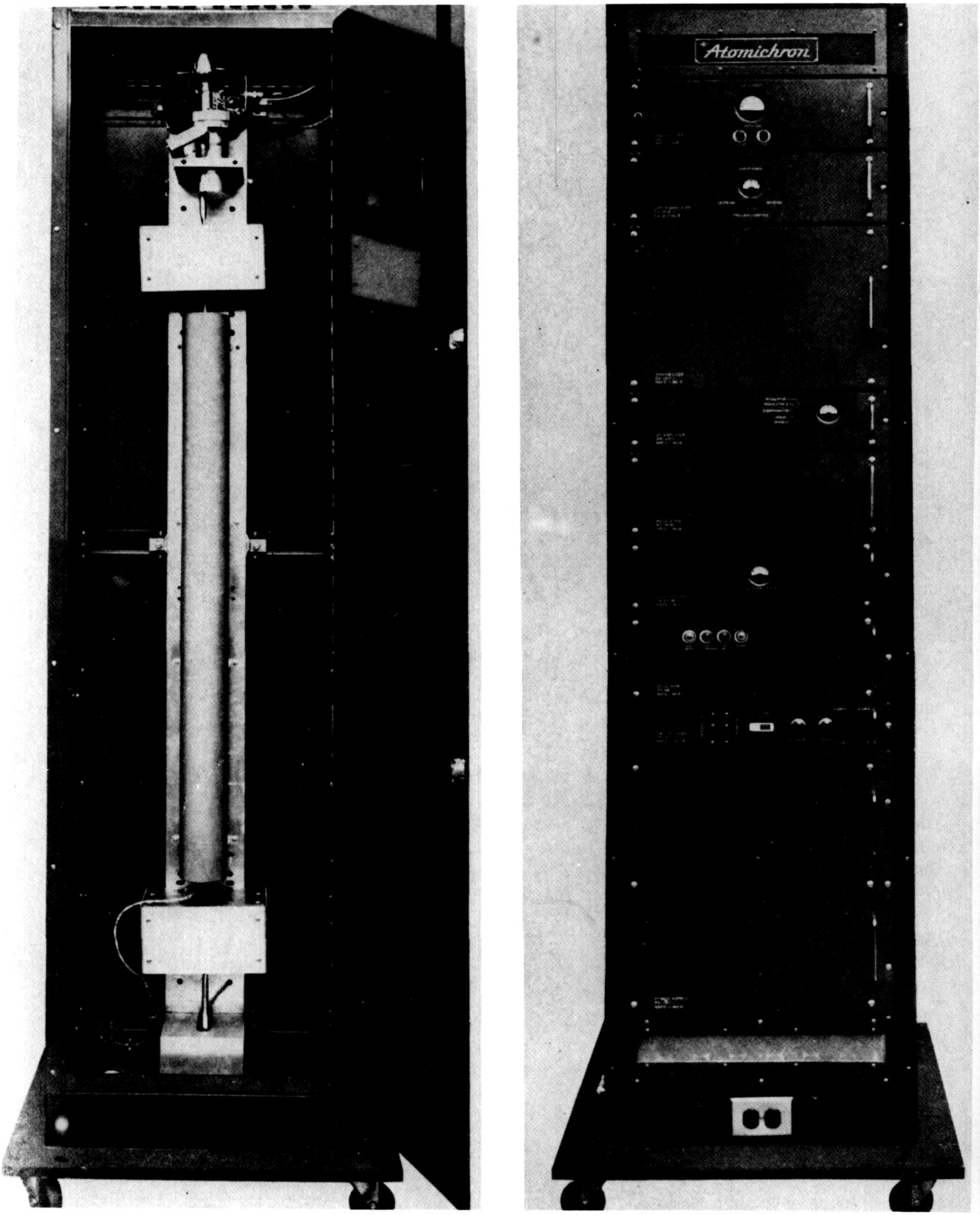


Figure 5–First commercial atomic beam frequency standard. National's Atomichron [33]. © 1972 by the IEEE.

This was National's Atomichron developed [33] by Daly and Orenberg in collaboration with Zacharias and further improved by Holloway and McCoubrey. This device used Ramsey's separated oscillatory field method for increased precision, a special design of cesium oven that could be operated several years without exhaustion, titanium pumping to permit permanent sealing off of the evacuated beam tube, and many other features generally necessary for an effective commercial device. The first commercial Atomichron is shown in figure 5. The development of the Atomichron was supported financially largely by the U.S. Signal Corps at Ft. Monmouth, NJ, and the Office of Naval Research, although some support came from the Air Force. A purchase order by the Signal Corps for a relatively large number of Atomichrons made possible the development of mass-production techniques and improved engineering to permit sufficient reliability and reductions in price to assure commercial success.

The early atomic beam frequency standards were subject to various frequency shifts dependent on the amplitude of the radio-frequency power and on other variables. To account for these results, Ramsey, with the aid of computational analysis supported by the National Co., investigated the various possible distortions that would occur in an atomic beam resonance [34]. The elimination of radio-frequency phase shifts and other sources of distortion made possible the marked increases in accuracy that have been obtained with the atomic beam frequency standards.

From 1956 on, atomic beam frequency standards developed rapidly. Mockler, Beehler, Barnes, Hellwig, Wineland and others [33,35,36] developed an atomic cesium frequency standard at the National Bureau of Standards in Boulder, CO. Other commercial organizations such as TRG, Bomac, Varian, Hewlett-Packard, Frequency and Time Systems, Inc., and Frequency Electronics, Inc. became involved. Many laboratories throughout the world either constructed or purchased atomic beam frequency standards including those in Canada, France, Germany, USSR, China and the laboratories of Kartaschoff [35] and Bonanomi [35] in Switzerland. Reder, Winkler and others [35] at Ft. Monmouth and Markowitz at the Naval Observatory sponsored various worldwide studies of the comparison of atomic clock frequencies and the synchronization of clocks. Extensive studies were made of other atoms such as thallium for use in the atomic beam tubes, and various molecular resonances were studied for possible use in a molecular beam electric resonance apparatus for frequency control purposes. A Tl^{205} frequency measurement accurate to 2 parts in 10^{11} was reported by Bonanomi [37]. However, atomic cesium remains the most widely used substance in molecular or atomic beam frequency control devices. Particularly effective atomic beam cesium clocks were developed and sold by Hewlett-Packard which also developed a "flying clock" particularly suitable for the intercomparison of atomic clocks in different laboratories. A typical beam tube for an atomic cesium frequency standard is shown in figure 6. Accuracies as high as 1 part in 10^{13} have been claimed for some laboratory cesium standards [35].

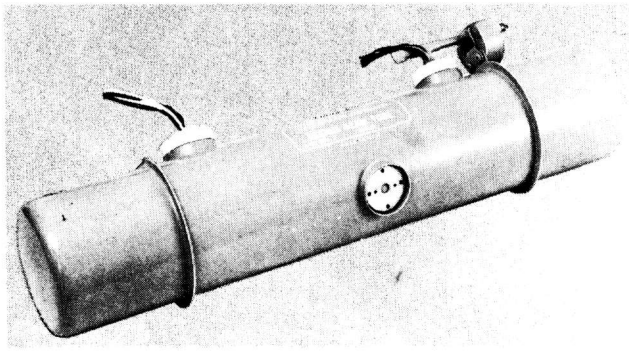


Figure 6–Beam tube for atomic cesium standard manufactured by Varian Associates for Hewlett-Packard [33]. © 1972 by the IEEE.

In 1967, the 13th General Conference on Weights and Measures resolved that the unit of time in the International System of Units should be the second defined as follows: "The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium atom 133," a definition that is still retained.

Microwave Absorption Spectroscopy

Microwave absorption spectroscopy had an early start in the experiments of Cleeton and Williams [38,39] who observed the absorption of microwave radiation at the NH_3 inversion frequency. However, research on microwave absorption was inhibited at that time by the lack of suitable microwave oscillators and circuits so there was no further development of microwave absorption spectroscopy until after the development of microwave oscillators and waveguides for radar components in World War II. Immediately following World War II there was a great burst of activity in microwave absorption spectroscopy. Although there were no publications on experimental microwave spectroscopy in 1945, in the single year of

1946 there were a number of important publications from many different laboratories including reports by the following authors [40]: Bleaney, Penrose, Beringer, Townes, Dicke, Lamb, Becker, Autler, Strandberg, Dailey, Kyhl, Van Vleck, Wilson, Dakin, Good, Coles, Hershberger, Lamont, Watson, Roberts, Beers, Hill, Merritt, and Walter, and in 1947 there were more than 60 published papers on this subject including a number of publications by Gordy and Jen, by authors with reports the previous year, and by others. A typical microwave absorption experiment at this time is shown schematically in figure 7.

Microwave absorption techniques were quickly recognized to be of potential value for frequency standards. In 1948 a group of workers [25] at the National Bureau of Standards built an ammonia clock that was completed in 1949 and is shown in figure 8. It eventually achieved an accuracy of 1 part in 10^8. Rossell [25] in Switzerland and Shimoda in Japan devised an improved ammonia absorption clock good to a few parts in 10^9.

In 1951 Townes [41] analyzed frequency stabilization of microwave spectral lines and the application to atomic clocks. In the references he lists other early contributors to the field. At the 1951 Fifth Frequency Control Symposium, Dicke [42] reported on microwave absorption molecular frequency standards. In the seventh, eighth and ninth symposia, Dicke, Carver, Arditi, and others described the continuation of this work at both Princeton and the Radio Corporation of America (RCA) with the financial support of the Signal Corps and the Office of Naval Research [35,42]. The microwave absorption studies soon merged with the optical pumping techniques, described in the next section, since the intensities of the resonances were greatly enhanced by the use of optical pumping.

Optical Pumping

The starting point of all research on optical pumping was a paper by Bitter [43] in 1949, which showed the possibility of studying nuclear properties in optically excited states. Kastler [44,45] showed the following year that this technique could be effectively combined with the double resonance method he and Brossel [44] had developed. Both optical pumping and optical detection techniques served the purpose of increasing the signal-to-noise ratio of the resonator output signal. The optical pumping greatly enhances the population of certain states so the signal is not weakened by stimulated emission nearly cancelling absorption and the optical detection increases the signal-to-noise ratio because of the lower noise level of optical detectors over microwave detectors.

The combination of optical pumping techniques with the buffer gas method for reducing Doppler shift developed by Dicke [33,42] provided gas cells of real value as frequency-control devices. Although many

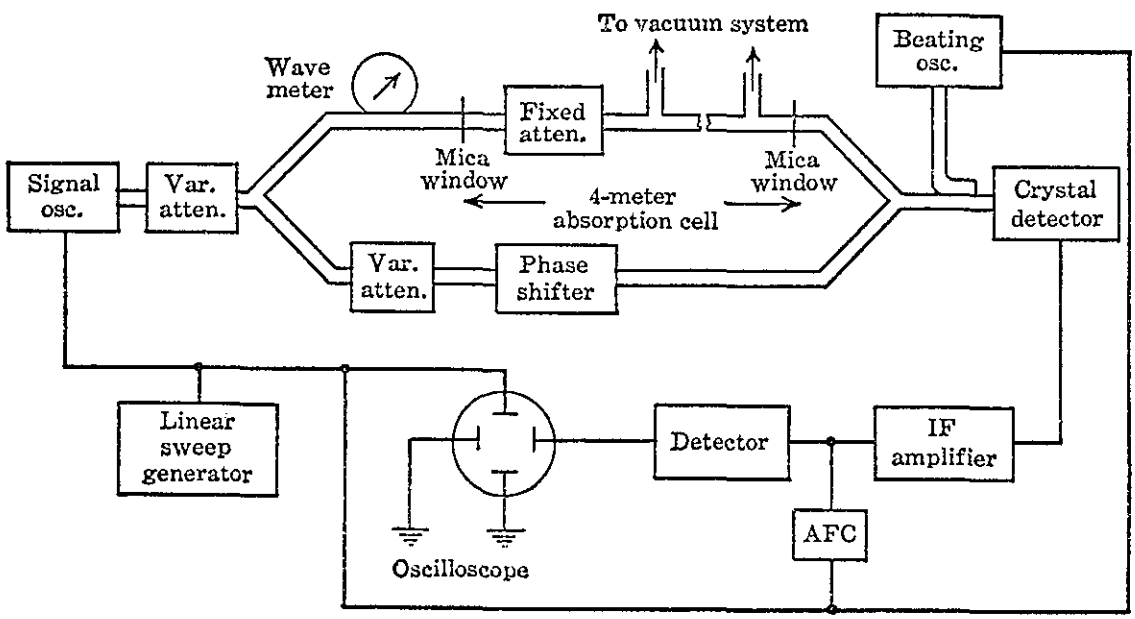


Figure 7–A typical microwave absorption experiment using a radio-frequency bridge and heterodyne detection.

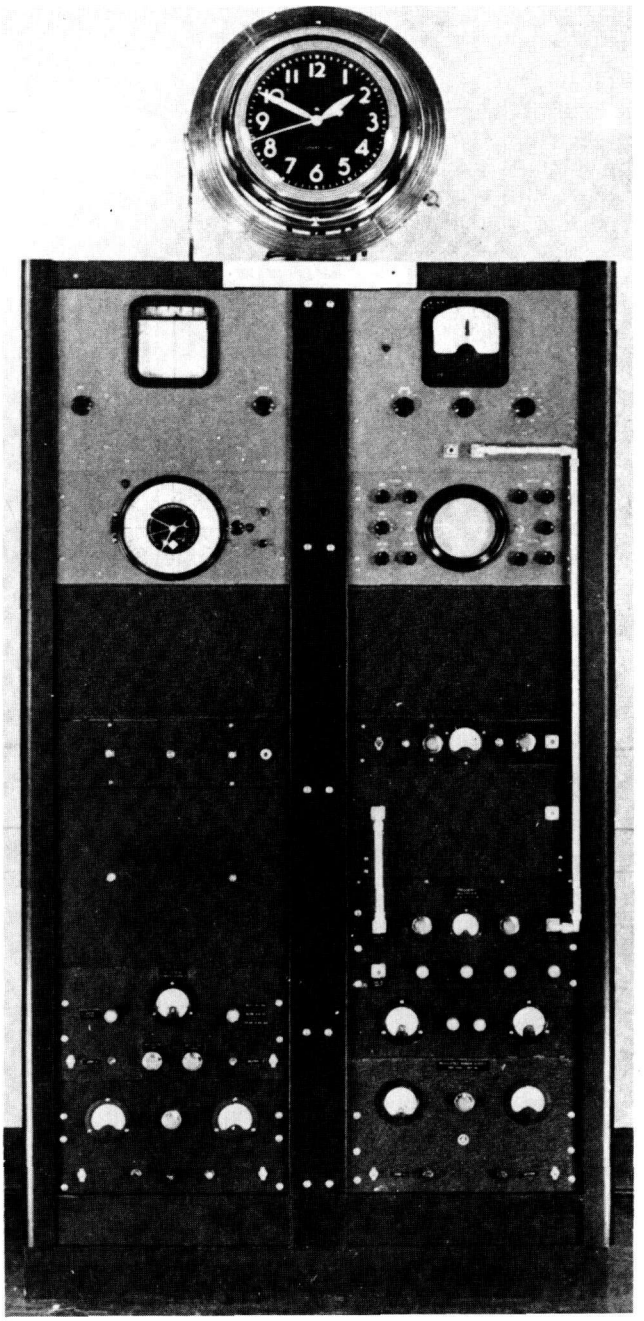


Figure 8–National Bureau of Standards ammonia clock [25]. © 1972 by the IEEE.

different atoms have been used in such gas cells, Rb^{87} soon became the favorite in most such devices. Extensive work in optically pumped gas cells for frequency control has been done at Princeton, RCA, International Telephone & Telegraph (ITT), Space Technology Laboratory, the National Bureau of Standards, Clauser Technology Corporation, Varian Associates, and many other commercial, university, and government organizations in the United States and abroad. Figure 9 shows a typical optically pumped rubidium frequency standard.

The optically pumped gas cells have the advantages of simplicity, relatively low cost, large signal-to-noise ratio, and good spectral purity. Unfortunately, the relatively large shift in frequency due to numerous buffer gas collisions is dependent on purity, pressure, and temperature. Changes in the light intensity shift due to variations in the pumping lamp intensity or spectrum may also be a problem. As a result, the stability of rubidium gas cells over a period of several months is ordinarily no better than a few parts in 10^{10}. These shifts prevent the optically pumped gas cells from being primary time and frequency standards, but the gas cells are used as frequency control devices when their accuracy is sufficient. Research is currently in progress in a number of laboratories to improve the stability of optically pumped gas cells; Bouchiat, Brossel [46], and others, for example, have eliminated the buffer gases as was done earlier in the hydrogen maser [47] and have used suitable coated walls to retain the atoms and reduce the effect of the first-order Doppler shift.

Molecular Masers

In 1951, Pound, Purcell and Ramsey [48] studied nuclear spin systems with inverted populations and noted that such systems in principle were intrinsic amplifiers rather than absorbers. The first suggestions actually to use systems with inverted populations as practical amplifiers or oscillators were made independently in the early 1950's by Townes [49] and independently but somewhat later by Weber [50] and Basov and Prokhorov [27]. The first such amplifier was successfully constructed in 1954 by Gordon, Zeiger, and Townes [49] and called a maser (Microwave Amplifier by Stimulated Emission of Radiation). The device used inhomogeneous electric fields to focus the higher energy molecular inversion states of ammonia molecules in a molecular beam. These molecules then emitted coherent stimulated emission radiation in passing through a cavity tuned to the 24-GHz ammonia inversion transition. A schematic diagram of the first ammonia maser is shown in figure 10. A report by Gordon on the new ammonia maser was a major attraction at the special meeting on atomic and molecular resonances sponsored by the Signal Corps Engineering Laboratory in 1956. In that year Bloembergen [29] proposed the three-level solid-state maser and in 1958 Townes and Schawlow [30] pointed out the possibility of masers at the infrared and optical frequencies.

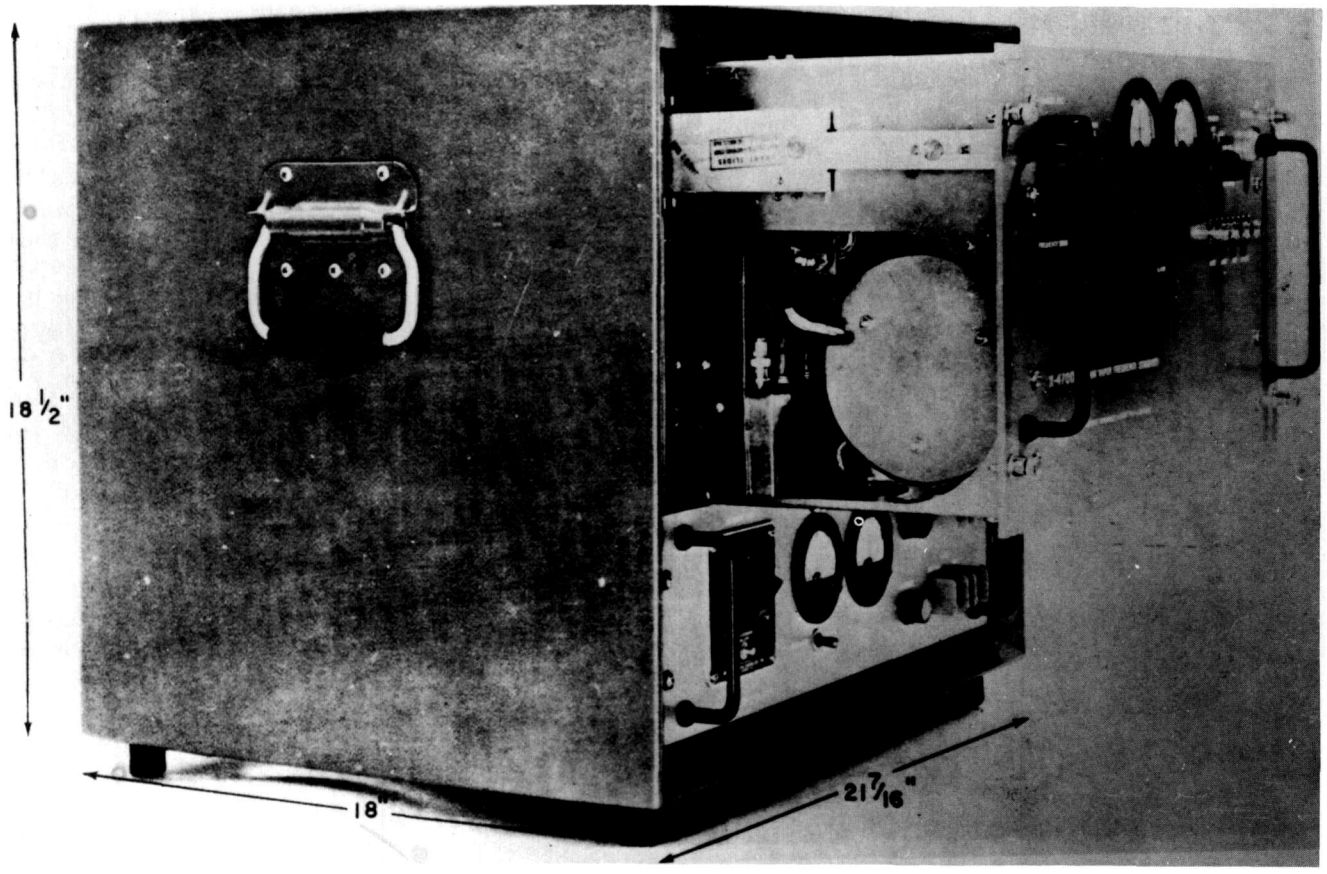


Figure 9–Rubidium frequency standard. © 1972 by the IEEE.

Since the announcement of the first successful ammonia maser in 1955 there has been tremendous research and development activity by scientists and engineers in many countries. Masers at infrared or optical frequencies (lasers) have great potential for frequency control. Further discussions of lasers will be deferred to a later section of this report. Molecular maser developments for the purposes of time and frequency control soon became intense and went in many directions including the search for more suitable molecules than ammonia, the development of two cavity masers analogous to the separated oscillatory field method [20] for molecular beams, the use of ammonia of different isotopic composition, and so forth. A value of the $N^{15}H_3$ frequency accurate to 5 parts in 10^{11} has been obtained by de Prins and confirmed by Barnes [35]. However, after a few years of intense molecular maser activity, the interest in such masers for time standards and frequency control waned since the molecular masers on the one hand lacked the simplicity and low cost of optically pumped rubidium gas cells and on the other hand lacked the high stability of either atomic cesium beams or atomic hydrogen masers.

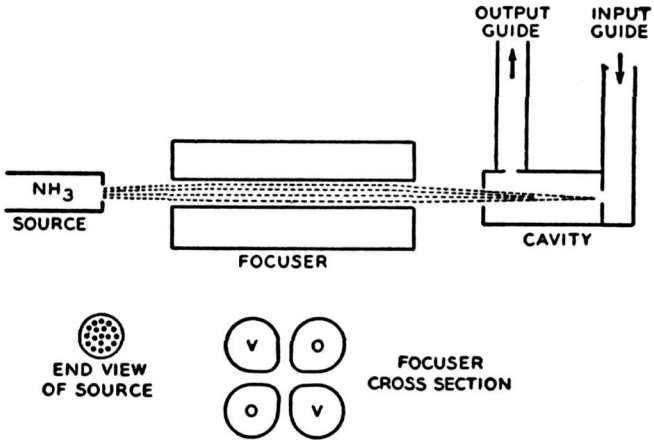


Figure 10–Schematic diagram of original ammonia maser [49].

Atomic Masers

In 1957 Ramsey [35] proposed to increase the accuracy of the atomic beam magnetic resonance method by retaining the atoms for a much longer time

between the two separated oscillatory fields, thereby obtaining much narrower resonances. His first thought was to confine the atoms with inhomogeneous magnetic fields in a large ring. However, it soon became apparent that the inhomogeneous confining magnet fields, which acted on the atoms for long periods of time, would hopelessly broaden the resonances. In fact, it became clear that the frequencies would be much less perturbed by a confinement force that was present for only a short fraction of the time even though the force might be stronger when it was applied. The obvious limit of such a device was confinement of atoms in a box with suitably coated walls. Although many wall bounces would be required to achieve marked narrowing of the resonance by a long storage time, the first experiment involved only a few wall collisions, since most scientists at that time believed that even atoms in the S state would undergo hyperfine transitions at a single wall collision. The first experiment of Kleppner, Ramsey, and Fjelstadt [51] involved only a few wall collisions and the experiment was appropriately called a "broken atomic beam resonance experiment." Cesium atoms and a Teflon-coated [52] wall were used in these first experiments.

Goldenberg, Kleppner, and Ramsey [53] then made an atomic beam resonance apparatus that stored atoms of cesium for a longer time and investigated alternate wall-coating material. They found that when the storage bulb was coated with a paraffin-like substance called Paraflint [54,53] resonances could be observed after as many as 200 wall collisions. It was recognized that atomic hydrogen would probably be a more suitable atom than cesium because of the low electric polarizability and the low mass of hydrogen, but cesium could be much more efficiently detected than hydrogen.

Kleppner and Ramsey [51,53,47] therefore proposed detection of the emitted radiation rather than of the atom. In particular, they noted that atoms of hydrogen in the higher energy hyperfine state could be focused into a suitably coated storage bulb by a six-pole magnet while atoms in the lower state would be defocused. They showed that if such a storage bulb were surrounded by a microwave cavity tuned to the 1420 MHz hyperfine transition frequency, then maser oscillation should occur. In 1960, Goldenberg, Kleppner, and Ramsey [47] constructed and operated the first atomic hydrogen maser. This apparatus is shown in figure 11. Although the total microwave power was small—approximately 10^{-12} W—the stability was so high that the output was concentrated into an extremely narrow band with a consequently favorable signal-to-noise ratio.

Although the first hydrogen maser used a wall coating of Paraflint or of Dri-Film [55] (dimethyldichlorosilane [53]), it was soon found that with atomic hydrogen, in contrast to cesium, Teflon-coated [52] walls gave longer storage times and smaller frequency shifts from wall collisions [56]. Bender [57] soon pointed out that spin exchange collisions of hydrogen atoms could not be neglected and might produce a significant frequency shift, but Crampton [58] noted that the normal tuning technique would cancel out such an effect. Later Crampton [59] pointed out the existence of a smaller additional spin exchange effect that would not be cancelled by the normal tuning method. This effect was omitted in earlier theories due to their neglect of the contribution of the hyperfine interaction during the short duration of the collision. Crampton [59] developed a technique for measuring the spin exchange effect. Crampton also pointed out the existence of a small frequency shift [59] due to magnetic field inhomogeneities: this small shift is often called the Crampton effect. Both of these effects are so small they did not affect past measurements and they can be further reduced by suitable apparatus design. Kleppner, Ramsey and their associates have developed and reviewed the theory and techniques for the atomic hydrogen maser [56,60].

A commercial hydrogen maser [60] was developed by Vessot, Peters, Vanier, McCoubrey, Levine and Cutler. The work was started at Bomac and successively transferred to Varian Associates and Hewlett-Packard. The H-10 maser developed by Vessot and his associates is shown in figures 12 and 13. Since then hydrogen masers have been developed and constructed at the Smithsonian Astrophysical Observatory by Vessot and his associates; at the Goddard Space Flight Center by Peters, Reinhardt, and others; in France by Audoin [61] and Grivet [35] in Canada by Vanier and by scientists at the NRC; in Switzerland by Kartaschoff [35]; at Harvard by Ramsey, Kleppner, Crampton, Myint, Brenner, Uzgiris, Zitzewitz, Debely, and Larson; at the National Bureau of Standards by Hellwig; at Williams College by Crampton; at the Jet Propulsion Laboratory; at the firm, ASUAG; at Hughes Aircraft by Wang; at Efratom; at the Applied Physics Laboratory; in the USSR at the Lebedev Institute and the Moscow State University; and in the Peoples Republic of China at Shanghai and Shianxi Observatory and at the National Institute of Metrology [62].

The hydrogen maser eliminates first-order Doppler shifts and photon recoil effects by virtue of the

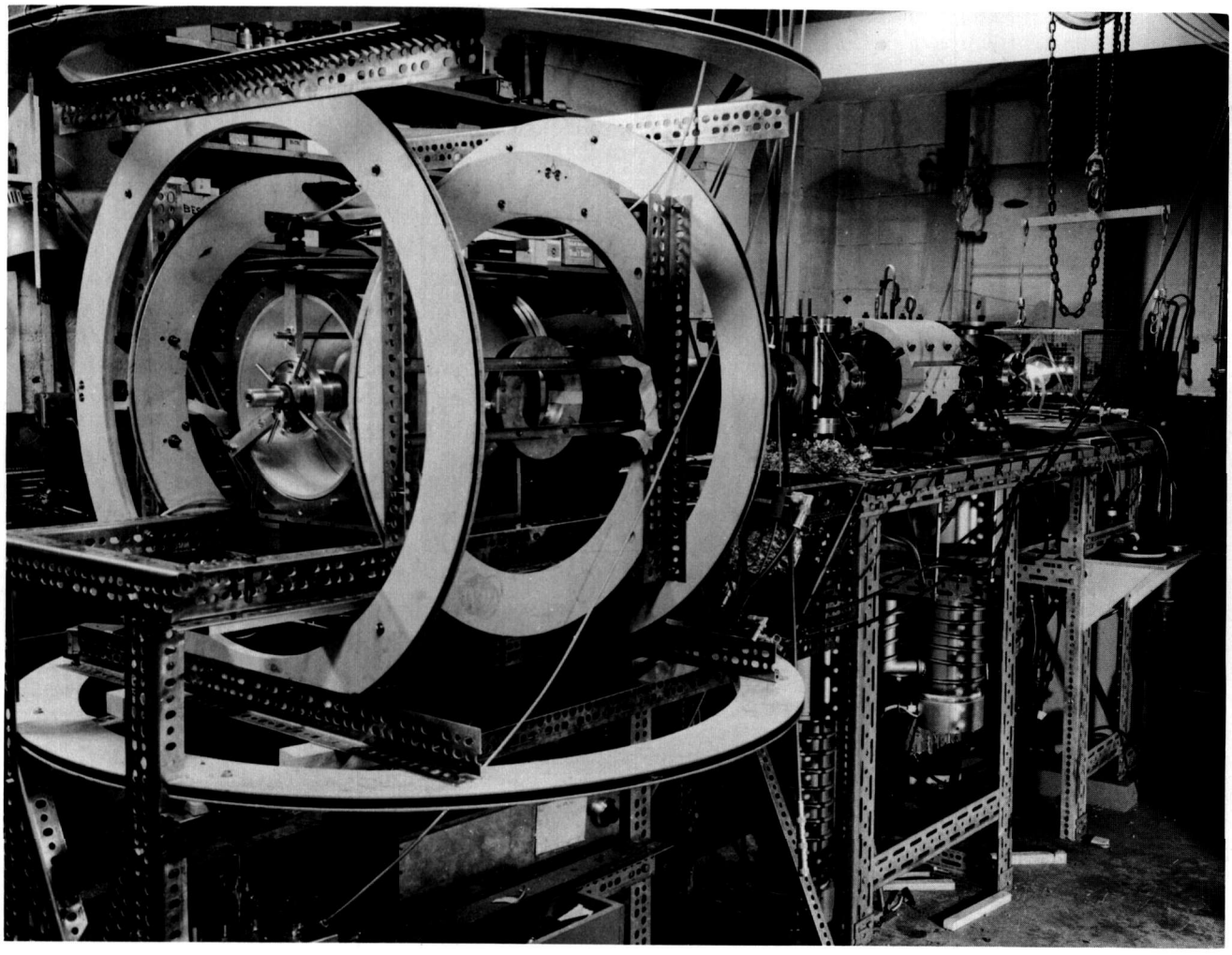


Figure 11–Original hydrogen maser [47]. The large coils are to cancel external magnetic fields. In later hydrogen masers, these were replaced by two or three concentric cylinders of high-permeability magnetic shielding.

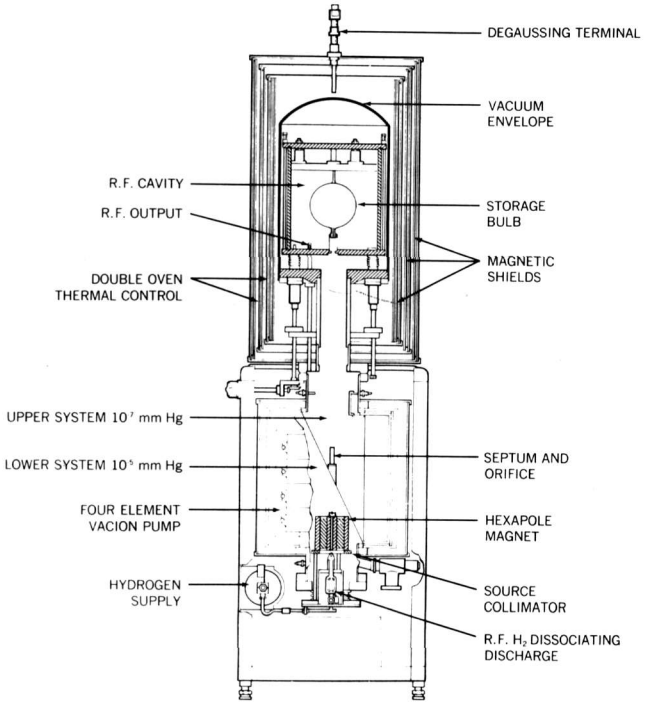


Figure 12–Schematic diagram of a commercial hydrogen maser developed by Vessot and his associates [60]. ⟶

confinement of the atoms in a box where the *average* velocity is essentially zero and by absorption of recoil momentum by the confining box. The hydrogen maser also benefits from the relatively long storage time with a resulting narrow resonance line and from the low noise characteristic of maser amplification. It shares with most other atomic or molecular frequency standards the need for correcting for the small second-order Doppler shift.

The chief disadvantage of the hydrogen maser for time and frequency control has been the existence of a

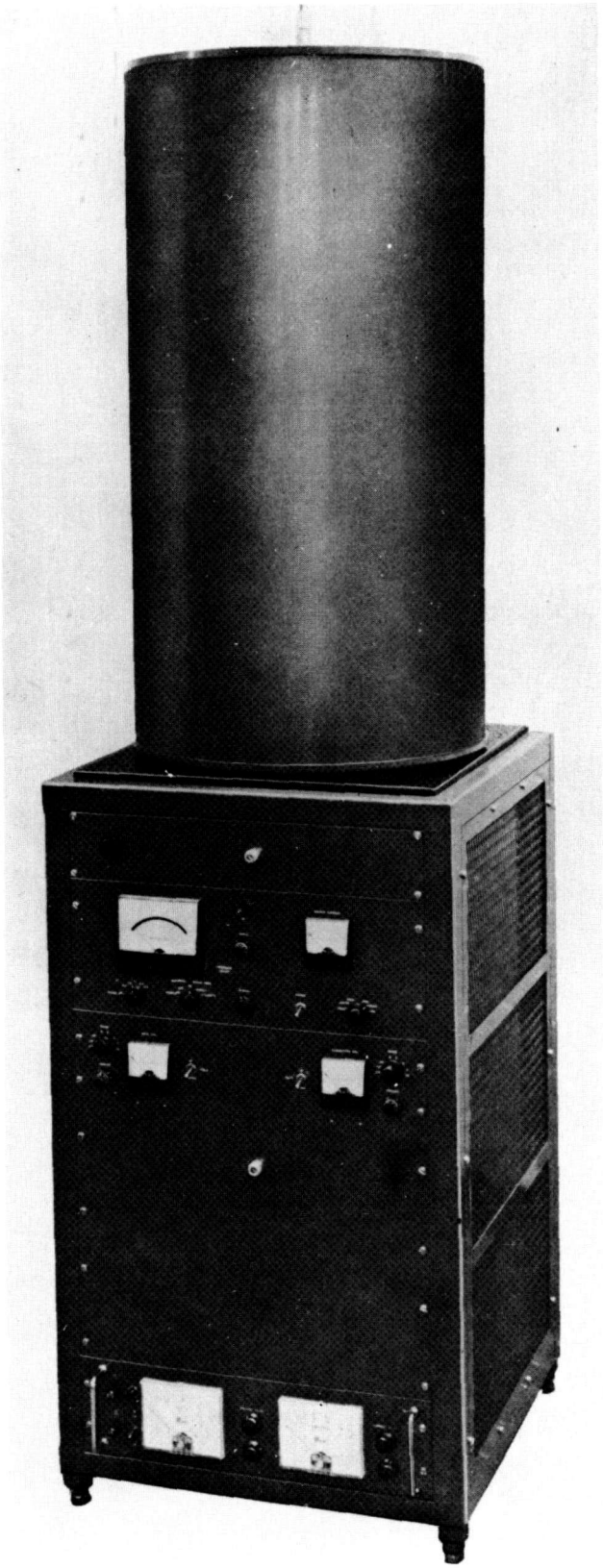


Figure 13–Commercial hydrogen maser [60]. © 1972 by the IEEE.

small frequency shift due to collisions of the atoms with the Teflon-coated walls of the storage bulb. With a 16-cm-diameter bulb this wall shift is about 2 parts in 10^{11} and can be measured by using bulbs of two different diameters. Until recently the measurements of the wall shifts have been limited to accuracies of a few percent by variations in different wall coating. However, Uzgiris and Ramsey [63] at Harvard have reduced the wall shift by a factor of 10 by the use of an atom storage vessel 10 times larger in diameter (1.5 μm). In the same laboratory, Brenner [64] and Debely [65] have developed a technique to measure the wall shift in a single storage bulb by making a change in the bulb's volume by deforming its shape. Since a single bulb is used in this method, it is free from the uncertainties in the nonreproducibility of the wall coatings of different bulbs. This method was first used on hydrogen masers with normal-size storage vessels, but Reinhardt [66] later applied it to a maser with a large storage vessel as well. Zitzewitz [67] has shown that at a temperature of about 80 °C the wall shift passes through zero; it is thus possible to operate the hydrogen maser at a temperature such that the wall shift vanishes and to select this temperature by the deformable bulb technique. Recent experiments with spin-polarized hydrogen atoms confined to a low temperature chamber covered with a helium film [68–71] suggest the possibilities of such confining chambers with atomic hydrogen masers. With these new methods, absolute accuracy better than 1 part in 10^{13} should be attained. Stabilities better than 1 part in 10^{15} have been achieved [72].

Although the hydrogen maser is the most stable atomic maser over long periods of time, Novick, Vanier, and others [35] have developed a high-power optically pumped atomic Rb^{85} maser whose relatively high output power is useful because it provides good short-term stability over periods of 10 seconds or less.

Hydrogen masers have been extensively used in long baseline interferometry in radio astronomy. Vessot and his associates [73] have used a hydrogen maser in a high altitude rocket to confirm the relativistic gravitational red shift to an accuracy of 0.007%.

Lasers

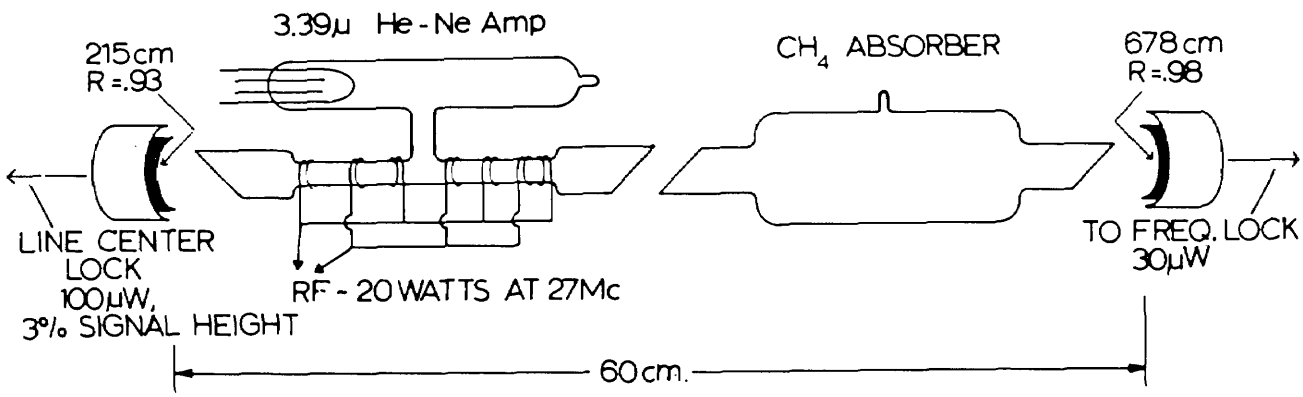
Townes and Schawlow [30] pointed out that masers could be produced at infrared and optical frequencies. The first optical maser or laser was successfully made from ruby by Maiman [74]. Subsequently there was a great burst of activity in this field and lasers were made of a wide variety of materials, some of which

permitted operation at high pulsed power. From the point of view of frequency control the laser using a helium-neon gas mixture developed by Javan [75] and his associates was the first one of interest as a time standard because of its potential stability.

As absolute time standards, most early lasers suffered from the fact that the output frequency was primarily determined by the distance between two mirrors since the first-order Doppler broadening of the atomic or molecular resonance exceeds the resonance width of the interferometer. This characteristic contrasts with a microwave maser where the frequency is determined primarily by the atomic transition with only a relatively small amount of pulling from mistuning of the microwave cavity. However, various methods for diminishing the first-order Doppler spread and thereby for determining the laser frequency more by atomic or molecular properties have been developed. These methods usually depend upon nonlinear effects.

One method that has been particularly effective is laser-saturated absorption spectroscopy developed by J. L. Hall [76,77], Schawlow [78], Hansch [78], and others [79,80,81–84]. In such a device, laser light is passed in opposite directions through, say, a CH_4 or I_2 absorption cell. There is a minimum of absorption at a frequency corresponding to no first-order Doppler shift, since stationary molecules absorbing at that frequency absorb the light from both directions equally well and hence are more readily saturated than are the moving molecules which respond at most to light from a single direction. A schematic view of a laser-saturated absorption device is shown in figure 14. An absorption cell containing methane is included in the optical path between the two mirrors of a helium-neon laser so the methane molecules are subjected to two opposite beams of light. Stabilities of a few parts in 10^{14} have been achieved by Hall and others [76] with 3.39 μm He-Ne laser-saturated absorption in CH_4 but the reproducibility was only about 1 in 10^{11}. Although this technique markedly reduces first-order Doppler broadening, it does not automatically remove all shifts associated with molecular recoil. Also power shifts and second-order Doppler effects remain. A combination of saturated absorption with an atomic beam of calcium has recently given encouraging results [85].

Double-resonance [80,86,87] and two-photon



$$(\Delta\nu_{HWHM})^2 = \left\{ \left[70\text{kc} + (16\text{kc}/_{m\,\text{Torr}}) \times P_{CH_4} \right]^2 + \left[.5\text{kc}/_{\mu W} \times P_{ext.} \right]^2 \right\}$$

$$\text{so } \Delta\nu_{HWHM} \sim 240\text{kc at } 300^\circ\text{K, 10 milli Torr \& 100}\mu\text{W}$$

$$S/N\ (1\,\text{sec}) \sim \frac{3\,\mu\text{W signal}}{1.0 \times 10^{-11}\,\text{W noise}} = 3 \times 10^5\ (=109\,\text{Db!})$$

$$\delta f = 2.3\ \frac{\Delta\nu_{HWHM}}{S/N} \simeq 1.9\,\text{Hz RMS}$$

$$\delta f/f \simeq 2 \times 10^{-14} \times \frac{1}{\sqrt{T(\text{sec})}}$$

Figure 14–Laser-saturated methane frequency reference [76].

Doppler-free absorption spectroscopy [88] eliminated first-order Doppler broadening by requiring the absorption of two photons. If these photons come from opposite directions and are at different frequencies appropriate to an intermediate real energy level, the different first-order Doppler shifts would prevent simultaneous absorption of both photons except for the absorption by molecules moving with approximately zero velocity along the direction of the laser beam, since for these molecules the first-order Doppler shifts are approximately zero. Two-photon Doppler-free absorption spectroscopy is particularly effective when the two photons moving in opposite directions are at the same frequency, even though in this case the intermediate transition is to a virtual level since it is unlikely that a real level will fall exactly halfway between the initial and final states. Since the Doppler shift in one direction is equal and opposite to that in the opposite direction, the sum of the two frequencies is independent of the molecular velocity, so molecules at all velocities can contribute to the two-photon Doppler-free spectrum. Since the two photons move in opposite directions with equal momentum, there is no recoil of the molecule and hence no Doppler or recoil broadening. High laser power levels, however, may be required so power shifts may be a problem but they can be reduced with a suitable experimental arrangement. In common with most other methods, the second-order Doppler shift is not eliminated in two-photon spectroscopy.

A major advance in recent years has been the development of frequency multiplying techniques in the optical region by Javan [75]. Evenson and his associates [81,91], Hall and Borde [77] and others. With these techniques it is possible to compare laser frequency standards with the cesium beam standard used in the definition of the second. With such devices, both the frequency and the wavelength of CH_4(and CH_2) stabilized lasers have been measured and thereby a precision value for the velocity of light of 299 792 458.3$\pm$1.2 m/s has been obtained [76,89,90]. This work has now become so accurate that it is expected soon to become the basis for the definition of the meter. The Advisory Committee to the General Conference on Weights and Measures has recommended for the new definition of the meter that "the meter is the length equal to the distance travelled in a time interval 1/299,792,458 of a second by plane electromagnetic waves in a vacuum." When this recommendation is adopted the unit of length will be defined in terms of the second, which in turn is determined by atomic clocks. The velocity of light will cease to be a measurable quantity and instead will be determined by definition.

Trapped Ions

Dehmelt [91] in 1959 first used electromagnetic ion traps in radio-frequency resonance studies. The intrinsic width of the resonances as determined by the uncertainty principle can be very narrow since the ions are retained in the apparatus for very long periods of time. However, until recently these devices have suffered from the relatively high velocity of the ions in the trap (approximately of 1 eV of kinetic energy) with the correspondingly larger broadening due to the second-order Doppler shift. Initial efforts by Dehmelt et al. [92] to diminish this were only partially successful and trapped-ion devices have not as yet provided frequencies as stable as those of the best alternative frequency standards. However, as will be discussed in the section on Doppler broadening, Wineland and Dehmelt [93] have proposed an ingenious technique for resonant radiation cooling of trapped ions and have shown [94,95] that spectacular line narrowing can be achieved. The resonance widths achieved correspond to temperatures below 50 mK [94,95]. If this technique is fully successful, trapped-ion resonance devices should become highly promising frequency standards, although their stability is degraded by the low signal-to-noise ratio which results from the space charge limitation on the number of ions that can be studied simultaneously. Ion recoil ordinarily causes no difficulty since the recoil momentum is absorbed by the trapping field.

Superconducting Cavities

High-stability superconducting cavity oscillators have recently been made by Stein and others [96] at Stanford University. Although such oscillators do not strictly come within the scope of this report, their stability, especially for short times, is sufficiently great that they should be discussed here at least briefly even though they are not suitable as absolute standards since the frequency depends on cavity dimensions instead of a characteristic atomic or molecular transition frequency. A schematic view of such a superconducting cavity is shown in figure 15. Stabilities of the order of 10^{-15} have been attained over 100 second intervals or more. Since short-term stability increases with oscillator power and since superconducting cavities can be operated at a

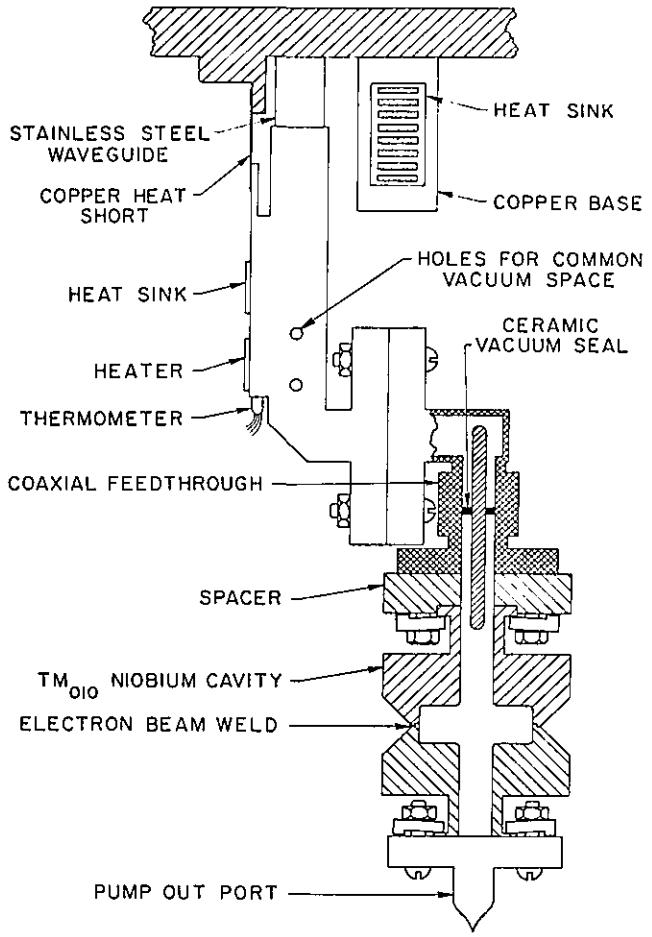


Figure 15–View of superconducting cavity showing the mounting [20].

relatively high power level, they have particularly favorable short-term stability.

Doppler Broadening

The atoms or molecules in atomic and molecular frequency standards are in thermal motion and hence subject to both first- and second-order Doppler shifts or broadening. The first-order Doppler shift—the familiar increase in the frequency received from an approaching radiation source—is proportional to $v/c \sim 3 \times 10^{-7}$ so any competitive frequency standard must provide a means for eliminating the first-order Doppler shift. Consequently, the essential features of the different frequency standards can most simply be given by describing the way that each one eliminates both frequency shifts and resonance broadening from first-order Doppler shifts.

In cesium and molecular beam devices the first-order Doppler shift is eliminated by the use of two separated oscillatory fields of coherent radiation of the same phase. In the hydrogen maser, the first-order Doppler shift is eliminated by confining the hydrogen atoms to a small volume which is traversed many times during the radiation process of each atom so the velocity averages to zero. In trapped-ion spectroscopy, the first-order Doppler shift is eliminated for the same reason. With laser-saturated molecular absorption devices, double-resonance spectroscopy, and two-photon spectroscopy, the first-order Doppler shift is removed by the requirements of the absorption of two or more photons moving in opposite directions, as discussed in the sections on these devices.

However, even after the first-order Doppler shift is eliminated, there remains in atomic and molecular oscillators a second-order Doppler shift whose magnitude is of the order $(v/c)^2 \sim 10^{-13}$. If much progress is to occur beyond the present accuracy of a few parts in 10^{13}, means must be found to reduce the magnitude of the second-order Doppler shifts, i.e., to reduce the velocities. New possible techniques for reducing the magnitudes of the velocities have been proposed by Hansch and Schawlow [97] and by Wineland and Dehmelt [93]. The basic idea is to cool, say, trapped ions by shining on them intense laser light at a frequency slightly below the resonance frequency. This light can be absorbed by an ion whose motion provides the appropriate first-order Doppler shift. The subsequent emission, however, is in all directions and hence on the average at the normal resonance frequency. By conservation of energy the ion must therefore lose kinetic energy. In this fashion the trapped ions can be cooled by many successive absorptions and reemissions. Dehmelt [94] Wineland [95] and their associates have already demonstrated impressive line narrowing by the laser cooling technique. It will be of great interest during the coming years to see if these techniques for reducing the second-order Doppler width lead to marked increases in the accuracy of clocks and frequency standards.

Accuracy, Reproducibility, and Stability

In discussions of time and frequency standards it is necessary to distinguish between three different but related properties of the standards: accuracy, reproducibility, and stability. Accuracy measures the degree to which a standard independently agrees with the value specified in the definition of the unit of time.

Reproducibility is a measure of the extent to which properly adjusted independent devices of the same design agree. Stability is a measure of the degree to which the same device gives the same result in successive intervals of time. The stability is conventionally measured by the parameter $\sigma_y(\tau)$ which is the square root of the two-sample Allan variance [98] for adjacent samples which in turn is one-half of the mean square of the fractional differences of the frequencies measured in adjacent intervals of time duration τ.

For different applications, different characteristics are the most relevant. Thus, for absolute standards of frequency, the accuracy is the most important property. On the other hand, for many measurements, such as long baseline interferometry in radio astronomy, stability is of primary concern.

The stability parameter $\sigma_y(\tau)$ is plotted as a function of the time interval τ for a number of different kinds of oscillators in figure 16.

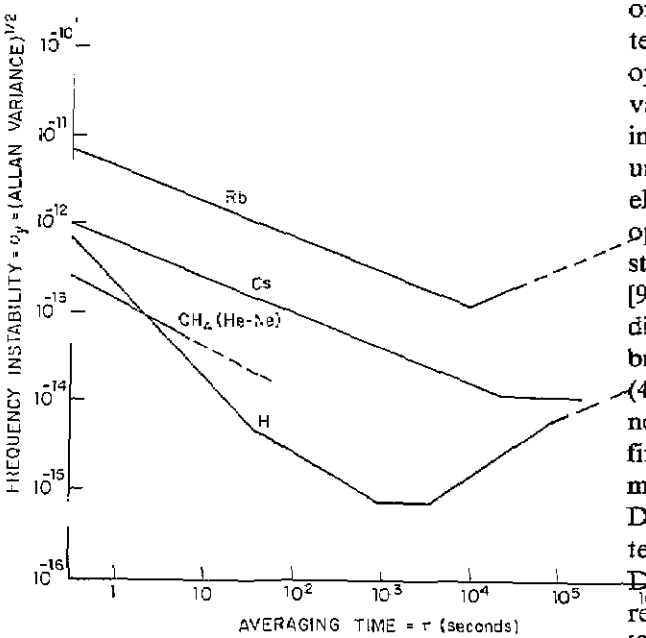


Figure 16-Stabilities of various atomic clocks [99,96,36,72].

The need for accurate and stable timing has been recognized for many centuries and the development of better clocks has been vigorously pursued throughout that time. However, the truly spectacular advances in that field have occurred only in the past few decades, as is illustrated in figure 17, which shows the development of the accuracy of timing through history.

Future Prospects

Although atomic and molecular frequency and time standards have been a reality for a number of years, developments are occurring at a relatively rapid rate. As a result it is impossible to forecast reliably the future developments that will lead to the most major subsequent advances. However, a number of prospective developments for the different devices are included in the above discussions of these devices. For highest stability and reproducibility the most promising of these prospects are probably the following: (1) Further improvements on the existing atomic and molecular beam methods such as the widely used cesium frequency standard, including better velocity definition to reduce uncertainties due to second-order Doppler shifts and including possible new molecules with higher frequency resonances. (2) Hydrogen maser improvements including new confinement surfaces such as low temperature helium or neon films and combinations of the deformable bulb techniques with either the large box maser or operation at a temperature where the wall shift vanishes. The use of electronic cavity tuning provides increased stability, and interesting studies have been undertaken at the National Bureau of Standards and elsewhere of the use of atomic hydrogen as a passively operating frequency standard [99]. (3) Improved stored-ion devices, especially the new techniques [93,94,95] for cooling the trapped ions to markedly diminish the first- and second-order Doppler broadening. Development of monoion oscillators [85]. (4) The use of lasers especially when combined with nonlinear spectroscopy techniques which eliminate first-order Doppler broadening, such as saturated molecular absorption and especially two-photon Doppler-free spectroscopy; the use of laser cooling techniques [97]. (5) The reduction of the second-order Doppler broadening for any of the methods by the resonance cooling techniques discussed earlier [93,97,94,95]. (6) Improvements in frequency multiplying techniques to connect the microwave and optical regions. (7) Improvements in superconducting cavity oscillators. (8) Combinations of various techniques such as saturated absorption laser spectroscopy with atomic beams or use of a superconducting cavity as a slave oscillator for an atomic resonance device. (9) For durations longer than a month measurements on selected pulsars may provide high stability.

If past precedents are followed, there will in the

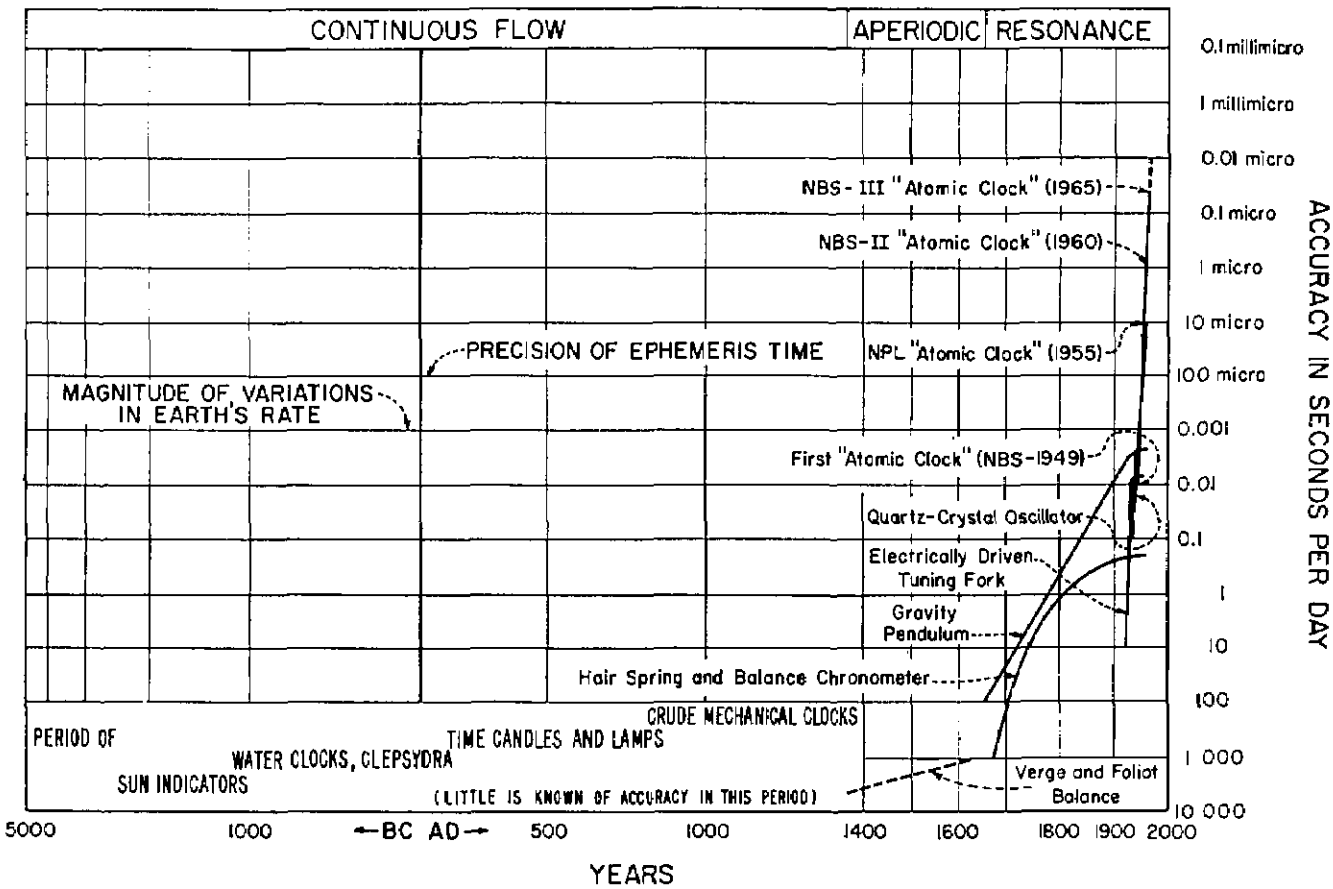


Figure 17–The accuracy of timing through history [25].

future also be many unexpected new ideas and developments that drastically improve existing techniques or lead to totally new kinds of highly stable clocks.

References

[1] Dunoyer, L. Comptes Rendu **152**, 594 (1911) and L. Radium **8**, 142 (1911).

[2] Stern, O. Z. Phys. **39**, 751 (1926); Knauer, F. and Stern, O., Z. Phys. **39**, 764 and 780 (1926).

[3] Darwin, C. Proc. R. Soc. **117**, 258 (1927).

[4] Phipps, T. E.; Stern, O. Z. Phys. **73**, 185 (1931).

[5] Guttinger, P. Z. Phys. **73**, 169 (1931).

[6] Majorana, E. Nuovo Cimento **9**, 43 (1932).

[7] Frisch, R. O.; Segre, E. Z. Phys. **80**, 610 (1933).

[8] Rabi, I. I. Phys. Rev. **49**, 324 (1936).

[9] Motz, L.; Rose, M. Phys. Rev. **50**, 348 (1936).

[10] Rabi, I. I. Phys. Rev. **51**, 652 (1937).

[11] Schwinger, J. Phys. Rev. **51**, 645 (1937).

[12] Rabi refers to the visit of Gorter in a footnote to the first paper on the molecular beam magnetic resonance method [10] and Gorter 29 years later published an article giving his own somewhat different recollections of the same visit [Physics Today **20**, 76 (Jan. 1967).]

[13] Gorter, C. J. Physica **3**, 503 and 995 (1936).

[14] Rabi, I. I.; Zacharias, J. R.; Millman, S.; Kusch, P. Phys. Rev. **53**, 318 (1938) and **55**, 526 (1939).

[15] Kellogg, J. M. B.; Rabi, I. I.; Ramsey, N. F.; Zacharias, J. R. Phys. Rev. **55**, 729 (1939); **56** 728 (1939); and **57**, 677 (1940).

[16] Rabi, I. I. Phys. Rev. **67**, 199 (1945).

[17] Kusch, P.; Millman, S.; Rabi, I. I. Phys. Rev. **57**, 765 (1940).

[18] Millman, S.; Kusch, P. Phys. Rev. **57**, 438 (1940).

[19] Kusch, P.; Taub, H. Phys. Rev. **75**, 1477 (1949).

[20] Ramsey, N. F. Phys. Rev. **76**, 966 (1949); *Molecular Beams*, New York; Oxford, 1956 (1969); and IEEE Trans. Instrum. Meas. **IM-21**, 90 (1972).

[21] Ramsey, N. F.; Silsbee, H. B. Phys. Rev. **84**, 506 (1951).

[22] Lundeen, S. R.; Jessop, P. E.; Pipkin, F. M. Phys. Rev. Lett. **34**, 377 and 1368 (1975).
[23] Kolsky, H. G.; Phipps, T. E.; Ramsey, N. F.; Silsbee, H. B. Phys. Rev. **80**, 483 (1950).
[24] Sherwood, J. E.; Lyons, H.; McCracken, R. H.; Kusch, P. Bull. Am. Phys. Soc. **27**, No. 1, 43 (1952).
[25] Lyons, H. Ann. N.Y. Acad. Sci. **55**, 831 (1952) and Sci. Am. **196**, 71 (Feb. 1957).
[26] Zacharias, J. R. private communication and Phys. Rev. **94**, 751 (1954). (R. Weiss and R. Vessot were associated with Zacharias in the experimental work on the "Fountain" experiment.)
[27] Zacharias, J. R.; Yates, J. G.; Haun, R. D. M.I.T., Res. Lab. Electron., Cambridge, MA, Quart. Prog. Rep. 30, Jan. 1955, and "An Atomic Frequency Standard," Proc. IRE (Abstract) **43**, 364 (Mar. 1955).
[28] Basov, N. G.; Prokhorov, A. M. Zh. Eksp. Teor. Fiz. **27**, 431 (1954) and **28**, 249 (1955); or JETP Lett. **1**, 184 (1955).
[29] Bloembergen, N. Phys. Rev. **104**, 324 (1956).
[30] Schawlow, A. L.; Townes, C. H. Phys. Rev. **112**, 1940 (1958).
[31] Baird, J. K.; Miller, P. D.; Dress, W.; Ramsey, N. F. Phys. Rev. **179**, 1285 (1969).
[32] Essen, L.; Parry, V. L. Nature **176**, 280, 284 (1955).
[33] Reder, F. H. "Atomic Clocks and Their Applications," USASRDL Tech. Rep. 2230 (AD 265452), 1961.
[34] Ramsey, N. F. Phys. Rev. **100**, 1191 (1954); **109**, 822 (1958); J. Phys. (Paris) **19**, 809 (1958); I. Esterman, Editor, *Recent Research in Molecular Beams*, **107**, Academic Press, New York, 1959.
[35] Proc. Frequency Control Symposia 1955–1982; IEEE Trans. Instrum. Meas. **IM-13** (1964); IEEE Trans. Instrum. Meas. **IM-15** (1966); IEEE Trans. Instrum. Meas. **IM-19** (1970); also IEEE Trans. Quantum Electron. **QE-5** (1969). R. E. Beehler, Ann. Freq. Control Symp. **25**, (1971).
[36] Hellwig, H.; Evenson, K. M.; Wineland, D. J. Physics Today **31**, 23 (Dec. 1978).
[37] Bonanomi, J. Quantum Electronics **III**, Columbia Univ. Press, New York, 1964.
[38] Cleeton, C. E.; Williams, N. H. Phys. Rev. **45**, 234 (1934).
[39] Condon, E. V.; Odishaw, H. *Handbook of Physics*, McGraw-Hill, New York, 1967.
[40] Townes, C. H.; Schawlow, A. L. *Microwave Spectroscopy*, McGraw-Hill, New York, 1955.
[41] Townes, C. H. J. Appl. Phys. **22**, 1365 (1951).
[42] Dicke, R. H. Phys. Rev. **89**, 472 (1953).
[43] Bitter, F. Phys. Rev. **76**, 833 (1949), and Pryee, M. H. T. Phys. Rev. **77**, 136 (1950).
[44] Brossel, J.; Kastler, A. C. R. Acad. Sci. (Paris) **229**, 1213 (1949).
[45] Kastler, A. J. Phys. (Paris) **11**, 225 (1950), and J. Opt. Soc. Am. **47**, 460 (1957).
[46] Bouchiat, M. A.; Brossel, J. Phys. Rev. **147**, 41 (1966).
[47] Goldenberg, H. M.; Kleppner, D.; Ramsey, N. F. Phys. Rev. Lett. **8**, 361 (1960).
[48] Pound, R. V.; Purcell, E. M.; Ramsey, N. F. Phys. Rev. **81**, 156, 278, 279 (1951) and **103**, 20 (1956).
[49] Gordon, J. P.; Zeiger, H. Z.; Townes, C. H. Columbia Rad. Lab. Prog. Rep., Dec. 1951; J. Commun. Eng. Japan **36**, 650 (1953); and Phys. Rev. **95**, 282 (1954); also Phys. Rev. **99**, 1264 (1955).
[50] Weber, J. "Amplification of Microwave Radiation by Substances Not in Thermal Equilibrium," Trans. IRE Electron Devices **ED-3**, 1–4 (June 1953).
[51] Kleppner, D.; Ramsey, N. F.; Fjelstadt, P. Phys. Rev. Lett. **1**, 232 (1958).
[52] Teflon is a registered trademark of the duPont Corporation for certain fluorohydrocarbons. The FEP Teflon was found to provide a more uniform coating for the hydrogen maser storage vessels.
[53] Goldenberg, H. M.; Kleppner, D.; Ramsey, N. F. Phys. Rev. **123**, 530 (1961).
[54] Paraflint is a registered trademark of Moore and Munger of New York City for certain long chain paraffins.
[55] Dri-Film is a registered trademark of the General Electric Co. for dimethyldichlorosilane.
[56] Kleppner, D.; Goldenberg, H. M.; Ramsey, N. F. Phys. Rev. **126**, 603 (1962); Berg, H. C.; Kleppner, D. Rev. Sci. Instrum. **33**, 238 (1962).
[57] Bender, P. L. Phys. Rev. **132**, 2154 (1963).
[58] Crampton, S. B. Phys. Rev. **158**, 57 (1967).
[59] Crampton, S. B.; Duvivier, J. A.; Read, G. S.; Williams, E. R. Phys. Rev. **A5**, 1952 (1972), Crampton, S. B.; Wong, H. T. M. Phys. Rev. **A12**, 1305 (1975); Bull. Am. Phys. **18**, 709 (1973) and **19**, 83 (1974).
[60] Kleppner, D.; Berg, H. C.; Crampton, S. B.; Ramsey, N. F.; Vessot, R. F. C.; Peters, H. E.; Vanier, J. Phys. Rev. **138**, A972 (1965).
[61] Audoin, C. Rev. Phys. Appl. **1**, 2 (1966) and **2**, 309 (1967); Phys. Lett. **28A**, 373 (1968); Audoin, C.; Desaintfuscien, M.; Petit, P.; Schermann, J. P. Nucl. Instrum. Methods **69**, 1 (1969); Design of a double focalization in a hydrogen maser, IEEE Trans. Instrum. Meas. **IM-17**, 351–358 (Dec. 1968). (This work utilizes a useful double-focusing method to eliminate the undesired F=1 m F=1 state from the focused beam.); Electron. Lett. **5**, No. 13 (1969); C. R. Acad. Sci. (Paris) **264**, 698 (1967) and **270**, 906 (1970); Double-resonance method for determination of level populations, IEEE J. Quantum Electron. **QE-5**, 431–434 (Sept. 1969); Haroche, S.; Cohen-Tanoudji, C.; Audoin, C.; Schermann, J. P. Phys. Rev. Lett. **24**, 861 (1970).
[62] Laboratories that have engaged in hydrogen maser studies include Harvard University, Massachusetts Institute of Technology, Bomac Laboratories, Varian Associates, Hewlett-Packard, the National Bureau of Standards, Goddard Space Flight Center, the Jet Propulsion Laboratory, U.S. Electronics Command, Hughes Research Laboratory, Laboratoire de l'Orloge Atomique, Orsay (France), PTB (Braunschweig, Germany), the National Research Council and Laval University (Canada), R. R. L. (Tokyo, Japan), LSRH (Neuchatel, Switzerland), and the Lebedev Institute (Moscow, U.S.S.R.), ASUAG, Efratom, Applied Physics Laboratory, Moscow State University, Shanghai and Shianxi Observatory and the National Institute of Metrology in China.
[63] Uzgiris, E.; Ramsey, N. F. Phys. Rev. **A1**, 429 (1970).
[64] Brenner, D. J. Appl. Phys. **41**, 2942 (1970).
[65] Debely, P. E.; Rev. Sci. Instrum. **41**, 1290 (1970).
[66] Reinhardt, V.; Lavanceau, J. Proc. Annu. Symp. on Freq. Control **28**, 379 (1974).
[67] Zitzewitz, P. W.; Ramsey, N. F. Phys. Rev. **A3**, 51 (1971).
[68] Silvera, I. F.; Walraven, J. T. M.; Mathey, A. P. M. Phys. Rev. Lett. **44**, 164 (1980) and **45**, 449 (1980).
[69] Cline, R. W.; Smith, D. A.; Greytak, T. J.; Kleppner, D. Phys. Rev. Lett. **45**, 2117 (1980).
[70] Crampton, S. B.; Greytak, T. J.; Kleppner, D. Phillips, W. D.; Smith, D. A.; Weinrib, A. Phys. Rev. Lett. **42**, 1039 (1979).
[71] Hardy, W. N.; Berlinsky, A. J.; Whitehead, L. A. Phys. Rev. Lett. **42**, 1042 (1979).
[72] Sydnor, R. L.; Kuhorle, P. F.; Kirk, A.; Endsley, E. F.; Meyer, R. F.; Zygielbaum, A. I. Final Report Hydrogen

Maser Comparison Test, Jet Propulsion Laboratory (1982).
[73] Vessot, R. F. C. Levine, M. W.; Matison, E. M.; Blomberg, E. L.; Hoffman, T. E.; Nystrom, G. V. Farrel, B. F.; Decher, R.; Eby, P. B.; Baugher, C. R.; Watts, J. W.; Teuber, D. L.; Wills, F. D. Phys. Rev. Lett. **26**, 208 (1980).
[74] Mainman, T. H. Nature **187**, 493 (1960).
[75] Javan, A.; Bennett, W.; Herriott, D. R. Phys. Rev. Lett. **6**, 106 (1961); Hocker, L. O.; Small, J. G.; Javan, A. Phys. Rev. Lett. **29A**, 321 (1969).
[76] Barger, R. L.; Hall, J. L. Phys. Rev. Lett. **22**, 4 (1969); Appl. Phys. Lett. **22**, 196 (1973); *Atomic Masers and Fundamental Constants* **5**, 322 (1976) (Plenum Press).
[77] Hall, J. L.; Borde, C. Phys. Rev. Lett. **30**, 1101 (1973).
[78] Hansch, T. W.; Levenson, M. D.; Schawlow, A. L. Phys. Rev. Lett. **26**, 946 (1971).
[79] Evenson, K. M. et al., Phys. Rev. Lett. **29**, 1346 (1972).
[80] Brewer, R. G. Science **178**, 247 (1972).
[81] Mungall, A. S. et al., Metrologia **17**, 123 (1981).
[82] Becker, G. Metrologia **18**, 17 (1982).
[83] Arditi, M. Metrologia **18**, 59 (1982).
[84] de Marchi, A. Metrologia **18**, 103 (1982).
[85] Barger, R. Z.; English, T. C.; West, J. B. Annu. Symp. on Freq. Control **29**, 316 (1975) (U.S. Army Signal Corps. Ft. Monmouth, NJ).
[86] Schlossberg, H. R.; Javan, A. Phys. Rev. Lett. **17**, 1242 (1966).
[87] Hansch, T. W.; Shahin, I. S.; Schawlow, A. L. Phys. Rev. Lett. **27**, 707 (1971).
[88] Visikenko, L. S.; Chebotaev, V. P.; Shishaev, A. V. JETP Lett. **12**, 113 (1970); Pritchard, D.; Apt, J.; Ducas, T. W. Phys. Rev. **32**, 641 (1974); Levenson, M. D.; Bloembergen, N. Phys. Rev. Lett. **32**, 645 (1974); Birahan, F.; Cagnac, B.; Grynberg, G. Phys. Rev. Lett. **32**, 643 (1974); Hansch, T. W., et al., Opt. Comm. **11**, 50 (1974).
[89] Evenson, K. M.; Wells, J. S.; Peterson, F. R.; Danielson, B. L.; Day, G. W. Appl. Phys. Lett. **22**, 192 (1973) and **20**, 296 (1972); Phys. Rev. Lett. **31**, 573 (1973).
[90] Jolliffe, B. W.; Rowley, W. R. C.; Shotton, K. C.; Wallard, A. J.; Woods, P. Z. Nature **251**, 46 (1974).
[91] Dehmelt, H. G. Phys. Rev. **109**, 381 (1959); Advances in Atomic Molecular Physics **3**, 53 (1967) and **5**, 109 (1959).
[92] Dehmelt, H. G.; Major, F.; Fortson, E. N.; Schuessler, H. A. Phys. Rev. Lett. **8**, 213 (1967); Phys. Rev. **170**, 91 (1968) and **187**, 5 (1969).
[93] Wineland, D.; Dehmelt, H. Bull. Am. Phys. Soc. **18**, 1521 (1973) and **20**, 60, 61, 637 (1975).
[94] Neuhauser, W.; Hohenstatt, M.; Toschek, P. E.; Dehmelt, H. Phys. Rev. **A22**, 1137 (1980).
[95] Wineland, D. J.; Itano, W. M.; Bergquists, J. C.; Dryllinger, R. E. Phys. Lett. **82A**, 75 (1981); Optics Letters **5**, 245 (1980) and Phys. Rev. **A20**, 152 (1980).
[96] Stein, S. R.; Turneauve, J. P. Proc. Annu. Symp. Freq. Control **27**, 414 (1973), and HPL 741, Stanford High Energy Physics Laboratory, Stanford, CA.
[97] Hansch, T. W.; Schawlow, A. L. Opt. Commun. Netherlands **13**, 68 (1975) and review by U.S. Letokhov, Comments on At. Mol. Phys. **6**, 119 (1977).
[98] Allan, D. W. Proc. IEEE **54**, 221 (1966).
[99] Hellwig, H. W. Proc. IEEE **63**, 212 (1975); Metrologia **6**, 56 (1970); and NBS Tech. Note 616, 1 (1972) and 662, 1 (1975).

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 88, No. 5, September-October 1983

Size and Refractive Index Determination of Single Polystyrene Spheres

Egon Marx and George W. Mulholland
National Bureau of Standards, Washington, DC 20234

Accepted: June 7, 1983

The intensity of the light scattered from individual dielectric spheres was measured as a function of the scattering angle, for light polarized parallel and perpendicular to the scattering plane. These sets of data were used to determine the radius and refractive index of the spheres by fitting the data to the scattering function obtained from Mie theory. The light was produced by a He-Cd laser (λ=441.6 nm). Measurements were performed on particles of six discrete sizes with radii in the range 117–1175 nm. Several different measures of the quality of fit were examined, and the least-squares fit, unweighted or weighted with a factor $\sin^2(\theta/2)$, was found to be the best. The values obtained for the index of refraction were found to be within 1% of the published bulk value 1.615. The measured radii differed by several percent from those specified by the manufacturer. Several sources of error were analyzed, and their effects were simulated in numerical experiments. The largest source of error in the instrument was a discrepancy between the actual scattering angle and the reading provided by the instrument. Less significant was the noise introduced by the motion of the particle within the laser beam. A calibration eliminated most of the error in the scattering angle. The precision and accuracy of this technique are estimated to be 0.2% and 0.9%, respectively, of the particle size for particles of a nominal radius of 457 nm.

Key words: dielectric spheres; inverse electromagnetic scattering; light scattering; Mie scattering; particle size; polystyrene latex; refractive index.

1. Introduction

Recently developed instrumentation has the capability of recording the intensity of the light scattered from a single micrometer-size particle as a function of angle. One such instrument has as its key component an electrostatic particle levitator with automatic servocontrol and a periscope detector that moves in a horizontal arc of nearly 180° [1][1]. Phillips et al. [2] reported the determination of particle radius r and refractive index n of a single polystyrene sphere with a diameter of 1099 nm. Subsequent studies of polystyrene spheres with this type of instrument include analysis of the size distribution of spheres of diameter of 796 nm [3], and measurements of size and refractive index for spheres of diameters of 1011, 794, and 600 nm [4].

Our study represents an extension of these investigations and includes the following improvements: the spectra for both vertically and horizontally polarized light are recorded, an accurate angle calibration is performed, the output is automatically converted to digital form by a high-precision A/D converter, and several sizes in a wide range—six groups of particles with radii between 117 and 1175 nm—were studied. A major effort of our study was a detailed error analysis, enabling us to make a quantitative statement regarding the accuracy of the technique. Our analysis applies specifically to the

About the Authors: Egon Marx is a physicist with the Micro and Optical Metrology Group in the NBS Center for Manufacturing Engineering. George W. Mulholland is a chemist with the Smoke and Toxic Gas Group in the NBS Center for Fire Research.

[1] Figures in brackets indicate the literature references at the end of this paper.

Differential II light scattering photometer[2], which was designed by Wyatt and Phillips, but the principles apply to such other single-particle instruments as the Gucker photometer for rapid measurement of angular scattering patterns for individual particles [5].

We describe the instrument used to measure the scattered light, the procedure used to collect the data, and the preliminary manipulation of the data to prepare it for analysis in section 2. Of special importance is the calibration of the readings of the scattering angles, which greatly affects the results of the analysis.

We discuss the data analysis procedure in section 3. Although the principle is straightforward, the effect of noise and other errors introduce a number of complications familiar in inverse scattering problems. The choice of the measure of fit, resulting in different weights being assigned to errors, influences the results significantly, especially for smaller particles that show more noise in the data.

To understand the sensitivity of the results to different errors, we conducted several numerical experiments in which theoretically computed intensities are altered with simulated errors to generate the "data" that are analyzed by the same procedures as the real data. The results of the sensitivity analysis are presented in section 4.

The experiments were performed on particles with radii between 117 and 1175 nm. We took data for a number of particles in each nominal size sample, both for horizontal and vertical light polarizations, with several repeat scans for the same particle and polarization. In section 5 we discuss the different sources of errors and compare our results to those obtained by other authors.

2. Experimental Method

2.1 General Characteristics

We use a Differential II light scattering photometer, manufactured by Science Spectrum, to measure the intensity of scattered light versus angle for a single particle. An aerosol made up of polystyrene spheres is generated by nebulizing a suspension of the spheres in distilled water; the water evaporates rapidly leaving charged spheres. The spheres are passed into the optical cell, which has two plate electrodes and a pin electrode as shown in figure 1. The pin electrode is

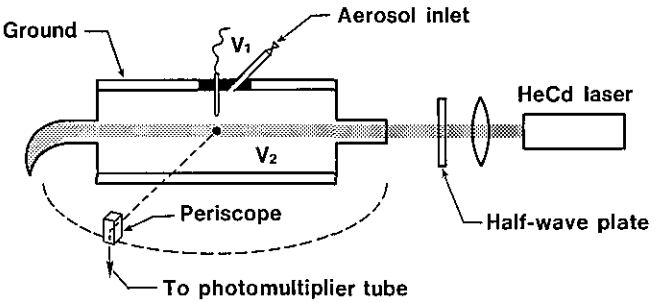


Figure 1–Single particle scattering instrument (Differential II). The pillbox-shaped scattering cell has a pin electrode (V_1) separated by an insulator from the ground plate electrode and opposite to the base plate electrode (V_2).

located just above the laser beam. We move a sphere into the field of view by use of a hydrostatic bellows, turn on the voltage to the pin electrode, and "walk" the particle to the center of the laser beam by adjusting the differential voltage between the pin and base plate. Catching a particle turns out to be analogous to playing a pinball game in terms of the required manipulations and, especially, in regard to the psychological mood that evolves. Once the particle is in the center of the laser beam, we engage the servo-control that keeps the particle at a fixed location by comparing the intensity of the acattered light reflected by a mirror edge with that transmitted above; the two intensities differ when the particle is at the edge of the beam. A more detailed account of the instrument design can be found in the literature, see ref. [1].

The scattered light reaches the photomultiplier with an S-21 response via a traversing periscope, which has an acceptance angle of about 2°. Data are typically recorded over the range of 20° to 160°. The stepping motor is operated at the highest scanning speed, which corresponds to about 1 scan in 50 s. The output of the angle encoder and the photomultiplier tube is displayed on an x-y recorder and is also processed by a Hewlett-Packard scanner, A/D converter, and an HP 9845T desktop calculator. The switching rate of a scanner used to feed the intensity and angle data to a single A/D converter limits the data acquisition rate to a pair of readings every 0.3 degrees. The data pair closest to each integer angle is stored on magnetic tape.

A He-Cd laser (λ=441.6 nm) is used as the light source. The short wavelength is an advantage when

[2] Certain commercial equipment, instruments, and materials are identified in this paper to specify adequately the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

scattering experiments from submicron particles are performed, because the scattered intensity is a function of r/λ and higher values of r/λ increase the structure in the scattering pattern. We find that inserting a 15-cm focal length coated achromat 20 cm from the pin electrode increases the scattered light intensity and reduces the stray light reflection from the entrance tube to the scattering cell and from the light trap. This choice of lens is actually a compromise between ease of catching the particle (no lens at all is best) and optimum signal-to-noise ratio (the laser beam should be focused near the particle). We have also inserted a half-wave plate to allow measurements of the scattering intensity with the plane of polarization of the light perpendicular to the scattering plane as well as in the same plane (vertical and horizontal polarizations, respectively). The availability of both polarizations helps in the accurate determination of the values of n and r, described in section 3.

A typical sequence of measurements begins with a scan, with no particle in the beam, to determine the background intensity versus angle. After catching a particle, the voltage is adjusted to minimize noise due to the particle motion in the laser beam (the lowest noise is obtained for highly charged particles and low electric fields). We then take one or more scans with a particle in the beam, rotate the half-wave plate by 45° to change the polarization of the incident beam, and take one or more scans for this configuration. Finally, we remove the particle and take another background scan with the second polarization state. We subtract the appropriate background scan I_B from the particle scan I_P by setting

$$I_D(\theta_i)=I_P(\theta_i)-I_B(\theta_i'). \quad (1)$$

The angles θ_i and θ_i' at which measurements are made are close to each other but not equal. Due to the flatness of the background scan from 20° to 160°, the difference between these angles is of no consequence.

Scattering measurements were made for six discrete particle sizes spanning a range of 117 to 1175 nm in radius. The largest particles were polyvinyltoluene; the others were polystyrene. The material was obtained from Dow Diagnostics. The nominal particle size and lot number, as stated on the labels, are given in table 1. A few drops of the packaged material, which is nominally 10% particles by volume suspended in water, were diluted with filtered distilled water to a convenient concentration for catching the particle in the scattering cell. For nominal particle size of 551 nm radius, a volume concentration of 100 ppm, corresponding to a number concentration of about 2×10^6 particles/cm^3, provided an adequate concentration for ease in catching the particle. For other particle sizes, the material was diluted so that the number concentration was constant at about 2×10^6 particles/cm^3.

An effort was made to estimate the amount of impurities in the distilled water, which can affect the measured particle size and index of refraction. We determined by a gravimetric technique that the nonvolatile impurities in the filtered distilled water were of the order of 1 to 2 ppm. The effect of these impurities as well as impurities coming from the packaged material are discussed below.

2.2 Calibration

We found that the angle reading, θ_{inst}, deviates from the true angle, θ_t, by as much as 3°. We performed an

Table 1. Comparison of our results for particle size with electron microscope results of Dow Chemical Company.

Sample[a]	No. of Particles	NBS Light Scattering $r\pm\sigma$, nm	C.I.[b], nm	Dow Electron Microscopy $r\pm\sigma$[c], nm	% Diff.
PVT, 8C2J	6	1129±2	±10.4	1175±9.7	+4
PSL, 7G3L	7	587±4	± 2.6	551±2.8	—7
PSL, 1A74	8	450±5	± 2.5	457±3.0	+2
PSL, 2F8E	10	295±4	± 1.1	300±1.5	+2
PSL, 4N1A	9	236±4	± 1.8	230±2.4	—3
PSL, 4N6H	6	108±3	± 1.5	117±1.4	+8

[a] The abbreviations PVT and PSL are used for polyvinyltoluene and polystyrene. The four character designation is the Dow Diagnostics lot number.

[b] The confidence interval, C.I., is calculated for the 95% confidence level using the relation, C.I.$=t_{m-1}(0.95)\sigma/\sqrt{m}$, where m refers to the number of runs, σ is the standard deviation as measured by Dow, and $t_{m-1}(0.95)$ is Student's t-value corresponding to a 95% confidence level.

[c] This is the nominal radius as specified on the label on the sample.

angular calibration using an accurately indexed protractor 20 cm in diameter with a pointer connected to the traversing periscope, as shown in figure 2. The protractor agreed within 0.09° with an indexing table. Instrument readings and protractor readings were made every 10°. As shown in figure 3, the angle θ_c defined by

$$\theta_c = \theta_t - \theta_{inst} \quad (2)$$

is essentially linear with respect to the instrument angle. From a linear least square fit of the data, we obtain

$$\theta_t = 1.0172\theta_{inst} + 0.24. \quad (3)$$

A small correction, 0.18°, must be added to the constant term on the right-hand side of eq (3) to account for the change in angle due to the scanner time delay in going from an angle reading to an intensity reading. We found that the angle correction given by eq (3) changed after performing experiments on 50 particles. We think that this drift is primarily a result of electronic instability in the encoder.

In addition to the scattering angle calibration, we also checked the polarization of the light relative to the scattering plane by using an accurately oriented

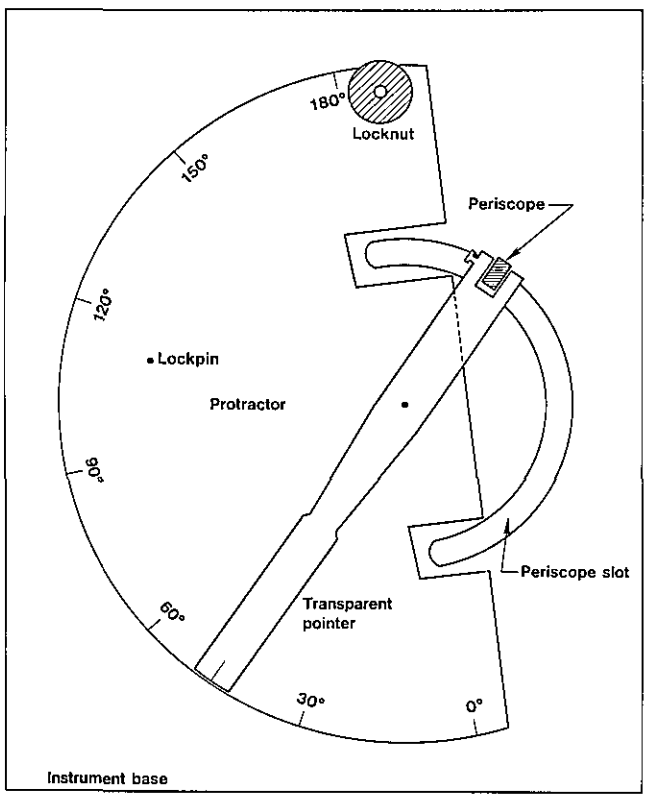


Figure 2–Angle calibration device. Marks are etched every half degree.

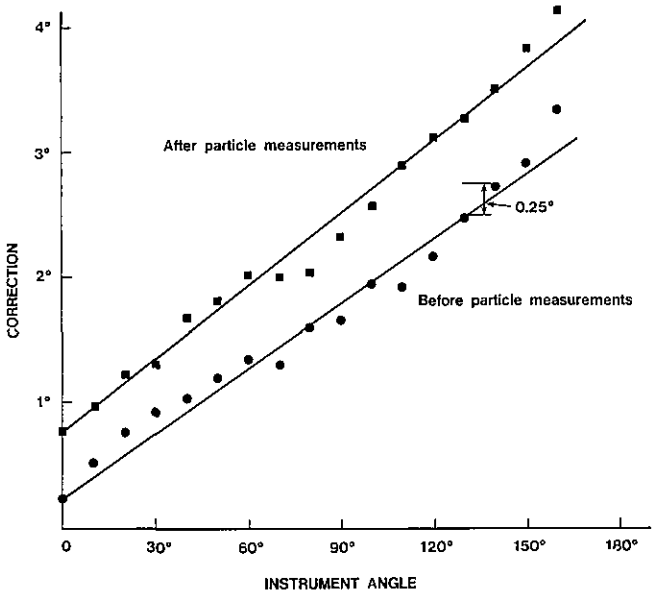


Figure 3–Correction to the angle read by the instrument as a function of that angle before and after measurements done on 50 particles.

polarizer mounted in a rotator. The ratio of the intensities of the vertically and (unwanted) horizontally polarized components of the light was about 200:1. This ratio was reversed when the light passed through a half-wave plate to change the polarization. In section 4.2, we estimate the effects of the polarization mixture and the scattering angle correction on the fitted value of size and refractive index via numerical experiments.

One other slight misalignment results from the elevation of the periscope above the plane where the scattering angle is measured, defining a new scattering plane. This offset is most apparent when the periscope is rotated to the zero scattering angle, because the laser impinges on the lower portion of the periscope slit. The offset is about 1 mm, which corresponds to a 1.3° inclination from the intended scattering plane. This offset leads to both a correction $\Delta\theta$ in the scattering angle and a small angle α between the direction of polarization and the perpendicular to the physical scattering plane for vertical polarization. These angles are given by

$$\sin(\Delta\theta) = \sin\theta\ \cos\theta\ [(1+\cot^2\theta\ \sin^2\phi)^{1/2} - \cos\phi], \quad (4)$$

$$\sin\alpha = \tan\phi\ (\sin^2\theta + \tan^2\phi)^{-1/2}, \quad (5)$$

where θ is the scattering angle measured by the instrument and ϕ is the elevation angle of the periscope. The change in the polarization direction is the same for both horizontal and vertical polarizations. These equations give $\Delta\theta = \phi$, $\alpha = 90°$ for $\theta = 0°$; $\Delta\theta = 0°$, $\alpha = \phi$, for $\theta = 90°$; and $\Delta\theta = -\phi$, $\alpha = 90°$ for $\theta = 180°$.

We can see that the corrections are largest for the endpoints of the angle range; for $\theta=20°$ and $\phi=1.3°$, we obtain $\Delta\theta=0.04°$ and $\alpha=3.8°$. The correction in the scattering angle is negligible, and the polarization effect is comparable to that produced by the laser beam (about 200:1).

3. Data Analysis

After the light scattering data are collected, they are transmitted to an Interdata 7/32 minicomputer for processing. We adapted the program supplied with the Differential II instrument to our needs, and wrote a number of ancillary FORTRAN programs to facilitate the preparation of the input files and the presentation of the results.

We can specify a range of values of the refractive index n and the radius r, and a number of points for each range; the computer finds a quality of fit Q for each of the points in the corresponding rectangular grid. The quality of fit can be defined in a number of ways, as discussed in section 3.3. The values of n and r that correspond to the minimum value of Q give us the best estimates of n and r for that particle. When data are taken for the horizontal and vertical polarizations, a separate best fit is provided for both. Usually we obtain plots showing the theoretical and experimental curves for both polarizations for each best fit, allowing visual evaluation of the data and the fit. We also used three dimensional plots of Q as a function of n and r to get a better understanding of the results. To obtain information about the sensitivity of the results to different types of errors, we also carried out the same procedure with simulated data generated from a theoretical curve by addition of random noise and other errors.

3.1 Mie Theory

The theory originally developed by G. Mie [6–8] allows us to express the intensity of the electromagnetic field scattered by a uniform sphere when a monochromatic plane wave falls on it.

Due to the symmetry of the configuration, the scattered intensity depends only on the angle θ between the incident and scattered directions and the direction of polarization. It suffices to consider separately two polarizations, with the incident field either perpendicular or parallel to the scattering plane defined by the incident and scattered light directions of propagation.

The intensity of the scattered field at a large distance R from the sphere (R large compared to the wavelength λ) is given by

$$I_{\parallel}(\theta) = \frac{I_0}{k^2R^2}\left|\sum_{l=1}^{\infty}\frac{2l+1}{l(l+1)}\left[{}^{e}B_l\ \frac{P_l^1(\cos\theta)}{\sin\theta} + {}^{m}B_l\ \frac{d}{d\theta}P_l^1(\cos\theta)\right]\right|^2, \tag{6}$$

$$I_{\perp}(\theta) = \frac{I_0}{k^2R^2}\left|\sum_{l=1}^{\infty}\frac{2l+1}{l(l+1)}\left[{}^{e}B_l\ \frac{d}{d\theta}P_l^1(\cos\theta) + {}^{m}B_l\ \frac{P_l^1(\cos\theta)}{\sin\theta}\right]\right|^2, \tag{7}$$

where I_0 is the intensity of the incoming beam, P_l^1 is an associated Legendre function, the wave number is $k=2\pi/\lambda$, and the scattering coefficients ${}^{e}B_l$ and ${}^{m}B_l$ are given by

$${}^{e}B_l = \frac{n\psi_l'(kr)\psi_l(nkr)-\psi_l(kr)\psi_l'(nkr)}{n\zeta_l^{(1)\prime}(kr)\psi_l(nkr)-\zeta_l^{(1)}(kr)\psi_l'(nkr)}, \tag{8}$$

$${}^{m}B_l = \frac{n\psi_l(kr)\psi_l'(nkr)-\psi_l'(kr)\psi_l(nkr)}{n\zeta_l^{(1)}(kr)\psi_l'(nkr)-\zeta_l^{(1)\prime}(kr)\psi_l(nkr)} \tag{9}$$

with ψ_l and ζ_l related to Bessel and Hankel functions by

$$\psi_l(\rho)=(\tfrac{1}{2}\pi\rho)^{\frac{1}{2}}J_{l+\frac{1}{2}}(\rho), \tag{10}$$

$$\zeta_l^{(1)}(\rho)=(\tfrac{1}{2}\pi\rho)^{\frac{1}{2}}H_{l+\frac{1}{2}}^{(1)}(\rho). \tag{11}$$

We see that the coefficients ${}^{e}B_l$ and ${}^{m}B_l$ have an intricate dependence on n and r. The fact that at least in some cases the shape of the scattering curves changes little if the product nr remains constant is not apparent from the equations. The effects of this dependence on the data analysis are discussed in section 4.

3.2 Procedure

The data from the experiment are stored on the HP microcomputer, formatted as required for the input to the analysis program, and transmitted to the Interdata minicomputer. We then add the remaining input information required by the program, choose a range of values for n and r and the number of grid points for each variable, and compute the corresponding values of the quality of fit. In most cases we also eliminate some of the readings that are obviously incorrect, especially at the beginning and at the end of the curves; we attribute these errors to saturation of the photomultiplier or to uncertainties due to a large background intensity and a small signal. Normally we plot the theoretical and experimental curves for the best fits (both polarizations for each best fit), and decide whether it is necessary to shift or refine the

grid. Another helpful tool is the production of three-dimensional plots of Q as a function of n and r to represent and understand the shape of this surface better. Actually, we plot $1/Q$ and $-\log Q$ and look for a maximum, because the resulting plots are more easily visualized.

In most cases, this procedure is repeated for one or more other measures of the quality of fit, as discussed below.

3.3 Quality of Fit

Although often the quality of the fit of a theoretical curve to the experimental data can be evaluated visually from the plots, a numerical value of the quality of fit is of great help and can be used for automated computer searches for a minimum.

A simple least-squares fit is based on the expression

$$Q=\frac{1}{N}\sum_{i=1}^{N}(E_i-\alpha T_i)^2 \tag{12}$$

where N is the number of data points, E_i is the experimental value, T_i is the theoretical value for the same angle, and α is an overall scaling factor required because we do not measure absolute intensities. The theoretical value is computed for each degree, and then averaged over three consecutive values to take into account the two-degree acceptance angle of the detector. We carry out a further interpolation to find a better approximation to the theoretical value at the precise angle where the intensity was measured. The original program uses integer values for the angles, and this correction becomes important if the angles are entered digitally and then corrected.

We compute the value of α by finding the value that minimizes the corresponding Q, which is

$$\alpha=\left(\sum_{i=1}^{N}E_iT_i\right)/\sum_{i=1}^{N}T_i^2. \tag{13}$$

Originally we used the expression for Q as given in the program supplied with the instrument, which differs from eq (12) in that the experimental values of the intensities are scaled to the theoretical ones, not the other way around. This choice proved to be troublesome because the normalization of the theoretical values can change abruptly as peaks in the curve appear or disappear; that is, the normalization becomes strongly dependent on n and r. Although in most cases the values of n and r for the best fit change little if at all, spurious minima may appear when the normalization changes abruptly near those values, as described in section 4.2 for a numerical experiment. Also the sensitivity analysis can be affected by this difficulty.

Alternatively, the value of α can be chosen by matching a particular value of the measured intensity to the theoretical value at the same angle. This least-squares fit gives much weight to errors in the measured intensity in the region where these values are large and the slope of the curve is steep; a small error in the measured value for the scattering angle leads to a large deviation between E_i and αT_i, which is then squared.

We tried several other methods to define the quality of fit. The most interesting alternative resulted from weighting the intensities by a function of the angle. Specifically, we considered

$$I_1(\theta)=I(\theta)\sin^2(\theta/2), \tag{14}$$

which reduces the contributions to the quality of fit from the large peaks at small angles; conversely, it overemphasizes the error at angles close to 180°, where the amplitudes are small and the relative errors are large. The factor $\sin^2(\theta/2)$ was chosen because it tends to equalize the size of the peaks.

We also tried a least-squares fit to the function

$$I_2(\theta)=\log I(\theta)\ u(I-I_{\min}), \tag{15}$$

where u is a unit step function we introduced to avoid problems with values of the intensity smaller than a selected minimum value $I_{\min}$.

For the previous functions, we can find a value of α that minimizes Q. This is not the case for Q defined as the area between the experimental and theoretical scattering curves, which is approximately equal to the sum of the absolute values of the differences of the intensities at each angle. The value of α was selected by matching the first peak or, preferably, by setting

$$\alpha=\left(\sum_{i=1}^{N}E_i\right)/\sum_{i=1}^{N}T_i, \tag{16}$$

which makes the average deviation between the curves vanish; these choices do not necessarily minimize Q.

For good quality data, all methods naturally lead to the same or very similar values for n and r. For noisy data, there were examples where the least-squares fit would not lead to the best value for n and r as seen from the curves. Other methods also failed in particular examples, so that no overall recommendation can be made other than to try several methods in difficult

cases, and always check the plots of the curves for the selected best fits.

4. Results

4.1 Actual Experiments

The primary results that we obtained from this series of experiments were the size and refractive index of the particles. In table 1 we present the average values and standard deviations for the radius obtained from a number of particles of each nominal size. In table 2 we give our results for the index of refraction together with values found by other authors. For each particle, the chosen values are those that provide the best fit for both polarizations (as reflected in the harmonic mean of the values of Q for vertical and horizontal polarizations); we chose the scan with the lowest Q if we repeated scans of the same particle. The nature of the theoretical curves and the quality of the data vary considerably with the size of the particles. One sample curve for each size and polarization is shown in figure 4. The number of features (peaks and valleys) increases with the radius of the sphere, while the observed noise, due mainly to the particle motion in the laser beam, is larger for the small particles. In figure 5 we use actual data to illustrate the difference in the quality of fit before and after the angle correction. The errors in n and r are almost 2% if the angle correction is not made.

When the same data are analyzed with the weight function $\sin^2(\theta/2)$, the results vary slightly for the different particles; a comparison of the averages for both types of fit is shown in table 3. In figure 6 we compare a curve obtained with this weight factor to the unweighted curve.

To show the variability of the results within a group of particles, we present in table 4 the best values of the radius and the index of refraction for those particles of nominal radius of 457 nm, including the best scans for each polarization when multiple scans were made. The particle size is not perfectly uniform in each sample, and we expect to see a spread in the measured radii. On the other hand, the refractive index should be the same for all particles of equal composition, and we attribute the variations to experimental error.

For one particle we took nine scans with vertically polarized light and three more for horizontal polarization. The results of the analysis of these data are shown in table 5. The standard deviation for the radius is smaller than the one previously calculated for groups of particles, but that for the refractive index is about the same, suggesting that this parameter does not change at least for particles of the same size.

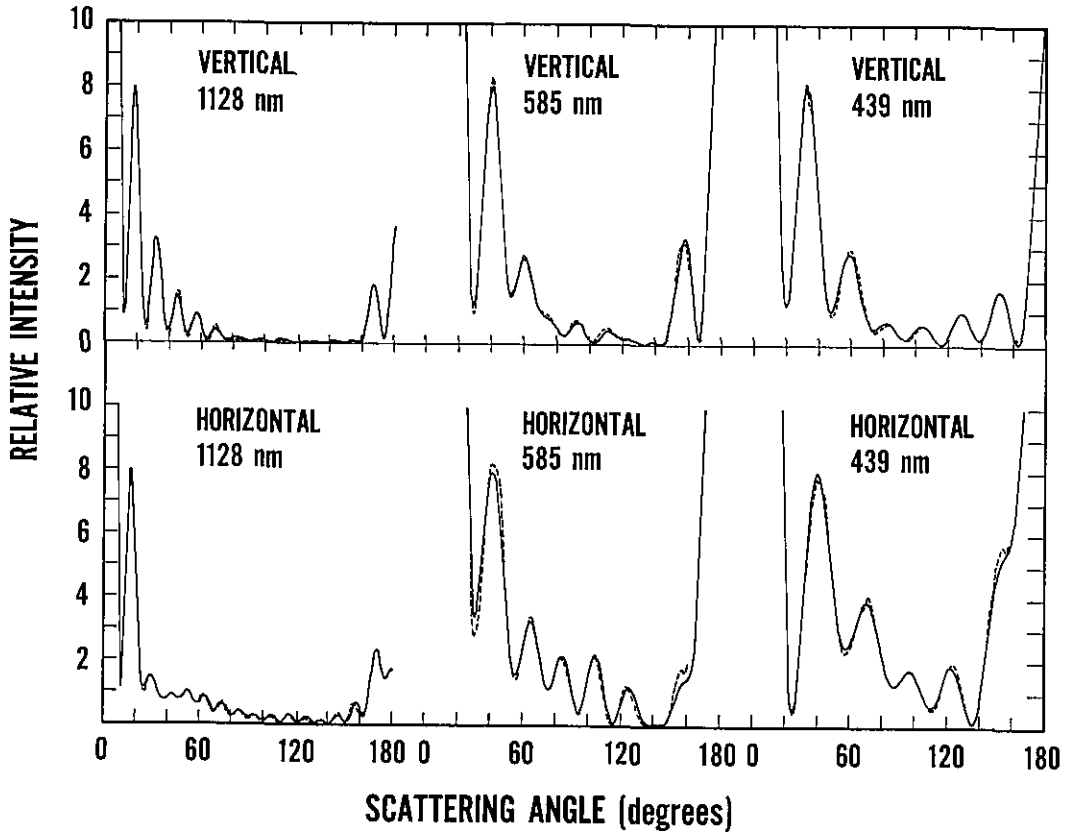
If we choose the value of r that gives the best fit for a fixed value of the refractive index, for instance the best-fit value 1.608, we find that the standard deviation σ of the radius is lowered to a value of 0.47 nm. The quantity σ provides a measure of the random component of the uncertainty in the single-particle measurements. If we choose instead the value 1.615 obtained from bulk measurements, we obtain a radius of 437.4 nm with a σ of 0.53 nm. Cook and Kerker [3] reported results from nine repeated scans for a single particle of radius 398 nm; they obtain a σ of the radius equal to 1.6 nm. Our smaller value of σ is probably in part a result of the automatic conversion of data to the digital form; Cook and Kerker point out that the component of σ resulting from reading the data off the chart recordings is approximately 0.8 nm.

To obtain the parameters of the best fit, we examined a table of values of $Q(n,r)$ for a rectangular grid to find the minimum. The nature of the surfaces representing $1/Q$ and $-\log_e Q$ as functions of n and r vary mainly with the size of the particle and the polarization. Some of these surfaces are rather flat, while others show a ridge that cuts diagonally across the surface in such a way that Q varies little when we increase the radius and decrease the index of refraction simultaneously. We show part of such a table near the minimum of Q for a particle of nominal radius of 551 nm in table 6, and we show three-dimensional plots of the surface in figure 7 for both polarizations.

Table 2. Comparison of our results for the refractive index with those obtained by others

Material	Nominal Radius, nm	Refractive Index This study $n\pm\sigma$	Refractive Index Literature[a] $n\pm\sigma$
PSL	bulk		1.615[b]
PSL	66–148		1.620±0.008[c]
PSL	117	1.66±0.05	
PSL	230	1.603±0.029	
PSL	279		1.610±0.002[d]
PSL	300	1.599±0.013	1.64±0.019[e]
PSL	457	1.612±0.004	
PSL	505		1.58±0.012[e]
PSL	551	1.619±0.002	
PVT	23–190		1.603±0.009[d]
PVT	1175	1.613±0.002	

[a] Bateman's dispersion relation used to obtain refractive indices for λ=441.6 nm
[b] Boundy and Boyer
[c] Heller and Pugh
[d] Smart and Willis
[e] McRae



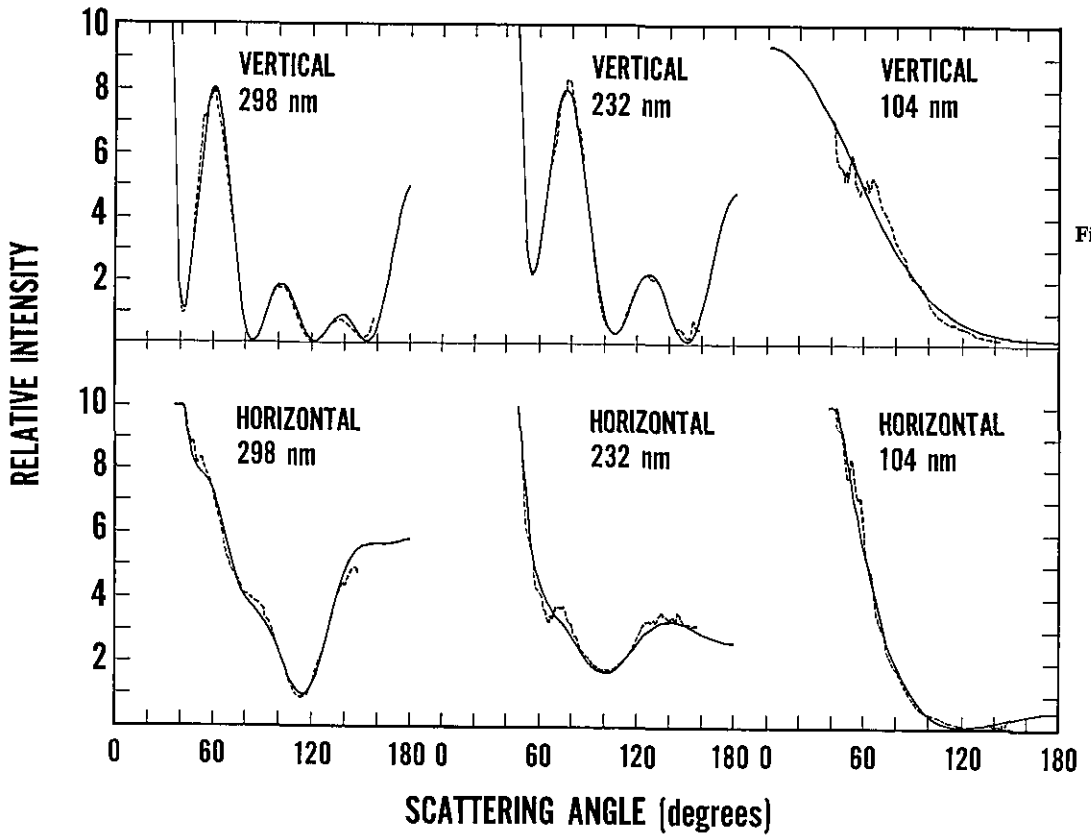


Figure 4–Typical curves of the scattered intensity as a function of angle for both polarizations and for each particle size. The computed curve (solid line) is a best fit to the data (dashed line).

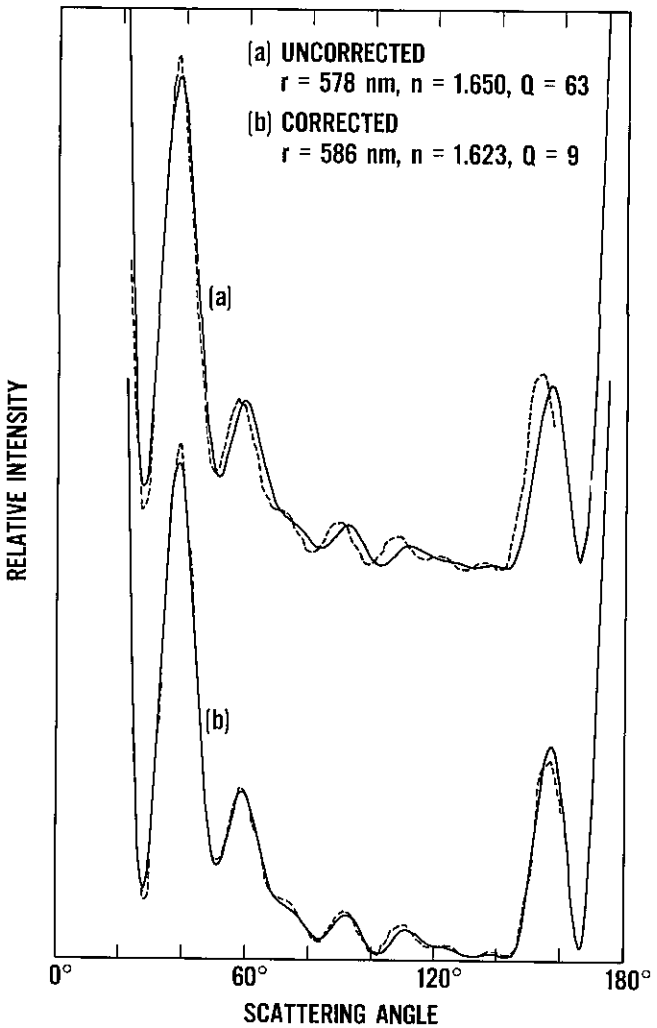


Figure 5–Best fit to experimental data (a) before and (b) after angle correction.

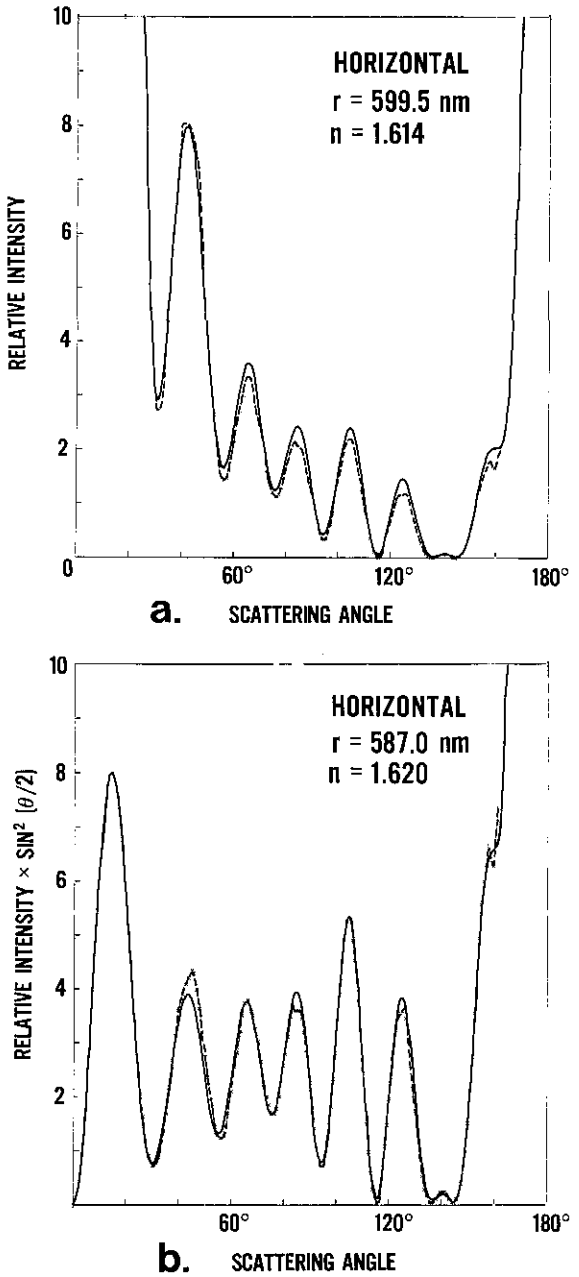


Figure 6–Best fits for curves of intensity of the light scattered by a 551-nm particle as a function of angle by the least squares method (a) without, and (b) with a weighting factor. ⟶

Table 3. Comparison of radius and refractive index obtained by unweighted and weighted least-squares fits.

Nominal radius, nm	Radius, nm (unweighted)	Radius, nm (weighted)	Refractive index (unweighted)	Refractive Index (weighted)
117	108	108	1.664	1.669
230	236	237	1.603	1.607
300	295	293	1.599	1.616
457	450	451	1.612	1.609
551	587	587	1.619	1.622
1175	1129	1118	1.613	1.627

Table 4. Radius and index of refraction for eight particles of 457 nm nominal radius.

	Harmonic mean[a]		Vertical polarization		Horizontal polarization	
	Radius	Refractive Index	Radius	Refractive index	Radius	Refractive index
	454.5	1.614	455.0	1.614	453.5	1.610
	457.0	1.613	457.0	1.613	453.0	1.623
	449.0	1.618	453.0	1.605	449.0	1.618
	446.5	1.612	451.0	1.600	446.0	1.614
	448.0	1.613	451.5	1.610	448.0	1.613
	450.5	1.613	453.5	1.610	450.5	1.613
	453.5	1.608	453.5	1.608	453.5	1.610
	439.4	1.606	440.2	1.603	439.4	1.610
Average	449.8	1.612	451.8	1.608	449.1	1.614
σ	5.1	0.004	4.8	0.005	4.5	0.004

[a] Harmonic mean refers to the determination of the best fit radius and refractive index by taking the harmonic mean of the Q's for the cases of vertically and horizontally polarized light.

Table 5. Results for repeated scans for one particle of 457 nm nominal radius.

	Unweighted fit		Weighted fit			
Run	Radius	Refractive index	Radius	Refractive index	$n=1.608$[a] Radius	$n=1.615$ Radius
1-V	440.2	1.601	438.0	1.612	438.4	436.6
2-V	436.4	1.620	437.6	1.614	439.8	437.8
3-V	439.2	1.606	436.6	1.618	438.6	437.2
4-V	440.2	1.603	439.8	1.605	438.8	437.2
5-V	439.4	1.606	441.0	1.601	439.0	437.2
6-V	438.0	1.610	439.8	1.604	438.6	436.8
7-V	439.8	1.604	438.6	1.610	438.8	437.0
8-V	439.4	1.606	438.8	1.609	439.0	437.2
9-V	439.4	1.607	440.0	1.604	439.2	437.4
1-H	439.4	1.610	440.2	1.601	439.8	438.4
2-H	438.0	1.612	440.2	1.612	438.6	437.4
3-H	438.4	1.614	436.4	1.620	439.4	438.2
Average	439.0	1.608	438.9	1.609	439.0	437.4
σ	1.10	0.0053	1.50	0.0063	0.47	0.53

[a] The last two columns correspond to values of the radii obtained for a fixed value of the refractive index (unweighted fit). The first value is obtained for the best fit to all particles of this size, and the second one is the published bulk value.

Table 6. Normalized values of Q near the minimum for a particle of 551 nm nominal radius.

	Vertical polarization					Horizontal polarization				
n	1.618	1.622	1.626	1.630	1.634	n	1.614	1.618	1.622	1.626	1.630
r						r					
582.5	384	147	28	4.8	1.1[a]	581.5	6.3	5.0	3.5	2.1	1.5
583.5	204	40	7.3	1.3	1.8	582.5	5.3	3.6	2.1	1.3	1.6
584.5	57	11.1	1.8	1.2	3.4	583.5	4.0	2.4	1.3	1.1	2.2
585.5	16.6	3.0	1.0	2.4	5.4	584.5	3.0	1.7	1.0	1.5	3.2
586.5	4.7	1.1	1.8	4.3	7.9	585.5	2.3	1.3	1.2	2.3	4.5
587.5	1.5	1.4	3.3	6.4	10.5	586.5	1.9	1.3	1.9	3.5	6.2
588.5	1.3	2.6	5.1	8.8	13.5	587.5	1.9	1.8	2.9	5.1	8.1

[a] The minimum values of Q in each column are underlined to illustrate the slow variation of Q on a diagonal.

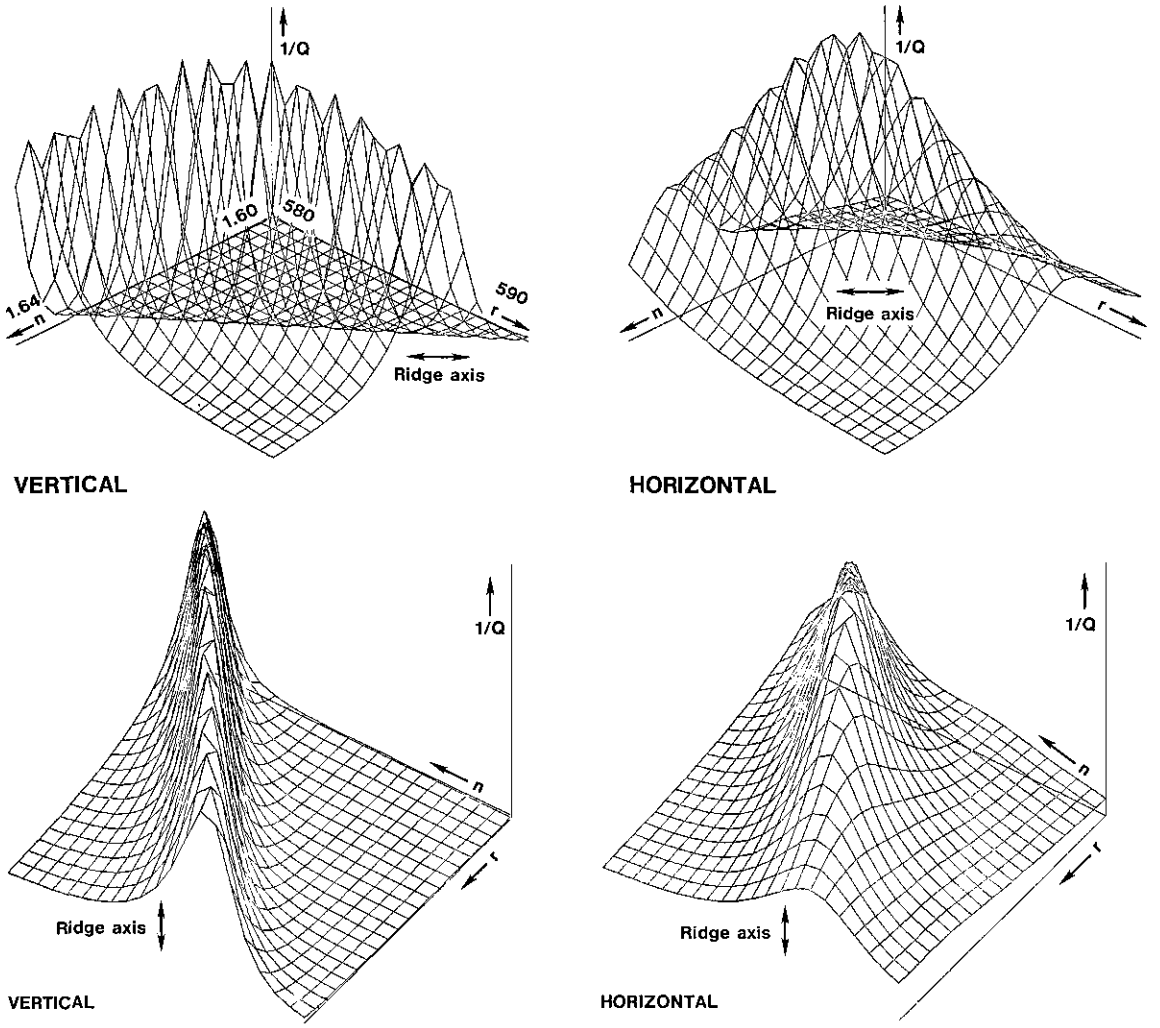


Figure 7–Two views of the three-dimensional surface of the inverse of the quality of fit, $1/Q$, as a function of the index of refraction n and the radius r, over a range of a few percent from the best values. The ridges cause difficulties in accurately determining the correct values of n and r.

4.2 Numerical Experiments and Sensitivity Analysis

To study the effects of experimental errors on the best fit parameters, we modified the program that produces the theoretical values of the scattered intensities to add simulated errors and generate a file or "data" to be processed by the same procedure described above. Noise obtained using a random number generator with a maximum amplitude of from 1 to 3% of the intensity at each angle was added to all curves.

The error in the angle calibrations was simulated by computing an "experimental" angle

$$\theta' = 0.98\theta - 0.5, \qquad (17)$$

the errors is of the order of 2%, and the relative error in the refractive index is comparable in magnitude to where θ is the true angle used in the computation of the scattered intensity. The results of the analysis of these "data" are shown in table 7. The magnitude of that in the radius, and they are of opposite signs. In this table we also show the results obtained by minimizing the average instead of the harmonic mean of the Q's for the vertical and horizontal polarizations. In a number of cases we obtain a better result when we use the average value. The reason for this behavior is that the quality of the "data" is essentially the same for both polarizations, so that larger values of Q reflect a greater variability of Q with n and r; thus, a mean that emphasizes the larger set of Q's leads to a better result. This property is even more evident in the second group of results, which simulate the drift in the angle correction between the beginning and the end of

a series of runs made over a period of a month. The formula we used here is

$$\theta' = 0.998\theta - 0.5, \tag{18}$$

and the resulting errors in n and r can be as large as 1.5% when we use the harmonic mean, and 0.5% when we use the average of the Q's.

We simulated the noise in the values of the scattered light intensities by superimposing: (1) a random noise of maximum amplitude of 2% computed independently at each degree, and, (2) a larger noise of maximum amplitude of 5% that remains constant over periods of 10° in the scattering angle. For the two smallest particles, which are less stable in the beam than the larger ones, we increased the amplitudes of the noise to 3 and 10% respectively. For particles of a radius of 450 nm and larger, the error in the results was at most 0.3%, while for the smallest particles the error can reach 1%.

We also added 1% of the intensity for the horizontal polarization to the intensity for the vertical polarization and vice versa to simulate the detector displacement and the imperfection of the half-wave plate. The effect on the results was less than 0.2%, and we do not show them in table 7.

We also processed all these simulated experiments using the program that includes a weighting factor of $\sin^2(\theta/2)$. The results obtained this way were better than those obtained with a straight least-squares fit in many cases, and worse in some.

A numerical experiment with simulated noise illustrates the problem that can arise when the experimental intensities are the ones to be scaled. The high sharp ridge in the first graph of figure 8 is a consequence of a sudden change in the normalization of the theoretical intensities as a peak in the curve disappears; the corresponding values for n and r lead to a poor fit of the curves. The accompanying figure obtained by scaling the theoretical intensities shows only the physically meaningful ridge.

We also show an example of the pitfalls of an uncritical use of the value assigned to the quality of fit. In figure 9 we show two curves that correspond to

Table 7. Results of numerical experiments[a].

Assumed radius, nm	Best: Radius, nm	Best: Refractive index	Best (weighted): Radius, nm	Best (weighted): Refractive index
Angle error ($\theta' = 0.98\theta - 0.5$), Harmonic mean Q				
1150	1114(-3.1%)	1.660(+3.1%)	1098(-4.5%)	1.695(+5.2%)
580	560(-3.4%)	1.660(+3.1%)	576(-0.7%)	1.600(-0.6%)
450	434(-3.6%)	1.655(+2.8%)	438(-2.7%)	1.660(+3.1%)
300	288(-4.0%)	1.635(+1.6%)	292(+2.7%)	1.615(+0.3%)
230	228(-0.9%)	1.640(+1.9%)	226(-1.7%)	1.645(+2.2%)
115	119(+3.5%)	1.560(-3.1%)	119(+3.5%)	1.560(-3.1%)
Angle error, average Q				
1150	1114(-3.1%)	1.665(+3.4%)	1134(+1.4%)	1.635(-1.6%)
580	572(-1.4%)	1.635(+1.6%)	580 -	1.610 -
450	442(-1.8%)	1.630(+1.2%)	446(-0.9%)	1.610 -
300	292(-2.7%)	1.645(+2.2%)	300 -	1.625(+0.9%)
230	228(-0.9%)	1.640(+1.9%)	226(-1.7%)	1.645(+1.2%)
115	119(+3.5%)	1.560(-3.1%)	119(+3.5%)	1.560(-3.1%)
Angle calibration drift ($\theta' = 0.998\theta - 0.5$), Harmonic mean Q				
1150	1132(-1.6%)	1.634(+1.5%)	1148(-0.2%)	1.608(-0.1%)
580	570(-1.7%)	1.632(+1.4%)	578(-0.3%)	1.608(-0.1%)
450	444(-1.3%)	1.626(+1.0%)	444(-1.3%)	1.630(+1.2%)
300	300 -	1.620(+0.6%)	300 -	1.618(+0.5%)
230	228(-0.8%)	1.630(+1.2%)	228(-0.9%)	1.626(+1.0%)
115	117(+1.7%)	1.590(-1.2%)	115 -	1.606(-0.2%)

(continued)

Table 7. Results of numerical experiments -continued.

Assumed radius, nm	Best: Radius, nm	Best: Refractive index	Best (weighted): Radius, nm	Best (weighted): Refractive index
Angle calibration drift, average Q				
1150	1140(-0.7%)	1.626(+1.0%)	1150 -	1.608(-0.1%)
580	578(-0.3%)	1.616(+0.4%)	582(+0.3%)	1.604(-0.4%)
450	448(-0.4%)	1.616(+0.4%)	446(-0.9%)	1.620(+0.6%)
300	296(-1.3%)	1.628(+1.1%)	300 -	1.614(+0.2%)
230	230 -	1.616(+0.4%)	228(-0.9%)	1.628(+1.1%)
115	117(+1.7%)	1.586(-1.5%)	115 -	1.612(+0.1%)
Random noise[b], Harmonic mean Q				
1150	1150 -	1.610 -	1150 -	1.610 -
580	582(+0.3%)	1.604(-0.4%)	580 -	1.610 -
450	450 -	1.610 -	450 -	1.610
300	300 -	1.616(+0.4%)	300 -	1.612(+0.1%)
230	230 -	1.608(-0.1%)	230 -	1.610 -
115	115 -	1.594(-1.0%)	115 -	1.614(+0.2%)
Random noise, average Q				
1150	1150. -	1.610 -	1150. -	1.610
580	582.(+0.3%)	1.604(-0.4%)	580. -	1.610 -
450	450. -	1.610 -	450. -	1.610 -
300	300. -	1.618(+0.5%)	300. -	1.614(+0.2%)
230	232.(+0.9%)	1.596(-0.9%)	230. -	1.608(-0.1%)
115	115. -	1.592(-1.1%)	115. -	1.614(+0.2%)

[a] Refractive index is 1.610 in all cases. The numbers in parentheses are the deviations from the assumed values.

[b] Random noise of maximum amplitude of 2% every degree plus a larger noise of maximum amplitude 5% that remain constant for 10 ° is used. For the smallest two particle sizes the corresponding numbers are 3% and 10% respectively.

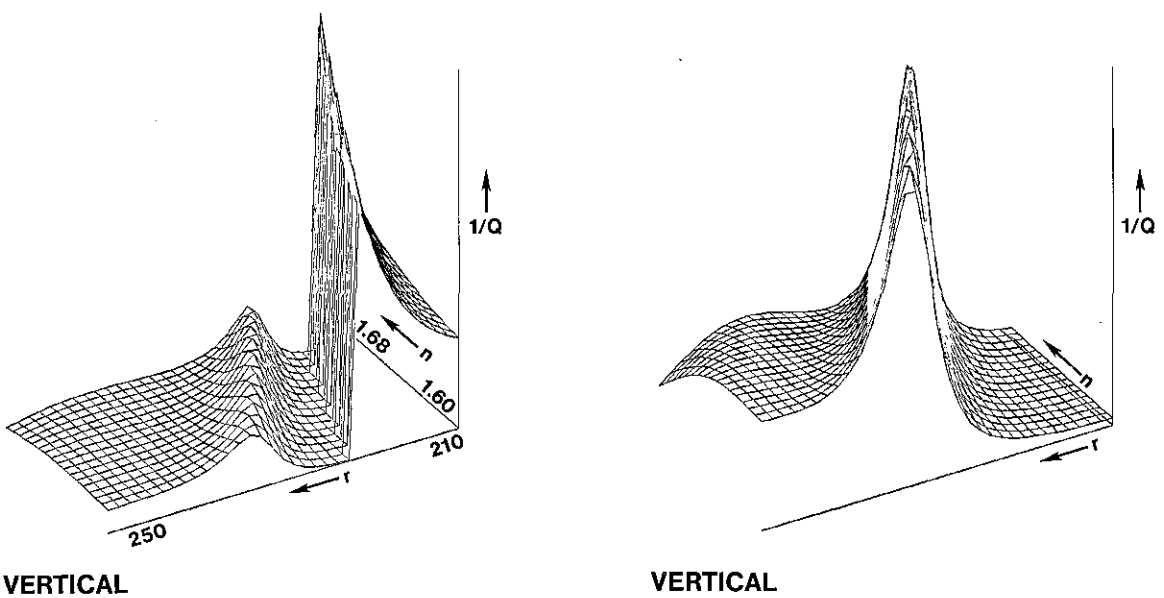


Figure 8–Plots of $1/Q$ as a function of n and r that show a ridge of spurious "best" fits for vertical polarization when the experimental intensities are scaled, a ridge that is absent when the theoretical intensities are scaled instead. Simulated data obtained by adding random noise to the calculated intensity.

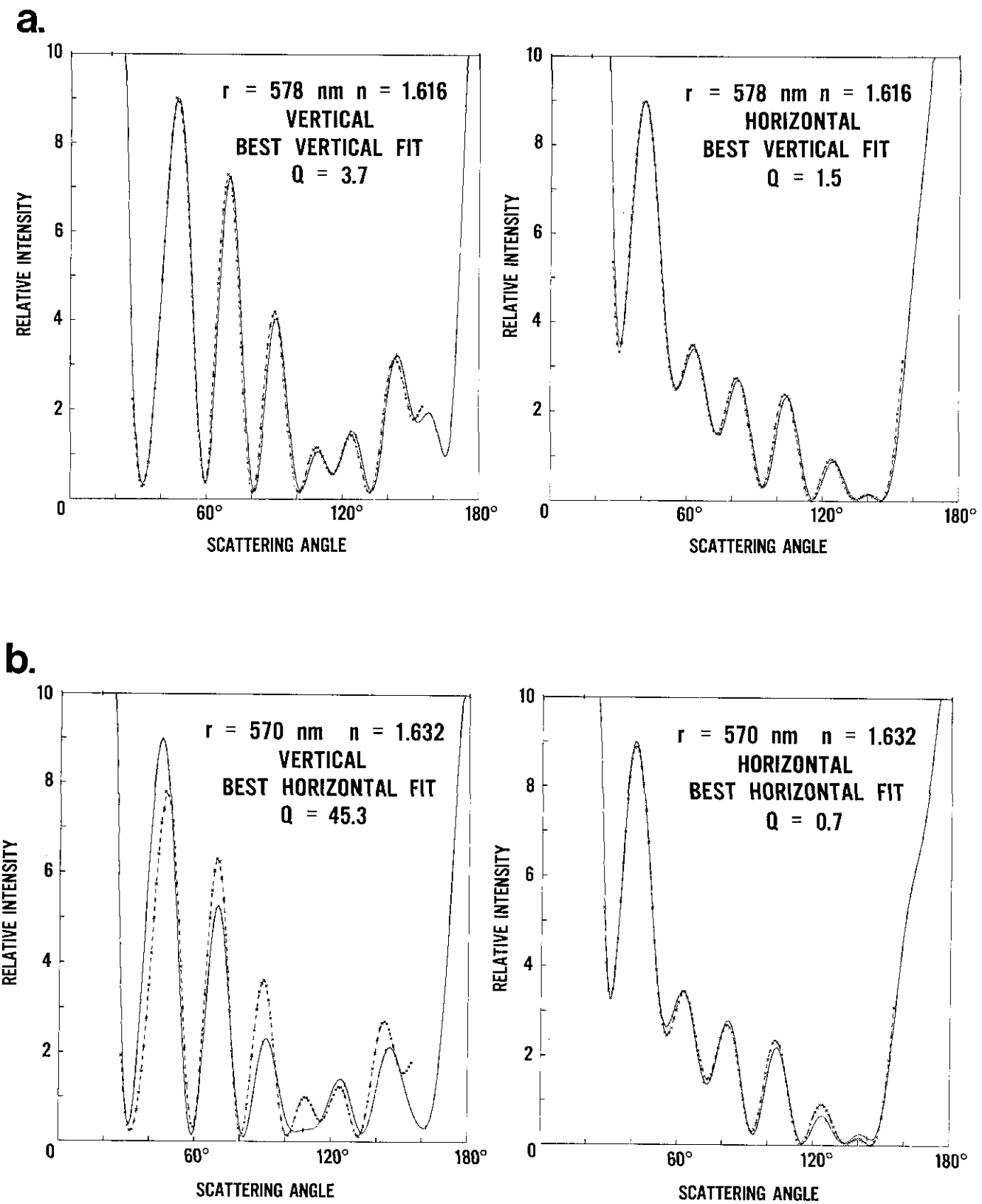


Figure 9–Overall best fit using (a) the average of Q_V and Q_H, which emphasizes the larger of the two, and (b) the harmonic mean, which emphasizes the smaller Q.

the horizontal and vertical polarizations for the parameters that give the best fit for the vertical polarization (a), and two more that do so for the best fit for the horizontal polarization (b). If we choose the pair (b) with the lowest harmonic mean between Q_V and Q_H, that is, the one that has $Q_V=45.3$ and $Q_H=0.7$, we obtain a very poor fit to the data for vertical polarization. The other pair (a), with $Q_V=3.7$ and $Q_H=1.5$, which has the lowest average of Q_V and Q_H, is a much better compromise. In this case, the larger values of Q_V relative to Q_H are not due to lower quality data, but to the presence of more variation in the intensity curve. A third method to choose between different pairs of Q_V and Q_H is to normalize both to the corresponding minimum value and then take the average. In numerical experiments, each of these methods gave the best values of n and r in comparable numbers of instances, so that no clear choice of method was evident.

5. Discussion

5.1 Error Analysis

We obtained information about the magnitude of the errors both from numerical experiments and from the actual data.

The size of the error in the index of refraction and in the radius when the angle errors are not corrected, a decrease of about 2% in the r and an increase of the same magnitude in n, indicate that the calibration of the angle readout is a necessity. We found an additional problem of a drift in this calibration, as the correcting factors came out somewhat different at the beginning and at the end of a run of experiments. The effect of this drift, a relative error of about 1%, has to be included in the uncertainty of our results.

The noise in the intensity of the scattered light can be observed better by fixing the angle and recording a trace of the intensity versus time. The general characteristics of such a trace show that there are two components of this noise. The smaller component, of about 2–3% of the intensity, affects each angle reading independently; the larger component, which we think comes from the motion of the particle within the laser beam when the servocontrol is activated, varies between 5% for the larger particle to 10% for the smaller ones and has a lower frequency so that it affects intensity readings in groups of about 10° for the data collection speed that we used.

Errors in the polarization of the laser beam come mainly from two sources, imperfections in optical components and misalignment of the periscope. The imperfections in the half-wave plate and other optical components allow the introduction of some light of the wrong polarization into the incident beam at a 200:1 intensity ratio. The error from the inclination of the periscope with respect to the assumed scattering plane varies with the scattering angle, as shown in section 2, and it remains within about 0.2% for the range of angles that we used. The effects of these errors on the size and refractive index are small compared to those introduced by the angle errors and the noise discussed above. A combined estimate of the effects of these errors on n and r comes to less than 0.2%.

Variability in the experimental results can be found in the analysis of the data for different particles within the same sample, and also from repeated scans of the same particle.

Within a group of particles from the same sample, we expect the index of refraction to have a common value, while the radius can vary significantly. (For instance, we excluded a particle of radius 417 nm from the statistics for the particles of nominal radius of 457 nm.)

When we take a fixed value for the index of refraction for all particles from a sample, the best values for the radius under this restriction show a smaller variation but the value of this radius depends on which value is chosen from the index of refraction. We can see this from the results shown in table 5.

Tables 4 and 5 show that the values of the standard deviations for the index of refraction are similar in the particle group and for repeated scans of the same particle, which confirms that the index of refraction does not vary significantly within particles of a given nominal size. For particles of different sizes we find a variation of the index of refraction from 1.599 to 1.619, with the exception of the smallest particle. For this particle of nominal radius of 117 nm, the data are too noisy to find a reliable value of the refractive index; also the curves show very little structure and the quality of fit varies slowly with the radius and index of refraction.

The magnitudes of the various errors affecting the overall uncertainty in the measurement of the radius of a single particle are summarized in table 8. The quantity σ obtained for the 12 repeat measurements of one particle with nominal radius of 457 nm is a measure of the random error. The numerical experiments simulating the noise in the scattered light intensity yield similar values of σ. Sources of systematic errors associated with the instrument include the error in particle size due to the angle drift,

δ_θ, to the polarization misalignment, δ_P, and to the uncertainty in the laser wavelength, δ_λ. Results in table 7 indicate that the percentage change in r and n are comparable in magnitude. Consequently, our observed variation in n of 1% leads to a δ_θ of 1% also. The minimum value of δ_θ, estimated to be 1.4 nm, was obtained for the 457 nm particle for which an angle calibration was performed at the beginning and at the end of 2 days of measurements. Numerical experiments give a magnitude of 0.2% for δ_P. The least important error is the uncertainty in the wavelength of the He-Cd laser, $\delta_\lambda=0.07$ nm. An overall estimate of the instrument uncertainty for five measurements on a single particle, u_s, can be obtained from

$$u_s=t_4(0.95)\sigma/\sqrt{5}+|\delta_\theta|+|\delta_P|, \qquad (19)$$

where $t_4(0.95)$ is the value of the limit obtained from Student's t-distribution for four degrees of freedom. For the case above we find u_s equal to 3.8 nm. As indicated in table 8, for all other sizes there is a greater error associated with the angle drift and a correspondingly larger value of u_s.

For comparison with other studies of particle size, it is of interest to obtain the uncertainty associated with the determination of the average particle size. In eq (19) the quantity $\sigma/\sqrt{5}$ is replaced by the standard error, $\sigma_P/\sqrt{m}$, where σ_P is the standard deviation of the m particles measured. The formula for the standard error assumes the particle to be randomly selected. The use of the Differential II for an aerosol with a broad distribution will probably lead to a biased result because of the operator's tendency to choose inadvertently, say, the particle with the brightest appearance. In our case we expect little bias because the size distribution is narrow.

In addition to instrumental errors and statistical uncertainty due to the small sample size, there are uncertainties associated with the nature of the particle. While the polystyrene spheres are often considered to be spherical and homogeneous, we expect that slight imperfections might exist. One source of nonuniformity of the particle can be the residual (nonvolatile) impurities in the distilled water. If the 2 ppm of impurities coat the particle when the water droplet containing the particle evaporates, the thickness of this coat would vary between 0.4% of the radius for the smallest particles and 0.003% for the largest one, based on the average droplet diameter of 4.2 μm. This calculation also includes the impurities coming from the liquid in the undiluted sample. If we assume, on the other hand, that all the emulsion stays on the particle instead of being diluted, we find that another coating of between 0.4% for the smallest particles and 0.09% for the largest one is produced. The measured index of refraction can also be affected by such a coating, but visual observation of the residue indicates that these coats are translucent and not likely to change the results significantly. Table 8 includes the uncertainty associated with the impurity, δ_I, and an estimate of the overall uncertainty in the mean particle size, u_T, which is given by

$$u_T=t_{m-1}(0.95)\sigma_P/\sqrt{m}+|\delta_\theta|+|\delta_P|+|\delta_I|. \qquad (20)$$

The relative error in n is similar in magnitude to that in r.

Another imperfection of the particle is the slight deviation from spherical shape. Careful measurements by transmission electron microscope of the particles of nominal radius of 457 nm indicated a deviation of

Table 8. Uncertainty in the particle radius.

Nominal radius, nm	Precision σ[a], nm	δ_θ[a], nm	δ_P[a], nm	δ_I[a], nm	u_s[b], nm	u_T[c], nm
1175	1.1	+11.0	2.0	-1.0	14.4	24.4
551	1.1	+ 6.0	1.0	-0.5	8.4	10.1
457	1.1	+ 1.4[d]	1.0	-0.4	3.8	5.3
300	1.2	+ 3.0	0.5	-0.3	5.0	4.9
230	2.0	+ 2.0	0.3	-0.9	4.8	5.0
117	2.0	- 1.0	0.2	-0.6	3.7	3.5

[a] σ was estimated from repeat measurements of a single 457 nm particle. The quantities δ_θ, δ_P, and δ_I represent systematic errors resulting from angle calibration drift, polarization misalignment, and impurities in the water, respectively.

[b] The total uncertainty for a single particle based on five experiments is $u_s=t_4(0.95)\sigma/\sqrt{5}+|\delta_\theta|+|\delta_P|$, where $t_4(0.95)$ is the Student t-value for four degrees of freedom and for 95% confidence level.

[c] The total uncertainty for the mean particle size is given by $u_T=t_{m-1}(0.95)\sigma_P/\sqrt{m}+|\delta_\theta|+|\delta_P|+|\delta_I|$, where σ_P is the size distribution determined by Dow Chemical Company.

[d] For this particle size, the uncertainty resulting from angle calibration drift is small, because an angle calibration was performed before and after the experiment.

0.6% from sphericity at a precision of about 0.3%. If the particles were nonspherical, then over the course of our measurements the particle would have undergone extensive rotation as a result of Brownian motion. We conclude from Yeh's treatment of light scattering from spheroids with a slight deviation from sphericity [9] that the effective radius measured by light scattering will satisfy the inequality $r_{min}<r<r_{max}$, where the corresponding D_{min} and D_{max} are the length of the minor and major axes of an ellipsoid of revolution. So we expect that our measurement r would be within 0.6% of the average of r_{min} and r_{max}, probably closer to 0.3 or 0.4%. This possible error is not included in eq (20). There may also be inhomogeneity of the particle resulting from strain as the water evaporates from the particle, but we have no estimate of this effect.

5.2 Comparison With Other Experiments

In table 1 we compare our results with the information provided by Dow Chemical Co. on the samples. The size and standard deviation they provide are obtained by transmission electron microscopy on a population of the order of 100 particles [10]. We find a difference that is outside the limits of the statistical variation within a lot.

Particles of the same sample of nominal radius 457 nm were examined at the National Bureau of Standards by other methods to determine the size distribution of the particles. Preliminary results obtained by light scattering from a suspension give a mean radius of 448 nm; array sizing by optical microscopy gives 450 nm, and transmission electron microscopy gives between 441 to 466 nm depending on the diffraction grating replica used to determine the magnification.

McRae [4] obtained a radius of 288 nm for the same lot of particles of nominal radius of 300 nm, for which we determined a radius of 295 nm.

In table 2 we compare our results for the refractive index to those reported in the literature. We used Bateman's dispersion relation [11].

$$n=a+b/\lambda^2 \tag{21}$$

where a and b are constants and λ is the wavelength of the light, to reduce results obtained with light of other wavelengths to a common wavelength, namely the one we used. The value of b is determined in Bateman's article [11] to be 1.0087×10^{-10} when λ is in centimeters. The constant a is found from the original pair of values of n and λ. Our results are fairly close to the bulk value of 1.615 and not too different from the results obtained in other experiments.

McRae [4] used the same type of instrument that we used and found a stronger size dependence of the refractive index than we did. Before performing the correction of the angle readings, we also found that the values of the refractive index that we obtained varied more with the size of the particle. Heller and Pugh [12] used an interferometer to measure the refractive index of a suspension of polystyrene spheres in water. Their results were most reliable for particle sizes up to 200 nm radius. They also determined the refractive index of polyvinyltoluene spheres.

Smart and Willis [13] made refractive index measurements of different mixtures of glycerol and water and then used a suspension of polystyrene spheres in these liquids to determine the transmission of light through this medium. Then they extrapolated to 100% transmission in a liquid that would have the same refractive index of the polystyrene spheres. The extrapolation method depends on the particle size, and it is simplest for small particles.

6. Concluding Remarks

We have demonstrated that both the size and the refractive index of a single polystyrene sphere with a nominal radius in the neighborhood of 500 nm, comparable to the wavelength of light, can be determined with an error of less than 1% with the Differential II light scattering photometer. The accuracy of this method is better than that achieved by electron microscopy and other procedures.

To obtain accurate results, we find that the angle calibration of the instrument must be made to an accuracy of 0.2° to 0.3°. Using the instrument readout without correction can lead to errors as large as 3 or 4% in the determination of size and refractive index of polystyrene spheres. Furthermore, the magnitude of the errors varies with particle size, so that the refractive index can appear to be size dependent.

The determination of the best estimates of the particle size and refractive index requires a considerable amount of care when the quality of fit surface $Q(n,r)$ has a deep and narrow valley in a diagonal direction (see fig. 7), where correlated changes in both n and r lead to very small changes in Q. This problem occurs frequently in nonlinear least-squares fit procedures. The measurement of the scattered light intensities for both polarizations for the same particle allows for a more reliable estimate of the size and refractive index.

For the smaller values of the radius the surface of fit

is rather flat with a shallow trough, a situation that also leads to difficulties in the determination of n and r, especially with noisy data.

If we establish that the true index of refraction obtained by this method is independent of the particle size, this quantity could be determined accurately from data for a large number of particles or from bulk measurements. We can then use the data from light scattering to make a much faster and more accurate determination of the radius of a particle, especially when a deep valley exists.

Rapid data collection and processing are essential for obtaining the large amount of data necessary for an accurate determination of the size and refractive index of dielectric spheres. Even with the instrument operated at its maximum scan speed, it takes about 10 min to obtain the vertical and horizontal polarization scans and the associated background scans, to do some minor processing on the data, and to store it on tape. The data reduction using a minicomputer (Interdata 7/32) requires about 30 min per particle when Q is computed over a rectangular grid of values of n and r. We have tried a Marquardt-Levenberg algorithm to speed up the search for the minimum value of Q, but preliminary results indicate that this method is difficult to apply and sometimes fails to find the minimum when the surface of fit has a deep narrow valley or when this surface is almost flat. We found it best to perform first a grid search and start a more efficient algorithm from the best fit to refine the results. In any case, it is always best to inspect the resulting curves for the actual fit between the theoretical and measured values.

Noise in the intensity of the scattered light limits the accuracy of the determination of size and refractive index, especially for the smaller particles. The noise might be reduced by filtering out the high-frequency structure in the laser beam to obtain a smooth Gaussian profile. If the noise is a result of the motion of the particle within the laser beam, the ratio of the intensity to that at a fixed angle would be insensitive to this noise. To obtain this ratio, one must add a detector system at a fixed angle to measure this intensity for each pair of intensity-angle readings.

If the Differential II is calibrated to measure absolute intensity by means of monodisperse spheres of known size, the additional information might alleviate the difficulties related to the valleys in the surface of fit. Measurements of absolute intensities may also be helpful to characterize particles with an additional parameter such as a nonnegligible imaginary component of the refractive index or a layered structure.

The authors acknowledge R. Bukowski and E. Braun for their assistance in interfacing the Differential II to a data acquisition system, and R. Young for his advice in designing an angle calibration device.

References

[1] Wyatt, P. J.; Phillips, D. Q. J. Colloid Interface Sci. **39**: 125–135; 1972.

[2] Phillips, D. Q.; Wyatt, P. J.; Bergman, R. M. J. Colloid Interface Sci. **34**: 159–162; 1970.

[3] Cooke, D. D.; Kerker, M. J. Colloid Interface Sci. **42**: 150–155; 1973.

[4] McRae, D. D. J. Colloid Interface Sci. **87**: 117–123; 1982.

[5] Marshall, Q. R.; Parmenter, C. S.; Seaver, M. J. Colloid Interface Sci. **55**: 624–636; 1976.

[6] Mie, G. Ann. Physik **25**: 337–445; 1908.

[7] Born, M.; Wolf, E. *Principles of Optics*, 1980; Oxford: Pergamon Press.

[8] Kerker, M. *The Scattering of Light and Other Electromagnetic Radiation*, New York: Academic Press; 1969.

[9] Yeh, C. Physical Rev. **135**: A1193–1201; 1969.

[10] Bradford, E. B.; Vanderhoff, J. W. J. Appl. Phys. **26**: 864–871; 1955.

[11] Bateman, J. B.; Weneck, E. J.; Eshler, D. C. J. Colloid Sci. **14**: 308–329; 1959.

[12] Heller, W.; Pugh, L. L. J. Colloid Sci. **12**: 294–307; 1957.

[13] Smart, C.; Willis, E. J. Colloid Interface Sci. **25**: 577–583; 1967.

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 88, No. 5, September-October 1983

The Force-Elongation Curve of a Thin Fibrous Network

Jack C. Smith
National Bureau of Standards, Washington, DC 20234

Accepted: May 13, 1983

Specimens from low-density weblike handsheets were tested in a tensile tester. In a test the direction of extension was frequently reversed and the specimen reextended to obtain a series of force-elongation curves. For Kraft woodpulp specimens the force-elongation behavior was well represented by an exponential equation involving three parameters: a modulus of elasticity C_2, a length parameter x_c related to average segment length between network bonds, and an elongation value x_0 at which the curve starts. The unstrained length of a specimen l increases, and the parameters x_c and C_2 tend to decrease with each successive reextension curve. For a series of specimens of increasing area density representative values of x_c/l tend to decrease and C_2 to increase. For a series of specimens made from pulps beaten increasing amounts representative values of x_c/l tend to decrease and C_2 to increase. Some features of the tensile behavior can be modeled by a system of parallel filaments of equal length to which longer parallel filaments with an exponential length distribution have been added. Upon extension the filaments assume load successively, thus simulating the force-elongation behavior of a paper network. By thinking in terms of this model it is possible to anticipate intuitively much of the behavior of a paper network.

Key words: nonwoven fabrics, tensile behavior; paper, tensile behavior; paper fibers, adhesion; paper fibers, bonding; paper pulps, characterization.

1. Introduction

In the making of paper, a mat of fibers usually obtained from woodpulp is deposited from a slurry onto a screen. The fibers in the mat bond together to form paper, the quality of which depends upon the kind of pulp and additives used, the beating treatment to which the pulp is subjected, and the subsequent processing of the mat after it has been formed. Such important paper properties as strength and durability are dependent upon the strength and permanence of the interfiber bonds. Therefore, a knowledge of pulp characteristics associated with good bonding could be very useful in making paper from recycled pulps, for example, or in producing papers with improved durability.

In order to estimate bond strength, it is necessary to determine the behavior of a paper specimen subjected to mechanical strain and to interpret this behavior in terms of a suitable theory. To this end Dodson [1][1] has reviewed literature on the nature of bonds in paper and the behavior of paper under mechanical strain. The experimental and theoretical behavior of nonwoven fabrics subjected to tensile stress has been studied by Backer and Petterson [2,3] and by Hearle et al. [4–8]. Dent [9] has reviewed existing theories for calculating Young's modulus of filament webs.

The "two-dimensional" fiber network discussed by Kallmes and Corte [10] is especially suitable for studies of bond strength. In such a network the fibers lie in a single plane except at the fiber-fiber contacts; most of the contacts involve only two fibers and the total area of the crossings is a small fraction of the sheet area.

About the Author, Paper: A physicist, Jack C. Smith is a guest worker in NBS' Polymer Standards and Science Division. The work on which he reports was sponsored in part by the U.S. Department of Energy and the U.S. Department of Treasury.

[1] Figures in brackets indicate literature references at the end of this paper.

Such a network can be approximated, for instance, by a handsheet of 2.5 g/m^2 mass per unit area made from spruce sulfite pulp. The structure of this network can be subjected to theoretical analysis, and handsheets approximating this network are sufficiently durable for testing.

If a specimen cut from such a handsheet is stretched to break in a sensitive tensile tester, a jagged force-elongation curve is obtained in which the force rises and dips through one or more peaks, and eventually diminishes to zero. Such a curve is exemplified by the recorder trace, figure 1, obtained by testing a specimen from a 2.5 g/m^2 handsheet made from Southern Kraft pulp that had been beaten 5,000 revolutions in a laboratory beater. The specimen, 2 cm

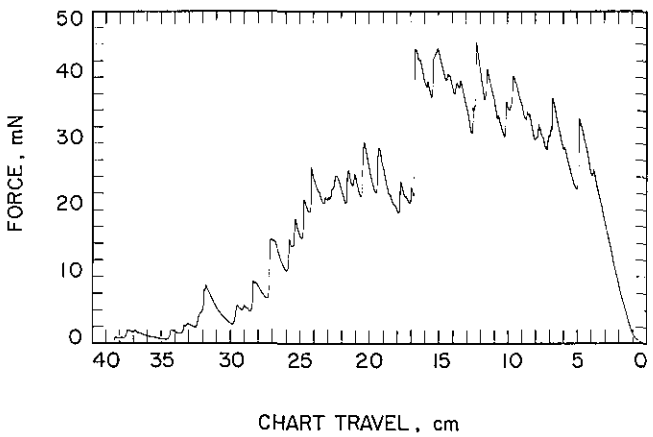


Figure 1–Force-elongation curve for a handsheet specimen of beaten Southern Kraft woodpulp, 2.5 g/m^2 mass per unit area.

long and 1 cm wide, was extended at a rate of 0.2 cm/min; 1 cm of chart travel corresponds to an extension of 0.01 cm, and full-scale force for the recorder chart is 49 mN (5g).

Each jag in the curve signifies the breaking of a bond between fibers. (It is assumed that the fibers are much stronger than the bonds between them and that the fibers themselves do not break.) When a bond break occurs, energy previously stored in the network is dissipated, and Dodson [11] has suggested that the average strain energy dissipated per bond break could be used as an energy parameter to characterize bonding. Smith [12–15] has evaluated this and other parameters from tests on a number of different handsheets, but the resulting values were of low precision. These values were obtained by a data analysis partly dependent upon subjective judgment.

It is believed that much of this subjectivity could be eliminated were it possible to characterize the force-elongation behavior of a test specimen at various stages during a tensile test. A semiempirical method for accomplishing this was discovered, but the method must be checked by additional experiments before its applicability can be established. This cannot be done at the present time inasmuch as this research has been discontinued. The method, however, seems promising and is presented herewith.

2. Characterization of the Force-Elongation Curve

During a tensile test the following behavior sequence is observed in the fibrous network: the fibers tend to align themselves in the direction of the extension with the straightest and most perfectly aligned fiber segments bearing most of the load. As the specimen is elongated it contracts laterally forming a fluted structure in the portions not restrained by the end clamps. Bond breaks often occur at random points, causing only local damage and leaving the surrounding network essentially intact. However, a breakage of some bonds, usually in the vicinity of network flaws or irregularities, results in more extensive damage. Often a succession of breaks occurs in the same region because of stress concentrations that develop. In the regions of more extensive damage, parts of the network lose their strain and sag away. As the extension continues, further deterioration occurs. Portions of the network midway between the ends of the specimen tend to collapse into threadlike structures in which the longitudinal forces are concentrated. These forces in turn are diffused through the uncollapsed portions of the network at either end. In the final stages of the breaking process, often only one threadlike structure remains. The specimen then sustains relatively large elongations before breaking.

Consider the recorder trace, figure 2, obtained by testing a specimen from a 2.5 g/m^2 handsheet made from Northern Kraft woodpulp that had been beaten 5,000 revolutions in a laboratory beater. The specimen, 2 cm long by 1 cm wide, was extended at a rate of 0.2 cm/min. Chart travel of 1 cm corresponds to a speciman extension of 0.01 cm, and full-scale force for the specimen is 196 mN (20 g). At intervals the test was stopped, the extension was reversed until the tensile force in the specimen was zero, and the specimen then reextended to obtain a new force-elongation curve. Recording traces during the reversal of the crosshead travel are not shown.

The general condition of the specimen at various stages is noted in figure 2. For the first eight reextensions the specimen did not develop any large holes or tears, and the state of the specimen at the

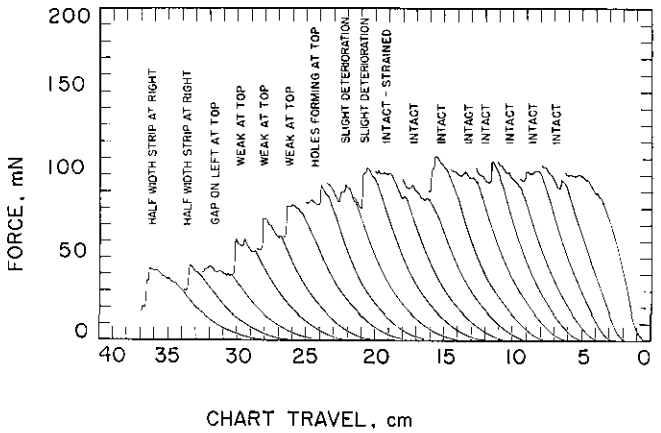


Figure 2–Cycled force-elongation curves for a handsheet specimen of Northern Kraft woodpulp, 2.5 g/m^2 mass per unit area.

beginning of the reextension is described as "intact." At later stages of the test the specimen developed holes and tore at the edges as it deteriorated.

In figure 3 are shown plots of F', or force per unit elongation, versus force F for the first 11 reextended force-elongation curves of figure 2. The origin of the F' scale has been shifted vertically for each of these plots, and arbitrary units have been selected to display the plots to best advantage. The curve corresponding to each plot is identified by its region of chart travel. The plots are seen to be linear and to have positive intercepts on the F' axis. Plots similar to these have resulted from most of the tests on specimens formed from woodpulps.

It will be noticed that the early points corresponding to small values of F usually fall below the straight line. At these beginning force values the slope of the force-elongation curve increases rapidly as the network fibers start to orient. It is only after the initial orientation that the network behaves so that F' increases linearly with F. Also it has been found in general that F, F' data from the initial force-elongation curve in a test do not fit well to a straight line. In this instance F' usually rises rapidly, becomes constant until F is moderately large, and then increases linearly with F. Data from the first two reextension curves often tend to plot in a similar fashion. This behavior is thought to arise from an initial maldistribution of forces in a network. It is only after the network has been "broken in" that F' has a good linear correspondence with F. In the plot of figure 3, data from the initial force-elongation curve therefore have been omitted.

Sometimes when a test specimen is unloaded the recorder trace does not return to the original zero force level. This behavior is usually an effect of

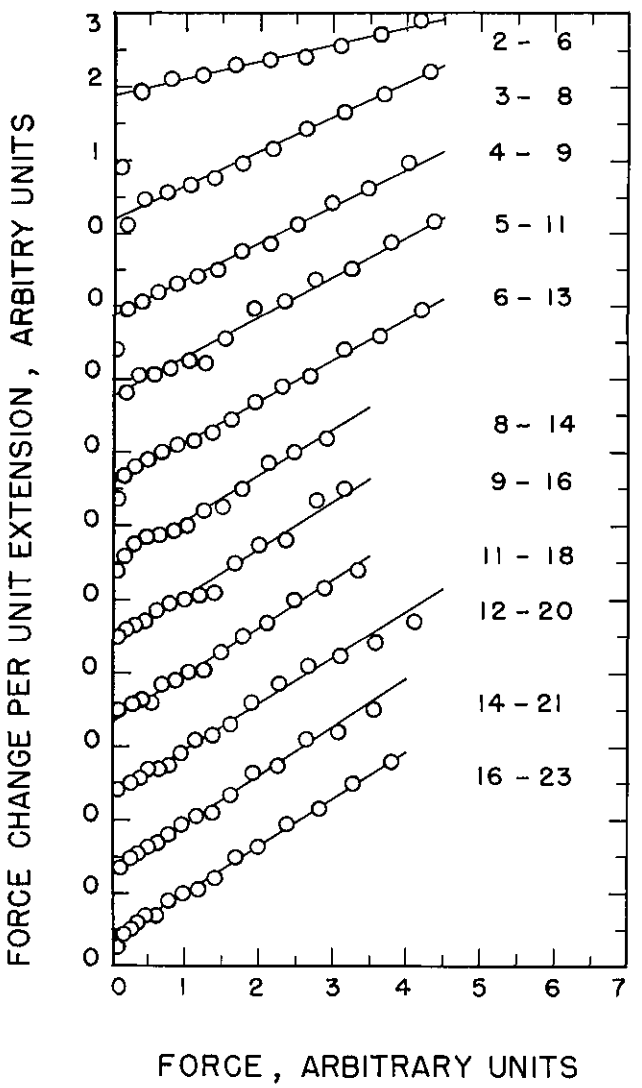


Figure 3–Plots of force change per unit extension versus force for curves of figure 2. Numbers at right give the region of chart travel in cm for each associated force-elongation curve.

hysteresis. Its consequences can be minimized by waiting a few minutes before reextending the specimen, but it is more practical to reextend immediately and to correct data from the new curve for the change in baseline. Sometimes a shift of baseline is due to a change in the recorder system, a circumstance especially likely because of the high sensitivities required by these tests. Thus a correction for baseline shift should be made in any event.

Force-elongation curves of these paper network specimens exhibit viscoelastic, or hysteretic behavior. Thus the slopes of the force-elongation curves will depend upon the rate of extension. The rate of extension therefore must be maintained at some standard value in all of the tests in order to obtain comparable results. This precaution should minimize

the influence of viscoelasticity, but the possibility of other hysteretic effects should be kept in mind in the interpretation of data.

Let the equation of a straight line faired through a plot such as those shown in figure 3 be

$$\frac{dF}{dx} = \frac{1}{x_c}(C_2+F) \tag{1}$$

where x_c is the reciprocal of the slope of the straight line, and C_2/x_c is the intercept on the F' axis. This differential equation can be integrated to obtain

$$F=C_1e^{x/x_c} - C_2 \tag{2}$$

In eqs (1) and (2) F is the tensile force sustained by the specimen, and x is a distance variable measured in the direction of the elongation. C_1, C_2 and x_c are constants characterizing the force-elongation curve.

Ordinarily the values of C_1, C_2 and x_c would be obtained by fitting eq (2) to a curve, using the least squares technique. This method, however, involves solving a complicated transcendental equation, so a procedure providing a good approximate solution is used instead. The x, F data from the curve to be fitted is differentiated to obtain values of F', and eq (1) is fitted to the F, F' data to evaluate C_2 and x_c. Equation (2) with these known values of C_2 and x_c is then fitted to the curve to evaluate C_1.

According to eq (2), F becomes zero when

$$x=x_0=x_c\ln\ (C_2/C_1) \tag{3}$$

Thus eq (2) can be put in the form

$$F=C_2[e^{(x-x_0)/x_c}-1] \tag{4}$$

x_0 is the value of the distance variable at which the force just increases from zero, so $x-x_0$ is the elongation of the network. As the specimen fibers become better aligned, and as the specimen deteriorates, the unstrained length of the specimen increases. Thus for each successive reextension curve of figure 2, successively increasing values of x_0 should be expected.

Equation (2) or (4) usually fits test data on woodpulp specimens very well. For instance, values of C_1, C_2, x_c and x_0 were found for each of the force-elongation curves of figure 2, and corresponding curves, eq (4), were calculated. Plots of these calculated curves superposed the original data for practical purposes. The only noticeable deviations occurred at small values of F. The calculated curves started to rise at slightly larger values of x, and had larger initial slopes.

It should be emphasized that the force-elongation curves characterized here were obtained from tests on specimens formed from Kraft woodpulps. In these pulps strong bounds are formed only between adjacent fibers that are compatible in shape and are in intimate contact. Bondability is enhanced by a beating treatment that causes the fibers to become pliable and to assume a ribbonlike shape. In some other types of pulps beating causes fibrillation, and the bonds formed consist of many microbonds between the fibrils and adjacent fibers. Similar strong bonds extending over a large area are formed when the beaten pulp contains appreciable amounts of fines which act as a glue.

Test data obtained on network specimens formed from these latter pulps differ from data obtained on woodpulp specimens. In the plots of F' versus F, F' rises slowly at first or stays at a constant value, and only increases linearly with F after F has become moderately large. A straight line through this late linear portion usually has a negative intercept on the F' axis. Force-elongation data for these pulps cannot be fitted by eq (4). A more complicated equation involving an additional linear or exponential term is required.

3. A Mechanism for Force-Elongation Behavior

When a test specimen from a fibrous network is stretched, segments of the fibers composing the network tend to orient themselves in the direction of the extension and to resist the extension through the combined action of forces along their lengths. As the extension proceeds, more and more of these segments between bonds become oriented and bear load. Other fiber segments have a curved configuration between bond points, and cannot bear load until the network is stretched sufficiently to straighten them out. As a result of these processes the force-elongation curve has a slope that increases with increasing elongation. This reaction of the network to extension can be simulated by a system of parallel filaments of unequal lengths, each filament adding to the resistive force of the system as the extension is increased. The flexural stiffness of the fibers and the restraining action of the clamps at the ends of the specimen also contribute to the network's reaction to stretching, but these and

other effects are considered secondary in importance and can be disregarded.

Consider as a model a system of parallel filaments of unequal length. Let the unstrained length of the system be l, and let there be ν filaments of this length. Let the length of any other filament be $l+\eta$, and let the number of filaments having length increments between η and $\eta+d\eta$ be $\nu f(\eta)d\eta$. Let each of the filaments be linearly elastic with spring constant k, neglecting the decrease in k with increase in unstrained length. If this system of parallel filaments is extended from the length l to a length $l+\xi$, the sum S of the tensile forces in the filaments is

$$S=\nu k\xi+\nu k\int_0^{\xi} f(\eta)(\xi-\eta)d\eta \tag{5}$$

To find the proper value of $f(\eta)$, S can be set equal to the force F, as defined by eq (4). Thus if it is assumed that

$$f(\eta)=\frac{1}{x_c}e^{\eta/x_c} \tag{6}$$

eq (5) becomes

$$\begin{aligned} F&=\nu k\left\{\xi+\frac{1}{x_c}\int_0^{\xi}e^{\eta/x_c}(\xi-\eta)d\eta\right\}\\ &=\nu k\left\{\xi+x_c\left[e^{\xi/x_c}-\frac{\xi}{x_c}-1\right]\right\}\\ &=\nu kx_c\left[e^{(x-x_0)/x_c}-1\right] \end{aligned} \tag{7}$$

upon substituting the elongation $x-x_0$ for ξ. Thus $f(\eta)$ is properly defined by eq (6), and comparison of eqs (7) and (4) yields the relationship

$$C_2=\nu kx_c \tag{8}$$

The model is thus seen to consist of ν filaments of unstrained length l in parallel with additional filaments of length $l+\eta$ distributed such that the number of filaments with length increments between η and $\eta+d\eta$ is

$$\frac{\nu}{x_c}e^{\eta/x_c}d\eta.$$

If such a model is elongated to a length $l+\xi$, the number of filaments should be equal to $\nu e^{\xi/x_c}$. It would be unrealistic to expect that this model would apply in situations were ξ becomes very large because too many stretched filaments would be required. In fact it is often observed that at large elongations the curvature of the force-elongation curve does not increase as rapidly as eq (1) would require. At these elongations, however, bond breaks occur; thus the exponential distribution just derived for filaments having lengths greater then l is adequate for practical purposes.

The response of a fibrous network to tensile strain may now be understood more clearly. Suppose that a typical test specimen (1 cm wide by 2 cm long) is gathered together at each end, and extended to form a threadlike structure. In an ideal network in which there are no fluctuations in density and the fibers have no flexural stiffness, no force would be required during this part of the extension. The threadlike structure would consist of a number of interconnected fiber pathways of perfectly oriented fiber segments. Most of the fiber segments interlinking the pathways would hang in slack loops, and would not sustain force until the structure is extended or until bonds maintaining the shorter pathways are broken. A test specimen with the original width maintained at the clamped ends would approximate the response just described for this idealized threadlike structure.

The tensile response of the threadlike structure when elongated would be similar to that of the parallel filament model, but the mechanism of the response is different. Instead of independent filaments there are interconnected pathways, and instead of additional filaments that assume load, interconnections between pathways become oriented and assume load as the elongation progresses. The number ν of initially strained filaments and the distribution function $f(\eta)$ for the additional filaments no longer have precise meanings, but are still useful for interpretations of behavior in terms of the parallel filament model.

From an inspection of eq (4) it is seen that x_c has dimensions of length. If the magnitude of x_c is small, the force-elongation curve will rise rapidly as $x-x_0$ increases. This is the kind of behavior that might be expected to result from testing a network in which the fiber segments between bonds are small. This small segment size insures that additional parallel pathways that sustain force are rapidly formed during extension of the network. This suggests that x_c is a characteristic length having a value that increases or decreases with average segment size of the network. For a given network the value of x_c found from a test also depends upon the length l of the test specimen. In order that force-strain curves be comparable for specimens of different length taken from the same handsheet, it is necessary that the characteristic strain $\epsilon_c=x_c/l$ have a

constant value characteristic of the network. Thus the value of x_c should vary proportionally with the specimen length.

It is seen from eq (4) that C_2 is a parameter having dimensions of force; therefore its value should not depend upon specimen length. From eq (8) the value of C_2 for the parallel filament model is expressed as the product of ν initially strained filaments and the quantity kx_c. But k, the spring constant of the filament, can be set equal to m/l, where m is the tensile modulus of elasticity of a filament. Thus eq (8) becomes

$$C_2=\nu\frac{m}{l}x_c=\nu m\epsilon_c \tag{9}$$

Although the value of C_2 may be independent of specimen length, it is expected to have a value that increases as the width increases. For long specimens the value of C_2 should be directly proportional to the width.

4. Effect of Deterioration

During a test the unstrained length of the specimen just before each reextension is given by

$$l=l_0+x_0-x_s \tag{10}$$

where l_0 is the original unstrained length, x_0 is calculated for each reextension curve using eq (3), and x_s is the value of x at which the initial force-elongation curve just increases from zero. For the curves of figure 2 the test specimen had an initial unstrained length l_0 of 20.0 mm. The value of x_s, 5.0 mm on the chart, was 0.05 mm after multiplying by a scale factor of 0.01. In eq (10) the values of l_0 and x_s are constant for each of the successive reextension curves obtained in a test, but the value of x_0 increases with each successive curve because of progressive deterioration. Thus l, the unstrained length of the specimen just before each reextension, also increases with progressive deterioration, and the fractional increase in unstrained length l/l_0-1 might be taken as a measure of deterioration. The values of the parameters C_2 and x_c characterizing the successive reextension curves also change with progressive deterioration. These changes of C_2 and x_c as functions of deterioration, measured by l/l_0-1 are studied here.

Table 1 gives the values of constants characterizing the force-elongation curves of figure 2. According to this table the characteristic strain x_c/l for the specimen decreases in value as the specimen deteriorates. This behavior is demonstrated in figure 4 where values of x_c/l for the 2nd through the 11th reextension curves are plotted versus the corresponding values of l/l_0-1. The value of x_c/l for the first reextension curve was exceptionally high and was not plotted. A straight line approximating the dependence of x_c/l upon l/l_0-1 is passed through the data points determined from the third through the eighth reextension curves. Evidently the characteristic strain x_c/l decreases rapidly at first relative to deterioration as measured by l/l_0-1, but as the test progresses x_c/l decreases more slowly and in a more linear fashion. Similar behavior is almost always observed in tests on specimens formed from woodpulp.

According to table 1 the constant C_2 decreases very rapidly as the specimen deteriorates during a test. From eq (9) part of this decrease in value is attributable to decrease in value of x_c/l and the remainder to decrease in value of $\nu m=lC_2/x_c$. The decrease in value of lC_2/x_c is still very rapid but may

Table 1. Values of constants C_2, x_c and x_0 for curves of figure 2.

Curve	x_0 (mm)	l (mm)	l/l_0-1 (%)	x_c (mm)	C_2 (mN)	x_c/l (%)	$\log_e lC_2/x_c$ ([1])
2–6	0.21840	20.168	0.842	0.84484	153.66	4.1889	8.2075
3–8	0.31207	20.262	1.310	0.42984	49.456	2.1214	7.7542
4–9	0.41506	20.365	1.825	0.38783	31.634	1.9044	7.4152
5–11	0.53146	20.482	2.407	0.35375	24.750	1.7272	7.2675
6–13	0.65520	20.605	3.026	0.35872	21.343	1.7409	7.1115
8–14	0.79018	20.740	3.701	0.33553	16.005	1.6178	6.8970
9–16	0.93808	20.888	4.440	0.33543	14.312	1.6058	6.7926
11–18	1.0905	21.041	5.203	0.32164	11.572	1.5287	6.6293
12–20	1.2613	21.211	6.057	0.33259	11.854	1.5680	6.6280
14–21	1.4338	21.384	6.919	0.31462	9.500	1.4713	6.4703
16–23	1.6796	21.630	8.148	0.30249	10.861	1.3985	6.6549

[1] C_2 is expressed in mN units, x_c/l as a fraction.

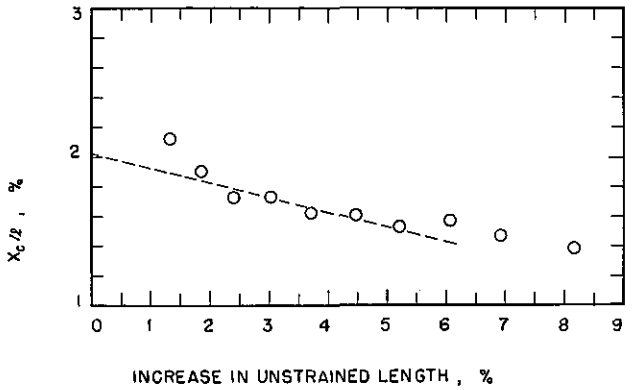


Figure 4–Plot of x_c/l versus percent increase in specimen length, data of table 1.

be mitigated for comparison purposes by plotting the natural logarithm of this quantity versus the fractional increase in length l/l_0-1, as is done in figure 5. A straight line approximating this dependence is passed through data points determined from the third through the eighth reextension curves. The trend of this plot is similar to that observed for test data from other specimens formed from woodpulp.

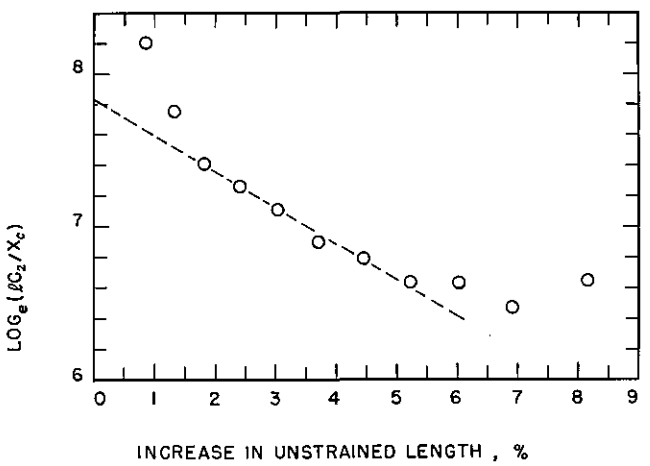


Figure 5–Plot of $\log_e(lC_2/x_c)$ versus percent increase in specimen length, data of table 1. C_2 is expressed in mN units, x_c/l as a fraction.

A mechanism accounting for the decrease of x_c/l and $lC_2/x_c=\nu m$ as the specimen deteriorates is not intuitively evident. However one might speculate that the following process occurs: assume that the specimen network is formed of fiber segments of size distribution such that longer segments occur less frequently than shorter segments. Assume also that the longest segments are the most likely to sustain forces and thereby dominate the force-elongation behavior of the network. Thus for force-elongation curves obtained early in a test the fiber pathways sustaining force are channeled through the long segments. In this instance the network might react to an elongation as though it were a coarse network composed of a small number of fiber threads with long segment lengths between bonds. The measured value of $lC_2/x_c=\nu m$ would be proportional to the modulus of one of these threads, and thus be large. The measured value of x_c/l would be proportional to the segment length, and thus be large also. Bonds associated with the long segments are subjected to large forces and tend to break preferentially, so that as the test proceeds the fiber pathways sustaining force are channeled through the more numerous fiber segments of shorter length. Thus the values of x_c/l and $\nu m=lC_2/x_c$ tend to decrease progressively as the test continues.

The effects of the mechanism just described are observed at that stage of the test when the specimen is "broken in" but still intact. For the initial and first one or two reextension curves extremely high values of x_c/l and lC_2/x_c are observed, and these values decrease rapidly as the test proceeds. Much of this initial behavior may attributed to fluctuations in density of fiber distribution in the specimen, which causes the forces in the fiber segments to be distributed nonuniformly. The bonds subjected to the high force concentrations are rapidly broken and the force distribution becomes more uniform.

In many specimens data points obtained from successive reextension curves tend to plot in a scattered fashion so that it is not always possible to fit a smooth curve to them with good confidence. This behavior may also be attributed to density distributions which cause successive bond breaks to be localized during a test, the locality of the breaks shifting as the test progresses.

In the later stages of a test, holes form in the specimen and deterioration is apparent visually. Data points for x_c/l and lC_2/x_c obtained when the specimen is in this condition tend to fluctuate and to plot above the trend line fitted to "intact" data. The last three data points in figures 4 and 5 demonstrate this effect. Conceivably at this stage of a test the force distribution in a specimen has become sufficiently uniform that values of x_c/l and lC_2/x_c do not change much until gross deterioration occurs.

5. Materials Studied and Experimental Procedures

Low-density open-web handsheets were prepared from Northern and Southern Kraft woodpulps. Fibers from the Northern woodpulp were thin and easily

assumed a ribbonlike configuration after a moderate amount of beating. The Southern woodpulp fibers were coarser and needed more beating to become pliable and attain a ribbonlike texture. For handsheets of the same mass per unit area made from the two pulps the Southern pulp handsheets had fewer fibers per unit area and fewer bonds per unit area than did the handsheets of Northern pulp. Methods used to prepare the handsheets are described elsewhere [12,13].

Force-elongation curves were obtained for a series of Northern Kraft handsheets having densities varying from 1.5 to 3.5 g/m^2, and for a series of Southern Kraft handsheets having densities varying between 1.75 and 3.5 g/m^2. Force-elongation data were also obtained for handsheets prepared from Northern and Southern Kraft woodpulps that had been subjected to various amounts of beating. In addition, data were obtained on specimens of different dimensions taken from the same handsheet in order to study the effects of specimen size and shape.

Tensile tests were conducted on specimens 2 cm long and 1 cm wide except for several specimens of 5 cm length. Crosshead speed of the tensile tester was 0.2 cm/min for the specimens of 2 cm length and 0.5 cm/min for the specimens of 5 cm length. Chart speed was 20 cm/min. During the tests the direction of extension was frequently reversed to unload the specimens and then resumed to obtain a series of force-elongation curves such as those shown in figure 2. The constants C_2, x_c, x_0 and l associated with each of these curves were determined by the methods described previously.

Tests were performed in a laboratory maintained at 20 °C, 50 percent R.H. on specimens subjected to this environment for at least 12 hours before testing. The handsheets from which the specimens were obtained were not always stored under these controlled conditions, and may have changed somewhat during the time interval between preparation and testing.

In the experimental verifications to be discussed it is expedient to characterize the tensile behavior of each test specimen by representative values of x_c/l and lC_2/x_c. These values were obtained by plotting x_c/l and $\log_e(lC_2/x_c)$ versus the corresponding values l/l_0-1, and extrapolating to the origin where $l/l_0-1=0$, as was done in figures 4 and 5. The extrapolated values thus obtained are quite imprecise and the extrapolation process involved some degree of subjective judgment. The precision could have been improved by additional testing especially devoted to this purpose, but the improvement was not deemed necessary for these demonstrations where the intent is merely to show a qualitative trend of the values of x_c/l and lC_2/x_c that is consistent with expected behavior.

6. Effect of Specimen Size and Shape

A short series of tests was made to see if the values of C_2 and x_c/l depend upon the size and shape of the specimen. For these tests specimens of width × length 1×2, 1×5 and 2×5 cm were cut from the same handsheet. The handsheet was of 2.5 g/m^2 density formed from Northern Kraft woodpulp beaten 5,000 revolutions in a laboratory beater. Extrapolated values of x_c/l and $\log_e(lC_2/x_c)$ derived from these tests are given in table 2. Values of C_2 calculated from these derived values are also given.

Table 2. Extrapolated values of x_c/l and $\log_e(lC_2/x_c)$ for different size specimens from the same handsheet

Specimen (w×l cm)	x_c/l [1] (%)	$\log_e(lC_2/x_c)$ [2]	C_2 (mN)
1×2	2.16±0.52 [3]	8.24±0.30 [3]	82
1×5	2.04±0.34	7.81±0.22	50
2×5	1.85±0.13	8.70±0.22	111

[1] Average of extrapolations from five specimens.
[2] C_2 is expressed in mN units, x_c/l as a fraction.
[3] Standard deviation.

Values of x_c/l for the three specimens are approximately the same, indicating that the value of x_c/l is not dependent upon the width of the specimen. This observation is consistent also with the idea that x_c/l is a strain parameter independent of specimen length and characteristic of the handsheet network. It is interesting to note that the standard deviations of the values of x_c/l decrease markedly as the areas of the specimens increase. This suggests that as the area of the specimen increases the effects of textural inhomogeneities are smoothed out.

The value of C_2 for the 2×5 cm specimen is approximately twice as large as that for the 1×5 cm specimen, in agreement with the expectation that for long specimens C_2 has a value directly proportional to specimen width. One might expect the value of C_2 for the 1×2 cm specimen to be closer to that for the 1×5 cm specimen; however the length of this specimen is quite short relative to the width, so that specimen shape can also influence the value of C_2 in this case.

The tests just described indicate that specimen dimensions of 2×5 or 1×5 cm yield the most precise results. But the 1×5 cm specimen might be preferable for certain tests, because with it the force drops that

occur during a test are more clearly delineated. In addition, the long narrow configuration tends to minimize edge effects at the clamped ends, which prevent proper orientation of the network fibers and an even sharing of the tensile force throughout the width of the specimen.

7. Effect of Handsheet Density

According to the previous discussion it is expected that lC_2/x_c should increase as the density of the test specimen is increased, because the higher density specimens have more fibers per unit area and hence more pathways to sustain a tensile force. It is also expected that x_c/l should decrease as the density of the test specimen is increased, because the higher density specimens have smaller segement lengths between bonds. In order to verify these expectations, extrapolated values of $\log_e(lC_2/x_c)$ and x_c/l were obtained for a series of Northern and a series of Southern Kraft woodpulp handsheets of various densities. These values and their standard deviation limits are plotted versus handsheet density in figures 6 through 9.

The extrapolated values of $\log_e(lC_2/x_c)$ for the Northern and Southern pulps tend to increase and the values of x_c/l tend to decrease as the handsheet density increases, in agreement with expectations. Linear regression lines have been drawn through the data points to make this more obvious. Although not demonstrated directly, values of C_2 for the Northern and Southern pulps also increase with handsheet density. As the values of lC_2/x_c increase exponentially and the values of x_c/l decrease only linearly, the product of these quantities, C_2, should therefore

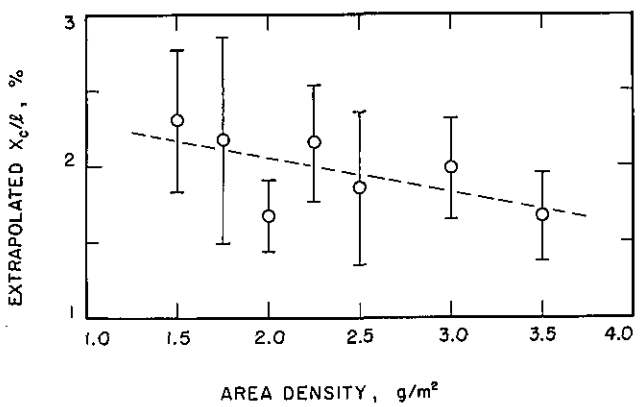


Figure 6–Plot of extrapolated x_c/l values versus area density for handsheets formed from Northern Kraft woodpulp.

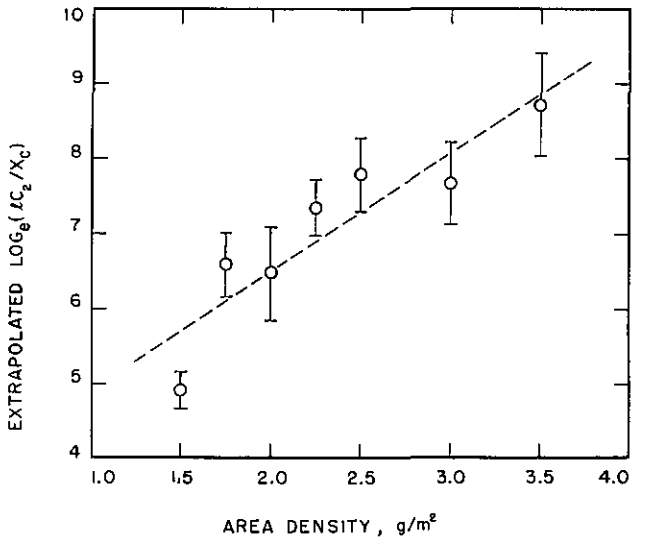


Figure 7–Plot of extrapolated $\log_e(lC_2/x_c)$ values versus area density for handsheets formed from Northern Kraft woodpulp. C_2 is expressed in mN units, x_c/l as a fraction.

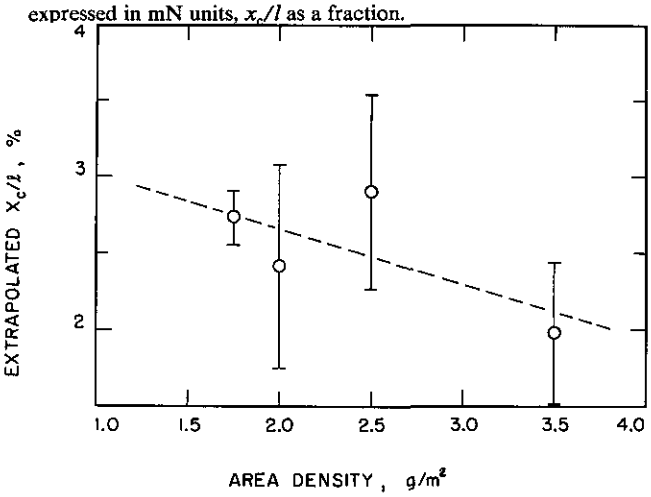


Figure 8–Plot of extrapolated x_c/l values versus area density for handsheets formed from Southern Kraft woodpulp.

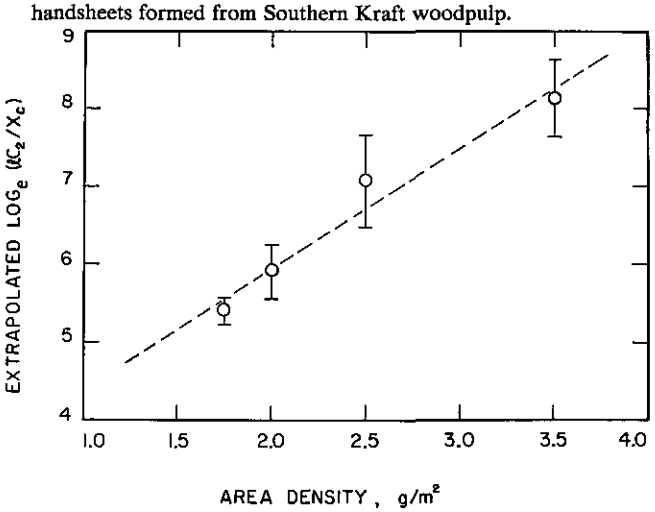


Figure 9.–Plot of extrapolated $\log_e(lC_2/x_c)$ values versus area density for handsheets formed from Southern Kraft woodpulp. C_2 is expressed in mN units, x_c/l as a fraction.

increase. This increase of C_2 with handsheet density has been confirmed by direct calculation.

The fibers in the Southern pulp furnish are coarser than the fibers in the Northern pulp furnish. Thus a handsheet formed from the Southern pulp would have fewer fibers per unit area than would a Northern pulp handsheet of the same area density. The segment lengths and hence x_c/l for the Southern pulp specimen thus would be larger. The data of figures 6 and 8 confirm these expectations. According to figure 6 the value of x_c/l read from the regression line is 0.0194 for a Northern pulp handsheet specimen of density 2.5 g/m^2. Similarly, from figure 8, the value of x_c/l is 0.0248 for a Southern pulp handsheet specimen. This larger value for the Southern pulp handsheet is in agreement with expectations.

A similar intuitive estimate of the relative values of C_2 for Northern and Southern pulp handsheets does not seem possible. However values of $\log_e(lC_2/x_c)$ for 2.5 g/m^2 handsheets can be read from figures 7 and 9. The corresponding values of lC_2/x_c are 1480 mN for the Northern pulp handsheet and 829 mN for the Southern pulp handsheet. Multiplying by the values of x_c/l found from figures 6 and 8 yields C_2 values of 28.7 mN for the Northern pulp handsheet and 20.5 mN for the Southern pulp handsheet.

8. Effect of Beating

Extrapolated values of x_c/l and $\log_e(lC_2/x_c)$ were determined for a series of handsheets of the same density prepared from Northern Kraft woodpulp beaten various amounts in a laboratory beater. These values were also determined for a similar series of handsheets prepared from Southern Kraft woodpulp. The results are plotted in figures 10–13. The effect of beating should be to increase the number of bonds per unit area of the handsheet, so that the number of pathways sustaining tensile force should increase, and the average segment length between bonds should decrease with the degree of beating. Thus the values of x_c/l should decrease and those for $\log_e(lC_2/x_c)$ should increase with the degree of beating. Inspection of figures 10 and 11 shows that these expectations are borne out for the Northern Kraft pulp handsheets. The dependence of x_c/l and $\log_e(lC_2/x_c)$ upon beating is not so pronounced for the Southern pulp. The linear regression lines plotted in figures 12 and 13, however, show the expected trend.

9. Summary

It has been suggested that the strength and durability of bonds between fibers in paper can be studied by preparing low-density "two-dimensional" handsheets from paper pulps and subjecting specimens cut from these handsheets to tensile tests. If such a specimen is stretched to break, a jagged force-elongation curve is obtained which rises and dips through one or more peaks, and eventually diminishes to zero. This behavior is interpreted to mean that the specimen deteriorates slowly during a test as the interfiber bonds break successively. Thus if during a test the extension were frequently reversed until the tensile force in the specimen were zero, and the specimen then reextended, a series of force-elongation curves would be obtained. By studying the changes in these curves, deterioration in the specimen could be evaluated and some means of compensating for this

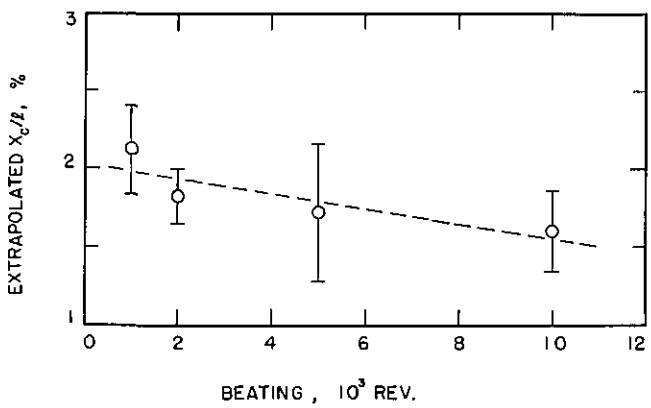


Figure 10.–Plot of extrapolated x_c/l values versus degree of beating for handsheets formed from Northern Kraft woodpulp.

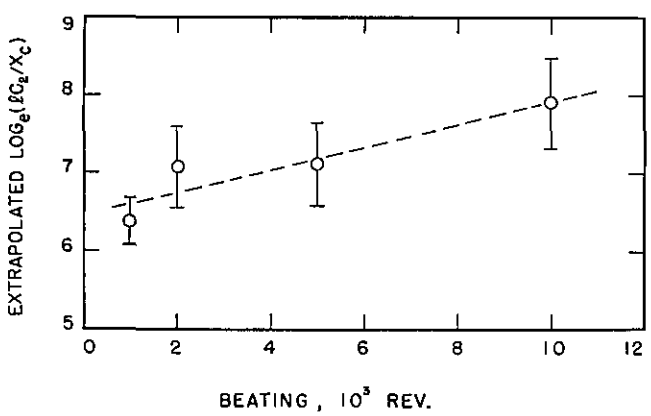


Figure 11.–Plot of extrapolated $\log_e(lC_2/x_c)$ values versus degree of beating for handsheets formed from Northern Kraft woodpulp. C_2 is expressed in mN units, x_c/l as a fraction.

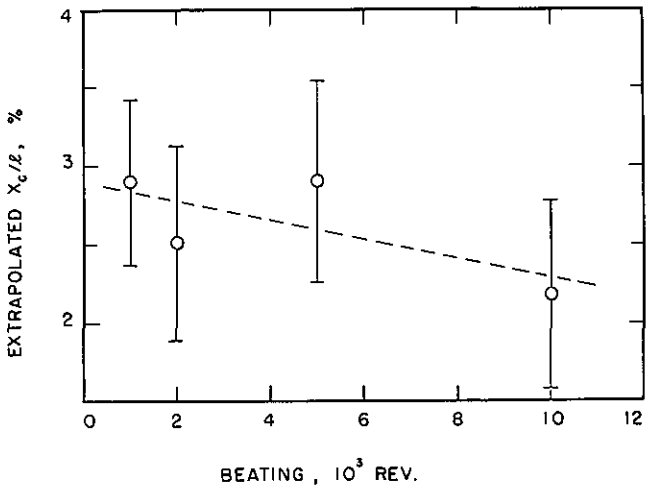


Figure 12.–Plot of extrapolated x_c/l values versus degree of beating for handsheets formed from Southern Kraft woodpulp.

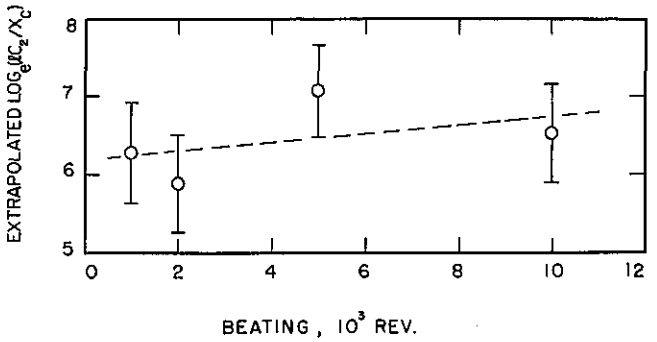


Figure 13.–Plot of extrapolated $\log_e(lC_2/x_c)$ values versus degree of beating for handsheets formed from Southern Kraft woodpulp. C_2 is expressed in mN units, x_c/l as a fraction.

deterioration devised. It might then be possible to reduce data taken at various stages of a test to a comparable basis.

The tensile behavior of low-density specimens formed from Northern and Southern Kraft woodpulps was studied. The force-elongation curves were found to be well represented by an equation of the type $F=C_2[\exp(x-x_0)/x_c-1]$, where F is the force and C_2 is a parameter having the dimensions of force, x is an elongation variable, x_0 is the elongation at which the force just rises from zero, and x_c is a parameter having the dimensions of length. The unstrained length of a specimen at the start of a test is l_0, but the unstrained lengths l corresponding to successive force-elongation curves have increasing values. Thus the value of l/l_0 increases as the test progresses, and can be used as a measure of deterioration. It is found that the value of x_c/l tends to decrease gradually, and the value of C_2 decreases rapidly as the specimen deteriorates. If representative values of x_c/l and C_2 are determined for handsheets of various densities, it is found that x_c/l decreases and C_2 increases as the density of the handsheet increases. Also x_c/l decreases and C_2 increases for handsheets made from pulps subjected to increasing amounts of beating.

Some features of the tensile behavior just described can be modeled by a system of parallel filaments of equal length to which longer parallel filaments with an exponential length distribution have been added. As this system is extended the filaments assume load successively, and the force-elongation behavior is governed by the same mathematical relationship as that found for the "two-dimensional" handsheet specimens. By thinking in terms of this model it is possible to anticipate intuitively much of the behavior just described for the handsheet specimens.

The author is indebted to Dr. Edmond L. Graminski (Bureau of Engraving and Printing) for helpful discussions, and to Mrs. Elizabeth E. Toth (National Bureau of Standards) for preparing the samples tested.

References

[1] Dodson, C. T. J. Rep. Prog. Phys. **33**: 1–43; 1970.
[2] Backer, S.; Petterson, D. R. Text. Res. J. **30**: 704–11; 1960.
[3] Petterson, D. R.; Backer, S. Text. Res. J. **33**: 809–16; 1963.
[4] Hearle, J. W. S.; Stevenson, P. J. Text. Res. J. **33**: 877–88; 1963.
[5] Hearle, J. W. S.; Stevenson, P. J. Text. Res. J. **34**: 181–91; 1964.
[6] Hearle, J. W. S.; Newton, A. Text. Res. J. **37**: 778–97; 1967.
[7] Hearle, J. W. S.; Newton, A. Text. Res. J. **38**: 343–51; 1968.
[8] Hearle, J. W. S.; Ozsanlav, V. J. Text. Inst. **70**: 19–28; 1979.
[9] Dent, R. W. J. Engr. Ind., Trans. ASME **102**: 360–65; 1980.
[10] Kallmes, O.; Corte, H. TAPPI **43**: 737–52; 1960; Errata, TAPPI **44**: 448; 1961.
[11] Dodson, C. T. J. "The Failure of bonds during the fracture of a random fibrous network," in Structure, Solid Mechanics and Engineering Design, Proceedings of the Southampton 1969 Civil Engineering Materials Conference, M. Te'eni, ed., London: John Wiley; 1971. 357–366.
[12] Smith, J. C. Characterizing the interfiber bond strengths of paper pulps in terms of a breaking energy. Natl. Bur. Stand. (U.S.) NBSIR 76-1148. Available from National Technical Information Service, PB-264689.
[13] Smith, J. C.; Graminski, E. L. Characterizing the interfiber bond strength of paper pulps in terms of a breaking energy: effect of beating. Natl. Bur. Stand. (U.S.) NBSIR 77-1286. Available from National Technical Information Service, PB-276473.
[14] Smith, J. C.; Graminski, E. L. Characterizing the interfiber

bonding of paper pulps: effect of preparation pressure on tensile test specimens. Natl. Bur. Stand. (U.S.) NBSIR 78-1459. Available from National Technical Information Service, PB-280291.

[15] Smith, J. C. Characterizing the interfiber bonding of paper pulps: rationale for bonding parameters derived from tensile test data. Natl. Bur. Stand. (U.S.) NBSIR 79-1722. Available from National Technical Information Service, PB80-150329.