Journal of Research of the National Institute of Standards and Technology

Volume 99 Number 2 March–April 1994

U.S. Department of Commerce—**Ronald H. Brown,** Secretary
Technology Administration—**Mary L. Good,** Under Secretary for Technology
National Institute of Standards and Technology—**Arati Prabhakar,** Director

The Journal of Research of the National Institute of Standards and Technology features advances in measurement methodology and analyses consistent with the NIST responsibility as the nation's measurement science laboratory. It includes reports on instrumentation for making accurate and precise measurements in fields of physical science and engineering, as well as the mathematical models of phenomena which enable the predictive determination of information in regions where measurements may be absent. Papers on critical data, calibration techniques, quality assurance programs, and well-characterized reference materials reflect NIST programs in these areas. Special issues of the Journal are devoted to invited papers in a particular field of measurement science. Occasional survey articles and conference reports appear on topics related to the Institute's technical and scientific programs.

ISSN 1044-677X **Coden: JRITEF** **Library of Congress Catalog Card No.: 89-656121**

United States Government Printing Office, Washington: 1994

Volume 99, Number 2, March–April 1994
Journal of Research of the National Institute of Standards and Technology

Contents

Articles

News Briefs

Volume 99, Number 2, March–April 1994
Journal of Research of the National Institute of Standards and Technology

Volume 99, Number 2, March–April 1994
Journal of Research of the National Institute of Standards and Technology

Volume 99, Number 2, March–April 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 121 (1994)]

A Sealed Water Calorimeter for Measuring Absorbed Dose

Volume 99 Number 2 March–April 1994

Steve R. Domen

National Institute of Standards and Technology,
Gaithersburg, MD 20899-0001

The NIST sealed water calorimeter is intended for direct measurement of absorbed dose to water. This calorimeter was used for a series of approximately 3700 measurements to investigate the so-called heat defect, that is, anomalous endothermic or exothermic effects caused by dissolved gases. The three systems investigated were "high-purity" water saturated with N_2, H_2, and mixtures of H_2/O_2. The repeatability of the measurements of absorbed dose rates for the ^{60}Co teletherapy beam was studied with different water fillings and accumulated absorbed dose. Measurements with the H_2/O_2 system varied with accumulated absorbed dose. Based on the measurements and theoretical considerations, it appears that the H_2-saturated system is the best choice for eliminating the heat defect. Measurements with both the N_2- and H_2-saturated systems are in good agreement with those determined with a graphite and graphite-water calorimeter (for which there is no heat defect).

Key words: absorbed dose; calorimeter; convective barrier; heat defect; thermistor; water.

Accepted: December 2, 1993

1. Introduction

Absorbed dose[1] is widely used to quantify the medical or biological use of ionizing radiation. The output of a radiation therapy accelerator is calibrated by determining the absorbed dose rate in water, which has been chosen as the standard reference material [1] because it has absorption and scattering properties similar to tissue. Traditionally, absorbed dose is determined from measurements with an ionization chamber and using a protocol [2], which gives the procedure along with numerous correction factors for determining the absorbed dose to water. It has been shown [3–5] that the most nearly direct approach is with a water calorimeter. In principle (when the heat defect is negligible), the absorbed dose at the measurement position is simply the product of the temperature rise and the specific heat capacity of water. Yet this direct approach was not considered possible because of technical difficulties [1].

The possible success of the water calorimeter was viewed mainly as involving two investigative phases [4] as to whether: (1) an accurate small temperature rise could be measured with a sufficiently good signal-to-noise ratio, and (2) the heat defect could be dealt with by its elimination or a correction applied for its effect. In the present work, these effects were investigated in three systems of "high-purity" water saturated with high-purity gases of (1) nitrogen, (2) hydrogen, and (3) mixtures of hydrogen and oxygen. They were investigated as a function of accumulated absorbed dose

[1] Absorbed dose is the energy per unit mass imparted to matter at a point by ionizing radiation. The SI unit of absorbed dose is the gray (symbol Gy). 1 Gy = 1 J/kg (= 100 rad).

over long periods of time and with different water fillings.

The temperature rise was measured with two calibrated thermistors in opposite arms of a Wheatstone bridge to double the output signal. For negligible heat defects and changes in thermistor power, the absorbed dose D is:

$$D = (1/2)\,(\Delta R/R)\,|\bar{S}^{-1}|\,c, \qquad (1)$$

where D is the absorbed dose,

1/2 is the result of using two thermistors to measure a temperature rise,

$\Delta R/R$ is the measured fractional change in the Wheatstone bridge balancing resistor,

$|\bar{S}^{-1}|$ is the absolute value of the reciprocal of the mean sensitivity of the thermistors determined from the calibration data, and

c is the specific heat capacity of water at the calorimeter operating temperature.

The product 1/2 $(\Delta R/R)$ $|\bar{S}^{-1}|$ is the temperature rise.

2. The Calorimeter

2.1 Design

Figure 1 shows the general features of the sealed water (SW) calorimeter. "High-purity" water (HPW) is sealed within a thin-wall cylindrical glass container, 110 mm long and 33 mm in diameter. It has a dual purpose: (1) to enclose the sealed water, and (2) to act as a convective barrier. The assembly is immersed in a 30 cm cube acrylic container filled with once-distilled water (the same system used for the calorimeter described in Ref. [4]). The electrical resistivity of the water was about 0.4 MΩ·cm. A thermistor was mounted in the water to display its temperature to a resolution of 0.01 °C.

The name HPW is used in this paper merely to refer to the relatively cleaner stagnant water in the glass container. The water was prepared in a system consisting of a filter, deionizer, and an organic absorber, which gave the water an electrical resistivity of 30 MΩ·cm. The water was withdrawn and handled in glassware which had been cleaned and placed overnight in a furnace at 450 °C. The water in the glass container was saturated with the high-purity gases.

A square collimated ^{60}Co beam (145 mm × 145 mm at the 50 percent dose points) produced a temperature rise, that was measured at a 5 cm depth from the water surface. The temperature probes consisted of the calibrated thermistors (0.25 mm in diameter) enclosed within and near the end of thin glass capillaries positioned along the axis of the tube. Figure 2 shows schematic constructional details of the sensor end of the probe. The distance between the thermistors can be varied. It was set at 9 mm within a uniform field.

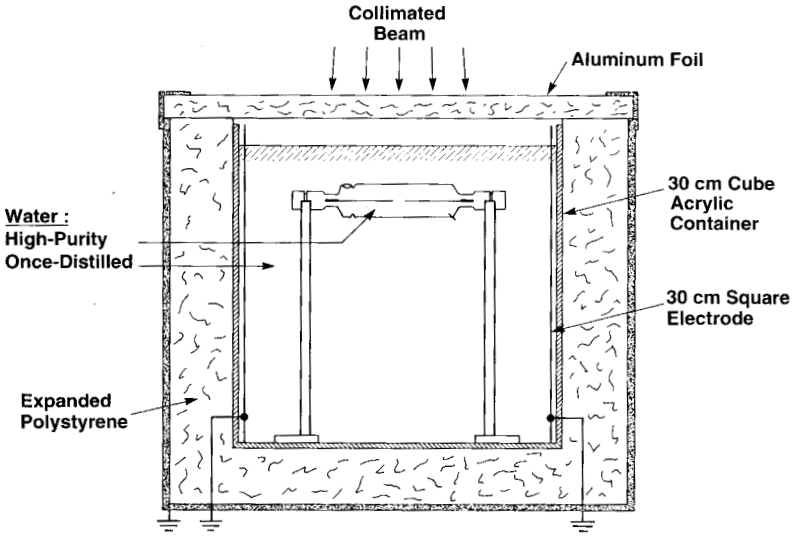


Fig. 1. Essential features of the SW calorimeter for measuring absorbed dose or absorbed dose rate.

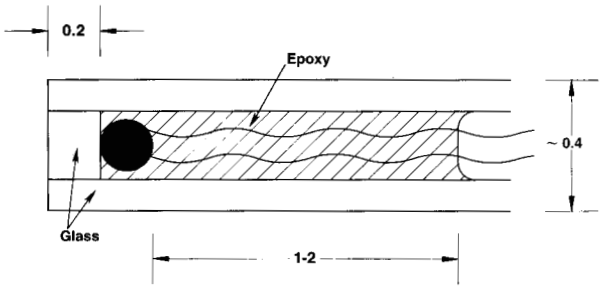


Fig. 2. Schematic diagram showing the essential constructional details of the temperature probe consisting of an embedded thermistor near the end of a long thin capillary. The dimensions are in millimeters.

Figure 3 is a photograph showing some constructional details of the mounted detector assembly. The entire length of the glass structure is 175 mm. The outside diameter around the mid-plane is 33 mm, and the wall thickness varies from about 0.25 mm to 0.33 mm. The volume of the enclosed water is 90 cm^3. Shown at the upper left within the tube (and indicated in Fig. 1) is a bubble of the saturating gas. Its purpose is to expand or contract, to prevent the tube from fracturing when the assembly is subject to large room temperature variations. The bubble is entrapped at that position by the circumferential depression in the tube, which prevents the bubble from migrating to the beam axis directly in front of the sensing thermistors. Shown at the lower right of the tube in Fig. 3 (and indicated in Fig. 1) is a platinum lead. It was in contact with the HPW and the once-distilled water, which was electrically grounded. The platinum lead was included as a precaution against possible charge buildup within the HPW, which could cause large disturbances in the small temperature signals that were to be measured. The large diameter portions (7 mm) of the temperature probes passed through snugly fitting holes in bushings of *low density*

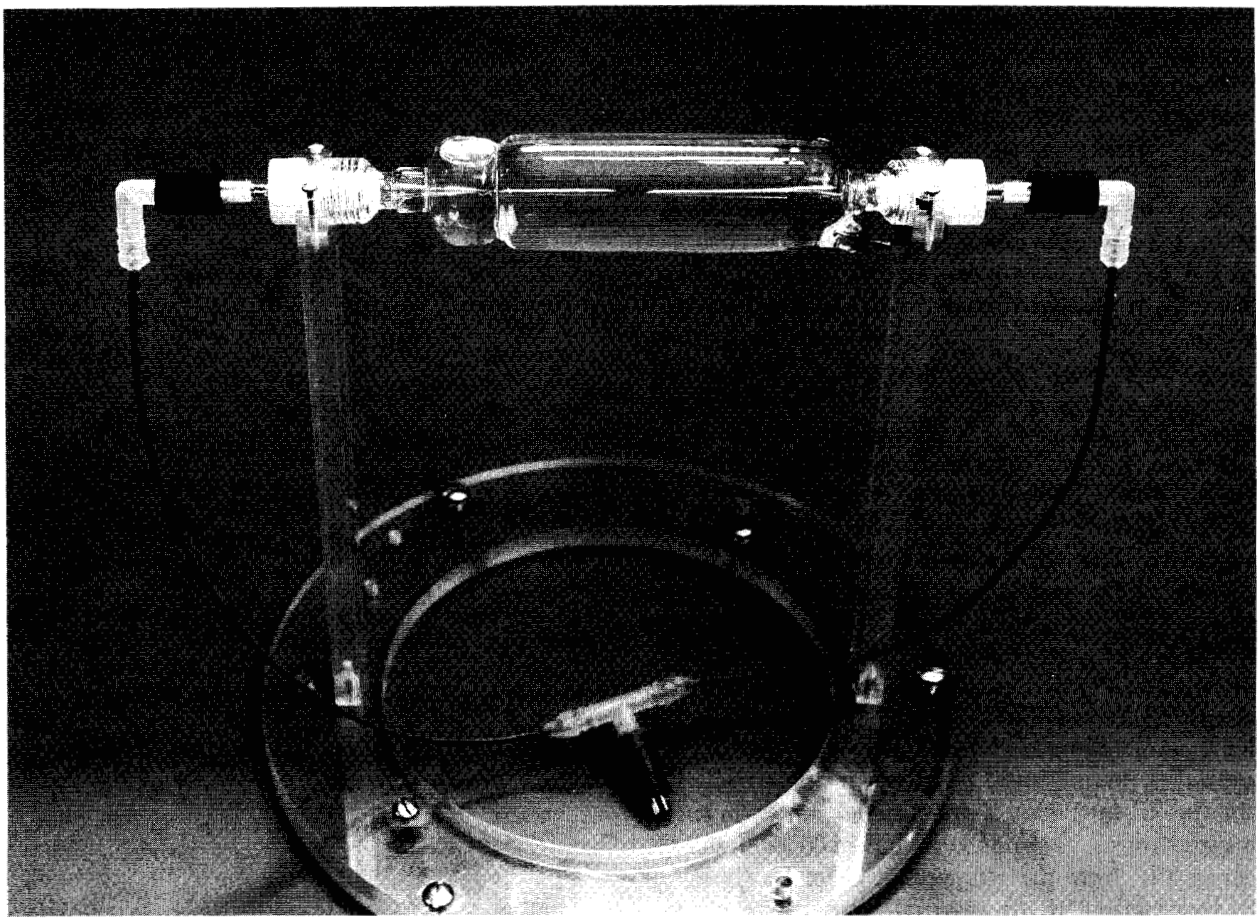


Fig. 3. Photograph showing some details of the mounted thin-wall glass tube enclosing the HPW and temperature probes connected to external waterproof leads.

[(0.91–0.93) g/cm^3] polyethylene, which is relatively elastic. High density [(0.94–0.97) g/cm^3] polyethylene should *not* be used, because it is relatively rigid and stiff. The low density polyethylene, according to preliminary helium-leak tests, helped to make vacuum-tight seals between the polyethylene and glass when the bushings were screwed into position against glass apertures (shown below). This sealed the water from exposure to the atmosphere containing oxygen, which would otherwise be reabsorbed in the water to cause a heat defect. The total area of polyethylene in contact with the water was about 5 mm^2. Small holes were drilled through the electrical sockets so that the air within the capillaries would vary with atmospheric pressure. If troublesome moisture condensed within the probes, it could be evacuated through the holes by placing the probes in a vacuum chamber.

2.2 Measurement and Drift Balancing Circuit

The upper part of Fig. 4 is a Wheatstone bridge circuit containing thermistors 1 and 2 in the temperature probes. It is used for calibrating the thermistors and measuring the irradiation temperature response. The resistor R′ is preset so that $R' \cong r_2$ (the resistance of thermistor 2), when $V_2 = V_{R'}$. When the bridge is balanced, the resistances of the four arms are approximately equal. These conditions will give near minimal corrections for changes in thermistor power during irradiation [4]. The resistor R_1 is preset to give a desired average electrical power dissipation in the thermistors. Irradiation causes the bridge to be unbalanced, but is restored near the balance position by adjusting the bridge balancing resistor, R. A computer program analyzed the runs and calculated slight variable corrections (<0.03 percent for powers up to 30 μW) for effects of slight changes in thermistor powers during a run. This correction is included in the value $\Delta R/R$ [in Eq. (1)], the fractional change in the bridge balancing arm.

The lower part of Fig. 4 is a resistance-capacitance (RC) circuit. Its function was described previously in detail [5–7]. Briefly, its purpose is to balance out small initial drifts in signal across PP′ caused by temperature gradients in the vicinity of the thermistors. This is done by moving R_2 in such a direction as to produce a change in drift signal across PP′ from the RC circuit, which is opposite to that caused by the internal temperature gradients. When used for the measurements described in this paper, R_3 was set on 0.5 GΩ to give a 33 min time constant, large in comparison to the radiation time. Irradiation times were varied and were nominally 1 min and 1.5 min.

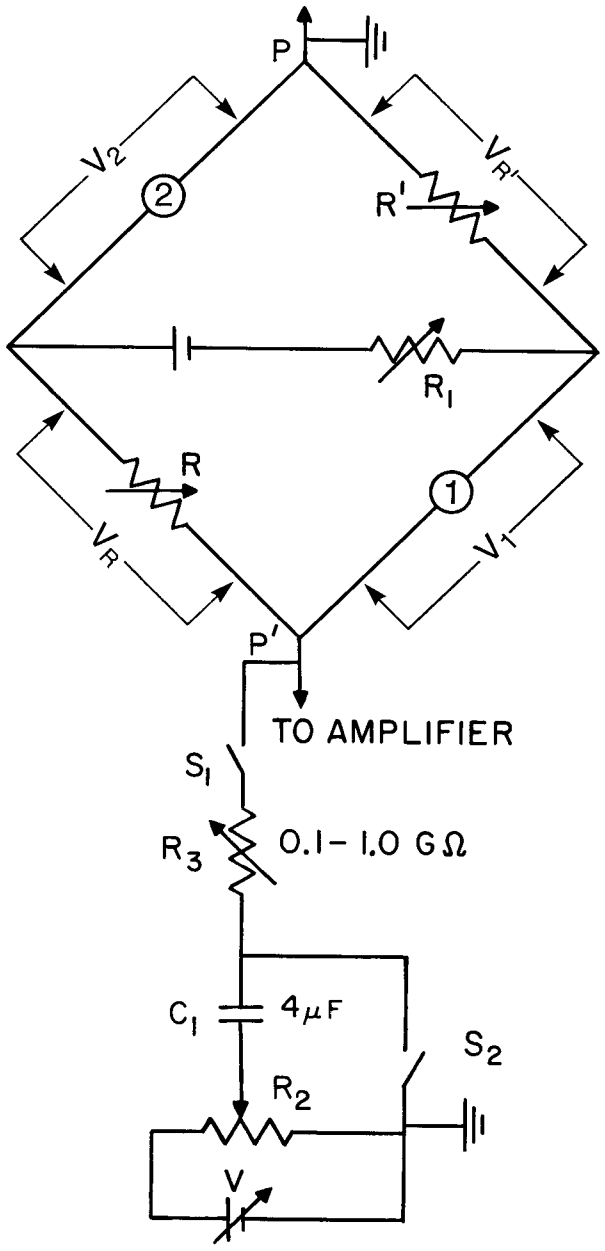


Fig. 4. Wheatstone bridge circuit for detecting temperature changes and a resistance-capacitance circuit (temperature drift balancer) to counteract drifts in signal across PP′ as a result of temperature gradients in the vicinity of the thermistors.

2.3 Thermistor Calibration

The thermistor resistance value (r) at a given absolute temperature (T) is given by the well-known empirical expression:

$$r = r_0 \, e^{\beta(1/T - 1/T_0)}, \quad (2)$$

where r_0 is the resistance at temperature T_0. β is the "material constant," which has the dimension of temperature (K).

The sensitivity of a thermistor (S) is defined as $(1/r)(dr/dT)$, which gives

$$S = |-\beta/T^2|. \quad (3)$$

Equation (2) can be reduced to linear form:

$$y = \beta x + \theta, \quad (4)$$

where $y = \ln r$, $x = 1/T$, and $\theta = \ln r_0 - \beta/T_0$, a constant. Least-squares fits of the data are applied to Eq. (4).

Calibration of a thermistor is the determination of S, which is its fractional change in resistance per degree change in temperature. The temperature was measured with a calibrated mercury thermometer (0.01 °C per division) and a quartz thermometer. The thermistors used had a resistance of about 3.3 kΩ at 22 °C with a negative coefficient of resistance of about 3.7%/K. The temperature probes were removed from the glass container and placed in the once-distilled water (Fig. 1) so that they would rapidly change with the water temperature which was varied at intervals of 1 °C from 15 °C to 29 °C. The temperature of the water was raised with four immersion heaters (total 100 W), and the water was circulated to attain uniform temperature. Then the water was allowed to become stagnant before measuring the thermistor resistances. Their resistances as a function of temperature are shown in Fig. 5. The bridge was balanced at each temperature. The thermistor resistances (r_1 and r_2) plus the external lead resistances, which were about 0.6 percent of r_1 or r_2, could be determined from the known resistances (R and R') and the four measured potentials across the bridge arms as indicated in Fig. 4. The thermistors, however, were calibrated one at a time (for the sake of safety in handling), which required replacing a thermistor with a known resistance. This then gave two methods of determining a thermistor resistance, from (1) two measured potentials and a known resistance, and (2) three known resistances.

The effects of the external lead resistances decrease the sensitivity of the Wheatstone bridge arms containing the thermistors. This decreases the observed calorimetric signal. These are nullifying effects that lead to negligible measurement errors.

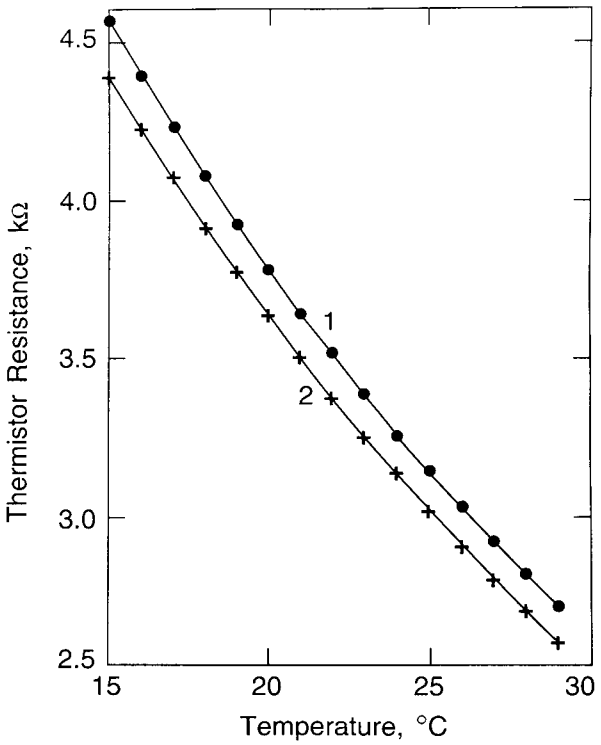


Fig. 5. Calibration data of thermistor resistances as a function of temperature.

The latter effect appears in the numerator of Eq. (1), and the former appears in the denominator. Therefore, small lead resistances can effectively be considered as an integral part of r, if no correction is applied to the observed signal. Otherwise, essentially the same correction factor must be applied to determine the true thermistor sensitivity and the calorimetric signal.

Figure 6 shows interesting calibration results measured eight times over a period of 184 weeks. They were first measured soon after their construction. The second calibration was done 62 weeks later (in preparation for irradiation measurements described below). Although they received no irradiation during this period, the β and r values had average increases of 0.45 percent and 3.5 percent, respectively. During the first phase of the experiment (5.6 kGy accumulated absorbed dose in N_2-saturated water), they increased again by 0.03 percent and 0.34 percent, respectively. With no further increase in absorbed dose and 53 weeks later, the increases were 0.22 percent and 0.21 percent, respectively. Over the next 55 week period [(5.6–17.8) kGy accumulated absorbed dose in H_2- and H_2/O_2-saturated water] the increases were 0.20 percent and 0.14 percent, respectively.

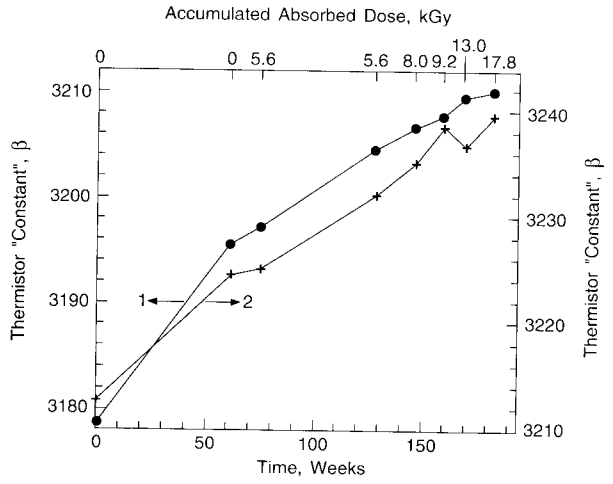


Fig. 6. Thermistor "constants," β, as a function of accumulated absorbed dose and time.

The changes appear not to have been caused by irradiation. This is consistent with other reported results of accumulated absorbed dose up to 620 kGy [8]. In another investigation [9] an accumulated absorbed dose of 4.3 MGy at a rate of 25 kGy/min from 3 MeV electrons produced a resistance change of only about 0.01 percent. A possible cause may have been a gradual release of strains within the thermistors as a result of their manufacture or handling their delicate leads during the fabrication of the temperature probes.

The above points out the precautions that must be taken to assure that the β values remain essentially constant during an extended experimental period. If they change significantly, it must be detected immediately and recalibrated. The method used was the observation that a change in β was accompanied by a change in r. The resistances were routinely measured before and after a daily set of measurements from the known potential and resistance values of the Wheatstone bridge arms as described above. The β-value changes between the dates of calibration were assumed to be linear for determining slight corrections to the date of absorbed dose measurements.

Table 1 lists a set of measured β values determined from different temperature ranges, which

Table 1. Thermistor material "constants," β

Temperature Range (°C)	β_1 (K)	β_2 (K)
15-21-27	3206.64	3233.54
16-21-26	3206.93	3234.17
17-21-25	3206.30	3234.20
18-21-24	3205.92	3234.56
	3206.45 ± 0.007%	3234.12 ± 0.006%
16-22-28	3209.29	3237.01
17-22-27	3210.12	3236.57
18-22-26	3210.46	3236.83
19-22-25	3208.88	3236.55
	3209.69 ± 0.001%	3236.74 ± 0.004%
17-23-29	3213.33	3239.68
18-23-28	3212.42	3239.78
19-23-27	3213.25	3238.94
20-23-26	3214.35	3239.00
	3213.34 ± 0.012%	3239.35 ± 0.007%
19-24-29	3216.57	3242.35
20-24-28	3215.56	3242.34
21-24-27	3217.35	3241.98
	3216.49 ± 0.016%	3242.22 ± 0.004%
21-25-29	3220.47	3245.60
22-25-28	3218.70	3245.88
23-25-27	3221.26	3245.02
	3220.14 ± 0.024%	3245.50 ± 0.008%

vary from three to six degrees above and below mid-temperature ranges from 21 °C to 25 °C. The average values are listed at each set of mid-range temperature and the experimental standard deviations of the mean, which average about 0.01 percent. The average values are plotted in Fig. 7. From all sets of measurements, the average increase in the β values was 0.11%/°C increase in mid-temperature calibration. The temperature of the thermistors was known during the measurements of absorbed dose (near 22 °C or 23 °C). Small corrections were made to the β values because the temperatures differed from those nominal values.

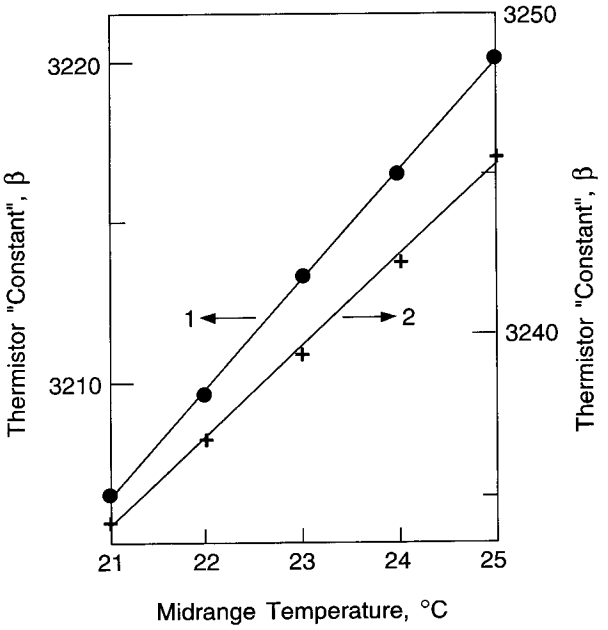


Fig. 7. Thermistor "constants," β, as a function of mid-range temperature calibration.

3. Heat Flow Calculations

Ideally, the temperature rise at a point in water would be made with a massless sensor. But this is not realizable in practice. Thermistors are made of metallic oxides, which are covered with glass beads in the manufacturing process. The thermistor and its bare leads are then embedded in a thin glass capillary to electrically insulate it from the grounded water. Temperature rises, therefore, are measured primarily in glass.

The immediate rate of temperature rise of the irradiated glass, over a wide range of photon and electron energies, is about four times that of the surrounding water. Reference [10] gives useful information on the temperature rises of various materials in the form of a foil, wire, or small sphere, such as a thermistor.

Irradiation causes excess heat to be generated in the temperature probe. The amount of this heat is proportional to the mass of glass, which therefore must be made as small as possible. If measurements are to be accurate, the excess heat must be rapidly conducted away during irradiation and reduced to negligible amounts within a few seconds after irradiation. This is the start of the calorimetric afterdrift, which is used in the analysis for determining the temperature rise. Some excess heat will remain during this period. Heat-flow calculations must be made to determine if the non-water materials can be made small enough, so that the effects of the remaining excess heat do not significantly affect the accuracy of measurements.

With reference to Fig. 1, there are three sources of excess heat if the assumption is made that the HPW within the glass container has a zero heat defect with a relative absorbed dose rate of 1. These are convenient assumptions for the sake of carrying out the calculations described below. The sources of excess heat are the (1) once-distilled water, (2) thin glass wall, and (3) glass capillaries. Calculations were performed by use of the well-known Schmidt numerical method described in many publications as, for example, in Ref. [5]. The water was imagined to consist of concentric cylindrical shells (0.25 mm thick) with a common axis along the axis of the temperature probes.

3.1 Effect of Once-Distilled Water

The once-distilled water was found to have an exothermic effect of 3.5 percent [4]. Therefore, the relative absorbed dose rate in this water is 1.035. This will cause excess heat to be conducted cylindrically toward the thermistors. Figure 8 shows the excess temperature rise (in percent) of the thermistors as a function of time after a 60 s irradiation run. The effect is greatly reduced when the tube diameter is increased from 25 mm to 30 mm.

3.2 Effects of Glass Wall

The glass wall is an effective barrier against convection, which can occur relatively easily in the large external volume of water at room temperature [11,12]. Its diameter (d) must not be too small, because significant excess heat from it may arrive at the thermistors before the irradiation run is

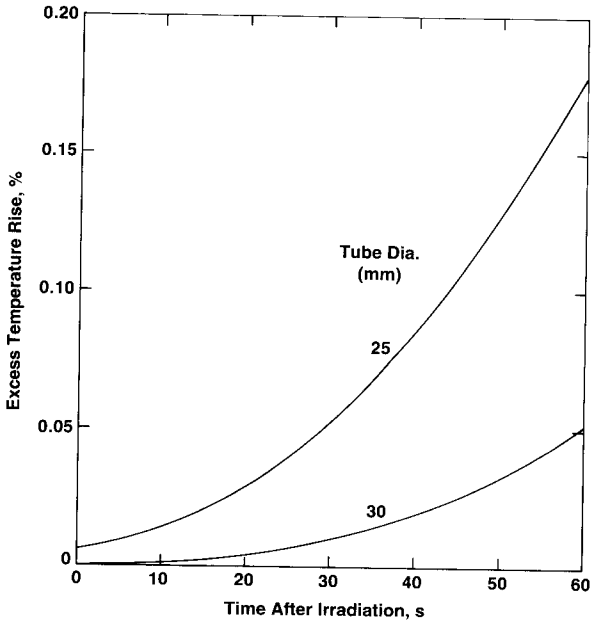


Fig. 8. Excess thermistor temperature rise caused by heat conduction from once-distilled water outside a tube (25 mm and 30 mm in diameter) as a function of time after a 60 s irradiation run.

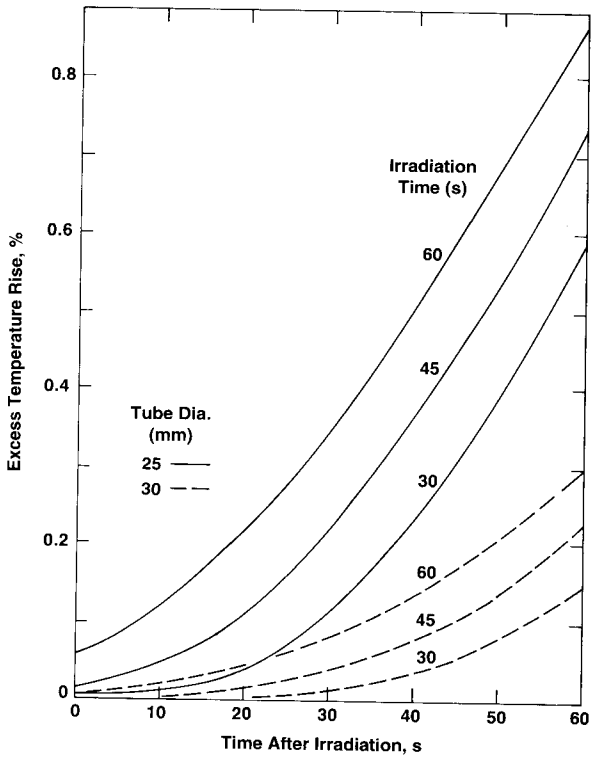


Fig. 9. Excess thermistor temperature rise caused by heat conduction from a 0.25 mm thick glass wall tube (25 mm and 30 mm in diameter) as a function of time after irradiation.

completed. It must not be too large, because conditions will eventually occur when the onset of convection will take place (when the Rayleigh number ≥ 1000) [4,13,14]. Results from the previous and following numerical calculations were valuable in deciding on an appropriate diameter.

A uniform glass wall thickness of 0.25 mm was assumed for the calculations. A further simplification was made to facilitate the calculations. Considering the mass energy-absorption coefficients of glass and water, their densities, and specific heat capacities, it can be shown that the glass can be imagined as being replaced with water where the relative absorbed dose rate is about 2.4 for the case of ^{60}Co irradiation. Excess heat will be conducted cylindrically away and toward the thermistors. Figure 9 gives the excess temperature rise (in percent) of the thermistors as a function of time after a 30 s, 45 s, and 60 s irradiation run. The results show that the tube diameter should be 30 mm or somewhat greater. The constructed tube shown in Fig. 3 has an outside diameter of 33 mm in the region of the thermistors. It appears that the diameter should not be larger than 35 mm. At that diameter the calculated excess temperature rise is 0.09 percent, at 60 s after a 60 s irradiation run. In practice, extrapolation of the final drift to the mid-run will *tend* to correct for the variable increase of the small amount of excess heat sensed by the thermistors; or the calculated curves shown in Figs. 8 and 9 could be included in a computer program and subtracted from a measured post-irradiation drift.

3.3 Maximum Effect of Temperature Probe

The basic detailed construction of the temperature probe is illustrated in Fig. 2. The thermistor is located near the end of the probe and embedded in epoxy. The excess heat is conducted cylindrically and spherically away from the region near the end of the probe; and while this is happening, heat is conducted axially along the capillary toward the end of the probe. The combined effect of this geometry and of the densities and thermal conductivities of the glass and epoxy is a complex situation that is under detailed study at the time of this writing. However, a maximum effect (greater by a factor of about 7) was determined by calculating the excess temperature rise at the center of a long solid glass rod 0.5 mm in diameter. There is then only cylindrical flow of heat. This is easily solved by the numerical method mentioned above. The results are shown in Fig. 10. Note that if significant excess heat remains at the end of the irradiation period, there will be a noticeable sharp drop in signal

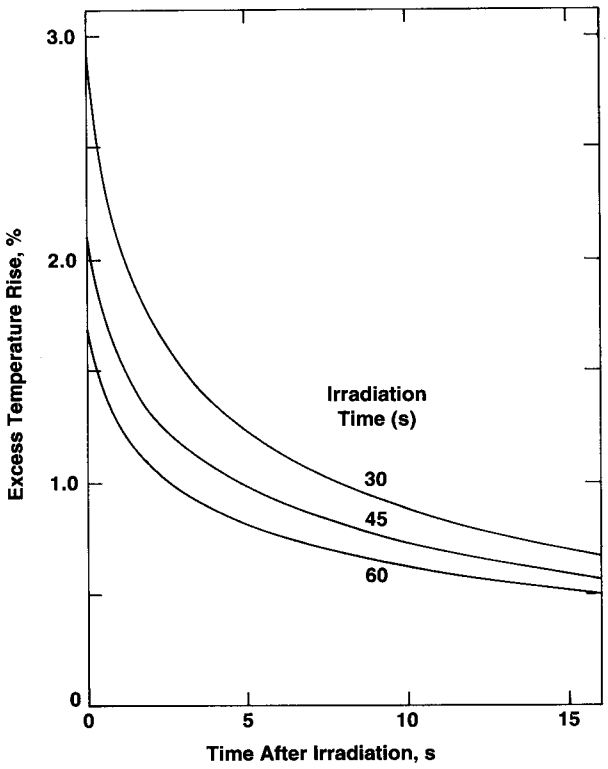


Fig. 10. Excess temperature rise at the axis of a long solid glass rod (0.25 mm radius) immersed in water as a function of time after irradiation.

within 5 s after irradiation. During actual measurements (shown below) this was too small to be observable.

3.4 Orientation of Probes

To avoid concerns and possible errors in measurement, it appears that positioning the axes of the temperature probes perpendicular to the beam (as shown in Fig. 1) is better than positioning a pair of probes parallel to the beam [11]. The concerns and errors would depend on the absorbed dose (temperature) profile along the probe, its geometry, and the thermal diffusivity of the pyrex glass probe which is about four times that of water. Measurements are commonly made at the peak of the absorbed dose curve, where the radial temperature gradients are usually much smaller than the axial gradients. This might cause a component of heat to flow along the temperature probes (parallel to the beam) significantly different from that through water.

4. Probe Fabrication

4.1 Capillary Formation

The temperature sensing probes were fabricated from commercially available pyrex pipets. Figure 11 illustrates the first step. Weight A (about 3 g) was attached near the lower end, which had nominal inside and outside diameters of 1 mm and 1.5 mm, respectively. These dimensions are too large. The inside diameter was to be formed by a high-melting-point wire, 0.30 mm in diameter. This would barely allow the thermistor to be inserted into the completed capillary. Weight B (about 300 g) was attached to the lower end of the wire, to keep it straight. The guide hole held the assembly in a vertical position. The flame shield offered

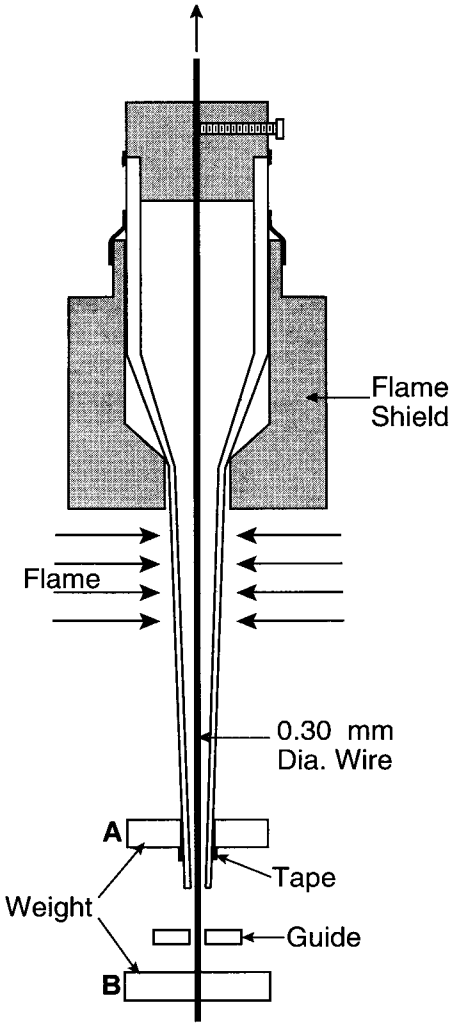


Fig. 11. Illustrative method of reducing the diameter and wall thickness of a pipet to desirably small dimensions.

some protection from distortion of the enclosed glass surface, which was to be part of the sealing surface for enclosing the HPW. Two torches were used to direct flames along opposite sides of the pipet. When the glass softened, weight A caused the pipet to stretch and collapse around the wire. The formed capillary and enclosed wire were cut. This shattered a small section of the capillary and blunted the end of the wire, which was filed to a point to permit its removal. It was successfully removed, roughly about 75 percent of the time without breaking the thin-wall capillary. The glass did not fuse to the wire, because it was pre-blackened over a candle flame. It was also pre-strained to remove its curvature, which would otherwise cause a force against the thin capillary and possibly cause it to break when weight B was removed. Although the capillary diameters over a distance of at least 1 cm from the end ranged only 0.38 mm to 0.44 mm (measured with a graduated microscope disk), they were sufficiently sturdy. Capillaries with larger diameters were rejected.

4.2 Grinding

Figure 12 shows the method used for grinding the end of the capillary. The shattered end rested on a microscope slide containing a mixture of fine grinding powder and water. After the shattered end was ground off, the unit was heated in a furnace at 565 °C for several minutes to remove possible strains in the glass.

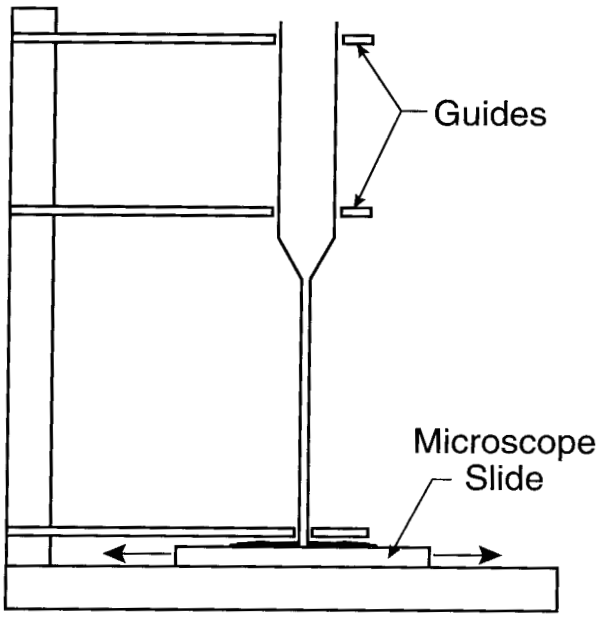


Fig. 12. Illustrative method of grinding the end of a capillary.

4.3 Aligning

Figure 13(a) indicates that the XX′ axis of the probe large diameter (7 mm) did not necessarily coincide with the axis of the small capillary. This happened even though (as illustrated in Fig. 11) the assembly was held in a vertical position. The angle of offset was random. A significant offset would cause the thermistor to be unnecessarily too close to the glass wall when it was screwed into the assembly as shown in Fig. 3. This would cause the excess heat from the wall to arrive sooner at the thermistor. The most important part of the offset (toward the thermistor) was removed by the method illustrated in Fig. 13(b). The large diameter end was held in a lathe collet. The long thin capillary was flexible enough so that part of it was made to pass through a hole in a guide held with the lathe chuck. The lathe was turned at a slow speed (~25 rpm) while a flame was directed on the capillary as shown. When the glass softened, the strain along the capillary was relieved and then the flame was removed. This caused the thermistor to be aligned close to the XX′ axis.

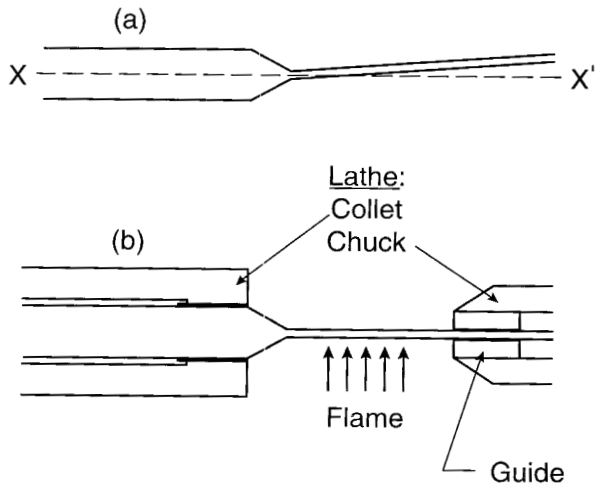


Fig. 13. Illustrative method of making the sensor end of the capillary axis coincide with the larger tube axis.

4.4 Encapsulation

4.4.1 Thermal Coupling A desirable characteristic of a temperature probe is to have good thermal coupling between the thermistor and its surroundings. Figures 14(a) and 14(b) show the results of preliminary work to determine the method to be used and its effectiveness in enclosing a thermistor [15]. When enclosed and surrounded by air as indicated in Fig. 14(a), the immersed probe had

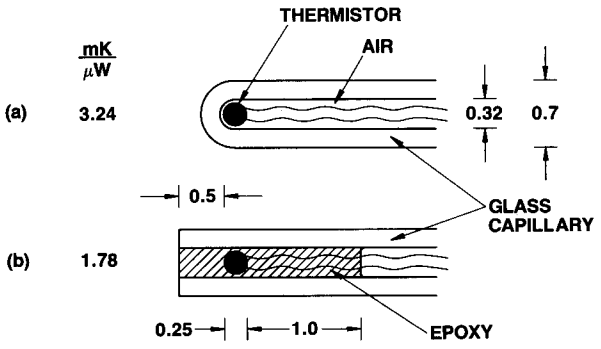


Fig. 14. Results of measurements of the temperature rise in a thermistor per unit electrical power dissipation when it was surrounded directly by: (a) air; (b) epoxy. The dimensions are in millimeters.

a temperature rise of 3.24 mK/μW of electrical power dissipation. The end of the long capillary was then ground off and the thermistor was embedded as shown in Fig. 14(b). Its temperature rise was then 1.78 mK/μW. This lower value is an improvement and is attributed mainly to the higher thermal conductivity of the epoxy, about 30 times that of air.

Further preliminary work removed possible troublesome features shown in Fig. 14(b), which indicates that the surface of the epoxy would be in contact with the HPW that would absorb organic impurities. Furthermore, prolonged immersion would cause the electrically grounded water to be absorbed in the epoxy and be in contact with the thermistor and its leads, which would cause large erratic signals. This possible source of trouble was eliminated by closing the capillary end with glass as illustrated in Fig. 2.

4.4.2 Lead Insulation Figure 15 illustrates the procedure for embedding, enclosing, and ensuring that its bare Pt-Ir leads (25 μm in diameter and ~1 cm long) would not be electrically shorted. The leads were soldered to 25 μm diameter copper wires coated with polyurethane. To prevent electrical shorts, particularly in the congested region of the soldered joints, observations through a microscope revealed the necessary locations on which to apply beads of fast drying epoxy. Then the leads and thermistor were withdrawn and the epoxy beads were applied. Several cycles like this had to be made before observations revealed that the epoxy beads separated the bare portions of the wires.

4.4.3 Embedment Figure 15(b) shows the application of a viscous slow-drying epoxy at the opening of the capillary. A thin wire was used to transport several applications of the epoxy into

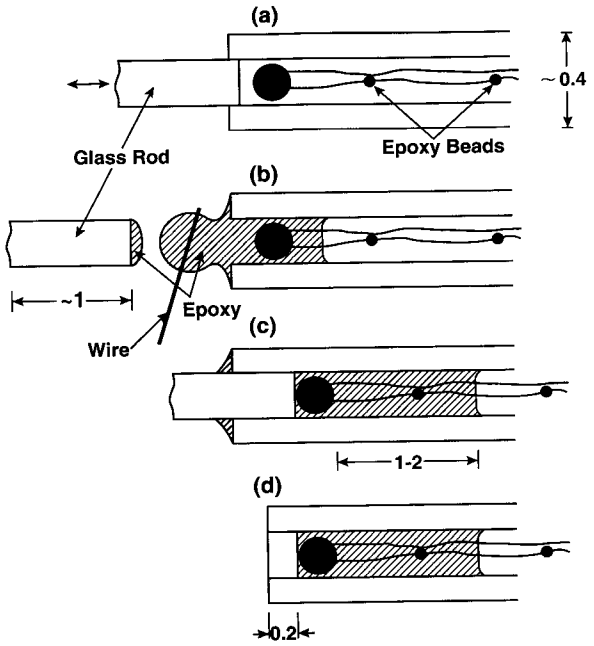


Fig. 15. Illustrative procedure for embedding and enclosing a thermistor and ensuring that its bare leads would not be electrically shorted. The dimensions are in millimeters.

position with the aid of a microscope. The epoxy was drawn into the capillary by the surface tension caused by those materials. The amount of epoxy applied was such that its final length behind the thermistor was 1 mm to 2 mm, as indicated in Fig. 15(c). This provided increased thermal conductivity from the heated wires close to the thermistor with only a slight addition of a non-water material near the thermistor.

4.4.4 End Closure The glass rod (~1 mm long) was attached to an apparatus containing three micrometer movements that were necessary for maneuvering the glass rod into the capillary opening. The glass rod was made slightly tapered, and its ground right surface had a diameter only about 10 μm to 15 μm smaller than the opening of the capillary. The inserted rod pushed back the thermistor, making a direct contact. The taper of the rod was such that the glass rod was also in firm contact with the capillary. The capillary end shown in Fig. 15(c) was ground down to the 0.2 mm dimension shown in Fig. 15(d). Placement of the thermistor close to the capillary end causes it to be in the "cool" region of the excess heat, because of the added and strongest heat conductivity (spherical) from the capillary end.

Figure 16 is a photograph of the sensor end of the completed temperature probe. The scale markings are in millimeters. Reflections and refractions

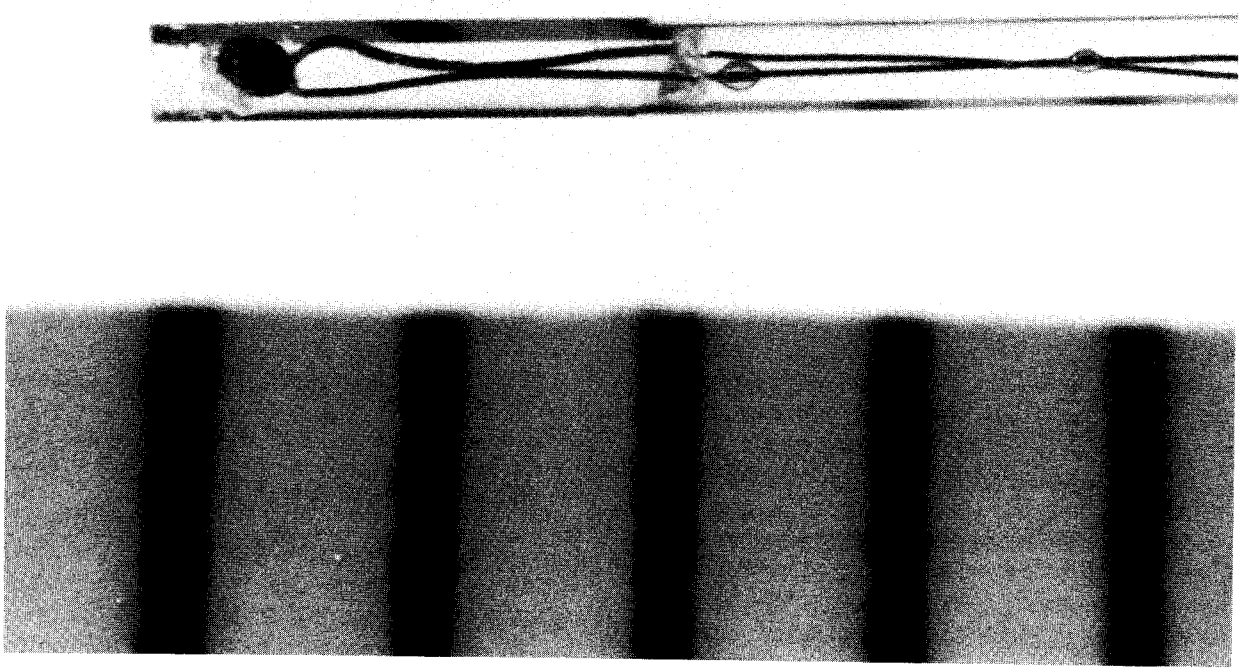


Fig. 16. Photograph of the constructed temperature probe. Reflections make it appear that there is an abrupt change in capillary diameter near the right end to the enclosed epoxy. The scale is in millimeters.

make it appear that there is an abrupt change in capillary diameter at a position near the right end of the enclosed epoxy. Figure 17 shows the entire probe assembly. The electrical plug was epoxied into position. Two holes were drilled parallel to its axis in order to not entrap the air within the capillary. Six probes were constructed. When immersed in water, their temperature rises ranged from (1.34–1.56) mK per μW of electrical power dissipation.

The setup shown in Fig. 1 was used to measure the leakage resistances between the thermistors 1 and 2 in the temperature probes and the surrounding water, which remained in the glass container for 221 d. The results are shown in Fig. 18. The resistances were always high and stable over long periods compared to the irradiation time. A calculation based on an assumed maximum and rapid short time change showed that the resultant contribution to the noise level would be lower by a factor of about 500 compared to the noise observed during the measurements.

5. Heat Defect

The heat defect, k_{hd}, is defined by

$$k_{hd} = (E_a - E_h)/E_a, \tag{5}$$

where E_a is the energy imparted to a material and E_h is the energy released as heat. The reaction is endothermic when $E_a > E_h$ and exothermic when $E_h > E_a$.

The heat defect was perhaps the greatest obstacle to initiating the investigation of the water calorimeter. The initial design consisted of once-distilled water open to air and in contact with plastic materials [3,4]. Exothermic results reported by investigators ranged from 1 percent to 5 percent. It

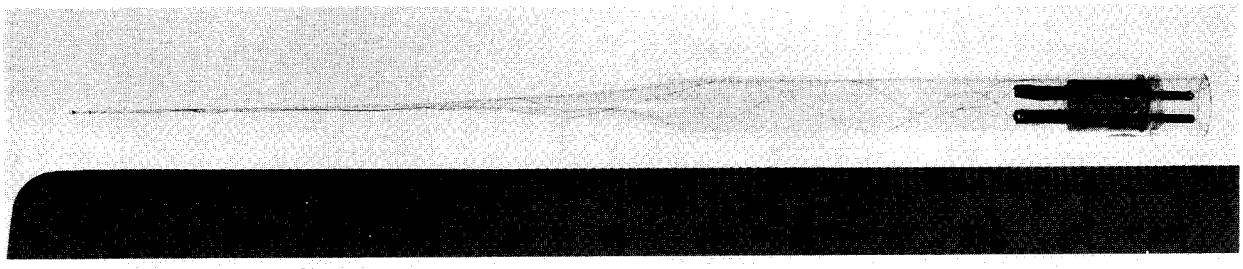


Fig. 17. Photograph of the temperature probe assembly.

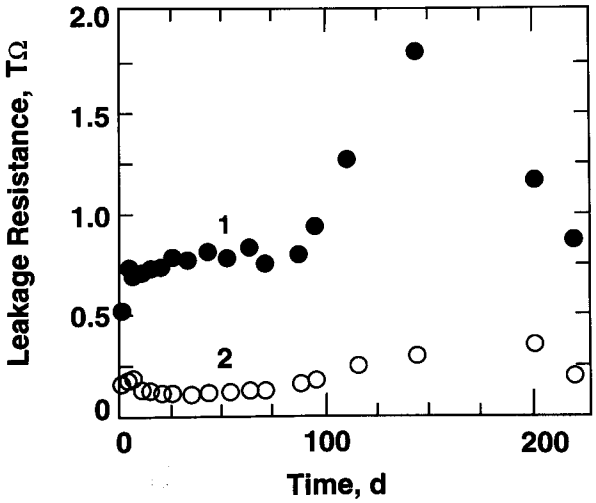


Fig. 18. Measurements of leakage resistance between thermistors 1 and 2 in the temperature probes and the surrounding water.

is difficult to determine how much of this can be attributed to differences in water quality. Part of it can be attributed to real errors in the comparative method of ionization chamber measurements and the use of protocols to determine absorbed dose.

Significant progress has since been made in carrying out the above mentioned second and relatively longer phase (2) in the development of the water calorimeter [4]: theoretical and experimental investigations aimed at giving more assurance concerning the heat defect.

Therefore, the new generation of calorimeters are to be relatively clean, to improve the water quality. The water is to be contained in clean glass containers, and sealed from the atmosphere [16,17]. It is also to be prepared in such a way as to cause an essentially zero heat defect [16–20], or to determine if a particular type of heat defect could be reproduced and accurately corrected [20]. These studies in the present experiment required saturating the HPW with high purity gases. The stated commercial minimum purities of the N_2, H_2, and O_2 gases used were 99.9995 percent, 99.9995 percent, and 99.997 percent, respectively.

Figure 19 illustrates the method of using the high-purity N_2 (<0.2 parts per million O_2) to replace absorbed gases in the HPW, which is then transferred to the glass detector assembly. A vacuum pump was initially used to evacuate possible trapped air pockets within the gas gauges and valves (not shown). The detector assembly was initially flooded with N_2. Polyethylene tubing (3 mm inside diameter) connected the components shown. A rotameter indicated that the N_2 flow rate was about 30 cm^3/min. The gas passed through a fritted disk at the bottom of a glass column containing the HPW. The column had a volume six times that of the glass detector assembly. A vigorous flow of many tiny bubbles of N_2 ascended through the water. The gas easily passed through the assembly, because the polyethylene bushings had been partly unscrewed to make the threads loosely fitting. The flow rate continued for 40 min. Then the column and detector assembly were rotated about 1/2 turn around the horizontal axis, while the gas pressure continued. The water ascended, entrapping a bubble of N_2 in the region of the platinum grounding lead. When almost all the water was gone from the column, the bushings were screwed in to make the seals against the circular glass openings.

6. Measurement Procedure

The following describes (1) tests which show that the absorbed dose measurements were made in the absence of convection, (2) a series of radiation runs resulting in increasing temperature gradients and temperature drifts, and (3) a method of rapid restoration to equilibrium.

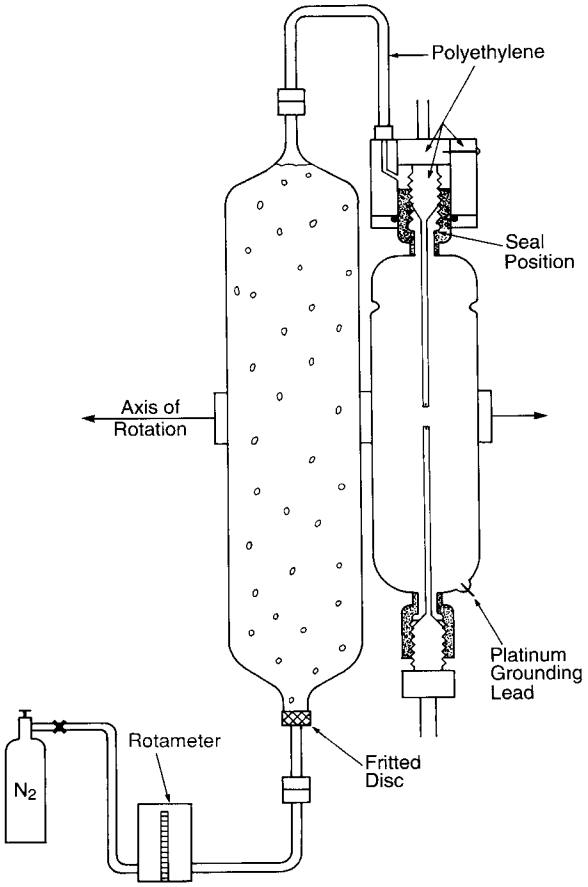


Fig. 19. Illustrative method of gas saturating the HPW, which was then transferred to and sealed in the shown thin-wall glass tube. The drawing is not to scale.

Figure 20 shows results of absorbed dose rate measurements as a function of thermistor power. The uncertainties are the experimental standard deviations of the mean for the number of indicated measurements. Reference [21] describes convective velocity effects on an electrically insulated thermistor (0.25 mm in diameter) in water, free of convective barriers. It was shown that convection began around the heated thermistor when the electrical power was about 50 μW. Its measured temperature rise per μW of power was 1.41 mK. Therefore, convection began when its temperature rise was about 70 mK. The threshold of convection also will depend on thermistor size [12] and details of construction. The two temperature probes used in the present investigation had measured temperature rises of 1.33 mK and 1.53 mK per μW of power. Therefore, convection would begin when the average thermistor power would be raised to about 50 μW in water free of convective barriers. But the probes were mounted within the glass tube (Fig. 3)

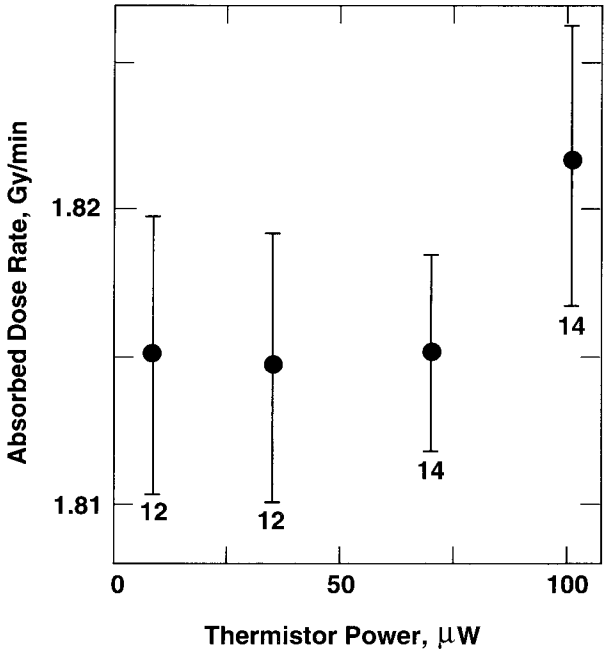


Fig. 20. Absorbed dose rate measurements as a function of thermistor power.

where its wall was a convective barrier. It is uncertain if this barrier raises the electrical power at which convection begins. Therefore, in the operational procedure it is safer to assume that the wall has no effect, for this case, and to use power levels significantly below 50 μW. Although Fig. 20 shows that there is no significant difference in the measurements up to 100 μW, the maximum power used was 30 μW. This is significantly below the threshold of 50 μW at which convection begins around a thermistor as described above. Many of the absorbed dose measurements were made at 9 μW. The range of powers from 9 μW to 100 μW caused the thermistors to rise to average equilibrium temperatures of 13 mK to 143 mK above the background water temperatures, while the temperature rise during irradiation caused them to rise an additional temperature of only about 0.5 mK (shown below). Therefore, slight disturbances of the equilibrium temperatures caused by irradiation produced convection (if present) would have resulted in significant changes in the measurements as a function of thermistor power.

Figure 21 is a typical series of radiation runs. Time increases from right to left. A small initial heating drift was allowed. This permitted five runs to be made in a series of runs where the drifts were small. The ^{60}Co beam produced a dose rate of about 1.8 Gy/min. The spikes shown are caused by

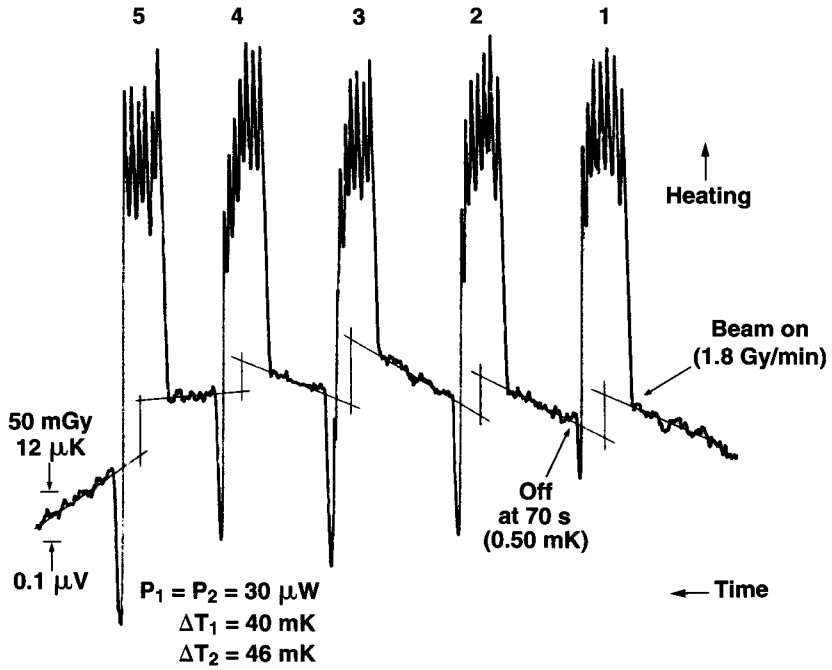


Fig. 21. Typical consecutive runs showing measurement parameters and changes in temperature drifts as a result of heat conduction caused by collimated beam irradiation. Time increases from right to left.

the heating and manual adjustments in the bridge balancing resistor, R. The duration of the runs was varied around 70 s to produce a temperature rise of only about 0.5 mK. The runs were analyzed in the usual way by extrapolating the initial and final drift to the mid-run, and determining the offset by calibrating the chart deflection from a known resistance change in R. A computer calculated the resistance value of R that would rebalance the bridge.

The first three temperature drifts are essentially unchanged. The fourth drift shows a small but significant change. The fifth drift shows a slight cooling, followed by the sixth drift, which shows significant cooling. This typical behavior is a result of outward conduction of heat from the penumbra of the collimated square beam, approximately 145 mm at the 50 percent dose level. Initially, the thermistors did not sense the outward flow of heat. Each thermistor had an electrical power dissipation of 30 μW, which caused an equilibrium temperature rise of 40 mK in thermistor 1 and 46 mK in thermistor 2. Superimposed on these temperature rises were the 0.5 mK temperature rises as a result of beam irradiation. The results shown in Fig. 20 and the predictable behavior of the successive drifts shown in Fig. 21 are indications that the glass wall is an effective convective barrier against external convection and that it is unnecessary to operate the calorimeter at 4 °C [11].

The drift balancer circuit shown in Fig. 4 can balance out large drifts. But generally, large temperature gradients within the calorimeter should be removed and equilibrium restored. Equilibrium was restored after eliminating the temperature gradients by circulating the once-distilled water shown in Fig. 1 (in most cases this circulating procedure was done after only two consecutive runs). The water was slightly heated with the immersion heaters and circulated with an aquarium pump, which forced air to rise along the four vertical corners of the acrylic container. After several minutes the pump was turned off and the water gradually became stagnant. The drift continued because of a temperature difference between the circulated water and that within the glass container. The heating or cooling drift gradually became smaller as conduction continued, and within approximately 20 min (depending of the initial temperature difference) the drifts were small enough to continue another series of runs. Approximately 40 runs could be made in a period of 5.5 h.

7. Results of Absorbed Dose Measurements

7.1 N_2-Saturated Results

All measurements refer to a linear depth of 5 cm and were normalized to a particular date to correct for ^{60}Co decay. Slight corrections normalized the measurements to a calorimeter operating temperature of 22.0 °C, where the density of water is 0.9978 g/cm^3. The specific heat capacity of water at this temperature was taken to be 4.1808 J/(g·°C) [22], which is assumed to be the same for air-saturated and air-free water [23]. A small estimated correction (mentioned below) was made for the effect of the thin glass wall.

Figure 22 shows the number of daily measurements with the first water filling irradiated with an accumulated absorbed dose of 2.9 kGy. Measurements were made on 14 d over a time period of 58 d. The vertical and horizontal bars indicate, respectively, the standard deviations of the mean and the accumulated absorbed dose. It is assumed that the variations shown are statistical. On the first day there was an initial variation, which was not noticeable on subsequent days. The first measurement was about 6.5 percent higher than those after the 7th run (accumulated absorbed dose of about 20 Gy), where the measurements thereafter appeared to show statistical variations. This was the same general type of behavior observed on the first day of measurement after other water fillings, described below. The initial measurements showed exothermicity, which rapidly vanished and were ignored, as discussed below.

On some days, measurements began with a group of measurements followed by continuous unrecorded irradiation from 30 min to 3 h and then repeating the measurement cycle. On other days, the first two steps were reversed. On a few days only measurements were made. The numbers shown in Fig. 22 indicate a total of 484 measurements. The average of the 14 daily average values is shown as the first point in Fig. 23. The experimental standard deviation of the mean of the 14 average values is 0.07 percent.

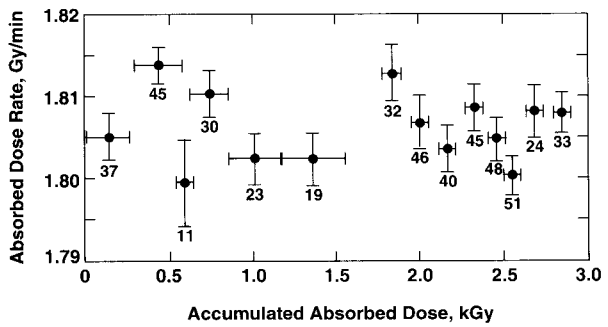


Fig. 22. Results of 14 d of measurements of absorbed dose rate vs accumulated absorbed dose in the first N_2-saturated water filling. The numbers indicate the number of measurements made on a day. There were 58 elapsed days from the first to the last day of measurement.

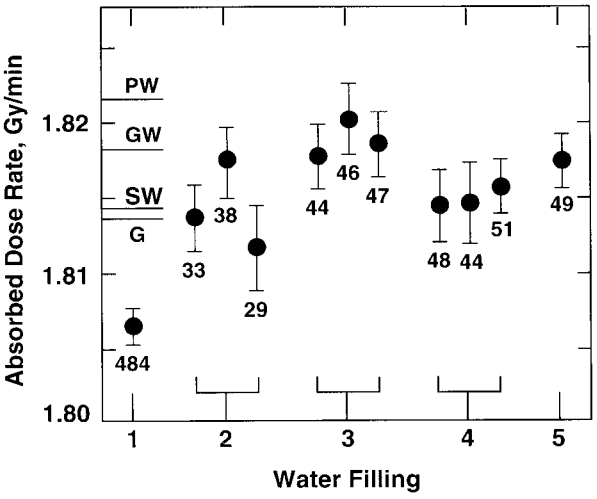


Fig. 23. Results with five different N_2-saturated water fillings. The average result is indicated (SW) and compared with results determined with other calorimeters: PW (polystyrene-water), GW (graphite-water), and G (graphite).

The water in the glass container was discarded and filled four other times. Figure 23 shows the average results of the daily measurements. Initial variations were again observed on the first day of measurement, after the second, third, and fourth filling. After the second filling, the first measurement was about 5.5 percent higher than those after the 12th run (accumulated absorbed dose of about 34 Gy). After the third filling, the first measurement was about 7 percent higher than those after the sixth run (accumulated absorbed dose of about 10 Gy). After the fourth filling, the first measurement was about 3.5 percent higher than those after the second run (accumulated absorbed dose of about 5 Gy). After the fifth filling, the calorimeter was pre-irradiated for 45 min (accumulated absorbed dose of 70 Gy). Subsequent measurements showed only statistical variations.

For the second, third, and fourth fillings, Fig. 23 shows three average results from measurements made on three days with each of those fillings. The average of each of those three values (for a particular filling) was assigned as a single value for that

filling. The average of the five values is 1.814 Gy/min, which is indicated as SW. The standard deviation of the five averages is 0.27 percent, which indicates the reproducibility of measurements with the water fillings. This is in good agreement with the 0.25 percent standard deviation determined by Schulz et al. [17], who used a similar water purifying system and saturated the water with high-purity nitrogen.

Figure 23 shows results of comparable absorbed dose determinations converted to water as measured with a polystyrene-water (PW) calorimeter [7] shown in Fig. 24, a graphite-water (GW) calorimeter where the schematic detector details are shown in Fig. 25, and a graphite (G) calorimeter [24] shown in Fig. 26. Measurements with the PW and GW calorimeters were converted to absorbed dose to water by use of the mass energy-absorption coefficients. Conversion of the G calorimeter measurements is described by Pruitt et al. [25]. The estimated combined standard uncertainties of the measurements with the PW, GW, and G calorimeters are 1.1 percent, 0.7 percent, and 0.6 percent, respectively.

The PW result is based on the specific heat capacity of polystyrene, which was calculated from an empirical equation [26]:

$$c = (104.15)^{-1}\,(7.7551\cdot10^{5}\,T^{-2} + 0.53447\,T - 41.58)\ \mathrm{J/(g\cdot K)}, \qquad (6)$$

where T is the absolute temperature. One mole of polystyrene = 104.15 g.

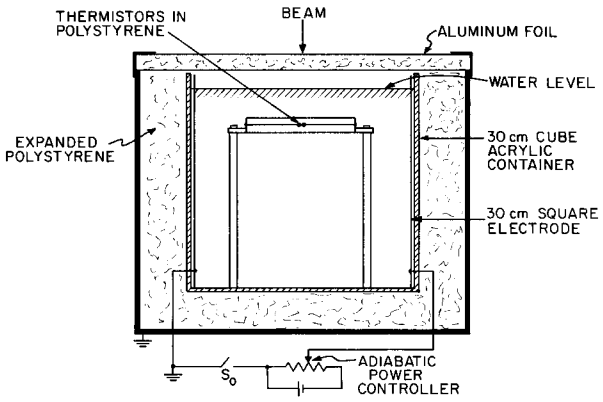


Fig. 24. Essential features of the absorbed dose polystyrene-water calorimeter. The "adiabatic" power controller permits a potential to be applied across the electrodes so that in the vicinity of the 10 mm thick polystyrene disks the rate of temperature rise of the water, as a result of electrical power and beam irradiation, is as nearly as possible the same as that of the disks.

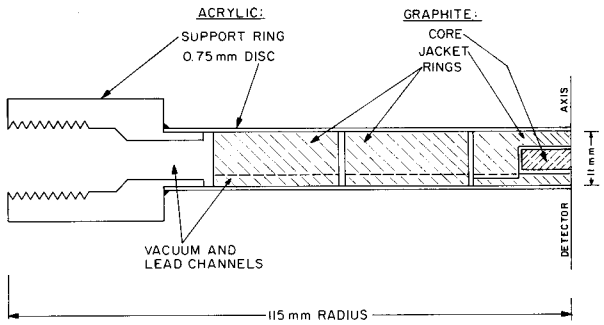


Fig. 25. Essential features of the detector assembly for the absorbed dose graphite-water (GW) calorimeter.

Recently, accurate measurements were made of the specific heat capacity of polystyrene [27] of sample pieces cut from a different slab, 50 mm thick. The result was 2.0 percent higher [8,27] than that calculated by the use of Eq. (6). The result determined with the GW and the G calorimeter do not depend on knowing the specific heat capacity of graphite, because the temperature responses of those calorimeters are electrically calibrated. Therefore, these two results are considered to have less uncertainty than that determined with the PW calorimeter; and although this latter result is in reasonably good agreement with the three others shown, it is not included with the calorimetric calibration of the ^{60}Co source.

Other comparisons of absorbed dose to water standards (calorimetric, Fricke dosimetry, and ionometric) are generally in agreement to about 1 percent. These are reported in Refs. [28–31].

7.2 H_2-Saturated Results

Figure 27 shows the results obtained with five fillings of water saturated with hydrogen. In nearly all cases, each point is the average of 32 measurements made in a day. The uncertainties are experimental standard deviations of the mean. Measurements for the first filling were made on 14 d. Measurements for the second, third, fourth, and fifth fillings were made on 8 d. The accumulated absorbed dose for each filling was about 1.25 kGy. Daily measurements began with or without a 30 min, 45 min, or 90 min pre-irradiation. On some days, a continuous irradiation period of 45 min or 90 min was applied after measurement No. 16 (midway of a daily set of measurements). In contrast to the N_2-saturated water, the initial measurements in the H_2-saturated water showed little or no initial exothermic effect.

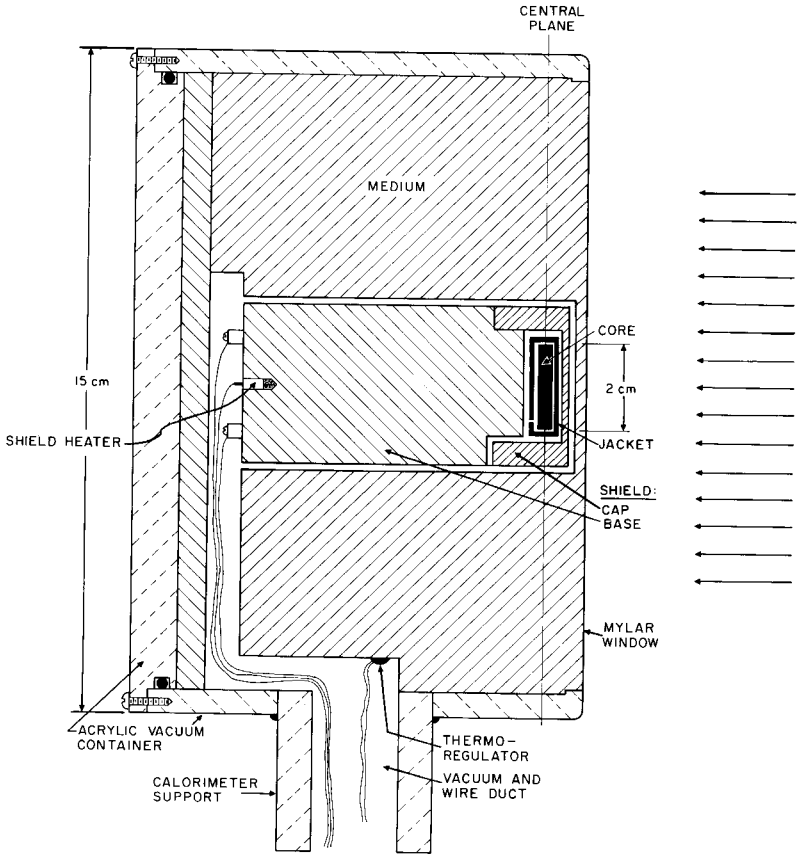


Fig. 26. Side-view cross section of the graphite (G) absorbed dose calorimeter.

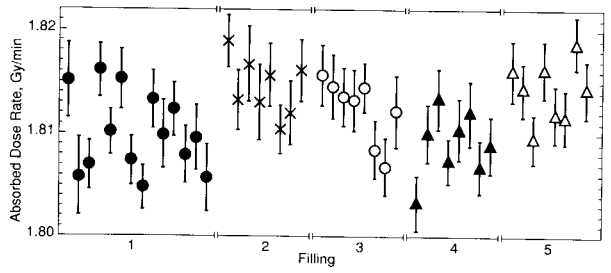


Fig. 27. Daily averages of absorbed dose rate measurements with five fillings of water saturated with H_2.

A single value for each filling was calculated as the average of the points shown for that filling in Fig. 27. The results are shown in Fig. 28. The average value and the experimental standard deviation of the mean was calculated for those five points and are shown in Fig. 29. The results with the H_2- and N_2-saturated water are in good agreement and within the uncertainties. The 0.13 percent difference is in good agreement with, and in the same direction as the 0.3 percent difference reported by Klassen and Ross [20].

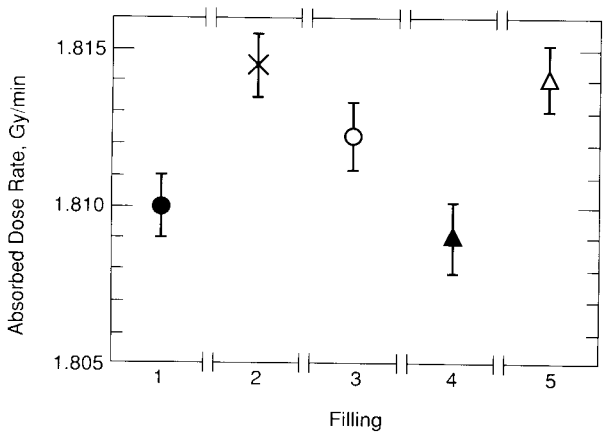


Fig. 28. Average results of the five fillings saturated with H_2.

7.3 H_2/O_2-Saturated Results

Figure 30 shows the results with mixtures of H_2/O_2 and compared with the H_2 result (from Fig. 29). The 2.4 percent exothermic effect is that reported [20] for saturating "high-purity" water with equal flow rates of H_2 and O_2, and where there was

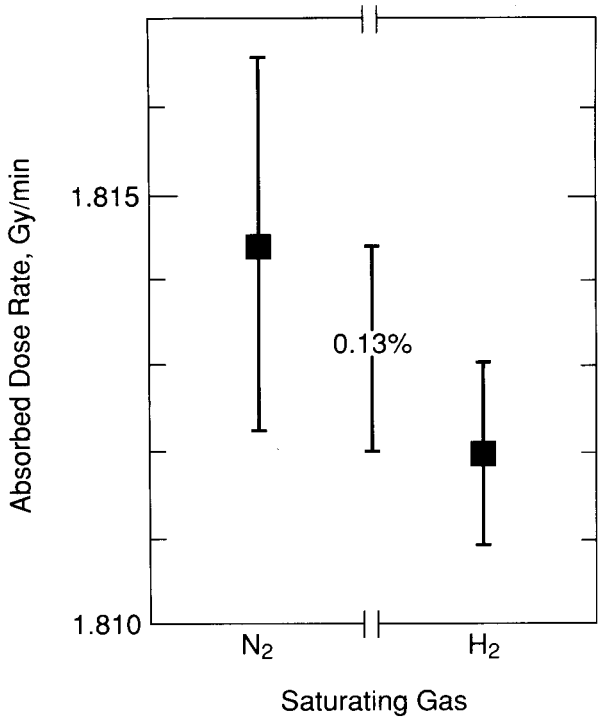


Fig. 29. Comparison of the results of the N_2- and H_2-saturated systems.

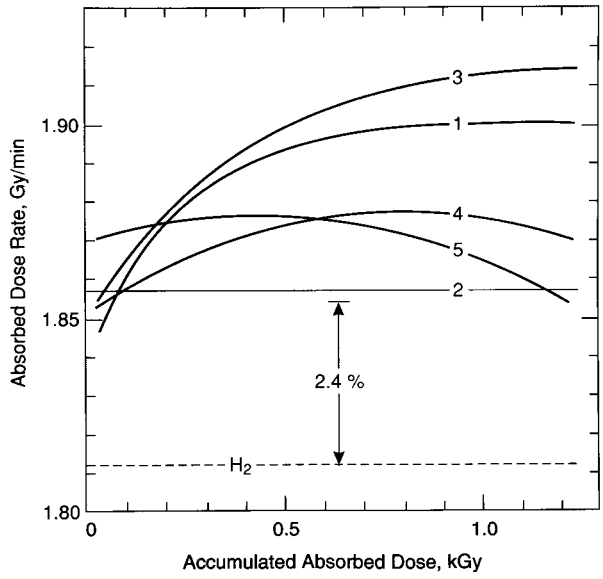


Fig. 30. Measured results with five fillings of water saturated with mixtures of H_2/O_2 and comparison with the H_2-saturated system.

a large volume of those gases above the surface of *continuously circulated* water. The absorbed gases were quickly distributed throughout the water to keep their concentrations (and chemistry during irradiation) relatively constant. This is a different and relatively open system compared to the closed system described in this paper where only a small bubble of the gas remained in the glass detector assembly (Figs. 1 and 3) containing the *stagnant* water, a necessary requirement for measuring absorbed dose at a position. In such a system, the H_2/O_2 concentrations (and chemistry) change during irradiation.

In Fig. 30, the first filling shows that the initial exothermic effect is in the region of 2.4 percent, but increases further. The second filling shows a constant exothermic effect near 2.4 percent. The third filling shows the same general behavior as the first filling. The fourth filling has the same initial exothermic effect, but increases more slowly before showing a decrease. The fifth filling shows an initial exothermic effect of about 3 percent that increases slightly before slowly decreasing to about 2.4 percent at an accumulated absorbed dose of 1.2 kGy.

Subsequent to the measurements, a theoretical calculation was made for the H_2/O_2 closed system described as a function of accumulated absorbed dose and for different concentrations of the H_2- and O_2-saturating gases [32]. Subsequently, the gas flow rates measured with rotameters (Fig. 19) were checked with a more accurate method by use of a mass flowmeter [32] and by measuring the rate at which the gases displace a known volume of water. The flow rates as measured with the rotameters were significantly different. The tests showed that the detector assembly was saturated with a variable preponderance of H_2 in the five fillings for the measurements of absorbed dose. The measured results as shown in Fig. 30 are, therefore, in general agreement with theory.

In retrospect, it was fortunate that the water was not saturated with equal flow rates of H_2 and O_2, because a preponderance of H_2 "magnified" the variation of the heat defect with accumulated absorbed dose. The measurements, therefore, initiated the theoretical calculations [32] which led to the general conclusion that the closed H_2/O_2-saturated system is not recommended as a standard for measuring absorbed dose in *stagnant* water.

8. Corrections and Uncertainties

Negligible uncertainties were assigned to the following: (1) the specific heat capacity of water [22]; (2) measurements of source-detector distance and the 5 cm *linear* depth from the water surface to the

horizontal plane where the thermistors were located, because of the uncertainty (~0.02 percent) in the optical sighting and the micrometer measurements; (3) measurements of the thermistor sensitivities; because of the reproducibility of their values before and after the experiment (from Table 1) and because the average of those values were used in the analysis; (4) beam exposure timing; and (5) excess heat generated in the temperature probes, because of the calculated heat flow analysis.

Ionization measurements in a water phantom showed that the aluminum foil and expanded polystyrene of the calorimeter lid decreased the measurements by 0.34 percent. This correction was made and a 0.05 percent uncertainty of one standard deviation was assigned. The effect of the thin glass wall was to cause the measurement depth to be slightly greater than if the wall had been "water equivalent." A rough estimate based on ionization measurements and a simple geometry showed that the absorbed dose measurements had to be corrected by an increase of about 0.14 percent. A 0.1 percent estimated uncertainty was assigned.

The most difficult correction and uncertainty to assign is that related to the heat defect, if any. Figure 29 indicates that the difference of 0.13 percent is not significant for the N_2 system, even though the measurements and theoretical calculations show that the H_2 system is cleaner. The N_2 system showed significant initial exothermic effects after a filling, which was not observed or relatively small for the H_2 system. Also, there is good agreement with measurements made with the G and GW calorimeters, which have no heat defect. No correction is applied to the H_2 system, but a 0.3 percent heat defect uncertainty is assigned. Table 2 lists the standard uncertainties with the H_2 system. The combined standard uncertainty is 0.4 percent.

9. Conclusions and Future Plans

Based on the measurements and theoretical calculations, it appears that the H_2-saturated system is the best practical choice in eliminating the heat defect in water. The flow rate of the gas need not be accurately known so long as there is an abundant supply (roughly about 30 cm^3/min) bubbling through the water column such as shown in Fig. 19 for about 40 min. Measurements and tests indicate that the design of the calorimeter is sufficient for operation at room temperature.

Future plans are to significantly improve the operational efficiency by increasing the rapidity of making successive irradiation runs. This can be done by decreasing the time from the moment the water circulation is turned off to the time that the temperature drift is sufficiently small to start another irradiation run.

The motion and variation in temperature of the water in the region of the detector assembly was studied during the calibration of a thermistor removed from the glass container. Therefore, its response to the motion and temperature of the water was essentially instantaneous. About 10 min was required for the motion to subside. During this period there was a heat exchange between the water and its surroundings. This caused the water to be not strictly uniform in temperature. The variations in temperature could be significant compared to the small temperature rises during irradiation. It is required that the glass container be surrounded by motionless water so that if there is a temperature difference between that water and that within the container, heat conduction will take place to reduce the drifts to sufficiently small and predictable values during the irradiation period.

The above would require (1) a means for quickly stopping the motion of the water around the detec-

Table 2. Uncertainties with the H_2 system

Source	Estimated standard uncertainty (%)	
	Statistical, s_i (degrees of freedom, v_j)	Other, μ_j
Heat defect		0.3
Reproducibility of measurement groups (Fig. 28)	0.15 (4)	
Beam attenuation from glass wall		0.1
Beam attenuation of calorimeter lid	0.05 (5)	

Combined standard uncertainty $= \sqrt{\Sigma s_i^2 + \Sigma \mu_j^2} = 0.4\%$

tor assembly after circulation, and (2) that the circulated water (immediately after power turn off) is essentially at the same temperature as that of the stagnant water in the detector assembly. The means for (1) would simply be a thin horizontal baffle positioned at some distance below the detector assembly and attached to its two supports (Fig. 1) and to investigate if the baffle does not significantly interfere with the water circulation. The means for accomplishing (2) is by making a slight modification in the Wheatstone bridge shown in Fig. 4. During circulation, a switch would replace R′ with a thermistor (not shown) in the circulated water, and replace thermistor 1 with a resistor. The bridge would then respond to temperature differences between the two bodies of water. This would indicate if the circulation needed to be continued or the water slightly heated with the immersed heaters (not shown) to produce a bridge null condition. This would reduce the time for the final drift to subside to small values.

10. Recommendation

A recommendation pointed out in Ref. [4] needs to be repeated concerning protection against significant variations in ambient temperature. Although this was not the case during the measurements, a structure was built for that purpose. It is a four-sided acrylic structure having an expanded polystyrene lid with a polyethylene film covering the beam entrance window. A temperature sensor, heater, and fan were mounted on an inside wall. The circulated air was controlled to about 0.02 °C and set to match that of the water in the calorimeter.

11. References

[1] ICRU Report No. 14, Radiation Dosimetry: X-Rays and Gamma Rays with Maximum Photon Energies Between 0.6 and 50 MeV, International Commission on Radiation Units and Measurements, Washington, D.C. (1969).
[2] Task Group 21, Radiation Therapy Committee, AAPM, A Protocol for the Determination of Absorbed Dose from High-Energy Photon and Electron Beams, Med. Phys. **10**, 741 (1983).
[3] S. R. Domen, Med. Phys. **7**, 157 (1980).
[4] S. R. Domen, J. Res. Natl. Bur. Stand. (U.S.) **87**, 211 (1982).
[5] S. R. Domen, in The Dosimetry for Ionizing Radiation, Vol. II, K. R. Kase, B. E. Bjarngard, and F. H. Attix, eds., Academic Press, New York (1987) p. 245.
[6] S. R. Domen, Int. J. Appl. Radiat. Isot. **34**, 927 (1983).
[7] S. R. Domen, J. Res. Natl. Bur. Stand. (U.S.) **88**, 373 (1983).
[8] S. R. Domen and W. Z. Ba, Nucl. Instr. Meth. **B24/25**, 1054 (1987).
[9] B. Petree and P. J. Lamperti, unpublished data.
[10] ICRU Report No. 34, The Dosimetry of Pulsed Radiation, International Commission on Radiation Units and Measurements, Washington, D.C. (1982) p. 22.
[11] R. J. Schulz and M. S. Weinhous, Phys. Med. Biol. **30**, 1093 (1985).
[12] R. B. Barnett, Water calorimetry for radiation dosimetry, Ph.D. Thesis, Univ. of Calgary, Calgary, Alberta, Canada (1986).
[13] E. McLaughlin, Chem. Rev. **64**, 389 (1964).
[14] M. G. Velarde and C. Normand, Sci. Am. **243**(1), 93 (1980).
[15] S. R. Domen, in Proc. NRC Workshop on Water Calorimetry, C. K. Ross and N. V. Klassen, eds., National Research Council of Canada, Ottawa, (1988) p. 85.
[16] C. K. Ross, N. V. Klassen, and G. D. Smith, Med. Phys. **11**, 653 (1984).
[17] R. J. Schulz, C. S. Wuu, and M. S. Weinhous, Med. Phys. **14**, 790 (1987).
[18] H. J. Selbach, K. Hohlfeld, and H. M. Kramer, An experimental method for measuring the heat defect of water using total absorption of soft x-rays, Metrologia **29**, 341–347 (1992).
[19] M. Roos, B. Grosswendt, and K. Hohlfeld, An experimental method for determining the heat defect of water using total absorption of high-energy electrons, Metrologia **29**, 59–65 (1992).
[20] N. V. Klassen and C. K. Ross, Radiat. Phys. Chem. **38**, 95 (1991).
[21] S. R. Domen, J. Res. Natl. Bur. Stand. (U.S.) **93**, 603 (1988).
[22] N. S. Osborne, H. F. Stimson, and D. C. Ginnings, Values listed in the familiar Handbook of Chemistry and Physics, Chemical Rubber Publishing Co.
[23] T. H. Laby and E. O. Hercus, Proc. Phys. Soc. (London) **47**, 1003 (1935).
[24] S. R. Domen and P. J. Lamperti, J. Res. Natl. Bur. Stand. (U.S.) **78**, 595 (1974).
[25] J. S. Pruitt, S. R. Domen, and R. Loevinger, J. Res. Natl. Bur. Stand. (U.S.) **86**, 495 (1981).
[26] U. Gaur and B. Wunderlich, J. Phys. Chem. Ref. Data, **11**, 313 (1982).
[27] R. A. Robie, private communication.
[28] C. K. Ross, N. V. Klassen, K. R. Shortt, and G. D. Smith, in NRC Workshop on Water calorimetry, C. K. Ross and N. V. Klassen, eds, National Research Council of Canada, Ottawa (1988) p. 69.
[29] O. Mattsson, H. Svensson, G. Wickman, S. R. Domen, J. S. Pruitt, and R. Loevinger, Acta Oncologica **29**, 235 (1989).
[30] J. S. Pruitt, Absorbed-dose calibration of ionization chambers in a ^{60}Co gamma-ray beam, NIST Special Publication 250-40 (1990) p. 44.
[31] K. R. Shortt, S. R. Domen, and D. L. Bensen, Comparison of Dosimetric Measurement Standards of the NRCC and the NIST, manuscript in preparation.
[32] N. V. Klassen, manuscript in preparation.

About the author: *Steve R. Domen is a Guest Researcher in the Ionizing Radiation Division of the National Institute of Standards and Technology. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 2, March–April 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 143 (1994)]

Planar Near-Field Measurements of Low-Sidelobe Antennas

Volume 99 Number 2 March–April 1994

Michael H. Francis and Allen C. Newell

National Institute of Standards and Technology,
Gaithersburg, MD 20899-0001

Kenneth R. Grimm
Nichols Research Corporation,
Vienna, VA 22182-2222

John Hoffman
System Engineering Corporation,
Columbia, MD 21046

and

Helmut E. Schrank
Hunt Valley, MD 21030

The planar near-field measurement technique is a proven technology for measuring ordinary antennas operating in the microwave region. The development of very low-sidelobe antennas raises the question whether this technique can be used to accurately measure these antennas. We show that data taken with an open-end waveguide probe and processed with the planar near-field methodology, including probe correction, can be used to accurately measure the sidelobes of very low-sidelobe antennas to levels of −55 dB to −60 dB relative to the main beam peak. A special probe with a null in the direction of the main beam was also used for some of these measurements. This special probe reduced some of the measurement uncertainties but increased the uncertainties due to probe-antenna interactions. We discuss the major sources of uncertainty and show that the probe-antenna interaction is one of the limiting factors in making accurate measurements. The test antenna for this study was a slotted-waveguide array whose low sidelobes were known. The near-field measurements were conducted on the NIST planar near-field facility.

Key words: antenna measurements; far field; low-sidelobe antennas; planar near field.

Accepted: November 18, 1993

1. Introduction

This paper describes the efforts of several groups of scientists and engineers to determine the sidelobes of a very low-sidelobe array antenna using the near-field technique and to quantify the uncertainties. More than 20 years ago the National Institute of Standards and Technology (then the National Bureau of Standards) [1–3] and other laboratories [4] developed the theory and basic measurement methods for using the planar near-field (PNF) technique to determine the far field of antennas. Newell and Crawford [5] compared the far fields of array antennas obtained using near- and far-field measurement techniques. This comparison indicated that there was good agreement between the directly measured far field and the far field as determined using the near-field technique. Nevertheless, little has been reported on the effectiveness of using the near-field technique to determine the far-field patterns of a very low-sidelobe antenna.

In Sec. 2 we describe the experiment, including a description of the array, the design of a special probe, and a description of the near-field measurement facility. Section 3 describes the diagnostic tests and Sec. 4 describes the final measurement results. In Sec. 5 we evaluate the uncertainties. In Sec. 6 we discuss possible future research. Finally, in Sec. 7 a summary is given.

2. Design of Experiment

2.1 Purpose of Experiment

This experiment had three objectives. The first was to determine whether the PNF technique could be used to accurately measure low-sidelobe arrays. In order to determine this, two low-sidelobe arrays, the Ultra-Low-Sidelobe Array (ULSA) and the Airborne Warning and Control System Array (AWACS), were measured using the PNF technique.

Second, we wanted to show that the PNF analyses of Yaghjian [6] and Newell [7] were applicable to low-sidelobe arrays. We did this by purposely inducing certain errors in some of the measurements and comparing their measured effect with the effect predicted by theory. We also estimated the total expected uncertainties for this experiment. By comparing the final far field determined from the PNF technique to that determined using the far-field technique, we show that the differences are within what is expected from these analyses.

Finally, we investigated methods of reducing the effects of uncertainties. In particular, we showed that a difference (Δ) probe with a null in the direction of the main beam of the antenna under test (AUT) can be used to reduce the effects of some uncertainties.

2.2 Description of the Arrays

An S-band ULSA was used as the primary test antenna for this project. This antenna consists of eight equal-length slotted waveguides, fed at one end by a resonant waveguide manifold and terminated at the other end by high quality matched loads. The waveguide feed manifold has eight identical slotted T-junctions, fed in groups of four from each end of the vertical manifold waveguide which has a short circuit at its center. As illustrated in Fig. 1, the ULSA has a rectangular aperture (4.8 m $\times$ 0.4 m), and is designed for a -60 dB Chebyschev azimuth pattern with uniform illumination in elevation.

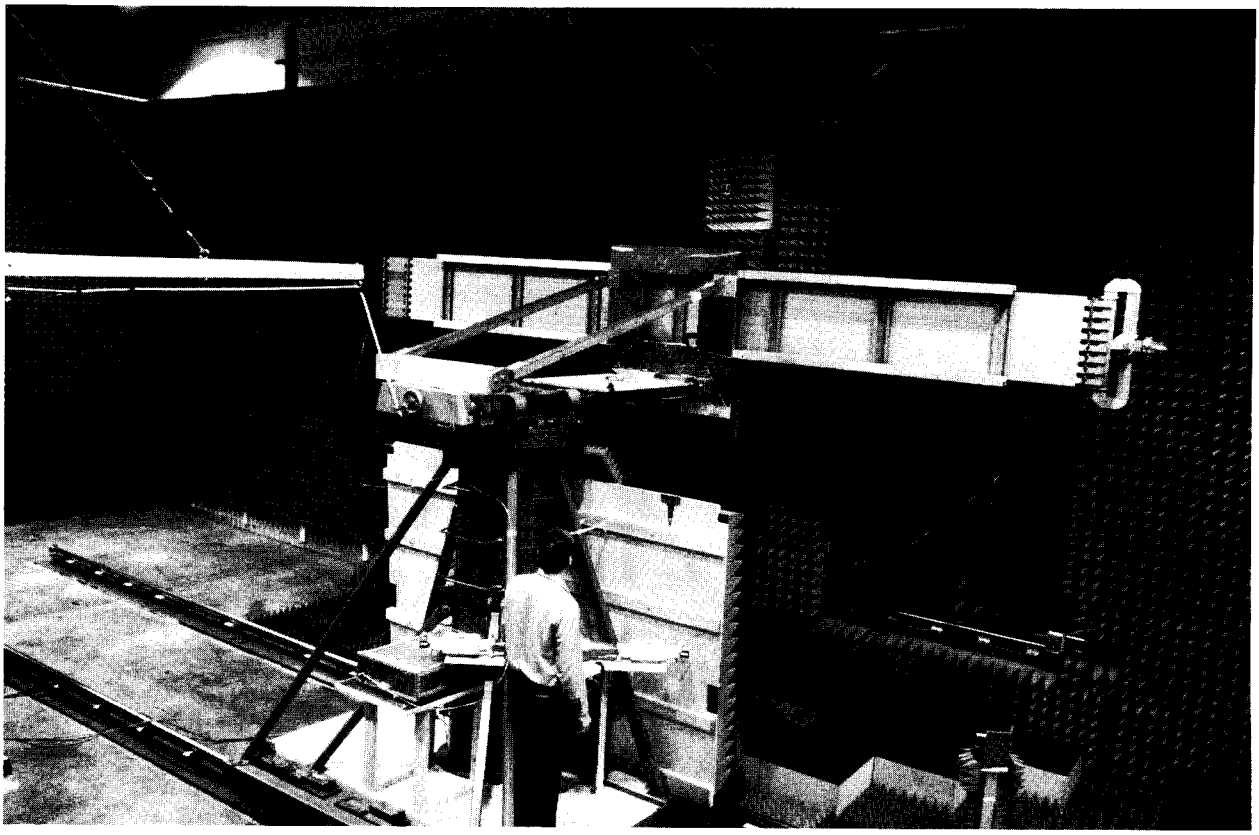


Fig. 1. The ULSA array.

An azimuth pattern, measured in 1978 at 3.025 GHz on a far-field range, is shown in Fig. 2. In this pattern, we note three things: (1) the main lobe is squinted off normal (toward the feed end) by an angle of $-13°$; (2) there is an anomalous lobe at $+44°$ about 31 dB below the main beam; and (3) the wide-angle sidelobes are well below -50 dB. After the array was refurbished for this project in 1983, the azimuth pattern was measured on the far-field range at 3.0 GHz (Fig. 3). Comparing Fig. 2 with Fig. 3 shows good agreement particularly in terms of the wide-angle sidelobe levels. In 1989, NIST measured the ULSA pattern at 2.9 GHz, 3.0 GHz, and 3.1 GHz. We will discuss these results in Sec. 4.

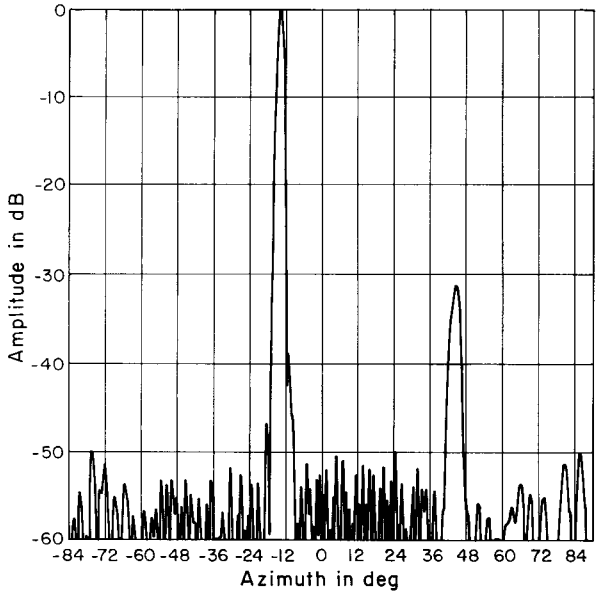


Fig. 2. Azimuth plane pattern of the ULSA taken in 1978 at 3.025 GHz.

The AWACS, which NIST measured in 1984, was built as an engineering prototype model. It has the same aperture dimensions (approximately 7.6 m × 1.5 m) and number of edge-slotted waveguides (28) as the production antennas, but uses manually driven ("trombone") phase shifters instead of the electrically controlled phase shifters to control the elevation pattern. The sidelobes in the azimuth pattern are actually slightly lower than those finally designed into the production antennas. These patterns cannot be shown because they are classified. However, we performed an analysis on these data and the uncertainties are approximately the same as those for the ULSA measurements. Figure 4 is a photograph of the AWACS array showing it mounted on the NIST PNF range.

The main beam radiated by both the ULSA and the AWACS is not normal to the array face, but is squinted toward the feed end by an angle of [8]

$$\sin\theta = \frac{\lambda}{2S} - \frac{\lambda}{\lambda_g}, \tag{1}$$

where λ is the free space wavelength, S is the slot spacing, and λ_g is the guide wavelength. If we substitute for λ_g for the TE_{10} mode using the results of Silver [9], we can rewrite Eq. (1) as

$$\theta = \sin^{-1}\left[\frac{\lambda}{2S} - \sin\left(\cos^{-1}\left(\frac{\lambda}{2a}\right)\right)\right], \tag{2}$$

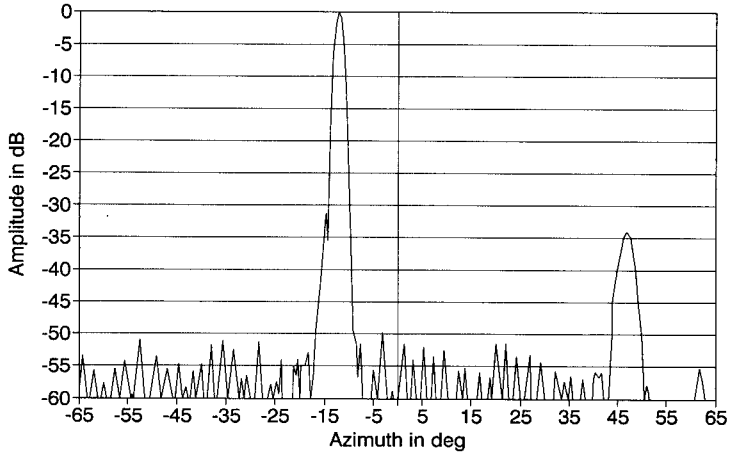


Fig. 3. Azimuth plane pattern of the ULSA taken after refurbishment at 3.0 GHz.

where a is the inside wide dimension of the waveguide. With this expression, we can predict the squint angles for the ULSA array, using $a = 7.214$ cm and $S = 5.32$ cm, yielding squint angles of 12.7° at 3.0 GHz and 9.7° at 3.1 GHz, respectively. Similarly, we can predict the squint angles of the AWACS using $a = 5.817$ cm and $S = 4.55$ cm.

The near-field scanning tests were then performed for fixed-beam steering directions in accordance with the selected cw test frequency. When we scanned with the Δ probe we steered the probe pattern null to be coincident with the test array squint angle at each test frequency. The use of such a probe and its design is described next.

2.3 Design of the Difference Probe

Huddleston has shown that an optimum probe pattern can minimize uncertainties in the measured test antenna spectrum due to truncation of a finite scan plane [10]. The Huddleston probe has a pattern of increasing directivity for conditions of decreasing scan plane area. For a particular planar scanning geometry, the optimum probe concentrates the received power within the available near-field sampling grid; that is,

$$\int\int |v_L(x,y)|^2 \, dx\, dy = P_0(1-\epsilon^2_{xy}), \tag{3}$$

where $v_L(x,y)$ is the scan plane voltage, P_0 is the total power radiated into the forward hemisphere, and ϵ_{xy} measures the concentration of $|v_L(x,y)|^2$.

The fractional power not concentrated by the probe over the finite scan area is $P_0\epsilon^2_{xy}$. Upon transform, the measured coupling product will be an aliased version of the true spectrum. The magnitude of the difference between the aliased and true spectra has a maximum uncertainty of

$$\int_{-k_{x\max}}^{k_{x\max}} \int_{-k_{y\max}}^{k_{y\max}} |A_L - A_L'|^2 \, dk_x \, dk_y \le \frac{P_0\epsilon^2_{xy}}{(2\pi)^2}, \tag{4}$$

where $|k_{x\max}| = \pi/\delta_x$, δ_x is the x sampling increment, $|k_{y\max}| = \pi/\delta_y$, δ_y is the y sampling increment, A_L is

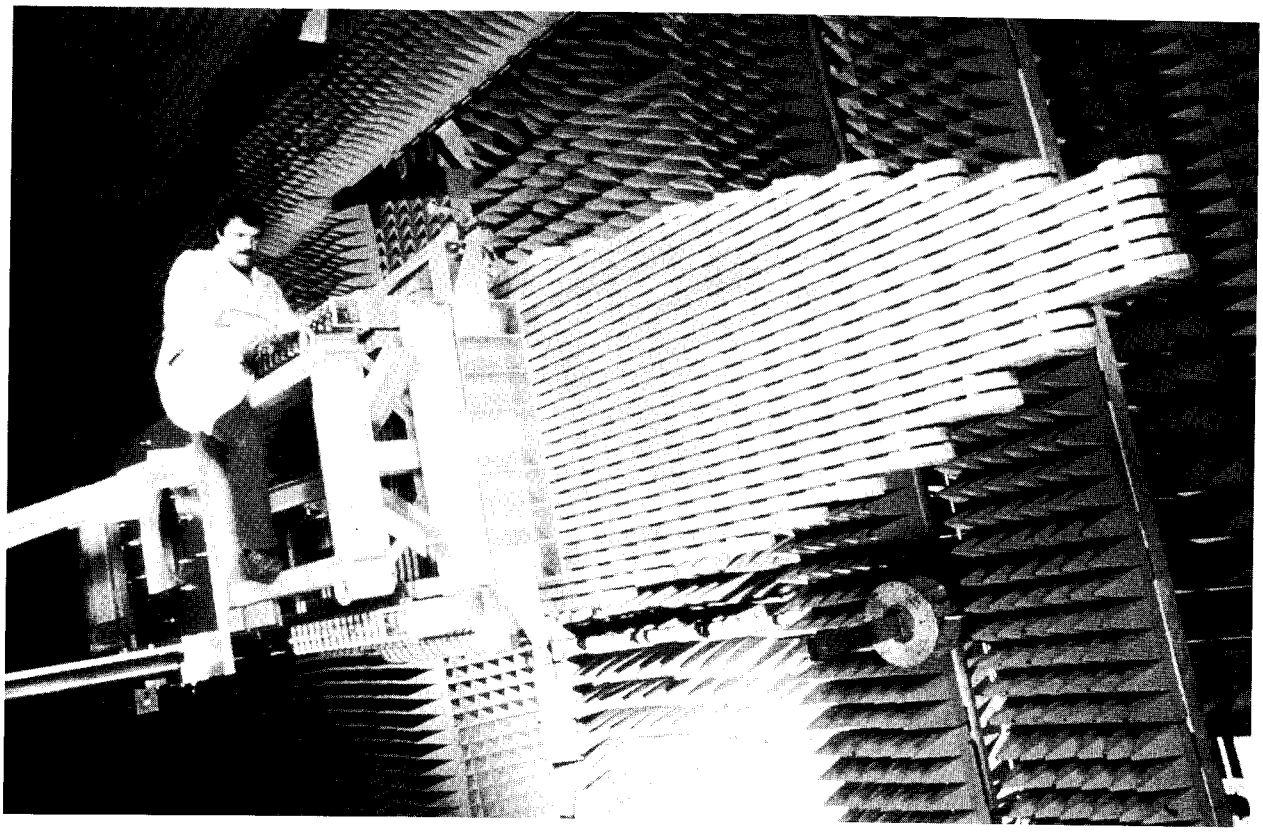


Fig. 4. The AWACS array.

the true spectrum, and A_L' is the measurement estimate of A_L.

Huddleston probes minimize $P_0\epsilon^2_{xy}$ by radiating with higher directivity as the ratio of test antenna area to scan plane area approaches 1. For low-sidelobe antennas, we can extend the optimum probe concept by requiring that Eq. (4) apply only over the test antenna's sidelobe region. This leads to an expression for a near-field scan voltage, based on a band-limited version of the reaction integral written by Joy, as [11]

$$v(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[k_xA\times$$

$$(Bw)]e^{-jk_zz}\,e^{(jk_xx+jk_y)}\,dk_x\,dk_y\,, \qquad (5)$$

where w is a spectral windowing function to be imposed on the coupling product spectrum via the probe's far-field pattern. For usual broad beam open-end waveguide probes, $W(k_x, k_y)=1$, and the coupling product spectrum $A\cdot(Bw)$ is band limited only by the high attenuation of the evanescent modes on the scan plane; that is, $A\cdot(Bw)\rightarrow 0$ for $k_z \geq [1-(k_x/k)^2-(k_y/k)^2]^{0.5}$ and $|z|>\lambda$.

However, if w is an intentional stop band in the probe pattern (that is, a null region), then the near-field coupling product spectrum will be band limited by both evanescent cutoff as well as real angle filtering. The ideal probe pattern weighting function for sidelobe testing is then given by

$$W(k_x,k_y)=1 \text{ if } k_c<|k_{x,y}|<k_{xmax,\,ymax}, \text{ sidelobe}$$

$$=0 \text{ if } 0\leq|k_{x,y}|\leq k_c, \text{ mainbeam} \qquad (6)$$

and is sketched in Fig. 5. This pattern is not realizable for any practical small aperture probe, but a useful approximation has been designed and built by combining the anti-phased outputs of two collinear waveguide elements. This probe has been called the difference or Δ probe. In-phase combining of elements produces a sum or Σ probe.

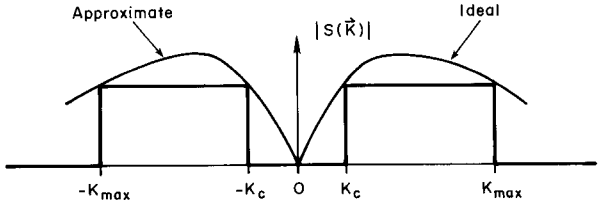


Fig. 5. The ideal probe pattern.

Figure 6 shows one of three two-element probes built for this project, and Fig. 7 shows a contour plot of this probe's measured far-field azimuth pattern at 3.1 GHz. The shifted null is accomplished by physically steering the Δ-probe by an appropriate angle or by using unequal cable lengths between the transition of each waveguide element and an integral coaxial hybrid coupler mounted on the probe carrier plate. The waveguide elements for the probe of Fig. 6 were spaced horizontally by 4.8 cm (0.5 λ at 3.1 GHz). The unequal length cables to the hybrid combiner provided a fixed null shift coincident with the test array's mainbeam pointing direction. More commonly, the null coincidence was maintained by mechanically rotating the probe carrier plate to fixed null pointing directions and then using equal length cables to the hybrid. Of course, we must measure two-dimensional vector probe patterns for all probe-frequency combinations prior to near-field scanning.

2.4 Description of the Measurement Facility

A typical near-field measurement system can be conveniently discussed in terms of three subsystems: (1) computer, (2) rf source and receiver, and (3) mechanical scanner and probe positioner. A great deal of variety is possible for each of these subsystems. We will describe only the essential features of each subsystem used at the NIST PNF measurement facility.

2.4.1 Computer Because of the large amounts of data involved, computer control of the measurement system is essential. In the NIST measurement facility, a special purpose microcomputer is used as the position controller. The position controller receives digital inputs from the x–y position encoders (in this case a laser interferometer), controls the motor drives moving the probe, and triggers the receiver to perform measurements at predetermined points. The NIST facility also has a minicomputer which records the data on a data storage device, (either magnetic tape or hard disk), monitors the receiver and position controller for errors, and performs some of the data analysis during the diagnostic tests. A mainframe computer (1983) or 386 PC (1989) is used for the bulk of the data analysis, especially the two-dimensional FFT.

2.4.2 RF System The basic components of the rf system are shown schematically in Fig. 8. They are: the transmitting and receiving antennas, isolators, mixers, variable attenuator and phase shifter, receiver, and synthesizer (signal source). The signal source must be stable in frequency and power level

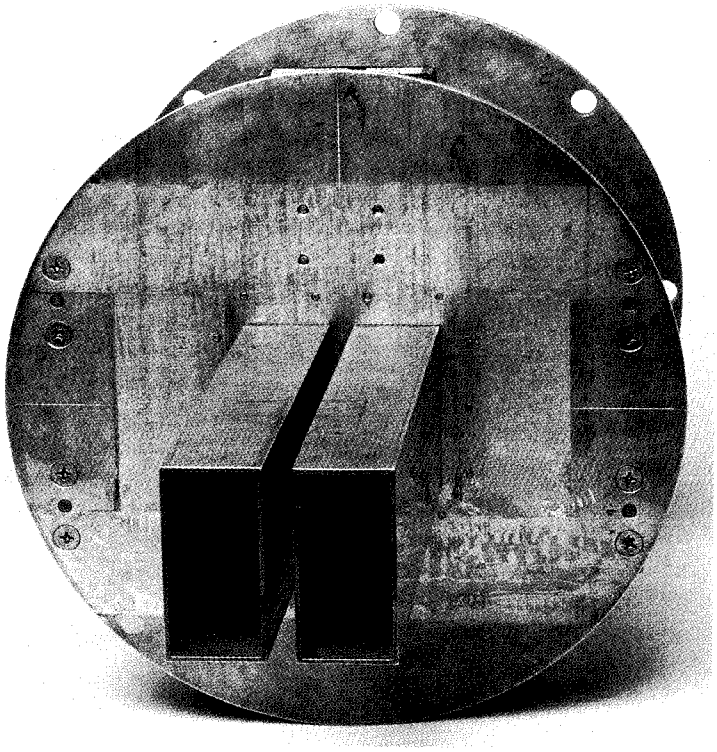


Fig. 6. The Δ probe showing both its elements.

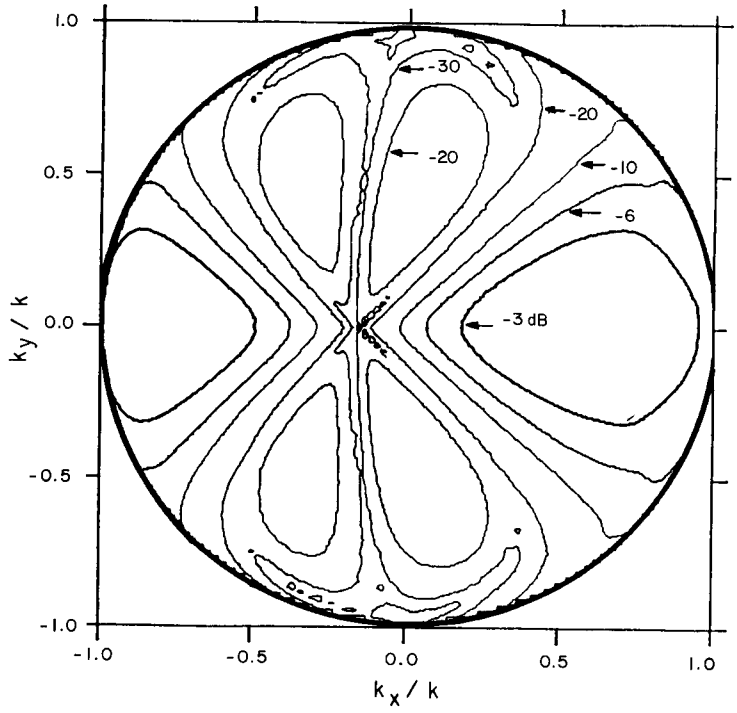


Fig. 7. Contour plot of the Δ probe azimuth pattern.

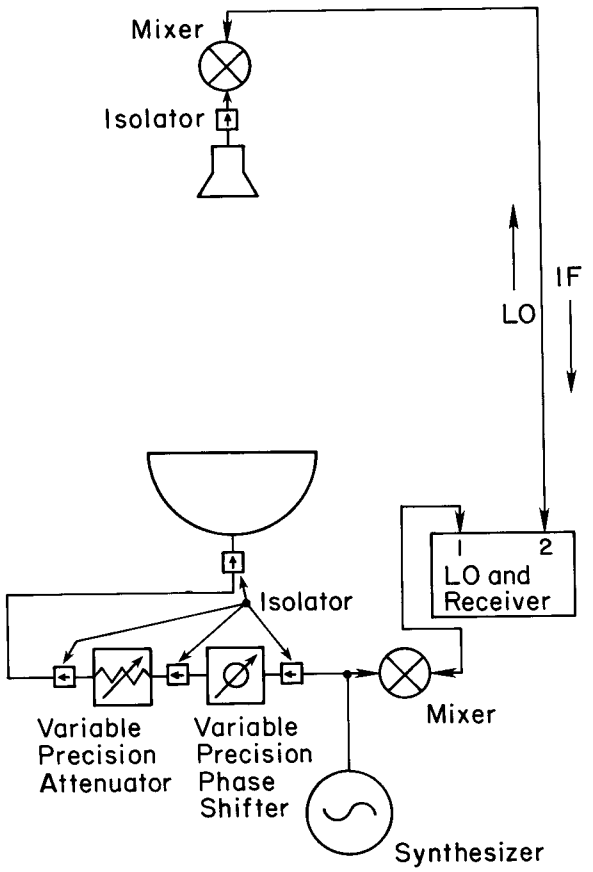


Fig. 8. Rf measurement system schematic.

to minimize its contribution to measurement uncertainty. Generally, a frequency stability of a few parts in 10^6 and amplitude variations of less than 0.1 dB over the total scan time are adequate. Tie scans can be used to correct for small amounts of drift which are on the order of a few degrees in phase and a few percent in amplitude over the total scan time. The source power output requirements depend on the gains of the AUT and the probe as well as the mixer sensitivity. The maximum magnitude of the near-field amplitude $|b_0'(P)|_{max}$ is approximately given by

$$|b_0'|^2_{max} \approx *(G_p/G_a)|a_0|^2, \quad (7)$$

where G_p and G_a are respectively the probe and AUT gains, and a_0 is the input amplitude to the transmitter. To reduce noise for the low-sidelobe measurements, $b_0'(P)|_{max}$ should be at least 65–70 dB above the noise but also be low enough to be within the linear range of the mixer. When the probe is cross polarized to the AUT, the maximum signal drops to about 30 dB below the maximum signal of the co-polarized case. It is desirable for the cross-polarized maximum signal to be of the same order as the co-polarized maximum signal. This can be achieved by increasing a_0, by increasing the receiver gain, or by some combination of the two.

A critical requirement of the source, and in fact for the complete rf system, is that it be well shielded. Signal leakage from the source, transmission lines, and input components, or signal pick-up by similar parts of the receiving system can cause significant measurement uncertainties in the near-field amplitude and phase. Tests must be performed to guarantee that the whole system is well shielded and that signals can only be transmitted and received via the AUT-probe transmission path.

The receiver is a very important part of the measurement system and must accurately measure the near-field amplitude and phase of the rf signal over a dynamic range of at least 80 dB. It must be stable over the time required for near-field scanning, which may be hours, and have good linearity and resolution in its conversion from rf signal to digital output.

2.4.3 Mechanical Scanner and Probe The mechanical scanner or probe positioner consists of the supports, guides and drive motors to move the probe over the planar area, the encoders to measure the probe position, and the rf transmission line to couple the probe output to the receiver. The "box frame" design of the NIST scanner, Fig. 9, was one of the first PNF range designs. A large, rigid frame, constructed from metal I beams, serves as the base and support for the two horizontal guide rails. Both rails are precision ground, stainless steel cylinders supported at intervals of about 30 cm. Precision linear ball bearings attach the vertical column to the horizontal guide rails. The box frame requires additional support to maintain a rigid vertical structure. This is accomplished by attaching it to a stable interior wall.

At the NIST facility an auxiliary linear translator has been attached to the AUT to extend the effective measurement area, as shown in Fig. 10. The essential requirement of this linear translator is that the translation be precisely known and controlled to correctly combine different segments of the measurement plane. Also, to correctly combine these different segments there should be some overlap between them. By comparing the amplitude and phase of the overlap regions of the different segments, we can correct for both amplitude and phase drift between segments.

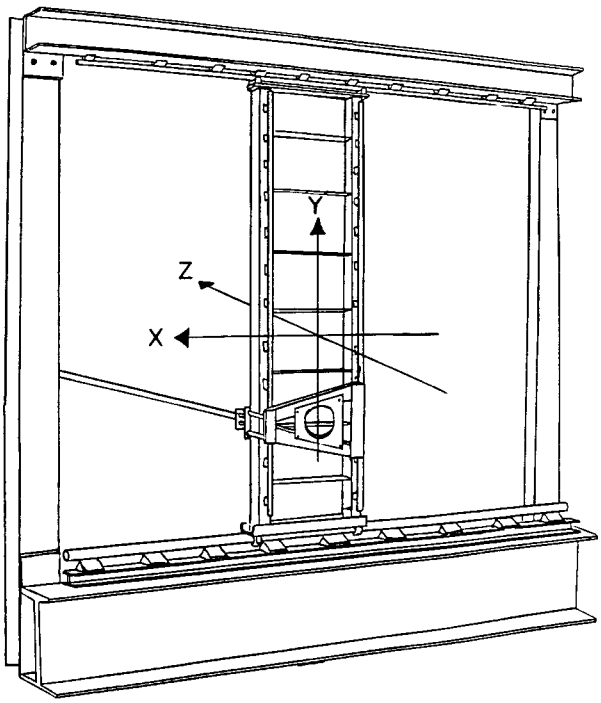


Fig. 9. Box frame planar scanner.

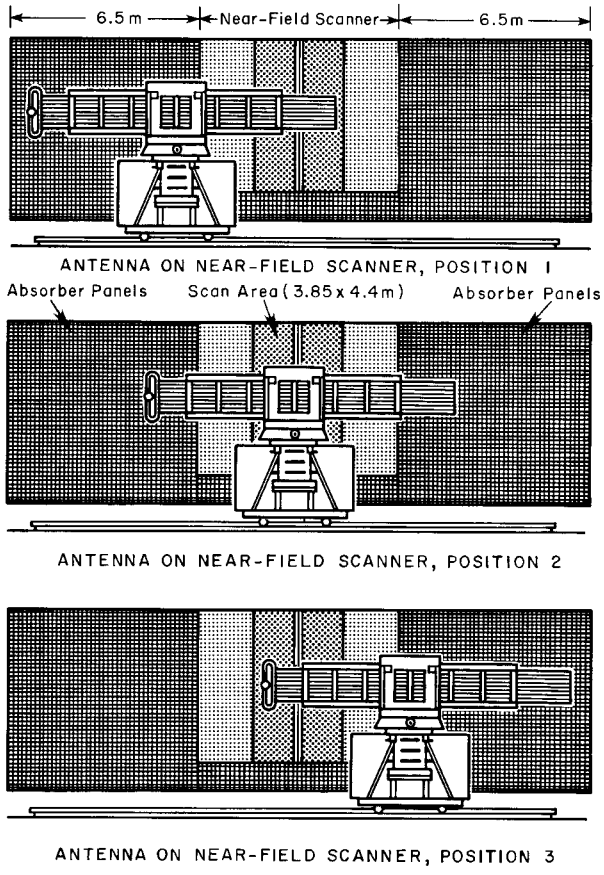


Fig. 10. Schematic of antenna in three positions relative to the scanner.

A common concern in any scanner is to ensure that the rf transmission line between the probe and receiver move without causing significant amplitude or phase change in the measured data. At the NIST facility, the transmission line is supported on a scissors-arm mechanism so that rotation is of concern at only three joints. We use service loops of semi-rigid or flexible coaxial cable at these three points to produce a stable transmission line.

3. Preliminary Tests

Several preliminary, one-dimensional tests are performed to determine the required near-field spacing, required scan size, the best value of the probe-antenna separation distance, and to determine the level of the leakage and reduce it if necessary. These tests can also be used to estimate some of the measurement uncertainties. However, two-dimensional tests (described in Sec. 4) provide better uncertainty estimates since one-dimensional tests assume that the pattern is separable into functions of X and Y. Usually, this is only approximately true.

3.1 Test Space

The test space procedure is used to determine the required near-field spacing. It consists of taking one-dimensional scans in X and Y with very fine spacing ($\approx 0.05\ \lambda$). First an FFT is performed on the full set of data, then using only every other point, then using every third point, and so forth. From this, the far-field spectra from various spacings can be compared. The smallest spacing is assumed adequate and when the spacing from the other FFTs is so large that the spectrum changes by more than the desired accuracy a spacing equal to or smaller than the next smallest spacing is chosen.

From these one-dimensional tests we concluded that the spacing between near-field data points should be about 0.4 λ_{min} in both X and Y, where λ_{min} is the wavelength at the highest measurement frequency.

3.2 Test Scan

The test scan procedure is similar to the test space procedure in that one-dimensional centerline scans in X and Y are performed. In this case, very long centerline scans are performed, and data are truncated from the edge in various amounts. As in the test space procedure the computed spectra are compared and a required scan length determined.

The choice of scan area was 10.4 m in X by 3.8 m in Y.

3.3 Multiple Reflection Tests

A Z-multiple-reflection test, where data are taken as a function of Z at several fixed X and Y values, is performed. A separation distance is chosen to minimize the peak-to-peak variations as a function of Z. From these tests the probe-antenna separation distance was chosen to be 65 cm for the ULSA and 35 cm for the AWACS.

A second multiple-reflection test is designed to estimate the effects on the far field of multiple reflection interactions between the probe and AUT. The test is implemented by taking centerline scans in both X and Y at several different separation distances (see Fig. 11). Consecutive separation distances differ by $\lambda/8$. One of these separation distances is equal to the separation distance determined using the first Z-multiple-reflection test.

An FFT is performed on the centerline scans to obtain a one-dimensional far field; these far fields are averaged together after correcting the phases for the different separation distances. This average is then subtracted in a complex manner from each one-dimensional centerline-scan far field to estimate the magnitude and character of the multiple reflections. Additionally, the average far field can be transformed back to the near field to obtain an average one-dimensional near field. This average can be subtracted from the individual centerline near-field scans to obtain the near-field character of the multiple reflections. The gain and one-dimensional patterns can be compared and the uncertainty due to multiple reflections estimated.

Sample results are shown in Figs. 12–15. Surprisingly, the character of the multiple reflections bears some resemblance to the average far-field pattern. We also find that there are two different periods in the near-field amplitude and phase multiple reflections. The predominant period is twice the spacing between elements and corresponds to the spacing between elements of the same slant direction (every other element). This period is clearly visible in Figs. 14 and 15. The second period corresponds to the spacing between consecutive elements and only becomes apparent in the invisible space part of the far-field spectrum (Fig. 16). Each period produces distinctive lobes in the computed spectrum. The location of each lobe depends on the corresponding periodicity in the near field. The first lobe is due to a near-field periodicity of approximately λ and occurs in the computed spectrum at $|k_x/k| \approx 1$ (relative to the main beam direction) or near $\theta = 90°$. The second lobe is due to a near-field periodicity of approximately $\lambda/2$ and therefore occurs in the computed spectrum at $|k_x/k| \approx 2$ (relative to the main beam direction). This is in the invisible part of the spectrum.

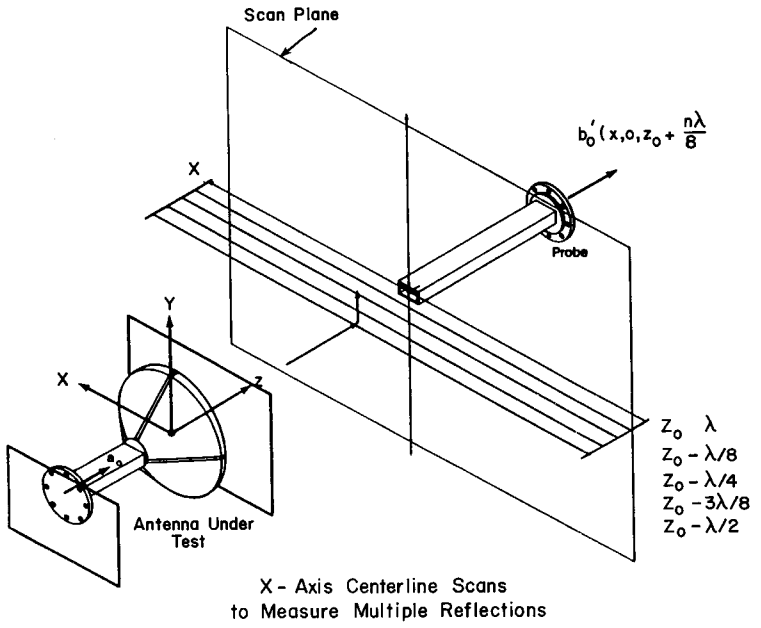


Fig. 11. Schematic of centerline multiple reflection tests.

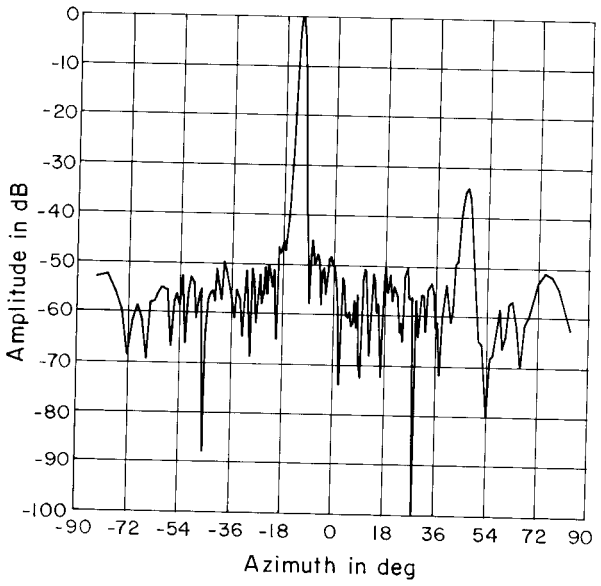


Fig. 12. Sample result for the far field after averaging for multiple reflections.

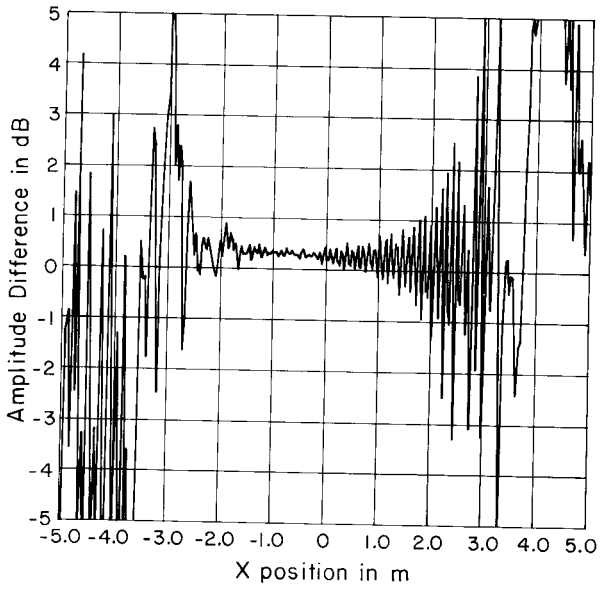


Fig. 14. Sample multiple reflection near-field amplitude using the sum probe.

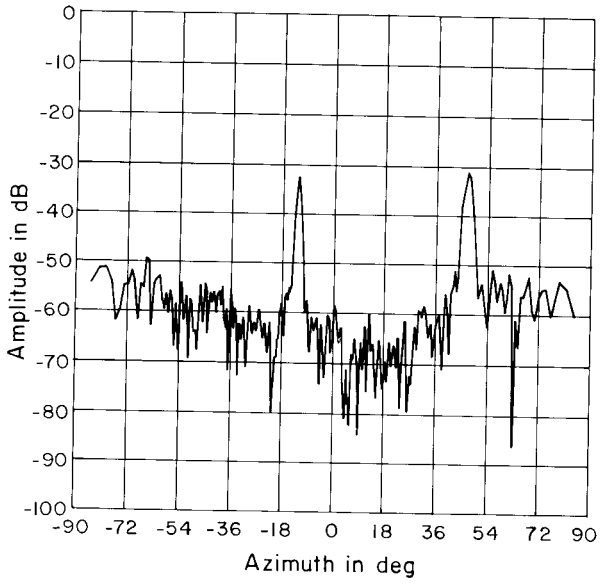


Fig. 13. Sample multiple reflection spectrum amplitude, using the sum probe, relative to the peak average far field for the sum probe.

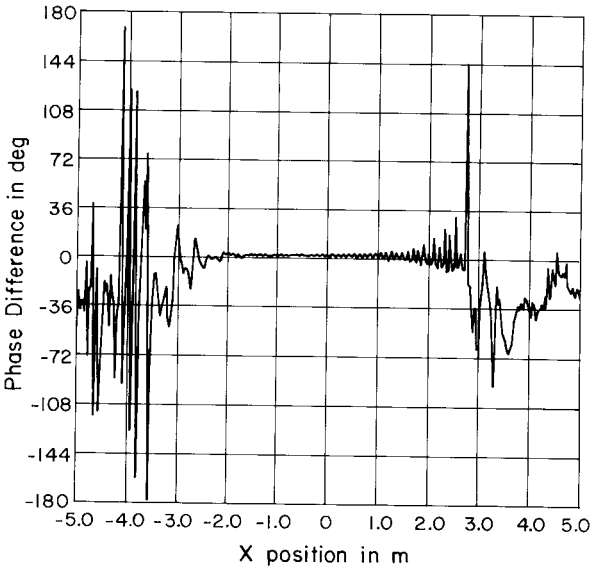


Fig. 15. Sample reflection near-field phase using the sum probe.

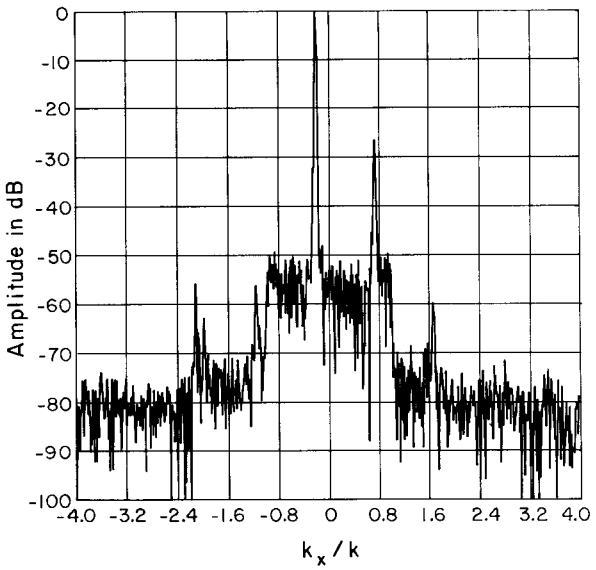


Fig. 16. Test space far-field spectrum showing multiple reflection lobes.

3.4 Leakage Tests

Leakage tests are performed by taking centerline scans with a termination on the receiving side in place of the probe to test for leakage on the receiving side and by doing centerline scans with a termination on the transmitting side in place of the AUT. The significant sources of leakage were located and shielded so as to reduce the leakage level to −75 dB relative to the near-field peak.

4. Two-Dimensional Measurement Results

4.1 Summary of Measurements

Some two-dimensional near-field measurements were performed with each of the probes. In order to compare the results and the uncertainties associated with each probe there was some overlap in these measurements. These measurements are summarized in Table 1.

Fewer measurements were performed with the Σ probe. This was because after a few measurements this probe was found to behave in a manner closely resembling the open-end waveguide.

Measurements were performed at different distances to estimate the effects of multiple reflections between the probe and the AUT.

Table 1. Summary of two-dimensional near-field measurements

AUT	Frequency (GHz)	Open-end Waveguide	Δ	Σ
ULSA	2.9		$M, Z = 65$ cm	$M, Z = 65$ cm
ULSA	3.0	$M, Z = 65$ cm		
ULSA	3.1	$M, Z = 65$ cm	$M, Z = 65$ cm	$M, Z = 65$ cm
ULSA	3.1	$C, Z = 65$ cm		
AWACS	F2	$M, Z = 35$ cm	$M, Z = 35$ cm	
AWACS	F2	$C, Z = 35$ cm		
AWACS	F2	$M, Z = 37.5$ cm	$M, Z = 37.5$ cm	

Δ = measurement with difference probe.
Σ = measurement with sum probe.
Z = the probe-antenna separation distance.
M = measurement of main component.
C = measurement of cross component.

4.2 Combining Segments

Both the ULSA and the AWACS have long dimensions which exceed the capability of the NIST scanner. It is, therefore, necessary to measure the near field of both these antennas in segments and to combine these segments into one near field. As discussed earlier, the purpose of the linear rails is to allow the AUT to be moved while preserving an accurate knowledge of the X-position.

In order to combine the three segments accurately we must account for the effects of both amplitude and phase drift. This is accomplished by requiring some overlap between adjacent segments. In particular, the measurement was performed with adjacent segments having an overlap of five scans which were taken vertically.

A computer program, STITCH, was used to combine adjacent segments using the following technique. The amplitude and phase of the center section which contained the near-field peak signal was used as the reference. The amplitudes and phases of the adjacent segments were compared to those of the center segment in the overlap region. This comparison was done by computing an amplitude ratio and phase difference at each point in the overlap region. These amplitude ratios and phase differences were averaged using a weighting which was proportional to the square of the amplitude. The amplitudes of each point in the adjacent segments were corrected by multiplying by the average of the ratio of the center segment amplitude to the adjacent segment amplitude. The phases were corrected by adding the average phase difference.

The amplitude correction was a few percent at most and the phase correction was a few degrees at most. A comparison of two typical overlap scans is shown in Fig. 17.

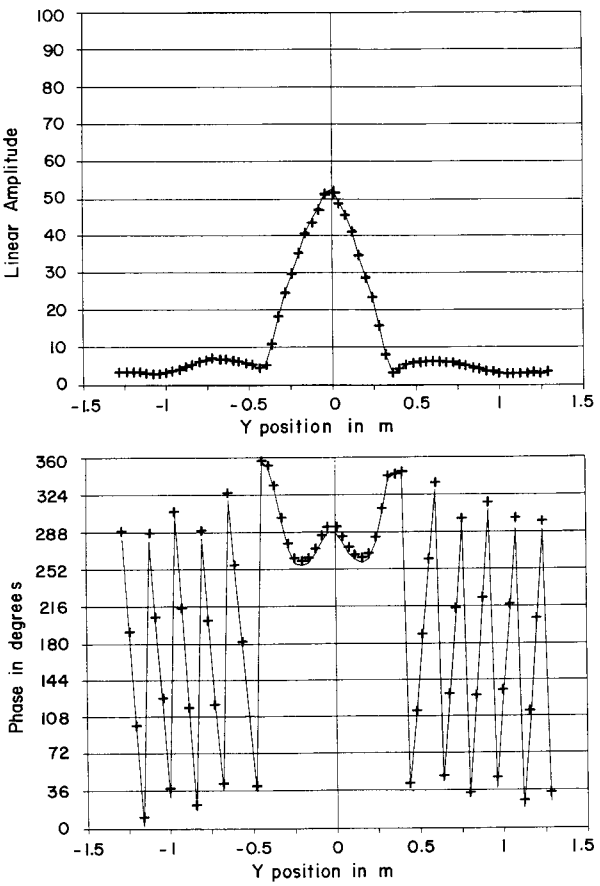


Fig. 17. A comparison of two vertical scans from adjacent segments which overlap, amplitude (top), phase (bottom), scan 1 (solid line), scan 2 (+ + +).

4.3 Room Scattering Test

To estimate the effects of room scattering NIST developed a room scattering test, SCAT, which takes advantage of the unique feature of the NIST facility of being able to move both the AUT and the probe in the X-direction.

Ideally, to measure room effects one would like to move the room while keeping the antennas fixed. This is obviously impractical! Instead, we set up a test procedure where the antenna and probe were moved together in the X-direction relative to the room so that the X-coordinate of the probe with respect to the AUT was fixed. A Y-scan was taken at each X-position. These scans would be identical if there were no room scattering. The room scattering can be estimated by obtaining a single average Y-scan and subtracting it from each of the others; an FFT is then performed on the resulting two-dimensional data to obtain an estimate of the room-scattering far-field spectrum.

The results of the tests indicated that the room-scattering far field was approximately random (see Fig. 18) and the peak value was about −70 dB relative to the peak of the AUT's far field. The root mean square (RMS) value of the room scattering was found to be −89 dB for the open-end waveguide and −94 dB for the Δ probe.

4.4 Truncation Tests

According to Yaghjian [6] we can estimate an upperbound to the uncertainty due to scan plane truncation by performing an integration of the data along the outside edge of the near-field scan area. We obtained this spectrum by setting the rest of the near-field data to zero and performing the usual FFT. The resulting truncation spectra for the azimuth cut are shown in Fig. 19 (for the open-end waveguide–AWACS combination) and Fig. 20 (for the Δ probe–AWACS combination). These plots are normalized relative to the AUT's peak far field. The Δ probe is on the whole about 7 dB better than the open-end waveguide. The magnitude of the truncation spectrum is generally between about 63 dB and 70 dB below the far-field peak for the open-end waveguide and between about 70 dB and 75 dB below the far-field peak for the Δ probe.

4.5 Simulated Position Error

To verify the analysis of Yaghjian [6] and Newell [7] with respect to position errors some measurements were done to simulate the effects of periodic Z-position errors. The analysis predicts that Z-position errors have a smaller effect on the spectrum obtained using the Δ probe than on the spectrum obtained using the Σ probe or open-end waveguide. Four PNF measurement scans were done for this simulation, two for the Σ probe and two for the Δ probe. Of the two scans for each probe, one had no errors introduced and the other had periodic errors purposely introduced. These errors had periods of 2.5 λ, 1.67 λ, and 1.11 λ. The magnitude of the error for each period can be found in Table 2. In addition, Table 2 contains a summary of the spectral location, predicted spectral error, and the actual measured errors.

Table 2 shows that the measured errors are in substantial agreement with the predicted. More important, as predicted by theory, the simulated position error has a smaller effect on the spectrum

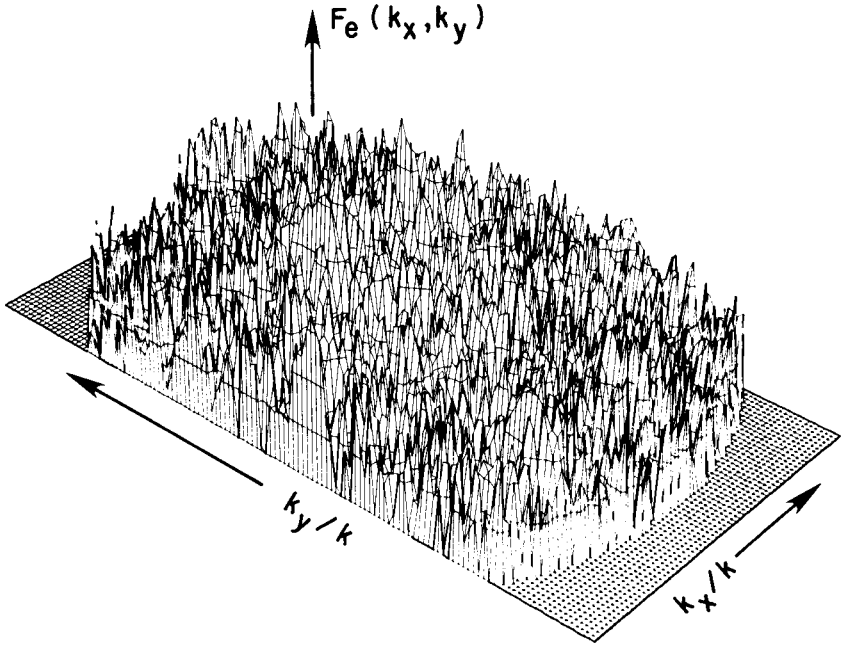


Fig. 18. Sample scattering far field using the open-end waveguide. Peak is −70 dB relative to the peak AUT far field.

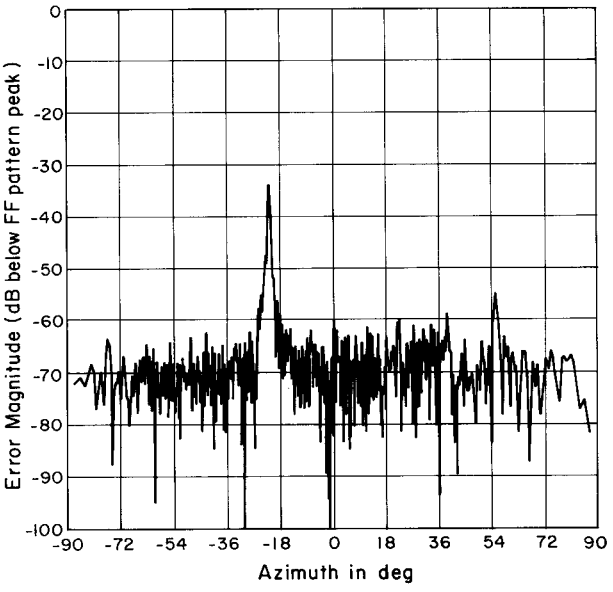


Fig. 19. Azimuth cut of the truncation spectrum using the open-end waveguide probe.

obtained using the difference probe than on the spectrum obtained from the Σ probe. Figure 21 compares the far field with and without errors for the Σ probe, and Fig. 22 is the same comparison for the Δ probe. As expected the position errors are smaller for the Δ probe than for the Σ probe.

Table 2. Predicted and measured errors due to simulated Z-position errors

Period	Amplitude	Probe	Measured location (k_x/k)	Predicted ΔD (dB)	Measured ΔD (dB)
2.5 λ	0.13 mm	Σ	0.23	9.5	9.5
			−0.57	7.9	8.6
1.67 λ	0.25 mm	Σ	0.43	9.6	11.9
			−0.77	7.0	8.6
1.11 λ	0.38 mm	Σ	0.73	7.7	8.8
2.5 λ	0.13 mm	Δ	0.23	0.3	0.2
			−0.57	0.25	0.3
1.67 λ	0.25 mm	Δ	0.43	0.3	0.2
			−0.77	0.2	0.3
1.11 λ	0.38 mm	Δ	0.73	0.2	0.3

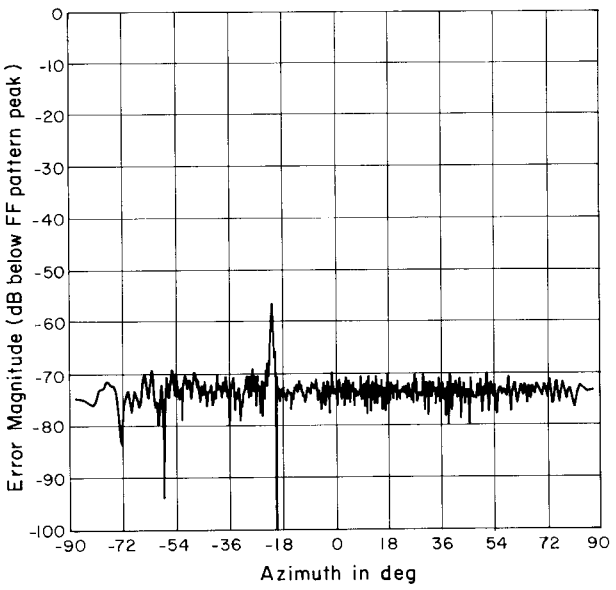


Fig. 20. Azimuth cut of the truncation spectrum using the Δ probe.

4.6 Probe-AUT Multiple Reflections

We can obtain a better estimate of the uncertainty due to probe-AUT multiple reflections by taking a full set of two-dimensional near-field data at two different separation distances which differ by $\lambda/4$. Then we average (in a complex manner) the two far-field spectra and subtract (again in a complex manner) the average from each of the individual spectra to obtain a multiple-reflection spectrum.

The resulting spectrum for the open-end waveguide-AWACS combination is shown in Fig. 23 and that for the Δ probe and AWACS is shown in Fig. 24. The multiple-reflection spectrum for the Δ probe is 3 dB to 5 dB higher than that for the waveguide probe. The multiple reflections associated with the Δ probe are greater because the Δ probe is larger.

4.7 Data Processing

For an arbitrarily polarized test antenna, the transmitting coefficient for the main and cross polarizations are given by Eq. (8a) and Eq. (8b). Since the D′ and D″ terms are scalar quantities, the measured far-field polarization of the test antenna will be determined by the polarization components of the probe calibration files. If the probe and test antenna have matched main-polarized components and the cross-polarized responses of the probe and test antenna are small, we then use Eq. (10a) to obtain the main component of the test antenna.

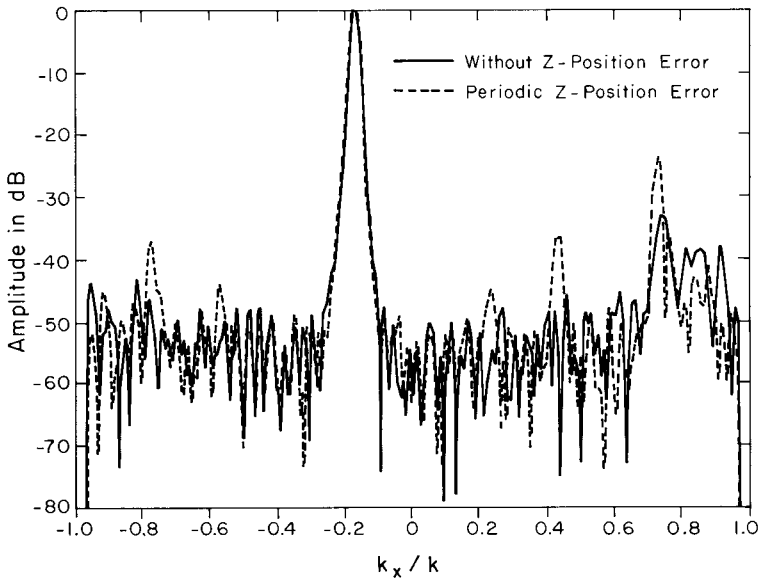


Fig. 21. Azimuth far-field cut for the sum probe with simulated position errors (dashed line) and without errors (solid line).

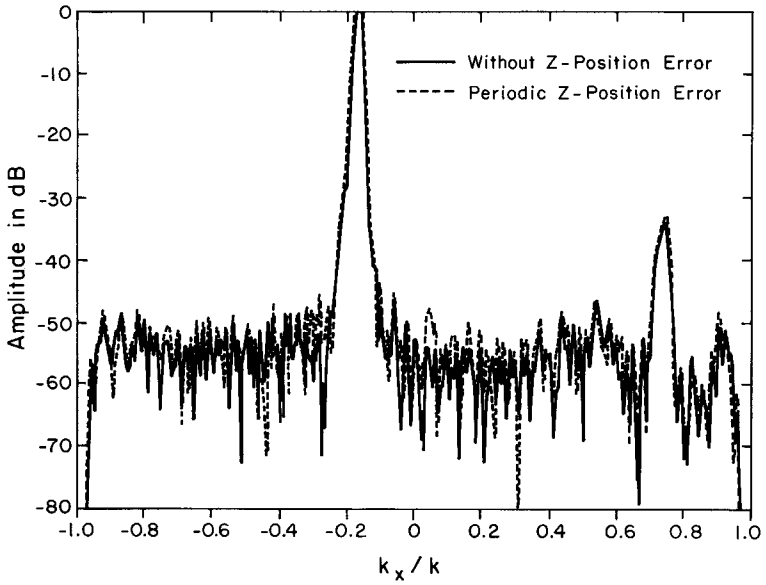


Fig. 22. Azimuth far-field cut for the Δ probe with simulated position errors (dashed line) and without errors (solid line).

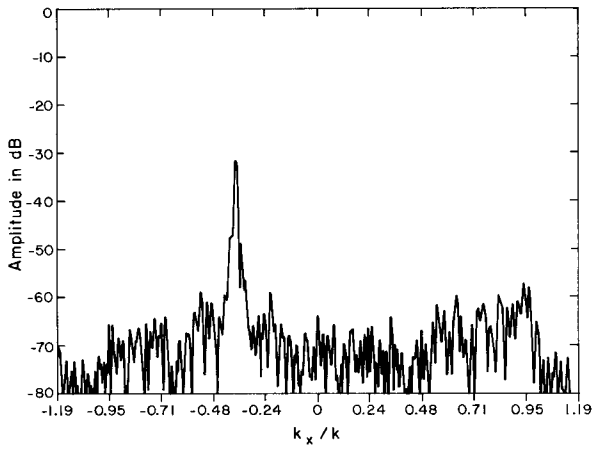


Fig. 23. Multiple reflection spectrum for the waveguide probe using data taken from two near-field scans separated by λ/4.

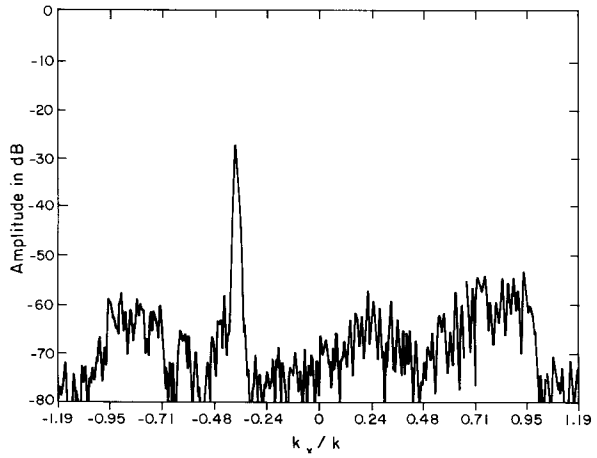


Fig. 24. Multiple reflection spectrum for the Δ probe using data taken from two near-field scans separated by λ/4.

Near-field centerline amplitude scans for the ULSA array as measured by the Δ probe, the Σ probe, and the open-end waveguide for a Z-separation of 65 cm are shown in Fig. 25. The filtering effect of the Δ probe is responsible for the compression of the near-field dynamic range as well as much of the high-frequency amplitude ripple. Following FFT processing, we show the corresponding coupling product transforms for the open-end waveguide and the Δ probe in Fig. 26, plotted over the real wavenumber space only ($|k_x/k| < 1.0$). The coupling product for the Σ probe differs little from that of the waveguide probe and we do not show it here. The filtering property of the Δ probe is clearly visible by noting the reduced main beam and suppressed sidelobes near the main beam. Similarly, the Σ probe filters the sidelobes at wide angles beyond the large anomalous lobe ($|k_x/k| > 0.8$) but is accurately measuring the mainbeam region. The coupling product spectrum can be probe-corrected to obtain the principal azimuth angle pattern of the ULSA array. Figure 27 shows the resulting probe-corrected far field for the waveguide probe at 3.0 GHz.

To certify the measurement uncertainty due to all random sources, two-dimensional evanescent scanning tests were conducted with the AWACS array. Near-field spacings of 3.81 cm ensured that the coupling product spectrum would extend to $(|k_x/k|, |k_y/k|) = 1.19$ at the selected test frequency, thus exposing the so-called evanescent spectrum whenever $[(k_x/k)^2 + (k_y/k)^2]^{0.5} \geq 1$. Because all evanescent antenna modes actually radiated by the test antenna are attenuated way below the dynamic range of the PNF instrumentation only a short distance from the antenna, we do not expect to intercept antenna evanescence during scanning with any probe at $Z = 35$ cm ($3.85\,\lambda$). Therefore, the magnitude of the coupling-product spectrum in this evanescent region is a direct measure of the far-field noise power.

These evanescent spectra generally show a random distribution of sidelobe peaks in the region beyond the visible space limits, at or below −80 dB for the open-end waveguide probe. For the Δ probe the spectrum is normalized to the peak of the open-end waveguide probe; the evanescent sidelobes have random peaks below −90 dB. However, in both spectra, the evanescent spectra also contain distinct sidelobe ridges at about −55 dB (Figs. 28 and 29), which cannot actually be radiated because, of course, the evanescent PNF are too highly attenuated to be measured at $Z = 35$ cm. Therefore, these evanescent sidelobe ridges must be the result of undiagnosed periodic scan-plane error—most probably due to unavoidable periodic multipath interactions. By excluding these ridges, we may compute the RMS over all remaining evanescent wavenumbers, and compare this spectral noise average to the peak coupling product mainbeam response. Table 3 contains these measured noise ratios for the AWACS test array. The table shows that the Δ probe AWACS signal-to-noise ratios are better by 12 dB when compared to scanning with a standard open-end waveguide probe.

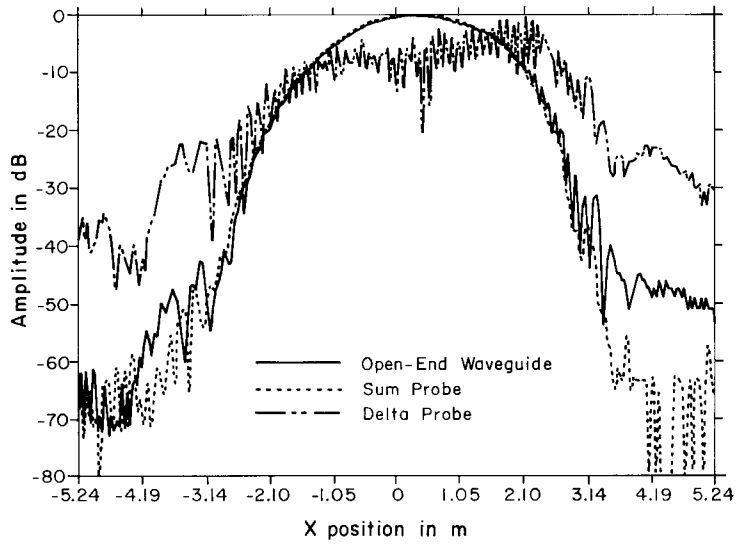


Fig. 25. Near-field centerline cut for the amplitude using the Δ probe (dashed line), the sum probe (dotted line), and waveguide probe (solid line).

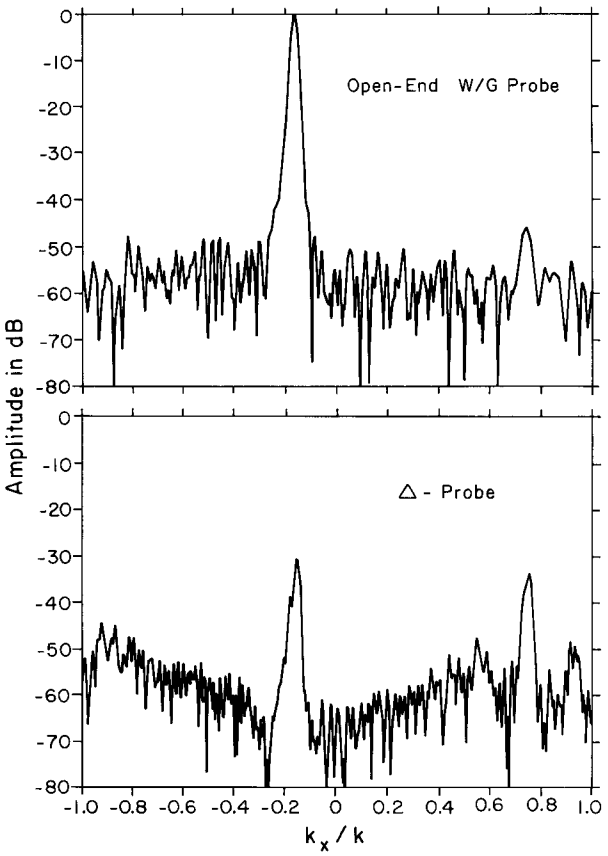


Fig. 26. Centerline coupling product amplitude using the waveguide probe (top) and using the Δ probe (bottom).

Table 3. Predicted and measured signal-to-noise ratios for AWACS

Probe type	Array	Signal/Noise ratio Measured[a] (dB)	Spectral ratio Measured[b] (dB)
Open-end W/G	AWACS	82.2	74.2
Delta	AWACS	94.0	89.7

[a] Excluding correlated evanescent lobe ridges.
[b] All evanescent space.

5. Measurement Accuracy

We used the mathematical analyses of Yaghjian [6] and Newell [7] to estimate the uncertainty for these measurements. These analyses allow us to estimate upperbound uncertainties for most of the possible uncertainty sources of Table 4 (from Table I in Newell [7]).

The sources of uncertainty listed in Table 4 fall into two broad categories. The first is uncertainties in the probe parameters arising from the measurements of the probe's gain, polarization, and pattern; second, uncertainties in the calculated spectra due to uncertainties in the measured near-field data and various reflection coefficients.

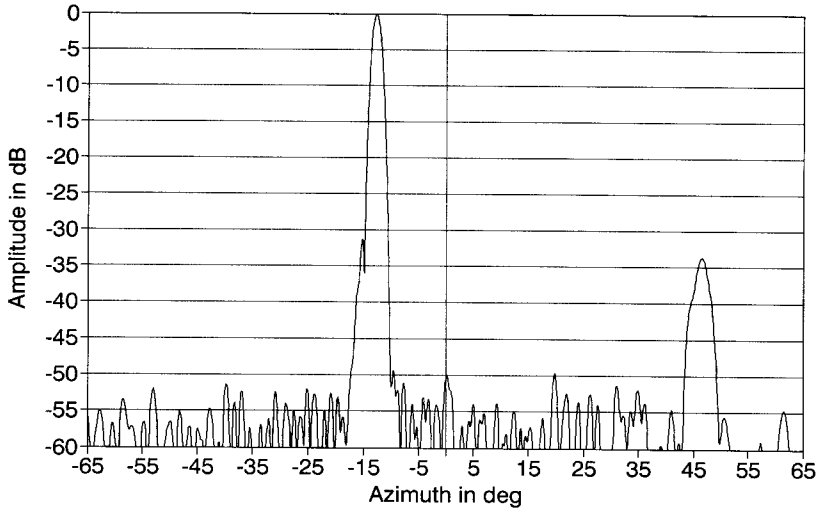


Fig. 27. Probe corrected centerline far-field pattern for the ULSA.

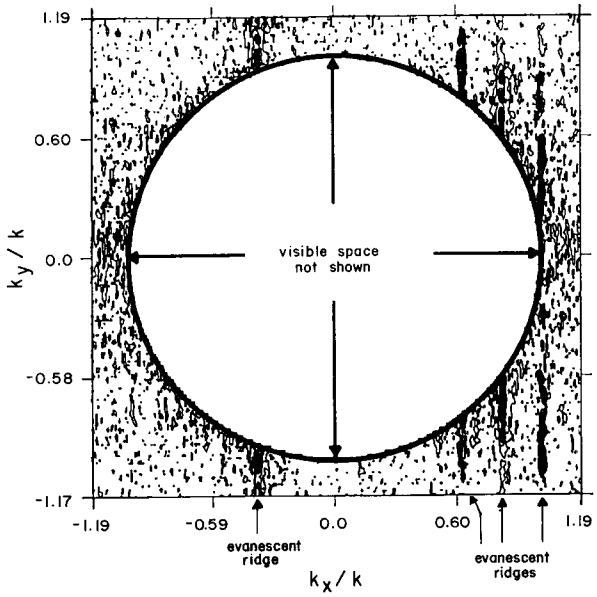


Fig. 28. Evanescent spectrum contour plot using the open-end waveguide. Note the ridges at $K_x/K = -0.37$, 0.64, 0.82, 1.02.

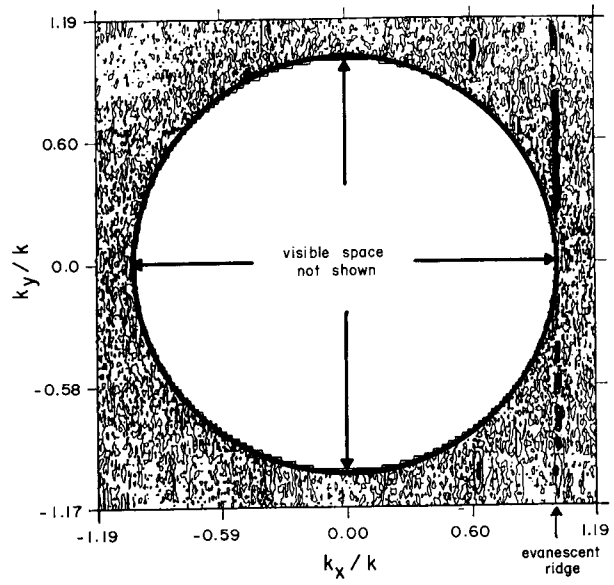


Fig. 29. Evanescent spectrum contour plot using the Δ probe. Note the ridge at $K_x/K = 1.02$.

Table 4. Uncertainty sources in planar near-field measurements

1.	Probe relative pattern
2.	Probe polarization ratio
3.	Probe gain measurement
4.	Probe alignment
5.	Normalization constant
6.	Impedance mismatch factor
7.	Antenna alignment
8.	Data point spacing (aliasing)
9.	Measurement area truncation

Table 4. Uncertainty sources in planar near-field measurements—continued

10.	Probe X-Y position uncertainties
11.	Probe Z-position uncertainties
12.	Multiple reflections (probe/AUT)
13.	Receiver amplitude nonlinearity
14.	System phase uncertainty Receiver phase nonlinearity Flexing cables and rotary joints Temperature effects
15.	Receiver dynamic region
16.	Room scattering
17.	Leakage and crosstalk
18.	Round-off in amplitude / phase

5.1 Probe-Parameter Uncertainties

Newell [7] has shown that the antenna transmitting coefficients, t_m (main component) and t_c (cross component), are given by

$$t_m(K) = \frac{\dfrac{D'(K)}{s'_m(K)} - \dfrac{D''(K)}{s''_c(K)}\,\rho'_s(K)}{1 - \dfrac{\rho'_s(K)}{\rho''_s(K)}} \tag{8a}$$

$$t_c(K) = \frac{\dfrac{D''(K)}{s''_c(K)} - \dfrac{D'(K)}{s'_m(K)\,\rho''_s(K)}}{1 - \dfrac{\rho'_s(K)}{\rho''_s(K)}} \tag{8b}$$

where D' is the uncorrected (for the probes) far field using probe 1 whose main component is in the same sense as the main component of the AUT, D″ is the uncorrected far field using probe 2 whose main component is in the same sense as the cross component of the AUT, $\boldsymbol{K}$ is the transverse part of the wave vector, and ρ' and ρ'' are the polarization ratios of probe 1 and probe 2 respectively. Hereafter we will drop the explicit use of $\boldsymbol{K}$. With these conventions, when

$$|\rho'_s/\rho''_s| \leq 1 \tag{9a}$$

and when

$$|\rho'_s/p_t| \leq 1, \tag{9b}$$

where $p_t = t_c/t_m$, the probe correction equations become [7]

$$t_m = \frac{D'}{s'_m} \tag{10a}$$

$$t_c = \frac{D''}{s''_c} - \frac{D'}{s'_m \rho''_s} \tag{10b}$$

When the conditions of Eq. (10a) and (10b) apply, we can calculate the differential of these equations to obtain equations for the fractional uncertainties. We find that the fractional uncertainties for t_m and t_c are [7]

$$\frac{dt_m}{t_m} = \frac{dD'}{D'} - \frac{ds'_m}{s'_m}, \tag{11a}$$

$$\frac{dt_c}{t_c} = \left(1+\frac{1}{p_t \rho''_s}\right)\left(\frac{dD''}{D''} - \frac{ds''_c}{s''_c}\right)$$

$$+ \left(\frac{1}{p_t \rho''_s}\right)\left(\frac{d\rho''_s}{\rho''_s} - \frac{dD}{D'} + \frac{ds'_m}{s'_m}\right). \tag{11b}$$

The uncertainties in the far field are caused by uncertainties in the values of D', D", s'_m, s''_c, and ρ''_s. When Eq. (9a) and Eq. (9b) are valid, the polarization ratio of probe 1 ρ'_s, has no significant effect on either component of the far field. In addition, the probe's effect on the AUT cross component depends on the relative polarization ratios of the AUT and probe 2. Since the main-component uncertainties are proportional to the uncertainties in s'_m, the uncertainties in the probe gain and pattern have a one-to-one correspondence to uncertainties in the AUT's main component. Typical uncertainties for the probes are given in Table 5.

Table 5. Typical probe uncertainties

Source of uncertainty	Standard uncertainty (in dB)
Probe gain	0.06
Relative probe pattern amplitude	
at −5 dB	0.09
at −15 dB	0.2
at −30 dB	0.9

If the sum probe or open-end waveguide is used to measure the AUT, we expect a standard uncertainty due to the probe of about 0.06 dB in the gain of the AUT. The difference probe, on the other hand, has a 30 dB null in the direction of the AUT main beam; therefore the standard uncertainty in its gain in the direction of the null is about 0.9 dB. In addition, Δ probe steering uncertainties introduce additional gain uncertainties. Because of the sharpness of the null, the steering uncertainty of 0.5° causes a standard gain uncertainty of 1.6 dB. Thus the standard gain uncertainty using the Δ probe is greater than 1.8 dB.

These probes all have their patterns down by about 20 dB from their peak at wide angles, hence the standard uncertainty in the amplitude of the AUT's pattern at wide angles is about 0.3 dB due to uncertainties in the relative patterns of the probes.

5.2 Near-field Measurement Uncertainties

In addition to the uncertainties in the probe parameters, there are uncertainties in the calculated far field due to uncertainties in the near-field measurements.

A number of these uncertainties cannot be estimated beforehand. They are: (1) the multiple reflection effects because there exists no theoretical method to determine the magnitude of the multiple reflections between probe and AUT; (2) the impedance uncertainties (which affect only the gain) because the reflection coefficients and their uncertainties are unknown until measured; (3) uncertainties due to leakage and crosstalk until they are actually measured; and (4) uncertainties due to room scattering. Room scattering has not played an important role at the NIST facility in the past because its magnitude is small but it could be a limiting factor in measuring low sidelobes. For the low-sidelobe measurement, NIST developed a test to estimate the effect of room scattering as was described earlier. The implications of the results of this test for the measurement accuracy will be discussed later.

The effect of multiple reflection can be estimated by taking measurements at several Z-distances, averaging the results, and subtracting the average from the individual measurements. The results of these tests were discussed in a previous section. In short, the multiple reflection uncertainties were greater for the difference probe than for the other two probes that were used in the measurement and were a function of far-field angle.

Impedance uncertainties were small and contributed at most 0.05 dB to the uncertainty in the gain of the AUT.

Leakage and crosstalk were estimated by making one-dimensional near-field measurements first with a termination on the transmitting side with the probe operating normally and then with a termination on the receiving side with the AUT operating normally. The leakage and crosstalk were 75 dB below the near-field peak and cause uncertainties in the far-field pattern as summarized in Table 6.

As indicated in Sec. 4, the peak room scattering is −70 dB, and the RMS room scattering is −89 dB. From this, we obtain the uncertainties in Table 7.

Table 6. Far-field pattern uncertainties due to leakage

Pattern amplitude (in dB)	Standard uncertainty (in dB)
0.	0.0004
−15.	0.003
−30.	0.02
−45.	0.1
−60.	0.2
−75.	2.4

Table 7. Uncertainties due to room scattering

Pattern amplitude (in dB)	Standard uncertainty (in dB)
0.	0.0003
−15.	0.002
−30.	0.01
−45.	0.06
−60.	0.3
−75.	1.9

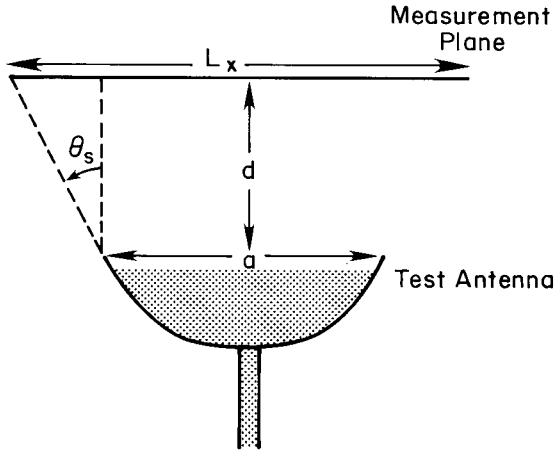
In principle, data point spacing can be chosen so that aliasing is arbitrarily small. However, noise and/or rapidly varying systematic errors (for example, multiple reflections) set the practical lower limit. If the data spacings in X and Y are δ_x and δ_y respectively, then the aliased Fourier transform of the data $F_e(K)$ in terms of the true FFT is [7]

$$F_e(K) = \sum_{m,n=-\infty}^{\infty} F\left(k_x + \frac{2m\pi}{\delta_x}, k_y + \frac{2n\pi}{\delta_y}\right). \quad (12)$$

Aliasing is contributed by terms for which $m \neq 0$, $n \neq 0$. The terms $m = \pm 1$, $n = \pm 1$ are usually the only terms which contribute significantly to this error. The magnitude of the aliasing uncertainty can be estimated using the test space procedure. In this test a centerline near-field scan is performed using very small spacing between data points. The FFT is then performed on the complete set of data, then on every other data point, then every third point, and so forth. Using these centerline tests we are able to determine a data point spacing such that the aliasing error is small in comparison to other errors for all three probes. The aliasing uncertainty is about the same for all of the probes.

Area truncation has two effects. First, the far-field pattern results obtained by Fourier transforming the PNF data are valid only within the angular region defined by the geometry of the antenna and the scan area, as shown in Fig. 30. The second effect produces uncertainties in the far-field pattern within the region of validity. Yaghjian [6] showed that we can obtain an upperbound uncertainty from a knowledge of the measured data on the boundary of the scan area. Denoting the plane-polar coordinates of the boundary by (p', ϕ_p), the normalized data on the boundary by $B(p', \phi_p)$, the spherical coordinates in the far field by (θ, ϕ), the magnitude of the electric field by $|E(r)|_{r\to\infty}$, and the maximum acute angle between the plane of the scan area and a line connecting the edges of the antenna aperture and the scan area by γ_{max}, we express the fractional uncertainty in the far field as

$$\left|\frac{\Delta D(\theta,\phi)}{D(\theta,\phi)}\right| \leq \frac{\left|\int_0^{2\pi} B(p', \phi_p) e^{-ikp'\sin\theta\cos(\phi-\phi)} p' \, d\phi_p\right|}{2\pi r \, |E(r)|_{r\to\infty} \cos\gamma_{max}}. \quad (13)$$



Maximum Angle for Accurate Far Field = θ_s

$$\theta_s \simeq \text{Tan}^{-1}\left(\frac{L_x - a}{2d}\right)$$

Fig. 30. Schematic relationship between the scan length and the maximum angle to which far-field patterns can be accurately computed.

In this test an FFT is done on the full set of data, then all the data are set to zero except the data at the edges and the FFT recomputed. The far fields from these two data sets can be compared and used to determine the uncertainty due to the truncation

of the scan plane. We find that the standard truncation uncertainty for the peak of the far field is 0.06 dB. At 40 dB below the far-field peak the standard truncation uncertainty is about 0.3 dB and at 60 dB below the far-field peak the standard uncertainty is 0.9 dB.

The remaining uncertainties, which are due to uncertainties in position, amplitude, phase, and alignment, were shown by Yaghjian [6] and Newell [7] to have a term in common: they all depend on the ratio $g(K)$ of the peak uncorrected far-field amplitude $D(K_0)$ to the far-field amplitude $D(K)$ in the direction K. Because the difference probe has a null in the direction of the AUT's main beam it leads to an uncorrected far field which has a peak value which is 25 dB to 30 dB below the peak of both the sum probe and the open-end waveguide. As a result these uncertainties will be 25 dB to 30 dB below the corresponding uncertainties for the other two probes.

Position uncertainties (and some other uncertainties) often concentrate their effects in certain directions because of periodicities in the measurement (for example, the structural supports of the scanner). If there is no periodicity then these uncertainties will add to the noise of the far field. Otherwise, the position uncertainties are concentrated in the directions sin (γ/τ_i) relative to the direction of the main beam, where τ_i is the corresponding periodicity. For antennas in which the main beam is steered off axis and for direction angles near the main beam, $(\theta - \theta) < \gamma/(10L)$ (L is the maximum antenna dimension and θ_b is the angle between the direction of the Z-axis and the direction of the main beam) the uncertainty in the far field is [7]:

$$\left|\frac{\Delta D(K)}{D(K)}\right|_{dB} \le \frac{344}{\sqrt{\eta}} \left(\frac{\Delta(K)}{\lambda}\right)^2 \sin^2\theta_b \, g(K) \qquad (14)$$

for X and Y uncertainties, and

$$\left|\frac{\Delta D(K)}{D(K)}\right|_{dB} \le \frac{43}{\sqrt{\eta}} \left(\frac{\delta_z(K)}{\lambda}\right)^2 \cos^2\theta_b \, g(K) \qquad (15)$$

for the Z uncertainties. For directions where $\lambda/(10L) < (\theta - \theta_b) < \pi/2$ the uncertainty is

$$\left|\frac{\Delta D(K)}{D(K)}\right|_{dB} \le 13.5 \left(\frac{\Delta(K)}{\lambda}\right) \sin\theta_b \, g(K) \qquad (16)$$

for the X and Y uncertainties, and

$$\left|\frac{\Delta D(K)}{D(K)}\right|_{dB} \le 13.5 \left(\frac{\delta_z(K)}{\lambda}\right) \cos\theta_b \, g(K) \qquad (17)$$

for the Z uncertainties. η is the aperture efficiency, $\Delta(K)$ is the FFT of the X and Y-position errors, and $\delta_z(K)$ is the FFT of the Z uncertainties. Table 8 shows the period and magnitude of the position errors that are observed for the NIST near-field scanner.

The period of 9.1 m in Table 8 corresponds approximately to twice the scanner's height, and 40.5 cm corresponds to the distance between supports in the X-direction. For L = 6 m, the largest dimension of the ULSA, $\eta = 0.5$, $\theta_b = 12.7°$ (the location of the ULSA main beam), we obtain the uncertainties found in Table 9 for the sum and open-end waveguide probes.

For the difference probe, a −40 dB sidelobe would have $g(K) = 3.2$ instead of 100 as is the case for the other two probes and a −60 dB sidelobe corresponds to $g(K) = 32$ instead of 1000. This is a consequence of the 30 dB null which the difference probe has in the direction of the AUT's main beam and which reduces the peak value of $D(K_0)$ by 30 dB. The resulting uncertainties for the difference probe are found in Table 10.

Table 8. Mechanical periods for the NIST scanner

Type of uncertainty	As a function of	Periods	Magnitude (in cm)
Z	X	none	0.04
	Y	9.1 m	0.06
Y	X	40.5 cm	0.01
		91.0 cm	0.01
X	Y	9.1 m	0.06

Table 9. Far-field uncertainties resulting from near-field position uncertainties using sum and open-end waveguide probes

Type	Function variable	Direction angle from main beam	Sidelobe level (in dB)	Standard uncertainty (in dB)
Z	X			noise
	Y	±0.6	0.0	0.05
Y	X	±6.0	−40.0	0.2
		±14.0	−60.0	1.7
X	Y	±0.6	0.0	0.01

Table 10. Far-field uncertainties resulting from near-field position uncertainties using the difference probe

Type	Function variable	Direction angle from main beam	Sidelobe level (in dB)	Standard uncertainty (in dB)
Z	X			noise
	Y	±0.6	0.0	0.08
Y	X	±6.0	−40.0	0.01
		±14.0	−60.0	0.09
X	Y	±0.6	0.0	0.02

Because of the difference probe's properties the position uncertainties for a −40 dB sidelobe and a −60 dB sidelobe are greatly reduced in comparison to both the sum probe and the open-end waveguide probe.

Amplitude and phase instrumentation uncertainties arise from receiver nonlinearity, flexing of cables and rotary joints, source and receiver drift, and temperature drift. The drift amounts to about 0.5°/h. However, we can partially correct for drift by the use of tie scans. There is also a 2° to 3° phase uncertainty associated with the flexing of cables to the probe and with the rotary joint. The periods for the cable flexing are associated with the X and Y dimensions of the scanner, about 4.5 m in each direction. We can partially correct for receiver nonlinearity by using a calibration of the receiver against a calibrated rotary vane attenuator. Using this calibration we can make a first-order correction for the nonlinearity of the receiver. The first-order correction coefficient for the frequency band of the ULSA is 0.01 ± 0.005. Thus 0.005 is the peak residual uncorrected nonlinearity. The periods associated with this uncertainty depend on the variation of the near-field amplitude and phase as functions of X and Y. The near-field amplitude variation depends partly on the properties of the probe. The expected periods for the sum and open-end waveguide probes are 8 m in X and 2 m in Y. The expected periods for the difference probe are 8 m in X and 1 m in Y. The upperbound amplitude uncertainty due to nonlinearity is [7]

$$\left|\frac{\Delta D(K)}{D(K)}\right|_{\mathrm{dB}} \leq 6U(K)g\,(K), \tag{18}$$

where $U(K)$ is the FFT of the residual receiver nonlinearity uncertainty. When $\lambda/(10L) \leq (\theta - \theta_b) \leq -\pi/2$ the uncertainty in the phase nonlinearity causes a far-field uncertainty which is [7]

$$\left|\frac{\Delta D(K)}{D(K)}\right|_{\mathrm{dB}} \leq 13.5\,\frac{V(K)}{360}\,g(K), \tag{19}$$

where $V(K)$ is the FFT of the residual phase uncertainties (including the uncertainty due to cable flexing). The resulting far-field uncertainties due to residual uncorrected receiver nonlinearity in the phase and amplitude are shown in Table 11. The uncertainties of Table 11 for the −20 dB and −40 dB sidelobes are less for the Δ probe but not much less because the largest uncertainties lie in the direction of this probe's null.

Table 11. Far-field uncertainties due to residual amplitude and phase uncertainties

Type	As a function of	Direction from main beam	Sidelobe level (in dB)	Standard uncertainty (in dB)
Amplitude	X	±0.7	0.	0.03
	Y	±1.3	−20.	0.3
		±3.0	−40.	3.0
Phase	X	±1.3	−20.	1.0
	Y	±1.3	−20.	1.0

In addition to the above mentioned phase and amplitude uncertainties, there is also round-off which will contribute to the noise. The near-field amplitude is measured on a 0 to 100 linear scale; the maximum round-off for the amplitude is 0.05 on this scale. The phase is measured in degrees and has a maximum round-off of 0.05°. The signal-to-noise ratio due to the amplitude round-off is given by Newell [7] as

$$\left(\frac{S}{N}\right)_a = \frac{\sqrt{N_e/2}}{(N/N_e)\,3\,\sigma_a}, \tag{20}$$

and due to the phase round-off as

$$\left(\frac{S}{N}\right)_\psi = \frac{\sqrt{N_e/2}}{3\,\sigma_\psi}, \tag{21}$$

where N_e is the number of measurement points within the effective area, N is the total number of measurement points, and the σ's are the standard deviation of the round-off distributions. Since we used a spacing of 3.81 cm in both X and Y and a

measurement area of about 10.4 m in x by 3.8 m in Y, then $N \approx 27\,500$ measurement points. Assuming a 50% efficiency for the antenna and using the dimensions of the ULSA we find $N_e \approx 2500$. With maximum round-off of 0.05 for the amplitude and 0.05° for the phase then $\sigma_a = 0.0003$ relative to the peak and $\sigma_\psi = 0.0005$ (in radians) and we find $(S/N)_a = 82$ dB and $(S/N)_\psi = 87$ dB. The total signal-to-noise ratio is therefore expected to be about 81 dB.

A summary of the approximate uncertainties as derived from the above tests and analyses is given in Table 12. Assuming these uncertainties are independent of each other the total standard uncertainty is the root sum square (RSS) of these errors. This implies that the total standard uncertainty at −55 dB for the waveguide probe is (+3.2 dB, −5.1 dB) and that for the difference probe is (+2.2 dB, −3.0 dB). Uncertainties will exceed this in some directions where periodicities are larger.

Our analysis for the NIST near-field facility leads us to conclude that it should be possible to accurately measure (within a few decibels) sidelobes down to about 55 dB below the AUT's peak far field.

5.3 Comparison of Near-Field and Far-Field Results

We can now compare the far-field patterns obtained from the NIST PNF range to those obtained by others on a far-field range. When we overlay the results of Fig. 27 to those in Fig. 3 we found that there is an offset between the two patterns of 0.5°. Accounting for this offset and overplotting the two results (Fig. 31) we see the two patterns agree quite well. At −55 dB the difference between the two patterns is generally less than 5 dB, which is within the NIST measurement uncertainty.

Table 12. Far-field uncertainties using the NIST planar scanner

Uncertainty source	Standard uncertainties in dB for a −55 dB sidelobe Waveguide probe	Difference probe
Truncation	+1.4 −1.6	±0.5
Aliasing	±0.5	±0.5
Multiple reflections	+1.4 −1.6	+2.2 −2.9
Room scattering	±0.2	±0.1
Position	+2.2 −2.9	±0.1
Amplitude	+1.4 −1.6	±0.1
Phase uncertainty	+1.4 −1.6	±0.1
Random noise	±0.4	±0.3
RSS uncertainty	+3.2 −5.1	+2.2 −3.0

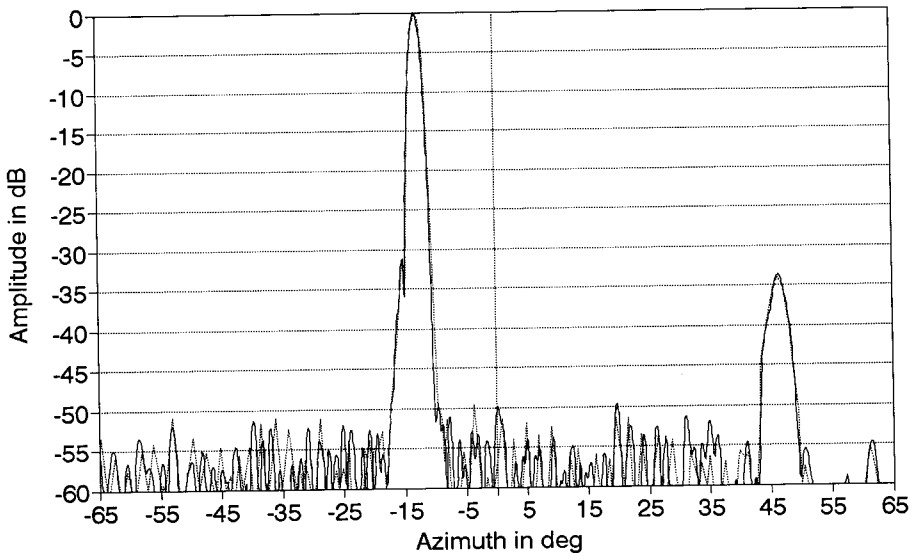


Fig. 31. Comparison of far-field patterns determined using the PNF range (solid) and the far-field range (dotted line).

6. Future Research

If we could ignore the multiple reflections caused by the difference probe, this probe would clearly be the best probe to use for these low-sidelobe PNF measurements (but not for the gain or main beam). However, the effects of multiple reflections are substantially worse for the Δ probe than for the other two probes so the Δ probe proves to be only slightly better (about 2 dB) in the level to which it is able to distinguish the low sidelobes of the AUT.

Why a probe or any antenna has particular reflection properties is poorly understood. There is only a rough approximation for correcting for this effect. The argument is usually made that if data are taken over enough scan planes and transformed to the far field while accounting for the different separation distances, and data from the various planes averaged, then the multiple reflection effects tend to cancel. There is no test for determining how many different scan planes are adequate for this procedure. In calculating the far field from near-field data, the terms involving multiple reflections between the AUT and the probe are ignored.

It would be desirable to approach the problem of multiple reflections from both the experimental and theoretical views. From the theoretical view the possibility of including the first-order multiple reflection term could be explored. According to Kerns [12] the full solution to the transmission equation is

$$\frac{b_0'}{a_0} = \hat{S}_{02}'\,\hat{T}_{21}\,(1-\hat{S}_{11}\,\hat{R}')\,\hat{S}_{10}, \qquad (22)$$

where $\hat{R}'$ describes the receiving probe as a passive scatterer in the transmitting coordinating system, and $\hat{S}_{11}$ describes the scattering properties of the transmitting antenna. In practice for near-field measurements we assume $\hat{S}_{11}\hat{R}' \sim 0$. Since multiple reflections are obviously present, often it would be desirable to know what can be done to solve equation (6-1) when $\hat{S}_{11}\,\hat{R}' \neq 0$.

From an experimental view point it would be desirable to know which kinds of probes and antennas produce larger multiple reflections. This might be determined by doing Z-multiple- reflection tests (as described in Sec. 3.3) for different combinations of AUT and probe.

7. Summary

The near-field measurement technique can be used to measure sidelobes of very low-sidelobe antenna arrays. Near 3 GHz the NIST measurement facility can measure sidelobes to about 55 dB to 60 dB below the AUT's peak far field. The main limitations to accurately determining sidelobes below 60 dB are multiple reflection effects between the AUT and the probe.

The Δ probe can be useful in reducing some uncertainties, which depend on the quantity $g(K)$. However, because higher multiple-reflection effects are associated with it, it is only marginally better than an open-end waveguide in measuring low sidelobes. If multiple reflections could be substantially reduced, the difference probe would be an extremely useful probe to measure sidelobes below 65 dB.

The Δ probe cannot be used to accurately determine the main beam region of the AUT far-field pattern. This is because in the region of the Δ probe null, which corresponds to the AUT main beam, small uncertainties in steering cause large uncertainties (1.5 dB or more) in the probe pattern amplitude which in turn causes large uncertainties in the probe correction for the AUT main beam.

Acknowledgments

We thank Douglas Kremer and Nino Canales for performing the measurements and many of the modifications to the NIST near-field facility which made this measurement possible. We also thank Katie MacReynolds for doing much of the data processing. The authors also acknowledge the support of the Department of Defense for providing the funding that made this research possible. Last, but not least, we wish to thank Westinghouse Electric Corporation for doing the far-field measurements on their far-field range near Baltimore, Maryland.

8. References

[1] D. M. Kerns, Correction of near-field antenna measurements made with an arbitrary but known measuring antenna, Electron. Lett. **6**, 346–347 (1970).

[2] D. M. Kerns, New method of gain measurement using two identical antennas, Electron. Lett. **6**, 347–349 (1970).

[3] R. C. Baird, A. C. Newell, P. F. Wacker, and D. M. Kerns, Recent experimental results in near-field antenna measurements, Electron. Lett. **6**, 349–351 (1970).

[4] E. B. Joy and D. T. Paris, Spatial sampling and filtering in near-field measurements, IEEE Trans. Antennas Propagat. **AP-20**, 253–261 (1972).

[5] A. C. Newell and M. L. Crawford, Planar near-field measurements on high performance array antennas, Natl. Bur. Stand. (U.S.) NBSIR 74-380, July 1974.

[6] A. D. Yaghjian, Upper-bound errors in far-field antenna parameters determined from planar near-field measurements, Part 1:analysis, Natl. Bur. Stand. (U.S.) Tech. Note TN 667, 1975.

[7] A. C. Newell, Error analysis techniques for planar near-field measurements, IEEE Trans. Antennas Propagat. **AP-36**, 754–768 (1988).
[8] H. Y. Yee, Slot-antenna arrays, Chapter 9, Antenna Engineering Handbook, McGraw-Hill Book Company (1984).
[9] S. Silver, in Microwave transmission lines, in Microwave Antenna Theory and Design, McGraw-Hill Book Company (1949).
[10] G. K. Huddleston, Optimum probes for near field antenna measurements on a plane, Doctoral Dissertation, Georgia Institute of Technology, University Microfilms International (#7823710) Ann Arbor, Michigan, August 1978.
[11] E. B. Joy, Spatial sampling and filtering in near field measurements, Doctoral Thesis, Georgia Institute of Technology, Georgia, November 1970.
[12] D. M. Kerns, Plane-wave scattering-matrix theory of antenna and antenna-antenna interactions. Natl. Bur. Stand. (U.S.) Monograph 162, June 1981.

About the authors: *Michael H. Francis is a physicist in the Electromagnetic Fields Division in the NIST Electronics and Electrical Engineering Laboratory. Allen C. Newell is the Division Chief of the Electromagnetic Fields Division. Kenneth R. Grimm is the Principal Engineer for RF Systems Research and Development with Nichols Research Corporation. John Hoffman is a Senior Scientist with System Engineering Group. Both Grimm and Hoffman were with Technology Service Corporation at the time of this work. Helmut E. Schrank is a retired engineer from Westinghouse Electric Corporation. He is a former President of the Antennas and Propagation Society of the IEEE. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 2, March–April 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 169 (1994)]

On the Physics Required for Prediction of Long Term Performance of Polymers and Their Composites

Volume 99 | Number 2 | March–April 1994

Gregory B. McKenna

National Institute of Standards and Technology,
Gaithersburg, MD 20899-0001

The long term performance of polymers and their composites is an important aspect of their increasing use in engineering applications. Temporal, thermal, and mechanical stresses can all contribute to the deterioration of performance. Here we examine the concepts of the physics of glassy polymers and how they are important in developing constitutive equations that describe their volume/temperature/stress time response. The understanding of such response forms the basis of the prediction of long term performance.

Key words: composites; constitutive equations; glasses; material clocks; nonlinear viscoelasticity performance; solid polymers.

Accepted: November 17, 1993

1. Introduction

When a glass forming liquid, such as a polymer melt or rubber, is cooled there is some temperature at which the molecular mobility decreases enough that the material's thermodynamic state or structure cannot attain equilibrium in the time scale of the experiment (determined by the cooling rate) [1–3]. Below this temperature, generally referred to as the glass transition temperature T_g, glassy materials are unstable with the result that their properties continuously evolve towards a temporally distant equilibrium. Efficient design with and use of polymeric materials requires an understanding of the physics underlying the structural recovery and the impact of the changing thermodynamic state on the mechanical properties of the material. In what follows we attempt to put into perspective the interrelationships between the physics of glasses, how it influences material performance and how we model material response using nonlinear constitutive equations. Additionally, we will describe some of the implications of structural recovery in polymers for composite materials performance. Finally, we will discuss the importance of developing methods to combine computer codes with nonlinear constitutive law modeling to material processing and performance prediction.

2. Phenomenology of Glasses

Referring to Fig. 1, the *glass transition event* can be defined in terms of a volume-temperature surface in a cooling experiment. At a given cooling rate, $q = -\mathrm{d}T/\mathrm{d}t$, the volume begins to depart from the equilibrium volume at a characteristic temperature $T = T_g$ referred to as the *glass transition temperature*. If the magnitude of the cooling rate decreases, T_g decreases $\{T_g(q_2) < T_g(q_1)$ for $q_2 < q_1\}$.

If one cools a glass as depicted in Fig. 1 and then keeps the temperature fixed the volume of the material evolves spontaneously towards equilibrium, as represented by the arrow at T_1. When the temperature is decreased rapidly in a *T-jump* experiment the *volume recovery* response looks like that depicted in Fig. 2 where the *volume departure*

from equilibrium $\delta_v = (v - v_\infty)/v_\infty$ is plotted against the logarithm of the time after the temperature change. Note that v is the (specific) volume at time t and v_∞ is the equilibrium volume. Similar results are obtained if one measures enthalpy instead of volume [4]. In general, one can refer to the evolution of the glassy polymer's thermodynamic state towards equilibrium as *structural recovery* [4-6].

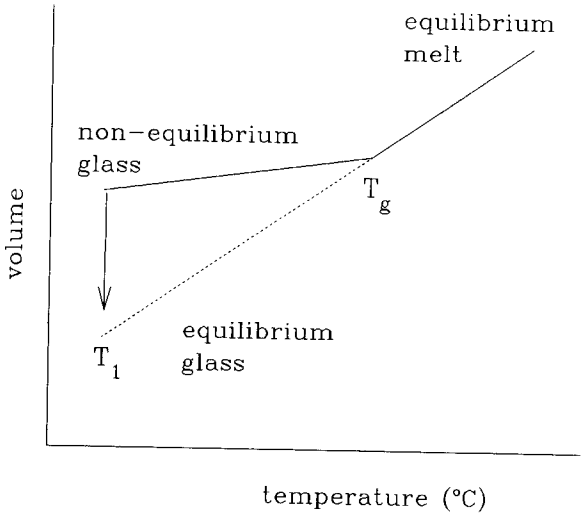


Fig. 1. Schematic representation of the volume-temperature behavior of a glass forming material.

Asymmetry of approach [1, 3–6] experiments demonstrate the inherent nonlinearity of the structural recovery process. In these experiments, referring to Fig. 3, one equilibrates the glass prior to performing T-jump experiments by either increasing or decreasing the temperature. When one jumps from $T_0 = T - \Delta T$ (up-jump) or $T_0 = T + \Delta T$ (down-jump) to T, one finds that the response from below T is significantly different from that when the jump is from above T. This is readily seen in Fig. 4 in which up- and down-jump histories are depicted for a polymer glass. The asymmetry of approach to equilibrium is obvious. The nonlinear behavior was observed early by Tool [6] in working with inorganic glasses and has been attributed to a dependence of mobility on the current structure of the glass. Thus, in down-jump experiments, where the rate of volume recovery decreases with increasing time after the quench, the response is *autoretarded*, while in the up-jump the rate of volume recovery increases with increasing time and the response is *autocatalytic* [1]. As described subsequently, the dependence of the molecular mobility on the current glassy structure is one of the "essential ingredients" required to model the phenomenology of glassy behavior [3].

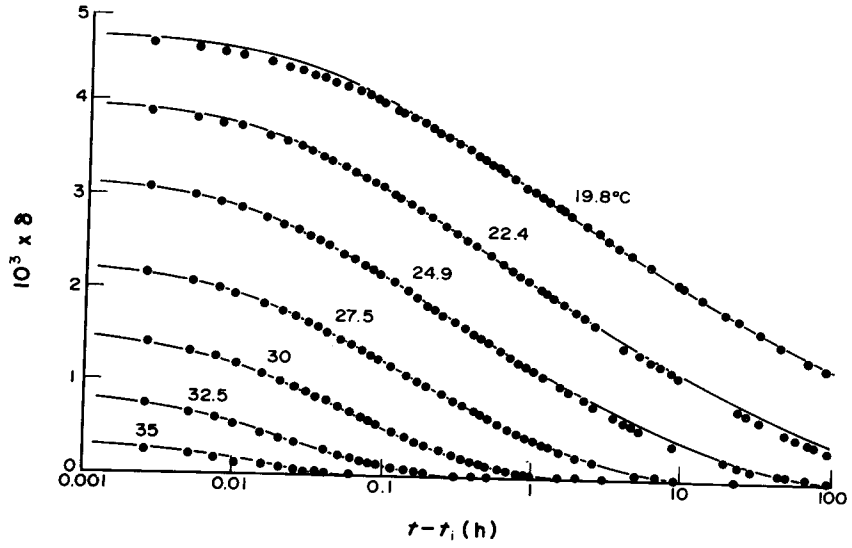


Fig. 2. Isothermal contraction of glucose glass after quenching from $T_0 = 40$ °C to different temperatures, as indicated (after Ref. [1], with permission).

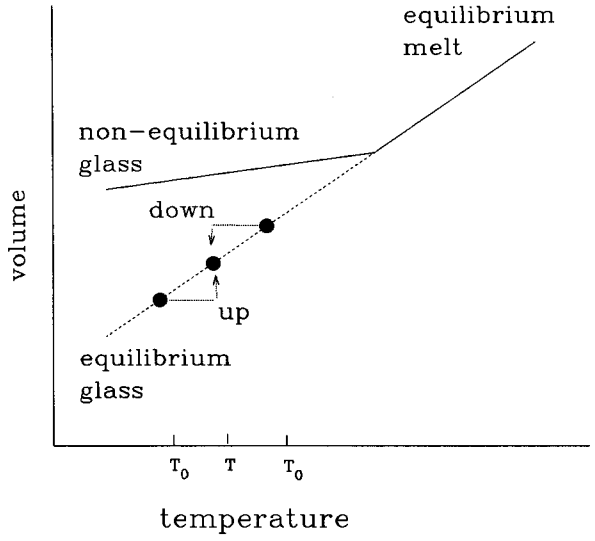


Fig. 3. Volume-temperature schematic of the asymmetry of approach experiment. (See text for discussion.)

The so-called *memory effect* is a manifestation of another essential ingredient required to describe glassy behavior. In such experiments one applies multiple temperature jumps to the sample. If we refer to Figs. 5 and 6 the most dramatic effect is seen when the experiment is performed such that the sample is first subjected to a down-jump to T_1 (A, B, C in Fig. 5) and allowed to "age" isothermally until the volume extrapolated along the glassy thermal expansion line is the same as the equilibrium volume at T_0 (D in Fig. 5). The sample is then subjected to an up-jump to T_0 and the resulting initial departure from equilibrium δ_v is near to zero. As seen in Fig. 6, the glass does not remain in equilibrium, rather, it "remembers" its prior history and δ_v goes through a maximum before it approaches the response of a down-jump directly to T_0 at long times. Such behavior cannot be explained without invoking a *nonexponential retardation function* to describe the behavior. The nonexponential function can either be represented as a *sum of exponentials* (distribution of retardation times) or using a Kolrausch [7]-Williams-Watts [8] (KWW) function (stretched exponential). The usefulness of these functions is described subsequently.

In the final paragraph of this section we discuss briefly the problems of performing *nonisothermal* types of experiments. The experiments described above, while using two or more temperatures were *isothermal* in the sense that the response of interest was observed at constant temperature. In those experiments we observed both the material nonlinearity (asymmetry of approach) and a manifestation of the nonexponentiality of the fundamental retardation response (memory effect). In nonisothermal experiments, one might expect that manifestations of these effects could lead to apparently interesting behaviors that truly were simple

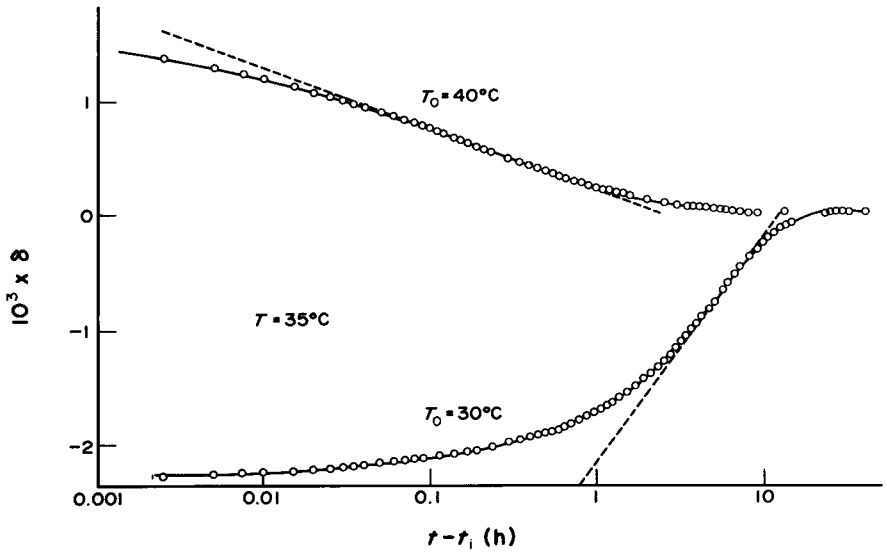


Fig. 4. Expansion and contraction isotherms for poly(vinyl acetate) glass after heating and cooling from $T_0 = T \pm 5$ °C. This plot shows the asymmetry of the expansion and contraction isotherms (after Ref. [1], with permission).

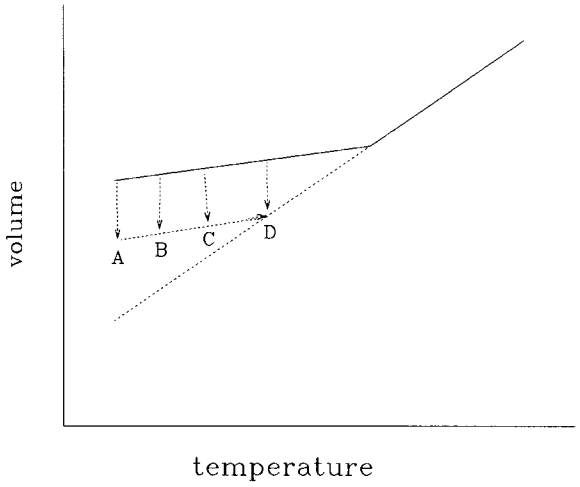


Fig. 5. Schematic of volume-temperature history used in memory experiment (see text for discussion).

evidences of the nonlinearity and nonexponentiality of behaviors. Thus, in a differential scanning calorimetry experiment, for example, one follows the heat capacity C_p as temperature is changed. Often one observes peaks, as shown in Fig. 7. The question then arises as to how to interpret such peaks and the possible effects of the nonlinear behavior of glasses need to be included in any interpretation of the results. Similar comments hold for other experimental methods such as thermomechanical analysis (TMA) or dynamical mechanical analysis (DMA) in which temperature is varied. In the following section we discuss the equations that are valuable in describing the nonlinear, nonexponential structural recovery of glass forming systems.

3. The Physics of Structural Recovery

Prior to being able to describe the viscoelastic response of polymeric glasses, it is essential that the underlying physics of the polymer glass be understood. In this section we present the phenomenological equations that have been developed and are generally agreed upon as describing reasonably well the structural recovery of glasses. It is anticipated that the physics incorporated into these equations are the minimum needed to successfully describe glassy behavior, at least to a first approximation. The ideas presented here will carry over to the viscoelastic response of polymers and the phenomenon of physical aging. These will be visited in a subsequent section.

3.1 Requirements of the Models

From the above discussion it is readily seen that the kinetics of structural (volume, enthalpy) recovery in glassy materials exhibits a richness of behaviors that offers a severe test of any model. In 1971 Narayanaswamy [9] developed a formalism

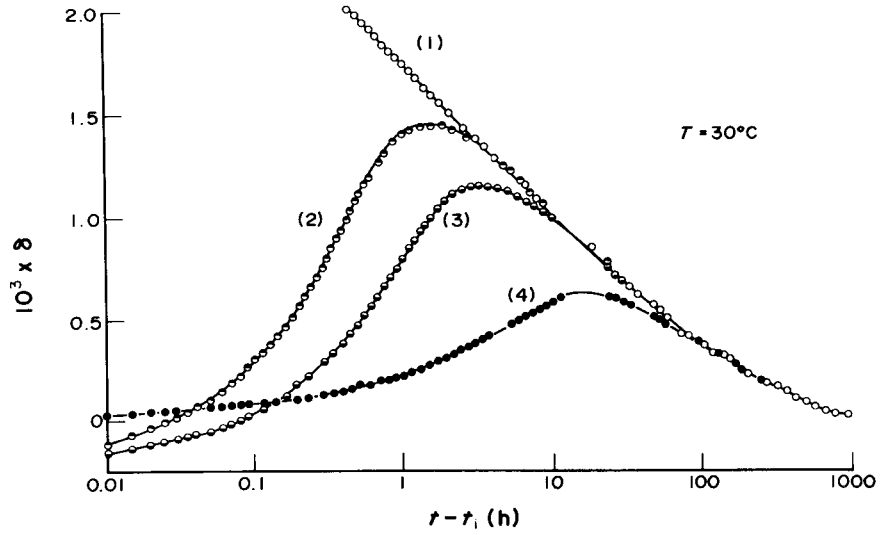


Fig. 6. Isothermal evolution at $T_0 = 30$ °C for poly(vinyl acetate) showing memory effect: (1) quench from 40 °C to 30 °C; (2) quench from 40 °C to 10 °C for 160 h followed by up-jump to 30 °C; (3) quench from 40 °C to 15 °C for 140 h followed by up-jump to 30 °C; (4) quench from 40 °C to 25 °C for 90 h followed by up-jump to 30 °C. Note that at short times $\delta \approx 0$. (After Ref. 1, with permission.)

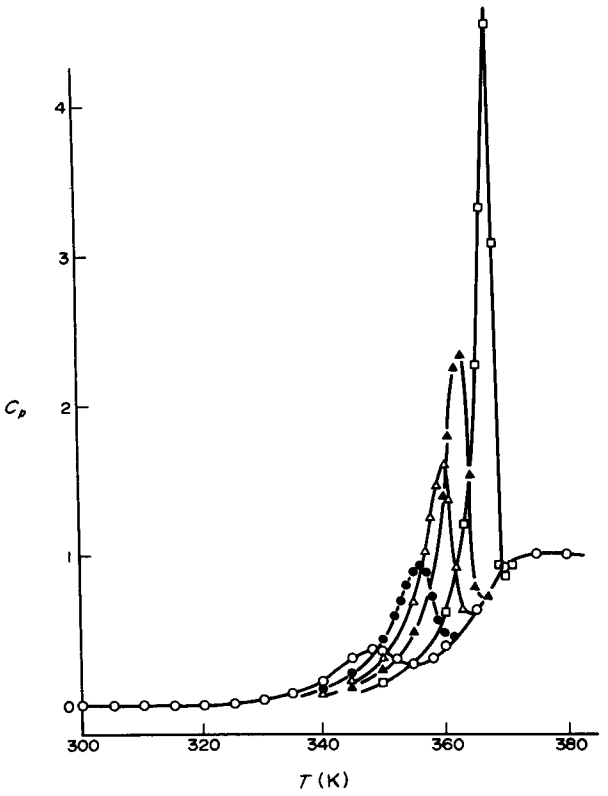


Fig. 7. DSC curves calculated using a Narayanaswamy [9]-Moynihan [10]-type phenomenological model for glasses subjected to difference annealing times t_e at 60 °C. (○) 1 h; (●) 1 day; (Δ) 1 week; (▲) 1 month; (□) 1 year. This demonstrates how complex features in DSC traces can occur due to the nonlinearity in the structural recovery response (after Ref. [46], with permission).

analogous to that of the reduced variables approach used in viscoelasticity theories to describe "time-temperature" superposition. Importantly, his model incorporated the concepts described above of a distribution of relaxation times (or nonexponential relaxation) and the dependence of the internal material "clock" on the current structure of the glass. Until then there was no model that incorporated the "essential ingredients" described above that are necessary to explain the major features of the phenomenology of glassy kinetics. In subsequent work Moynihan, et al. [10] enlarged upon the Narayanaswamy approach and Kovacs, et al. [11] arrived at similar results independently. The Narayanaswamy [9]-Moynihan [10] model is given an excellent review by Scherer [5] and the Kovacs, Aklonis, Hutchinson, Ramos (KAHR) [11] model is summarized by McKenna [3]. Here we treat them briefly.

3.2 The KAHR Model

Following the KAHR [11] approach, we use a multiple ordering parameter description of the (volume) departure from equilibrium. Then the rate of recovery at constant pressure is

$$-\mathrm{d}\delta_i/\mathrm{d}t = -q\Delta\alpha_i + \delta_i/\tau_i \quad (1 \leq i \leq n) \tag{1}$$

where δ is the total departure from equilibrium. $\Delta\alpha = (\alpha_l - \alpha_g)$ is the change in coefficient of thermal expansion from liquid (l) to glass (g) at the glass transition and the $\Delta\alpha_i = g_i\Delta\alpha$ are the weighted contributions of each ordering parameter to $\Delta\alpha$ with

$$\Sigma g_i = 1 \tag{2}$$

$q = \mathrm{d}T/\mathrm{d}t$, t is time and the τ_i are the retardation times associated with each retardation mechanism or ordering parameter. The model assumes that each retardation mechanism is independent of the others and depends on the total departure from equilibrium. Furthermore the total departure from equilibrium is written as

$$\delta = \Sigma\,\delta_i\,. \tag{3}$$

Similar equations can be written for the enthalpy and also for pressure dependence rather than temperature.

The solution to Eq. (1) now depends upon the specific temperature and structure dependences of the τ_i. These dependences are put into the KAHR model in a manner equivalent to the time-temperature superposition principles of viscoelasticity theory. Then, KAHR assume that, by a change in temperature or δ, each retardation time is shifted by the same amount and that the amount of shift due to a change in temperature is independent from that due to the departure from equilibrium, i.e., structure. As a result, the spectrum of retardation times simply shifts along the time axis but does not change in shape:

$$\tau_i(T,\delta) = \tau_{i,\mathrm{r}}\,a_T a_\delta \tag{4}$$

where a_T and a_δ are the appropriate shift factors for the spectrum at any T and δ relative to $\tau_{i,\mathrm{r}}$ in the reference state at equilibrium. Then simply

$$a_T = \tau_i(T,0)/\tau_i(T_\mathrm{r},0) \tag{5}$$

$$a_\delta = \tau_i(T,\delta)/\tau_i(T,0) \tag{6}$$

where T_r is a reference temperature and $\delta = 0$ denotes equilibrium. The thermal history dependence of δ can be written in a compact form by using a reduced time variable z

$$z = \int_0^t \frac{d\xi}{a_T a_\delta} . \tag{7}$$

The relevant expression for $\delta(t)$ is then given by the convolution integral

$$\delta(z) = -\Delta\alpha \int_0^z R\,(z - z')\, dz' \frac{dT}{dz'} \tag{8}$$

where $R(z)$ is a normalized retardation (recovery) function for the system

$$R(z) = \Sigma g_i \exp(-z/\tau_{i,r}) . \tag{9}$$

In the case of the Narayanaswamy [9]-Moynihan [10] development, the retardation function is expressed in the so-called Kolrausch [7]-Williams-Watts [8] (KWW) function

$$R(z) = \exp\,(-z/\tau_r)^\beta . \tag{10}$$

Importantly, these equations provide a formalism for the description of the behavior of the glass once one determines the functions for a_T, a_δ, $\Delta\alpha$, and $R(z)$. This is not a trivial experimental task and requires some further simplifications. However, the major features of the volume recovery response are reasonably well represented within this formalism. A thorough discussion of the quality of such models is presented by McKenna [3]. At this point we leave the reader to note the importance of Eqs. (1–10) lies in the ability to obtain the nonlinear behavior manifested in the asymmetry of approach experiment simply by including in Eq. (8) the fact that δ depends upon itself through the reduced time z expressed in Eq. (7). Similar concepts of a material clock that depends upon the response that one is attempting to measure has also become increasingly important in the description of the nonlinear viscoelastic properties of polymers and will be touched upon in a future section.

4. Physical Aging in Polymer Glasses

4.1 The "Classical" Picture of Aging

4.1.1 The Linear Viscoelastic Regime It has long been known that polymeric materials are not in equilibrium in the glassy state and that their properties resultingly changed with time. In the early 1960's Kovacs, Stratton and Ferry [12] attempted to use the ideas of volume recovery and reduced time to describe the behavior of a polymer glass when subjected to changes in temperature. Their results were inconclusive, possibly due to lack of precision in the measurements, but suggested that as a first approximation the idea that as the glassy structure changes, the mechanical viscoelastic response might shift in a fashion similar to the volume recovery response described above. Subsequently, this idea was exploited extensively by Struik [13] in the mid-1970s in his development of the "classical" picture of physical aging. We review this here.

If one subjects the glass to a temperature jump from above T_g to below T_g, the volume recovery described previously is accompanied by changes in the viscoelastic response of the glass. In Struik's work, when the mechanical deformations were small, he observed that the creep curves obtained at different aging times t_e after the T-jump could be superimposed by a shift a_{te} along the time axis. He thus established a time-aging time superposition principle similar to time-temperature superposition. Note that we have used the notation a_{te} here rather than a_δ as was done above. The reason for this is our feeling that the effects of structure on the structural (volume, enthalpy) recovery response need not be the same as for the viscoelastic (creep, relaxation) response. Thus, although the changing thermodynamic state obviously impacts the kinetics of both structural recovery and creep or stress relaxation, its impact need not be the same for the different processes [14–20].

Results for aging of epoxy glasses obtained in this laboratory following Struik's methods are shown in Figs. 8, 9, and 10. Figure 8 first schematisizes the typical experimental procedure in which the polymer is quenched to some temperature at $t_e = 0$ and then load-unload "probes" are applied sequentially to the sample at increasing aging times. The duration of the loading time t_1 is kept at less than 0.1 t_e in order that the changes that occur in polymer structure during the mechanical experiment be small and not influence the measurement. In Fig. 9 we show creep curves obtained from uniaxial extension measurements on an epoxy glass at different aging times in which the shifting of the responses to longer times with increasing t_e is clear. Typical shift factors required to superimpose such curves plotted against aging time are depicted in Fig. 10 for different temperatures of aging. The important feature that is shown here is that upon

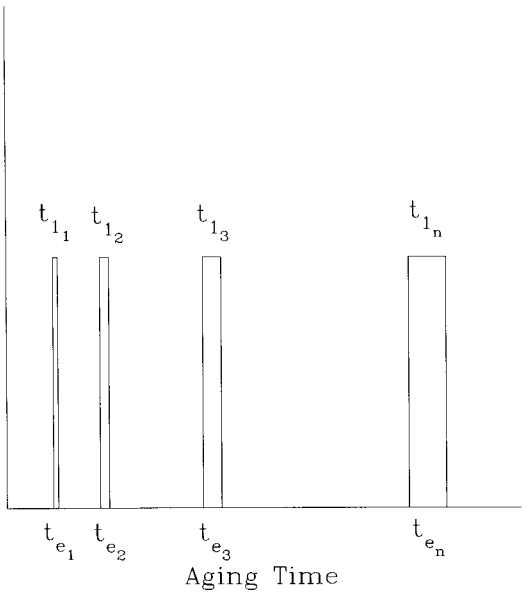


Fig. 8. Schematic of loading sequence to probe physical aging after a temperature jump. γ is the applied load or deformation. The t_{ei} are the aging times after the temperature change. The t_{1i} represent the duration of the load application. In general one selects $t_{1i}/t_{ei} \leq 0.10$.

aging below the glass transition two types of behavior are evident. First, far below T_g the aging continues for the duration of the experiments. Second, near to T_g the aging seems to nearly cease at some time "$t^{\dagger}$" that might be expected to be some equilibration time for the glass. This will be discussed subsequently.

In the linear viscoelastic regime Struik found that the double logarithmic shift rate $\mu = \mathrm{d}\ \log(a_{te})/\mathrm{d}\ \log(t_e) \leq 1$ in the aging regime which he defined as being between the α and β transitions in polymers. While this may be a predominant effect, there is little doubt that aging does occur below the β transition in glassy polymers, but the relative effects of the structural recovery on the β-transition and the glass transition are different and certainly the time-aging time superposition will break down in this regime. An example of such a problem arose in the work of McKenna and Kovacs [21] on PMMA in which it was found that the stress relaxation responses at different aging times could not be superimposed below 80 °C and this was attributed to the influence of the strong and broad β relaxation in PMMA.

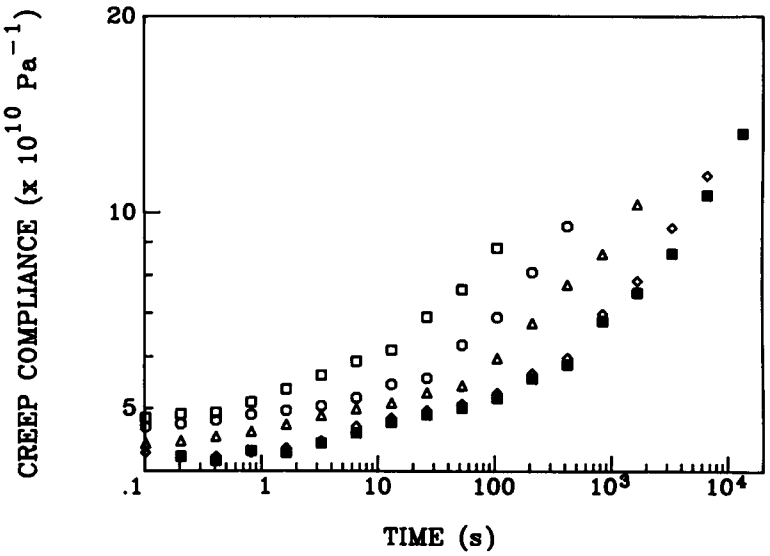


Fig. 9. Creep compliance curves at different aging times t_0 for an epoxy glass quenched from T_g + 22 °C to T_g − 9 °C. t_e: (□) 28 min; (○) 126 min; (Δ) 503 min; (◇) 2013 min; (■) 4026 min. (After Ref. [14b].)

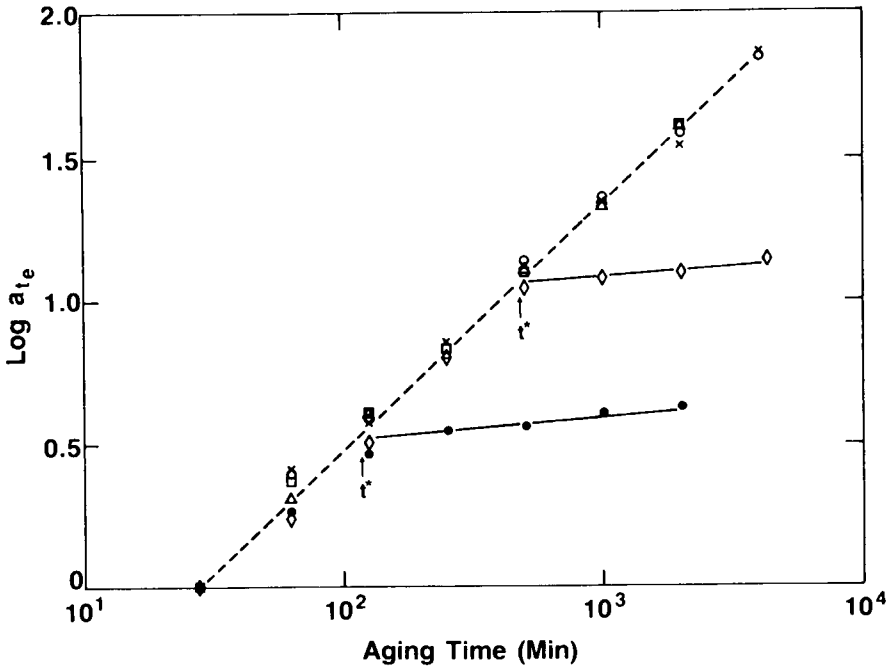


Fig. 10. Double logarithmic representation of aging time shift factor a_{te} vs aging time t_e for an epoxy glass aged at different temperatures below its T_g. $T_g - T$: (●) 30.1 °C; (X) 24 °C; (□) 20.8 °C; (◇) 10.3 °C; (●) 6.3 °C. (After Ref. [14a].)

The picture of physical aging that emerges from the work of Struik and others in the linear viscoelastic regime is relatively straight forward and indicates a "structural" dependence of the characteristic viscoelastic times that can be described by a time-aging time superposition principle—at least to first approximation. Struik [13], Kovacs, Stratton and Ferry [12], among others, explain the structural changes in terms of free volume concepts. Because these are well covered in the literature (see, e.g., Ferry [2]; McKenna [3]) we will not discuss them here.

4.1.2 The Nonlinear Viscoelastic Regime The picture of physical aging put forth above applies to the linear viscoelastic or small stress and deformation regime. In the case of the effects of structural recovery on the nonlinear viscoelastic response of the glass, the picture changes, and as will be seen is controversial. First we will present the views of Struik [13] as originally put forth in his book and early papers. This will be followed by a section describing work done in our laboratory that verify the experiments of Struik (and others) but that provide evidence that contradicts his interpretations.

Figure 11 depicts schematically the evidence that Struik had that indicates the differences in the aging behavior in the linear and nonlinear viscoelastic regimes. There are two things to be observed here. First, as observed previously by Matsuoka [22], at a given aging time the creep compliance at large stresses is shifted to shorter times relative to the small stress response. Second, upon increasing the aging time, the aging time shift factor is greater for the small stress experiment than for the large stress experiment. Thus, as shown in Fig. 12, when the stress is increased, the slope of the double logarithmic plot of a_{te} vs t_e decreases. Struik concluded from these (and some other experiments) that the large stress erases the prior aging. This has subsequently come to be known as rejuvenation. Figure 13 depicts schematically the sort of interpretation that Struik used to explain why the large stresses or deformations rejuvenate the glass. He hypothesized that the volume (or free volume) is increased by the input of mechanical work by the loading and unloading of the sample with the result that the stresses move the glass closer to the same structural state as the unaged or freshly quenched glass.

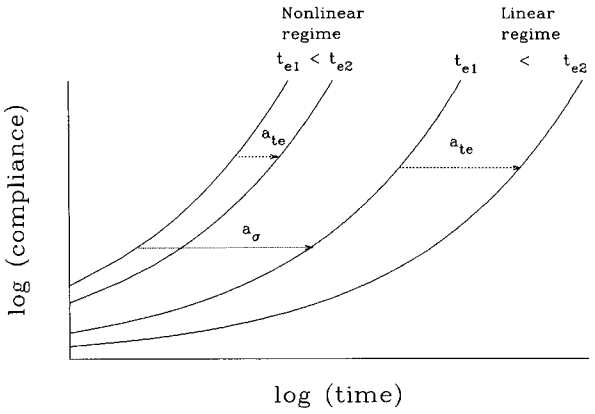


Fig. 11. Schematic representation of the relative effects of structural recovery on the small stress (linear regime) and high stress (nonlinear regime) creep compliance of a glassy polymer. The a_{te} represent aging time shift factors. The a_σ is a possible stress shift factor.

These results, and others presented by Struik and others, are not questioned. What is observed as the signature of rejuvenation is the decrease in the slope μ of $\log(a_{te})$ vs $\log(t_e)$. However, the hypothesis that the structure of the glass (as the nonequilibrium thermodynamic state) changes upon application of the large stresses or deformations, has specific implications for the aging behavior as manifested in the evolution of the viscoelastic properties as well as for the volumetric response after the material has been deformed. These are addressed in the next section.

4.2 Some "Non-Classical" Aging Experiments

In dealing with the aging problem, and in particular with the problem of the interactions of the large deformations or stresses with the volume or structural recovery, we have attempted to address the problem from the point of view of nonlinear viscoelasticity. In several papers [14–18, 23], we argued that the apparent rejuvenation of the polymer glass was simply a manifestation of the fact that the observed effects were either due to a "memory" effect or, as in the case of the above experiments, of the fact that the volume recovery impacted the large deformation response less than it did the small deformation response. In what follows we show results from two types of experiment that support our contention that aging affects the nonlinear response less than it does the linear response, as evidenced by the Struik observation of a lower value of the double logarithmic derivative $\mu = \mathrm{d}\log(a_{te})/\mathrm{d}\log(t_e)$. First, we look at results

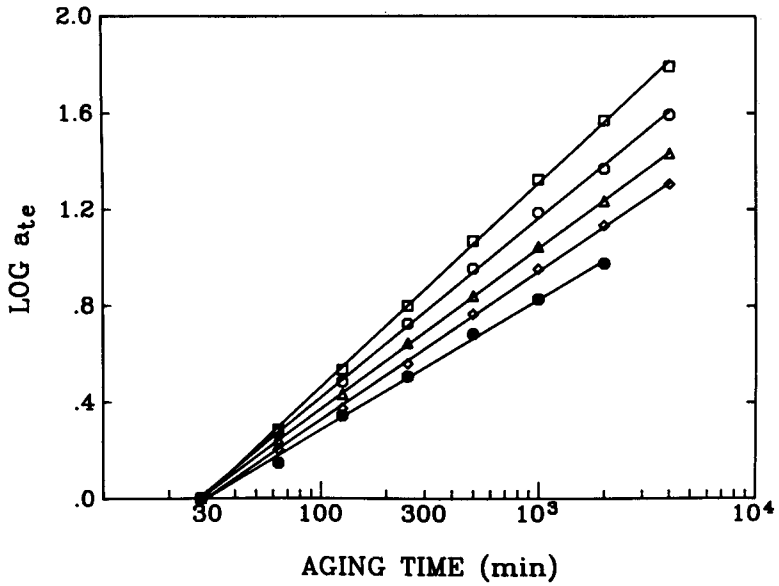


Fig. 12. Double logarithmic representation of a_{te} vs t_e for an epoxy glass aged at $T_g - 13.2$ °C. Symbols represent results from tests in which aging response was "probed" at different levels of applied stress: (□) 1 MPa; (○) 5 MPa; (Δ) 10 MPa; (◇) 15 MPa; (●) 20 MPa. (After Ref. [14b].)

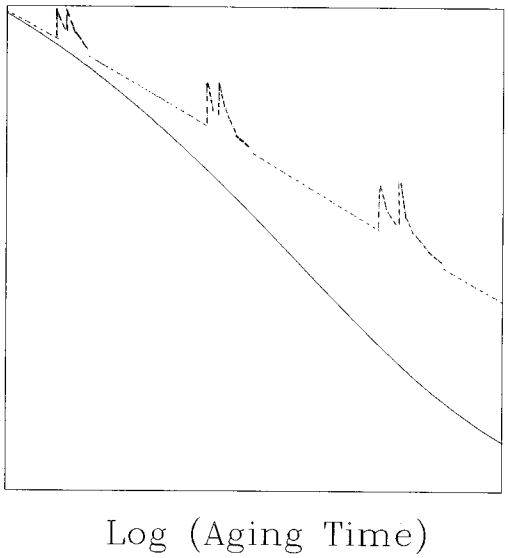


Fig. 13. Schematic of anticipated effect of applied stress on the volume recovery response if "rejuvenation" or "erasure" hypotheses are correct. Solid line, volume recovery in no-stress experiment. Dashed line represents residual effect due to erasure or rejuvenation. (See text for discussion.)

from experiments performed close to the glass transition and ask what does "$t^{\dagger}$" do as a function of stress magnitude? Second, in a similar temperature regime, we show results from experiments using the NIST torsional dilatometer [24], that show no change in the kinetics of volume recovery upon application of large deformations.

Recalling that $t^{\dagger}$ is a measure of the aging kinetics (how long until the mechanical response attains its equilibrium response)[1], the rejuvenation hypothesis suggests that $t^{\dagger}$ should increase as the level of the applied "probe" stress σ increases. With this in mind, Lee and McKenna [14] undertook a series of uniaxial extension creep experiments using epoxy glasses that could be readily tested near to their glass transitions. By changing the magnitude of σ in aging experiments after a temperature jump they were able to observe how $t^{\dagger}$ varies with σ and the change in μ at the same time. Figure 14 illustrates their dramatic finding that $t^{\dagger}$ is independent of the applied stress σ even though the signature of rejuvenation, that μ decreases with increasing σ,

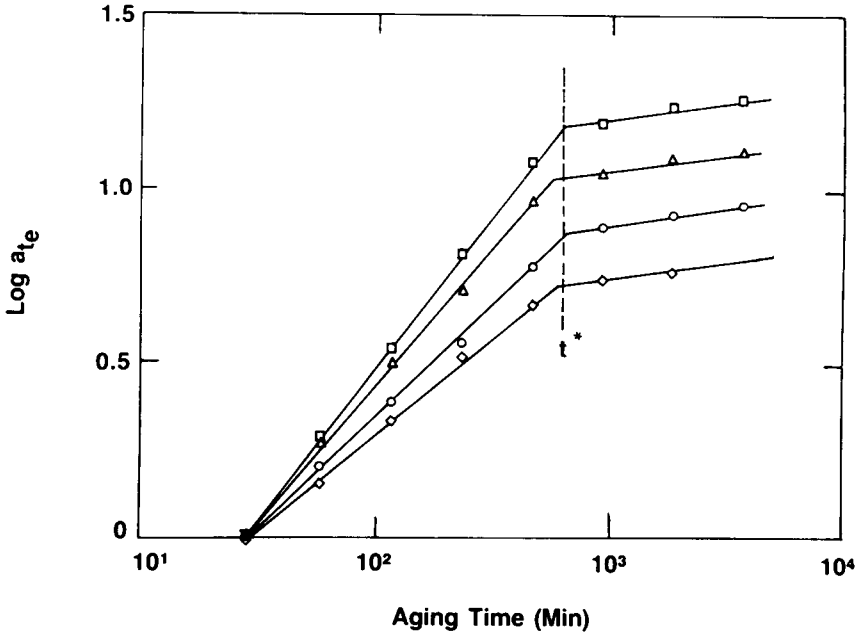


Fig. 14. Double logarithmic representation of a_{te} vs t_e for an epoxy glass aged near to the glass transition. Symbols represent tests performed at different applied stresses: (□) 1 MPa; (△) 5 MPa; (○) 10 MPa; (◇) 15 MPa. Note that t* (or t†) is insensitive to the level of applied stress. (After Ref. 14b.)

[1] In the prior work from this laboratory cited here we referred to $t^{\dagger}$ at t^*. Here we have adopted the notation $t^{\dagger}$ because it was brought to our attention that t^* is often used to mean reduced time in some other milieu. Hence we seek to avoid future confusion by introducing $t^{\dagger}$.

remains. This result provided strong evidence of the fallacy of the rejuvenation hypothesis, but was indirect in that $t^{\dagger}$ is not necessarily the equilibration point for the glass. Therefore, direct and simultaneous measures of the volume recovery and viscoelastic response after temperature jumps were perceived to be necessary. Such experiments are discussed next.

At NIST a torsional dilatometer was built [24] in which the torque, normal force and volume change in samples could be measured simultaneously in arbitrary thermo-mechanical histories. In this work we chose to first examine the down-jump aging experiment in mechanical histories similar to those defined by Struik [13], but in stress relaxation rather than creep. The torsional geometry of deformation was chosen because volume changes due to the mechanical deformation are second order effects in torsion and would therefore not swamp the small volume changes observed after small temperature jumps. (Had extension or compression been chosen, the Poisson's effect would have been very large compared to the volume recovery observed due to the temperature jump and the problem of taking small differences between large numbers would have magnified any uncertainties in the results). The NIST dilatometer has been described elsewhere [24]. It has a volume change sensitivity of approximately 10^{-7} cm^3/cm^3 for short term tests where temperature variations are unimportant. The apparatus now has a long term temperature stability of approximately ± 8 mK which leads to a long term volume stability of about 1.5×10^{-5} cm^3/cm^3.

Figure 13 shows the sort of dilatometric behavior one would expect during an aging experiment in which the deformations were large. In Fig. 15 typical results from experiments performed in the NIST torsional dilatometer at a large deformation (well into the regime where the "signature" of rejuvenation is seen) are shown. While the application of the torsion does temporarily increase the volume of the glass, as does the reduction of the torsion to zero, what is observed is that the underlying (baseline) volume recovery is unchanged by the application of the large deformations. Thus, the changing aging time shift factor with increasing stress is not explained by a change in the structure of the glass upon application of the deformation. Apparently the volume increases upon torsion do not contribute to a change in the structural recovery rate in the way anticipated by the rejuvenation model. These results indicate that, in some sense, the thermal volume and the mechanical volume are only coupled in the direction that the thermal volume (structure) impacts the mechanical response of the material. On the other hand, the mechanical deformations do not change the underlying structure of the glass in as much as the volume recovery kinetics are unaltered by them. An extensive discussion of this decoupling is presented in Lee and McKenna [14]. Two cautions need to be made about this interpretation of the results. First, volume is not the only state variable which is involved and it is conceivable that the results are only relevant to volume and not, e.g., entropy (enthalpy)—in which case free volume models become very tenuous indeed. Second, the results obtained to date are for behavior below the yield of the polymer and need to be extrapolated with caution beyond the yield.

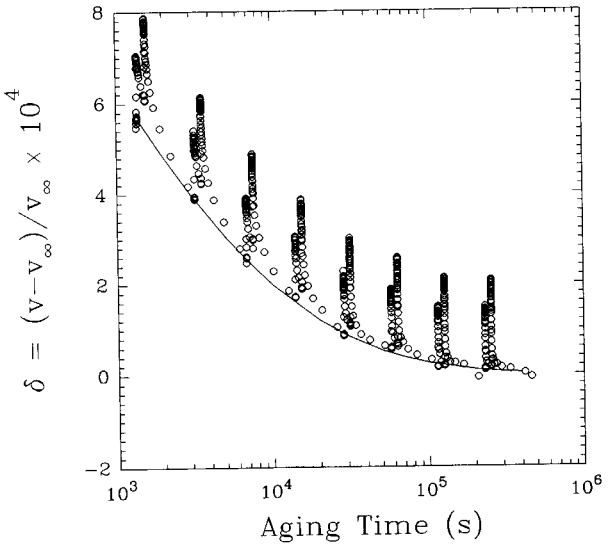


Fig. 15. Volume departure from equilibrium vs aging time for an epoxy glass at $T_g - 8.9$ °C after a down-jump from above T_g. Points are data obtained for a torsional "probe" strain of 3%, solid line represents data obtained for an undeformed sample. (Data from Ref. [9].)

5. Impact of Structural Recovery/Aging on Engineering Properties

Before going on to discuss some of the ways in which the above findings impact the development of methods for the description the mechanical response of glass forming polymers, we take an aside here to describe the impact of aging on some other properties of polymers. In particular, we are interested in the response of engineering properties that

often are assumed to be constant. The aging picture presented above provides a basis for understanding qualitatively what is happening. Much remains to be done to explain quantitatively the impact of aging on the yield and failure responses of polymers.

5.1 Yield

It has been shown by other workers that the yield stress of a glass forming polymer increases with increasing aging time after a quench [13, 25–26]. Here we describe work performed in our laboratory on epoxy resins that illustrates dramatically the impact of aging on material response [27]. Experiments were performed in uniaxial compression using a technique often used in metal plasticity in which the sample was loaded at one strain rate ϵ_1 until a maximum occurred in the stress strain curve. At the stress maximum the strain rate was increased by a factor of 10 to ϵ_2. The procedure allows one to obtain data at two strain rates using a single specimen. Figure 16 depicts the responses of an epoxy glass at increasing aging times after a quench from above T_g to a temperature near to the glass transition. It is obvious from this plot that the yield stress at both strain rates increases as the aging time increases. The magnitude of the increase is surprising, and as seen in Fig. 17 can attain a factor as great as 1.8 in aging from 0.1 h to 1000 h. Such behavior was also observed for a thermoplastic by Pan and McKenna [28] in separate experiments. Interestingly, in other experiments nearer to T_g one also observes a $t^\dagger$ break in the evolution of the yield stress with aging time. It turns out that this break occurs at longer times than for the viscoelastic properties. [16, 27]. This question of different time scales for different properties is one which is perplexing and unresolved. (A discussion of different time scales in polymer glasses is made in the paper of Santore, Duran and McKenna [16]). The major observation here is that the yield stress is obviously dramatically affected by the structural recovery of the glass and can evolve for long periods of time. This obviously can be important for engineering structures and composites.

5.2 Creep Rupture

The failure of materials is another time dependent property of polymers that has only partly been explained in spite of much work in the area. In the case of the impact of structural recovery on failure, there has been little work done on failure of bulk materials in creep rupture conditions. In studies performed by Crissman and McKenna [29, 30] on poly(methyl methacrylate) (PMMA) it was found that under certain conditions the creep rupture lifetime could be related to the viscoelastic properties of the material. In this case the impact of aging on both the creep response and the failure response was similar.

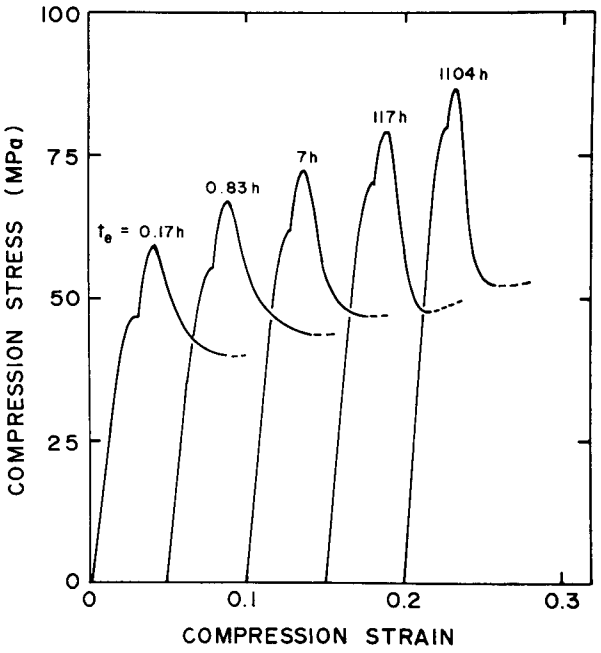


Fig. 16. Typical compression curves obtained at $T_g - 10$ °C for an epoxy glass for different aging times, as indicated. (After Ref. [27].)

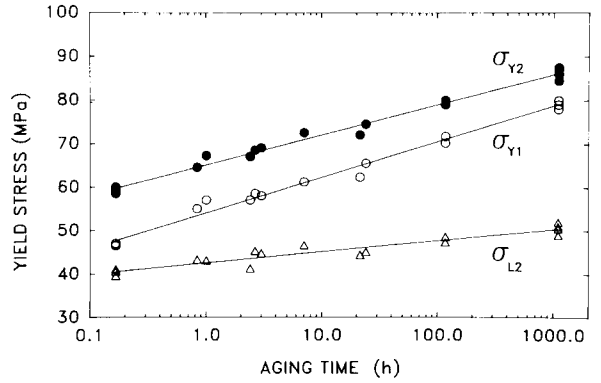


Fig. 17. Influence of aging time on the compression yield stress obtained at different strain rates for an epoxy glass tested at $T_g - 15$ °C. σ_{Y1} represents yield at a strain rate of 10^{-3} s^{-1}; σ_{Y2} represents a strain rate of 10^{-2} s^{-1}. σ_{L2} is the lower yield after σ_{Y2}. (After Ref. [27].)

Two studies were carried out. First, PMMA that had been aged at room temperature for approximately 5 years was compared with PMMA aged for one week at the same temperature. Second, PMMA samples that had been aged at 80 °C and then tested at 22.5 °C were compared for different aging times with the result that chemical and physical aging responses were both present.

In Fig. 18 we depict the times-to-failure t_f vs σ for the 5 year aged sample and for the 1 week aged sample. Two things should be noted from this figure. First, the stress dependence is a power law with an exponent of approximately 20. Second, the failure time for the 5 year aged sample is approximately 2.5–3 times that of the freshly quenched sample. In Fig. 19 we depict the time-stress shift factors a_σ vs σ for the same two polymers obtained from creep curves obtained at different stresses. Again the response is power law in stress and the power law exponent is again of magnitude 20. The similarity in the exponents is evidence that the viscoelastic and failure responses are related. In addition, we found that the aging time shift between the aged and unaged glasses was a factor of 2.5–3.

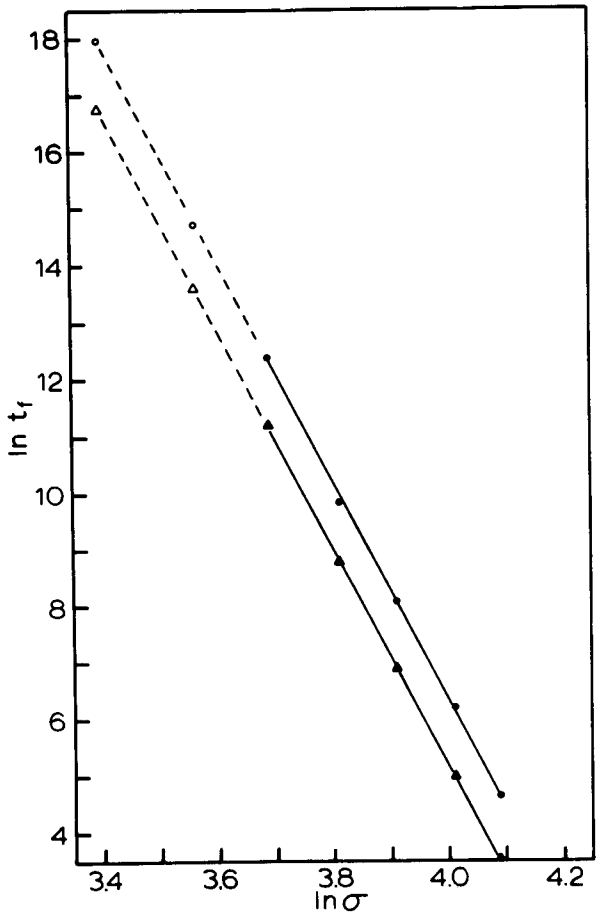


Fig. 18. Double logarithmic representation of the creep rupture time t_f vs applied stress for poly(methyl methacrylate) tested at 22.5 °C. Triangles represent data obtained for freshly quenched samples. Circles are for samples aged at room temperature for 5 years. (After Ref. [29].)

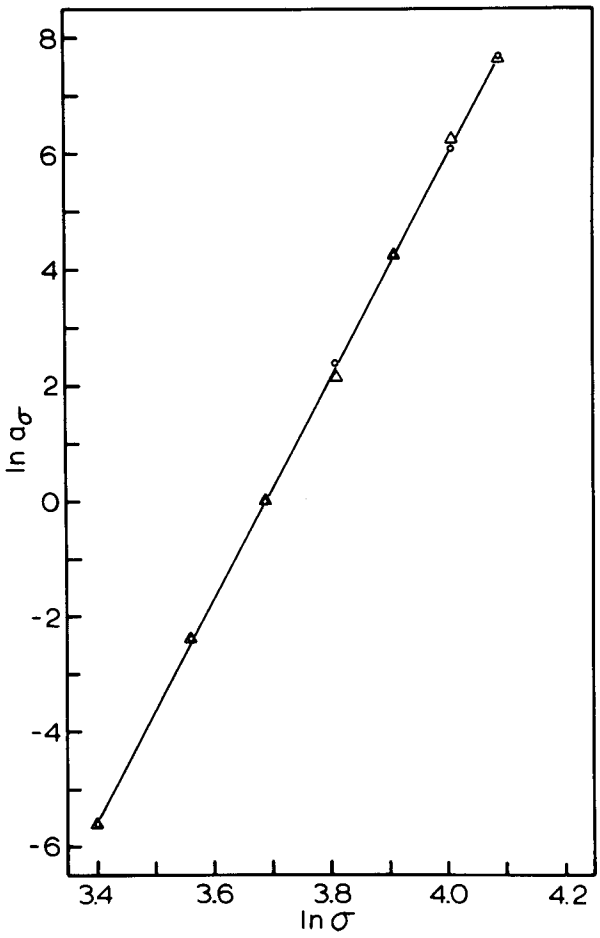


Fig. 19. Double logarithmic representation of the stress shift factor vs applied stress for poly(methyl methacrylate) tested at 22.5 °C. Symbols as in Fig. 18. (After Ref. [29].)

When time-aging time superposition also holds, as it did here, the failure and viscoelastic properties are readily described by the same shift factors. Thus, the failure and viscoelasticity of the PMMA are apparently impacted in the same way by the structural recovery.

In the instance of the PMMA we found that the strain to failure was virtually unchanged as a function of stress or whether the sample had been aged

or not. This led us to define some conditions in which a common failure criterion would be valid. The failure criterion is that the strain rate at failure multiplied by the time to failure are constant:

$$\dot{\epsilon}' t_f = c. \tag{11}$$

This was found to be true under the following conditions:

1. The time-stress superposition holds in a way such that the horizontal shift of the creep compliance curves is accompanied by a vertical shift $b_\sigma = \sigma/\sigma_r$ where σ_r is the reference stress.

2. The strain at failure is a constant.

This work was extended [30] by examining the aging of the same PMMA at 80 °C. There were three important results that came from this work. First, for a given aging time at 80 °C, the same type of time-stress superposition at 22.5 °C was observed as above. Also, the 22.5 °C creep responses of all of the samples aged at 80 °C could be superimposed by time-aging time superposition. However, the samples aged at 80 °C and tested at 22.5 °C did not show superposition with the samples aged and tested at 22.5 °C. This result is not fully understood. The second point is that at each aging time the failure criterion of Eq. (11) was followed. However, c was a function of aging time. Therefore, the aging time shift factors did not describe the changes in failure times, as they had in the samples aged at 22.5 °C. Finally, as shown in Fig. 20 the chemical degradation of the samples began to have an impact not only on the failure strain, but also on the shift factors. Thus, $\log(a_{t_e})$ vs $\log(t_e)$ goes through a maximum that corresponded reasonably well with the changes in molecular weight, which was felt to possibly result in some plasticization of the material. The maximum also corresponded to changing density for samples that had been aged for a long time. Molecular weight and density changes for the samples are shown in Table 1.

It is obvious that the complete understanding of the impact of structural recovery on the failure of polymers is not currently at hand. Recent work by Arnold [31] has shown that the increasing lifetime observed on aging of the PMMA (at 22.5 °C) is not repeated for polystyrene. This may not be surprising given the strong correlation between the viscoelastic and failure responses of PMMA, which have yet to be demonstrated for other polymers. However, this is an important effect and the physics governing it are not nearly as well defined as are those that describe the structural recovery itself. Finally, to our knowledge there has been little work performed to examine the impact of aging on the fatigue lifetimes of polymers under dynamic loading conditions.

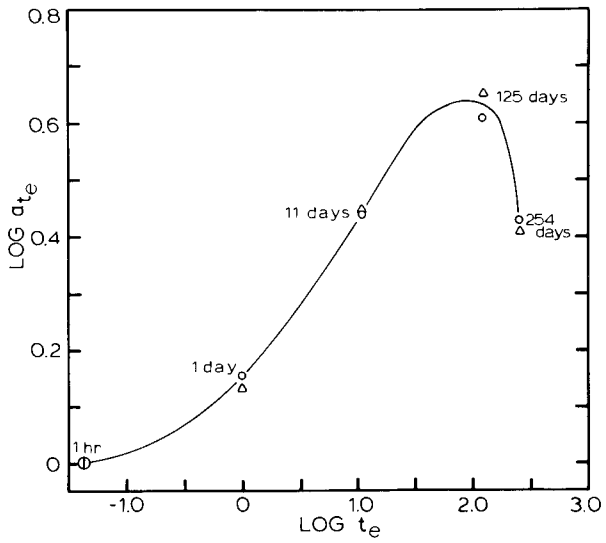


Fig. 20. Double logarithmic representation of the aging time shift factor a_{t_e} vs aging time t_e for a poly(methyl methacrylate) glass aged at 80 °C and tested at 22.5 °C. Peak in data correlates with onset of chemical degradation (see text for discussion; after Ref. [30]).

Table 1. Apparent molecular weights and densities of PMMA samples after thermal treatment

Thermal treatment	M_n[b]	M_w[b]	Density (g/cm³)[c]
As received	4.85×10^5	8.22×10^5	1.1887
Quenched from 120 °C to 23 °C	4.42×10^5	8.13×10^5	1.1887
Aged 1 h at 80 °C	4.52×10^5	8.18×10^5	1.1884
Aged 11 d at 80 °C	4.26×10^5	7.8×10^5	1.1889
Aged 125 d at 80 °C	3.02×10^5	5.62×10^5	1.1905
Aged 254 d at 80 °C	2.54×10^5	5.03×10^5	1.1888

[a] The samples were initially heated to 120 °C for 1 h followed by quenching to 80 °C.
[b] The relative error in the determination of these quantities is estimated to be about 10%.
[c] Estimated relative error in measurement is 0.0003 g/cm³.

5.3 Implications for Performance of Composite Materials

The performance of composite materials is the result of a complex interaction among the reinforcing fiber, the polymer matrix and the fiber/matrix

interface [32]. In fact, in many senses composites are not materials but rather material systems or structures. Then good design must include a thorough understanding of the behaviors of the individual components as well as the ways in which they interact. Thus, for example, the interlaminar properties of a composite may be resin or matrix dominated, but restraints imposed by the stiff fibers will influence the matrix and interface properties. Because the composite structure is complex and behavior depends upon the specifics of each system, in the following paragraphs we will simply make a general case that structural recovery or aging of the resin can be expected to significantly impact the performance of many, if not all, composite systems.

In Fig. 21 we depict the possible interactions between the matrix properties and the potential influence of the aging process on composite behavior. As can be seen, with the exception of chemical aging, which to a first approximation may not be significantly affected by the physical aging process, there are many possible routes for the matrix properties to change with a resulting change in the performance of the composite. Of course, the changes will not always be detrimental. During the course of structural recovery the creep compliance shifts to longer times, which can be expected to improve interlaminar creep resistance and off-axis creep resistance. Similarly, an increase in matrix yield stress may enhance transverse tensile strength. At the same time, interlaminar fracture toughness might be expected to be adversely affected. Finally, as noted in the figure, the fact that the dimensions of the matrix vary with the entire thermal history, can lead to residual stress build-up in composite systems. Understanding how processing histories affect the residual stresses can improve the manufacturing procedure to improve performance. Furthermore, changing the residual stresses and their relaxation behavior can lead to better control of system dimensions over the lifetime of the composite. Below we outline an example in glass to metal seals that illustrates how the understanding of the structural recovery of glasses was used to understand a failure in a hermetic seal.

In glass to metal seals used in electronics applications for the Department of Energy it is desirable to have the glass that seals the inner metallic core be under a residual compressive stress to prevent long term cracking of the seal. In the case reported by Chambers [33], tensile cracking was occurring in the seal in spite of the thermoelastic analysis prediction that the glass should be under a residual compression. At that time Chambers implemented a finite element analysis that took into account the structural recovery of the glass seal during the cooling stage of the seal formation. For the analysis, the structural recovery was modeled using the Narayanaswamy [9]-Moynihan [10] formulation of the structural recovery physics described above. The nonlinear material response that results upon cooling through the glass transition range was found

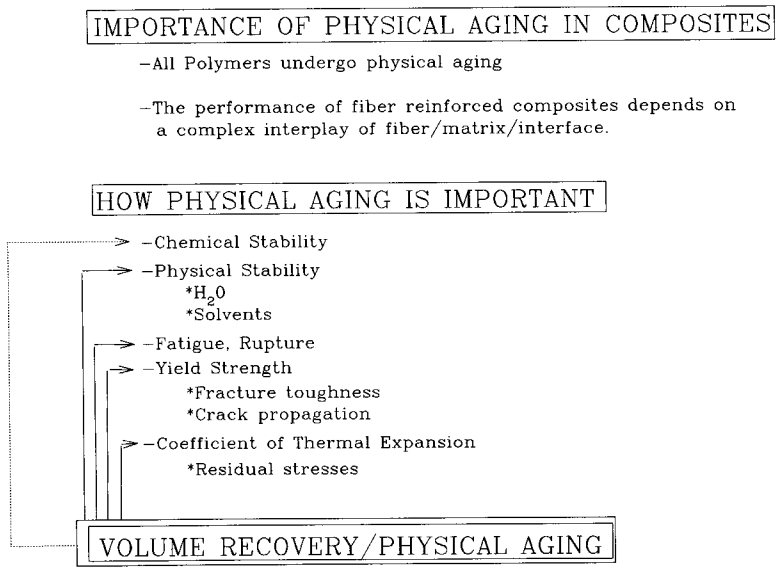


Fig. 21. Depiction of importance of structural recovery or physical aging upon the performance of composite materials.

to cause a large residual tensile stress in the glass at the temperature at which cracking was observed to occur. This is depicted in Fig. 22. The important point here is that a lack of understanding of the phenomenology of the structural recovery that tells that the volume changes are more complex than those given by a simple thermo-elastic analysis (i.e., simply taking the differences in coefficient of thermal expansion) lead to a failed part. Similar effects can be expected for fiber reinforced composites. In the case of the Chambers [33] analysis it was possible to model different processing conditions (cooling histories) that would lead to a significant reduction of the tensile stress and a reduction of seal failure.

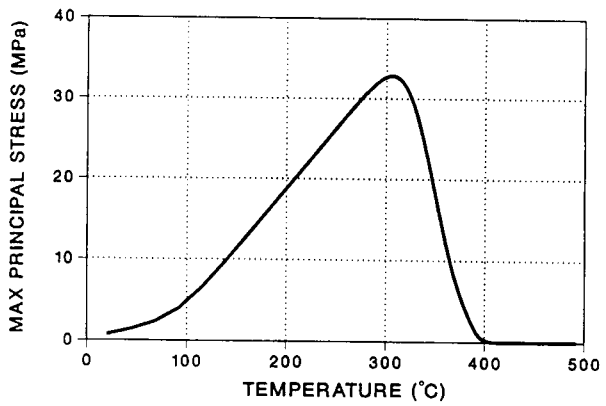


Fig. 22. Finite element model predictions of stress history upon cooling of a glass-metal seal. Tensile stress arises due to structural recovery nonlinearities. Thermo-elastic analysis resulted in a compressive stress of −6.7 MPa. (After Ref. [33], with permission.)

6. Glassy Physics and the Modeling of the Behavior of Polymers and Composites

From the above discussions it is readily apparent that there is a significant body of knowledge that is available to model the structural recovery of polymeric glasses. Such knowledge is important in the ability to develop computational models that are useful in predicting polymer material and composite system behaviors. In this section we will briefly discuss the development of such models and the ingredients necessary to implement them. Our perspective will begin with the constitutive models that describe the mechanical response of the material. (The reader is reminded that the Narayanaswamy [9]-Moynihan [10] and KAHR [11] models provide the constitutive equations to describe the structural recovery). From there we will discuss some physical evidence that exists to suggest that certain classes of constitutive equations are not correct. We will suggest some areas of work that should be vigorously pursued and finally we will argue that, although the computer power currently available is insufficient to handle the full modeling problem, it soon will be and the technological community needs to be ready to implement current and future knowledge relevant to the models we describe when computing power catches up with our knowledge.

6.1 A Brief Look at Some Nonlinear Constitutive Equations

Although there is a considerable body of knowledge concerning the nonlinear response of polymer melts and solutions, much of the work dealing with the behavior of engineering polymers has been devoted to plasticity and yield. Below the yield point these materials are, from this writer's view, viscoelastic and the correct constitutive description of their behavior will need to fall into that category. Therefore, we will not consider plasticity equations and will limit ourselves to the nonlinear viscoelastic constitutive equations.

The development of nonlinear constitutive equations is a very sophisticated field that we do not intend to thoroughly survey. Furthermore, the general multiple integral equations, while they have been used with success in some cases [34, 35] are very cumbersome to use both experimentally in the number of experiments required for the determination of material properties and computationally. Rather, we want to look at a class of single integral nonlinear constitutive laws that we refer to as reduced time or clock type equations and comment on their potential and the need to further evaluate those that appear to be promising.

The type of equation that describes the nonlinear structural recovery that was discussed previously falls into the category of reduced time equations. The material clock in the Narayanaswamy [9]-Moynihan [10] and KAHR [11] equations depends upon the instantaneous structure of the glass. (See Eqs. [7–10].) Similar classes of equations have been developed for polymer viscoelasticity and we describe those here.

6.2 The Schapery Model

One of the first nonlinear viscoelastic models of polymer solids to use reduced time variables was developed by Schapery [36, 37] in the 1960's and this model is still valuable today because the material response functions can be reasonably easily

evaluated in the laboratory and there is reason to believe that computationally it will be relatively easy to implement in finite element codes. Furthermore, the Schapery model has a formulation for both strain as a function of stress history and stress as a function of strain history—a problem that is often difficult to treat in some of the constitutive equations that have been developed. Thus, both creep and stress relaxation can be handled, albeit with different material property functions.

The creep formulation (strain as a function of stress) is as follows:

$$\epsilon = g_0 D_0 \sigma + g_1 \int_0^t d\tau \Delta D\,(\Psi - \Psi') \frac{dg_2\sigma}{dt}, \quad (12)$$

where ϵ is the strain, σ is the stress, D_0 is the zero time compliance, ΔD is the time dependent part of the compliance, τ is the dummy time variable, the g_i's are material parameters. The reduced time arguments are defined by

$$\Psi = \int_0^t dt'/a_\sigma \quad (a_\sigma > 0) \quad (13)$$

$$\Psi' = \Psi(\tau) = \int_0^\tau dt'/a_\sigma, \quad (14)$$

where a_σ is the stress shift factor. A similar set of equations was developed for the relaxation formulation (stress as a function of strain) as

$$\sigma = h_e E_e \epsilon + h_1 \int_0^t d\tau\, \Delta E\,(\xi - \xi') \frac{dh_2\epsilon}{d\tau}, \quad (15)$$

where the h_i's are material parameters and the reduced time arguments are defined as

$$\xi = \int_0^t dt'/a_\epsilon \quad (a_\epsilon > 0), \quad (16)$$

$$\xi' = \xi(\tau) = \int_0^\tau dt'/a_\epsilon, \quad (17)$$

a_ϵ is the strain shift factor.

In the Schapery formulations the reduced times are introduced in a way that is similar to that for time-temperature superposition in the sense that the strain shift factor and the strain or the stress shift factor and the stress both appear on the same side of the equation. Thus, unlike the structural recovery equations in which δ depends on temperature history and itself through a_δ [Eqs. (7) and(8)], the nonlinearity is simpler and the equations easier to handle. Also, material nonlinearity is introduced through the material parameters h_i or g_i. These equations have been used with some success and should be further evaluated because of their potential computational ease. There have also been some recent results [38] that indicate the possibility that the Schapery equations may be inadequate without modifications to treat all deformation or stress histories.

6.3 The Zapas Strain-Clock Model

One of the most successful nonlinear viscoelastic models used in melt and solution rheology was the BKZ theory of Bernstein, Kearsley and Zapas [39]. In the mid-1970s Zapas [40] proposed a modification of this BKZ model that would include a strain clock reduced time and attempted to apply the equation to the description of solid polymers. The formulation for the modified BKZ is tensorial in nature, for simplicity here we will deal with that in simple shear:

$$\sigma_{12}(t) = \int_{-\infty}^t \{[\gamma(t) - \gamma(\tau)]\, G_2[\gamma(t) - \gamma(\tau),$$

$$\Phi(t, \tau)]\}\, \dot{\Phi}(t, \tau), \gamma(\tau), t - \tau)]\, d\tau, \quad (18)$$

where σ_{12} is the shear stress, γ is the shearing strain, t is current time and τ is past time. $G_2(\gamma,t)$ is the derivative of G at the second argument, where $G(\gamma,t)$ is the nonlinear relaxation modulus. The strain clock which gives the reduced time is defined as

$$\Phi[\gamma(t), \gamma(\tau), t - \tau)] = \int_\tau^t \dot{\Phi}$$

$$[\gamma(t), \gamma(\Omega), t - \Omega)]\, d\Omega. \quad (19)$$

The strain-clock in the modified BKZ theory is a function of the entire strain history, as defined by Eq. (19). Note that again the strain-clock function (shift function) appears on the same side of the equation as does the strain—again a different form of nonlinearity from that seen in the Narayanaswamy [9]-Moynihan [10] and KAHR [11] equations for structural recovery. Additionally, the modified BKZ theory shown here does not have a

ready creep formulation, which can limit its versatility. As formulated it is also limited to incompressible materials. Both of these problems can be overcome by further development if the model is demonstrated to be useful in laboratory tests. The model will be more difficult to implement than was the Schapery [36, 37] model. It may, however, prove to be more general, although there has been little work done to demonstrate its utility. McKenna and Zapas [41] showed that it could be used to describe the two step behavior of a glassy PMMA in torsion and found that the clock function required to describe the shearing response also described the normal stress response in the same experiment.

6.4 The Bernstein-Shokooh Stress-Clock Model

In the Bernstein-Shokooh [42] stress-clock model, again the BKZ theory serves as the starting point. Now, however, the material time is assumed to depend upon the stress. The equations in simple shear can be written as

$$\sigma_{12}(t)=\int_{-\infty}^{t}\{G\cdot[\gamma(t)-\gamma(\tau),\beta(t,\tau)]\}\, b_\sigma(\tau)\,\mathrm{d}\tau\,. \tag{20}$$

Where now b_σ is the stress shift factor and the reduced time $\beta(t,\tau)$ is defined as

$$\beta(t,\tau)=\int_{\tau}^{t} b_\sigma(s)\,\mathrm{d}s. \tag{21}$$

As in the BKZ theory [39] itself, the Bernstein-Shokooh [42] stress-clock model is formulated for stress relaxation. However, unlike either the Schapery [36, 37] model or the Zapas [40] strain-clock model, the shift factor depends upon the dependent variable as is the case for the Narayanaswamy [9]-Moynihan [10] and KAHR [11] models. Thus, in Eq. (20) we find that the stress depends upon itself through the reduced time. Bernstein and Shokooh [42] showed that if one linearizes the equations in the strain (take the small strain limits), then one can invert Eq. (20) to arrive at a creep formulation as well as the relaxation formulation. Nonlinearities still arise because of the dependence of the time on the stress. The creep equation for simple shear is

$$\gamma_{12}(t)=1/2\,\sigma_{12}(t)\,J_0+1/2\int_0^t J[\beta(t,\tau)]\,\sigma_{12}(t)\,b_\sigma(\tau)\,\mathrm{d}\tau) \tag{22}$$

and the symbols have the same meaning as above. Note that this is the "linearized" form for the creep, therefore the compliance J depends not on strain but on the reduced time only.

The stress-clock formulation gives qualitative behavior that one observes in solid polymers. However, to this author's knowledge the constitutive law has not been tested for data from creep or relaxation in solid polymers. Laboratory evaluation of the model is obviously required.

6.5 Volume Clock Models

There have been several attempts in the literature to use volume (or free volume) clocks to describe the nonlinear viscoelastic behavior of polymers [43–45]. In this paragraph we present the constitutive model developed by Knauss and Emri [45].

Knauss and Emri [45] define a reduced time in terms of a shift factor $\phi[T, C, \sigma]$ where the functional forms of temperature T, concentration C and stress σ dependences are formulated from "free" volume considerations. Then one can write the isothermal, nondiluted response as:

$$S_{ij}=2\int_{-\infty}^{t}\mu[Z(t)-Z'(\xi)]\frac{\partial e_{ij}(\xi)}{\partial\xi}\,\mathrm{d}\xi \tag{23}$$

$$\tau_{KK}=3\int_{-\infty}^{t}K[Z(t)-Z'(\xi)]\frac{\partial\theta(\xi)}{\partial\xi}\,\mathrm{d}\xi, \tag{24}$$

where Z is the reduced time, S_{ij} are the deviatoric stresses, e_{ij} are the deviatoric strains, τ_{KK} is the first stress invariant and $\mu(t)$ and $K(t)$ are material functions. $\theta(t)$ is the dilatational strain in the sample and is related to τ_{KK} through the bulk creep compliance $M(t)$ and determines the reduced time Z:

$$Z(t)-Z'(\xi)=\int_{\xi}^{t}\frac{\mathrm{d}s}{\phi[\theta(s)]} \tag{25}$$

$$\theta(t)=1/3\;M(t)*\mathrm{d}\;\tau_{KK}\,, \tag{26}$$

where the * denotes a Stieltjes Convolution operation.

Again we see that material nonlinear behavior can be modeled by the use of a material time that runs differently from the laboratory time. In this case, the volume clock depends upon the hydrostatic pressure [Eq. (26)] which affects both the deviatoric (distortional) components of the stress [Eq. (23)] as well as the dilatational components

[(Eq. 24)]. Upon applying Eqs. (23–26) Knauss and Emri [45] show that yield-like behavior can be achieved in simple extension. However, in isochoric motions, such as shear, or volume decreasing motions, such as compression, these equations would not give a yielding response since the material time would either be the same as the laboratory time (shear) or the material time would shift to longer times upon decreasing the volume (compression).

The equations of Shapery [36, 37], Zapas [40, 41], Bernstein [42] and Knauss and Emri [45] all have some similar structures and can reproduce some of the features of nonlinear behavior of glassy polymers. The question arises, of course, of how one distinguishes among these potentially useful material descriptions. Obviously laboratory tests can be used to test them directly but these can be tedious and one would like to have some physical basis on which to select which equations to evaluate. We cannot make this evaluation today. However, in the next section we discuss some of the physics of clocks with the intent of providing some ideas and thoughts that should serve to guide such evaluations.

6.6 The Physics of Clocks

In a sense the physics of material clocks is not well developed, although obviously the mathematics is. Schapery's [36, 37] model is developed from considerations of nonequilibrium thermodynamics and the clock functions or shift factors can be related to second and third order terms in the entropy production and the Gibbs free energy. Obviously, further developments in the relaxation of complex systems need to be made before the full physics is understood.

From a phenomenological point of view, the clock idea is straightforward. The molecular mobility that determines the structural recovery, stress relaxation, or creep (i.e., of the system dynamics) is dependent on a state variable, applied stress or strain, strain history, etc., in a specific way. In particular, the relaxation or retardation function that describes the material behavior shifts rigidly along the time axis. If such constitutive laws are to be reasonable descriptions of material behavior, the correct variable that shifts the mobility or time scale of the material must be chosen and the constitutive law must incorporate the shift in the correct way. Because of the complicated nature of the nonlinear phenomena, attempting to validate the clock functions is difficult and there have been few systematic attempts to determine how well these functions work. In fact, here is not the place to make that evaluation. However, we do want to present some data that suggest that the volume clock, while a decidedly attractive approach to nonlinear viscoelastic constitutive law development, incorporates the wrong physics. This is not to say that the others are more correct, it is just that it is a simpler task to test the physics of the volume clock than the others where the physics is more obscure and less readily tested.

6.7 Some Evidence Against Volume Clocks

In fact the evidence against the volume clock was presented earlier in this paper. Recalling the measurements in which we observed the physical aging of an epoxy glass near to the glass transition at the same time that we measured the volume recovery, one can replot the data as $\log(a_{te})$ vs δ. This is done in Fig. 23. As seen in the plot, $\log(a_{te})$ reaches a (nearly) constant value before the volume reaches equilibrium at $\delta = 0$. If the volume clock approach were correct, then equilibrium would be attained simultaneously. Obviously there is a lack of one-to-one correspondence between the volume and the evolving mechanical properties. Our suggestion has been that a strain or stress clock might be a better description of the response. These would also explain why such phenomena as the apparent rejuvenation observed by Struik [13] occur in tension, compression and shear where the volume changes are very different. In addition, the stress and strain clocks allow for the description of large deformation behavior such as yield in compression and shear. Quantitative descriptions remain to be tested.

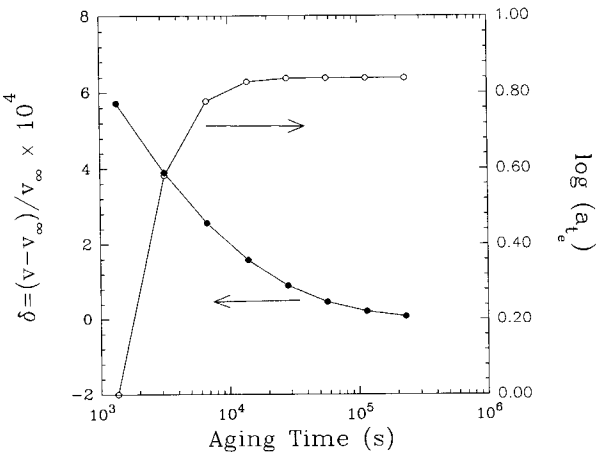


Fig. 23. Comparison of aging time shift factor ($\log a_{te}$) evolution towards equilibrium with that of volume departure (δ). See text for discussion. (Data from Ref. [16].)

7. The Importance of Understanding the Physics of Glassy Polymer

In the above we have reviewed the phenomenology of the structural recovery of polymer glasses (and inorganic glasses as well). The equations that describe the nonlinearities incorporate two major premises. First, the concept of a material clock in which the mobility of the glass (hence the structural recovery kinetics) depends explicity on the thermodynamic state through a shift factor a_δ. Second, there exists a material retardation function that is nonexponential in nature. It has historically been represented as a distribution of retardation times or a stretched exponential KWW function. We have seen how the understanding of how the material properties, through a reduced time, depend upon themselves leads to a description of otherwise complicated nonlinear behaviors.

The nonequilibrium nature of glassy materials impacts significantly the performance of polymer glass formers. Not only does the volume evolution result in a material whose dimensional stability is very problematical, it also leads invariably to changes in the engineering properties of the material. These changes can result in improved or deteriorated performance. Also, we have shown how the inherent instability of glass forming systems can lead to unexpected consequences in composite systems. Unexpected residual stresses, dimensional changes, and the like can all result from the nonequilibrium nature of glassy matrices in composites and other two component systems. Understanding the underlying phenomenology of the glass formers can lead to better process control to minimize problems.

Finally, we discussed the use of material clocks to describe the nonlinear viscoelastic behavior of solid polymers. The equations incorporate material time functions that depend upon stress, strain or volume. They have been shown to mimic, at least qualitatively, many of the observed nonlinear behaviors of polymeric materials in mechanical tests. However, there has been little in the way of systematic evaluation of either the physics behind the clock functions or the range of usefulness of the constitutive equations that result. In the last paragraphs of this paper we argue that the time is right to make those evaluations.

In the example above that showed that the nonlinear behavior of glass forming systems can lead to surprising tensile stresses in glass-metal seals, we presented a vision of the future. In this instance, Chambers [33] and his co-workers were able to use known physics and apply them to a specific problem in a finite element code. Going the next step to more complicated geometries and incorporating the material clock functions into the mechanical behaviors in addition to the volume recovery behavior of the glassy material will require a major effort. If that effort is successful, however, a large benefit will be reaped. In almost all manufacturing of polymers into two component systems, be they composites or electronic packaging, the polymer starts off in a "liquid" state and passes through a glass transition—either thermally or chemically. The parts then are cycled thermally and internal stresses (residual stresses) build up in the systems. These stresses can lead to failure (as in the glass-metal seal example) or to problems with dimensional stability or both. Successful implementation of computer codes to describe this behavior is important and possible. It will provide the ability to design processes in a way that reduces the number of iterations required to manufacture a component with "zero" defects. The barriers to the implementation of codes for complex parts are the following: 1) Computer power; 2) Acceptance and validation of appropriate nonlinear constitutive laws; 3) Lack of physical data (measurements) for the materials.

As long as the computer power has been insufficient to implement codes with sufficient complexity to reflect the true nature of the problem, we have been reluctant to move forward in the full development of the appropriate constitutive models. Now the computer power is catching up. The glass-metal seal problem is a simple one, but incorporates part of what is required—the nonlinear phenomenology of the Narayanaswamy [9]-Moynihan [10]-KAHR [11] models of structural recovery. The nonlinear stress and strain viscoelastic constitutive laws could be incorporated into codes for simple systems as well. However, the constitutive laws need to be validated. This means testing the categories described above in the range of temperature, stress and strain relevant to processing of polymers and their composite systems. From the point of view of computing these cannot be fully implemented for complex systems at this point in time. However, it would be a shame to begin the evaluation of such constitutive laws only after ten years or so when computers will be able to do the calculations. Therefore, we argue that there is a need for a coordinated effort to 1) examine the physics of material clocks; 2) establish experimentally their ability to describe the nonisothermal, nonlinear creep and relaxation of glass forming polymers (both amorphous and semicrystalline) in multiple deformation geometries

(tension, compression and shear); 3) develop efficient computer codes that can be validated in simple two component systems. Then, when computer power reaches its next levels, it will be a relatively straight forward task to implement finite element codes for complex problems. Current knowledge is now sufficient to make this breakthrough. The only thing that is lacking is the will and coordination to do it.

8. References

[1] A. J. Kovacs, Fortschr. Hochpolym.-Forsch. **3**, 394 (1964).
[2] J. D. Ferry, Viscoelastic Properties of Polymers, 3rd. edn., Wiley, New York, 1980.
[3] G. B. McKenna,Glass Formation and Glassy Behavior, in Comprehensive Polymer Science, Vol. 2, Polymer Properties, ed. by C. Booth and C. Price, Pergamon, Oxford (1989).
[4] A. J. Kovacs, J. J. Aklonis, J. M. Hutchinson, and A. R. Ramos, J. Polym. Science, Polymer Physics Ed., 1979, **17**, 1097.
[5] G. W. Scherer, Relaxation in Glass and Composites, Wiley, New York, 1986.
[6] A. Q. Tool, J. Res. NBS **37**, 73 (1946); J. Amer. Ceram. Soc. **29**, 240 (1946).
[7] F. Kolrausch, Pogg. Ann. Phys. **12**, 393 (1847).
[8] G. Williams and D. C. Watts, Trans. Faraday Soc. **66**, 80 (1970).
[9] O. S. Narayanaswamy, J. Am. Ceram. Soc. **54**, 491 (1971).
[10] C. T. Moynihan, P. B. Macedo, C. J. Montrose, P. K. Gupta, M. A. DeBolt, J. F. Dill, B. E. Dom, P. W. Drake, A. J. Easteal, P. B. Elterman, R. P. Moeller, H. Sasabe, and J. A. Wilder, Ann. N.Y. Acad. Sci. **279**, 15 (1976).
[11] A. J. Kovacs, J. J. Aklonis, J. M. Hutchinson, and A. R. Ramos, J. Polym. Sci., Polym. Phys. Ed. **17**, 1097 (1979).
[12] A. J. Kovacs, R. A. Stratton and J. D. Ferry, J. Phys. Chem. **67**, 152 (1963).
[13] L. C. E. Struik, Physical Aging in Amorphous Polymers and other Materials, Elsevier, Amsterdam, 1978.
[14a] A. Lee and G. B. McKenna, Polymer **29**, 1812 (1988).
[14b] A. Lee and G. B. McKenna, Polymer **31**, 423 (1990).
[15] Y. Leterrier and G. B. McKenna, Polymeric Materials: Science and Engineering **68**, 218 (1993).
[16] M. M. Santore, R. S. Duran, and G. B. McKenna, Polymer **32**, 2377 (1991).
[17] G. B. McKenna, Y. Leterrier, and C. R. Schultheisz, Polymer Engineering and Science (In press, 1994).
[18] R. J. Roe and G. M. Millman, Polym. Eng. Sci. **23**, 318 (1983).
[19] J. M. Hutchinson, Aberdeen University, Personal communication.
[20] G. B. McKenna, M. M. Santore, A. Lee, and R. S. Duran, J. Non-Crystalline Solids **131-133**, 497 (1991).
[21] G. B. McKenna and A. J. Kovacs, Polym. Eng. Sci. **24**, 1138 (1984).
[22] S. Matsuoka, S. J. Aloisio, and H. E. Bair, J. Appl. Phys. **44**, 4265 (1973).
[23] G. B. McKenna and L. J. Zapas, Polymer **26**, 543 (1985).
[24] R. S. Duran and G. B. McKenna, J. Rheology **34**, 813 (1990).
[25] R. A. Bubeck, S. E. Bales, and H.-D. Lee, Proc. 12th Annual North American Thermal Analysis Society Meeting, Williamsburg, VA (1983) pp. 38–43.
[26] C. Bauwens-Crowet and J.-C. Bauwens, Polymer **23**, 1599 (1982).
[27] C. G'Sell and G. B. McKenna, Polymer **33**, 2103 (1992).
[28] D. H. Pan, T. S. Chow, and G. B. McKenna, Bull. Ameri. Phys. Soc. **37**, 676 (1992).
[29] J. M. Crissman and G. B. McKenna, J. Polym. Sci., Phys. Ed. **25**, 1667 (1987).
[30] J. M. Crissman and G. B. McKenna, J. Polym. Sci., Phys. Ed. **28**, 1463 (1990).
[31] J. C. Arnold, J. Polym. Sci. Phys. Ed. **31**, 1451 (1993).
[32] G. B. McKenna, Polymer-Plast. Technol. Engr. **5**, 23 (1975).
[33] R. S. Chambers and E. B. Becker, J. Amer. Ceram. Soc. **72**, 929 (1989).
[34] C. P. Buckley, J. Phys. D. Appl. Phys. **10**, 2135 (1977).
[35] C. P. Buckley and A. E. Green, Phil. Trans. Royal Soc. **281**, 543 (1976).
[36] R. A. Schapery, Polym. Eng. Sci. **9**, 295 (1969).
[37] Y. C. Lou and R. A. Schapery, J. Comp. Matls. **5**, 208 (1971).
[38] D. Colucci and J. R. Caruthers, Bull. Am. Phys. Soc. **38**, 415 (1993).
[39] B. Bernstein, E. A. Kearsley, and L. J. Zapas, Trans. Soc. Rheol. **7**, 391 (1963).
[40] L. J. Zapas, in Deformation and Fracture of High Polymers, ed. by H. H. Kausch, J. A. Hassell and R. I. Jaffee, Plenum, New York (1974).
[41] G. B. McKenna and L. J. Zapas, J. Rheology **23**, 151 (1979).
[42] B. Bernstein and A. Shokooh, J. Rheology **24**, 189 (1980).
[43] R. M. Shay, Jr. and J. M. Caruthers, J. Rheology **30**, 781 (1986).
[44] W. K. Waldron, Jr., Influence of Normal Stress Effects on Finite Shear Deformations of Compressible Nonlinear Isotropic Solids, Ph.D. Dissertation, University of Michigan, Ann Arbor, MI, 1992.
[45] W. G. Knauss and I. Emri, Polymer Eng. Sci. **27**, 86 (1987).
[46] I. M. Hodge and A. R. Berens, Macromolecules **15**, 762 (1982).

About the author: *Dr. Gregory B. McKenna has been an active researcher in the Polymers Division at NIST since 1976, when he received an NRC Post Doctoral Research Accociateship. His research work has covered nonlinear rheological behavior of melts and solids, rubber elasticity and physics of glassy polymers. He is currently leader of the Structure and Mechanics Group in the Polymers Division. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 2, March–April 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 191 (1994)]

The Measurement and Uncertainty of a Calibration Standard for the Scanning Electron Microscope

Volume 99 | Number 2 | March–April 1994

J. Fu, M. C. Croarkin, and T. V. Vorburger

National Institute of Standards and Technology,
Gaithersburg, MD 20899-0001

Standard Reference Material 484 is an artifact for calibrating the magnification scale of a Scanning Electron Microscope (SEM) within the range of 1000× to 20000×. Seven issues, SRM-484, and SRM-484a to SRM-484f, have been certified between 1977 and 1992. This publication documents the instrumentation, measurement procedures and determination of uncertainty for SRM-484 and illustrates with data from issues 484e and 484f.

Key words: interferometer; precision; random error; systematic error; SRM; uncertainty.

Accepted: December 16, 1993

1. Introduction

Standard Reference Material (SRM)-484 is a sample for calibrating the magnification scale of a Scanning Electron Microscope (SEM). The SRM consists of nickel layers separated by thin layers of gold. Individual samples are mounted so that the layers are viewed in cross-section and appear as thin gold lines in a nickel substrate. The distances or spacings between the gold lines are certified for calibrating the magnification scale of an SEM. The latest issue, SRM-484f, was certified for five spacings, nominally of 0.5 μm, 0.5 μm, 1 μm, 3 μm, and 5 μm. A micrograph is shown in Fig. 1. However, issues prior to 484f contained spacings between 1 μm and 50 μm.

SRM-484 was developed by Ballard at the National Bureau of Standards (now NIST) in 1977 [1]. The first three issues, SRM-484, SRM-484a, and SRM-484b, were calibrated using an electron microscope by comparing the micrograph of each sample against the micrograph of a master sample which had been measured with the NIST line-scale interferometer system [2]. In 1980, studies by Swyt, Jensen, and Hembree [3,4,5] led to the development of a system for direct calibration. This system, which combines a field emission scanning electron microscope with an interferometer system, has been used for calibrating SRM-484 since 1981 with only slight modifications over the years. It has also been used for confirming the diameters of polystyrene spheres with nominal values of 3 μm, 10 μm, and 30 μm. The approach is a pitch measuring technique as shown in Fig. 2. More in-depth detail can be found in Refs. [6], [7], and [8].

2. Physical Properties

The fabrication process developed by John Young and Fielding Ogburn at NIST involves electroplating a layer of bright nickel onto the surface of a thin Monel sheet and subsequently electroplating alternate layers of gold and nickel to produce layers of desired thickness between the gold layers. The gold layer thicknesses are controlled to 200 nm or less. The sheet is vacuum heat treated at 265 °C for 16 h to relieve the residual stress in the layers.

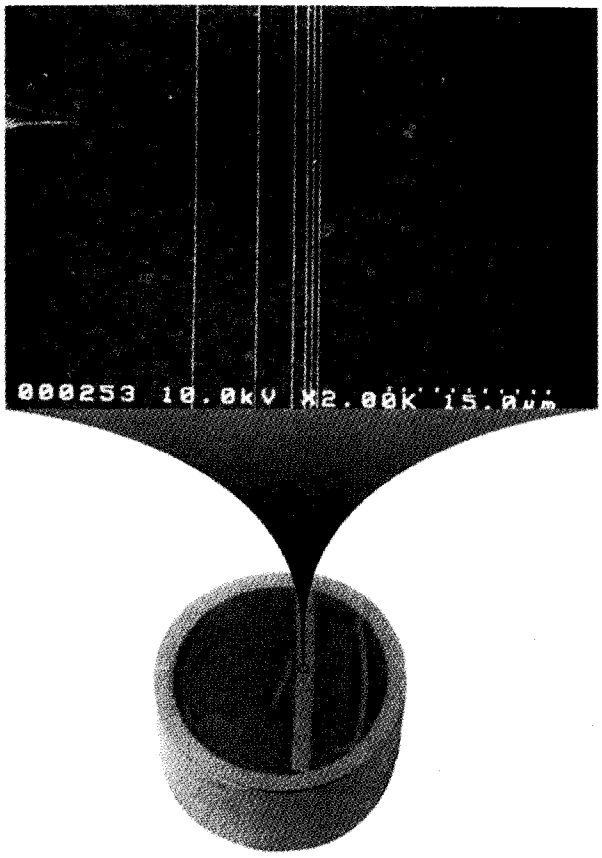


Fig. 1. A micrograph of SRM-484f.

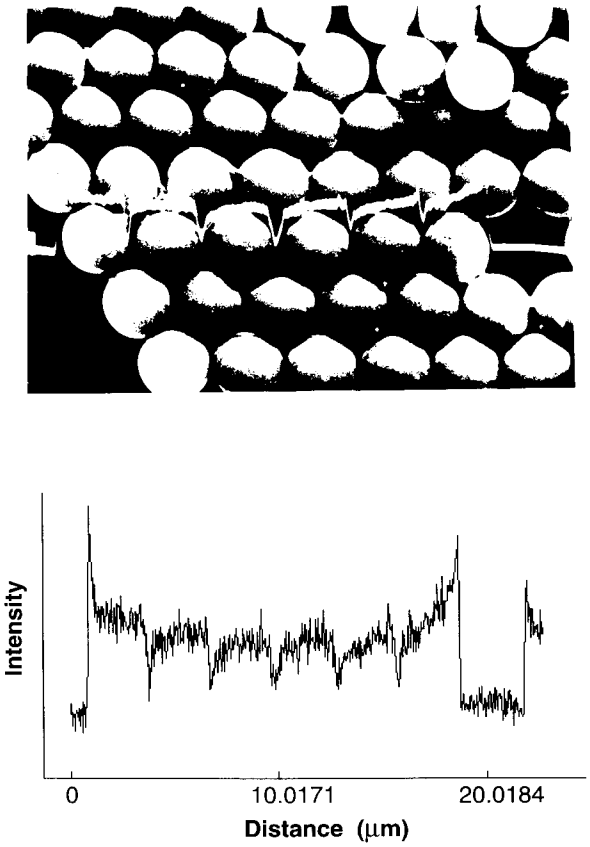


Fig. 2. Micrograph of 3 µm latex spheres and the measured data of backscattered electron intensity versus distance.

Several composite sheets have been produced by this process, each containing between five and ten spacings in the range of 0.5 µm to 50 µm. For each issue, individual samples of size 9 mm × 9 mm are sheared from a sheet and mounted sideways in a steel holder for metallographic polishing. Normally in a polishing process, soft material is removed faster than hard material, but images from a scanning tunneling microscope (STM), an atomic force microscope (AFM), and a stylus profilometer all reveal that, after polishing, the gold lines of the SRMs protrude about 30 nm above the nickel surface (Fig. 3). We speculate that a hard gold-nickel alloy may have been formed by the heat treatment, or the removal rate of nickel is faster than that of gold due to the chemo-mechanical effects in the polishing.

3. Instrumentation

The measuring system consists of a VG-HB50A field emission scanning electron microscope, HP5526A laser interferometer, piezoelectric displacement stage, wave generator and Digital MINC/DECLAB 23 computer[1]. A system diagram is shown in Fig. 4. A 30 kV, 5×10^{-9} A beam is used. Such a low beam current is necessary to minimize contamination marks on the SRMs. A scintillator back-scattered electron (BSE) detector,

Certain commercial equipment, instruments, or materials are identified in this paper to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

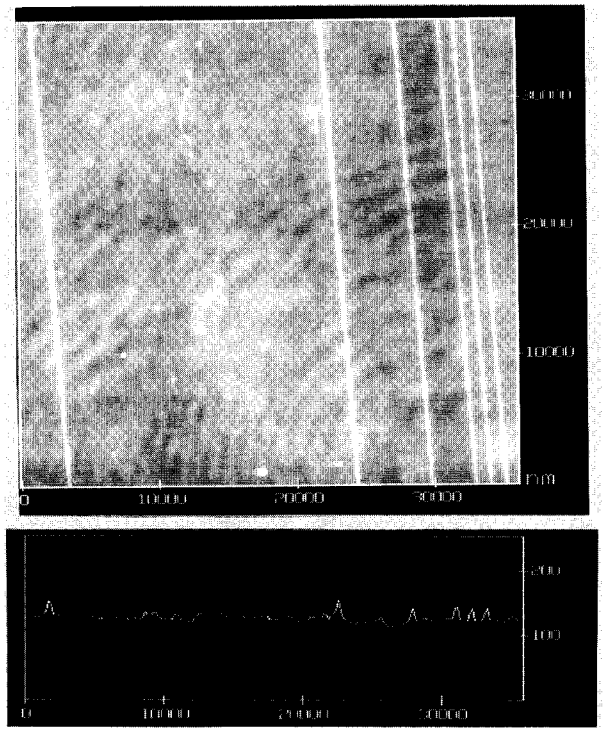


Fig. 3. An AFM image of one of the SRM-484e samples. The section profile shows the height of the lines. The height of the left most line is ≈24.5 nm. The height of the third line from the left is ≈23.6 nm.

rather than a secondary electron detector, is used because of the strong material contrast between gold and nickel [9] for backscattered electrons. The vacuum in the sample chamber of the SEM is maintained at 3×10^{-7} Pa during calibration.

In ordinary SEM operation, the electron beam is scanned across the sample to obtain the image; however, for SRM-484, the electron beam is held stationary. The SRM is carried by the piezoelectric displacement stage [10] and scanned across the beam. The stage motion is driven by a high voltage power supply (not shown in Fig. 4) whose output is controlled by a programmable wave function generator. The reflector of the measuring leg of the interferometer is mounted on the stage. A polarizing beam splitter and the optics of the reference leg are held stationary on the platform adjacent to the stage. A schematic diagram of the optical system is shown in Fig. 5.

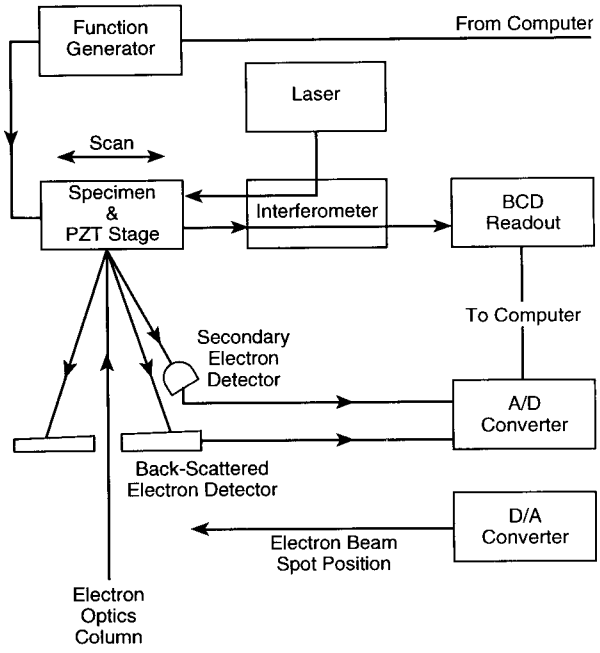


Fig. 4. System diagram of the NIST metrology electron microscope.

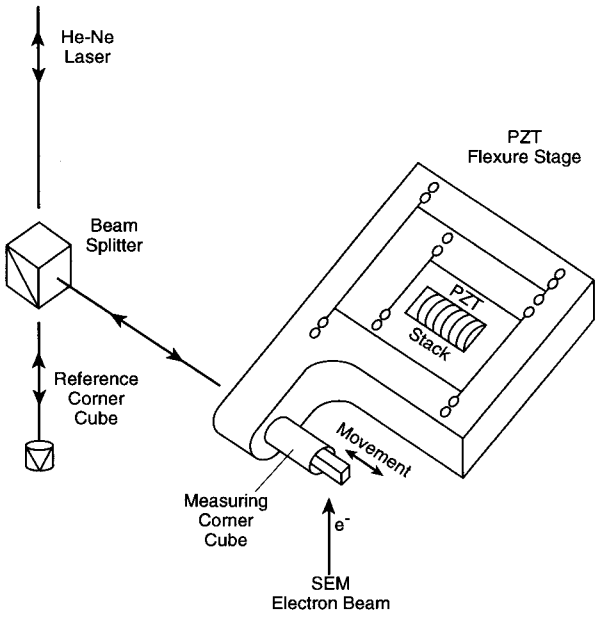


Fig. 5. Schematic drawing of the optical system of the metrology electron microscope.

As the stage scans, the computer records both the displacement of the stage as measured by the interferometer, and the current of the backscattered electrons collected by the BSE detector. Each scan takes 48 s to complete during which time 2000 pairs of distance measurements and intensities are collected. The time interval between data points is approximately 25 ms. This is equivalent to a distance of approximately 6 nm for SRM-484f and 30 nm for the other issues which were measured at higher scanning speeds. The distance between the centers of any two gold lines is computed as the difference between position readings under the two corresponding intensity peaks. A typical plot of SRM-484f data is shown in Fig. 6. The peak-to-peak measurement technique produces an unbiased estimate of the spacing between the two gold lines. The measurement problem is quite different from the linewidth measurement problem [3,4,11] where an unbiased estimate of width requires a left-edge to right-edge determination. The pattern of lines for SRM-484 prior to 484f is shown at the top of Fig. 7, and SRM-484f is shown at the bottom. An example of certified spacings is shown in Table 1 for SRM 484f.

Table 1. A typical set of certified values for SRM-484f

Spacing	Nominal value	Certified value
lines 1-0	0.5 μm	0.568 μm
lines 2-1	0.5 μm	0.579 μm
lines 3-2	1 μm	1.108 μm
lines 4-3	3 μm	3.308 μm
lines 5-4	5 μm	5.392 μm

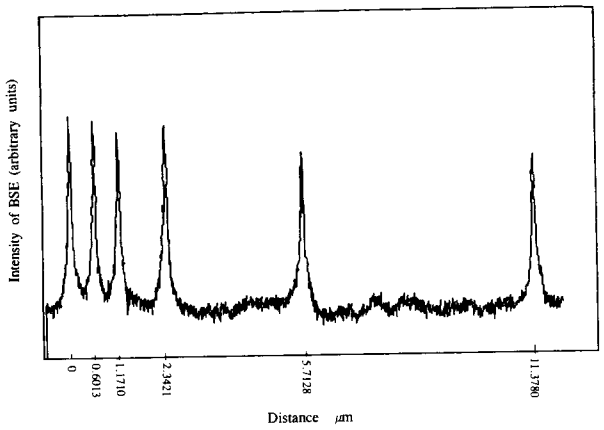


Fig. 6. A plot of measured BSE signal versus distance for one SRM from issue 484f.

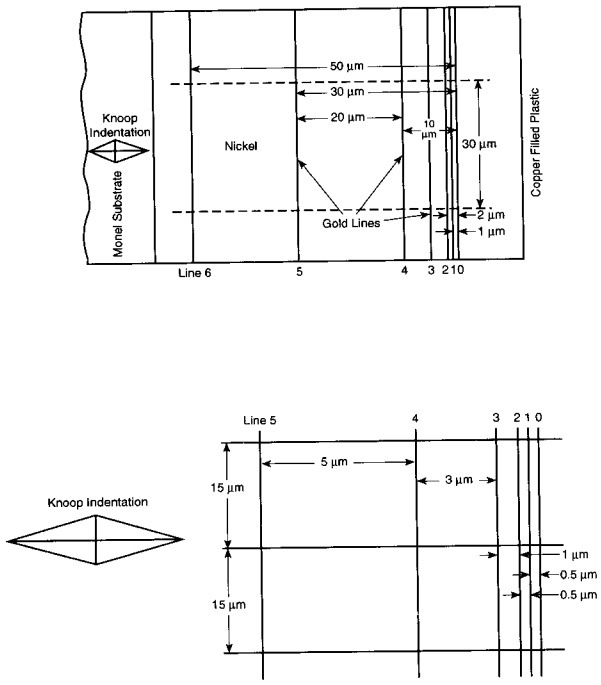


Fig. 7. The lines pattern for SRM-484a-e (Top) and 484f (Bottom).

4. Measurement Procedure

Although the measurement procedure has varied slightly from issue to issue, for the last four issues it has always involved multiple scans for estimating the effects of both instrumental error and non-parallelism of the gold lines. For issue 484e, each sample was scanned across the beam three times at each of three positions; for issue 484f, the number of positions was increased in order to check on parallelism, each sample being scanned across the beam nine times. The first three scans were at the same location along the Knoop indentation mark; the remaining six scans were at evenly spaced intervals that span 15 μm above and below the indentation mark. Approximately 150 samples were individually certified for each issue. During the period of calibration of the issue, which lasts approximately 3 months, a master sample, previously measured by the NIST line-scale interferometer system, and a control sample, selected at random from the samples in the current issue, are measured periodically in exactly the same manner as the SRM samples. The measurements on the control sample are made with sufficient regularity to cover the SRM certification in fairly even time increments and sample the range of experimental

conditions. The history of measurements on spacing 0-1 of the control sample for SRM-484f is shown in Fig. 8.

5. Components of Uncertainty

The uncertainties quoted in the calibration certificate for SRM-484f and its predecessors were based upon an analysis of uncertainty arising from random and systematic components, as described by Eisenhart, Ku, and Collé [12]. Since the issuance of the calibration certificate, a new NIST policy and associated guideline document [13] were issued for the evaluation and expression of the uncertainties of NIST measurement results. This new policy, among other things, classifies uncertainty components according to whether they are evaluated by statistical methods or by other means, and deemphasizes the use of the terms "random" and "systematic." Future calibration certificates and publications concerning the SRM-484 series will follow the new NIST guideline. However, the analysis here is based on the discussion of uncertainty in previous calibrations.

The total uncertainties for the measured line spacings are classified into random and systematic components. The random component of uncertainty depends on (1) instrumental precision, (2) the raggedness or lack of parallelism for the gold lines which causes disparities among positions on the sample, and (3) long-term measurement precision. The systematic component of uncertainty depends on the relationship between the measured value of line spacing and that realized through line-scale interferometry. A detailed description for each component is given below.

1) Instrumental precision is estimated from repetitive scans made at the same position on each sample. For SRM-484e, there are three scans at each of three positions and for SRM-484f, the only repetitive scans are at the center position on each sample. The measurement of the ith scan at the kth position is denoted by x_{ik} ($i = 1,\cdots,I$; $k = 1,\cdots,K$). Estimates of instrumental precision are made according to

$$s_{\text{inst}} = \left(\frac{1}{(I-1)} \sum_{i=1}^{I} (x_{ik} - x_{\cdot k})^2\right)^{1/2},$$

where

$$x_{\cdot k} = \frac{1}{I} \sum_{i=1}^{I} x_{ik}.$$

2) Between-position precision is estimated from the measurements at K distinct positions by

$$s_{\text{position}} = \left(\frac{1}{(K-1)} \sum_{k=1}^{K} (x_{\cdot k} - x_{\cdot\cdot})^2\right)^{1/2},$$

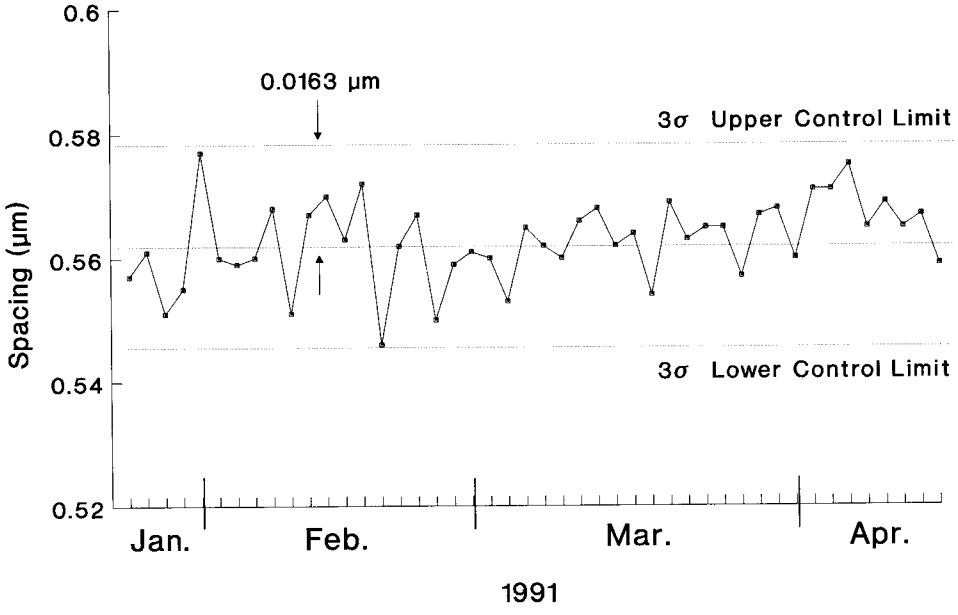


Fig. 8. Control chart showing the 0.5 μm spacing measurements on the control sample for SRM-484f.

where

$$x_{\cdot\cdot} = \frac{1}{K}\sum_{k=1}^{K} x_{\cdot k}.$$

The estimates of instrumental precision are pooled [14] over positions and the n samples to obtain an overall estimate, $\hat{\sigma}_{\text{inst}}$, and the estimates of between-position precision are pooled similarly to obtain an overall estimate $\hat{\sigma}_{\text{position}}$ (the caret ˆ symbolizes a statistical estimate from data). Any lack of parallelism is revealed in the standard deviation,

$$\hat{\sigma}_{\text{lack}} = \left(\hat{\sigma}^2_{\text{position}} - \frac{1}{I_0}\hat{\sigma}^2_{\text{inst}}\right)^{1/2}.$$

For issue 484e, the value of I_0 equals 3; for issue 484f, the value of I_0 equals 1.2 to account for the unequal number of scans at the seven positions [15]. The results of the calculations for both SRM-484e and SRM-484f are shown in Table 2. A comparison of $\hat{\sigma}_{\text{inst}}$ and $\hat{\sigma}_{\text{position}}$ shows that $\hat{\sigma}_{\text{lack}}$ is significant for issue 484e, especially for the larger spacings, indicating a general lack of parallelism for the SRMs in this issue. For issue 484f, $\hat{\sigma}_{\text{lack}}$ is negligible indicating that, in general, parallelism is not a problem for this issue. However, an individual sample from issue 484f could have parallelism problems if s_{position} for that sample is excessively large; i.e., if

$$s_{\text{position}} > \sqrt{F_{0.05}(6,2n)}\,\frac{\hat{\sigma}_{\text{inst}}}{1.2},$$

where $F_{0.05}(6,2n)$ is the upper 5 percent point of the F distribution with 6 degrees of freedom in the numerator and $2n$ degrees of freedom in the denominator. Only 5 percent of issue 484f fell into this category as should happen by chance; nonetheless, samples with large standard deviations were examined and remeasured for confirmation.

3) A total standard deviation, s_t, is estimated from the M calibrations on the control sample where the average value over all scans for the mth calibration is denoted by y_m $(m = 1,\cdots,M)$. For SRM-484f, the standard deviation is given by

$$s_t = \left(\frac{1}{M-1}\sum_{m=1}^{M}(y_m - y_{\cdot})^2\right)^{1/2},$$

where

$$y_{\cdot} = \frac{1}{M}\sum_{m=1}^{M} y_m.$$

The quantity s_t, as shown in Table 2, accounts for both instrumental precision and long-term effects which cannot be controlled in the laboratory. For SRM-484e, the total standard deviation was estimated for each position and then pooled over positions so as not to include parallelism problems.

Table 2. Component standard deviations, μm

SRM-484e						
	Instrument		Position		Total	
Spacing	$\hat{\sigma}_{\text{inst}}$	df	$\hat{\sigma}_{\text{position}}$	df	s_t	df
1	0.029	1164	0.020	350	0.0167	117
2	0.031	1164	0.020	350	0.0193	117
5	0.031	1164	0.021	350	0.0199	117
10	0.031	1164	0.022	350	0.0210	117
30	0.032	1164	0.032	350	0.0239	117
50	0.032	1164	0.072	350	0.0361	117
SRM-484f						
	Instrument		Position		Total	
Spacing	$\hat{\sigma}_{\text{inst}}$	df	$\hat{\sigma}_{\text{position}}$	df	s_t	df
0.5	0.012	318	0.012	954	0.0071	44
0.5	0.014	318	0.014	954	0.0067	44
1	0.013	318	0.014	954	0.0085	44
3	0.016	318	0.015	954	0.0117	44
5	0.018	318	0.017	954	0.0172	44

4) Systematic error relative to the defined unit of length was studied with measurements on a master magnification sample which has been measured several times with the SEM and several times with the NIST line-scale interferometer system. The line-scale system has an estimated (3σ) uncertainty of 0.01 μm. Differences between values obtained with the line-scale system and values obtained with the SEM help to identify systematic errors in either system. The uncertainty (3σ) of measured differences between the SEM and line-scale system is approximately 0.1 μm. The differences, as shown for issues 484e and 484f in Fig. 9, are within this uncertainty. For issue 484e, the differences are fairly randomly distributed about zero with a maximum of 0.012 μm and a minimum of −0.018 μm. For issue 484f, the differences are larger with a maximum of 0.046 μm for the 10 μm spacing.

The fact that the majority of the differences are in one direction for issue 484f may result from the treatment of the master sample. The SEM measurements are made on a smooth surface and the line-scale system requires an etched surface to increase contrast. The master sample was etched and measured with the line-scale system and then polished and measured with a SEM. The polishing process may remove enough material to slightly change the spacings relative to the etched surface. Although we do not treat the differences as being significant for the uncertainty statement, research in this area is continuing.

6. Certification and Uncertainty

Distances between the centers of gold lines are individually certified, and the certified region is located relative to a Knoop indentation. The certified value for each spacing is an average of all measurements, and the certification is valid within 15 μm of either side of an imaginary line extending from the Knoop indentation normal to the gold lines. The random component of uncertainty takes into account instrumental variability, long-term measurement fluctuation and any parallelism problem which affects the 15 μm region to either side of the center.

For issue 484e, the standard deviation of the certified value is

$$s = (s_t^2 + \frac{1}{3}\hat{\sigma}_{\text{lack}}^2)^{1/2}.$$

Because of the parallelism problems with the SRMs in this issue, the random component of

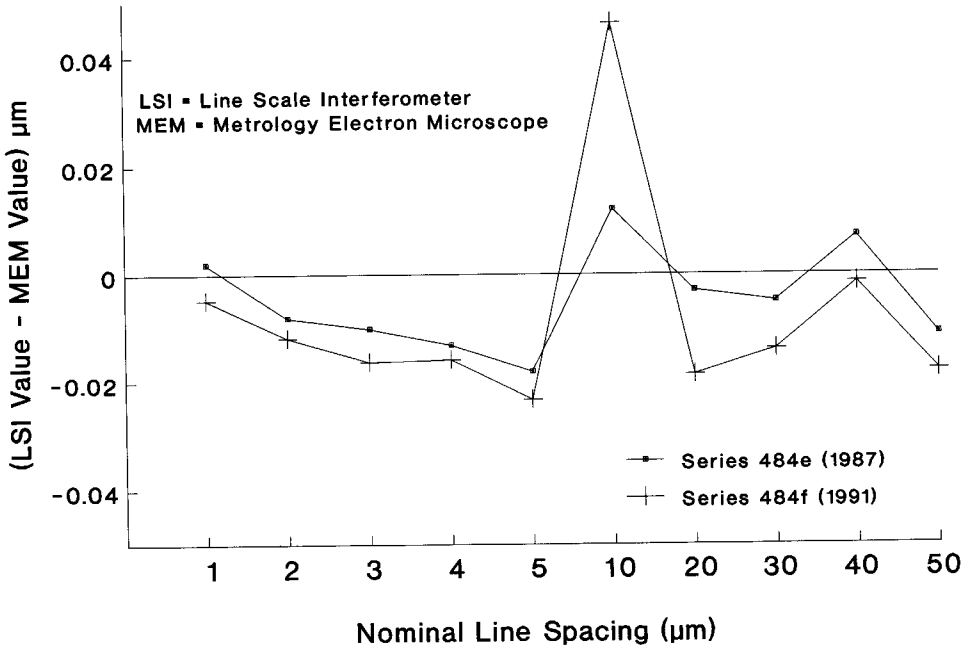


Fig. 9. Difference between measurements made by line-scale interferometer and SEM on the master sample.

uncertainty is reported as a statistical tolerance interval using an approximate Bonferroni [16] limit,

$$L = 2s + \frac{\nu^{1/2}}{\chi^2_{0.05}} \hat{\sigma}_{\text{lack}} Z_{p/2}.$$

The first term in L represents a confidence limit and the second term represents an upper bound due to $\hat{\sigma}_{\text{lack}}$ where $\chi^2_{0.05}$ is the lower 5 percent point of the χ^2 distribution with ν degrees of freedom, and ν, the degrees of freedom associated with $\hat{\sigma}_{\text{lack}}$, is obtained from the Welch-Satterthwaite approximation [17]. The quantity $Z_{p/2}$ is the upper $p/2$ percent point of the normal distribution with $p = 0.05$. The interpretation is that the interval defined by the certified value of spacing $\pm L$ provides coverage for at least 95 percent of the spacings within the certified region at a confidence level greater than or equal to 95 percent.

For SRMs without parallelism problems (issue 484f), the random component of uncertainty is reported as 3 standard deviations, $L = 3s_t$.

Historical uncertainties for all issues of SRM-484 are listed in Table 3. In all cases, the total uncertainty was reported as $\pm U$ where U is the linear sum of the random and systematic components of uncertainty or $U = L +$ systematic uncertainty. The systematic uncertainty is considered negligible for issues 484e and 484f.

As an example, we also consider the uncertainty that would be quoted for SRM-484f under the new NIST guideline [13]. As discussed above, lack of parallelism and the uncertainties arising from other systematic effects have been considered as negligible. The combined standard uncertainty u_c is then purely statistical and given by $u_c = s_t$. The quoted uncertainty under the new NIST guideline is then the expanded uncertainty $U = 2u_c = 2s_t$. Therefore, the expanded uncertainties U would be 2/3 of those given in Table 3 for SRM-484f.

7. Conclusion

It is essential for all SEM users to know the correct magnification of their instruments; the SRM-484 is a sample for calibrating the magnification scale of SEMs. The certified values of spacings between gold lines in a nickel matrix are measured with a metrology electron microscope and compared with line-scale interferometry. Properties of the SRM and the measurement system result in uncertainties of approximately 5% for 0.5 μm spacings and 1% for 50 μm spacings.

Table 3. Historical uncertainties for SRM-484

SRM issue	Nominal spacing (μm)	Uncertainty (μm)
SRM-484	1	±0.039
	2	±0.039
	3	±0.039
	5	±0.039
	50	±0.710
SRM-484a	1	±0.039
	2	±0.039
	3	±0.039
	5	±0.039
	50	±0.476
SRM-484b	1	±0.032
	2	±0.032
	3	±0.056
	5	±0.056
	50	±0.580
SRM-484c	1	±0.022
	2	±0.028
	3	±0.034
	5	±0.045
	10	±0.078
	20	±0.140
	30	±0.200
	50	±0.360
SRM-484d	1	±0.027
	2	±0.033
	3	±0.038
	5	±0.048
	10	±0.085
	20	±0.140
	30	±0.200
	50	±0.330
SRM-484e	1	±0.058
	2	±0.056
	5	±0.061
	10	±0.079
	30	±0.102
	50	±0.251
SRM-484f	0.5	±0.021
	0.5	±0.020
	1	±0.026
	3	±0.035
	5	±0.052

Acknowledgments

The authors would like to thank David B. Ballard, John Beers, Mike Postek, and Sam Jones for their assistance and discussion.

8. References

[1] D. B. Ballard, A procedure for calibrating the magnification of scanning electron microscope using NBS SRM-484, NBSIR 77-1248, Natl. Bur. Stand. (U.S.) (June 1977).

[2] J. S. Beers and K. B. Lee, Interferometric measurement of length scales at the National Bureau of Standards, Precision Eng. **4** (4), 205 (1982).

[3] S. Jensen and D. Swyt, Sub-micrometer length metrology: problems, techniques and solutions, Scanning Electron Microscopy, I SEM Inc., AMF O'Hare, Chicago (1980) p. 393.

[4] D. A. Swyt and S. W. Jensen, An electron-microscope-based system for accurate microdimensional measurements, Precision Eng. **3** (1), 11 (1981).

[5] G. G. Hembree, S. W. Jensen, and J. F. Marchiando, Submicrometer linewidth measurement by electron microscopy, G. W. Bailey (ed.), 38th Ann. Proc. Electron Microscopy Soc. Amer., San Francisco, CA. (1980) p. 312.

[6] A. W. Hartman, T. D. Doiron, and G. G. Hembree, Certification of NIST SRM 1961: 30 μm diameter polystyrene spheres, J. Res. Natl. Inst. Stand. Technol. **96** (5), 551 (1991).

[7] T. R. Lettieri, A. W. Hartman, G. Hembree, and E. Marx, Certification of SRM 1960: nominal 10 μm diameter polystyrene spheres, J. Res. Natl. Inst. Stand. Technol. **96** (6), 669 (1991).

[8] A. W. Hartman, T. D. Doiron, and J. Fu, Certification of NIST SRM 1962: 3 μm diameter polystyrene spheres, J. Res. Natl. Inst. Stand. Technol. **97** (2), 253 (1992).

[9] K. F. J. Heinrich, X-ray optics and microanalysis, 4th Intl. Cong. on X-ray Optics and Microanalysis, R. Castaing, P. Deschamps, and J. Philibert, eds., Hermann, Paris (1966), p. 1509.

[10] F. E. Scire and E. C. Teague, Piezodriven 50 μm range stage with subnanometer resolution, Rev. Sci. Instrum. **49** (12), 1735 (1978).

[11] M. T. Postek, R. D. Larrabee, and W. J. Keery, Scanning electron microscope linewidth measurement standards program at the National Bureau of Standards, EMSA Bulletin **17** (2), 59 (1987).

[12] Churchill Eisenhart, Harry H. Ku, and R. Collé, Expression of the uncertainties of final measurement results, NBS Special Publication 644, Natl. Bur. Stand. (U.S.) (Jan. 1983).

[13] B. N. Taylor and C. E. Kuyatt, Guidelines for evaluating and expressing uncertainty of NIST measurement results, NIST Technical Note 1297, Natl. Inst. Stand. Technol. (Jan. 1993).

[14] George W. Snedecor and William G. Cochran, Statistical method, 6th edition, The Iowa State University Press, Ames, Iowa (1967) p. 101–102.

[15] Franklin A. Graybill, An introduction to linear statistical models, Vol. 1, McGraw-Hill Co., New York (1961) p. 351–354.

[16] Rupert G. Miller, Jr., Simultaneous statistical inference, McGraw-Hill Co., New York (1966), p. 124.

[17] K. A. Brownlee, Statistical theory and methodology in science and engineering, 2nd edition, John Wiley & Sons, Inc., New York (1966) p. 236.

About the authors: *J. Fu is a materials engineer in the Precision Engineering Division of the NIST Manufacturing Engineering Laboratory. M. C. Croarkin is a mathematical statistician in the Statistical Engineering Division of the NIST Computing and Applied Mathematics Laboratory. T. V. Vorburger is a supervisory physicist in the Precision Engineering Division of the NIST Manufacturing Engineering Laboratory. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 2, March–April 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. **99**, 201 (1994)]

Letter to the Editor

New Values for Silicon Reference Materials, Certified for Isotope Abundance Ratios

Volume 99 | Number 2 | March–April 1994

P. De Bièvre and S. Valkiers

Institute for Reference Materials and Measurements[1],
Commission of the European Communities,
Joint Research Centre,
B-2440, Geel, Belgium

and

H. S. Peiser

National Institute of Standards and Technology,
Gaithersburg, MD 20899-0001

New isotope abundance and relative atomic mass (atomic weight) values—with low, hitherto unattained uncertainty—are reported for two previously described silicon reference materials using a well-known method with an improved isotope-ratio mass spectrometer. These new values are directly traceable to the SI, more specifically to the unit for amount of substance, the mole, and independent of the SI unit of mass and of the Avogadro constant. Besides the residual mass-spectrometric uncertainties, these new values depend in effect only on a recently published direct comparison of the cyclotron frequency in a Penning trap of $^{28}Si^+$ with that of $^{12}C^+$.

Key words: absolute abundances; atomic weights; cyclotron frequencies; isotope abundances; isotope ratio mass spectrometry; isotope reference materials; Penning trap; relative atomic masses; reference material; silicon; silicon tetrafluoride.

Accepted: February 10, 1994

Two new silicon isotope reference materials (RMs) have previously been announced [1]: IRM-017[2] in chips of a silicon crystal (distributed in samples of about 2 mmol) and IRM-018[2] in the chemical form of SiO_2 (distributed in samples of about 0.1 mol) [2]. An independent set of absolute measurements with an improved mass spectrometer [3] is now reported. Some further refinement of the gas mass-spectrometric measurement has been achieved. In that method SiF_3^+ ions generated from a synthetic mixture of highly enriched specimens of Si isotopes are compared with these ions from the RMs of proven internal homogeneity [4]. Improved data analysis has in the meantime also been developed and adopted for all "standard atomic weight" evaluations by the Commission on Atomic Weights and Isotopic Abundances of the International Union of Pure and Applied Chemistry [5].

The uncertainties of the measurements here described still depend mostly on the mass-spectrometrically derived isotope abundance ratios. However, these uncertainties now approach those of the best values of the relative atomic masses of the silicon isotopes. For these masses we have based our calculations on the recent direct comparison in a Penning trap of the cyclotron frequency of $^{28}Si^+$ with that of $^{12}C^+$ [6]. The derived ^{28}Si mass is uncertain by only 1×10^{-8} and only 3×10^{-9} higher than the value recently published by Audi and Wapstra [7] with a marginally lower estimated uncertainty by taking other measurements into consideration. Uncertainties in recognized values for the atomic masses of ^{29}Si and ^{30}Si are negligible because of their low abundances in terrestrial silicons.

Table 1 summarizes the new results. Their significance rests on:

[1] Formerly: Central Bureau for Nuclear Measurements.
[2] After the renaming of the Institute, these are now labelled IRMM-017 and IRMM-018.

Table 1. Isotopic composition and A_r(Si) for both RMs. Expanded uncertainties, U, are indicated under the digits to which they relate and are computed on a two standard deviation basis.

Abundance ratio	IRMM-017 New values	IRMM-017 Previous values [1]	IRMM-018 New values	IRMM-018 Previous values [1]
$n(^{29}Si)/n(^{28}Si)$	0.050 771 5 6 6	0.050 69 12	0.050 844 2 4 8	0.050 83 12
$n(^{30}Si)/n(^{28}Si)$	0.033 488 9 7 8	0.033 52 10	0.033 585 1 6 6	0.033 60 10
Amount (of substance) Fraction				
^{28}Si	0.922 287 7 8 6	0.922 33 14	0.922 144 0 7 0	0.922 14 14
^{29}Si	0.046 825 9 5 8	0.046 75 11	0.046 885 7 4 2	0.046 88 11
^{30}Si	0.030 886 4 7 0	0.030 92 8	0.030 970 3 5 8	0.030 98 8
Relative mean atomic mass (atomic weight) A_r(Si)	28.085 408 15	28.085 40 19	28.085 635 12	28.085 65 19

1) the reduction of the uncertainties by one order of magnitude;
2) the direct traceability of these values to the mole, the SI unit for amount of substance, involving only relative mass measurements of the enriched samples, without any absolute mass measurement based on the kilogram;
3) the ability with these RMs to compare reliably silicon specimens at the 10^{-7} level of relative atomic mass (atomic weight) A_r(Si), a level at which many geological sources and processed materials can be differentiated;
4) the reliable intercomparisons of these isotopic RMs with Si SRM 990 of NIST [8] enable the reduced uncertainty to be transferred to the latter without much loss. A value of $A_r(Si) = 28.085\ 538 \pm 0.000\ 018$ might be indicated compared with the NIST certified value from 1975 measurements of $A_r(Si) = 28.085\ 526 \pm 0.000\ 056$.

References

[1] S. Valkiers, P. De Bièvre, G. Lenaers, and H. S. Peiser, J. Res. Natl. Inst. Stand. Technol. **96**, 617–619 (1991).
[2] S. Valkiers, G. Lenaers, and P. De Bièvre, Characterization of Two Silicon Isotope Reference Materials: IRMM Internal Report GE/R/MS/11/89.
[3] P. De Bièvre, S. Valkiers, H. S. Peiser, F. Schäfer, and P. Seyfried, P.T.B.-Mitteil. in press (1994).
[4] P. Seyfried, P. Becker, A. Kozdon, F. Lüdicke, F. Spieweck, J. Stümpel, H. Wagenbreth, D. Windisch, P. De Bièvre, H. H. Ku, G. Lenaers, T. J. Murphy, H. S. Peiser, and S. Valkiers, Z. Phys. B **87**, 289–298 (1992).
[5] F. Schäfer, S. Valkiers, P. D. P. Taylor, P. De Bièvre, Int. J. Mass Spectrom. Ion Phys., in press (1994).
[6] R. Jertz, D. Beck, G. Bollen, J. Emmes, H.-J. Kluge, E. Schark, S. Schwarz, Schweighard, and P. Sonne, Physica Scripta **48**, 399–404 (1993).
[7] G. Audi and A. H. Wapstra, Nucl. Phys. **A565**, 1–65 (1993).
[8] I. L. Barnes, L. J. Moore, L. A. Machlan, T. J. Murphy, and W. R. Shields, J. Res. Natl. Bur. Stand. (U.S.) **79A**, 727–735 (1976).

About the authors: *P. De Bièvre is a PhD from Gent University and also Professor in Isotope Mass Spectrometry at Antwerpen University in addition to his position at IRMM. S. Valkiers received a degree in chemical engineering in 1976 from HIK College of Technology, Geel, Belgium. H. S. Peiser, now retired, was at one time Chief of the Crystal Chemistry Section at the National Bureau of Standards (now the National Institute of Standards and Technology). The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*