Journal of Research of the National Institute of Standards and Technology

Volume 99 Number 3 May–June 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 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

Contents

Articles

Errata

Conference Reports

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

News Briefs

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. **99**, 227 (1993)]

Sources of Uncertainty in a DVM-Based Measurement System for a Quantized Hall Resistance Standard

Volume 99 | Number 3 | May–June 1994

Kevin C. Lee, Marvin E. Cage, and Patrick S. Rowe

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

Transportable 10 kΩ standard resistors have become fairly widespread in industrial, university, and government standards laboratories because of their low temperature coefficient of resistance, ease of transportation, and convenient value. The values of these resistors, however, tend to drift with time, requiring periodic recalibration against an invariant standard such as the quantized Hall resistance. The availability of a simple, inexpensive measurement system for calibrating 10 kΩ resistors against such an invariant standard would be of great benefit to primary standards laboratories. This paper describes a simple automated measurement system that uses a single, high accuracy, commercially available digital voltmeter (DVM) to compare the voltages developed across a 10 kΩ standard resistor and a quantized Hall resistor when the same current is passed through the two devices. From these measurements, the value of the 10 kΩ standard resistor is determined. The sources of uncertainty in this system are analyzed in detail and it is shown that it is possible to perform calibrations with relative combined standard uncertainties less than 1×10^{-7} (0.1 ppm).

Key words: digital voltmeter; DVM method; electrical metrology; electrical reference standards; quantized Hall resistance; quantum Hall effect; resistance calibration.

Accepted: March 3, 1994

1. Introduction

Resistors composed of coils of wire wound around suitable forms have been used as standards of resistance for many years [1]. Such devices are even today widely used as working standards of resistance by primary and secondary standards laboratories in industry, university, and government. Due to aging of the wire and other effects, however, the values of these resistors tend to drift with time, requiring periodic recalibration against a known standard. Because wire-wound resistors drift with time, many national standards laboratories have adopted a standard of resistance based on the quantum Hall effect [2]. When a sample containing a thin, two dimensional conducting layer known as a two-dimensional electron gas (2 DEG) is cooled to liquid helium temperatures in the presence of a very strong magnetic field, the resistance of the device becomes quantized, assuming well defined values given by

$$R_H(i) = \frac{V_H(i)}{I} = \frac{h}{ie^2} \equiv \frac{R_K}{i} \qquad (1)$$

where V_H is the voltage across the Hall device, I is the current through the device, h is the Planck constant, e is the elementary charge, i is an integer, and R_K, the von Klitzing constant, has been defined by international agreement to be exactly 25 812.807 Ω for the purposes of practical electrical metrology [3]. The resistance is

time invariant, and is, under appropriate conditions of measurement, independent of the measurement conditions, such as current, temperature, and magnetic field. Because of these properties, the quantized Hall resistance, by international agreement, has been used since January 1, 1990 as a practical representation of the ohm.

The measurement systems in use at national standards laboratories are quite complex and are capable of achieving relative combined standard uncertainties [4] of 1×10^{-8} (or 0.01 ppm where 1 ppm = 1×10^{-6}) or less [5]. Many government and industrial standards laboratories do not need such small uncertainties in their work. Indeed, the uncertainty required by many laboratories is sufficiently large that the drifts in the values of wire-wound artifact standards are less than their measurement resolution and are therefore ignored. Such laboratories are well served by wire-wound artifacts which can be sent periodically to national standards laboratories to be calibrated or to participants in NIST's Measurement Assurance Program, or MAP. Generally, the relative combined uncertainty achieved in the MAP is about 0.1 ppm to 0.2 ppm.

Some laboratories, however, require smaller uncertainties. At such levels, drifts of the values of wire-wound artifacts require that they be frequently recalibrated. For these laboratories, the availability of a simple and inexpensive invariant standard of resistance would be of great benefit. The fairly recent introduction of high accuracy commercial digital voltmeters (DVMs) with 8 1/2 digit resolution has made it possible to conceive of a simple and fairly inexpensive measurement system [6] that would enable government and industrial standards laboratories to perform calibrations of their wire-wound resistors directly against a quantized Hall resistor with uncertainties of 0.1 ppm or less.

Such a calibration system has three distinct parts to it: a quantized Hall resistance device, a cryogenic system in which a superconducting solenoid and the Hall device are cooled to liquid helium temperatures, and a measurement system for comparing the standard resistor to this quantized Hall resistor. While the selection of the sample and cryogenic system are beyond the scope of this paper, a few words must be said about them, for they affect the design of the measurement system [7]. The sample and the cryogenic system must be such that the conditions for accurate measurement of the Hall resistance are met [8]. The conditions pertinent to this discussion are that the plateaus in the Hall voltage extend over as broad a range of magnetic field as possible, and that the voltage drop along the sample in the direction of the current flow, V_x, be as small as possible under the conditions of measurement.

While in theory any Hall plateau [any value of i in Eq. (1)] can be used for resistance calibration, in practice, the plateaus corresponding to higher values of i ($i > 4$) tend, except in unusual cases, to have values of V_x which are unacceptably large for precision resistance calibrations [9]. In general, the plateaus corresponding to lower values of i (i=4,2,1) occur at higher magnetic fields, are broader, and have values of V_x that are small enough to permit accurate resistance calibrations. The value of magnetic field at which any given plateau occurs is a function of the electron concentration in the 2 DEG, which in turn is a function of the design of the sample: samples can be designed to exhibit the i=4 plateau, for example, at very low fields of only a few tesla, or very high fields of 10 T or more. In choosing a sample design, one must balance several factors: if the plateau occurs at lower field, it will be more accessible with an inexpensive magnet and cryogenic system, but it will be narrower, and possibly V_x may be too large; on the other hand, if the plateau occurs at a high magnetic field, the plateau will be broader, V_x will be smaller, but the cryogenic system and magnet will have to be much larger, and consequently more expensive. It is the opinion of the authors that the optimum sample exhibits the i=4 plateau [$R_H(4)$=6 453.201 75 Ω] in a magnetic field range of 4.5 T to 6.5 T, and the i=2 plateau between 9.0 T and 13.0 T. These fields are easily attainable with commercially available superconducting magnets. The discussion of the measurement system in this paper therefore assumes that the resistance of the Hall device, $R_H(i)$, will be 6 453.201 75 Ω, 12 906.403 5 Ω, or 25 812.807 Ω, corresponding to the i=4, 2, or 1 plateau.

2. Description of Measurement System

This measurement system is shown schematically in Fig. 1. The standard resistor to be calibrated (R_S) is placed in series with the Hall device (R_H). A constant current is passed through both resistors, and the potential difference across each resistor is measured with the DVM. All measurements of resistance are four-terminal measurements. The current source and DVM are connected to the resistors using switches, so that the direction of current flow can be easily reversed, and the meter can be connected to either of the resistors. The measurements are performed in the order described by Marullo-Reedtz and Cage [5]. The potential drop V_R across the standard resistor R_S is first measured with the current flowing in the "positive" direction. The current direction is then reversed, and the potential drop across the standard resistor is measured again. These two

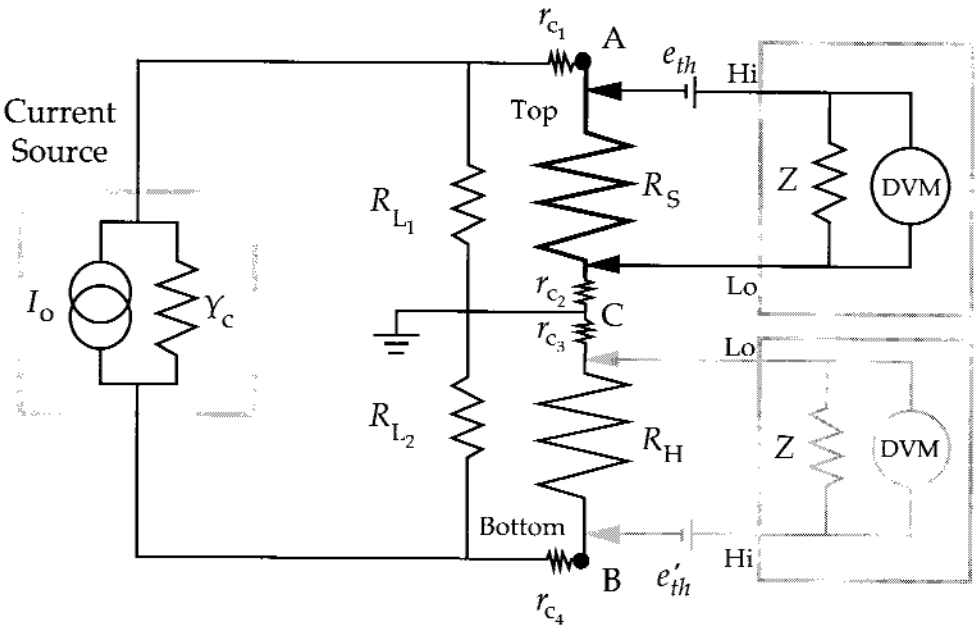

Fig. 1. DVM-based measurement system for comparing wire-wound standard resistors (R_s) with a quantized Hall resistor (R_H). R_{L_1} and R_{L_2} are the net leakage resistances to ground of the cables, current source, resistors, and the rest of the measurement system. The contact resistances are denoted by $r_{c_1}, \ldots, r_{c_4}$. The voltage across each resistor is measured with the digital voltmeter (DVM). The thermal voltages generated at the connections between the DVM and the resistors are denoted by e_{th} and e'_{th}. Z is the internal impedance of the DVM and Y_c is the Norton equivalent admittance of the current source.

measurements are repeated in the reverse order, to give a set of four values:

$$V_R(+I),\ V_R(-I),\ V_R(-I),\ V_R(+I). \tag{2}$$

The DVM is then connected to the Hall device, and this same sequence of four measurements is made on it twice. Finally, the standard resistor is measured again. The entire sequence of measurements is:

$$(1)\quad V_R(+I),\ V_R(-I),\ V_R(-I),\ V_R(+I) \Rightarrow \langle V_{R1}\rangle \tag{3a}$$

$$(2)\quad V_H(+I),\ V_H(-I),\ V_H(-I),\ V_H(+I) \Rightarrow \langle V_{H2}\rangle \tag{3b}$$

$$(3)\quad V_H(+I),\ V_H(-I),\ V_H(-I),\ V_H(+I) \Rightarrow \langle V_{H3}\rangle \tag{3c}$$

$$(4)\quad V_R(+I),\ V_R(-I),\ V_R(-I),\ V_R(+I) \Rightarrow \langle V_{R4}\rangle \tag{3d}$$

In all, a group of sixteen measurements is taken. The four individual values of each of these sets is averaged to yield a group of four values shown at the right. It will be shown later that this eliminates the effects of thermal voltages that are either constant or vary linearly with time. Finally, the two voltage drops across the standard resistor, $\langle V_{R1}\rangle$ and $\langle V_{R4}\rangle$ are averaged, as are the two measurements of the voltage drop across the Hall resistor, $\langle V_{H2}\rangle$ and $\langle V_{H3}\rangle$:

$$\langle V_R^{TOP}\rangle = \frac{\langle V_{R1}\rangle + \langle V_{R4}\rangle}{2}, \tag{4a}$$

$$\langle V_H^{BOT}\rangle = \frac{\langle V_{H2}\rangle + \langle V_{H3}\rangle}{2}. \tag{4b}$$

The above measurement sequence is then repeated with the positions of the Hall resistor and the standard resistor interchanged: i.e., the Hall resistor is connected in the top position of the circuit in Fig. 1, and the standard

resistor is placed in the bottom position. The average voltages in Eqs. (4a) and (4b) are then computed for this interchanged position, yielding $\langle V_R^{BOT} \rangle$ and $\langle V_H^{TOP} \rangle$.

In theory, if the current through the two resistors is constant,

$$\frac{\langle V_R^{TOP} \rangle}{\langle V_H^{BOT} \rangle} = \frac{\langle V_R^{BOT} \rangle}{\langle V_H^{TOP} \rangle} = \frac{R_S}{R_H} . \quad (5)$$

In practice, the situation is complicated greatly by a large number of factors, including thermally generated voltages, leakage resistances, contact resistances, instabilities in the current source, noise in the voltmeter, etc. It is shown in Sec. 3 that by application of the measurement procedure outlined above, the effects of several of these factors can be eliminated or minimized. Other factors cannot be eliminated by design of the measurement system, and their effects must be independently minimized. Some of these factors contribute to the error of the determination of the value of the standard resistor. At this point it should be noted that the value of the standard resistor R is usually expressed as a deviation from its nominal value, denoted R_S^{nom}:

$$\left(\frac{R_S - R_S^{nom}}{R_S^{nom}}\right) = \frac{R_S}{R_S^{nom}} - 1.$$

This ratio can be expressed in terms of the voltage ratios in Eq. (5) as:

$$\frac{R_S}{R_S^{nom}} - 1 = \frac{\langle V_R \rangle}{\langle V_H \rangle} \frac{R_H}{R_S^{nom}} - 1.$$

As a result, the correction factors and uncertainties in the ratio of the voltages in Eq. (5) must be divided by the ratio $\frac{R_S^{nom}}{R_H}$ in order to obtain the correction factors and uncertainties in the deviation of the resistor from its nominal value.

In this paper, relationships between the magnitudes of the various effects and the resulting error in the value of the standard resistor are determined. It is shown in Sec. 3 that if certain limits are placed on the magnitudes of the various systematic effects, the magnitudes of the factors required to correct for their effect on the value of the standard resistor are significantly less than the uncertainty in the value of the standard resistor due to random effects, and the corrections can be neglected while still maintaining a relative combined standard uncertainty of 0.1 ppm. In Sec. 4, the uncertainties due to random effects, such as nonlinear drifts in the output of the current source and in the gain and offset of the DVM, etc., are estimated. Limits are derived for the maximum values of these factors to ensure that the relative combined standard uncertainty in the value of the standard resistor is of the order of 0.1 ppm. It is shown that an uncertainty of 0.1 ppm or less can be obtained when calibrating resistors with R_S as much as a factor of 4 different from the quantized Hall resistance, $R_H(i)$ [this holds true whether the $i = 1$, 2, or 4 plateau is used for $R_H(i)$, i.e., whether the value of $R_H(i)$ is 25 812.807 Ω, 12 406.403 5 Ω, or 6 453.201 75 Ω].

3. Corrections Arising From Systematic Effects

Systematic effects that can contribute significantly to the error of the determination of the value of the standard resistor are associated with four main parts of the measurement system. These are:

i) the wires, cables, and switches used to connect the Hall resistor and the standard resistor;

ii) the Hall device and standard resistor;

iii) the current source used to supply the current through the resistors; and

iv) the DVM used to measure the voltages across the two resistors.

The magnitudes of the errors arising from each of these parts of the measurement system are estimated in this section.

3.1. Measurement System

In the measurement system shown in Fig. 1, the Hall resistor and the standard resistor to be calibrated are connected in series, and are connected, by means of cables and switches, to a current source. In principle, the same current flows through each resistor, so the ratio of the voltage measured by the DVM across the standard resistor and the Hall resistor should equal the ratio of the values of the resistors. In practice, however, there are thermally generated voltages in the wires, switch contacts, and various connections in the circuit which cause the measured voltages to differ from the actual voltage drops across the resistors. It is shown in Sec. 3.1.1 that the averaging technique described in Eq. (3) above eliminates the effect of thermal voltages that are either constant or vary linearly with time. Variable

contact resistances in the switches connecting the resistors to the current source can cause variations in the current supplied by the current source. In Sec. 3.1.2, an upper limit is derived for the magnitude of the permissible variations in the contact resistances. Leakage currents in the cables, switches, and other components of the system can also result in errors in the measurements of the voltages. The errors due to the leakage resistances can be corrected for, as described in Sec. 3.1.3.

3.1.1 Thermal Electromotive Forces Thermally generated voltages in the switch contacts, the wires, and the connections to the DVM result in erroneous determinations of the voltage drops across the resistors in the circuit. These thermal voltages are denoted by e_{th} and e'_{th} in Fig. 1, and arise when there are temperature variations between various parts of the circuit and when there are junctions between dissimilar materials, such as in switches or at solder connections. Thermal voltages are therefore generated at the connections between the current source and each resistor, and at the connections between the DVM and each resistor. The thermal voltages generated in the contacts to the current source add to the current produced by the current source. As is shown in the next section, if the equivalent admittance of the current source (Y_c in Fig. 1) is low enough, the current source will adjust its output to maintain a constant current through the circuit, and these thermal voltages have no effect on the measurement. Thermally generated voltages in the contacts to the DVM, on the other hand, are significant, and must be corrected for. Since the thermal voltages are a function of the temperature differences in the circuit, they do not change sign when the current is reversed. Thus, if one averages the voltage across one of the resistors measured with the current in one direction and the voltage across the same resistor obtained with the current in the opposite direction, the thermal voltage does not contribute to the average. Specifically, if the Hall resistor is in the BOTTOM position of the circuit of Fig.1, the voltage drop across it with the current flowing in one direction (which will be denoted as the "positive" direction) is

$$V_H^{BOT}(+I) = \frac{+I}{G_{L_2} + G_H} + e'_{th}, \qquad (6)$$

where $G_{L_2}=1/R_{L_2}$ is the leakage conductance in parallel with the Hall resistor (described in more detail in Sec. 3.1.3) and $G_H = 1/R_H$ is the Hall conductance. With the current flowing in the opposite direction (denoted as the "negative" direction) one obtains

$$V_H^{BOT}(-I) = \frac{-I}{G_{L_2} + G_H} + e'_{th}. \qquad (7)$$

The average value of the voltage across the Hall resistor is then independent of the thermal voltage:

$$\langle V_H^{BOT} \rangle = \frac{1}{2}\left[V_H^{BOT}(+I) - V_H^{BOT}(-I)\right] = \frac{I}{G_{L_2} + G_H}. \qquad (8)$$

It has been assumed that the magnitudes of the thermal voltages are the same and independent of the current direction. Practically, the current is reversed using switches, but the connections between the resistor and the DVM are not broken when the direction of the current is changed. Thus, the thermal voltages in the DVM contacts should not change when the current is reversed, and the assumption that the thermal voltages remain constant or vary linearly with time during the measurements should be a good one. It is also important to note that it was assumed that the current does not drift with time. This is discussed in more detail in Sec. 3.2.

3.1.2 Contact Resistances Contact resistances occur at all of the junctions in the system, including solder connections, switches, and other connectors in the circuit. In addition, the ohmic contacts to the 2 DEG in the quantized Hall device also exhibit contact resistance. Voltages develop across only those contacts that have current flowing through them. The effects of the resistances of the contacts between the resistors and the DVM are therefore minimized by making 4-terminal measurements of the voltages across each resistor, i.e., the voltage across a resistor is measured between two terminals that are separate from the terminals that carry the current.

The contact resistances $r_{c_1}, \ldots, r_{c_4}$ between the resistors and the current source shown in Fig. 1 change the total resistance of the circuit, but if they are constant and reproducible, as would be the resistances at solder connections, then it is apparent that they have no effect on the measurement. The resistances of the contacts in the switches connecting the resistors to the current source may change, however, for the switch contacts are constantly being opened and closed. If these contact resistances are *reproducible*, then they also have no effect on the total measurement. To see this, consider the case where the total contact resistance for one current direction always differs by an amount δZ from that for the opposite current direction. If the equivalent admittance of the current source is zero, then obviously this has no effect. In practice, however, the admittance of the

current source is nonzero, and this difference in contact resistance changes the total load impedance, and slightly affects the current through the resistors by an amount δI_L. In this case, the average in Eq. (8) is multiplied by the factor

$$\left(1 + \frac{\delta I_L}{2I_L}\right), \tag{9}$$

where I_L is the current through the resistors. If the difference δZ in the contact resistances is reproducible, then the averages of both the voltages across the Hall resistor and the standard resistor are multiplied by this same factor, and it cancels when ratios of the voltages are taken (as described in Sec. 2, and in more detail in Sec. 3.1.3).

If the contact resistances of the switches vary in a random, uncorrelated, and irreproducible manner every time they are opened and closed, each voltage has a different correction factor of the form given by Eq. (9). These correction factors do not cancel when voltage ratios are taken. As a result, it is necessary to keep $\delta I_L/I_L$ of the order of 10^{-8} or less to assure uncertainties of 1 part in 10^7 in the determination of the value of the standard resistor. This requirement places a limit on the variation in contact resistance δZ which depends on the admittance of the current source: the smaller the admittance, the larger the permissible variations in δZ. Specifically, if the admittance of the current source is Y_c, it can be shown that [10][1]

$$\frac{\delta I_L}{I_L} = \frac{-Y_c \delta Z}{[1+(R_H+R_S)Y_c]}. \tag{10}$$

If the admittance of the current source is about 50 μS, $R_H(i=4) = 6\,453.201\,75\ \Omega$, and R_S=10 kΩ,

$$\frac{\delta I_L}{I_L} = 2.7 \times 10^{-5}\ \delta Z,$$

and δZ must be less than 0.36 mΩ for $\delta I_L/I_L$ to be less than 10^{-8}. If a low admittance current source is used, δZ can be larger, but in general, the contact resistances in the switches and contacts of the circuit should be kept very small, and should be reproducible to within a few tenths of a milliohm, in order for them not to affect the accuracy of the measurement. This should not be difficult if care is taken to ensure that all of the contact surfaces are very clean and not covered with a thin oxide or a layer of organic contaminant.

[1] The current source is represented by its Norton equivalent, i.e., an ideal current source I_0 in parallel with an equivalent internal admittance Y_c. See Ref. [10] for more information.

The problem posed by the resistances of the contacts to the quantized Hall device is a vastly more subtle one. As described above, the *difference* between the resistances of the current carrying contacts to the 2 DEG with forward and reverse directions of current must be less than a few tenths of a milliohm. It is very difficult to produce contacts to the 2 DEG with such small contact resistances [11], but fortunately, the actual contact resistances need not be this low: the important point is that the contact resistances must be independent of the direction (and magnitude) of the current. Nevertheless, it is necessary that the current carrying contacts have contact resistances of less than 10 mΩ, or they generate substantial amounts of noise which prevents accurate resistance comparisons. One would think that the resistance of the contacts used to measure the Hall potential are unimportant, for no current flows through them. These contact resistances, however, must also be in the range of a few milliohms, for reasons which are beyond the scope of this paper [7, 12].

3.1.3 Leakage Resistances Leakage resistances arise from the noninfinite resistance of the electrical insulations used in constructing the system. As a result, the leakage resistance is distributed throughout the system: there are contributions from the current source, the cables, the DVM, and even the standard resistor and the wires leading to the quantized Hall resistor. The leakage resistances from the high terminal of the current source, the current reversal switch, the cables, the standard resistor, and the DVM have been combined in the idealized "leakage resistance" R_{L_1} shown in Fig. 1. The leakage resistances from the low terminal of the current source, cables, etc., and the quantum Hall resistor have been combined in the idealized "leakage resistance" R_{L_2}. The circuit is grounded between the Hall resistor and the standard resistor to minimize the effects of the leakage resistance between the low terminal of the DVM and ground: the low terminal of the DVM is always connected to this point throughout the entire measurement sequence. The high terminal is alternately connected to point A in Fig. 1 to measure the standard resistor, or point B to measure the Hall resistor. If it can be assumed that the leakage resistances R_{L_1} and R_{L_2} remain constant throughout the measurement cycle, and that the Hall resistor and the standard resistor have nominally the same values, it can be shown that the error due to leakage resistance is eliminated by averaging the ratio of the voltage across the standard resistor to the voltage across the Hall resistor with the resistors in the standard configuration shown in Fig. 1 and the same ratio obtained when the positions of the resistors in the measurement circuit are interchanged. With the resistors connected as shown in Fig. 1 (the standard position), the average

voltages across the resistors [as determined from Eq. (8)] are:

$$\langle V_R^{TOP}\rangle = \frac{I}{G_{L_1} + G_S} \tag{11a}$$

$$\langle V_H^{BOT}\rangle = \frac{I}{G_{L_2} + G_H}, \tag{11b}$$

where $G = 1/R$ for each of the resistors in Fig. 1. The ratio of these two voltages is then

$$\frac{\langle V_R^{TOP}\rangle}{\langle V_H^{BOT}\rangle} = \frac{G_{L_2} + G_H}{G_{L_1} + G_S}$$

$$= \frac{G_H}{G_S}\frac{1 + \frac{G_{L_2}}{G_H}}{1 + \frac{G_{L_1}}{G_S}} \approx \frac{G_H}{G_S}\left(1 + \frac{G_{L_2}}{G_H} - \frac{G_{L_1}}{G_S}\right). \tag{12}$$

If it is assumed that the leakage resistances are independent of the positions of the resistors, the positions of the Hall resistor and the standard resistor can be usefully interchanged. In this case, the Hall resistor is in parallel with R_{L_1} and the standard resistor is in parallel with R_{L_2}. The ratio of Eq. (12) is then

$$\frac{\langle V_R^{BOT}\rangle}{\langle V_H^{TOP}\rangle} = \frac{G_{L_1} + G_H}{G_{L_2} + G_S} \approx \frac{G_H}{G_S}\left(1 + \frac{G_{L_1}}{G_H} - \frac{G_{L_2}}{G_S}\right). \tag{13}$$

Averaging Eqs (12) and (13) yields:

$$\frac{1}{2}\left(\frac{\langle V_R^{TOP}\rangle}{\langle V_H^{BOT}\rangle} + \frac{\langle V_R^{BOT}\rangle}{\langle V_H^{TOP}\rangle}\right)$$

$$= \frac{G_H}{G_S}\left(1 + \frac{G_{L_1}+G_{L_2}}{2G_H} - \frac{G_{L_1} + G_{L_2}}{2G_S}\right). \tag{14}$$

If the standard resistor has exactly the same value as the Hall resistor, then the last two terms on the right side of Eq. (14) cancel, and the leakage resistances have no effect. If the resistors are not equal in value, then there is a small correction factor. If the leakage resistances change when the current is reversed or the resistors are interchanged, then the above analysis does not hold. Even in this case, however, the error in the determination of the value of the standard resistor caused by the leakage resistances will be of the order of the ratio of the leakage conductance $G_L = G_{L_1} + G_{L_2}$ to the conductance of the larger of the two resistances R_H or R_S. Typically, cables insulated with PTFE Teflon, and carefully constructed current sources will have leakage resistances greater than about $10^{12}\ \Omega$ ($G_L < 10^{-12}$ S). If $R_S = 10\ k\Omega$, and $R_H = 6\ 453.201\ 75\ \Omega$ or $12\ 906.403\ 5\ \Omega$, the correction to the value of the standard resistor will be of the order of 0.003 ppm, which is more than an order of magnitude less than the uncertainty due to random effects, and can therefore be neglected. For resistance ratios of 4 or more, as would result from the comparison of a $6\ 453.201\ 75\ \Omega$ and a $25\ 812.807\ \Omega$ resistor, however, the correction to the value of the standard resistor can be as large as 0.02 ppm, which is comparable to the uncertainty due to random effects and cannot be neglected. In this case, a correction factor can be estimated by measuring the leakage resistance between the point C in Fig. 1 and earth which will be approximately equal to $1/(G_{L_1}+G_{L_2})$ (in Fig. 1 point C is connected to earth, but for this leakage resistance measurement this connection must be removed).

The uncertainty associated with the assessment of this leakage resistance, however, will be quite large, possibly as large as the correction factor itself, so one can treat this correction factor as a component in the assessment of the combined uncertainty. If lower uncertainties are required, however, it will be necessary to increase the leakage resistance of the system another order of magnitude, something which is quite difficult to do.

3.2. Current Source

Several factors determine the optimum current for these measurements. The higher the current, the larger the voltages across the resistors, and the smaller the averaging time required to obtain voltage measurements with a given experimental standard deviation of the mean. Too large a current, on the other hand, can cause self-heating of the standard resistor, which will change its value and, more importantly, can cause breakdown of the dissipationless current flow in the Hall device, rendering it unusable for resistance calibrations [13]. Typically, currents between 10 μA and 50 μA satisfy these conditions. It should be noted that when performing calibrations of 10 kΩ resistors against a quantized Hall resistor, it is the maximum current that the quantized Hall device can sustain without breaking down that limits the maximum current at which measurements can be performed. Generally, this critical current is far below the current at which even a typical 10 kΩ resistor would start to show self-heating effects.

It should be recognized, however, that a reduction in averaging time due to a higher current is only realized if the primary factor limiting the accuracy of the voltage measurement is the signal-to-noise ratio and *not* the resolution of the meter. If the primary limiting factor is

the resolution of the meter, then the benefits of increasing the current are somewhat limited. For example, if the resolution of the meter is 10 nV on the 100 mV range, then one benefits from choosing a current that produces a voltage near the top of this range, as in this case the voltage measurement has the minimum relative uncertainty. Increasing the current so that the voltage is so large that the meter must use the next range, e.g., the 1 V range, however, may not result in any benefits if the resolution of the 1 V range is 100 nV. In this case, the relative uncertainty of the voltage measurement will be the same or worse than at the lower current. In practice, some benefit may be obtained by using the 1 V range in this example, for the internal DVM noise is often less on the higher voltage ranges than on the 100 mV range, so lower uncertainties in the measurement of the voltage can be achieved with shorter measurement times, even though the resolution of the meter is poorer on the 1 V range.

The method for determining the ratio of the value of the standard resistor to the Hall resistance described in Sec. 2 assumes that the current through the resistors is constant during the time that the voltage measurements are made. Any variations in the current, such as drifts or noise, will appear as drifts or noise in the voltages across the resistors and will affect the accuracy of the calibration of the standard resistor. In order to obtain a relative combined standard uncertainty in the calibration of 10^{-7}, it is necessary to keep the noise in the current source and nonlinear deviations in the current below 0.01 ppm; for currents between 10 μA and 50 μA, this dictates that the current variations be less than 0.1 pA to 0.5 pA. Such low noise levels are rather difficult to achieve with active current sources containing operational amplifiers, transistors, or other solid state components which usually have noise levels of the order of a few ppm, and are therefore generally unacceptable for use with this method. Current sources made using mercury batteries and current-limiting wire-wound resistors are capable of meeting these stringent requirements, even though the output of these current sources tends to decrease with time in a predictable, linear manner.

Fortunately, such stable linear drifts in current do not affect the measurements if the measurement sequence described in Sec. 2 is used. If the current decreases at a constant rate B, and the current at the beginning of the first measurement is I, then the current at a time t after the first measurement was begun is

$$I(t) = \mathrm{I} - Bt. \tag{15}$$

If each individual voltage measurement described in Sec. 2 takes a time Δt, then the current at the start of each voltage measurement in the group is given by the expression in column 2 of Table 1. As described above in Sec. 2, the first four voltage measurements (of the standard resistor with positive and negative current) are averaged to eliminate thermal voltages, yielding the value shown in the third column of Table 1. The average voltages across the standard resistor are then averaged, as are the average voltages across the Hall device, as described in Sec. 2, Eq. (4a), to give

$$\langle V_{\mathrm{R}}^{\mathrm{TOP}} \rangle = \frac{\langle V_{\mathrm{R1}} \rangle + \langle V_{\mathrm{R4}} \rangle}{2} \approx R_{\mathrm{S}} \left[I_0 - \left(\frac{30}{4} B \, \Delta t \right) \right]. \tag{16a}$$

$$\langle V_{\mathrm{H}}^{\mathrm{BOT}} \rangle = \frac{\langle V_{\mathrm{H2}} \rangle + \langle V_{\mathrm{H3}} \rangle}{2} \approx R_{\mathrm{H}} \left[I_0 - \left(\frac{30}{4} B \, \Delta t \right) \right]. \tag{16b}$$

The effective current is the same in both of these equations, and is eliminated when the ratio of V_{H} to V_{R} is taken. Note that the effective current in both of these equations is the current at the exact mid-point of the measurement cycle.

3.3 Digital Voltmeter (DVM)

The quality of the digital voltmeter used to measure the voltages across the resistors is the ultimate factor limiting the accuracy of this technique. In order to obtain resistance calibrations with relative combined standard uncertainties of 10^{-7}, the DVM must be capable of measuring voltages with uncertainties about a factor of 5–10 less than this. If a 20 μA current is used, the voltages across the resistors will be of the order of 200 mV, and the DVM must be capable of resolving voltages of the order of 0.2×10^{-7} V, or 20 nV [14]. Commercial DVMs are now available from several manufacturers that have such a high resolution. In addition to the high resolution, however, the DVM must have very high accuracy, a high degree of linearity, a high input impedance, high stability, and very low noise. Commercial 8 1/2 digit multimeters from several manufacturers are on the market that meet these specifications. In this section, various systematic effects

Table 1. Assuming a constant rate of change in the current produced by the current source and that each voltage measurement takes a time Δt, the current is calculated at the beginning of each voltage measurement on the standard resistor (R_S) and and the Hall resistor (R_H). Each set of four measurements on a resistor is averaged to eliminate thermal voltages and other constant offsets, resulting in the average voltages shown in column 3.

Resistor	I	Mean voltage
$R_S(+)$	$+I_0$	
$R_S(-)$	$-(I_0-B\Delta t)$	$R_S\left[I_0-\left(\frac{3}{2}\right)B\Delta t\right]$
$R_S(-)$	$-(I_0-2B\Delta t)$	
$R_S(+)$	$+(I_0-3B\Delta t)$	
$R_H(+)$	$+(I_0-4B\Delta t)$	
$R_H(-)$	$-(I_0-5B\Delta t)$	$R_H\left[I_0-\left(\frac{11}{2}\right)B\Delta t\right]$
$R_H(-)$	$-(I_0-6B\Delta t)$	
$R_H(+)$	$+(I_0-7B\Delta t)$	
$R_H(+)$	$+(I_0-8B\Delta t)$	
$R_H(-)$	$-(I_0-9B\Delta t)$	$R_H\left[I_0-\left(\frac{19}{2}\right)B\Delta t\right]$
$R_H(-)$	$-(I_0-10B\Delta t)$	
$R_H(+)$	$+(I_0-11B\Delta t)$	
$R_S(+)$	$+(I_0-12B\Delta t)$	
$R_S(-)$	$-(I_0-13B\Delta t)$	$R_S\left[I_0-\left(\frac{27}{2}\right)B\Delta t\right]$
$R_S(-)$	$-(I_0-14B\Delta t)$	
$R_S(+)$	$+(I_0-15B\Delta t)$	

associated with the DVM that contribute to the total uncertainty are analyzed. In Sec. 3.3.1, a correction term accounting for offsets and nonlinearities in the response of the DVM to applied voltages is derived. In Sec. 3.3.2, the effect of the small current source between the input terminals of the DVM is considered. The effect of noninfinite input impedance is considered in Sec. 3.3.3.

While this measurement system is quite similar to potentiometric measurement systems achieving smaller ultimate uncertainties [5], the accuracy, range, resolution, and linearity requirements on the DVM used with this system are greater. In the potentiometric measurement systems, a potentiometer is used to cancel the Hall voltage, so the detector is only used to measure very small differences between the Hall voltage and the voltage drop across the standard resistor being calibrated. Therefore, the detector need not have a very great range, but must have very low noise, high resolution, and high accuracy. Furthermore, since the detector is only measuring small deviations from zero, the linearity of the detector over large ranges is not crucial. In the measurement system described in this paper, the DVM is used to measure voltages that differ widely from zero, and that are both positive and negative. This requires the DVM to have a very high degree of linearity.

Offsets and nonlinear responses of the DVM can be determined by calibrating the DVM against a Josephson array. The Josephson array produces a time-invariant voltage that is related to fundamental constants, and, by international agreement, provides a practical metrological representation of the volt. If the Josephson array produces a defined voltage $\mathcal{V}$, the voltage indicated by the DVM will be:

$$V_{DVM} = A + g\mathcal{V} + N(\mathcal{V}), \tag{17}$$

where A is the offset, g is the gain of the DVM, and $N(\mathcal{V})$ is a nonlinear correction. For most modern high quality meters,

$$g = 1 + S, \tag{18}$$

where S is a small number. The values of A, g, and N should be determined by measuring V_{DVM} with applied array voltages in the neighborhood of the values expected to be encountered in the resistance measurements. A least-squares fitting procedure should be used to determine the gain, offset, and nonlinear corrections for *both* positive and negative voltages.

If the offset voltage A is the same for both positive and negative voltages, then the same averaging procedure that eliminates the thermal voltages will eliminate the offset voltage: the offset will cancel when $V(+I)$ and $V(-I)$ are subtracted, as in Eq. (8). In practice, however, neither A nor g need be the same for positive and negative voltages.

As described in Sec. 2, the ratio of the resistor values is determined from the arithmetic mean of the ratios of the averaged voltages across each resistor in the standard and interchanged position, as in Eq. (14). Each voltage in Eq. (14) must be corrected for the offset, nonunity gain, and nonlinearities in the DVM before the ratios are taken. A factor taking all of these corrections into account can be derived for Eq. (14) as follows: the voltage indicated by the DVM when it is connected to the Hall resistor in the bottom position in the circuit of Fig. 1, with positive current, is given by (neglecting thermal voltages)

$$V_H^{BOT}(+I) = A^+ + g^+\mathcal{V}_H^{BOT} + N\left(\mathcal{V}_H^{BOT}\right), \tag{19}$$

where $\mathcal{V}_H^{BOT}$ is the "true" voltage across the Hall resistor in the bottom position in the circuit of Fig. 1. Likewise, with the current reversed,

$$V_H^{BOT}(-I) = A^- - g^-\mathcal{V}_H^{BOT} + N\left(-\mathcal{V}_H^{BOT}\right), \tag{20}$$

Averaging as in Eq. (8) gives:

$$\langle V_H^{BOT}\rangle = \alpha + \gamma\, \mathcal{V}_H^{BOT} + \delta N_H^{BOT}\,, \tag{21a}$$

where $$\alpha = \frac{A^+ - A^-}{2}\,, \tag{21b}$$

$$\gamma = \frac{g^+ + g^-}{2}\,, \tag{21c}$$

and

$$\delta N_H^{BOT} = \frac{N\left(\mathcal{V}_H^{BOT}\right) - N\left(-\mathcal{V}_H^{BOT}\right)}{2}\,. \tag{21d}$$

If leakage resistances and other systematic effects are ignored, the ratio of the "true" voltages across the resistors is equal to the ratio of the resistors, i.e.,

$$\frac{1}{2}\left(\frac{\mathcal{V}_R^{TOP}}{\mathcal{V}_H^{BOT}} + \frac{\mathcal{V}_R^{BOT}}{\mathcal{V}_H^{TOP}}\right) = \frac{G_H}{G_S} = \frac{R_S}{R_H}\,. \tag{22}$$

Inverting Eq. (21a) above gives

$$\mathcal{V}_H^{BOT} = \frac{\langle V_H^{BOT}\rangle - \alpha - \delta N_H^{BOT}}{\gamma}\,. \tag{23}$$

Repeating the calculation for each of the voltages and substituting into the left-hand side of Eq. (22) yields

$$\frac{1}{2}\left[\frac{\left(\frac{\langle V_R^{TOP}\rangle - \alpha - \delta N_R^{TOP}}{\gamma}\right)}{\left(\frac{\langle V_H^{BOT}\rangle - \alpha - \delta N_H^{BOT}}{\gamma}\right)}\right.$$

$$\left.\frac{\left(\frac{\langle V_R^{BOT}\rangle - \alpha - \delta N_R^{BOT}}{\gamma}\right)}{\left(\frac{\langle V_H^{TOP}\rangle - \alpha - \delta N_H^{TOP}}{\gamma}\right)}\right] = \frac{R_S}{R_H}\,. \tag{24}$$

Generally, the corrections α and δN are very much smaller than the values of the voltages $\langle V\rangle$, so that the ratio of $\alpha+\delta N$ to $\langle V\rangle$ is of such small magnitude that terms of second order in this quantity can be neglected. Equation (24) can then be simplified to

$$\frac{R_S}{R_H} = \frac{1}{2}\left(\frac{\langle V_R^{TOP}\rangle}{\langle V_H^{BOT}\rangle} + \frac{\langle V_R^{BOT}\rangle}{\langle V_H^{TOP}\rangle}\right) +$$

$$\frac{1}{2}\left[\left(\frac{\langle V_R^{TOP}\rangle}{\langle V_H^{BOT}\rangle}\right)\left(-\frac{\alpha + \delta N_R^{TOP}}{\langle V_R^{TOP}\rangle} + \frac{\alpha + \delta N_H^{BOT}}{\langle V_H^{BOT}\rangle}\right) +\right.$$

$$\left.\left(\frac{\langle V_R^{BOT}\rangle}{\langle V_H^{TOP}\rangle}\right)\left(-\frac{\alpha + \delta N_R^{BOT}}{\langle V_R^{BOT}\rangle} + \frac{\alpha + \delta N_H^{TOP}}{\langle V_H^{TOP}\rangle}\right)\right]. \tag{25}$$

The first term on the right side of Eq. (25) is the ratio of the voltages across each resistor [averaged as in Eq. (3)], while the second term on the right side of Eq. (25) is the correction factor that must be applied to correct for nonzero offsets and nonlinearities in the DVM. This expression can be considerably simplified by noting that both α and δN are very much smaller than the voltages $\langle V\rangle$, so the denominators can be replaced by their nominal values. In addition, the δN are the *differences* between the nonlinearity corrections for positive and negative voltages, and since the magnitudes of the voltages across a given resistor are all essentially the same, little error will be incurred by making the approximation that δN is a constant for each resistor. Thus, with the approximations

$$\begin{aligned}
\langle V_R^{TOP}\rangle &\approx \langle V_R^{BOT}\rangle = \langle V_R\rangle \\
\langle V_H^{BOT}\rangle &\approx \langle V_H^{TOP}\rangle = \langle V_H\rangle \\
\delta N_R^{TOP} &\approx \delta N_R^{BOT} \approx \delta N_R \\
\delta N_H^{BOT} &\approx \delta N_H^{TOP} \approx \delta N_H \\
\frac{\langle V_R^{TOP}\rangle}{\langle V_H^{BOT}\rangle} &\approx \frac{R_S}{R_H}\,,
\end{aligned} \tag{26}$$

the correction term in Eq. (25) can be written

$$\frac{1}{\langle V_H\rangle}\left(\frac{R_S}{R_H}(\alpha + \delta N_H) - (\alpha + \delta N_R)\right). \tag{27}$$

It is important to remember that α is the *difference* between the offsets for positive and negative voltages, as given by Eq. (21) above. Usually, the offsets are negligibly small and are the same for positive and negative voltages, so that α in the above equation can be neglected. In this case, Eq. (27) can be written

$$\frac{1}{\langle V_H\rangle}\left(\frac{R_S}{R_H}\,\delta N_H - \delta N_R\right). \tag{28}$$

The quantity δN in the above equation is the *difference* between the nonlinearity correction for positive voltage and for negative voltage. This need not be zero and must be determined by calibrating the meter against a Josephson array. It should be noted that if both resistors have nominally the same value, then even though δN_R and δN_H may not be zero, they will have essentially the same value, and the correction term in Eq. (28) will vanish.

Corrections for the nonlinearity of the meter therefore need only be made in the case that the resistors being compared have nominally different values. In the case that the ratio of the resistances is 4, this correction term can be quite significant. Using the DVM calibration data in Fig. 2 of Cage et al. [6], and assuming R_S=25 812.807 Ω and R_H=6 453.201 75 Ω, the correction term of Eq. (28) would be 0.40 ppm ± 0.28 ppm. It is important to note that the correction to the value of the resistor R_S, expressed as a relative deviation from its nominal value, $(R_S-R_S^{nom})/R_S^{nom}$, due to nonlinearities in the DVM is obtained by dividing Eq. (28) by the ratio R_S^{nom}/R_H (see Sec. 2), which in the case of this example is 4. The correction to the value of $(R_S-R_S^{nom})/R_S^{nom}$ due to the DVM nonlinearity is therefore only 0.1 ppm ± 0.07 ppm in this example.

Since the nonlinearity correction term above may change with time, the DVM should be calibrated before each resistance calibration is performed. It should be noted that if either the offsets or the gain vary randomly with time, they will not cancel in the above equations, and will give rise to an uncertainty in the final determination of the value of the standard resistor as discussed in Sec. 4.

3.3.2 Internal Current Source Modern digital voltmeters inject a small current into the circuit to which their input terminals are attached. This current is often very small: for a Hewlett Packard 3458A DVM[2] it was measured to be 10^{-14} A (with no bias applied to the meter input terminals) by connecting a Keithley 602 electrometer directly across the input terminals. If this current does not change sign or magnitude with varying applied bias, it should have no effect on the average voltage measurements, as any offset produced by it would cancel when the voltages measured with positive and negative current are averaged. If it changes sign and magnitude with changing applied bias, but generally retains a magnitude in the range of 10^{-14} A, and if the measurement current is in the range of 10 μA to 50 μA, this small current would result in an error in the ratio of the resistors of only 0.0002 ppm to 0.0010 ppm.

3.3.3 Input Impedance Whenever a DVM with finite input impedance is connected across one of the resistors, some current will be shunted through the DVM. If the input impedance of the DVM is $Z_{DVM}=1/G_{DVM}$, and if the admittance of the current source is negligibly small (that is, the current source is assumed to be ideal), then the correction to the resistance ratio with the resistors in the standard position will be:

$$\frac{\langle V_R^{TOP}\rangle}{\langle V_H^{BOT}\rangle}=\frac{G_H+G_{DVM}}{G_R+G_{DVM}}=\frac{G_H}{G_R}\left(1+\frac{G_{DVM}}{G_H}-\frac{G_{DVM}}{G_R}\right)=$$

$$\frac{R_S}{R_H}\left(1+\frac{R_H-R_S}{Z_{DVM}}\right). \qquad (29)$$

In the event that the two resistors being compared are of nominally equal value, the correction factor in Eq. (29) will vanish. If however, the resistors have different nominal values, the factor must be evaluated, and can be quite appreciable, particularly if the input impedance of the DVM is less than 10^{12} Ω. In this case, if R_S=$R_H(i=1)$=25 812.807 Ω and $R_H(i=4)$=6 453.201 75 Ω (a 4-to-1 resistance ratio) the correction to the value of R_S will be as much as 0.019 ppm. In practice, the input impedances of modern DVMs are somewhat higher than 10^{12} Ω. The input impedances of two Hewlett Packard 3458A DVMs were measured using a method described by Cage et al. [6]. The impedances, measured at 22 °C with a relative humidity of 43%, were 5.9×10^{12} Ω for one DVM and 3.6×10^{12} Ω for the other DVM. Such input impedances would lead to only a 0.001 ppm correction to the resistance ratio when calibrating a 10 kΩ resistor against $R_H(i=4)$=6 453.201 75 Ω, and a 0.005 ppm correction to the resistance ratio when calibrating a 10 kΩ resistor against $R_H(i=1)$= 25 812.807 Ω.

4. Uncertainties Arising from Random Effects

The discussions in the previous sections have concerned only errors and sources of uncertainty in the determination of the value of the standard resistor due to systematic effects. As can be seen from the summary in Table 2, the errors associated with most of these systematic effects can be reduced to the level of 0.01 ppm by appropriate design of the measurement system. Random variations in the various parts of the measurement system, such as nonlinear drifts in the current, thermal voltages, contact resistances, and even the offset and gain of the DVM can also contribute to the uncertainty in the determination of the value of the standard resistor. Many of these have been discussed in Sec. 3 and their effects are summarized in Table 3. Appropriate design and construction of the measurement system, as noted in Sec. 3, generally results in contributions to the combined standard uncertainty from these effects of a few parts in 10^8. External noise picked up by the cables in

[2] Certain commercial equipment, instruments, or materials are identified in this paper in order to adequately specify 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.

the measurement system, and circulating currents arising from multiple ground connections in the circuit (commonly referred to as "ground loops") can also contribute to the combined standard uncertainty. Fortunately, the influence of these effects on the measurement can be greatly reduced by adequately shielding the measurement cables and equipment, preferably with two layers of shielding, and checking the circuit to ensure that there is only a single connection to ground (see Ref. [5]). Often this latter task is complicated considerably by the fact that some instruments have various internal connections to ground that are not obvious, so that while it may appear that the meter is isolated from ground, it in fact is either connected internally to ground through its power cable, or through the chassis, which may be screwed to a rack, which itself may be grounded.

Even if the effects of external noise, random drifts in the system, etc., can be eliminated, the fundamental factor limiting the accuracy of this technique is the internal noise and resolution of the DVM. Modern commercial 8 1/2 digit multimeters are available that have resolutions as fine as 10 nV out of 1 V, as mentioned in Sec. 3.3, so the DVM resolution does not really limit the accuracy of this technique. The internal noise in the DVM, however, can have rms values as high as 0.04 ppm to 0.1 ppm. This, coupled with random, long-term drifts in the gain, offset, and nonlinearity, are primarily responsible for limiting the accuracy of this technique.

In order to minimize the effect of such noise on the measurement, we must first examine the nature of the noise and how it affects the measurement. What we measure are the voltages across the resistors. These may be regarded as fixed voltages, upon which are super-imposed noise voltages, which can have different frequencies and amplitudes. Very high frequency noise, with a mean period less than the time required for a single measurement, will not greatly affect the measurement, for it will average to zero during the measurement time. Thus, the uncertainties in the voltage measurements in Eq. (3) can be greatly reduced by averaging each voltage measurement for a long time. This can be done by making repeated measurements of each voltage over an extended period of time and then averaging these measurements to produce a mean voltage. The uncertainty in this mean will be given by [see Ref. 14]

$$s = \sqrt{\frac{\sum_{i=1}^{N} (V_i - \langle V \rangle)^2}{N(N-1)}} \tag{30}$$

It would appear from Eq. (30) that the more measurements of each voltage are made, the smaller the uncertainty in the mean. This, however, is not the case, for in addition to the obvious high-frequency noise that gives rise to scatter in the individual measurements (each of which are assumed to be made over a short time interval of a second or less), there are long term drifts in the current, thermal voltages, and components of the measurement system. These drifts can be considered to be "noise" with very low frequency. In addition, there may be long-term, predictable drifts in the system: an example of this would be the slow, linear decrease in current produced by a mercury battery-powered current source as the batteries get depleted. Averaging a voltage measurement for longer periods of time will decrease

Table 2. Systematic effects that give rise to corrections to the value of the standard resistor determined using this method. If the value of each effect is kept within the limits shown in the second column, the value of the correction to the deviation of the value of R_S from its nominal value, given by the quantity $(R_S - R_S^{nom})/R_S^{nom}$, will be shown in the third column. The first value was calculated assuming that R_S=10 kΩ and $R_H(i=4)$=6 453.201 75 Ω; the second value was calculated assuming R_S=10 kΩ and R_H=25 812.807 Ω. Since the values of these corrections are generally very small in comparison with the uncertainty due to random effects, these correction factors are not applied, but for simplicity are included in the uncertainty of the final result.

Effect	Value	Absolute value of correction to result (in ppm)	Section in which calculated:
Leakage resistance	$>10^{12}$ Ω	0.004 to 0.016	3.1.3
DVM input impedance	$>10^{12}$ Ω	<0.003 to 0.015	3.3.3
DVM nonlinearity	<0.05 μV ±0.01 μV	0 to <0.10±0.07	3.3.1

Table 3. Randomly varying effects that contribute to the uncertainty in the determination of the value of the standard resistor. In the first row, the first number in the third column was calculated assuming R_S=10 kΩ and R_H=6 453.201 75 Ω; the second number was calculated assuming R_H=25 812.807 Ω.

Effect	Value	Standard uncertainty (in ppm)	Section in which calculated
Contact resistance reproducibility	<0.001 Ω	0.017 to 0.046	3.1.2
Nonlinear drifts in current source	$<10^{13}$ A	<0.005	3.2
DVM current source	$<10^{-12}$ A	0.000 to 0.001	3.3.2
Noise in DVM	5nV to 15 nV for 30 s measurements	0.015 to 0.046 for single group of measurements	4

the effects of higher frequency noise with periods less than the measurement time, but if these long-term linear drifts in current (for example) are present, there is some measurement time beyond which the effects of short term noise become negligible, and the primary contribution to the uncertainty given by Eq. (30) will be the linear drift in measurement current. For a drift of about 0.2 ppm in 10 min, this time can be as short as 10 h or as long as several days, depending on the magnitude of the noise voltage.

For optimum results, the individual voltage measurements indicated in Eq. (3) should be averaged for short enough times that long term drifts do not contribute to the uncertainty of the measurement at all. Then, as was shown in Sec. 3.2, appropriate averaging of sets of such measurements causes the effect of the long-term drifts to vanish. With modern 8 1/2 digit DVMs, averaging each of the voltage measurements in Eq. (3) for a period of about a minute results in uncertainties in the determination of the mean value of each voltage of between 5 nV and 15 nV, depending on the internal noise of the DVM. This results in a relative standard uncertainty in the determination of the resistance ratio from a single group of measurements of between 0.015 ppm and 0.046 ppm.

It is important to understand that this estimate of the uncertainty in the resistance ratio only considers the contributions from the noise, and does *not* include other random effects, such as irreproducible contact resistances, and various random drifts in the system. Therefore, while the uncertainty due to noise may be quite small, the *actual* value of the ratio may be considerably different from the value determined from a single group of measurements. In order to reduce the combined uncertainty due to all random effects in the determination of the resistance ratio, it is therefore necessary to perform numerous groups of measurements, and then average the resistance ratios r determined from each group to obtain a final average $< r_i >$:

$$< r > = \frac{1}{N} \sum_{i=1}^{N} r_i \, . \tag{31}$$

Because long-term drifts and other random effects tend to cause fairly large fluctuations in the values of the r_i determined from different groups (usually larger than the uncertainty in each r_i due to noise), the uncertainty in the final resistance ratio determined from Eq. (31) is determined from the set of group means $< r_i >$

$$s = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} (r_i - < r >)^2} \, . \tag{32}$$

The uncertainty calculated from Eq. (32) includes the effects of noise, and of random variations in the system, so there is no need to include these effects explicitly. The magnitude of the uncertainty will depend on the magnitudes of these various random effects. When the internal noise in the DVM is sufficiently low that each voltage measurement has an uncertainty on the order of 5 nV to 7 nV, noise in the DVM will actually be only a minor factor contributing to the final uncertainty; more significant contributions will come from the irreproducibility of the contact resistances and other factors shown in Table 3. If these random effects are kept within the limits shown in Table 3, however, it is usually possible to obtain relative uncertainties (due to random effects) in the determination of the mean resistance ratio [Eq. (31)] of less than 0.01 ppm after 20 groups of measurements.

5. Conclusion

Standards laboratories requiring resistance calibrations with relative combined standard uncertainties of less than 0.1 ppm could benefit from a simple, inexpensive, intrinsic resistance standard. While the quantum Hall effect provides such a standard of resistance, the measurement systems used at most national standards laboratories are far too expensive, complex, and time consuming to construct and use in government and industrial standards laboratories. This paper has analyzed the sources of uncertainty arising from both systematic and random effects in a simple quantized Hall resistance measurement system that uses a modern commercially available 8 1/2 digit digital voltmeter to compare the voltages developed across a standard resistor placed in series with a quantized Hall resistor. A measurement sequence has been presented which minimizes the effects of thermal voltages and linear drifts in the current on the final determination of the unknown resistance. Criteria have also been presented for minimizing the effects of the contact resistances and leakage resistances.

Table 2 summarizes the systematic effects that cause errors in the determination of the resistance using this technique. Most of these errors are so small as to be negligible in comparison with the uncertainty in the final resistance ratio due to random effects, and these corrections are therefore neglected and for simplicity are included in the uncertainty of the final result. In the case of the nonlinearity of the DVM, however, the error can be fairly large, and a correction factor [derived in Eq. (25)] must be applied to the final result. Table 3 summarizes the permissible magnitudes of random variations and drifts in the various components of the measurement system.

The combined standard uncertainty in the value of the standard resistor is the square root of the sum of the squares of the standard uncertainties arising from both random and systematic effects in the voltage measurements. If the uncertainty from random effects is less than 0.06 ppm (easily achievable with just a single group of measurements), the relative combined standard uncertainty in the determination of the resistance can be as low as 0.06 ppm if the DVM nonlinearity correction in Table 2 is negligible (including for simplicity the correction factors in Table 2 as sources of uncertainty). If the DVM nonlinearity correction is as high as that shown in Table 2, the combined uncertainty will be as high as 0.09 ppm.

If the uncertainty due to random effects is less than 0.01 ppm, which can be achieved by averaging up to 20 groups of data as described in Sec. 4, the relative combined standard uncertainty in the value of the standard resistor can be less than 0.03 ppm (again assuming the DVM nonlinearity correction to be negligible, as would be the case if the resistors being compared had the same nominal value). This DVM-based measurement system can, therefore, be used to compare wire-wound standard reference resistors with quantum Hall resistors with a relative combined uncertainty of less than 0.1 ppm, and in the most favorable cases, with uncertainties less than 0.06 ppm. Since the quantized Hall resistance does not drift with time, this measurement system can be used to calibrate wire-wound resistors, the values of which tend to drift with time.

6. Acknowledgements

The authors gratefully acknowledge the support of the Calibration Coordination Group of the Department of Defense for this work. They also wish to thank Drs. B. F. Field, B. N. Taylor, R. F. Dziuba, and A. F. Clark of NIST and Prof. C. A. Lee of Cornell University for helpful discussions.

7. References

[1] See, for example, J. L. Thomas, Precision Resistors and their Measurement, in Precision Measurement and Calibration: Electricity and Electronics, Natl. Bur. Stand. (U.S.) Handbook 77, Vol. I (1961), pp. 111–142.

[2] K. v. Klitzing, G. Dorda, and M. Pepper, New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance, Phys. Rev. Lett. **45**, 494-497 (1980).

[3] BIPM Proc.-Verb. Com. Int. Poids et Mesures **56**, 20 (1988).

[4] B. N. Taylor and C. E. Kuyatt, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, Natl. Inst. Stand. Technol. Tech. Note 1297 (January 1993).

[5] See, for example, G. Marullo-Reedtz and M. E. Cage, An Automated Potentiometric System for Precision Measurement of the Quantized Hall Resistance, J. Res. Natl. Bur. Stand. (U.S.) **92**, 303–310 (1987).

[6] M. E. Cage, D. Y. Yu, B. M. Jeckelmann, R. L. Steiner, and R. V. Duncan, Investigating the Use of Multimeters to Measure Quantized Hall Resistance Standards, IEEE Trans. Instrum. Meas. **40**, 262–266 (1991).

[7] For more information on the selection of samples, see W. van der Wel, E. G. Haanappel, J. E. Mooij, C. J. P. M. Harmans, J. P. André, G. Weimann, K. Ploog, C.T. Foxon, and J.J. Harris, Selection Criteria for AlGaAs-GaAs Heterostructures in View of Their Use as a Quantum Hall Resistance Standard, J. Appl. Phys. **65**, 3487–3497 (1989).

[8] F. Delahaye, Technical Guidelines for Reliable Measurements of the Quantized Hall Resistance, Metrologia **26**, 63–68 (1989).

[9] See, for example Fig. 6 in K. Jaeger, P. D. Levine, and C.A. Zack, Industrial Experience with a Quantized Hall Effect System, IEEE Trans. Instrum. Meas. **40**, 256–261 (1991).

[10] P. Horowitz and W. Hill, The Art of Electronics (Cambridge Univ. Press, Cambridge, England, 1980), p. 8.

[11] D. Jucknischke, H. J. Buhlmann, R. Houdré, M. Ilegems, M. A. Py, B. Jeckelmann, and W. Schwitz, Properties of Alloyed AuGeNi Contacts on GaAs/GaAlAs Heterostructures, IEEE Trans. Instrum. Meas. **40**, 228-230 (1990).

[12] H. Hirai and S. Komiyama, A Contact Limited Precision of the Quantized Hall Resistance, J. Appl. Phys. **69**, 655–662 (1990).

[13] M. E. Cage, R. F. Dziuba, B. F. Field, E. R. Williams, S. M. Girvin, A. C. Gossard, D. C.Tsui, and R. J. Wagner, Dissipation and Dynamic Nonlinear Behavior in the Quantum Hall Regime, Phys. Rev. Lett. **51**, 1374-1377 (1983).

[14] The uncertainty of the DVM measurement is $1/2\sqrt{3}$ of its resolution. The derivation of this result is given in Annex F2 of the Guide to the Expression of Uncertainty in Measurement (International Organization for Standardization, Geneva, Switzerland, 1993), p. 53.

About the authors: *Kevin C. Lee and Marvin E. Cage are physicists in the Electricity Division at NIST. Patrick S. Rowe, a student from the Massachusetts Institute of Technology, was a Guest Researcher in the Electricity Division when this work was done. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 241 (1994)]

A dc Method for the Absolute Determination of Conductivities of the Primary Standard KCl Solutions from 0 °C to 50 °C

Volume 99 | Number 3 | May–June 1994

Y. C. Wu, W. F. Koch, D. Feng, and L. A. Holland

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

and

E. Juhasz, E. Arvay, and A. Tomek

National Office of Measures, Budapest, Hungary

A new method for the absolute determination of electrolytic conductivity based on direct current and potentiometric measurements is described. The unique design of the cell uses a removable center section whose length and cross-sectional area are accurately known. Two pairs of matched Ag, AgCl electrodes are used in a four terminal mode of resistance measurement. Measurements of the electrolytic conductivity of primary standard potassium chloride solution using this novel dc conductance cell are compared with the currently adopted IUPAC and OIML recommendations. In addition, measurements have been made of the electrolytic conductivity of a solution of potassium chloride having a molality of 1 mol/kg (mole KCl per kilogram H_2O). The values so obtained over the temperature range of 0 °C to 50 °C are recommended as the new primary standards for electrolytic conductivity.

Key words: absolute measurement; conductance; electrolytic conductivity; potassium chloride; primary standards; resistance standard.

Accepted: April 20, 1994

1. Introduction

The measurement of conductivity is based on the relation

$$\kappa \equiv 1/\rho = G/R, \qquad (1)$$

where κ is the electrolytic conductivity (specific conductance) of an electrolyte solution; ρ is the resistivity; $G \equiv l/A$ is the cell constant of a cell having a length l between electrodes of area A; and R is the resistance. Of these three quantities, G is predetermined, and R is readily measurable, so that κ of an unknown solution can be determined. The determination of G is either by mechanical means (absolute measurement of the physical dimensions) or through an accurately and independently known κ (calibration measurement). When the primary standards for aqueous electrolytic conductivity are determined absolutely, the prime concern is accuracy; therefore an optimization of the three quantities is necessary. From Eq. (1) it is clear that when the κ to be determined is large, then either G needs to be large or R, the measured resistance will be small. However, there is a limitation on how large or how small either G or R can be, because the accuracy of both G and R depends upon the measuring instruments. Another practical restriction for G is space (i.e., the size of the controlled-temperature bath and the limit of the size of the connector) which limits the variation in G to a factor of about three (i.e., l greater than 30 cm is impractical). Two methods can be chosen for determining R: an ac bridge or a dc digital voltmeter (DVM). For solutions of low conductivity where R

is high, the ac method is preferred because of lower experimental uncertainties. For solutions of high conductivity (low R), the dc method offers several advantages. Of these, the most obvious is the avoidance of the complications due to capacitance effects in the ac circuit. In addition, the lead resistance in a high conductivity (low resistance) ac measurement is a significant quantity, whereas for a four terminal dc measurement, the lead resistance is eliminated. The dc method was originated at the National Office of Measures in Hungary and was carried out at NIST.

The conductivities of a 1 demal[1] (D) solution and a solution of potassium chloride having a molality[2] of 1 mol/kg are relatively high ($\kappa > 10$ S/m) and, therefore, it is desirable to use the dc method to determine the relatively lower R (R about 100 Ω for $G = 1000\ \mathrm{m}^{-1}$). The operating principle of the direct current method is based on Ohm's law ($E = IR$). The potential difference of an IR drop can be measured accurately to 10^{-7} V to 10^{-8} V with a calibrated voltmeter. It is important that the current be kept relatively low to avoid heating the solution (I^2R) and to minimize electrolysis (Faraday's law). In addition, reversible electrodes are required to eliminate polarization effects. For KCl solutions, Ag, AgCl electrodes are suitable.

2. The Cell Design and the Assembly

For an absolute determination of conductivity the cell constant needs to be determined with basic standards, (e.g., length measurement). To accomplish this requirement, a precision bore pyrex tubing was cut into three pieces, in order that the center section of the cell could be removed and that the internal diameter and length of this center section could be measured accurately by physical methods. The two ends of the center section were polished to flatness, while the surfaces of the ends of the other two pieces which will join with the center section were ground to a roughness of about 25 μm. In this way, a slight gap was formed between the joined ends of the tubing to effect a miniscule flow of solution, when the sections are rejoined using a nylon union. The other two ends of the cut tubing are fused to pyrex tubing comprising the current compartments of the cell. A U-shaped capillary separates the current-electrode chamber from the connecting chamber to minimize the effect of electrolysis products on the measurement. A third tube is joined at each of the junctions of the center section with the end sections using a nylon-union tee. These compartments contain the potential leads for the measurements. The final assembly is shown schematically in Fig. 1.

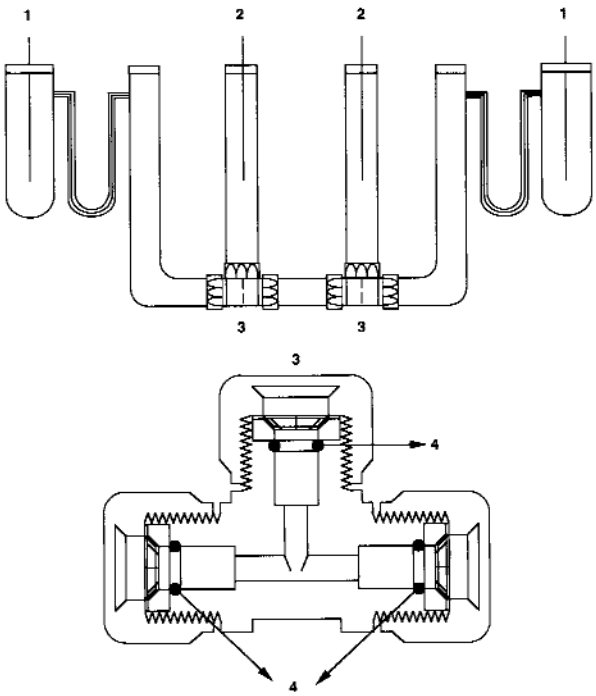

Fig. 1. Schematic representation of the dc conductance cell.
1 – Pt or Ag-AgCl electrode
2 – Ag-AgCl or calomel electrode (double junction)
3 – nylon union tee
4 – O-ring

The length, l, and internal diameter, d, of the center section were measured precisely at 25 °C by the Fabrication and Technology Division at NIST. Accordingly, at the 95 percent level of confidence, they are:

$$l = (19.8500 \pm 0.0002)\ \mathrm{cm}$$

$$d = (0.90010 \pm 0.00008)\ \mathrm{cm},$$

[1] The demal (D) is a designation unique to electrolytic conductivity and is defined at only three points: a 1 D solution, as adopted by OIML, is defined as 71.1352 g KCl per kg solution; a 0.1 D solution is defined as 7.41913 g KCl per kg solution; and a 0.01 D solution is defined as 0.745263 g KCl per kg solution.

[2] Molality is a designation in common usage in physical chemistry and thermodynamics. The molality of solute B is the amount of substance of solute B in a solution divided by the mass of the solvent. A solution of potassium chloride having a molality of 1 mol/kg is prepared by dissolving 1 mole of KCl in 1 kg of H_2O.

whereupon,

$$A = (0.63631 \pm 0.00016)\ \text{cm}^2, \text{ and}$$

$$G \equiv l/A = (31.195 \pm 0.0053)\ \text{cm}^{-1}.$$

The temperature effect on the cell constant is similar to that reported previously [1], i.e.,

$$dG/dt = -\alpha_g \cdot G, \tag{2}$$

where α_g, the thermal expansion coefficient of pyrex glass, is equal to 3.6×10^{-6}/°C. Thus

$$G(t) = G(25\ °\text{C}) + (dG/dt)(t - 25\ °\text{C}) \tag{3}$$

or

$$G(t) = 31.195\ \text{cm}^{-1} - 0.000112\ \text{cm}^{-1}(t - 25\ °\text{C}).$$

The effect is less than 0.01% in the range of temperature of interest which is smaller than the uncertainty of the mechanical measurements. The value $G(25\ °\text{C}) = 31.195\ \text{cm}^{-1}$ is used throughout the experiments without corrections.

3. Experimental

3.1 Apparatus

A high precision digital voltmeter (DVM), an accurate standard resistor, and a constant current supply are needed for the measurement of R. Both the DVM and the standard resistor have been calibrated by the Electricity Division at NIST. The constant current supply is capable of delivering current from 25 μA to 250 mA. Its stability is better than ±0.005% over 24 h. A schematic diagram of the operating circuit is shown in Fig. 2.

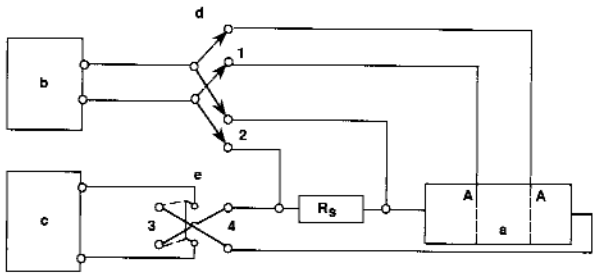

Fig. 2. Schematic drawing of the operating and switching circuitry: a) The cell with gaps A,A, b) DVM, c) constant current supply, d) and e) double pole double throw switches; R_s) standard resistor.

With switch "d" in the "2" position, and switch "e" in the "3" position, the current flows through the standard resistor and the conductance cell, and the *IR* drop across the standard resistor is measured by the DVM as $E_s = IR_s$. *I* is thus determined. Moving switch "d" to the "1" position while switch "e" remains in the "3" position effects the measurement of the voltage between the potential electrodes in the conductance cell with current flowing, i.e., $E_c = IR_c$. (The subscripts s and c refer to the standard resistor and the conductance cell, respectively.) Switching "e" to the "4" position reverses the flow of current and the polarity of the current electrodes. If there is any disparity between the potential electrodes in the conductance cell, it will manifest itself in the voltage measurement by reversing the current. Therefore, the constancy of the current is established. Hence, a data point at a single current consisted of four measurements as shown in Table 1.

Table 1. Measurement Sequence

Measurement No.	Switch "d"	Switch "e"
1	2	3
2	1	3
3	1	4
4	2	4

Two pairs of matched Ag, AgCl electrodes of the type used in the emf/pH research effort at NIST were used in the conductance cell, one pair for the current electrodes, the other for the potential electrodes. The preparation and calibration of these electrodes have been described in conjunction with the pH studies [1]. All four electrodes were within ±0.05 mV of each other. Ag, AgCl electrodes were used instead of Pt electrodes, even for the current electrodes, to minimize or eliminate the effects of electrolysis and polarization.

Three types of constant temperature baths were used in this study: oil, water, and ice. The oil bath has been described elsewhere [2]. An old water bath was modified to make more room for the conductivity cell. Its temperature was controlled to ±0.001 °C to ±0.002 °C for temperatures above 0 °C. At 0 °C, an ice bath was used.

3.2 Materials

SRM 999 KCl was used to prepare the solutions. It was ignited at 500 °C for 4 h and stored in a desiccator before use. Distilled and deionized water with conductivity <0.2 μS/cm was employed. Solutions were prepared by weight and buoyancy corrections were made.

3.3 Procedure

The glassware of all sections of the cell was cleaned with chromic acid, then, HCl, rinsed with water, and soaked in water overnight. After the sections were dried, the cell was assembled. Solution was added and the electrodes were placed in the appropriate chambers. The whole cell assembly was put into the bath, and its temperature was adjusted and controlled at the designated point. Usually, the starting temperature was 25.000 °C, and a steady state was reached after about 30 min. The wiring to the electrodes was connected according to the scheme shown in Fig. 2, and the measuring sequence commenced. It took approximately 3 s for a reading. There were four readings for a selected current, two each for each direction (E_s and E_c) as described above. Each data point was taken as the mean of three different currents. At each temperature, the current flowed through the cell for approximately 30 s. Thus, the power dissipated in the cell was less than 1 mW, and its effect was much smaller than a millidegree per second change for the solution. Therefore, there was no significant heating effect.

3.4 Results

According to Eq. (1), the determination of the conductivity of the solution requires the measurement of resistance, R. For the dc method, R is determined by Ohm's law as E/I. The current, I, is determined by means of a standard resistor, R_s, where E, the potential difference (or voltage drop) is read from a digital voltmeter, DVM. The results for R and the electrolytic conductivity of a 1 D solution and a solution of potassium chloride having a molality of 1.0 mol/kg at various temperatures are listed in Tables 2 and 3; the temperature coefficients are also given.

Table 2. Electrolytic conductivity values for a solution of potassium chloride having a molality of 1.0 mol/kg at temperatures from 0 °C to 50 °C

		κ/(S/cm)		
t/°C	R/Ω[a]	Observed	Smoothed[b]	Dev. %[c]
0	491.32	0.063 493	0.063 487	0.009
5	433.17	0.072 016	0.072 030	−0.020
10	385.96	0.080 825	0.080 844	−0.023
15	347.06	0.089 884	0.089 900	−0.018
20	314.54	0.099 175	0.099 170	0.005
25	287.14	0.108 64	0.108 62	0.019
30	263.83	0.118 24	0.118 24	0.000
35	243.74	0.127 99	0.127 97	0.014
40	226.40	0.137 79	0.137 81	−0.017
45	211.16	0.147 73	0.147 72	0.005
50	197.82	0.157 69	0.157 67	0.015

[a] Each R value is the mean of six values measured at different currents with a standard deviation of the mean no greater than 0.035%.
[b] Equation: $\kappa/(\mathrm{S/cm}) = 0.063\,488\,2 + 1.679\,13\times10^{-3}t + 6.007\,81\times10^{-6}t^2 - 3.837\,02\times10^{-8}t^3$.
[c] Dev. $\% = 100\,(\kappa-\kappa_{\mathrm{smoothed}})/\kappa_{\mathrm{smoothed}}$.

Table 3. Electrolytic conductivity values for 1 D KCl solution at temperatures from 0 °C to 50 °C

		κ/(S/cm)		
t/°C	R/Ω[a]	Observed	Smoothed[b]	Dev. %[c]
0	478.95	0.065 132	0.065 135	−0.005
5	422.44	0.073 845	0.073 860	−0.021
10	376.43	0.082 870	0.082 871	−0.002
15	338.56	0.092 140	0.092 136	0.005
18	318.95	0.097 805	0.097 804	0.001
20	306.97	0.101 62	0.101 62	0.001
25	280.28	0.111 30	0.111 30	−0.002
30	257.58	0.121 11	0.121 11	−0.002
35	238.10	0.131 02	0.131 05	−0.026
40	221.18	0.14 104	0.141 08	−0.031
45	206.43	0.151 12	0.151 15	−0.022
50	193.40	0.161 29	0.161 24	0.033

[a] See footnote "a" to Table 2.
[b] Equation: $\kappa/(\mathrm{S/cm}) = 0.065\,135 + 1.714\times10^{-3}t + 6.414\,1\times10^{-6}t^2 - 4.502\,8\times10^{-8}t^3$.
[c] See footnote "c" to Table 2.

4. Discussion

The values shown in Tables 2 and 3 are the results of three series of measurements. In each series the cell was disassembled, cleaned, dried and reassembled. In each of the processes, the gaps between the ends of the center tube and the end tubes could be affected. This limits the reproducibility to ±0.01%.

Additional sources of uncertainty are derived from (1) the cell constant, (2) the solution, and (3) the electrical measurement. These sources are examined as follows:

4.1 The Cell Constant

There are three limitations that restrict the improvement of accuracy, viz., the size of the precision bore tubing, the union-tee fitting that holds the tubing together, and the accuracy of the measuring instrument. For the latter, the uncertainty (2σ estimate) of the dimensional measurement at the Fabrication and Tehnology Division at NIST is 1 μm. Therefore, a larger bore and longer tubing would be needed to reduce the relative uncertainty. However, at present, the largest commercially available nylon union-tee limits the outside diameter of the tubing to 1.25 cm. This alone limits the reduction of the relative uncertainty for the cell constant. In this study, the cell constant from 11 measurements was $(31.195 \pm 0.0053)\ \mathrm{cm}^{-1}$, at the 95% level of confidence.

4.2 The 1 D KCl Solution and KCl Solution Having a Molality of 1.0 mol/kg

The KCl and water have been proved to be adequately pure [2]. The relative expanded uncertainty (2σ estimate) associated with the weighing was 0.002%; the balances were calibrated and checked with NIST weights. Another source of uncertainty was the evaporation of solution in the transferring process and in rotating the cell to get rid of the trapped air bubble(s). The relative expanded uncertainty (2σ estimate) associated with the preparation of the solution is estimated to be 0.003%.

4.3 Electrical Measurements

In this category, there should only be negligible uncertainties (2σ estimate) from systematic effects (0.001%), because the constant current source is very stable (approximately 0.0005%), and the DVM (uncertainty specified by the manufacturer 0.0005%) and the standard resistor have been calibrated with primary standards, as discussed above.

To sum up: the relative expanded uncertainty at the 95% level of confidence is estimated to be $U = 0.018\%$.

5. Conclusion

The electrolytic conductivity values for the 1 D solution and the solution of potassium chloride having a molality of 1.0 mol/kg have been determined absolutely from 0 °C to 50 °C in intervals of 5 °C by the dc method.

These values are listed in Table 4, together with the 1 D values recommended by the Organisation Internationale de Metrologie Légale (OIML) [4] and converted to ITS-90 [3]. They agree to ±0.02%, which is within the experimental uncertainty.

Table 4. Recommended electrolytic conductivity values for solutions of potassium chloride having a molality of 1.0 mol/kg, and for 1 D solutions

	κ/(S/cm)		
t/°C	KCl solution 1.0 mol/kg	KCl solution 1 D (present work)	KCl solution 1 D (OIML (4))
0	0.063 487	0.065 135	0.065 144
5	0.072 030	0.073 860	
10	0.080 844	0.082 871	
15	0.089 900	0.092 136	
18		0.097 804	0.097 82
20	0.099 170	0.101 62	
25	0.108 62	0.111 30	0.111 32
30	0.118 24	0.121 11	
35	0.127 97	0.131 05	
40	0.137 81	0.141 08	
45	0.147 72	0.151 15	
50	0.157 67	0.161 24	

Because of the improvement in accuracy, the enhancement of the temperature range from 0 °C to 50 °C, and the inclusion of the molality scale, the values listed in Table 4 are recommended as the primary electrolytic conductivity standards.

6. References

[1] Y. C. Wu, W. F. Koch, and G. Marinenko, J. Res. Nat. Bur. Stand. **89**, 395 (1984); Y. C. Wu, W. F. Koch, and R. A. Durst, NBS Special Publ. 260-53 (1988).
[2] Y. C. Wu and W. F. Koch, J. Soln. Chem. **20**, 391 (1991). Y. C. Wu, W. F. Koch, and K. W. Pratt, J. Res. Natl. Inst. Stand. Technol. **96**, 191 (1991).
[3] Y. C. Wu, W. F. Koch, W. J. Hamer, and R. L. Kay, J. Soln. Chem. **16**, 985 (1987); *ibid*. **19**, 1053 (1990).
[4] Organization Internationale de Metrologie Legale (OIML). International Recommendation No. 56, June 1980.

About the authors: *Young Chi Wu is a senior research chemist in the Electroanalytical Research Group, Inorganics Analytical Research Division (IARD), Chemical Science and Technology Laboratory (CSTL) at NIST. Lisa A. Holland is a summer employee in the Electroanalytical Research Group, and a graduate student in the Department of Chemistry, University of North Carolina at Chapel Hill. William F. Koch is the Chief of the Inorganic Analytical Research Division. Daming Feng was a Guest Scientist in the Electroanalytical Research Group from the Guangzhou Research Institute of Nonferrous Metals, Wushan, Guangzhou, People's Republic of China. Endre Juhàsz (retired) and Anna Tomek (retired) are former Heads of the Chemistry Department, and Eva Arvay (deceased) was a senior research chemist at the National Office of Measures, Budapest, Hungary. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 247 (1994)]

Three-Axis Coil Probe Dimensions and Uncertainties During Measurement of Magnetic Fields from Appliances

Volume 99 Number 3 May–June 1994

Martin Misakian and Charles Fenimore

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

Comparisons are made between the average magnetic flux density for a three-axis circular coil probe and the flux density at the center of the probe. The results, which are determined assuming a dipole magnetic field, provide information on the uncertainty associated with measurements of magnetic fields from some electrical appliances and other electrical equipment. The present investigation extends an earlier treatment of the problem, which did not consider all orientations of the probe. A more comprehensive examination of the problem leaves unchanged the conclusions reached previously.

Key words: appliance; coil probe; magnetic field; measurement; measurement uncertainty; power frequency.

Accepted: April 13, 1994

1. Introduction

This paper reconsiders a problem related to the measurement of power frequency magnetic fields from electrical appliances using three-axis circular coil probes. Specifically, it reexamines the differences between the average magnetic flux density as determined using a magnetic field meter with a three-axis circular coil probe and the magnetic flux density at the center of the probe, B_0, assuming the field is produced by a small loop of alternating current, i.e., a magnetic dipole. The "average" arises as a consequence of the averaging effects of the coil probes over their cross sectional areas when placed in a nonuniform magnetic field. The differences between the average magnetic field and B_0 can be regarded as measurement errors because the center of the probe is normally considered the measurement location. The magnetic dipole field is chosen as the relevant field because its geometry provides a good approximation of the magnetic field produced by many electrical appliances [1].

The average magnetic flux density measured by a three-axis magnetic field meter, B_{av3}, is also referred to as the resultant magnetic field and is defined as [2]

$$B_{av3} = \sqrt{B_1^2 + B_2^2 + B_3^2}, \quad (1)$$

where B_1, B_2, and B_3 are average root-mean-square (rms) magnetic field components determined by each of three orthogonally oriented coil probes.

Differences between B_{av3} and B_0 are calculated as a function of r/a where r is the distance between the magnetic dipole and the center of the probe, and a is the radius of the three-axis probe. In addition, differences between B_{av3} and B_0 are examined for different orientations of the magnetic dipole and rotations of the three-axis probe. Because the relative orientation of the dipole and three-axis probe is not known during most measurement situations, there is a distribution of possible differences between B_{av3} and B_0, and these differences

collectively represent a source of measurement uncertainty for a given r/a. What will be of interest in this paper is the largest difference that occurs between B_{av3} and B_0 as a function of r/a (for all possible orientations of the dipole). This largest difference is designated ΔB_{max3}.

This investigation extends an earlier treatment of the problem which considered different orientations of the dipole, but not all possible orientations of the three-axis probe [3]. The maximum difference between B_{av3} and B_0, ΔB_{max3}, is found by a numerical search during which B_{av3} is determined by numerical integration. The major advance over the earlier study is the development of an expression giving the average magnetic flux density for a circular coil probe for any position and orientation of the probe in the dipole magnetic field. This development allows the search for ΔB_{max3} to consider "all" possible rotations of the three-axis probe. The extended search is shown to leave unchanged the values of ΔB_{max3} that were determined by the earlier treatment.

2. Expression for Average Magnetic Field

In the derivation given below, it is assumed that the cross sectional areas of the wire in the coil probes and the opposing magnetic fields produced by currents induced in the probes are negligible. We also assume that the three orthogonally oriented coils of the three-axis probe have circular cross sections of equal area. These assumptions either can be met in practice or can be taken into account by a calibration process.

The average magnetic flux density, B_{av}, for a single circular coil probe with cross sectional area A is given by

$$B_{av}=\frac{1}{A}\int\int_A \boldsymbol{B}\cdot\boldsymbol{n}\,dA, \tag{2}$$

where dA is an element of probe area, $\boldsymbol{n}$ is a unit vector perpendicular to A, and $\boldsymbol{B}$ is the magnetic flux density. In spherical coordinates, the magnetic flux density for a small current loop of radius b is [4]

$$\boldsymbol{B}=\frac{\mu_0 I b^2}{2r^3}\cos\theta\boldsymbol{u}_r+\frac{\mu_0 I b^2}{4r^3}\sin\theta\boldsymbol{u}_\theta, \tag{3}$$

where μ_0 is the permeability of vacuum, I is the alternating current, and $\boldsymbol{u}_r$ and $\boldsymbol{u}_\theta$ are unit vectors in the directions of increasing r and θ, respectively. The assumption is made that $b \ll r$, and the sinusoidal time dependence of the field has been suppressed. The value of B_0 is given by the magnitude of B [Eq. (3)]. Figure 1 shows the spherical coordinates r and θ, a small current loop at the origin of the coordinate system, and a sketch of a three-axis probe. The center of the probe coincides with the origin of the prime coordinate system x', y', and z'. The coil probes are labelled P1, P2, and P3, have unit normal vectors $\boldsymbol{n}_1$, $\boldsymbol{n}_2$, and $\boldsymbol{n}_3$, respectively, and are shown in Figure 1 (inset) for illustrative purposes as being in the directions of prime coordinates. The orientation of the magnetic dipole with respect to the position of the probe is characterized by the angle θ.

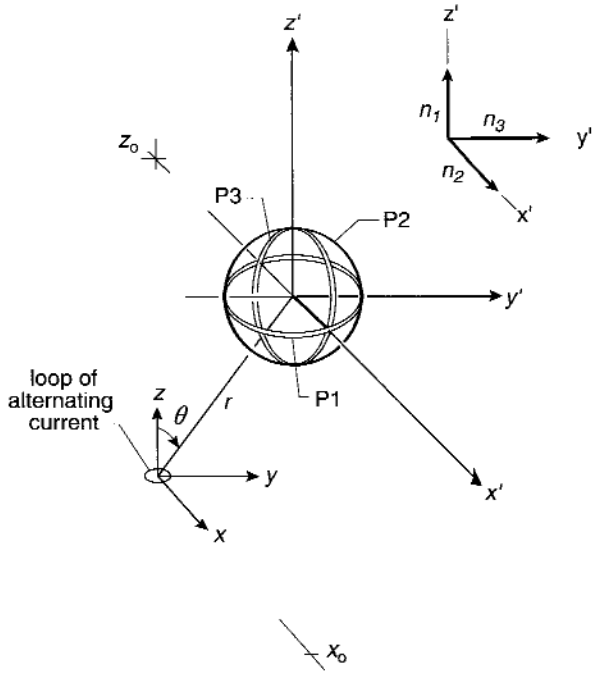

Fig. 1. Three-axis magnetic field probe with its center at $x=x_0$, $y=0$, and $z=z_0$. A small current loop producing a dipole magnetic field is located at the origin of the unprimed coordinate system. The unit vectors $\boldsymbol{n}_1$, $\boldsymbol{n}_2$, and $\boldsymbol{n}_3$ are normal to the areas of probes P1, P2, and P3, respectively. Changes in the angle θ correspond to varying the orientation of the dipole with respect to the probe.

For our purposes, it is convenient to express $\boldsymbol{B}$ in terms of Cartesian coordinates. The magnetic flux density is then [2]

$$\boldsymbol{B}=i\frac{3Cxz}{2r^5}+j\frac{3Cyz}{2r^5}+k\frac{C}{2r^3}\left(\frac{3z^2}{r^2}-1\right), \tag{4}$$

where $r=[x^2+y^2+z^2]^{1/2}$, $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ are unit vectors for the Cartesian coordinates, and C is the constant $\mu_0 I b^2/2$.

The goal is to develop an expression for B_{av} at an arbitrary point which can be evaluated for any orientation of the coil probe. The value of B_{av3} can then be found by combining the rms values of B_{av} from three orthogonal directions according to Eq. (1). The approach described below for obtaining the desired expression for B_{av} is to transform the problem into the coordinate system of the coil probe. In this coordinate system, the unit vector normal to the plane of the coil coincides with the "z-axis", $\boldsymbol{B}$ is expressed in terms of the probe coordinates, and the integration over the area of the circular coil probe is carried out numerically in polar coordinates.

We begin by considering, without loss of generality, a three-axis coil probe with its center at $x=x_0$, $y=0$, and $z=z_0$, where $x_0=r\sin\theta$ and $z_0=r\cos\theta$ (Fig. 1). We then focus on coil probe P1 and its unit vector $\boldsymbol{n}_1$ after it is rotated through angles α_1 and α_2 with respect to the prime coordinate system as shown in Fig. 2. The unit vectors $\boldsymbol{n}_2$ and $\boldsymbol{n}_3$ will also change in orientation to maintain their orthogonal relationship, but are not shown for purposes of clarity. In the prime coordinate system, $\boldsymbol{n}_1$ is given by

$$\boldsymbol{n}_1=\boldsymbol{i}\sin\alpha_1\cos\alpha_2+\boldsymbol{j}\sin\alpha_1\sin\alpha_2+\boldsymbol{k}\cos\alpha_1. \tag{5}$$

By examination,

$$\begin{aligned}\boldsymbol{n}_2&=\boldsymbol{i}\sin(\alpha_1+90°)\cos\alpha_2+\boldsymbol{j}\sin(\alpha_1+90°)\sin\alpha_2\\&\quad+\boldsymbol{k}\cos(\alpha_1+90°)\\&=\boldsymbol{i}\cos\alpha_1\cos\alpha_2+\boldsymbol{j}\cos\alpha_1\sin\alpha_2-\boldsymbol{k}\sin\alpha_1.\end{aligned} \tag{6}$$

The remaining unit vector, $\boldsymbol{n}_3$, is given by

$$\begin{aligned}\boldsymbol{n}_3&=\boldsymbol{n}_1\times\boldsymbol{n}_2\\&=-\boldsymbol{i}\sin\alpha_2+\boldsymbol{j}\cos\alpha_2,\end{aligned} \tag{7}$$

but it also can be determined by examination, i.e., α_1 is replaced by 90° and α_2 is replaced by $\alpha_2+90°$ in Eq. (5). For this case, $\boldsymbol{n}_3$ is constrained to lie in the x'-y' plane. Later this constraint is removed.

The coordinate system of the probe is reached by the following transformations:

(i) translation of the origin to the origin of the prime coordinates shown in Fig. 1,

(ii) rotation of the prime coordinates through angle α_2 about the z'-axis, yielding the double-prime coordinates x'', y'', z'', as shown in Fig. 2, and

(iii) rotation of the double-prime coordinates through angle α_1 about the y''-axis yielding the triple-prime coordinates x''', y''', z''' (Fig. 2). In the triple-prime coordinate system, the normal vector, $\boldsymbol{n}_1$, given by Eq. (5), is along the z'''-axis as desired.

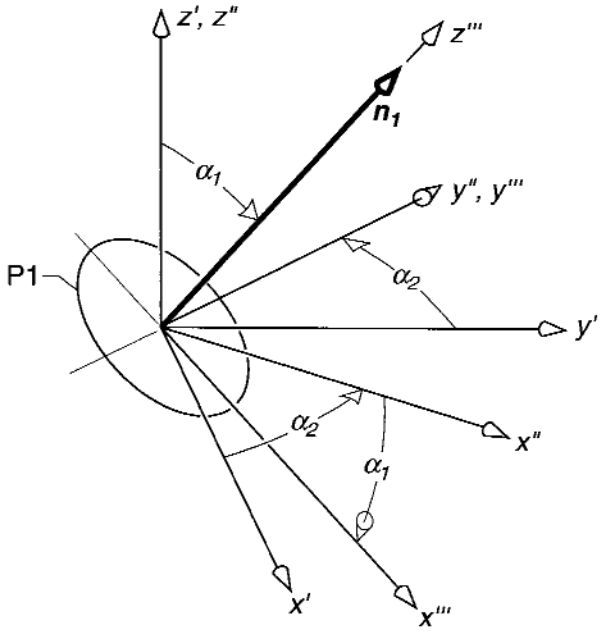

Fig. 2. (a) Geometry of unit vector $\boldsymbol{n}_1$ and coordinates after rotation of the prime coordinates through angle α_2 about the z'-axis and after rotation of double-prime coordinates through angle α_1 about y''-axis.

Transformation (i) is given by

$$\begin{aligned}x&=x_0+x'\\y&=y'\\z&=z_0+z'.\end{aligned} \tag{8}$$

The first rotation of coordinates (ii) is given by

$$\begin{aligned}x'&=x''\cos\alpha_2-y''\sin\alpha_2\\y'&=x''\sin\alpha_2+y''\cos\alpha_2\\z'&=z'',\end{aligned} \tag{9}$$

and the second rotation (iii) is given by

$$\begin{aligned}z''&=z'''\cos\alpha_1-x'''\sin\alpha_1\\x''&=z'''\sin\alpha_1+x'''\cos\alpha_1\\y''&=y'''.\end{aligned} \tag{10}$$

From Eqs. (8)–(10), we have

$$\begin{aligned}x&=x_0+(z'''\sin\alpha_1+x'''\cos\alpha_1)\cos\alpha_2-y'''\sin\alpha_2\\y&=(z'''\sin\alpha_1+x'''\cos\alpha_1)\sin\alpha_2+y'''\cos\alpha_2\\z&=z_0+z'''\cos\alpha_1-x'''\sin\alpha_1,\end{aligned} \tag{11}$$

which, when substituted into Eq. (4), expresses $\boldsymbol{B}$ in terms of the probe coordinates. A simplification of Eq. (11) is had by noting that the integration over the area of the probe [Eq. (2)] occurs in the x'''-y''' plane, i.e., $z'''=0$. Contributions to B_{av3} from each

of the coil probes is found by using the appropriate normal vector [Eqs. (5)–(7)] during the integrations. As noted above, the integration is carried out numerically (using a double Simpson's Rule) in polar coordinates, i.e.,

$$x''' = \rho \cos\psi, \quad 0 \leq \rho \leq a, \quad 0 \leq \psi < 2\pi$$
$$y''' = \rho \sin\psi,$$
$$dA = dx''' dy''' = \rho \, d\rho \, d\psi. \tag{12}$$

The accuracy of the numerical integrations was checked by increasing the number of divisions between the limits of integration for ρ and ψ. The results reported below were not affected by further refinements of the intervals used during the integrations.

In the search for ΔB_{max3}, it will be necessary to perform rotations of the three-axis probe about the z'''-axis or $\boldsymbol{n}_1$ direction (see Search Protocol below), i.e., the unit vectors $\boldsymbol{n}_2$ and $\boldsymbol{n}_3$ for probes P2 and P3 are rotated about $\boldsymbol{n}_1$. This removes the constraint noted earlier that $\boldsymbol{n}_3$ lies in the x'-y' plane. Because the integrand for B_{av} is in terms of the α angles, relationships must be found between the angle of rotation about the z'''-axis, designated as ϕ, and the α values that appear in the integrands for P2 and P3. These relationships are found by examining the unit vectors for the probes $\boldsymbol{n}_2$ and $\boldsymbol{n}_3$ as they rotate about the z'''-axis or $\boldsymbol{n}_1$ direction.

In Figs. 3 and 4, consider $\boldsymbol{n}_1$ in a direction characterized by the angles α_{10} and α_{20}, i.e.,

$$\boldsymbol{n}_1 = \boldsymbol{i}\sin\alpha_{10}\cos\alpha_{20} + \boldsymbol{j}\sin\alpha_{10}\sin\alpha_{20} + \boldsymbol{k}\cos\alpha_{10}. \tag{13}$$

From Eqs. (6) and (7), we have

$$\boldsymbol{n}_2 = \boldsymbol{i}\sin(\alpha_{10}+90°)\cos\alpha_{20} + \boldsymbol{j}\sin(\alpha_{10}+90°)\sin\alpha_{20} + \boldsymbol{k}\cos(\alpha_{10}+90°), \tag{14}$$

and

$$\boldsymbol{n}_3 = \boldsymbol{i}\sin(90°)\cos(\alpha_{20}+90°) + \boldsymbol{j}\sin(90°)\sin(\alpha_{20}+90°) + \boldsymbol{k}\cos(90°). \tag{15}$$

The trigonometric expressions in Eqs. (14) and (15) are not simplified in order to aid the reader in seeing the relationships between the three unit vectors.

Following a counterclockwise rotation of ϕ degrees about the z'''-axis, the α_2's will increase in value and the α_1's will decrease in value in the expressions for $\boldsymbol{n}_2$ and $\boldsymbol{n}_3$. These changes also occur in the expression for the magnetic flux density $\boldsymbol{B}$.

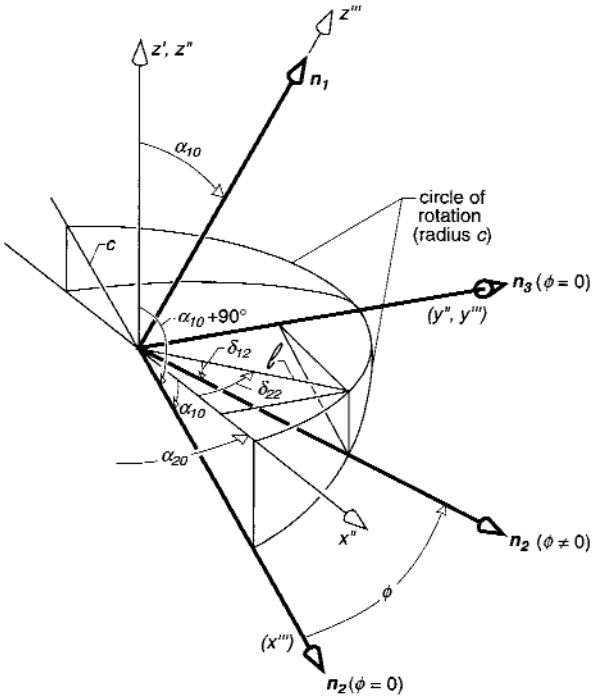

Fig. 3. Geometry of coordinates and unit vectors after unit vector $\boldsymbol{n}_2$ is rotated ϕ degrees about $\boldsymbol{n}_1$ or z'''-axis. The rotation of $\boldsymbol{n}_3$ is not shown for purposes of clarity (see Fig. 4).

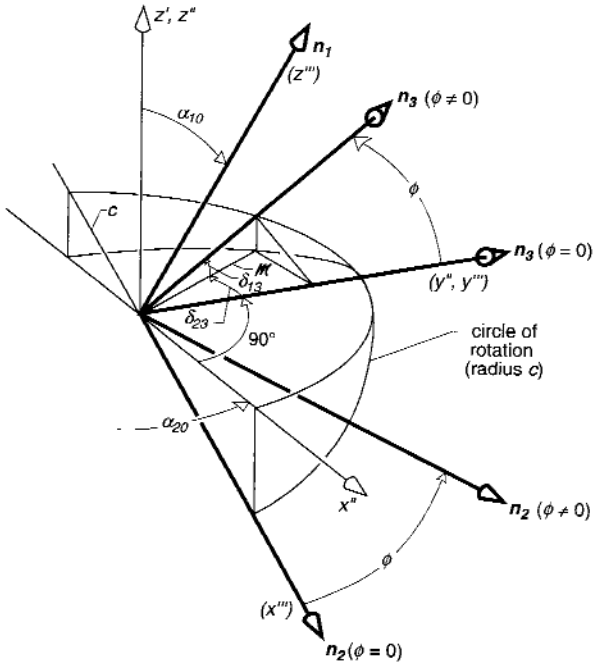

Fig. 4. Geometry of coordinates and unit vectors after unit vectors $\boldsymbol{n}_3$ and $\boldsymbol{n}_2$ are rotated through angle ϕ about $\boldsymbol{n}_1$ or z'''-axis.

After a rotation of ϕ degrees, a line along the unit vector $\boldsymbol{n}_2$ will intersect the circle of rotation in the x'''-y''' plane at a point given by (Fig. 3)

$$\begin{aligned} x''' &= c\cos\phi \\ y''' &= c\sin\phi \\ z''' &= 0, \end{aligned} \tag{16}$$

where the radius for the rotation has been arbitrarily taken to be some constant c. From Eqs. (10) and (16), the same point in the double prime coordinate system is

$$\begin{aligned} z'' &= -x'''\sin\alpha_{10} = -c\cos\phi\sin\alpha_{10} \\ x'' &= x'''\cos\alpha_{10} = c\cos\phi\cos\alpha_{10} \\ y'' &= y''' = c\sin\phi. \end{aligned} \tag{17}$$

The increment to α_{20} for unit vector $\boldsymbol{n}_2$ after rotation ϕ, δ_{22} (Fig. 3), can be found from the expression for its tangent, i.e.,

$$\begin{aligned} \tan\delta_{22} &= \frac{y''}{x''} = \frac{\tan\phi}{\cos\alpha_{10}} \\ \delta_{22} &= \tan^{-1}\left(\frac{\tan\phi}{\cos\alpha_{10}}\right). \end{aligned} \tag{18}$$

Prior to the rotation, the angle with respect to the z'-axis for the unit vector $\boldsymbol{n}_2$ is given by $\alpha_{10}+90°$ (Fig. 3). After the rotation, the corresponding angle will be $\delta_{12}+90°$ where $\delta_{12}<\alpha_{10}$. The value of δ_{12} is found by noting that after the rotation ϕ, the line l from the origin to the projection of the circle of rotation onto the x''-y'' plane is given by

$$l = \sqrt{(x'')^2+(y'')^2}, \tag{19}$$

and that the tangent for δ_{12} is just $|z''/l|$ (Fig.3). From Eqs. (17) and (19),

$$\delta_{12} = \tan^{-1}\left|\frac{z''}{l}\right| = \tan^{-1}\left(\frac{\cos\phi\sin\alpha_{10}}{\sqrt{\cos^2\phi\cos^2\alpha_{10}+\sin^2\phi}}\right). \tag{20}$$

Thus, following a rotation of ϕ degrees about $\boldsymbol{n}_1$, α_{20} and $\alpha_{10}+90°$ will be replaced by $\alpha_{20}+\delta_{22}$ and $\delta_{12}+90°$, respectively, in the expressions for $\boldsymbol{n}_2$ and $\boldsymbol{B}$ during the calculation of B_{av} for probe P2.

The α values for $\boldsymbol{n}_3$ can be determined with a similar analysis. Following a rotation of ϕ degrees about the z'''-axis (Fig. 4), a line along the unit vector $\boldsymbol{n}_3$ will intersect the circle of rotation in the x'''-y''' plane at the point

$$\begin{aligned} x''' &= -c\sin\phi \\ y''' &= c\cos\phi \\ z''' &= 0. \end{aligned} \tag{21}$$

From Eqs. (10) and (21), the same point in the double prime coordinate system is

$$\begin{aligned} z'' &= c\sin\phi\sin\alpha_{10} \\ x'' &= -c\sin\phi\cos\alpha_{10} \\ y'' &= c\cos\phi. \end{aligned} \tag{22}$$

Following rotation ϕ, the angle $\alpha_{20}+90°$ for $\boldsymbol{n}_3$ [Eq. (15)] will increase by an amount δ_{23} as shown in Fig. 4. The increment, δ_{23}, can be found from the expression for the absolute value of its tangent, i.e.,

$$\begin{aligned} \tan\delta_{23} &= \left|\frac{x''}{y''}\right| = \tan\phi\cos\alpha_{10}, \\ \delta_{23} &= \tan^{-1}(\tan\phi\cos\alpha_{10}). \end{aligned} \tag{23}$$

Prior to the rotation, the unit vector $\boldsymbol{n}_3$ makes an angle of 90° with respect to the z'-axis (Fig. 4). After the rotation, this angle will decrease by an amount δ_{13}. The value of δ_{13} is found by noting that after the rotation ϕ, the line m from the origin to the projection of the circle of rotation onto the x''-y'' plane is given by

$$m = \sqrt{(x'')^2+(y'')^2}, \tag{24}$$

and that the tangent for δ_{13} is just z''/m (Fig. 4).

From Eqs. (22) and (24) we have

$$\begin{aligned} \tan\delta_{13} &= \frac{z''}{m} = \frac{\sin\phi\sin\alpha_{10}}{\sqrt{\sin^2\phi\cos^2\alpha_{10}+\cos^2\phi}} \\ \delta_{13} &= \tan^{-1}\left(\frac{\sin\phi\sin\alpha_{10}}{\sqrt{\sin^2\phi\cos^2\alpha_{10}+\cos^2\phi}}\right). \end{aligned} \tag{25}$$

Thus, following a rotation of ϕ degrees about $\boldsymbol{n}_1$, $\alpha_{20}+90°$ and 90° are replaced by $\alpha_{20}+90°+\delta_{23}$ and $90°-\delta_{13}$, respectively in the expressions for $\boldsymbol{n}_3$ [Eq. (15)] and $\boldsymbol{B}$ during calculation of B_{av} for coil probe P3.

3. Search Protocol

The search for the largest difference between B_{av3} and B_0, ΔB_{max3}, for a given distance r from the dipole proceeds as follows:

(i) For a fixed distance r away from the dipole, and with $\theta=\alpha_1=\alpha_2=0$, the three-axis probe is rotated about the z'''-axis or $\boldsymbol{n}_1$ direction in 2° steps (i.e., ϕ is incremented in 2° steps). Then B_{av} for each coil probe is evaluated and combined according to Eq. (1) for each value of ϕ to obtain B_{av3}, B_{av3} is compared with B_0, and the largest difference is

saved. Because of the symmetry of the problem, a total rotation of 90° is required to cover all the cases (with 2° increments).

(ii) The angle α_1 is advanced in 5° steps and the above comparisons are repeated as the probe is rotated about the z'''-axis or $\boldsymbol{n}_1$ direction. The maximum value of α_1, without duplication of results is 90°.

(iii) For each value of α_1, α_2 is incremented from 0° in steps of 5° and the above comparisons are repeated. Because of symmetry arguments, a total rotation of 180° is required to consider all the cases without duplication.

(iv) Following the above calculations, different orientations of the magnetic dipole are considered by changing the angle θ in 15° steps and repeating steps (i) through (iii). The choices of increments indicated above were found to provide adequate sensitivity for determining ΔB_{av3}.

(v) Steps (i) through (iv) are repeated for different values of r.

A diagram schematically indicating several positions for $\boldsymbol{n}_1$, and rotations about $\boldsymbol{n}_1$, as the above protocol was carried out for a fixed value of r is shown in Fig. 5.

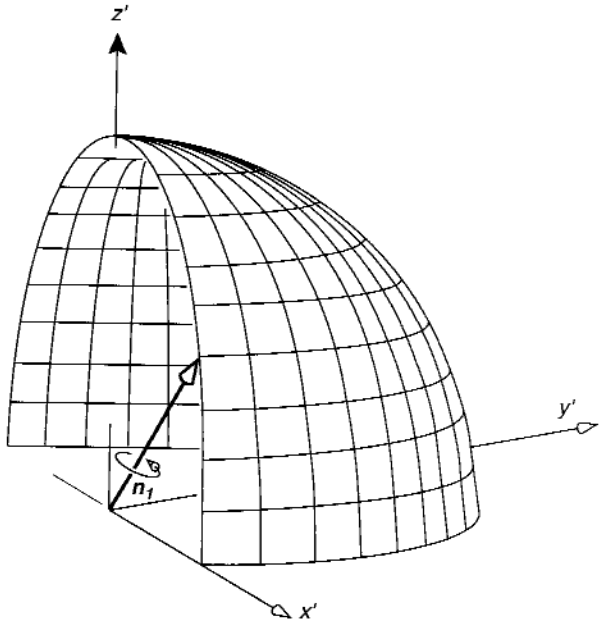

Fig. 5. Trajectories of $\boldsymbol{n}_1$ during search for ΔB_{max3}. The unit vectors $\boldsymbol{n}_2$ and $\boldsymbol{n}_3$ are not shown but maintain their orthogonal relationship with $\boldsymbol{n}_1$ throughout the search.

4. Results and Discussion

As already noted, an earlier search for ΔB_{max3} [3] was not as comprehensive as the one described in this paper. While the ratio r/a and θ could be varied without restriction, the rotations of the three-axis probe were limited to simple rotations about the x'-, y'-, or z'-axis. That is, it was not possible to consider differences between B_{av3} and B_0 for combinations of rotations about two or three axes. This problem has been overcome with a more generalized expression for B_{av} compared to the ones used in the earlier calculation. What is perhaps surprising, however, is that the ΔB_{max3} values obtained with the more comprehensive search protocol are the same as previously calculated. That is, ΔB_{max3} is negative and occurs for all r/a values when $\theta = 90°$, as previously found, and correspond to the ΔB_{max3} values determined earlier following simple rotations about the y'-axis (referred to as "α rotations" in Ref. [3]). Numerical values of ΔB_{max3} are provided in Table 1 as a function of r/a.

Table 1. Values of ΔB_{max3} as a function of normalized distance r/a from magnetic dipole

r/a	ΔB_{max3}(%)
3	−19.6
4	−10.8
5	−6.9
6	−4.8
7	−3.5
8	−2.7
9	−2.1
10	−1.7
11	−1.4
12	−1.2
13	−1.0
14	−0.9
15	−0.8

5. Conclusions

The present calculations have determined the largest differences between the resultant magnetic field, B_{av3}, and the field value at the center of the probe B_0, assuming a dipole magnetic field. These largest differences, designated ΔB_{max3}, are reported in Table 1 as a function of normalized distance, r/a, from the center of the dipole and agree with values previously found after a far less comprehensive search [3]. The quantity, ΔB_{max3}, can be regarded as the largest error due to instrumental

averaging effects. As noted earlier, because the relative orientations of the dipole and three-axis probe are not known for a given r/a under typical measurement conditions, there will be a range of possible differences between B_{av3} and B_0. Thus, ideally, it would be desirable to determine the distribution of differences between B_{av3} and B_0 and treat the problem using a statistical approach, but that has been left to a future calculation.

Because the dipole field is a good approximation of fields produced by many electrical appliances, the information in Table 1 should be taken into account when total uncertainties are being determined during measurements of magnetic fields from appliances. For example, if the resultant magnetic field is to be measured at a distance r from an appliance with a combined relative standard uncertainty [5] of less than $\pm 10\%$, magnetic field meters with three-axis probes having radii a such that $r/a \approx 3$ should be considered unsuitable. Three-axis probes having radii such that $r/a = 5$ would conservatively be considered suitable if the combined relative standard uncertainty from all other sources (e.g., calibration process, frequency response) amounted to about 3% or less, since $6.9\% + 3.0\% = 9.9\%$, where 6.9% is taken from Table 1 for $r/a = 5$.

Acknowledgments

The authors are pleased to acknowledge the assistance of Edward Kelley in enhancing the speed of the computer program used for the calculations. Support was received from the Office of Energy Management of the U.S. Department of Energy.

6. References

[1] D. L. Mader and S. B. Peralta, Residential Exposure to 60-Hz Magnetic Fields From Appliances, Bioelectromagnetics **13**, 287–301 (1992).

[2] IEEE Magnetic Fields Task Force, A Protocol for Spot Measurements of Residential Power Frequency Magnetic Fields, IEEE Trans. Power Delivery **8**, 1386–1394 (1993).

[3] M. Misakian, Coil Probe Dimensions and Uncertainties During Measurements of Nonuniform ELF Magnetic Fields, J. Res. Natl. Inst. Stand. Technol. **98**, 287–295 (1993).

[4] D. Corson and P. Lorrain, Introduction to Electromagnetic Fields and Waves, W. H. Freeman, San Francisco, CA (1962) p. 210.

[5] B. N. Taylor and C. E. Kuyatt, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, NIST Technical Note 1297 (1993).

About the authors: *M. Misakian is a physicist and C. Fenimore is a mathematician in the Electricity Division of NIST. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 255 (1994)]

Improved Automated Current Control for Standard Lamps

Volume 99 | Number 3 | May–June 1994

James H. Walker and Ambler Thompson

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

As radiometric lamp standards improve, the need to set lamp current to specific values becomes more important. Commercially available power supplies typically provide 12 bit internal digital-to-analog logic which permits current control with a relative expanded uncertainty of about 1 part in 4096, corresponding to an expanded uncertainty of the current of about 2 mA at 8 A (in this paper, expanded uncertainties are given as 2 standard deviations). For an FEL-type standard spectral irradiance lamp, this corresponds to a spectral irradiance difference of 0.12% at 655 nm. We have developed a technique using 16 bit digital-to-analog conversion which permits current control with a relative expanded uncertainty of about 1 part in 65536, corresponding to an expanded uncertainty of the current of about 0.1 mA at 8 A. This corresponds to a spectral irradiance difference of approximately 0.006% for an FEL lamp at 655 nm. We describe the technique used to achieve this improvement and we show data from a lamp demonstrating the effect of the improvement. We also describe the limitation provided by the uncertainty of the resistance of the current measuring shunt.

Key words: control; current; lamp; irradiance; radiance; radiometry; standards.

Accepted: **April 20, 1994**

1. Introduction

Transfer standard lamps (tungsten and tungsten-halogen) for the near ultraviolet, visible, and infrared spectral regions are approximately Planckian thermal sources. This means that the spectral distribution and brightness are functions of the filament temperature and hence the electrical current flowing through the filament. Current changes effect short-wavelength radiation more than long-wavelength radiation. A current change that causes a 0.1% change in radiance at 800 nm causes a 0.2% change at 400 nm. To specify the spectral distribution, the NIST certification of spectral radiance or irradiance of a standard lamp is at a designated current. For radiometric measurements of the highest accuracy it is imperative that the lamp current be known accurately and match the current at which the spectral radiance or irradiance was calibrated.

For spectral radiance and spectral irradiance in the visible, an approximate relationship for determining the change in luminous flux from a lamp due to a current change in the lamp is

$$\phi/\phi_r \approx (I/I_r)^{6.24} ,$$

where ϕ and I are the changed luminous flux and current, respectively, and ϕ_r and I_r are the rated values of luminous flux and current, respectively [1] [2]. Using this approximate relationship, an increase in current of 1 mA in a lamp operating at 8.000 A would give $(8.001/8.000)^{6.24} = 1.00078$ or a change of about 0.078% in luminous flux. By way

of comparison, measurements made on our FEL-type standard spectral irradiance lamps operating at 8.000 A give a value of about 0.06% per 1 mA in irradiance at 655 nm. Further, measurements made on standard spectral radiance lamps operating at 40.000 A give a value of about 0.012% per 1 mA in radiance at 655 nm. The approximate relationship gives a value $(40.001/40.000)^{6.24} = 1.00016$ or a change of about 0.016% in luminous flux.

The current state of the art in spectroradiometry is about one percent relative expanded uncertainty [3] for spectral radiance and spectral irradiance for standard sources in the visible. For tungsten ribbon spectral radiance lamps the component of relative expanded uncertainty attributed to lamp current ranges from 0.01% to 0.14% for gas-filled lamps and from 0.01% to 0.12% for vacuum lamps. For spectral irradiance lamps the component of relative expanded uncertainty attributed to lamp current ranges from 0.01% to 0.08%. These uncertainties would be even larger if the lamps were operated at lower currents to achieve lower spectral radiance and irradiance levels. Depending on the radiance level, the current, and the wavelength being observed, a 1 mA change in lamp current can cause a change in spectral radiance of 0.012% to 0.7% or more. These seemingly small uncertainties will become significant when the overall expanded uncertainty of spectral radiance and irradiance calibrations is reduced by a factor of three or more. Therefore, accurate control of the lamp current is required, especially for measurements in the ultraviolet.

The requirement to control the current in a standard lamp to 1 mA necessitates power supplies with a current stability of 0.01% or less. While many of the currently available commercial power supplies have the required current stability, most use twelve bit digital-to-analog converters to program the current and thus have a current settability of about 4 mA for the operating current of a typical lamp. This fails to meet the requirement. To solve this problem, we have developed a computer-controlled measurement and control system for a constant current power supply. This system utilizes commercially available components to set and maintain the current at the 0.0010% level for both types of standard lamps. We report on the design of this system and the results of spectral radiance measurements using this system with a tungsten vacuum strip lamp.

2. Description of Automated Current Control System

Figure 1 shows our automated current control system. It is a feedback control system which uses a 16 bit digital-to-analog converter (D-A) located on a board in the computer to produce a control voltage to a voltage controlled constant-current power supply. The power supply current is proportional to this voltage. The load consists of the standard lamp (V40) and a stable shunt resistor. The voltage drop across the shunt resistor is measured by a six-and-a-half-digit digital volt meter (DVM) and the result is acquired by the computer. A data acquisition and control program written in Basic calibrates the circuits and then ramps the current up or down for large current changes or adjusts the control voltage in small amounts to provide accurate constant current operation.

The measurement shunt is nominally rated at 0.01 Ω, 100 A. The shunt was calibrated by the electricity Division at NIST and has a resistance of 0.0099998 Ω with a relative expanded uncertainty of 0.0033% at 8 A. Lamp V40, used as the source for all of the measurements described in this paper, is from a group of specially-constructed, high-stability vacuum lamps of the Quinn-Lee design [4]. This lamp was used because of its extremely small radiance drift rate, on the order of 0.1% per 500 h at 655 nm. The use of this stable lamp permits us to distinguish the effect of small lamp current changes on the spectral radiance. This lamp was calibrated for spectral radiance at a current of 7.600 A in the NIST Facility for Automated Spectroradiometric Calibrations (FASCAL) [5].

A Burr-Brown[1] PCI-20001C general purpose control board with a Burr-Brown PCI-2006M 16 bit digital-to-analog module was used to produce an output voltage for controlling the current setting of the power supply. This D-A module possesses several programmable output voltage ranges. We used the 0 V to 10 V range. A Hewlett Packard 6030A power supply (0 V to 200 V, 0 A to 17 A, 1000 W, 0.01% line and load regulation) operated in the constant current mode with voltage programming was used to supply the current for the lamp circuit.

[1] 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.

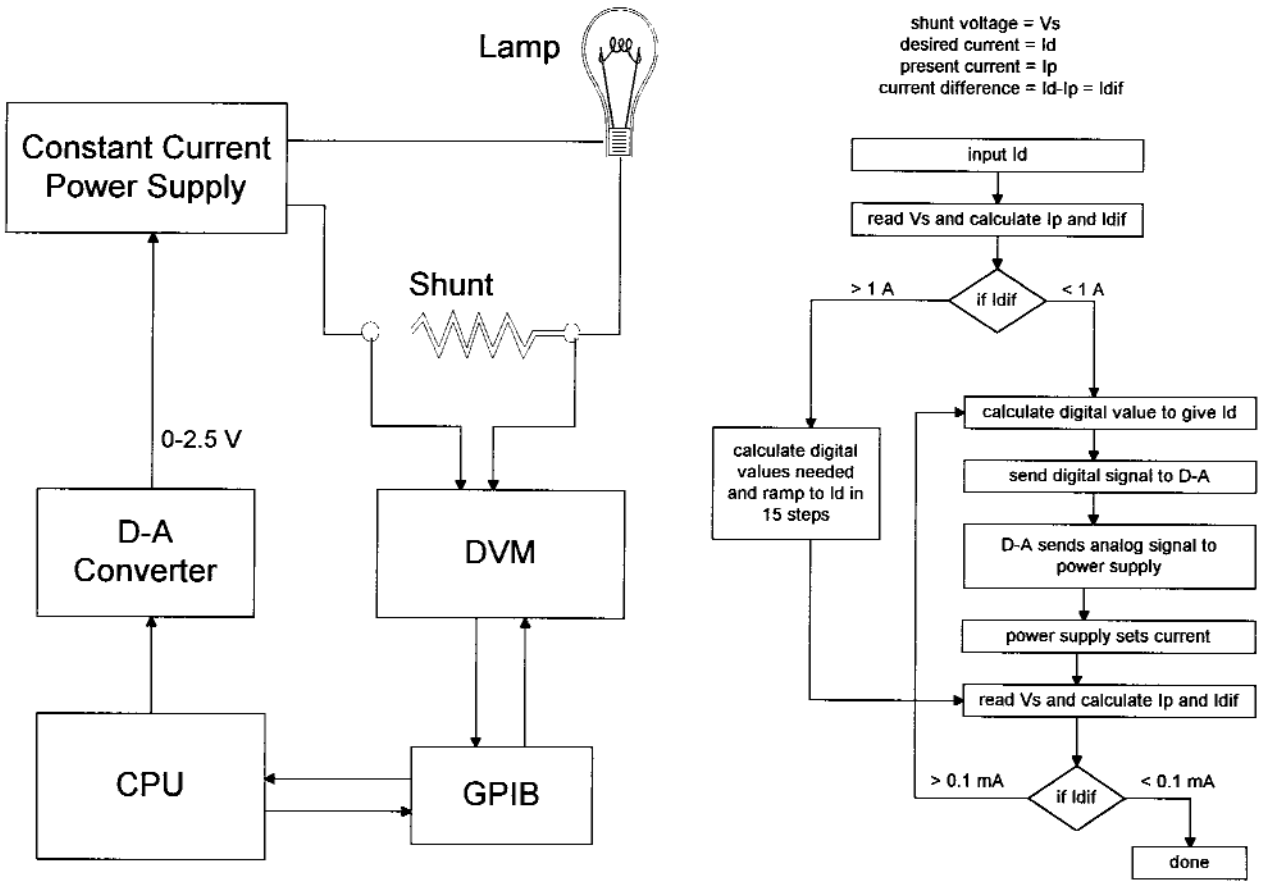

Fig. 1. Schematic diagram of the lamp current feedback control system and a programming flowchart.

Since this power supply requires 0 V to 5 V of external voltage control to provide 0 A to 17 A and the operating lamp current was about 7.6 A, an appropriate scaling of voltage output from the D-A board was required. This scaling also limits the voltage output from the D-A board so that the maximum current possible from the power supply does not damage the lamp, and utilizes the full 16 bit programmability of the D-A converter. Initially, an inline 4/1 voltage divider was used to scale the output voltage from the D-A converter. Later it was found that the output could be scaled by bridging of jumpers on the D-A board by a single appropriate resistor. This is the method we used in our applications.

The D-A module is set to a fixed memory address by DIP switches on the control board. A memory address must be used which does not conflict with other components of the computer. The fullscale output range of the D-A converter is set by a jumper on the D-A module. The voltage output of the D-A converter (0 to fullscale voltage) is controlled by writing an integer from -32768 to $+32768$ to the module memory address using the POKE command. When the computer is turned on, the D-A memory data address is zero, and since this would result in the power supply producing a current of 4 A (half maximum current), we use the autoexec.bat file to run a small program which resets the D-A board to zero voltage output when the computer is turned on. Another subroutine calibrates the power supply current as a function of the signal to the D-A converter with the power supply output short circuited. This information is used by a subroutine which ramps the current up or down to the desired value, within 0.1 mA. The subroutine performs the following steps:

1. The desired lamp current is set by the operator.
2. The present lamp current is read and the difference from the desired current is calculated.

3. If the difference is greater than 1 A, the current is slowly ramped up (or down) to approximately the desired setting by gradually changing the setting of the digital-to-analog converter based on the calibration factor determined for the circuit.
4. If the difference is less than 1 A, a calculation is done using the calibration factor to determine the bit change needed to give the desired current.
5. The lamp current is read again and a new difference is calculated.
6. If necessary, one bit at a time is changed to set the current to within 0.1 mA.

The stability of the D-A output voltage is better than the resolution of one part in 65536 (2^{16}) or 0.0015% which corresponds to 0.12 mA at 7.6 A. The gain drift is 25×10^{-6}/°C or 0.19 mA/°C at 7.6 A. The D-A output voltage replaces the power supply internal program voltage which only has a resolution one part in 4096 (2^{12}), corresponding to 4.15 mA at 7.6 A. The use of an external, high-stability shunt to monitor changes in the lamp current improves the sensitivity of the current measurement.

3. Results

The radiance measurements described in this paper (except for the calibration of lamp V40) were performed in the NIST Low-Level Spectroradiometric Calibration Facility (LLR). The LLR spectroradiometer uses a 0.67 m McPherson scanning monochromator (f/4.7) that is equipped with a prism predisperser. Since the radiance levels to be measured were relatively high, a silicon detector with an integral amplifier was used to measure the spectral radiance (the LLR normally uses a photomultiplier with a photon counting system). The output of the silicon amplifier was measured by a six-and-one-half-digit DVM. The target viewed by the spectroradiometer for these radiance measurements was 0.6 mm wide by 0.8 mm high, similar to that used in FASCAL. The bandpass of the spectroradiometer was about 2 nm over the wavelength range of 400 nm to 1000 nm.

Continuous measurements were made of the spectral radiance at 900 nm of lamp V40 at 7.600 A over a period of about 43 min. Before each measurement, the lamp current was reset to within 0.1 mA of 7.600 A. Figure 2 shows a plot of the current difference (in mA) and the radiance difference

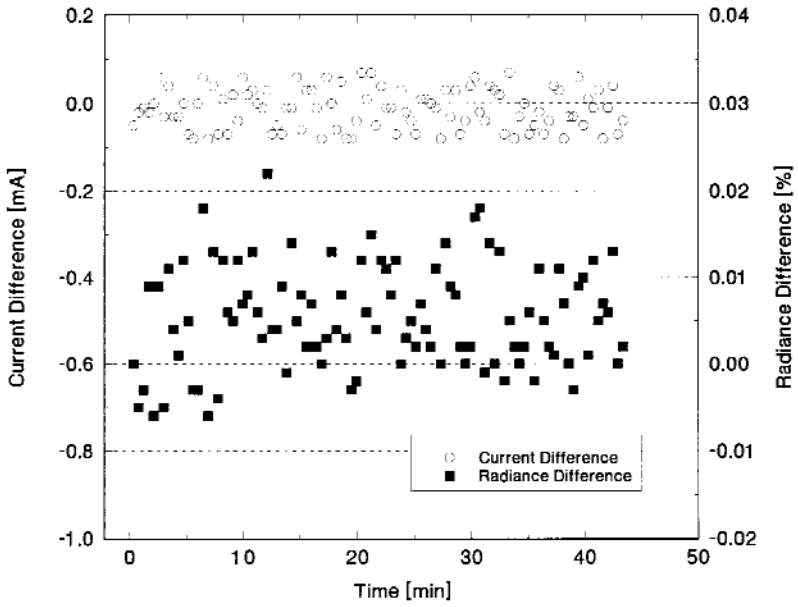

Fig. 2. Current and radiance stability as a function of time for lamp V40 operating at 7.6 A. The current was reset to 7.6 A prior to each radiance measurement. The lamp current is expressed as the difference from 7.6 A in milliamperes. The radiance at 900 nm is expressed as the percent difference from the initial radiance measurement.

(in%) as a function of time (in minutes). The average current over this period was 7.59999 A with an expanded uncertainty of 0.10 mA. The radiance stability relative expanded uncertainty was 0.012% with no discernable drift over the observed interval. The radiance differences plotted in Fig. 2 are relative to the initial radiance reading.

Additional measurements were carried out under the same conditions as above, except the lamp current was set only at the beginning and the radiance and current were continuously monitored over the observation period. Figure 3 shows the results of these measurements as a plot of the current difference (in mA) and the radiance difference (in%) as a function of time (in minutes). The current decreased 0.4 mA to 0.5 mA with respect to the initial setting over the approximately 43 min time interval. The radiance decreased 0.04% over this time interval. From Fig. 3 it can be seen that a 1 mA change in lamp current corresponds to about a 0.08% change in radiance at 900 nm. The radiance drift would be about twice as large at 400 nm to 500 nm.

To examine the wavelength dependence of radiance as a function of lamp current, spectroradiometric scans were made from 400 nm to 1000 nm at 100 nm intervals. The measurement sequence for each wavelength consisted of taking initial measurements at 7.600 A and then varying the lamp current from 7.540 A to 7.660 A in steps of 10 mA and repeating the spectroradiometer scan at the new lamp current setting. A plot of the measured radiance difference (in %) as a function of the measured current difference (in mA) for each of the measured wavelengths is shown in Fig. 4. The two standard deviation radiance repeatability for these measurements was less than 0.01% from 600 nm to 1000 nm, was 0.03% at 500 nm and was 1.30% at 400 nm. In this current range the lamp radiance appears to be directly proportional to the lamp current. A linear fit was applied to the data for each wavelength to determine the slope and these results are summarized in Table 1. The results demonstrate that this slope, which is the rate of change of the lamp radiance with respect to the current, is approximately inversely proportional to the wavelength. This behavior is predicted by the Planck equation. The last column in Table 1 was calculated by taking the slope at 1000 nm and predicting the slope at the other measured wavelengths. The differences between the observed and predicted values for the slope when analyzed in this manner are about 0.1%.

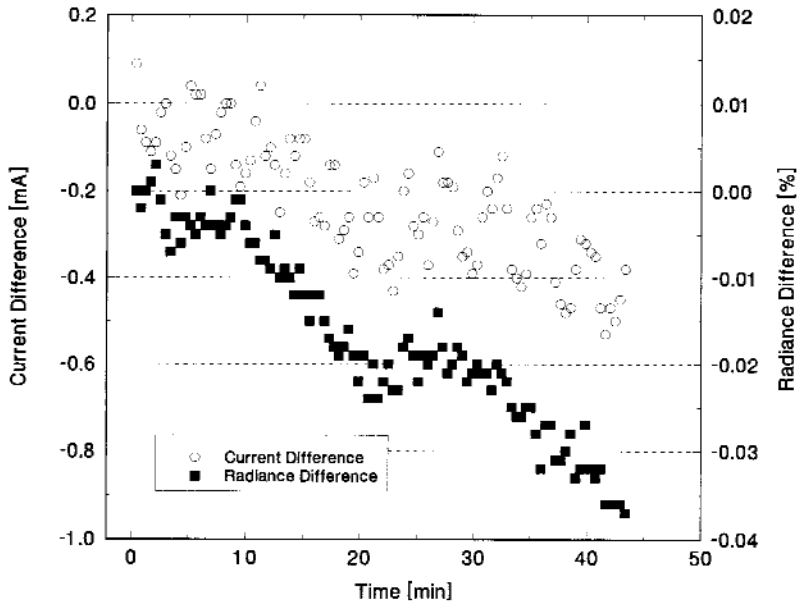

Fig. 3. Current and radiance stability as a function of time for lamp V40 operating at 7.6 A. The current was set to 7.6 at the beginning of the time period and the current value was monitored prior to each radiance measurement. The lamp current is expressed as the difference from 7.6 A in milliamperes. The radiance at 900 nm is expressed as the percent difference from the initial radiance measurement.

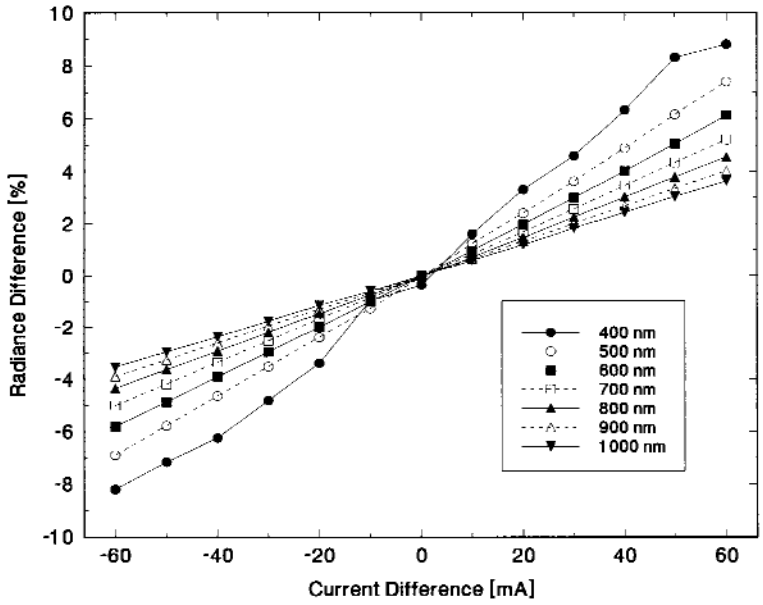

Fig. 4. Lamp V40 radiance difference (in percent) as a function of the current difference (in mA from 7.6 A) and wavelength. Wavelengths measured were (400, 500, 600, 700, 800, 900 and 1000) nm.

Table 1. The fitted slope and its uncertainty from the data in Fig. 4 and the predicted slope for lamp V40 operating at 7.6 A. The Wien approximation of the Planck equation predicts that the slope is inversely proportional to wavelength. The predicted slopes are based upon the observed slope at 1000 nm

Wavelength (nm)	Fitted slope (%/mA)	Standard uncertainty of fitted slope (%/mA)	Predicted slope (%/mA)
400	0.15056	0.00323	0.14933
500	0.11934	0.00081	0.11946
600	0.09933	0.00052	0.09955
700	0.08510	0.00037	0.08533
800	0.07406	0.00035	0.07466
900	0.06588	0.00021	0.06637
1000	0.05973	0.00017	0.05973

4. Discussion

Table 1 shows that the spectral radiance level of lamp V40 would be in error by 0.15% at 400 nm and 0.06% at 1000 nm if the current was in error by 1 mA. If this lamp was operated at 5.000 A, the radiance differences for a 1 mA error would be significantly larger. We have shown that, in order to keep the relative expanded uncertainty in the radiance due to the lamp current below 0.1%, it is necessary to control the current to a fraction of a milliampere. This can be accomplished with our current control technique. However, there is another component of uncertainty that becomes significant when control is done to a fraction of a milliampere. As was stated earlier, the shunt calibration has a relative expanded uncertainty of 0.0033%. For V40 at 7.600 A, this would amount to a 0.25 mA expanded uncertainty in the current setting and a relative expanded uncertainty in the radiance level of 0.015% at 1000 nm and 0.038% at 400 nm. The uncertainty of the shunt resistance is a conservative estimate by the NIST Electricity Division and is due to variation in room temperature causing small changes in the shunt resistance. Since we use a shunt rated at 100 A and calibrate and use the shunt at 8 A, the expected temperature changes due to power dissipation in the shunt will be small, provided it has a low temperature coefficient. For the manganin type shunt we used, the typical temperature coefficient is about 10×10^{-6}/°C (or about 0.1 mA/°C at 7.6 A). The expanded uncertainty of the shunt calibration due to temperature fluctuations could be reduced substantially by immersing the shunt in a temperature controlled oil bath. It is also important to use a DVM with sufficient resolution to measure the voltage drop across the shunt. Since we

are measuring 7.6 A across a 0.01 Ω shunt, the expected voltage reading is approximately 0.076 V, so a six-and-a half-digit DVM is necessary to achieve the required resolution.

In summary, we have employed an easy-to-use, automated technique for setting and controlling standard lamp current with significantly reduce uncertainty. We have incorporated a 16 bit D-A converter to provide us with the resolution we need to set lamp current to within 0.1 mA and we have used a high stability shunt and a six-and-a-half digit DVM to monitor and feed back the current to the control circuit. These improvements will be more valuable as the expanded uncertainty of spectral radiance and irradiance standards is reduced. In conclusion, it is important to point out that the utilization of standard lamps is dependent on several measurements: the calibration of the lamp for spectral radiance or irradiance and the electrical measurements of resistance and voltage.

5. References

[1] Lighting Handbook Reference and Application, Illuminating Engineering Society of North America, Eighth Edition (1993).

[2] J. W. T. Walsh, Photometry, Constable & Company LTD (1958) p. 532.

[3] B. N. Taylor and C. E. Kuyatt, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, NIST Technical Note 1297 (1993).

[4] T. J. Quinn and R. D. Lee, Vacuum Tungsten Strip Lamps with Improved Stability as Radiance Temperature Standards, presented at the Fifth Symposium on Temperature, Washington, DC, June 21–24, 1971, Instrument Society of America, Pittsburgh, PA (1972).

[5] J. H. Walker, R. D. Saunders, and A. T. Hattenburg, Spectral Radiance Calibrations , NBS Special Publication SP 250-1, (1987).

About the Authors: *James H. Walker and Ambler Thompson are members of the Radiometric Physics Division of the NIST Physics Laboratory. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 263 (1994)]

Refractive Indices of Fluids Related to Alternative Refrigerants

Volume 99 Number 3 May–June 1994

Thomas J. Bruno, Marcelo A. Wood, and Brian N. Hansen

National Institute of Standards and Technology,
Boulder, CO 80303

As part of a comprehensive program to develop suitable methods of chemical analysis for alternative refrigerants and their products, we have compiled a database of spectral, chromatographic, and physical property data that can aid in compound identification. As a small part of this effort, we have measured the refractive indices of a number of such fluids for which data were unavailable. The measurements were performed on a commercially available, digital Abbe refractometer that was modified for the relatively low temperature measurements (0 °C to 20 °C) that are sometimes required with these samples.

Key words: alternative refrigerants; brominated/chlorinated ethanes; ethenes; ethers; propanes; propynes; refractive index.

Accepted: May 5, 1994

1. Introduction

The threat of atmospheric ozone depletion has lead to a great deal of research in many laboratories worldwide to find suitable substitutes for the fully halogenated fluids. These fluids have been used for many years as refrigerants, propellants, and blowing/foaming agents. Since the production of many of the older fluids is being phased out by law in most industrialized nations in the near future, there is a pressing need to thoroughly characterize the most promising substitutes. The Thermophysics Division of NIST has been a major force in this effort, with comprehensive experimental and theoretical thermophysical properties research.

Along with the efforts in thermophysical properties, an effort in the chemical analysis of these materials was developed out of necessity. This need arose because it is clearly impossible to understand the thermophysics of fluids of unknown or unreliable purity. Numerous analytical methods [1–4] and devices [5–16] have been introduced, and a comprehensive database of analytical data has been compiled [17–23]. This database contains spectral, chromatographic, and physical property information that is of value in the identification and analysis of alternative refrigerant fluids and byproducts. In this respect, the database covers fluid reaction products and common impurities—in short, any material that might have to be identified and quantified as part of the thermophysical properties work.

One of the most valuable physical properties for the identification of a material is the refractive index n_D. Because it can be readily determined with a relative expanded uncertainty of a few parts in 10 000 (coverage factor $k = 2$) [24], it is very useful and reliable in providing confirmatory evidence of the identity of a compound, especially in the liquid state. A large number of the materials of interest in alternative refrigerant research are newly syn-

thesized, and therefore published values of refractive indices are often unavailable. As part of our efforts to provide as complete an analytical database as possible, we have measured the indices of refraction of 23 fluids for which no data were available.

We wish to emphasize that not all of these fluids are considered potential refrigerants. Indeed, many of these materials are heavily chlorinated or brominated, and are thus unsuitable from an ozone depletion point of view. It is important, however, to have the capability to readily identify these materials, since they may occur as reaction/decomposition products, or perhaps as residual impurities in field installations.

2. Experimental

The measurement of the refractive index of liquid samples is most often done with either an immersion refractometer or an Abbe refractometer. The immersion type is the most accurate instrument for use with liquids, typically producing measurements having standard deviations of 0.000 03. It is usable for refractive indices in the range 1.32 to 1.54. This instrument is somewhat inconvenient to handle, however, and requires a relatively large sample (10 mL to 15 mL) that must be maintained at constant temperature. Moreover, it can sometimes be plagued by sample viscosity effects. The Abbe refractometer, on the other hand, is an easily-used laboratory instrument amenable to much smaller sample sizes, usually just a few drops [25]. The smaller sample size makes temperature control of the sample much easier. This refractometer can be used for refractive indices in the range 1.3 to 1.7. It produces measurements of somewhat larger uncertainty than the immersion refractometer, with typical experimental standard deviations of 0.0001 [25–28]. In addition, it is a bit more complicated in construction than the immersion instrument.

Because of the relatively small quantities available for most of our samples, we used a digital Abbe refractometer for the measurements reported in this work. Since many of the fluids required measurement at a lower temperature than the nearly universal 25 °C of most reported liquid measurements, some simple modifications to the standard commercial instrument were required. The Amici prisms of the refractometer were thermostatted with a circulating low temperature bath that used ethanol as the working fluid. In addition, the optical housing of the refractometer was purged with a gentle flow of dry nitrogen to prevent condensation on the interior optics of the instrument. Some areas of the instrument were provided with glass-wool insulation. This was done for temperature control and as an added precaution against condensation of ambient moisture on critical surfaces. The samples were generally cooled in an ice bath prior to their being placed on the lower prism. The temperatures of the samples were measured before and after the measurement of each refractive index. Temperature was measured with a thermistor located in the lower Amici prism and had an expanded uncertainty of 0.05 °C.

The samples that were used for these measurements were either obtained commercially or were synthesized in other laboratories, and were of the highest available purity. They were used without further purification.

3. Results

The refractive indices of the fluids that were measured in this study are provided in Table 1, along with their respective refrigerant code numbers [29] and the temperatures at which the measurements were taken. The fluids are divided into the following classifications: ethanes, ethenes, brominated ethanes, propanes, propynes, and ethers.

4. Discussion

The repeatability of the measurements reported in Table 1 was assessed by performing multiple measurements in a relatively short time period under the same instrumental conditions. In general, only very slight variations (on the order of 0.01%) were noted between replicate measurements for each sample. In order to assess the longer term stability and reliability of the data, a large number of measurements were performed on one sample: 1,2,2-trichloropentafluoropropane, R-215a. During the course of several hours, 21 measurements were taken at 20.0 °C, and 21 were taken at 25.0 °C. The lower Amici prism was cleaned after each measurement, and sample was reapplied to the surface. The results are shown in Table 2, where the quoted uncertainty is the expanded uncertainty with a coverage factor $k = 2$ based solely on the experimental standard deviation of the mean of the 21 measurements. Probability plots constructed from both sets

Table 1. Refractive indices of the fluids measured in this study

Code number	Fluid	n_D	Temperature (°C)
R-121	1,1,2,2-tetrachlorofluoroethane	1.4487	20.0
R-122	1,1-difluoro-1,2,2-trichloroethane	1.3922	20.0
R-131	2-fluoro-1,1,2-trichloroethane	1.4396	20.0
R-131a	1-fluoro-1,1,2-trichloroethane	1.4252	20.0
R-1112aB2	1,1,-dibromodifluoroethene	1.4489	20.0
R-142B1	2-bromo-1,1-difluoroethane	1.3871	20.0
R-113B2αβ	2-chloro-1,2-dibromo-1,1,2-trifluoroethane	1.4281	20.0
R-114B2	1,2-dibromotetrafluoroethane	1.3708	20.0
R-123B2	1,2-dibromo-1,1,2-trifluoroethane	1.3720	20.0
R-123B1α	1-bromo-2-chloro-1,1,2-trifluoroethane	1.3721	20.0
R-133aB1	2,2,2-trifluoroethyl bromide	1.3429	5.0
R-215a	1,2,2-trichloropentafluoropropane	1.3497	25.0
		1.3525	20.0
R-215ba	1,2,3-trichloropentafluoropropane	1.4570	20.0
R-216ba	1,2-dichlorohexafluoropropane	1.3114	5.0
R-225ca	3,3-dichloro-1,1,1,2,2-pentafluoropropane	1.3248	20.0
R-225cb	1,3-dichloro-1,1,2,2,3-pentafluoropropane	1.3265	20.0
R-243db	2,3-dichloro-1,1,1-trifluoropropane	1.3677	20.0
R-253fb	3-chloro-1,1,1-trifluoropropane	1.3298	20.0
R-262da	2-chloro-1,3-difluoropropane	1.3810	20.0
R-216B2	1,2-dibromohexafluoropropane	1.3596	20.0
R-2240	3-chloro-1-propyne	1.4362	20.0
R-E150a	α,α-dichloromethyl methyl ether	1.4070	20.0
R-E270b	2-chloroethyl methyl ether	1.4370	20.0
R-E280	2,2-dichloroethyl methyl ether	1.4165	5.0

Table 2. Results of extended index of refraction measurements taken for R-215a

Temperature, °C	n_D
(20.0 ± 0.05) °C:	1.3525 ± 0.0001
(25.0 ± 0.05) °C:	1.3497 ± 0.0001

of data were linear, indicating that the deviations that were measured were normally distributed. We think that this level of reproducibility (approximately 0.01%) is indicative of that of the data provided in Table 2. The data are therefore of sufficient reliability for qualitative identification purposes. The experimental standard deviations are small compared to typical differences in refractive index that one observes from fluid to fluid. Performing such a number of multiple measurements for all of the samples was impossible because of the very limited supply available for most of these fluids.

The combined standard uncertainty of the measurements was assessed by measuring the refractive indices of several halocarbons having well established values of n_D, as reported in the literature. On the basis of these comparisons, we estimate the final relative expanded uncertainty of the measurements presented here at 0.02%.

5. References

[1] T. J. Bruno, Strategy of Chemical Analysis of Alternative Refrigerants, Natl. Inst. Stand. Technol. Tech. Note 1340, (1990).
[2] T. J. Bruno, in Handbook for the Identification and Analysis of Alternative Refrigerants, T. J. Bruno, ed., Chapt. 1. Chemical Analysis Methods for Alternative Refrigerants and Related Products, CRC Press, Boca Raton, FL, in press.
[3] T. J. Bruno, ASHRAE Trans. **98**(2), 204 (1992).
[4] T. J. Bruno, ASHRAE Trans. **98**(2), 210 (1992).
[5] T. J. Bruno, Rev. Sci. Instr. **63**(10), 4459 (1992).
[6] T. J. Bruno, Method and Apparatus for the Recording of Infrared Spectra at Low Temperature, United States Patent No. 5,280,177.
[7] T. J. Bruno, Proc. Contr. Qual. **3**, 195–207 (1992).
[8] T. J. Bruno, Am. Lab. **25**(12), 15 (1993).
[9] T. J. Bruno, Am. Lab. **25**(14), 16 (1993).
[10] B. H. Hansen, A. F. Lagalante, R. E. Sievers, and T. J. Bruno, Rev. Sci. Instr., in press.
[11] T. J. Bruno, Anal. Chem. **58**, 1596 (1986).

[12] T. J. Bruno, Apparatus and Method for the Evaporative Concentration of Liquid Samples, United States Patent No. 5,217,904.
[13] T. J. Bruno, J. Chem. Educ. **69**(10), 837 (1992).
[14] T. J. Bruno, J. Chromatogr. Sci. **32**, 112 (1994).
[15] T. J. Bruno, J. Chromatogr. Sci. **23**, 325 (1985).
[16] T. J. Bruno and G. L. Hume, J. Res. Natl. Bur. Stand. (U.S.) **90**(3), 255–257 (1985).
[17] T. J. Bruno and M. Caciari, J. Chromatogr. **672**(1,2), 149 (1994).
[18] T. J. Bruno and M. Caciari, Retention of halocarbons on a hexafluoropropylene epoxide modified graphitized carbon black, Part 2: ethane-based compounds, J. Chromatogr., in press.
[19] T. J. Bruno and M. Caciari, Retention of halocarbons on a hexafluoropropylene epoxide modified graphitized carbon black, Part 3: ethene-based compounds, J. Chromatogr., in press.
[20] T. J. Bruno, Spectroscopic Library for Alternative Refrigerant Analysis, Natl. Inst. Stand. Tech. Special Publication 794 (1990).
[21] T. J. Bruno and P. D. N. Svoronos, Basic Tables for Chemical Analysis, Natl. Bur. Stand. (U.S.) Technical Note 1096 (1986).
[22] T. J. Bruno and P. D. N. Svoronos, CRC Handbook of Basic Tables for Chemical Analysis, CRC Press, Boca Raton, FL (1989).
[23] T. J. Bruno, B. N. Hansen, M. Caciari, P. D. N. Svoronos, R. F. X. Klein, and T. C. Waidner, in Handbook for the Identification and Analysis of Alternative Refrigerants, T. J. Bruno, ed., Chapt. 2. Analytical Data for the Analysis of Alternative Refrigerants and Related Materials, CRC Press, Boca Raton, FL, in press.
[24] B. N. Taylor and C. E. Kuyatt, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, Natl. Inst. Stand. Technol. Technical Note 1297 (1993).
[25] D. P. Shoemaker, C. W. Garland, and J. I. Steinfeld, Experiments in Physical Chemistry, 3rd. ed., McGraw-Hill Book Co., New York (1974).
[26] L. E. Maley, J. Chem. Educ. **45**, A467 (1968).
[27] McGraw-Hill Encyclopedia of Science and Technology, 7th. ed., Vol. 15, McGraw-Hill Book Co., (1992) pp. 253–254.
[28] P. D. N. Svoronos and E. Sarlo, Organic Chemistry Laboratory Manual, William C. Brown Publishers, Dubuque, Iowa (1993).
[29] Number Designation and Safety Classification of Refrigerants, ASHRAE Standard, ANSI/ASHRAE 34-1992, American Society of Heating Refrigerating and Air Conditioning Engineers, Atlanta, GA.

About the authors: *Thomas J. Bruno is Group Leader of the Process Separations Group (in the Thermophysics Division), Marcelo Wood is a PREP (Professional Research Experience Program) student in the Process Separations Group, and Brian N. Hansen is an ASEE (American Society for Engineering Education) postdoctoral research associate in the Process Separations Group at NIST. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 267 (1994)]

A Theoretical Analysis of the Coherence-Induced Spectral Shift Experiments of Kandpal, Vaishya, and Joshi

Volume 99 Number 3 May–June 1994

John T. Foley and M. Wang

Department of Physics and Astronomy,
Mississippi State University,
Mississippi State, MS 39762

The optical system used by Kandpal, Vaishya, and Joshi in their experiments on coherence-induced spectral shifts is analyzed theoretically. An approximate form for the cross-spectral density in the secondary source plane is obtained, and it is shown that, contrary to the assertions of Kandpal, Vaishya, and Joshi, the corresponding complex degree of spectral coherence in this plane is wavelength dependent. After making some assumptions about the behavior of the interference filter used in the system, an approximate form for the spectrum of the light on-axis in the observation plane is obtained. It is shown that the peak wavelengths of this spectrum do not agree with those reported by Kandpal, Vaishya, and Joshi. Possible reasons for this disagreement are discussed.

Key words: correlation-induced spectral shifts; optical coherence theory; radiometry; spectroscopy; Wolf shifts.

Accepted: April 25, 1994

1. Introduction

In 1989 Kandpal, Vaishya, and Joshi of the National Physical Laboratory (India) published results of experiments in which they observed spectral shifts caused by a simple optical system [1]. In their experiments light from the exit aperture of an integrating sphere was imaged by a two lens system which had an interference filter between the lenses. They found that when a small aperture was placed in the image plane, the peak wavelength of the spectrum of the light measured on-axis in the far zone of the aperture was shifted away from the peak wavelength which occurred when no aperture was used.

They interpreted these shifts as being "Wolf shifts," i.e., coherence-induced shifts of the type predicted by Wolf [2-3]. Their explanation [4] for the occurrence of the shifts was based on their assertion (which was neither proven directly experimentally nor justified theoretically) that the filter-lens combination eliminated the wavelength dependence of the complex degree of spectral coherence of the light in the image plane. If this assertion is true, then a coherence-induced change of the spectrum is to be expected when the light propagates from the image plane to the far zone, as was shown experimentally by Morris and Faklis [5] and Faklis and Morris [6]. Kandpal et al. [4] then argued that the introduction of a circular aperture in the image plane helped in modifying the spectrum in the far zone.

In this paper, we will analyze theoretically the optical system used in Ref. [1]. Many details of the system which did not appear in Ref. [1] were provided to us by the group at NPL [7]. The basic outline of our paper is as follows. In Sec. 2 the optical system is described, and the basic assumptions to be made throughout the paper are stated. In Sec. 3 an approximate form for the cross-spectral density of the light in the image plane is obtained, and it is shown that the corresponding

complex degree of spectral coherence is *not* wavelength independent. In Sec. 4 the spectrum of the light on-axis in the observation plane is investigated, and an approximate form for it is obtained, both for the case in which no aperture is used, and for the case in which the small aperture is used. It is shown that, for each interference filter used, the peak wavelength in the latter case is shifted with respect to the peak wavelength in the former case. However, these shifts do not agree with those observed by the group at NPL. The shifts predicted by our analysis are much smaller, so small as to be unobservable to within the accuracy of their experiments. In Sec. 5 our conclusions are presented.

2. Optical System, Notation and Assumptions

The optical system used by the group at NPL is pictured in Fig. 1. Just to the left of plane I there is an integrating sphere of radius 25 cm which has a 450 W tungsten halogen lamp at its center. Plane I contains the exit aperture, σ, of the integrating sphere. σ is a circular aperture of radius $a_0 = 0.12$ cm, and it is in the front focal plane of a lens, L_1, whose focal length is $f_1 = 5$ cm and radius is $a_1 = 0.45$ cm.

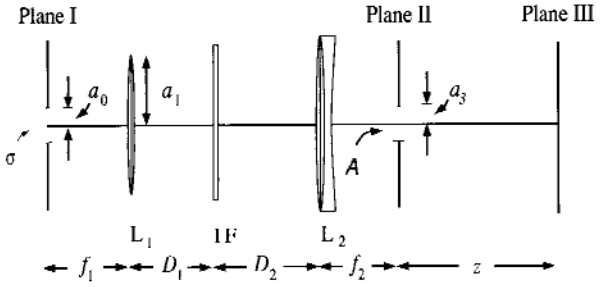

Fig. 1. The optical system.

There is an interference filter, IF, a distance $D_1 = 20$ cm behind L_1, and a second lens, L_2, a distance $D_2 = 30$ cm behind the interference filter. L_2 is an achromatic doublet, and its focal length is $f_2 = 20$ cm. The radius of L_2 and the transverse dimensions of the interference filter are all larger than a_1. It is therefore a straightforward matter to show that L_1 is the aperture stop for the system. In their experiments the group at NPL used six different interference filters; the shortest peak transmission wavelength used was 422 nm and the longest was 652 nm.

The exit aperture of the integrating sphere is imaged onto plane II. It is a straightforward matter to show that the magnification, M, for this imaging is:

$$M = -f_2/f_1 = -4\,. \tag{1}$$

Two different kinds of things were done as regards plane II. In one case an aperture A of radius a_3 was placed in plane II, and experiments were done for the values $a_3 = 0.012$ cm and 0.50 cm.[1] In the second case no aperture was used. For notational convenience, we will treat the latter case as if an aperture A of radius $a_3 = \infty$ were in plane II. Also, since no shift was observed in the 0.50 cm aperture case, we will not consider that case. Plane III, where the measurements were made, is a distance $z = 100$ cm from plane II.

The following notation will be used throughout this paper. The locations with respect to the optical axis of points in the planes I, II, and III will be specified by, respectively, the two-dimensional position vectors $\boldsymbol{\rho}''$, $\boldsymbol{\rho}'$, and $\boldsymbol{\rho}$. The locations of points in the planes occupied by L_1 and L_2 will be specified by, respectively, the two-dimensional position vectors $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. The lengths of vectors will be denoted by the corresponding non-boldfaced symbols, e.g., $\rho'' = |\boldsymbol{\rho}''|$.

The following assumptions will be made. As concerns the properties of the light coming from the integrating sphere, it will be assumed that the exit aperture of the integrating sphere radiates as a uniform lambertian source. Let the spectral radiance of this source, at angular frequency ω, be denoted by $B_0(\omega)$. The cross-spectral density, $W^{(1)}(\boldsymbol{\rho}_1'', \boldsymbol{\rho}_2'', \omega)$ of the light in plane I can be written as [9–11]:[2]

$$W^{(1)}(\boldsymbol{\rho}_1'', \boldsymbol{\rho}_2'', \omega) = 2\pi B_0(\omega)$$

$$\times j_0(k|\boldsymbol{\rho}_2'' - \boldsymbol{\rho}_1''|) P_0(\boldsymbol{\rho}_1'') P_0(\boldsymbol{\rho}_2''), \tag{2}$$

where k is the wavenumber of the light,

$$k = \omega/c = 2\pi/\lambda\,, \tag{3}$$

j_0 is the spherical Bessel function of the first kind of order zero,

[1] In Ref. [1] the diameter of the smaller aperture was incorrectly reported as 0.24 cm. The correct value is 0.024 cm [8].

[2] In order for Eq. (2) to be appropriate, the radius of the source must be much larger than the wavelength of the light. In the NPL experiments the wavelengths of interest are visible and the radius of the exit aperture of the integrating sphere is 0.12 cm, so this condition is fulfilled.

$$j_0(u)=\frac{\sin(u)}{u}, \tag{4}$$

and P_0 is the pupil function for the exit aperture of the integrating sphere,

$$P_0(\rho'')=\mathrm{circ}(\rho''/a_0). \tag{5}$$

In Eq. (3) c is the speed of light in vacuo and λ is the wavelength of the light. In Eq. (5) circ is the circle function,

$$\begin{aligned}\mathrm{circ}(u)&=1, \text{ if } |u|\leq 1,\\ &=0, \text{ if } |u|>1.\end{aligned} \tag{6}$$

Secondly, since L_1 is the aperture stop for the system, it will be assumed that the finiteness of the transverse sizes of L_2 and the interference filter can be neglected. Thirdly, it will be assumed that the transmittance of the interference filter may be described by the Lissberger-Wilcock model [12–13], and that its behavior is such that the approximation described in Sec. 4.1 is valid. In addition, both lenses will be treated as thin lenses, and the possibility of chromatic aberration introduced by either L_1 or L_2 will not be taken into account. The paraxial approximation will be used throughout our work.

3. Cross-Spectral Density of the Light in Plane II

In this section, the cross-spectral density in plane II will be investigated. First the coherent impulse response function for the propagation from plane I to plane II will be determined. This result will then be used to obtain an "exact" (within the paraxial approximation) expression for the cross-spectral density incident upon plane II. Approximations appropriate to the NPL experiments will then be used to obtain an approximate expression for the cross-spectral density in plane II. The corresponding spectrum and complex degree of spectral coherence in plane II will then be calculated and discussed.

3.1 Coherent Impulse Response Function for the Propagation from Plane I to Plane II

Let P'' be a point in plane I located by position vector $\boldsymbol{\rho}''$, and let P′ be a point in plane II located by position vector $\boldsymbol{\rho}'$. The coherent impulse response function, $h(\boldsymbol{\rho}', \boldsymbol{\rho}'', \omega)$, for the propagation from plane I to plane II is the field at P′ due to a monochromatic, unit amplitude point source of angular frequency ω at P″. In order to calculate this function we must first discuss the effect of the interference filter.

The effect of the interference filter on a plane wave is depicted in Fig. 2. Here a unit amplitude, monochromatic plane wave of angular frequency ω propagating in the direction specified by the unit vector $\boldsymbol{s}$ is incident upon the filter at an angle of incidence θ. If, as we have assumed, the finiteness of the transverse size of the filter can be neglected, the effect of the interference filter is to change the amplitude of the plane wave by the factor $t(\theta, \omega)$, the amplitude transmission function of the interference filter.

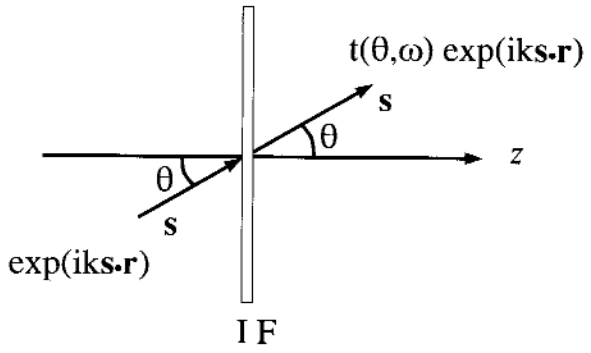

Fig. 2. The effect of the interference filter on a plane wave.

When this effect is taken into account, it can be shown (see Appendix A) that in the paraxial approximation $h(\boldsymbol{\rho}', \boldsymbol{\rho}'', \omega)$ is the product of $t(\rho'/f_2, \omega)$, the amplitude transmission function of the interference filter evaluated at $\theta=\rho'/f_2$, and $h_N(\boldsymbol{\rho}', \boldsymbol{\rho}'', \omega)$, the coherent impulse response function (paraxial form) for the case when there is no interference filter present, i.e.,

$$h(\boldsymbol{\rho}', \boldsymbol{\rho}'', \omega)=t(\rho'/f_2, \omega)h_N(\boldsymbol{\rho}', \boldsymbol{\rho}'', \omega), \tag{7}$$

where

$$h_N(\boldsymbol{\rho}', \boldsymbol{\rho}'', \omega)=C(\omega)\exp\{ik[\phi(\rho'')+\varkappa(\rho')]\}$$

$$\times \mathrm{Besinc}\left(\frac{ka_1}{f_1}\left|\boldsymbol{\rho}''-\frac{1}{M}\boldsymbol{\rho}'\right|\right) \tag{8}$$

and

$$C(\omega)=\frac{1}{M}\frac{\pi a_1^2}{\lambda^2 f_1^2} \tag{9}$$

$$\phi(\rho'')=f_1+(\rho''^2/2f_1), \tag{10}$$

$$\varkappa(\rho')=f_2+D+\left(1-\frac{D}{f_2}\right)(\rho'^2/2f_2), \tag{11}$$

$$\text{Besinc}\,(u)=2J_1(u)/u\ . \tag{12}$$

In Eq. (11) D is the distance from L_1 to L_2, i.e., $D=D_1+D_2$. In Eq. (12) J_1 is the Bessel function of the first kind of order one.

3.2 Cross-Spectral Density of the Light Incident Upon Plane II

The cross-spectral density of the light incident upon plane II, $W^{(\mathrm{II}-)}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$, is related to the cross-spectral density in the plane I by the expression [14]

$$W^{(\mathrm{II}-)}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)=\int_\sigma\int_\sigma h^*(\boldsymbol{\rho}_1', \boldsymbol{\rho}_1'', \omega)\, h(\boldsymbol{\rho}_2', \boldsymbol{\rho}_2'', \omega)$$

$$\times W^{(\mathrm{I})}(\boldsymbol{\rho}_1'', \boldsymbol{\rho}_2'', \omega)\,\mathrm{d}^2\rho_1''\,\mathrm{d}^2\rho_2'', \tag{13}$$

where $h(\boldsymbol{\rho}', \boldsymbol{\rho}'', \omega)$ is the coherent impulse response function for the propagation from plane I to plane II. Upon substituting Eqs. (2) and (7) into Eq. (13), and using Eq. (8) in the result, we find that

$$W^{(\mathrm{II}-)}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)=2\pi B_0(\omega)t^*(\rho_1'/f_2, \omega)t(\rho_2'/f_2, \omega)$$

$$\times F(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)\ , \tag{14}$$

where

$$F(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)=\mathrm{C}^2(\omega)E(\rho_1', \rho_2'', \omega)$$

$$\times\int_\sigma\int_\sigma \bar{E}\,(\rho_1'', \rho_2'', \omega)\ \text{Besinc}\left(\frac{ka_1}{f_1}\left|\boldsymbol{\rho}_1''-\frac{1}{M}\boldsymbol{\rho}_1'\right|\right)$$

$$\times\text{Besinc}\left(\frac{ka_1}{f_1}\left|\boldsymbol{\rho}_2''-\frac{1}{M}\boldsymbol{\rho}_2'\right|\right)j_0\left(k\left|\boldsymbol{\rho}_2''-\boldsymbol{\rho}_1''\right|\right)$$

$$\times\mathrm{d}^2\rho_1''\,\mathrm{d}^2\rho_2'', \tag{15}$$

and $E(\rho_1', \rho_2', \omega)$ and $\bar{E}(\rho_1'', \rho_2'', \omega)$ are phase factors:

$$E\,(\rho_1', \rho_2', \omega)=\exp\left[ik\left(1-\frac{D}{f_2}\right)(\rho_2'^2-\rho_1'^2)/2f_2\right]\ , \tag{16}$$

$$\bar{E}(\rho_1'', \rho_2'', \omega)=\exp\left[ik\left(\rho_2''^2-\rho_1''^2\right)/2f_1\right]. \tag{17}$$

Equation (14), with $F(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$ given by Eq. (15), is an exact (within the paraxial approximation) expression for the cross-spectral density incident upon plane II. However, the integral on the right hand side of Eq. (15) cannot be evaluated analytically. Nevertheless, for the NPL system certain approximations are appropriate which simplify $F(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$ considerably.

First, it is shown in Appendix B that for the NPL parameter values we may set $\bar{E}(\rho_1'', \rho_2'', \omega)\approx 1$ in the integrand on the right hand side of Eq. (15). Hence, to a good approximation

$$F(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)=C^2(\omega)E(\rho_1', \rho_2', \omega)\int_\sigma\int_\sigma$$

$$\times\text{Besinc}\left(\frac{ka_1}{f_1}\left|\boldsymbol{\rho}_1''-\frac{1}{M}\boldsymbol{\rho}_1'\right|\right)\text{Besinc}\left(\frac{ka_1}{f_1}\left|\boldsymbol{\rho}_2''-\frac{1}{M}\boldsymbol{\rho}_2'\right|\right)$$

$$\times j_0(k|\boldsymbol{\rho}_2''-\boldsymbol{\rho}_1''|)\,\mathrm{d}^2\rho_1''\,\mathrm{d}^2\rho_2''\ . \tag{18}$$

Next, it is shown in Appendix C that for the NPL parameter values,

$$\int_\sigma\int_\sigma \text{Besinc}\left(\frac{ka_1}{f_1}\left|\boldsymbol{\rho}_1''-\frac{1}{M}\boldsymbol{\rho}_1'\right|\right)$$

$$\times\text{Besinc}\left(\frac{ka_1}{f_1}\left|\boldsymbol{\rho}_2''-\frac{1}{M}\boldsymbol{\rho}_2'\right|\right)j_0(k|\boldsymbol{\rho}_2''-\boldsymbol{\rho}_1''|)\mathrm{d}^2\rho_1''\,\mathrm{d}^2\rho_2''$$

$$\approx\frac{1}{2C^2(\omega)}\left(\frac{a_1}{f_2}\right)^2\text{Besinc}\left(\frac{ka_1}{f_2}\left|\boldsymbol{\rho}_2'-\boldsymbol{\rho}_1'\right|\right)$$

$$\times\text{circ}(\boldsymbol{\rho}_1'/|M|a_0)\ \text{circ}(\boldsymbol{\rho}_2'/|M|a_0)\ . \tag{19}$$

It then follows from Eqs. (14), (18), and (19) that the cross-spectral density of the light incident upon plane II is given, to a good approximation, by the expression

$$W^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega) = \frac{\pi a_1^2}{f_2^2} B_0(\omega) t^*(\boldsymbol{\rho}_1'/f_2, \omega)$$

$$\times t(\boldsymbol{\rho}_2'/f_2, \omega) E(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega) \operatorname{Besinc}\left(\frac{k a_1}{f_2} \left|\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1'\right|\right)$$

$$\times \operatorname{circ}(\boldsymbol{\rho}_1'/|M|a_0) \operatorname{circ}(\boldsymbol{\rho}_2'/|M|a_0). \quad (20)$$

The spectrum, $S^{(\mathrm{II}-)}(\boldsymbol{\rho}', \omega)$, of the light incident upon plane II at the point P′ located by the vector $\boldsymbol{\rho}'$ and at the angular frequency ω is obtained from the corresponding cross-spectral density by setting $\boldsymbol{\rho}_1' = \boldsymbol{\rho}_2' = \boldsymbol{\rho}'$.[3] It therefore follows from Eq. (20) that

$$S^{(\mathrm{II}-)}(\boldsymbol{\rho}', \omega) =$$

$$\frac{\pi a_1^2}{f_2^2} B_0(\omega) |t(\boldsymbol{\rho}'/f_2, \omega)|^2 \operatorname{circ}(\boldsymbol{\rho}'/|M|a_0). \quad (21)$$

It can be seen from this expression that this spectrum is nonuniform (since t depends on $\boldsymbol{\rho}'$) and occupies the region $\rho' \leq |M|a_0$, i.e., the region which corresponds to the geometrical optics image of σ, the exit aperture of the integrating sphere.

3.3 Cross-Spectral Density, Spectrum, and Complex Degree of Spectral Coherence in Plane II

The cross-spectral density, $W^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$ in plane II (with the effect of the aperture A taken into account) is given by the expression

$$W^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega) =$$

$$W^{(\mathrm{II}-)}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega) \operatorname{circ}(\boldsymbol{\rho}_1'/a_3) \operatorname{circ}(\boldsymbol{\rho}_2'/a_3). \quad (22)$$

Upon substituting the approximate Eq. (20) into Eq. (22), we find that the corresponding approximate expression for the cross-spectral density of the light in plane II is

$$W^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega) = \frac{\pi a_1^2}{f_2^2} B_0(\omega) t^*(\boldsymbol{\rho}_1'/f_2, \omega)$$

$$\times t(\boldsymbol{\rho}_2'/f_2, \omega) E(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega) \operatorname{Besinc}\left(\frac{k a_1}{f_2} \left|\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1'\right|\right)$$

$$\times \operatorname{circ}(\boldsymbol{\rho}_1'/a) \operatorname{circ}(\boldsymbol{\rho}_2'/a), \quad (23)$$

where a is the smaller of $|M|a_0$, and a_3, i.e., $a = \min(|M|a_0, a_3)$.

The physical significance of the parameter a is that it is the radius of the "secondary source." If $a_3 < |M|a_0$ part of the incident light is blocked by the aperture A, and therefore the radius of the secondary source is a_3. However, if $a_3 > |M|a_0$ the aperture does not (within the approximation we are using) block any of the light, therefore $|M|a_0$ is the radius of the secondary source. The distinction between these two cases is necessary because when the small aperture is used $a_3 < |M|a_0$, whereas when no aperture is used $a_3 > |M|a_0$.

The spectrum, $S^{(\mathrm{II})}(\boldsymbol{\rho}', \omega)$, of the light in plane II can be calculated by using the result in Eq. (23):

$$S^{(\mathrm{II})}(\boldsymbol{\rho}', \omega) = W^{(\mathrm{II})}(\boldsymbol{\rho}', \boldsymbol{\rho}', \omega),$$

$$= \frac{\pi a_1^2}{f_2^2} B_0(\omega) |t(\boldsymbol{\rho}'/f_2, \omega)|^2 \operatorname{circ}(\boldsymbol{\rho}'/a). \quad (24)$$

We see from this expression that the secondary source occupies a circular domain of radius a, and that the spectrum of the light inside the secondary source is proportional to the product of spectral radiance in plane I, $B_0(\omega)$, and transmittance of the interference filter, $|t(\boldsymbol{\rho}'/f_2, \omega)|^2$.

The complex degree of spectral coherence, $\mu^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$, of the light in the plane II is defined by the formula [16]:

$$\mu^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega) = \frac{W^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)}{\sqrt{S^{(\mathrm{II})}(\boldsymbol{\rho}_1', \omega)} \sqrt{S^{(\mathrm{II})}(\boldsymbol{\rho}_2', \omega)}}. \quad (25)$$

Upon using Eqs. (23) and (24) in Eq. (25), we find that when $\rho_1' < a$ and $\rho_2' < a$, i.e., for pairs of points located inside the secondary source,

[3] The spectrum S is also sometimes referred to as the spectral density of the light. It is the optical intensity per unit angular frequency at the position and angular frequency in question [15].

$$\mu^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega) =$$

$$E(\rho_1', \rho_2', \omega)\,\frac{t^*(\rho_1'/f_2, \omega)}{|t(\rho_1'/f_2, \omega)|}\,\frac{t(\rho_2'/f_2, \omega)}{|t(\rho_2'/f_2, \omega)|}$$

$$\times \mathrm{Besinc}\left(\frac{ka_1}{f_2}\,|\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1'|\right). \tag{26}$$

This result shows that *the group at NPL was incorrect in stating that the filter-lens combination eliminates the wavelength dependence of the complex degree of spectral coherence of the light in plane II.* In particular, since the first three factors on the right hand side of Eq. (26) are each unimodular, it follows that the modulus of $\mu^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$ is equal to the absolute value of $\mathrm{Besinc}(ka_1|\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1'|/f_2)$. Hence, the assertion by the group at NPL that the wavelength independence of $\mu^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$ causes the spectrum to change upon propagation from plane II to plane III is not correct. There may be a change of the spectrum (we will investigate that in the next section), but it is not caused by a wavelength independence of $\mu^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$.

In the next section we will refer to the effective correlation length, $L(\omega)$, of the light inside the secondary source. The correlation length of a Besinc type correlation function is typically taken as the smallest separation, $|\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1'|$, for which the complex degree of spectral coherence takes on the value zero. Since Besinc has its first zero when its argument is equal to 3.832, it follows from Eqs. (26) and (3) that

$$L(\omega) = 0.610\,\frac{\lambda f_2}{a_1}. \tag{27}$$

4. Spectrum of the Light On-Axis in Plane III

Let P be a point in plane III located by position vector $\boldsymbol{\rho}$. When the Fresnel approximation is used to propagate the field from plane II to plane III, the spectrum of the light at P is given by the expression [17]

$$S^{(\mathrm{III})}(\boldsymbol{\rho}, \omega) =$$

$$\frac{1}{(\lambda z)^2}\iint W^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)\,\exp[ik(\rho_2''^2 - \rho_1''^2)/2z]$$

$$\times \exp[-ik(\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1')\cdot\boldsymbol{\rho}/z]\,\mathrm{d}^2\rho_1'\,\mathrm{d}^2\rho_2', \tag{28}$$

where $W^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$ is given by Eq. (23), and the integrations extend formally from $-\infty$ to ∞ in each variable.[4]

If the secondary source obeys the Leader condition [18]

$$\frac{kaL(\omega)}{z} << 1, \tag{29}$$

then we are in the far zone of the secondary source and the quadratic exponential term in Eq. (28) can be dropped. We will restrict our considerations to the on-axis observation point, $\boldsymbol{\rho} = 0$.

4.1 No Aperture in Plane II

In this case $a = |M|a_0$, and the Leader condition is not fulfilled for the NPL parameter values. The on-axis spectrum is therefore given by the expression

$$S^{(\mathrm{III})}(0, \omega) = \frac{1}{(\lambda z)^2}\iint W^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$$

$$\times \exp[ik(\rho_2''^2 - \rho_1''^2)/2z]\,\mathrm{d}^2\rho_1'\,\mathrm{d}^2\rho_2', \tag{30}$$

where $W^{(\mathrm{II})}(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega)$ is given by Eq. (23), with $a = |M|a_0$.

In order to investigate $S^{(\mathrm{III})}(0, \omega)$, let us first introduce average and difference vectors, $\bar{\boldsymbol{\rho}}'$ and $\tilde{\boldsymbol{\rho}}'$, in plane II:

$$\bar{\boldsymbol{\rho}}' = \tfrac{1}{2}(\boldsymbol{\rho}_2' + \boldsymbol{\rho}_1'), \tag{31a}$$

$$\tilde{\boldsymbol{\rho}}' = \boldsymbol{\rho}_2' - \boldsymbol{\rho}_1'. \tag{31b}$$

In terms of these variables, it follows from Eqs. (30), (23), (16) and (27) that

$$S^{(\mathrm{III})}(0, \omega) = \frac{\pi a_1^2}{\lambda^2 z^2 f_2^2}\,B_0(\omega)\iint G(\bar{\boldsymbol{\rho}}', \tilde{\boldsymbol{\rho}}', \omega)$$

$$\times \exp(-ik\bar{\boldsymbol{\rho}}'\cdot\tilde{\boldsymbol{\rho}}'/d)\,\mathrm{d}^2\bar{\boldsymbol{\rho}}'\,\mathrm{d}^2\tilde{\boldsymbol{\rho}}', \tag{32}$$

where

$$G(\bar{\boldsymbol{\rho}}', \tilde{\boldsymbol{\rho}}', \omega) = t^*(|\bar{\boldsymbol{\rho}}' - \tfrac{1}{2}\tilde{\boldsymbol{\rho}}'|/f_2, \omega)$$

$$\times t(|\bar{\boldsymbol{\rho}}' + \tfrac{1}{2}\tilde{\boldsymbol{\rho}}'|/f_2, \omega)\ \mathrm{Besinc}[3.832\tilde{\rho}'/L(\omega)]$$

$$\times \mathrm{circ}[(\bar{\boldsymbol{\rho}}' - \tfrac{1}{2}\tilde{\boldsymbol{\rho}}')/|M|a_0]\,\mathrm{circ}[(\bar{\boldsymbol{\rho}}' + \tfrac{1}{2}\tilde{\boldsymbol{\rho}}')/|M|a_0], \tag{33}$$

[4] In practice, the circ functions in Eq. (23) limit the domains of integration to the area occupied by the secondary source.

and

$$\frac{1}{d} = \frac{1}{f_2}\left(\frac{D}{f_2} - 1\right) - \frac{1}{z}. \tag{34}$$

Let us now examine the behavior of $G(\bar{\rho}', \tilde{\rho}', \omega)$ as a function of the difference vector $\tilde{\rho}'$ for a fixed value of the average vector $\bar{\rho}'$. The region over which the Besinc function is significant is the domain $\tilde{\rho}' \leq L(\omega)$. As $\tilde{\rho}'$explores this region the first arguments of t^* and t in Eq. (33) each change by at most the angle

$$\frac{L(\omega)}{f_2} = 0.610\,\frac{\lambda}{a_1}. \tag{35}$$

For visible light this is an extremely small angle. For example, for $\lambda = 550$ nm and $a_1 = 0.45$ cm, $L(\omega)/f_2 = 75$ μrad. Therefore we will assume that the behavior of the interference filter is such that when evaluating the integral on the right hand side of Eq. (32) we may, to a good approximation, neglect the $\tilde{\rho}'$ dependence of both t^* and t, i.e., that we may approximate G as

$$G(\bar{\rho}', \tilde{\rho}', \omega) = |t(\bar{\rho}'/f_2, \omega)|^2\,\mathrm{Besinc}[3.832\tilde{\rho}'/L(\omega)]$$

$$\times \mathrm{circ}[(\bar{\rho}' - \tfrac{1}{2}\tilde{\rho}')/|M|a_0]\mathrm{circ}[(\bar{\rho}' + \tfrac{1}{2}\tilde{\rho}')/|M|a_0]. \tag{36}$$

Furthermore, for the NPL system parameter values and the wavelengths used in their experiments, $L(\omega) << |M|a_0$. Therefore we will now use the quasihomogeneous approximation [19] which corresponds to replacing Eq. (36) by the expression

$$G(\bar{\rho}', \tilde{\rho}', \omega) =$$

$$|t(\bar{\rho}'/f_2, \omega)|^2\,\mathrm{Besinc}[3.832\tilde{\rho}'/L(\omega)]\,\mathrm{circ}(\bar{\rho}'/|M|a_0). \tag{37}$$

Upon substituting the approximate form Eq. (37) for G into Eq. (32), it can be shown, after some straightforward calculations, that the on-axis spectrum in plane III is given by the expression

$$S^{(\mathrm{III})}(0, \omega) = \frac{\pi b^2}{z^2} B_0(\omega) M(\omega), \tag{38}$$

where

$$M(\omega) = \frac{1}{\pi b^2}\int |t(\bar{\rho}'/f_2, \omega)|^2\,\mathrm{circ}(\bar{\rho}'/|M|a_0)$$

$$\times \mathrm{circ}(\bar{\rho}'/b)\mathrm{d}^2\bar{\rho}', \tag{39}$$

and $b = a_1 d/f_2$. For the NPL parameter values, $|M|a_0 = 0.48$ cm and $b = 0.35$ cm. Therefore, Eq. (39) simplifies to

$$M(\omega) = \frac{1}{\pi b^2}\int |t(\bar{\rho}'/f_2, \omega)|^2\,\mathrm{circ}(\bar{\rho}'/b)\mathrm{d}^2\bar{\rho}',$$

$$= \frac{2}{\beta^2}\int_0^{\beta} |t(\theta, \omega)|^2\,\theta\,\mathrm{d}\theta, \tag{40}$$

where $\beta = b/f_2$.

Let us now discuss the transmittance, $|t(\theta, \omega)|^2$, of the interference filter. We will assume that it can be described by the Lissberger-Wilcock model [12-13]. This model works well for Fabry-Perot and all-dielectric interference filters for angles of incidence which are less than 20°. It can also be used to describe, less accurately, double halfwave and induced transmission filters [20].

Let λ_0 denote the peak wavelength transmitted by the interference filter at normal incidence ($\theta = 0$), $(\Delta\lambda)_0$ denote its bandwidth (FWHM) at normal incidence, and T_0 denote its maximum transmittance. According to the Lissberger-Wilcock model the transmittance of the interference filter is given by the expression

$$|t(\theta, \omega)|^2 = T_0\left\{1 + \left[\frac{2(\lambda - \lambda_0)}{(\Delta\lambda)_0} + \frac{\lambda_0}{(\Delta\lambda)_0}\frac{\theta^2}{\eta^2}\right]^2\right\}^{-1}, \tag{41}$$

where η is the effective index of refraction of the interference filter.

The integral in Eq. (40) can be evaluated (see Ref. [12]) without any approximation, for the transmittance function Eq. (41). The result is that

$$M(\omega) = \frac{T_0}{2X}\tan^{-1}\left[\frac{2X}{1 + \xi(\xi + 2X)}\right], \tag{42}$$

where $X = \lambda_0\beta^2/[2\eta^2(\Delta\lambda)_0]$, and $\xi = 2(\lambda - \lambda_0)/(\Delta\lambda)_0$.

Some comments about the interference filter parameters λ_0, $(\Delta\lambda)_0$, and η are now in order. The effective indices of refraction, i.e., the values of η, for the filters used in the NPL experiments are not known. However, typical values of η are in the range $1.4<\eta<3.4$, with the lower values being more typical [21–25]. We will therefore use $\eta=2$ in our calculations. The precise values of λ_0 and $(\Delta\lambda)_0$ are also not known. For reasons which will be explained in Sec. 4.3, we will take the values of λ_0 for the six filters to be 421.9 nm, 484.0 nm, 512.4 nm, 566.0 nm, 609.1 nm, and 652.0 nm, with their bandwidths being, respectively, 9 nm, 9 nm, 5 nm, 13 nm, 8 nm, and 8 nm.

4.2 Small Aperture in Plane II

In this case $a_3=0.012$ cm and the Leader condition is fulfilled. It therefore follows from Eqs. (28) and (23) that the spectrum can be written as

$$S^{(\mathrm{III})}(0,\omega)=$$

$$\frac{\pi a_1^2}{\lambda^2 z^2 f_2^2}B_0(\omega)\int_{\mathrm{A}}\int_{\mathrm{A}} t^*(\rho_1'/f_2,\omega)t(\rho_2'/f_2,\omega)$$

$$\times E(\rho_1',\rho_2',\omega)\,\mathrm{Besinc}\left(\frac{ka_1}{f_2}|\boldsymbol{\rho}_2'-\boldsymbol{\rho}_1'|\right)\mathrm{d}^2\rho_1'\,\mathrm{d}^2\rho_2', \quad (43)$$

where A is the circular aperture of radius a_3. For the same reasons as were discussed in the Sec. 4.1, we will assume that the product of t^* and t on the right hand side of Eq. (43) can be replaced by $|t|^2$ evaluated at the average position, i.e., that to a good approximation,

$$S^{(\mathrm{III})}(0,\omega)=$$

$$\frac{\pi a_1^2}{\lambda^2 z^2 f_2^2}B_0(\omega)\int_{\mathrm{A}}\int_{\mathrm{A}}|t(|\boldsymbol{\rho}_1'+\boldsymbol{\rho}_2'|/2f_2,\omega)|^2\,E(\rho_1',\rho_2',\omega)$$

$$\times\,\mathrm{Besinc}\left(\frac{ka_1}{f_2}|\boldsymbol{\rho}_2'-\boldsymbol{\rho}_1'|\right)\mathrm{d}^2\rho_1'\,\mathrm{d}^2\rho_2', \quad (44)$$

For the NPL parameter values, it can be shown from Eq. (41) that, to a very good approximation, $|t(|\boldsymbol{\rho}_2'+\boldsymbol{\rho}_1'|/2f_2,\omega)|^2\approx|t(0,\omega)|^2$, for all $\boldsymbol{\rho}_1'$ and $\boldsymbol{\rho}_2'$ which are in A. The on-axis spectrum in plane III is therefore given, to a good approximation, by the expression

$$S^{(\mathrm{III})}(0,\omega)=\frac{\pi a_3^2}{z^2}B_0(\omega)|t(0,\omega)|^2\,N(\omega), \quad (45)$$

where

$$N(\omega)=\frac{a_1^2}{\lambda^2 a_3^2 f_2^2}\int_{\mathrm{A}}\int_{\mathrm{A}}E(\rho_1',\rho_2',\omega)$$

$$\times\,\mathrm{Besinc}\left(\frac{ka_1}{f_2}|\boldsymbol{\rho}_2'-\boldsymbol{\rho}_1'|\right)\mathrm{d}^2\rho_1'\,\mathrm{d}^2\rho_2'\,. \quad (46)$$

This integral can be evaluated by substituting the Fourier integral representation of the Besinc function [see Eq. (88)] into the integral, interchanging the orders of integration, and then recognizing the resultant integrals as familiar diffraction integrals. The result is that

$$N(\omega)=1-\sum_{s=0}^{\infty}\frac{(-1)^s}{2s+1}\left(\frac{u}{v}\right)^{2s}Q_{2s}(v)\,, \quad (47)$$

where

$$u=\frac{ka_3^2}{f_2}\left(\frac{D}{f_2}-1\right), \quad (48)$$

$$v=\frac{ka_3a_1}{f_2}\,, \quad (49)$$

and the $Q_{2s}(v)$ functions are those introduced by Wolf [26] and simplified by Petersen [27]:

$$Q_0(v)=J_0^2(v)+J_1^2(v)\,, \quad (50)$$

$$Q_{2s}(v)=-\frac{2s+1}{2s(s+1)}\{v[J_2(v)J_{2s+1}(v)$$

$$-J_1(v)J_{2s+2}(v)]+2sJ_1(v)J_{2s+1}(v)\}\,,\ \text{if } s>0. \quad (51)$$

4.3 Numerical Investigation of the Spectrum On-Axis in Plane III; Comparison to the NPL Experimental Values

Let $s_{\mathrm{N}}^{(\mathrm{III})}(0,\lambda)$ be the spectrum, as a function of wavelength, at the on-axis position in plane III, for the case in which there is no aperture in plane II. Let the wavelength at which this spectrum peaks be

denoted by λ_P and the bandwidth (FWHM) of this spectrum be denoted by $\Delta\lambda$. Let $s_A^{(III)}(0, \lambda)$ be the spectrum, as a function of wavelength, at the on-axis position in plane III, for the case where the small aperture is in plane II. Let λ_P' denote the peak wavelength for this spectrum. The group at NPL measured λ_P, $\Delta\lambda$, and λ_P' for six different interference filters [1]. Their results are shown in Table 1, along with the shift,

$$\delta\lambda_P = \lambda_P' - \lambda_P \,, \tag{52}$$

which occurred due to the insertion of the aperture.

Table 1. NPL experimental results

λ_P, in nm	$\Delta\lambda$, in nm	λ_P', in nm	$\delta\lambda_P$, in nm
422.0	9	421.0	−1.0
484.1	9	483.6	−0.5
512.4	5	514.1	1.7
566.1	13	564.1	−2.0
609.1	8	610.3	1.2
652.0	8	653.2	1.2

There are several things which should be noticed about the results shown in Table 1. First, the shifts are, in absolute value, of the order of 0.5 nm to 2.0 nm. Secondly, some of the shifts are blueshifts and some are redshifts. Furthermore, there are no obvious trends in the behaviors of the shifts as functions of either λ_P or $\Delta\lambda$.

Let us now investigate the shifts that our theory predicts. It follows from Eqs. (38) and (42) that the spectrum when there is no aperture in plane II is given by the expression

$$s_N^{(III)}(0, \lambda) = \frac{\pi b^2}{z^2} b_0(\lambda) m_N(\lambda), \tag{53}$$

where $b_0(\lambda)$ is the spectral radiance, as a function of wavelength, at the exit aperture of the integrating sphere and

$$m_N(\lambda) = \frac{T_0}{2X} \tan^{-1}\left[\frac{2X}{1+\xi(\xi+2X)}\right]. \tag{54}$$

Here ξ and X are given by the expressions which follow Eq. (42). It follows from Eqs. (45), (47), and (41) that the spectrum when the small aperture is in plane II is given by the expression

$$s_A^{(III)}(0, \lambda) = \frac{\pi a_1^2}{z^2} b_0(\lambda) m_A(\lambda), \tag{55}$$

where

$$m_A(\lambda) = T_0 N (2\pi c/\lambda)\left\{1+\left[\frac{2(\lambda-\lambda_0)}{(\Delta\lambda)_0}\right]^2\right\}^{-1}. \tag{56}$$

In order to simulate the experimental conditions, we did the following things. The color temperature of the lamp was 3200 K [7]. We therefore took the spectral radiance, $b_0(\lambda)$, to be a Planck distribution [28]:

$$b_0(\lambda) = \frac{2hc^2}{\lambda^5} \frac{1}{\exp(hc/\lambda k_B T) - 1} \,, \tag{57}$$

with a temperature of $T = 3200$ K. In Eq. (57), h is Planck's constant and k_B is Boltzmann's constant. Furthermore, we chose the center frequencies, λ_0, and the bandwidths, $(\Delta\lambda)_0$, of the interference filters to be such that the λ_P and $\Delta\lambda$ values obtained for the spectrum $s_N^{(III)}(0,\lambda)$ [from Eq. (53)] agreed with the experimental values of Table 1 to the number of decimal places being displayed there.

Table 2 lists the interference filter parameters, and λ_0 and $(\Delta\lambda)_0$, we used in our calculations and gives the shifts, $\delta\lambda_P$, predicted by our theory. These shifts were obtained by using a search routine to find λ_P' and λ_P from, respectively, Eqs. (55) and (53), and then subtracting the two values.

Table 2. Theoretical values for the shift

λ_0, in nm	$(\Delta\lambda)_0$, in nm	$\delta\lambda_P$, in nm
421.9	9	0.007
484.0	9	0.009
512.4	5	0.009
566.0	13	0.010
609.1	8	0.011
652.0	8	0.012

These shifts disagree with those measured by the group at NPL. The theoretical shifts are approximately two orders of magnitude smaller than the experimental shifts. Furthermore, the theoretical shifts are always redshifts, whereas the experimental shifts are in some cases redshifts and in other cases blueshifts.

At this point some further comments about the shifts predicted by our theory are in order. First, let us consider λ_P'. $s_A^{(III)}(0, \lambda)$ is the product of two wavelength dependent factors, $m_A(\lambda)$ and $b_0(\lambda)$. Due to the presence of the factor $N(2\pi c/\lambda)$, $m_A(\lambda)$ is not centered at λ_0, it is blueshifted from it by an amount of the order of 0.0001 nm. Multiplication by the Planck spectrum then causes $m_A(\lambda)$ to be redshifted by about 0.1 nm. Now, let us consider λ_P. $s_N^{(III)}(0, \lambda)$ is the product of two wavelength dependent factors, $m_N(\lambda)$ and $b_0(\lambda)$. $m_N(\lambda)$ is not centered at λ_0, it is blueshifted from this value by an amount of the order of 0.01 nm. Multiplication by the Planck spectrum then causes $m_N(\lambda)$ to be redshifted by about 0.1 nm. When we subtract λ_P from λ_P' the two Planck shifts cancel. Since the blueshift of $m_N(\lambda)$ is much larger than the blueshift of $m_A(\lambda)$, the resulting shift, $\delta\lambda_P$, is a redshift of the order of 0.01 nm.

The physical origin of this redshift is as follows. The field incident upon the interference filter can be represented as a superposition of polychromatic plane waves traveling in different directions, and the wavelength at which the transmittance of the interference filter peaks decreases as the angle of incidence of a plane wave increases [see Eq. (41)]. Therefore, for each polychromatic plane wave incident upon the interference filter at an oblique angle of incidence, the spectrum of the transmitted light has a peak wavelength which is less than it would be for a normally incident plane wave. In the case in which no aperture was used, the contributions from a significant set of such waves arrive at the on-axis observation point in plane III. However, in the case in which the small aperture was used, the contributions from a large subset of these waves were blocked at plane II and did not arrive at plane III. As a result, the spectrum in plane III when the small aperture is used peaks at a longer wavelength than it does when no aperture is used.

5. Conclusion

Our paper contains two separate sets of results. (1) In Sec. 3 an approximate form for the cross-spectral density of the light in plane II was obtained [Eq. (23)], and it was shown that the corresponding complex degree of spectral coherence contradicts the explanation for the shifts given in Ref. [1]. (2) In Sec. 4 further approximations were made to propagate the cross-spectral density obtained in Sec. 3 from plane II to plane III, and an approximate expression for the on-axis spectrum was obtained, both for the case in which no aperture is used, and for the case in which the small aperture is used. It was found that the peak wavelength of the spectrum in the latter case is shifted with respect to the peak wavelength for the former case. However the shifts predicted by our analysis are much smaller than those reported in Ref. [1], so small as to be unobservable to within the accuracy of their experiments.

This brings us to an important question. Our analysis is predicting no observable shift (to within the accuracy of the measurements of the group at NPL), and yet the group at NPL observed shifts; so where is this shift coming from? In our opinion there are three possibilities. One possibility is that their interference filters did not behave in the manner we have assumed in our calculations, and that shifts arose as a result. A second possibility has to do with the spatial coherence properties of the light in plane III. Since the secondary source created in plane II when no aperture is used has a radius which is forty times as large as in the case in which the small aperture is used, the spatial coherence properties of the light in plane III will be quite different in the two cases. If this difference is significant enough, it may be that the monochromator used in the detection process responds differently in the two cases. A third possibility, and in our opinion the best one, is that the shift is caused by chromatic aberration introduced by L_1, or L_2, or both. L_1 has a short focal length and is not an achromatic doublet, and is therefore an obvious candidate for introducing significant chromatic aberration. L_2 is an achromatic doublet. However, it can only be perfectly achromatic for two wavelengths, whereas filters with six quite different center wavelengths were used in the NPL experiments.

Finally, let us comment on the relevance of these possibilities for typical spectral irradiance measurement systems. Typical spectral irradiance measurement systems do not use interference filters. Therefore, if the shift is caused by a nonideal behavior of the interference filter, it is irrelevant for typical spectral irradiance measurements. As concerns the second possibility, it has been known for some time that the spatial coherence properties of the light incident upon a monochromator do effect its response, viz., its slit scattering function [29–30]. Hence any comparison of experimentally measured spectra which have significantly different spatial coherence properties at the entrance aperture of the monochromator should take this into account. Finally, if the shift is caused by chromatic aberration, it will not occur in spectral irradiance systems which use only mirrors, and is an effect which should obviously be taken into account in any system which uses lenses.

6. Appendix A: Coherent Impulse Response Function for the Propagation from Plane I to Plane II

6.1 Field Exiting L_1

Let Q be a point in the plane occupied by L_1, and let $\boldsymbol{\alpha}_1$ be the two-dimensional position vector in the plane which locates Q. The field, $U^{(1-)}(\boldsymbol{\alpha}_1, \omega)$, incident upon L_1 at the point Q due to a monochromatic, unit amplitude, point source at position P″ in plane I is given by the expression

$$U^{(1-)}(\boldsymbol{\alpha}_1, \omega) = \frac{1}{i\lambda} \frac{\exp(ik|\boldsymbol{r}_1 - \boldsymbol{\rho}''|)}{|\boldsymbol{r}_1 - \boldsymbol{\rho}''|}, \qquad (58)$$

where $\boldsymbol{r}_1 = (\boldsymbol{\alpha}_1, f_1)$. Since the angles involved are small, this expression may be approximated (paraxial approximation) as

$$U^{(1-)}(\boldsymbol{\alpha}_1, \omega) = \frac{\exp(ikf_1)}{i\lambda f_1} \exp[ik(\boldsymbol{\alpha}_1 - \boldsymbol{\rho}'')^2/2f_1]. \qquad (59)$$

The field exiting the lens, $U^{(1+)}(\boldsymbol{\alpha}_1, \omega)$, is given by the expression [31]

$$U^{(1+)}(\boldsymbol{\alpha}_1, \omega) =$$

$$U^{(1-)}(\boldsymbol{\alpha}_1, \omega)\exp(-ik\alpha_1^2/2f_1)P_1(\boldsymbol{\alpha}_1), \qquad (60)$$

where $P(\boldsymbol{\alpha}_1)$ is the pupil function for L_1, $P_1(\boldsymbol{\alpha}_1) = \mathrm{circ}(\boldsymbol{\alpha}_1/a_1)$. Using Eq. (59) in Eq. (60) yields

$$U^{(1+)}(\boldsymbol{\alpha}_1, \omega) =$$

$$\frac{\exp[ik\phi(\rho'')]}{i\lambda f_1} \exp(-ik\boldsymbol{\alpha}_1 \cdot \boldsymbol{\rho}''/f_1)P_1(\boldsymbol{\alpha}_1), \qquad (61)$$

where

$$\phi(\rho'') = f_1 + (\rho''^2/2f_1). \qquad (62)$$

In the next subsection we will need $\tilde{U}^{(1+)}(\boldsymbol{K}, \omega)$, the two-dimensional spatial Fourier transform of $U^{(1+)}(\boldsymbol{\alpha}_1, \omega)$,

$$\tilde{U}^{(1+)}(\boldsymbol{K}, \omega) =$$

$$\int U^{(1+)}(\boldsymbol{\alpha}_1, \omega) \exp(-i\boldsymbol{K} \cdot \boldsymbol{\alpha}_1)\mathrm{d}^2\alpha_1, \qquad (63)$$

evaluated at $\boldsymbol{K} = k\boldsymbol{s}_\perp$. It follows from Eqs. (61) and (63) that

$$\tilde{U}^{(1+)}(k\boldsymbol{s}_\perp, \omega) = \frac{\exp[ik\phi(\rho'')]}{i\lambda f_1} \tilde{P}_1[k(\boldsymbol{s}_\perp - \boldsymbol{s}_{0\perp})],$$

$$= \frac{\pi a_1^2}{i\lambda f_1} \exp[ik\phi(\rho'')]$$

$$\times \mathrm{Besinc}[ka_1|\boldsymbol{s}_\perp - \boldsymbol{s}_{0\perp}|], \qquad (64)$$

where $\boldsymbol{s}_{0\perp} = -\boldsymbol{\rho}''/f_1$.

6.2 Field Incident Upon L_2

Let P be a point in-between L_1 and the interference filter. Let $\boldsymbol{r}$ be a position vector which goes from the center of L_1 to P. The field at P can be represented as a superposition (angular spectrum) of plane waves [32]:

$$U(\boldsymbol{r}, \omega) = \int a(\boldsymbol{s}_\perp, \omega)\exp(ik\boldsymbol{s} \cdot \boldsymbol{r})\mathrm{d}^2 s_\perp. \qquad (65)$$

where $\boldsymbol{s} = (\boldsymbol{s}_\perp, s_z)$ is a unit vector,

$$s_z = \sqrt{1 - s_\perp^2}, \text{ if } s_\perp^2 \le 1,$$

$$= i\sqrt{s_\perp^2 - 1}, \text{ if } s_\perp^2 > 1, \qquad (66)$$

and

$$a(\boldsymbol{s}_\perp, \omega) = \frac{1}{\lambda^2} \tilde{U}^{(1+)}(k\boldsymbol{s}_\perp, \omega). \qquad (67)$$

In Eq. (66), $s_\perp = |\boldsymbol{s}_\perp|$.

In spherical polar coordinates, as measured from the center of L_1, $\boldsymbol{s} = (1, \theta, \phi)$. In this notation $\boldsymbol{s}_\perp = (s_x, s_y) = (\sin\theta\cos\phi, \sin\theta\sin\phi)$. Therefore the angle of incidence of the plane wave $\exp(ik\boldsymbol{s} \cdot \boldsymbol{r})$ with respect to the normal to the surface of the IF is θ [see Fig. 2], where $s_\perp = \sin\theta \approx \theta$.

Let $t(\theta, \omega)$ be the amplitude transmission function of the filter for a plane wave of angular frequency ω and angle of incidence θ. If we neglect the finiteness of the transverse size of the IF, the effect of the interference filter is to change the amplitude of each plane wave, $\exp(ik\boldsymbol{s} \cdot \boldsymbol{r})$, by the factor $t(s_\perp, \omega)$. The amplitude, call it $\hat{a}(\boldsymbol{s}_\perp, \omega)$, of the plane wave exiting the filter is therefore

$$\hat{a}(s_\perp, \omega) = t(s_\perp, \omega) a(s_\perp, \omega) ,$$

$$= \frac{1}{\lambda^2} t(s_\perp, \omega)\, \tilde{U}^{(1+)}(k s_\perp, \omega) ,$$

$$= \frac{\pi a_1^2}{i\lambda^3 f_1} t(s_\perp, \omega) \exp[ik\phi(\rho'')]$$

$$\times \text{Besinc}[k a_1 |s_\perp - s_{0\perp}|] . \quad (68)$$

Let $r_2 = (\alpha_2, D)$ be the position vector which goes from the center of L_1 to the point in L_2 located by the two-dimensional position vector α_2. The field, $U^{(2-)}(\alpha_2, \omega)$ incident on L_2 can be represented as a superposition of plane waves:

$$U^{(2-)}(\alpha_2, \omega) = \int \hat{a}(s_\perp, \omega) \exp(iks \cdot r) d^2 s_\perp ,$$

$$= \int \hat{a}(s_\perp, \omega) \exp(iks_\perp \cdot \alpha_2)$$

$$\times \exp(iks_z D) d^2 s_\perp . \quad (69)$$

Since the angles involved are small, we may make the approximation $s_z = \sqrt{1 - s_\perp^2} \approx 1 - \frac{1}{2} s_\perp^2$ in Eq. (69), and we find that

$$U^{(2-)}(\alpha_2, \omega) = \exp(ikD) \int \hat{a}(s_1, \omega) \exp(iks_\perp \cdot \alpha_2)$$

$$\times \exp(-ikDs_1^2/2) d^2 s_\perp . \quad (70)$$

6.3 Final Expression

The coherent impulse response, $h(\rho', \rho'', \omega)$, for the propagation from plane I to plane II is the field in plane II due to the point source in plane I, i.e.,

$$h(\rho', \rho'', \omega) = U^{(II)}(\rho', \omega) . \quad (71)$$

If we neglect the finite transverse size of L_2, this equation can be written as [33]

$$h(\rho', \rho'', \omega) = \frac{\exp[ik\psi(\rho')]}{i\lambda f_2} \tilde{U}^{(2-)}(k\rho'/f_2, \omega), \quad (72)$$

where $\psi(\rho') = f_2 + (\rho'^2/2f_2)$. It follows from Eq. (70) that

$$\tilde{U}^{(2-)}(k s_\perp, \omega) =$$

$$\lambda^2 \hat{a}(s_\perp, \omega) \exp(ikD) \exp(-ikDs_1^2/2). \quad (73)$$

Upon substituting Eq. (73) in Eq. (72) and Eq. (68) in the result we find that

$$h(\rho', \rho'', \omega) =$$

$$-\frac{\pi a_1^2}{\lambda^2 f_1 f_2} \exp\{ik[\phi(\rho'') + \chi(\rho')]\}$$

$$\times \text{Besinc}(k a_1 |(\rho'/f_2) - (\rho''/f_1)|)\, t(\rho'/f_2, \omega) , \quad (74)$$

where

$$\chi(\rho') = f_2 + D + \left(1 - \frac{D}{f_2}\right) (\rho'^2/2f_2) . \quad (75)$$

The magnification of the imaging from plane I to plane II is $M = -f_2/f_1$. It therefore follows from Eq. (74) that the coherent impulse response function can be written as

$$h(\rho', \rho'', \omega) = t(\rho'/f_2, \omega) h_N(\rho', \rho'', \omega), \quad (76)$$

where $h_N(\rho', \rho'', \omega)$ is the coherent impulse response function if there is no interference filter,

$$h_N(\rho', \rho'', \omega) = C(\omega) \exp\{ik[\phi(\rho'') + \chi(\rho')]\}$$

$$\times \text{Besinc}\left(\frac{k a_1}{f_1} \left| \rho'' - \frac{1}{M} \rho' \right|\right) . \quad (77)$$

Here $\phi(\rho'')$ and $\chi(\rho')$ are given by Eqs. (62) and (75) respectively, and

$$C(\omega) = \frac{1}{M} \frac{\pi a_1^2}{\lambda^2 f_1^2} . \quad (78)$$

7. Appendix B: Derivation of $\overline{E}(\rho_1'', \rho_2'', \omega) \approx 1$

Let us first define average and difference variables in plane I as, respectively,

$$\bar{\rho}'' = \tfrac{1}{2} (\rho_2'' + \rho_1''), \quad (79a)$$

$$\tilde{\rho}'' = \rho_2'' - \rho_1'' . \quad (79b)$$

In terms of these variables $\overline{E}(\rho_1'', \rho_2'', \omega)$ [see Eq. (17)] can be rewritten as

$$\overline{E}(\rho_1'', \rho_2'', \omega) = \exp\left(\frac{2\pi i}{\lambda f_1} \bar{\rho}'' \cdot \tilde{\rho}''\right) . \quad (80)$$

As ρ_1'' and ρ_2'' each explore the domain σ, the maximum value of $\bar{\rho}''=|\bar{\rho}''|$ is a_0. In the integrand of Eq. (15) the range of values of $\tilde{\rho}''=\rho_2''-\rho_1''$ over which j_0 is significant is $\tilde{\rho}''\leq\lambda/2$. Hence as ρ_1'' and ρ_2'' range over the domain where the integrand is significant,

$$\left|\frac{2\pi}{\lambda f_1}\bar{\rho}''\cdot\tilde{\rho}''\right|\leq\frac{\pi a_0}{f_1}. \tag{81}$$

For the NPL system $a_0=0.12$ cm and $f_1=5$ cm; therefore over the domain where the integrand is significant the real and imaginary parts of $\bar{E}(\rho_1'', \rho_2'', \omega)$ fulfill, respectively, the inequalities:

$$0.997\leq \mathrm{Re}[\bar{E}(\rho_1'', \rho_2'', \omega)]\leq 1, \tag{82a}$$

$$-0.075\leq \mathrm{Im}[\bar{E}(\rho_1'', \rho_2'', \omega)]\leq 0.075 \tag{82b}$$

8. Appendix C: Derivation of Eq. (19)

Let

$$B(\rho_1', \rho_2', \omega)=\int_\sigma\int_\sigma \mathrm{Besinc}\left(\frac{ka_1}{f_1}\left|\rho_1''-\frac{1}{M}\rho_1'\right|\right)$$

$$\times\mathrm{Besinc}\left(\frac{ka_1}{f_1}\left|\rho_2''-\frac{1}{M}\rho_2'\right|\right)j_0(k|\rho_2''-\rho_1''|)\,\mathrm{d}^2\rho_1''\mathrm{d}^2\rho_2'', \tag{83}$$

and let us make the change of variables

$$u_j=\rho_j''/a_0,\quad (j=1,2) \tag{84a}$$

$$v_j=\rho_j'/Ma_0,\ (j=1,2) \tag{84b}$$

In terms of these new variables, Eq. (83) can be rewritten as

$$B(\rho_1', \rho_2', \omega)=a_0^4\int_{u.d.}\int_{u.d.}\mathrm{Besinc}(\kappa_1|u_1-v_1|)$$

$$\times\mathrm{Besinc}(\kappa_1|u_2-v_2|)j_0(\kappa|u_2-u_1|)\mathrm{d}^2u_1\mathrm{d}^2u_2, \tag{85}$$

where $u.d.$ is the unit disk, $u.d.=\{u: u\leq 1\}$, and

$$\kappa_1=\frac{ka_1a_0}{f_1}, \tag{86a}$$

$$\kappa=ka_0. \tag{86b}$$

If $\kappa_1>>1$ then the first and second Besinc functions in the integrand on the right hand side of Eq. (85) are, respectively, very sharply peaked about the values $u_1=v_1=\rho_1'/Ma_0$ and $u_2=v_2=\rho_2'/Ma_0$. As a result, if v_1 or v_2 lies outside the unit disk, then the integral on the right hand side of Eq. (85) is approximately zero, and if v_1 and v_2 both lie within the unit disk the integrals in Eq. (85) may be approximated by extending the limits to $\pm\infty$, i.e.,

$$B(\rho_1', \rho_2', \omega)\approx a_0^4\,\mathrm{circ}(\rho_1'/Ma_0)\,\mathrm{circ}(\rho_2'/Ma_0)$$

$$\times\iint\mathrm{Besinc}(\kappa_1|u_1-v_1|)\,\mathrm{Besinc}(\kappa_1|u_2-v_2|)$$

$$\times j_0(\kappa|u_2-u_1|)\mathrm{d}^2u_1\mathrm{d}^2u_2. \tag{87}$$

The right hand side of Eq. (87) may be simplified by using the Fourier representations of the functions in the integrand:

$$\mathrm{Besinc}(\kappa_1|u_j-v_j|)=\frac{1}{\pi\kappa_1^2}\int\mathrm{circ}(K_j/\kappa_1)$$

$$\times\exp[-iK_j\cdot(u_j-v_j)]\,\mathrm{d}^2K_j,\ (j=1,2), \tag{88}$$

$$j_0(\kappa|u_2-u_1|)=$$

$$\frac{1}{2\pi\kappa^2}\int\frac{\mathrm{circ}(K/\kappa)}{\sqrt{1-(K/\kappa)^2}}\exp[-iK\cdot(u_2-u_1)]\mathrm{d}^2K. \tag{89}$$

Upon substituting Eqs. (88) and (89) into Eq. (87), we find, after performing some simple integrations, that

$$B(\rho_1', \rho_2', \omega)=D(\omega)\,\mathrm{circ}(\rho_1'/Ma_0)\,\mathrm{circ}(\rho_2'/Ma_0)$$

$$\times\int\frac{\mathrm{circ}(K/\kappa_1)\,\mathrm{circ}(K/\kappa)}{\sqrt{1-(K/\kappa)^2}}\exp[iK\cdot(v_2-v_1)\mathrm{d}^2K,$$

$$=D(\omega)\,\mathrm{circ}(\rho_1'/Ma_0)\,\mathrm{circ}(\rho_2'/Ma_0)\int\frac{\mathrm{circ}(K/\kappa_1)}{\sqrt{1-(K/\kappa)^2}}$$

$$\times\exp[iK\cdot(v_2-v_1)]\mathrm{d}^2K, \tag{90}$$

where

$$D(\omega)=\frac{(2\pi a_0)^4}{(\pi\kappa_1^2)^2\,2\pi\kappa^2}, \tag{91}$$

and the step performed in Eq. (90) follows from the fact that $\kappa_1/\kappa = a_1/f_1 < 1$.

If $(a_1/f_1)^2 << 1$, Eq. (90) can be approximated as

$$B(\boldsymbol{\rho}_1', \boldsymbol{\rho}_2', \omega) = D(\omega)\ \mathrm{circ}(\boldsymbol{\rho}_1'/Ma_0)\ \mathrm{circ}(\boldsymbol{\rho}_2'/Ma_0)$$

$$\times \int \mathrm{circ}(\boldsymbol{K}/\kappa_1)\exp[i\boldsymbol{K}\cdot(\boldsymbol{v}_2-\boldsymbol{v}_1)]\mathrm{d}^2K,$$

$$= D(\omega)\ \mathrm{circ}(\boldsymbol{\rho}_1'/Ma_0)\ \mathrm{circ}(\boldsymbol{\rho}_2'/Ma_0)\pi\kappa_1^2$$

$$\times \mathrm{Besinc}(\kappa_1|\boldsymbol{v}_2-\boldsymbol{v}_1|),$$

$$= \frac{1}{2C^2(\omega)}\left(\frac{a_1}{f_2}\right)^2 \mathrm{Besinc}\left(\frac{ka_1}{f_2}|\boldsymbol{\rho}_2'-\boldsymbol{\rho}_1'|\right)\ \mathrm{circ}(\boldsymbol{\rho}_1'/|M|a_0)$$

$$\times \mathrm{circ}(\boldsymbol{\rho}_2'/|M|a_0)\ . \qquad (92)$$

For the NPL system, $a_0 = 0.12$ cm and $a_1/f_1 = 0.09$, and the peak wavelengths of the interference filters used in their experiments were in the range 422 nm to 652 nm. For these parameter values, $1.04\times10^3 \le \frac{ka_1a_0}{f_1} \le 1.61\times10^3$ and $(a_1/f_1)^2 = 8.10\times10^{-3}$; therefore, $\kappa_1 >> 1$ and $(a_1/f_1)^2 << 1$. It then follows that the approximation Eq. (92) is appropriate for the NPL system.

Acknowledgments

We would like to thank Dr. H. C. Kandpal (NPL, New Delhi, India) for providing us with the details of the optical system. This work was supported by NIST under contract number P043NANB210406.

9. References

[1] H. C. Kandpal, J. S. Vaishya, and K. C. Joshi, Opt. Commun. **73**, 169 (1989).
[2] E. Wolf, Opt. Commun. **62**, 12 (1987).
[3] E. Wolf, Phys. Rev. Lett. **56**, 1370 (1986).
[4] Ref. [1], Sec. 3, second paragraph.
[5] G. M. Morris and D. Faklis, Opt. Commun. **62**, 5 (1987).
[6] D. Faklis and G. M. Morris, Opt. Lett. **13**, 4 (1988).
[7] H. C. Kandpal, private communication.
[8] H. C. Kandpal, J. S. Vaishya, and K. C. Joshi, to be published in Opt. Eng., June 1994.
[9] A. Walther, J. Opt. Soc. Am. **58**, 1256 (1968).
[10] W. H. Carter and E. Wolf, J. Opt. Soc. Am. **65**, 1067 (1975).
[11] J. T. Foley and E. Wolf, J. Mod. Opt. **38**, 2053 (1991), Sec. 3.1.
[12] P. H. Lissberger and W. L. Wilcock, J. Opt. Soc. Am. **49**, 126 (1959).
[13] P. H. Lissberger, J. Opt. Soc. Am. **58**, 1586 (1968).
[14] E. Wolf and J. R. Fienup, Opt. Commun. **82**, 209 (1991).
[15] L. Mandel and E. Wolf, J. Opt. Soc. Am. **66**, 529 (1976), Eqs. (1.2b) and (2.14).
[16] Ibid., Eq. (2.10).
[17] A. S. Marathay, Elements of Optical Coherence Theory, John Wiley & Sons, New York, (1982). See the equation on the top of p. 188 and set $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \boldsymbol{\rho}$.
[18] J. C. Leader, J. Opt. Soc. Am. **68**, 1332 (1978).
[19] W. H. Carter and E. Wolf, J. Opt. Soc. Am. **67**, 785 (1977).
[20] P. H. Lissberger, Rep. Prog. Phys. **33**, 197 (1970), Sec. 5.1.
[21] I. H. Blifford, Appl. Opt. **5**, 105 (1966).
[22] D. J. Hemingway and P. H. Lissberger, Appl. Opt. **6**, 471 (1967).
[23] R. H. Eather and D. L. Reasoner, Appl. Opt. **8**, 227 (1969).
[24] R. M. Lerner, Appl. Opt. **10**, 1914 (1971).
[25] M. Petitdidier, Appl. Opt. **22**, 1797 (1983).
[26] E. Wolf, Proc. R. Soc. A **204**, 533 (1951).
[27] R. W. Petersen, J. Opt. Soc. Am. **65**, 1519 (1975).
[28] R. W. Boyd, Radiometry and the Detection of Optical Radiation, John Wiley and Sons, Inc., New York (1983) p. 54.
[29] J. B. Shumaker, Introduction to Coherence in Radiometry, Natl. Bur. Stand. (U. S.) Tech. Note 910-6 (1983) pp. 37–44.
[30] K. D. Mielenz, J. Opt. Soc. Am. **57**, 66 (1967).
[31] J. W. Goodman, Introduction to Fourier Optics, McGraw-Hill, New York (1968) p. 83.
[32] C. J. Boukamp, Rep. Prog. Phys. **17**, 35 (1954).
[33] Ref. [31], p. 86.

About the authors: John T. Foley is Professor of Physics at Mississippi State University and a Fellow of the Optical Society of America. Ming Wang is a lecturer in the Physics Department at Mississippi State University.

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 281 (1994)]

Comments on the paper "Wolf Shifts and Their Physical Interpretation Under Laboratory Conditions" by K. D. Mielenz

In a recent paper [1] in the Journal of Research of the National Institute of Standards and Technology, K. D. Mielenz has criticized the generally accepted interpretation of a phenomenon discovered a few years ago, regarding frequency shifts of spectral lines due to coherence properties of sources. In these comments we show that much of the criticism is invalid.

The phenomenon was discovered in 1986 [2] and has been discussed extensively in about 100 papers published since then. In its broader sense the phenomenon concerns the effect of source correlations on the spectrum of the emitted radiation. The paper [1] contains a number of serious errors and misinterpretations, which make much of the criticism invalid or misleading at best.

In the third paragraph of Ref. [1], a passage is quoted from one of my papers on this subject [3] where I suggested some possible mechanisms for producing source correlations. However Mielenz does not state that several such mechanisms have been found since then. Some, in fact, are discussed in several references which Mielenz quotes (his Refs. [8–11]); others are discussed in Refs. [4] and [5] cited below.

Near the beginning of Sec. 2 of Ref. [1] Mielenz poses the question "Does the spectrum of partially coherent light change on propagation in free space, and are such changes consistent with the principle of energy conservation?" In spite of Mielenz's hedging on these questions, the answer to the first question is emphatically "yes, in general," and the answer to the second question is "yes, always." That spectral changes can take place on propagation in free-space has been demonstrated most dramatically by experiments of G. Indebetouw [6], in which a partially coherent planar secondary source of essentially uniform spectrum generated light which after propagating in free space exhibited a spectrum which at some points had highly oscillatory behavior. That changes of this kind are consistent with the principle of conservation of energy has been shown in Ref. [7].

In the same paragraph of Ref. [1] Mielenz states "The experiments cited here did not pertain to free-space propagation in a literal sense, but were diffraction or interference experiments...". This statement is simply untrue. Diffraction and interference have played a role in *constructing* the partially coherent secondary sources, but *the propagation from the secondary sources was in free space and has given rise to spectral changes.*

In Sec. 2 Mielenz also states the "spectral distribution...does not change along the path of a ray." I find it rather astonishing that Mielenz chooses to criticize the interpretation of a rather subtle optical phenomenon by using the most primitive model for light propagation[1]; and what does Mielenz mean by a "ray" in a partially coherent field?

In Sec. 3 Mielenz introduces terms, symbols and notions which follow an International Lighting Vocabulary. Useful as these concepts undoubtedly are for the purposes of practical radiometry, they are largely inappropriate for the analysis of correlation-induced spectral changes in which interference of partially coherent light, even in free space, demands analysis in terms of statistical wave theory. In this connection it may perhaps not be inappropriate to recall that the foundations of radiometry on the basis of modern theory of radiation has not been fully clarified to this day, in spite of numerous attempts that have been made over a period of several decades[2]. It is also to be noted the term "coherence" is not listed in the International Lighting Vocabulary.

In Sec. 4 Mielenz refers to Newton's famous experiment on recombination of "colors" dispersed by a prism. Newton's experiments, classic as they are, are irrelevant to elucidating the effect of correlation-induced spectral changes. The wave theory of light was not even formulated at Newton's time to explain his observation; and he did not use partially coherent sources.

Near the end of Sec. 4 there appears the statement that "The Helmholtz equations [Eq. (5)] apply to individual frequencies and thus appear to imply that spectra do not change on free-space propagation...". Evidently appearances can be deceptive, because already in the first publication on this subject (Ref. [2]) and in numerous subsequent publications, it was shown that the two Helmholtz equations which govern the propagation of the cross-spectral density do, in fact, predict that, in

[1] As early as 1909 the great physicist H. A. Lorentz stressed on p. 25 in his famous book "The Theory of Electrons" that "...in general it will not be possible to trace the path of parts of energy in the same sense in which we can follow in their course the ultimate particles of which matter is made up."

[2] Some of the problems encountered in attempts to clarify the foundations of radiometry are discussed in Ref. [8]. Many of the numerous papers on this subject are reprinted in an excellent collection edited by A. T. Friberg [9].

general, the spectrum of light changes in free-space propagation; and as already mentioned above, such changes have been demonstrated by experiments which have been carried out in several laboratories.

In Sec. 5 Mielenz criticizes an analysis and interpretation given in Refs. [10] and [11] relating to spectral changes that may arise in Young's interference experiment. He quotes two formulas for the spectrum in the interference pattern [his Eqs. (24) and (27)], which pertain to interference of light from a spatially incoherent source. For this particular case it is true, as Mielenz implies, that the same results can be derived without the use of coherence theory. But he does not quote the more general formulas (Eq. (6) of Ref. [10] and Eq. (3) of Ref. [11]) which cannot be derived in this way. For example, if the light incident on the two pinholes is generated by a laser operating on several modes or has emerged from a rotating ground glass-plate that was illuminated by spatially coherent light, the spectrum of the light in the Young's interference pattern could not be predicted the way Mielenz suggests but would require the use of the more general formulas.

For essentially the same reasons, Mielenz reference towards the end of Sec. 5 to the well-known textbooks by Jenkins and White and by Strong are inappropriate, as they do not discuss Young's interference experiments with partially coherent sources.

There is actually much more in Ref. [1] that one might rightly question, but the preceding remarks should suffice to show that Mielenz's criticism is largely unfounded and unsound, or is confined to special situations involving spatially incoherent sources for which alternative, although equivalent, interpretations of the phenomenon of correlation-induced spectral changes can be given.

I find it regrettable that some members of the radiometry community have adopted a rather negative attitude towards the recent developments relating to radiation from sources of different states of coherence. This is probably due to the fact that they are mainly accustomed to dealing with traditional thermal sources, which are effectively spatially completely incoherent. As already mentioned, the radiation fields produced by such sources can indeed be analyzed by the traditional methods, which do not require the use of coherence theory. However, the radiation properties of many other sources which are used today, such as multimode lasers and various x-ray sources [12–14] can only be fully understood within the framework of the theory of partial coherence. The discussion of Ref. [1] indicates an attitude which can only inhibit real progress in the development and understanding of such sources and of the radiation which they generate.

Acknowledgment: Some of our publications to which reference is made in this comment were supported by the Department of Energy, under grant DE-FG02-90ER 14419. The views expressed in this article do not constitute an endorsement by the Department of Energy.

Emil Wolf

Department of Physics and Astronomy, and the Institute of Optics,
University of Rochester, Rochester, NY 14627

Received: February 9, 1994

References

[1] K. D. Mielenz, J. Res. Nat. Inst. Stand. Technol. **98**, 231 (1993).
[2] E. Wolf, Phys. Rev. Lett. **56**, 1370 (1986).
[3] E. Wolf, Nature **326**, 363 (1987).
[4] G. V. Varada and G. S. Agarwal, Phys. Rev. **A44**, 7626 (1992).
[5] D. F. V. James, Phys. Rev. **A47**, 1336 (1993).
[6] G. Indebetouw, J. Mod. Opt. **36**, 251 (1989).
[7] E. Wolf and A. Gamliel, J. Mod. Opt. **32**, 927 (1992).
[8] E. Wolf, J. Opt. Soc. Amer. **68**, 6 (1978).
[9] A. T. Friberg, ed., Selected Papers on Coherence and Radiometry, SPIE Optical Engineering Press, Milestone Series, Vol. 69, Bellingham, WA (1993).
[10] D. F. V. James and E. Wolf, Optics Commun. **81**, 150 (1991).
[11] D. F. V. James and E. Wolf, Phys. Lett. **A157**, 6 (1991).
[12] K. J. Kim, Nuclear Inst. Meth. Phys. Res. **A246**, 71 (1986).
[13] J. E. Trebes, K. A. Nugent, S. Mrowka, R. A. London, T. W. Barbee, M. R. Carter, J. A. Koch, B. J. MacGowan, D. L. Matthews, L. B. Da Silva, G. F. Stone, and M. D. Feit, Phys. Rev. Lett. **68**, 588 (1992).
[14] D. Atwood, Phys. Today **45**, 24 (August 1992).

About the author: Emil Wolf is the Wilson Professor of Optical Physics at the University of Rochester. He is the co-author, with Max Born, of the book Principles of Optics and is the editor of the series Progress in Optics.

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 283 (1994)]

Reply to Professor Wolf's comments on my paper on Wolf shifts

Since Professor Wolf's comments [1] on my paper on Wolf shifts [2] are rather general in scope, I will attempt to answer them in the order of the issues addressed in my paper.

1. *Does the spectrum of partially coherent light change on propagation in free space?*

Having reconsidered the paper by Indebetow, my answer remains "no." Indebetow constructed a source that incorporated prisms and, therefore, changed color when viewed at different angles. But the spectrum depended only on the angle of observation, not on the distance traveled by the light. How can this demonstrate "most dramatically" that "the propagation from the source...has given rise to spectral changes"? A proof that the spectrum does not change may be found in my Eqs. (11) and (12).

I used the theory of partial coherence, not a "most primitive model," to derive Eqs. (11) and (12). The word "ray" was a figure of speech, not unlike Wolf's remark that "Newton did not use partially coherent sources." Does it not follow from the van Cittert-Zernicke theorem that a hole in a window shutter is partially coherent?

2. *Do the theory of partial coherence and the classical Huyghens-Fresnel-Kirchhoff diffraction theory give different results in situations that involve incoherent physical sources? Which of them should be applied for solutions of practical problems?*

According to my Eq. (21) the theory of partial coherence is a generalization of classical wave optics. "Wolf shifts" produced by apertures illuminated by incoherent sources are equivalent to diffraction phenomena that can be explained classically. Coherence theory is indispensible for the explanation of diffraction phenomena produced by partially coherent sources. As may be seen from the final paragraphs of his "Comments," Professor Wolf agrees with these answers.

3. *Are traditional radiometric practices afflicted by previously unknown errors due to the partial coherence of light?*

The "attitude" of radiometrists can be attributed to papers by Professor Wolf and others that were cited in my paper. These papers quoted the results of a doubtful experiment performed in India as evidence that the spectroradiometric scales of national standards laboratories suffer from large "Wolf-shift" errors. The Indian experiment was suspect because it conflicted with classical results quoted in Sec. 5.2 of my paper. This has now been confirmed by Foley and Wang on the basis of coherence theory (see their paper in this issue of the Journal of Research of the National Institute of Standards and Technology).

Klaus D. Mielenz

Alpine Lake Resort,
Terra Alta, WV 26764-0402

Received: February 24, 1994

References

[1] Emil Wolf, J. Res. Natl. Inst. Stand. Technol. **99**, 281 (1994).
[2] Klaus D. Mielenz, J. Res. Natl. Inst. Stand. Technol. **98**, 231 (1993).

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 285 (1994)]

Errata

Erratum: Precision Comparison of the Lattice Parameters of Silicon Monocrystals

E. G. Kessler, A. Henins, R. D. Deslattes, L. Nielsen, and M. Arif

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

[J. Res. Natl. Inst. Stand. Technol. Volume 99, Number 1, January-February 1994, p. 1]

The figures over figure captions 2, 3, and 4 have been permuted. The figure over figure caption 4 belongs with figure caption 2. The figure over figure caption 2 belongs with figure caption 3. The figure over figure caption 3 belongs with figure caption 4.

Volume 99, Number 3, May–June 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 287 (1994)]

Conference Report

WORKSHOP ON CHARACTERIZING DIAMOND FILMS III Gaithersburg, MD February 24–25, 1994

Report prepared by

Albert Feldman

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

Sandor Holly

Rockwell International,
6633 Canoga Ave.,
Canoga Park, CA 91304

Claude A. Klein
C.A.K. Analytics,
9 Churchill Lane,
Lexington, MA 02173

and

Grant Lu

Norton Diamond Film,
Goddard Rd.,
Northboro, MA 01532

1. Introduction and Conclusions

The third in a series of workshops was held at NIST on February 23–24, 1994, to discuss, in depth, specific topics deemed important to the characterization of diamond films made by chemical vapor deposition (CVD diamond) and to address the need for standards in diamond technology. The topics chosen for this workshop were based on feedback from the attenders of the previous workshop. The audience targeted for this workshop consisted of producers and potential users of CVD diamond technology. University scientists and scientists from government laboratories were invited as experts in properties measurements. There were 55 attenders at the workshop.

We focussed on three technical topics for discussion: characterizing brazing and polishing, standardization of thermal conductivity measurements, and characterizing stress strain, and fracture.

The principal conclusions of the workshop include:

- The development and availability of reliable post growth processes, such as polishing and brazing, will contribute in a substantial way to realizing new applications of CVD diamond. It is important to characterize and evaluate the results of these processes in a uniformly accepted way.
- The first results of the interlaboratory round-robin comparison of thermal conductivity and thermal diffusivity measurements were presented. They showed considerable laboratory to laboratory variations, although much of this variability may be due to the manner in which the different measurement methods employed are affected by specimen inhomogeneity and specimen anisotropy. Additional laboratories will be making measurements that will add to the present measurement data base.
- CVD diamond exhibits lower resistance to fracture than other forms of diamond. This is impeding the use of CVD diamond in applications where the material is subjected to high stresses,

such as those induced by thermal shock. The strength deficiency is attributed to surface flaws and high internal stresses that originate during growth. The relationship between growth conditions and internal state of stress has not yet been resolved.

2. Characterizing Brazing and Polishing

There were six presentations in this session. The first three presentations centered on techniques for polishing and surface figuring CVD diamond. The last three talks discussed issues and techniques for bonding diamond to other materials.

The development and availability of reliable post-growth processes, such as polishing and brazing, will contribute in a substantial way to realizing new applications of CVD diamond. Being able to characterize and evaluate the results of these processes in a uniformly accepted way is very important. The quality of a brazed joint between a CVD diamond component and a ceramic or metal support structure may be defined in a number of different ways, depending on the specific application. In one application the most important feature of a joint might be its heat transfer characteristics at cryogenic temperatures or across a broad temperature range; in another application its most important characteristic may be its mechanical strength, which might be determined by the level of success in compensating for the thermal expansion mismatch between the substrate and the diamond. Optical applications will be driven by other requirements, such as how a specific polishing process controls the surface micro-roughness or how closely an accurate surface figure can be attained.

The first presentation was by R. Miller of Raytheon who gave an extensive overview of diamond polishing technology. In his talk, he grouped the currently pursued polishing methods into contact type and noncontact type processes. Contact methods may be further divided into nonreactive and reactive methods. The non-reactive polishing methods discussed were the conventional and the abrasive jet techniques. Among the reactive polishing techniques, the hot metal polishing approach was emphasized. Polishing diamond against iron, manganese, lanthanum, and cerium metals has had variable success.

The iron plate method is usually a dynamic method in which the diamond is rubbed against the iron at an elevated temperature. It results in a smooth finish (20 nm peak-to-valley) when the plate temperature is low ($\approx$750 °C) and the polishing is done in a hydrogen environment; the material removal rate is relatively low ($\approx$0.5 μm/h). The polishing process and the material removal rate may be speeded up ($\approx$7 μm/h) by increasing the plate temperature to $\approx$950 °C. However, this results in a rough finish. An appropriate sequence of polishing temperatures appears to result in an acceptably smooth surface in a reasonable length of time. However some intergranular cracks may form during the processing which may or may not be acceptable, depending on the application. A method for characterizing the cracks in the polished CVD diamond surface must be used in order to evaluate the seriousness of such cracks.

Molten lanthanum or cerium metals, used in a static procedure, minimize the intergranular cracking in CVD diamond during polishing. In this procedure, the diamond is placed in contact with the metal, but no relative motion takes place, that is, no rubbing occurs. This process results in higher material removal rates than the previous method, but at the cost of increased surface roughness.

Other reactive contact polishing methods mentioned were float polishing employed at the Naval Air Warfare Center and amorphous silicon oxide techniques used by Edge Technologies, Inc.

Noncontact polishing methods fell into two groups, those that use laser beam etching and those that use ion beam etching. Polishing with a YAG laser beam at either 1.06 μm or at 0.53 μm and with an excimer laser beam was discussed. Two different ion beam techniques have been used to polish of CVD diamond. In the first technique, diamond is removed by sputter-etching with an argon or oxygen ion beam. A second novel technique has been developed by the Spire Corporation. It uses ion implantation to soften a top layer of diamond about 1 μm thick, which is then removed by conventional polishing methods.

The second talk was given by T. S. Sudarshan of Materials Modification Inc. He summarized recent advances in the use of lasers for polishing diamond films. Special emphasis was placed on the difficulty associated with polishing large areas and on the interpretation of the characterization of the polished diamond surfaces. Technical areas in which surface roughness plays an important role, such as tribological applications, were discussed.

The last talk on polishing was presented by S. Holly of Rockwell International, Rocketdyne Division. He highlighted the importance of surface shape control during polishing. Several applications require polished surfaces that are very flat. These include thermal applications, such as large

area diamond plate heat-sinks for multichip module technology and optical applications such as infrared windows. Future applications of CVD diamond will require accurate control of non-flat surfaces, such as spherical, aspherical, cylindrical, and toroidal shapes, that might be used in optical domes and lenses.

A polishing procedure that utilizes a combination of several polishing methods was then described. The procedure is based on interferometrically controlled ablation of diamond to produce a flat surface. The material removal proceeds in a manner similar to that used in single-point diamond turning; however, in this process, material removal is controlled in real time by continuous interferometric monitoring of the surface being ablated. The present status and recent results of this polishing procedure were discussed.

The first talk on CVD diamond brazing was presented by T. P. Thorpe of the Naval Research Laboratory (NRL). He focused on the importance of reliable brazed joints for producing high quality homoepitaxial diamond. These joints provide a thermal path between the diamond substrate and an underlying heat sink. The joints must survive high temperatures (up tp 1500 °C) for extended periods of time. Because the temperatures are so high, the conventional commercial brazing material (from Drukker International), a gold-tantalum composite which melts at ≈1060 °C, could not be used. As a result, several alternative brazes were investigated.

For NRL's purposes the selection criteria for a braze material were: 1) it should contain a refractory carbide-forming element in order to achieve satisfactory bonding to the diamond substrate. 2) It should form a refractory compound or alloy with the underlying molybdenum heat sink. 3) All products created during braze formation should have a melting point above 1300 °C. Among the brazes tried were several variations on the Drukker braze recipe, including substitution or alloying of gold with different members of the platinum group of metals. Greatest success was achieved with a Pd substitution. This was primarily due to its relatively low melting point of 1552 °C which helped to minimize graphitization of the diamond surface during braze formation. Platinum-gold alloys were also found to be effective but they were inherently more difficult to fabricate due to an unexpectedly complex Au-Pt interaction. Both of the above mentioned brazes were found to withstand growth conditions for several hours at temperatures up to 1400 °C without degradation.

Diamond-metal brazes employing Nichrome were also fabricated. They were able to withstand temperatures up to 1200 °C. Details of all of the fabrication techniques were discussed.

The second talk on brazing was given by R. Meilunas of the Grumman Corporate Research Center. The talk emphasized the importance of rigorous joint design procedures. A practical approach to braze design would include: 1) Consider the temperature of operation of the joint; select the melting point of the braze alloy to be 100 °C to 200 °C above the operation temperature. 2) Select the joint geometry. 3) Determine the mechanical properties of each component to be brazed. 4) Determine the thermoelastic/thermoplastic properties of the braze alloy (such as stress/strain vs temperature, strain hardening coefficient, elastic modulus, Poisson's ratio). 5) Determine the wetting/bonding properties of the braze alloy to the components to be brazed. 6) Perform a stress analysis of the joint. 7) Perform experimental validation of the stress analysis. 8) Determine the environmental stability of the joint.

A list of practical braze alloys was presented that covered brazing temperatures between 280 °C and 1300 °C. The properties of several substrate materials, such as W, Be and Cu, were related to the brazing process. The value and importance of a finite element analysis of brazed joints was described. The following statements concluded the presentation: 1) The stability of diamond-metal joints depends critically upon the residual stresses developed during cool-down from the brazing temperature. 2) A finite element stress analysis can provide insights into the residual stress distribution and potential failure modes in the joints. 3) The design of CVD diamond-metal brazed joints requires knowledge of wetting, chemical bonding, and metallurgical interactions during brazing.

The final talk of the session was given by J. Intrater, Oryx Technology Corporation. The talk focused on specimen surface preparation and selection of methods for analyzing the brazed joint. In considering sample surface preparation, special emphasis was placed on the challenges of sectioning and polishing of diamond and the problems that arise from the vast differences in hardness between metals and diamond. Important to characterization of metallization on diamond is the determination of the elemental distribution throughout the interfacial region and an assessment of the carbon phases present (i.e., the amount of sp^2 and sp^3 carbon present). The talk discussed use of a scanning electron microscope (SEM) equipped

with an x-ray microprobe and a wavelength dispersive spectrometer and Raman spectroscopy for evaluating the joint.

3. New Developments

In order to foster international cooperation in the development of standards for CVD diamond, we invited Yoichiro Sato of the National Institute for Research in Inorganic Materials (also known as NIRIM) to discuss the need for standardized methods of characterization and property measurement in Japan. In the area of CVD diamond, no standards development is occurring in Japan at present. Dr. Sato expressed the view that the large property variations that occur in CVD diamond make standards development difficult at this time. He has agreed to inquire whether other laboratories in Japan might want to participate in the round robin measurements of thermal conductivity.

K. V. Ravi of Lockheed Missiles & Space Company discussed the optical transmission properties of diamond produced by a novel combustion flame method. Some in the audience questioned whether the transmissivity was as good as was claimed, based on certain features seen in the absorption spectrum. This issue was not resolved.

E. Etz of NIST compared conventional Raman spectra of CVD diamond (obtained with visible light excitation) with Fourier Transform Raman (FT-Raman) spectra obtained with an excitation wavelength of 1.06 μm. While no significant difference was found between the two types of spectra obtained from a type IIa single crystal diamond, radical spectral differences were observed between the two types of spectra from CVD diamond. CVD diamond specimens exhibiting extremely clean Raman spectra in the visible show a large number of spectral features in the infrared region. Thus, FT-Raman spectroscopy appears to be a much more sensitive probe of diamond quality than visible Raman spectroscopy. Of note was a pronounced spectral feature in the FT-Raman spectrum of CVD diamond that was not observed in the visible Raman spectrum.

4. Standardizing Thermal Conductivity Measurements

Four presentations were made during this session.

J. Graebner of AT&T reported the results of thermal conductivity measurements on an isotopically enriched CVD diamond plate [amount-of-substance (or atom number) fraction of ^{12}C, 0.055%]. At room temperature, the in-plane thermal conductivity was 22 W/(cm · K) and the perpendicular-to-plane thermal conductivity was 26 W/(cm · K). Not only are these values higher than any previously reported for CVD diamond, but the perpendicular value is higher than that for the best natural single crystal diamond with the normal abundance of ^{13}C. An analysis of the temperature dependence of the thermal conductivity was consistent with a reduction in point-defect scattering of phonons in the isotopically enriched CVD diamond. This would account for the higher thermal conductivity.

O. Käding of Daimler-Benz described thermal conductivity measurements using photothermal displacement spectroscopy at transient thermal gratings (PDS-TTG). In this method, a polished diamond surface is heated by two pulsed laser beams that interfere to form a transient periodic heating pattern (grating) on the specimen surface. The relaxation time for the heating pattern to dissipate is governed by the period of the grating and the lateral thermal diffusivity. The relaxation time is determined by measuring the change in the deflection angle of a cw laser beam that is reflected from an undulation in the sample surface caused by the heating. The depth dependence of the thermal diffusivity near the surface can be determined by varying the grating period.

A. Feldman of NIST presented the results of round robin measurements of thermal conductivity. Measurements where performed by six laboratories on ten specimens provided by four manufacturers. Three geometries were provided: squares, disks, and long/narrow strips. Most laboratories could not test every specimen because each measurement procedure usually required a specific specimen geometry. Typically, the highest value and the lowest value for a given specimen differed by a factor of two. However, such a comparison is not necessarily meaningful because the thermal conductivities of the specimens were inhomogeneous and anisotropic. Each of the techniques sampled the specimen properties in different ways so that different results were to be expected. However, even measurements that should have yielded similar results showed variations up to 80%.

G. Lu of Norton Diamond Film presented the results of a survey that had been sent to users of CVD diamond for thermal management inquiring about their need for a standardized measurement method. When asked if only one type of measurement could be performed, what would they prefer

it to be, a clear majority preferred in-plane thermal conductivity rather than a perpendicular-to-plane measurement or an approximate average of the two. A clear majority also chose ±10% as the minimum uncertainty needed. There was a strong preference for a less accurate, quality control tool that would be used on every lot rather than a more accurate measurement that sampled every fifth or tenth lot. However, there was no consensus on other questions such as: was it preferable to determine both the in-plane and perpendicular-to-plane thermal conductivity on an occasional basis; or, should the thermal conductivity of every lot be measured in one direction.

5. Characterizing Stress, Strain, and Fracture

Recent advances in the art of producing CVD diamond now make it possible to obtain free-standing millimeter-thick deposits that have elastic, optical, and thermal properties comparable to those of type-IIa natural diamond crystals. These deposits, however, exhibit fracture strengths that are substantially lower than anticipated judging from the reported strength values of single crystals (natural and synthetic) or polycrystalline high-temperature/high-pressure produced diamond compacts. Since many of the potential applications of CVD diamond, for example infrared transmitting windows or domes, critically depend on the ability to withstand stresses generated in a thermal shock environment, it has become essential to properly assess the fracture behavior of this material. Furthermore, the production of thick CVD diamond plates is often hindered by cracking of the deposit, which suggests that the CVD process gives rise to internal stresses that may initiate crack propagation. The purpose of this session was to explore how a proper description of the development of strains and stresses in CVD diamond may help us understand its strength and fracture characteristics. In this regard, it has been established that the deposition process itself generates highly localized stress concentrations at the grain boundaries, thus weakening the structure. These stress concentrations appear to result from the intrinsic elastic anisotropy, from the formation of high-order twins boundaries, and (perhaps) from the defects associated with the presence of residual hydrogen impurities. Still, the issue of how to control process-induced stresses, in effect, establishing the relationship between growth conditions and state of stress, has not yet been resolved. Ten presentations at this session addressed these questions.

Evidence of internal strains in CVD diamond films was first reported by Japanese workers [1] based on Raman spectroscopy work. Y. Sato of the National Institute for Research in Inorganic Materials indicated that measurements performed on films 10 μm to 30 μm thick deposited on silicon showed peak-position displacements up to 3 cm^{-1} in addition to substantial line broadening, which points to inhomogeneous strain distributions. Microscopic birefringence patterns seen on polished surfaces also demonstrate that, besides gradual stress variations between top and bottom layers, there were steep stress gradients, perhaps up to 5 GPa/μm, within regions comparable to the grain size. At Norton, K. Gray obtained Raman spectra at various locations on an optical quality 0.5 mm thick CVD diamond bar uniaxially stressed in a three-point bending fixture and observed that the Raman peak shifts in accord with the hydrostatic pressure coefficient, 2.9 cm^{-1}/GPa, previously recorded for single crystals. Because of the large beam spot-size (30 μm), the Norton spectra do not provide the resolution required to detect highly localized stresses as was done at the Naval Research Laboratory (NRL). J. Butler reported that the micro-Raman instrument used at NRL, which operates at a wavelength of 514.5 nm, provides spatial resolutions as low as 1 μm in addition to exceptional spectral resolution. Thus, the stress can be assessed not only in terms of the shifting and the broadening of the diamond Raman line but also from the splitting of the line that occurs due to the lifting of the longitudinal-optic/transverse-optic phonon degeneracy. In accord with the NIRIM observations, this important work leads to the conclusion that residual stresses in CVD diamond films are highly localized and exhibit a spatial distribution that correlates well with surface morphological features; peak stresses exceeding 3 GPa have been recorded.

At this point, it becomes desirable to examine what causes the stress concentrations, and three papers addressed that issue.

S. Kurtz and collaborators at the Pennsylvania State University are making use of the Poisson-Voronoi tessellation model, in conjunction with finite-element methods, to calculate both the effective elastic modulus of random aggregates of diamond grains and the microstresses that develop at the periphery of individual grains in textured

diamond deposits. Their result for Young's modulus was shown to be in remarkable agreement with the value obtained elsewhere [2] upon using the Hershey-Kröner-Eshelby averaging procedure. More importantly, these numerical simulations show that, near grain triple points in textured thin films, there are very large stress concentrations, or stress singularities, which derive from the elastic anisotropy of the diamond lattice and depend on the degree of preferred orientation.

On a more visual level, D. Shechtman of the Technion (at present, guest scientist at NIST and NRL) considered the impact of low-order twin boundaries that evolve during the deposition of CVD diamond and that play a role in providing suitable nucleation sites. The intersection of these low-order twin boundaries leads to the formation of higher-order twin boundaries that may contribute to creating locally stressed regions. The columnar structure of CVD diamond, specifically, the presence of grain boundaries as well as twin boundaries, thus can provide a direct fracture path and, therefore, contribute to lowering the fracture resistance.

Very comprehensive investigations of the relationships between growth conditions, microstructure, and residual stresses in microwave-plasma CVD diamond deposits were reported on by A. Harker of Rockwell International. The characterization tools he uses to examine microstresses include angle-resolved x-ray diffraction, channeling electron microscopy, and stimulated fluorescence. One of the conclusions of his work is that even if there are wide fluctuations in local stress levels, the average residual bulk stress appears to be rather insignificant with the growth surface generally under tension and the nucleation surface under minor compression.

The latter work corroborates some of the results of R. Hallock and C. Klein of Raytheon, who performed wafer-curvature experiments on diamond/silicon laminates for the purpose of characterizing the average intrinsic strain of CVD diamond as a function of the diamond thickness and deposition temperature. In this connection, they emphasized that if the proper thin-film conditions are not satisfied, commonly used formulas for obtaining the coating stress from the wafer curvature involve semi-empirical approximations of questionable validity. The model they are using to interpret their data is based upon a general theory of elastic interactions in multi-layer laminates [3] that assumes isotropic relationships in the layer planes, in other words, texture-free polycrystalline deposits on (111) or (100) Si substrates. Errors originate primarily from some unavoidable uncertainty in assessing the growth temperature and, hence, in determining the thermal mismatch strain; in addition, Si creep may cause problems at growth temperatures in excess of ~950 °C. For these reasons, wafer-curvature experiments were also carried out with diamond on diamond systems, which not only eliminates the thermal mismatch but yields direct evidence on intrinsic strains since the wafer curvature is related to the coating strain through a very simple equation. In both types of experiments (diamond on silicon as well as diamond on diamond), the average process-induced strain was found to be relatively small but strongly dependent upon the deposition temperature, the sense of the strain changing from compressive at lower temperatures to tensile at temperatures above 1030 °C.

C. Klein of C.A.K Analytics commented that in evaluating the fracture-strength of CVD diamond it is essential to keep in mind that the crack propagation mechanism appears to be controlled by surface flaws. Consequently, the stress at failure cannot be considered indicative of an intrinsic fracture strength because it obeys a probability law. In effect, the failure probability at a given applied stress obeys the Weibull distribution. An analysis of fracture-strength measurements performed on CVD diamond films [4] made by hot filament CVD, 10 μm to 100 μm thick, yielded a characteristic strength of ~485 MPa at the growth surface and 1030 MPa at the nucleation surface.

J. Mecholsky of the University of Florida confirmed that polycrystalline diamond indeed fails from the surface and outlined current approaches to describing the fracture behavior, which include fracture mechanics considerations, the fractal geometry description of a fractured surface, and molecular dynamics modeling. Fracture mechanics experiments were used to determine the critical fracture energy and the fracture toughness. Using the characteristic strength values reported above, one can estimate the flaw sizes to range from 30 μm at the nucleation surface to 150 μm at the growth surface; these dimensions are roughly equal to the estimated grain sizes at these surfaces. Also of interest is Mecholsky's observation, using the $\sin^2(\psi)$ x-ray diffraction technique, that the stress pattern in diamond coatings deposited on silicon shows evidence of growth stresses, i.e., stresses that cannot be accounted for by thermal expansion mismatch alone. This conclusion is in accordance with the wafer-curvature work done at Raytheon.

Finally, M. Drory of Crystallume obtained a fracture toughness value of 5.3 $MPa \cdot m^{1/2}$ for CVD plates thicker than 200 μm that had been subjected to a uniaxial tensile stress. This value, which is typical of high-strength ceramic materials, agrees with Drory's earlier value derived from indentation testing but appears to be lower than values obtained by other workers and may reflect specimen-to-specimen variations in microstructural features. How such variations would affect the fracture toughness is not really understood.

6. References

[1] S. Matsumoto, Y. Sato, M. Kamo, and N. Setaka, Jpn. J. Appl. Phys. **21**, L183 (1982).
[2] C. Klein, Mater. Res. Bull. **27**, 1407 (1992).
[3] P. Townsend, D. Barnett, and T. Brunner, J. Appl. Phys. **62**, 4438 (1987).
[4] G. Cardinale and C. Robinson, J. Mater. Res. **7**, 1432 (1992).