

In This Issue:

Vol. 92, No. 3, May-June 1987

Departments***News Briefs*****DEVELOPMENTS**

157

Infrared Systems Developed to Test Gallium Arsenide Wafers
 Data Encryption Standard to be Reviewed
 NBS Step Generator Offers Improved Features
 Spectral Irradiance Scale Using Silicon Detector
 NBS/Israeli Agreement on Cancer, Blood Serum Research
 NBS-Developed Fire Measurement Tool Considered by ASTM
 NBS Develops New Polymer Electrolyte for Batteries
 Attenuation Measurements on Deformed Optical Fibers
 Cycle-Counting Methods for Fatigue Analysis
 NBS Using Sound Waves to Detect Flaws in Concrete
 Evaluating HVAC Performance in "Smart" Buildings
 Report Lists 23 Years of Semiconductor Publications
 First Report on NBS Large-Scale Seismic Project Available
 Optical Fiber Measurement Symposium Proceedings
 Electromechanical Properties of Superconducting Magnets
 Report Details 1986 NBS Analytical Chemistry Activities
 NBS Studies Ability of AEC Industry to Exchange CAD Data

STANDARD REFERENCE DATA

161

Chemical Kinetics Data for Combustion Chemistry

NBS SERVICES

161

NBS Surveying Time/Frequency Service Users
 New Watthour Meter Calibration Facility Offers Higher Accuracy, Measurement
 of Reactive Power
 Software Test Center Moved to NBS
 NVLAP Required by NRC for Processing Dosimeters
 New Steam Turbine Technology: Large Energy-Saving Potential

News Reports

New Small-Angle X-ray Scattering Facility Available for Cooperative,
 Proprietary Research 164
 New Lab Accreditation Program for Construction Services to Help U.S. Companies 164

Conferences/Events

Measurement Uncertainties: Report of an International
 Working Group Meeting **R. Collé,** 243
Pekka Karp

CALENDAR

245

Articles

Two Theories of Experimental Error	A. R. Colclough	167
Submicrometer Linewidth Metrology In the Optical Microscope	Diana Nyssonen, Robert D. Larrabee	187
Submicrometer Microelectronics Dimensional Metrology: Scanning Electron Microscopy	Michael T. Postek, David C. Joy	205
Instrument-Independent CAD Spectral Databases: Absolute Cross-Section Measurements In QQQ Instruments	Richard I. Martinez, Seksan Dheandhanoo	229
Note on the Choice of a Sensitivity Weight In Precision Weighing	R. S. Davis	239

News Briefs

Developments

INFRARED SYSTEMS DEVELOPED TO TEST GALLIUM ARSENIDE WAFERS

Detecting flaws in gallium arsenide (GaAs) semiconductor materials should be easier with the two polarized infrared light systems developed by NBS researchers. Both are nondestructive methods that wafer manufacturers can use to screen materials before marketing.

One system can examine an entire wafer, while the other employs a 75- to 600-X microscope to view isolated wafer portions. Both systems allow digital storage of images and the use of false-color graphics to represent wafer characteristics, such as variations in the transmitted infrared intensity, which could indicate potential problems.

GaAs wafer applications in high-speed electronic and optoelectronic devices are growing rapidly, but production of the near-perfect GaAs crystals needed for optimum performance is not as advanced as with the older silicon technology. The two NBS systems can aid in production control by pinpointing wafer flaws and inhomogeneities. Bureau researchers are using the infrared techniques in-house, but will also assist industries in setting up their own systems.

For further information, contact Dr. Michael Bell, National Bureau of Standards, Gaithersburg, MD 20899.

DATA ENCRYPTION STANDARD TO BE REVIEWED

The Data Encryption Standard, issued in 1977 as a Federal Information Processing Standard (FIPS), is again being reviewed to determine its adequacy to protect computer data. (FIPS are developed by NBS by the federal government.) Comments are invited and must be received on or before June 4,

1987. Send to the Director, Institute for Computer Sciences and Technology, National Bureau of Standards, ATTN: Second Review of FIPS 456, B154 Technology Building, Gaithersburg, MD 20899. The standard provides an algorithm to be used in electronic hardware devices to protect unclassified computer data. It was first reviewed and reaffirmed in 1983.

For further information, contact Dr. Dennis Branstad at the above address, or telephone 301/975-2913.

NBS STEP GENERATOR OFFERS IMPROVED FEATURES

Three NBS researchers have developed a calibration standard for transient waveform recorders that is an improvement over existing commercial instruments in either accuracy or variability of voltage levels. Transient waveform recorders are used to measure rapid voltage pulses in applications such as research into automotive engine performance or the testing of electrical power equipment for its vulnerability to lightning. These instruments also play key roles in nuclear fusion research and weapons testing. Developed by NBS researchers, the device generates precise, fully programmable voltage steps which exhibit fast transitions and exceptionally smooth settling to the final voltage value. Technical details about the step generator's design are available from T. Michael Souders, National Bureau of Standards, Gaithersburg, MD 20899.

SPECTRAL IRRADIANCE SCALE USING SILICON DETECTOR

NBS researchers have established an experimental scale of spectral irradiance for the wavelength range 400-700 nanometers based on an absolute silicon photodetector. Spectral irradiance and the re-

lated scale of luminous intensity are widely used in the photographic, lighting, defense, and aerospace industries for measuring the output of light sources. The new scale makes use of interference filters and a 100 percent quantum efficient light detector invented by researchers from NBS and United Detector Technology, Inc. The new scale is independent of—and has the potential to be easier to use than—either of the traditional scales based on the thermal physics of blackbodies or absolute thermal detectors. From 500 through 700 nanometers in the visible, the new silicon-detector-based scale agrees with the traditional blackbody scale used at NBS to within +0.5 percent to -0.7 percent with a standard deviation of 0.3 percent. When converted to luminous intensity (the scale used to relate irradiance to human eyes), the agreement between the two methods is even better.

For further information, contact Robert Bruening, National Bureau of Standards, Gaithersburg, MD 20899.

NBS/ISRAELI AGREEMENT ON CANCER, BLOOD SERUM RESEARCH

The National Physical Laboratory (NPL) in Jerusalem, Israel, has joined forces with NBS in a cooperative research program to develop and validate methods for chemical analysis of blood serum. An Israeli researcher is collaborating with NBS in providing a quality assurance program to enhance the accuracy of data from laboratories participating in the National Cancer Institute's (NCI's) Cancer Chemoprevention Program.

This program, which aims to determine if certain vitamins and minerals play a role in preventing some types of tumors, requires accurate measurements of the nutrients in blood serum before a link can be made to cancer prevention. The NPL researcher will validate methods based on liquid chromatography with electrochemical detection for measuring fat-soluble nutrients (beta-carotene and Vitamins A and E) in serum. He will help NBS establish accurate values for concentration of these "micronutrients" in serum samples which will be sent routinely to NCI-sponsored labs for testing the accuracy of their analyses. He also will assist in developing a standard reference material certified for nutrient concentration which will be offered for sale to laboratories interested in calibrating their instruments and validating their analytical methods. The NBS/NPL agreement is expected to last 2 years.

For further information, contact Willie May, National Bureau of Standards, Gaithersburg, MD 20899.

NBS-DEVELOPED FIRE MEASUREMENT TOOL CONSIDERED BY ASTM

The American Society for Testing and Materials has published a proposed voluntary fire hazard test method based on an instrument developed at NBS. The instrument, known as the NBS Cone Calorimeter can be used to predict how much heat a burning object such as a piece of furniture will release, by testing a small sample of material. This rate of heat release is one of the most important measures of fire hazard and figures prominently in predicting the course of a fire and its effects. The bench-scale instrument is based on a method of measuring the amount of oxygen consumed. The NBS Cone Calorimeter also can be used to make other fire hazard measurements such as the time it takes for the material to ignite and the amount of soot and smoke produced. Five instrument suppliers are manufacturing the device. The International Organization for Standardization (ISO) also is considering adopting the Cone Calorimeter as a standard measurement device and method.

For further information, contact Vytenis Babrauskas, National Bureau of Standards, Gaithersburg, MD 20899.

NBS DEVELOPS NEW POLYMER ELECTROLYTE FOR BATTERIES

NBS scientists have developed a new lightweight polymeric electrolyte material that has wide potential use in solid-state, high energy density batteries for weapon systems, satellites, and consumer products that require a lightweight energy source. The polymeric electrolyte is based on a design of interpenetrating polymer networks (IPN's) with two co-continuous phases, like a sponge with holes.

One phase is a cross-linked epoxy that provides strength and dimensional stability; the other is a low-molecular weight poly (ethylene oxide) with dissolved salt that gives high conductivity. Polymers that dissolve salts and conduct ions are an important class of materials because they are inert to lithium, the lightest metal that can be used in batteries, and they also can be fabricated in thin films for high technology applications. The NBS research was partially funded by the Office of Naval Research. A patent application has been filed for the new polymeric electrolyte material.

For further information, contact T. George Davis, National Bureau of Standards, Gaithersburg, MD 20899.

ATTENUATION MEASUREMENTS ON DEFORMED OPTICAL FIBERS

The optical attenuation of lightguides is one of the most important parameters to system designers [1]. In addition to the inherent attenuation of a given fiber, designers must take into account the added losses caused by the perturbations of bending, twisting, stretching, and overlapping of uncabled fibers. NBS has investigated these losses in short lengths of a variety of multimode fibers, using optical time domain reflectometry. The results of the studies showed that bending and microbending are the sources of most of the losses, that tension causes up to 4 dB/N·km of loss (depending on the type of fiber), that twisting losses are negligible, and that overlapping causes losses somewhat less than bending does.

Reference

- [1] NBSIR 86-3052, Attenuation Measurements on Deformed Optical Fibers, available from the National Technical Information Service, Springfield, VA 22161 (\$26.90 prepaid, order by PB# 87-132-289).

CYCLE-COUNTING METHODS FOR FATIGUE ANALYSIS

Structures such as bridges, aircraft, vehicles, pressure vessels, or offshore platforms are subjected to random cyclic loads that may result in structural failure. The designer or engineer needs information about the fatigue life of a structure in order to create a design that won't fail. Current approaches to random load fatigue analysis need the determination of stress cycles and associated stress ranges that are ambiguous in a random load history. NBS has devised Fortran IV programs that use the rain-flow counting method and the mean crossing-range technique to determine the stress ranges and cycles of random load histories. These techniques reduce the complexity of the history so that results can be compared to non-random test results. A new NBS publication [1] presents the concepts and complete Fortran IV programs for these analysis methods.

Reference

- [1] NBSIR 86-3055, Cycle-Counting Methods for Fatigue Analysis with Random Load Histories: A Fortran User's Guide, National Technical Information Service, Springfield, VA 22161 (\$13.95 hardcopy or \$6.50 microfiche prepaid, order by PB# 87-104758).

NBS USING SOUND WAVES TO DETECT FLAWS IN CONCRETE

NBS researchers have developed a nondestructive method to detect flaws in concrete [1]. Known as "impact-echo," the technique works on the same principle as the sonar pings used to locate and determine the depth of a submarine. An impact on the concrete generates sound waves which are reflected by flaws inside the concrete. A receiver mounted on the surface of the concrete picks up the reflections, or echoes. The location of a flaw is determined by measuring how long it takes to receive the reflected echo. So far, the NBS researchers have successfully used the technique to find artificial flaws embedded in a number of different concrete slabs. In addition, they have been able to detect pockets of unconsolidated concrete and the depth of cracks which are perpendicular to the surface. They also have been able to distinguish a hollow metal duct from one that is filled with grout. The NBS team plans to test the technique on other structural elements such as rectangular beams and round columns. Also, they want to investigate its potential for detecting other types of flaws, such as voids, beneath pavements.

Reference

- [1] NBSIR 86-3452, Impact-Echo: A Method for Flaw Detection in Concrete Using Transient Stress Waves, National Technical Information Service (NTIS), Springfield, VA 22161 (\$24.95 prepaid, order by PB# 87-104444/AS).

EVALUATING HVAC PERFORMANCE IN "SMART" BUILDINGS

In a report [1] for the General Services Administration, NBS researchers describe specifications for evaluating the thermal and environmental performance of advanced-technology buildings. These buildings, sometimes called "smart" buildings, have sophisticated controls for heating, ventilation, and air-conditioning systems. The specifications will help avoid design, construction, and operation errors which may result in buildings which are unsuitable for occupants or which have excessive operating costs. They include ways to measure airtightness and infiltration rates and to determine insulation effectiveness. Also included are methods to evaluate indoor pollutant levels. The report describe various diagnostic tests and the materials and equipment needed. It also contains work statements which describe how each test should be conducted, how data should be analyzed, and how the results should be presented. Although the specifications were developed for use on advanced-technology

buildings, most of the methods can be used to evaluate other buildings as well.

Reference

- [1] NBSIR 86-3462, Specifications for Thermal and Environmental Evaluations of Advanced-Technology Office Buildings, available from the National Technical Information Service, Springfield, VA 22161 (\$13.95 prepaid, order by PB# 87-134326/AS).

REPORT LISTS 23 YEARS OF SEMICONDUCTOR PUBLICATIONS

Looking for resources on such semiconductor-related topics as spreading resistance, linewidth measurement, test structures, or dopants? Do you need information on Raman backscattering, Schottky barriers, or the Monte Carlo method? If so, these and many other subjects are listed in an NBS bibliography now available [1].

Cataloging Bureau publications on semiconductor measurement technology produced between 1962 and 1985, the document is categorized by year, author, and subject. Other useful information is included, such as how to obtain an NBS videotape on safe operating area limits for power transistors and one on laser scanning of active semiconductor devices.

Reference

- [1] NBSIR 86-3464, Semiconductor Measurement Technology: A Bibliography of NBS Publications for the Years 1962-1985, available from National Technical Information Service, Springfield, VA 22161 (\$16.95 prepaid, order by PB# 87-112298).

FIRST REPORT ON NBS LARGE-SCALE SEISMIC PROJECT AVAILABLE

As part of their studies on how full-scale bridge columns perform during earthquakes, NBS researchers constructed and tested six one-sixth scale replicas of the columns. The results will be used to determine if the behavior of full-scale columns can be extrapolated from model behavior. The first in a series of reports on the project [1] is now available and gives a detailed description of the design, fabrication, testing, and evaluation of the model columns. Included among the findings for small-scale columns is that recent California Department of Transportation specifications were sufficient to prevent the longitudinal reinforcing bars from pulling out of the footings for all the specimens. Testing of the first full-scale column was completed in July 1986. A second full-scale column is being fabricated at NBS and will be tested this year.

Reference

- [1] NBSIR 86-3494, Behavior of 1/6-Scale Model Bridge Columns Subjected to Cyclic Inelastic Loading, available from the National Technical Information Service, Springfield, VA 22161 (\$24.95 prepaid, order by PB# 87-152245).

OPTICAL FIBER MEASUREMENT SYMPOSIUM PROCEEDINGS

The 1986 Optical Fiber Measurements Symposium, held in September 1986 in Boulder at NBS, brought together over 300 representatives from 17 countries to present 34 papers. Topics of the 29 contributed papers spanned the full range of measurements necessary to specify an optical fiber, with a heavy emphasis on dispersion and mode-field diameter measurements in single-mode fibers. The five invited papers summarized the state of the art and looked to related and future measurement problems in the characterization of sources, detectors, specialty fibers, and planar waveguide devices. Summaries of the papers are presented in the proceedings of the symposium [1].

Reference

- [1] NBS SP 720, Technical Digest: Symposium on Optical Fiber Measurements 1986, Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402 (\$8 prepaid, order by stock no. 003-003-02772-3).

ELECTROMECHANICAL PROPERTIES OF SUPERCONDUCTING MAGNETS

Accurate data on superconductor performance under mechanical load is essential to the design of the magnets used in various fusion energy devices. NBS is conducting a 3-year research program sponsored by the Department of Energy to obtain this data and to quantify the electromechanical properties of promising new superconductor materials. A new NBS publication [1] gives the results of the first 18 months of the program. Included are data on the properties of experimental (PbMo_6S_3 and liquid-infiltrated Nb-Ta/Sn) and commercial (Nb_3Sn) superconductors and stainless steel conductor-sheathing materials. A new electron-tunneling diagnostic tool for probing the energy gap of practical superconductors is described.

Reference

- [1] NBSIR 86-3044, Electromechanical Properties of Superconductors for DOE Fusion Applications, National Technical Information Service, Springfield, VA 22161 (\$40.90 prepaid, order by PB# 87-125-753 A-06).

REPORT DETAILS 1986 NBS ANALYTICAL CHEMISTRY ACTIVITIES

Analysis of foreign particles in semiconductor devices, determination of the origin of ancient objects, and the measurement of human exposure to pesticides are just a few of the activities described in a new NBS report [1]. The document gives capsule summaries of 1986 NBS research in analytical chemistry. Major publications and staff talks are listed, along with tallies of the analytical chemistry Standard Reference Materials produced in 1986. (These materials, sold by NBS, are widely used by industries such as metals and chemicals, as well as in clinical and environmental laboratories, to improve analytical measurements.) The 184-page report also spotlights basic research on new measurement techniques and analytical chemistry's continuing role in providing data for highly complex sample types—samples such as foods, high-technology materials, hazardous chemical and nuclear wastes, plant and animal tissues, and body fluids.

Reference

- [1] NBSIR 86-3486, Technical Activities 1986—Center for Analytical Chemistry, available free of charge from Center for Analytical Chemistry, A309 Chemistry, National Bureau of Standards, Gaithersburg, MD 20899, telephone: 301/975-3143.

NBS STUDIES ABILITY OF AEC INDUSTRY TO EXCHANGE CAD DATA

The use of computer-aided design (CAD) systems in the architecture, engineering, and construction (AEC) industry is increasing steadily. But a recent study [1] at NBS confirms that it is difficult and often impossible to transfer data and drawings from one CAD system to another.

NBS researchers found the primary constraint is a lack of dependable and verifiable methods of exchanging data between different CAD systems. Intermediate formats and translators are available, but, say the researchers, they are not adequate for comprehensive AEC CAD operations.

To help resolve the problem, a committee was formed in 1984 to develop specifications for exchanging AEC information as part of the Initial Graphics Exchange Specification (IGES). IGES has been a national standard since 1981. The IGES/AEC committee currently is developing additional exchange specifications to be incorporated into the next version of IGES. For further information on this committee, contact Kent Reed, co-chair, IGES/AEC Committee, B306 Building Research Building, National Bureau of Standards, Gaithersburg, MD 20899. Telephone: 301/975-5852.

Reference

- [1] NBSIR 86-3476, The Current Ability of the Architecture, Engineering, and Construction Industry to Exchange CAD Data Sets Digitally, National Technical Information Service, Springfield, VA (\$11.95 prepaid, order by PB# 87-134334/AS).

Standard Reference Data

CHEMICAL KINETICS DATA FOR COMBUSTION CHEMISTRY

NBS has published the first volume in an important new series of compilations of evaluated data on chemical kinetics for combustion chemistry [1]. Computer models of the complex chemistry of combustion have become important tools in the development of combustion systems of high efficiency and low pollution, but such models can never be better than the data fed to them. The new publication includes critical kinetic and thermochemical data for over 300 reactions involving the combustion of methane. Later papers will expand the coverage to include the data needed to model the combustion of all the alkanes, and, over the long range, other organic compounds typical of fossil fuels. The work is sponsored by the Department of Energy and NBS.

Reference

- [1] W. Tsang and R. F. Hampson, Chemical kinetic data base for combustion chemistry, Part I. Methane, *J. of Phys. and Chem. Ref. Data*, V. 15, No. 3, pg. 1089.

NBS Services

NBS SURVEYING TIME/FREQUENCY SERVICE USERS

The National Bureau of Standards is conducting a survey of users of all of its time and frequency services, such as WWV and WWVH shortwave broadcasts, WWVB 60-kHz broadcasts, or telephone time-of-day service. It requests users to participate in this Time and Frequency Services Users Survey. The survey form is available from:

Time & Frequency Survey, Div. 524.00
National Bureau of Standards
Boulder, CO 80303

or call 303/497-3294 between 8 a.m. and 5 p.m. MDT to request a copy.

The survey results will help NBS provide the best mix of services and levels of service to the broad spectrum of users who depend on them. Feedback from all kinds of users is needed to assure that the Bureau's resources for these services are allocated in the most effective way.

NEW WATTHOUR METER CALIBRATION FACILITY OFFERS HIGHER ACCURACY, MEASUREMENT OF REACTIVE POWER

A facility for calibrating watthour meters that offers a fivefold improvement in accuracy over the existing calibration system is now in operation at NBS. It incorporates digital technology and will ensure the accuracy of meter standards used in laboratories by utilities and meter manufacturers.

The facility, designed and built at NBS, enables routine meter calibrations to be made with a 0.01 % uncertainty. Previously, typical NBS calibrations were in the 0.05 % uncertainty range.

Traditionally, the Bureau has performed calibrations by maintaining accurate standard watthour meters that are compared to ones sent to NBS for calibration. The new digital system is more accurate because it features a well-characterized, electronically generated signal source. The signal has a known value which is fed into the meter being calibrated. The meter error is then displayed in digital form. The updated facility also is faster, allowing measurements that previously took 12 to 14 minutes to be made in just a few seconds. Additionally, NBS now will have the capability to calibrate meters for measurement of reactive power, a service not previously available. Initially, the NBS facility will provide calibrations at 60 Hz which is the frequency domestic utilities use, and at 50 Hz to support the export market. Later, other frequencies will be added, such as 400 Hz, used on spacecraft as well as on commercial and military aircraft. Frequencies as high as 20 KHz are also being considered.

For further information, write John Ramboz, National Bureau of Standards, Gaithersburg, MD 20899, or phone 301/975-2434.

SOFTWARE TEST CENTER MOVED TO NBS
NBS and the General Services Administration have agreed to transfer the GSA testing service for programming language compilers to the NBS. The service allows vendors who want to sell compilers to the federal government to validate that the product conforms to a particular Federal Information Processing Standard (FIPS). In addition, NBS

plans to work with GSA, other Federal agencies, and industry to develop a government-wide policy for testing products and services for conformance to Federal standards. The policy would establish criteria for identifying the products and services for which conformance testing would be required.

NBS has played an active role in developing conformance tests. Most recently, NBS is working with industry to develop conformance tests for software standards, including standards for graphic systems, database management systems, office systems/document interchange, and operating systems.

For further information, write the Software Standard Validation Group, National Bureau of Standards, Building 225, Room A266, Gaithersburg, MD 20899, or phone 301/975-3247.

NVLAP REQUIRED BY NRC FOR PROCESSING DOSIMETERS

The Nuclear Regulatory Commission (NRC) has amended regulations to require all NRC-licensed organizations to have their personnel radiation dosimetry device readings performed by a processor that is accredited by the NBS National Voluntary Laboratory Accreditation Program (NVLAP). The NVLAP dosimetry program, which provides for periodic evaluations of dosimeter processors, was established in 1984 at NRC request to improve the accuracy of measurements on ionizing radiation that may be received by workers.

Under the personnel dosimetry program, accreditation is limited to personnel services for types or models of dosimeters that document whole body and skin dose radiation. All participating organizations are required to demonstrate that they are able to process each dosimeter type in accordance with ANSI N13.11, Criteria for Testing Personnel Dosimetry Performance.

Currently, the NVLAP has 45 accredited processors with evaluations under way for an additional 13.

For further information, contact Robert Gladhill, National Bureau of Standards, Gaithersburg, MD 20899.

NEW STEAM TURBINE TECHNOLOGY: LARGE ENERGY-SAVING POTENTIAL

Electric utilities could save up to \$200 million annually in generating costs by using a new packing ring designed to reduce leakage in steam turbines, according to the Department of Energy (DOE).

The improved ring was developed by Ronald Brandon of Schenectady, NY, with partial funding from a grant by DOE, after NBS reviewed and recommended Brandon's proposal to DOE. Packing rings, located on the turbine drive shaft of an electric generator, typically become worn from vibration during start-up and shut-down, allowing steam to escape and wasting energy. The new ring reduces wear and allows a tighter seal. It was tested recently at a power plant in Maryland.

NBS evaluates, free of charge, ideas for energy-saving inventions. To date, nearly 400 promising ideas have been recommended to DOE for possible support. For further information, contact George Lewitt, National Bureau of Standards, Gaithersburg, MD 20899.

News Reports

NEW SMALL-ANGLE X-RAY SCATTERING FACILITY AVAILABLE FOR COOPERATIVE, PROPRIETARY RESEARCH

An NBS research facility for studying the microstructures of polymers, metals, ceramics, and biological materials is now available to scientists in industry, government, and universities for cooperative and proprietary research. The small-angle x-ray scattering (SAXS) facility at the Gaithersburg, MD, site will help researchers to better understand the performance of existing materials and will aid them in developing new ones with different properties.

Research opportunities with the SAXS method are very broad. Metallurgists can use the SAXS technique to study crystal structure, void formation and growth, and phase separation in metals and alloys. The SAXS method also can be used to study pore characteristics in ceramics, and it can be used to measure molecular arrangements in biological materials.

For polymers research, the SAXS method can be used to study the phase separation of molecules, crystallite morphology, melting and crystal growth, molecular dimensions, and polymer networks.

NBS polymer scientists plan to use their new SAXS instrument to obtain information on the microstructures of polymer blends to develop polymer alloy phase diagrams. It is estimated that polymer blends currently account for 20 percent of the 2.5 billion-pound worldwide consumption of new engineering polymer materials.

The new SAXS instrument, which measures 10 meters in length, uses a 12-kilowatt rotating anode generator to produce a highly collimated pin hole beam of x rays. This extremely narrow beam is used as a probe to characterize the internal structures of materials on a size scale in 1- to 100-nanometer range.

For the material characterizations, samples are placed in a special chamber in the beam line where an area approximately 1 millimeter in size is exposed. The x-ray scattering pattern is picked up on a two-dimensional detector located behind the sample chamber. The pattern is recorded by a computer and displayed on a color graphics system. This information may be recovered for later use.

The new SAXS facility will be equipped with a sample chamber to take measurements on materials from room temperature up to 400 C. Measurements also can be taken on materials being deformed under stress.

The NBS facility was designed by polymer physicist Dr. John D. Barnes with assistance from Dr. Frederick I. Mopsik, a research chemist, and mechanical engineer Manfred Osti.

For information on the new NBS SAXS facility, or to schedule cooperative, independent, or proprietary research time, contact: Dr. John D. Barnes, B210 Polymers Building, National Bureau of Standards, Gaithersburg, MD 20899, or call 301/975-6786.

NEW LAB ACCREDITATION PROGRAM FOR CONSTRUCTION SERVICES TO HELP U.S. COMPANIES

A new voluntary laboratory accreditation program (LAP) for organizations that perform construction testing services on concrete, aggregates, cement, soil, rock, asphalt, and geotextiles has been established by NBS. The program will be particularly useful to laboratories testing construction materials for export.

The LAP, established at the request of industry under the procedures of the National Voluntary Laboratory Accreditation Program (NVLAP), automatically assures accredited laboratories of acceptance of their test data by foreign laboratory

systems that have international agreements with NBS.

For example, the United States has an agreement with the Australian National Association of Testing Laboratories (NATA). As a consequence, building materials tested for fire resistance by a NVLAP-accredited U.S. laboratory meet an Australian building-code requirement for testing by a NATA-accredited laboratory.

In addition to Australia, NBS has agreements with the National Testing Laboratory Accreditation Scheme in the United Kingdom, and the Testing Laboratory Registration Council of New Zealand. Negotiations are under way to establish a similar agreement with Canada.

NBS was requested to establish the new construction testing services program because many engineering decisions are made solely on the basis of laboratory test data. The new program will help the buyers of private and public buildings, both here and abroad, to more easily determine if the products in structures meet specific national building codes and insurance requirements.

"Quality assurance in the testing of materials will reduce overall construction costs by eliminating the need to remove products that lack adequate test data," said Harvey W. Berger, NVLAP program manager.

Berger also pointed out that accurate, reliable test data can contribute to the construction of safer, more durable residential, commercial, and public structures.

Under NVLAP procedures, the new construction testing services program can be expanded to include the accreditation of laboratories to test other construction materials such as wood and steel.

NVLAP, established in 1976 and managed by NBS, is a voluntary system whereby organizations and individuals request NBS to establish a laboratory accreditation program. On an individual basis, laboratories seek accreditation for having the competence to use specific test methods.

"Competence" is determined by evaluating applicant laboratories to assure that they have the equipment, staff, and procedures necessary to perform recognized tests in accordance with nationally or internationally accepted standards or test methods.

NVLAP-accredited laboratories pay annual fees, go through on-site reassessment every 2 years, and participate in scheduled proficiency testing to maintain accredited status. The laboratories are listed in the NVLAP directory that is distributed worldwide.

Currently, approximately 170 laboratories are accredited in programs administered by NBS for thermal insulation, carpet, concrete, solid-fuel room heaters, acoustical testing services, personnel radiation dosimeters, commercial products (paint, paper, and mattresses), building seals and sealants, and electromagnetic compatibility and telecommunications equipment. Other LAPs have been proposed for asbestos abatement and electrical and safety testing.

For information on the new construction testing services program, or the public workshop that will be held at the Bureau to develop the technical requirements for the accreditation of laboratories under the LAP, contact: Harvey W. Berger, Manager, Laboratory Accreditation, Room A531 Administration Building, National Bureau of Standards, Gaithersburg, MD 20899, or call 301/975-4016.

Two Theories of Experimental Error

Volume 92

Number 3

May-June 1987

A. R. Colclough

National Physical Laboratory
Teddington, Middlesex
TW11 OLW, England

The reader . . . will have seized my meaning if he perceives that the different situations in which uncertain inferences may be attempted admit of logical distinctions which should guide our procedure.
Sir Ronald Fisher [1]¹

Following the widespread adoption of new approaches to the combination of experimental uncertainties, two theories of error are identified and their possible

justifications assessed. They are the "orthodox theory" based on the familiar distinction between random and systematic errors and the "randomatic theory" which dispenses with the distinction and treats all errors as the orthodox theory treats random errors. The orthodox theory suffers from a number of important confusions about the nature of its central distinction, about the combination of uncertainties, and about which populations of results can correctly be said to contain random errors. These confusions are clarified and the central distinction is argued to be objective. Three justifications are developed for the randomatic theory: 1) that it is implied by the generally accepted law of error propagation, 2) that all so-called systematic errors belong to populations characterized by hitherto unnoticed frequency-based distributions, and 3) that they belong to subjectivist prior distributions. But, upon examination, the ar-

gument in terms of the law of error propagation is found to beg key controversial questions, the frequency-based distributions are found not always to be of suitable form, and the subjectivist distributions are found to be unrealistic. Thus the randomatic theory remains unjustified by objective standards. Moreover, its use could lead to the underestimation of uncertainties in the usual sense of the maximum possible or conceivable error in the result of a particular specified experiment. The concept of systematic error is argued to be indispensable and new recommendations are formulated which are orthodox in general character.

Key words: error theory; experimental errors; frequency-based distributions; orthodox theory of errors; randomatic theory of errors; random errors; subjectivist distributions; systematic errors.

Accepted: January 20, 1987

1. Introduction

Foundational questions in statistics are notoriously controversial.² Nowhere is this more true than in error theory which presents special problems not usually encountered in other fields of statistical practice. In particular, it often invites the

¹ Numbers in brackets indicate literature references.

² Editor's note: As the author readily acknowledges, error theory is a controversial field and his views are at odds with some recent thinking on uncertainty evaluation intended to secure international consensus. Yet it is believed that this paper is a valuable contribution to the literature on measurement uncertainties presenting a closely argued case for one of the standard positions on the subject. An alternative view will be presented by R. Collé in a subsequent issue of this Journal.

experimenter, when estimating experimental uncertainties, to assess probabilities in the absence both of statistical data and the prospect of data. In view of this it is, perhaps, not surprising that confusion on the subject is widespread among experimental scientists, most of whom are specialists in fields other than statistics and unfamiliar with its foundational controversies.

About the Author: A. R. Colclough serves with the National Physical Laboratory's Division of Mechanical & Optical Metrology.

Yet there is no dearth of advice about what simple procedures are to be followed in estimating uncertainties. The problem is rather that the experts disagree with one another. One reason for their disarray is of course the deeply confusing nature of the questions involved. Another is that their advice has to serve a variety of needs. Metrologists and others making fundamental physical measurements arguably require a rigorous and objective (i.e., demonstrably realistic) theory of errors on which to base accurate estimates of uncertainty. Without this it would be impossible to rule on the existence of discrepancies between the results of primary experiments. In contrast, many scientific and industrial activities require only rough-and-ready "uncertainty" estimates, and simple methods of arriving at them may be preferred to reliable ones based on objective principles. Thus the national calibration services of a number of countries employ procedures that are quite different from the traditional ones based on the distinction between random and systematic errors which most students still learn. Books, standards, codes of practice, offi-

cial directives, and the like could now be cited describing them and they have much in common (e.g. [2-5]).

In recent years a great service has been done for error theory by the Bureau International des Poids et Mesures (BIPM) which has consulted very widely on the matter and produced a set of clear and concise recommendations for the estimation of experimental uncertainties that appear to be broadly in line with the widespread new procedures mentioned above [6-9]. These have provided a stimulus and a clear focus for renewed discussion of error theory which, however, remains as controversial as ever. They are set out in display box 1 for ease of reference and may be compared with the representative set of orthodox recommendations in box 2 [10].

Controversy arising from the different needs and interests of various users is probably inevitable. Its origin lies in practical considerations which are bound to be subject to value judgments of all kinds and this must preclude an uncontroversial ruling on universally appropriate procedures for the esti-

Box 1. The BIPM's Recommendations for the Combination Of Experimental Uncertainties

1. The uncertainty in the result of a measurement generally consists of several components which may be grouped into two categories according to the way in which their numerical value is estimated:

- A — those which are evaluated by statistical methods,
- B — those which are evaluated by other means.

There is not always a simple correspondence between the classification into categories A or B and the previously used classification into "random" and "systematic" uncertainties. The term "systematic uncertainty" can be misleading and should be avoided.

Any detailed report of the uncertainty should consist of a complete list of the components, specifying for each the method used to obtain its numerical value.

2. The components in category A are characterized by the estimated variances, s_i^2 , (or the estimated "standard deviations" s_i) and the number of degrees of freedom, ν_i . Where appropriate, the estimated covariances should be given.

3. The components in category B should be characterized by quantities u_j^2 , which may be considered as approximations to the corresponding variances, the existence of which is assumed. The quantities u_j^2 may be treated like variances and the quantities u_j like standard deviations. Where appropriate, the covariances should be treated in a similar way.

4. The combined uncertainty should be characterized by the numerical value obtained by applying the usual method for the combination of variances. The combined uncertainty and its components should be expressed in the form of "standard deviations."

5. If, for particular applications, it is necessary to multiply the combined uncertainty by a factor to obtain an overall uncertainty, the multiplying factor used must always be stated.

mation of uncertainties. But, given any clear, natural, and physical definition of "experimental uncertainty" it should be possible to state unequivocally what the *correct* procedures are for its estimation even though they may not be well suited to all practical needs. The lack of consensus about what these are ought to be a matter of concern, and the purpose of this paper is to attempt to define a philosophically-neutral, physically correct and rigorous error theory without regard to practicability. In fact it is arguable that the principles to be identified for uncertainty estimation are not markedly less practicable for most purposes than are those now becoming widely adopted.

The concept of uncertainty to which the following discussion relates is, in informal terms, the range within which the result of a particular specified experiment is uncertain as defined by its maximum possible or conceivable error. This entails that, when the experiment is faithfully repeated by different workers using the apparatus and procedures specified, their uncertainty bounds always or nearly always embrace the true value of the quantity to be determined and overlap each other. Moreover, when discrepancies do occur, they

should always or nearly always be small compared to the uncertainties themselves. These requirements are intended to be met in a literal physical sense and uncertainty estimation procedures which do not guarantee this will be regarded as failing to implement the chosen concept. This is the concept required when judging the consistency of results. It also has the merit of not presupposing any philosophical position as would, for example, a definition in terms of standard deviations intended for application to all error types. Moreover, when "uncertainty" is defined in terms of quantities such as expected error or standard deviation, it is often because these terms are thought to give an easily calculated order-of-magnitude estimate for uncertainty as defined above and not because they express the most relevant concept themselves.

The recommendations of boxes 1 and 2 raise a number of fundamental questions, the more important being:

- What is the nature of the distinction between "random" and "systematic" quantities and how does it relate to that between Types "A" and "B"?
- Is it objective and useful or merely a distinction without a significant difference?

Box 2. Representative Orthodox Recommendations for the Combination Of Experimental Uncertainties [10]

The uncertainty on a measurement should be put into one of two categories depending on how the uncertainty is derived: a random uncertainty is derived by a statistical analysis of repeated measurement while a systematic uncertainty is estimated by nonstatistical methods.

When combining the uncertainties on individual measurements in a complex experiment involving measurements on several physical quantities the two categories of uncertainties should be kept separate throughout.

In such an experiment the total random uncertainty should be obtained from the combination of the variances of the means of the individual measurements together with those associated with any constants, calibration factors, etc.

The component systematic uncertainties should be estimated in the form of maximum values or overall limits to the uncertainties.

In reporting measurements of the highest accuracy, a full statement of the result of an experiment should be in three parts, the mean corrected value, the random uncertainty, and the systematic uncertainty . . . The components that have contributed to the final uncertainty should be listed in sufficient detail to make it clear whether they would remain constant if the experiment were repeated . . . The estimate of the total systematic uncertainty should be stated . . . Each component of the systematic uncertainty should be listed, expressed as the estimated maximum value of that uncertainty . . . The method used to combine these component (systematic) uncertainties should be made clear.

The combination of random and systematic uncertainties to give an "overall uncertainty" is deprecated, but if in a particular case this is thought to be appropriate then it should be given in addition to the two uncertainties, together with the method of combination.

- Is it legitimate to represent *all* uncertainties, including those evaluated by other-than-statistical means, by statistical or quasi-statistical quantities?

- Is it legitimate to combine uncertainties of different types as though combining variances of random variables of zero mean?

Different answers to these questions will be obtained depending on the general theory of errors that is adopted and how the various concepts referred to are defined within it. Two general theories will be identified below, clearly formulated and used to derive answers to these and related questions. They will be referred to as the “orthodox” theory which retains the distinction between random and systematic quantities and the “rANDOM” theory which dispenses with it and treats all errors and uncertainties on an equal footing. It is on the latter theory that the BIPM recommendations appear to be based.

In formulating the two theories, it will be necessary to be clear about the objectivity of various error populations to be considered and the probability distributions to be defined over them. No serious controversy need arise about the physical status of populations of results and their corresponding errors when they are produced by repeatable measurements according to a well-defined experimental specification that permits random variations about nominal conditions. Nor is there a problem about errors which are sampled at random in some physical sense from a pre-existing error population (perhaps when an instrument with a certain zero error is chosen from a population of similar instruments in which a corresponding population of zero errors exists). But the objectivity of populations cannot be countenanced when an experimental specification is too loose to produce properly controlled results or when pre-existing populations are not unambiguously identified.

All schools of philosophy accept the notion that probability *evaluations* based on the frequencies observable in an objective statistical population are themselves objective. This is true irrespective of the particular view held of the concept of probability itself (for example, that it is a long-run frequency or a subjective “degree of belief”). Thus frequency-based probability evaluations are philosophically neutral and so unproblematic in error theory as in other fields of statistics.

But there is one school of statistical thought and practice of particular relevance to error theory where probability evaluations that are not frequency-based are employed freely with those that are. In subjectivist statistical method a “prior probability distribution” describing a subject’s “degrees

of belief” in the various possible outcomes of some trial *before* results are obtained is estimated, perhaps in a vague way on the basis of general experience. When statistical data are subsequently gathered this prior probability distribution is “conditionalized” by the application of Bayes’ theorem to produce a frequency-based posterior distribution which, given sufficient data and not-too-wild a choice of prior distribution, agrees closely with that obtained by other statisticians (see section 3.3). If the subjectivist’s prior degrees of belief are based on correct, but approximate, physical information, his prior distribution will be approximately physically objective as well as being, presumably, psychologically objective. If, on the other hand, it is a mere unfounded intuition or a guess, it will not generally be physically objective. Nor will it, if, in acknowledgment of ignorance, the subjectivist assigns equal probabilities to all possible values because he has no reason to prefer one to another (Bayes’ postulate). Where the latter types of probability evaluation are employed in error theory, there will be a serious question about the objectivity of uncertainty evaluations calculated from them.

2. The Orthodox Theory of Errors

Although orthodox error theory is characterized by a central distinction between “random” and “systematic” errors, the exact definitions of these key terms are vague, and confusion exists about what *methods of combination* are correct for the corresponding types of uncertainty and about the correct classification of error types arising from various causes. A clearer statement of the orthodox theory will therefore be formulated with which most orthodox error theorists would be in general agreement. Those adhering to it will be referred to, purely for convenience and in a narrow sense only, as “conservatives.”

2.1 The Formulation of the Orthodox Distinction Between Systematic and Random Uncertainties

There are three possible approaches to the classification of errors as systematic or random. Definitions may be cast in terms of . . .

- 1) how they would behave if an experiment were repeated (e.g., in terms of the forms which their distributions would take),
- 2) how their causes would behave upon repetition of the experiment or the nature of their causes (e.g.,

scale errors, rounding, fluctuations of one kind or another, mistakes), and
 3) the way they are treated (e.g., by statistical means or on the basis of a theoretical estimate).

Confusion often arises in elementary accounts of the nature of errors because these various approaches are not clearly distinguished. In this section the classification of errors will be based initially on their behavior when an experiment or, perhaps, some associated "trial," is repeated [approach 1) above]. The combination of uncertainties will be dealt with mainly in the following section. The important practical question of how error types as defined by their behavior are to be identified in terms of their known causes will be discussed in section 2.3. In the interest of brevity, the term "experiment" will stand in what follows either for a single measurement; a set of measurements, some of which may be repetitions; or for a whole experiment as usually understood. The term "result" will be used for the value obtained from an experiment in any of these three senses.

The Fundamental Four-Fold Error Classification

When an experiment is repeated many times, four types of behavior are possible in the observed results as shown in figure 1 . . .

1) each result may differ from the true value by the

- same amount and with the same sign, i.e. the error is constant,
- 2) each error may vary randomly realizing a stable random distribution with a non-zero mean,
- 3) each error may vary randomly realizing a stable distribution with a zero mean, or
- 4) each error may vary non-randomly (e.g., cyclically or by failing to produce convergent frequencies).

These four classes of error are doubtless capable of further division, but the classification as it stands is obviously unique in any given case and exhaustive since it consists of successive *dichotomies* or disjunctions of logical complements: constant error or varying error (non-randomly or randomly (non-zero mean or zero mean)). In other words, *there are no errors which do not belong to one or another of these four classes and none belonging to more than one*. Since the classification is exhaustive, any other classification of error-related concepts, including that in terms of systematic and random types, must embrace all four classes if it is also to be complete.

The Definition of 'Random' and 'Systematic' Errors

Although the exact nature of the distinction between systematic and random errors is often a matter of confusion, the practical motive behind it is clear enough. It arises from the perception that some errors, the "random" ones, can be treated

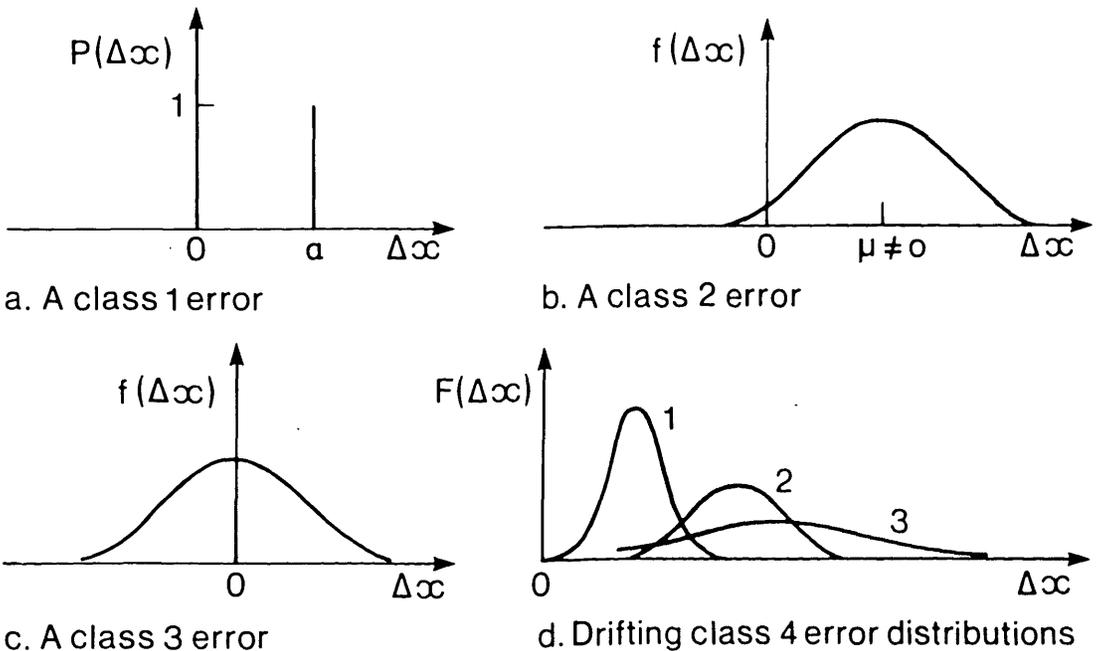


Figure 1—The four classes of error defined in the text.

statistically and in principle reduced to any desired level *solely on the basis of results*, while others, because of a tendency to act in one particular direction, cannot. The latter group of errors, the "systematic" ones, must therefore be assessed, and perhaps corrected for, independently of results.

But however clear the motivating ideas may appear, there is a widespread and crucial confusion in orthodox error theory about what types of population of results may be said to contain random errors. Must the results be *actually observed* by the experimenter when repeating his experiment before the existence of random errors can be contemplated? Or is it sufficient that the results *could be* observed repeatedly, though the experimenter chooses to conduct a measurement just once? Can errors in the repeatedly observed or single results of others be regarded as random when the results are used to calculate that of one's own experiment? Could it ever be correct to regard the error in the single result of some "trial" associated with an experiment, but not usually regarded by the conservative as part of it, as random (e.g., a scale error in an instrument "randomly" chosen for use)? These questions will be addressed later, but here definitions of "random" and "systematic" will be formulated which could be applied to any population of results accepted as "statistical." For simplicity the initial discussion is cast in terms of results obtainable by repeating an experiment.

Clearly a class 1 error could never be evaluated by contemplating a sample of results, however large, since, being the same for each result, it would lead to no differences in successive values from which its magnitude and sign could be inferred. An error of this kind, caused perhaps by a constant unwanted and uncorrected physical effect, is often regarded by the conservative as the standard case of a systematic error. As such, it is contrasted with a class 3 error which can be assessed in detail and reduced to any desired level by taking the average of a sufficiently large sample of results. This is the standard case of a random error.

The relation of class 2 errors to the random versus systematic distinction is less straightforward. The conservative frequently likes to oppose systematic to random errors, yet here is a randomly distributed error which nevertheless introduces on average a non-zero error into results which cannot be reduced indefinitely solely by averaging a large sample. However, while it may not be obvious how to classify class 2 errors themselves, every class 2 error can clearly be said to consist of a class 1 systematic component and a class 3 random one, the former component being identified with its mean

or expected value.³ Indeed, since the evaluation and treatment of uncertainties is always carried out separately for each component, there is no practical necessity to classify class 2 errors themselves. Definitions of "systematic" and "random" might therefore be adopted which result in class 2 errors being one, the other, neither, or both according to taste.

The above view of the mixed composition of class 2 errors need not, of course, imply an assumption that their constant or systematic component corresponds to any single physical cause or group of causes different from those giving rise to the random variation. Although they can be produced by distinct class 1 and class 3 errors, generally both components will have a cause or causes in common and in that sense are part of the same error. For this reason formal definitions of "random" and "systematic" would need to refer both to errors and error components.

Cast in terms of a result y instead of an error $\Delta y \equiv y - y_0$ where y_0 is the true result, the definition of systematic errors as class 1 errors or error components is equivalent to that sometimes offered in terms of statistical bias: $E(y) - y_0$.⁴

Class 4 errors are probably far more common than is generally realized. For example, any error that increases uniformly with time, even if "sampled" at random intervals, would be of this kind. In spite of this their existence is not usually recognized. Class 4 errors cannot, of course, be counted as random, but it is of little practical importance whether they are held to be systematic or are neither random nor systematic.

In the light of these considerations "systematic" and "random" errors might be defined by the scheme set out in box 3 or by equivalent definitions which would not necessarily be cast in terms of the four-fold error classification. Class 2 errors, the categorization of which was seen to have no practical significance, have been arbitrarily taken to be neither random nor systematic and class 4 errors to be systematic.

The definitions of errors of classes 1 to 4 were physical ones cast in terms of what behavior would be observed if an experiment were repeated many

³ Such a resolution is always possible for any class 2 error Δx . Setting $\Delta x = \mu_{\Delta x} + \Delta r$, where $\mu_{\Delta x} \equiv E(\Delta x)$ and the random variable $\Delta r \equiv \Delta x - \mu_{\Delta x}$, one has for its expected value $E(\Delta x) = \mu_{\Delta x} + E(\Delta r)$. Thus $E(\Delta r) = \mu_{\Delta x} - \mu_{\Delta x} = 0$ so that $\Delta x = \mu_{\Delta x} + \Delta r$, Δr being a class 3 error as asserted.

⁴ Systematic error defined in terms of a class 1 error or error components $\equiv \mu_{\Delta y} \equiv E(\Delta y) \equiv E(y - y_0) = E(y) - y_0 \equiv$ systematic error defined in terms of bias.

Box 3. A Possible Definition of 'Random' and 'Systematic' Errors

Orthodox Category	Error or Error Component
Systematic	class 1 error
	class 1 component of class 2 error
	class 4 error
Random	class 3 error
	class 3 component of class 2 error
Neither	class 2 error

times according to a clear experimental specification. Thus what class an error belongs to is a completely objective matter when it arises in results of repeatable measurements. Since the definitions of "random" and "systematic" error of box 3 are cast in terms of error classes 1 to 4, they too are objective categories applicable to all such errors.

It should also be noted that the subjunctive or "counterfactual" nature of the definitions ("... would be ... if an experiment were ...") enables single-reading errors to be called "random" or "systematic" even though the concepts are defined relative to a large population of errors. This should not be a matter for concern, of course; physical properties are typically "dispositional" in this way. That is, they are manifested only under appropriate conditions, but are held to persist in their absence. This important point will be discussed further in section 2.3.

The Definition of 'Systematic Uncertainty'

Once "systematic error" is defined, "systematic uncertainty" can be defined in terms of it. There are several ways of doing this of which the simplest is the following:

The "systematic uncertainty" in a given direction in the result of an experiment is *the magnitude of the range of its possible values as defined by knowledge of its maximum possible systematic error or error component in that direction.*

This concept of systematic uncertainty has been criticized because limits which are literally the maximum possible are often absurdly large and in

most practical cases there is an ineliminable element of "subjective" judgment in assessing plausible ones. Indeed, it is at this point that some experimenters abandon orthodoxy and introduce probability concepts to confine the range of the error to lie within conceivable rather than possible limits (cf. the definition of "random uncertainty" below).

But the conservative does not concede that it is appropriate to treat all errors as random errors. He usually prefers to abandon the definition of "systematic uncertainty" in terms of maximum possible error, but maintains that there are systematic errors which are not randomly distributed in *his* experiment (e.g., errors due to the use of biased theoretical corrections required by its specification). Uncertainty is therefore to be treated in terms of what Eisenhart has called "credible bounds" [11]. These are often said typically to be less than the maximum possible bounds, but if probabilities are employed in judging them they are held not to contribute to the random uncertainty of the final result. The conservative may also wish to maintain that there are some practical cases where admissible probabilistic information is lacking and where credible bounds are best replaced by maximum possible bounds.

The Definition of 'Random Uncertainty'

The expression "random uncertainty" is conventionally defined in terms of "random error" as follows:

The "random uncertainty" in a given direction in the result of an experiment is *the magnitude of the range of its values as defined by a knowledge of its maximum conceivable random error or error component in that direction.*

The use of "conceivable" here where "possible" was used in the previous definition, is in recognition of the common necessity of choosing a confidence level of less than 100% probability which for many distributions corresponds to the range \pm infinity. The justification of this procedure, apart from necessity, is that everyone is prepared to discount possible exceptions at some low level of probability.

2.2 The Orthodox View of the Combination Uncertainties

The law of error propagation states how various errors in an experiment combine to produce the error in its final result. Unfortunately the combining errors are not usually known, else they could be corrected for at source. What are usually known instead are their estimated maximum possi-

ble values, credible bounds, variances, or other quantities related to their respective uncertainties. How does the conservative use this information to estimate the uncertainty of his final result?

Orthodox 'Combination' of Random Uncertainties

The estimation of the random uncertainty resulting from the combined effect of two independent random errors is unproblematic in principle. The distributions of the errors convolute and their standard deviations combine in quadrature (i.e., their variances add) to produce those of the resultant error. Resultant random uncertainty is to be estimated from the resultant distribution relative to some choice of a confidence level close to one. It should be noted that in general random uncertainties, as opposed to standard deviations, do not combine in quadrature to yield a correct resultant random uncertainty. This may easily be demonstrated by consideration of the combination of two similar, but independent, uniformly distributed random errors, for example, which yields a resultant with a triangular distribution. The only exception to this rule arises from the combination of normally distributed errors which interact to form another normally distributed error; here uncertainties do combine in quadrature. But in general, unlike the expression "combination of errors," the phrase "combination of uncertainties" can be misleading.

Orthodox Combination of Systematic Uncertainties

Wavering conservatives sometimes entertain the notion that systematic uncertainties can be combined in quadrature to obtain a resultant systematic uncertainty [10,11]. This view may arise from feelings that it would be improbable that many systematic errors would all pull in the same direction or, more specifically, that $p(+)=p(-)=0.5$ (Bayes' postulate applied to signs); that in ignorance of their values they are uniformly distributed between bounds (Bayes' postulate applied to errors); and that credible bounds must be something like standard deviations because they are assessed from probabilistic considerations. However, combination in quadrature of systematic uncertainties is fundamentally inconsistent with the orthodox theory one of the first principles of which is that there exist constant errors and error components. That constant *error*-like quantities combine in a linear way is accepted by everyone. There is no dispute that when the distributions of combining random variables are convoluted to produce a resultant distribution, the mean of this distribution is simply the arithmetical sum of those of the combining variables. This is true in error theory as in other fields

of statistics and applies in particular to class 2 errors. To the consistent conservative the rationale for this is that the means are to be regarded not as random variables but as constants of the experiment of unknown sign and magnitude (or "constants of nature" in general statistical parlance). This is because they are parameters of *particular* error populations explicitly or implicitly identified by any complete experimental specification. As such, the means cannot be said in any physical sense to be drawn from a population and are undistributed except, perhaps, in the form of a delta function at some unknown location between credible bounds. Since no probabilities can be assigned to their various possible values the upper limit to be placed on the sum of the means can only be obtained from the sum of their individual upper limits, however defined. This becomes a simple point of logic where upper limits are defined to be maximum possible values. But even when credible bounds are employed, they are still intended to confine the conceivable values of unknown, undistributed constants which are agreed to combine in a linear way and could all pull in the same direction. Thus the consistent conservative permits himself no recourse to statistical procedures in such cases and must recommend that systematic uncertainties be combined in a linear way. Even if it were thought that systematic quantities were randomly distributed, uncertainties as opposed to systematic "standard deviations" would not be the appropriate quantities to combine in quadrature as argued above (cf. box 1, recommendation 3).

Orthodox Combination of Random With Systematic Uncertainties

How are uncertainties corresponding to "mixed" (class 2) errors to be evaluated on the orthodox view? Little guidance on this important matter is to be found in conservative literature, but a procedure is easily devised. In the case of a class 3 error, uncertainties u_+ and u_- are obtained from a confidence level p_L applied to a *single* class 3 distribution. In calculating u_+ and u_- for a class 2 error, the consistent conservative must consider not one distribution, but *two* different worst-case distributions as shown in figure 2. These arise in the following way:

- the form of the distribution of the purely random component of the error is observed or inferred as it might be for the case of a class 3 error, but its mean μ_{Δ} is unknown,

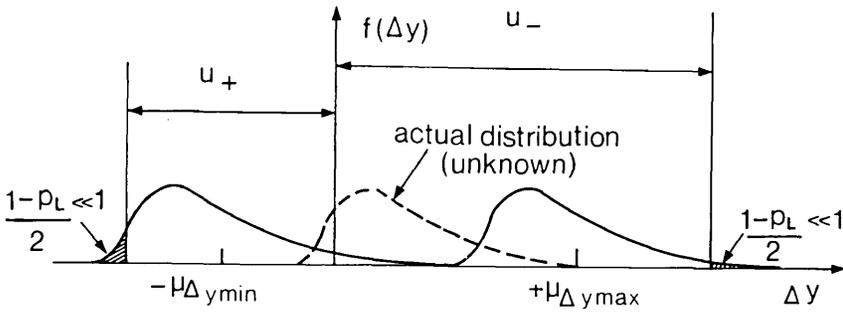


Figure 2—The orthodox method of calculating uncertainties corresponding to class 2 errors applying a confidence level to two worst-case distributions.

- the maximum positive limit on its mean, $\mu_{\Delta y_{max}}$, is obtained by summing its component limits in the way argued for above,
- similarly, a minimum negative limit on its mean, $-\mu_{\Delta y_{min}}$, is calculated,
- one worst-case distribution is obtained by setting $\mu_{\Delta y} = +\mu_{\Delta y_{max}}$ simply because this represents one of the two worst conceivable situations.
- similarly, the other worst-case distribution is obtained by setting $\mu_{\Delta y} = -\mu_{\Delta y_{min}}$.

The probable presence of large positive errors implies the necessity of a large negative uncertainty. Thus in order to obtain a value for u_- , the conservative now “slides” a vertical boundary out along the positive error axis until a small fraction $(1-p_L)/2$ of error values is enclosed beneath the curve to the right of the line. A similar process conducted in the opposite direction will yield u_+ . (The convention of choosing a value of p_L to exclude a fraction $(1-p_L)/2$ rather than $(1-p_L)$ ensures continuity with the usual convention for class 3 errors as $\mu_{\Delta y_{max}}$ and $\mu_{\Delta y_{min}}$ both approach zero.)

This procedure covers cases where positive and negative systematic uncertainties or the random components of errors or both are disposed asymmetrically. But it does not, of course, allow probabilities to be associated with u_+ and u_- as with the uncertainties corresponding to class 3 errors because none was associated with $\mu_{\Delta y_{max}}$ and $\mu_{\Delta y_{min}}$. It might be said that *at least* an estimated fraction p_L of the results of a repeated experiment would lie between $y-u_-$ and $y+u_+$, but not that an estimated fraction $1-p_L$ would lie outside this range. For this reason no probabilities can be associated with the compatibility of two experimental results y_a and $y_b (>y_a)$ where either or both have mixed errors. They agree if $u_{+a}+u_{-b} > y_b-y_a$. If $u_{+a}+u_{-b} < y_b-y_a$, they disagree. On the orthodox view, there is no more to be said.

No analogous analysis of error-related quantities other than uncertainty is offered here. The estimation of expected values of errors, of their expected absolute values or of rms values requires that $\mu_{\Delta y}$ or $E(\mu_{\Delta y})$ is known. The conservative believes that they are equal because the former is a constant and not a distributed random variable. But since it is an unknown constant he is bound to regard the derivation of expressions for expectations to be of no practical use. It will be seen later that supporters of the randomatic theory take a different view and that $E(\mu_{\Delta y})$ is assumed to be calculable even when $\mu_{\Delta y}$ is not known with certainty.

2.3 Error Types and Their Identification on the Orthodox View

It has already been noted that there is often confusion about what populations can legitimately be said to contain random errors for the purpose of estimating uncertainties in experimental results. Since the “combination” of uncertainties depends on the identification of the corresponding error types, this is a matter of some practical importance and the confusion needs to be resolved.

At the beginning of section 2.1, it was pointed out that the distinction between systematic and random errors was sometimes based upon error causes [approach 2)], rather than upon the behavior of errors as experiments were repeated [approach 1)]. It will be clear that equivalent definitions of “systematic” and “random” could be cast in terms of causes, provided allowance was made for any non-linear dependence upon them of the resulting errors.

While the above possibility is widely intuited, it has never been developed to the author’s knowledge. It nevertheless seems tacitly to underpin a different and much less satisfactory type of exercise intended to define systematic errors in terms of their causes. Here systematic errors in results are defined by an enumeration of systematic causes.

Some typical cases are shown in box 4. But such enumerations amount only to definition by example and so cannot be fundamental. Without a prior criterion stating how causes are to be related to one category or the other, the classification would not be possible. In the absence of a statement of the criterion the procedure remains obscure and, equally important, there is no way of telling if it is complete. It seems clear, however, that in each case the vague underlying notion is that such errors are class 1 or constant errors. Unfortunately the errors listed in box 4 do not always behave as class 1 errors when the *relevant* trial is repeated, as will be argued below. The enumeration is based on simplistic rules-of-thumb which are no substitute for a physical analysis of the way their causes operate. In what follows, errors defined in this way will be referred to as "so-called" systematic errors to distinguish them from those defined in terms of their behavior (cf. box 3).

Also mentioned at the beginning of section 2.1 were definitions of "systematic" and "random" cast in terms of how errors or uncertainties are actually estimated or treated in a given case, rather than in terms of what is possible or proper in view of their nature [approach 3)]. When such definitions are offered it is sometimes unclear whether it is intended that the method of evaluation or treatment determines what category errors fall under or *vice versa*. Here it will be assumed in the interest of objectivity that it is the nature of the error which

determines the correct method of evaluating its corresponding uncertainty, so that no definition purporting to be fundamental need ever mention the actual methods of evaluation employed in given cases. The opposite view is again related to the aforementioned confusion about which populations of results can be said to contain random errors. The errors of box 4 frequently arise from random processes and so arguably contain a random component. But such components will have a constant effect when combining with errors in the results of other measurements, no matter how often, the latter are repeated. From this it is sometimes concluded that the errors of box 4 cannot be held to contain a random component for the purpose of calculating an uncertainty in one's own result. Thus one author, having given examples of systematic errors, writes, "There is no strict definition of systematic errors, since what is systematic for one experiment may not be for another" [12]. Another states, "One has to remember that some errors are random for one person and systematic for another" [13]. This outlook may have led to the mistaken view that the central orthodox distinction is mutable and that the labeling of errors as "random" or "systematic" is somehow conventional. Credence has thereby been lent to the notion that what is actually done is as important as why it is done. Fortunately the confusion can be removed by resolving certain more fundamental ones about probabilities and it can be stated definitively what

**Box 4. An Orthodox Definition of 'Systematic Error'
By Enumeration of Causes
(‘So-called’ Systematic Errors)**

"Systematic errors" are those owing to:

- Single readings
 - rounded
 - interpolated
- Instrument errors
 - calibration errors
 - other scale errors
 - errors due to "subclinical" malfunction
 - errors due to bad practice
- Residual correction errors arising from inexactness in correcting for known systematic effects
- External errors arising from results taken from other experiments

populations can correctly be said to contain random errors. It will then be clear that the distinction between random and systematic errors is an objective one and that the nature of any given error is fully determined once a complete experimental specification has been formulated.

The confusion is well illustrated by the much-discussed problem case of external errors. Two possibilities exist for their treatment. The error in an external quantity might be taken to be entirely systematic, even where the worker producing the result can be said to have correctly assessed it as being entirely random or part random and part systematic. The alternative is to take over the error assessment, assumed correct, of the original worker in deriving the uncertainty of one's own result. The justification of the former view is that no matter how often the main experiment is repeated, errors in the external result will always affect the answer in the same direction and to the same degree; i.e., it is a class 1 error. Thus one experimenter's random error is another's systematic error. But the opposing view notes that if the external measurement had been conducted by the "borrower," it would be regarded as an ancillary measurement in his own experiment and no question would arise of changing any random component in its uncertainty to a systematic component. Who did what is held to be an unphysical consideration which could not change the nature of an error and so the original worker's analysis is to be retained.

The position will be adopted here that the latter argument is the correct one. While the random component of an external error certainly will affect the result of the experiment in hand with a definite sign and magnitude (what other way is there?), this is true only in the sense that it is true for its own internal random error. The best estimate of the internal error introduced by an external random error component is zero with an uncertainty based on the latter's distribution (cf. *Campion et al.*, [14]).

To see this more clearly some may find it helpful to consider a simple gaming analogy of a class 3 error. Suppose a die, possibly loaded, is thrown repeatedly to estimate the unknown expected value $\mu_n \equiv E(n)$ of the outcome n (1 to 6). Here μ_n is analogous to a true value to be determined by measurements, n to the observed (digital) results, and $n - \mu_n$ to a class 3 error. After a single throw the "error" $n - \mu_n$ is, like a measurement error in a single reading, physically determined, but unknown. Nevertheless, everyone would accept that some unknown, but objective, probability was in principle to be associated with it and that this would be the same whether one threw the die oneself or

someone else threw it. For example, in the case of a fair die $p(n - \mu_n) = 1/6$ for any n where $\mu_n = 21/6$.

Suppose further that it is desired to assess the uncertainty in an estimate \bar{n} of μ_n obtained as the mean of 100 outcomes, one of which was provided by an outsider. Would it be correct to calculate random uncertainties $\pm u$ in $\Sigma n/100$ at, say, the 95% level of confidence for 99 throws and then to augment these by maximum error limits, $u_+ = (\bar{n} - 1)/100$ and $u_- = (6 - \bar{n})/100$, corresponding to the single "external" result? Of course no one would proceed in such a way. The uncertainties would be calculated at the 95% level of confidence for the full 100 throws. Thus "external" random quantities are to be treated no differently from "internal" ones as asserted.

This justification for the ruling that external random errors are not to be distinguished from internal ones is easily generalized and so provides a basis for the resolution of the question of which populations of results can legitimately be said to contain random errors. It means that in principle *any* error determined by a random process, whether under the management of the experimenter or not, is to be treated as a random or part-random error even if it occurs in a single result. This accords with universally accepted statistical principles as applied in the above gaming example and licenses experimenters to treat many so-called systematic errors, or components of them, as random. Thus random errors in the result of an experiment can arise from external results, from calibrations and, if an experimenter's instrument can realistically be said to be sampled at random from some population, from instrumental imperfections which do not change in the course of his measurements. Similarly, if a conservative were, unusually, in a position to assess "credible bounds" for a so-called systematic error in his experiment using knowledge of the form of its random component, these bounds too should be treated as partly or wholly random uncertainties. As a result of these reforms of orthodox practice, reductions can be made in many overall experimental uncertainties conservatively estimated on the incorrect assumption that some of their components were purely systematic. This is so because combination in quadrature is permitted for the standard deviations of the newly identified random error distributions. However, as will become clear in section 3, the fact that all assessable external random error components are to be treated as such does not imply that all external error components are random.

From these considerations it is clear that the identification of an error as random or systematic

or a mixture of both should be based on an analysis of the way its causes would operate upon repetition of the experiment or some associated trial. In particular it is necessary to identify all random mechanisms which can affect it even though they are normally regarded as being outside the experiment. The correct identification of error types will be ensured if attention is directed to the *whole* experiment including those parts conducted by others and the random trials performed, perhaps unwittingly, by oneself (for example, the choice of an instrument). Each repeatable operation in the whole experiment, whether actually repeated or executed just once, should have exactly the same status as one's own repeated measurements. This broad and rational outlook may be contrasted with the uncritical use of rules of thumb such as those illustrated in box 4.

Six common related failings of conservative argument and practice have been encountered in this section. To summarize, they are . . .

- vagueness about the meaning and objectivity of the basic distinction between random and systematic quantities,
- a confusion about scope: which populations can be said to contain random errors?
- vagueness about the correct method of combining systematic uncertainties.
- vagueness about the correct method of calculating uncertainties corresponding to mixed (class 2) errors.
- the misidentification of error types by the naive use of rules of thumb, and
- failure to notice the widespread existence of random errors.

The realization that the role of random errors in experiments is much wider than orthodox assumptions sometimes allow has doubtless been a stimulus to the alternative view offered by the randomatic theory. And the fact that conservatives appear to have forgotten the reasons for orthodox practice has made the task of the randomatic theory's proponents an easier one.

3. The Randomatic Theory of Errors

To those nurtured on the orthodox view, the randomatic theory of errors seems initially to be very radical. As the first main tenet of the theory, the distinction between random and systematic errors is held either to be a merely conventional distinction without an objective difference or to be a real, but irrelevant, distinction for the purposes of determining uncertainties in practical cases. How-

ever, which of these views is held by any given proponent of the randomatic theory (or "randomaticist," for ease of reference) may not always be clear. The second main tenet of the theory is that all uncertainties are to be calculated by statistical techniques, for example by combining "standard deviations" in quadrature irrespective of how a conservative error theorist would classify their corresponding errors.

There can only be three main types of justification of the new theory. The first, presented in section 3.1, takes as its starting point the generally agreed law of error propagation and uses it to attempt to show that random and systematic standard deviations, so-called, are logically required to be combined in quadrature [15-17]. That being the case, the distinction, whether originally valid or not, is shown to be irrelevant for arriving at an overall assessment of uncertainty. On this approach, the randomaticist's second tenet would appear to be the more fundamental.

The second and third justifications of randomatic procedures depend not on the law of error propagation, but on the assumption or perception (depending on one's position) that all errors of interest classified by the conservative as systematic can be associated with random variables having a parent population over which a probability distribution can be defined. The conservative, on this view, has simply failed to notice something useful; namely that all errors are random and can, even according to his own beliefs, be treated statistically. It is thus implied that the central conservative distinction corresponds to no real difference and the first randomatic tenet plays the more fundamental role by providing a justification for the second. This type of argument can be fundamentally different when cast in frequency-based terms (Justification 2 presented in section 3.2) from that cast in subjectivist terms (Justification 3 presented in section 3.3). A number of authors have proposed procedures for uncertainty estimation based on the assumption that all errors can be represented by random variables, but it is generally unclear whether Justification 2 or 3 is intended [18-21].

3.1 Justification 1: Randomatic Theory via the Law of Error Propagation

The law of error propagation, which is quite uncontroversial, states that the error Δy in the result y of an experiment is given by

$$\Delta y = \Sigma (\partial y / \partial x_j) \Delta x_j$$

where the Δx_j are the errors in the various individual or repeated measured values of x_j of the experiment. To illustrate why it is thought that both random and systematic "standard deviations" are to be combined in quadrature the simple case will be considered where the required result of an experiment is the mean of n similar results x_j : $y = \sum x_j/n$. Let the x_j suffer systematic errors $+a$ and $\times b$ with a random error R_j so that $\Delta x_j = a + (b-1)x_j + R_j$. Approximate corrections $-(a-\Delta a)$ and $\div(b-\Delta b)$ would generally be made to the observed x_j whereupon $\Delta x_j = \Delta a + \Delta b x_j + R_j$. It is easily shown that, provided the expectation $E(\Delta a \Delta b x_j) = 0$, the law of error propagation implies:

$$E(\Delta y^2) = \Delta a^2 + \Delta b^2 \bar{x}^2 + \sigma_x^2/n$$

where $\sigma_x^2 \equiv E(R_j^2)$ i.e., "standard deviations" of residual correction errors and random errors combine in quadrature. It is assumed that in any well-designed experiment significant systematic errors will always be corrected for and that this therefore provides a general rationale for practical random procedures of the kind which those recommended in box 1 appear to be.

Problems with Justification 1

The conservative, is unlikely to find Justification 1 convincing for two reasons. Firstly he would not accept that systematic errors were always corrected for in well-designed experiments. There are many cases where a systematic error is tolerably small and where a reliable correction is difficult to estimate. Here the experimenter will often prefer to leave it uncorrected and to estimate the uncertainty in terms of bounds. For the argument to work in such a case it would be necessary to assume not that $E(\Delta a \Delta b x_j) = 0$, but that $E(a(b-1)x_j) = 0$, which is only true in general if the errors $+a$ and $+(b-1)x_j$ are uncorrelated class 3 ones like the R_j . But this is exactly what the conservative denies; it will be recalled that the existence of class 1 errors is the first principle of his theory. If anyone repeating the experiment according to the same experimental specification could be expected to encounter the same constant values of a and b , then it would be the case that $E(a(b-1)x_j) = a(b-1)\bar{x}$. More generally, if the errors $+a$ and $\times b$ were always drawn from the same two respective populations with unknown non-zero means μ_a and μ_b , then $E[a(b-1)x_j] = \mu_a(\mu_b - 1)\bar{x} \neq 0$.

The conservative's second objection would be that, even where corrections are made for systematic errors, different residual correction errors are

not generally statistically independent class 3 errors either. For example if a and b arose from two corrections made for systematic effects on the basis of simplified theoretical models which all experimenters following the specification would be expected to use, then Δa and Δb could be constant class 1 errors in which case $E(\Delta a \Delta b x_j) = \Delta a \Delta b \bar{x}$. From this it follows that

$$E(\Delta y^2) = (\Delta a + \Delta b \bar{x})^2 + \sigma_x^2/n$$

which is the usual conservative formula with systematic errors combining together in a linear way. The enlightened conservative believes that typically residual correction errors, like most so-called systematic errors, would be of class 2 so that $E(\Delta a \Delta b x_j) = \mu_{\Delta a} \mu_{\Delta b} \bar{x} \neq 0$ as before. It is therefore the case that Justification 1, though perfectly correct given certain random presuppositions, cannot be used to establish those presuppositions on pain of circularity. The conservative will see the argument as begging certain key questions as controversial as Tenet 2 itself. The same applies to any justification employing a statistical proof that standard deviations combine in quadrature and which implicitly assumes that all errors are random variables of zero mean (e.g., [22]). It will become clear that Justification 1 is implicitly dependent on the reliability of Justification 2 or 3.

3.2 Justification 2: Ransomatic Theory via Frequency-Based Statistical Distributions

This justification depends upon the assumption that every systematic error belongs to a well-defined stable population which can be generated by repeated measurements or by some other repeatable trial associated with the experiment. For example a barometer zero error might be said to be long to and be sampled from the population of zero errors realizable by constructing an infinite population of barometers to the same engineering specification and perhaps subjecting them to the same calibration procedure. The error would thus be fixed for any given experimenter executing the experimental procedure, but would be a random variable analogous to a single reading (cf. the discussion of conservative attitudes to such quantities in section 2.3). From this it might be argued that all errors were random errors.

Problems with Justification 2

The results of measurements repeated according to a clear experimental specification, and the corresponding errors, belong to well-identified populations; those defined in advance by the specification.

But what population do those systematic errors "outside" the experiment belong to? It might be argued that if a systematic error could be assigned to more than one population equally naturally with no means of identifying the "right" one, different but equally correct standard deviations and uncertainties could be derived. They could not therefore be objective quantities (cf. Ayer [23]). For example, does a barometer zero-error belong to 1) the population of zero-errors realized by repeated constructions to the same specification or to 2) the different population of zero-errors to be found in barometers available for use in (say) British laboratories? Since randomaticists do not identify their populations, but simply invoke distributions or even just standard deviations, their calculated uncertainties cannot in practice be objective frequency-based ones.

However, though this may be true of informal practice, there is no deep problem of principle here for the randomatic theory. So-called systematic errors really can belong to several natural populations from which they are simultaneously sampled. The experimenter may use his approximate knowledge of these to choose or define the population characterized by the smallest errors as the basis of his calculation of uncertainties, provided that the population involved really would be randomly sampled by repetition of the error selection procedure actually employed in his experiment (e.g., through the purchase of a barometer by his organization). If, for example, he judges that zero errors of British barometers in general would only very rarely exceed ± 30 Pa, but that his particular design would limit this to ± 10 Pa, then it is legitimate to use the latter information ignoring the former. Different experimenters may draw their barometers (and their zero errors) from the same or different populations. But if these are properly identified and their corresponding distributions or bounds plausibly assessed, uncertainties will be correctly estimated in each case. Because the population sampled is a determinant of the experimental result and its error, it will be supposed in all that follows that it must be explicitly or implicitly identified in a complete experimental specification and is not to be regarded as a matter "outside" the experiment (cf. section 2.3).

But there is a more serious objection to Justification 2 than the charge that randomatic populations of systematic errors are not uniquely identifiable. This states that their distributions are not generally of class 3 having zero means. As noted in the previous section, unknown non-zero means for residual correction errors are only to be

expected. And this is true in general of the frequency-based distributions characterizing genuinely physical error populations. For example, there is no reason to suppose that the populations 1) or 2) above have zero means. Indeed there exist many errors which can only have one particular sign and for which corrections are not made. Thus if randomatic procedures are to be justified, it cannot be in terms of frequency-based distributions.

To avoid this conclusion it would have to be demonstrated that the means of class 2 systematic error populations (so-called) were themselves class 3 random variables appearing with different frequencies in some physical population sampled by the experiment. Then it would arguably be appropriate to convolute the distribution of the mean with that representing the purely random variation of the so-called systematic error to yield a frequency-based class 3 resultant distribution as required by the randomatic theory. But, the view that means (systematic errors proper) are distributed in the sense of appearing with different frequencies in some physical population would betray a misconceived identification of the relevant experiment and population. It has been noted that complete experimental specifications must identify, albeit implicitly, a particular so-called systematic error population as an essential feature of the experiment. While the corresponding distribution and mean are not known, they are determined through the definition of the experiment and not through some external random trial. Different workers independently following exactly the same experimental specification will therefore sample the same error population, producing results with a random variation, but all sharing the same bias from the true value. Thus the mean of the error is clearly sampled from a population of just one value. Since experimenters are interested in estimating the maximum possible or conceivable error in a *particular* specified experiment, the conservative is right to regard the mean as an undistributed quantity or as being "distributed" as a delta function at some unknown location. Of course, if the error population were investigated statistically, an estimate for its mean could be obtained and the error in the estimate characterized by a random distribution. The mean would be corrected for and the error in the mean would be treatable as random. However, the mean would then, by definition, not be a systematic error, but a measured quantity.

The randomaticist, if he seriously invokes a frequency-based distribution for the uninvestigated mean, is implying that it is in a literal sense singly sampled from some wider population once-and-for-

all on behalf of all the independent experimental repetitions which could ever be conducted. Perhaps the population envisaged would be that of systematic errors, positive and negative, encountered in experiments in general, with the credible or maximum bounds re-scaled and re-dimensioned in each case to match those of the experiment in hand. Apart from the problematic question of whether this "super-population" of means itself has a zero mean, there would be no objection to randomatic procedures if this experiment were like that for which an uncertainty is required. But it is quite different from the conception of the experiment normally held. If *this* experiment were repeated, there would be a grand prefatory sampling of the error mean on each occasion followed by repetitions of the experiment as normally conceived. The results and the errors would then be different from those of the experiment for which an uncertainty is sought.

The constancy of the unknown mean which the conservative takes so seriously is therefore of quite a different nature from that of the determined outcome of a prefatory single sampling. It is built into the common concept of an experiment as a definite specified trial. As such there is no frequency-based rationale for treating uninvestigated systematic error means statistically and any justification of randomatic principles must hang on subjectivist arguments.

3.3 Justification 3: Randomatic Theory via Subjectivist Statistical Distributions

Modern subjectivist statisticians frequently identify probabilities with rational "degrees of belief." Their general method is 1) to assign prior probabilities $p(x_j)$ to the possible results x_j , of some trial reflecting their beliefs prior to making observations, and 2) to modify these in the light of evidence E (observed frequencies) using Bayes' theorem to yield posterior probabilities:

$$p(x_j|E) = p(E|x_j)p(x_j) / \sum_k p(E|x_k)p(x_k)$$

In this way posterior values converge with those evaluated conventionally and they "realize" the same distributions as others. These will of course be different in general from their prior distributions.

In many subjectivist treatments of statistics, the psychological concept of a "degree of belief" is defined in terms of the betting odds which a subject would just be prepared to accept on such-and-such being the case. This notion, together with certain weak rationality constraints, for example betting in

such a way as to avoid becoming the victim of a Dutch book, are held to be sufficient for deriving the axioms of probability theory.⁵

A familiar example of intuitive subjectivist practice is afforded by the situation where an experimenter has no information on whether a so-called systematic error is positive or negative and knows nothing about its magnitude except that it cannot exceed $|a|$. Having no reason to believe any value in the range $\pm a$ more or less probable than another, he invokes a uniform Laplacian distribution of magnitude $1/2a$ between $\pm a$. Such a distribution is of course of class 3.

It has already been noted that it is essential to the randomatic theory that any distribution used to calculate uncertainties is of class 3. This is because a standard deviation, the only recognized "measure" of uncertainty, is defined relative to its distribution mean and so cannot reflect uncertainty arising from an unknown and unobservable non-zero distribution mean. (Covariances too, like those invoked in box 1, are defined relative to means and so can only allow for bias arising from correlation between purely random components of errors.) Unlike frequency-based distributions for systematic errors, subjectivist distributions are generally of class 3 because the sign of a systematic error is typically unknown. Where the distribution is not of class 3 corrections are sometimes applied to make it so. That subjectivist distributions are of class 3 is the great strength of Justification 3 compared to Justification 2.

However, because class 3 prior distributions which are not frequency-based are by definition undetermined by evidence, they are not objective. Different subjectivists will invoke different prior distributions and so calculate inconsistent standard deviations and uncertainties. More importantly, they will disagree in general from those which could be realized by repetition of the relevant trial. If they were to agree, it would be because there was sufficient knowledge to calculate approximate

⁵ The most detailed foundational development of subjectivist theory in terms of rational betting and Bayesian conditionalization has been undertaken by de Finetti [24,25]. Useful reviews of his work may be found in Gillies [26,27]. Clear statements of subjectivist ideas have also been presented by Savage [28,29] who in the latter gives a schematic subjectivist account of a physical measurement (the weighing of a potato). Two wide-ranging collections of subjectivist and Bayesian papers are "Studies in Subjective Probability," edited by Kyberg and Smokler [30], and "Bayesian Analysis in Econometrics and Statistics," edited by Zellner [31]. A modern mathematical text extending the tradition in a philosophically conscious way is "Statistical Decision Theory" by Berger [32]. A well-known critique of subjectivist methodology was presented by Fisher in "Statistical Methods and Scientific Inference." [1]

frequencies in advance. But then frequency-based distributions would have been invoked which are not generally of class 3 (cf. the preceding section). Randomaticists sometimes depend on class 3 subjectivist distributions to justify their recommended procedures for calculating uncertainties, but then make the incorrect assumption that their uncertainties are objective as they would be had their invoked distributions been frequency-based.

The lack of objectivity of prior distributions appears to some to be a fatal flaw in subjectivist statistics. In contrast, subjectivists see it as no great problem. They accept nowadays that such prior distributions are often non-objective best guesses or unbiased starting points for a Bayesian inference and are prepared to engage in mathematical analyses of "robustness" with respect to their uncertain features (roughly, how insensitive the posterior distribution is to any lack of realism in them) [e.g., 33].⁶ Objectivity is achieved through evidence and the process of Bayesian conditionalization. Unfortunately systematic error distributions are by their nature never investigated statistically and conditionalized. In this respect error theory differs crucially from other fields of statistical practice where Bayesian methods are employed. *Thus even subjectivists would regard subjectivist prior error distributions and a randomatic theory based on them as non-objective.*

With the failure of this justification, it is seen that in spite of its attractive features the randomatic theory lacks an objective foundation. Moreover, if uncertainties are defined to be maximum possible or conceivable errors in the results of particular specified experiments, the lack of objectivity can be expected to result in underestimated uncertainties on occasions. For example, two equal systematic uncertainties $\pm a$ would combine to yield a resultant uncertainty of $\pm 2a$. By treating the corresponding errors as being normally distributed, say, on the grounds that smaller errors are generally to be expected more often than larger ones, and associating the credible bounds $\pm a$ with some confidence level close to one, the randomaticist will calculate an uncertainty corresponding to the resultant error of $\pm 1.41a$ at the same level of confidence. Under unfavorable circumstances both combining errors could be close to the same bound so that their resultant would virtually always lie outside the randomatic uncertainties. Although some experi-

menters have a compelling intuition that such unfavorable occurrences are "improbable," especially where larger numbers of systematic errors combine, it is impossible to provide a physical rationale for this if systematic errors cannot be random variables in any objective sense. After all, none betting on the joint outcome of several particular dice known to have various undetermined biases would assume that the biases had zero expectations unless, unlike systematic errors, they had been randomly selected from a population in which this was true. The psychological origin of the intuition is no doubt the desire to believe that probabilities will always provide a basis for rational inference and action in the face of uncertainty. (cf. Fisher [35]). But if the preceding arguments are correct, this would appear not to be so.

4. Conclusions and Recommendations

It has been argued that the distinction between random and systematic errors is, when properly formulated, a clear and objective one applicable to all error populations. Moreover, which category an error falls into should determine whether it is treated statistically or not. Thus there is no room for a different fundamental distinction between error types A and B based on the method of treatment employed in given cases (cf. box 1). Frequently, conservatives have automatically taken so-called systematic errors like those of box 4 to be entirely systematic when they have in fact contained an assessable random component. Conversely, randomaticists have implicitly assumed that all so-called systematic errors are class 3 random errors having zero means. Thus both typical conservative and randomatic practices are based on unrealistic principles.

Given that many so-called systematic errors do contain both a random and systematic component, how are their corresponding uncertainties to be assessed? The formal answer to this question has already been given in section 2.2 where a procedure was recommended for estimating the uncertainty corresponding to a class 2 error (cf. fig. 2). However, this procedure is very often difficult to apply because insufficient information is available to characterize separately the class 1 and class 3 components of class 2 errors.

In dealing with such cases, experimenters of all persuasions often feel able to judge maximum possible or credible bounds beyond which the unknown distribution is certain to cover zero or a negligible probability. The bounds will often be

⁶ Early subjectivists sometimes regarded Bayes' postulate or some variant thereof as an unchallengeable axiom; this view led to well-known logical difficulties and is not generally held by modern subjectivists [34].

symmetrical about zero, but it should not be supposed that they correspond to equal confidence intervals of the real distribution or license the experimenter to invoke a symmetrical or other class 3 distribution spanning them. As they are estimated from the worst possible or conceivable combination of physical effects it may well be that the incidental physical influences on the parent error population cause one or both tails of the distribution to become negligible well within their respective bounds (cf. the three curves shown in fig. 2). Thus many distributions are consistent with the choice of bounds and the mean of the real distribution could in principle lie anywhere between them.

Typically the experimenter will be unable to partition his uncertainty as defined by the bounds exactly into random and systematic components. Guided by the definition of uncertainty as maximum possible or conceivable error, the rigorous worker will adopt as the basis of further calculations a model derived from the maximum apportionment of uncertainty to the systematic category judged possible or conceivable. This is because resultant uncertainties calculated on the basis of an overestimated random component would be too small as random components combine in quadrature rather than additively. His judgment of the maximum apportionment of uncertainty to the systematic category will require him to assess the maximum range in which the mean of the actual distribution could lie. Thus, just as credible bounds were initially placed on the so-called systematic error distribution itself, so narrower credible bounds are placed on its mean, the systematic error as properly defined. Like the outer bounds, the inner bounds will not correspond to confidence limits and do not confine the mean in that sense; there is no pre-existing statistical sample to enable an estimate of the mean and the standard deviation of the estimate to be made. (If there were the mean would be like a measured quantity and it would be proper to correct for it and treat the remaining component as a class 3 random error). However, much vaguer information, perhaps that *in the given experimental context* smaller errors are more common than larger ones, can sometimes justify restricting the range of possible values of the mean. Where the information required for this is lacking, the inner credible bounds will become coincident with the outer which then become the systematic uncertainties they have so often been taken to be.

It may at first sight be thought that making such judgments is hopelessly arbitrary and the problems have to be acknowledged. But experimenters usually design their experiments so that the difficult

uncertainties to evaluate are the least significant. Then some inaccuracy in judgments is tolerable. The difficulties are in any case largely common to all theories of error: judgments of maximum limits or credible bounds for the conservative and of systematic standard deviations for the randomistic are often arbitrary in problem cases. However, their simple rules of procedure only disguise the difficulties without removing them. The approach recommended above brings them into the light and, while it calls for an additional judgment apportioning uncertainty between random and systematic categories, this is not markedly more difficult than those already required. More fundamentally, if the approach is realistic, as has been argued, it should be physically correct.

With these points in mind the following procedures are recommended for the estimation of experimental uncertainty.

Recommendations for the Evaluation of Experimental Uncertainties

1 The *whole* experiment should be defined (cf. section 2.3). All measurements, corrections, calibrations, external results, and single random samplings contributing to the final result of the experiment should be listed. All significant sources of error in the experimenter's own part of the whole experiment should be identified. The nature and magnitude of uncertainties in all other results should be ascertained.

2 Choose a confidence level (e.g., 95 or 99% probability) beyond which possibilities are regarded as being inconceivable. This level should be clearly stated.

3 Decide to which class, 1 to 4, each error belongs. This decision should be made irrespective of whether the measurement or trial was actually repeated or not; the definition of these classes is in terms of what *would* be observed on repetition (cf. section 2.1). Where measurements have not been repeated, it should be possible to identify the class of any error from the specification of the measurement concerned giving the nominal conditions and procedures required for its execution and their permitted variations. In the case of single trials associated with an experiment (e.g., the selection of an incompletely characterized instrument or material) relevant error populations should be identified and one chosen as a basis for uncertainty estimation which minimizes the uncertainty (cf. section 3.2).

4 If some subsidiary result of the experiment is observed to be subject to a significant class 4 error (so introducing a class 4 error into other

results which it is used to calculate), attempt to identify the weak aspect of the control of the experiment which allowed it to occur. This may be done by experimental tests or by analysis of the experimental specification or both. When identified, repeat the experiment with better control if practicable. Alternatively, estimate the maximum range of values which the uncontrolled causative condition or conditions could possibly or conceivably take, and use these to compute maximum possible or credible errors in the quantities concerned. Treat these as systematic uncertainties according to the procedure of paragraph 5. If the source of the class 4 error cannot be identified, then of course no final uncertainty may be calculated.

5 Estimate the maximum possible or credible absolute values in each direction of the class, 1 errors and of the constant components of the class 2 errors. Again, this may be done by reference to presumed measurement specifications or identified pre-existing error populations. Multiply each uncertainty by the coefficient $\partial y/\partial x_j$ in the law of error propagation to obtain corresponding uncertainty components in the final result of the experiment. Add these together to obtain an overall systematic uncertainty in the final result.

6 Identify those class 2 and 3 sources of random error which contribute directly to the final result. Multiply the observed or estimated standard deviation of each by the coefficient $\partial y/\partial x_j$ to obtain corresponding components for the final result. Combine these in quadrature to yield a standard deviation for its random component of error.

7 Having observed or inferred the form of the random component of error in the final result of the experiment, use the systematic uncertainties of paragraph 5 to define upper and lower limits for its mean, $\mu_{\Delta y_{\max}}$ and $-\mu_{\Delta y_{\min}}$, thus obtaining two "worst-case" distributions. Use the confidence level of paragraph 2 to calculate corresponding uncertainties, u_+ and u_- , according to the procedure of section 2.2 (see fig. 2). These overall uncertainties should be quoted together with 1) their systematic components, 2) their common random component, 3) the confidence level, and 4) any useful additional statistical information e.g., the number of degrees of freedom in calculated means or fitted curves (cf. Campion et al. [10]).

Although the responsibility for the views expressed above remains his alone, the author gratefully acknowledges his debt to many others with whom he has agreed or disagreed on the subject

including Dr. K. G. Birch, Mr. J. E. Burns, Dr. P. J. Campion, Dr. E. Richard Cohen, Mrs. Mary C. Croarkin, Prof. Jon Dorling, Dr. K. R. Eberhardt, Mr. D. R. Fisher, Prof. P. Giacomo, Dr. Donald A. Gillies, Dr. Harry H. Ku, Dr. K. T. Maslin, Dr. J. W. Mueller, Dr. B. W. Petley, Prof. H. R. Post, Dr. K. C. Shotton, Prof. J. Skakala, Dr. T. J. Quinn, Mr. F. J. Roberts and Prof. Dr. K. Weise.

References

- [1] Fisher, Sir Ronald A., *Statistical Methods and Scientific Inference*. Oliver and Boyd: Edinburgh and London, p. 35 (1956).
- [2] Dietrich, C. F., *Uncertainty, Calibration and Probability*. Adam Hilger: London (1973).
- [3] DIN 1319, Part IV, Basic Concepts of Measurements; Treatment of Uncertainties in the Evaluation of Measurements, Beuth, Berlin (1985).
- [4] British Calibration Service. Guidance Publication 3003, *The Expression of Uncertainty in Electrical Measurements* (1980).
- [5] Servizio di Taratura in Italia. Procedura di Taratura di Blocchetti di Riconcontro Piano Paralleli con il Metodo di Confronto Meccanico (unpublished) paragraph 13.
- [6] Giacomo, P., News from the BIPM, *Metrologia* 17 69-74 (1981).
- [7] Report of the BIPM Working Group on the Statement of Uncertainties (1st meeting—21 to 23 October 1980) to the Comite International des Poids et Mesures, Kaaris, R., rapporteur, BIPM (1981).
- [8] Report on the BIPM Enquiry on Error Statements, Report BIPM-80/3, BIPM (1980).
- [9] Procès-Verbaux des Seances, Comite International des Poids et Mesures 48 pp. A1-A10 (1980).
- [10] Campion, P. J.; J. E. Burns and A. Williams, *A Code of Practice for the Detailed Statement of Accuracy*, National Physical Laboratory, HMSO, London, page 23 (1980).
- [11] Eisenhart, Churchill, Realistic evaluation of the precision and accuracy of instrument calibration systems, *J. Res. Natl. Bur. Standards* 67C pp. 21-161 to 47-187 (1963).
- [12] Barford, N.C., *Experimental Measurements: Precision, Error and Truth*, Addison-Wesley Publishing Co. Inc.: London, Reading (Massachusetts), Menlo Park (California), Don Mills (Ontario) (1967).
- [13] Vigoureux, P., in *Precision Measurements and Fundamental Constants, Proceedings of the First International Conference on Precision Measurements and Fundamental Constants*, National Bureau of Standards, Gaithersburg, MD. N. Langenberg and B. N. Taylor, eds. (NBS Special Publication 343, 1971), pp. 493-525.
- [14] Campion, P. J.; J. E. Burns and A. Williams, Ref. 10: *ibid.*, page 2.
- [15] Mueller, J. W., The assignment of uncertainties to the results of experimental measurements (Review), in *Precision Measurement and Fundamental Constants II, the Proceedings of the Second International Conference on Precision Measurement and Fundamental Constants*, National Bureau of Standards, Gaithersburg, MD, B. N. Taylor and W. D. Phillips, eds. (NBS Special Publication 617, pp. 375-381 1984).
- [16] Mueller, J. W., Some second thoughts on error statements,

- Nuclear Instruments and Methods **163** 241-251 (1979).
- [17] Mueller, J. W., Some Second Thoughts on Error Statements, BIPM Report 78/4 (1978).
- [18] Wagner, S., Zur Behandlung systematischer Fehler bei der Angabe von Messunsicherheiten, Physikalisch-Technische Bundesanstalt Mitteilungen **5** 343-347 (1969).
- [19] Wagner, S. R., On the Quantitative Characterization of the Uncertainty of Experimental Results in Metrology, Physikalisch-Technische Bundesanstalt Mitteilungen **89/2** 83-89 (1979).
- [20] Woeger, W., Randomization of systematic errors and its consequences for the evaluation of measurements, in Quantum Metrology & Fundamental Physical Constants, proceedings of the NATO Advanced Studies Institute Conference. Erice, Sicily, Paul H. Cutler and Armand A. Lucas, Eds., Series B, Vol. 98, Plenum, pp. 613-621 (1983).
- [21] Weise, K., Vorschlag zum Begriff der Messunsicherheit, Physikalisch-Technische Bundesanstalt Mitteilungen **93** 161-167 (1983).
- [22] Cohen, E. Richard, The unification of random and systematic uncertainties, in 1986 Measurement Science Conference, proceedings, Irvine, CA (1986).
- [23] Ayer, A. J., Note (ii) of Two notes on probability, in *The Concept of a Person and other Essays*, Macmillan and Co. Ltd.: London (1963). Reprinted from *Revue Internationale de Philosophie*, No. 58, Fasc. 4 (1961).
- [24] de Finetti, B., La Prevision: ses lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincaré*, Vol. 1, page 1 (1937) (Reprinted in English in reference 30.)
- [25] de Finetti, B. *Teoria delle Probabilita. Sintesi Introduttiva con Appendice Critica*, Turin (1970). (Now available in an English edition.)
- [26] Gillies, D. A., The subjective theory of probability, *Brit. J. for the Philosophy of Science* **23** 138-157 (1972).
- [27] Gillies, D. A., *An Objective Theory of Probability*, Methuen: London, Introduction (1975).
- [28] Savage, L. J., *The Foundations of Statistics*. Wiley: New York (1954).
- [29] Savage, L. J. et al., *The Foundations of Statistical Inference*. Methuen: London (1962).
- [30] *Studies in Subjective Probability*. H. E. Kyburg and M. E. Smokler, Eds., Wiley: New York and London (1964).
- [31] *Bayesian Analysis in Econometrics and Statistics*. A. Zellner, Ed., North-Holland: Amsterdam, New York, Oxford (1980).
- [32] Berger, J. O., *Statistical Decision Theory*. Springer-Verlag: New York, Heidelberg, Berlin (1980).
- [33] Berger, J. O., reference 32: *ibid.*, *passim*, but especially section 4.2.
- [34] Hacking, I., Subjective probability, *Brit. J. for the Philosophy of Science* **16** 334-339 (1966).
- [35] Fisher, Sir Ronald A., reference 1: *ibid.*, pages 43, 44.
- [36] Campion, P. J.; J. E. Burns and A. Williams, Ref. 10: *ibid.*, page 15.

Submicrometer Linewidth Metrology In the Optical Microscope

Volume 92

Number 3

May-June 1987

Diana Nyssonen

CD Metrology, Inc.
Germantown, MD 20874

Robert D. Larrabee

National Bureau of Standards
Gaithersburg, MD 20899

The recent impetus of the semiconductor industry toward submicrometer feature sizes on integrated circuits has generated an immediate need for measurement tools and standards suitable for these features. Optical techniques have the advantages of being nondestructive and of having high throughput, but the disadvantage of using wavelengths comparable to feature size which results in complex scattered fields and image structures that are difficult to interpret. Although submicrometer optical linewidth measurement is possible for 0.3 μm feature sizes, current instrumentation and linewidth standards, particularly for wafers, will have to radically improve

in accuracy as well as in precision to meet the anticipated needs of the integrated circuit (IC) industry for submicrometer dimensional metrology. This paper discusses the effects of inadequate precision and accuracy on process control in IC fabrication and suggests some ways of circumventing these limitations until better instrumentation and standards become available.

Key words: critical dimension measurement; linewidth; micrometrology; microscopy; optical metrology; process-control metrology.

Accepted: January 9, 1987

Introduction

Until relatively recently, optical linewidth measurement systems were the only practical tools for monitoring feature sizes produced by lithographic processes. With the shrinking of feature dimensions to the submicrometer level, and the concern over diffraction and wavelength limitations of optical tools, many fabrication lines jumped to scanning

electron microscope (SEM) measurement tools as the panacea to all of the problems and limitations of existing optical systems. In response, new optical systems have appeared including ultraviolet and laser scanning systems. This paper and an accompanying paper on SEM systems in this issue of the *Journal of Research* [1]¹, assess the capabilities and limitations of each of these technologies and look at how well they will be able to meet the measurement needs of present and future semiconductor processing technologies.

In the optical arena, diffraction effects due to the wavelength of light being comparable to the feature sizes of interest are the major limitation. With the use of shorter ultraviolet wavelengths (e.g., as

About the Authors: Diana Nyssonen joined NBS as a physicist in 1969 and in 1985 left to form her own engineering and R&D company in optical and dimensional metrology, CD Metrology, Inc. Robert D. Larrabee heads the Microelectronics Dimensional Metrology Group, part of the Precision Engineering Division in the NBS Center for Manufacturing Engineering.

¹Figures in brackets indicate literature references.

as low as the 366 nm line of mercury), optical measurements are possible for linewidths down to about $0.3 \mu\text{m}$ (Airy disk diameter is $0.45 \mu\text{m}$ for $f/1$ optics at 366 nm). However, to go to this narrow a linewidth, it is necessary to model the effects of diffraction in the image and develop a meaningful criterion of which point on the image profile corresponds to the edge of the line. This modeling becomes increasingly difficult as the feature height becomes larger than about one-quarter wavelength and as the aspect ratio (feature height/width) approaches and becomes larger than unity. This difficulty is partly mathematical (e.g., the feature cannot be treated as planar using scalar theory and, for small linewidths and large aspect ratios, diffraction effects from adjacent edges interact). The difficulties are also partly due to the fact that the effects of diffraction become more pronounced and propagate further from the edge as the feature height increases and the geometry of the edge departs more from an ideal vertical shape. Indeed, for large aspect ratios and nonvertical walls, the very definition of "linewidth" is open to interpretation.

Definition of Linewidth

With linewidth standards such as the NBS photomask linewidth Standard Reference Materials,

SRMs 474 and 475 [2], the fundamental limitation on the quoted uncertainty in linewidth is not due to the precision (standard deviation, s) of the calibration system, but rather to the definition of linewidth for sloping edges when the slope angle is not under tight control and not easily measured. The current statement of uncertainty accompanying these SRMs is based on the sum of two contributions, one from the nonreproducibility of the measurement system (approximately $0.01 \mu\text{m}$, $3s$) and a larger contribution from a systematic error in edge detection due to the variation in edge slope which occurs during fabrication of the photomasks (see fig. 1a). This latter contribution for this standard (i.e., Δ) is estimated to be $0.05 \mu\text{m}$ and is based on the fact that the NBS photomask calibration system (using transmitted light, broad spectral bandwidth peaked at 530 nm, and coherent edge detection) cannot detect the difference between a vertical edge and a $\Theta=70^\circ$ edge slope. (They both produce the same signal.) Hence, for a 150 nm thick chromium oxide/chromium layer, a 70° edge slope produces a $0.05 \mu\text{m}$ "edge width." Since it is impossible to say what point on this portion of the edge corresponds to the measured linewidth, a systematic error $\pm 0.05 \mu\text{m}$ (worst case) is assigned to the measurement.

In order to reduce this systematic error, the contribution from the uncertain edge position must be

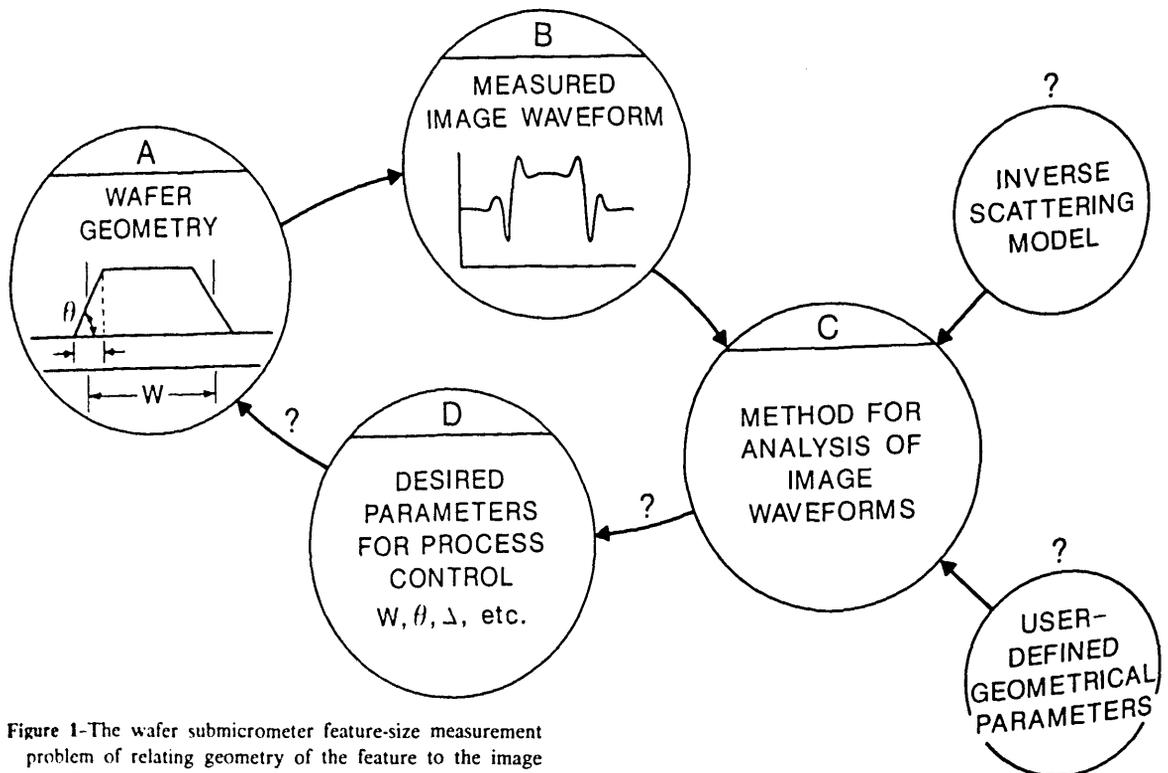


Figure 1-The wafer submicrometer feature-size measurement problem of relating geometry of the feature to the image waveform and developing accurate algorithms for analysis.

reduced. There are two possible ways of achieving this: either the photomask edge slopes must be maintained at angles closer to the vertical (where linewidth is unambiguously defined); or, if the edge slope cannot be adequately controlled, the slope angle, "edge width" or other equivalent parameter must be measured and used to characterize the edge geometry. Currently, the measured quantity of "linewidth" reveals nothing about the true geometry of the line edge in either an optical or SEM measurement. Therefore, the real concern for future technology is whether either SEM or optical technology will be able to go beyond the vague concept of "linewidth" and yield more accurate information about the true edge geometry.

Precision and Accuracy

In metrology [3], *precision* or repeatability is defined as the spread in values associated with repeated measurements on a given sample. That is, the measurement of a given quantity will produce measurements which can be averaged to produce a mean value

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

where x_i is the result of the i -th measurement and n is the total number of measurements. The *precision* or repeatability is characterized by the standard deviation

$$s = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \right]^{1/2} \quad (2)$$

These general formulas assume that n is large and that the errors are random and result in a Gaussian or "normal" distribution centered about the true mean. In many cases, such as length metrology, this may not be true. For example, a common source of error in dimensional measurements is misalignment of the target to be measured to the axis of the measuring instrument which varies randomly with each insertion of the target into the instrument. In this case, misalignment always causes linewidth measurements that are too large.

There is also the question of what is meant by a given standard deviation or statement of precision for a measurement system. For example, a system may be highly repeatable over a period of a few minutes but be extremely temperature dependent

such that temperature fluctuations during the course of the day may produce much larger measurement variations. The first time period refers to short-term precision which is what is usually quoted by the instrument manufacturer, while the latter refers to long-term precision which is heavily dependent not only on the long term stability of the instrument but also on its environment. If the system is expected to hold calibration over periods of weeks or months without recalibration taking only control chart measurements, then the only meaningful precision statement must be based on measurements taken over that same long period of time, i.e., weeks or months.

Accuracy, on the other hand is a more ambiguous and elusive concept. Usually, there is some agreed-upon quantity which one is trying to measure. However, when examined in detail, this quantity and its definition frequently become fuzzy and may escape clear definition.

For example, linewidth on integrated circuit (IC) features seems a clear enough idea until one begins to look at real structures. In figure 2(a), the line has an ideal structure with vertical walls and smooth edges and linewidth can be unambiguously defined as indicated by W . Real structures, like that shown in figure 2(b), do not have well-defined edges. They may have an asymmetric, nonvertical geometry with raggedness along their length. The only meaningful measurement on specimens with ragged edges may be an average along some specified length of the line. In different applications, the basic quantity that is desired to be measured and called linewidth may be different, e.g., the width at the bottom when either etching or doping will be the next process step or the mean width if comparisons with electrical linewidth measurements will be made.

Therefore, a more refined definition of linewidth is needed. For instance, we may agree, as has been proposed, to measure the line structure at some distance above the interface, averaged along a $1 \mu\text{m}$ length as illustrated in figure 2(c). The problem then becomes one of determining how well a given instrument can measure the agreed-upon quantity. If the system measures such a quantity with a systematic error, i.e., it always measures too large or too small, the average error or offset can be determined from measurements on a reference standard with known values. This average error is defined to be the accuracy of the measurement. The ideas of accuracy and precision can be combined [3] into what we here call the uncertainty (see fig. 3):

$$U = E + 3s. \quad (3)$$

Figure 2-Definition of linewidth:

(a) Ideal line geometry with width W unambiguously defined; (b) real line structure showing asymmetric, non-vertical edge geometry and edge raggedness; (c) proposed definition of linewidth as width W defined at some height h above the interface (between the patterned layer and sublayer) and averaged along some length of the line. The height h is selected appropriate to the application, e.g., near the interface for patterned resist.

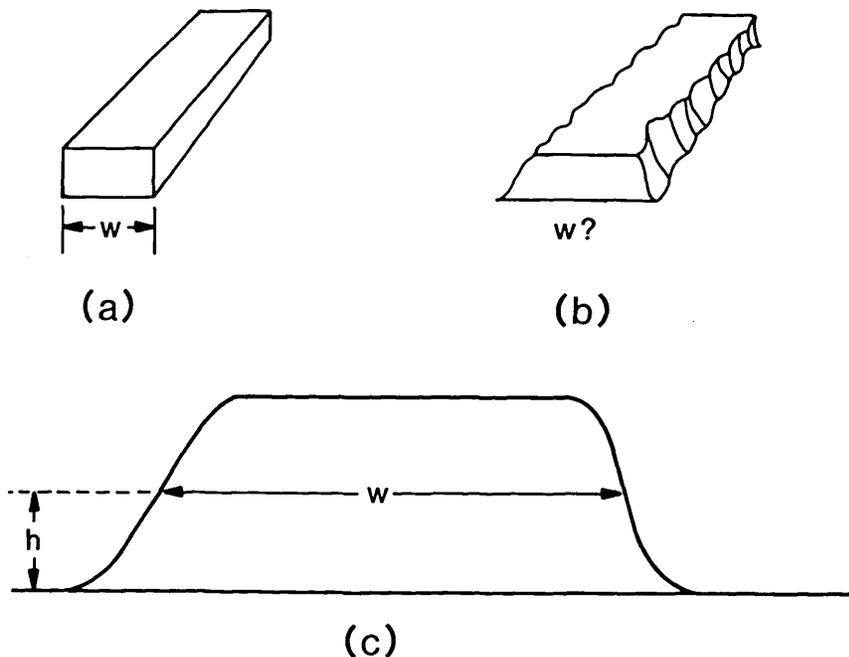
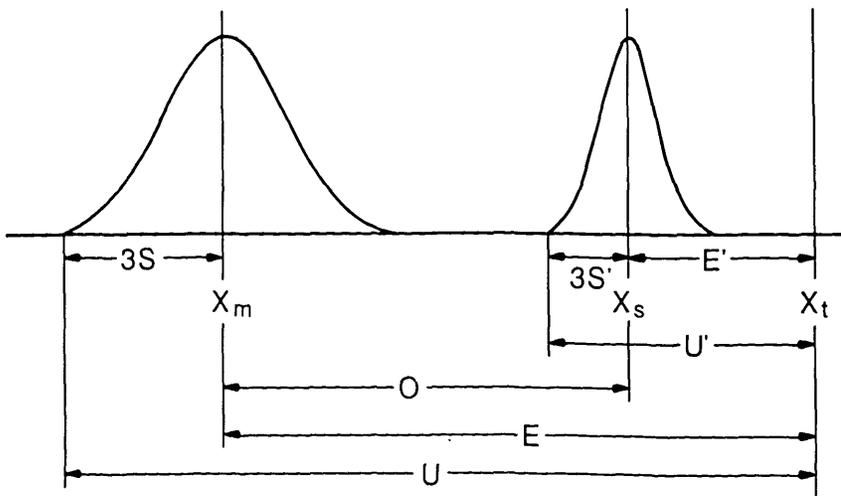


Figure 3-Definition of uncertainty U and standard deviation s .

In this figure: X_t is the "true" value or desired value of the measurement, X_s is the value assigned to the standard with its precision given by $3s'$ and total uncertainty U' , X_m is the result of measurement on another system with precision $3s$. If the measurement offset O is eliminated by correction to the value of the standard X_s , the uncertainty U associated with X_m is still at least $U' + 3s$. Note that X_t is frequently ill-defined and that when the characteristics of the standard used to determine the offset O do not match those of the part to be measured, the uncertainty in X_m may actually be larger than indicated.



When a measurement is given as $x \pm U$, the desired quantity may lie anywhere in the interval defined by $\pm U$. If more measurements are made, they can be averaged and the precision improved in eq (2) by dividing by \sqrt{n} . The accuracy E will remain the same unless a systematic correction to the data can be determined by calibration to a reference standard. Notice that, even when a reference standard is used, E cannot be reduced to zero. The calibration standard has some stated accuracy

and precision associated with it as illustrated in figure 3:

$$U' = E' + 3s'. \quad (4)$$

When measurements are corrected by subtracting the average difference between the known values and results of measurements on the reference standard, the total uncertainty becomes

$$U = E' + 3\sqrt{s^2 + (s')^2}. \quad (5)$$

The rule here is that random errors are added in quadrature, but systematic errors must be added linearly [4]. Even when measurements are not corrected to the reference standard because the measured values lie within the stated uncertainty of the standard, the measurement system cannot be stated to have an uncertainty less than the standard to which it is compared.

Everyone wants accuracy in measurements but, unfortunately, accuracy is only achieved with expenditure of time and effort. A necessary, but not sufficient, condition for accuracy is precision (reproducibility). There are at least four main causes of imprecision in submicrometer optical metrology: 1) variations in the conditions of measurement (e.g., focus), 2) perturbing environmental conditions (e.g., vibration), 3) variations in human judgment (e.g., deciding where the feature edges are located), and 4) variations in the characteristics of the specimens being measured (e.g., thickness of features). Some of these factors can be eliminated (e.g., automation can eliminate some sources of human-induced imprecision) and some can be minimized (e.g., sources of vibration can be identified and remedial measures taken). Some of the remaining (perhaps unknown) sources of imprecision are random (e.g., noise) and thus reducible to acceptable levels by averaging repeated measurements. However, if the remaining errors are not random, (e.g., variations in the image profile resulting from contamination deposited on the surface) no amount of averaging will reduce them! Therefore, a well thought out procedure of measurement based on sound metrological principles can significantly improve precision (e.g., specifying that measurements be taken in the center of the field of view to minimize off-axis aberrations of the optical system). It is not the purpose of this paper to list all possible sources of imprecision or to recommend a universal procedure for obtaining the best precision. However, one purpose is to point out that one very important step toward accuracy is to recognize and control all known or suspected sources of imprecision.

One does not need an official standard to measure instrument precision. A typical specimen of the type to be measured (test wafers or a sample of the product) and known to be stable in time will suffice. One determines long-term precision by repeated (e.g., hourly, daily, weekly, or monthly) measurements on this type of control specimen and then applies well known quality control charting [4] or equivalent procedures to determine control limits and thereby ascertain that the measurement is under control. However, the attainment of the

required degree of long-term precision does not guarantee accuracy. Given precision, there are two main sources of inaccuracy in optical submicrometer metrology; 1) lack of a generally accepted standard of comparison, and 2) improper use of standards. If suitable standards are not available, there are probably good technical reasons for their unavailability, and that reason will probably determine what can, or cannot, be done about it. The temptation is to use the best in-house control specimen as a standard. This may be acceptable as a temporary expedient if done correctly.

Accuracy may be achieved only if the instrument is sufficiently precise and if the specimens of interest exactly match the standard in all important ways except the dimension(s) being measured. However, for linewidth metrology on wafers, one usually cannot guarantee that the specimens to be measured will match the standard in feature height, in substrate properties, in edge geometry and irregularity, in complex index of refraction of the feature material, etc. Recent modeling efforts [5,6] indicate that all these things should be considered to be of prime importance. Clearly, it is inappropriate to use a thin-layer metal-on-glass photomask standard such as NBS 474 or 475 [2] to "calibrate" a system which subsequently will be used on other types of specimens (e.g., photoresist lines on silicon wafers). As mentioned above, it may be appropriate to use in-house control specimens as temporary calibration standards but, if and only if, they closely match the specimens, or range of specimens, to be measured and are known to be stable in time. Some of the material to follow in this paper is intended to be a guide to those factors that must be considered in matching such a standard to the specimens to be measured. This is important because, until accurate measurement systems and standards become available, instrument precision may be the best that can be achieved.

Effect of Measurement Errors On Process Control

It is generally accepted that some linewidth measurements are a necessary form of process control. When accuracy and precision satisfy the gauge-makers rule, i.e., that the measurement system be 3 to 10 times better than the system that generated the part, the argument is incontrovertible. However, when the accuracy and precision of the measurement system are on the same level as the tool being monitored, the situation changes radically. Suppose, as shown in figure 4, the nominal desired

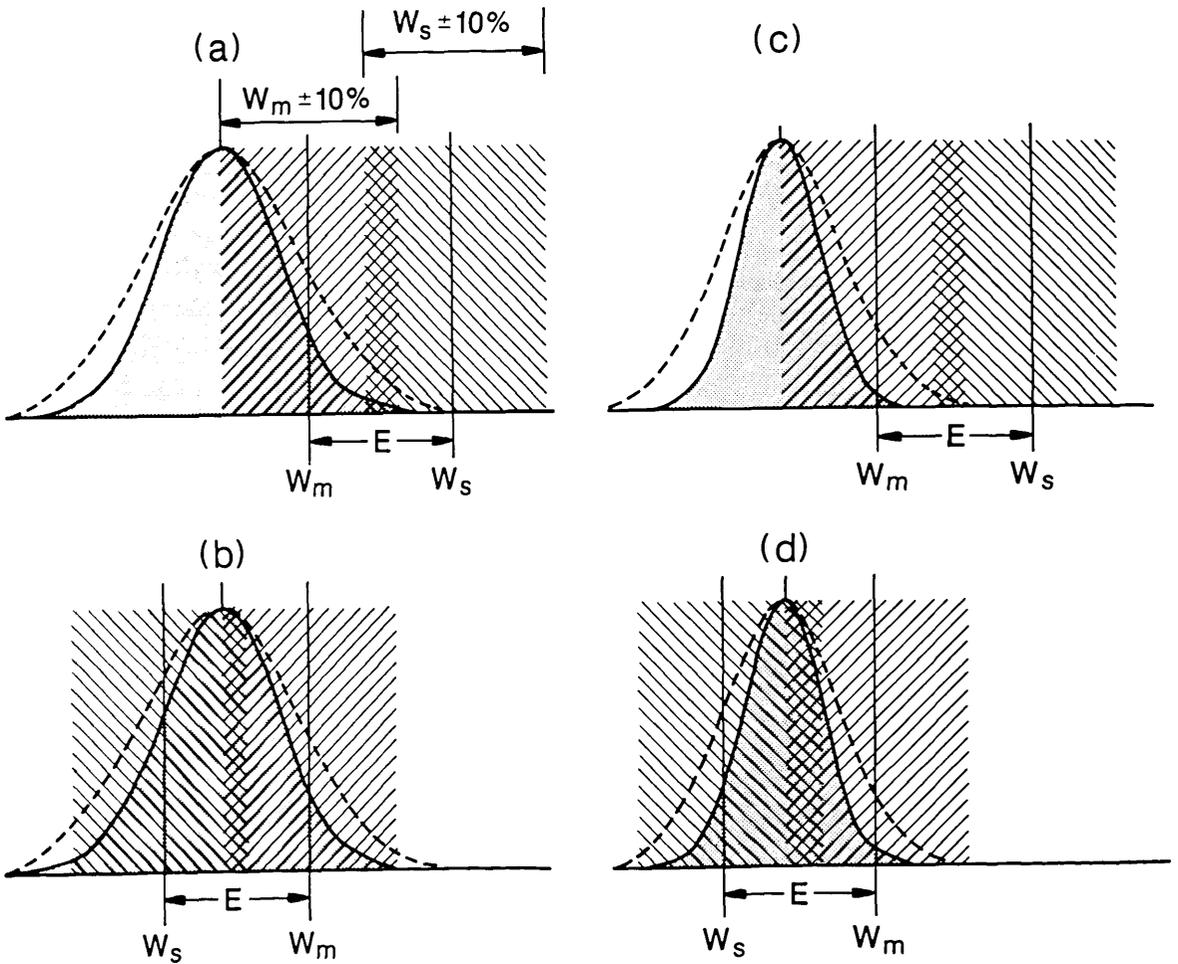


Figure 4—Effect of measurement errors $\pm E$ on acceptance and rejection of measured parts with normal distribution of measured values. Shaded area represents Gaussian distribution of measurements made on a batch of parts. Shaded \square area represents tolerance window for acceptable parts as defined by measurement system. Shaded \square is true tolerance window defined with respect to an accurate standard. Offset is, therefore, the difference between the measured value W_m and that of the standard W_s . When an accurate standard doesn't exist, the value of E is unknown.

linewidth W_s is $1.0 \mu\text{m}$ with a 10% tolerance ($\pm 0.1 \mu\text{m}$) specified. Suppose that a single measurement is made on a sampling of parts and that the resulting parts are found to have widths that are normally distributed centered $0.1 \mu\text{m}$ off the nominal W_m with a spread of $\pm 0.20 \mu\text{m}$ ($3s$, where s is the standard deviation). Assume the long term precision of the measurement system is $0.1 \mu\text{m}$ ($3s$) and that an unknown offset of $\pm 0.15 \mu\text{m}$ exists. The instrument precision when taken into account increases the actual spread in the dimensions of the parts (dashed curve in figure 4) as compared to the standard deviation of the measurements. An offset E typically occurs when the parts being measured differ in form or substance from the standard used for daily calibration of the measurement system. In figure 4a, a negative offset is shown, such

that if 50% of the parts are assumed to be within tolerance, 42% of them are bad and only 8% of them meet specifications. If the part was initially overspecified, the situation may not necessarily raise alarm, if the 50% rejection rate could be tolerated.

In figure 4(b), the offset is of the same magnitude, but positive rather than negative and again, unknown to the process control engineer. If, as before, 50% of the parts are accepted based on belief in the measurements, 32% bad parts will have been accepted and 42% of good parts would have been rejected. In this case, it would have been better to accept all the parts without testing. This would result in no rejected parts (no losses) and the percentage of good parts accepted would have been higher, 59% as compared to 36%.

If the process is in tighter control so that the variation in dimensions as given by $3s$ is only $0.1\ \mu\text{m}$ rather than $0.2\ \mu\text{m}$, the results are even worse. The offset of $-0.15\ \mu\text{m}$, [fig. 4(c)] results in the same acceptance rate (50%). However, none of the parts really meet spec because the unknown offset is larger than the spread in linewidths. Even if the tolerance was overspecified, the situation would not improve greatly; there would still be a large percentage of bad parts accepted. Furthermore, demanding improved precision of the measurement system by itself would not improve the situation.

If, as shown in [fig. 4(d)], the offset is in the opposite direction, tighter process control improves the situation somewhat. However, acceptance of all parts without testing would still result in a higher percentage of good parts accepted, 76% as compared to 52%, and no parts would be lost through rejection. However, since the offset is unknown, the actual situation may be any of the above and the process is clearly not in control.

The conclusion, here, is that the concept of achieving quality by using process control measurements only works when the measurement system has accuracy and precision much better than the variation in the parts being measured (the gauge-makers rule again). This above analysis leads one to the conclusion that at least half of the linewidth measurements currently made at micrometer and smaller dimensions on wafers during manufacture are probably useless if not downright damaging. It becomes obvious, therefore, that to make linewidth measurements an effective process control tool for submicrometer and future technologies, the accuracy as well as precision must be radically improved.

Optical Linewidth Metrology

Many of the potential sources of measurement error mentioned above can be eliminated, circumvented, or minimized by the use of high quality optical and electronic systems in a suitable environment coupled with the use of sound metrological techniques of measurement and data reduction. One of the least understood and most often encountered sources of error is that associated with edge detection, that is, the location of the edge on the image profile of the feature. Ideally, as illustrated in figure 1, one would like to determine the actual geometrical shape of the edge of the feature from its measured optical image profile, decide what point on that shape should be taken as the "edge," and then determine what point on the optical im-

age profile corresponds to this definition of edge. The first of these steps (determination of the actual geometrical shape of the edge) from an appropriate optical signal is the most difficult, and to date, has not been treated adequately for any but the most simple cases (e.g., for photomasks). This is not an exercise of image analysis in computer programming, but a fundamental inverse scattering problem in optical imaging theory [7]. Until this problem is solved under assumptions appropriate for some real instrument, it will be impossible to accurately measure the dimensions of any thick features by optical techniques. As pointed out in the companion paper [1], there is also an analogous problem in scanning electron microscopy that must be solved before feature dimensions can be measured accurately by that technique. This is, in fact, the reason why NBS has not issued linewidth standards for anything but thin layers of metal lines on glass (i.e., photomasks). NBS is working on the problem for both the optical and scanning electron microscope cases, but the magnitude of the problem and the generality of the solution needed (e.g., applicable to a wide variety of structures, feature materials, and measuring systems) will require first the development of practical solutions, and then their application to standards.

Pitch measurements are not particularly sensitive to the accuracy of edge detection because, if two lines have identical geometrical edges and thus identical image profiles, the distance between them can be measured as the distance between corresponding "edges" irrespective of the edge detection criterion used (see fig. 5). For line and space widths, however, any errors in edge detection of left and right feature edges do not cancel by symmetry, but add because of asymmetry, and produce a result with twice the individual edge detection error. Therefore, use of a pitch standard for linewidth measurements will not lead to an adequate calibration for either an optical or an SEM system.

Features with heights larger than approximately $1/4$ the wavelength of light (thick layer) cannot be approximated as thin layers and the image profile of such features will depend on all of the parameters mentioned above. Therefore, it is not sufficient to use a thin-layer (photomask) standard to determine the edge detection criterion for an instrument and then use this criterion for anything but similar thin layers. This is true in spite of the fact that the width of the line on the standard may be known and, perhaps, traceable to NBS. At the present time, there is no way to provide traceability of optical linewidth measurements of any feature dimen-

$$\text{LINEWIDTH: } W_M = (X_2 - \Delta) - (X_1 + \Delta) = X_2 - X_1 - 2\Delta$$

$$\text{PITCH: } P = (X_1 + \Delta) - (X_3 + \Delta) = X_1 - X_3$$

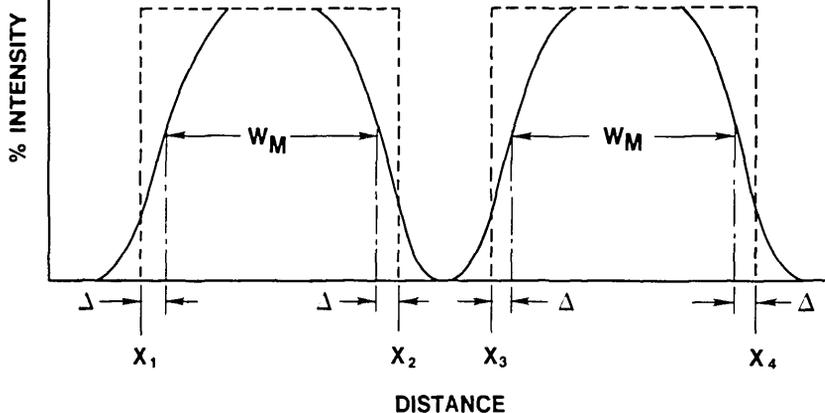


Figure 5—The effect of the edge-detection error (Δ) caused by use of an incorrect threshold on pitch and linewidth measurements.

sions on silicon wafers to NBS. The situation is even worse for two-dimensional features such as vias at micrometer and smaller dimensions, where the specimen cannot be taken as uniform in one dimension and both lateral dimensions must be modeled.

All of these factors combine to produce a situation where it is not possible, at present, to attain accuracy or traceability to NBS in critical dimension measurements for most features of interest to semiconductor device and circuit processing. Precision and a crude assessment of the accuracy may be attainable if the time and effort is taken to do the critical dimension measurement carefully and correctly. The first step in doing this carefully and correctly is to understand the metrologically important factors in the measurement and their reduction to practice in the instrument used for the measurement.

Measurement of Small Feature Dimensions

In order to assess the capabilities of a linewidth measurement system, it is necessary to separately consider two parts of the system: 1) the portion of the system that acquires the signal which is usually called the image profile or waveform, and 2) the edge detection algorithm used to extract a linewidth from this signal. Even when the image profile is not digitized and stored in the system, this profile is the basic signal whose reproducibility must be considered when discussing instrument precision. Many factors effect the reproducibility of this waveform. The factors related to instrument

quality have been previously discussed in the literature including coherence, aberrations, focus, alignment, vibration, etc. [8]. The basic factors to be discussed here are concerned with spectral bandwidth, mode of illumination and mode of collection of the light. The signal waveforms produced are different when the same diffraction-limited high numerical aperture (N.A.) optics are used in imaging systems differing in these respects. Therefore, the edge detection algorithm used must be tailored to the system and include its spectral bandwidth, mode of illumination, and mode of collection.

In addition, the geometrical characteristics of the sample affect this signal waveform. Some of these characteristics are illustrated in figure 6. Figure 6(a) is the image of a polysilicon line with vertical edges. In figure 6(b), the thickness of the patterned layer has changed, in (c) the thickness of the oxide sublayer has changed, and in (d) the edge geometry has changed, all resulting in changes in the image waveform. These figures were computed for the case of a narrow illumination cone angle (0.17 N.A.), $\lambda = 530$ nm, and a 0.85 objective N.A. For larger illuminating cone angles and broader spectral bandwidth, these effects are still present but must be treated by integration over the appropriate cone angle and spectral bandwidth for the instrument. This integration can add considerable complexity to the task of modeling the image waveform for such instruments.

In terms of controlling the precision of the measurement system, a great deal of effort has been spent by instrument manufacturers on focus control and repeatability. With a given sample, it is very apparent that small amounts of defocus

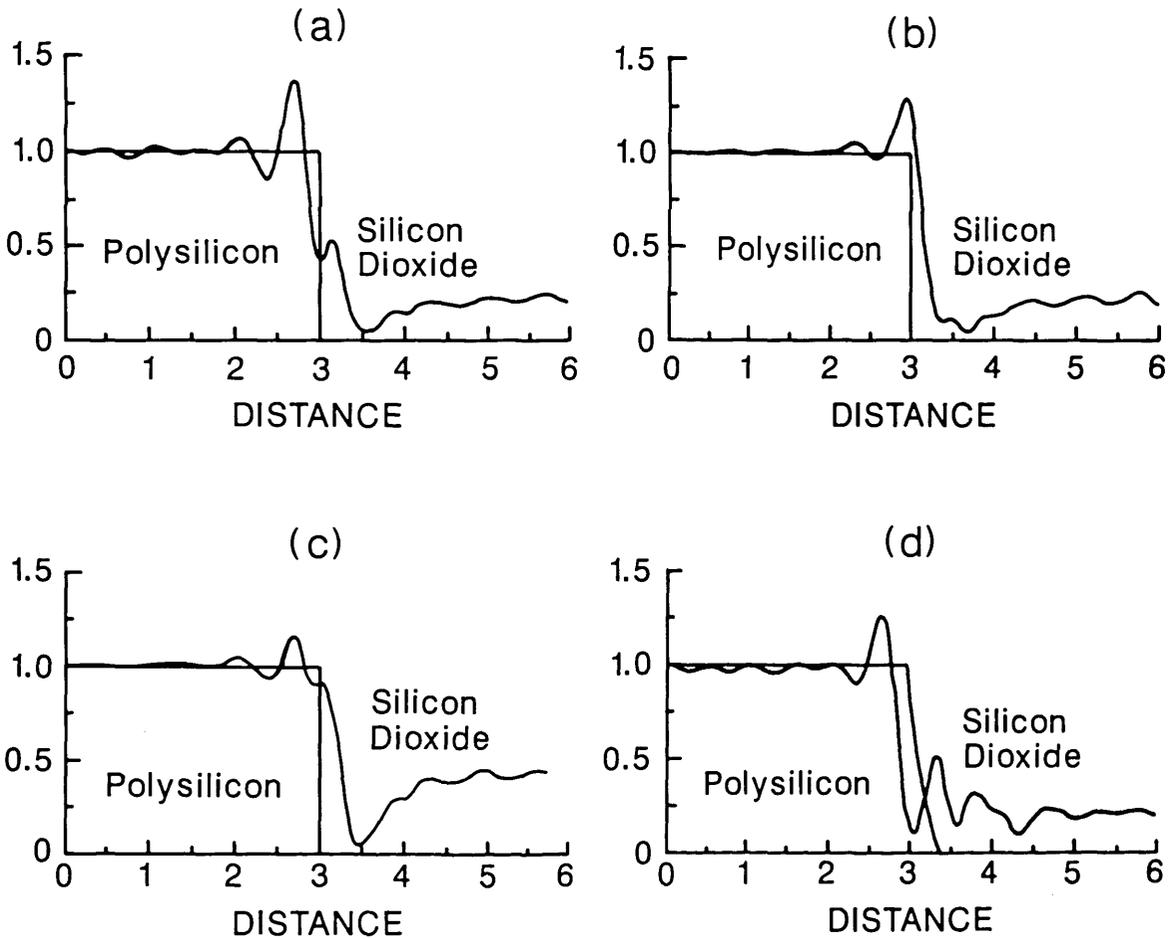


Figure 6—Factors which affect the image waveform: example is calculated for a $0.6\ \mu\text{m}$ thick polysilicon line on a $0.105\ \mu\text{m}$ thick oxide layer on silicon for (a) vertical edges, (b) with change in thickness of polysilicon layer to $0.65\ \mu\text{m}$, (c) with change in oxide thickness to $0.125\ \mu\text{m}$, and (d) with change in edge geometry. Edge geometry is shown superimposed on image profile for reference. The lines are assumed to be symmetric about their centers and only the right half is shown.

change the image waveform and, therefore, introduce an error in linewidth measurement. Similarly, small changes in the sample (e.g., thicknesses or edge geometry) also discernibly effect the image waveform and, unless taken into account by the measurement system, contribute to imprecision and loss of control. This change which results in the image waveform no longer matching that of the standard used for calibration, introduces an unknown offset or error in the measurement. These offsets vary and are estimated to be as large as $0.3\ \mu\text{m}$ or more on processed wafers [9].

The major cause of this offset is the use of a fixed edge detection algorithm, which does not take into account changes in the characteristics of the sample being measured. Instrument manufacturers generally leave the choice of edge detection algorithm to the user with little, if any, guidance as to what is appropriate for a given sample. The edge detection

algorithms available are few, often not based on sound metrological principles, and usually not able to adapt to (or detect) changes in sample geometry, thereby turning linewidth metrology into a poorly practiced black art rather than a science!

Instrument Design

The most important factors which influence the signal waveform are those of the instrument itself including 1) spectral bandwidth, 2) mode of illumination, and 3) mode of collection or imaging of the specimen. Although most optical microscopes are designed for use at specific wavelengths, the quality of the optics is not the primary reason for restricting spectral bandwidth. Virtually all of the materials of concern in IC manufacture are patterned layers and the restriction on spectral band-

width is done for the same reason that a single wavelength source is desirable in ellipsometry: the waveforms of interest vary with wavelength.

Figure 7 illustrates the calculated effect of wavelength on normalized reflectance (inverse of contrast) and phase step at the edge (optical path difference plus phase difference on reflection from thin films) for a $0.6\ \mu\text{m}$ thick patterned dielectric layer. The thicker the layer, the more rapidly the image waveform changes with wavelength. Use of broad spectral bandwidth integrates the response over the bandwidth of the system resulting in a loss of sensitivity and edge acuity.

The variation of spectral response is obvious to one who has observed the rich color variations of processed wafers or has used a color chart to determine oxide thickness. The effect of angle of illumination can similarly be observed by tilting the wafer and noting the change in color with viewing angle. Wavelengths that have a high reflectance at one angle of incidence will have a lower reflectance at another angle allowing a different wavelength to determine the observed color. Therefore, it would not be surprising to see a parallel beam of laser light cause the contrast of the patterned wafer to change with variation in the angle of incidence (see fig. 8). What is perhaps less obvious, is that a

focused laser beam has the same general effect to that of broad spectral bandwidth. The focused laser beam may be viewed as the sum of plane (collimated) waves corresponding to each differential element of solid angle within the cone angle of the illuminating lens [10], with each angle producing a different image waveform. Again the effect of integrating (here over the cone angle of the lens) is one of poorer edge acuity and loss of sensitivity to edge geometry resulting in a larger uncertainty in the measurement. That is, the system may produce the same signal waveform and linewidth measurement for a range of object thicknesses, edge geometry and geometrical widths. This could result in an apparent increase in precision (i.e., a decrease in the standard deviation) but, because of the insensitivity to geometry, the uncertainty in the measurement increases, and the ability of the system to detect changes affecting device performance may be lost.

The mode of collection of the signal energy coupled with the mode of illumination determines the resolution and coherence properties of the system which are separate from the effects discussed above. The most commonly used configurations for illumination and collection are illustrated in figure 9, i.e., bright-field, (a) and (b), focused-beam [11], (c), and confocal [12], (d), microscopes. All of

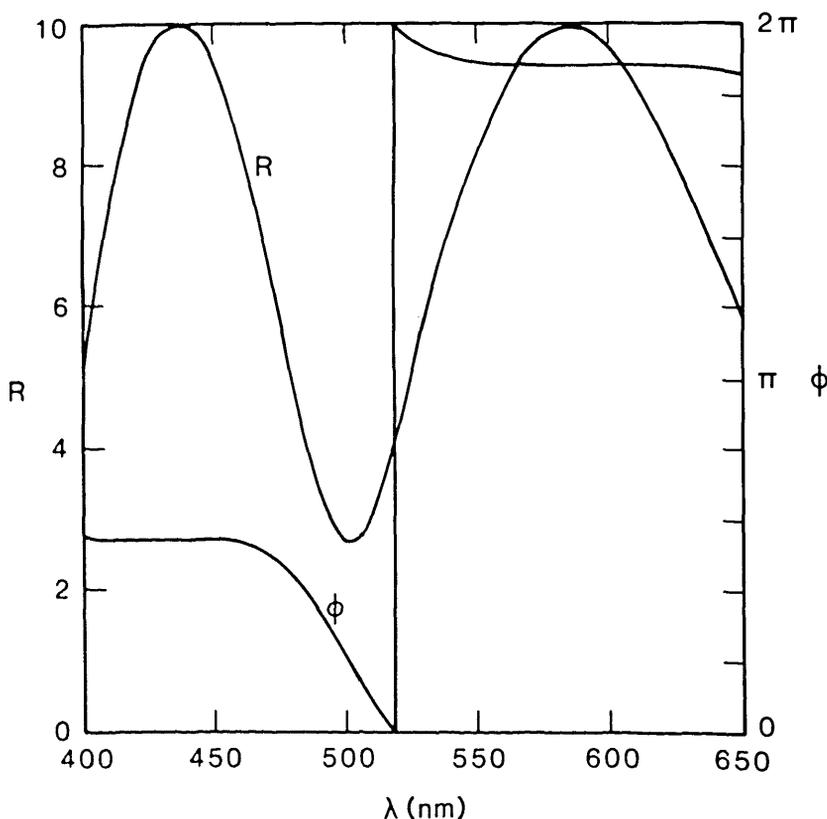


Figure 7—Relative reflectance R and phase differences ϕ for a $0.6\ \mu\text{m}$ thick layer of silicon dioxide on silicon calculated from the Fresnel equations for varying wavelength. Curves are normalized with respect to the R and ϕ parameters of the bare silicon substrate.

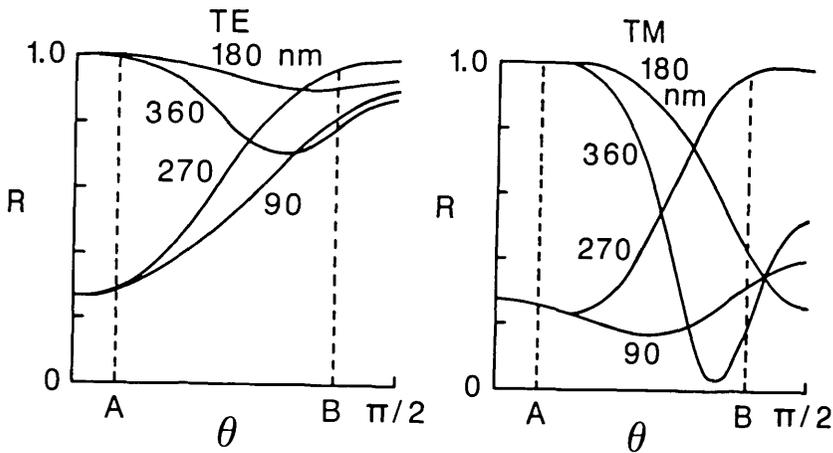


Figure 8—Variation of relative reflectance R as a function of angle of incidence θ for SiO_2 patterned layers of varying thickness (on silicon), $\lambda = 530$ nm. Dashed lines correspond to a cone angle of (A) 0.22 N.A. and (B) 0.95 N.A. Curves are shown for both transverse electric (TE) and transverse magnetic (TM) directions of polarization.

these systems are partially coherent or effectively coherent imaging systems and are, therefore, sensitive to uniformity of phase as well as intensity across the field of view. Nonuniform phase is one source of asymmetry in the image profile. Accuracy of alignment and optical quality of the illuminating system, therefore, become more demanding than for a conventional microscope imaging system (less coherent) with the most stringent demands made by the single wavelength, narrow angle of incidence systems.

For planar objects ($< \lambda/4$ thick), these four configurations of systems would be expected to produce similar image waveforms for the same numerical apertures. However, for patterned thick layer materials such as those found in IC manufacture, they do not. Characteristic image waveforms, expected for these systems are shown in figure 10. The choice of the type of image waveform becomes important when discussing edge detection algorithms for linewidth measurement. Each of these signals represents a different response to an edge discontinuity in the same material. Therefore, accurate edge detection algorithms will be different for each of these systems. To date, the only system that has been well-characterized and for which any accurate edge detection algorithms exist, is the effectively coherent (narrow illuminating cone angle) version of the bright-field microscope [fig. 10(b)] [6]. It has been shown that, as is the case with ellipsometry, it is much easier to analyze the system for single wavelength and single angle of incidence, and easier to develop accurate measurement algorithms. As in the case with ellipsometry [13], there is additional information to be gained from other angles of incidence and other wavelengths. However, integrating over a broad spectral bandwidth or a wide cone angle is not the best method to extract that additional information.

Sensitivity, Ease of Operation, And Resulting Uncertainty In the Measurement

There are three major reasons (not necessarily advantages) for using broad spectral bandwidth and a large illuminating cone angle: 1) there is an increase in available energy and improved signal-to-noise ratio (S/N), 2) the resulting waveforms have simpler structure, and 3) there is less sensitivity to system alignment and sample differences.

Linewidth instrument manufacturers have always preferred white light sources or focused laser beam systems (lower power requirements) because of improved signal-to-noise and lower cost. In addition, manufacturing tolerances, particularly alignment of microscope parts, is less demanding. That is, integrating over a large cone angle makes the system less sensitive to both variations in sample response due to angle of incidence and errors in alignment of the optical system parts.

Recently, with the move toward automated systems and automated signal processing, the argument has been put forth that coherent (narrow angle of incidence) image waveforms are "too complex" for automated signal processing. However, the simpler waveforms (more nearly monotonic) are gained only with an accompanying loss of accuracy and sensitivity, and larger measurement uncertainties. An analogous problem exists in electron-beam lithography. With a relatively large beam diameter, the lithography system is less sensitive to an array of problems including beam stability, beam cross-section variation, vibration, positional errors, proximity effects, etc. These problems become more apparent as the beam diameter and least-countable address are reduced. Yet, no one doubts that the smaller beam produces

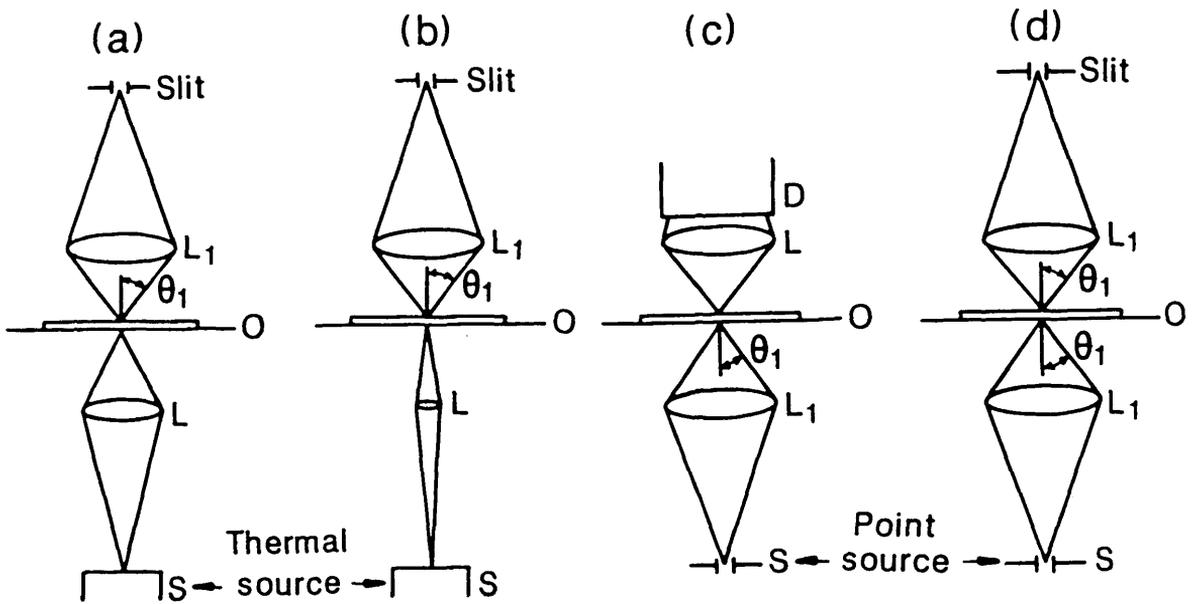


Figure 9—Basic optical system designs used for feature-size measurement: (a) conventional bright-field (partially coherent) with broad spectral bandwidth, (b) narrow cone-angle, bright-field (effectively coherent with single wavelength laser source), (c) focused laser beam, and (d) confocal microscopes. The systems are shown in transmission (unfolded for reflection) and in each case, the source S illuminates the line object O through a lens L. A laser source is considered to be a point source located infinitely far away from the lens L₁ in (c) and (d). The scattered light is collected through a second lens and images onto the detector. In (a), (b) and (d) the slit at the image plane is unresolved when projected back onto the object plane. In (c), the detector D collects all of the light; it, therefore, may be placed in either the image plane, at the lens as shown, or the lens L may be eliminated altogether. Although the schematics are shown for critical illumination, Kohler illumination may be used in (a) and (b) without changing the system response.

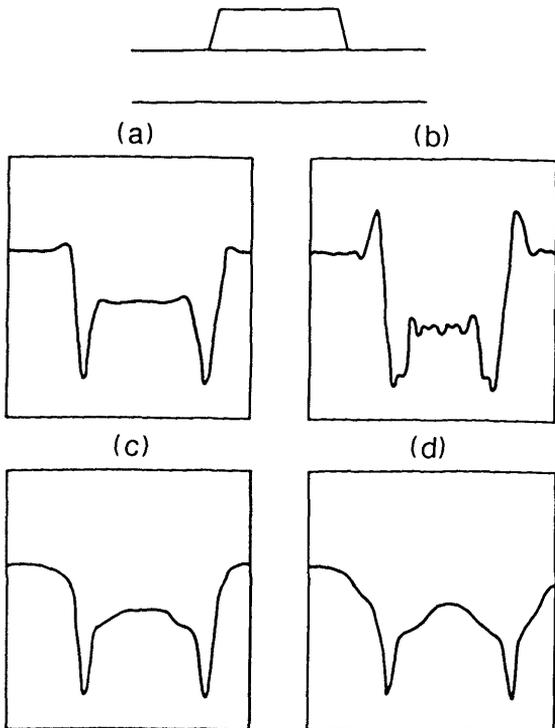


Figure 10—Characteristic image waveforms for line object shown at top corresponding to systems in figure 9; (a) through (d) are same as in figure 9.

smaller, more accurate pattern definition. It is easy to see that the less sensitive system with the larger beam diameter is coupled to a loss of accuracy and resolution. The same is true in linewidth measurement with respect to angle of incidence and spectral bandwidth.

Single wavelength, narrow cone angle, reflected-light optical systems are also more sensitive to surface contamination than systems not having these features. Although the optical imaging mechanism is different from that of an SEM system, and optical systems inherently do not deposit contamination on the surface, the problem of contamination is potentially as serious for optical systems as it is for an SEM in terms of image profile distortion. Surface contamination can result from residues of fabrication processes and airborne particles or improper handling and storage. This is one of the reasons that the use of the less sensitive transmitted-light system is recommended for measurement of photomasks [14].

Resolution

The differences between the optical configurations shown in figure 9 with respect to resolution

are small at the high N.A.'s used for linewidth measurement at micrometer and submicrometer dimensions. Bright-field and focused beam systems have the same response for the same N.A.'s and equivalent coherence parameters (ratio of condenser to objective N.A. for bright-field; ratio of collecting to illuminating N.A. for focused-beam systems [15]). Confocal microscope systems which have double the resolution at low N.A.'s show only a slight improvement in resolution at high N.A.'s [16]. There is greater potential for improvement in resolution to be gained by using shorter wavelengths. This is due to several factors: 1) the nonapplicability of small angle approximations ($\sin \Theta \approx \tan \Theta$) at high N.A.'s, and 2) the loss in diffracted energy at high angles of incidence for line objects that are thick compared to $\lambda/4$. The chief advantage of the confocal system is its $(\sin x/x)^4$ impulse response (1-D) rather than the conventional $(\sin x/x)^2$ [12].

Comparable edge profiles are shown in figure 11 for a planar (i.e., thin) object. The fourth power function reduces the magnitude of the coherent edge ringing while still producing the minimum or dark interference band at the line edge. This type of response produces signals with less detailed structure. At the present time, however, insufficient analysis has been done to produce accurate edge detection algorithms for either focused-beam or confocal microscope systems due to the required integration over the angle of incidence for thick-layer line objects discussed above.

Waveform Analysis—Edge Detection

Submicrometer lithography puts stringent requirements on the reproducibility of the image waveform and the accuracy of subsequent analysis of the waveform for linewidth measurement. High precision can only be achieved by controlling those factors which affect the image waveform including focus, etc. Noise also affects precision and accuracy; the two most significant sources being photon noise due to inadequate illumination levels for a given detector and vibration. As has been shown in lithography and in the SEM [1] vibration increases the apparent line dimension.

Smoothing is a frequently taken alternative to reduction of photon noise by increased source output or elimination of vibration by use of isolation systems. With "white" noise, excessive smoothing (over distances greater than the desired precision and accuracy of the measurement) results in loss of sensitivity, and changes in linewidth dimensions,

the equivalent of using a larger beam size in electron-beam lithography. When noise sources such as vibration have characteristic frequencies and are not "white," the effect of excessive smoothing is signal distortion with an accompanying loss of accuracy and precision. The best method of improvement in accuracy and precision at the nanometer level is achieved by use of brighter light sources and better vibration isolation.

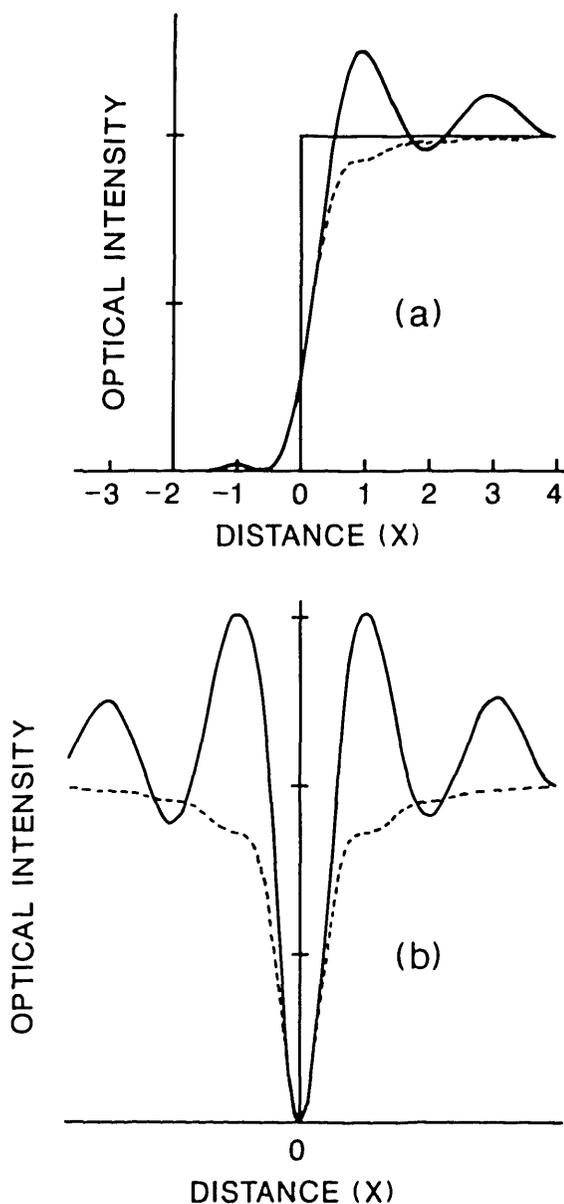


Figure 11—Comparison of calculated edge profiles for coherent bright-field (solid line) vs. confocal microscope (dashed line) for planar line object: (a) high contrast (opaque) with no phase discontinuity at the edge; (b) low contrast, π -phase discontinuity.

Standards

The only national or international standards currently available for linewidth measurement are photomask standards whose dimensions will shortly be extended by NBS down to $0.5\ \mu\text{m}$. Methods of reducing the present $0.05\ \mu\text{m}$ uncertainty of the NBS photomask standard by taking into account the variable edge geometry due to process variations are currently being considered at NBS. No optical or SEM linewidth standards currently exist for features on silicon wafers. This section would not be complete without some discussion of what can be done to improve silicon-wafer process control in the absence of traceability to national or international standards.

The methods of improperly applying standards to process control most often seen on IC fabrication lines include: 1) use of photomask standards for calibration of systems used to measure wafers, 2) measuring a single in-house specimen in an SEM by conventional techniques and, assuming that the results are representative of a given process step, subsequently adjusting all measurements (by addition of a "fudge factor"), or 3) measuring a single in-house standard in cross section in an SEM and similarly adjusting subsequent measurements. Each of these methods will introduce some level of variable and unknown offsets in subsequent product measurements.

The worst method is the use of a photomask standard for other than pitch or line scale calibration. In addition to poor signal (or visibility) when viewed in reflected light, the chief problem is that the image profile (except in rare cases) does not match that of the wafer being measured. Therefore, any edge detection threshold or other criterion based on the certified photomask linewidths is guaranteed to be in error by an unknown amount. These errors may be as large as $0.5\ \mu\text{m}$. The second choice, the use of an in-house standard measured by conventional SEM techniques also has problems. It is usually assumed that, regardless of the respective measurement and edge detection techniques used, all of the difference between the optical and SEM measurements is due to error in the optical measurement. This is an unwarranted assumption which is discussed in the accompanying paper [1]. Some error is associated with both the SEM measurement W_e and the optical measurement W_o . At the present time, there is no technically sound way of apportioning the difference between SEM and optical measurement errors. It is also possible as illustrated in figure 12 that both are

in error in the same direction so that the error in the optical measurement is larger (or smaller) than the measured difference, D .

Measurements using SEM-viewed cross sections of the lines on product wafers, while likely to reduce some of the SEM error (principally that due to interaction with the substrate and shadowing effects), do not eliminate the SEM contribution entirely and again leave the process control engineer unable to assess the true magnitude of the optical and SEM errors. Thus, if the SEM measurement is assumed arbitrarily to be accurate, some unknown offset will still be present.

In addition, all of these methods have another problem in common resulting from use of a single product sample. Both SEM and optical measurements (to different degrees) are sensitive to characteristics of the specimen such as layer thickness, edge geometry, and contamination. These variables cause changes in the image waveform resulting in a variable error or offset whenever a fixed threshold (or other edge detection criteria) is used which does not take into account changes in the material or geometry. Currently, SEM and optical edge detection criteria used are unable to adapt to changes in the image profile by appropriate corrections to the edge detection criteria used.

Thus, the examples shown in figure 4 represent actual situations which might arise in process control situations given the present state-of-the-art measurement systems and standards. There are, however, several things that the process control engineer can do to improve the situation. First, the measurement system must be under the best control possible and its long-term precision established by accepted control chart techniques. The system should be calibrated to a pitch or magnification standard. This standard need not match the material characteristics of the product to be measured since pitch or line-scale is not sensitive to edge detection errors as long as the line geometry is symmetric. Next, a test pattern or sample, characteristic of the product, can be used to determine precision for the range of linewidths of interest. This de facto product standard should be measured initially to form a data base and then repeated measurements made over many days or longer to establish the long-term reproducibility [4].

Once satisfactory precision is established, there are two remaining concerns that must be addressed: sensitivity of the instrument to changes in geometry of the lines that might effect product performance especially at submicrometer dimensions, and the relationship of the measured linewidths to

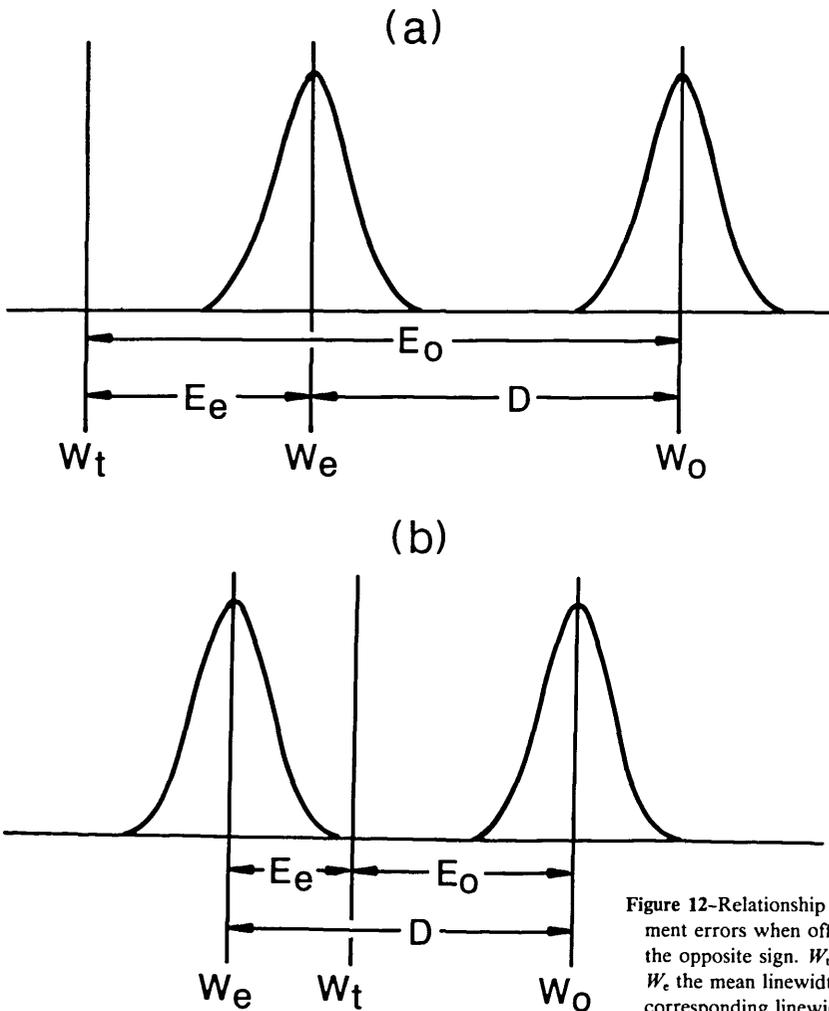


Figure 12-Relationship of optical (E_o) and SEM (E_e) measurement errors when offsets are (a) of the same sign, and (b) of the opposite sign. W_t is the desired (user-defined) linewidth; W_e the mean linewidth as measured in the SEM; and W_o the corresponding linewidth measured optically.

product performance to establish an acceptable window for process control since the measured linewidths (because of unknown offsets) cannot be assumed to relate directly to engineering tolerances.

Sensitivity of the measurement system can be determined by examining good and bad product by other methods such as ellipsometry, profilometry, and SEM inspection in cross-section (after coating or other means to eliminate charging). For example, do two resist lines with different edge slopes, (which, as determined from SEM inspection, should show differing line widths) actually show differences in linewidth in the optical measurement system? and, are such differences proportional to the differences seen in the SEM? Since neither system can be assumed to be accurate per se, the only concern here is whether the measurement system has the sensitivity to distinguish good from bad product. This should be determined by correlating

data from samples of the product showing thickness and edge geometry differences representative of what is expected in a production run.

The most difficult step is establishing the window of acceptable linewidths for a given process step and measurement instrument. The idea is to correlate measured linewidths on both good and bad product (which has been determined to have failed because linewidths are out of specification) with some performance characteristic. The innovative techniques necessary here are based on knowledge of the particular device and fabrication process. For example, where linewidth of the fabricated feature can be correlated to the operational speed of the device, electrical data on the completed devices can be used to define an acceptable range of measured linewidths. Similarly, for diffusion lines, electrical data on linewidth test patterns can be correlated with optically measured linewidths. Establishing an acceptable window for

an arbitrary resist patterning step in the fabrication process presents the most complex problem. The optical linewidth measurement system is very sensitive to thickness and edge geometry changes in thick dielectric layers such as resist. The desired bottom width of the resist line is one of the most difficult to measure and thus to correlate with device performance. The acceptability of the resist profile must be determined from the acceptability of the resulting patterned layer with a window allowed for variability in the patterning process.

Although the approach described above is more demanding, it is likely to yield more satisfactory results than blind faith in the linewidth values produced by any one state-of-the-art optical or SEM linewidth measurement system when the gauge-makers rule is not met by the measurement system. If the measurement results at some stage of processing can be shown to be a valid predictor of yield, then the need for accuracy is somewhat circumvented. This approach requires the continued processing of measured specimens, and the tracking of specific specimens as they are subsequently processed and ultimately tested. However, this approach is not a substitute for accuracy because its success depends on unknown and, perhaps, uncontrolled factors besides the measured parameter(s) that affect yield. These factors lower the correlation, diminish the value of the critical dimension measurement and, if they get out of control, can dominate the yield and destroy the previously determined correlation. It is, however, something that should be done even after accuracy is achieved to validate the importance and justify the cost of the critical-dimension measurement in question.

A major problem still remains, it does little good to generate a standard with a very small uncertainty if the measurement system does not satisfy the gauge-makers rule: the accuracy and precision associated with *both* the standard and the measurement system must be 3 to 10 times better than the variations produced by the lithography tool which generated the wafer for the most effective process control. Better measurement tools and standards are, therefore, needed for submicrometer lithography.

Alternative Linewidth Measurement Techniques

There are several alternative linewidth measuring techniques that have been suggested and, in some cases implemented, that overcome one or more of the disadvantages of the optical imaging

techniques discussed above. Some, like the scanning electron microscope [1], use a different form of "illumination." Others, like the scanning tunneling microscope [17] directly probe the feature surface topography and produce a profile. Still others, like electrical test patterns [18], do not produce images or profiles, but directly measure an average linewidth.

Although the SEM potentially has better resolution in terms of beam characteristics, the complex interaction of the electron beam with the specimen [19] currently limits the accuracy and precision available in feature-size measurement. This and other problems associated with the use of SEMs for linewidth measurement are discussed in the accompanying paper by Postek and Joy [1].

Electrical techniques based on test patterns [18] have the advantage of simplicity for both the measurement system and interpretation of the data. However, the test patterns require significant area on the integrated circuit and it is usually not possible to measure the actual lines of interest in the circuit. In addition, only conductive lines can be measured. However, the measurement is self calibrating, fast, simple to understand, and implementable with standard testing hardware. Clearly, the electrical test pattern approach has something to offer and will find its niche in semiconductor processing. Newer electrical techniques such as profiling by use of the tunnel effect [17], have potential in some applications and, at the present time, are being researched for dimensional and other applications.

Optical approaches inherently have some distinct advantages over alternatives (e.g., they are nondestructive and applicable to all materials regardless of their electrical conductivity). Because of this, there is a continual search for new ways to exploit optics to circumvent or eliminate the problems created by the relatively long wavelength of visible light. Use of shorter wavelengths is based on sound principles, but is currently limited by the availability of good quality optical elements at the shorter ultraviolet and soft x-ray wavelengths [20]. Scanning aperture (or near-field microscopy) [21,22] appears to have the advantage of circumventing the diffraction limitations of dry optics, thereby, providing greater resolution over conventional microscopy. These systems are difficult to implement and not well understood, nor have they been analyzed sufficiently for use in metrological applications.

A number of modifications of more conventional optical microscopy have been proposed (e.g., confocal microscopy [12] and phase-measuring systems

[16]) but, these techniques do not circumvent diffraction effects and, presently have no metrologically sound criteria of edge detection. The authors' believe that, as is the case with optical versus electron-beam lithography, the inherent advantages of optics will assure its niche in submicrometer critical dimension metrology for some time to come. However, that niche can only be filled by methods that are backed by sound theoretical analysis, predictions that agree with experiment, and meaningful edge detection criteria. Current research has shown that single wavelength, narrow angle of incidence microscopy is extremely sensitive to edge geometry and that, through inverse scattering, the possibility exists of extracting line geometry parameters from an optical signal. However, the development of such schemes and their analysis will take an investment in time and resources, and that is unfortunate because the rapid progress to submicrometer dimensions made by the semiconductor industry in recent years has led to needs for submicrometer metrology today. Perhaps this is the price that must be paid by an industry that tends to take metrology for granted and, in the past, has not supported metrological research and development to the extent needed to meet its future demands.

Conclusions

Accurate measurement of submicrometer feature sizes on integrated circuits is a problem of primary importance to the semiconductor industry and one that is not likely to have an effective and efficient solution in the near future. Although optical techniques offer the advantages of nondestructive testing and relative simplicity of use coupled with high throughput, they presently are incapable of the needed precision (reproducibility), and accuracy for any but the simplest of specimens (i.e., photomasks). Suitable optical systems and associated edge-detection criteria will be developed and applied to integrated-circuit features. But until then there will be no acceptable linewidth standards for silicon wafers, and there will be no universally accepted accuracy in linewidth measurements on these wafers. With the development of suitable edge-detection criteria and the use of ultraviolet (or shorter) wavelengths, most of the submicrometer linewidth region above $0.3 \mu\text{m}$ may be measurable optically. However, for the present, the semiconductor industry will, of necessity, have to use in-house standards for instrument set-up, maintenance, and quality control to gain reproducibility.

The best that can be done under such circumstances is a crude assessment of accuracy based on the most accurate alternative measurements available. This unfortunate situation is due, in part, to the rapid progress of the industry in achieving ever smaller feature sizes—that progress has been faster than the developments in dimensional metrology needed to keep pace with it.

The authors wish to thank Dr. Michael Postek for reviewing the manuscript and for his helpful suggestions.

References

- [1] Postek, M. T., and D. C. Joy, Microelectronics dimensional metrology in the scanning electron microscope, *J. Res. Natl. Bur. Stand.* **92** 3 (1987).
- [2] National Bureau of Standards SRM 475, available from the Office of Standard Reference Materials, Room B311, Chemistry Building, Gaithersburg, MD 20899.
- [3] ASTM Manual on Presentation of Data and Control Chart Analysis, STP15D, American Society for Testing and Materials, 1916 Race Street, Philadelphia, PA 19103.
- [4] Croarkin, C., and R. N. Varner, Measurement Assurance for Dimensional Measurements on Integrated-Circuit Photomasks, NBS Tech. Note 1164 (August 1982).
- [5] Nyyssonen, D., and C. P. Kirk, Modeling the Optical Microscope Images of Thick Layers for the Purpose of Linewidth Measurement, *Proc. SPIE Vol. 538, Optical Microlithography IV*, 179-187 (1985).
- [6] Nyyssonen, D., Theory of optical edge detection and imaging of thick layers, *J. Opt. Soc. Am.* **72** 1425-1436 (1982).
- [7] See, for example, feature issue on Inverse problems in propagation and scattering, *J. Opt. Soc. Am.* **11** 1901 ff (1985).
- [8] Bullis, W. M., and D. Nyyssonen, Optical linewidth measurements on photomasks and wafers, Chapter 7 in *VLSI Electronics: Microstructure Science Vol. 3* (January 1982).
- [9] Grant, P., and T. Fahey, Optical measurement of uncertainties and process tolerances, *SPIE* **480** 30-32 (1984).
- [10] Richards, B., and E. Wolf, Electromagnetic Diffraction in Optical Systems, II. Structure of the Image Field in an Aplanatic System, *Proc. Roy. Soc. (London)* **A253** 358-379 (1959).
- [11] Wilke, V., Optical scanning microscopy—the laser scan microscope, *Scanning* **7** 88-96 (1985).
- [12] Sheppard, C. J. R., and A. Choudhury, Image formation in the scanning microscope, *Optica Acta* **24** 1051-1073 (1977).
- [13] Humlicek, J., Sensitivity extreme in multiple-angle ellipsometry, *J. Opt. Soc. Am. A* **2** 713-722 (1985).
- [14] Nyyssonen, D., Linewidth Calibration for Bright-Chromium Photomasks, NBSIR 86-3357 (May 1986).
- [15] Nyyssonen, D., Focused-beam vs. conventional bright-field microscopy for integrated circuit metrology, *Proc. SPIE 565 Micron and Submicron Integrated Circuit Metrology* (1985).

- [16] Hobbs, P. C. D.; R. L. Jungerman and G. S. Kino, A phase sensitive scanning optical microscope, *Proc. SPIE* **565** 71-80 (1985).
- [17] Baratoff, A., Theory of scanning tunneling microscopy-methods and approximations, *Physica* **127B** 143-150 (1984).
- [18] Yen, D.; L. W. Linholm and M. G. Buehler, A cross-bridge test structure for evaluating the linewidth uniformity of an integrated circuit lithography system, *J. Electrochem. Soc.* **129** 2313-2318 (1982).
- [19] Newbury, D. E.; D. C. Joy, et al., *Advanced Scanning Electron Microscopy and X-ray Microanalysis*, Plenum Press: New York, pp. 3-38 (1986).
- [20] Howells, M.; J. Kirz, et al., Soft x-ray microscopes, *Physics Today* **38** No. 8 22-32 (Aug., 1985).
- [21] Lewis, A.; M. Isaacson, et al., Development of a 500 Å^o spatial resolution light microscope, *Ultramicroscopy* **13** 227 (1984).
- [22] Pohl, D. W.; W. Denk and M. Lanz, Optical stethoscopy: image recording with resolution $\lambda/20$, *App. Phys. Lett.* **44** 651 (1984).

Submicrometer Microelectronics Dimensional Metrology: Scanning Electron Microscopy

Volume 92

Number 3

May-June 1987

Michael T. Postek

National Bureau of Standards
Gaithersburg, MD 20899

David C. Joy

AT&T Bell Laboratories
Murray Hill, NJ 07974

The increasing integration of microelectronics into the submicrometer region for VHSIC and VLSI applications necessitates the examination of these structures both for linewidth measurement and defect inspection by systems other than the optical microscope. The low beam-voltage scanning electron microscope has been recently employed in this work due to its potentially high spatial resolution and large depth of field. This paper discusses applications of the scanning electron microscope to microelectron-

ics inspection and metrology in light of the present instrument specifications and capabilities, and relates the scanning electron microscope to the controls required for submicrometer processing.

Key words: critical dimension; dimension; linewidth; measurements; metrology; micrometer; SEM; standards; submicrometer.

Accepted: January 9, 1987

Introduction

The scanning electron microscope (SEM) has become an important tool in the inspection and measurement of microelectronics for the Very Large Scale Integration (VLSI) and Very High Speed Integrated Circuit (VHSIC) programs. As the feature dimensions on integrated circuits reach into the submicrometer region (fig. 1), inspection techniques using scanning electron microscopes are

becoming commonplace. Many processing facilities are presently working at a 10%, or even 5% tolerance, in order to produce the precise structures needed for submicrometer circuits. The effect on the process precision of the linewidth measurement is shown in table 1. Application of the "gauge maker's rule" to the necessary tolerances means that soon the goal for process precision will be in the nanometer range. Even though optical microscopes can be useful for critical linewidth measurement and inspection to about 0.3 μm [1],¹ many fabrication lines, in anticipation of future needs, are integrating SEMs into the production sequence at chip levels of 1.25 μm geometry and smaller (table 2). Advanced scanning electron beam instruments are presently being developed to facilitate this work and to do automated linewidth measurement and inspection [2-4].

About the Authors: Michael T. Postek is a project leader in SEM metrology in the Microelectronics Dimensional Metrology Group, a part of the Precision Engineering Division in the NBS Center for Manufacturing Engineering. David C. Joy is involved in the application of electron beam instruments to the characterization of semiconductor materials and devices and the development of the theory of electron-solid interactions. His current address: EM Facility, Walters Life Sciences Building, University of Tennessee, Knoxville, TN 37996.

¹Figures in brackets indicate literature references.

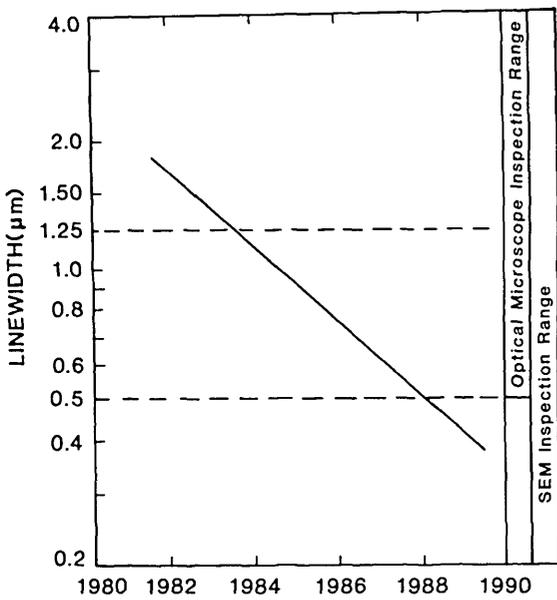


Figure 1—Projected decrease in the size of the linewidth of VHSIC and VLSI circuits through the 1980s and the relationship to optical and scanning electron microscope inspection instrumentation.

Use of the scanning electron microscope for semiconductor device inspection has several advantages over optical microscopy (table 3), the major advantage being the increased potential resolution due to the much shorter wavelength of the electrons and thus, the ability to circumvent the diffraction effects prevalent in the optical microscope. But, as with anything good, there are also limitations and compromises that complicate the choice.

The scanning electron microscope is often thought of as a panacea for the measurement needs

Table 1. Relationship of process tolerance to the linewidth edge uncertainty.

Feature Size	Process Tolerance (Micrometers)			
	10% Control	LWM Edge Uncertainty	5% Control	LWM Edge Uncertainty
1.25	0.125	0.0625	0.0625	0.03125
1.00	0.100	0.0500	0.0500	0.02500
0.75	0.075	0.0375	0.0375	0.01875
0.50	0.050	0.0250	0.0250	0.01250
0.25	0.025	0.0125	0.0125	0.00625
0.10	0.010	0.0050	0.0050	0.00250

of the semiconductor community. This is not true today for *accurate* linewidth measurement, but it may ultimately fill that niche as the instrument matures. Unlike the optical microscope which traces its history back to the 1600s and in which optical theory has had a great deal of time to mature, the SEM has only been on the scene as a production instrument since the early-to-mid 1960s and electron optical theory presently is limited by this infancy. The SEM was not originally developed to do the very precise critical dimension measurement required today by the semiconductor manufacturing industry, but as an analytical and picture taking instrument. The mystique surrounding the SEM found its way into semiconductor manufacturing via this route and soon SEM-based measurement followed. In this transition, an attitude developed and was fostered that anything photographed in an SEM was correct. Since the SEM is considered the ultimate authority, measurements made using this instrument are also thought to be indisputably correct. Figure 2 demonstrates a scanning electron micrograph of a common object that everyone should immediately recognize and be able to measure. The

Table 2. Typical inspection instrument allocation scheme for a semiconductor processing facility.

Basic Material	Device Type	Minimum Linewidth		Measurement Instrument	
		Production	R&D	Production	R&D
Silicon	Large Scale Integrated Circuit	1.5 μm	1.2 μm	Optical	Optical
	High Speed Bipolar Integrated Circuit	1.0 μm	0.5 μm	Optical	SEM
	Transistor	0.8 μm	0.5 μm	SEM	SEM
Gallium Arsenide	Integrated Circuit	0.8 μm	0.3–0.5 μm	SEM	SEM
	Field Effect Transistor	0.3 μm	0.25–0.3 μm	SEM	SEM

Table 3. Comparison of some of the advantages and disadvantages afforded by the use of the scanning electron microscope for semiconductor linewidth measurement and inspection.

SEM VS. OPTICAL MICROSCOPE

Comparative Advantages

High Resolution Potential (2-20 nm)
 Excellent Depth of Focus (Field)
 Flexible Viewing Angles
 X-Ray Characterization
 Readily Interpreted Image

Comparative Disadvantages

High Vacuum Required
 Lower Throughput
 Electron Beam/Sample Interactions
 Sample Charging
 No Linewidth Standard Available
 Expensive

magnification is indicated in the lower left corner of the micrograph and, in the center, a line scale or micrometer marker indicates the size of the structure as scaled to the magnification. If the philosophy that everything that comes from a scanning electron microscope is correct, then so is that micrograph. This *could* be typical of any micrograph obtained in a standard instrument. This micrograph seems correct as it resides in the frame of reference of the reader (a dime is small; it easily fits in a pocket, so 7.1x seems proper) therefore, a rough

measurement based on the information provided on the micrograph would set the size of the dime as being about 12 mm in diameter. This measurement is smaller than the actual size, for a dime is about 18 mm in diameter. The actual magnification displayed on the micrograph should be about 4.6x. Because it seems reasonable to the reader, the magnification of 7.1x is acceptable. Many micrographs taken of micrometer and submicrometer structures in fabrication facilities also seem reasonable (a submicrometer line is small so such a measurement seems correct) but that does not make them accurate. One has no real firsthand experience in this microscopic world and thus most anything can seem reasonable given the right circumstances. This *especially prepared* micrograph of the dime is designed to prove a point, which is that the SEM does not always tell the truth. The scanning electron-beam instrument *as with any instrument being used for metrology*, will only provide correct data to the observer if it is adjusted to a proper calibration standard, its limitations are understood and strict controls are established and maintained. Without these controls, precise measurements using the SEM are impossible. The engineer, using the SEM to control a process, must look as critically at the micrographs obtained as we now look at the previous figures and he must also ask specific questions of the operator to ensure that the data obtained are really significant and accurate.



Figure 2—Scanning electron micrograph of a dime demonstrating the importance of proper SEM calibration procedures. Note that the magnification is displayed in the lower left corner and the accelerating voltage displayed in the center. See text for full explanation.

For the purpose of discussing SEM metrology, a clear distinction between the terms precision and accuracy must be made at the onset. This is necessary because in many instances these two terms have been erroneously used and treated as if they were synonymous.

Precision and Accuracy

In metrology [5] the term precision, often referred to as repeatability, is defined as the spread in values associated with the repeated measurements on a given sample using the same instrument under the same conditions. The assumption is that the number of measurements is large, the sample is stable over time and that the errors introduced are random. This is essentially a measure of the repeatability of the instrumentation. Precision relates directly to at least four distinct factors: 1) instrument; 2) operator; 3) environment, and 4) sample. Many of the factors affecting SEM measurement precision will be discussed in later sections of this paper. In order to measure precision, it is not necessary to use an official standard. It is only necessary to use a sample that is of good quality and stable with time. This provides a measure of precision that is locally traceable, and is related to that particular instrument and sample. Furthermore, in the SEM, due to the higher inherent resolution attainable, this precision may only relate to a given section or area of that sample because a sample may vary from location to location. Due to the need for stability with time, the sample materials chosen for these samples may not be identical to the typical product sample of interest (i.e., photoresist). To compare precision between more than one site or instrument would require the particular sample to be carefully transported to the other location and then the test repeated. An adjunct to this would be that an organization (such as NBS) make up and test (with a single instrument) a series of precision test samples which then could be taken to the various sites of interest and the sample precision of the instruments at those sites tested and compared with the measurements made on the original instrument.

Accuracy, on the other hand, is a far more ambiguous concept usually relating to the measurement of some agreed upon quantity (or quantities). Accuracy for SEM metrology is one goal of the program at NBS. This goal is not necessarily identical in principle, or practice, to the goals of the present semiconductor industry, but the results are the same. That is, the production of an accurate SEM standard that can be used to determine the

accuracy of semiconductor product measurements. Not only must the above factors affecting precision be considered as limitations of measurement accuracy but also the manner by which a given structure is being measured. Thus, a program similar to that employed for the NBS optical microscope linewidth mask standard must also be undertaken [1]. This program utilizes computer modeling of the electron beam/sample interactions in order to obtain the necessary measurement accuracy. Many of those factors necessary to effectively model linewidth measurements in the SEM are not fully understood at this time [6] and approaches are being developed to quantify them [7,8].

In practice, accuracy may be achieved only if the instrument making the measurement is sufficiently precise and the specimen of interest exactly matches the standard in all important ways (materials, substrate, etc.) except the dimension or dimensions being measured. One complication for linewidth metrology of thick lines (i.e., photoresist, etc.) on wafers is that, even if an acceptable standard were available composed of one set of particular materials, there is no guarantee that a given production sample will match precisely the characteristics of the standard. This is especially true because of the vast number of possible combinations of substrate and resist being used in semiconductor technology today. What may become feasible is the development of an accurate linewidth standard of well established geometry and the parallel development of a computer program to handle the sample and instrumental differences between this standard and the product being measured. This problem is similar in concept to that required for the development of the Z, A, F factors for quantitative x-ray microanalysis and the programs developed at NBS (and other laboratories) to undertake this problem [9]. A program to undertake this challenge is also being implemented.

The Scanning Electron Microscope Metrology Instrument

The architecture of a typical scanning electron microscope wafer inspection instrument is similar to any modern SEM designed for low accelerating voltage operation with the exception that it is modified to accept and view large semiconductor wafers. The instrument may also have cassette to cassette capabilities to facilitate wafer loading and unloading and a computer-based video profile analysis or "linewidth" measurement system. An example of a generalized instrument is shown in figure 3. In this instrument, a finely focused beam of elec-

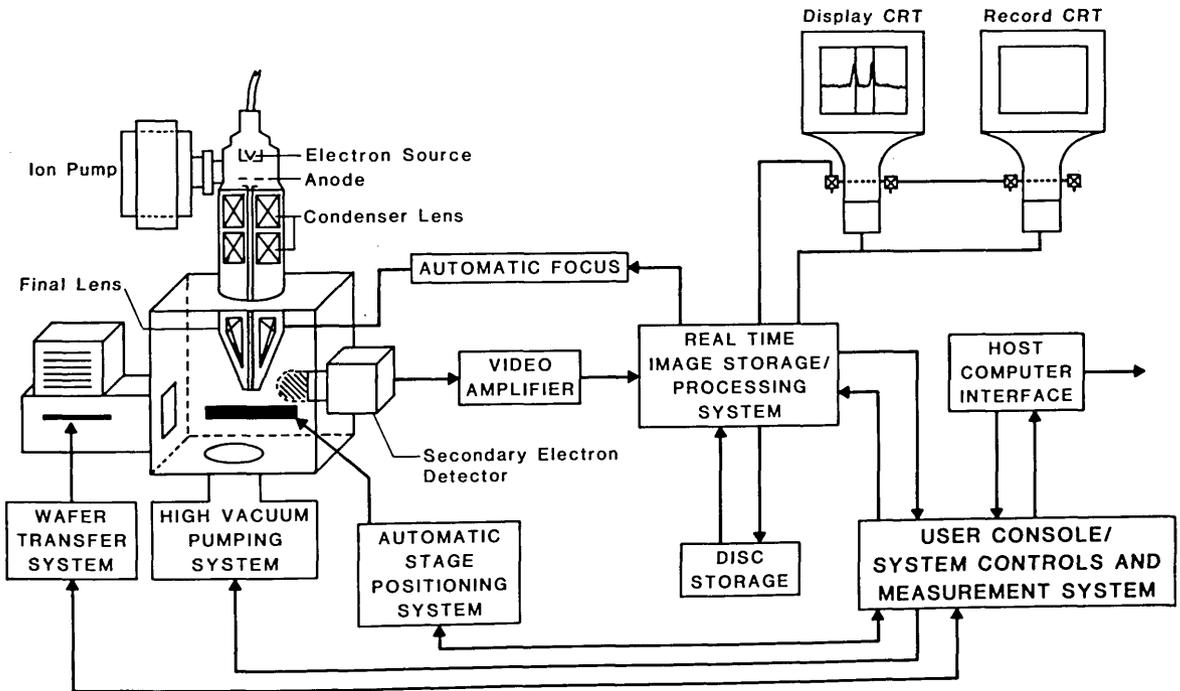


Figure 3—Schematic of a typical scanning electron microscope based wafer inspection instrument. The electron source and column design will vary with manufacture.

trons is moved, or scanned, from point to point on the specimen surface in a precise rectangular motion called a raster pattern. The electrons originate from a filament that may either be heated to a high temperature (thermionic emission), extracted at room or near room temperature (cold field emission) or a combination of both (thermally assisted field emission). Table 4 compares the operational characteristics of the different electron sources presently in use in instruments designed for wafer inspection. The electron gun is the "heart" of the SEM and the overall performance of the instrument ultimately relates to the current density of electrons emitted from the source. The larger this density the better the signal-to-noise ratio and hence the higher the limiting resolution. One measure of the performance characteristics of the electron gun is the measure of brightness (β). Brightness is the current density of the electron beam per unit solid angle and is defined by the following:

$$\beta = \frac{4i}{\pi^2 d^2 \alpha^2} \quad (1)$$

where i is the beam current; d is the diameter of the electron beam and α is the beam divergence (all measured at the specimen). Brightness is proportional to the current density of the source and it also increases linearly with accelerating voltage

[9]. The electron beam, once generated, travels down the column where it undergoes a multistep demagnification with magnetic lenses so that when it impinges on the sample, the beam diameter can range between about 1 nm and 1 micrometer (at 30 keV). Depending upon the particular application and specimen composition, the operator optimizes the proper conditions for magnification range, by adjustment of accelerating voltage, beam current and spot diameter.

The electron beam is precisely deflected in the raster pattern either in an analog or digital manner depending upon the design of the particular instrument. Most newer instruments employ digital scanning so that they can use frame storage and also incorporate auto-focus and auto-astigmatism correction [10,11]. This deflection is synchronized with the deflection of the display cathode ray tube (CRT) so there is a point by point visual representation of the specimen on the CRT screen as the electron beam scans the specimen. The smaller the area scanned by the electron beam, in the raster pattern relative to the display CRT size, the higher the magnification. The theory of the operation of the scanning electron microscope has been covered by several authors [9,12-14] and the reader is directed there for more in-depth coverage of this topic.

Table 4. Comparison of the four types of electron emitters presently in use in wafer inspection instruments. Data is for 20 keV operation.

COMPARISON OF TRADITIONAL ELECTRON EMITTERS USED IN SCANNING ELECTRON MICROSCOPY				
	Tungsten Hair Pin	Lanthanum Hexaboride	Cold Field Emitter	ZR-W (100) Emitter
Type of Emission Source	Thermionic	Thermionic	Field	Field
Temperature (K)	2650-2900	1750-2000	300	1800
Brightness (A/Cm ² SR)	10 ⁴ -10 ⁵	10 ⁵ -10 ⁶	10 ⁷ -10 ⁹	10 ⁷ -10 ⁹
Virtual Source Size (Angstroms)	1,000,000	200,000	50-100	50-100
Energy Spread (eV)	2-5	1-3	0.2-0.3	0.28-0.36
Vacuum (Torr)	10 ⁻³ -10 ⁻⁵	10 ⁻⁵ -10 ⁻⁷	10 ⁻⁹ -10 ⁻¹¹	<10 ⁻⁸

Electron Signals Used for Metrology

The primary electron beam, as it traverses the sample, interacts directly with the sample resulting in a variety of signals being generated that are useful for semiconductor inspection, analysis and metrology [15]. For historical reasons the major signals of interest to microelectronics dimensional metrology are divided into two groups, backscattered and secondary electrons, even though it must be remembered that this distinction is often arbitrary, especially at low beam energies.

Backscattered Electrons

Backscattered electrons are those which have scattered within the specimen and have been re-emitted from the specimen surface with energies which are a significant fraction (50% or more) of the incident beam energy. On a typical specimen, between 10% and 30% of the incident electrons ultimately become backscattered electrons. This fraction varies with the atomic number and surface geometry of the specimen but it is relatively independent of the beam energy. Because these electrons have relatively high energies they can travel significant distances through the sample and emerge from the whole area defined by the beam interaction volume. Thus, in silicon at 15 keV a

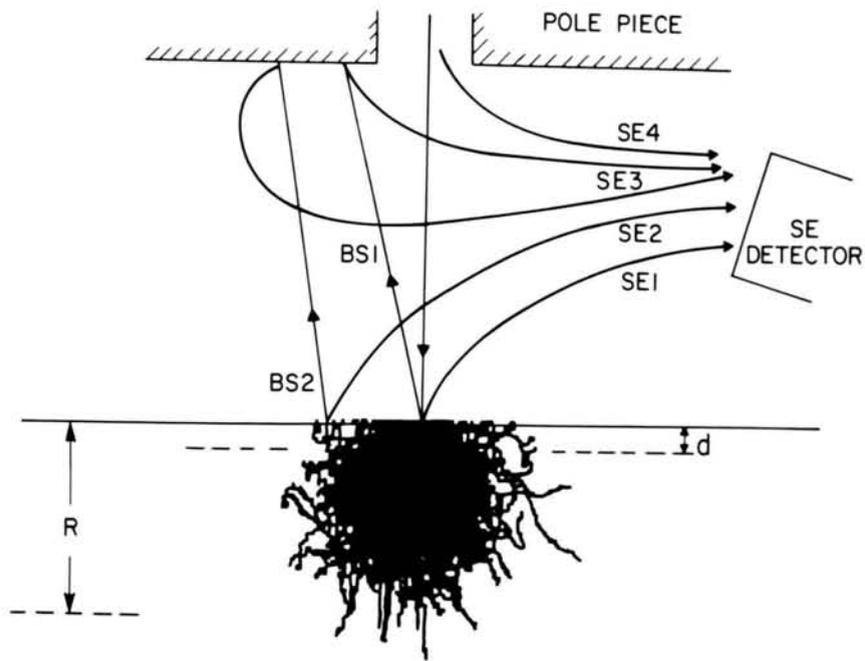
backscattered electron may escape from an area which is about one micron in radius and from depths of up to one and a half microns beneath the surface (fig. 4). The maximum range of electrons in a sample, can be approximated using the expression derived by Kanaya and Okayama [16]

$$Range(\mu m) = 0.0276AE_0^{1.67}/Z^{0.889}\rho \quad (2)$$

where E_0 is the primary electron beam energy (keV), A is the atomic weight, ρ is the density of the material (g/cm³) and Z is the atomic number. The calculated range of electrons in silicon for a variety of changes in accelerating voltage is shown in table 5. If one considers that the calculated range approximates the boundaries of the electron trajectories as a region centered on the beam impact point (fig. 4), then it can be seen that the backscattered electrons which emerge from approximately the upper one-third to one-half of this region do not, in general, carry much information about the high resolution details making up the surface topography of the specimen. But, at low magnifications (less than 1000x) where features on the scale of microns are being viewed, significant and useful signal information is carried by these electrons.

Because the backscattered electrons are energetic they are re-emitted away from the sample surface in straight lines. Consequently, they are usually collected by placing a detector in their path rather than by using a collecting (attracting) field.

Figure 4—The origins of various components of the secondary (SE) and backscattered (BS) electrons in the specimen chamber of the SEM. The electron range in the specimen is R , and the secondary electron escape depth is shown as d .



The size, sensitivity and position of the detector drastically affect its collection efficiency and thus the appearance of the image and, of course, the results of any measurements made from it. A large detector placed above the sample will give a high quality, low noise, image that appears evenly illuminated but in which the topography is of low contrast. A small detector, placed to one side of the sample, will collect fewer electrons (yielding a noisier image), but will produce topographic contrast that is much stronger and is marked by what appear to be strongly directional shadows. Metrology schemes must, therefore, take into account the characteristics of the detector and its effect on the observed signal.

Secondary Electrons

Secondary electrons are another signal of interest in the SEM. These electrons are defined as those with energies between about 1 and 50 eV. At an incident energy of 15 keV each 100 incident

Table 5. Approximate Kanaya/Okayama electron range in micrometers for silicon computed using eq 2 for several accelerating voltages.

Kanaya/Okayama Electron Range in Micrometers For Silicon								
keV	1.0	1.5	2.0	5.0	10.0	15.0	20.0	30.0
μm	0.032	0.062	0.101	0.466	1.48	2.92	4.72	9.29

electrons will produce, on average, 10 to 20 secondary electrons. This number, however, increases rapidly as the beam energy is reduced until at some energy E_2 (fig. 5) the total secondary plus backscattered yield ($\eta + \delta$) becomes one (unity); that is to say each incident electron produces on average one emitted electron. Since the secondaries are low in energy, their trajectories are readily deflected by local electric or magnetic fields. High efficiency collection of secondaries is therefore possible even with a physically small detector since this can be made efficient by applying a suitable electron-attracting (biasing) voltage to it. This convenience plus the higher signal-to-noise ratio

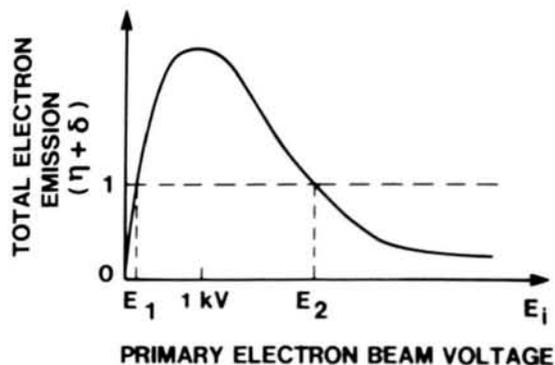


Figure 5—Variation of total secondary plus backscatter electron yield from a specimen plotted as a function of incident beam energy. The total yield is unity for two energies E_1 and E_2 called the cross-over points.

has led to secondary electrons being the preferred mode of operation for most purposes in the SEM.

Because of their low energy, secondaries cannot reach the surface from deep in the specimen, and typically they escape from a region only 5 to 10 nanometers beneath the surface. They, therefore, carry surface-specific information. Several different types of secondary electrons can be distinguished [17], as shown in figure 4. The most desirable for metrology and imaging are called the SE1 electrons, which are generated as the beam enters the sample. These secondary electrons are produced at the beam impact point and therefore carry the highest resolution information. The secondary electrons that are produced by backscattered electrons as they again pass through the surface escape region are called SE2 electrons. These secondaries are emitted from a surface area as large as that from which the backscattered electrons emerge, and the number of these electrons will depend directly on the number of backscattered electrons. Thus, the SE2 signal carries the same contrast information, and displays the same spatial resolution, as the backscattered signal. Typically, the SE2 component is as large, or larger than, the SE1 signal.

Finally, secondary electrons can also be produced external to the specimen by backscattered electrons which have been emitted from the specimen that hit the polepiece or walls of the specimen chamber (SE3), or from the impact of the incident electrons on the electron-optical defining apertures (SE4). The SE3 electrons carry information similar to that of the SE2 electron signal. The SE4 electrons contribute no contrast information, but, simply act as a "background" to the wanted signal, reducing its visibility and signal-to-noise ratio. Thus, in an SEM designed for metrology, attention must be given to reducing the relative magnitudes of the SE3 and SE4 components. In an unoptimized instrument, as much as 60% of the total secondary signal collected can be attributed to these unwanted emissions.

Since the secondary electron signal is easily influenced by the application of local electrical or magnetic fields, it is readily understood that the collection efficiency of a detector can relate directly to its position and potential. Detectors that have a location at some off-axis angle, as in many instruments also equipped to do x-ray microanalysis, show preferentiality of detection. In these cases, it is not possible to achieve the symmetrical waveforms necessary for precise linewidth metrology. To compensate for an off-axis position of the secondary electron detector, on a sample normal to

the electron beam, the sample must be physically rotated toward the detector until the video waveform of the line becomes symmetrical, then the structure can be straightened on the display CRT by adjusting the raster pattern with digital raster rotation. Since error can be introduced using this technique during the measurement of a tilted sample, it is much more desirable to have an on-axis detector [6] or two similar detectors on either side of the sample and the signals balanced and summed [18].

Low Accelerating Voltage SEM Operation

Historically, scanning electron microscopy was done at relatively high accelerating voltages (typically 20–30 keV) in order to obtain the best signal-to-noise ratio and best resolution. Nonconducting or semiconducting samples required an overcoating of gold or a similar material to provide conduction to ground of the electrons and to improve the secondary electron generation of the sample. In semiconductor device processing, this procedure is considered a destructive technique because the device cannot be processed further. On-line inspection during the production process of semiconductor devices is designed to be nondestructive which requires that the specimen be viewed in the scanning electron microscope uncoated. A thin insulating film on a conducting substrate can be viewed at a high accelerating voltage with an absence of electrical charging since most of the electrons are deposited in the substrate, but not all films are sufficiently thin for this technique. High accelerating voltages can also damage a semiconductor sample or device [19]. Low accelerating voltage inspection is thought to eliminate, or at least minimize, charging and device damage. In order to accomplish this in the SEM, the sample is viewed at accelerating voltages in the range of about 0.2–2.5 keV. Further advantages derived by operating the SEM at low accelerating voltages are that the electrons impinging on the surface of the sample have less energy, penetrate into the sample a shorter distance and have a higher cross section for the production of secondary electrons near the surface where they can more readily escape and, thus, be collected.

The secondary electrons are the most commonly detected signal carrier for low accelerating voltage inspection since their signal is much stronger than any of the others. The behavior of the total emitted electrons from a sample, shown in figure 5, is extremely significant to low accelerating voltage op-

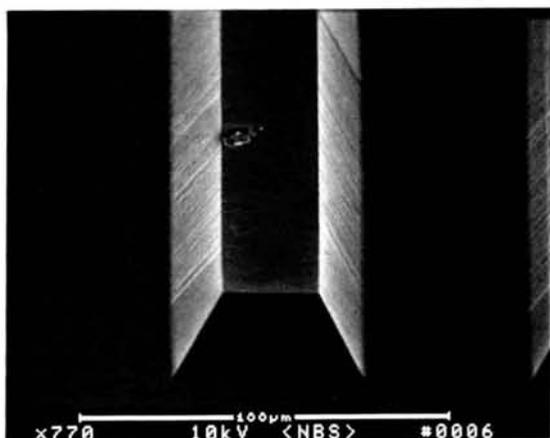
eration as those points where the curve crosses unity (i.e., E-1 and E-2) are the points where no electrical charging of the sample will occur. During irradiation of an insulating sample such as photoresist or silicon dioxide viewed normal to the electron beam, a negative charge can develop causing a reduction in the primary electron beam energy incident on the sample. If the primary electron beam energy is 10 keV and the particular sample has an E-2 of 2.0 kV then the sample will charge to about -8 kV so as to reduce the effective incident energy to 2 keV and bring the yield to unity. This charging phenomenon will have detrimental effects on the electron beam and degrade the observed image (to be discussed later). If the primary electron beam energy is chosen between E-1 and E-2 then there will be more electrons emitted than are incident in the primary beam, and the sample will charge positively. Positive charging is not detrimental as it is only limited to a few electron volts because of the resulting barrier to the continued emission of the low energy secondary electrons. This reduction in the escape of the secondaries stabilizes the surface potential but reduces the signal as these electrons are now lost to the detector. The closer that the accelerating voltage approaches to the unity yield point, the less the charging effects. Each material component of a specimen being observed has its own total emitted electron/keV curve and so it is possible that in order to completely eliminate sample charging a compromise must be made to accommodate the different specimen materials. For most materials used in present semiconductor processing an accelerating voltage in the range of about 1.0 keV (± 0.5 keV) is sufficient to reduce charging and minimize device damage. Tilting the sample increases the total electron emission and thus, is also useful in decreasing sample charging (to be discussed later).

Although operation at low beam energies is useful for the inspection of delicate samples with a minimum of charging, the filament brightness is lower leading to reduced signal-to-noise ratio. This results in a loss in apparent sample detail. High brightness electron sources and digital frame storage techniques for signal integration over short periods of time at TV rates minimize this problem [20]. The more abiding problem with low accelerating voltage operation is the lower spatial resolution (as compared to the higher beam energy operation) characteristic of this operational mode. If a contemporary instrument, equipped with a high brightness lanthanum hexaboride filament is capable of 4 nanometers resolution at 30 keV accelerating voltage it may be only able to achieve

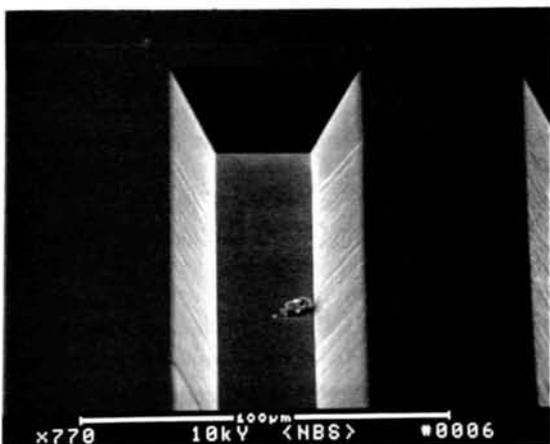
about 10–12.5 nanometer resolution at 1.0 keV. This limitation must be understood and factored into the precision requirements for submicrometer measurement applications.

Specimen Beam Interactions

While it is often true that the appearance of a scanning electron micrograph is such that its interpretation seems simple, this may not always be the case (figs. 6a and 6b). Care must always be taken so as not to become confused by "obvious" interpretations. When quantitative feature-size measurements



a



b

Figure 6—Scanning electron micrographs showing an illusion possible in the SEM that demonstrates that an understanding of the sample is often necessary to facilitate proper interpretation of the images. (a) In this micrograph, the image appears to be a line standing above the substrate. (b) In this micrograph, the structure appears as a trench. The only difference between these micrographs is 180 degrees of raster rotation.

are to be made it is even more necessary to be able to unambiguously relate signal variations to the details of the surface morphology. Because the interaction of electrons with a solid is such a complex affair (e.g., each electron may scatter several thousand times before escaping or losing its energy, and a billion or more electrons per second may hit the sample) statistical techniques are an appropriate means for attempting to mathematically model this situation. Although transport theory [21] provides an elegant solution for simple systems, it is of little value when considering complex device geometries. The most adaptable tool, at the present time, is the "Monte Carlo" simulation technique. In this technique, the interactions are modeled and the trajectories of individual electrons are tracked through the solid. Because many different scattering events may occur, and because there is no a priori reason to choose one over another, algorithms involving random numbers are used to select the sequence of interactions followed by any electron (hence the name, Monte Carlo). By repeating this process for a sufficiently large number of incident electrons (usually 5000 or more) the effect of the interactions is averaged, thus giving a useful idea of the way in which electrons will behave in the solid.

The Monte Carlo technique has many benefits as well as several limitations [6,22]. Because each electron is individually followed, everything about it (its position, energy, direction of travel, etc.) is known at all times. Therefore, it is straightforward to take into account the sample geometry, the position and size of detectors, and other relevant experimental parameters. The computer required for these Monte Carlo simulations is modest and, in fact, even current high performance personal computers can produce useful data in reasonable times.

In its simplest form [23,24], the Monte Carlo simulation allows the backscattered signal to be computed, since this only requires the program to count what fraction of the incident electrons subsequently re-emerge from the sample for any given position of the incident beam. By further subdividing these backscattered electrons on the basis of their energy and direction of travel as they leave the sample, the effect of the detection geometry and detector efficiency on the signal profile can also be studied. However, while this information is a valuable first step, under most practical conditions it is the secondary electron signal that is most often used for metrology in the low accelerating voltage applications. Simulating this is a more difficult problem because two sets of electron trajectories—1) those of the primary (incident) electron,

and 2) those of the secondary electron that it generates—must be computed and followed. While this is possible in the simplest cases [7,25] it is a more difficult and time consuming approach when complex geometries are involved.

For this reason, a new approach has been proposed [8,22] and is currently undergoing further development. In this method, a simple diffusion transport model for the secondary electrons is combined with a Monte Carlo simulation for the incident electrons. This procedure allows both the secondary (SE1+SE2) and the backscattered signal profiles to be modeled simultaneously with very little increase in computing time. Once that data are available, the effect of other signal components, such as the SE3 signal, can also be estimated. All the computed results discussed below are generated using this method.

The importance of being able to model signal profiles for some given sample geometry is that it provides a quantitative way of examining the effect of various experimental variables (such as beam energy, probe diameter, choice of signal used, etc.) on the profile produced, and gives a way of assessing how to deal with these profiles and determine a criterion of line edge detection for given edge geometries and thus, a linewidth [6]. However, at the present time, the Monte Carlo technique is not useful for deducing the line-edge geometry from the acquired SEM video profiles.

SEM-Based Metrology

The basic premise underlying the use of the scanning electron microscope for critical dimension measurement for semiconductor research and production applications is that the video image acquired, displayed, and ultimately measured reflects accurately the structure of interest. However, the secondary electrons detected do not necessarily originate at the point of impact of the primary electron beam. Indeed the effects of the four types of electron contributions to the actual image or linewidth measurement (see fig. 4) have not been fully evaluated. Errors in measurement are also introduced by sample charging and environmental influences (e.g., stray magnetic fields and vibration). In measurement applications, error due to the actual location of signal origination usually will not affect pitch measurements because the errors cancel [1,26,27]. However, in linewidth measurement, many potential errors are additive and thus will give twice the edge detection error to the mea-

sured width. The imprecision of any SEM-based metrology system is composed of two basic components: the imprecision of the actual instrument itself assuming an ideal sample, and the imprecision introduced by variations in the actual sample [28]. Some of the factors that today limit the precision of the SEM metrology instrument will now be discussed.

Definition of Linewidth

Scanning electron microscope metrology and optical metrology have one thing in common at the present time; that is except for vertical edges, there is no well-defined definition of the meaning of linewidth [1]. The first consideration that must be developed and defined when describing the term linewidth is what is actually being physically measured. Depending upon the lithographic process, the definition of linewidth may vary relative to the structural importance to subsequent steps. Figure 7a shows an idealized structure in cross section. In this case, D_1 and D_2 are not equal and hence the sidewall has some angle from normal. Linewidth could be defined as D_1 or D_2 or their average. Due to the large depth of field of the SEM inspection instrument, this distinction becomes significant since, if the conditions are properly chosen, both regions could be simultaneously in acceptable focus. Another situation for linewidth definition error occurs when an undercut sample is being observed (fig. 7b). In this case, D_1 is smaller than D_2 , but D_1 may not be readily observed unless the sample is highly tilted. Either of these two cases can result in difficulties in deducing where the edge is located and errors in precision. As the sidewall approaches 90 degrees (fig. 7c) this definition problem diminishes as $D_1 = D_2$ and precision (reproducibility) problems relate only to edge and sidewall irregularities and not misinterpreted edge location. A further confusion to any of the above instances would be introduced if the line was asymmetrical in cross section. In addition, the improved resolution of the SEM, as compared to the optical microscope, can also lead to deceptively imprecise data due to small irregularities in edge and sidewall structure that can be resolved and measured by the SEM. This discussion of the definition of linewidth has been limited to the description of where on the particular structure the measurement is to be made and not how to make the measurement. Further work modeling the structures and relating it to the physical edge is necessary before the actual linewidth can be defined and accurately measured.

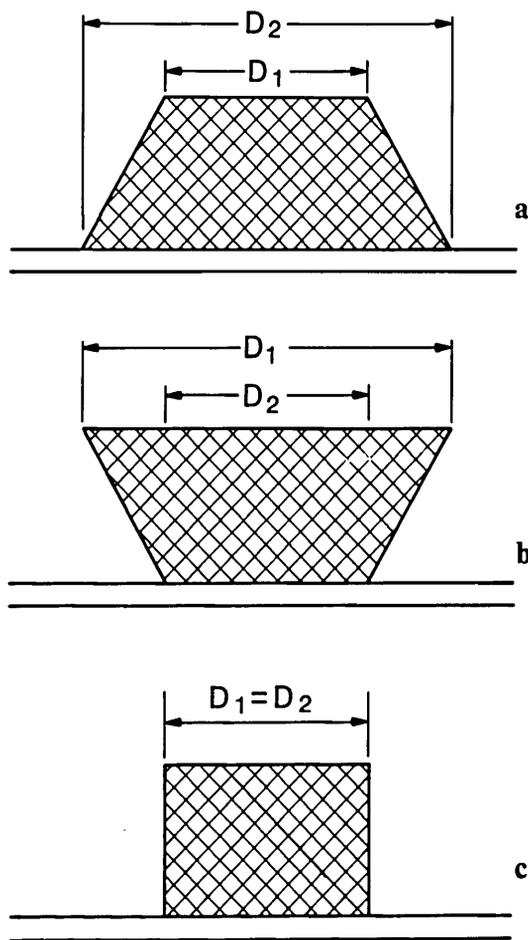


Figure 7—Drawing of a line structure as viewed in cross section showing the confusion possible in determining what edge is, in fact, being measured in the scanning electron microscope. (a) Trapezoidal structure where the upper width D_1 is smaller than the base width D_2 . (b) Undercut structure where D_1 is larger than D_2 . (c) Structure with vertical sidewalls where D_1 and D_2 are approximately equal.

Sources of Instrumental Error

Methods of Measurement. In commercial SEMs, used for critical dimension (CD) or linewidth metrology, two basic techniques of measurement are presently employed: beam scanning and frame storage. The two techniques are, in principle, similar. The beam scanning technique digitally acquires one scan line of video information from a sample positioned perpendicular to the x direction (horizontal scanning axis; the y-scan direction is commonly the vertical axis) with some pixel point resolution, and measurement algorithms are arbi-

trarily applied to that single line scan to obtain the width. Multiple acquisition of these linescans enables averaging over the field of view. In the frame storage imaging and measurement technique, an entire raster of information is stored digitally at some pixel point resolution depending upon the hardware design of the particular instrument. With this technique, since many individual line scans of data are actually stored (generally in about 512 positions along the line in the y direction) measurement algorithms can be applied anywhere in the field to data acquired in the x direction. Under both of these conditions, the precision of the measurement is severely influenced by the factors previously discussed such as electron beam effects, sample irregularities and the definition of linewidth. The instrumentation design and limitations must also be considered as a factor adding uncertainty to the measurement. For example, scan linearity, magnification compensation, and lens hysteresis are serious influences that must be considered, understood and compensated for, if possible, to name a few. Jensen 1980, Jensen and Swyt 1980, Seiler and Sulway 1984 and Nyssonen and Postek 1985, discuss these and other instrumental limitations (e.g., CRT linearity) and the reader is directed to these references for further information. The overall precision of the metrology system is also limited by the pixel point resolution of the measurement system. Table 6 demonstrates the linewidth measurement uncertainties associated with a 512×512 pixel point resolution system. Many commercial linewidth measurement systems at the present time acquire approximately 512 pixel points of information for linewidth measurement although some of the newer "dedicated" systems can acquire up to 2048 pixel points of information [4]. These techniques, even with their limitations, are of value due to their speed as throughput is a major concern for the production engineer. However, limitations on the pixel point resolution must

also be understood in order to properly interpret the measurement results.

Measurements can also be done by moving the stage/sample rather than the electron beam [6,26]. In this technique, the beam remains stationary (or oscillated slightly in the y direction to integrate slight sample irregularities) and the sample is driven in the x direction on a piezo stage. As the sample is moved, its position is precisely monitored using laser interferometry. Both the sample position and video intensity data for each point are stored for analysis. Using this technique, most of the errors in the SEM focusing and scanning system are minimized if not eliminated (but not the electron beam/sample interaction problems) and the measurement can be referenced to an accepted standard of length traceable to national standards [29]. Unfortunately, this technique although extremely accurate requires an elaborate laser-interferometer piezo-scanned specimen stage. Consequently, the procedure is relatively slow, thus making it unattractive for most production situations where throughput is of paramount importance.

Environmental Influences. The scanning electron microscope metrology system used for on-line inspection is usually located in a clean room. A great mass of literature is available on the air scrubbing aspects of the clean room and the mechanisms necessary to ensure that particle counts are low. However, little attention has been paid to the consequences of these actions on the metrology instrumentation. The SEM metrology instrument is an imaging system and as such the problems posed by the clean room environment are readily observable by these systems with excellent resolution. It should be noted that these problems can also detrimentally affect *other* clean room instrumentation but their effects are not directly observable in time and so the significance is lost. In most cases surveyed, the SEM metrology instruments presently operating in the typical clean room are not performing optimally. This is usually due to two main reasons: excessive vibration and stray electromagnetic fields.

Vibration. The effect of vibration on linewidth metrology, while obvious, is unfortunately, often overlooked. Clearly, vibration can originate from either the instrument or the environment, but their effects on the measurement of linewidth are similar (figs. 8a and 8b). Vibration, of the specimen relative to the electron beam, broadens the measurement and yields a linewidth uncertainty of twice

Table 6. Relationship between the pixel point resolution of a measurement system and the linewidth resolution for several magnification ranges.

LINewidth MEASUREMENT RESOLUTION (512 PIXEL POINT RESOLUTION)			
Magnification	Typical Field of View	Maximum Possible Pixel Point Resolution	Maximum Possible Linewidth Resolution
10,000X	10 μm	0.02 μm	0.04 μm
50,000X	2 μm	0.004 μm	0.008 μm
100,000X	1 μm	0.002 μm	0.004 μm

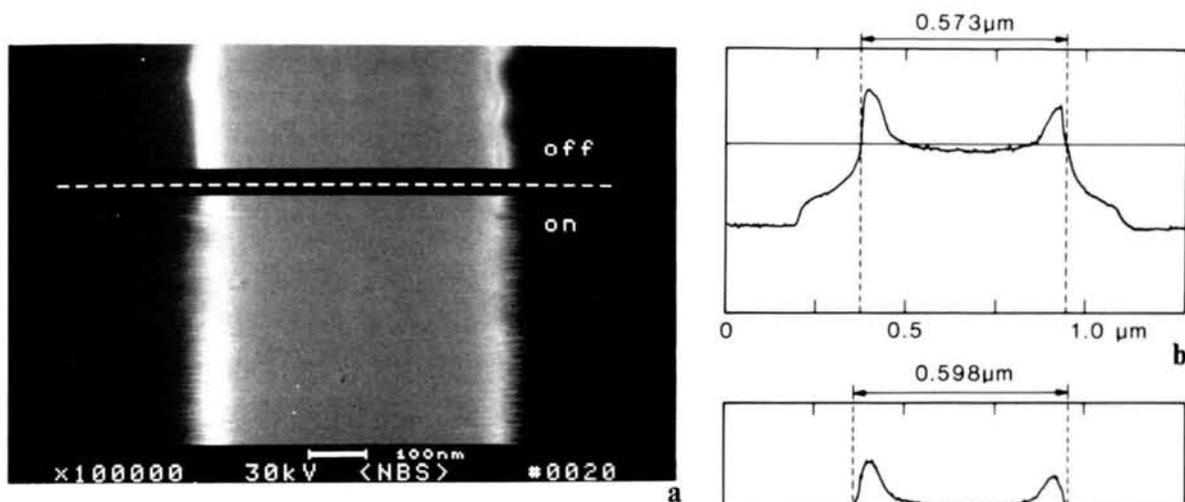
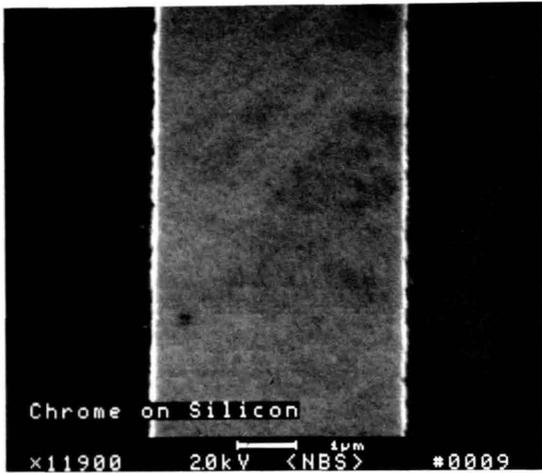


Figure 8—The effect of deliberately induced vibration on the image and measured linewidth. (a) Scanning electron micrograph showing the effect of vibration induced by a small cooling fan on the image; source off (top) and on (bottom) on the image. (b) Typical linewidth measurement taken with an arbitrary 40% positive automatic threshold crossing algorithm under ambient vibration levels typical for proper SEM operation. (c) Similar measurement, using the same threshold crossing algorithm, of the same sample position after vibration was induced. (100,000x; 30 keV)

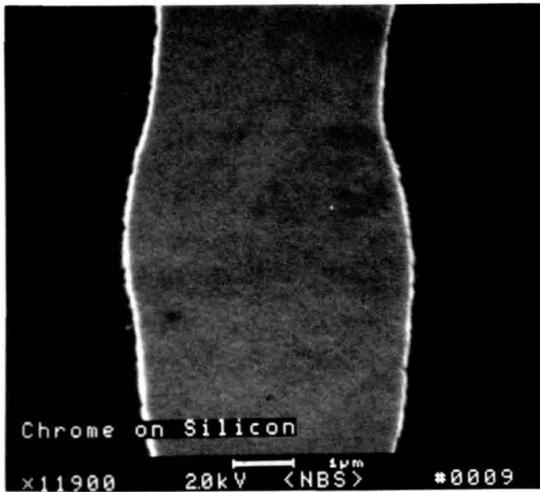
that of each edge. The sources of vibration in the particular installation must be identified and eliminated or steps taken to isolate the instrument from them. Some of the typical sources of vibration in the clean room are: undampened floor vibration, blower fans, vacuum pumps and air flow across the instrument. One solution to the vibration problem is to decouple the clean room from the measurement instrument either by placing the instrument on a vibration isolation unit, or a massive concrete pillar sunk to bedrock, or both. Of the two possibilities the latter is preferred wherever possible. The concrete instrument pad can then be properly vibration isolated from the clean room floor. It is recommended, that the entire instrument including the area used by the operator be on the concrete pad and not just the column section as vibrations can be transferred via the operator and umbilicals to the column section. Unfortunately, there is some cost to this modification but at some point decisions to optimize the metrology instrumentation must be made to ensure that the required measurement precision be met. Vibration induced by air flow can be minimized or eliminated by instrument shrouding or shielding. One consequence of unrecognized vibration is deceptively good measurement system precision since the continuous vibration is being continually integrated into the image, obscuring the actual sample detail and, smoothing the measurement data. Probably the best solution to the metrology problems is to design clean rooms

that have the SEM metrology instrumentation in an optimized external, but adjacent, area to the actual clean area and the product transferred to it in a controlled manner.

Stray Magnetic Fields. The SEM metrology instrument is an electronic instrument in an electronically hostile environment [30]. Both ac and some dc fields can affect this instrument in an undesirable manner. Studies have shown that many problems involve a field effect induced by the improper wiring and grounding of the instrumentation [31,32]. This not only refers to the SEM but also to any other instrument or wiring in the immediate environment including the lighting system. Ground looping in the clean room of equipment and lighting can result in ac fields in excess of 30 milligauss at the SEM column. This is equivalent to operating a small transformer a few centimeters from an SEM column. Many SEM manufacturers specify that external fields not exceed 1–3 milligauss for proper instrument operation. Newer instruments operating at TV rates synchronize the scan to the 60 Hertz (i.e., power line) thus concealing the observable field effects of that component on the CRT screen. Further, not all of the interference is generated at that frequency and problems can still be induced in the image and the measurement (figs. 9a and 9b). The most effective approach to the problem of undesirable stray fields is to identify the source and to eliminate it there. Supplemental



a



b

Figure 9—The effects of non-synchronized ac field induction on the image in an SEM. (a) Micrograph taken under ambient conditions where the instrument is synchronized to line (60 Hertz). (b) Micrograph taken after induction of a non-synchronized field that is 1/4 cycle off of 60 Hertz. Note the broadening of the structure.

shielding should only be used afterwards if the sources cannot be identified or eliminated. The shield may only prove to be a temporary solution since the overall complexion of the situation may be altered as other equipment is moved in and out of the clean room environs over time.

Operator Factors. Scanning electron microscopes, especially those equipped as metrology instruments, are complex, expensive investments. One area that has been severely neglected by many semiconductor companies is the role the metrology instrument operator plays in the success or failure

of the on-line inspection program. Even the simplest of the SEM systems are far more technologically involved than their optical microscope counterparts (although in both instances highly trained individuals should be used). This is especially true where routine instrument maintenance is concerned. Not every applicant is suited to become an SEM metrologist, and once an appropriate candidate is selected, a substantial amount of training must be invested in order for that individual to become confident with the particular instrument or instruments under his supervision. Further, once an individual has proved to be an asset in that position he must be encouraged to remain in that area and not be transferred out. Once an operator leaves the SEM metrology area his real experience value is lost. Experience cannot be taught, only gained! The trend toward automation of the SEM inspection processes may minimize the need for a large number of trained operators at some point in the future; however, this will not be for some time.

Instrument Maintenance. The SEM requires periodic electron optical column maintenance in order to maintain proper performance. Proper maintenance is especially important to low accelerating voltage operation. The maintenance period varies with instrument design, application and the types of specimens observed. It must be noted that in all instruments the components that directly interact with the electron beam (e.g., apertures) do become dirty due to deposition of residual hydrocarbons and oxidation products [33]. In a clean vacuum system, the majority of these contaminants are out-gassing products of the sample. Contaminant build-up can result in charging in the electron gun or in the column resulting in poor performance [34]. Asymmetrically deposited contamination, especially on apertures, increases astigmatism levels and may ultimately lead to the point where it becomes uncorrectable. Also, heavy build-up of contamination on an aperture can dislodge and either block the beam path or develop a charge and deflect the beam. The instrument operator must be experienced enough to recognize this condition and suspend work and take corrective actions so as not to compromise the measurement work. Some maintenance downtime must be expected on a periodic, or on an as-needed, basis in all production situations. Instrument manufacturers consider routine maintenance to be a user responsibility; however, in recent years this has been relaxed somewhat due to extended service policies and improved instrument performance. In order to regain the original performance level, only trained, expe-

rienced personnel fully understanding the work should undertake routine maintenance. Otherwise, extended and costly downtime may result.

One problem associated with the SEM in the production environment has been the lack of unified instrument standardization techniques that ensure that an instrument is operating optimally or, once an instrument has been dismantled for routine maintenance, that it is brought back to the same optimum level of performance where it was once running. Further, the data taken during the interface time between routine maintenance periods or while a decision was being made to service an instrument may, or not may not, be characteristic of the actual product, but a reflection of the condition of the instrument. Clearly, critical decisions must be made by the operator, based on the experience with the particular instrumentation in place that affects product acceptance. This is especially troublesome in locations where multiple instruments are in place (especially if they are from several different manufacturers) and the data is fed into a central data base for real-time analysis. Techniques for this purpose must be developed and diagnostics must be implemented into the SEM metrology instrument for this purpose. Each day, or at the beginning of each shift, diagnostic procedures must be done to ensure that the instrument is performing properly.

Sample Charging. The effects of sample charging on measurements made in the SEM have been studied [35-37]. Negative charging resulting when the electron beam voltage exceeds E-2 (fig. 5) can affect the video profile (fig. 10) and thus the measurement. The foremost effect is the possible deflection of the electron beam as the sample builds up an appreciable charge with its accompanying electric field. This may either manifest itself as a catastrophic and obvious beam deflection where the image is lost or a more subtle and less obvious effect on the beam. The subtle effects are the most damaging to metrology as they may manifest themselves either as a beam deceleration or a small beam deflection. All instrument compensations directly relate to the accelerating voltage applied and all instrument adjustments (e.g., magnification) depend on this beam energy. A slight beam deflection around a line structure can move the beam a pixel point or two, thus invalidating the critical dimension measurement. One pixel point deflection of a 1 μm line measured at 10,000x with a 512 pixel point digital scan corresponds to about 38-40 nm linewidth error (less at higher magnification). Positive charging may also have detrimental effects on the measurements as a positively charging struc-

CHROME ON GLASS MASK 10,000x; 0° TILT

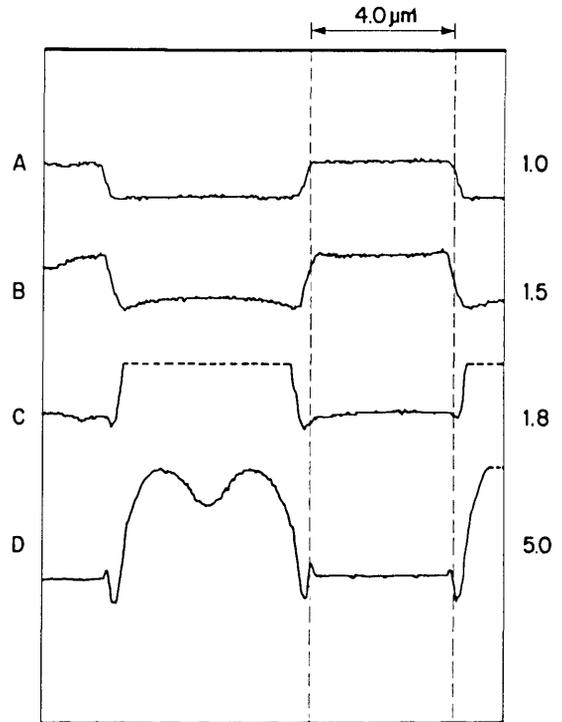


Figure 10—Sample charging and the effect on the video profile.

This chrome-on-glass mask was viewed and measured at increasing accelerating voltages. At 1.0 keV (A) no apparent sample charging occurs, as the voltage was increased to 1.5 keV (B) charging in the glass area begins to occur. The increase of accelerating voltage through 1.8 keV (C) to 5.0 keV (D) results in apparent sample charging and over-ranging of the video signal (dotted line). Note how the profile in the measured area of the chrome also changes with accelerating voltage.

ture can attract secondary electrons from adjacent pixel points, thus altering the measurement waveforms.

Sample charging can be reduced, if not completely eliminated, by adjustment of the accelerating voltage to the appropriate points on the total electron emission curve (fig. 5). Rapid TV-rate or near-TV-rate scanning is also being employed by several manufacturers to further reduce charging. Under these conditions, the electron beam dwells on the sample for less time per point than in slow scan, thus the charge has less time to develop. Another possible charge reducing technique which offers some improvement, is to tilt the sample toward the detector. Tilting the sample permits operation at higher accelerating voltages without charging effects by increasing the total electrons emitted. A sample viewed at 45 degrees of tilt may not demon-

strate charging with an accelerating voltage as high as 2.5 keV whereas the same sample will charge at about 1.3–1.4 keV viewed normal to the electron beam [37]. However care must be taken during the critical dimension measurements to minimize possible errors that tilting may introduce [37].

Signal Detection and Accelerating Voltage. The magnitude of the errors introduced to the linewidth measurement relative to the mode of signal detection and of beam acceleration voltages has been studied [38]. Figure 11 shows a silicon wafer sample with a silicide layer patterned with micrometer and submicrometer lines. This sample was observed and measured under controlled conditions at a variety of accelerating voltages and electron detection modes. A micrograph showing the effect of the choice of signal detection (secondary and backscattered electron imaging) is demonstrated in figure 12. In that micrograph, the actual width of the line is not changing dimension as the beam scans it to the extent indicated, only the manner of perceiving it in the instrument changed. The results of repeated measurements with a pixel point resolution of approximately 9 nanometers demonstrate that, depending upon accelerating voltage applied and the electron detection mode used to image and measure the structure of interest, a variety of results can be obtained. Further, measurement broadening affects of the beam penetration and beam/specimen interactions are apparent. Figure 13 shows the video profiles of the line measured at

two accelerating voltages and table 7 shows the measurement data. The SEM magnification was calibrated against an NBS standard and any processing irregularities present in the sample were well within the pixel resolution of the system and were also averaged over the field of view during the measurement process. Data was obtained from an average of 40 scans over a field of about $4.0\ \mu\text{m}$ and the measurements between accelerating voltage changes were adjusted to give the pitch. This clearly demonstrates that measurement criteria for each accelerating voltage must be established so that electron beam effects can be properly accounted for. Changes in apparent dimension can be attributed to the uncertainties contributed by: electron beam interaction effects, solid angle of electron detection, detector sensitivity, and the criterion used to determine the edge location in the computation of linewidth. These data further suggest that if several instruments are operating on a production line, care must be exercised to insure that all are working with the same accelerating voltages, instrument and measurement conditions.

Sample Contamination Effects. Semiconductor samples introduced into the SEM vary greatly in their surface cleanliness. For SEM inspection cleanliness, in this context, is not as much a lack of particles as a chemical cleanliness. This is as much of a concern in the SEM as it is in the optical microscope. The surface contamination levels present on the sample will vary with the preceding processing steps. Residual hydrocarbons adhering to

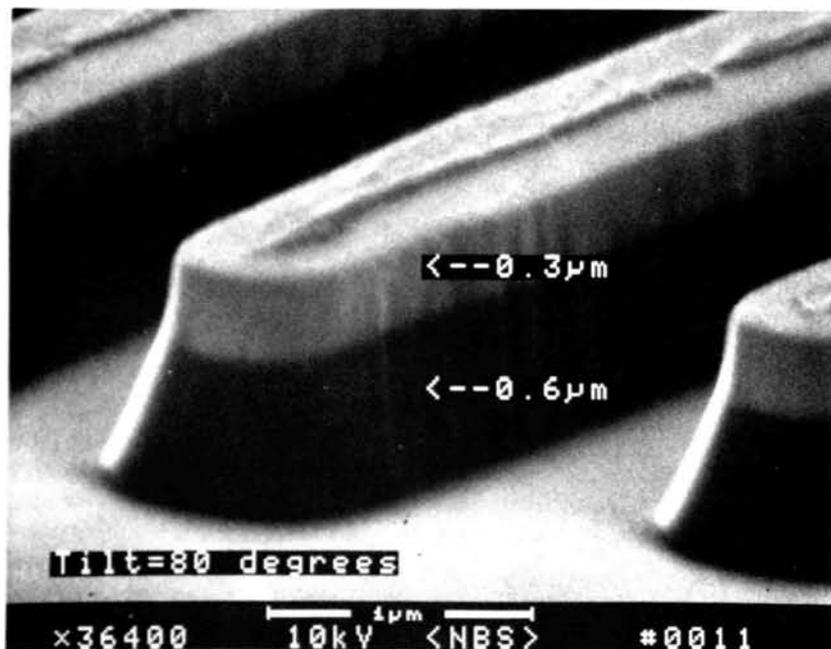


Figure 11—Micrograph of a nominal $0.75\ \mu\text{m}$ line showing the silicide and the etched silicon layers.

Figure 12—Effect of the mode of signal detection on the scanning electron microscope image. In this split field image, the effect of signal detection strategies on the image and thus the measurement, can be seen between secondary electron collection (SEC) and backscattered electron detection (BSE).

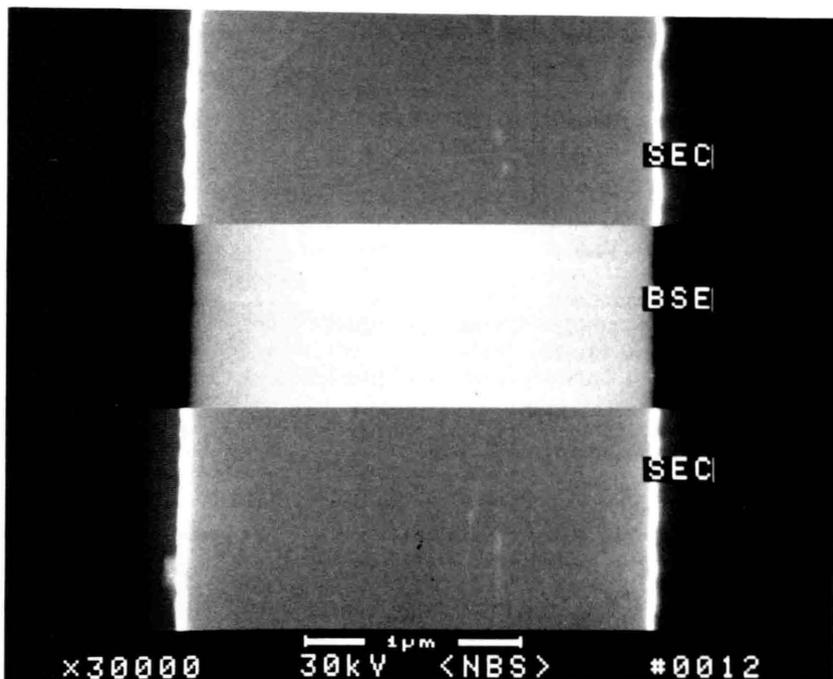
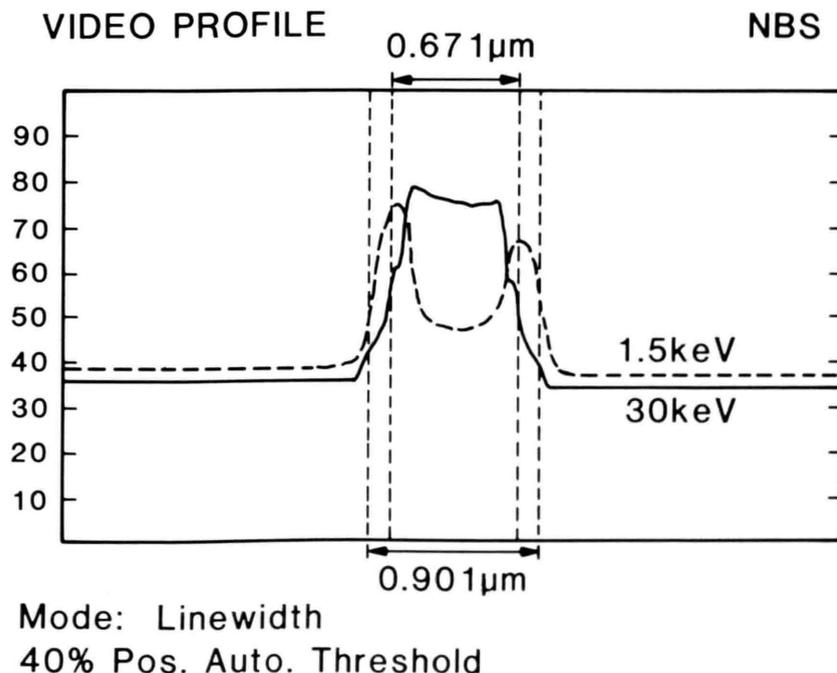


Figure 13—Overlay comparison of two digitally acquired video profiles of the 0.75 μm nominal line. One profile was taken at 1.5 keV and the other one taken at 30 keV. This comparison shows the reason for measurement discrepancies between accelerating voltages as the automatic threshold algorithm, arbitrarily set at 40%, is not appropriate for both the measurement conditions.



the surface or diffusing from within the structures in the vacuum can ionize as the beam scans resulting in beam deflection or beam broadening. The electron beam can also act to decompose these hydrocarbons at the surface in the area of the raster pattern effectively depositing a layer of carbon (figs. 14a and 14b). At higher accelerating voltages the electron beam penetrates this contamination and shows little effect (fig. 14a). At low accelerat-

ing voltages used for non-destructive inspection this contamination can severely alter signal generation and thus compromise data.

Sample Dimensional Changes. Electron beam irradiation can induce dimensional changes in photoresist structures [39–41]. A high resolution SEM image demonstrating a good signal-to-noise ratio can expose that sample to total electron beam dosages higher than that typical for electron beam

Table 7. Data from the measurement of a nominal 0.75 μm silicide on silicon line showing measurement variation as a function of accelerating voltage and signal detection mode.

NOMINAL 0.75 MICROMETER LINEWIDTH (AVERAGE OF 40 SCANS)				
keV	SEC	SD	BSE	SD
1.5	0.916	+ -0.0140	NA	NA
3.0	0.891	+ -0.0092	NA	NA
5.0	0.856	+ -0.0098	NA	NA
10.0	0.774	+ -0.0224	0.564	+ -0.0054
20.0	0.703	+ -0.0125	0.556	+ -0.0073
30.0	0.669	+ -0.0178	0.563	+ -0.0052
AVERAGE	0.802		0.561	
SD	+ -0.102		+ -0.004	

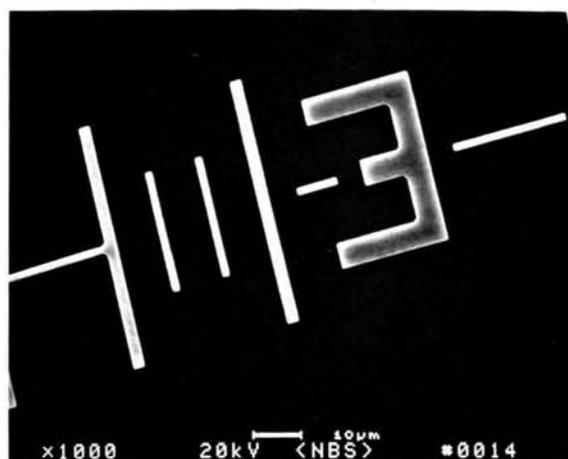
NA=NOT APPLICABLE

SD=STANDARD DEVIATION OF THE INDICATED AVERAGE AND IS A MEASURE OF THE VARIABILITY

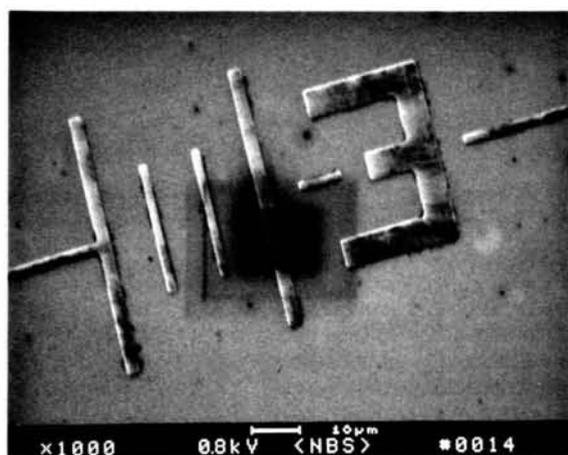
lithography. This can have a pronounced effect on the critical dimensions by either causing the resist to swell or shrink. Erasmus (1986), recently studied the dimensional stability of several commonly employed resists. This work demonstrated that even with a beam operated at 1.0 keV accelerating voltage resist shrinkage can be induced. Figure 15 reproduces some of the results found for an easily damaged resist such as PMMA. The rate of resist shrinkage is greatest when the electron range is approximately equal to the thickness of the resist because, under irradiation, all of the beam energy is deposited in the resist. Clearly, this is an interesting and controversial topic and further work on this and other materials needs to be done. The possibility of dimensional changes of the sample occurring during the measurement process must be explored and care must be exercised to determine the optimum conditions where radiation damage and instrument operating conditions are optimized.

Monte Carlo Modeling and Measurement

The above discussion demonstrates that many factors contribute positively or negatively to scanning electron microscope metrology. Many of the previously identified influences can be modeled using the Monte Carlo technique in an effort to develop increased measurement accuracy and precision. The type of information to be gained from a Monte Carlo simulation of the linewidth profile is best illustrated using a real example. Figure 16a shows the experimental line profile obtained from a chromium strip, 4.0 μm wide and 0.2



a



b

Figure 14—Sample contamination and the effects on the secondary electron image of a chrome on silicon wafer. (a) Sample viewed and photographed at high accelerating voltage (20 keV). (b) Sample viewed at low accelerating voltage (0.8 keV). Note the contamination on the sample apparent at low accelerating voltage operation is not apparent in the high accelerating voltage micrograph.

μm thick, deposited on a silicon substrate (fig. 16b). The profile was recorded in the secondary electron detection mode at 10 keV beam energy, with the beam sampling the specimen at intervals of approximately 10 nanometers.

Using this geometrical information, and the relevant physical parameters (such as the density, atomic weight and atomic number) of the materials making up the structure, the expected signal profiles for the secondary and backscattered electron signals can be estimated using the Monte Carlo model. The profile is built up by performing the simulation for beams incident at points separated by 10 nm, in order to match the experimental pixel

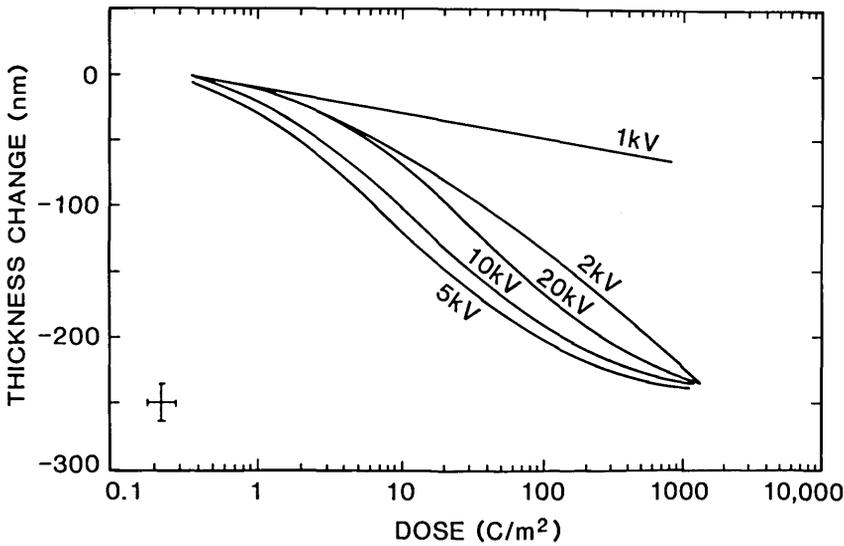


Figure 15—Experimental results showing the change in thickness of a $0.5\ \mu\text{m}$ thick PMMA film under electron beam irradiation for several accelerating voltages. (Figure re-drawn from Erasmus [39].)

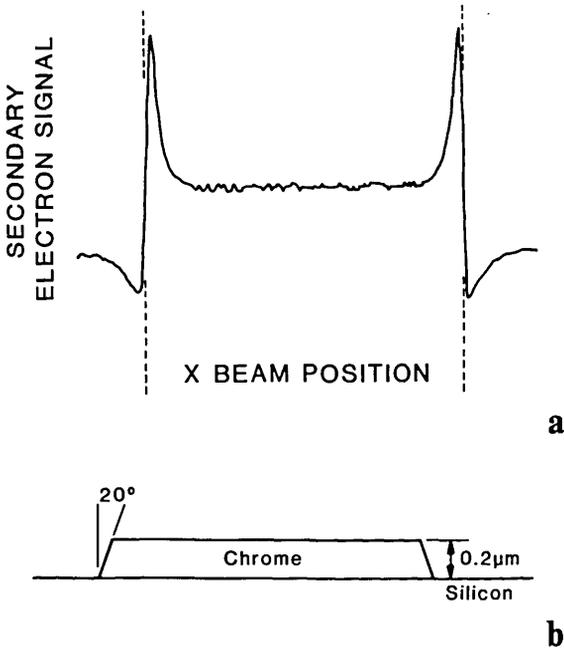


Figure 16—Monte Carlo modeling of SEM images. (a) Experimental video profile of the secondary electron image of the structure shown in (b) a $4.0\ \mu\text{m}$ chrome strip on silicon. The incident beam energy was 10 keV.

spacing. At each point 5000 trajectories are computed to ensure that the statistical error of the computation is kept to an acceptably low level. In order to generalize the simulation as much as possible, the profile is initially calculated for idealized conditions. Any given set of experimental conditions can then be matched by appropriate correction to this ideal profile.

The secondary and backscattered electron profiles obtained from the calculations are shown in figs. 17a and 17b. A comparison of these profiles with the experimental profile reveals several features of interest. The most important of these is the fact that the experimental profile, although recorded on the secondary electron mode, actually more closely resembles the computed backscattered profile. Compare, for example, the variation in signal just before the rapid rise at the edges of the chrome strip. The reason for this is that, as mentioned above, there are many sources of secondary electrons in the specimen chamber of the SEM. While, in principle, it is desirable to collect only those secondary electrons (SE1 and SE2) generated directly by the incident beam, in practice a contribution from the SE3 secondaries which are produced by the impact of backscattered electrons on the final lens and chamber walls are also included. These secondary electrons carry the information of the backscattered electrons that created them. The detected secondary electron signal is therefore actually a mixture of the secondary and backscattered components, the ratio of the mixture being determined by the exact geometrical arrangement of the sample in the chamber at any given time. For the data shown here, it is necessary to mix in about a 30% contribution from the backscattered electrons to match the experimental data.

Second, it is obvious that the features in the computed profiles are much sharper than those observed experimentally. One reason for this is that a real SEM has a finite probe diameter, while the computer model assumes a probe of zero size. The effect of a finite beam size can easily be simulated

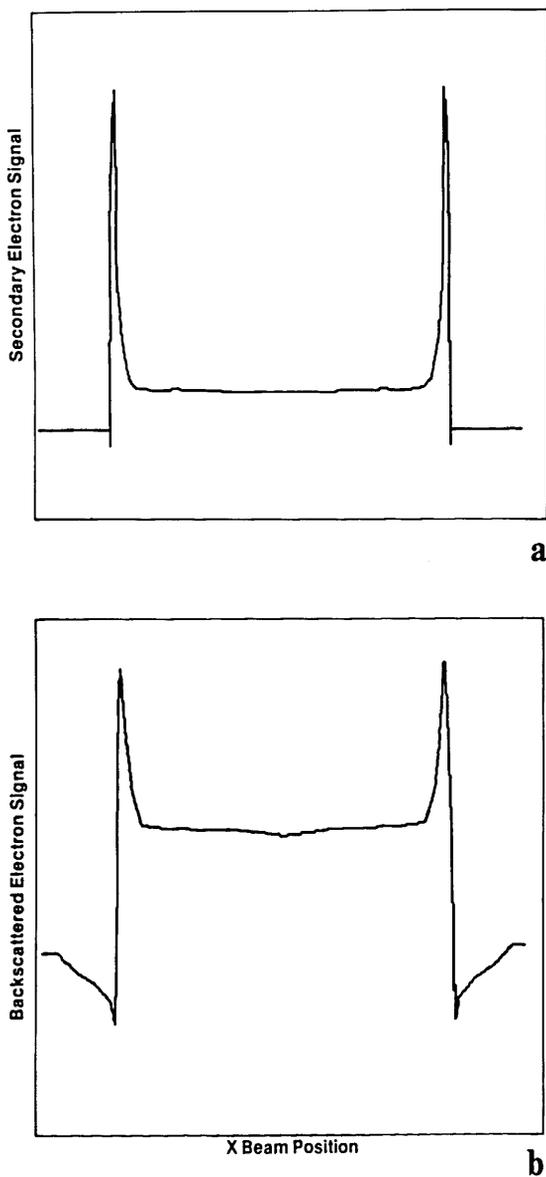


Figure 17—Monte Carlo modeling of SEM images. Idealized (a) secondary electron and (b) backscattered electron line profiles computed using the Monte Carlo technique for the structure shown in figure 16b.

by convolving the computed profiles with a function such as a gaussian, representing the size and intensity distribution of the incident electron probe. The computation also takes no account of the statistics of the signals detected. Because the measurement must be made in finite time, with a restricted beam current, the experimental data are shot-noise limited to a relatively poor signal-to-noise ratio. This can be modeled in the computed profiles by adding in an appropriate level of random noise. Finally, the computed profiles take no account of the properties of the electron detectors

or the associated electronics. The effect of the behavior of these components can be mathematically modeled and then used to modify the simulated profiles.

The final result of these modifications is shown in figure 18. The mixed secondary and backscattered signals have been convolved to an effective probe diameter of 25 nm full width half maximum (FWHM), adjusted to a signal-to-noise ratio of 10:1, and the detector efficiencies matched to those of the microscope. The resultant profile is now in good agreement with the experimental data. The advantage of proceeding in this systematic way from the idealized data to the fully corrected data is that it is possible to investigate the importance of different aspects of the experimental arrangement, by examining their effect on the linewidth "measured" from the computed profiles. For example, using an arbitrary 40% threshold crossing measuring criterion, the uncorrected secondary and backscattered profiles of figure 17 give widths that are, respectively, 0.45% and 0.95% smaller than the nominal expected width. After allowing for such factors as the finite probe size, the signal-to-noise ratio, and the detectors, the secondary profile now measures a value 0.5% larger than the nominal width, while the backscattered profile corresponds to a width 0.65% smaller. This significant discrepancy arises because the secondary and backscattered profiles are affected in opposite ways by the corrections applied. Although, for a line several micrometers in width the percentage error is not large, for a narrow line the effect would be proportionally much greater. Another result of this difference in width between the two profiles is that in situations where the experimental signal is actually a mixture of secondary and backscattered components, as in the case here, the measured linewidth will be a function of the ratio between the signals, and this may vary across the sample.

The sample discussed here is, in many ways, suited for SEM metrology since the feature is relatively large, has sharp edges, and is of high contrast. The fact that even in this case many sources of error are present indicates that the problems of more complicated specimens will be more challenging, and the requirement for modeling even greater.

Automated Wafer Inspection

It is apparent the SEM metrology instruments will follow the direction of the present optical instruments and fully automatic inspection systems will become available. It would seem that all the

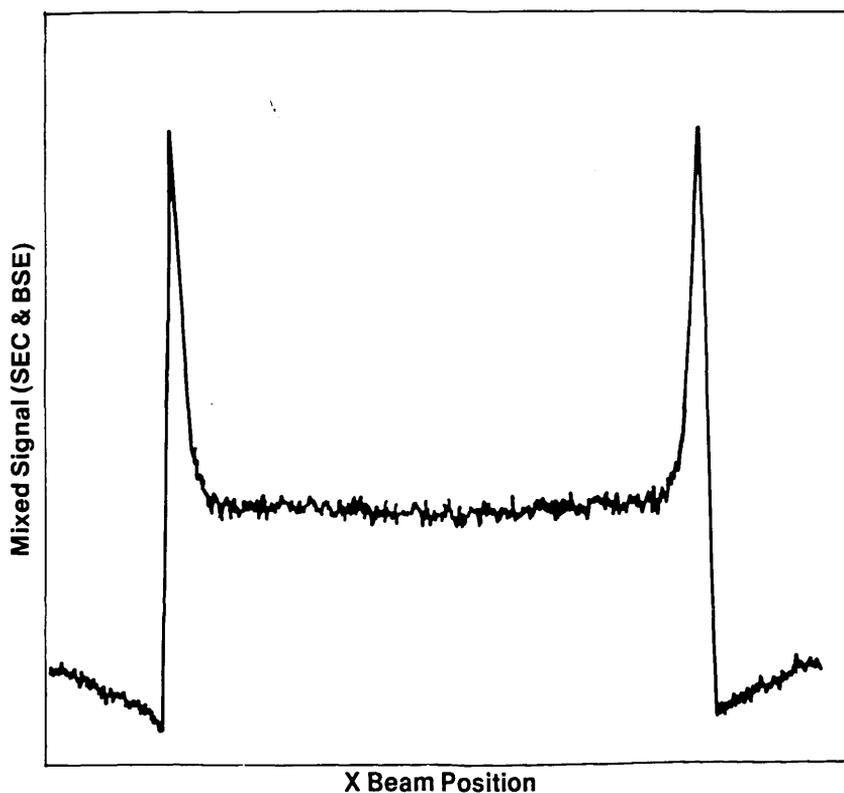


Figure 18—Stimulated line profile from the structure of figure 16b including corrections for finite beam diameter, detector efficiency, and signal-to-noise ratio. Components of both SEC and BSE signals have been added to allow for SE-3 contributions in the experimental profile (fig. 16a).

components for such a system are presently available: electron beam column and components from SEM manufacturers, and high speed wafer and data handling systems from the optical instrument manufacturers. A joining of the two is inevitable. One must not be lulled into thinking that the two system strategies are directly interchangeable. There are serious differences in the physics of the two types of instruments that must be understood and dealt with before image analysis can acquire and decipher meaningful metrology data from the acquired electron image. From what has been shown in earlier parts of this paper the problems are not trivial.

A desirable feature in a fully automated wafer inspection/SEM metrology instrument is the ability to compare the acquired image to some stored image or image-generating data base and undertake linewidth measurement and analysis. It would be folly to think that an image acquired in an SEM could be directly compared to a CAD database until the electron beam/sample effects were fully understood. An image overlay based on the stored image of a good device at high pixel point density with that of the unknown could be implemented, however extremely tight controls on the instrumental data acquisition conditions (as discussed above) must be maintained otherwise false image differences would result.

The automated inspection tool, while computing linewidth, could also undertake particle and defect analysis. The SEM images with electrons. The ability to see a feature is a function of the contrast produced. If the contrast of the structure is not adequate it is not observed. Signal is directly related to the number of electrons provided by the electron gun and, in this instance, the image contrast is derived from at least two main sources: atomic number contrast and topographic contrast. The electron beam must supply sufficient electrons in a small enough gaussian spot to resolve the structure of interest and the particle must be observed at sufficient magnification so that it is clearly discernable from the background. Further, the measurement must be made at a magnification adequate to resolve the structural detail necessary to meet the precision specifications desired in table 1. For the modern IC processing applications, particulate matter with sizes down to the submicrometer region must be considered. Table 8 demonstrates a projected throughput vs. magnification for the analysis of a submicrometer particle for a typical chip size of 1 cm^2 . This analysis also assumes that there is sufficient atomic number contrast to image the particle, a pixel point resolution adequate to resolve it to the analysis system and sufficient beam current focused into a spot size less than a pixel point. It is clear that new data acquisition and data

Table 8. Comparison of the total data acquisition time necessary for a typical 1 cm² die, using present techniques, as a function of the instrument magnification and the acquisition frame rate.

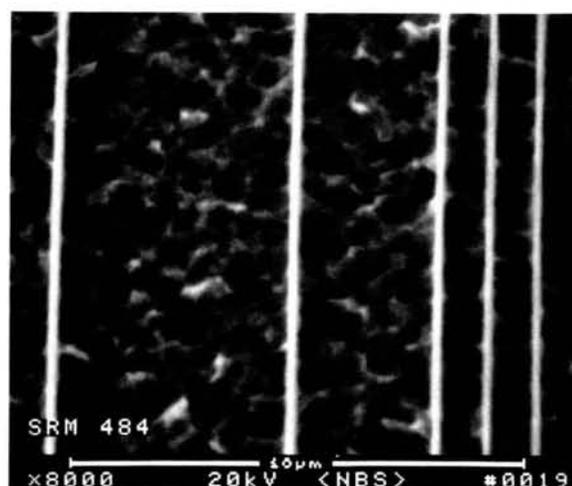
Magnification	Field Size (μm)	Number Of Fields	Pixel Size (nm)		Overall Data Acquisition Time (hr)		
			512	1028	Frame Rate (s)		
					1.0	0.5	0.25
10,000	10	1000×1000	19.5	9.7	277.0	139.0	69.5
5,000	20	500×500	39.1	19.5	69.0	35.0	17.5
2,500	40	250×250	78.1	38.9	17.4	8.7	4.4
1,250	80	125×125	156.2	77.8	4.3	2.2	1.1
600	160	63×63	312.5	155.6	1.1	0.55	0.28

handling techniques necessary for this work will need to be developed in order that the SEM instrumentation compete with the throughput of present optical inspection instruments.

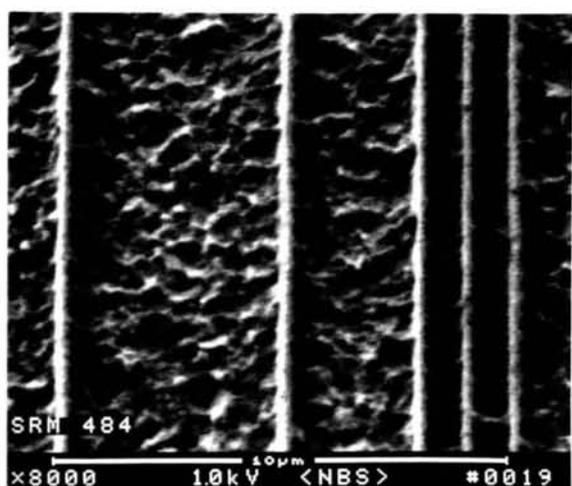
SEM Measurement Standards

A major project being undertaken at the National Bureau of Standards at the present time is the development of national standards for SEM linewidth metrology. The only magnification standard reference material (SRM) presently available for calibrating scanning electron microscopes is, SRM 484. This standard has served well for several years and is still useful for many SEM applications, but it was developed prior to the recent interest in low accelerating voltage operation and wafer inspection. SRM 484, in its present form, is unsuitable for use in new SEM inspection instruments for two main reasons: a lack of suitable contrast in the 1.0 keV accelerating voltage range and the overall size which is not compatible with newly introduced wafer inspection instrumentation. Presently, a project has been initiated at NBS to physically modify this sample without altering its calibration or certification procedures to make it suitable for low accelerating voltage operation (figs. 19a and 19b). The linewidth measurement standard developed for the optical microscope SRM 474 is not designed or recommended for use in the SEM and it should not be used for this purpose [42].

The optical theory and modeling for the SRM 474 is not directly adaptable to the SEM and therefore the criteria used to determine the edge location is not applicable and should not be considered as such. From the above discussions of the electron beam effects and the requirement for modeling, this should be apparent as the two types of instruments are totally independent of each other in both the



a



b

Figure 19—Scanning electron micrographs of the SEM magnification standard SRM 484 following the procedure used to enhance the contrast of the sample for low accelerating voltage use. (a) 20 keV accelerating voltage (b) 1.0 keV accelerating voltage.

underlying physics and in operation. SRM 474 could, however, be used to measure pitch at low accelerating voltage under conditions where the sample is not charging. However, such use may damage the SRM (e.g., contamination) and render it useless for optical microscopy. In this mode, the magnification of the instrument could be calibrated to pitch. However, again the reader is warned that continuing this adjustment process to include linewidth measurements is not recommended, as a general calibration procedure, because the edge criterion so obtained would only be valid for a similar chrome-on-glass mask.

For the present time, product precision is a prime concern to the semiconductor industry, and until such national standards for SEM linewidth measurement on integrated circuit wafers are available, the best that can be done now is the development of a series of internal "golden" samples within a particular organization for each level of processing [43]. The development of such samples referenced between the SEM and the optical microscope has been discussed [1]. Using the established national standards to properly adjust the magnification of an instrument, this series of well characterized internal standards is then used to develop offsets to the instrument for each level and also to periodically check the measurement drift of the instrument.

Conclusions

Proper metrology with any type of instrument is not a trivial matter, the SEM is no different. For the precise metrology required in the manufacture of integrated circuits for submicrometer processing, an understanding of the areas that can be a problem associated with the scanning electron microscope is even more important than in any other commercial application of this instrument. The uncertainties associated with each instrument in each environment must be assessed and understood for proper metrology to be done. It has been our goal in this paper to outline some of these problems to the reader in order to put into perspective what can actually be expected from this type of instrumentation at this time. We are confident that given the necessary attention, the SEM can do the job required. As this instrument matures further in this field and research is done to improve the theoretical understanding of the physical processes going on in this instrument, the entire field of scanning electron microscopy in all its diverse applications will be furthered.

The authors would like to acknowledge and thank Mr. William Keery for his assistance in setting up some of the experimental examples used and his suggestions during the preparation of this manuscript. We would also like to thank Drs. Don Atwood, Robert Larrabee, Steve Erasmus, and Diana Nyssonen for useful technical discussions and critical review of this manuscript. NBS also acknowledges the co-sponsoring of the low acceleration voltage standard research by Harry Diamond Laboratories, Adelphi, MD (MIPR no. R86-086).

References

- [1] Nyssonen, D. E., and R. D. Larrabee, Submicron linewidth measurement in the optical microscope, *Solid State Technology* (in press).
- [2] Ohtaka, T.; S. Saito, T. Furuya and O. Yamada, Hitachi S-6000 field emission CD-measurement SEM, *SPIE Proceedings* **565** 205-208 (1985).
- [3] Russell, P. E.; T. Namac, M. Shimada, and T. Someya. 1984. Development of SEM-based dedicated IC metrology system. *SPIE Proceedings* 480:101-108.
- [4] Yamaji, H.; M. Miyoshi, M. Kano, and K. Okumura, High accuracy and automatic measurement of the pattern linewidth on very large scale integrated circuits, *SEM/1985/1 SEM Inc.* 97-102.
- [5] ASTM Manual on presentation of data and control chart analysis, STP 15D American Society for Testing and Materials, 1916 Race Street, Philadelphia, PA 19103.
- [6] Nyssonen, D., and M. T. Postek, SEM-based system for the calibration of linewidth SRM's for the IC industry, *SPIE Proceedings* **565** 180-186 (1985).
- [7] Joy, D. C., Resolution in low voltage scanning electron microscopy, *J. Microsc.* **140** 283-292 (1985).
- [8] Joy, D. C., Image modeling for SEM-based metrology, *Proceeding Joint Annual Meeting EMSA/MAS* (G.W. Bailey, ed.) San Francisco Press 650-651 (1986).
- [9] Goldstein, J. I.; D. E. Newbury, P. Echlin, D. C. Joy, C. Fiori, and E. Lifsin, *Scanning Electron Microscopy and X-Ray Microanalysis*, Plenum Press: New York 673 pp. (1984).
- [10] Erasmus, S., and K. C. A. Smith, An automatic focusing and astigmatism correction system for the SEM and CTEM, *J. Microsc.* **127** 185-199 (1982).
- [11] Smith, K. C. A., On-line digital computer techniques in electron microscopy, general introduction, *J. Microsc.* **127** 3-16 (1982).
- [12] Newbury, D. E.; D. C. Joy, P. Echlin, C. Fiori, and J. I. Goldstein, *Advanced Scanning Electron Microscopy and X-Ray Microanalysis*, Plenum Press: New York 454 pp. (1986).
- [13] Postek, M. T.; K. S. Howard, A. Johnson, and K. McMichael, *Scanning Electron Microscopy—A Students Handbook*, Ladd Research Industries, Burlington VT. 305 pp. (1980).
- [14] Wells, O.C., *Scanning Electron Microscopy*, McGraw-Hill: New York 421 pp. (1984).

- [15] Postek, M. T., The scanning electron microscope in the semiconductor industry, *Test and Measurement World* 54-75 (1983).
- [16] Kanaya, K., and S. Okayama, Penetration and energy-loss theory of electrons in solid targets, *J. Phys. D. Appl. Phys.* 5 43-58 (1972).
- [17] Drescher, H.; L. Reimer and H. Seidel, Rückstreuoeffizient und Sekundärelektronen-Ausbeute von 10-100 keV-electronen und Beziehungen zur Raster-elektronenmikroskopie. *Z. f. angew. Physik* 29 331-336 (1970).
- [18] Volbert, B., and L. Reimer, Advantages of two opposite Everhart-Thornley detectors in SEM. *SEM/1980/IV SEM, Inc.* 1-10.
- [19] Keery, W. J.; K. O. Leedy and K. F. Galloway, Electron beam effects on microelectronic devices. *SEM/1976/IV IITRI Research Institute, Chicago Ill.* 507-514.
- [20] Erasmus, S., Reduction of noise in TV rate electron microscope images by digital filtering, *J. Microsc.* 127 29-37 (1982).
- [21] Rez, P. 1983. A transport equation theory of beam spreading in the electron microscope. *Ultramicroscopy* 12:29-38.
- [22] Joy, D. C., A model for calculating secondary and backscattered electron yields, *J. Microsc.* (in press).
- [23] Lin, Y.-C., Alignment signals from electrons scattering near an edge for electron beam microfabrication, PhD Thesis, University of California, Berkeley 192 pp. (1981).
- [24] Hembree, G. G.; S. W. Jensen and J. F. Marchiando, Monte Carlo simulation of submicrometer linewidth measurements in the scanning electron microscope, *Microbeam Analysis 1981* (Roy Geiss, ed.) San Francisco Press 123-126 (1981).
- [25] Joy, D. C., Beam interactions, contrast and resolution in the SEM, *J. Microsc.* 136 241-258 (1984).
- [26] Jensen, S., and D. Swyt, Sub-micrometer length metrology: problems techniques and solutions, *SEM/1980/I SEM Inc.* 393-406 (1980).
- [27] Jensen, S. 1980, Planar metrology in the SEM. *Microbeam Analysis* (David Wittry, ed.) San Francisco Press 77-84 (1980).
- [28] Seiler, D. G., and D. V. Sulway, Precision linewidth measurement using a scanning electron microscope, *SPIE Proceedings Vol. 480* 86-93 (1984).
- [29] The official international standard for the meter is the distance travelled by light in a vacuum during $1/299,792,458$ second. The official international standard for the second is defined as the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the Cesium-133 atom.
- [30] Monahan, K., and D. S. Lim, Nanometer-resolution SEM metrology in a hostile laboratory environment, *SPIE Proceedings* 565 173-175 (1985).
- [31] Pawley, J. B., Strategy for locating and eliminating sources of mains frequency stray magnetic fields, *Scanning* 7 43-46 (1985).
- [32] Pawley, J. B., Low voltage scanning electron microscopy, *J. Microsc.* 136 45-68 (1984).
- [33] Echlin, P., Contamination in the scanning electron microscope. *SEM/1975/III IITRI Research Institute Chicago, Ill* 679-686 (1975).
- [34] Anger, K.; B. Lischke and M. Sturm, Material surfaces for electron-optical equipment, *Scanning* 5 39-44 (1983).
- [35] Brunner, M., and R. Schmid, Charging effects in low-voltage SEM metrology, *SEM/1986 SEM Inc.* 377-382 (1986).
- [36] Frosien, J., and B. Lischke, *Micrometrology With Electron Probes Microcircuit Engineering 85*, Academic Press: London 441-450 (1985).
- [37] Postek, M. T., Low accelerating voltage inspection and linewidth measurement in the scanning electron microscope, *SEM/1984/III SEM Inc.* 1065-1074 (1985).
- [38] Postek, M. T., Electron detection modes and their relation to linewidth measurement in the scanning electron microscope, *Proceeding Joint Annual Meeting EMSA/MAS* (G. W. Bailey, ed.) 646-649 (1986).
- [39] Erasmus, S., Damage to resist structures during SEM inspection, *J. Vac. Sci. Technol.* B5(1) 409-413 (1987).
- [40] Samoto, N.; R. Shimizu and H. Hashimoto, Changes in the volume and surface compositions of Polymethylmethacrylate under electron beam irradiation in lithography, *Jap. J. Applied Phys.* 24 482-486 (1985).
- [41] Kotani, H.; M. Kawabe and S. Namba, Direct writing of gratings by electron beam in Poly(methyl methacrylate) wave guides, *Jap. J. Applied Phys.* 18 279-283 (1979).
- [42] The handbook accompanying NBS Standard Reference Material 474 or 475.
- [43] Stein, R.; D. H. Cummings and J. Schaper, Calibrating microscopic linewidth measurement systems, *Semiconductor International* 132-136 (1986).

Instrument-Independent CAD Spectral Databases: Absolute Cross-Section Measurements In QQQ Instruments

Volume 92

Number 3

May-June 1987

Richard I. Martinez,
Seksan Dheandhanoo

National Bureau of Standards
Gaithersburg, MD 20899

The energy dependence of the cross section, $\sigma(E)$, for the symmetric (resonant) charge transfer reaction $\text{Ar}^+(\text{Ar},\text{Ar})\text{Ar}^+$ was measured in our triple quadrupole (QQQ) tandem mass spectrometer.¹ Our $\sigma(E)$, for $P \approx 0.04 - 0.43$ mtorr and $E \approx 5 - 60$ eV (LAB) [the range of collision energies

used for collisionally activated dissociation (CAD)], agrees to within 10% with the Rapp-Francis theory (impact parameter method in the two-state approximation), as corrected by Dewangan. We measured identical $\sigma(E)$ from both the rate of reactant ion decay and the rate of product ion formation; i.e., our instrument is kinetically well behaved. The measurement of these $\sigma(E)$ in other QQQ instruments can be used to validate whether or not a QQQ instrument has been properly designed to be kinetically well behaved. This is essential if generic, instrument-independent CAD spectral databases are to be developed on the basis of the absolute cross sections for the CAD of *known* ionic substructures. That is, since tandem mass spectrometry (MS/MS) exploits the ion fragmentation patterns characteristic of

ionic substructures, the characteristic profiles ["breakdown curves"] of ion abundance versus target thickness (or collision energy) correspond uniquely to the sequence: (parent), $\xrightarrow{\sigma_{ij}}$ (daughter), $\xrightarrow{\sigma_{jk}}$ (granddaughter)_i, etc. Hence, computer simulation of experimentally observed breakdown curves enables the structure of an *unknown* species to be assigned on the basis of the absolute cross sections σ_{ij} , σ_{jk} , etc. for CAD of *known* ionic substructures i, j, k , etc. Thus, if the calculated and experimental breakdown curves agree, the structure would be characterized.

Key words: calibration; cross sections; tandem mass spectrometry; target thickness.

Accepted: February 5, 1987.

Introduction

Triple quadrupole (QQQ) tandem mass spectrometry (MS/MS) is an analytical tool which can be used for rapid, direct speciation of complex multicomponent mixtures [1].² The analysis makes use of the collisionally-activated dissociation (CAD) of "parent" ions.³ A "parent" ion selected by the first quadrupole (Q1) is interacted with a target gas

within the second quadrupole (Q2). Q2 channels undissociated "parent" ions and "progeny" fragment ions into the third quadrupole (Q3) for mass analysis. The instrument thus produces a CAD spectrum of each initially-selected "parent" ion.

In principle, standard CAD spectra of a variety of ions (fragment ions, molecular ions, protonated molecules, etc.) could be generated and collected as reference libraries, to be used for comparison

¹Standard physics notations: $\text{A}^+(\text{B,C})\text{D}^+$ represents the reaction $\text{A}^+ + \text{B} \rightarrow \text{C} + \text{D}^+$; σ_E represents the value of the reaction cross section when measured at a particular interaction (collision) energy E ; $\sigma(E)$ is the functional form of the energy dependence observed when σ_E values are plotted versus their respective E values.

²Figures in brackets indicate literature references.

³A "parent" ion may be a molecular radical cation, a protonated molecule, or a "progeny" fragment ion (daughter, granddaughter, etc.) produced by the CAD of a larger precursor parent ion.

About the Authors: Richard I. Martinez, a research chemist, is with the NBS Center for Chemical Physics. Seksan Dheandhanoo, a guest scientist at NBS from Georgetown University during the work described, is a physicist. The work was funded by the U.S. Air Force, Environics Division.

against unknown spectra in a manner analogous to the use of reference libraries in the data handling systems of ordinary electron impact mass spectrometry. Further, it should be possible to infer the identity of an unknown complex molecule by identifying the ionic substructures of fragment ions generated in its CAD spectrum. However, to date reference libraries of CAD spectra have not been collected because of a lack of standardization of operating conditions of such instruments [2].

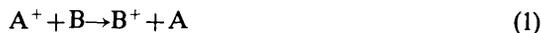
There are several instrument parameters which can cause significantly different CAD spectra to be observed for any given molecule. The key parameters are: 1) the number of collisions undergone by a "parent" ion within Q2, a parameter usually characterized in terms of "target thickness," which is defined as [(actual path length traversed by the ion through the gas target) \times (effective number density of the CAD target gas)]; 2) the duration of the interaction between the "parent" ion and the target gas, which is determined by the collision energy for "parent" ions entering Q2; and 3) the energy level of the analyzing quadrupole Q3 relative to that of Q2 which, because of the translational energy distribution of the "progeny" ions, determines whether or not some progeny ions can enter Q3.

Results of a recent international round robin [2] demonstrated that the target thickness is not a well-controlled parameter, with estimated target thicknesses differing by factors of 2-4 from apparent actual values. The problem of determining target thickness is complicated in QQQ instruments because of the complex oscillatory trajectories of ions within a quadrupole mass filter [3-7]; the actual path length traversed by the ion through the CAD gas can be significantly longer than the nominal gas target length [6]. Moreover, in QQQ instruments utilizing a molecular beam target (Type A configuration [2]) the problem is further complicated because of a lack of information about the extent of overlap of the projectile ion beam and the molecular beam target. On the other hand, in QQQ instruments utilizing a collision chamber (Type B configuration [2]), the actual target thickness can be significantly greater than an estimated value based on the length of the Q2 collision chamber and the pressure within it if the gas plume extends beyond the confines of the Q2 collision chamber into Q1 and Q3.

Kinetic Method

In a recent study from this laboratory [8] it was suggested that these problems can be circumvented by using a kinetic method to measure the effective target thickness within a QQQ instrument. That is,

if a reaction can be identified for which the cross section (or rate coefficient) is well established as a function of collision energy, then a simple measurement of the intensity of the reactant ion and/or product ion in the absence and presence of CAD target gas at known collision energy leads to an experimental determination of the target thickness. For example, for the charge transfer reaction:



under pseudo-first order conditions ($[B] \gg [A^+]$),⁴

$$Ln Y \equiv Ln [A^+]_0 / [A^+] = \sigma_E L_{eff} [B] \equiv \beta P_B \quad (2)$$

where σ_E is the value of the reaction cross section at a collision energy E , L_{eff} = effective path length of the oscillatory trajectory traversed by a projectile ion through the CAD target gas, $L_{eff} [B]$ = effective target thickness for A^+ in B, β = proportionality constant, and P_B = pressure of target gas B corresponding to $[B]$. Hence, in the absence of other loss processes for A^+ , measurement of $Ln [A^+]_0 / [A^+]$ provides in-situ calibration of the effective target thickness if σ_E is known. Moreover, if there are no other production processes for B^+ , if there is no mass discrimination within the QQQ mass filters between the m/z of A^+ and the m/z of B^+ , and if the ion collection efficiency approaches 100%, then $[B^+]_\infty = [A^+]_0$, and

$$Ln W = \sigma_E L_{eff} [B] \equiv \beta P_B \quad (3)$$

where $W = [B^+]_\infty / \{[B^+]_\infty - [B^+]\} \equiv [A^+]_0 / \{[A^+]_0 - [B^+]\}$. Hence, obtaining the same result from reactant ion loss and product ion formation experiments (i.e., $Ln Y$ and $Ln W$ measurements, respectively) provides strong assurance that a QQQ instrument is kinetically well behaved. That is, it provides a very important test that the instrument parameters and the reaction kinetics are well controlled (no back reactions, no impurity reactions, no scattering losses, no fringing fields, well-confined gas target, etc.).

In our earlier study [8], the symmetric (resonant) charge transfer reaction $Ne^+(Ne, Ne)Ne^+$ was used as a calibrating reaction for the validation of the target thickness measurements in our QQQ instru-

⁴Standard kinetic notation: $[A^+]_0$ and $[A^+]$ are, respectively, the intensities of the reactant ion A^+ when measured in the absence and presence of CAD target gas; $[B^+]$ is the intensity measured for the product ion B^+ when the target thickness is that used for the $[A^+]$ measurement; $[B^+]_\infty (= [A^+]_0)$ is the intensity of the product ion at "infinite" reaction time and/or target thickness when all of A^+ has been converted to B^+ .

ment. Abundant experimental and theoretical results had been previously reported for this reaction. Furthermore, because the NBS instrument had been constructed to incorporate the design considerations detailed by Dawson and coworkers [3-7], eq (4) [6] could be used to estimate $L_{\text{eff}} = R L_{\text{actual}}$.

$$R = [1 + (0.0738 r_0^2 F^2 M / E)]^{0.5}. \quad (4)$$

Here L_{actual} is the actual rectilinear pathlength for a well-confined CAD gas target; M = mass of projectile ion (in amu), E = axial ion energy (in eV), r_0 = field radius (in cm), F = rf frequency (in MHz). Equation (4) is based on operation of Q2 with the Mathieu parameters [3,4] at $a_2=0$, $q_2=0.28$ [6].⁵ It was shown [8] that when the effective target thickness was estimated by using eq (4), values for the absolute reaction cross section derived from eq [2] were in excellent agreement with theoretical predictions, as well as with previous experimentally-determined values. Furthermore, identical values for the reaction cross section were derived from

⁵It is important to note that eq (4) depends on M/E . Hence, for a given collision energy E , the effective target thickness L_{eff} [B] will be different for different projectile ions, and must be corrected accordingly.

reactant ion loss [eq (2)] and product ion formation [eq (3)] experiments, thus confirming that the NBS instrument is kinetically well behaved.

This paper reports results of an analogous exercise carried out using the $^{40}\text{Ar}^+(\text{Ar},^{40}\text{Ar})\text{Ar}^+$ reaction⁶ for $Ln Y$ measurements and the $^{36}\text{Ar}^+(\text{Ar},^{36}\text{Ar})^{40}\text{Ar}^+$ reaction for $Ln W$ measurements.⁶ The $\text{Ar}^+(\text{Ar},\text{Ar})\text{Ar}^+$ reaction is of special interest because argon is a target gas commonly used for CAD. Thus, this reaction may provide a convenient calibrant species for target thickness determinations in other laboratories. Since reference spectra for CAD libraries can be utilized only if they were obtained under conditions such that the target thickness is specified, the results reported here may permit the easy standardization of operating conditions for the determination of such reference spectra.

Experimental

Our specially designed QQQ instrument can be configured to use either a molecular beam (Type A) or collision chamber (Type B) configuration (see schematic, fig. 1). All experiments reported here utilized the Type B configuration.

⁶ ^{36}Ar is the sum of $^{40}\text{Ar} + ^{38}\text{Ar} + ^{36}\text{Ar}$. The natural abundance of ^{36}Ar ($0.3365 \pm 0.0006\%$ [9]) is sufficient to permit the measurement, with good signal-to-noise ratios, of the product ion growth ($Ln W$ measurements).

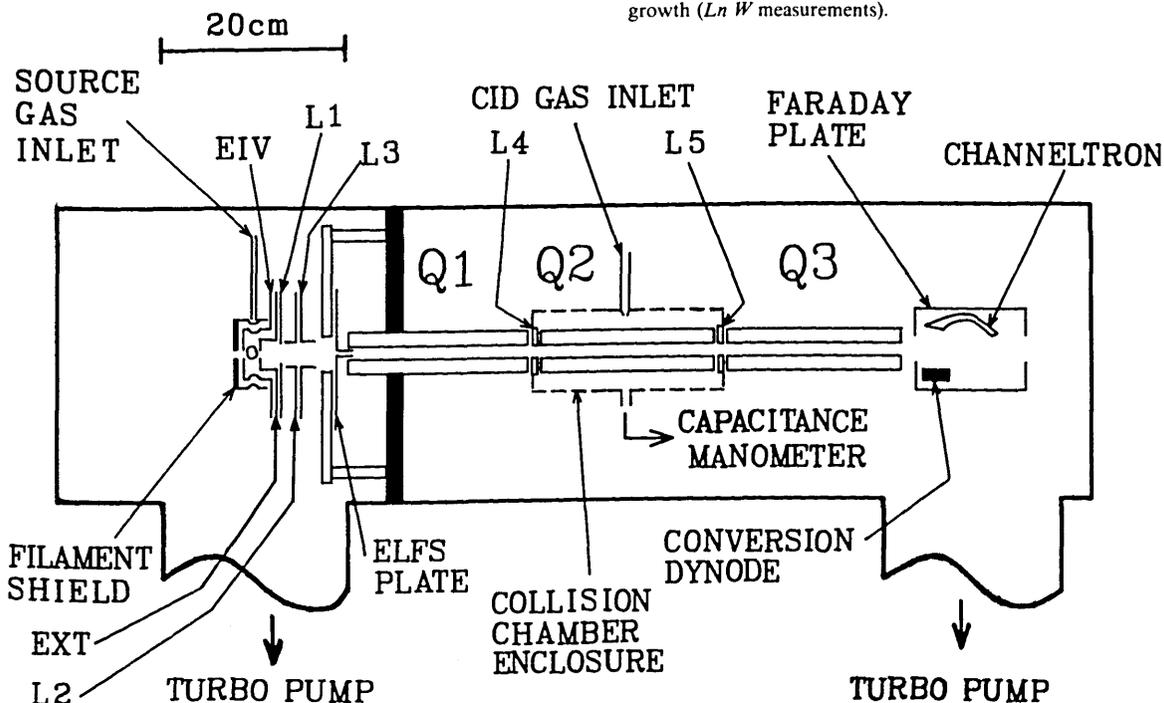


Figure 1-Schematic of QQQ instrument. EIV, EXT, L1-L5, etc. are ion optics lens elements; ELFSTM and CHANNELTRONTM are registered trademarks of Extrel and Galileo Electro-Optics, respectively.

An abbreviated description of the instrument follows (a detailed description will be published elsewhere [10]). The instrument was manufactured by Extrel, Inc.⁷ to conform to the design considerations stipulated by Dawson and coworkers [3-7]. It consists of three standard 7-270-9 quadrupole rod assemblies (Q1, Q2, Q3) mounted in tandem on a special multipurpose track. Each mass filter assembly is operated at 1.2 MHz, controlled by a 300-watt Model 150-QC quadrupole power supply and associated quadrupole control electronics. A C-50-IC controller regulates the standard Extrel electron impact ionizer mounted on the differential pumping wall. This ionizer has a filament perpendicular to the cylindrical quadrupole axis and has been modified to accommodate crossed molecular and laser beams. Each QQQ system parameter is computer controlled via its respective 16-bit DAC by the standard 8086-based Extrel Triple Quad Data System used for instrument control and data acquisition.

For the Type B configuration, Q2 is surrounded by a collision chamber enclosure while Q1 and Q3 are completely nude (no housing), and are adequately pumped by four 1200 l/s turbomolecular pumps, ensuring a well-confined collision region. The actual length of the collision region from the front face of the L4 aperture to the rear face of the L5 aperture is $L_{\text{actual}} = 21.74_5 \pm 0.07_5$ cm. All kinetic measurements were based on operation of Q2 with the Mathieu parameters [3,4] at $a_2 = 0$, $q_2 = 0.28$ [6]. For our instrument, $r_0 = \text{field radius} = 0.684$ cm (quadrupole rod diameter = 1.59 cm), $F = \text{rf frequency} = 1.2$ MHz, and the R correction factor from eq (4) is ca. 1.02 at $E = 60$ eV and 1.18 at $E = 5$ eV. Furthermore, the diameter of our L4 and L5 inter-quadrupole lens apertures is 1.27 ± 0.025 cm $\{ > 1.4r_0 \}$ [6], and thus conforms to the requirements for closely-coupled quadrupole fields [6]. Pressure measurements in the center of the collision chamber were made with a 1 torr MKS 310CA Baratron capacitance manometer [appropriate corrections were made for thermal transpiration ($\approx 3\%$) etc.].

Ar^+ ions were generated by 70 eV electron impact [11], and the Ar^+ projectiles were selected by Q1 [19]. The energy spread of the projectiles entering Q2 was determined to be ≤ 1.8 eV for 90% of

the ions [≤ 3 eV for 99% of the ions] when measured by using the Q2 pole bias (rod offset) to generate a stopping potential curve (see fig. 2). $E_{90\%}$ is the Q2 potential required to stop 90% of the ions. The collision energy E_{coll} was selected by setting the Q2 pole bias = $E_{90\%} - E_{\text{coll}}$.

Projectile decay experiments (cf. fig. 3) were performed at each selected collision energy by setting the Q3 pole bias more positive relative to the Q2 pole bias (e.g., $Q3 - Q2 \approx 3$ to 40 V for $E_{\text{coll}} \approx 5$ to 60 eV) to ensure only *unreacted* projectiles were able to enter Q3 [25]. Product growth experiments (cf. fig. 4) were performed by setting the Q3 pole bias sufficiently negative relative to the Q2 pole bias (e.g., $Q2 - Q3 \approx 110$ to 140 V for $E_{\text{coll}} \approx 40$ to 10 eV) to ensure that all ions (products *and* unreacted projectiles) were drawn out of Q2 into Q3 [25]. The typical ion collection efficiency is $\geq 97\%$; i.e., the total ion current for products + unreacted projectiles (i.e., with CAD gas on) $\geq 97\%$ of the initial projectile ion current (i.e., with CAD gas off). This high ion collection efficiency allows one to set

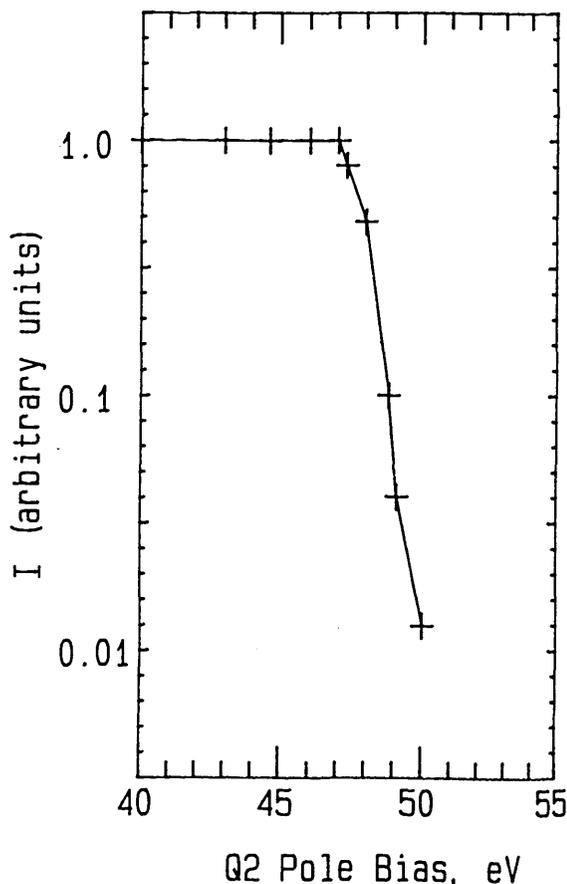


Figure 2—Energy distribution of Ar^+ projectiles entering Q2. I = Ion current in arbitrary units.

⁷Certain commercial equipment, instruments, and materials are identified in this paper in order to adequately specify the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the material, instruments, or equipment identified is necessarily the best available for the purpose.

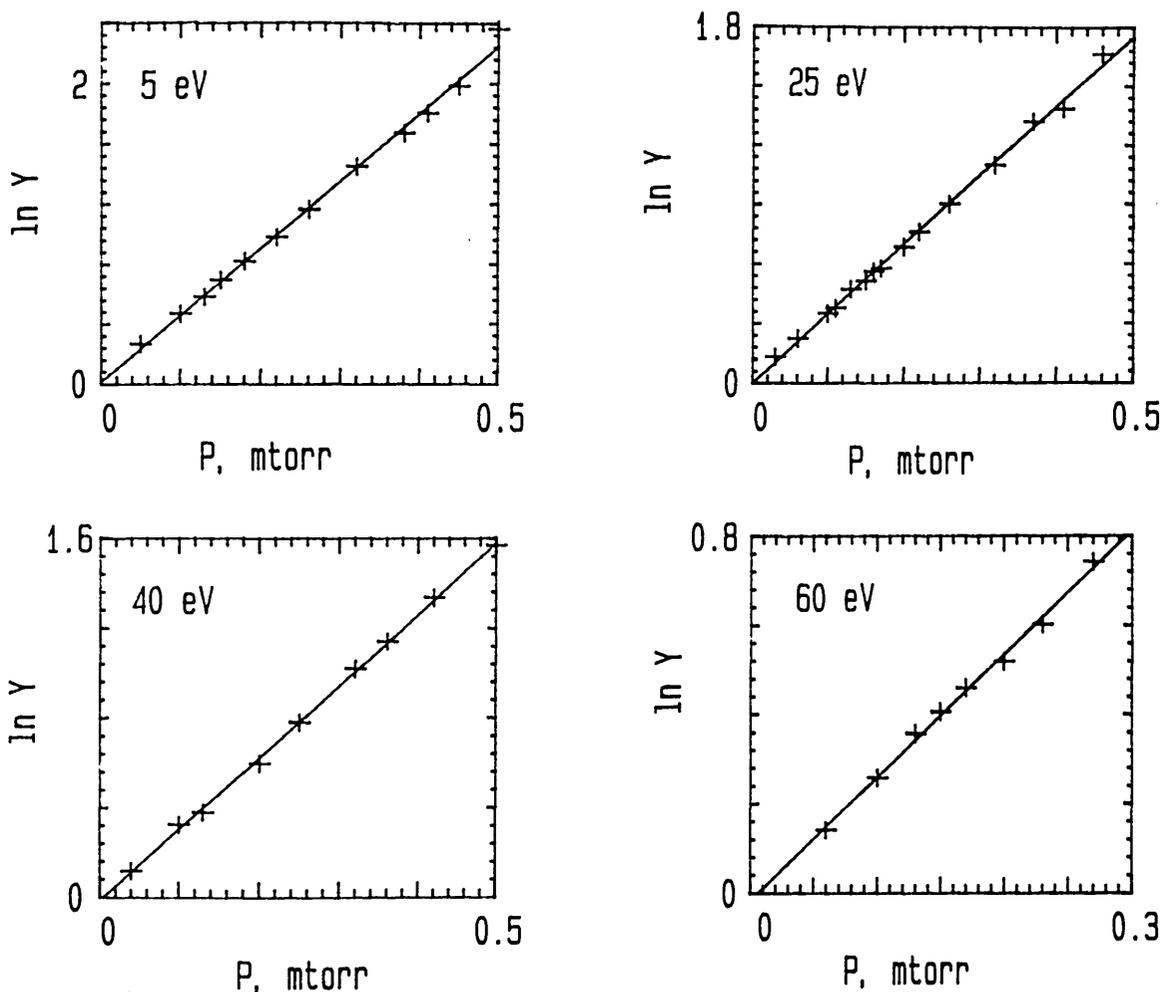


Figure 3—Projectile ion decay experiments. Plots of $\ln Y$ versus Ar target pressure P at fixed values of E_{coll} . $Y = [{}^{40}\text{Ar}^+]_0 / [{}^{40}\text{Ar}^+]_1$ and $\ln Y = k_F L_{\text{eff}} (m/2E_{\text{coll}})^{1/2} [Ar] = \sigma L_{\text{eff}} [Ar] \equiv \beta P$, L_{eff} —effective path length traversed by ion within Q2 collision chamber (corrected for rf effects [6]). m —mass of ${}^{40}\text{Ar}^+$ projectile.

$[{}^{40}\text{Ar}^+]_{\infty} = [{}^{36}\text{Ar}^+]_0$ when used in $\ln W$ [as defined in eq (3) and in fig. 4]. For both types of experimental measurements [viz., projectile ion decay (i.e., $\ln Y$ measurements) and product ion growth (i.e., $\ln W$ measurements)], several CAD target gas pressures were used (see figs. 3 and 4).

Results

Figures 3 and 4 show typical data for projectile ion decay and product ion growth experiments, respectively. Here P is the total Ar target gas pressure ($P = P_{40} + P_{38} + P_{36}$; where P_{40} , P_{38} , and P_{36} are, respectively, the partial pressures of ${}^{40}\text{Ar}$, ${}^{38}\text{Ar}$, and ${}^{36}\text{Ar}$). The well-established isotopic abundance of ${}^{40}\text{Ar}$ ($99.6003 \pm 0.0006\%$ ${}^{40}\text{Ar}$; $0.0632 \pm 0.0001\%$

${}^{38}\text{Ar}$; $0.3365 \pm 0.0006\%$ ${}^{36}\text{Ar}$ [9]) was used to determine P_{40} from the measured P .

Figure 5 shows the energy dependent cross sections for $\text{Ar}^+(\text{Ar}, \text{Ar})\text{Ar}^+$ in the format commonly used for resonant charge transfer reactions; viz. $\sigma^{0.5}$ vs. $\ln v$, where v is the projectile ion velocity. For $E_{\text{coll}} \approx 5\text{--}60$ eV [corresponds to $v \approx 0.5\text{--}1.8$ ($\times 10^6$) cm s^{-1}], the $\sigma(E)$ shown as (●) in figure 5 were derived from $\ln Y$ vs. P measurements for the ${}^{40}\text{Ar}^+$ projectile ion reacting with ${}^{40}\text{Ar} + {}^{38}\text{Ar} + {}^{36}\text{Ar}$ in the target gas (see fig. 3). For $E_{\text{coll}} \approx 10$ and 40 eV, $\ln W$ vs. P_{40} measurements of the rate of production of ${}^{40}\text{Ar}^+$ in ${}^{36}\text{Ar} + ({}^{40}\text{Ar}, {}^{36}\text{Ar}){}^{40}\text{Ar}^+$ [see fig. 4] led to the $\sigma(E)$ shown in figure 5 as (○). These were substantially the same as the $\sigma(E)$ determined from the $\ln Y$ vs. P measurements for the ${}^{40}\text{Ar}^+$ projectile.

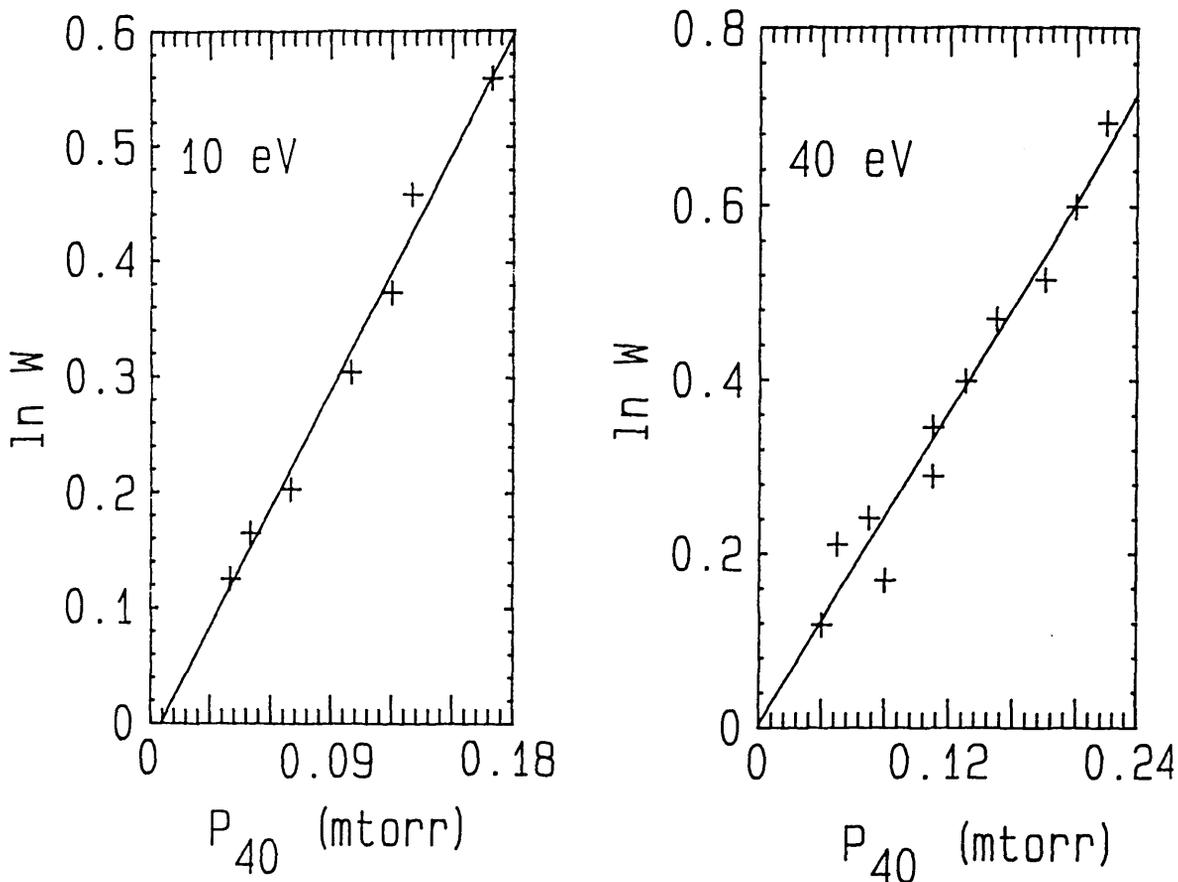


Figure 4—Product growth experiments. Plots of $\ln W$ versus P_{40} at fixed values of E_{coll} . $W = \frac{[^{36}\text{Ar}^+]_0}{\{[^{36}\text{Ar}^+]_0 - [^{40}\text{Ar}^+]\}}$ and $\ln W = \beta P_{40}$ where β is same as that of figure 3 and P_{40} is partial pressure of ^{40}Ar .

Discussion

Together with our results for $\text{Ar}^+(\text{Ar},\text{Ar})\text{Ar}^+$, figure 5 also summarizes experimental [28–41] and theoretical [42–47] results for this reaction from the literature [48]. Prior to our work it was not clear which theoretical model one could or should use to obtain reliable estimates of σ_E values for use in target thickness calibrations in the 5–60 eV range of collision energies, the range typically used for CAD experiments. The results reported here for $\sigma(E)$ (see fig. 5) are in excellent agreement with the $\sigma(E)$ predicted by the Rapp-Francis theory (impact parameter method in the two-state approximation) [42] as corrected by Dewangan [43] (solid line D in fig. 5), as well as with the experimental $\sigma(E)$ of other workers (see fig. 5, data labeled HES [34], Z [28], H [35], KPS [37], DSEG [29], FS [36]). For the data labeled C [31], the σ_E values are sig-

nificantly lower than those of the Dewangan line (labeled D) [43] and of other workers; however, the slope of his σ_E vs. E plot shows substantially the same $\sigma(E)$ as that of the Dewangan line. On the other hand, the $\sigma(E)$ of the data labeled HK [33] clearly differs from that of the Dewangan line and of other workers, even though some of the σ_E values labeled HK overlap some of the σ_E values of other workers. Hence, the data of figure 5 labeled C [31] and HK [33] are not considered further.

Our results show excellent agreement between the σ_E values derived from $\ln Y$ measurements (reactant ion loss; ● in fig. 5) and the corresponding values derived from $\ln W$ measurements (product ion formation; ○ in fig. 5). This concordance establishes 1) that our instrument is kinetically well behaved, and 2) the validity of Dawson's design considerations (closely-coupled quadrupole fields, properly filled acceptance, etc.) [3–7]. Similar agreement between $\ln Y$ and $\ln W$ measure-

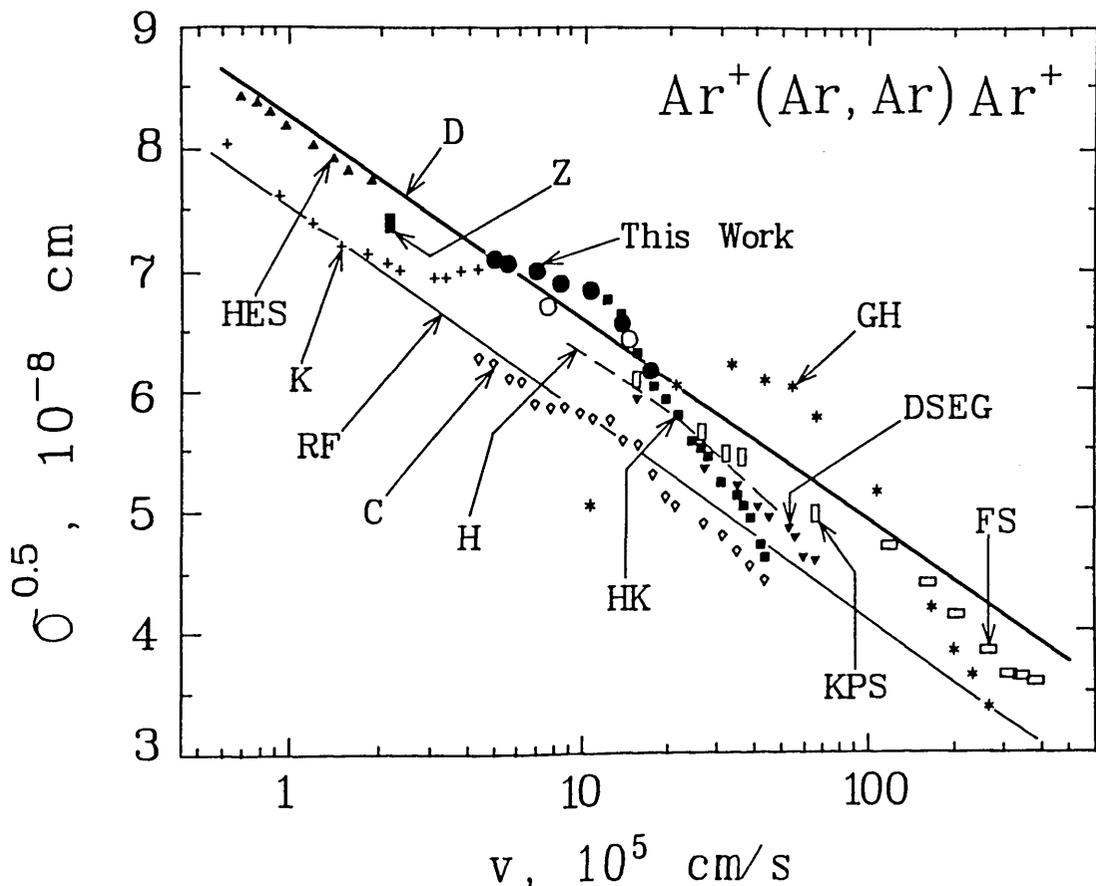


Figure 5—Plot of $\sigma^{1/2}$ (reaction cross section) $^{1/2}$ versus projectile ion velocity v . Comparison of our results with other workers and with theoretical models. ●—our projectile decay experiments, ○—our product growth experiments. Label [reference]: Z [28], DSEG [29], GH [30], C [31], K [32], HK [33], HES [34], H [35], FS [36], KPS [37], RF [42], D [43].

ments has also been observed in our Type B configuration for $\text{Ne}^+(\text{Ne}, \text{Ne})\text{Ne}^+$ [8] and $\text{Ar}^+(\text{N}_2, \text{Ar})\text{N}_2^+$ [49], further confirming that our instrument is kinetically well behaved. Thus we can use the σ_E values measured in our Type B configuration to determine the effective target thickness of Ar in our Type A configuration. However, similar performance is expected only in kinetically well behaved QQQ instruments which incorporate Dawson's design considerations [3-7].

Conclusions

The *kinetic method* described in the *introduction* potentially can provide a means whereby absolute target thicknesses for any gas can be accurately calibrated in-situ in kinetically well behaved QQQ instruments (in Type A or Type B configurations) for collision energies in the 5-60 eV range. Moreover, since the σ_E values for $\text{Ar}^+(\text{Ar}, \text{Ar})\text{Ar}^+$ are not strongly dependent on E over the range of in-

terest for CAD experiments $\{\sigma_{5eV} \approx 1.3 \sigma_{60eV}\}$, the kinetic method should provide fairly accurate target thickness calibrations even if the projectile energy distribution in other QQQ instruments is not as narrow as in the NBS instrument.⁸

⁸Interlaboratory round-robin testing of our kinetic method in various types of QQQ instruments is essential to confirm its reliability as a generic target thickness standard. Moreover, it will provide much-needed information about which QQQ instrument designs are not kinetically well behaved and therefore not well suited for the generation of standardized reference CAD spectra. The round-robin test will involve the experiment associated with figure 1(a) of [2] after first having completed in-situ target thickness calibrations of the participants' QQQ instruments by using our kinetic method with our $\sigma(E)$ for $\text{Ar}^+(\text{Ar}, \text{Ar})\text{Ar}^+$ (this work). A test protocol is being formulated. It will address how to set E_{coll} and q . Several investigators have agreed to participate. However, many more participants would be desirable to establish the degree of variability one encounters when using a standardized protocol with different operators on the same and/or different instruments of several types. Letters of inquiry from prospective participants may be sent to NBS.

The measurement of the $\sigma(E)$ for $\text{Ar}^+(\text{Ar},\text{Ar})\text{Ar}^+$ in other QQQ instruments can be used to validate whether or not a QQQ instrument has been properly designed to be kinetically well behaved. This is essential if generic, instrument-independent CAD spectral databases are to be developed on the basis of the absolute cross sections for the CAD of *known* ionic substructures. That is, since MS/MS exploits the ion fragmentation patterns characteristic of ionic substructures, the characteristic profiles ["breakdown curves"] of ion abundance versus target thickness (or collision energy) correspond uniquely to the sequence: (parent)_i $\xrightarrow{\sigma_{ij}}$ (daughter)_j $\xrightarrow{\sigma_{jk}}$ (granddaughter)_k, etc. Hence, computer simulation of experimentally observed breakdown curves should enable the structure of an *unknown* species to be assigned on the basis of the absolute cross sections σ_{ij} , σ_{jk} , etc., for CAD of *known* ionic substructures *i*, *j*, *k*, etc. Thus, if the calculated and experimental breakdown curves agree, the structure would be characterized. Dawson, et al. [50] demonstrated that computer simulation of breakdown curves is plausible. Hence, one can envision a CAD spectral database of critically-evaluated cross sections σ_{ij} , σ_{jk} , etc. for CAD of *known* ionic substructures measured in kinetically well-behaved instruments under standardized operating conditions. The advantages of such a database are: 1) the cross sections would uniquely characterize the CAD spectra of both known and unknown species (so long as the unknown species contain ionic substructures for which the CAD cross sections are known); 2) characterization of an unknown is *not* limited by the number of compounds in a "library"; 3) the format is compatible with its use in expert systems; and 4) end users are involved directly in its evolution by using critically-evaluated cross sections already in the database and by submitting new cross sections for inclusion in the database.

We gratefully acknowledge the many helpful discussions with Drs. Peter H. Dawson and Sharon G. Lias. S. Dheandhanoo is pleased to acknowledge Professor R. D. Bates, Jr. of Georgetown University for his support (DoC Grant No. 6H0613).

References

- [1] Yost, R. A., and D. D. Fetterolf, *Mass Spectrom. Revs.* **2** 1 (1983).
- [2] Dawson, P. H., and W.-F. Sun, *Int. J. Mass Spectrom. Ion Processes* **55** 155 (1983/1984).
- [3] Dawson, P. H., *Quadrupole Mass Spectrometry and Its Applications*, Elsevier, Amsterdam (1976).
- [4] Dawson, P. H., *Adv. Electronics Electron Phys.* **53** 153 (1980).
- [5] Dawson, P. H., *Int. J. Mass Spectrom. Ion Phys.* **20** 237 (1976).
- [6] Dawson, P. H., and J. E. Fulford, *Int. J. Mass Spectrom. Ion Phys.* **42** 195 (1982).
- [7] Dawson, P. H.; J. B. French, J. A. Buckley, D. J. Douglas, and D. Simmons, *Org. Mass Spectrom.* **17** 205 (1982).
- [8] Martinez, R. I., and S. Dheandhanoo, *Int. J. Mass Spectrom. Ion Processes* **74** 241 (1986).
- [9] Holden, N. E.; R. L. Martin and I. L. Barnes, *Pure & Appl. Chem.* **56** 675 (1984).
- [10] Martinez, R. I., *Rev. Sci. Instr.* (to be submitted).
- [11] Metastable ion production should account for less than 2% of the projectile ion beam [12] and should not be a problem [12-18].
- [12] Hagstrum, H. D., *Phys. Rev.* **104** 309 (1956).
- [13] Hagstrum, H. D., *J. Appl. Phys.* **31** 897 (1960).
- [14] Kadota, K., and Y. Kaneko, *J. Phys. Soc. Japan* **38** 524 (1975).
- [15] Kobayashi, N.; T. Matsuo and Y. Kaneko, *J. Phys. Soc. Japan* **49** 1195 (1980).
- [16] Matsuo, T.; N. Kobayashi and Y. Kaneko, *J. Phys. Soc. Japan* **50** 3482 (1981).
- [17] Rosner, S.D.; T. D. Gaily and R. A. Holt, *J. Phys. B: At. Mol. Phys.* **9** L489 (1976).
- [18] Durup, M., and G. Parlant, *Abst. 10th Int. Conf. Phys. Elect. Atom. Coll.* (North-Holland, Amsterdam 1977) p. 468.
- [19] For an Ar^+ projectile over the collision energy range $E_{\text{lab}} \approx 1-4000$ eV, $\sigma_{1/2}/\sigma_{3/2}$ is ca. 0.9-1.0 [20-22]. Hence, the total cross section measured for $\text{Ar}^+(\text{Ar},\text{Ar})\text{Ar}^+$ is not significantly affected by the composition of $\text{Ar}^+(^2P_{3/2})$ and $\text{Ar}^+(^2P_{1/2})$ in the projectile ion beam. Here, $\sigma_{3/2} = (\sigma_{3/2 \rightarrow 3/2} + \sigma_{3/2 \rightarrow 1/2})$, and $\sigma_{1/2} = (\sigma_{1/2 \rightarrow 1/2} + \sigma_{1/2 \rightarrow 3/2})$, where $\sigma_{3/2 \rightarrow 3/2}$, $\sigma_{3/2 \rightarrow 1/2}$, $\sigma_{1/2 \rightarrow 1/2}$, and $\sigma_{1/2 \rightarrow 3/2}$ are, respectively, the integrated cross sections for $\text{Ar}^+(^2P_{3/2})[\text{Ar},\text{Ar}]\text{Ar}^+(^2P_{3/2})$, $\text{Ar}^+(^2P_{3/2})[\text{Ar},\text{Ar}]\text{Ar}^+(^2P_{1/2})$, $\text{Ar}^+(^2P_{1/2})[\text{Ar},\text{Ar}]\text{Ar}^+(^2P_{1/2})$, and $\text{Ar}^+(^2P_{1/2})[\text{Ar},\text{Ar}]\text{Ar}^+(^2P_{3/2})$. Note also that $\sigma_m/\sigma_{3/2}$ is ca. 0.98-1.00 [22], where $\sigma_m = 1/3 \sigma_{1/2} + 2/3 \sigma_{3/2}$ if $\text{Ar}^+(^2P_{3/2})$ and $\text{Ar}^+(^2P_{1/2})$ are generated in the statistical 2:1 ratio. The fine-structure transitions $1/2 \rightarrow 3/2$ (0.178 eV exoergic) and $3/2 \rightarrow 1/2$ (0.178 eV endoergic) have been observed directly [23,24].
- [20] Hishinuma, N., *J. Phys. Soc. Japan* **32** 227 (1972).
- [21] Campbell, F. M.; R. Browning and C. J. Latimer, *J. Phys. B: At. Mol. Phys.* **14** 1183 (1981).
- [22] Liao, C.-L.; C.-X. Liao and C. Y. Ng, *J. Chem. Phys.* **82** 5489 (1985).
- [23] Itoh, Y.; N. Kobayashi and Y. Kaneko, *J. Phys. Soc. Japan* **50** 3541 (1981).
- [24] McAfee, K. B., Jr.; W. E. Falconer, R. S. Hozack, and D. J. McClure, *Phys. Rev.* **A21** 827 (1980).
- [25] These parameter settings are consistent with the knowledge that for $E_{\text{coll}} = 1-200$ eV (Lab) charge transfer reactions take place at large impact parameters with near-zero momentum transfer [26, 27, 41].

- [26] Kaiser, E. W.; A. Crowe and W. E. Falconer, *J. Chem. Phys.* **61** 2720 (1974).
- [27] Paulson, J. F.; F. Dale and S. A. Studniarz, *Int. J. Mass Spectrom. Ion Phys.* **5** 113 (1970).
- [28] Ziegler, B., *Z. Physik* **136** 108 (1953).
- [29] Dillon, J. A., Jr.; W. F. Sheridan, H. D. Edwards, and S. N. Ghosh, *J. Chem. Phys.* **23** 776 (1955).
- [30] Gilbody, H. B., and J. B. Hasted, *Proc. Roy. Soc.* **A238** 334 (1956).
- [31] Cramer, W. H., *J. Chem. Phys.* **30** 641 (1959).
- [32] Kobayashi, N., *J. Phys. Soc. Japan* **38** 519 (1975).
- [33] Hamilton, P. A., and P. F. Knewstubb, *Int. J. Mass Spectrom. Ion Processes* **57** 329 (1984).
- [34] Hegerberg, R.; M. T. Elford and H. R. Skullerud, *J. Phys. B: At. Mol. Phys.* **15** 797 (1982).
- [35] Hasted, J. B., *Proc. Roy. Soc.* **A205** 421 (1951).
- [36] Flaks, I. P., and E. S. Solov'ev, *Soviet Phys.—Tech. Phys.* **3** 564 (1958).
- [37] Kushnir, R. M.; B. M. Palyukh and L. A. Sena, *Bull. Acad. Sci. USSR Phys. Series* **23** 995 (1959).
- [38] Dawson, P. H., *Int. J. Mass Spectrom. Ion Processes* **63** 305 (1985).
- [39] Potter, R. F., *J. Chem. Phys.* **22** 974 (1954).
- [40] Neynaber, R. H.; S. M. Trujillo and E. W. Rothe, *Phys. Rev.* **157** 101 (1967).
- [41] Amme, R. C., and H. C. Hayden, *J. Chem. Phys.* **42** 2011 (1965).
- [42] Rapp, D., and W. E. Francis, *J. Chem. Phys.* **37** 2631 (1962).
- [43] Dewangan, D. P., *J. Phys. B: At. Mol. Phys.* **6** L20 (1973).
- [44] Johnson, R. E., *J. Phys. B: At. Mol. Phys.* **3** 539 (1970).
- [45] Duman, E. L., and B. M. Smirnov, *Teplofizika Vysokikh Temperatur* **12** 502 (1974).
- [46] Hodgkinson, D. P., and J. S. Briggs, *J. Phys. B: At. Mol. Phys.* **9** 255 (1976).
- [47] Dewangan, D. P., and S. B. Karmohapatro, *Phys. Letters* **29A** 115 (1969).
- [48] The data of [38–41] and the theoretical results of [44–47] are not shown in figure 5. With reference to figure 5, [38] $\{7.3 \times 10^{-8}$ cm at 4.1×10^5 cm s $^{-1}$ $\}$ and [41] both lie close to the theoretical line labeled D [43]; [39] lies approximately parallel to and about 8–14% lower than the data labeled C [31]; and [40] $\{6.9 \times 10^{-8}$ cm at 1.2×10^5 cm s $^{-1}$ $\}$ lies about 4% below the theoretical line labeled RF [42]. The theoretical results of [44] lie midway between those labeled RF [42] and D [43]; [45] and [46] agree with RF [42].
- [49] Martinez, R. I., and S. Dheandhanoo, *Int. J. Mass Spectrom. Ion Processes* (to be submitted).
- [50] Dawson, P. H.; J. B. French, J. A. Buckley, D. J. Douglas, and D. Simmons, *Org. Mass Spectrom.* **17** 212 (1982).

Note on the Choice of a Sensitivity Weight In Precision Weighing

Volume 92

Number 3

May-June 1987

R. S. Davis

National Bureau of Standards
Gaithersburg, MD 20899

Good weighing practice usually dictates that, when using double-substitution weighing to determine the mass difference between two weights, the nominal value of the sensitivity weight used to calibrate the optical scale of the mass comparator be at least four times greater than the difference of the two weights being compared. However, there are times when other considerations must override this rule. We examine the theoretical basis for the rule and the penalty for violating it. Finally, we propose a

modified weighing scheme which imposes a much less stringent rule for the size of the sensitivity weight. The new scheme requires an additional balance reading, but does not increase the overall measurement time significantly.

Key words: mass metrology; precision weighing; sensitivity weight; substitution weighing; transposition weighing; weighing.

Accepted: November 28, 1986

1. Introduction

Many precision mass comparisons, especially in the realm of metrology, still rely on mechanical balances. These balances may be either one-pan or two-pan. In both cases, however, weighing is done by double substitution between the unknown and an external standard. The procedure in use in most metrology laboratories is shown in table 1.

Table 1. Four observation scheme.

Operation	Load on Balance	Balance Indication
1	Y	I_1
2	X	I_2
3	$X+d$	I_3
4	$Y+d$	I_4

where Y represents the standard, X the unknown, and d the sensitivity weight. We are assuming that for two-pan balances double substitution has been used rather than double transposition. The arguments that follow apply with modification to the latter technique.

The difference in mass between Y and X , ΔM , (ignoring buoyancy corrections) is sometimes computed as [1]¹:

$$\Delta M = \frac{I_1 - I_2 - I_3 + I_4}{2(I_3 - I_2)} m_d = \frac{\Delta I}{\Delta I_d} m_d. \quad (1)$$

We may think of eq (1) as the product of the difference between Y and X in scale units, $(I_1 - I_2 - I_3 + I_4)/2$, multiplied by the balance sensitivity, $m_d/(I_3 - I_2)$. The sensitivity is the proportionality factor which converts differences in scale indication to units of mass. Here m_d is the known mass of d .

About the Author: R. S. Davis is a physicist in the Length and Mass Division of NBS' Center for Basic Standards.

¹Figures in brackets indicate literature references.

Most balance indications drift with time. Often the time dependence of the drift can be assumed to be linear. Based on this assumption, one usually tries to make the time intervals between the four weighing operations equal. If this is done, the estimate of ΔI , the difference between Y and X in scale units, will be unbiased by the drift. This will not be true of $I_3 - I_2$ however. The latter quantity estimates ΔI_d , the value of the sensitivity weight in scale units.

In order to remove the bias, a modified equation is used:

$$\Delta M = \frac{I_1 - I_2 - I_3 + I_4}{I_1 - 3I_2 + 3I_3 - I_4} m_d. \quad (2)$$

This is the equation found in the NBS MASSCODE [2] and has been advocated for general use if the added computational complexity can be handled by computer [3].

2. Variance of ΔM

There is a general rule [1,3] which states that the metrologist should take care that

$$\frac{\Delta I}{\Delta I_d} < 0.25. \quad (3)$$

If the rule is violated, the NBS MASSCODE prints a warning message along with the final calculation [2]. Since the author has not found a rigorous theoretical basis for the rule in the literature, one will now be given.

Each reading of scale indication is subject to random error. Let us assume this error can be characterized by a variance σ_I^2 which is the same for all measurements. Then the variance of ΔM as computed by eq (2) using first order propagation of error techniques is

$$\text{var}(\Delta M) = S^2 \sigma_I^2 \left[1 + 5 \left(\frac{\Delta I}{\Delta I_d} \right)^2 \right] \quad (4)$$

where $S = 2m_d / (I_1 - 3I_2 + 3I_3 - I_4)$ is the nominal value of the balance sensitivity (the quantity m_d is treated as a constant in this computation since its variance is usually much smaller than σ_I^2). Therefore, the rule represented by eq (3) implies that the variance in a single measurement of ΔM should not be allowed to increase by more than a factor of 1.31 above its minimum value. The choice of 1.31 is, of course, somewhat arbitrary. Reasonable people might all agree that a factor of 2, for instance, would be intolerably large, while a factor of 1

would be impractically small. We choose 1.31 because it is the *de facto* choice of the NBS MASSCODE. The important point is that we now have a rational criterion by which to compare various weighing procedures with respect to their demands on the value of the sensitivity weight.

The absence of a term linear in $\Delta I / \Delta I_d$ in eq (4) shows that the estimate of ΔI is uncorrelated with the estimate of ΔI_d . It is also evident that the variance of ΔM increases monotonically as the ratio $\Delta I / \Delta I_d$ becomes larger. In particular, if $\Delta I / \Delta I_d$ is of the order of 0.5 then the variance of ΔM increases to 2.25 times its minimum value. This is unacceptably large in many cases. A value of 0.5 for $\Delta I / \Delta I_d$ was the unavoidable case, however, for a series of important measurements made several years ago on our best kilogram comparator [4]. In order to cope with such a large value of $\Delta I / \Delta I_d$ it was necessary to use a modified weighing scheme.

3. The Five Observation Scheme

The weighing scheme used is identical to that of table 1 except for the addition of a fifth operation which is a repeat of the first.² The scheme is shown in table 2.

Table 2. Five observation scheme.

Operation	Load on Balance	Balance Indication
1	Y	I_1
2	X	I_2
3	$X + d$	I_3
4	$Y + d$	I_4
5	Y	I_5

The apparent difference in mass between Y and X is then estimated as follows:

$$\Delta M = \frac{I_1 - I_2 - I_3 + I_4}{-I_2 + I_3 + I_4 - I_5} m_d. \quad (5)$$

Equation (5) is also unbiased for a linear drift between measurements (though eq (5) is not the least squares solution for a linear drift model). The real virtue of eq (5) is that it is also an unbiased solution for a model which assumes only that the drift between operations 1 and 2 equals the drift between operations 3 and 4; and that the drift be-

² To the author's knowledge, the first reported use of this weighing scheme was in a 1967 paper by Bowman, Schoonover, and Jones [8]. These authors used a five-observation scheme to compare an external object with the built-in weights of a single-pan, mechanical balance.

tween operations 2 and 3 equals the drift between operations 4 and 5 [5]. The first drift occurs between operations which exchange the test weights on the balance pans. The second drift occurs when the sensitivity weight is added or removed. This model frees the operator from having to wait equal times between all measurements. Since the addition or removal of the sensitivity weight is a faster operation than the exchange of test weights, it is usually possible to accomplish the scheme of table 2 (where one need not wait equal intervals between operations) in about the same time as it takes to carry out the scheme of table 1 (where one must take measurements at equally spaced intervals).

When one computes the variance of ΔM based on eq (5) one discovers a remarkable result:

$$\text{var}(\Delta M) = S^2 \sigma_I^2 \left[1 - \frac{1}{2} \frac{\Delta I}{\Delta I_d} + \left(\frac{\Delta I}{\Delta I_d} \right)^2 \right] \quad (6)$$

where $S = 2m_d / (-I_2 + I_3 + I_4 - I_5)$.

The appearance of a term linear in $\Delta I / \Delta I_d$ indicates that, unlike eq (2), the estimate of ΔI in eq (5) is not independent of the estimate of ΔI_d . The result of a negative term in eq (6) is that the variance of ΔM is insensitive to the ratio $\Delta I / \Delta I_d$ for ratios between 0 and 0.5. Within this range, the variance of ΔM is actually below what it would be if the ratio $\Delta I / \Delta I_d$ were zero (table 3). The minimum

Table 3. Comparison of variances with respect to $\Delta I / \Delta I_d$ for results derived from eqs (1, 2, and 5).

$\Delta I / \Delta I_d$	$\text{var}(\Delta M) = k S^2 \sigma_I^2$		
	eq (1)	eq (2)	eq (5)
0	1.00	1.00	1.00
1/4	1.12	1.31	0.94
1/3	1.22	1.54	0.94
1/2	1.50	2.25	1.00
1	3.00	6.00	1.50

value for the variance of ΔM occurs for the ratio $\Delta I / \Delta I_d = 0.25$, although this minimum is only 6 percent below the variance for a ratio of zero. Finally, if we want to ensure the variance of ΔM not exceed $(1.31) \cdot S^2 \sigma_I^2$, we would make the rule that

$$\frac{\Delta I}{\Delta_d} < 0.86$$

This should be compared with eq (3).

4. Averaging

One of the ways to lessen the dependence of results obtained from table 1 on the ratio $\Delta I / \Delta I_d$ is by averaging. For N double substitutions at the same nominal load, one can average the N estimates of sensitivity and use the average value in the calculations of the various ΔM 's. The NBS MASSCODE takes this approach and amends the rule for the ratio of $\Delta I / \Delta I_d$ to:

$$\frac{1}{N^{1/2}} \frac{\Delta I}{\Delta I_d} < 0.25 \quad (7)$$

to cover cases where $N > 1$.

The amended algorithm leads to the following variance:

$$\text{var}(\Delta M) = S^2 \sigma_I^2 \left[1 + \frac{5}{N} \left(\frac{\Delta I}{\Delta I_d} \right)^2 \right]. \quad (8)$$

There are two possible objections to this approach. First, although the quadratic term in eq (8) is a factor of $1/N$ smaller than the same term in eq (4), it has been converted from a "within" to a "between-time" component [6]. Second, and more serious, the sensitivity of precision mechanical balances may be a function of time. This is certainly the case for NBS-2, the kilogram comparator which was designed and built at NBS and is now in use at the International Bureau of Weights and Measures (BIPM) [7]. In such cases, use of an average value for the sensitivity is unjustified.

5. Conclusion

The usual admonition that the ratio $\Delta I / \Delta I_d$ not exceed 0.25 ensures that the variance of a double substitution does not grow by more than 31 percent above its minimum value. We have examined a five-operation weighing scheme and have shown that use of this scheme relaxes the rule to the ratio $\Delta I / \Delta I_d$ not exceeding 0.86. We have also argued that the five-operation scheme can usually be performed in the same amount of time as the more usual four-operation scheme.

As a final comment, we emphasize that this analysis applies to un-servoed mechanical balances. For balances under servo control, the linear range of the scale is usually so large that it is never a problem to meet the conventional ratio rule. In addi-

tion, the sensitivities of servo-controlled balances are usually very stable over the course of a series of measurements.

References

- [1] Almer, H. E., Method of Calibrating Weights for Piston Gages. NBS Tech. Note 577 (1971 May).
- [2] Varner, R. N., and R. C. Raybold, National Bureau of Standards Mass Calibration Software. NBS Tech. Note 1127 (1980 July).
- [3] Jaeger, K. B., and R. S. Davis, A Primer for Mass Metrology. NBS Spec. Publ. 700-1 (1984 November).
- [4] Davis, R. S., Recalibration of the U.S. national prototype kilogram. *J. Res. Natl. Bur. Stand.* **90** (4) (1985 July–August) pp. 263–283.
- [5] Carré, P., and R. S. Davis, Note on weighings carried out on the NBS-2 balance. *J. Res. Natl. Bur. Stand.* **90** (5) (1985 September–October) pp. 331–339.
- [6] Pontius, P. E., Measurement Philosophy of the Pilot Program for Mass Calibration. NBS Tech. Note 288 (1966 May).
- [7] Almer, H. E., National Bureau of Standards kilogram balance NBS no. 2. *J. Res. Natl. Bur. Stand.* **76C** (1,2) (1972 January–June) pp. 1–10.
- [8] Bowman, H. A. and R. M. Schoonover (with appendix by M. W. Jones), Procedure for high precision density determinations by hydrostatic weighing. *J. Res. Natl. Bureau Stand.* **71C** (3) (1967 July–August) pp. 179–198.

Conferences / Events

MEASUREMENT UNCERTAINTIES: REPORT OF AN INTERNATIONAL WORKING GROUP MEETING

Experts representing the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), the International Organization of Legal Metrology (OIML), and the International Bureau of Weights and Measures (BIPM), met at the International Bureau of Legal Metrology in Paris, October 1–3, 1986, to initiate the development of a guidance document for the treatment and reporting of measurement uncertainties.

The need for such a guidance document has been long felt throughout the international measurement community. In 1980, the BIPM convened a meeting of experts from eleven national measurement laboratories for the purpose of arriving at a uniform and generally acceptable way of assigning uncertainties to measurement data. This BIPM Working Group on the Statement of Uncertainties agreed on a recommendation (Annex 1) which was subsequently adopted by the International Committee of Weights and Measures (CIPM) in October 1981 (Annex 2). The Recommendation consists of five points which provide a general philosophy for reporting uncertainties. In large part, the points are more in the nature of a briefly outlined approach, rather than an explicit specification of algorithms and methods. At the time of the formulation of the Recommendation, it was believed that many further details would have to be addressed and

resolved before the recommended approach could be routinely, uniformly, and widely used.

In the past year, the CIPM referred this matter to the ISO since it was felt that this was a more logical international body for trying to achieve agreement and uniformity on the statement of uncertainties within international standardization and metrology organizations. Responsibility was assumed by the ISO Technical Advisory Group (TAG) 4 since it serves as a coordinating mechanism for addressing measurement issues of common interest to the two worldwide standardization bodies, the ISO and the IEC, and the two worldwide metrology organizations, BIPM and OIML. The present working group (ISO TAG 4/WG3) was thus constituted under the terms of reference of ISO TAG 4, and consists of 11 experts nominated by the represented organizations. The Chairman of the working group is Dr. R. Collé of the National Bureau of Standards.

The terms of reference of the working group, as defined by the ISO TAG 4, is:

to develop a document based upon the recommendation of the BIPM Working Group on Uncertainty which provides guidance on the expression of measurement uncertainty for use within standardization, calibration, laboratory accreditation and metrology services. The purpose of such guidance is to promote full information on how uncertainty statements are arrived at and to provide a basis for the international comparisons of measurement results.

At the October meeting, the TAG 4 working group concluded that its task is to produce a document which will be firmly based on the BIPM recommendations of 1980, but will be more specific and usable. The document will be directed towards two primary user groups: national primary standards laboratories and secondary level standards

and calibration laboratories. The working group meeting resulted in the completion of a detailed outline for the organization and general contents of a guidance document, as well as a schedule and plan for producing a draft of this document. It is envisaged that a first draft of the document will be discussed at the next TAG 4/WG 3 meeting in May, 1987.

A complete report of the first meeting may be obtained from Mr. David E. Ederly, Standards Management Program, Building 101, Room A625, National Bureau of Standards, Gaithersburg, MD 20899.

ANNEX 1

RECOMMENDATION of the Working Group on the Statement of Uncertainties *presented to Comité International des Poids et Mesures*

Assignment of experimental uncertainties

RECOMMENDATION INC-1 (1980)

1 The uncertainty in the result of a measurement generally consists of several components which may be grouped into two categories according to the way in which their numerical value is estimated:

A - those which are evaluated by statistical methods,

B - those which are evaluated by other means.

There is not always a simple correspondence between the classification into categories A or B and the previously used classification into "random" and "systematic" uncertainties. The term "systematic uncertainty" can be misleading and should be avoided.

Any detailed report of the uncertainty should consist of a complete list of the components, specifying for each the method used to obtain its numerical value.

2 The components in category A are characterized by the estimated variances, s_i^2 , (or the estimated "standard deviations" s_i) and the number of degrees of freedom, ν_i . Where appropriate, the estimated covariances should be given.

3 The components in category B should be characterized by quantities u_j^2 , which may be considered as approximations to the corresponding variances, the existence of which is assumed. The quantities u_j^2 may be treated like variances and the quantities u_j like standard deviations. Where appropriate, the covariances should be treated in a similar way.

4 The combined uncertainty should be characterized by the numerical value obtained by applying the usual method for the combination of variances. The combined uncertainty and its components should be expressed in the form of "standard deviations."

5 If, for particular applications, it is necessary to multiply the combined uncertainty by a factor to obtain an overall uncertainty, the multiplying factor used must always be stated.

ANNEX 2

RECOMMENDATION CI-1981

The Comité International des Poids et Mesures

Considering

- the need to find an agreed way of expressing measurement uncertainty in metrology,
- the effort that has been devoted to this by many organizations over many years.
- the encouraging progress made in finding an acceptable solution, which has resulted from the discussions of the Working Group on the Expression of Uncertainties which met at BIPM in 1980.

Recognizes

- that the proposals of the Working Group might form the basis of an eventual agreement on the expression of uncertainties,

Recommends

- that the proposals of the Working Group be diffused widely;
- that BIPM attempt to apply the principles therein to international comparisons carried out under its auspices in the years to come;
- that other interested organizations be encouraged to examine and test these proposals and let their comments be known to BIPM;
- that after two or three years BIPM report back on the application of these proposals.

R. Collé

National Measurement Laboratory
National Bureau of Standards
Gaithersburg, MD 20899

Pekka Karp

Measurement Services Office
Technical Inspection Centre
Helsinki, Finland