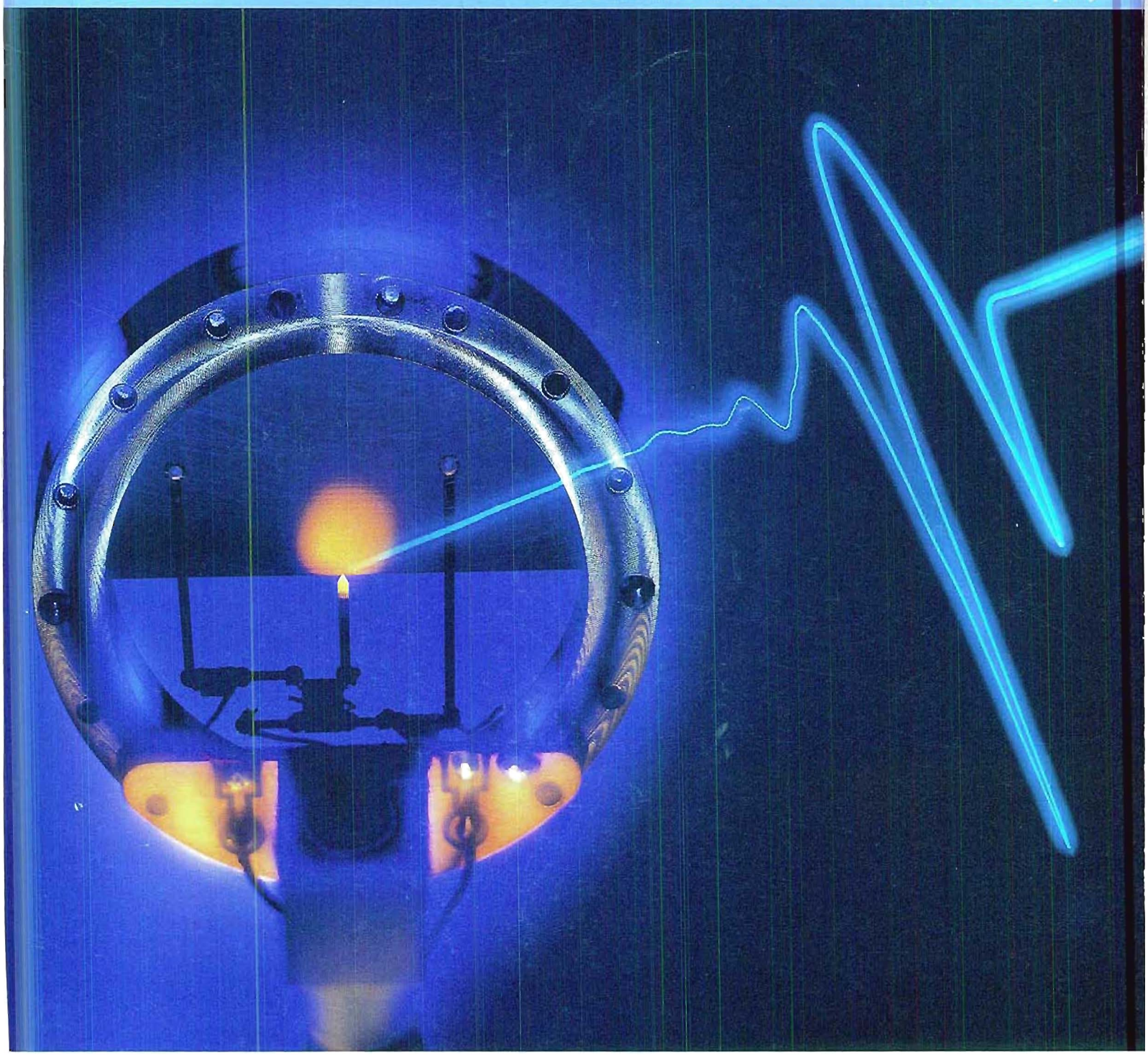
HEWLETT-PACKARD Journal

AUGUST 1998 Technical Information from the Hewlett-Packard Company

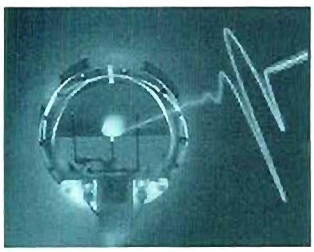


HEWLETT PACKARD

HEWLETT-PACKARD Journal

AUGUST 1998

THE COVER



The 150-MHz-bandwidth membrane hydrophone described on page 6. The signal is generated by a 20-MHz focused ultrasound transducer driving water into a nonlinear state. See page 11 to identify the parts on this hydrophone.

Highlights

The heart of an ultrasound imaging system is the electroacoustic transducer, a device that converts electrical signals into a focused mechanical wave and reconverts reflected mechanical echoes from organs and tissue for real-time image construction. Small, calibrated transducers called hydrophones are used to measure the acoustic output of the transducers used in these systems.

The first article in this issue describes a hydrophone developed by HP that has a spot diameter of 50 micrometers and a bandwidth greater than 150 MHz, enabling it to characterize medical imaging transducers with operating frequencies exceeding 20 MHz.

Measurement accuracy for HP optical power meters and other optical instruments is the main theme in the next article. The article also contains an overview on the theory of measurement.

Hewlett-Packard holds several internal conferences each year to allow HP scientists and engineers to share such things as best practices and research activities. We have five papers from the 1997 conference sponsored by engineers from HP's integrated circuit R&D community.

Improvements in simulation and verification tools for IC design are the main topics of the first two of these articles. The first describes the development

Volume 49 • Number 3

The Hewlett-Packard Journal is published quarterly by the Hewlett-Packard Company to recognize technical contributions made by Hewlett-Packard personnel.

Printed on recycled paper.

of a high-performance equivalence checker for gate-level simulation. The next article describes the development of a model for estimating cross talk between signal lines in submicrometer ULSI interconnects. The model has an accuracy comparable to SPICE.

The last three articles from the conference cover reducing transmission line reflections with HP's HSTL (high-speed transceiver logic) controlled impedance I/O pads, providing low-reflection transitions and high electrical isolation in a low-cost RF multichip module packaging family, and testing mixed-signal ASICs with the HP 9490 mixed-signal LSI tester.

Solutions to manufacturing and reliability problems with the epoxy resin used to attach silicon chips to substrates are discussed in the last two articles. First, we have a description of a new casting epoxy formulation curing condition used for HP surface mount LEDs. Second, we have an analysis that was done to determine the best way to inhibit epoxy bleeding, which causes yield loss in ceramic pin grid array packaging.

C.L. Leath
Managing Editor

M. Shahid Mujtaba, HP Laboratories, Palo Alto, California
Steven J. Narciso, VXI Systems Division, Loveland, Colorado
Danny J. Oldfield, Electronic Measurements Division, Colorado Springs, Colorado
Garry Orsolini, Software Technology Division, Roseville, California
Ken Poulton, HP Laboratories, Palo Alto, California
Günter Riebesell, Böblingen Instruments Division, Böblingen, Germany
Michael B. Saunders, Integrated Circuit Business Division, Corvallis, Oregon
Philip Stenton, HP Laboratories Bristol, Bristol, England
Stephen R. Undy, Systems Technology Division, Fort Collins, Colorado
Jim Willits, Network and System Management Division, Fort Collins, Colorado
Koichi Yuangawa, Kobe Instrument Division, Kobe, Japan
Barbara Zimmer, Corporate Engineering, Palo Alto, California

WHAT'S AHEAD

Articles for the November issue include discussions on LED packaging for automotive applications, OTDR application programming interfaces, the JetSend device-to-device protocol, porting a UNIX® application to the Windows NT® environment, connecting technical and management information systems together, and testing internationalized software. We will also have three articles from three HP internal engineering conferences.

The Hewlett-Packard Journal Online

The Hewlett-Packard Journal is available online at:

http://www.hp.com/hpj/journal.html

A quick tour of our site:

Current Issue—see it before it reaches your mailbox.

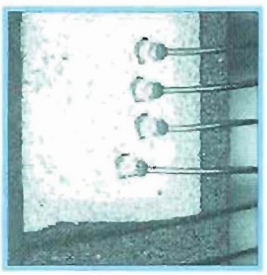
Past Issues—review back to February 1994, complete with links to other HP sites.

Index—search back to our first issue in 1949, complete with an **Order** button for issues or articles you'd like for your library.

Subscription Information—guidelines for U.S. and international subscriptions and a form you can fill out to receive an e-mail message about upcoming issues and changes to our Website.

Previews—contains an early look at future articles.

Articles



The Hewlett-Packard Journal Online

http://www.hp.com/hpj/journal.html

What's new?

- The Previews section contains the following new articles:

  Comparison of Finite-Difference and SPICE Tools for Thermal Modeling of the Effects of Nonuniform Power Generation in High-Power CPUs

  A Low-Complexity, Fixed-Rate Compression Scheme for Color Images and Documents

E-Mail Registration

- Use E-Mail Notification to register your e-mail address so that you can be notified when new articles are published.

The HP Journal Website is evolving rapidly—we invite you to keep up!

A 150-MHz-Bandwidth Membrane Hydrophone for Acoustic Field Characterization

Paul Lum
Michael Greenstein
Edward D. Verdonk
Charles Grossman, Jr.
Thomas L. Szabo

To measure the beam parameters of intravascular ultrasound imaging transducers with operating center frequencies exceeding 20 MHz and beamwidths below 200 μm, a hydrophone with a spot diameter less than 50 μm and a bandwidth greater than 150 MHz is required. The hydrophone described in this article is a step towards meeting these requirements.

Diagnostic ultrasound imaging is used routinely in a growing number of medical applications. Hewlett-Packard manufactures a range of ultrasound imaging systems that use digital beamformers for cardiology and multipurpose imaging and mechanical beamformers for cardiology, general-purpose, and intravascular imaging.

At the heart of an ultrasound system is a transducer, an electroacoustic device that converts electrical signals into a focused mechanical wave and reconverts reflected mechanical echoes from organs and tissue for subsequent real-time image construction. The transducer is a resonant device that has a bandpass filter frequency response. Currently available transducers have a center frequency in the range of 2 to 30 MHz. A phased array of individual transducers, typically consisting of 30 to 300 elements, is electronically focused and steered to provide a beam with the dimensions desired for a selected medical application.

Current trends in imaging are to the following higher-frequency applications:

- Intravascular imaging for plaque detection in blood vessels such as the coronary arteries
- Contrast-agent-assisted harmonic frequency imaging to view blood flow and perfusion in the heart

- Small parts imaging for near-surface high-definition examination of burn injuries, skin lesions, and cancers
- Laparoscopic surgery (ultrasound-guided interventional surgery)
- Ocular imaging (high-resolution visualization of anomalies and repairs of the eye).

These trends pose challenges in the characterization of the transducers used. By law, ultrasound imaging system manufacturers are required to measure the acoustic output of their systems at the transducer. To characterize the acoustic waves of these systems, small calibrated transducers called hydrophones are used. To avoid altering the acoustic fields they are measuring, hydrophones must be as nonperturbing as possible and must far exceed the bandwidth and the spatial resolution of the transducers being measured. This paper describes a new membrane hydrophone that provides the performance needed for these exacting applications.

The hydrophones presently used in the industry have −3-dB bandwidths of 15 to 20 MHz and effective spot sizes of 500 μm. They are appropriate for characterizing acoustic medical imaging transducers up to about 7 MHz. However, even transducers at these frequencies generate, through nonlinear propagation effects in water, higher harmonics that extend in frequency beyond the −3-dB bandwidths of the available hydrophones. To fully characterize a transducer, detection of the fifth harmonic is needed. Furthermore, there are new ultrasonic imaging modalities, such as intravascular ultrasound (IVUS) in the 30-MHz frequency range with 50-μm wavelengths. These transducers cannot be adequately characterized by 15-to-20-MHz-bandwidth hydrophones. (IVUS gives a cross-sectional view of the interior of coronary arteries to assess coronary atherosclerosis.)

Peak pulse parameters calculated from hydrophones with inadequate frequency response show large errors. Hydrophones with effective spot sizes that are too large underestimate the critical parameter of peak pressure because they average the pressure over the hydrophone's active area.

The IEC[1] and NEMA[2] standards regulate the characterization of medical ultrasound transducers by specifying the parameters of the hydrophones used to measure these transducers. The required effective spot diameter is 50 μm for a hydrophone to measure a 30-MHz IVUS transducer with a typical 1-mm aperture and a transducer-to-hydrophone range of less than 2 mm. For more details on transducer characterization, see the box on page 8.

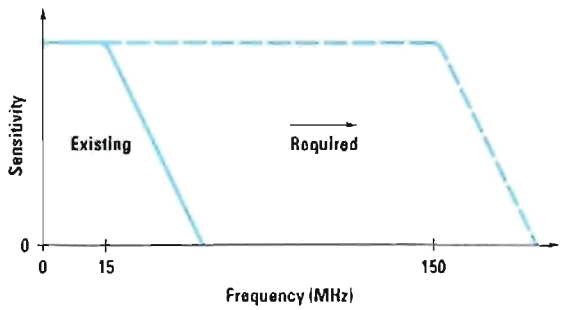
In view of these considerations, there is an important need for hydrophones capable of characterizing transducers with frequency components in the range of 150 MHz and a spatial resolution of typically less than 50 μm. This is illustrated in **Figure 1**, which is a plot of sensitivity as a function of frequency for both existing and required hydrophones.

This article describes the modeling, fabrication, and initial characterization of a membrane hydrophone capable of meeting these more exacting bandwidth and spatial requirements. The hydrophone is fabricated from a 4-μm-thick film of spot-poled* PVDF-TrFE piezoelectric polymer material. It has on-membrane electronics, a −3-dB bandwidth in excess of 150 MHz, and a measured effective spot diameter of less than 100 μm.[3]

* Poling refers to the process of aligning the directions of the ferroelectric domains in a ferroelectric material so that a net polarization occurs. The polymer material used for the hydrophone comes with random ferroelectric polarization. By applying an electric field at an elevated temperature we can rotate the directions of the individual polarization vectors so that they are all aligned along the electric field direction. The higher temperature lowers the work required. When the material returns to room temperature the domains stay aligned, thus creating an active region. Spot-poling is poling confined to selected areas

Figure 1

Bandwidth requirements for hydrophones.



The Hewlett-Packard Medical Products Group Acoustic Output Measurement Laboratory

The control and display of acoustic output have important performance and quality implications for medical diagnostic ultrasound equipment. The acoustic output of these devices should be optimized to provide image quality sufficient for a clinician to make a diagnosis, while at the same time, must be limited to U.S. Food and Drug Administration (FDA) approved levels. The FDA has set limits for acoustic output, and requires reporting of maximum output levels before marketing these devices and as part of the device labeling. Output measurements on production units are performed to ensure manufacturing process control and to establish compliance with FDA Quality System Regulations (QSR). Recognizing that measurement of acoustic output is an essential part of the design, manufacture, and marketing of this equipment, the HP Medical Products Group (MPG) in 1985 established a dedicated, state-of-the-art, acoustic output measurement laboratory at Andover, Massachusetts.

Laboratory Description. The laboratory has instruments required to perform measurements of acoustic pressure, intensity, frequency, and power in the range of 1 to 20 MHz, and is staffed by managers, support engineers, and highly skilled measurement technicians.

The primary measurement system used to measure pressure, intensity, and frequency consists of a water tank, hydrophone, motorized precision positioning system, high-speed sampling oscilloscope, system controller, and associated measurement software. The positioning system provides repeatable positioning of the acoustic beam relative to the hydrophone in the water tank and allows automatic scanning of the acoustic field. The oscilloscope captures the hydrophone output and transfers the sampled data to the system controller for parameter computation (pressure, intensity, frequency).

A radiation force balance is used to measure total acoustic power. This device consists of a small sound-absorbing target attached to one arm of a microbalance suspended in a water column. To perform a power measurement, the source transducer is coupled to the water column and the acoustic beam is directed at the target. The resulting force on the target, as measured by the microbalance, is proportional to the total acoustic power.

The methodology for performing these measurements conforms to those specified in the *NEMA UD-2 Acoustic Output Measurement Standard*, the *NEMA/AIUM Standard for Real-Time Display of Thermal and Acoustic Output Indices on Diagnostic Ultrasound Equipment*, and international standards.

Laboratory Mission. The laboratory's mission is fourfold. The first mission is to support the development of new ultrasound products. Before clinical studies, measurements are performed on prototype equipment to characterize the acoustic field, and maximum acoustic output is verified to be at or below FDA approved limits for pressure and intensity. Later, after more systems are manufactured, output control and display are optimized and validated, and additional measurements are performed to establish the output variability of newly designed products. Finally, production test protocols and test limits are established.

The second mission is to collect acoustic output data required for regulatory labeling. This includes collection of data for premarket notification and other international requirements.

The third mission is to support manufacturing engineering. The laboratory is responsible for addressing all aspects of acoustic output measurement related to FDA-GMP and ISO requirements, including the establishment of production test procedures and test limits and calibration and maintenance procedures of unique test equipment.

Finally, the laboratory is committed to advancing ultrasonic exposimetry, acoustic output measurement science, and associated measurement standards. The laboratory provides a real-world environment for the development of new measurement devices, such as the HP wideband hydrophone described in the accompanying article, and laboratory engineers are actively involved in national and international measurement standards committee work (NEMA, AIUM, IEC, etc.).

To accomplish these missions, the laboratory maintains close, cooperative working relationships with MPG's ultrasound product design team as well as MPG's manufacturing engineering and regulatory staffs.

Charles Grossman, Jr., Thomas L. Szabo, Kathleen Meschisen, and Katharine Stohlman
Imaging Systems Division

Acoustics

The basic structure of a membrane hydrophone is shown in **Figure 2**. Here, a portion of a thin membrane is shown with an electrode and a trace on each side of the membrane. The center frequency is inversely proportional to the thickness of the piezoelectric membrane. The spatial resolution of the hydrophone improves as the diameter of the electrode decreases, and the sensitivity and bandwidth are determined by the piezoelectric coupling of the membrane. The active area of the hydrophone is determined by the overlap of the top and bottom electrodes. In practice, one of the electrodes is extended over a large portion of the membrane to serve as a ground plane.

Acoustic modeling was used to characterize the effects of spot size, film thickness, mass loading from the thin-film electrodes, and directional sensitivity (directivity). The major differences in the properties of the piezoelectric polymer materials between the copolymer PVDF-TrFE used here and the more commonly used PVDF are the increased dielectric constant, the increased effective coupling constant, and the decreased electrical loss tangent of the copolymer.

The thickness-mode resonant frequency, f_o, of the membrane hydrophone is given by:

$$f_o = \frac{c}{2t}, \qquad (1)$$

where c is the acoustic velocity and t is the thickness of the membrane. This relation comes from the requirement that the thickness dimension of the membrane film must be a half wavelength. In general, it is important to have the thickness resonance of the membrane beyond the measurement range to maximize the flatness of the sensitivity. A 4-μm film satisfies this requirement by placing the thickness resonance frequency at 150 MHz. The thickness resonance is also affected by the mass of the metal electrodes. For a 4-μm membrane, conventional 3000Å electrodes degrade the peak frequency of the sensitivity and the fractional bandwidth. A choice of 1000Å for the electrode thickness is a good compromise between a corrosion-resistant electrode with adequate conductivity and adequate bandwidth.

Figure 2

Basic structure of a membrane hydrophone.

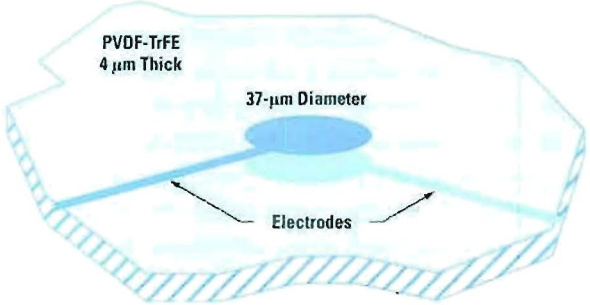
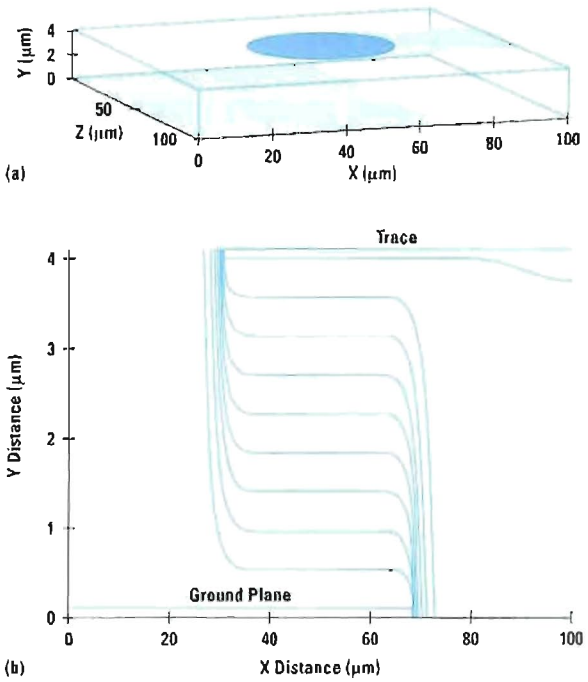


Figure 3

(a) Hydrophone structure for 3D electrostatic modeling. (b) Contours of constant potential at Z = 50 μm.



Electrostatic Fringe Fields

During the spot-poling process, the applied electric fields fringe beyond the edge of the spot electrode and can pole areas of the piezoelectric polymer beyond the intended spot electrode. Electrostatic field modeling was used to model the electrical field patterns to estimate the extent of the fringe field. Voltages on the three-dimensional structure are specified and Poisson's equation is solved iteratively. The geometry of the model is shown in **Figure 3a**. A 37-μm-diameter electrode spot and trace

are patterned over a semi-infinite ground plane. These are on opposite sides of a 4-μm-thick polymer film. In **Figure 3b**, 10% incremental contour plots of the electrical potential are shown for a central cross section in the X-Y plane, through the trace at the top, the polymer film, and the bottom ground plane. An additional 10 μm is poled at greater than 50% of the maximum potential on the electrode. Therefore, this model predicts a larger total active diameter of about 50 μm.

Electrical Matching

The effects of spot size and film thickness are seen by approximating the electrical impedance by capacitive reactance. This reactance is:

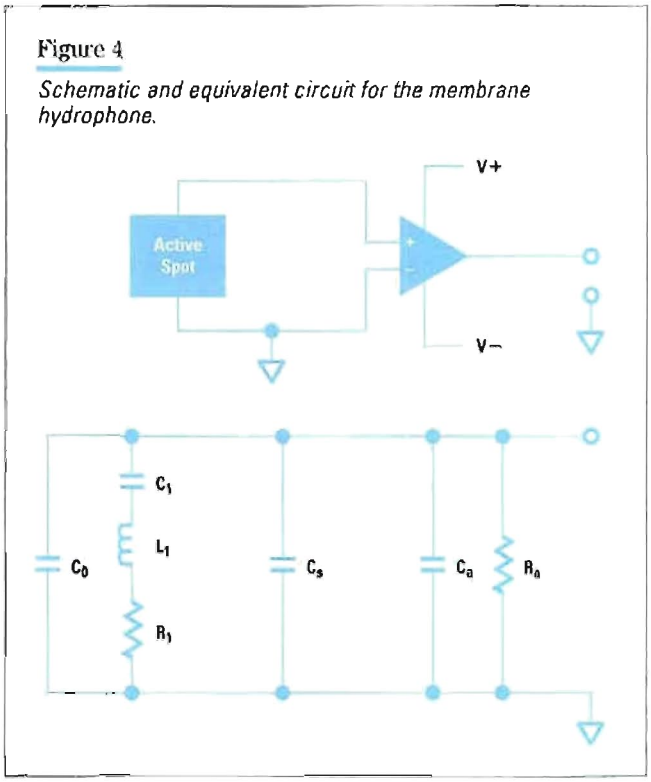
$$Z = \frac{-j}{2\pi f_0 C_0} = \left(\frac{-j}{c\varepsilon}\right)\left(\frac{2t}{\pi D}\right)^2, \qquad (2)$$

where t is the thickness, D is the diameter of the spot, ε is the clamped dielectric constant, and c is the speed of sound in the membrane. Thus the electrical impedance of the membrane hydrophone increases with the square of the membrane thickness and inversely with the square of the spot diameter. For a 4-μm-thick membrane with a 37-μm-diameter spot, Z is 100,000 ohms. This high impedance presents a challenge for matching the transducer to the 50-ohm impedance of the cable used to connect the hydrophone to an oscilloscope.

Figure 4 shows the schematic diagram and equivalent circuit for a membrane hydrophone with an adjacent amplifier. At the far left of the overall equivalent circuit is the equivalent circuit for the piezoelectrically active area resonator. This circuit yields two conditions for resonance. A series resonance condition exists when C_1, L_1, and R_1 resonate to produce an electrical impedance minimum, and a parallel resonance condition exists when the C_1-L_1-R_1 branch is inductive and tunes with C_0 to produce an impedance maximum. In the middle, C_s represents the stray capacitance of the connecting electrode. The adjacent amplifier is represented by a capacitance C_a and a real impedance R_a. Not shown is the 50-ohm coaxial cable that connects the amplifier output to additional electronic circuitry. On-membrane electronics are used to avoid corrupting the frequency characteristics and to match a 50-ohm cable properly.

Figure 4

Schematic and equivalent circuit for the membrane hydrophone.



Spatial Resolution

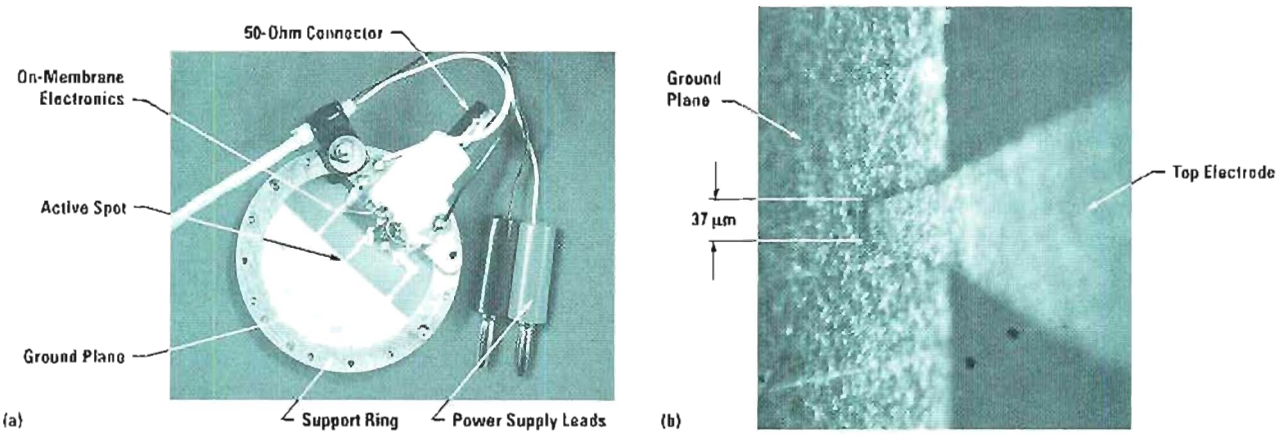
The spatial resolution of a hydrophone is determined by the effective spot size. The effective spot size is in turn determined by the geometric spot size and other electrical and acoustic factors. If the spot size is significantly larger than the size of the acoustic beam being measured, spatial averaging occurs (a hydrophone measures pressure, not energy). Averaging by the hydrophone results in overestimating the size of the beam and underestimating the absolute pressure levels of the beam. These two effects will be illustrated and discussed later in this article for both large-diameter and small-diameter hydrophones.

Fabrication

The fabrication of a membrane hydrophone begins with the raw PVDF-TrFE film in a roll form. An evaporator is used to deposit metal electrodes onto both sides of the film. Shadow masks establish the electrode patterns. One side is patterned for the active spot electrode and the on-membrane electronics connections. The other side is patterned for the ground plane. Alignment of the ground plane relative to the spot electrode is necessary to correctly establish the overlap of the spot electrode.

Figure 5

(a) Photograph of the rear side of the hydrophone. (b) Photomicrograph of the active spot, showing the 37-μm geometric diameter.



The membrane is then mounted onto a ring support structure. In its raw state, as received from the vendor, the piezoelectric PVDF-TrFE film is unpoled with unaligned ferroelectric domains. To align these domains and create the piezoelectric active area, the film is spot-poled at a temperature of 130°C and an electric field strength of 70V/μm.

A film thickness of 4 μm and a spot size of 37 μm result in a device with an electrical impedance of about 100 kilohms. Such a hydrophone is not well-matched to the 50-ohm cable typically used to connect the hydrophone to an oscilloscope input. This substantial electrical impedance mismatch issue is resolved by mounting a wideband, low-distortion buffer amplifier directly on the membrane. The selected amplifier, a surface mount device, has a frequency response that is flat within ±1 dB from dc to 500 MHz and an input impedance on the order of 450 kilohms. To avoid distorting the acoustic field, the amplifier is surface mounted directly onto the membrane at a distance of 10 mm from the spot electrode. The electronics are encapsulated with a cast backing of silicone resin that also acts to protect the fragile 4-μm-thick film. After the backing is cured, the hydrophone is characterized.

A photograph of the rear surface of the hydrophone is shown in **Figure 5a** with several of the key structures highlighted. Normally the back side of the hydrophone is encapsulated in a silicone gel to support the membrane and protect the electronics from the water environment. The membrane is stretched taut over the 60-mm diameter stainless steel support ring. The ground plane can be seen on the bottom left half of the membrane. The active spot is at the end of the trace pointing toward the center of the membrane. The on-membrane electronics are located in the upper right portion of the photo, with the two power supply leads shown on the right side of the photo. The hydrophone is terminated into a standard BNC 50-ohm coaxial connector. **Figure 5b** is a photomicrograph of the active area. Here the top surface electrode is seen on the right as it tapers down to a 37-μm diameter. The ground plane is located in the left half of the picture, under the piezoelectric film, just overlapping the 37-μm-diameter spot. The spatial overlap of the top electrode and the ground plane defines the active spot, once it has been poled.

Bandwidth

One of the key design goals is a −3-dB bandwidth of at least 150 MHz. Several potential methods for measuring the bandwidth include calibration against a known standard hydrophone, interferometry, reciprocity, and the nonlinear distortion method. The first three are generally accepted calibration methods but are currently limited to a maximum frequency of 50 MHz. The nonlinear distortion method, although not considered a standard method, is capable of measurements above 150 MHz, and was chosen to evaluate the new hydrophone.

In the nonlinear distortion method, a source transducer is used to produce nonlinear effects in the water propagation medium. In a fully developed shock wave, such as the familiar one generated by a jet airplane, a classic N wave, or N-shaped waveform, may be formed. The N wave gets its name from its very steep compressional rise time and its more gradual rarefaction fall time. As the waveform propagates, a shock front develops, which generates new frequency components not present in the original waveform. The frequency spectrum of an ideal N-shaped shock wave has frequency harmonics at multiples n of the fundamental, where $n = 1, 2, 3, \ldots, \infty$. In the classic N wave, each harmonic amplitude falls off as $1/n$. Perfect N waves are rarely achieved in the fields of medical ultrasound transducers because of diffraction phase effects in the beam. However, nonlinear distortion methods can still provide an extremely broadband signal to aid in the evaluation of hydrophone bandwidth.

An imperfectly shaped N waveform is shown in **Figure 6a**. This waveform is from a 20-MHz, 6.4-mm-diameter transducer excited in a tone-burst mode, as recorded by an HP 54720A digital oscilloscope and an HP 54721A plug-in amplifier with a 1-GHz analog bandwidth. The hydrophone is placed at the focal plane distance of 19 mm. In the focal plane, the intensity is sufficient to exploit the nonlinear properties of water to generate useful nonlinear distortion. The waveform exciting the transducer is a 20-MHz sinusoidal tone burst. Because of the nonlinear properties of the water propagation medium, the positive compressional half-cycles exhibit very fast rise times, while the negative rarefaction half-cycles exhibit slower fall times. The higher frequencies generated this way do not suffer significant propagation-related attenuation because they are generated right where they are detected.

The −3-dB bandwidth can be estimated from the sharp compressional portion of this waveform. From the 10%-to-90% rise time, a bandwidth of 150 MHz is calculated. The related frequency spectrum is shown in **Figure 6b**. The harmonics from the fundamental at ~20 MHz up to ~800 MHz can be seen. The peaks, corresponding to the harmonics at 20-MHz intervals, show the extent of the nonlinear distortion of the waveform. Without the nonlinear distortion, most of the energy would be located at the fundamental frequency of 20 MHz. At ~700 MHz, the −3-dB bandwidth of the buffer amplifier limits the measured frequency response of the hydrophone. Thus, this hydrophone can detect acoustic frequency components out to at least 800 MHz.

The superior bandwidth of this membrane hydrophone relative to hydrophones currently in use can be demonstrated by performing comparative measurements on acoustic fields generated by medical diagnostic ultrasound equipment. Comparative waveform measurements were made with a Hewlett-Packard SONOS 2500 diagnostic imaging system and a 5-MHz phased array transducer as the source. Two hydrophones were used to measure the acoustic waveform at focus. A calibrated reference

Figure 6

(a) Received nonlinear waveform from a 20-MHz transducer. (b) Spectrum of the nonlinear waveform with harmonics up to 800 MHz.

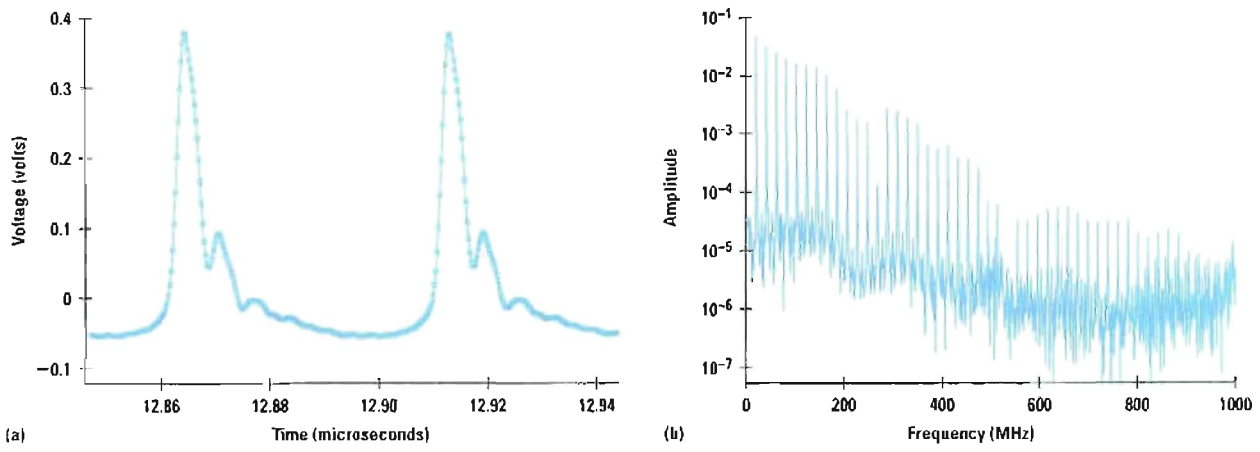
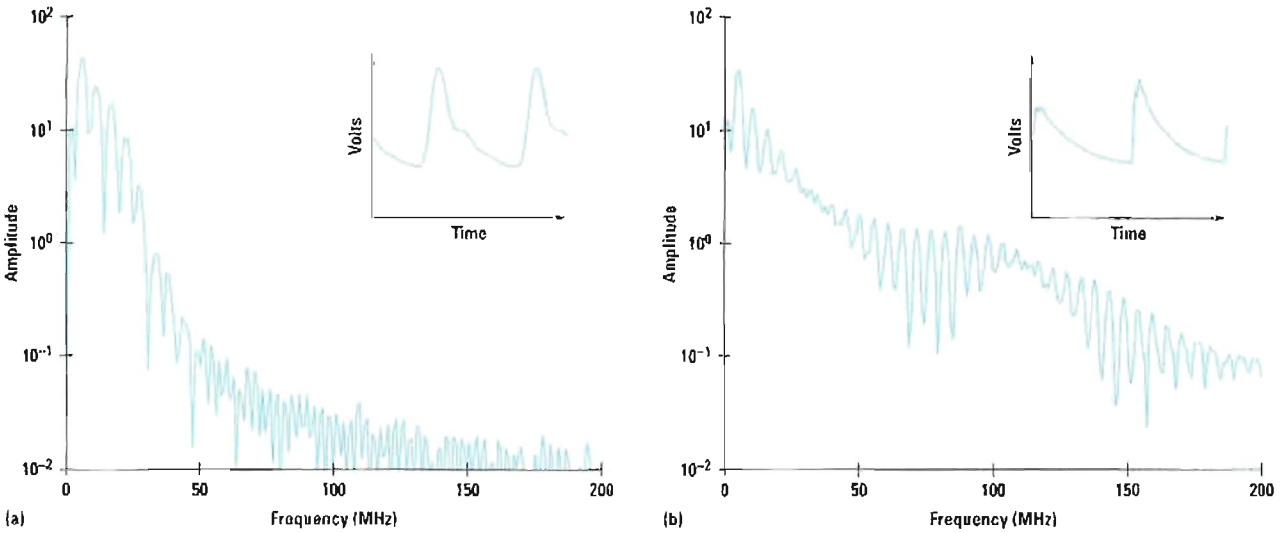


Figure 7

Waveforms (insets) and spectra for (a) 500-μm-diameter hydrophone and (b) new HP 37-μm-diameter hydrophone.



membrane hydrophone with a 500-μm spot diameter on a 25-μm-thick bilaminar PVDF membrane was compared with the new HP high-frequency membrane hydrophone with a 37-μm spot on a 4-μm-thick membrane. **Figure 7a** shows the frequency spectrum using the 500-μm-diameter hydrophone, and the inset shows the measured nonlinear waveform. The spectrum shows the fundamental at 5 MHz and three harmonics at 10, 15, and 20 MHz. **Figure 7b** shows the frequency spectrum using the new HP 37-μm-diameter hydrophone, and the inset shows the respective waveform that has greater detail and a sharper rise time because of the greater bandwidth. The spectrum shows the fundamental at 5 MHz and the subsequent 40 harmonics out to 200 MHz. This frequency data correlates well with the spectra shown in **Figure 6**, given the lower excitation frequency of the SONOS system.

Sensitivity

The sensitivity of a membrane hydrophone can be determined by two fundamental parameters. The first is the thickness resonance of the membrane film. The second is the response determined by the electrical and piezoelectric properties of the material. For a hydrophone the receive voltage sensitivity, M, is given by the ratio of the developed voltage to the incident acoustic pressure, that is, $M = V/P$.

Although the open-circuit voltage sensitivity is the most direct measure of the sensitivity of a hydrophone, it is difficult to measure with the new HP hydrophone and its required on-membrane amplifier. Consequently, the loaded-end-of-cable sensitivity, M_L (relative to 1 V/MPa), was measured and found to be -23 ± 2 dB, within -1.5 dB of the reference 500-μm-diameter hydrophone. This comparison was done with the reference hydrophone configured with a 6-dB external amplifier and the wideband hydrophone driving an external 25-dB amplifier. In each case the circuits are optimally tuned with the external amplifiers in place. Although sensitivity calibration measurements are needed out to 150 MHz, there is not yet a satisfactory calibration procedure for the required range.

Effective Spot Size

When the hydrophone diameter is significantly larger than the beam it measures, spatial averaging occurs. To compare the spatial resolution of the HP high-frequency hydrophone with a reference hydrophone with a 500-μm spot size, the tightly focused beam of a 20-MHz, 6.4-mm-diameter transducer was measured by both hydrophones close to the focal plane. The measured beam of this transducer at its focal length is close to the ideal predicted by theory provided that a sufficiently small hydrophone is used to make the measurement. The measured field can be

modeled to a good approximation as the spatial average of the theoretical field over the hydrophone area. Because of radial symmetry, the area averaging simplifies to a running mean across the theoretical beam.

Measured beamwidths for the two hydrophones are compared to theoretical calculations of the spatially averaged and unaveraged beamwidths in **Figure 8**. In **Figure 8a** the data from the 500-μm reference hydrophone is shown, and the comparable data for the HP hydrophone is given in **Figure 8b**. These curves demonstrate that the reference hydrophone underestimates the on-axis pressure at the beam peak by 40% and overestimates the beamwidth by 50%, whereas the new HP high-frequency hydrophone provides a faithful replica of the beam shape and pressure.

Angular measurement techniques can also be used to estimate the effective spot size of a hydrophone. In the angular response method, the hydrophone is swept through an arc about the reference transducer. In this way, directional response can be measured and used to estimate the effective diameter. In this method the actual beamwidths at −3 dB and at −6 dB are compared to those expected from theory. From the theoretical beamwidth, for a given frequency, the effective hydrophone spot diameter can be inferred from the beamwidth data. The same 20-MHz focused transmitter was used to perform a directivity measurement. The resulting measured half angles, when applied to the theoretical model, indicate a spot diameter upper bound of 100 μm for the device measured.

Intravascular Ultrasound Application

As an example of the application of this new hydrophone, a 30-MHz IVUS catheter transducer excited by an HP IVUS system under normal system settings was characterized. In use, a stationary sheath is used to protect the intimal lining of the blood vessel (one monolayer of cells thick) from the rotating catheter and transducer. The sheath both attenuates and focuses the beam. In **Figure 9a**, the waveform is shown as received by the hydrophone after a 1.9-mm path in water. The hydrophone was carefully placed on the axis of sound propagation. The thick line is the waveform with the 110-μm-thick polymer sheath, and the thin line is the waveform without the sheath. The beam that is focused by the sheath arrives at the hydrophone earlier. The disparity in the rise and fall times of these signals is an indication that nonlinear distortion has developed in the water medium. The frequency spectra of these waveforms in **Figure 9b** indicate the presence of nonlinear distortion both with a sheath and without a sheath. There are frequency components of this waveform out to at least 150 MHz, which cannot be measured by a conventional 20-MHz-bandwidth hydrophone.

Figure 8

Linear scans of a 20-MHz ultrasound transducer beam with two hydrophones: (a) a 500-μm-diameter reference hydrophone and (b) the 37-μm-diameter HP hydrophone. In each case, the theoretical transducer beam shape, the theoretical beam shape as spatially averaged by the hydrophone spot size, and the discrete hydrophone measurement data are shown.

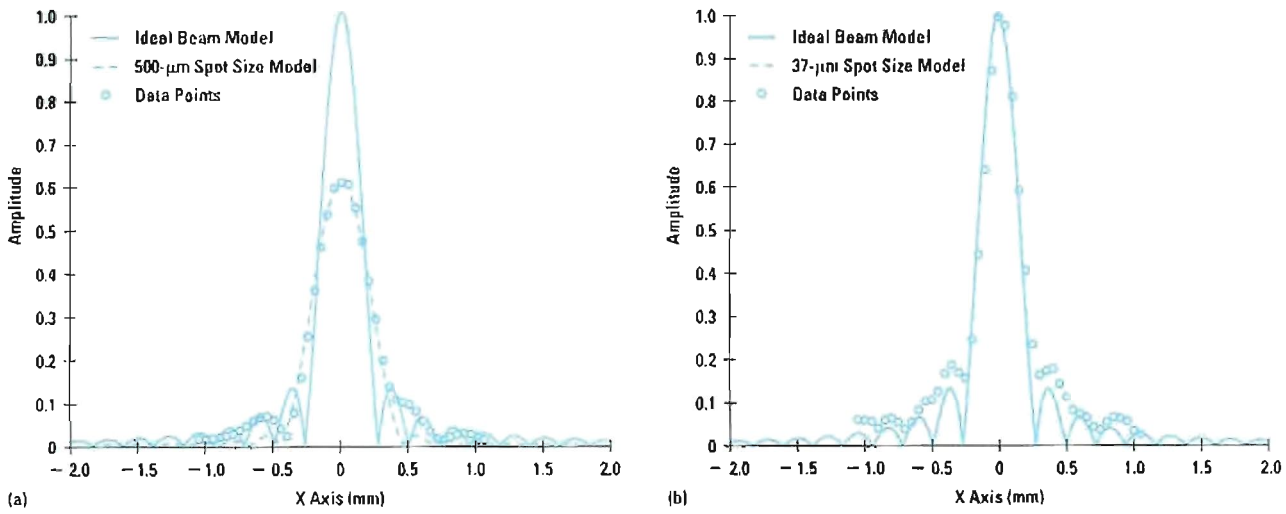
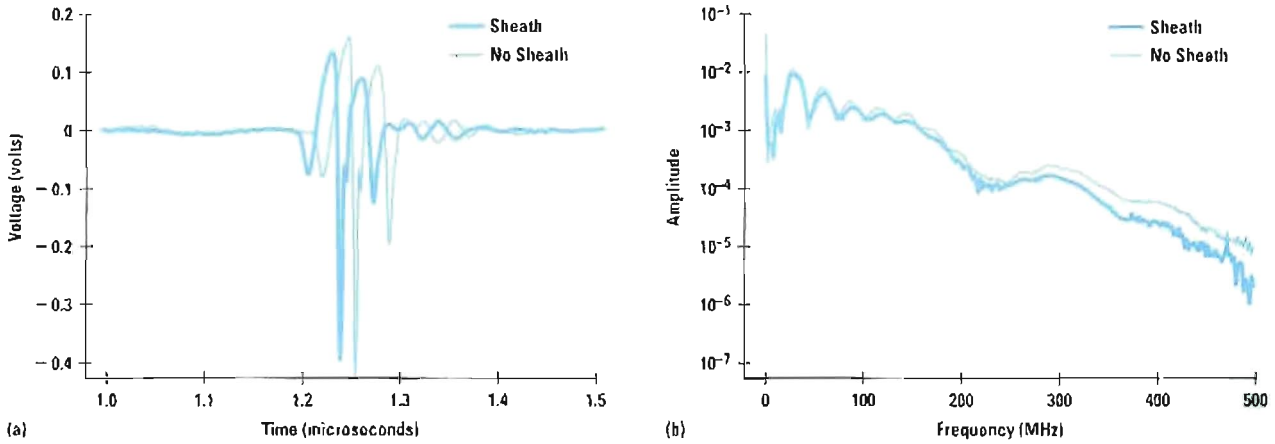


Figure 9

HP membrane hydrophone measurements of nonlinear data from a 30-MHz IVUS catheter. (a) Time-domain waveforms. (b) Frequency-domain spectra.



Summary

With the increasing use of intravascular ultrasound imaging transducers with operating center frequencies exceeding 20 MHz and beamwidths below 200 μm, smaller-spot-size and higher-frequency hydrophones are needed. Characterizing transducer acoustic pressure fields according to the AIUM/NEMA standards requires a hydrophone with a spot diameter less than 50 μm and a bandwidth greater than 150 MHz. The hydrophone described in this article is a step towards meeting these requirements.

Acoustic modeling was performed to provide a general guideline for selecting the membrane thickness, electrode diameter, electrode conductor metal thickness, electrode interconnect lead length, and placement of the amplifier. Fabrication of the hydrophone required characterization of the material, electrode patterning, spot poling, and assembly. The nonlinear distortion method was used to evaluate the bandwidth. Directivity measurements were used to determine an upper limit for the effective spot diameter. The substitutional method was used to evaluate the hydrophone sensitivity up to 20 MHz. As an example of the capability of the hydrophone, the on-axis signal from a 30-MHz IVUS transducer in water was characterized.

The result of this work is an acoustic hydrophone fabricated on a 4-μm-thick membrane film of PVDF-TrFE, having an effective active spot less than 100 μm in diameter, an on-membrane buffer amplifier within 10 mm of the active spot, and a loaded end-of-cable sensitivity of −23 dB relative to 1V/MPa in the range 5 to 20 MHz. Additional characterization is needed to determine the absolute hydrophone response in the range of 20 to 150 MHz, and to measure the effective spot size more accurately. Further details on this hydrophone and a more complete set of references are given in reference 3. This hydrophone has been licensed to an external vendor for commercialization, and is expected to be available in the spring of 1998.

Acknowledgments

The authors would like to acknowledge Belinda Kendle for help with hydrophone fabrication, Said Bolorforosh for many useful discussions, Henry Yoshida for the poling system, Jerry Zawadzki for numerous fixtures, Pete Melton and Kate Stohlman for their continual support and encouragement, and Paul Magnin for his consistent level of support.

References

1. *Measurement and Characterization of Ultrasonic Fields Using Hydrophones in The Frequency Range 0.5 MHz to 15 MHz*, IEC 1102:1991.

2. *Acoustic Output Measurement Standard For Diagnostic Ultrasound Equipment*, NEMA Standards Publication No. UD 2-1992, National Electrical Manufacturers Association, 2101 L Street, N.W. Washington, D.C. 20037.

3. P. Lum, M. Greenstein, C. Grossman, and T. Szabo, "High-Frequency Membrane Hydrophone," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 43, no. 4, July 1996, pp. 536-544.

Paul Lum

Paul Lum is a member of the technical staff of the analytical medical laboratory of HP Laboratories, specializing in medical sensors, actuators, transducers, and instrumentation. With HP since 1977, he holds an MS degree in electrical engineering and computer science from the University of Santa Clara. A native of San Francisco, he is married and has two children.

Michael Greenstein

Michael Greenstein received his PhD degree in experimental solid-state physics from the University of Illinois Urbana-Champaign in 1981 and joined HP Laboratories in 1982. He is currently a project manager for diagnostic sensors and devices. His work has resulted in 18 patents and 26 publications. Born in Oakland, California, he is married, has a daughter, raises bonsai trees, and plays soccer.

Edward D. Verdonk

Ed Verdonk is a member of the technical staff of the analytical medical laboratory of HP Laboratories, specializing in ultrasound imaging. He received his PhD degree in physics in 1992 from Washington University of St. Louis and joined HP in 1993. He is married, has a daughter, and comes from Edmonton, Alberta, Canada.

Charles Grossman, Jr.

Charlie Grossman is a manufacturing development engineer at the HP Imaging Systems Division and supports the Acoustic Output Measurement Laboratory. He joined HP in 1979 after ten years in circuit design with GTE. His BSEE degree (1969) is from the State University of New York at Buffalo. Born in Munich, Germany, he enjoys photography and travel.

Thomas L. Szabo

Tom Szabo joined HP in 1981 after 12 years of research activities with the U.S. Air Force. A senior R&D engineer with the HP Imaging Systems Division, he is involved in transducer array design, beamforming, and acoustic output characterization. He received his PhD degree in physics from the University of Bath (UK) in 1993. He has published 65 papers and his work has resulted in three patents. A native of Budapest, he is married, has a son, and enjoys photographing megalithic sites and gargoyles and training cats to do tricks.

Units, Traceability, and Calibration of Optical Instruments

Andreas Gerster

This article presents a short and comprehensive overview of the art of units measurement and calibration. Although the examples focus on optical instruments, the article may be of interest to anyone interested in metrology.

The increasing number of companies using quality systems, such as the ISO 9000 series, explains the growing interest in the validation of the performance of measurement instruments. For many customers it is no longer sufficient to own a feature-rich instrument. These customers want to be sure that they can test and measure in compliance with industrial and legal standards. Therefore, it is important to know how it can be guaranteed that a certain instrument meets specifications.

This article is intended to give an overview of the calibration of optical power meters and other optical instruments at HP. Along with the specific instruments, common processes and methods will be discussed. The first section will deal with some aspects of the theory of measurement. Then, processes and methods of calibration and traceability will be discussed. These first two sections give a general and comprehensive introduction to the system of units. Finally, the calibration procedures for certain HP optical instruments will be described.

Theory of Measurement

Measurement has long been one of the bases of technical, economical, and even political development and success. In the old Egyptian culture, surveying and trigonometry were important contributors to their prosperity. Religious and political leaders in these times founded their power on, among other things, the measurement of times and rotary motions, which allowed them to predict solar and lunar eclipses as well as the dates of the flood season of the Nile river. Later in history, weights and length measurements were fundamental to a variety of trading activities and to scientific progress. Improvements in

Andreas Gerster

Andreas Gerster received his Diplom-Physiker from the University of Stuttgart in 1996 and joined HP the same year. An engineer in the HP Böblingen Instrument Division optical standards laboratory, he is responsible for calibration tool development and uncertainty analysis. He is a member of DPG (Physical Society of Germany). Andreas is married, has a daughter, and enjoys sailing and biking.

measurement techniques, for example, allowed Keppler to set up his astronomical laws. Keppler used the measurement results of his teacher Tycho Brahe, who determined the orbits of the planets in the solar system with an unprecedented accuracy of two arc minutes.[1] Because of the strong impact of a homogenous system of measurements, all metrologic activities even today are controlled by governmental authorities in all developed countries.

Let us consider first the measurement itself. Measurement is the process of determining the value of a certain property of a physical system. The only possibility for making such a determination is to compare the unknown property with another system for which the value of the property in question is known. For example, in a length measurement, one compares a certain distance to the length of a ruler by counting how often the ruler fits into this distance. But what do you use as a ruler to solve such a measurement problem? The solution is a mathematical one: one defines a set of axiomatic rulers and deduces the practical rulers from this set.

At first, rulers were derived from human properties. Some of the units used today still reflect these rulers, such as feet, miles (in Latin, mile passuum = 1000 double steps), cubits (the length of the forearm), or seconds (the time between two heartbeats is about one second). As one can imagine, in the beginning these axiomatic rulers were anything but general or homogenous—for example, different people have different feet. Only a few hundred years ago, every Freie Reichsstadt (free city) in the German empire had its own length and mass definitions. The county of Baden had 112 different cubits at the beginning of the nineteenth century, and the city of Frankfurt had 14 different mass units.[2] In some cities the old axiomatic ruler was mounted at the city hall near the marketplace and can still be visited today.

With the beginning of positivism, about the time of the revolution in France, people were looking for absolute types of rulers that could make it easier to compare different measurements at different locations. Thus, in 1790, the meter was defined in Paris to be 1/40000000 of the length of the earth meridian through Paris. Because such a measurement is difficult to carry out, a physical artifact was made out of a special alloy, and the standard meter was born. This procedure was established by the international treaty of the meter in 1875, and although the unit definitions have changed, the contract is still valid today.

Related to this search for suitable axiomatic rulers is the question of how many different rulers are really necessary to deduce all practical units. Among others, F. Gauss delivered valuable contributions to the answer. He proposed a system consisting of only three units: mass, time and length. All other mechanical and electrical and therefore optical units could be deduced. As Lord Kelvin showed in 1851, the temperature is also directly related to mechanical units. Therefore, along with the meter, two other axiomatic rulers were defined. Mass was defined by the international kilogram artifact which was intended to have the mass of one cubic decimeter of pure water at 4°C (in fact it was about 0.0028 g too heavy). The definition of time finally was related to the duration of a certain (astronomical) day of the year 1900.

For various reasons, these definitions were not considered suitable anymore in the second half of the 20th century, and metrologists tried to find natural physical constants as bases for the definition of the axiomatic units. At first the time unit (second) and the length unit (meter) were defined in terms of atomic processes. The meter was related to the wavelength of the light emitted from krypton atoms due to certain electronic transitions. In the case of the time unit, a type of cesium oscillator was chosen. The second was defined by 9,192,631,770 cycles of the radiation emitted by electronic transitions between two hyperfine levels of the ground state of cesium 133.

These new definitions of the axiomatic units had a lot of advantages over the old ones. The units of time and length were now related to natural physical constants. This means that everybody in the world can reproduce these units without having to use any artifacts and the units will be the same at any time in any place.

For practical reasons, more units were added to the base units of the Système International d'Unités, or International System of Units (SI). Presently there are seven base units, two supplementary units, and 19 derived units in the SI. The base units are listed in Table I. There is no physical necessity for the selection of a certain set of base units, but only practical reasons. In fact, considering the definitions, only three of the base units—the second, kelvin, and kilogram—are independent, and even the kelvin can be derived from mechanical units.

Table I
Definitions of the Seven Base Units of the Système International d'Unités (SI)

Unit	Name and Symbol	Definition
Length	meter (m)	One meter is now defined as the distance that light travels in vacuum during a time interval of 1/299792458 second.
Mass	kilogram (kg)	The kilogram is defined by the mass of the international kilogram artifact in Sèvres, France.
Time	second (s)	A second is defined by 9,192,631,770 cycles of the radiation emitted by the electronic transition between two hyperfine levels of the ground state of cesium 133.
Electrical Current	ampere (A)	An ampere is defined as the electrical current producing a force of 2×10^{-7} newtons per meter of length between two wires of infinite length.
Temperature	kelvin (K)	A kelvin is defined as 1/273.16 of the temperature of the triple point of water.
Luminous Intensity	candela (cd)	A candela is defined as the luminous intensity of a source that emits radiation of 540×10^{12} hertz with an intensity of 1/683 watt per steradian.
Amount of Substance	mole (m)	One mole is defined as the amount of substance in a system that contains as many elementary items as there are atoms in 0.012 kg of carbon 12.

Nevertheless, some small distortions remain. Of course, all measurements are influenced by the definitions of the axiomatic units, and so the values of the fundamental constants in the physical view of the world, such as the velocity of light c, the atomic fine structure constant a, Plank's constant h, Klitzing's constant R, and the charge of the electron e, have to be changed whenever improvements in measurement accuracy can be achieved.

This has led to the idea of relating the axiomatic units directly to these fundamental constants of nature.[3] In this case the values of the fundamental constants don't change anymore. The first definition that was directly related to such a fundamental constant of nature was the meter. In 1983 the best known measurement value for the velocity of light c was fixed. Now, instead of changing the value for c whenever a better realization of the meter is achieved, the meter is defined by the fixed value for c. One meter is now defined as the distance that light travels in a vacuum during a time interval of 1/299792458 second. The next important step in this direction could be to hold the value e/h constant and define the voltage by the Josephson effect (see the Appendix).

Calibration and Traceability

According to an international standard, calibration is "the set of operations which establish, under specified conditions, the relationship between the values indicated by the measuring instrument and the corresponding known values of a measurand."[4] In other words, calibration of an instrument ensures the accuracy of the instrument.

Of course, no one can know the "real" value of a measurand. Therefore, it must suffice to have a best approximation of this real value. The quality of the approximation is expressed in terms of the uncertainty that is assigned to the apparatus that delivers the approximation of the real value. For calibration purposes, a measurement always consists of two parts: the value and the assigned measurement uncertainty.

The apparatus representing the real value can be an artifact or another instrument that itself is calibrated against an even better one. In any case, this best approximation to the real value is achieved through the concept of *traceability*. Traceability means that a certain measurement is related by appropriate means to the *definition* of the unit of the measurand under question. In other words, we trust in our measurement because we have defined a unit (which is expressed through a standard, as shown later), and we made our measurement instrument conform with the definition of this unit (within a certain limit of uncertainty).

Thus, the first step for a generally accepted measurement is a definition of the unit in question that is accepted by everybody (or at least by all people who are relevant for our business), and the next step is an apparatus that can

realize this unit. This apparatus is called a (primary) *standard*. What does such an apparatus look like? It simply builds the definition in the real world. In the case of the second, for example, the *realization* of the unit is given by cesium atoms in a microwave cavity, which is used to control an electrical oscillator. The realization of the second is the most accurate of all units in the SI. Presently, uncertainties of 3×10^{-15} are achieved.[5] The easiest realization is that of the kilogram, which is expressed by the international kilogram artifact of platinum iridium alloy in Sèvres near Paris. The accuracy of this realization is excellent because the artifact *is* the unit, but if the artifact changes, the unit also changes. Unfortunately, the mass of the artifact changes on the order of 10^{-9} kg per year.[5]

Representation and Dissemination

National laboratories (like PTB in Germany or NIST in the U.S.A.) are responsible for the realization of the units in the framework of the treaty of the meter. Since there can be only one institution responsible for the realization, the national laboratories must disseminate the units to anyone who is interested in accurate measurements. Most of the realizing experiments are rather complicated and sometimes can be maintained for only a short time with an appropriate accuracy. Thus, for the *dissemination* of the unit, an easy-to-handle *representation* of the unit is used. These representations have values that are traceable to the realizing experiments through *transfer measurements*. New developments in metrology allow the representation of some units as quantum standards, so in some cases the realization of a unit has a higher uncertainty than its representation. In the Appendix this effect is discussed for the example of electrical units.

Once the representation of a unit is available, the calibration chain can be extended. The representations can be duplicated and distributed to institutions that have a need for such secondary standards. In some cases, a representation can be used directly to calibrate general-purpose bench instruments. In the case of electrical units, the representations are used to calibrate highly sophisticated calibration instruments, which allow fully automatic calibration of a device under test, including the necessary reporting.[6]

Although the calibration chain described above is a very common and the most accepted method of providing traceability, it is only one among others. Another method of providing traceability involves comparing against natural physical constants. A certain property of a physical system is measured with the device under test (DUT) and the reading of the device is compared against the known value of this property. For example, a wavelength measurement instrument can be used to measure the wavelength of the light emitted from a molecule as a result of a certain electronic transition. The reading of the instrument can easily be compared against the listed values for this transition. This method of providing traceability has some advantages. It is not necessary to have in-house standards, which have to be recalibrated regularly, and in principle, the physical constant is available everywhere in the world at all times with a fixed value. However, a fundamental issue is to determine what is considered to be a natural physical constant. Of course there are the well-known fundamental constants: the velocity of light, Planck's constant, the hyperfine constant, the triple point of water, and so forth, but for calibration purposes many more values are used. In literature some rather complicated definitions can be found.[3] We'll try a simple definition here: a natural physical constant is a property of a physical system the value of which either does not change under reasonable environmental conditions or changes only by an amount that is negligible compared to the desired uncertainty of the calibration. Reasonable in this context means moderate temperatures (-50 to 100°C), weak electromagnetic fields (order of nilliteslas), and so forth.

Somewhat different from the two methods described above are ratio-type measurements using self-calibrating techniques. An example of this technique will be described in the next section.

Calibration of Optical Instruments

In this section we describe the calibration procedures and the related traceability concepts of some of the optical measurement instruments produced by HP. In contrast to the calibration of electrical instruments like voltmeters, it is nearly impossible to find turnkey solutions for calibration systems for optical instruments. Because optical fiber communications is a new and developing field, the measurement needs are changing rapidly and often the standardization efforts cannot keep pace.

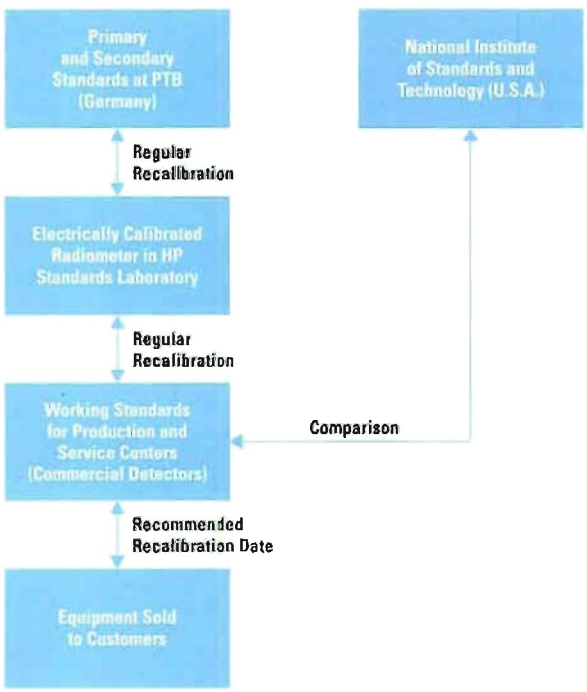
Calibration of Optical Power Meters

The basic instrument in optical fiber communications is the optical power meter. Like most commercial power meters used for telecommunications applications, the HP 8153A power meter is based on semiconductor photodiodes. Its main purpose is to measure optical power, so the most important parameter is the optical power accuracy.

Figure 1 illustrates the calibration chain for HP's power meters. This is an example of the traceability concept of an unbroken chain of calibrations. It starts with the primary standard at the Physikalisch Technische Bundesanstalt (PTB) in Germany. As discussed in the previous sections, the chain has to start with the definition of the unit. Since we want to measure power, the unit is the watt, which is defined to be $1W = (1\ m/s)(1\ kg \cdot m/s^2)$.

Figure 1

The calibration chain for the HP 8153A optical power meter. This is an example of calibration against a national standard (PTB). In addition, the HP working standards are calibrated at NIST in the U.S.A. Thus the calibrations carried out with these working standards are also traceable to the U.S. national standard. HP also participates in worldwide comparisons of working standards. This provides data about the relation between HP's standards and standards of other laboratories in the world.



Thus, the optical power must be related to mechanical power. Normally such a primary standard is realized by an electrically calibrated radiometer. The principle is sketched in **Figure 2**. The optical power is absorbed (totally, in the ideal case) and heats up a heat sink. Then the optical power is replaced by an applied electrical power that is controlled so that the heat sink remains at the same temperature as with the optical power applied. (The electrical power is related to mechanical units, as shown in the Appendix.) In this case the dissipated electrical power P_{el} is equal to the absorbed optical power P_{opt} and can easily be calculated from the voltage V and the electrical current I:

$$P_{opt} = VI = P_{el}.$$

As always, in practice the measurement is much more complicated. Only a few complications are mentioned here; more details can be found in textbooks on radiometry:[7,8]

- Not all light emitted by the source to be measured is absorbed by the detector (a true black body does not exist on earth).
- The heat transfer from the electrical heater is not the same as from the absorbing surface.
- The lead-in wires for the heaters are electrical resistors and therefore also contribute to heating the sink.

For these and other reasons an accurate measurement requires very careful experimental technique. Therefore,

Figure 2

Principle of an absolute radiometer (electrically calibrated radiometer). The radiant power is measured by generating an equal heat by electrical power. The heat is measured with the thermopile as an accurate temperature sensor. The optical radiation is chopped to allow control for equal heating of the absorber: the electrical heater is on when the optical beam is switched off by the chopper and vice versa.

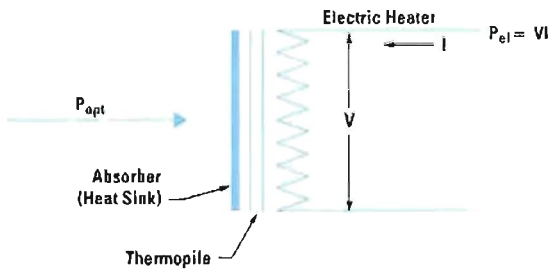
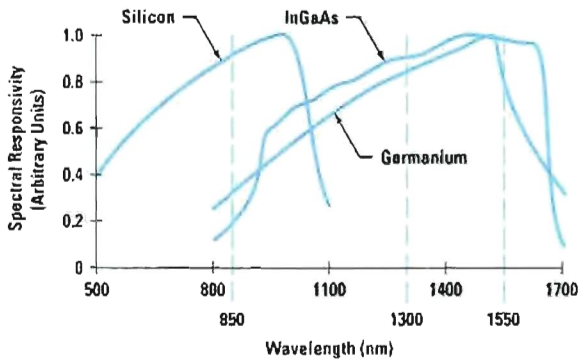


Figure 3

Typical wavelength dependence of common photodiodes.



for dissemination, the optical watt is normally transferred to a secondary standard, such as a thermopile. To keep the transfer uncertainty low it is important that this transfer standard have a very flat wavelength dependence, because the next element in the chain—the photodiode—can also exhibit a strong wavelength dependence (see **Figure 3**).

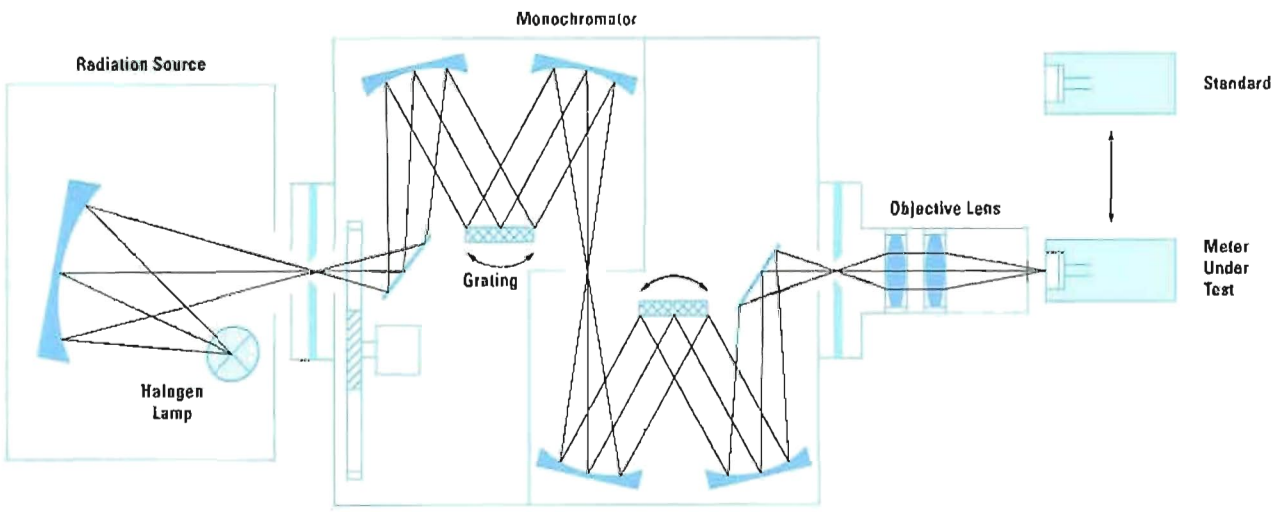
Normally, absolute power is calibrated at one reference wavelength, and all other wavelengths are characterized by their relative responsivities, that is, by the dependence of the electrical output signal on the wavelength of an optical input signal at constant power. Because of all the experimental problems related to traceability from optical to electrical (and therefore mechanical) power, an absolute power uncertainty of 1% is hard to achieve. To keep the transfer uncertainties from PTB to the HP calibration lab as low as possible, HP uses an electrically calibrated radiometer as the first device in its internal calibration chain.

For the selection of a certain wavelength, a white light source in combination with a grating-based monochromator is used (see **Figure 4**). This solution has some advantages over a laser-based method:

- A continuous spectrum is available over a very large wavelength range (from UV to the far IR). Lasers emit light only at discrete lines or the tuning range is limited to a few tens of nanometers.
- A tungsten lamp is a classical light source that exhibits almost no coherence effects.
- The power can be kept relatively flat over a large wavelength range.
- The output beam is only weakly polarized.

Figure 4

Calibration setup for absolute power calibration. The monochromator is used to select the wavelength out of the quasicontinuous spectrum of the halogen lamp. The output power as a function of wavelength is first measured with the working standard and then compared with the results of the meter under test. The deviation is used to calculate correction factors that are stored in the nonvolatile memory of the meter.



As always, in practice there are some disadvantages that make a monochromator system an unusual tool:

- A monochromator is an imaging system, that is, external optics are necessary to bring the light into a fiber or onto a large-area detector.
- Available power is rather low compared to the power levels available from telecommunications lasers. Typically 10 μW is achieved in an open-beam application, but the power level that can be coupled into a fiber may be 30 dB less.
- Because a monochromator is a mechanical tool, a wavelength sweep is rather slow. This can give rise to power stability problems.
- Often the optical conditions during calibration are quite different from the usual DUT's application. This leads to higher uncertainties, which must be determined by an uncertainty analysis.

Thus, setting up a monochromator-based calibration system is not without problems. Nevertheless, for absolute power calibration over a range of wavelengths it is the tool of choice. The calibration is carried out by comparing the reading of the DUT at a certain wavelength with the reading of the working standard at this wavelength. The deviation between the two readings yields a correction factor that is written into the nonvolatile memory of the DUT. In the case of the detectors for the HP 8153A power meter the wavelength is swept in steps of 10 nm over the whole wavelength range. The suitable wavelength range depends on the detector technology. For wavelengths between two calibration points, the correction factor is obtained by appropriate interpolation algorithms. After the absolute power calibration is finished, the instrument can deliver correct power readings at any wavelength.

Calibration of Power Linearity

The procedure described above calibrates only the wavelength axis of the optical power meter. How accurate are power measurements at powers that do not coincide with the power selected for the wavelength calibration? This question is answered by the linearity calibration.

State-of-the-art power meters are capable of measuring powers with a dynamic range as high as 100 dB or more. Ideally, the readings should be accurate at each power level. If one doubles the input power, the reading should also double. A linearity calibration can reveal whether this is really the case. The linearity of the power meter is directly related to the accuracy of relative power measurements such as loss measurements.

There are several reasons for nonlinearity in photodetectors. At powers higher than about 1 mW, the photodiode itself may become nonlinear because of saturation effects. Nonlinearities at lower powers are normally caused by the electronics that evaluate the diode signal. Internal amplifiers are common sources of nonlinearity; their gains must be adjusted properly to avoid discontinuities when switching between power ranges. Analog-to-digital converters can also be the reason for nonlinearities.

In a well-designed power meter the nonlinearities induced by the electronics are very small. Thus, the nonlinearity of a good instrument is near zero, which makes it quite difficult to measure with a small uncertainty. Indeed, often the measurement uncertainty exceeds the nonlinearity.

The linearity calibration of HP power meters is an example of a self-calibration technique.[9,10] The nonlinearity NL at a certain power level P_x is defined as:

$$NL = \frac{r_x - r_{ref}}{r_{ref}},$$

where $r = D/P$ is the power meter's response to an optical stimulation, with D being the displayed power and P the incident power. The subscript ref indicates a reference power level, which can be arbitrarily selected. Replacing the responsivity r by D/P, the nonlinearity can be written as:

$$NL = \frac{D_x/P_x}{D_{ref}/P_{ref}} - 1 = \frac{D_x}{D_{ref}}\frac{P_{ref}}{P_x} - 1.$$

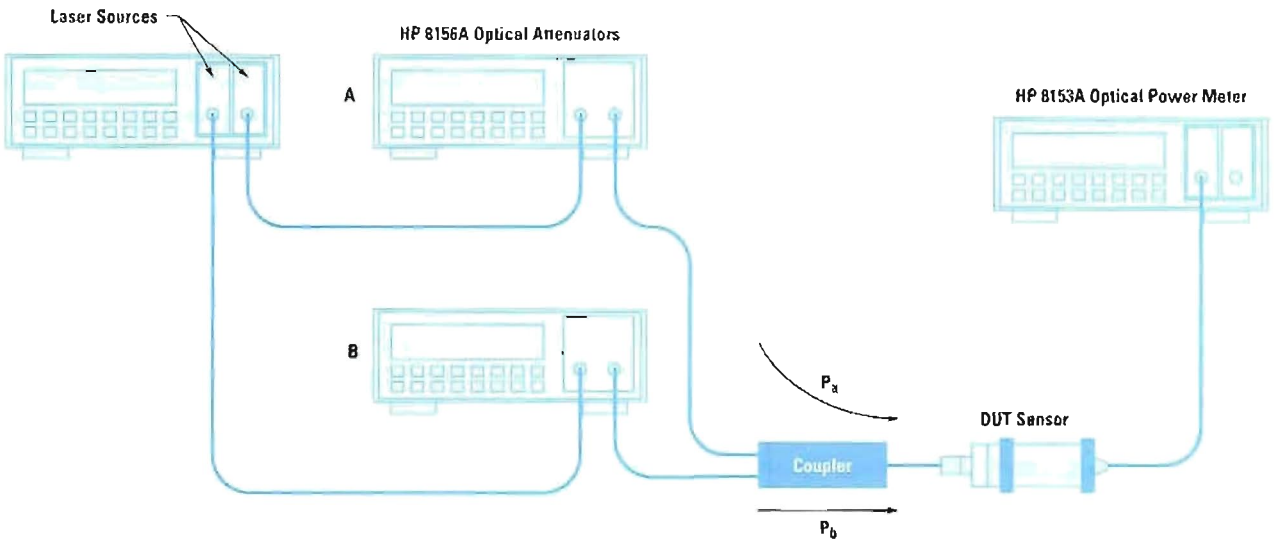
This can now be compared with the error for a loss measurement. Loss means the ratio between two power levels P_1 and P_2. Let D_1 and D_2 be the corresponding displayed power levels, and define the real loss $RL = P_1/P_2$ and the displayed loss $DL = D_1/D_2$. The loss error LE is given by the relative difference between RL and DL:

$$LE = \frac{DL - RL}{RL} = \frac{DL}{RL} - 1 = \frac{D_1P_2}{D_2P_1} - 1.$$

It is evident that if one selects P_1 as reference power, the loss error is given by the nonlinearity at $P_x = P_2$. If P_1 is different from the reference power P_{ref} the statement is still valid to a first-order approximation. The bottom line

Figure 5

Setup for the self-calibrating method of linearity calibration.



is that in relative power measurements like insertion loss or bit error rate tests, linearity is the important property.

The setup for this self-calibration technique is shown in **Figure 5**. First, attenuator A is used to select a certain power P_a, which is guided through optical path A to the power meter under test. The corresponding power reading is recorded. Then attenuator A is closed and the same power as before is selected with attenuator B, resulting in a power reading P_b. Now both attenuators are opened, and the resulting reading P_c should be nearly the sum of P_a and P_b (see **Figure 6**). Any deviation is recorded as nonlinearity. Using the same notation as before, the displayed loss DL is given by:

$$DL = \frac{P_c}{P_a} \approx \frac{P_c}{P_b}.$$

The real loss RL can be calculated by adding the two first readings:

$$RL = \frac{P_a + P_b}{P_a} \approx \frac{P_a + P_b}{P_b}.$$

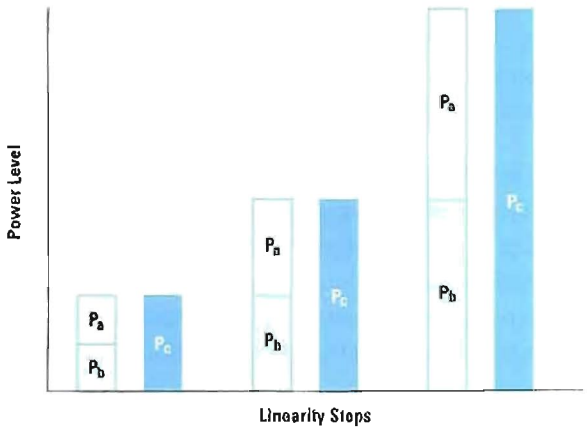
Because P_a and P_b are selected to be equal, the nonlinearity NL is:

$$NL = \frac{DL}{RL} = \frac{\frac{P_c}{P_a}}{\frac{P_a + P_b}{P_a}}$$

$$\approx \frac{\frac{P_c}{P_a}}{\frac{P_a + P_a}{P_a}} = \frac{P_c}{2P_a}.$$

Figure 6

The powers selected for the self-calibration procedure. The entire power range is calibrated in 3-dB steps.



The last term on the right side expresses the nonlinearity only in terms of values that are measured by the instrument under test. This means the calibration can be carried out without a standard instrument. Of course, it would also be possible to measure the real loss RL with a standard instrument that was itself calibrated for nonlinearity by a national laboratory. In any case, such a self-calibrating technique has a lot of advantages. There is no standard that must be shipped for recalibration at regular intervals, whose dependencies on external influences and aging contribute to the uncertainty of the calibration process, and that could be damaged, yielding erroneous calibrations.

Calibration of Laser Sources

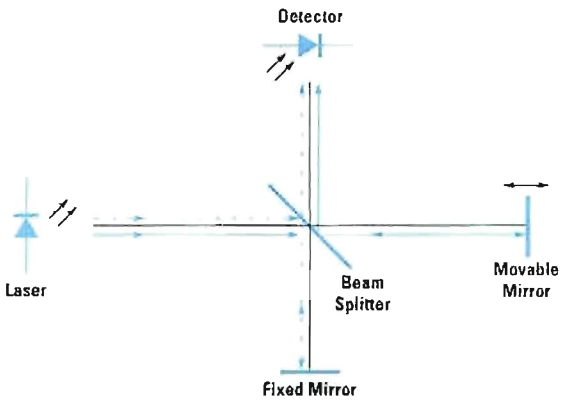
The last example will deal with calibration against natural physical constants and will be used to make some remarks about the determination of uncertainties. Having dealt with optical power sensors we will now focus on sources.

The most commonly used source in optical communications is the semiconductor laser diode. Only a few years ago, the exact wavelength emitted by the lasers was not so important. Three wavelength windows were widely used: around 850 nm, 1300 nm, and 1550 nm. 850 nm was chosen pragmatically; the first available laser diodes emitted at this wavelength. At 1300 nm, fiber pulse broadening is minimal in standard fibers, allowing the highest bandwidth, and at 1550 nm, fiber loss is minimum. As long as a fiber link or network operated at only one wavelength, as was mainly the case in recent years, and all its components exhibited only weak wavelength dependence, the exact wavelength was not of great interest. An accuracy of about 1 nm or even worse was good enough. Indeed, a lot of laser sources are specified with an accuracy of ± 10 nm.

The situation changed completely with the advent of wavelength-division multiplexing (WDM). This means that several different wavelengths (i.e., colors) are transmitted over one link at the same time, allowing an increase in bandwidth without burying new fibers. Since the single channels are separated by only 1.6 or 0.8 nm, one can easily imagine that wavelength accuracy becomes very important. Today wavelength accuracy on the order of a few picometers is required for WDM applications. The task of providing such wavelength accuracy for tunable laser sources like the HP 8168 Series is quite challenging.

Figure 7

Principle of the Michelson interferometer. A coherent beam is split by a beam splitter and directed into two different arms. After reflection at the fixed and movable mirrors, the superposition of the two beams is detected by a photodetector. Because the two beams are coherent, the superposition will give rise to an interference term in the intensity sum. Thus, moving the movable mirror will cause the intensity detected by the detector to exhibit minima and maxima. The distance between two maxima corresponds to a displacement of λ/2 of the movable mirror, where λ is the wavelength of the laser source. Measuring the necessary displacement to produce, say, ten maxima will allow a direct determination of the wavelength.



The accuracy should be the same over the whole wavelength range. The first question is how wavelength is measured.

First, it should be clear that wavelength means vacuum wavelength, which is effectively the frequency. Unfortunately, the wavelength of light varies under a change of the refractive index of the material it passes through. The tools for measuring nanometer distances are interferometers. Most of the commercially available wavelength meters are based either on the Michelson interferometer or the Fabry-Perot interferometer.[11] We will focus on the Michelson technique here.

The principle of the Michelson interferometer is shown in **Figure 7**. The challenge in the case of the Michelson interferometer is to measure the shift of the movable mirror. The required uncertainty of a few pm cannot, of course, be achieved by mechanical means. Instead, the interference pattern produced by light with a known wavelength is compared to the pattern of the unknown

source and by comparison the unknown wavelength can be calculated. The known light source is used here as a natural physical constant. Of course, the wavelength of this source must be independent of external influences. A stabilized He-Ne laser is often used because the central wavelength of its 633-nm transition is well-known and there are several methods of stabilizing the laser wavelength to an accuracy of fractions of 1 pm. Thus, such an interferometer-based wavelength meter is the ideal tool to calibrate a laser source for wavelength accuracy.

As mentioned above, every measurement consists of two parts: the measurement value and the corresponding uncertainty. The uncertainty of a specific measurement is estimated by a detailed uncertainty analysis.[12] As an example, **Table II** shows a fictitious uncertainty calculation for measuring the vacuum wavelength of a laser source with a Michelson interferometer in normal air. As shown, all relevant influences have to be listed and their individual contributions summed. The most difficult part is the determination of the values of the contributions. For statistical reasons, the sum is gained by a root-sum-square algorithm. This uncertainty calculation is the most important part of a calibration. The quality of the instrument to be calibrated is determined by the results of this analysis. In the case of a calibration process in a production environment, the specifications for all instruments sold depend on it.

Table II
Example of an Uncertainty Calculation

Component	Standard Deviation ±	Uncertainty ±
Internal Influences:		
Uncertainty of Reference Laser	0.08×10^{-6}	0.15×10^{-6}
Alignment, Diffraction	1.1×10^{-6}	2.2×10^{-6}
Fringe Counting Resolution	0.5×10^{-6}	1.0×10^{-6}
Total 1	$\mathbf{1.2 \times 10^{-6}}$	$\mathbf{2.4 \times 10^{-6}}$
Atmospheric Influences:		
Content of Carbon-dioxide (1/ppm)	0.17×10^{-9}	0.34×10^{-9}
Relative Humidity	1.95×10^{-9}	4.00×10^{-9}
Air Pressure	2.88×10^{-8}	5.76×10^{-8}
Temperature	5.71×10^{-9}	1.14×10^{-8}
Total 2	$\mathbf{2.94 \times 10^{-8}}$	$\mathbf{5.88 \times 10^{-8}}$
Uncertainty of Edlén Equation	$\mathbf{1.15 \times 10^{-8}}$	$\mathbf{2.30 \times 10^{-8}}$
Extension of Wavelength Limits	$\mathbf{1.15 \times 10^{-8}}$	$\mathbf{2.30 \times 10^{-8}}$
Influences from Source Under Test	$\mathbf{3.11 \times 10^{-10}}$	$\mathbf{5.38 \times 10^{-10}}$
Total	$\mathbf{1.2 \times 10^{-6}}$	$\mathbf{2.4 \times 10^{-6}}$

In **Table II**, the uncertainty caused by influences from the source under test is considered to be rectangularly distributed. Thus, the standard deviation is $1/\sqrt{3}$ times the uncertainty. In all other cases the standard deviation is known and a Gaussian distribution can be assumed. This leads to an uncertainty of two times the standard deviation at a confidence level of 95%. The Edlén equation is an analytical expression that describes the dependence of the refractive index of air on environmental conditions like temperature and humidity.

Wavelength or Frequency?

Finally, we'll consider the relation between wavelength and frequency. Because the definition of the meter is related to the definition of time and therefore frequency (and moreover, the frequency of light is invariant under all external conditions) it seems that it might be better to measure the frequency of the light emitted by a laser source rather than its wavelength.

It is now possible to measure a frequency of around 100 THz (10^{14} Hz). About two years ago, researchers from PTB succeeded in coupling the frequency of a laser emitting at 657 nm directly to the primary time standard (a Cesium clock as mentioned above) which oscillates at 9 GHz.[13] This coupled laser is a realization of a vacuum wavelength standard (and therefore a meter standard) which currently has an unequaled uncertainty of 9×10^{-13}.

Unfortunately, this method of directly measuring the frequency of light is very difficult, time-consuming, and expensive, so only a few laboratories in the world are able to carry it out. Thus, for the time being, wavelength will remain the parameter to be calibrated instead of frequency. Nevertheless, this is a good example of how lively the science of metrology is today.

Acknowledgments

I would like to thank my colleagues Horst Schweikardt and Christian Hentschel for their fruitful discussions.

References

1. W.R. Fuchs, *Bevor die Erde sich bewegte*, Deutsche Verlags Anstalt, 1975

2. W. Trapp, *Kleines Handbuch der Masse, Zahlen, Gewichte und der Zeitrechnung*, Bechtermuenz Verlag, Augsburg, 1996.

3. B.W. Petley, *The Fundamental Physical Constants and the Frontier of Measurement*, Adam Hilger, Ltd., 1995.

4. *Calibration of Fiber-Optic Power Meters*, IEC Standard 1315.

5. S. Ghezali, cited by E.O. Göbel, "Quantennormale im SI Einheitensystem," *Physikalische Blätter*, Vol. 53, 1997, p. 217.

6. *Calibration: Philosophy and Practice*, Fluke Corporation, 1994.

7. W. Budde, *Physical Detectors of Optical Radiation*, Academic Press, Inc., 1983.

8. W. Erb, Editor, *Leitfaden der Spektroradiometrie*, Springer Verlag, 1989.

9. C.L. Sanders, "A photocell linearity tester," *Applied Optics*, Vol. 1, 1962, p. 207.

10. K.D. Stock (PTB Braunschweig), "Calibration of fiber optical power meters at PTB," *Proceedings of the International Conference on Optical Radiometry*, London, 1988.

11. *Fiber Optics Handbook*, Hewlett-Packard Company, part no. 5952-9654.

12. *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization (ISO), 1993.

13. H. Schnatz, B. Lipphardt, J. Helmcke, F. Riehle, and G. Zinner, "Messung der Frequenz Sichtbarer Strahlung," *Physikalische Blätter*, Vol. 51, 1995, p. 922.

Bibliography

1. D. Derickson, *Fiber Optic Test and Measurement*, Prentice Hall PTR, 1998.

Online Information

Additional information about the HP lightwave products described in this article can be found at:

http://www.tmo.hp.com/tmo/datasheets/English/HP8153A.html

http://www.tmo.hp.com/tmo/datasheets/English/HP8168E.html

http://www.tmo.hp.com/tmo/datasheets/English/HP8168F.html

http://www.tmo.hp.com/tmo/datasheets/English/HP8156A.html

Appendix: Realization of Electrical Units

A lot of measurements today are carried out using electrical sensors. Therefore, it is important to know how the electrical units are realized and related to the mechanical quantities in the Système International d'Unités (SI). In optical fiber communication all power measurements use electrical sensors.

Historically, the electrical units are represented by the ampere among the seven SI base units. Unfortunately, it is not easy to build an experiment that realizes the definition of the ampere in the SI (see Table I on page 19). It is disseminated using standards for voltage and resistance using Ohm's law. Nevertheless, there is a realization for the ampere (see the current balance in **Figure 1**).

A coil carrying a current I exhibits a force F in the axial direction (z direction) if it is placed in an external inhomogeneous magnetic field H ($F \propto \partial H(z)/\partial z$). The force is measured by compensating the force with an appropriate mass on a balance. The uncertainty of such a realization is around 10^{-6}, the least accurate realization of all SI base units.

A similar principle is used for the realization of the volt, which is not a base but a derived SI unit. For realization one uses the following relation which can be derived from the SI definition of the voltage V (1 volt = 1 watt /ampere):

$$W = V \cdot I \cdot t$$

or

$$V = \frac{W}{I \cdot t}.$$

This expression results from the energy W that is stored in a capacitor that carries a electrical charge I·t. In this case one measures the force that is necessary to displace one capacitor plate in the direction perpendicular to the plate ($F = \partial W/\partial z$). Again the accuracy is about 10^{-6}. The approximate uncertainties in realization and representation of some SI units are listed in **Table I**.

Table I

Uncertainty of realization and representation of selected units of the SI system. Note that the representation of the voltage unit has a lower uncertainty than its realization.

Unit	Uncertainty of Realization	Uncertainty of Representation
kilogram	0	8×10^{-9}
meter	9×10^{-13}	3×10^{-11}
second	1×10^{-14}	1×10^{-14}
volt	1×10^{-7}	5×10^{-10}

However, in contrast to the ampere, for the volt there is a highly accurate method of representing the unit: the Josephson effect. The Josephson effect is a macroscopic quantum effect and can be fully understood only in terms of quantum physics. Only a brief description will be given here.

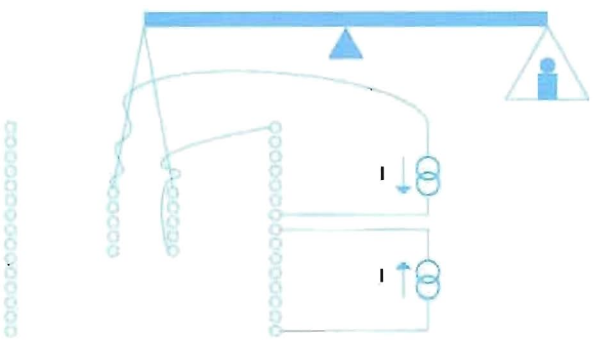
A Josephson element consists of two superconducting contacts that are separated by a small insulator. Astonishingly, there is an electrical dc current through this junction without any voltage drop across the junction. This is the dc Josephson effect. If one now applies an additional dc voltage at the junction, one observes an additional ac current with a frequency f that depends only on the applied dc voltage V and the fundamental constants e (charge of the electron) and h (Planck's constant):

$$f = \frac{2eV}{h}.$$

This effect can be used to reproduce a dc voltage with very high accuracy. One applies a dc voltage V and a microwave frequency f at the junction and observes a superconducting mixed current with ac and dc components. For certain voltages, and only for these

Figure 1

Principle of a current balance. The force experienced by the inner coil is compensated by an appropriate mass on the balance.



voltages, there is a resulting dc current in the time average. The condition is:

$$V = nf\frac{h}{2e}, \qquad n = 0, 1, 2, \ldots .$$

The uncertainty in reproducing a certain voltage V thus depends only on the uncertainty of f. As shown above, time and therefore frequency can be reproduced very accurately. One only has to control the microwave oscillator with a Cesium time standard. With this setup, an accuracy of 5×10^{-10} is achievable in the representation of the voltage unit, which is about 10,000 times better then the accuracy in realization—really a strange situation.

One solution would be to fix the value for e/h, similar to what was done for the new definition of the meter by holding the value for c constant. It would then be possible to replace the ampere as an SI base unit by the volt and define the volt using the Josephson effect.

Techniques for Higher-Performance Boolean Equivalence Verification

Harry D. Foster

The techniques and algorithms presented in this paper are a result of six years' experience in researching, developing, and integrating Boolean equivalence verification into the HP Convex Division's ASIC design flow. We have discovered that a high-performance equivalence checker is attainable through careful memory management, the use of bus grouping techniques during the RTL-to-equation translation process, hierarchical to flat name mapping considerations, subequivalence point cone partitioning, solving the false negative verification problem, and building minimal binary decision diagrams.

In 1965 Gordon Moore observed that the complexity and density of the silicon chip had doubled every year since its introduction, and accompanying this cycle was a proportional reduction in cost. He then boldly predicted—what is now referred to as Moore's Law—that this technology trend would continue. The period for this cycle was later revised to 18 months. Yet the performance of simulators, the main process for verifying integrated circuit design, has not kept pace with this silicon density trend. In fact, as transistor counts continue to increase along the Moore's Law curve, and as the design process transitions from a higher-level RTL (Register Transfer Language) description to its gate-level implementation, simulation performance quickly becomes the major bottleneck in the design flow.

In 1990, the HP Convex Division was designing high performance systems that used ASICs with gate counts on the order of 100,000 raw gates. During this period the HP Convex Division used a third-party simulator for both RTL and gate-level verification. This simulator had the distinction of being the fastest RTL simulator on the market, but it also suffered the misfortune of being one of the market's slowest gate-level simulators in our environment. Hence,

Harry D. Foster
Harry Foster joined the HP Convex Division in 1989 after receiving his MSCS degree from the University of Texas. An engineer/scientist, he has been responsible for CAD tool research and development, format verification, resynthesis, and a clock tree builder. Before coming to HP, he worked at Texas Instruments. He was born in Bethesda, Maryland, and likes traveling, weight training, and inline skating.

running gate-level simulations quickly became a major bottleneck in the HP Convex Division system design flow. In addition, these regression simulations could not completely guarantee equivalence between the final gate-level implementation and its original RTL specification. This resulted in a lack of confidence in the final design.

To address these problems, the HP Convex Division began researching alternative methods to regression simulation, resulting in a high-performance equivalence checker known as lover (LOgic VERifier).

Today, the HP Convex Division is delivering high-performance systems built with 0.35-μm CMOS ASICs having on the order of 1.1 million raw gates.[1] lover has successfully kept pace with today's silicon density growth, and has completely eliminated the need for gate-level regression simulations. Our very large system-level simulations are now performed entirely at the faster RTL level when combining the various ASIC models. This has been made possible by incorporating our high-performance equivalence checker into our design flow. We now have confidence that our gate-level implementation completely matches its RTL description.

Boolean Equivalence Requirements

Our experience has been that last-minute hand tweaks in the final place-and-route netlist require a quick and simple verification process that can handle a complete chip RTL-to-gate comparison. Such hand tweaks, and all hand-generated gates, are where most logic errors are inadvertently inserted into the design. The following list of requirements drove the development of the HP Convex Division's high-performance equivalence checker:

- Must support RTL-to-RTL (flat and hierarchical) comparisons during the early development phase.
- Must support both synthesizable and nonsynthesizable Verilog RTL constructs for RTL-to-gate comparisons.
- Must support a simple one-step process for comparison of the complete chip design (RTL-to-gate, gate-to-gate, hierarchical-to-flat).
- Must support the same Verilog constructs and policies defined for the entire HP Convex Division design flow (from our cycle-based simulator to our place-and-route tools), along with standard Verilog libraries.

Boolean Equivalence Verification

Boolean equivalence verification is a technique of mathematically proving that two design models are logically equivalent (e.g., a hierarchical RTL description and a flat gate-level netlist). This is accomplished by the following steps:

1. Compile the two designs. Convert a higher-level Verilog RTL specification into a set of lower-level equations and state points, which are represented in some internal data structure format. For a structural or gate-level implementation, the process includes resolving instance references before generating equations.

2. Identify equivalence points. Identify a set of controllability and observability cross-design relationships. These relationships are referred to as equivalence points, and at a minimum consist of primary inputs, primary outputs, and register or state boundaries.

3. Verify equivalence. Verify the logical equivalence between each pair of observability points by evaluating the following *equivalence equation*:

$$\neg\left(m_1(x_i) \oplus m_2(x_i)\right) \vee \left(x_i \cap D_e(x)\right) = 1. \qquad (1)$$

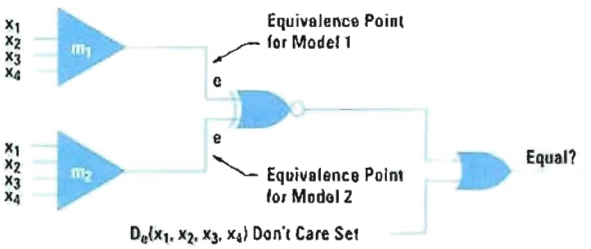
In the equivalence equation (equation 1), $\neg$ is the propositional logic NOT or negation operator, $\vee$ is the propositional logic OR or disjunction operator, $\oplus$ is the Boolean XOR operator, $\cap$ is the set intersection operator, m_1 represents the logic expression (or cone of logic*) for an observability equivalence point found in design model 1, and m_2 is the expression for the corresponding point found in design model 2. The variables x_i are the cone's input or controllability equivalence points for both models' logic expressions. Finally, $D_e(x)$ is known as the *don't care set* for the equivalence point e, and consists of all possible values of x for which the logical expressions m_1 and m_2 do not have to match. **Figure 1** graphically illustrates the process of proving equivalence between two observability equivalence points.

Another way to view **Figure 1** is to observe that if a combination of x_i can be found that results in $m_1(x_i)$ evaluating to a different value than $m_2(x_i)$, and x_i is not contained within the don't care set $D_e(x)$, then the two models are

* A cone of logic is the set of gates or subexpressions that fan into a single point, either a register or an equation variable

Figure 1

Proving equivalence.



formally proved different by the following *nonequivalence equation*:

$$\{m_1(x_i) \oplus m_2(x_i)\} \wedge \neg\{x_i \cap D_e(x)\} = 1, \qquad (2)$$

where $\wedge$ is the propositional logic AND or conjunction operator. Finding an x_i that will satisfy the nonequivalence equation, equation 2, is NP-complete.* In general, determining equivalence between two Boolean expressions is NP-complete. This means, as Bryant[2] points out, that a solution's time complexity (in the worst case) will grow exponentially with the size of the problem.

Instead of trying to determine inequality by finding a value for x that satisfies the nonequivalence equation, a better solution is to determine the equality of m_1 and m_2 through a specific or unique symbolic or graphical representation, such as an *ordered-reduced binary decision diagram* (OBDD).

Ordered-Reduced Binary Decision Diagrams

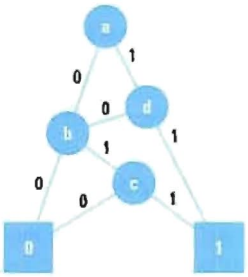
An ordered-reduced binary decision diagram (OBDD) is a directed acyclic graph representation of a Boolean function.[2] The unique characteristic of an OBDD is that it is a canonical-form representation of a Boolean function, which means that the two equations m_1 and m_2 in our previous example will have exactly the same OBDD representation when they are equivalent. This is always true if a common ordering of the controllability equivalence points is used to construct the OBDDs for m_1 and m_2.

Choosing an equivalence point input ordering can, in some cases, influence the resulting size of the OBDD. In addition, finding an optimal ordering of equivalence points in an attempt to minimize an OBDD is itself a co-NP-complete problem.[2,3] However, for most cases it is only necessary to find a good ordering, or simply one that works. Techniques for minimizing the size of an OBDD will be described later in this paper.

* An NP-complete or co-NP-complete problem is a relatively intractable problem requiring an exponential time for its solution. See reference 3.

Figure 2

OBDD (ordered-reduced binary decision diagram) with good ordering.



Figures 2 and 3 show two examples of an OBDD for the Boolean function (a & d) | (b & c), where & is the Boolean AND operator and | is the Boolean OR operator.

Performance Techniques

This section provides details and techniques for achieving high performance in the Boolean equivalence verification process. Some of the techniques can be incorporated into a user's existing design flow to achieve higher performance from a commercial equivalence checker.

Figure 3

OBDD with not-so-good ordering.

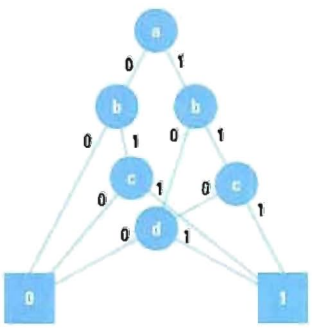
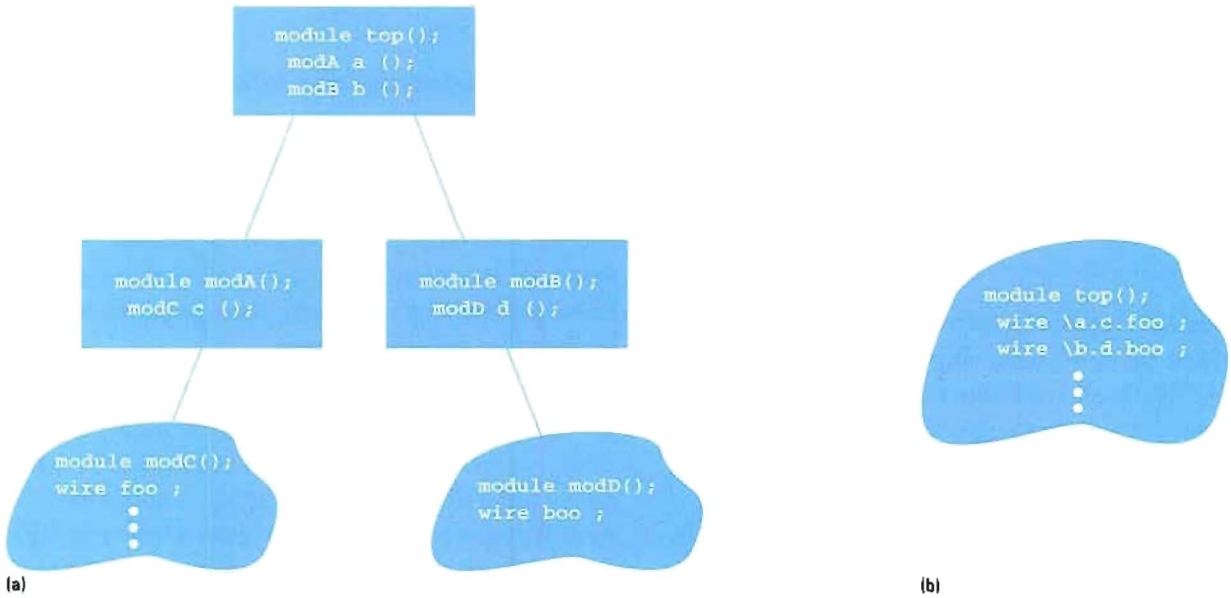


Figure 4

Hierarchical to flat name mapping. (a) Hierarchical RTL and gates. (b) Flat gate description.



Don't Throw Away Useful Information. When establishing an ASIC design flow, it is a benefit to view the entire flow globally—not just the process of piecing together various CAD tools (simulators, equivalence checkers, place-and-route tools, etc.). Why throw out valuable information from one process in the design flow and force another process to reconstruct it at a significant cost in performance?

At the HP Convex Division, we've designed our flow such that the identification of registers and primary inputs and outputs is consistent across the entire flow. The same hierarchical point in a Verilog cycle-based simulation run can be referenced in a flat place-and-route Verilog netlist without having to derive these common points computationally.

Name Mapping. lover will map the standard cross-design pairs of controllability and observability equivalence points directly as a result of the the HP Convex Division flow's naming convention. **Figure 4** helps illustrate how name mapping can be preserved between a hierarchical RTL Verilog tree of modules and a single flat gate-level description.

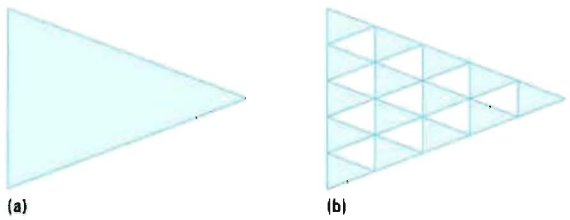
To support non-name-based mapping of equivalence points, that is, designs that violate the HP Convex Division flow naming conventions, lover will accept equivalence mapping files containing the two designs' net pair relationships. In addition, the user can provide a special filter function, which will automate the process of resolving cross-design name mapping for special cases.

For example, let f_1 be a special mapping function for design model 1 and f_2 be a special mapping function for design model 2. Then an equivalence point will be established whenever $f_1(\xi) = f_2(\zeta)$. This allows the two models' Verilog wire and register names (or strings ξ and ζ) to differ, but resolve to the same equivalence point through their special mapping functions.

Cone Partitioning. A technique known as *cone partitioning* is used to minimize the size of the OBDDs built during the verification process, since smaller OBDDs require significantly less processing time and consume much less memory than larger ones. Cone partitioning is the process of taking a large cone of logic and dividing it into a set of smaller sized cones. **Figure 5** helps illustrate this concept.

Figure 5

Cone partitioning.(a) Large cone of logic. (b) Set of partitioned cones.



The two designs verified by lover are stored internally in a compact and highly efficient net/primitive relational data structure. OBDDs are built from the relational data structure only on demand for the specific partitioned cone of logic being proved, and then immediately freed after their use. This eliminates the need to optimize the specific cone's OBDD into the entire set of OBDDs for a design. In addition to achieving higher performance during the verification process, this ensures that any differences found between the partitioned cones tends to be isolated down to either a handful of gates or a few lines of RTL code. This greatly simplifies the engineer's debug effort.

Another advantage of cone partitioning is that it becomes unnecessary to spend processing time minimizing equations for large cones of logic, since they are automatically decomposed into a set of smaller and simpler cones.

In general, cones of logic are bounded by *equivalence points*, which consist of registers and ASIC input and output ports. However, lover takes advantage of a set of pairs of internally cross-design equivalent relationships (e.g., nets or subexpressions), which we refer to as *subequivalence points*, to partition large cones. Numerous methods have been developed to compute a set of cross-circuit subequivalence points based on the structural information or modular interfaces.[4] Costlier ATPG (automatic test pattern generator) techniques are commonly used to identify and map subequivalence points between designs lacking a consistent naming convention. lover determines subequivalence points directly without performing any intensive computations by taking advantage of the consistent name mapping convention built into the the HP Convex Division design flow. Numerous subequivalence points are derived directly from the module's hierarchical boundaries.

In addition, lover attempts to map the Verilog module's internal wire and register variable names between designs. For example, Synopsys will unroll the RTL Verilog wire and register bus ranges as follows:

```
wire [0:3] foo;
```

will be synthesized into gates with the following unrolled names:

```
wire \foo[0], \foo[1], \foo[2], \foo[3];
```

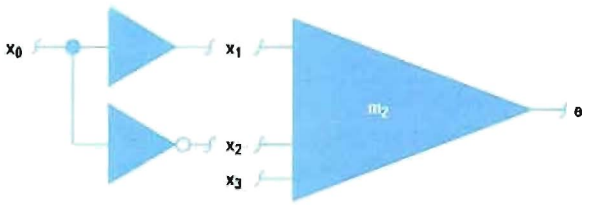
lover recognizes Synopsys' unrolled naming convention and will map these points back to the original RTL description. This results in a significantly better cone partition than limiting the subequivalence points to only the structural or module interface.

The performance gains achieved through cone partitioning are highly variable and dependent on a circuit's topology. Some modules' measured performance gains have been on the order of 20 times, while other modules have suffered a performance loss of 1.5 times when applying a maximum cone partition. The performance gains tend to increase as the proof moves higher up the hierarchy of modules (due to larger cones). In general, we've observed that cone partitioning contributes to about a 40% increase in performance over the entire chip. More important, cone partitioning allows us to prove certain topologies containing large cones of logic that would be impractical to prove using OBDDs.

OBDD Input Ordering. Cone partitioning has the additional advantage of simplifying the ordering of OBDDs by reducing the size of the problem. lover uses a simple topological search through these smaller partitioned cones, which yields an excellent variable ordering for most cases. This search method was first proposed by Fujita,[5] and consists of a simple depth-first search through a circuit's various logic levels, starting at its output and working backwards towards its controllability equivalence points. The equivalence points encountered first in the search are placed at the beginning of the OBDD variable ordering.

Solving the False Negative Problem. Cone partitioning techniques used to derive cross-circuit subequivalence points can lead to a proof condition known as a *false negative*. This quite often forces the equivalence checker into a more aggressive and costlier performance mode to complete the proof. We have developed a method of identifying a false negative condition while still remaining in the faster

Figure 6
False negative.



$m_1(x_1, x_2, x_3) \neq m_2(x_1, x_2, x_3)$
when $x_1 = 1$ and $x_2 = 1$

name-mapping mode throughout the entire verification process.

One type of false negative can occur when the RTL specifies a don't-care directive to the synthesis tool, and the equivalence checker does not account for the don't care set $D_c(x)$ in equation 1.

Another more troublesome type of false negative can occur when the synthesis process recognizes a don't care optimization opportunity not originally specified in the RTL. This can occur when the synthesis step is applied to a subexpression, which then takes an optimization advantage over the full cone of logic (e.g., generating gates for a subexpression with the knowledge that a specific combination of values on its inputs is not possible).

Figure 6 provides a simple example of this problem. It is possible that the partitioned cone m_2's synthesized logic will be optimized with the knowledge that x_1 and x_2 are always mutually exclusive. This can lead to a false negative proof on the partitioned cones $m_1(x_1, x_2, x_3)$ and $m_2(x_1, x_2, x_3)$ for the impossible case $x_1 = 1$ and $x_2 = 1$.

However, if the subequivalence points that induced the false negative condition are removed from the cone partition boundaries of m_1 and m_2 (e.g., x_1 and x_2), the resulting larger cone partition $m_1'(x_0, x_3)$ is easily proved to be equivalent to $m_2'(x_0, x_3)$. **Figure 7** helps illustrate how the false negative condition can be eliminated by viewing a larger partition of the cone.

There are numerous situations that can induce a false negative condition, and most are much more complex than the simple example provided in **Figure 6**. Iover has algorithms built into it that will detect and remove all false negative conditions. These algorithms are invoked only when nonequivalence has been determined between observability points.

The algorithm used to solve a false negative performs a topological or breadth-first search through the levelized cone of logic, removing the first input or controllability equivalence point encountered. It then reproves the resulting larger cone. The following steps describe the algorithm used by Iover to remove a false negative condition:

1. Levelize the cone of logic by assigning the observability equivalence point the number 0. Then, step back to the next level of logic in the cone, assigning a logic level number to each logic primitive and net. Continue this process back through all levels of logic until reaching the cone's controllability equivalence points.

2. Remove the lowest level of logic (or numbered) controllability subequivalence point, resulting in a larger cone partition.

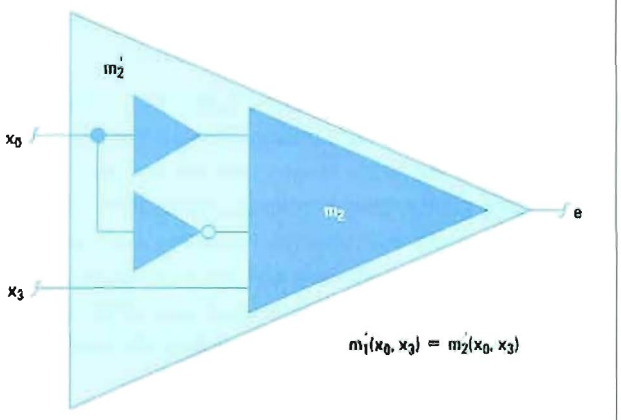
3. Relevelize the new large cone of logic.

4. Identify and order the new set of input controllability equivalence points.

5. Try reproving the new larger cone partition with the new set of controllability subequivalence points.

6. If the new cone partition is proved nonequivalent and we can continue removing subequivalence points (i.e., we haven't reached a register or ASIC port boundary) go to

Figure 7
Solving the false negative problem.



step 2. Otherwise, we are done. If step 5 is proven, the two cones are equivalent.

Process Memory Considerations. Most high-performance tools require their own memory management utility to reduce the system overhead time normally associated with searching a process's large memory allocation table. lover implements three methods of managing memory: (1) recycle high-use data structure elements during compilation, (2) ensure that memory is unfragmented when building OBDDs (i.e., at the start of each cone's proof), and (3) maintain and manipulate a single grouped structure representation for equations containing buses.

- High-Use Data Structure Recycling. The various data structures (or C typedefs) used during the compilation process can be recycled when it becomes necessary to free them. Recycling is a technique of linking the specific structure types together into a *free list*. Later compilation steps can tap into these lists of high-use data structure elements and not incur any of the system overhead normally associated with allocating or freeing memory. We have observed performance gains in the order of 1.25 times for small designs and up to 2 times for larger designs by using data structure recycling techniques.
- OBDD Unfragmented Memory Management. A block of memory should be reserved for use by the OBDD memory management utilities. Once a proof is complete for a partitioned cone, its memory block can be quickly reset to its original unfragmented state by simply resetting a few pointers. Working with a block of unfragmented memory increases the chances of fitting a cone's OBDD into the system's cache. The performance gains achieved by controlling the fragmentation of memory are significant, but hard to quantify. In general, we have observed that the performance of manipulating OBDDs degrades linearly as memory fragmentation increases.
- Grouping Structures for Verilog Buses. Care should be taken to retain the buses within an equation as a single grouped data structure element for as long as possible. Expanding an equation containing buses into its individual bits too soon will result in a memory explosion during the compilation process and force unnecessary elaboration on the equation's replicated data structures (i.e., process duplication while manipulating the individual bits for a bused expression).

The Don't Care Set. As a final comment on performance techniques, we need to point out that the HP Convex Division design flow has a requirement of simulating the ATPG vector set on the RTL. This helps flush out any ATPG model library problem and provides an additional sanity check and assurance that Boolean equivalence verification was performed on the entire design. An additional benefit of verifying the ATPG vectors at the RTL level is a potential tenfold speedup in simulation performance compared to a gate-level simulation of the vector set.

To support RTL simulation of the ATPG vectors, the Verilog control structures (e.g., case and if) must be fully specified or defaulted to a known value. lover takes advantage of this flow requirement and makes no attempt to gather and process the don't-care set $D_c(x)$ in equation 1. We have developed linting tools within our flow to ensure that these control structures are fully specified. In addition, lover detects violations of this ATPG RTL requirement.

Performance Gains

This section demonstrates the performance gains that can be achieved through techniques described in this paper. The benchmarks for **Tables I** and **II** were performed on an HP S-Class technical server (a 16-way symmetric multiprocessing machine based on the HP PA 8000 processor with 4G bytes of main memory). The single-threaded performance times provided for lover include a composite of the RTL and gate-level model compilation times, as well as the verification step (unlike most commercial tools, compilation and verification are performed within a single process).

Table I describes four recently designed 0.35-µm CMOS ASICs built for Hewlett-Packard's Exemplar S-Class and X-Class technical servers. These times represent a comparison of a full hierarchical RTL Verilog model to its complete-chip flat gate-level netlist.

Table II presents the gate-to-gate run-time performance for the same four ASICs. These times represent a complete chip hierarchical gate-level model (directly out of synthesis) compared to its final hand-tweaked flat place-and-route netlist.

Table I
RTL-to-Gate lover Results

Chip Name	Size (kgates)	Minutes	GBytes
Processor Interface	550	116	1.3
Crossbar	500	26	1.1
Memory Interface	570	68	1.3
Node-to-Node Interface	300	56	1.0

Table II
Gate-to-Gate lover Results

Chip Name	Size (kgates)	Minutes	GBytes
Processor Interface	550	20	1.2
Crossbar	500	9	0.9
Memory Interface	570	20	1.2
Node-to-Node Interface	300	10	0.9

Additional Performance Gains

The Hewlett-Packard Convex Division is in the business of researching and developing symmetric multiprocessing high-performance servers. Historically, most vendors' CAD tools have lagged behind the design requirements for our next-generation systems. To solve the vast and escalating problems encountered during the design of these systems, the HP Convex Division has begun research in the area of parallel CAD solutions.[6] A prototype multithreaded equivalence checker (p-lover) has been developed to investigate the potential performance gains achievable through parallel processing.

The prototype multithreaded equivalence checker is based on lover's single-threaded proof engine. A front-end input compiler and data structure emulation engine was developed to feed the parallel threads with partitioned cones of logic for verification.

Figure 8 shows the speedup factors we obtained with p-lover when launching two, four, and six threads. Note the superlinear performance we were able to achieve with two threads (see reference 6 for a discussion of superlinear behavior). Each thread only contributed a 4% increase in the overall process memory size. This can be attributed to a single program image for the compiled net and primitive relational data structures, which were stored in globally shared memory.

The following is a list of multithreaded tool design considerations we've identified while developing our prototype equivalence checker:

- Long Threads. To reduce the overhead incurred when launching threads, the full set of observability equivalence points needs to be partitioned into similar-sized subsets or lists for each thread (i.e., threads should be launched to prove a list of cones and not a single point). **Figure 9** helps illustrate this idea.
- Thread Balancing. When a thread finishes proving its list of observability equivalence points, the remaining threads should rebalance their list of unproven cones to maintain maximum tool performance.
- Thread Memory Management. Each thread must have its own self-contained memory management and OBDD utility.

Figure 8
Multithreaded performance.

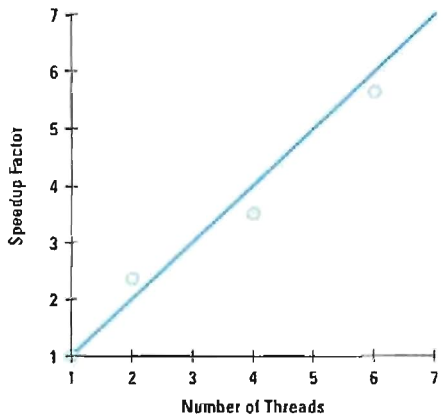
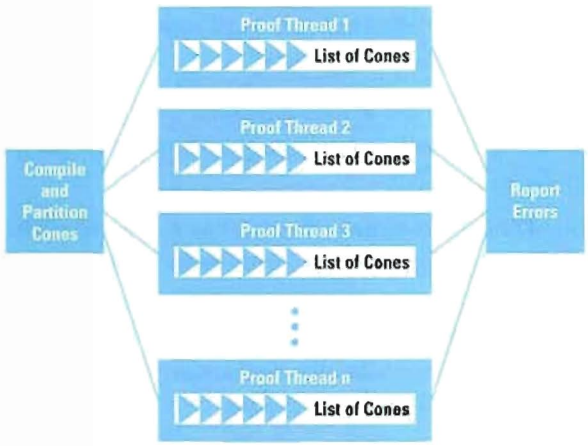


Figure 9
Multithreaded equivalence checker.



- System Resources and Locking. All I/O and logging must be eliminated from the individual launched threads. Otherwise, the locking and unlocking schemes built into the system resource's critical sections will dramatically degrade the tool's performance. Errors can be flagged in internal data structures and reported after all threads have finished processing their individual lists of equivalence points.

A production version of p-lover will require additional research to eliminate some of the locking requirements necessary when addressing globally shared memory. In particular, solving the false negative problem in a multithreaded environment will require some additional thought. However, the potential performance gains obtainable through a multithreaded equivalence checker are attractive.

Conclusion

Boolean equivalence verification, an integral process within the HP Convex Division's ASIC design flow, bridges the verification gap between an ASIC's high-level RTL used for simulation and its place-and-route gate-level netlist. We have presented techniques in this paper that have contributed to the development of a Boolean equivalence checker with performance on the order of 100 times faster than many currently available commercial tools. Even a commercial equivalence checker will benefit substantially if its users understand a few of the techniques we have presented and apply them directly to their design flow (e.g., name mapping, subequivalence points, and cone partitioning concepts). Finally, we have presented data from a prototype multithreaded equivalence checker to illustrate that an even higher performance level is attainable through a parallel solution.

References

1. L. Bening, T. Brewer, H. Foster, J. Quigley, B. Sussman, P. Vogel, and A. Wells, "Physical Design of 0.35-µm Gate Arrays for Symmetric Multiprocessing Servers," *Hewlett-Packard Journal*, Vol. 48, no. 2, April 1997, pp. 95-103.

2. R. Bryant, "Graph-Based Algorithms for Boolean Function Manipulation," *IEEE Transactions on Computers*, August 1986, pp. 677-691.

3. M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, 1979.

4. E. Cerny and C. Mauras, "Tautology Checking Using Cross-Controllability and Cross-Observability Relations," *IEEE Transactions on Computers*, January 1990, pp. 34-37.

5. M. Fujita, H. Fujisawa, and N. Kawato, "Evaluation and Improvements of Boolean Comparison Method Based on Binary Decision Diagrams," *International Conference on Computer-Aided Design*, November 1988, pp. 2-5.

6. L. Bening, "Putting Multi-Threaded Simulation To Work," accessible on the World Wide Web at URL: http://www.hp.com/wsg/tech/supercon96/index.html

On-Chip Cross Talk Noise Model for Deep-Submicrometer ULSI Interconnect

Samuel O. Nakagawa
Dennis M. Sylvester
John G. McBride
Soo-Young Oh

A simple closed-form model for calculating cross talk noise on signal lines in deep-submicrometer interconnect systems has accuracy comparable to SPICE for an arbitrary ramp input rate. Interconnect resistance, interconnect capacitance, and driver resistance are all taken into account. The model is suitable for rapid cross talk estimation and signal integrity verification.

Interconnect geometry in deep-submicrometer integrated circuit technologies is being aggressively scaled down for wiring density, leading to high aspect ratios in metal lines.[1-3] For example, according to the Semiconductor Industries Association roadmap, metal aspect ratio is expected to reach 2:1 in the 0.25-μm technology generation and 3:1 by the year 2004. As a result of the increasing metal aspect ratio, capacitive coupling between neighboring signal lines increases and more cross talk noise is generated. With increasing edge rates and ground bounce in advanced technologies, cross talk will become a pervasive signal integrity issue.

Traditionally, SPICE simulations have been used to estimate cross talk noise in the signal lines. Although accurate, these simulations are time-consuming. When the number of signal lines easily exceeds one million as it does in today's advanced microprocessors, SPICE simulations are too inefficient to carry out for each line. A rapid and accurate cross talk noise estimation alternative is needed to ensure acceptable signal integrity in a limited design cycle time. In reference 4, a closed-form model based on RC transmission line analysis is presented. However, the driver modeling is not discussed and the analysis is limited to step response. Another model approximates the driver with a resistor and a ramp voltage source,[5] but signal line resistance is neglected. These approaches lack the accuracy needed in deep-submicrometer interconnect analysis.

In this paper we present a closed-form cross talk noise model with accuracy comparable to that of SPICE for an arbitrary ramp input rate. Interconnect resistance, interconnect capacitance, and driver resistance are all taken into account.

Model for Timing-Level Analysis

First, we derive a closed-form expression for cross talk noise when the rise time at the aggressor *output* is known. A circuit schematic of this model is shown in **Figure 1**. In a typical electronic design automation environment, circuit timing simulators can provide a rapid and accurate estimate of the signal rise time at the output of a driver. This information significantly simplifies our driver modeling. An aggressor transistor is treated as a ramp voltage source, V_s ($= V_{dd}/T_r$). A victim transistor is modeled as an effective resistance, R_{vd}. This value is taken to be the linear resistance for the p- or n-channel MOSFET, depending on the victim line's logic state. This driver resistance and the victim line resistance, R_{vi}, are lumped into a single resistance, R_v. R_a is the line resistance of the aggressor. C_a and C_v are the lumped capacitance for the aggressor line and victim line, respectively, and C_c is the coupling capacitance between the lines (**Figure 2**).

Based on the circuit in **Figure 1**, the cross talk noise voltage V_x as a function of time t is expressed as:

$$V_x = \frac{R_v C_c V_{dd}}{\tau_0 T_r}\left(\tau_0 + \tau_1 e^{-t/\tau_1} - \tau_2 e^{-t/\tau_2}\right) \quad (1)$$

for $0 \le t \le T_r$, and as:

$$V_x = \frac{R_v C_c V_{dd}}{\tau_0 T_r}\left\{\tau_1\left[e^{-t/\tau_1} - e^{-(t-T_r)/\tau_1}\right]\right\} - \frac{R_v C_c V_{dd}}{\tau_0 T_r}\left\{\tau_2\left[e^{-t/\tau_2} - e^{-(t-T_r)/\tau_2}\right]\right\} \quad (2)$$

for $T_r \le t$, where V_{dd} is the supply voltage, T_r is the rise time at the output of the aggressor driver, and

$$\tau_0^2 = \left[R_a(C_a + C_c) + R_v(C_v + C_c)\right]^2 - 4R_vR_a(C_vC_c + C_vC_a + C_cC_a) \quad (3)$$

$$\tau_1 = \frac{\left[2R_vR_a(C_vC_c + C_vC_a + C_cC_a)\right]}{\left[R_a(C_a + C_c) + R_v(C_v + C_c) + \tau_0\right]} \quad (4)$$

$$\tau_2 = \frac{\left[2R_vR_a(C_vC_c + C_vC_a + C_cC_a)\right]}{\left[R_a(C_a + C_c) + R_v(C_v + C_c) - \tau_0\right]}. \quad (5)$$

The peak voltage, $V_{x,max}$, always occurs when $T_r \le t$. Therefore, by differentiating equation 2 with respect to t, we obtain:

$$V_{x,max} = \frac{R_vC_cV_{dd}}{\tau_0 T_r}\left[\varphi_1\tau_1\left(\frac{\varphi_1}{\varphi_2}\right)^{\tau_2/(\tau_1-\tau_2)} - \varphi_2\tau_2\left(\frac{\varphi_1}{\varphi_2}\right)^{\tau_1/(\tau_1-\tau_2)}\right], \quad (6)$$

where $\varphi_1 = \exp(-T_r/\tau_1) - 1$ and $\varphi_2 = \exp(-T_r/\tau_2) - 1$. For a sufficiently slow rise time ($T_r \gg \tau_2$), $V_{x,max}$ approaches the limit of $R_vC_cV_{dd}/T_r$. Also, for a special case where $R_a = R_v$, $C_a = C_v$ and $T_r = 0$, equation 6 reduces to a simple model presented by Sakurai.[4]

Figure 1

A circuit diagram for the cross talk models in this paper. In the timing-level model, the aggressor driver is modeled as a ramp input, V_s ($= V_{dd}/T_r$), and R_a is the line resistance of the aggressor. In the transistor-level model, V_s is the ramp input to the aggressor driver, and R_a is the sum of aggressor driver resistance R_{ad} and the aggressor line resistance R_{ai}. In both models R_v is the sum of the line and driver resistances. C_a and C_v are the lumped ground capacitances for the aggressor line and victim line, respectively, and C_c is the lumped coupling capacitance.

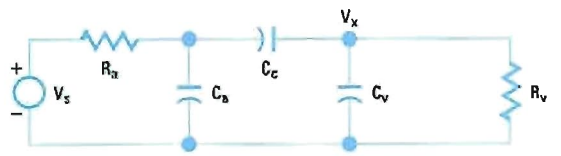


Figure 2

A cross-sectional view of two lines above a ground plane considered in this study. The coupling capacitance C_c is the source of on-chip cross talk noise. It is a significant fraction of total interconnect capacitance in deep-submicrometer interconnect technology.

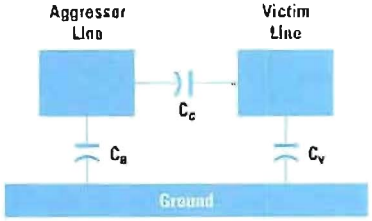
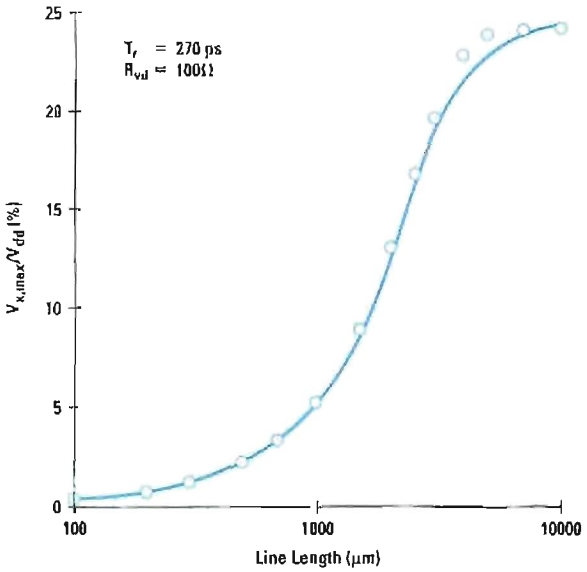


Figure 3

Normalized cross talk noise voltage as a function of interconnect length. The model prediction is represented by a solid line and the SPICE simulations are represented by circles. The error of the model compared with SPICE is less than 10%. Cross talk noise increases sharply for line lengths over 1000 μm before reaching a saturation value.



$$V_{x,max} = \frac{V_{dd}}{2}\frac{C_c}{C_a + C_c}. \quad (7)$$

The accuracy of the model of **Figure 1** is demonstrated in **Figure 3** and **Figure 4** for a representative cross-sectional geometry of a global line in 0.25-μm technology.[6] To account appropriately for the distributed nature of the interconnect RC network, the lumped ground capacitances C_a and C_v are scaled by a factor of 0.5 based on the Elmore delay model.[7] The lumped coupling capacitance C_c, on the other hand, is scaled by a semi-empirical, technology independent factor, $\alpha = (1-\beta)\{\exp(-T_r/\tau_0)\} + \beta$. The parameter β accounts for the presence of the victim driver resistance, and is given by $\beta = 0.5[1 + R_{vd}/(R_{vi} + R_{vd})]$. β is unity for a device-dominated case in which shielding resulting from interconnect resistance is negligible, and it decreases monotonically to 0.5 as interconnect becomes more dominant. The scaling factor α is equal to β for a slow rise time, but monotonically approaches unity for a sufficiently fast rise time. In **Figure 3** line length is varied to cover both the device-dominated case (interconnect length ≤ 1000 μm) and the interconnect-dominated case (interconnect length ≥ 3000 μm). The model prediction matches the SPICE results very well. The agreement is also excellent in **Figure 4**, where the rise time varies over a wide range.

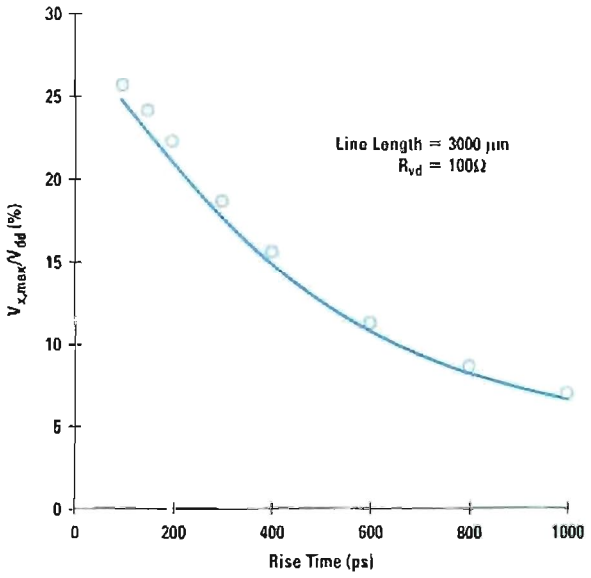
Since all parameter values in equations 1 through 7 are readily available from the timing analysis tools, this model forms an excellent basis for a cross talk screening tool at the timing level. The nonproblematic signal lines can be quickly identified and filtered with this model. Only those lines that potentially violate noise margin need further detailed simulations. The efficiency of signal integrity verification can be significantly improved by this scheme.

Model for Transistor-Level Analysis

Next, we consider a case in which the rise time to the *input* of the aggressor transistor is known. In this case the rise time at the *output* of the aggressor transistor is first computed as a function of the input rise time using a technology dependent function. Then equation 6 is used to calculate the maximum cross talk noise.

Figure 4

Normalized cross talk noise voltage as a function of rise time. The model prediction is represented by a solid line and the SPICE simulations are represented by circles. The error of the model compared with SPICE is less than 10%. Cross talk noise is a strong function of rise time and is a serious concern when rise time becomes less than 200 ps in deep-submicrometer technologies.



The rise time at the output of the aggressor transistor, T_r, is expressed as:

$$T_r = T_{ri} + T_{rw} + T_{rc}, \quad (8)$$

where T_{ri}, T_{rw}, and T_{rc} account for the intrinsic delay, input slope, and interconnect loading dependencies, respectively.

Intrinsic Delay Dependency. The intrinsic delay dependency of the aggressor output rise time, T_{ri}, is empirically expressed as:

$$T_{ri} = k_i \frac{V_{dd}}{I_{d,sat}} C_j, \quad (9)$$

where V_{dd} is the supply voltage, $I_{d,sat}$ is the saturation source-to-drain current, and C_j is the junction capacitance. The T_{ri} term is usually small (~5 ps) and is independent of the aggressor input rise time. It is also independent of device size; both $I_{d,sat}$ and C_j increase as the driver size increases, canceling each other. The term k_i is a fitting parameter. Our study shows that $k_i = 0.4$ for many different technology generations. The T_{ri} term is important only for the following cases:

- Older technology generations for which the RC of a device is significant
- A transistor with extremely small loads
- Very fast input rise time (<35 ps).

None of these cases is of practical interest in deep-submicrometer technologies.

Input Slope Dependency. The input slope dependency of the aggressor output rise time, T_{rw}, is a linear function of the aggressor input rise time, T_{ra}:

$$T_{rw} = k_w T_{ra}, \quad (10)$$

where k_w is a technology dependent fitting parameter and is typically between 0.1 and 0.2 for deep-submicrometer technologies. This linear relationship holds extremely well for the practical values of T_{ra} ranging from 50 ps to 500 ps, as shown in **Figure 5**.

This input slope dependency term can be very significant, especially for slower input signals and small load capacitances. For instance, for a 1-mm line with T_{ra} = 160 ps, T_{rw} can be as high as 30% of T_r.

Interconnect Loading Dependency. The third term in equation 8 results from the charging and discharging of the interconnect through the aggressor driver. Since the driver goes through both the saturation and linear modes of operation during the charging and discharging, T_{rc} has two corresponding terms:[8]

$$T_{rc} = \gamma \xi C_i \left[\frac{V_t - 0.1V_{dd}}{I_{d,sat}} \right] + \gamma \xi C_i \left[\frac{1}{k(V_{dd} - V_t)} \ln \left(\frac{19V_{dd} - 20V_t}{V_{dd}} \right) \right], \quad (11)$$

where C_i is the interconnect capacitance, V_t is the threshold voltage of the driver, and k is the device transconductance, which is given by:

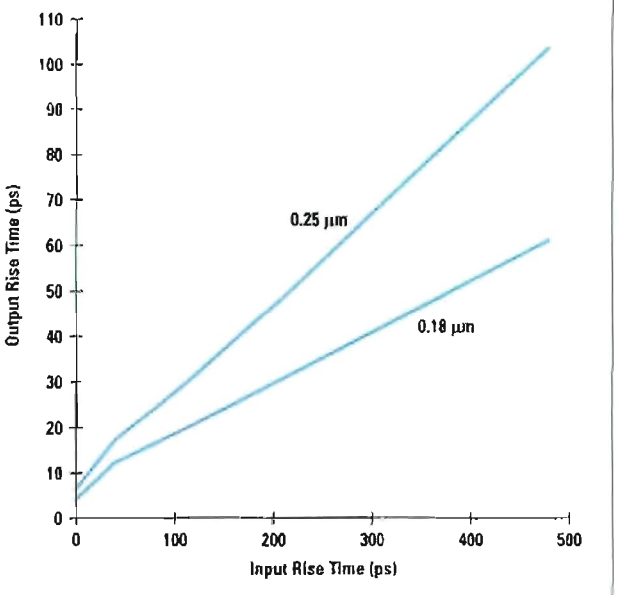
$$k = \frac{2I_{d,sat}}{(V_{dd} - V_t)^2}. \quad (12)$$

The term γ is an empirical expression to account for capacitance shielding caused by interconnect resistance, and is given by:

$$\gamma = 1 - \left[\frac{R_{ai}}{R_{ai} + R_{ad}\sqrt{3}} \right]^{-1}, \quad (13)$$

Figure 5

Unloaded output rise time as a function of input rise time of the aggressor driver for 0.25-μm and 0.18-μm technologies. A linear relationship holds well for input rise time above 50 ps.



where R_{ai} and R_{ad} are the aggressor line resistance and driver resistance, respectively. The term ξ is an empirical constant accounting for the loss due to short-circuit current and is typically equal to 1.2. Short-circuit current does not serve to charge or discharge the line.

The first term in equation 11 describes the transient in the saturation region, but is typically much smaller than the second term because of the large current drive and the small voltage swing in the saturation region. The second term is for the transient in the linear region, and is technology dependent only on the ratio of V_t/V_{dd}.

Benchmark of Model. Rise time values at the output of the aggressor driver calculated based on equations 8 through 13 for a wide range of interconnect lengths are compared with SPICE simulations in **Figure 6**. The model predictions are in good agreement with SPICE simulations. The modeling error compared with SPICE is shown to be less than 10% in **Figure 7**.

As a comparison, the rise time estimation based on a previously published model[9] is also shown in **Figure 6**. This model neglects T_{ri} and T_{rw}. Also, interconnect capacitance shielding and short-circuit current in T_{rc} are not considered. As a result, this model significantly underestimates T_r for short lines and overestimates T_r for long lines.

Figure 7

Rise time estimation error of models compared with SPICE. Error for the model in this paper is ± 10%. A model based on reference 9 produces a significant error.

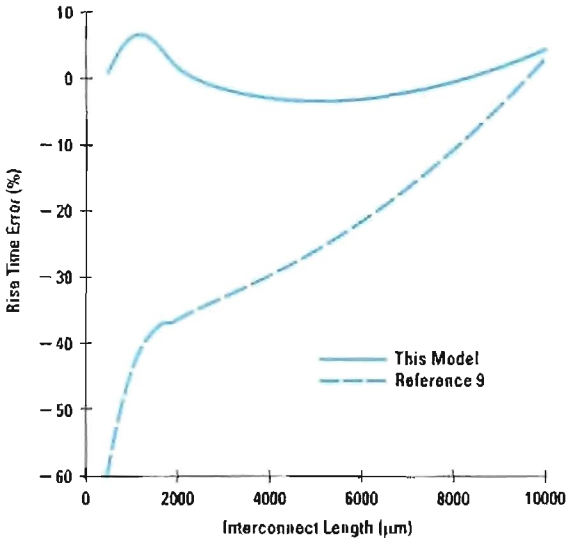
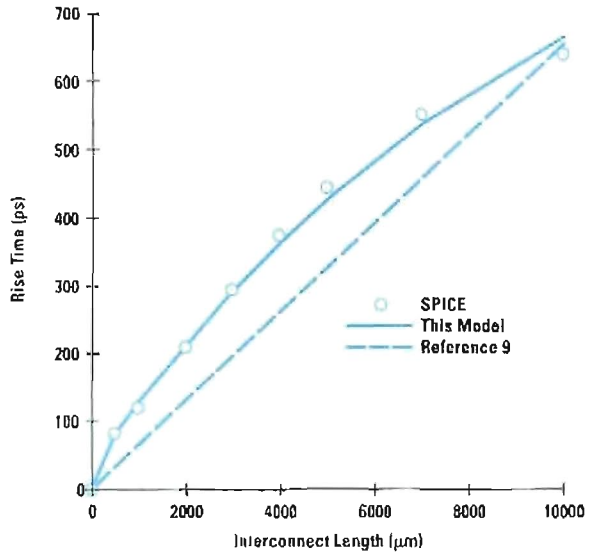


Figure 6

Comparison of rise time estimates. The model in this paper is in excellent agreement with SPICE results. A model in reference 9, which neglects T_{ri} and T_{rw} in equation 8 as well as interconnect capacitance shielding and short-circuit current in equation 11, exhibits large error over a wide range of interconnect lengths.

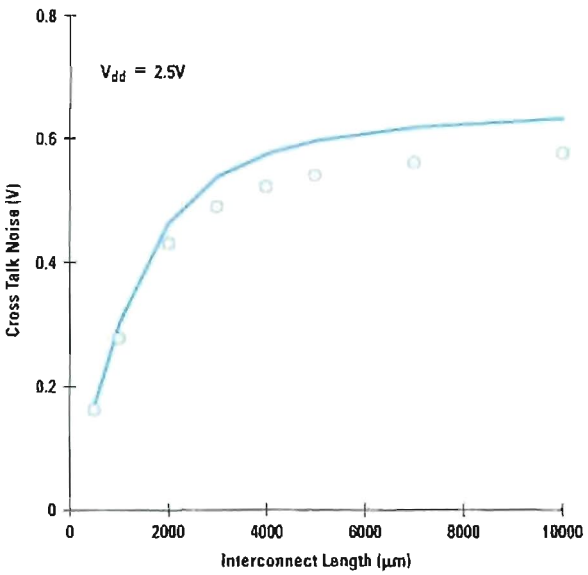


Once the rise time at the output of the aggressor driver is calculated, the corresponding peak cross talk noise can be computed based on equation 6. In **Figure 8**, modeled and SPICE peak cross talk noise values are plotted as a function of interconnect length. Our model provides a very smooth curve and matches the SPICE result within 10% over a wide range of interconnect lengths.

The technology dependent fitting coefficients in equations 9 through 11 can be found easily by running SPICE for several calibration cases. With the calibrated coefficients, this model rapidly generates accurate cross talk noise estimation for various driver sizes, interconnect loads, and rise times. The model is an attractive alternative to SPICE when many transistor-level simulations for cross talk noise are needed, including the case of quick screening mentioned earlier.

Figure 8

Estimated cross talk noise voltage as a function of interconnect length. The model prediction is represented by a solid line and the SPICE simulations are represented by circles. The model is accurate (less than 10% error) over a wide range of interconnect lengths.



Conclusion

In this paper we have analyzed the accuracy and applicability of a simple closed-form model for calculating cross talk noise on signal lines in deep-submicrometer interconnect systems. With appropriate scaling and calibration of the model coefficients, it was shown that the model is sufficiently accurate for cross talk analysis. All model parameters and coefficients are readily available. Therefore, the model is suitable for rapid cross talk estimation and signal integrity verification.

References

1. M. Bohr, S.U. Ahmed, L. Brigham, R. Chau, R. Gasser, R. Green, W. Hargrove, E. Lee, R. Natter, S. Thompson, K. Weldon, and S. Yang, "A high performance 0.35μm logic technology for 3.3V and 2.5V operation," *International Electron Devices Meeting Technical Digest*, 1994, pp. 273-276.

2. *The National Technology Roadmap For Semiconductors*, Semiconductor Industry Association, 1994.

3. K. Rahmat, O.S. Nakagawa, S.-Y. Oh, and J. Moll, "A scaling scheme for interconnect in deep-submicron processes," *International Electron Devices Meeting Technical Digest*, 1995, p. 245.

4. T. Sakurai, "Closed-form expression for interconnection delay, coupling, and cross talk in VLSIs," *IEEE Transactions on Electron Devices*, Vol. 40, 1993, pp. 118-124.

5. E. Sicard and A. Rubio, "Analysis of cross talk interference in CMOS integrated circuits," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 34, 1992, pp. 124-129.

6. M. Bohr, S.S. Ahmed, S.U. Ahmed, M. Bost, T. Ghani, J. Greason, R. Hainsey, C. Jan, P. Packan, S. Sivakumar, S. Thompson, J. Tsai, and S. Yang, "A high performance 0.25μm logic technology optimized for 1.8V operation," *International Electron Devices Meeting Technical Digest*, 1996, pp. 847-850.

7. M. Shoji, *CMOS Digital Circuit Technology*, Prentice-Hall, 1988.

8. N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*, Addison-Wesley, 1993.

9. T. Sakurai and R. Newton, "Alpha-power-law MOSFET model and its applications to CMOS inverter delay and other formulas," IEEE Journal of Solid-State Circuits, Vol. 25, 1990, pp. 584-593.

Samuel O. Nakagawa

Sam Nakagawa is a member of the technical staff of HP Laboratories. He received a BSEE degree from the University of Illinois Urbana-Champaign in 1985 and MSEE and PhDEE degrees from The Pennsylvania State University in 1987 and 1993. He joined HP Laboratories in 1994.

John G. McBride

John McBride is a member of the technical staff of the HP VLSI Technology Center, writing netlist quality tools for custom VLSI designs. Born in Vernal, Utah, he received his MSEE degree in 1991 from Stanford University. He is married, has four children, and enjoys white water rafting and tinkering with cars. He is also active in his church and the Boy Scouts of America.

Dennis M. Sylvester

Dennis Sylvester is a graduate student researcher at the University of California at Berkeley, working for his PhDEE degree. He works part-time in the ULSI laboratory of HP Laboratories. He is a member of the IEEE and a native of Midland, Michigan.

Soo-Young Oh

Soo-Young Oh is a project manager in the ULSI laboratory of HP Laboratories. He received his PhDEE degree from Stanford University and joined HP in 1980. He is a member of the IEEE and has published more than 50 papers. Born in Seoul, Korea, he is married, has two children, and enjoys swimming and playing guitar.

Theory and Design of CMOS HSTL I/O Pads

Gerald L. Esch, Jr.

Robert B. Manley

To control reflections, the impedance of integrated circuit output pad drivers must be matched to the impedance of the transmission lines to which the pads are connected. HP's HSTL (high-speed transceiver logic) controlled impedance I/O pads use an on-chip impedance matching network that compensates for process, voltage, and temperature (PVT) variations.

Transmission line reflections are one of the major factors limiting high-speed I/O performance. These reflections can be controlled by matching the driver output impedance to that of the transmission line. Traditional solutions require the use of off-chip components to implement matching termination networks. This adversely impacts board density, reliability, and cost. Integration of the termination network on-chip removes these negative attributes while providing additional advantages.

In this paper, we review a solution for an on-chip impedance matching network. Our HSTL (high-speed transceiver logic) family of controlled impedance I/O pads includes single-ended and differential drivers and receivers, along with compensation circuitry for process, voltage, and temperature (PVT) variations. Measured HSTL signal integrity in a large, complex board environment is presented.

Gerald L. Esch, Jr.
Jerry Esch is a hardware design engineer with the HP Integrated Circuit Business Division, specializing in high-speed graphics ASIC design, high-speed CMOS I/O pad design, and real-time MPEG encoding. Born in Kearney, Nebraska, he received his BSEE degree in 1992 from New Mexico State University and joined HP in 1994. He is a volunteer with the Special Olympics.

Robert B. Manley
Bob Manley is a VLSI design engineer with the HP Integrated Circuit Business Division. He holds BS degrees in physics (1975) and electrical engineering (1979). He joined HP in 1979. Born in South Bend, Indiana, he is married, has two children, and is a pilot and a competitive rifle and pistol shooter.

Parallel versus Series Termination

When I/O signal integrity and speed are of utmost importance, many pad designers turn to parallel termination networks. Parallel termination eliminates transmission line reflections. However, parallel termination exacts a costly toll on power dissipation because a dc component is added to power consumption. An alternative termination approach is source series termination. In a point-to-point environment, series termination provides an output driver with a means to absorb incident waves, effectively damping any reflections in the transmission line. Matching a driver's output impedance to that of the board impedance increases signal integrity and speed while keeping power dissipation to a minimum.

Source series termination is easily implemented with two components: a low-impedance output driver and a precision series resistor. To decrease factory costs and conserve board space, it is desirable to replace the printed circuit board precision series resistor with an on-chip, PVT-compensating resistor.

Output Driver

The single-ended output driver, shown in **Figure 1**, has two major components: a push-pull driver and an on-chip series termination resistor. The three components that form the termination resistor are the driver NFET resistance R_{DS}, an n-well resistor R_{ESD} for ESD protection, and R_{PROG}, a programmable resistor between the nodes PRE and POST. Controlled by on-chip calibration circuitry, the programmable resistor takes on a range of resistances to ensure that the driver's output impedance R_o matches the transmission line impedance Z_o. This allows reflections to be completely absorbed in the driver regardless of process, temperature, and voltage fluctuations. Thus:

$$R_o = R_{DS} + R_{PROG} + R_{ESD} = Z_o.$$

R_{PROG} is tuned by turning on and off various combinations of transfer NFETs with a six-bit binary word. Each bit in the binary word, PROG[5:0], controls a transfer gate in the programmable resistor array. The NFETs have conductances corresponding to their binary weighted bit positions in PROG[5:0]. For example, if PROG[0] controls a transfer gate with conductance of G, then PROG[1] controls a transfer gate with a conductance of 2G. The resistance of R_{PROG} decreases as the binary count PROG[5:0] increases. In effect, as the binary count increments, more resistors are added in parallel in the NFET array.

Calibration Circuitry

The calibration circuitry (**Figure 2**) is designed to program all HSTL output driver impedances, R_o, to match that of an external precision resistor, R_{EXT}. During normal operation, an enable signal, CAL, causes an NFET equivalent in size to an HSTL output driver pull-up NFET to conduct. Current begins to flow through the I/O pad through R_{EXT}. This current path forms a voltage divider, where:

$$V_{PAD} = V_{DDQ}\frac{R_{EXT}}{R_{EXT} + (R_{DS} + R_{PROG} + R_{ESD})},$$

or

$$V_{PAD} = V_{DDQ}\frac{R_{EXT}}{R_{EXT} + R_o}.$$

Figure 1

Single-ended output driver.

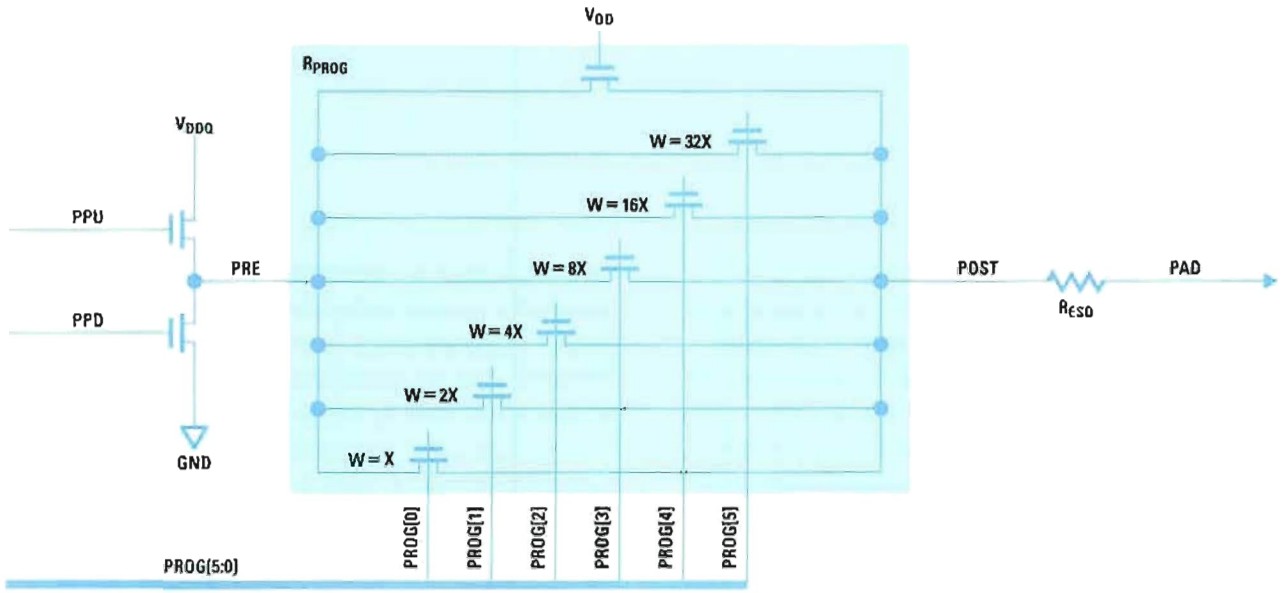
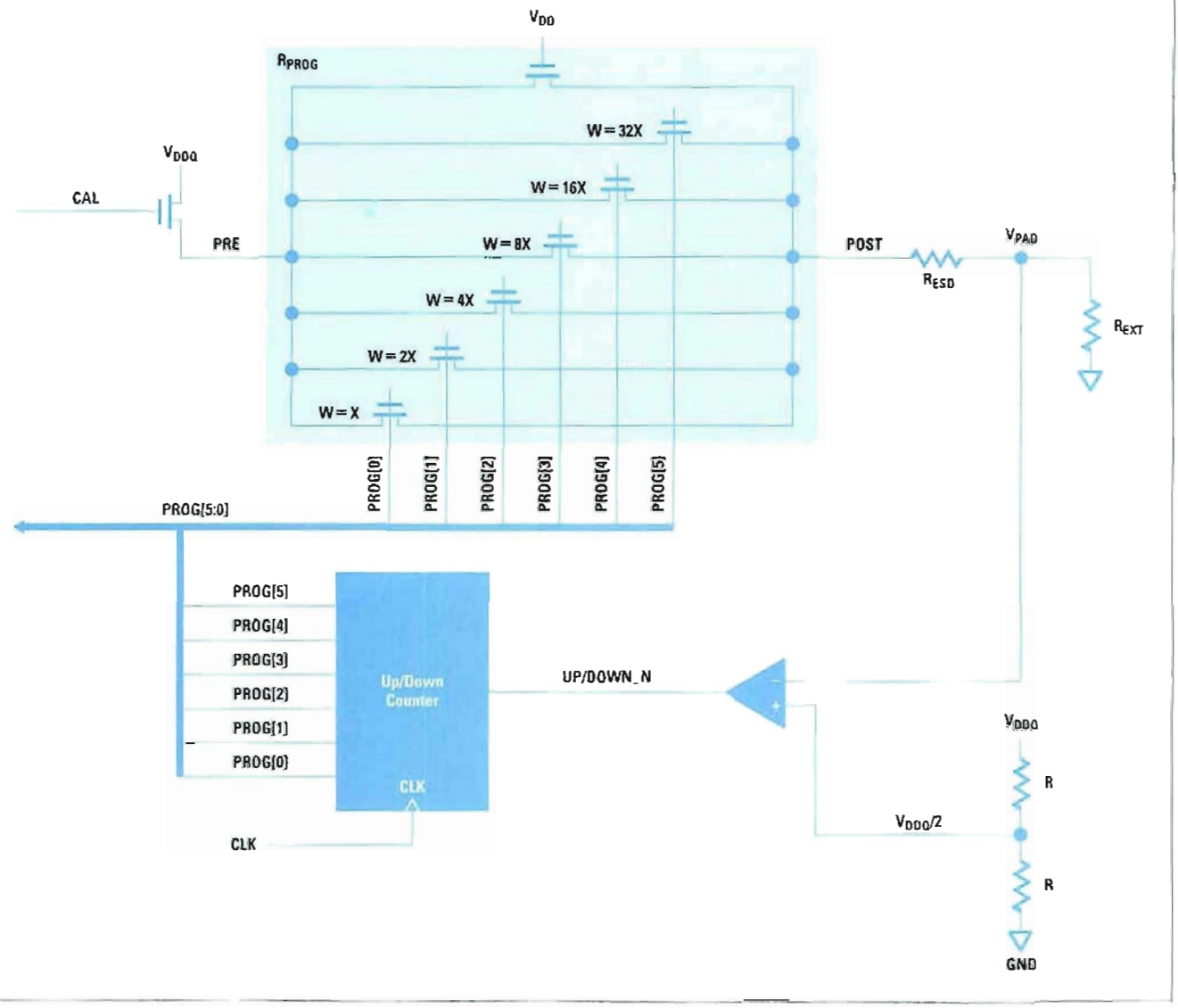


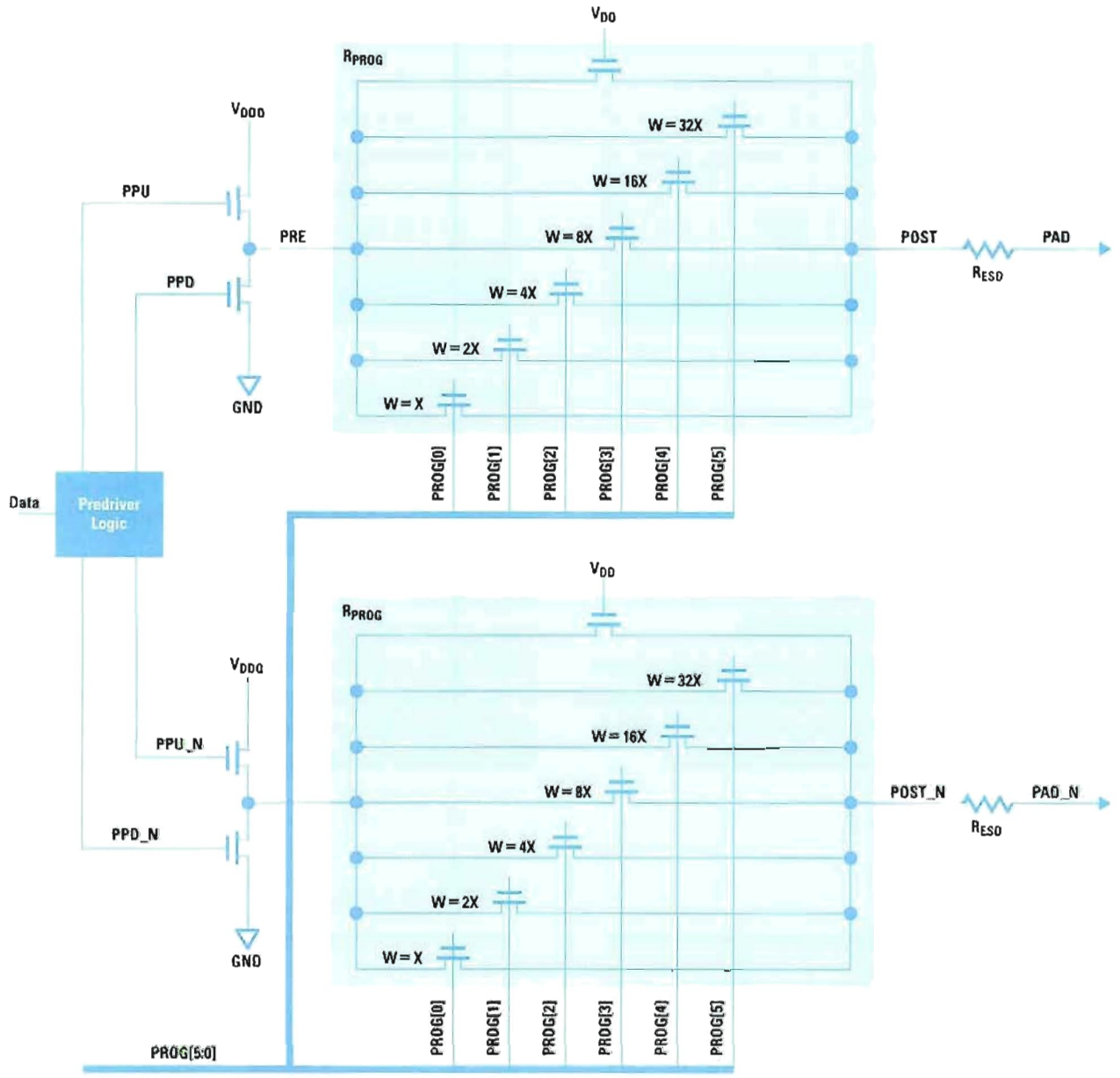
Figure 2

Calibration circuitry.



V_{PAD} serves as an input to the inverting terminal of the differential amplifier. The noninverting terminal's input voltage is $V_{DDQ}/2$. This reference voltage is generated on-chip via a voltage divider. Any difference between the input voltages of the differential amplifier is perceived as a resistance mismatch between R_O and R_{EXT}. The ΔV causes the differential amplifier's output to program an up/down counter to increment or decrement its six-bit output. Upon receiving a clock edge, the up/down counter drives a new six-bit binary count, PROG[5:0]. This calibration word is used by the calibration circuitry's programmable resistor and distributed to other HSTL driver programmable resistors. Incremental binary changes in PROG[5:0] cause incremental resistance changes in the programmable resistor. Because R_{PROG} is now programmed to a new value, V_{PAD} obtains a new analog value. V_{PAD} again acts as an input to the differential amplifier and the impedance matching process starts over. The calibration action is continuous and transparent to normal chip operation.

Figure 3
Differential driver.



Differential Driver

The differential driver in **Figure 3** is the combination of two HSTL drivers. By knowing the driver's output resistance, an external parallel termination network can set the dc operating points to comply with the HSTL differential specification. The predriver logic is responsible for keeping the differential clock signals in conformance with the ac specification.

The predriver logic also performs two other important tasks. The single-ended-to-differential conversion preserves the input duty cycle while minimizing transients in the supply currents.

Figure 4
Single-ended waveform.

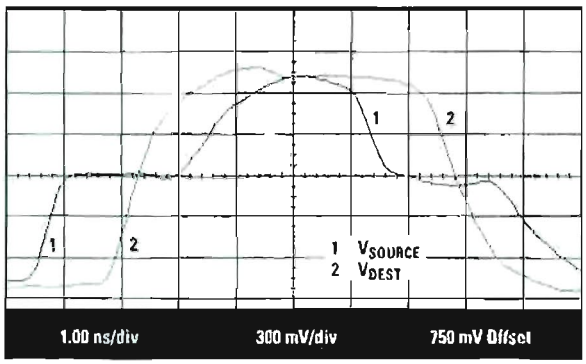
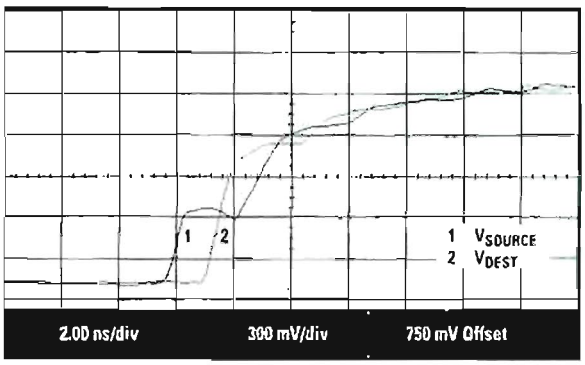


Figure 6
Overdamped signal.



Measured Results

Figure 4 depicts a single-ended signal traversing a 6-inch transmission line. V_{DEST} is the signal received at the end of the line. Signal integrity can be monitored by examining the location of the inflection point at the driver (V_{SOURCE}). An inflection point near half the high logic level ($V_{DDQ}/2$ for HSTL) indicates that the driver output impedance matches the transmission line impedance.

Figure 5 shows a differential signal after traveling down a 6-inch transmission line. V_{DEST} is consistently contained within the tight HSTL differential specifications because of the known differential driver output resistance.

Figures 6, 7, and 8 illustrate the effects of varying the driver's output impedance by changing the external calibration resistor, R_{EXT}. In **Figure 6** $R_{EXT} > Z_o$, in **Figure 7** $R_{EXT} = Z_o$, and in **Figure 8** $R_{EXT} < Z_o$. Signal integrity is maintained when the driver output resistance matches the transmission line impedance.

Future Work

One potential shortcoming of implementing a parallel NFET array to mimic a source series termination resistor is the variation in driver output resistance R_o. **Figures 9 and 10** show driver output resistance as a function of output voltage V_o. Ideally, R_o should be constant over the

Figure 5
Differential waveform.

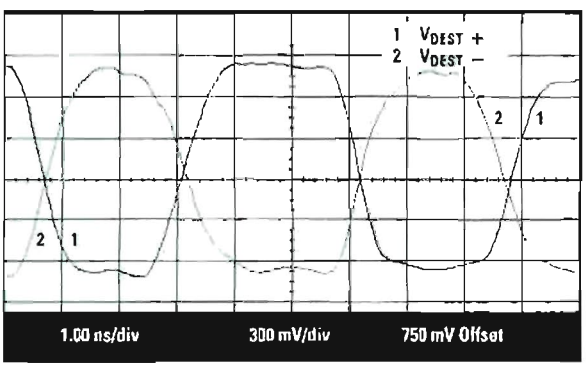


Figure 7
Critically damped signal.

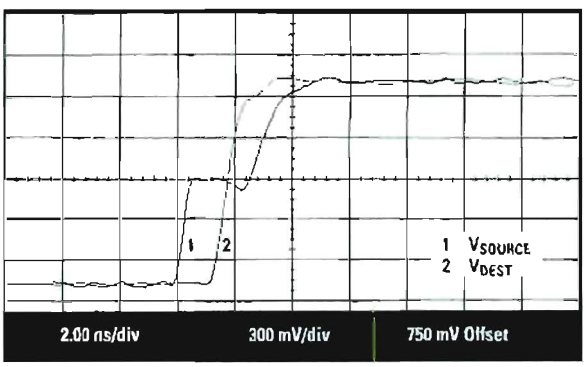
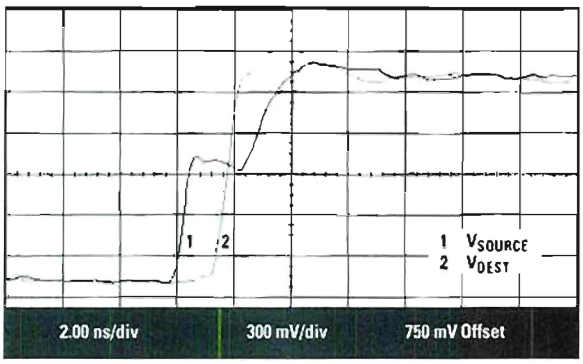


Figure 8

Underdamped signal.



range of V_o. However, R_o changes with variations in the values of V_{GS}, V_{DS}, and the back gate voltage of the transfer and driver NFETs. **Figure 9** shows the output resistance changing as the output voltage swings from 0.1V to 1.4V (10% to 90%). **Figure 10** illustrates the output resistance as a function of the output voltage swing from a logic high to a logic low. Notice in **Figure 9** that the point at which $R_o = Z_o$ (50Ω) occurs when $V_o = V_{DDQ}/2$. This is because the calibration circuitry tunes the programmable resistor with the pull-up portion of the output driver at $V_{DDQ}/2$. With proper driver width ratioing, the pull-down driver NFET would also have $R_o = 50\Omega$ at $V_o = V_{DDQ}/2$.

Figure 9

R_o as a function of V_o as V_o goes from a logic low to a logic high.

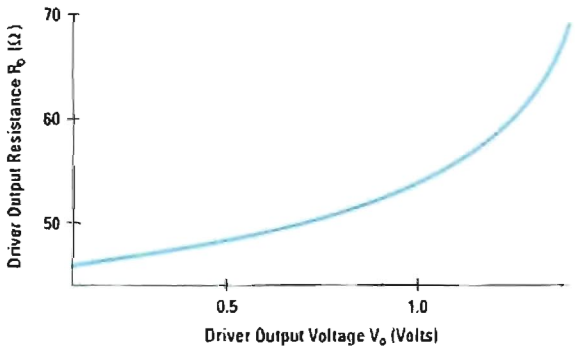
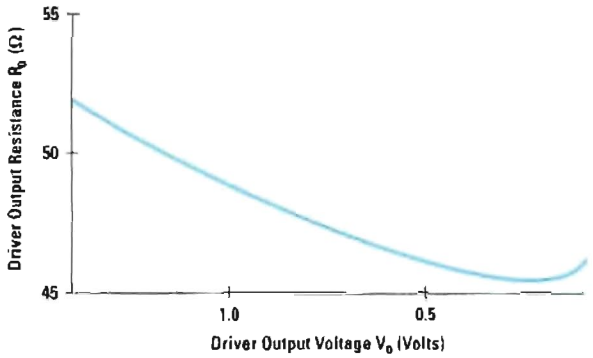


Figure 10

R_o as a function of V_o as V_o goes from a logic high to a logic low.



However, to help reduce variations in R_o, the crossover point was shifted towards the high logic level of V_o. For HSTL I/O applications, this inherent deviation in output resistance minimally affects signal integrity. For future applications, it may be worthwhile to investigate alternative series termination schemes for tighter impedance matching environments.

Conclusion

A parallel NFET array can be a simple and effective way of controlling a driver's output resistance. With an on-chip source series termination resistor, a chip can communicate at higher frequencies and board space that would normally contain termination networks is freed. Our family of HSTL pads has been proven to work at 200 MHz in large, complex board environments.

Acknowledgments

The authors wish to thank Gordon Motley for his valuable pad design and ESD advice, Rick Luebs for offering the original design concept, Rob Martin for his technical support, Ron Larson and Wally Wahlen for their characterization effort, Holly Manley for making the paper readable and Amy Jahnke for supplying direction. We also thank Gary Wetlaufer for his support.

Bibliography

1. G. Esch, Jr. and R. Manley, "HSTL CMOS I/O Pads for 200-MHz Data Rates," *Proceedings of the 1996 Hewlett-Packard Design Technology Conference*, pp. 51-58.

2. R. Canright, Jr., "The Impact of Driver Impedance upon Transmission Line Impedance," *Proceedings of the 44th Electronic Components and Technology Conference*, pp. 669-674.

3. *HSTL (High-Speed Transceiver Logic)*, JEDEC Standard 16.3, Draft Revision 3, JEDEC Council, July 14, 1994.

4. R. Wilson, "JEDEC ready to approve very-high-speed I/O spec," *Electronic Engineering Times*, December 5, 1994, p. 1.

A Low-Cost RF Multichip Module Packaging Family

Lewis R. Dove
Martin L. Guth
Dean B. Nicholson

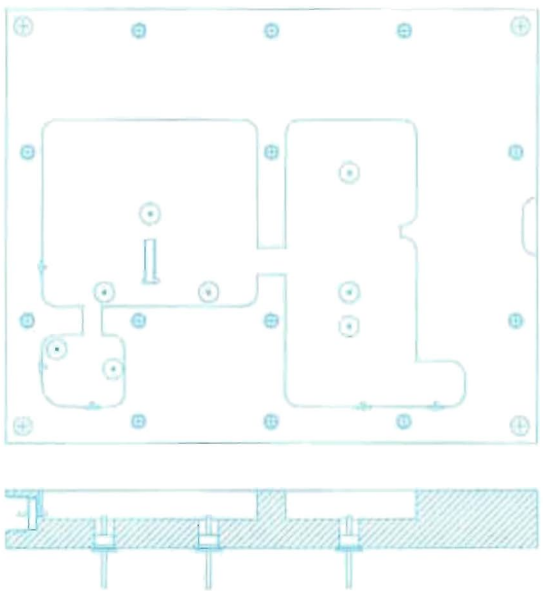
These packages provide much lower cost than traditional high-frequency packaging, shielding, and interconnects, while still providing low-reflection transitions and high electrical isolation.

Packaging of RF and microwave microcircuits has traditionally been very expensive. The packaging requirements are extremely demanding: very high electrical isolation and excellent signal integrity to frequencies of 10 GHz and above, as well as the ability to accommodate GaAs ICs dissipating significant amounts of heat. The traditional approach has been to start with a machined metal package and solder in dc feedthroughs and RF glass-to-metal seals (**Figure 1**). Thin-film circuits on ceramic or sapphire are then attached to the floor of the package using electrically conductive epoxy. The channels machined into the package form waveguides beyond cutoff, which provide isolation from one circuit section to another. Next, GaAs or Si ICs are attached to the floor of the package in gaps between the thin-film circuits. After wire bonding is done to interconnect the ICs and the thinfilm circuits to each other and to the package, the RF connectors are screwed on over the RF glass-to-metal seals to form the finished assembly (**Figure 2**).

Traditionally, both the placement of the thin-film circuits in the microcircuit package and the bonding to connect them have been done manually. The microcircuit typically plugs into a small bias printed circuit board that is connected to the rest of the instrument through a dc cable, while RF connections to the instrument are made by semirigid coaxial microwave cables, which screw onto the RF connectors on the microcircuit. This solution is capable of providing excellent electrical performance and quick turnaround times from design to implementation. However, traditional microcircuits do have the substantial drawbacks of being very expensive, bulky, and heavy.

Figure 1

Traditional microcircuit package with dc and RF glass-to-metal seals. (a) Top view. (b) Cross section.



Until the end of the cold war, much of Hewlett-Packard's test and measurement equipment was bought for defense-related applications, for which performance was often of paramount importance. Today, the demand for RF and microwave test and measurement instrumentation is driven primarily by consumer applications, such as the satellite, wireless, or fiber-optic communications markets. In these extremely competitive markets, keeping costs down and getting good value are crucial. To lower the cost of RF and microwave instruments, the cost of their microcircuits must be reduced, since this is often a significant portion of the manufacturing cost of the instrument. It is important when reducing costs, however, that the predictable performance and quick turnaround time of the traditional microcircuit be preserved.

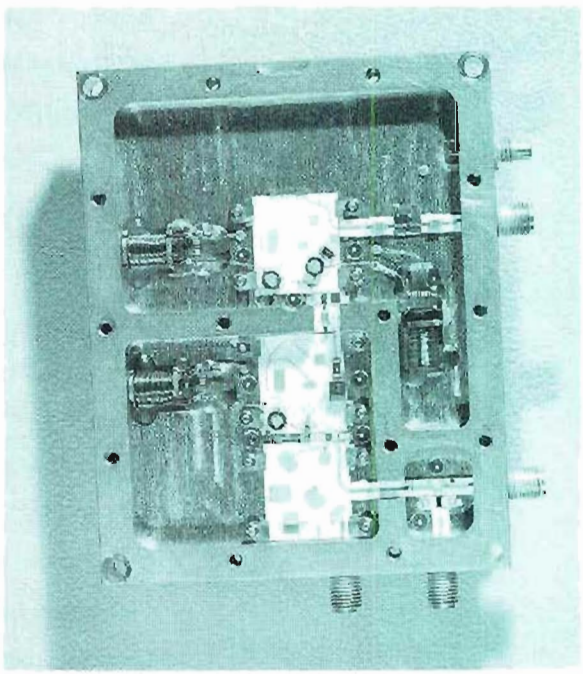
MIPPS Modules

We have developed a family of high-performance microcircuit packages to be used for RF and microwave applications. Within HP, the new family is called MIPPS, which stands for *multichip integral-substrate PGA (pin grid array) package solution.*

The MIPPS modules meet the requirements discussed above by providing high RF and microwave performance at a lower cost per function than previously available through a new design that replaces many expensive, traditional microwave packaging components with simpler, less-expensive alternatives. The MIPPS modules have been designed to work well up to 10 GHz, which addresses most of the present consumer-driven applications where cost is critical. To be inexpensive, the MIPPS modules were designed from the outset for manufacturability and to use high-yielding assembly processes. High-frequency electrical modeling using HP's High-Frequency Structure Simulator (HFSS) 3D modeling tool was used to synthesize the RF transition into the MIPPS module, resulting in an electrical design that worked the first time it was prototyped. Leveraging the MIPPS design and manufacturing processes for new MIPPS microcircuits continues to shorten the time needed to introduce each succeeding MIPPS design.

Figure 2

Finished traditional microcircuit assembly.

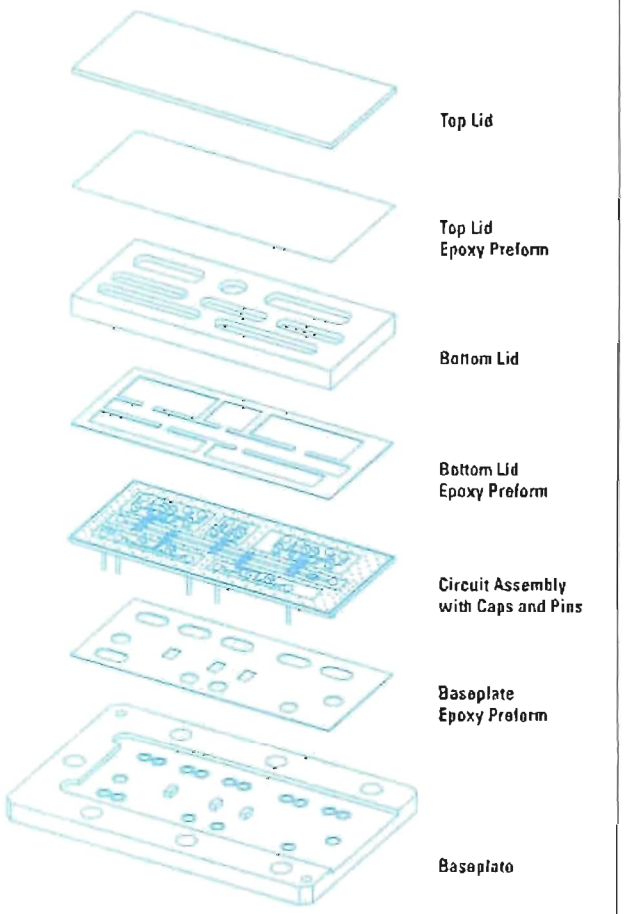


Reducing Cost

Figure 3 shows an exploded view of the basic MIPPS concept. The expensive portions of the traditional microcircuit have been changed to produce a lower-cost assembly. The first example of cost reduction is the way the module is electrically connected to the rest of the instrument. All connections from the instrument to the MIPPS module are made by way of 0.020-inch-diameter pin grid array (PGA) pins that cost pennies each, and are connected to the thick-film metallization on top of the ceramic by a hemisphere of solder. This PGA connection is inexpensive compared to the filtered dc feedthroughs and RF glass-to-metal seals used in a traditional microcircuit.

Figure 3

Exploded view of the MIPPS (multichip integral-substrate PGA package solution) assembly.



For dc connections, the unwanted RF energy on the bias lines in the MIPPS module can be stripped off with low-cost surface mount components to an adequate degree in place of the coaxial capacitor used for the traditional microcircuit feed. The PGA pins used as dc feeds can either be attached to the printed circuit board by inexpensive single-contact sockets, or by soldering the pins directly to the backside of the printed circuit board. Experience has shown that plugging the MIPPS module into single-contact sockets is desirable because it allows much faster assembly or replacement of the MIPPS assembly.

RF connections are made using the same 0.020-inch-diameter PGA pins used for the dc feeds. Good RF performance is obtained by choosing the diameter of the hole in the baseplate that the PGA pin goes through. The proper hole diameter compensates for the capacitive discontinuity of the metal pad on the top surface of the alumina to which the PGA pin is soldered. The RF transitions can be made either directly to the printed circuit board if a low-RF-loss printed circuit board is being used, or to an SMA connector (either barrel or flange mount) if the RF signal is to be sent or received through a coaxial cable.

The next cost savings come from replacing the multiple small thin-film circuits used in traditional microcircuits with a single, large, thick-film-on-ceramic circuit. A thick-film circuit that has both conductors and resistors printed on it generally costs only 10% to 25% as much as a similar thin-film circuit on ceramic, and additional savings are accrued in assembly. Assembly of the substrate to the metal baseplate needs to provide mechanical rigidity as well as electrical shielding. Two things must be done to make this attachment reliable (no attachment failures over temperature excursions). The first is to minimize the TCE (temperature coefficient of expansion) mismatch between the ceramic and the baseplate. Type 416 stainless steel was chosen as the best compromise of raw material cost, ease of machinability to minimize cost, and TCE matching with the ceramic, which is 96% purity alumina. The next step to minimize the stress on the conductive epoxy joint is to choose a material that forms a somewhat flexible bond between the substrate and the baseplate. This helps absorb TCE mismatch induced stress. MIPPS modules with 0.800-by-1.700-inch circuits have been temperature cycled ten times over the −55°C to +150°C temperature range with no failures.

Using a single large ceramic substrate instead of multiple smaller substrates as in traditional microcircuits also provides substantial savings during assembly. Instead of having to manually place and line up multiple circuits in a package, a single substrate is assembled to a baseplate that has the epoxy preform already tacked onto it. The alignment between the thick-film circuit and the baseplate is accomplished by using tooling, instead of "eyeballing it" as in traditional microcircuits. The problem of conductive epoxy wicking up between circuits and shorting out the conductive traces to ground, a perennial issue with traditional microcircuits, is completely avoided in the MIPPS module design.

Another technique for reducing costs in the MIPPS module was to design it to take advantage of automated assembly techniques. With a planar ceramic circuit that has all the circuitry accurately aligned to the other circuitry to within the tolerance of the thick-film printing process, automated die attach and automated wire bonding are very straightforward. This is contrasted with a traditional microcircuit containing small thin-film circuits placed down inside deep channels, requiring careful manual wire bonding. Autobonding is significantly faster and more repeatable than the manual bonding process typically used in RF microcircuits.

Maintaining Traditional Microcircuit Performance

Traditional microcircuits have good RF transitions into and out of the package, low loss and low reflections along transmission lines in the package and when transitioning between transmission lines and ICs, and high isolation between circuit functions in the package. All of these elements need to be preserved in any proposed alternative packaging scheme. The first performance challenge to be tackled was the RF transition into and out of the MIPPS module. Our target for MIPPS modules was to have good RF performance up to 10 GHz. The starting point was choosing the diameter of the PGA pin as 0.020 inch. The 0.020-inch-diameter pin is standard for connecting to an SMA connector. Thus, this choice allows us to make RF transitions to the MIPPS module either from printed circuit board or semirigid coaxial cable. The thickness of the baseplate was set by structural rigidity concerns at 0.090 inch. The diameter of the thick-film solder pad used for pin attach was chosen as a best compromise between mechanical strength and parasitic capacitance.

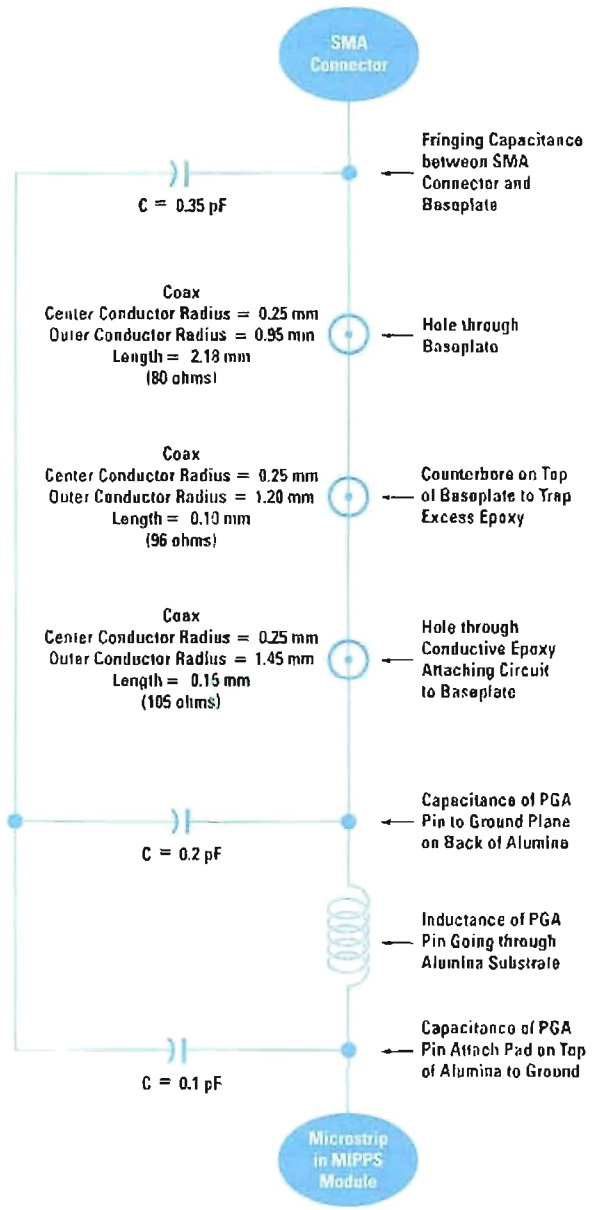
HP's High-Frequency Structure Simulator (HFSS) was used to determine the hole diameter through the baseplate and backside groundplane relief on the alumina circuit that would give the best RF transition to 10 GHz. These simulations eventually led to a design that correctly electrically compensates the launch and gives better than 40-dB return loss up to 3 GHz and better than 15-dB return loss to 10 GHz for a transition from an SMA connector to a 50-ohm line in the MIPPS module. HFSS was also used to design the RF transition to the MIPPS module from an inner-layer printed circuit board stripline. The HP Microwave Design System (MDS) was then used to create a physically based model of the transition, shown in **Figure 4**. New users of the MIPPS module family can use this MDS model in circuit simulations to determine how well the package will work for their application.

After the RF signal enters into the MIPPS module, it needs to be able to propagate down the transmission line with low loss and minimal reflections. The 0.025-inch-thick 96% alumina substrate used for the MIPPS modules allows the use of gold conductor traces 0.024-inch wide for 50-ohm signal lines. This is a wide enough line to have low resistive losses for RF signals (< 0.6 dB/inch at 10 GHz), yet the substrate is thin enough that the impedance of the transmission line stays fairly constant from dc to 10 GHz. The next-higher-order microstrip mode that could cause unwanted impedance discontinuities is well above 10 GHz. Finally, to have low loss and low reflections as an RF signal propagates down a transmission line, the edges of the line must be smooth. The HP thick-film process used provides the necessary edge smoothness for good performance to 10 GHz.

Many measurements need much greater than 100 dB of dynamic range, so high isolation is an extremely important criterion for microwave instrumentation packaging. In the traditional microcircuit package, high isolation between circuit functions is maintained by keeping the circuitry in narrow channels, which act as waveguides beyond cutoff for the RF signals, being too narrow to pass radiated emissions. The same function is performed in a MIPPS module by having hundreds of conductor-filled vias that connect the bottom ground plane of the circuit with the metallization on the top. When the lid with its machined channels and cavities is conductively epoxied to the topside metallization, the lid is very effectively

Figure 4

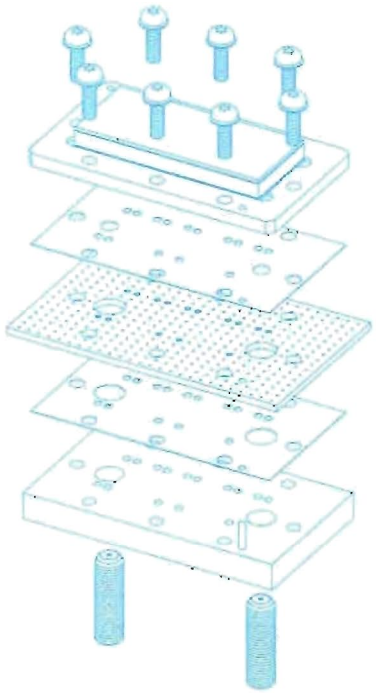
Electrical model for an SMA-connector-to-MIPPS microstrip transition.



grounded, and the waveguide beyond cutoff channels in the lid then provide the needed isolation. To isolate the dc feedthroughs and RF transitions to the printed circuit board from stray RF signals, a metal waffle gasket is placed between the MIPPS module and the printed circuit board as shown in **Figure 5**. A similar waffle gasket is

Figure 5

MIPPS connection to a printed circuit board and SMA connectors.



placed between the printed circuit board and a backing plate to shield the backside of the printed circuit board.

Test Strategy

Early in the project a decision was made to have all of the PGA pins on 0.100-inch centers for compatibility with low-frequency PGA packages. This allows MIPPS modules to be easily tested at frequencies up to 20 MHz using industry-standard zero-insertion-force sockets. The first MIPPS product is a 130-dB electronic step attenuator used in RF signal generators at up to 4 GHz. Once the RF performance was initially characterized, a low-speed ac test was implemented, taking approximately 30 seconds to complete. A full RF test is done after the part is installed on the instrument printed circuit board, with better than 97% yield resulting. By providing a quick, low-speed screening at the module level, duplication of a lengthy instrument-level test on the module itself is avoided. In addition, the MIPPS assembly is tested before sealing the lid, allowing easy replacement of defective ICs. For more complex module designs, if a high yield at final RF test cannot be achieved

with a simple low-frequency module-level test, more extensive RF testing can be implemented before lid seal to get the desired yield at final RF test.

MIPPS Projects

The first MIPPS module designed was a 0-to-130-dB electronic step attenuator operating from dc to 4 GHz, with attenuation adjustable in 5-dB steps. This module is pictured in **Figure 6**, which shows a completely assembled package (top) and a view of the attenuator thick film attached to the baseplate. The module is used in the HP E4400A family of electronic signal generators. It is based on GaAs IC step attenuator chips manufactured by HP. Previous signal generators have used mechanically switched step attenuators, which have exceptionally good electrical performance (very low loss and low levels of reflected energy), but take approximately 50 milliseconds to switch and settle from one attenuation state to the next. Although they are guaranteed to work for greater than five million cycles, they will eventually wear out. With the more complex and higher-speed modulation formats used in new wireless communications protocols, as well as the higher volumes of these products, mechanically switched step attenuators were wearing out in HP's test instrumentation at an unacceptable rate. Electronically switched step attenuators switch much faster and are more reliable, but the challenge was to produce one at a cost similar to a mechanically switched step attenuator. While step attenuator chips in surface mount packages mounted on printed circuit boards would be able to meet the cost goals for the electronically switched step attenuator, the electrical performance at 4 GHz in terms of loss and reflections would be inadequate. For this reason, the bare GaAs chips are epoxied either to the thick-film substrate or onto pedestals machined into the baseplate. With GaAs switches that must have very high isolation in the off state, the inductance between the backside ground of the chip and true ground must be essentially zero. In this case, the grounding vias that go through the 0.025-inch-thick substrate to connect the pad under the IC to ground are too inductive. Therefore, these chips are mounted to a baseplate pedestal that protrudes through a hole in the ceramic substrate. This allows the chip backsides to be connected to an excellent ground.

Figure 6

The first MIPPS module, a 0-to-130-dB electronic step attenuator operating from dc to 4 GHz. (top) A completely assembled package. (bottom) A view of the attenuator thick film attached to the baseplate.

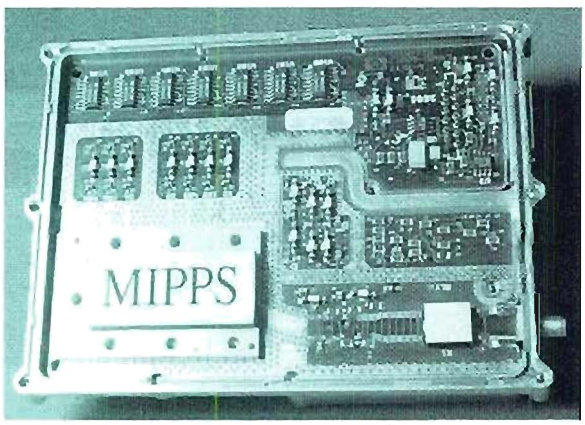
An additional constraint on the electronic step attenuator design was that some of the customers for the HP E4400A signal sources containing this step attenuator were expected to expose it to extremely humid conditions, necessitating that the modules' internal components be completely shielded from moisture. This was accomplished by developing a process to add low-dielectric-constant, low-RF-loss silicone gel to the interior of the MIPPS package, thereby protecting the GaAs ICs and reducing the risk of moisture-induced metal migration from the silver-filled electrically conductive epoxies within the assembly. With the current design, the MIPPS module will function without damage even if liquid water is introduced inside the cavity.

The final MIPPS electronic step attenuator has a somewhat higher insertion loss than a mechanically switched step attenuator. However, the MIPPS electronic step attenuator switches much more quickly and should be able to switch reliably indefinitely, since there are no moving parts to wear out. The finished assembly with the MIPPS module, the switching logic, the RF reverse power protection circuitry, and the additional EMI filtering on the control lines is shown in **Figure 7**.

An interesting addition to the standard MIPPS step attenuator module occurred when the HP instrument division

Figure 7

Finished assembly with MIPPS module, switching logic, RF reverse power protection circuitry, and additional EMI filtering.

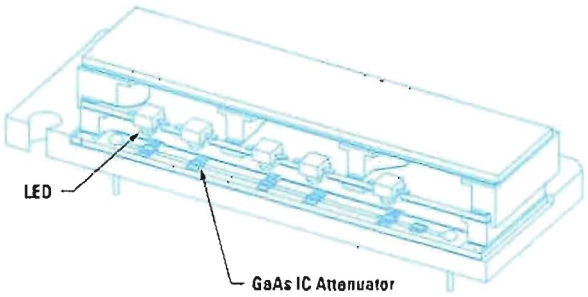


using the MIPPS module noticed that the output RF level through the module was not completely switched and settled in the 28 microseconds necessary for testing GSM (a European cellular phone standard) equipment. This problem is caused by the GaAs switches in the step attenuator having the *slow tail effect*: the initial switching takes place very quickly, but it can take up to 10 milliseconds to reach the final value. It was discovered that high-intensity light can virtually eliminate the slow tail effect in the GaAs switches, so a printed circuit board containing HP's new miniature high-intensity LEDs (one above each GaAs step attenuator) was incorporated on top of the MIPPS module as shown in **Figure 8**. The low-cost method by which these are integrated into the module also made it possible to do this for a relatively small incremental cost, and shows the flexibility of designing in the MIPPS format.

Another MIPPS module under development is used in the HP 8509X RF electronic calibration module. Previously, most network analyzer calibration has been done using mechanical calibration standards that were screwed onto the end of the network analyzer cables in a procedure that took five minutes or longer. Although electronic calibration modules have previously been offered by HP, they covered a limited frequency range and were relatively expensive. In designing the next-generation RF electronic calibration

Figure 8

Cross section of a MIPPS module with light-emitting diodes.



products, it was important to keep the costs down and provide an operating frequency range of 300 kHz to over 6 GHz. Both of these goals were achieved with the RF electronic calibration MIPPS module, incorporating eight GaAs IC switches to switch open circuits, short circuits, through lines, and 50-ohm loads in and out. It is possible to completely calibrate a network analyzer in about 30 seconds, using only a single set of RF connections.

Acknowledgments

The MIPPS module was a joint development effort between HP's Microwave Technology Center (MWTC) and HP's Colorado Springs Technology Center (CSTC). We'd like to thank the manager of MWTC, Jerry Gladstone, for giving us the time and guidance to start and continue this project. We would also like to thank Roger Graeber from the Microwave Instruments Division (the first customer of the MIPPS technology) for his many helpful suggestions on how to achieve high isolation in small areas, Don Estreich, who managed the project for the last two years at MWTC, for making sure that we had the necessary resources for the job, and Heidi Barnes of MWTC for continuing the MIPPS work on new projects and continuing to refine the concept. From CSTC, we'd like to thank Ron Baker and Vicki Foster for their invaluable help in assembling prototype hybrids and Rosa Clayton, Rosemary Johnson, Brenda Archuleta, and Al Baca for the fine work they have done in manufacturing the thick-film substrates. We would also like to thank Bob Kressin, Jeff Herrle, Chris Eddleman, and Scott Casmer for their continued product engineering support.

Lewis R. Dove

Lew Dove is an engineer-scientist at HP's Colorado Springs Technology Center, where he works on thick-film packages for high-frequency multichip modules. He has BSEE and MSEE degrees from the University of Arizona. He is married and has two sons, and his hobbies include flying, running, and skiing.

Marlin L. Guth

Marlin Guth is a product design and development engineer at HP's Colorado Springs Technology Center. He worked on the RF attenuator hybrid for the HP E4400A signal generator and the pin driver logic hybrid for the HP E330 Series of test systems. He has a BSEE degree from Florida Atlantic University, and he joined HP in 1979. He was born in Cleveland, Ohio, and his interests include photography, skiing, mountain climbing, electronic music, and scuba diving.

Dean B. Nicholson

Dean Nicholson has spent the last seven years designing multichip module products using GaAs ICs. He joined the HP Microwave Technology Division in 1980 after graduating from the University of California at Berkeley with a BS degree in electrical engineering and materials science (double major). Born in Orange, California, Dean is married with four children. If he had some free time, he would enjoy soccer, abalone diving, and hiking.

Testing with the HP 9490 Mixed-Signal LSI Tester

Matthew M. Borg

Kalwant Singh

The tester's features include a timing interval analyzer for statistical analysis of clock periods, synchronous generation of arbitrary waveforms with respect to master digital clocks, and a library of digital signal processing routines. These features have been applied to production measurements of key parameters like AGC loop bandwidth, phase-locked loop timing jitter, and ADC signal-to-noise ratio and distortion parameters.

In recent years, there has been significant theoretical work on defining a methodology for fault detection and classification in analog circuits.[1-4] However, because input-output relationships are more complex for analog circuits than for digital circuits, the development of a systematic, automated approach for detecting defects in analog circuits is far behind the digital counterpart. For this reason, implementations of analog test strategies remain largely functional.[3] This is also the case in the work described here.

The purpose of this paper is not to further the state of mixed-signal test theory and methodologies, but rather to share with the reader the state of mixed-signal testing within the HP Integrated Circuit Business Division (ICBD) today. We present descriptions of the test development processes for a partial response maximum likelihood (PRML) read channel ASIC and a charge-coupled device (CCD) signal processor ASIC, including specific examples of analog test implementation demonstrating some of the capabilities of the HP 9490 tester.

The read channel IC was designed for HP's DDS3 format DAT (digital audio tape) drive. The CCD signal processor is a three-channel interface chip designed for HP's scanner products. These chips contain significant analog functionality, including programmable and automatic gain control (AGC) amplifiers, switched capacitor filters, a clock recovery phase-locked loop, moderate- and high-resolution analog-to-digital converters (ADC's), and several

Matthew M. Borg
Matt Borg is an analog IC designer with the HP Integrated Circuit Business Division. He holds BSEE (1987) and MSEE (1990) degrees from the University of California at Davis. He joined HP in 1989.

Kalwant Singh
Kalwant Singh received his PhD degree in electrical engineering from Rensselaer Polytechnic Institute in 1993. A hardware design engineer with the HP Integrated Circuit Business Division, he joined HP Singapore in 1985. He is a native of Singapore.

digital-to-analog converters (DACs). The test strategy implemented for these blocks was largely functional. Many circuits could not afford the additional complexity and parasitics of embedded test access. However, provisions were made in the designs for accessing inputs and outputs of functional blocks, and to allow special test modes of operation. The HP 9490 tester has sufficient analog resources to efficiently execute detailed functional testing with high resolution for detecting subtle variations in performance resulting from process variations and defects (see "Tester Description" on page 66).

The emphasis of this paper is primarily on analog test, with specific examples given for the read channel AGC and phase-locked loop blocks and the CDD signal processor analog signal path.

Test Program Evolution

The test programs for the read channel and CCD signal processor chips both evolved in three distinct phases:

- Turn-on
- Performance verification and debug
- Consolidation and production worthiness.

The turn-on stage, typically lasting a few weeks before and after first silicon, involved putting together a very basic screen test consisting of continuity test, reference voltage verification, digital functional vectors, and tests for signs of life from the analog blocks. Simplicity of the initial analog tests was necessary to get screened parts into the customer's hands quickly. We discovered that development of complex analog tests required an intimate knowledge of the tester and the overall function of the chip, the main challenges being getting the chip into the desired state, constructing the correct analog stimulus, and synchronizing the analog and digital inputs.

The performance verification and debug stage of test development lasted from after the initial prototype shipments until artwork release for the final chip revision. During this stage, digital and analog static current tests were debugged, pad leakage tests were added, and any remaining digital functional tests were added, but the majority of time was spent adding complexity and refinements to the original analog tests and creating new analog tests to verify that all analog functions met the required specifications. Often this activity was interrupted by the need to create specific tests to debug unexpected behavior discovered either by the customer or the test development process. In the case of the read channel ASIC, the customer provided a test harness that could be used to power up the chip, write to registers, and view outputs while stimulating the analog inputs. This proved to be very useful for debug activities. However, there were several cases in which the HP 9490 tester's ability to control the timing of analog inputs and capture outputs on a cycle-by-cycle basis was invaluable in isolating design bugs or marginalities.

During the final test development phase, the many analog functional and debug tests were consolidated into fewer, more efficient tests. For both ICs, we retained the capability of putting the test programs in a debug mode in which additional data is saved in diagnostics files and captured waveforms are saved for viewing.

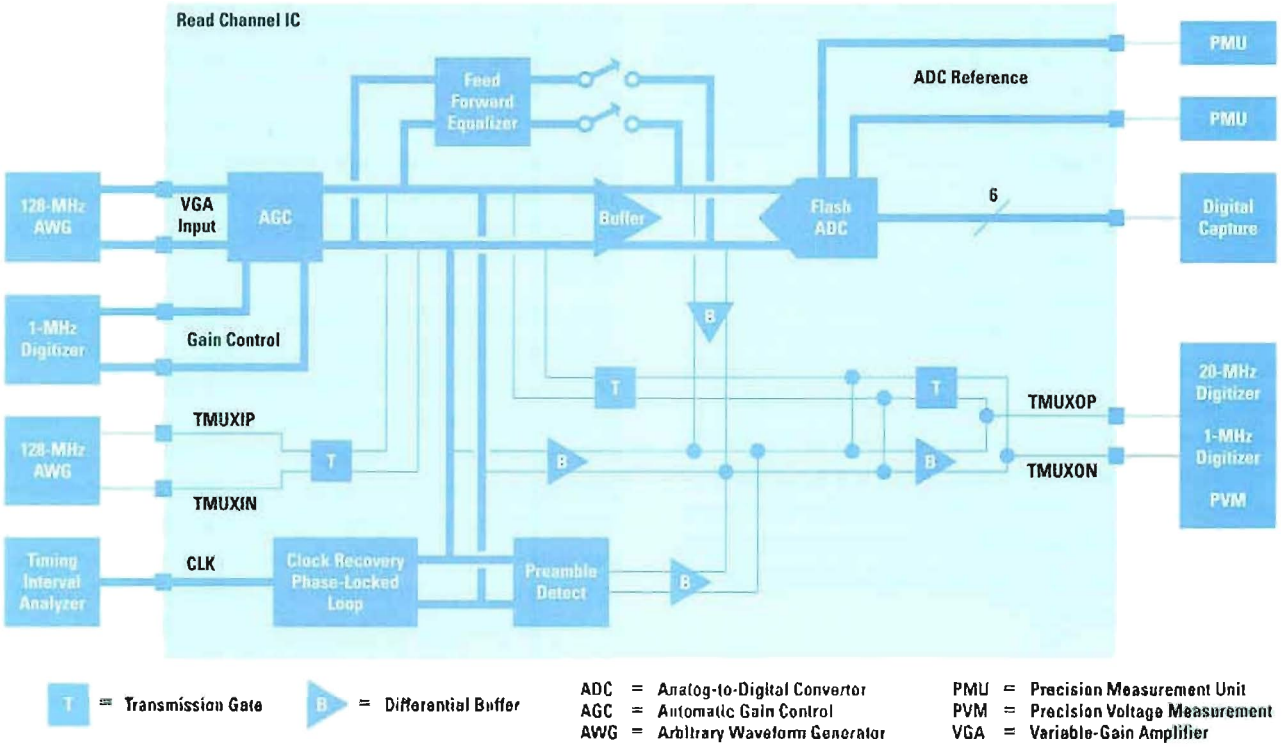
Analog Test Strategy

The strategy used for guaranteeing that the analog blocks were defect-free was first to measure the analog supply current in both the static (power-down) and power-on states, and second, to generate functional tests that both isolate subblocks and mimic customer use to verify all specifications listed in the ERS. The general premise was that a manufacturing defect would cause an individual subblock such as an op amp or comparator or a larger functional block to produce an unexpected output or compromised performance. An unexpected output might be an incorrect comparison, an incorrect dc level, excessive offset, instability, or an unavailable mode of operation. Compromised performance might be measured in terms of gain, linearity, dynamic range, resolution, settling time, bandwidth, acquisition range, detection threshold, or some other appropriate measure. Very often, individual subblocks cannot be specifically isolated, but their performance can be inferred from higher-level tests that exercise the subblocks in a variety of ways.

Any attempt to assess test coverage must first consider how a possible defect could manifest itself at one of the observation ports. The observable effect of a given type of defect will vary depending upon several factors, including the function of the block and the location of the defect. Generally speaking, in a fully differential circuit such as the read channel AGC, a defect that occurs in the differential signal path is likely to cause some type of offset, whereas a defect occurring in common-mode circuitry, such as bias

Figure 1

Read channel analog signal path with simplified test access circuitry and HP 9490 analog resource utilization.



circuits, or single-ended signal paths is likely to cause faults such as improper dc or common-mode voltages, excessive current, or reduced range of operation.

Specification limits are set based upon both the customer's performance requirements and the observed distribution of the test parameters during characterization testing. Test specification windows must be set wide enough to cover expected process variations, provided that there is adequate performance margin, but narrow enough to weed out defects that cause "soft" faults. Of course drawing the line between process variation and soft faults is a tricky business, which can be mitigated to some extent by multiple tests with overlapping coverage of potential defects.

Testing the Read Channel IC

For a complex analog system such as a PRML read channel, it was necessary to design in test access points that allowed the inputs and outputs of functional blocks to be stimulated and measured as directly as possible. Such access proved invaluable in debugging and verifying that the individual blocks met their respective design goals. It was also useful for achieving a level of test coverage adequate to meet quality goals. The read channel design includes an analog test multiplexer with programmable ac and dc test paths for bypassing analog blocks and observing internal nodes (**Figure 1**). The ac test paths are isolated from the main signal path by transmission gates (T) and wideband differential buffers (B). Two pads (TMUXOP and TMUXON) dedicated to observing analog voltages have a pair of wideband ac buffers capable of driving the tester load, including the 10-kΩ input impedance of the HP 9490 20-MHz digitizer. Auxiliary differential input pads (TMUXIP and TMUXIN) allow the AGC block to be bypassed and the ADC, preamble detector, and clock recovery block inputs to be driven directly. Individual control of all transmission gates in the test path allows calibration of the dc offset of the buffer paths. The read channel IC also includes a digital test multiplexer that directs digital outputs generated by analog blocks to test pads.

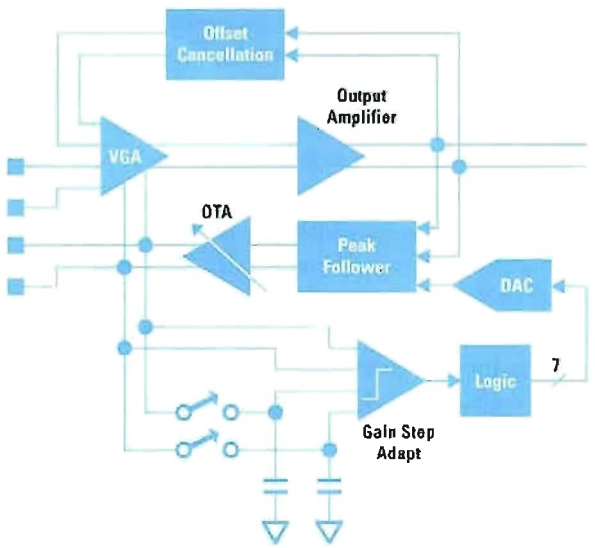
Additional constraints placed on the design of the analog blocks were that (1) they needed to be individually powered down to a state in which they drew no current from the analog supply, (2) analog outputs were tristated (in a high-impedance state), and (3) digital outputs were driven to the rails. Also, the current drawn by any block from the analog supply was required to be proportional to a reference current set by an on-chip bandgap reference and an external precision resistor.

Testing the Read Channel AGC

The AGC block, pictured in **Figure 2**, is one of the most complex analog systems tested on the ICBD HP 9490 testers. It includes a fully differential four-stage variable-gain amplifier (VGA) with a gain range of 12 dB to 32 dB, a fixed-gain linear output amplifier, continuous-time global offset cancellation, a peak follower for amplitude detection, a programmable operational transconductance amplifier (OTA) for multiple loop bandwidth selection, a 7-bit DAC for output amplitude setting, and circuitry that adaptively eliminates the gain step that occurs at the DAT preamble-data boundary.

Figure 2

Read channel AGC (automatic gain control) block diagram.



DAC = Digital-to-Analog Converter
OTA = Programmable Operational Transconductance Amplifier
VGA = Variable-Gain Amplifier

The specifications used to design this block were used as a guide for developing the test routines for debug and performance verification and finally for production screening tests. The three test input ports for the AGC block are the VGA input, the gain control input, and the digital input to the reference DAC. The test output ports include the amplifier chain output, the peak follower output, the DAC output, the loop filter output, and the output of the gain step adaption comparator. Needless to say, this leaves many internal circuit nodes that can only be observed through their interaction with the outputs of major subblocks.

The AGC block test is separated into several subtests which individually target the amplifier chain including offset cancellation, the peak detector, the AGC loop, the gain step adaption circuit, and the DAC. In many cases, a subtest will exercise a large portion of the AGC system to produce the stimulus needed to extract the performance measure of a particular subblock. This results in an overlap in coverage, which increases overall test coverage. The amplifier chain is tested by allowing the AGC loop to lock separately to three different ~2-MHz sine waves with input amplitudes covering the extremes and center of the VGA input dynamic range. The amplifier output is digitized for the three different inputs by an HP 9490 20-MHz digitizer through the analog test multiplexer output pads. A fast Fourier transform (FFT) is performed on the three sets of digitized data to extract the output amplitude, offset, total harmonic distortion (THD), and signal-to-noise ratio (SNR), which are all compared against pass/fail limits. Additional measurements are made of the amplifier output common-mode level and of the test buffer path offset for correction of the amplifier chain offset measurements. This series of tests provides wide coverage of potential defects in the amplifier chain and the rest of the AGC loop.

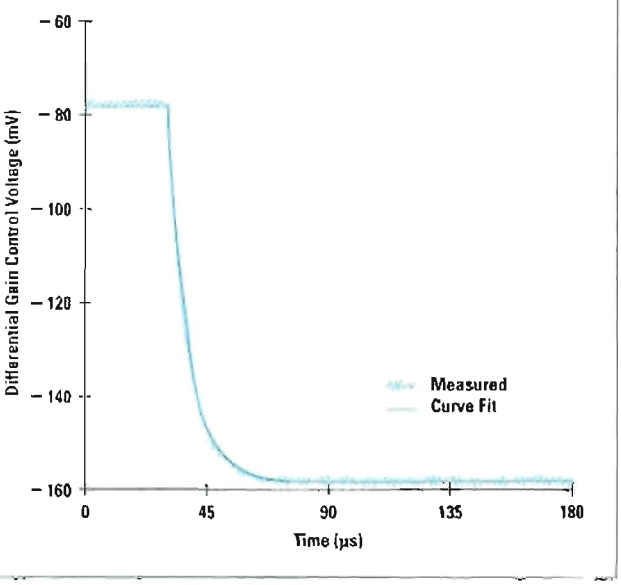
The decay rate of the peak follower output is an important parameter because it determines how the AGC loop will respond to the varied and sometimes sparse peaks of digital audio tape (DAT) data. For this reason, a specific test was written to extract the decay rate from a digitized peak follower output with the AGC loop locked to a 1-MHz sine wave. Another key parameter for the DAT read channel is the accuracy of the AGC loop bandwidth

settings, especially in the low-bandwidth mode, which is the primary mode of operation during a read cycle. The loop bandwidth is most easily measured by monitoring the VGA control voltage when a step in input amplitude is applied to the VGA input. A low-frequency (1-MHz), high-input-impedance (1-MΩ) digitizer was used for this task. Since the AGC loop bandwidth is directly proportional to the gain control sensitivity of the VGA, the measurement was performed three times with minimum, nominal, and maximum VGA input amplitudes. The input signal chosen was a 9-MHz sine wave (same frequency as the DDS3 format preamble) with a 2-dB amplitude step after the loop was initially settled. **Figure 3** shows the differentially digitized gain control voltage during an input step, as displayed by the HP 9490 waveform editor. Also plotted is the exponential curve fit extracted from the digitized data by a simple C routine in the test program. The curve fit is performed to approximate the loop time constant and hence the loop bandwidth.

As previously mentioned, the AGC loop bandwidth is programmable by selection of different values of transconductance of the loop transconductance amplifier. The faster-bandwidth modes are selected by on-chip state machines at the beginning of each track read when the AGC must quickly acquire lock from the absence of signal to a preamble segment at the beginning of each track. A test was written that simulates this condition by driving the VGA input with a low-amplitude noise signal followed by a maximum-amplitude preamble signal with an exponential turn-on envelope. This is the most taxing situation for the AGC loop because it must go from a condition in which the VGA gain control is railed to being settled in the minimum-gain condition in a short period of time (before the end of the preamble). The preamble detector block and on-chip state machines are enabled during this test so that the loop bandwidth switching occurs as it does in normal use. The gain control voltage is digitized and then processed by C code to verify that the gain has settled to within the acceptable limit before the end of the minimum length preamble.

Figure 3

Measured AGC gain control voltage during a 2-dB step in input amplitude.



Additional tests are implemented that fully exercise the reference DAC, the gain step adaption comparator, and the analog signal path from the VGA through the on-chip ADC.

Testing the Read Channel Phase-Locked Loop

The read channel IC includes a mixed analog and digital phase-locked loop block, which recovers the clock signal from the DDS-format data stream. The recovered clock period is quantized in increments of one-sixteenth the external system clock. The phase-locked loop maintains the phase relationship between the recovered clock and the data stream by interspersing short (15/16) or long (17/16) clock periods with nominal clock periods. Taking advantage of the capabilities of the HP 9490 timing interval analyzer, a test was written that measures the period of each recovered clock cycle with the phase-locked loop locked to a sine wave having a period 2.5% shorter than the nominal data period. **Figure 4** shows the resulting histogram of recovered clock periods (as displayed in the HP 9490 waveform editor). The overall mean clock period is calculated to verify that the recovered clock frequency is 2.5% higher than the system clock frequency. The standard deviation of the short clock periods is tested against pass/fail limits as a measure of the uniformity of the delay elements of the 16-tap analog delay line within the clock recovery block. Additional tests were written that (1) individually verify the thresholds of the 32 comparators in the phase-locked loop phase sampler, (2) verify that the phase-locked loop can acquire lock to sine waves with the

Tester Description

At the HP Integrated Circuit Business Division, HP 9490 mixed-signal testers are nominally equipped with the following resources:

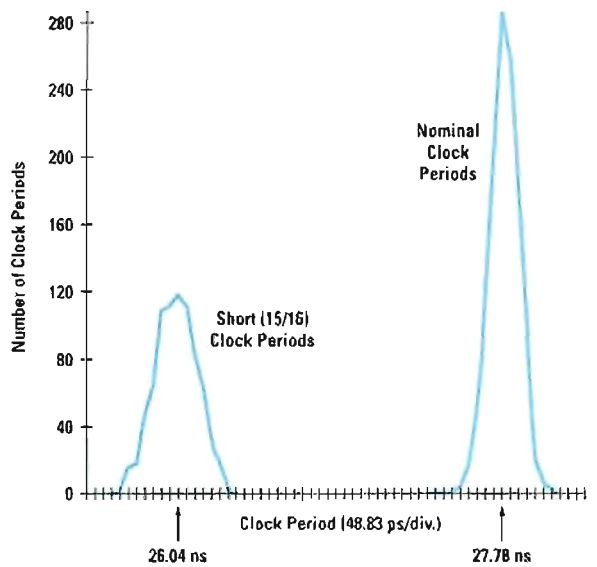
- Two 128-MHz dual-channel 12-bit arbitrary waveform generators (AWG)
- Two 1-MHz dual-channel 18-bit AWGs
- Two 20-MHz dual-input 12-bit digitizers
- Two 1-MHz dual-input 16-bit digitizers
- Two 1-GHz-bandwidth, 1-MHz-rate samplers
- One multiplexable dual-channel time measurement unit (TMU)
- One multiplexable dual-channel timing interval analyzer
- One precision voltage measurement unit
- One precision voltage source
- Two fixed and one multiplexed universal dc precision measurement units (PMU)
- Four dual-output DUT power supplies (DPS).

The test head includes 128 pins with per-pin dc function control. The maximum digital test frequency is 64 MHz (128 MHz with pin multiplexing) with edge and format changing on the fly. Vector depth is 4M bytes per pin. The digital test subsystem includes debugging tools such as shmoo plots, sequence debugger with fail mapping, and digital waveform display. Especially useful for testing of ADCs is the digital data capture capability with 500K-byte depth and special hardware for high-speed digital signal processing.

maximum expected frequency offset, and (3) measure the accuracy of the phase-locked loop phase offset setting by examining the phase at which the on-chip ADC samples a sine wave to which the phase-locked loop is locked.

Figure 4

Histogram of 2000 recovered clock periods.



Read Channel Test Summary

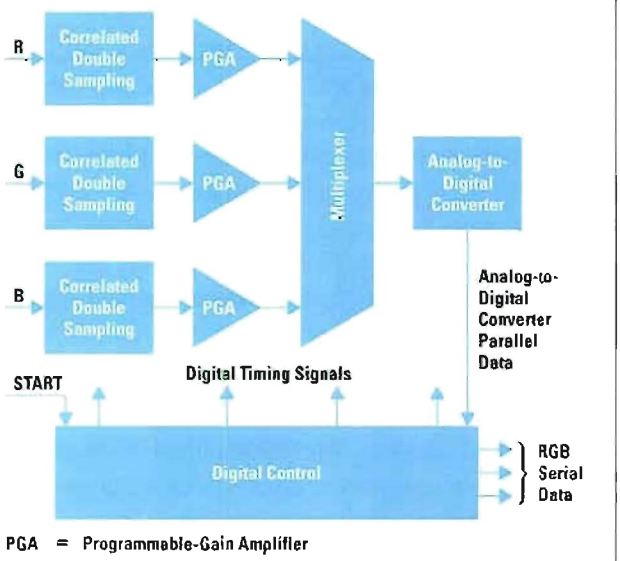
The final read channel production test is composed of 22 analog subtests, a digital scan subtest, 18 functional digital subtests, and three additional subtests for continuity, pad leakage, and static current. The overall test execution time is just under 7.5 seconds, with approximately 5.5 seconds for the analog tests, 0.5 second for the digital tests, and 1.5 seconds for the remainder.

Testing the CCD Signal Processor IC

The CCD signal processor is a CMOS-based monolithic IC that interfaces color (RGB) signal outputs from a CCD to a main ASIC.[5] The major components of this IC are three switched-capacitor 8-bit programmable gain amplifiers (PGAs), a 10-bit successive approximation ADC, and a 6-bit utility DAC (**Figure 5**, DAC not shown). The three PGAs perform 8-bit programmable offset compensation and 8-bit programmable amplification on the correlated double-sampled CCD signal. The ADC digitizes each of the PGA outputs in sequence, and the 10-bit converted codes for the RGB signals are serially output on three I/O pins. The correlated double sampling and amplification stage can be pipelined with ADC conversion and serial data output to maximize throughput.

Figure 5

CCD (charge-coupled device) processor signal path block diagram.



Signal Path

A typical CCD signal sampling sequence with this IC is as follows. The R (or G or B) signal is first held by the CCD at its black level, corresponding to the voltage output from the CCD pixel under no illumination. Then the CCD outputs transition to the video level, corresponding to the integrated illumination on that pixel. This complete single-pixel output cycle is initiated by the START pulse from the main controller IC.

To perform correlated double sampling, the black level from each pixel is sampled and subtracted from the sampled video level by the IC. The positions of the black sample point and the video sample point are 8-bit programmable in terms of the number of cycles from the end of the START pulse. The difference between the black and video levels for each pixel is amplified by the PGAs and then sampled and converted by the ADC.

Testing the CCD Signal Path

Because of test time limitations, we employed simplified versions of ramp and sinusoidal tests to test the IC signal path. We will describe the sinusoidal test here. It is assumed that the reader is familiar with concepts of ADC quantization noise.

Theory of Sinusoidal Test. The interested reader is referred to several excellent publications on ADC testing.[6-10] Briefly, the quantization noise of an ADC with a truly random input can be shown to have a mean square variance (or noise power) of $\Delta^2/12$, where Δ is the least-significant bit. By selecting a sine wave frequency that is not harmonically related to the sampling frequency, we can achieve quasirandom sampling over several cycles of the sine wave, and the ADC quantization noise power will approximate $\Delta^2/12$. We also assume that the sine wave exercises the full ADC range, that is, its peak amplitude is $2^N\Delta/2$, where N is the number of bits. Then:

$$\text{Signal Power} = 2^{2N-3} \times \Delta^2 \tag{1}$$

$$\text{Noise Power} = \frac{\Delta^2}{12} \tag{2}$$

$$\text{SNR (dB)} = 6.02N + 1.76. \tag{3}$$

The signal-to-noise ratio (SNR) provides a quantitative measure of the performance of the ADC. For example, an ideal quantization noise limited 10-bit ADC should yield an SNR of 61.96 dB. The practical IC signal path, however, will have its SNR limited by impairments such as random noise (fundamental and circuit-induced), distortion (from device nonlinearities), component mismatches, sampling time jitter, and so on. A measure of the actual performance of the IC signal path can be derived by calculating the effective number of bits (ENOB) from the measured signal-to-noise + distortion ratio (SNDR) as follows:

$$\text{ENOB} = \frac{\text{SNDR (dB)} - 1.76}{6.02}. \tag{4}$$

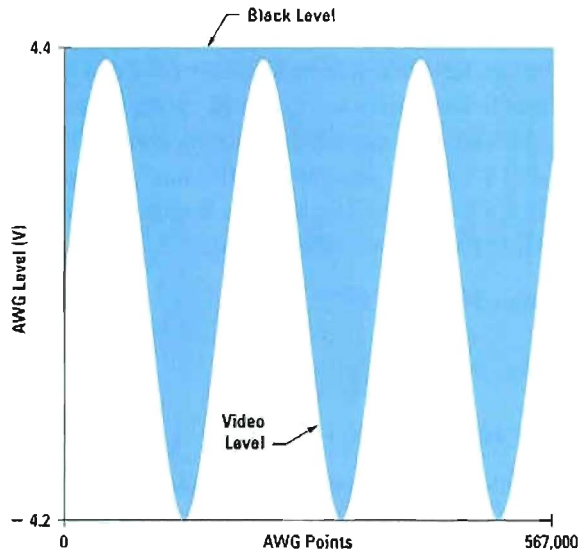
Alternatively, the ENOB can also be calculated by comparing the measured noise + distortion power, ND_{meas}, to its ideal value, $N_{ideal} = \Delta^2/12$:

$$\text{ENOB} = N - \log_2\left(\frac{ND_{meas}}{N_{ideal}}\right). \tag{5}$$

This method is chosen for calculating ENOB because it does not require the signal path to be driven over its full range, eliminating the possibility that the PGA and ADC might be overdriven and clipped, yielding an inaccurate SNDR.

Figure 6

Sinusoidal test AWG (arbitrary waveform generator) waveform.



Test Signal Generation. It is assumed that the black and video sampling positions are chosen such that all transients associated with the black and video level transitions are settled. As such, the signal path is insensitive to the frequency of the input signal (up to the Nyquist rate). A 128-MHz 12-bit arbitrary waveform generator (AWG) was used to generate a sine wave created from data points generated by a custom program. The AWG rate and the digital clock were set to 20 MHz, and a sine wave of about 1 kHz was used. The number of points digitally captured was limited to 1024 to reduce test execution time. This number must be a power of 2 to use the HP 9490's built-in FFT routines effectively.

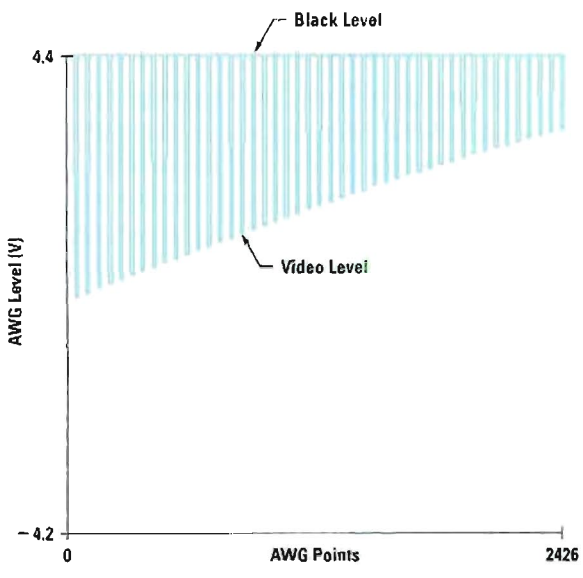
The IC's signal path range is 0 to 2.5V (ac coupled), while the AWG range is −4.4V to +4.4V. To maximize the resolution of the AWG waveform, the sine wave was generated over the full AWG output range, with the AWG internal attenuator set to −10.93 dB. This results in a maximum-resolution AWG signal that is within the input range. However, to avoid inadvertent clipping, which may result from PGA or ADC gain errors or offsets, we limited the sine wave amplitude to 95% of the full AWG range. In the initial stages of test development, we monitored the SNDR of the AWG waveform and found that its ENOB was nearly 11 bits. The noise input power of the AWG was subtracted from the measured signal path noise power to yield the effective measured noise power ND_{meas}, which was then used to calculate the ENOB. We assumed that the AWG noise was independent of the signal path noise.

Various low-pass filters (up to 132 MHz) are available to improve the SNR of the AWG output waveform, but they were not used in this test because of the sharp transitions required between black and video levels. **Figure 6** illustrates the 1-kHz waveform created for the AWG, while **Figure 7** shows the details of the waveform near its start. The difference between the black level (4.4V) and the sine wave envelope constitutes the input sinusoidal signal. The maximum rate of change of this waveform is about 8.6V/50 ns = 172V/μs, which is well within the slew rate of the AWG (600V/μs).[11]

Test and Data Analysis. The START pulse, which initiates the CCD signal processor conversion cycle, must be synchronized to the AWG waveform. The AWG must be started at the same time as the first START pulse, and subsequent START pulses must be synchronized with the beginning of each black level in the AWG waveform. The HP 9490 tester allows synchronization between the AWG, the digital pattern generator, and other mixed-signal

Figure 7

Zoomed-in sinusoidal test AWG waveform.



resources to within 1 ns when one master clock is used. This resolution limit increases to the master clock period (7.8 ns) when two master clocks are required to implement the test. This synchronization feature is fully exploited in the test. The serial data output of the IC is read into the HP 9490 digital capture memory and retrieved into the test workstation memory for analysis. A typical retrieved waveform and its Fourier spectrum are shown in **Figure 8**.

Fourier analysis of the retrieved waveform is performed by built-in digital signal processing (DSP) algorithms. The fundamental amplitude, phase, dc offset, SNDR, total harmonic distortion (THD), second-harmonic distortion (2HD), and third-harmonic distortion (3HD) are extracted by customized routines that call upon built-in discrete Fourier transform (DFT) routines using FFT methods. It is not necessary to capture an integral number of cycles, since the custom routine has a built-in Hanning windowing function.[8]

The total noise + distortion power is calculated from the fundamental amplitude and SNDR value, and is corrected for the noise power of the AWG. The ENOB is calculated using equation 5. The ENOB for the case shown in **Figure 8** is approximately 8.4 bits, corresponding to a PGA and ADC SNDR of about 52.3 dB. The THD measured is −53.1 dB, while the 2HD is −56.6 dB (**Figure 8**). This particular IC's PGA and ADC performance is thus distortion-limited. Although THD, 2HD, and 3HD are not tested parameters, they provided invaluable insight into the source of SNDR limitations during the debug phase. The integral nonlinearity (INL) error profile can be obtained by subtracting the retrieved waveform from the fundamental, and the maximum and root mean square INL error can be derived. A typical INL error profile is shown in **Figure 9**. We did not test all of the 1024 ADC codes. Differential nonlinearity (DNL) errors can be calculated from similar INL profiles obtained from high-resolution ramp tests.

The parameters tested in the sinusoidal production test are fundamental amplitude, ENOB, and maximum INL error. The dc offset, THD, 2HD, and 3HD are used for debugging purposes only, and are saved into diagnostic files during nonproduction debug modes together with fundamental amplitude, ENOB, maximum INL error, and the various waveforms illustrated here (e.g., **Figures 8 and 9**).

The HP 9490 mixed-signal tester allows optimization (minimization) of test time by pipelining digital pattern execution times with data analysis. The benefit of doing this in our case was minimal, since pattern execution

Figure 8

A waveform retrieved from the ADC and its spectrum.

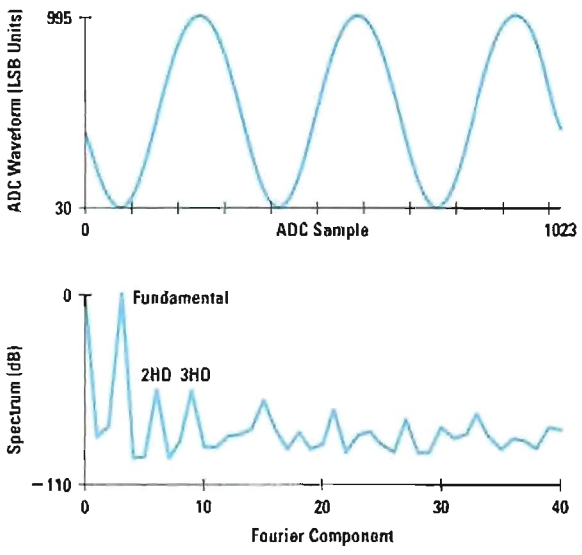
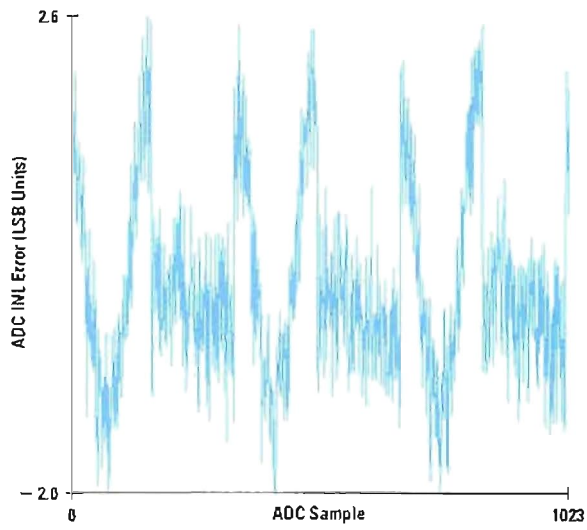


Figure 9

Sinusoidal test INL (integral nonlinearity) error.



times were extremely fast compared to data analysis times. The total production test time of this IC on the HP 9490 tester was about two seconds.

Conclusion

Mixed-signal test is a developing field within ICBD with many interesting challenges and room for advances in theory, methodologies, and standardization. The HP 9490 tester has been proven to be a very capable platform for testing complex analog circuitry as demonstrated by both the PRML read channel and CCD signal processor projects. Provisions made for controllability and observability of analog signals during the design process can yield highly testable designs. However, the development of mixed-signal tests continues to be a custom process requiring detailed knowledge of both the tester and the circuit under test.

Acknowledgments

The authors would like to acknowledge Tom Schmerge and Hal Cook for their help in test implementation and debug, the read channel team in Corvallis, Fort Collins, and Bristol, particularly Charles Moore, Rob Morling, and Chris Williams, for their help with test strategy, and the Fort Collins manufacturing engineers, technicians, and operators for providing HP 9490 support.

References

1. R. Harjani and B. Vinakota, "Analog Circuit Observer Blocks," *IEEE Transactions on Circuits and Systems*, Vol. 44, no. 3, March 1997, pp. 154-163.

2. M. Slamani and B. Kaminska, "Fault Observability Analysis of Analog Circuits in Frequency Domain," *IEEE Transactions on Circuits and Systems*, Vol. 43, no. 2, February 1996, pp. 134-139.

3. M. Sachdev, "A Realistic Defect Oriented Testability Methodology for Analog Circuits," *Journal of Electronic Testing*, Vol. 6, no. 3, June 1995, pp. 265-276.

4. J. Machado Da Silva, et al, "Mixed Current/Voltage Observation Towards Effective Testing of Analog and Mixed Signal Circuits," *Journal of Electronic Testing*, Vol. 9, no. 1/2, August/October 1996, pp. 75-88.

5. *CCD Signal Processor External Reference Specification* (HP Internal Document).

6. J. Doernberg et al, "Full Speed Testing of A/D Converters," *IEEE Journal of Solid-State Circuits*, Vol. SC-19, no. 6, December 1984, pp. 820-827.

7. *Dynamic Performance Testing of A to D Converters*, HP Product Note 5180A-2.

8. B. Peetz, et al, "Measuring Waveform Recorder Performance," *Hewlett-Packard Journal*, Vol. 33, no. 11, Nov. 1982, pp. 21-29.

9. M. Gordon, "Analog to Digital Converters," *Analog and Mixed Signal Conference Digest*, 1995, pp. 25.1-25.7.

10. B. Razavi, *Data Conversion System Design*, IEEE Press, 1995.

11. *HP 9490 TPL Manual*.

Online Information

More information about HP semiconductor test products can be found at:

http://www.hp.com/go/semiconductor

Reliability Enhancement of Surface Mount Light-Emitting Diodes for Automotive Applications

Koay Ban-Kuan
Leong Ak-Wing
Tan Boon-Chun
Yoong Tze-Kwan

Preencapsulation drying eliminates broken stitch bonds and reduces inconsistent reliability performance. A new casting epoxy formulation stops epoxy cracking, and optimization of the die-attach epoxy cure schedule solves lifted die-attach and delamination problems.

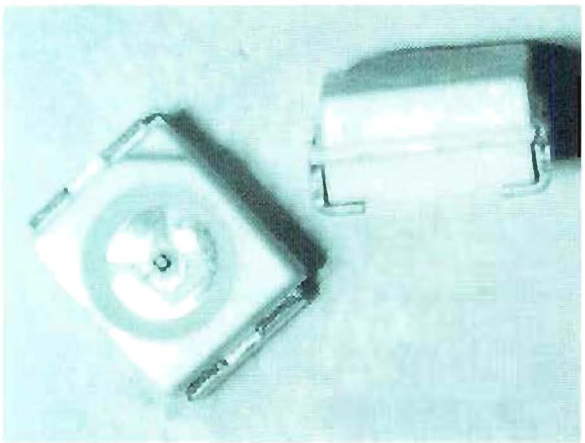
The current direction of the automotive lighting industry is to increase the use of printed circuit board surface area and to improve reliability to exceed that of the conventional light bulb. The two major categories of light-emitting diodes (LEDs) used in the automotive industry are exterior and interior. For interior use, HP's surface mount HSMx-Tnnn LEDs (36 products, e.g., HSMA-T425) had application potential, but their reliability needed to be enhanced to better suit the increasingly stringent automotive applications. In particular, the HP products had to conform to the European Cenelec Electronic Components Committee (CECC) standard.[1]

HP HSMx-Tnnn LEDs

Introduced in 1990, the surface mount HP HSMx-Tnnn LEDs (see **Figure 1**) occasionally experienced broken stitch bonds and epoxy-leadframe delamination when soldered. In September 1995, a second-generation product was released. This product resolved the broken stitch bond problems and improved the failure rate in temperature tests from an average of 120 ppm to 0 ppm after five temperature cycles. Extended temperature cycling between −40°C and 85°C for 20 cycles showed a significant improvement, from an average of 4500 ppm to 0 ppm.

On the downside, the product became somewhat more sensitive to moisture absorption and was not able to meet automotive market requirements for thermal performance (the CECC standard).

Figure 1
HP HSMx-Tnnn surface mount LEDs.



Infrared (IR) soldering per the CECC specifications (the upper temperature constraint of 260°C may be exceeded for a maximum of 10 seconds) resulted in failures after 168 hours of preconditioning at 85°C and 85% relative humidity (RH). The failures included broken stitch bonds, lifted die-attach, epoxy cracking, and epoxy-leadframe delamination. Therefore, an aggressive program was planned for the third generation of this surface mount LED to bring its quality to world-class automotive standards.

Third-Generation Surface Mount LED

The effects of moisture absorption by IC packages leading to electrical failures have been well-documented.[2] Delamination and package cracks during IR solder reflow are the predominant failure modes. Prebaking of packages to drive the moisture away before soldering and the use of moisture barrier bags are common practices but are not well-accepted by customers.

For the third-generation HSMx-Tnnn products, significant effort was put into understanding the failure mechanisms and the linkages to manufacturing processes and into raw material optimization. Implementation of preencapsulation drying solved the broken stitch bond problems and reduced inconsistent reliability performance.

A new casting epoxy formulation and curing conditions further enhanced the quality and reliability, improving the moisture sensitivity from level 3 to level 1 of the applicable JEDEC standard (test method A112).[3] To simulate the JEDEC level 1 standard, all the experimental units were subjected to 85°C/85% RH for 168 hours of preconditioning followed by two iterations of CECC IR soldering and extended temperature cycling between −55°C and 100°C or thermal shock from −40°C to 110°C. The cracked epoxy problem was solved by the new epoxy formulation, and the lifted die-attach and delamination problems were solved by optimizing the die-attach epoxy cure schedule. These enhancements significantly improved the robustness of the device and it was qualified by a major automotive supplier.

High-Temperature Epoxy Encapsulation Cure

The epoxy cure schedule was set at 125°C for eight hours based on the original product release qualification tests. However, after further discussion with the epoxy vendor, it was realized that a higher epoxy cure temperature could be used to improve the performance.

A 2^2 full factorial experiment was designed with one factor being the cure temperature at two levels—125°C and 150°C—and the other factor being the cure time at two levels—2 hours and 8 hours. The response was the failure rate after IR soldering followed by repeated thermal shocks between −40°C and +110°C, with 30 minutes dwell time and zero transfer time. The full set of data is shown in **Table I** for 500 units per cell.

Table I
Data from Epoxy Cure Temperature Experiment

Cell Number	Conditions	Cumulative Failure Rate (%) after 200 Cycles
1	125°C, 2 hr	5.49
2	125°C, 8 hr	3.17
3	150°C, 2 hr	0.22
4	150°C, 8 hr	0.65

The raw data showed that the 150°C cures gave very low failure rates compared to the 125°C cures. No factor was statistically different. Repeating this experiment gave similar results without highlighting any significant factor. It was suspected that other factors were influencing the behavior of the product.

Moisture Removal

The housing material, which is fiber-filled polyphthalamide, has a high affinity for moisture.[1] Water molecules from the ambient air form hydrogen bonds to polyamide linkages. This plasticizes the continuous matrix and causes dimensional changes, leading to increased mechanical stresses within the package over time.

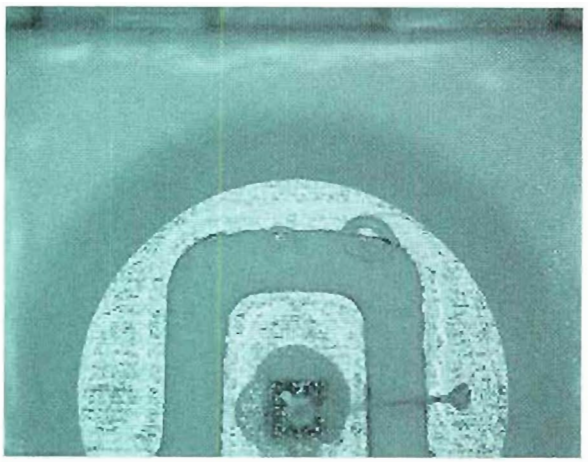
To quantify the effect of moisture on the overall package performance subsequent to epoxy encapsulation, an experiment was carried out by preconditioning the housing material (48 hr at 85°C/85%RH) before the casting process. One lot was used as the control cell. After the encapsulation cure, entrapped air bubbles were detected on the evaluation cell (**Figure 2**) but not on the control cell. The curing temperature was 135°C for 8 hours. The entrapped air bubbles may have been the result of moisture turning into steam during the gelation process.[5,6]

It was found that the glass transition temperature of the casting epoxy for the cell under evaluation was 116°C versus 139°C for the control cell.

The water molecules from the housing material can take part in ring-opening reactions with anhydride molecules competing with the –OH groups from DGEBA (diglycidyl ether of bisphenol A),[7] as illustrated in **Figure 3**. This results in a lower cross link density in the polymer matrix.

Figure 2

Entrapped air bubbles during the epoxy gelation process.



The moisture content in the housing material (polyphthalamide) is inconsistent and depends on the degree of exposure to the moisture in the ambient air. Therefore, drying or baking the housing material before casting is very critical.

To confirm the above hypothesis, an experiment was carried out with three factors: preconditioning after wire-bonding for 48 hours, drying after preconditioning, and different cast epoxy cure temperatures. The various cells and the cumulative failure rates are shown in **Table II**.

When there was no preconditioning after wire bonding, epoxy curing at 150°C gave an almost zero failure rate. When there was preconditioning after wire bonding, epoxy curing at 150°C still gave a lower failure rate than epoxy curing at 125°C. With no drying after preconditioning, the failure rates were greater than 80%. When there was preconditioning after wire bonding, no matter what drying condition was used the failure rate was not zero.

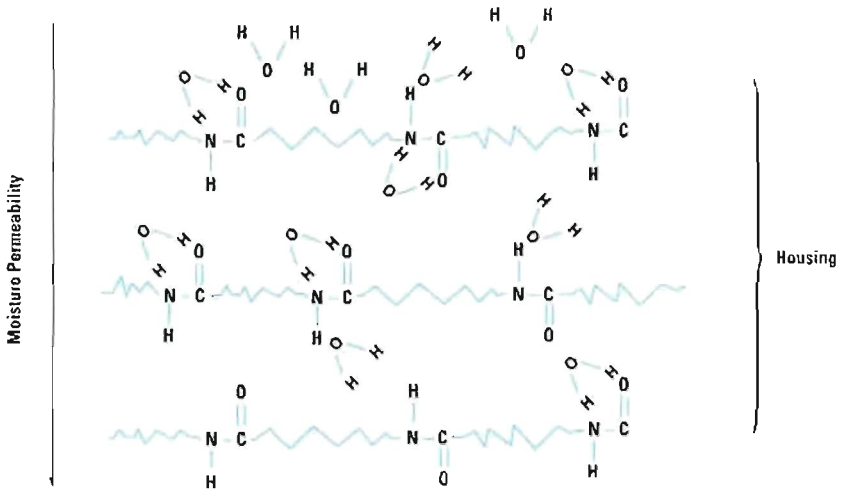
The preconditioning after wire bonding was hypothesized to be too severe and a more realistic experiment that simulated the production floor environment was done. After wire bonding, units were left exposed for 60 hours in an open environment where the room temperature reached 37°C and the relative humidity was between 60% and 80%. Having established that a 150°C epoxy cure temperature is superior, it was kept constant in the subsequent 2^3 full factorial experiment, which is summarized in **Table III**.

The analysis of variance revealed leadframe drying to be the biggest factor. The other factors had less than one tenth the significance of leadframe drying, so they were grouped together as residuals and another analysis of variance was done. This showed leadframe drying before dispensing to be significant at a 95% confidence level. Even though the oven ramp rate did not show up as a significant factor, the raw data indicated that the fast ramp gave fewer failures. The same can be said for the inline position of the magazine. This is probably due to better air circulation and better temperature control in the box oven.

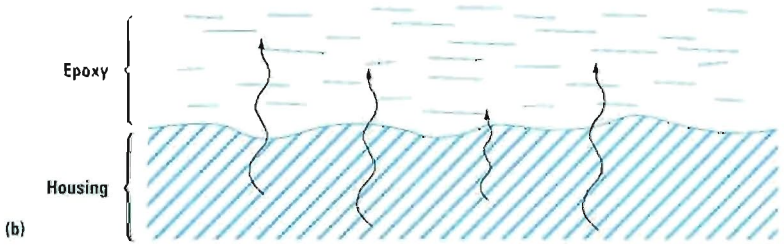
Physical analysis showed that the main failure mode for undried leadframes is broken stitch bonds. In those cells with dried leadframes, stitch bond failures were eliminated and reliability performance became consistent.

Figure 3

Schematic diagram showing water molecules (a) diffusing into the polymer matrix housing and (b) turning to water vapor during the cast epoxy curing process.



(a) Moisture from hydrogen bonding with –[–NHCO–]– reduces degree of crystalinity.



(b) During curing at high temperatures, hydrogen bonds dissociate and moisture diffuses into the epoxy mixture.

Table II

Failure Rates with and without Moisture Preconditioning and Drying

Cell Number	Preconditioning at 85°C/85% RH after Wire Bonding	Drying after Preconditioning	Epoxy Cure Temperature (°C)	Cumulative Failure Rate after 200 Temperature Cycles (%)
1	No	No	125	24.5
2	No	No	150	0.2
3	Yes	No	125	85.9
4	Yes	No	150	80.0
5	Yes	1 hr at 145°C	125	23.1
6	Yes	1 hr at 145°C	150	12.7
7	Yes	Vacuum	125	10.7
8	Yes	4 hr at 145°C	125	13.1
9	Yes	4 hr at 145°C	150	4.6

Table III

2^3 Full Factorial Experimental Design and Results

Run Number	Leadframe Baked before Dispensing Epoxy	Oven Ramp Rate*	Position of Magazine in Oven**	Cumulative Failure Rate after 300 Cycles*** (%)
1	No	Slow	Broadside	51.0
2	Yes	Slow	Broadside	0.6
3	No	Fast	Broadside	26.0
4	Yes	Fast	Broadside	0
5	No	Slow	Inline	25.5
6	Yes	Slow	Inline	0.2
7	No	Fast	Inline	11.7
8	Yes	Fast	Inline	0

* Two ovens were used. One was set up with a slow temperature ramp (0.5°C/minute) in the heating profile and the other with a fast ramp (5°C/minute) to a stable operating temperature of 150°C.

** The positions of the magazine where the leadframes were stored during the cure cycle were such that the airflow was blocked by the side plate (broadside) and the airflow was over the leadframes (inline).

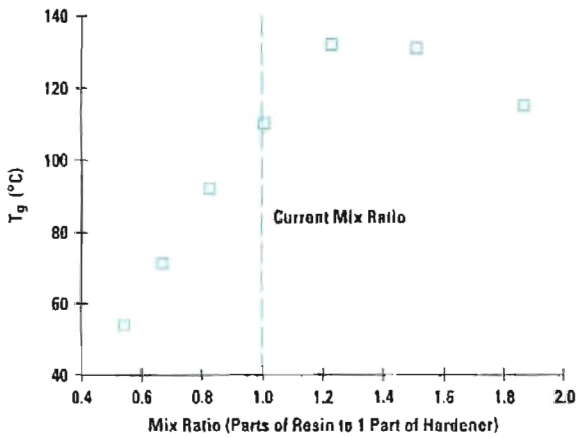
*** Temperature cycling between −55°C and +100°C.

Cast Epoxy Mix Ratio Considerations

An objective of this project was to make the product insensitive to moisture, thereby giving it unlimited shelf life for automotive applications. The previous product, if exposed to ambient conditions for more than a month, would typically fail after soldering, with severe epoxy cracks and delamination. Therefore, strengthening of the cast epoxy was crucial for this project.

Figure 4

Graph showing glass transition temperature at different mix ratios of resin to hardener.



The cast epoxy mix ratio recommended by the vendor is one part of resin to one part of hardener by weight. However, T_g (glass transition temperature) data indicates that the optimum dimensional stability for the epoxy system used falls in a range of resin ratios of 1.2 to 1.3 (**Figure 4**). The recommended mix ratio of 1.0 has a T_g from 110°C to 120°C. This is also the upper temperature experienced by the device during thermal cycling stress tests. Ideally, it is most desirable to have the material retain its glass-like behavior at a temperature much higher than the upper temperature extreme of the stress test (110°C for thermal shock).

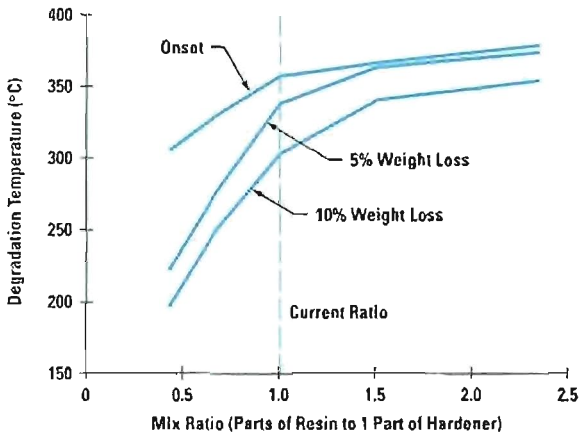
Thermogravimetric analysis (TGA) showed that when the resin content is increased the thermal stability of the epoxy system also improves (**Figure 5**). The optimum range of resin ratios is 1.1 to 1.5 and is a compromise between thermal stability and dimensional stability.

Epoxy Mix Ratio Optimization

A first pass experiment was done to determine if the package performance could be further optimized by modifying the epoxy resin-to-hardener mix ratio. Evaluation units were built using three different resin-to-hardener ratios of 1.0, 1.2, and 1.8 and then cured at 150°C and 125°C. The results after the preconditioning and stress tests are as shown in Table IV.

Figure 5

Graph showing thermal stability at different mix ratios of resin to hardener.



The resin ratio of 1.2 cured at 150°C gave a zero failure rate in all of the tests.

Spider cracks, as shown in **Figure 6** at the top of the wire loop, are caused by moisture introduced during preconditioning. These are seen in the control cell but not in the cell that has a resin ratio of 1.2. When the epoxy is saturated with moisture, spider cracks are normally observed after soldering. Cracking in epoxy resin induced by water absorption is a very well-known effect.[8] During the soldering process, the temperature exceeds the T_g of the cast epoxy causing it to become rubbery. The top of a gold wire loop will act as a stress initiator and a crack will propagate as a result of repeated temperature cycling.

A resin ratio of 1.8 is too high and leads to brittleness as evidenced by vertical cracks observed during the lead-forming operation (**Figure 7**). This shows that thermal stability alone is insufficient and mechanical stability is also needed to maintain package integrity.

It is possible that altering the mix ratio leads to a greater degree of cross-linking density in the epoxy matrix and results in higher T_g and increased modulus at temperatures above T_g, both commonly known effects that would enhance the dimensional stability of the epoxy with respect to temperature and thereby eliminate spider cracks.

Another advantage of a higher resin content in epoxy-anhydride systems is the reduction of the concentration of –COOH and –COO– hygroscopic linkages (**Figure 8**). This, coupled with higher cross-link density, reduces the moisture uptake capability, making the properties of the material less susceptible to change when exposed to moisture.

A fine-tuning of the resin ratio was done by using curing ratios of 1.0 through 1.4 at 150°C and checking the reliability through thermal shock and thermal cycling after 168 hours of preconditioning (**Figure 9**). Again, the results showed that using an epoxy mix ratio of 1.2 gave the best overall performance in thermal stress tests.

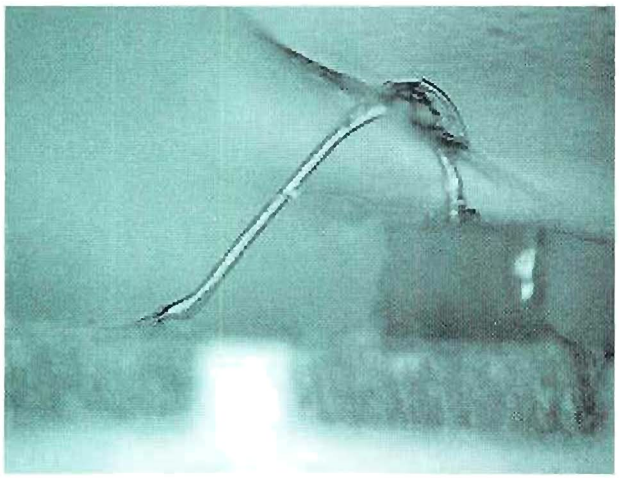
Table IV

Cumulative Failure Rates for Different Epoxy Resin-to-Hardener Mix Ratios

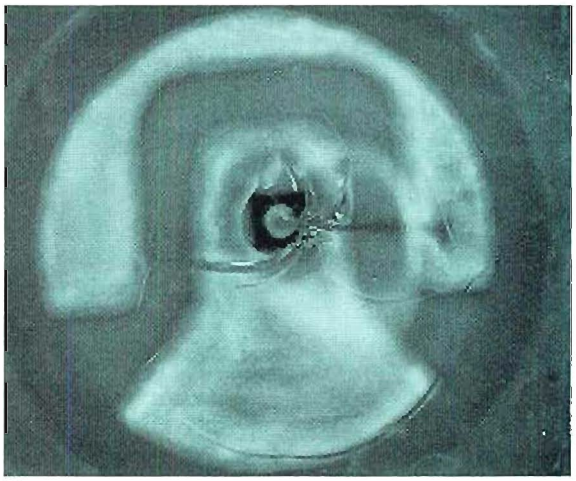
Mix Ratio	Cure Temperature (°C)	% Cracked during Forming	% Cracked after Preconditioning and Thermal Shock	% Cracked after Preconditioning and Thermal Cycling	Cumulative Failure Rate (%) after 300 Thermal Shocks	Cumulative Failure Rate (%) after 300 Thermal Cycles
1.0	125	1	53	68	2.0	1.8
1.2	125	1	0	1	1.2	0.2
1.8	125	20	52	43	75.4	64.4
1.0	150	0	43	38	0.4	1.6
1.2	150	0	0	0	0.0	0.0
1.8	150	13	51	42	20.0	3.0

Figure 6

Spider cracks.



(a)



(b)

Epoxy Mix Ratio and Cure Temperature

After the experimental noise had been eliminated, another experiment was run to check the significant factors. A 2^2 full factorial experiment was designed using two ovens. Each oven was stabilized at one temperature and the magazines of leadframes were loaded such that the temperature rise time was minimized. All the units were pre-encapsulation dried at the stipulated temperature and time. The results are summarized in **Table V**. The results proved beyond a doubt that the high cure temperature with a fast ramp and a 1.2 mix ratio is indeed the better process.

Figure 7

Vertical epoxy crack after the lead forming operation.

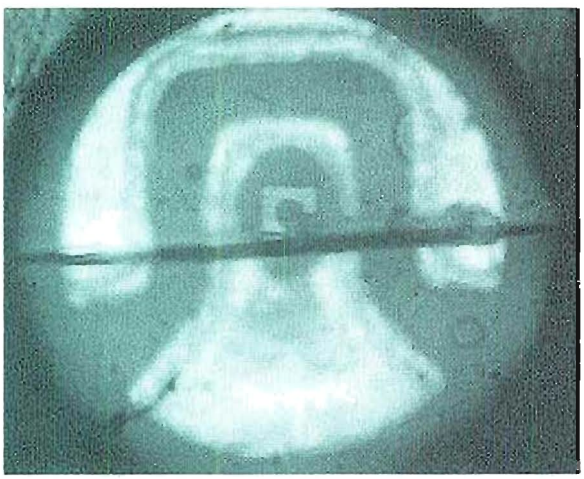


Figure 8

Anhydride esterification reaction with the oxirane ring.

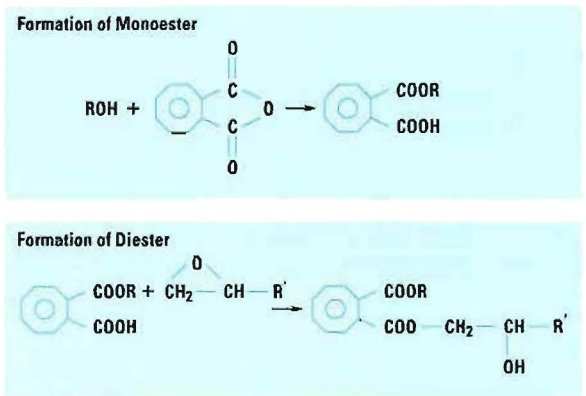


Figure 9

Epoxy mix ratio optimization reliability performance. All parts were preconditioned for 48 hr at 85°C/85% RH. (a) Thermal shock, −40°C to 110°C. (b) Thermal cycling, −55°C to 100°C.

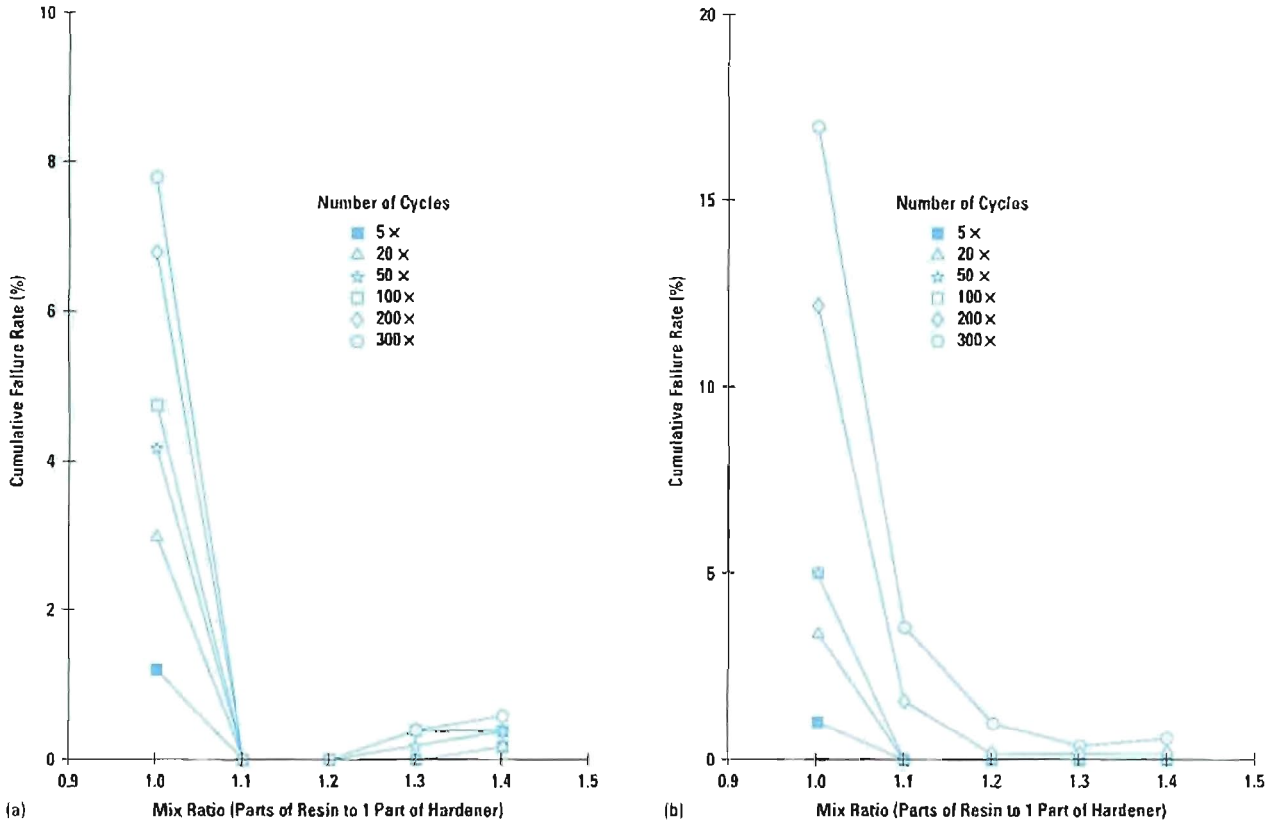


Table V

Failure Rates for Various Epoxy Mix Ratios and Cure Temperatures

Run Number	Epoxy Mix Ratio (Resin to Hardener)	Cure Temperature (°C)	Cumulative Failure Rate (%) after 500 Temperature Cycles
1	1.0	125	8.1
2	1.2	125	6.4
3	1.0	150	0.9
4	1.2	150	0.05

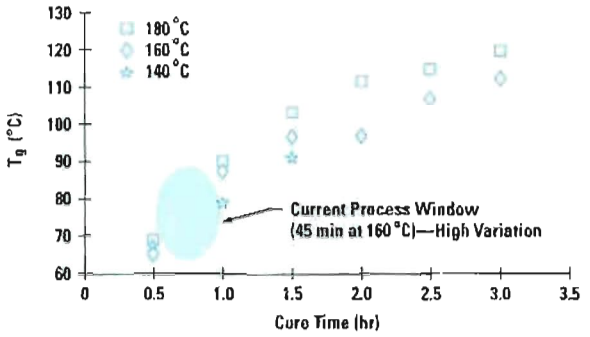
Die-Attach Silver Epoxy

The predominant failure mode of this surface mount package after solder reflow is lifted die-attach. There is a visible delamination between the leadframe and the silver epoxy die-attach material, causing an electrical discontinuity. Hence, there was a need to investigate the die-attach process and improve adhesion of the die to the leadframe using the current silver epoxy.

The T_g of the silver epoxy was measured as a function of cure time and temperature as shown in **Figure 10**. Using the current cure profile, a slight variation in the time or temperature of the cure will result in a high variation of T_g values. However, for this package there is a limit on the time and temperature of heat exposure because the housing material is prone to oxidation (color changes).

Figure 10

Glass transition temperature T_g for die-attach epoxy cured at different times and temperatures.



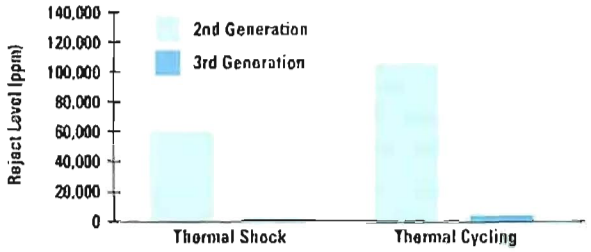
Ideally, the T_g should be as high as possible and in the stable region of the curve. As in the case of cast epoxy, the silver die-attach epoxy will be weak and rubbery if insufficiently cured. A higher cure condition will yield a higher T_g that gives a stronger and more stable adhesion.

To prove the benefits of high temperature and long cure time, a 2^2 factorial experiment was designed with thermal cycling reliability taken as the response. Die height was chosen as one factor because of a possible interaction between the die form factor and epoxy adhesion. The other factor was the cure condition, and the conditions investigated were 45 minutes at 160°C and 200 minutes at 180°C. The latter cure condition is known to discolor the housing material but is recommended by the vendor as optimum.[10]

The experimental results showed that the cure condition is more significant than the die height and that a higher cure temperature and a longer time will yield a higher T_g ($T_g = 114$°C for a cure of 200 minutes at 180°C) and better temperature cycle reliability. However, the high temperature and long time required for curing will oxidize the housing material. Curing in an inert atmosphere such as nitrogen will prevent the oxidation but is not cost-effective for this product. Characterization of the discoloration with different temperatures and times showed that the highest cure condition possible is 160°C for 90 minutes and represents the best compromise between silver epoxy T_g requirements and package cosmetics.

Figure 11

Third-generation product performance after preconditioning, IR soldering, and 1000 temperature cycles.



Conclusion

The reliability of HP's surface mount HSMx-Tmm LEDs was enhanced to better suit the increasingly stringent automotive applications. The preencapsulation drying of the package helped eliminate broken stitch bonds. A new casting epoxy formulation with new curing conditions improved the moisture sensitivity of the package, thereby eliminating epoxy cracking and raising the packaging standard from level 3 to level 1 of the relevant JEDEC standard. Optimization of the die-attach epoxy cure schedule eliminated the lifted die-attach and delamination problems. With these changes, the final test comparison showed that the third generation product is as much as 84 times more reliable than the second generation product (**Figure 11**).

Acknowledgments

The authors wish to thank their management, especially Bill Majkut, Steve Paolini, and Cheok Swee Beng, for supporting us throughout this project. Special thanks to Lim Chye Bee and Azlida Ahmad for coordination of the qualification and reliability builds. We are also grateful to Chris Togami, Cheng Ooi Lin, and Nayan Ashar for vendor liaison. Finally, we wish to thank Marina Chan, Teoh King Long, Lim Guat See, Hatijah Ahmad, and all of the production personnel, who have contributed greatly to the success of this project.

References

1. Cenelec Electronic Components Committee, *Method for the Specification of Surface Mounting Components (SMDs) of Assessed Quality, 1st Edition*, CECC 00802, 1990.

2. C.R. Totten, "Managing Moisture-Sensitive Devices," *Circuit Assembly*, September 1996.

3. Test Method A112, *Moisture-Induced Stress Sensitivity for Plastic Surface Mount Devices*, JEDEC Standard JESD22-A112, Electronic Industries Association, April 1994.

4. *Amodel PPA Resin Engineering Data*, Amoco, Reference AM-F-50060, 1990, pp. 44-45.

5. Aerospace Polymer Technology—Epoxies, http://www.corrosion.com/advanced/epoxies.html

6. C.S. Leech, Jr., "Removing Moisture from Electronic Components and Assemblies," *Circuit Assembly*, May 1994, pp. 34-37.

7. X. Sedlecek and J. Kahovec, *Crossedlinked Epoxies*, De Gryter, 1989, p. 117.

8. T.S. Ellies and F.E. Karasz, "Interaction of Epoxy Resins with Water: the Depression of Glass Transition Temperature," *Polymer*, Vol. 25, 1984, pp. 664-669.

9. H. Lee and K. Neville, *Handbook of Epoxy Resin*, McGraw-Hill, 1967, pp. 4-6.

10. *Ablebond 84-1lnLS Technical Data Sheet*, Ablestik, August 1993.

Online Information

More information about HP LEDs for automotive applications can be found at:

http://www.hp.com/go/automotive

Koay Ban-Kuan

Koay Ban-Kuan graduated in 1990 from the Universiti Sains Malaysia with a B.App.Sc. degree. He joined HP Malaysia in 1992 as an engineer. He has led and participated in various LED projects and is currently leading the efforts to make the HP HSMx-Tmm LEDs meet automotive reliability requirements and tolerate high-moisture conditions. He is a member of MINDS (Malaysian Invention and Design Society) and expects to complete his studies for the MBA degree this year.

Yoong Tze-Kwan

Yoong Tze-Kwan graduated from the University of Birmingham, UK, with a BSc degree in electronics and electrical engineering in 1986. He joined HP Malaysia in 1991 and has worked on process engineering for LEDs and high-pin-count hermetic ICs. His hobbies include tropical fish, aviation, and photography. Tze-Kwan recently left HP.

Leong Ak-Wing

Leong Ak-Wing has been an R&D engineer with HP Malaysia since the department was formed in 1991. He has been a project leader for automotive exterior lamps and for improving the reliability of the HP HSMx-Tmm surface mount lamps. Schooled in Kuala Lumpur's Technical College in radio and TV engineering, he joined HP in 1973. He enjoys reading and gardening. For exercise, he hikes up hills barefoot.

Tan Boon-Chun

Prior to his recent departure from HP, Tan Boon-Chun was a senior engineer at HP Malaysia specializing in epoxy encapsulation materials. He was working on a project to develop an epoxy formulation that has high heat resistance and good weathering properties for LED applications. Boon-Chun graduated from the Universiti Sains Malaysia majoring in polymer technology.

Engineering Surfaces in Ceramic Pin Grid Array Packaging to Inhibit Epoxy Bleeding

Ningxia Tan
Kenneth H. H. Lim
Bernard Chin
Anthony J. Bourdillon

Bleeding of epoxy resin around surfaces undergoing bonding during electronic packaging assembly has long caused sporadic yield loss. Previously, it was thought that vacuum baking reduced the yield loss resulting from surface contaminants. Although vacuum baking inhibits epoxy resin bleeding, it also produces coatings of hydrocarbons, which affect surface wettability and surface energy. Surfactant coating results in a surface chemistry similar to vacuum-baked substrates but is a better alternative because of its controllability.

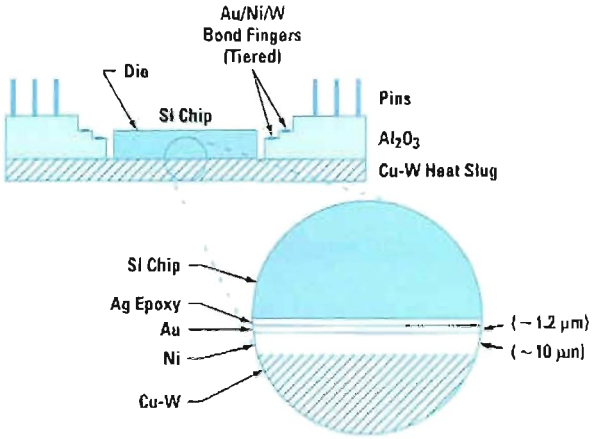
Epoxy bleeding is commonly observed in electronic packages around silicon chips attached with epoxy resin to substrates having gold or other metal surfaces. In severe cases, the bleeding contaminates the wire bond pads, causing failure. The effects of bleeding are often critical in advanced packaging components such as ceramic pin grid array (CPGA) substrates. In particular, when there is very little clearance between a bond finger tier and the die, minor resin bleedout can interfere with wire bondability.

During the time that CPGA technology has been used here at HP's Integrated Circuit Business Division in Singapore, we have experienced sporadic yield loss caused by epoxy bleeding. This is known to be an industry-wide problem. For several years the cause of the yield loss was unresolved. With increasing pin count and decreasing clearance between the die and the substrate cavity, resolving the problem has become more urgent.

Earlier studies[1] led us to resort to countermeasures like vacuum baking to reduce yield losses. However, the studies were not comprehensive enough to verify the effectiveness of vacuum baking, and yield losses have continued to occur from time to time for no apparent reason. We had not yet investigated the possibility that surface contamination, resulting from vacuum baking, can have

Figure 1

Schematic diagram of a CPGA substrate with silicon chip mounted on the die attach pad, which is the central portion of the heat slug. The epoxy bleeding occurs on the Al_2O_3 walls near the silicon chip.



the positive effect of reducing epoxy bleeding during die attachment.

In this article we describe how we used surface analysis and contact angle measurements to investigate the effects of vacuum baking on yield loss caused by epoxy bleeding. The same analysis techniques were employed to investigate the possibility of using surfactant-coated substrates to reduce epoxy bleeding.

For our analysis we used a CPGA (ceramic pin grid array) substrate[2] from one of our typical applications (see **Figure 1**). It consisted of a Cu-W heat slug onto which was electroplated a film of Ni (10 μm thick) followed by Au (typically 1.2 μm thick). The surface was cleaned and vapor dried with isopropyl alcohol. When the thickness of the surface gold film was varied from 1 μm to 2 μm on different specimens, the wettability was found to be independent of the gold thickness. The schematic in **Figure 1** shows the substrate after die attachment but before wire bonding.

Analysis of Vacuum-Baked Substrates

Investigations into the effects of vacuum baking focused on two areas: Auger analysis and wettability.

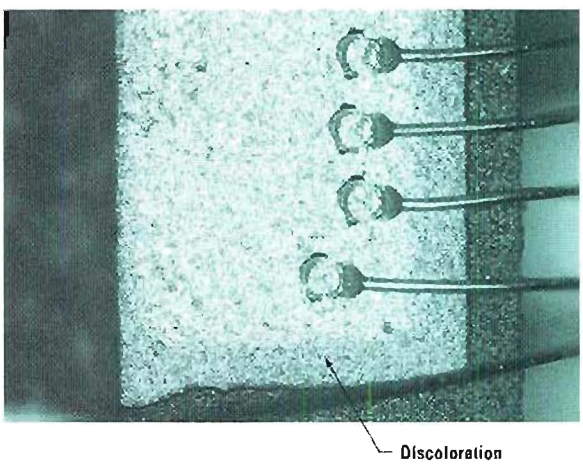
Auger Analysis. Auger analysis is an analytical technique that uses electron spectroscopy to examine the elemental composition of the outer atomic layer of a solid material (see "Auger Analysis" on page 83). Examining the elemental composition of surface contaminants on a solid is one application of this technique. Our analysis was performed on a JEOL JAMP-7100E Auger analyzer. The accelerating voltage was 5 kV, and the probe current was 1.52×10^{-7} A. Argon ion etching was applied with an etching speed estimated at 1.25 nm every 10 seconds on a silicon surface.

The analysis was carried out on substrates prepared for die attachment. **Figure 2** shows the discoloration caused by bleeding around a bonding pad. Auger spectra were recorded from the die attach pads (heat slugs), which consist of Cu electroplated with Ni and Au. Comparisons were made between samples in a vacuum at two stages during the process: before baking and after baking. The purpose of this was to monitor the contamination that results from baking. The samples consisted of the following:

- A raw CPGA substrate
- A CPGA substrate that was baked at 235°C for six hours in a conventional vacuum oven at a pressure of 0.1 mbar
- A CPGA substrate that was baked in an air convection oven at 235°C for six hours.

Figure 2

Photographic image of the die showing attached wires and edge discoloration (see arrow) caused by epoxy bleeding.



Glossary

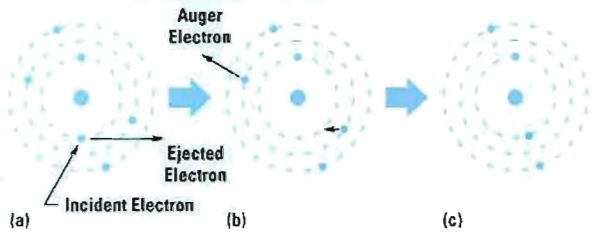
Auger Analysis

Auger (pronounced o-jay) analysis is a widely used electron spectroscopic method that is capable of providing information about the elemental composition of the outermost atomic layer of a solid. This technique examines the surface chemistry and interactions of elements on the surface of materials such as metals, ceramics, and organic matter. Auger analysis is named after its discoverer Pierre Auger.

The Auger process involves using a finely focused electron beam to bombard atoms on the surface of the sample being analyzed. When an atom on the sample is struck by a high-energy electron, there is a probability that a core-level electron will be emitted (**Figure 1a**). This collision puts the atom into an energetic ionic state with an electron missing from the core level. The atom can relax into a lower-energy state when another electron from the same atom falls from a higher-energy level to fill the core-level hole, releasing enough energy to eject a second electron—the Auger electron (**Figure 1b**). The state shown in **Figure 1c** is still excited and decay continues radiatively by another Auger process. Auger electrons will have an energy characteristic of the parent element. An energy spectrum of detected electrons shows peaks assignable to the elements present in the sample.

Figure 1

An illustration of the Auger process.



The ratios of the intensities of Auger electron peaks can provide quantitative information about the surface composition of a sample (for example, see **Figure 3** of the main article).

Surfactant

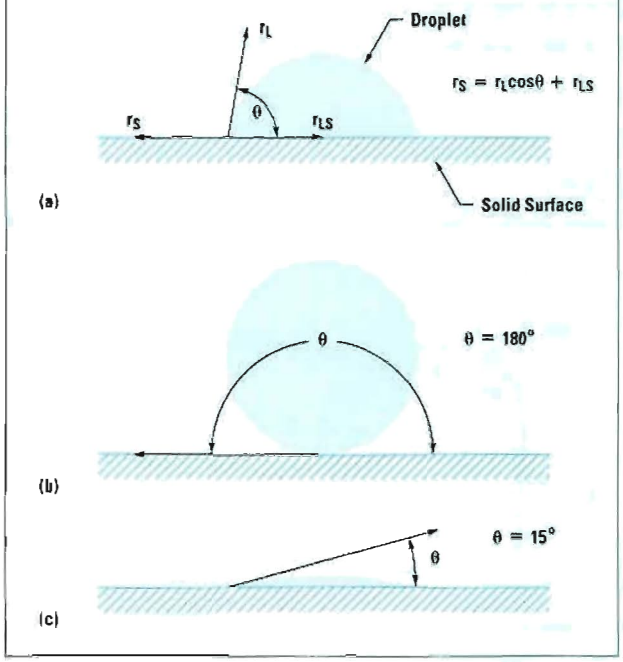
Some chemical materials have a special propensity to locate (adsorb) at interfaces or to form colloidal aggregates in solution at very low molar concentrations. Such materials are called surface-active agents, or surfactants. Surfactants are used to modify the wettability of solid surfaces.

Wettability

When a drop of liquid is placed on a solid surface, the liquid either spreads to form a thin film on the surface as in **Figures 2a** and **2c**, or remains a discrete drop as in **Figure 2b**. The measure of the degree of wetting is the *contact angle* (see **Figure 2a**). The contact angle is the angle made by the tangent to a droplet at the solid surface interface. High wettability is indicated by a small angle (**Figure 2c**). The extreme case of no wettability is shown in **Figure 2b**.

Figure 2

Contact angles. (a) Partial wetting. (b) No wetting. (c) Close to complete wetting.



Measurements were taken from several specimens of each type and found to be consistent. For comparison, additional measurements were taken from a second specimen (specimen B in **Table I**) which was subjected to evacuation at room temperature, without baking.

Wettability, Contact Angle, and Surface Energy. A liquid that spreads easily on a solid because of high surface energy has high wettability. This is quantified by measuring the angle that the liquid surface makes at the interface with the solid surface. Such wettability contact angles were measured through the sessile drop method[3] on a face contact anglemeter, model CA-A from Kyowa Interface Science Company.

Media used for calculating surface energies were deionized water and methylene iodide. Contact angle measurements were performed on raw substrates and on substrates after vacuum baking. To measure contact angles, the die pad area of a package was separated from the corresponding ceramic substrate by chipping away the side walls. A horizontal profile projector, at 20× magnification, was used to measure the equilibrium contact angle.

Repeated measurements from several specimens of each type of substrate were found to be consistent to within three degrees. The computation for the surface energy, r_S, is based on the method described by Wu and Brzozowski.[4] The surface energy of the solid is the sum of the surface dipole component, r_S^p, and the surface dispersion component, r_S^d. These components are related to the contact angle θ by the formula:

$$1 + (\cos\theta)r_L = 4\left(\frac{r_L^d\, r_S^d}{r_L^d + r_S^d} + \frac{r_L^p\, r_S^p}{r_L^p + r_S^p}\right) \qquad (1)$$

where r_L is the surface tension of the liquid used in wetting and consists of the sum of the dispersion and polar components. If all of the components of r_L are known for two liquids, then two corresponding measurements of the contact angle will make it possible to solve for r_S^d, r_S^p, and so on by solving the quadratic equation derived from equation 1.

Preliminary Results

Our quantitative analysis focused on Auger analysis as well as contact angle measurement and surface energy computation.

Auger Analysis. **Figure 3a** shows typical Auger spectra taken from a raw substrate. The etching times were zero seconds and five seconds. The elements detected on the surface were Au, S, C, and N. After five seconds of etching (**Figure 3b**), two of the contaminants, C and N, disappeared. A trace of S on the subsurface was estimated to be at a mean depth of 1.25 nm below the original surface.

Figure 3

Auger spectra generated from the die attach pad of a raw CPGA substrate after etching times of (a) 0 seconds and (b) 5 seconds.

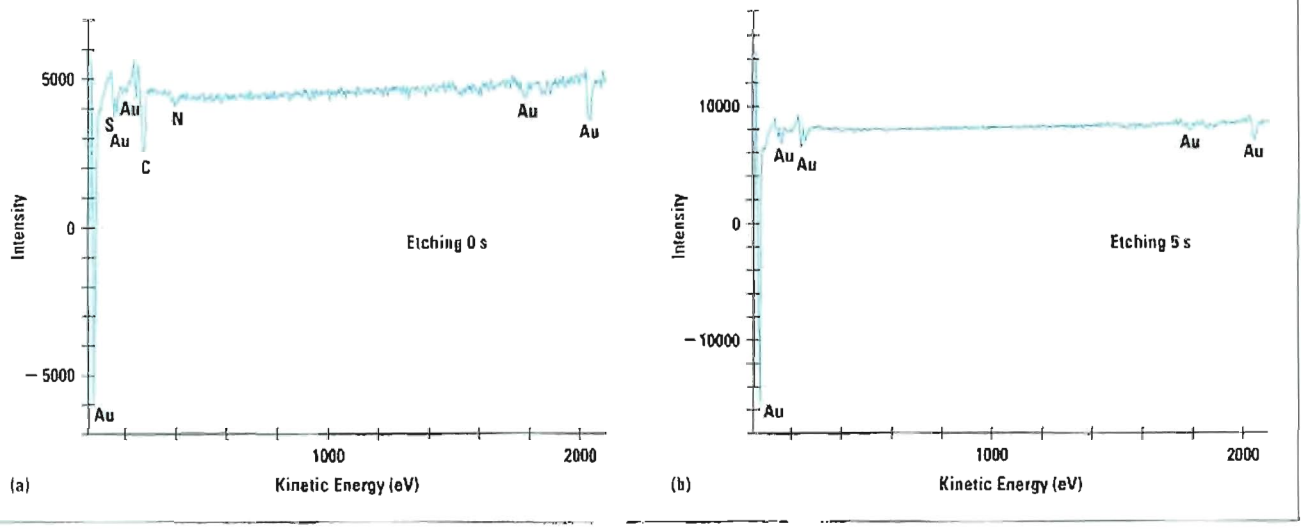
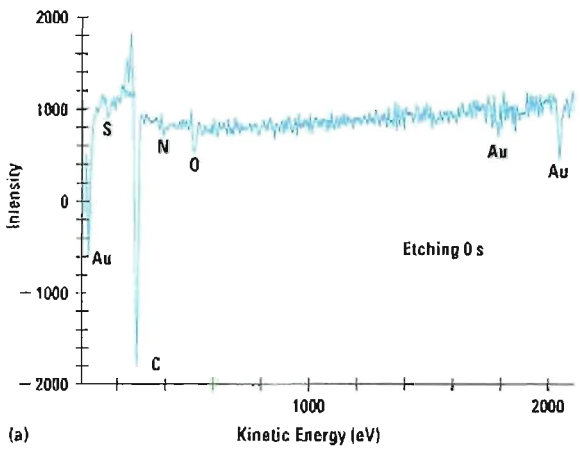
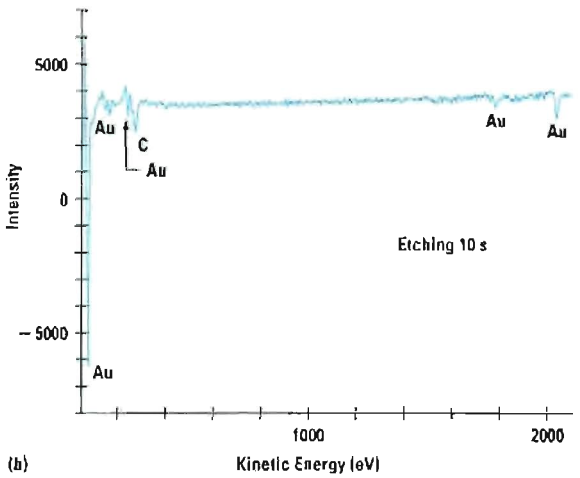


Figure 4

Auger spectra generated from the die attach pad of a CPGA substrate after vacuum baking and after etching times of (a) 0 seconds, (b) 10 seconds, and (c) 20 seconds.





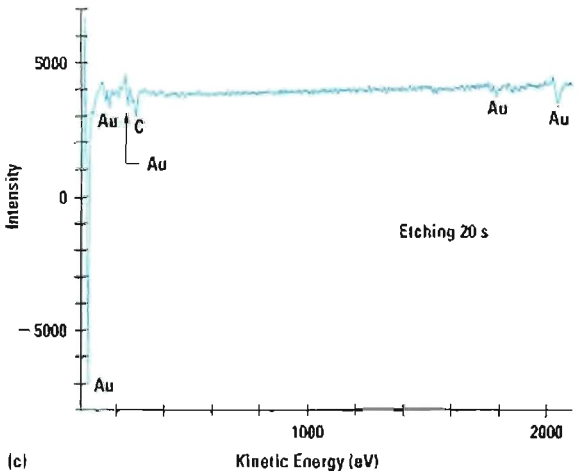


Figure 4 shows the Auger spectra taken on a substrate after vacuum baking. Etching times of zero seconds, ten seconds, and twenty seconds were used. Notice that, compared with the raw substrates, the C contamination is much greater. Even after a 20-second etching time (equivalent to an etching depth of about 5 nm), C continues to be observed, showing that the thickness of the carbon film is greatly increased by the baking procedure.

Auger spectra were taken on the substrate after baking in air only. Etching times of zero, five, and ten seconds were applied (see **Figure 5**). Notice that after five seconds of etching, C contamination was gone. The C contamination level is comparable with spectra for the raw substrate shown in **Figure 3b**. Although they were etched away before the plot in **Figure 5b** was made, traces of Ni and O were detected on the surface. These traces apparently result from diffusion of subsurface Ni during baking and from its oxidation by air.

Table I lists the results of the quantitative analysis. The results from specimen 2 show that evacuation at room temperature is not contaminating by itself. Some of the contamination comes from a dynamic process in which a cold specimen is immersed in a hot oven containing backstream oil vapor.

Contact Angle Measurement and Surface Energy Computation. The measurement and computational results[4] from a raw substrate before vacuum baking and after vacuum baking are listed in **Table II**. In these measurements, possible variations related to surface morphology were minimized by comparing measurements from specimens of the same lot. Furthermore, gold-plated surfaces are sufficiently dense to allow us to ignore surface morphology.

As described earlier, a significant increase in the carbon signal is observed after baking. At the same time, the solid surface energy decreases, especially the polar component. The observed decrease in surface energy is consistent with known surface energies from typical hydrocarbons, such as paraffin tetradecane (25.6 milliJoules per square meter, or mJ/m^2), so that the surface energy of the contaminated surface (33.6 mJ/m^2) lies between that value and the surface energy of the raw surface (41.1 mJ/m^2). Thus, the effective polarity of the substrate is reduced by the separation from the liquid drop provided by the insulating hydrocarbon film. This is also consistent with the knowledge that hydrocarbons prevent the spreading

Figure 5

Auger spectra generated from the die attach pad of a CPGA substrate after baking only in air and after etching times of (a) 0 seconds and (b) 5 seconds. After etching for 10 seconds, the spectrum was similar to (b).

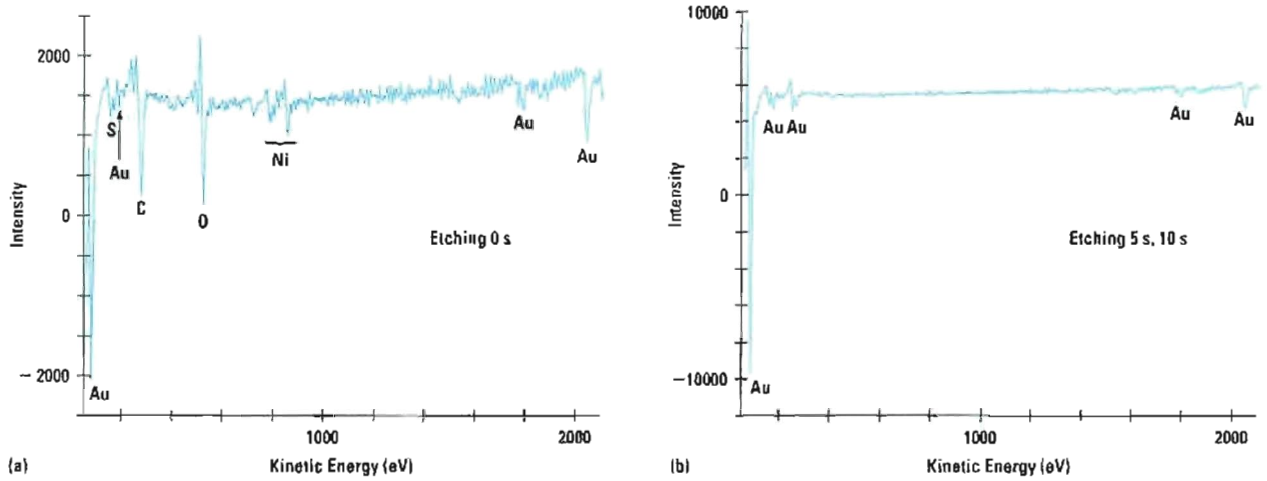


Table I

Elemental Analysis of CPGA Surface from Auger Spectra

	Etching Time (s)	Etched Thickness (nm)	Chemical Concentration (Atomic %)					
			Au	O	C	S	N	Ni
Specimen A								
Raw Substrate	0		58		34	4	4	
	5	1.25	100					
Vacuum Baked	0		33	4	57	1	5	
	10	2.5	66		34			
	20	5.0	76		24			
Baked in Air	0		45	20	27	2	2	4
	5	1.25	100					
	10	2.5	100					
Specimen B								
Raw Substrate	0		45		51	1	3	
	5	1.25	100					
	10	2.5	100					
Ambient Evacuated	0		45		51	2	2	
	5	1.25	100					
	10	2.5	100					

Table II
Measured Equilibrium Contact Angles and Computed Surface Energies

CPGA Substrate	Contact Angle (Degrees)		Surface Energy (mJ/m²)*			Polarity**
	Water	Methylene Iodide	r_S^d	r_S^p	r_S	
Raw substrate	76.3	44.8	28.6	12.5	41.1	0.30
Vacuum baked	92.4	54.1	27.9	5.7	33.6	0.17
Surfactant coated	94.4	49.3	31.7	4.1	35.8	0.11

* The surface energy r_S is the sum of the dispersion component r_S^d and the polar component r_S^p.

** The surface polarity is equal to the ratio r_S^p/r_S.

of water on gold,[5] even at a coverage of one monolayer. The surface energy, as measured, is much less than the known surface energy of Au (1510 mJ/m²).[6] The very high surface energy of gold demonstrates the importance of the surface film on wettability. Clearly, the polar interaction of the gold surface is reduced by atmospheric contamination (water vapor and organics). With baking, the nonpolar oil vapor further reduces the polar interaction.

Preliminary Discussion

The Auger results discussed above show that carbon contamination increases dramatically after standard procedures of vacuum baking. Typically, oil from the rotary vacuum pump backstreams to the oven and diffuses pore contaminants in the process. The samples are placed in a hot oven, which is then pumped down to create a vacuum. Because of thermal gradients, hydrocarbon vapor can then preferentially condense on cold substrate surfaces. This likelihood is consistent with the Auger analysis data on substrates that have been vacuum baked. In confirmation, surfaces that are only exposed to a high vacuum at room temperature, but without baking (that is, without thermal gradients) show an insignificant increase in contamination deposition.

Since hydrocarbon contamination inhibits epoxy bleeding, it is not necessarily advantageous to eliminate surface contamination. Hydrocarbon films, deposited on substrates during vacuum baking, have reduced surface energies (**Table II**) compared with clean Au surfaces. Surface energy and contact angle are related as in the Young-Dupré equation:[7,8]

$$r_S = r_L \cos\theta + r_{LS} \qquad (2)$$

where r_{LS} is the interfacial tension. When the surface energy is reduced, $\cos\theta$ is correspondingly reduced, resulting in high contact angles. A high contact angle implies a less wettable surface and therefore inhibits spreading phenomena when an adhesive material is applied to the solid surface. Consequently, the resin bleeding, caused by epoxy applied for die attachment in electronic packaging substrates, will be inhibited by the coatings. Since the electroplated Au surfaces have natural porosity, any degradation in adhesion caused by the coating is not critical. (Further discussion of this adhesion is given later.) Also, the dependence of the coating on various vacuum systems is discussed in reference 9.

The coating on vacuum-baked substrates, whether caused by oil backstreaming or by substrate contaminating residue, cannot be controlled. Neither the source of contamination nor the deposition conditions can be precisely determined. Inconsistency in surface quality explains why the sensitive effects of the contamination result in sporadic yield losses. A technique is needed that is compatible with other required properties, such as adhesion strength, that will result in hydrocarbon film coatings with near optimum contact angle.

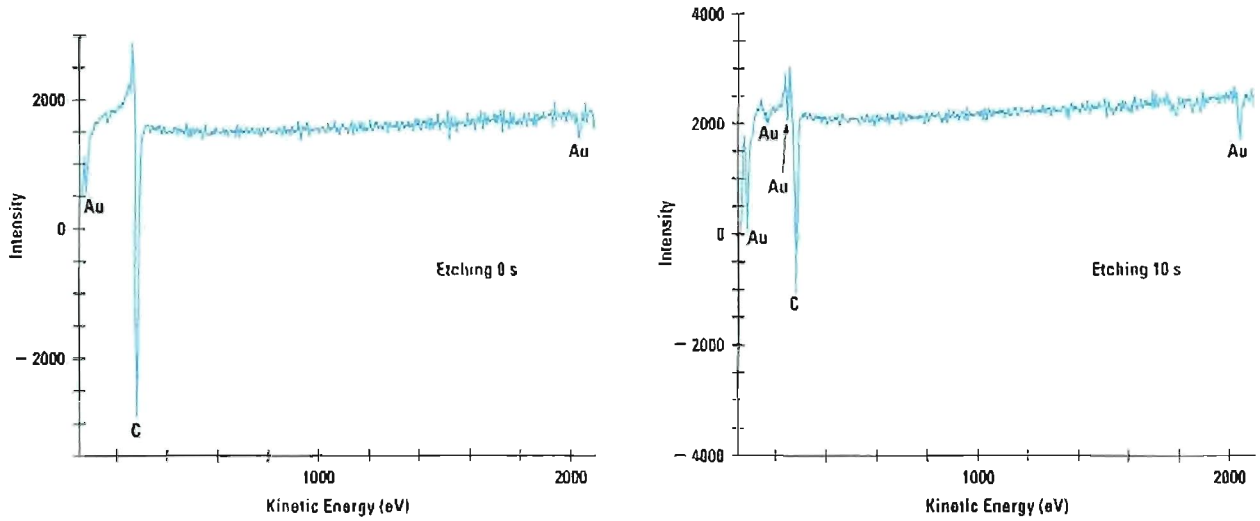
Analysis of Surfactant-Coated Substrates

A substrate was coated with a surfactant and subjected to the same Auger and contact angle characterization used on vacuum-baked substrates.

Auger Analysis. **Figure 6** shows the presence of carbon on the gold-plated layer. Initially, the carbon concentration is greater than in the vacuum-baked specimen shown in **Figure 4**. After 10 seconds of etching (**Figure 6b**),

Figure 6

Auger spectra generated from a surfactant-coated CPGA substrate after etching times of (a) 0 seconds and (b) 10 seconds.



there is still a considerable amount of surface carbon, showing that the film thickness of the hydrocarbon contaminants is greater than 1.2 nm.

As mentioned earlier, vacuum baking inhibits epoxy bleeding because of the hydrocarbon contamination it leaves on the surface of the substrate. Surfactant coating is also meant to contaminate a substrate's surface with a hydrocarbon film. From the results given in **Figures 4** and **6** it would seem that there is no difference between these two processes. However, from our study the shortcoming of the vacuum-baking process is that the hydrocarbon source is not controlled because of the oil backstreaming from the vacuum pump. On the other hand, surfactant coating can be controlled by mixing the right concentration of surfactant with water, forming the hydrocarbon film in the solution.

Wettability Contact Angle and Surface Energy. Equilibrium contact angles and computed surface energies on a surfactant-coated substrate are shown in **Table II** (last entry). The surface energy is considerably reduced from the raw substrate state and is close to the value calculated from the vacuum-baked substrate.

The coating consistency was evaluated by comparing contact angles of deionized water on 25 substrates sampled from five different lots. For an average measurement of 95°C the standard deviation was 1.9 percent. This deviation of less than 2 percent was about five times lower than was observed in vacuum-baked substrates and illustrates the satisfactory controllability of the surfactant coating process.

The coating stability was also investigated in two ways: by exposure to solvents and by exposure to air. After subjecting the coated substrate to typical cleaning processes, such as soaking in deionized water for 72 hours, Auger analysis showed no significant difference between substrates soaked 72 hours and substrates not soaked at all. Secondly, the contact angle was monitored in coated substrates exposed to a normal air-conditioned environment. With exposures of up to six months, no change was observed. However, at longer exposures, a small reduction in the contact angle was observed (for example, 5° after nine months). Thus, the surfactant coating process is compatible with production needs.

Discussion

Surfactant coating results in a surface chemistry similar to vacuum-baked substrates, but with the benefit of controllability. What effect do surfactants have on adhesion? The adhesion was tested by experiments involving die shear strength. Adhesive strengths greater than 20 kg/cm^2 were measured, consistent with U.S. military specifications.[10] Thus, although low wettability is generally accompanied by reduced adhesion, we found that the reduction

was not critical and within process margins. In these substrates, capillary effects, chemical bonding, and other factors ensure that the surfaces retain sufficiently strong adhesion, even when the surfaces are altered by the contamination we measured.

Similar tests were performed involving wire bond pull strength. In spite of the surface modification resulting from processing, pull strengths greater than 6 g for 1.2-μm-diameter gold wires were measured. Thus, the coated films were thin enough to have an insignificant effect on wire bond strength. The surfactant coated substrates passed Hewlett-Packard's general semiconductor qualification specification.

Finally, the surface tension is different in various adhesives, including various types of epoxy, polyimides, and similar materials. The selection of an adhesive depends on trade-offs between adhesion strength, curing times, moisture absorption, and other parameters. We have demonstrated that wettability can, in actual practice, be controlled.

Conclusion

This study illustrates the degree of cleanliness appropriate in processing of electronic packages. We had been striving for very clean ceramic surfaces, but that degree of cleanliness produces high surface energies that are susceptible to epoxy bleeding. It is not necessary to provide very clean substrates. On the contrary, it is preferable to engineer surfaces by treatment with a film of high wettability. Vacuum baking is not sufficiently controllable and is therefore not effective. A better alternative is an effective, controllable, simple, and reliable surfactant coating.

Acknowledgments

We are grateful to N.K. Chong, C.Y.S. Tan, and N.H. Lee for facilitating the project and to Y.C. Chong, J. Foo, and M. Dhaliwa for procurement support.

References

1. J.E. Ireland, "Epoxy Bleedout in Ceramic Chip Carriers," *International Journal for Hybrid Microelectronics*, Vol. 5, no. 1, 1982.

2. N.X. Tan, A.J.Y. Lee, A.J. Bourdillon, and C.Y.S. Tan, "Orientation Anomalies in Plating Thickness Measurements from Advanced Packaging Substrates," *Semiconductor Science and Technology*, Vol. 17, 1996, p. 437.

3. C.A. Miller and P. Neogi, *Interfacial Phenomena: Equilibrium and Dynamic Effects*, Marcel Dekker, New York, 1980, pp. 54-83.

4. S. Wu and K.J. Brzozowski, "Surface Free Energy and Polarity of Organic Pigments," *Journal of Colloid and Interface Science*, Vol. 37, 1971, p. 686.

5. V.A. Parsegian, G.H. Weiss, and M.E. Schrader, "Macroscopic Continuum Model in Influence of Hydrocarbon Contaminants on Forces Causing Wetting of Gold by Water," *Journal of Colloid and Interface Science*, Vol. 61, 1977, p. 356.

6. W.R. Tyson and W.A. Miller, "Surface Free Energies of Solid Metals: Estimation from Liquid Surface Tension Measurements," *Surface Science*, Vol. 62, 1977, p. 267.

7. A.W. Adamson, *Physical Chemistry of Surfaces*, second edition, Wiley Interscience, 1967.

8. J.J. Bikerman, *Physical Surfaces*, Academic Press, 1970.

9. N.X. Tan, K.H.H. Lim, and A.J. Bourdillon, "Analysis of Coatings Which Inhibit Epoxy Bleeding in Electronic Packaging," *Journal of Material Science: Materials in Electronics*, Vol. 8, 1997, p. 173.

10. *Test Methods and Procedures for Microelectronics*, Mil-Std-883D, Method 2019.5, 1993.

The Hewlett-Packard Journal

Subscriptions

The Hewlett-Packard Journal is distributed free of charge to HP research, design, and manufacturing engineering personnel, as well as to qualified non-HP individuals, libraries, and educational institutions.

To receive an **HP employee subscription** send an e-mail message indicating your HP entity, employee number, and mailstop to: **ldc_litpro@hp0000.hp.com**

Qualified non-HP individuals, libraries, and educational institutions in the **U.S.** can request a subscription by going to our website and following the directions to subscribe.

To request an **International subscription** locate your nearest country representative listed on our website and contact them directly for a subscription. Free subscriptions may not be available in all countries.

Back issues of the Hewlett-Packard Journal can be ordered through our website.

Our Website

Current and recent issues are available online at **http://www.hp.com/hpj/journal.html**

Submissions

Although articles in the Hewlett-Packard Journal are primarily authored by HP employees, articles from non-HP authors dealing with HP-related research or solutions to technical problems made possible by using HP equipment are also considered for publication. Before doing any work on an article, please contact the editor by sending an e-mail message to: **hp_journal@hp.com**

Copyright

Inquiries

Please address inquiries, submissions, and requests to:

Editor
Hewlett-Packard Journal
3000 Hanover Street, Mail Stop 20BH
Palo Alto, CA 94304-1185 U.S.A.

HEWLETT-PACKARD
Journal

AUGUST 1998 • Volume 49, Number 3

Technical Information from the Hewlett-Packard Company

5966-2131E