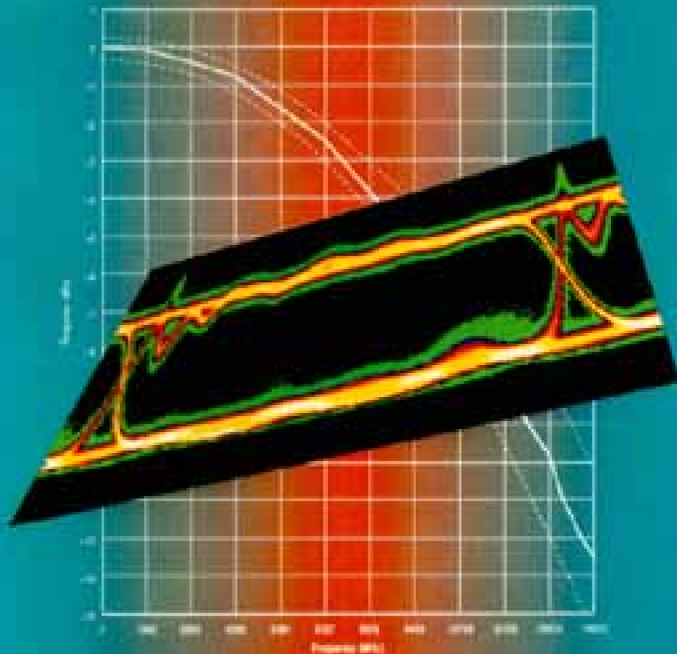
HEWLETT-PACKARD

JOURNAL

December 1996



HEWLETT PACKARD

table of contents

December 1996,
Volume 47, Issue 6

Articles

1. **A New Instrument for Waveform Analysis of Digital Communications Signals**
by Stephen W. Hinch, Michael J. Karin, and Christopher M. Miller

2. **Firmware Measurement Algorithms for the HP 83480 Digital Communications Analyzer**
by Michael G. Hart, Christopher P. Duff, and Stephen W. Hinch

3. **Design of Optical Receiver Modules for Digital Communications Analysis**
by Christopher M. Miller, Randall King, Mark J. Woodward, Tim L. Bagwell, Donald L. Faller, Jr., Joseph Straznicky, and Naily L. Whang

4. **Differential Time-Domain Reflectometry Module for a Digital Oscilloscope and Communications Analyzer**
by Michael M. McTigue and Christopher P. Duff

5. **Frequency Response Measurement of Digital Communications Analyzer Plug-in Modules**
by Rin Park and Paul D. Hale

6. **Radially Staggered Bonding Technology**
by Rajendra D. Pendse, Rita N. Horner, and Fan Kee Loh

7. **Implementation of Pad Circuitry for Radially Staggered Bond Pad Arrangements**
by Rita N. Horner, Rajendra D. Pendse, and Fan Kee Loh

8. **A Miniature Surface Mount Reflective Optical Shaft Encoder**
by Ram S. Krishnan, Thomas J. Lugaresi, and Richard Ruh

9 **The Global Positioning System and HP SmartClock**
by John A. Kusters

10 **The Third-Generation HP ATM Tester**
by Stewart W. Day, Geoffrey H. Nelson, and Thomas F. Cappellari

11 **Managed Objects for Internal Application Control**
by John P. Nakulski

12 **Developing a Design for Manufacturability Focus**
by John G. Fuller

13 **Production Test Strategy for the HP E5200A Broadband Service Analyzer**
by Cary J. Wright

14 **Usable Usability**
by Peter G. Tighe

A New Instrument for Waveform Analysis of Digital Communications Signals

The HP 83480 digital communications analyzer combines an optical reference receiver with an oscilloscope and communications measurement firmware. Its measurements meet the requirements of the SONET and SDH fiber-optic communications standards.

by Stephen W. Hinch, Michael J. Karin, and Christopher M. Miller

The telecommunications industry is currently experiencing a period of rapid growth and change. Just a few years ago a digital network needed a capacity of only a few hundred megabits per second to handle all the telephone conversations between such major metropolitan centers as San Francisco and New York. Today, to carry all the voice, fax, video, and data transmissions between these cities, a system needs a capacity of at least 2.5 gigabits per second (Gbits/s). Even this is not enough for the future. Ten-gigabit-per-second systems are already beginning operation in some areas, and researchers are hard at work increasing capacities into the range of 40 to 100 Gbits/s over the next several years.

This growth has spurred several fundamental changes in transmission systems. First is the rapid conversion from copper or microwave transmission media to fiber optic systems. Optical fiber offers many advantages over traditional copper: extremely wide bandwidth, low loss, high immunity to interference, and virtually no crosstalk between channels. Moreover, fiber does not suffer the annoying propagation delay typical of satellite-based communications. (It is easy to tell when a long distance call has been routed via satellite because of the noticeable delay between when you speak and when the listener hears your words.)

A second major change is the increasing importance of worldwide transmission standards. In years past, every equipment manufacturer used a separate proprietary transmission format, so the transmitter in one city and the receiver in another city had to be manufactured by the same vendor. With the deregulation of the U.S. phone system and the ever increasing importance of international communications, such proprietary schemes have become impractical.

SONET and SDH Standards

Two primary fiber-optic standards have emerged: SONET (Synchronous Optical Network), developed first by Bellcore and adopted by the American National Standards Institute (ANSI),[1] and SDH (Synchronous Digital Hierarchy), developed by the International Telecommunications Union (ITU).[2] SONET is primarily a North American standard and SDH is used in most of the rest of the world. An important objective of these standards is to ensure compatibility between equipment manufactured by different vendors. To achieve this the standards address such varied requirements as the physical properties of the optical signal and the transmission protocols and coding formats employed. It is a tribute to the cooperation between ANSI and ITU that in all practical respects, SONET and SDH are virtually identical.

Transmission System Design

A simple fiber-optic transmission system (Fig. 1) consists of a transmitter, a fiber-optic cable, and a receiver. The transmitter usually employs a digitally modulated laser diode operating at a wavelength of either 1300 or 1550 nm. The individual voice and data signals appear as low-rate electrical tributary signals that are time division multiplexed into a serial digital stream applied to the laser input. The laser's output is modulated in a simple nonreturn-to-zero (NRZ) format: it is turned on for the entire duration of a logical one pulse and turned off (or nearly off) for the entire duration of a logical zero pulse.

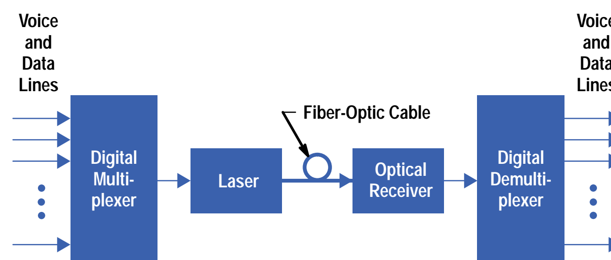


Fig. 1. *Fiber-optic transmission system.*

The fiber-optic cable is typically single-mode fiber with a 9-µm core diameter. The receiver consists of a p-i-n or avalanche photodiode serving as an optical-to-electrical (O/E) converter, with appropriate amplification, timing, and detection circuitry. A demultiplexer at the receiver output extracts the individual tributary signals. Real systems also often employ other network equipment such as digital cross-connects, optical amplifiers, and add-drop multiplexers.

The SONET and SDH standards place strict limits on the performance of each element in the system. At the physical level, transmitter specifications include output power, optical waveform shape, and extinction ratio. For the receiver, such parameters as sensitivity, electrical waveform shape, and output jitter are important. In the past, waveform characteristics of the system have been measured with a high-speed oscilloscope. Electrical signals could be measured directly, but to measure optical waveforms, a photodiode O/E converter was required in front of the oscilloscope.

The SONET and SDH standards impose new requirements on the optical waveform that are not easily measured with a conventional oscilloscope. One requirement, for example, is that the O/E converter be an optical "reference receiver" having a tightly controlled frequency response, specified as a fourth-order Bessel-Thomson filter whose 3-dB frequency is three quarters of the bit rate.[3] The resulting eye diagram (see subarticle "***Eye Diagrams and Sampling Oscilloscopes***") is compared against a specified mask that defines "keep out" regions for the waveform. The mask shape is designed to ensure that the quality of the waveform is sufficient to achieve satisfactory transmission performance. To make this measurement an oscilloscope must include a calibrated optical reference receiver, a way to generate the SONET/SDH mask automatically, and a way to compare the eye diagram accurately to the mask.

The Digital Communications Analyzer

The HP 83480 digital communications analyzer (Fig. 2) is the first commercial product to combine a SONET calibrated reference receiver with an oscilloscope and communications measurement firmware in a single package. It is based on proven HP digital sampling technology first introduced in the HP 54120 sampling oscilloscope in 1987. The HP 83480 is designed to address several limitations of existing technology. These include lack of overall optical calibration, incompletely calibrated SONET/SDH reference receiver path, and lack of flexible firmware measurement algorithms.

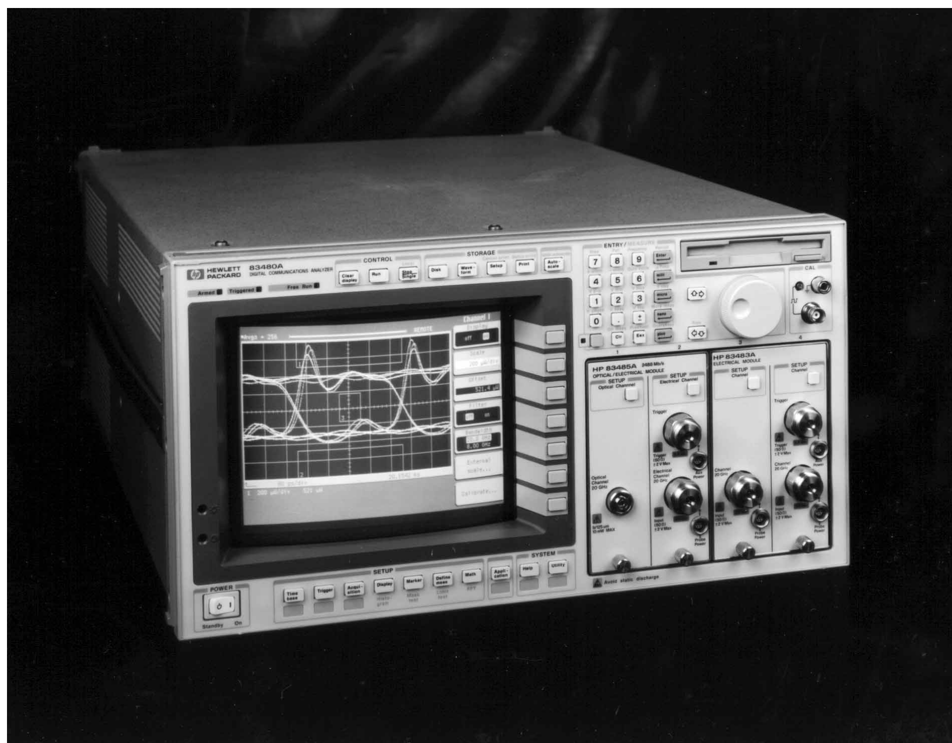


Fig. 2. *The HP 83480 digital communications analyzer combines a SONET calibrated reference receiver with an oscilloscope and communications measurement firmware.*

Overall Optical Calibration. To measure an optical signal before introduction of the HP 83480, a separate O/E converter had to be connected to the oscilloscope's vertical channel input. Rarely was the exact conversion gain of this combination known with any degree of accuracy, so the overall optical channel could not be considered a calibrated path. In the HP 83480, the optical channel is fully calibrated and displays a readout in optical watts. In addition, each optical channel includes a separate average power meter that approaches the accuracy of a dedicated optical power meter.

Calibrated SONET/SDH Reference Receiver Path. Current SONET/SDH standards place strict limits on the frequency response of the optical reference receiver but do not define requirements for the oscilloscope to which it is connected. The HP 83480 design team recognized early in the project that it made little sense to tightly control the reference receiver if the oscilloscope bandwidth and frequency response were insufficient to display the signal accurately. The team decided the answer was to apply the SONET/SDH frequency response tolerance requirement to the entire channel response, not just the O/E converter (see Fig. 3). Although this added to the design challenges and is more stringent than required by the standards, it vastly increases the user's confidence in the measurement.

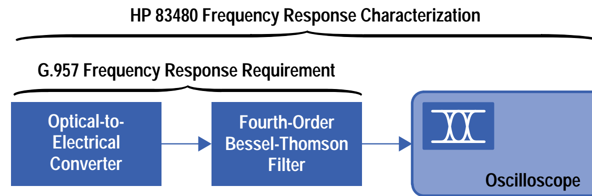


Fig. 3. *The SONET/SDH G.957 standard requires calibration of the optical receiver to a specified accuracy. In the HP 83480 digital communications analyzer, the entire instrument including the oscilloscope is calibrated.*

Flexible Firmware Measurement Algorithms. While other oscilloscopes have incorporated certain mask and parametric measurement capabilities, none has included all the measurements desired by users, and those provided have not always performed satisfactorily over the full range of expected waveforms. The HP 83480 design team placed a high priority on developing a complete set of features that correctly measure a wide range of waveform shapes.

Overall Design

The HP 83480 digital communications analyzer incorporates a modular design consisting of an oscilloscope mainframe together with various vertical channel plug-ins (Fig 4). The mainframe contains the analog-to-digital converter (ADC), time base, CPU, user interface, and display circuitry, while the vertical channel modules contain the O/E converters, samplers, IF amplifiers, and power monitoring circuitry. The various plug-ins cover different bandwidths, wavelengths, sensitivities, and fiber-optic media.

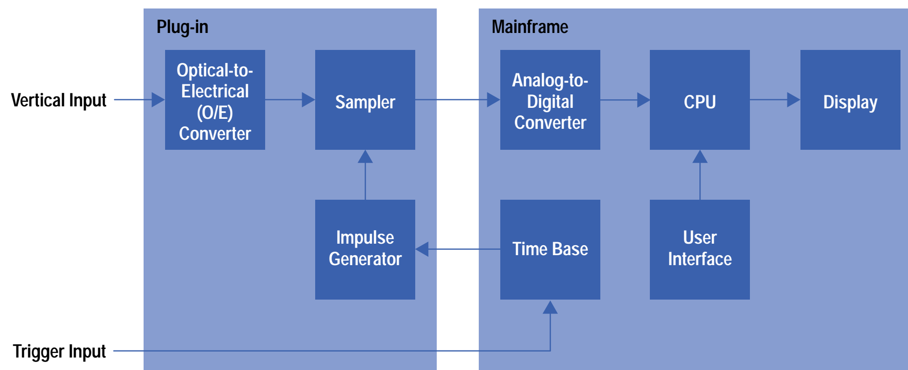


Fig. 4. *HP 83480 system block diagram.*

The mainframe/module approach offers maximum flexibility to the user. During initial market research, customers frequently expressed the desire to preserve their investment by being able to upgrade their oscilloscopes easily in the future.

The mainframe hardware is based on the modular HP 54720 oscilloscope platform introduced in 1992.[4] Major differences in the HP 83480 include the acquisition system, the internal firmware, and the front-panel design. The use of the HP 54720 mainframe provided two major opportunities for leverage of existing designs. First of all, it provided proven modular mainframe components complete with power supply, display, computer, and high-speed graphics. Secondly, it offered a software system that was designed with the ability to substitute different acquisition architectures with a minimum of engineering resources.

Acquisition System

The HP 83480 acquisition system has a repetitive sequential sampling architecture conceptually similar to that of HP's earlier high-speed sampling oscilloscope, the HP 54120. (Unlike the newer HP 54720, the HP 54120 did not use a modular design.) This approach provides the extremely high bandwidth necessary to display lightwave signals to beyond 10-Gbit/s rates. The acquisition system, split between the mainframe and plug-in, consists of a time base and sequential delay generator, a high-frequency trigger, microwave samplers, track-and-hold circuits, and analog-to-digital converters (Fig. 5). The plug-ins

for the HP 83480 are specific to this sampling architecture. They cannot be used in the HP 54720, nor can HP 54720 plug-ins be used in the HP 83480.

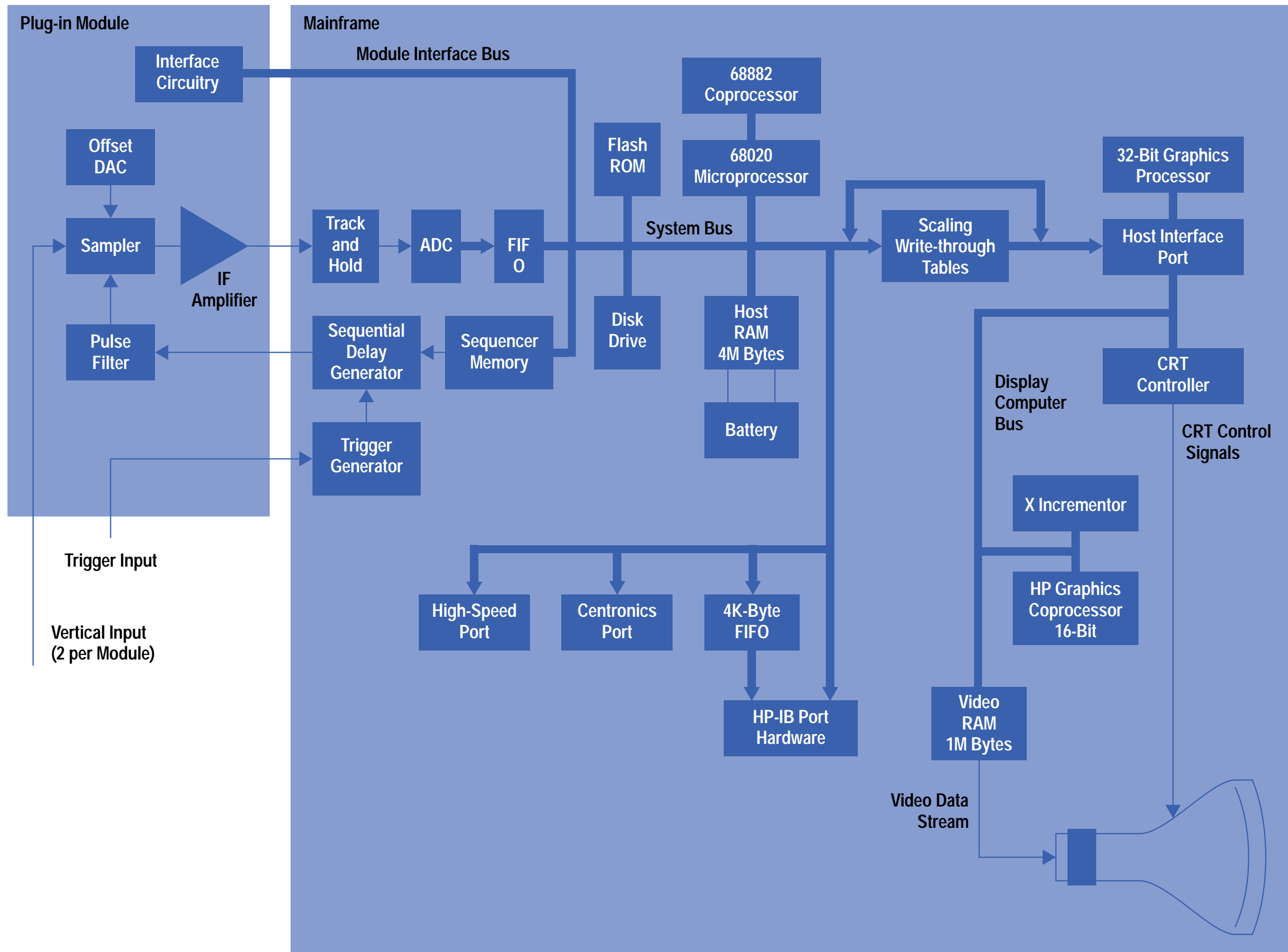


Fig. 5. *HP 83480 hardware block diagram.*

An acquisition cycle is defined as the sequence of events that must occur to acquire a single data point per enabled channel. Normally, many acquisition cycles are required to display a complete waveform. While the acquisition system in the earlier HP 54120 required the microprocessor to oversee each acquisition cycle, the HP 83480 is capable of running many acquisition cycles independently of the CPU. In addition, improvements in the design of the time base allow it to settle in less than 25 μs, compared to 100 μs for the HP 54120. These improvements result in a greater than tenfold throughput improvement over the HP 54120.

The cycle starts after the CPU has programmed a series of delay values into the sequencer memory based on the time-per-division setting of the oscilloscope. These delay values determine the time interval between the arrival of a trigger and the taking of a sample. Once the values have been programmed, the acquisition system continues to take samples at the programmed delay values until stopped or paused by the CPU.

The sampler in the plug-in measures the difference between the input signal's instantaneous amplitude at the sampling instant and a dc offset value programmed by a digital-to-analog converter DAC). This difference is amplified and converted to a bipolar pulse by charge amplifiers and IF filters located in the plug-in. This bipolar pulse signal is fed to the track-and-hold circuits in the mainframe. The charge amplifiers and IF filters in these circuits have been improved over those in the HP 54120 to achieve better than a factor of 2 improvement in noise floor.

The track-and-hold circuits follow the amplitude of the bipolar pulse. When the pulse reaches peak amplitude, the track-and-hold circuit holds the peak analog value. The ADCs convert each track-and-hold circuit's level to a 12-bit digital word. These 12-bit words are then put into a FIFO memory which is read and processed by the CPU.

Trigger and Time Base Systems

The oscilloscope's trigger system and time base play key roles in overall performance. To view the transition of a waveform, the oscilloscope must be syncronized to that transition. This syncronization must be very tight to achieve low jitter and high bandwidth. In the sequential sampling mode, the oscilloscope must trigger 500 times to capture and display 500 points. If the trigger has any jitter or uncertainty when it fires, this will be represented by a smeared waveform on screen. The need for a jitter-free trigger becomes all the more apparent given the fact that the HP 83480's fastest sweep time is 10 picoseconds per division.

The trigger system was leveraged intact from the HP 54120. The heart of the trigger mechanism is a threshold comparator—a custom circuit employing high-frequency hybrid technology. This circuit generates accurate, predictable, and programmable trigger performance that does not require delicate adjustment to yield a stable and jitter-free trigger point. The trigger detects the occurrence of a transition on the selected external trigger input. When the programmable threshold level is crossed, an acquisition cycle is started and the time base goes into action. Each plug-in module can have one trigger input, so up to two separate trigger inputs are available.

Firmware

The firmware system in the HP 83480 is highly leveraged from the HP 54720. Like any oscilloscope, the HP 83480 observes input signals, and in response to a user's request it produces data or screen images that describe these signals. These tasks are broken into five successive actions: signal conditioning, acquisition, data analysis, user interface, and display (Fig. 6).

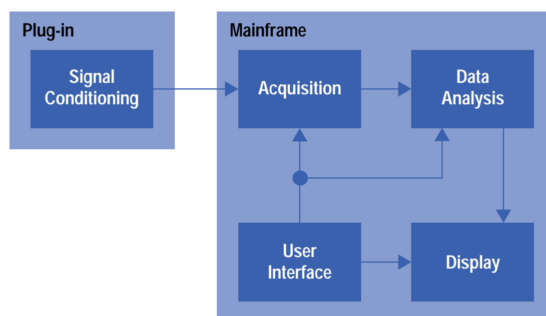


Fig. 6. *HP 83480 firmware block diagram.*

Signal conditioning in the HP 83480 includes the optical-to-electrical conversion in the optical channels and any signal amplification or attenuation that is built into the plug-ins. The sequential sampling acquisition system in the HP 83480 is completely different from the real-time sampling system in the HP 54720, so different drivers are required to control this acquisition system. After the acquisition drivers, the firmware system in the HP 83480 is identical to that of the HP 54720 except for the addition of the specialized measurements added for the communications industry. Measurements that were added for the HP 83480 include time and voltage histograms, more elaborate mask testing, and the ability to measure pulse parameters on multivalued waveforms such as eye diagrams. For details see the article on page 1. One testament to the flexibility of the software system is that histograms, while written for the HP 83480, were incorporated into a maintenance release of the HP 54720 over a year before the HP 83480 was introduced.

Plug-in Module Design

The plug-in modules house the O/E converters, samplers, and vertical channel signal conditioning circuitry. They also include the cabling necessary to interface the trigger hybrid in the mainframe to the instrument front panel. An early design decision the team faced was what sizes of plug-ins to adopt. The mainframe was leveraged from the HP 54720 oscilloscope, which had four input channels and accepted up to four single-slot plug-ins, two double-slot plug-ins, or one four-slot plug-in. Its plug-ins, however, were much simpler than those envisioned for the HP 83480. A single-slot module could not contain all the sampling circuitry, microwave hardware, and digital control circuitry necessary in even the simplest HP 83480 plug-in. The team decided that while the HP 83480 should still have four input channels, acceptable plug-ins would be restricted to two-slot and four-slot modules (no four-slot modules have been introduced to date).

Plug-ins come in two styles: optical and electrical. Optical plug-ins include one optical input channel and one electrical input channel, while electrical plug-ins consist of one or two electrical input channels. Optical measurement bandwidths range from 2 GHz to 30 GHz and electrical measurement bandwidths range from 12.4 GHz to 50 GHz. Plug-ins are also available that provide electrical TDR measurements (see ***Article 3***).

Three optical plug-in modules have been introduced: the HP 83485A, HP 83485B, and HP 83481A. Each optical plug-in module is optimized for particular SONET/SDH transmission rates from 155.52 Mbits/s to 9.95328 Gbits/s. A simplified block diagram of an optical channel is shown in Fig. 7. Each optical channel consists of an optical-to-electrical converter followed by a microwave transfer switch which routes the electrical signal either directly to the microwave sampler or first through

an electrical low-pass filter. The overall frequency response of the filtered path consisting of the O/E converter, transfer switch, filter, sampler, various semirigid cables, and the associated mismatch ripple meets the SONET/SDH requirements for a reference receiver. This is not a trivial objective; it requires careful design of the filter frequency response and low VSWR terminations in the associated circuits. In addition, each optical channel includes a built-in average optical power monitor. (For more information on the optical plug-in module design see ***Article 2***)

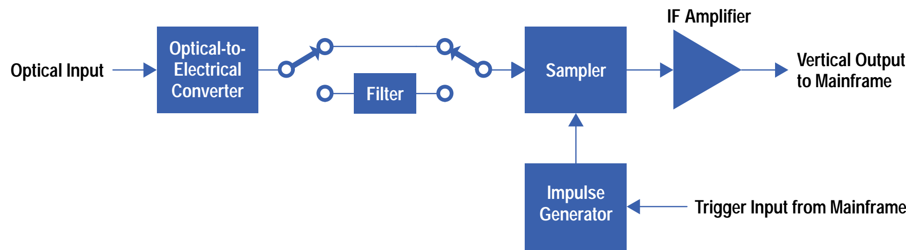


Fig. 7. *Optical plug-in module block diagram.*

Manufacturing

The manufacturing processes for the mainframe and plug-in modules are heavily leveraged from existing processes for the HP 54720 and HP 54120 oscilloscopes. The most significant manufacturing development challenge was the test processes for the optical plug-ins. Frequency response calibration was a particular challenge because of the extremely tight tolerance imposed on the reference receiver by the SONET/SDH standards. HP engineers worked closely with the U.S. National Institute for Standards and Technology (NIST) to implement a test system accurate enough to meet the intent of these standards (see ***Article 4***).

Acknowledgments

A task of this magnitude would have been impossible without the contributions of a great many people. Charlie Schaffer, Jim Egbert, Bryan Lord, and Greg LeCheminant provided valuable marketing inputs throughout the project. Rin Park and Jerry Townsend were responsible for the test system development. Rick Martinez, Kathy Albin, Denise O'Connor, Harry Schiff, and their manufacturing teams provided manufacturing support. Pamela Pitcher and Marj St. Clair were the project planners. Bob Bray, John Scharrer, and Dan Oldfield provided high-level R&D guidance for the project.

References

1. *Digital Hierarchy—Optical Interface Rates and Formats Specifications (SONET)*, ANSI T1.105-1991, American National Standards Institute, 1992.
2. *Digital Line Systems Based on the Synchronous Digital Hierarchy for Use on Optical Fibre Cables*, ITU G.958, International Telecommunications Union, 1990.
3. *Optical Interfaces for Equipments and Systems Relating to the Synchronous Digital Hierarchy*, ITU G.957, International Telecommunications Union, 1990.
4. J.A. Scharrer, "An 8-Gigasample-per-Second Modular Digitizing Oscilloscope System," *Hewlett-Packard Journal*, Vol. 44, no. 5, October 1993, pp. 6-10.

Eye Diagrams and Sampling Oscilloscopes

Most people are familiar with an oscilloscope display of repetitive waveforms such as sine, square, or triangle waves. These are known as *single-value* displays because each point in the time axis has only a single voltage value associated with it.

When analyzing a digital telecommunication waveform, single-value displays are not very useful. Real communications signals are not repetitive, but consist of random or pseudorandom patterns of ones and zeros. A single-value display can only show a few of the many different possible one-zero combinations. Pattern dependent problems such as slow rise time or excessive overshoot will be overlooked if they don't occur in the small segment of the waveform appearing on the display. For example, the amount of overshoot on the zero-to-one transition of a SONET laser transmitter depends on the exact pattern preceding it.

An eye diagram (Fig. 1) overcomes the limitations of a single-value display by overlapping all of the possible one-zero combinations on the oscilloscope screen. Eye diagrams are *multivalued* displays because each point in the time axis has multiple voltage values associated with it.

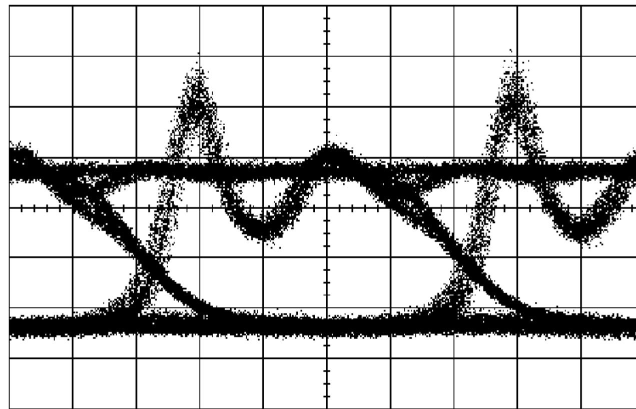


Fig. 1. *Eye diagram of a laser directly modulated at 2.48832 Gbits/s.*

An eye diagram is generated on an oscilloscope using the setup of Fig. 2. The data signal is applied to the oscilloscope's vertical input and a separate trigger signal is applied to its trigger input. Ideally, the trigger signal is a repetitive waveform at the clock rate of the data, although the data signal itself can also be used as a trigger signal in noncritical applications.

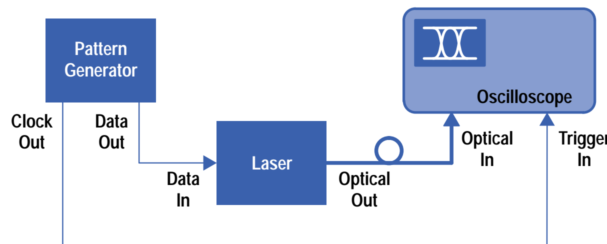


Fig. 2. *Block diagram for eye diagram testing.*

The oscilloscope triggers on the first clock transition after its trigger circuit is armed. Upon triggering, it captures whatever data waveform is present at the vertical input and displays it on the screen. The oscilloscope is set for infinite persistence so that subsequent waveforms will continue to add to the display.

For a short period of time after triggering, the oscilloscope is unable to retrigger while the circuitry resets. At the conclusion of this trigger dead time, the oscilloscope triggers on the next clock transition. The data pattern at this instant will probably be different from the previous one, so the display will now be a combination of the two patterns. This process continues so that eventually, after many trigger events, all the different one-zero combinations will overlap on the screen.

To analyze waveform properties accurately the oscilloscope must have a bandwidth at least three times the data rate, and preferably much more. At high data rates, this leads to an additional complicating factor. Wide-bandwidth single-shot oscilloscopes capable of capturing an entire waveform at once simply aren't economically available, so sampling oscilloscopes must be used instead. Sampling oscilloscopes make use of the concept of *equivalent time* to achieve effective bandwidths up to 50 GHz.

Instead of capturing an entire waveform on each trigger, the sampling oscilloscope measures only the instantaneous amplitude of the wave-form at the sampling instant (Fig. 3). On the first trigger event the oscilloscope samples the waveform immediately and displays the value at the very left edge of the screen. On the next trigger event, it delays slightly before sampling the data. The amount of this delay depends on the number of horizontal data points on the screen and on the selected sweep speed. It is determined by the relation:

$$D = \frac{1}{n-1} \times T,$$

where D is the delay time between successive points, n is the number of points on the screen, and T is the full-screen sweep time. For example, if the full screen sweep time is 10 nanoseconds and the display has 451 horizontal points, then the delay is (1/450) $\times$ 10 ns, or 22.2 ps. The amplitude sampled at this instant is displayed one point to the right of the original sample. On each subsequent trigger event the oscilloscope adds an ever-increasing delay before sampling, so that the trace builds up from left to right across the screen.

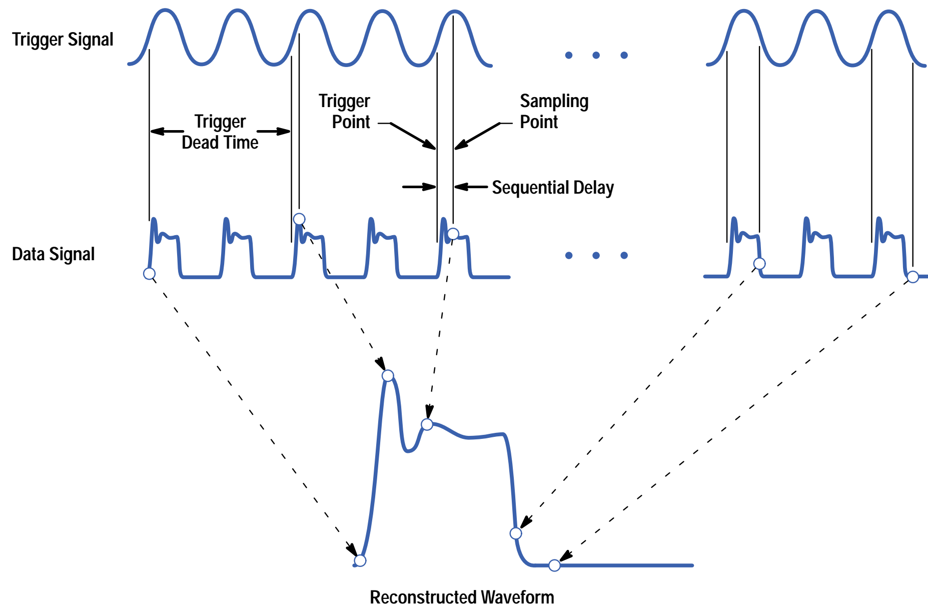


Fig. 3. *Concept of sequential sampling.*

The trade-off in using a sampling oscilloscope is a loss of information about the exact waveform characteristics. When sampling a repetitive waveform such as a sine wave, this doesn't usually pose a problem; the screen display shows a sine wave that is a sampled representation of the original waveform. When sampling a random data pattern, however, the eye diagram appears as a series of disconnected dots. All information about the exact nature of the individual waveforms is lost, so if the eye diagram shows excessive overshoot or slow rise time, the exact data pattern causing the problem can't be identified. This is why Hewlett-Packard incorporated the HP Eyeline display mode into the HP 83480 (see ***Article 2***). First developed for the HP 71501 eye diagram analyzer,[1] the eyeline display allows the individual data patterns making up the eye diagram to be seen.

Reference

1. C.M. Miller, "High-Speed Digital Transmitter Characterization Using Eye Diagram Analysis," *Hewlett-Packard Journal*, Vol. 45, no. 4, August 1994, pp. 29-37.

Firmware Measurement Algorithms for the HP 83480 Digital Communications Analyzer

Parametric measurements measure waveform properties such as rise time, fall time, overshoot, period, and amplitude on either a pulse waveform or an eye diagram. Mask measurements compare the shape of the waveform to a predefined mask. Eye parameter measurements measure properties that are unique to eye diagrams, such as eye height, eye width, jitter, crossing height, and extinction ratio.

by Michael G. Hart, Christopher P. Duff, and Stephen W. Hinch

An important part of the HP 83480 digital communications analyzer is its extensive set of built-in measurements designed especially for telecommunications applications. Internal firmware algorithms permit the user to quickly measure waveform properties that would be tedious and error-prone to measure manually. The algorithms also eliminate the subjectiveness inherent in a manual measurement. Of course, any built-in measurement is only as accurate as its firmware algorithm. A major objective for this instrument's design team was to develop robust algorithms capable of producing reliable results in virtually all situations.

The HP 83480's built-in measurements fall into three general categories:

- Parametric measurements measure waveform properties such as rise time, fall time, overshoot, period, and amplitude on either a pulse waveform or an eye diagram.
- Mask measurements compare the shape of the waveform to a predefined mask. If any part of the waveform intrudes into the mask, it is counted as a failure.
- Eye parameter measurements measure properties that are unique to eye diagrams, such as eye height, eye width, jitter, crossing height, and extinction ratio.

Parametric measurements can be made in either a real-time mode or a statistical mode. In the real-time mode the measurement is performed immediately on each acquired waveform. In the statistical mode, histograms are used to perform the measurement on a database representing multiple waveform acquisitions. Mask and eye parameter measurements are inherently statistical parameters and so are only measured using histograms.

Design Leverage

As much as possible, the firmware design of the HP 83480 was leveraged from existing code for the HP 54720 digitizing oscilloscope.[1] The HP 54720 incorporated an extensive set of very accurate pulse parameter measurements together with basic mask test capabilities. The parametric measurements did not function on eye diagrams, however, so this capability needed to be added. A number of additional eye parameter measurements were developed, and the existing mask test capability was extensively upgraded to include such features as a full set of standard telecom masks, built-in mask margins, and fixed voltage masks.

An important feature of these measurements is that they operate in real time, automatically updating on every acquisition of the waveform. Before the HP 54720, HP high-speed oscilloscopes made measurements in a single-shot mode. When the user pushed the button to determine rise time, for instance, the oscilloscope stopped acquiring data and made the measurement on a frozen waveform display. The real-time measurement mode was considered much more convenient, but it increased the complexity of the algorithm design. Not only did the algorithms need to be accurate, they also had to be fast enough not to slow down instrument operation perceptibly.

Parametric Measurements

From the outset, the design team wanted the parametric measurements to operate correctly on both pulse waveforms and eye diagrams. This proved to be a challenge in several ways. First, while the IEEE has developed standard definitions for such parameters as rise time, fall time, and overshoot for simple pulse waveforms, no similar standards exist for eye diagrams. To develop the appropriate definitions, team members spent many hours in discussions with potential users to understand their needs. Alternative algorithms were extensively tested to determine those that best met expectations.

A second challenge arose because of the different signal coding formats used on the various electrical tributary rates. For example, certain low-bit-rate transmission formats use a bipolar coding scheme known as alternate mark inversion (AMI). In this format, each successive logical one bit is represented by a pulse of polarity opposite to that of the previous pulse. In the case of a DS-1 signal at 1.544 Mbits/s, for instance, the amplitude of the first logic 1 pulse is specified as +3.0V while the amplitude of the next is −3.0V. Subsequent logic one pulses alternate between these two levels. The algorithm for measuring rise time on an AMI pulse waveform had to be versatile enough to produce the correct result whether the onscreen display was set to show only a single positive-going pulse or both positive-going and negative-going pulses (Fig. 1).

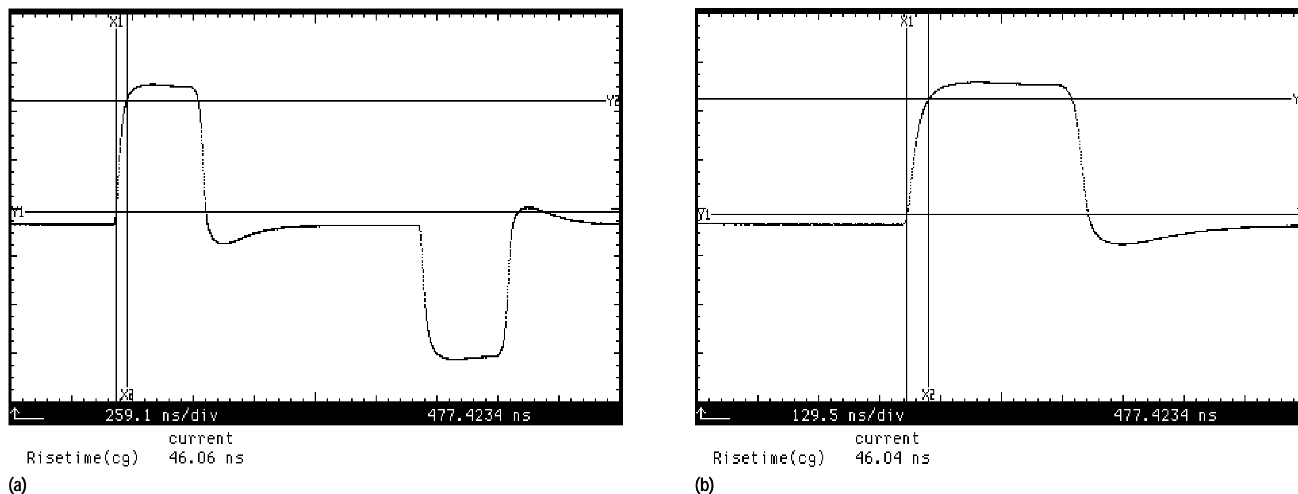


Fig. 1. *The HP 83480 measurement algorithms are designed to operate correctly on alternate mark inversion (AMI) waveforms regardless of whether the display shows both positive and negative pulses (a) or only a positive pulse (b).*

Histograms and Color-Graded Displays

Since eye diagrams are essentially a statistical representation of all possible one-zero waveform combinations, it makes sense to measure their waveform properties statistically. To do this, powerful database and histogram capabilities were developed, and these became the foundation for the new measurements.

The database in the HP 83480 corresponds to the size of the display—451 columns by 256 rows. Every waveform regardless of record length or number of vertical bits is mapped onto this database structure. Behind each database location, or pixel, is a 16-bit counter. Each time a waveform strikes a particular pixel, it is recorded as a “hit” at that location and the counter for that pixel is incremented by one. Up to 65535 hits can be counted at each pixel (Fig. 2).

To give the user an easy visual indication of the distribution of data in the database, the instrument uses a color-graded display mode. This mode maps the database into seven display colors, with each color representing a specific range of database hits. The algorithm that determines the range of hits corresponding to each color was carefully chosen to provide a meaningful display regardless of the maximum number of hits in the database. This display provides a very powerful three-dimensional view of the eye diagram (Fig. 3). It clearly shows the most prevalent paths of the signal as well as spurious noise and jitter.

Histogram measurements are derived directly from the database. A histogram window is set up to select a range of rows and columns to be included in the histogram as shown in Fig. 4. A vertical histogram is computed by windowing across a slice of time and plotting the frequency of occurrence for each database row. A horizontal histogram is computed by windowing across a range of voltages and plotting the frequency of occurrence for each database column.

The HP 83480 can display the histogram with either a linear or a logarithmic scale. With the linear scale, the histogram is plotted in hits per division. With the logarithmic scale, the histogram is plotted in dB per division. For each pixel, the displayed value is found from:

$$dB = 20\log_{10}\frac{N}{P},$$

where dB is the displayed value for the pixel in dB, N is the number of hits in the pixel, and P is the peak number of hits within the histogram window.

The linear scale is useful for looking at the peak and distribution of the histogram. The logarithmic scale is useful for observing the tails of the histogram.

In addition to the pictorial representation of the histogram, the HP 83480 computes a series of parametric measurements on the histogram data. The width of the histogram, the average value, the standard deviation, and the mean value plus and

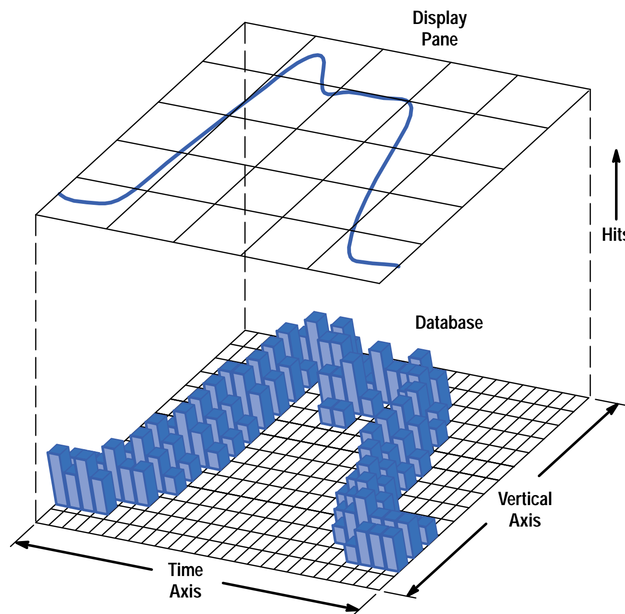


Fig. 2. *Waveforms are mapped onto a database for statistical analysis. The database consists of 115,456 pixels arranged in 451 columns by 256 rows. Each pixel can record up to 65,535 waveform hits.*

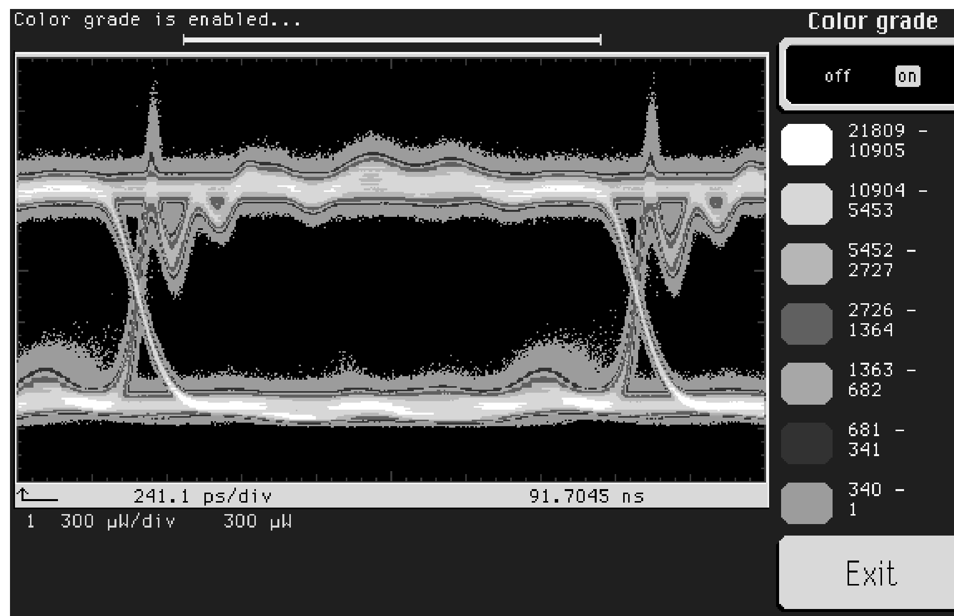


Fig. 3. *The color-graded display mode provides a visual representation of the waveform distribution.*

minus one, two, and three standard deviations are all useful parameters for understanding the noise and jitter of an eye diagram.

Eye Parameter Measurements

Measurements of eye parameters are made by automatically constructing histograms over selected regions of the database while in the color-graded display mode. Although these measurements could technically be made without activating the color-graded display mode, the design team decided that this was the best way to ensure that what the user sees on screen is correlated with the data in the database.

The measurement algorithms construct vertical and horizontal histograms to search for reference features in eye diagrams. Most eye parameters are referenced to four fundamental properties of the eye diagram:

- ***Top level***, V_{top}, is the mean logic one level.

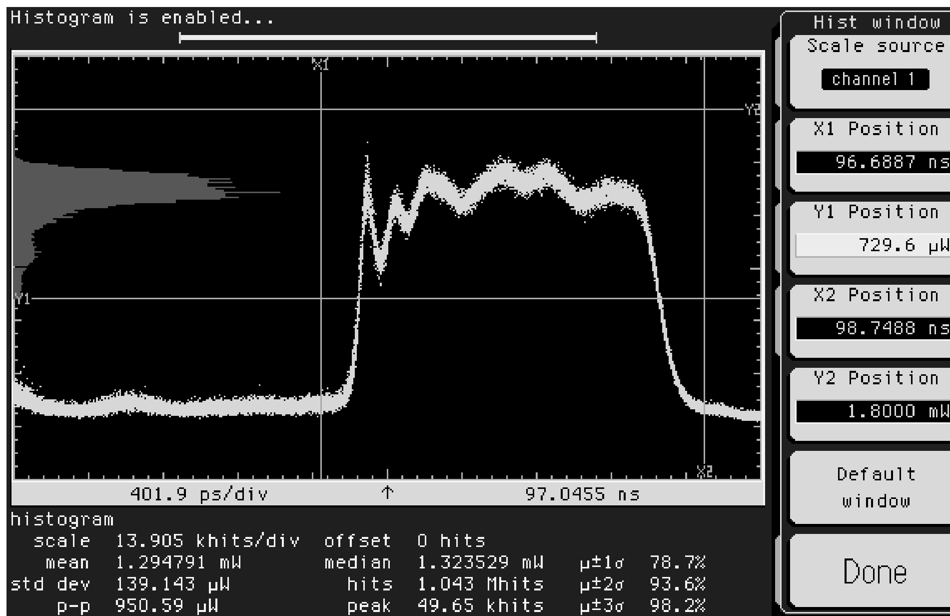


Fig. 4. *Histograms are used to analyze waveform statistics. In this example a vertical histogram has been constructed in a rectangular window inside the X1-X2, Y1-Y2 markers to determine the mean logic one level of the pulse.*

- ***Base level***, V_{base}, is the mean logic zero level.
- ***Eye crossing*** refers to the start of the bit period. By convention, this is defined as the point in time where the rising and falling edges of the eye intersect. Crossing time is defined as $t_{crossing}$, and crossing amplitude is $V_{crossing}$.
- ***Threshold crossings*** are defined as the times at which the signal crosses predefined threshold levels while making the transition between logic levels. Typical threshold levels used on an optical eye diagram are 20%, 50% and 80% of amplitude.

The algorithms for finding these values have been designed to accommodate a wide range of eye diagram shapes. V_{top} and V_{base}, for example, are determined by first finding the peaks of a vertical histogram constructed across the database as shown in Fig. 5. V_{top} is the mean value of the upper peak and V_{base} is the mean value of the lower peak. The standard deviations of these values are the rms noise levels on the logic one and logic zero levels.

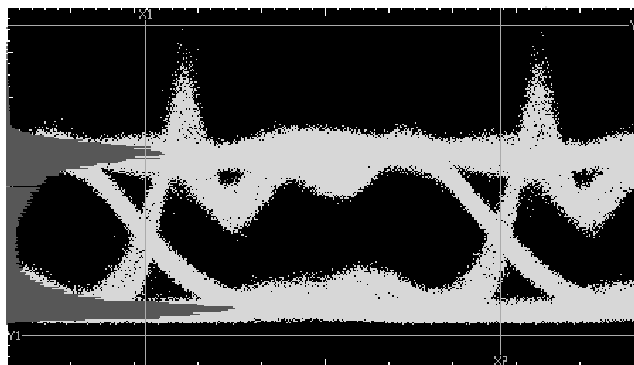


Fig. 5. *Vertical histograms are used to find the mean logic one and logic zero levels. In this example, the histogram extends the full width of the eye.*

Eye crossing times are located using an iterative algorithm on a horizontal histogram. The initial histogram is constructed on a window that just excludes the V_{top} and V_{base} data as shown in Fig. 6. Peaks in this histogram indicate the approximate locations of eye crossings. To find the crossings more precisely, a subsequent histogram is constructed in a narrow window about the approximate crossing point amplitude. The mean value is the crossing time, $t_{crossing}$, and the standard deviation, $\sigma_{crossing}$, is the rms jitter. Crossing amplitude is found by taking a vertical histogram on the same window.

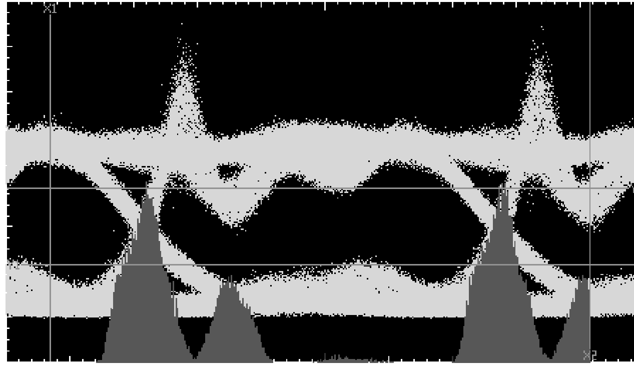


Fig. 6. *Horizontal histograms are used to find the eye crossing times. The window is set to exclude the logic one and zero level data.*

Rise time, fall time, overshoot, and duty cycle distortion measurements depend on finding rising and falling edges on the eye diagram. The HP 83480 finds threshold crossings by forming horizontal histograms using narrow windows centered at each threshold level.

More complex eye parameters are calculated from a set of constituent measurements, which in turn may have their own dependencies. An ordered list of measurements is performed by tracing dependencies using a technique first developed for the HP 54720.[2] The equations for eye parameter measurements are as follows:

- Eye Height:

$$(V_{top} - 3\sigma_{top}) - (V_{base} + 3\sigma_{base})$$

- Eye Width:

$$(t_{crossing2} - 3\sigma_{crossing}) - (t_{crossing1} + 3\sigma_{crossing})$$

- Q Factor:

$$\frac{(V_{top} - V_{base})}{(\sigma_{top} + \sigma_{base})}$$

- Jitter (rms):

$$\sigma_{crossing}$$

- Jitter (peak-to-peak):

$$6\sigma_{crossing}$$

- Crossing Level:

$$\frac{(V_{crossing} - V_{base})}{(V_{top} - V_{base})} \times 100\%$$

- Duty Cycle Distortion:

$$\frac{\left|t_{rising50\%} - t_{falling50\%}\right|}{(t_{crossing2} - t_{crossing1})} \times 100\%$$

- Overshoot:

$$\frac{(V_{top} + V_{95})}{(V_{top} - V_{base})} \times 100\%.$$

In the above equations, σ_{top} is the standard deviation of V_{top}, σ_{base} is the standard deviation of V_{base}, $\sigma_{crossing}$ is the standard deviation of $t_{crossing}$, $t_{rising50\%}$ is the time at which the rising edge reaches the 50% point between V_{top} and V_{base}, $t_{falling50\%}$ is the time at which the falling edge reaches the 50% point between V_{top} and V_{base}, and V_{95} is the amplitude of the 95th percentile of the data in a vertical histogram whose lower bound is V_{top} and whose upper bound is the top of the screen.

The user can control several measurement factors: the eye window over which V_{top} and V_{base} are measured, the signal type of the eye pattern, and the percentile for voltage thresholds. The industry has not yet adopted a consensus for how to measure V_{top} and V_{base}. Some users want the histogram window to extend across the full bit interval while others prefer only a narrow window about the center of the eye where the logic levels have settled to steady state (Fig. 7). With the HP 83480

the user can window on any time interval from 0 to 100% of the eye. The default window is ±10% around the center of the eye.

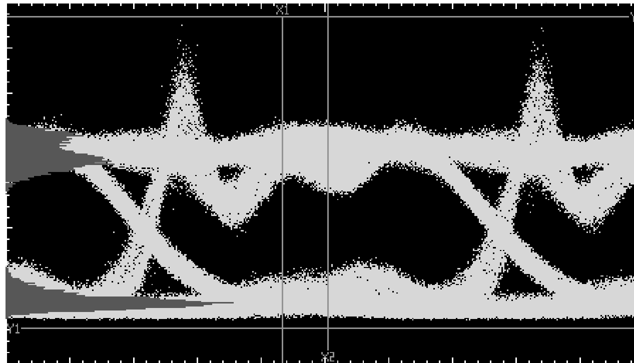


Fig. 7. *Mean one and zero levels on the same data as in Fig. 5 but found by windowing on only a narrow region about the center of the eye. Notice the difference in the logic one histogram compared to Fig. 5.*

The HP 83480 can measure parameters on three different signal types. Optical signals use a simple binary code called nonreturn-to-zero, or NRZ. With this code, the light is turned on for the full duration of a logic one pulse and turned off (or nearly off) for the full duration of a logic zero pulse. Electrical telecommunication signals use two different coding formats. Alternate mark inversion, or AMI, is a three-level format in which a logic zero is represented by a zero-volt signal and logic one pulses alternate between positive-going and negative-going voltages. Coded mark inversion, or CMI, is a binary signal in which a logic 0 includes a transition at the center of the bit period while the logic 1 does not. The HP 83480 can measure waveform parameters on all three signal types as well as unencoded periodic waveforms. Eye parameter measurements such as eye height, eye width, and extinction ratio are performed only on NRZ signals.

The threshold values used for characterizing rising and falling edges can be defined by the user. Rise times of electrical signals are traditionally measured from the 10% amplitude point to the 90% amplitude point. On optical signals, these points are often obscured in the noise, so 20%-to-80% thresholds are frequently used. The HP 83480 can use these or any other user-specified threshold values. Thresholds can also be defined as explicit voltage or power levels.

Extinction Ratio

Extinction ratio is a critical parameter for laser transmitters because it is a measure of the signal-to-noise level of the system. It is defined as the ratio of the logic one level amplitude (V_{top}) to the logic zero level amplitude (V_{base}). Although conceptually simple, both hardware and software considerations make the extinction ratio difficult to measure accurately.[3] The impact of hardware design is described in ***Article 3***. The most important software consideration is the removal of dc offset.

A principal source of dc offset comes from the optical-to-electrical (O/E) converter at the vertical channel input. Internal oscilloscope offsets can also play a role. While the HP 83480 has been designed to minimize these offsets, it is not possible to eliminate them completely. To correct for any residual offset, the HP 83480 requires an initial offset calibration. The user first removes the signal input from the optical channel, then pushes the Offset Cal softkey. The HP 83480 automatically measures and stores the offset, V_{offset}, by taking the mean of a large number of samples. Normally, once this calibration is done, it need not be repeated for the rest of the day unless the instrument's temperature changes by more than about five degrees. The HP 83480 allows extinction ratio results to be displayed in one of three formats:

- dB: $10\log[(V_{top} - V_{offset}) / (V_{base} - V_{offset})]$
- Ratio: $(V_{top} - V_{offset}) / (V_{base} - V_{offset})$
- Percent: $[(V_{base} - V_{offset}) / (V_{top} - V_{offset})] \times 100$.

Extinction Ratio Frequency Response Correction

Extinction ratio measurement accuracy can be heavily influenced by the hardware design of the vertical channel. One potentially serious source of error is the frequency response flatness of the channel. To measure extinction ratio accurately, the ac gain and the dc gain of the channel must be identical. A low-frequency gain increase of even 0.5 dB can lead to a large measurement error.

This is a very challenging design objective and is a major reason why the HP 83485A plug-in module (see ***Article 3***) employs a nonamplified optical channel. However, some users need to measure lasers operating at wavelengths or fiber diameters for which HP does not presently provide a solution. Other users need additional sensitivity to measure extremely low-level signals. In these cases an external O/E converter having a nonideal frequency response may have to be used.

Recent research has shown that it is possible to correct in software for frequency response errors in hardware.[4] The HP 83480 provides the ability to enter an extinction ratio frequency response correction factor. To determine this correction factor a signal with known extinction ratio is applied to the O/E converter's input and the measured extinction ratio is recorded. When both extinction ratios—known and measured—are expressed in percent, the frequency response error is a constant, independent of the actual extinction ratio. The value of this constant (in percent) can be entered into the instrument, which then automatically corrects the reading. Once entered, the correction factor is in effect regardless of whether the extinction ratio is displayed in percent, in dB, or as a ratio. With care, measurement accuracy of better than 1% is possible when using this technique.

While the use of frequency response correction considerably improves the measurement accuracy when using a nonideal O/E converter, it is not a universal solution. In general, the correction factor differs depending on the data rate, so a single number is only appropriate at the data rate for which it was determined. Depending on the frequency response characteristics of the external O/E converter, the correction factor can show slight sensitivity to varying data pattern characteristics even at a single data rate.

Mask Measurements

Mask tests are often used in production environments as an alternative to eye parameter analysis. By comparing an eye diagram against a predefined mask, the overall quality of the waveform can be assessed in one quick measurement. A mask consists of two parts, as shown in Fig. 8:

- A set of regions, or polygons, on the oscilloscope screen that define keep-out areas for the waveform. Waveforms that intrude into these polygons are counted as mask violations.
- Definitions of the time and amplitude scales for the mask. Many masks use an amplitude scale that is defined relative to the mean one and zero levels of the eye. Others require fixed voltage levels independent of measured signal levels.

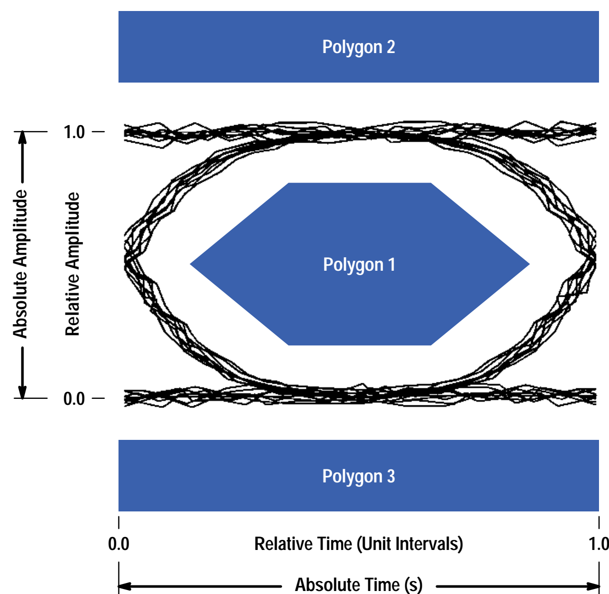


Fig. 8. *Concept of mask testing.*

The earliest masks were simply drawn on the oscilloscope screen with a grease pencil. Later oscilloscopes included rudimentary built-in mask drawing features. A limitation of these instruments was that the masks were drawn in screen coordinates that did not relate to the scale of the waveform. As the user adjusted the horizontal or vertical scales the mask remained fixed on the screen.

The mask measurement capabilities of the HP 83480 are far more powerful than in any previous instrument. The mask is referenced to true time and amplitude coordinates so that as the user changes the oscilloscope settings, the mask follows the waveform. It is also easy to rescale the mask for different data rates or amplitude levels.

There are two ways to create masks. A large number of standard telecommunications masks are built into the instrument (see **Table I**). These masks can be called onto the screen with the touch of a button. For nonstandard needs, the user can create custom masks.

Table I Standard Masks in the HP 83480 Digital Communications Analyzer	
Optical	Electrical
OC-1	STS-1 pulse
OC-3/STM-1	STS-1 eye
OC-12/STM-4	STS-3 pulse 0, 1
OC-24	STS-3 eye
OC-48/STM-16	DS-1
FC-133	DS-1C
FC-266	DS-2
FC-531	DS-3
FC-1063	PDH 2.048
FDDI	PDH 8.448
	PDH 34.4
	PDH 139.25

Custom Masks

Up to eight polygons can be created using the display as a drawing pad and the knob to control the vertical and horizontal coordinate of each polygon point. All polygon coordinates are normalized to the dimensions of the eye diagram. For example, (0,0) represents the zero level of the first crossing and (1,1) represents the one level of the second crossing. Special values are also used for the vertical axis minimum and maximum to represent the bottom and top of the screen regardless of the vertical scale. Absolute time and amplitude scales are defined separately from the relative coordinate system.

This coordinate system permits the entire mask definition to be compressed or expanded in either axis by assigning new amplitude or time values to the (0,0) and (1,1) coordinates. It also ensures that the mask tracks the signal when the time base or amplitude settings are modified.

Mask Margins

In a manufacturing environment it is often desirable to add a test line margin to industry-standard masks. At other times it is useful to reduce the size of the mask to determine by how much a waveform fails the test. The mask margin capability of the HP 83480 makes these tasks easy. This feature allows minimum and maximum margin limits to be defined as separate masks around the standard mask. The margin mask can be set as any percentage from −100% under to +100% over the standard mask (Fig. 9). Minimum and maximum mask margins are included in the definintions of the standard masks.

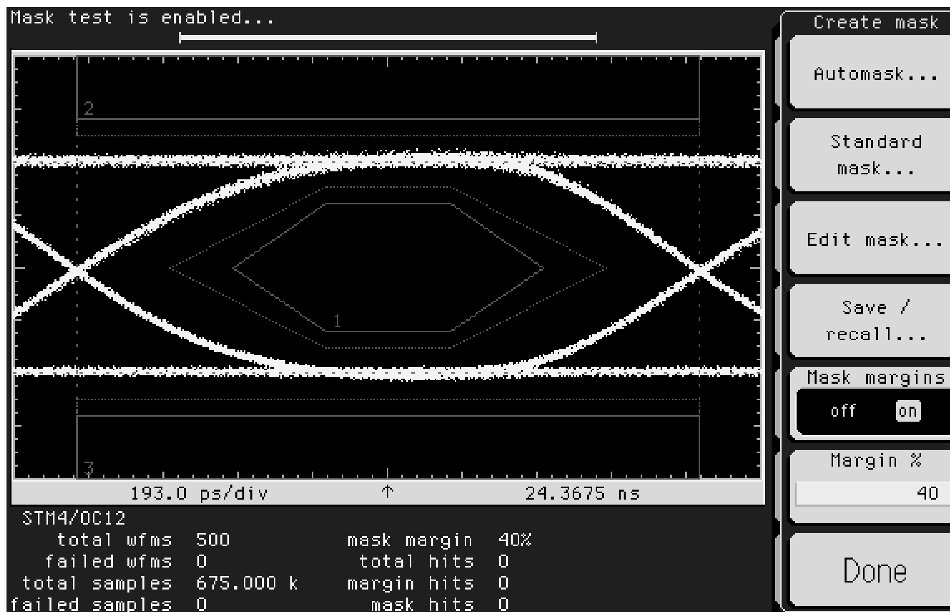


Fig. 9. *Mask margins are used to add guardbands for production testing. In this example a 40% margin has been added to an industry-standard OC-12 mask at 622.08 Mbits/s.*

Mask Alignment

Before a mask test is conducted, the mask must be properly aligned to the waveform. The HP 83480 provides two methods of automatic alignment. In the first, known as ***mask-to-waveform*** alignment, the mask is aligned directly to the displayed waveform and the instrument settings are left unchanged. When this mode is initiated the firmware measures reference points on the eye diagram and positions the mask to align it to the data.

The second alignment method is known as the ***fill display*** mode. In this method, the instrument scale settings are automatically adjusted to center one full eye diagram on the screen regardless of initial settings. Fill display makes the most efficient use of the data and so provides the most accurate test results. The mask-to-waveform mode, however, produces much faster results.

Most standard masks are defined with amplitudes relative to the signal amplitude. These masks automatically rescale to fit signals with different amplitudes. However, a few masks are defined using explicit voltage levels. For these fixed-voltage templates the instrument automatically aligns the time position of the mask but maintains a fixed vertical scale as defined in the standard.

Acknowledgements

A great many people in both Santa Rosa, California and Colorado Springs, Colorado made invaluable contributions to this project. Don St. Denis developed definitions for many of the mask test features that were eventually incorporated into the instrument. Dave Poppe and Caren Johnson wrote the initial mask test firmware. Jerry Kinsley did early work on mask margins. Chris Miller, Greg LeCheminant, Bernie Hovden, Mike Karin, and Charlie Schaffer provided many inputs on the functionality and usability of the firmware. Walter Coole did much of the development of the HP Eyeline software program, and Mark Woodward provided many valuable insights into the measurement of extinction ratio.

References

1. J.A. Scharrer, "An 8-Gigasample-per-Second Modular Digitizing Oscilloscope System," ***Hewlett-Packard Journal***, Vol. 44, no. 5, October 1993, pp. 6-10.
2. D.L. Johnson and C.J. Magnuson, "Architectural Design for a Modular Oscilloscope System," ***ibid***, pp. 51-58.
3. S. Hinch, M. Woodward, and C. Miller, "Accurate Measurement of Laser Extinction Ratio," ***1995 Lightwave Symposium Technical Papers***, Hewlett-Packard Company, 1995, pp. 1-12.
4. P.O. Andersson and K. Akermark, "Accurate Optical Extinction Ratio Measurements," ***Photonics Technology Letters***, Vol. 6, no. 11, November 1994, pp. 1356-1358.

HP Eyeline Display Mode

Sequential sampling oscilloscopes normally display eye diagrams as a series of disconnected points on the screen. While these points accurately represent all the combinations of digitized bit patterns, each point is obtained from a separate trigger event, so there is no way to determine exact characteristics of any specific bit combination (Fig. 1a).

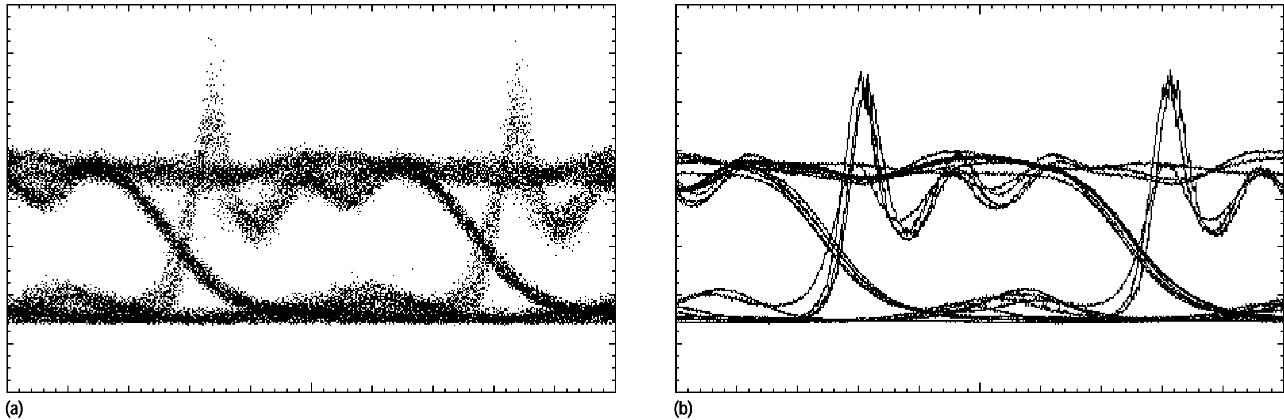


Fig. 1. *Sequential sampling oscilloscopes show eye diagrams as a series of disconnected dots (a). Using the HP Eyeline display mode (b), the individual bit patterns can be distinguished.*

When sampling live data there is no alternative, but when sampling repetitive waveforms such as pseudorandom binary sequence (PRBS) patterns it is often possible to show the individual bit sequences. This is done by synchronizing the oscilloscope trigger with the pattern repetition rate. In this case, the oscilloscope repeatedly triggers at the same point in the pattern, so the display is a sampled representation of that segment of the pattern.

This display mode, called pattern triggering, is commonly used in situations where the device under test can be stimulated with a digital pattern generator from a bit error rate tester (BERT). Pattern generators typically include a trigger output that can be set to produce a trigger pulse at the start of each pattern. The problem has been that until recently, it has not been possible to accurately display portions of the pattern far from the trigger point. While it is theoretically possible to show different parts of the pattern by increasing the oscilloscope delay, in reality this is impractical because of accuracy and jitter limitations in the oscilloscope time base.

The HP 83480 offers an optional *HP Eyeline display mode* (Fig. 1b), which overcomes this limitation. HP Eyeline display mode takes advantage of a new feature in the HP 71604B and HP 71612A pattern generators that allows the trigger point to be adjusted bit by bit within the pattern.

The HP Eyeline display mode runs as an application program that loads from the HP 83480's 3.5-in disk drive and allows the instrument to control the pattern generator over the HP-IB (IEEE 488, IEC 625). The equipment setup is shown in Fig. 2. (HP Eyeline display capability was first introduced in the HP 71501 Eye Diagram Analyzer using a different method to accomplish the same result.[1])

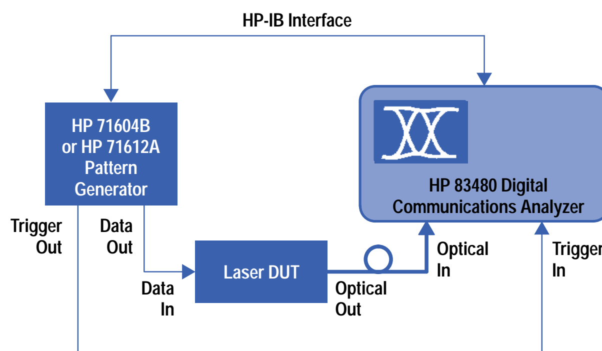


Fig. 2. *Equipment setup for HP Eyeline display mode.*

When the HP Eyeline program is running, the HP 83480 sets the pattern trigger location and samples one data point for each repetition of the pattern. The pattern generator transmits the entire data pattern between successive triggers. After an entire waveform record is taken (typically 500 to 4000 points, depending on the record length setting), the HP 83480 programs the pattern generator to delay the

trigger point by one bit and repeats the process. Eventually, the trigger point moves through the entire pattern, and the eye diagram shows all possible bit combinations.

One advantage of the HP Eyeline display mode is that it allows the use of signal averaging to reduce the effects of noise. (Averaging is not possible on an ordinary sampled eye because the result is the average between the two logic levels, causing the eye to collapse.) Signals too small to be seen without averaging can be readily identified using the HP Eyeline mode, as shown in Fig. 3. Another advantage is that it can aid troubleshooting by showing the bit sequence leading up to a mask test violation, as seen in Fig. 4. This can be used to identify the cause of pattern dependent errors.

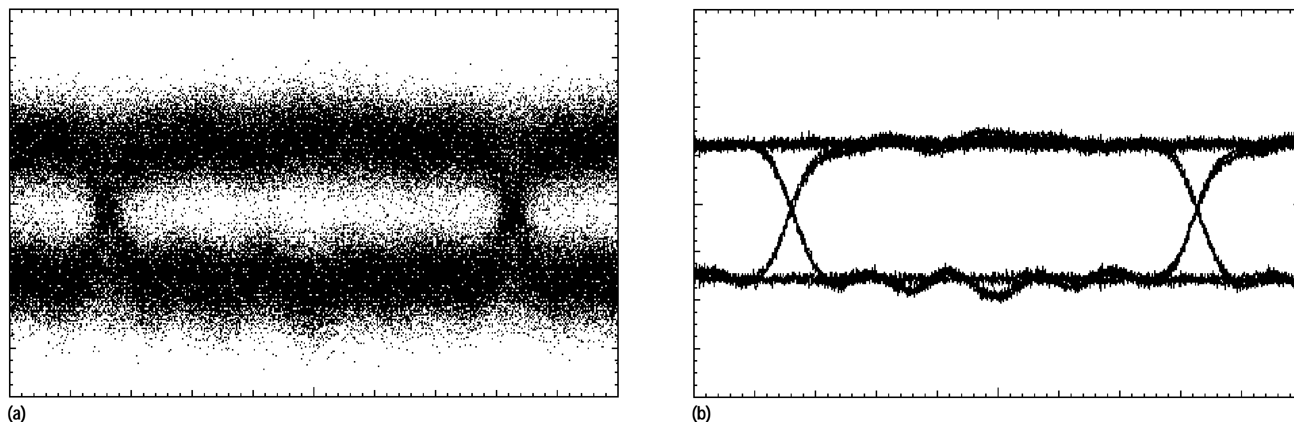


Fig. 3. *HP Eyeline mode allows the use of signal averaging to resolve signals from noise. The display in (a) shows how a low-level signal appears in the ordinary sampling oscilloscope display mode. The same signal is shown in (b) using Eyeline mode with 64-trace averaging applied.*

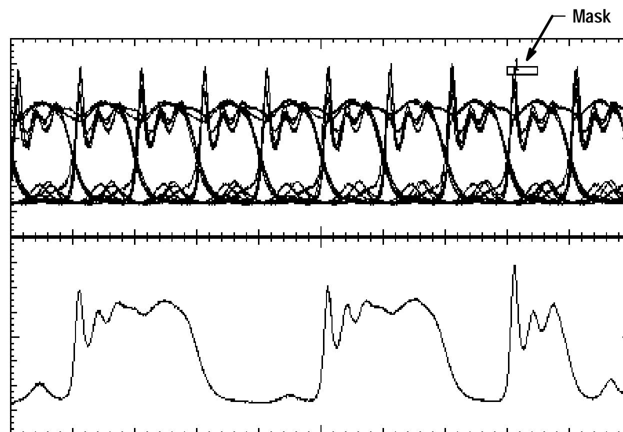


Fig. 4. *Error trace capture using HP Eyeline mode. The upper trace shows the complete eye diagram. The lower trace shows the bit pattern leading up to a mask violation. A custom mask was constructed to capture only those waveforms showing extreme overshoot.*

Despite its advantages, the HP 83480's HP Eyeline display mode is not the solution to all problems. It can only be used with HP pattern generators having a programmable trigger output, so it is not suitable for analyzing live traffic. And because it relies on multiple repetitions of the pattern to generate the eye, it is most suitable for short pattern sequences that repeat rapidly. At an OC-48 data rate (2.48832 Gbit/s), for instance, a complete eye showing all bit combinations of a $2^7 - 1$ PRBS pattern takes less than two seconds to generate. A complete $2^{23} - 1$ PRBS pattern at the OC-3 data rate (155.52 Mbits/s), however, would take 7.3 years!

Reference

1. C. Miller, "High-Speed Digital Transmitter Characterization Using Eye Diagram Analysis," *Hewlett-Packard Journal*, Vol. 45, no. 4, August 1994, pp. 29-37.

Design of Optical Receiver Modules for Digital Communications Analysis

These three bit-rate-specific optical plug-in modules are essential components of the HP 83480A Digital Communications Analyzer. They are for data rates of 155/622 Mbits/s, 2.488 Gbits/s, and 9.953 Gbits/s.

by Christopher M. Miller, Randall King, Mark J. Woodward, Tim L. Bagwell, Donald L. Faller, Jr., Joseph Straznicky, and Naily L. Whang

The transmission rates of telecommunication systems based on fiber-optic standards such as the Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) are at multiples of 51.84 Mbits/s. A significant amount of optical telecommunications network equipment operates at three, twelve, or forty-eight times this fundamental rate. Currently, new equipment is being deployed that operates at approximately 10 Gbits/s, or one hundred ninety-two times the fundamental rate.

Differing measurement requirements at these specific bit rates, along with the modularity available from the hardware architecture of the HP 83480 communications analyzer, allowed the design team to tailor plug-in modules for each application. This modularity also benefits customers, who can configure an instrument that best meets their current needs, then later modify it easily as their needs change. To date, three optical plug-in modules have been released. They are:

- HP 83481A 155/622-Mbit/s optical-to-electrical module
- HP 83485A 2.488-Gbit/s optical-to-electrical module
- HP 83485B 9.953-Gbit/s optical-to-electrical module.

Measurement Requirements

The measurement requirements inherent in the design and manufacturing of digital communications systems drove the specific design choices made for each plug-in module. The HP 83480 instrument family is used to characterize digital communications signals in the time domain. These signals are typically broadband and usually include a dc component. As an example, the inspection and analysis of eye diagrams are typical customer measurements.[1] Eye diagrams are constructed from multiple overlays of successive bit patterns with a synchronized trigger. To display the eye diagram properly, the measurement system must have sufficient bandwidth to show the fast transitions.

A measurement system is often most easily characterized in terms of its frequency response, or the magnitude and phase of the transfer function. This can be related to the time-domain impulse or step response performance by an inverse Fourier transform. The optimum frequency response of a measurement system depends on the waveform measurement parameter of greatest interest (rise and fall times, overshoot, etc.). A rule of thumb for reasonable measurements of rise and fall times is that the −3-dB bandwidth of the transfer function be at least three times and preferably five times the bit rate to be measured. Ideally, the transfer function should have a well-behaved roll-off and linear phase to prevent ringing or other measurement aberrations. The design target for the optical plug-in modules was to achieve a compromise between fast rise and fall times and excessive ringing. This was accomplished by controlling the amount of high-frequency peaking and striving for a Gaussian-like impulse response.

Very flat low-frequency performance is required for stable measurements of logic levels that extend over many bit periods, which occurs in both long-pattern-length pseudorandom binary sequences (PRBS) and live data transmissions. This is important, for instance, for accurate extinction ratio measurements. Extinction ratio is defined as the ratio of the signal power in the logic 1 state to the signal power in the logic 0 state. It is an important measurement of the distinction between logic states, the essential function of a digital communication system.

For a plug-in module to serve as a reference receiver, the frequency response must, at a minimum, comply with the low-pass Bessel-Thomson transfer function described in the SONET/SDH standards. In fact, mathematical simulations can demonstrate that even reference receiver frequency responses that technically meet the tight SONET/SDH standards can cause unacceptable time-domain artifacts that compromise the accuracy of extinction ratio measurements. Therefore, there is an advantage to a receiver frequency response that closely matches the ideal transfer function. It is critically important that the low-frequency transfer function, down to dc, be well-behaved. A low-frequency gain variation that either rises or droops will cause inaccurate extinction ratio measurements. For example, simulations have shown that a 0.2-dB low-frequency rise (which is within the SONET/SDH standard specifications) can cause an extinction ratio measurement of

10 to be in error by 10%.[2] In a manufacturing environment, such an error might cause the incorrect rejection of a good component.

Finally, the plug-in modules should have a dynamic range as large as possible. A high input power compression point extends the signal measurement range available to the user without adding an external optical attenuator. Noise considerations limit the low input signal range. Since these plug-in modules are often used for nonrepetitive waveform measurements (such as eye diagrams), waveform averaging often cannot be used to provide noise reduction.

Plug-in Module Overview

Each plug-in module contains an optical input channel with an optical-to-electrical (O/E) converter, at least one switchable SONET filter, and an electrical sampler with its associated pulse generation circuitry. In addition, each plug-in module has an electrical input channel and a trigger input to route the trigger signal to the mainframe. A generalized plug-in block diagram is shown in Fig. 1.

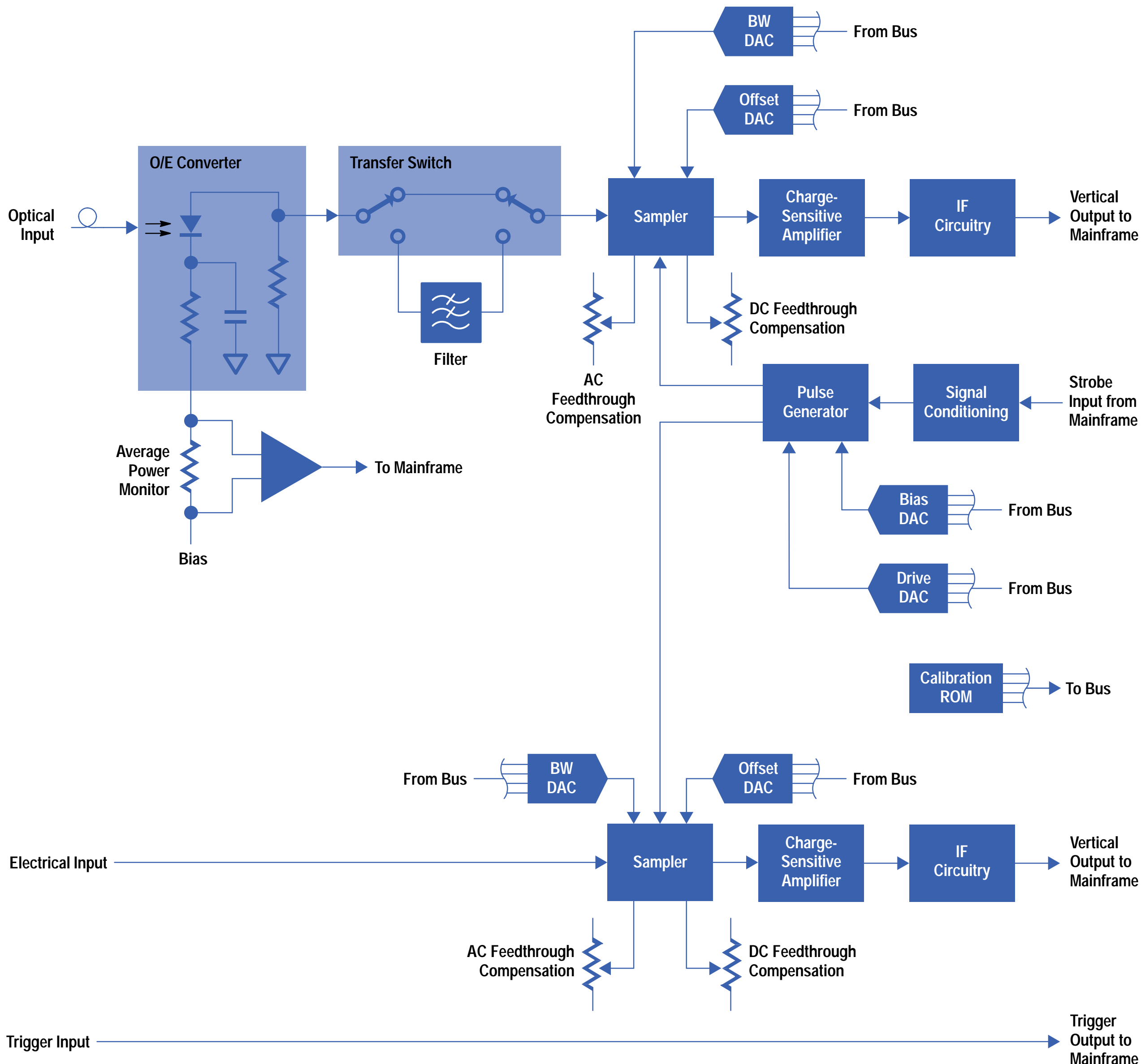


Fig. 1. *Generalized HP 83480 plug-in module block diagram.*

The O/E conversion starts with a photodiode that converts the incoming photons of light to a proportional electrical current. Because these receiver modules are intended to operate at the primary single-mode communication wavelengths of 1310 nm and 1550 nm, InP/InGaAs/InP p-i-n photodiodes are used. One of the major design choices is whether to add an electronic amplifier immediately after the O/E converter. This selection hinges mainly on a trade-off between the channel signal-to-noise ratio and frequency response. Amplified O/E converters can improve the sensitivity of the channel by

reducing the effective noise contributions of the subsequent electronic circuit stages. However, amplified O/E converters present more challenges in meeting the stringent frequency response requirements for optical reference receivers, especially at the higher data rates.

The detected electrical signal can be filtered to comply with the communication standard for reference receivers. Each plug-in module comes with either one or two SONET/SDH filters appropriate for the 155-Mbit/s to 10-Gbit/s transmission rates. These filters have a fourth-order Bessel-Thomson low-pass transfer function with a characteristic frequency (–3-dB frequency) at 0.75 times the transmitted bit rate. Around this transfer function there is a narrow tolerance window that depends on the bit rate. One of the key contributions of the HP 83480 is that the entire instrument meets this filtered response, not just the optical receiver. In addition, other communication standard filters can be installed to meet special customer requirements. Of course, all the filters can be switched out of the detected signal path to allow the maximum available measurement bandwidth.

Sampling Circuit Description. Signals at these bit rates cannot be digitized directly in real time. There are currently no analog-to-digital converters sufficiently fast to meet these measurement bandwidth requirements. Instead, a sampling technique is used that allows the display of signals that are both repetitive and have a stable trigger.[3] Many digital communications signals fit this description well enough to make the analyzer a very useful measurement tool. While many of the details of a sampling circuit are beyond the scope of this article, some insight into its function will provide an understanding of how the different plug-in modules have been optimized.

A sampler can be thought of as a very fast electrically controlled switch. A fast pulse is used to turn on the switch, which is connected to the analyzer input port. While the switch is on, a current related to the input signal flows into a capacitor. The amount of charge transferred during the sample interval is proportional to the signal present at that instant. If a stable trigger is available, we can eventually build up a representation of the input signal by scanning the time position of the sampling aperture relative to the trigger.

Some design complexity is required to build an electrical switch of sufficient speed. The sampling circuit consists of three main blocks: the sampler microcircuit, a step-recovery diode pulse generator which fires the sampler, and an amplifier and IF filter chain which reshapes the response of the sampler output into a bipolar pulse. The IF output thus generated is sent to the mainframe for further processing and display. The sampler and the step-recovery diode pulse microcircuits are leveraged from the previous-generation HP 54120 Series sampling oscilloscopes.[4]

It is not easy to generate a fast symmetric electrical pulse directly. Instead, the step-recovery diode pulse microcircuit provides a single falling edge with ~70-ps transition time. This waveform is sent to the sampler where it propagates on a slotline transmission line structure. The slotline is terminated by the package wall, effectively forming a short circuit. When the step-recovery diode pulse edge reaches the wall, an inverted reflection is generated and propagates in the reverse direction. The total voltage across any point along the slotline is the sum of both the incident step-recovery diode pulse and the inverted reflected pulse. At an electrical distance from the package wall equivalent to one-half the pulse fall time, the voltage waveform is roughly triangular with 70-ps rising and falling edges. Sampling diodes are placed on the slotline structure at this point. These diodes form the actual switch element.

The sampling diodes turn on when the total voltage provided by the step-recovery diode pulse and its reflection exceed the effective reverse voltage on the diodes. At that time the signal from the O/E converter is connected (switched) to a holding capacitor (see Fig. 2), which collects charge. By controlling the reverse bias voltage across the sampling diodes, the sampling aperture and therefore the bandwidth can be adjusted. A DAC (digital-to-analog converter) within the plug-in module provides this function. Fourier transform theory shows that the sampling aperture time is inversely related to the effective bandwidth. For a 20-GHz bandwidth, the sampling aperture is approximately 20 ps.

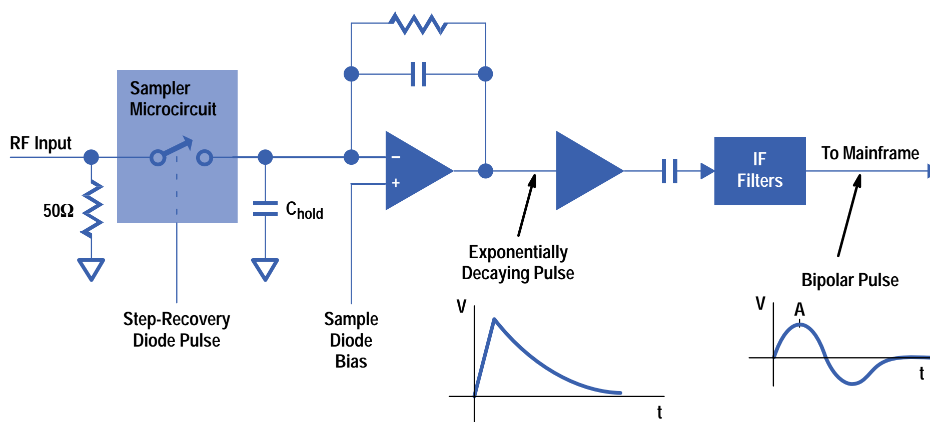


Fig. 2. *Generalized sampler circuitry.*

Why not make the sampling aperture as small as possible? As is usually the case, there is a trade-off involved. Here it is noise versus bandwidth. A larger sampling aperture means more charge can be collected per sample, resulting in a higher signal-to-noise ratio in the subsequent signal processing. It also means a reduction in measurement bandwidth. In reality, the trade-offs are not quite this simple. There are noise contributions from the sampling diodes that also increase at higher signal levels. In addition, larger signal levels also require more dynamic range in succeeding stages to avoid saturation effects. Therefore, selecting the appropriate sampler parameters is a key factor in the optimization of a plug-in module.

Once the charge has been collected, it must be amplified and shaped by filtering before being transferred to the mainframe for processing. The amplifier following the sampler is conceptually an operational amplifier with a resistor/capacitor feedback network. The feedback resistor needs to be as large as possible to minimize its noise contribution. Other system constraints and practical physical values limit the choice to a maximum of about 100 megohms. The input noise of the amplifier itself is minimized by using a discrete, low-noise, differential FET input followed by a precision, high-speed op amp. Significant design effort was expended on this circuit and the effective noise performance of the charge amplifier and IF circuitry is at least twice as good as the previous-generation products.

Because of the narrow sampling aperture, the sampled charge appears as a near-impulse to the much slower following amplifier. This amplifier stage eventually responds and produces an output pulse with an exponentially decaying tail that is proportional to the impulse amplitude. The subsequent filter sections in the plug-in module ac-couple the pulse, which prevents any drift in the dc level from altering the signal amplitude. These IF filter sections shape the exponentially decaying signal into a bipolar pulse, optimizing the signal-to-noise ratio before the analog-to-digital conversion is performed in the mainframe. In addition, the bipolar pulse provides a slowly changing peak amplitude so that the analog-to-digital conversion instant is less sensitive to timing variations in the mainframe.

By repetitively sampling the signal and using a synchronous trigger with a variable delay, a representation of the input signal can eventually be constructed. The sampler can only be triggered at a relatively low rate (40 kHz), so it takes multiple input signal cycles to complete the measurement process. If the IF signal is a faithful low-noise electrical representation of the input optical signal, the plug-in module hardware has done its job.

Electrical and Optical Calibration. The electrical sampling process is inherently nonlinear. To reproduce the electrical waveform accurately, a lookup table stored in the plug-in module provides the mapping from the input signal to the IF output for each sampler channel. For the best module performance, the user can perform an electrical calibration by selecting a softkey on the front of the mainframe. A built-in algorithm applies internal reference voltages to the sampler inputs, covering the allowable input range and automatically generating the calibration table.

The optical channel requires an additional calibration term to account for the optical-to-electrical conversion efficiency of the O/E converter. This conversion efficiency is dependent on the wavelength of the optical signal. Factory calibration constants are stored in the memory of each plug-in module for 1310-nm and 1550-nm wavelengths. The linearity of the O/E converters is excellent, so a single calibration constant is sufficient over the allowable optical input power range. Users can perform optical calibrations at other wavelengths by using an optical source with an accurately known power level. This also allows an operator to remove small insertion loss variations of the input connector for high-precision measurement applications.

A convenient feature of the plug-in module is the ability to monitor the average optical power accurately. The average current from a reverse-biased p-i-n photodiode is directly proportional to the average incident optical power. This proportionality constant is determined during the optical calibration process and is wavelength dependent. The average detected current is monitored by a variable-gain amplifier network, which provides a dynamic range of just over 30 dB.

Shown in Fig. 3 is a photograph of the three currently available optical plug-in modules. The side view of the plug-in module reveals the optical converter, samplers, step-recovery diode pulse generator, transfer switch, and filters.

HP 83485A Plug-in Module

The HP 83485A plug-in module is an integrated solution targeted for testing laser transmitters operating at rates up to 2.5 Gbits/s. For parametric measurements the optical channel offers a selection of either 12 or 20 GHz of well-behaved measurement bandwidth. As previously described, this bandwidth choice is made by changing the bias on the sampler diodes. The 12-GHz mode offers better sensitivity, having only about half as much channel noise as the 20-GHz mode. For SONET/SDH compliance testing at data rates of 622 Mbits/s or 2.5 Gbits/s (depending on which standard filter option is specified), the switchable low-pass filter shapes the overall frequency response of the channel. Filtered measurements are always made in the 12-GHz mode for best sensitivity, since the channel bandwidth is already limited by the narrower electrical filter bandwidth. In addition to the optical channel, a 20-GHz electrical channel and a 2.5-GHz trigger input are provided. Like the optical channel, the electrical channel can be switched to a 12-GHz mode for improved sensitivity.

Meeting the SONET/SDH specifications at 2.5 Gbits/s is a challenging task for optical transmitter manufacturers. These customers want a measurement system that contributes as little additional error as possible because measurement errors can reduce their manufacturing yields. Therefore, a significant effort in the design of the HP 83485A plug-in module was aimed at providing an accurate frequency response. The optical channel incorporates a 20-GHz unamplified O/E converter to maintain the highest integrity for eye diagram and mask compliance measurements. While many amplified optical receivers

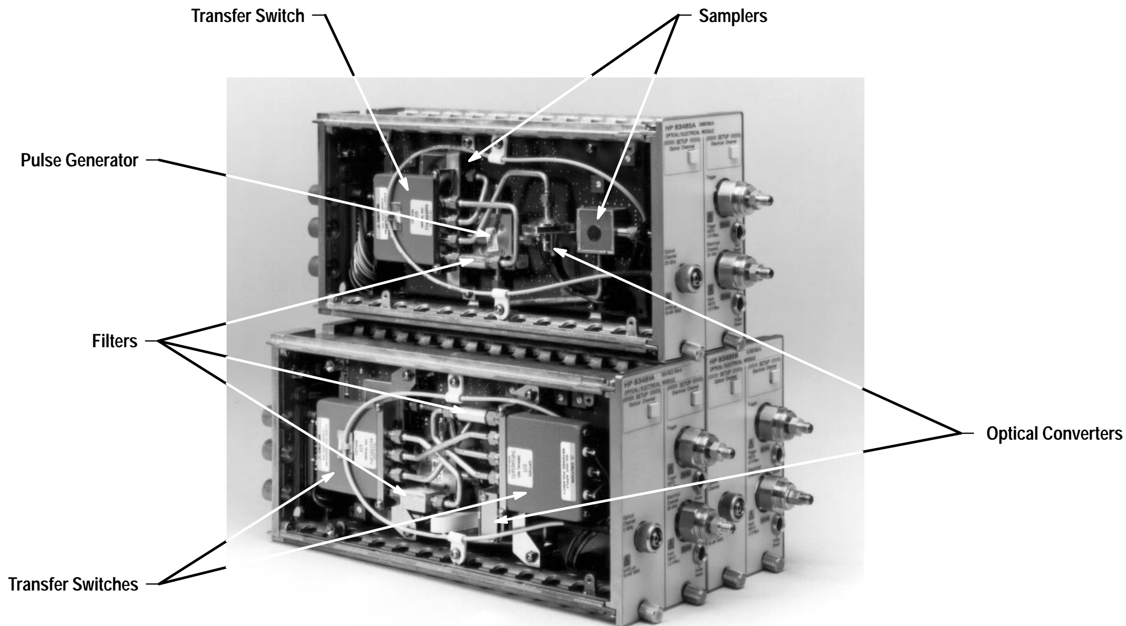


Fig. 3. *Three optical-to-electrical plug-in modules are available.*

now on the market claim compliance with the SONET/SDH standards, some still suffer from unacceptable frequency response variations. In the HP 83485A plug-in module, these variations are significantly reduced.

Optical Converter. The optical-to-electrical converter in the HP 83485A is similar to that of the HP 83440 family of stand-alone optical receivers.[5] The optical detection is performed by a custom 25-μm-diameter $InP/In_{0.53}Ga_{0.47}As/InP$ p-i-n top-illuminated mesa photodiode, which absorbs the incoming infrared light and converts it to an electrical current. The device allows light with wavelengths from 1200 to 1600 nm to pass through the antireflection coating and the top p-type InP layer and be absorbed in the intrinsic InGaAs layer below. The absorption of photons creates electron-hole pairs in the active layer. These carriers are then swept out by an electrical field formed by an applied reverse bias.[6] The carrier transit time across the InGaAs layer and the device capacitance determine the frequency response of the photodiode. A thicker intrinsic layer results in a lower device capacitance and a higher photodiode responsivity, which is the ratio of the detected photocurrent to the input optical power. This is true until the thickness is increased to a point where the quantum efficiency is unity or all the incident photons are absorbed. A thicker intrinsic layer also contributes to a longer transit time, so the thickness has to be carefully chosen to achieve the optimum photodiode bandwidth and responsivity.

The optical launch is required to have low optical reflections and maximum coupling to the photodiode. Optical reflections from a receiver can cause unpredictable measurement variations in some transmission systems. Any reduction in the power coupled to the photodiode produces a direct impairment in the signal-to-noise ratio of the measurement system. The optical launch in the HP 83485A O/E converter uses a single graded-index (GRIN) cylindrical lens to couple light into the small-area photodetector. The input fiber and the photodiode are tilted with respect to the optical axis to reduce reflections. The lens faces are antireflection-coated, resulting in optical return losses that are typically greater than 50 dB at each interface. The input optical fiber is locked in place using a specialized high-stability epoxy.

The optical converter package is hermetic. A mesa photodiode structure can be sensitive to moisture, resulting in increases in both dark current (reverse-bias photocurrent that flows without any light present) and noise. Water vapor can also cause a number of other semiconductor reliability problems. To provide hermeticity, the microcircuit package is sealed with glass and metal. The body and lid are made of gold-plated stainless steel. Glass-to-metal seals are soldered into the microcircuit body to provide electrical connections. A lens is soldered into the microcircuit lid to provide a hermetic optical port. This allows the input optical fiber to be attached with epoxy, since it is outside the hermetic wall. After the photodiode and other circuit elements are die-attached to the microcircuit floor and wire bonded, the lid is soldered to the body in a dry helium atmosphere.

The major modification made to the O/E converter used in the HP 83485A is the addition of a 50-ohm thin-film network at the photodiode output to provide a good electrical output match. This termination network is required to minimize reflections between components that can add frequency response ripple. Without the 50-ohm termination, the high output impedance of the photodiode would create a severe mismatch with the filter input, resulting in excessive ripple in the filtered response when used in the reference receiver mode. The two most important sets of reflection pairs are between the O/E converter and the low-pass filter and between the O/E converter and the sampler. The 50-ohm termination network in the O/E converter provides at least a 26-dB output match from dc up to 5 GHz, where interactions with the filter are critical. The termination network, however, does introduce some peaking in the O/E converter response near 20 GHz. The frequency

response of the sampler is adjusted accordingly to generate the best overall channel frequency response. Shown in Fig. 4 are the measured unfiltered and filtered frequency responses of an HP 83485A-based digital communications analyzer.

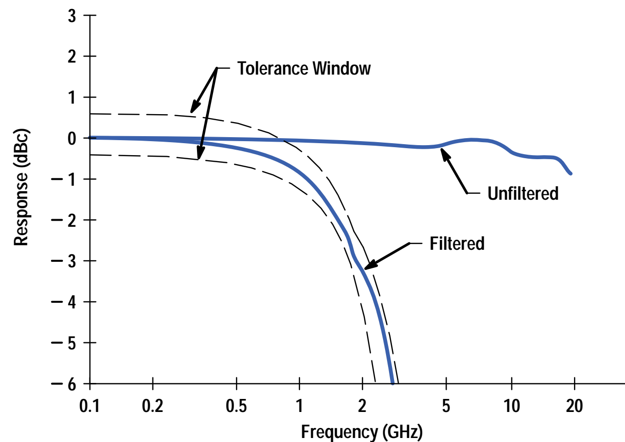


***Fig. 4.** Measured frequency responses of the HP 83480 digital communications analyzer with HP 83485A plug-in module.*

Because of the low optical conversion gain (~20V/W) that results from the unamplified O/E converter, there is a potential for poor plug-in module sensitivity. To alleviate this concern, adjustments were made to the sampling parameters. This resulted in a typical optical noise level of only 8 μW rms for the 12-GHz bandwidth setting.

HP 83485B Plug-in Module

The HP 83485B is a companion module to the HP 83485A and provides an integrated solution for testing laser transmitters operating at 10 Gbits/s. The HP 83485B also consists of an optical channel, an electrical channel, and a trigger channel. Its optical channel includes a fourth-order Bessel-Thomson low-pass filter, which can be switched into the measurement path for making the compliance measurements.

The design of the HP 83485B is leveraged from the HP 83485A, allowing reuse of most of the internal printed circuit boards and mechanical structures. To achieve an increased channel bandwidth, a higher-speed, 14-μm-diameter photodiode is used in the O/E converter, along with higher-bandwidth samplers in both the optical and electrical channels. The combination of these new components provides over 30 GHz of bandwidth in the optical channel. For the electrical channel, the higher-bandwidth samplers are biased to provide 40 GHz of electrical bandwidth. The amplifier and IF filter circuits following the sampler and the optical average power monitor circuits are identical to those in the HP 83485A plug-in module. Calibration of both the electrical and optical channels is also the same for this plug-in module.

Currently there is no SONET/SDH industry standard for the frequency response performance of reference receivers used in testing 10-Gbits/s transmitters. Expectations are, however, that the frequency response will have a shape similar to the fourth-order Bessel-Thomson transfer function used for the lower bit rates, but with wider tolerances. The HP 83485B filtered performance specifications give the best possible fourth-order performance while maintaining good manufacturability.

The higher bandwidth required for the HP 83485B forced several design and performance trade-offs. The most important trade-offs concerned the O/E converter and the sampler. At these higher frequencies, parasitic components in the O/E converter are more difficult to control, which can result in a frequency response that includes some peaking. The photodiode is followed by a termination circuit to provide a 50-ohm output match. Without the termination circuit, the photodiode frequency response is very well-behaved. If the termination circuit were ideal, this frequency response would be preserved. In practice, however, the bond wire required to connect the photodiode output introduces a parasitic inductance. This inductance resonates with the ~80-fF photodiode capacitance, resulting in a peaked frequency response.

In the time domain, this type of resonance would manifest itself as overshoot and ringing in response to an optical impulse at the input. These measurement artifacts are undesirable for precision time measurements, so it was necessary to make adjustments to both the unfiltered and filtered frequency responses. In the unfiltered mode, the sampler frequency response is adjusted to minimize the effect. In the filtered mode, the roll-offs of both the filter and the sampler were customized to compensate for the high-frequency peaking.

At 10 Gbits/s, mismatch ripple between the O/E converter and the filter can be significant. At these higher frequencies, the output impedance of the photodiode is dominated by its capacitance. By adding the termination circuit, the output impedance of the O/E converter is nominally 50 ohms at lower frequencies, but eventually becomes capacitive at higher frequencies. The circuit was designed to minimize this effect, but it could not be completely eliminated. Other potential

mismatch ripple contributions are minimized by selecting the best possible components in the filtered path, such as switches and connectors.

The filtered frequency response also requires good control of the hardware filter parameters. At lower bit rates, a tolerance of ±0.1 dB can be achieved for the passband region. At 10 Gbits/s, however, the filter elements become quite small in value, making manufacturing consistency more difficult. Correlation between the frequency response measurements of isolated filters and measurements of the optical channel in filtered mode is also important. A few tenths of a dB of absolute error is possible from the network analyzer systems used to make the filter measurements. The tolerance on the filter response is specified as ±0.5 dB in the passband. When assembled into the complete instrument, the plug-in module meets the transfer function specification with a tolerance of ±1.25 dB to 7.5 GHz, the characteristic frequency. While in principle controlling the mismatch ripple is important to avoid distorted eye diagrams and measurement errors, computer simulations show that the level of mismatch ripple present in the HP 83485B filtered path does not significantly affect extinction ratio measurements. Shown in Fig. 5 are the measured unfiltered and filtered frequency responses of an HP 83485B-based digital communications analyzer.

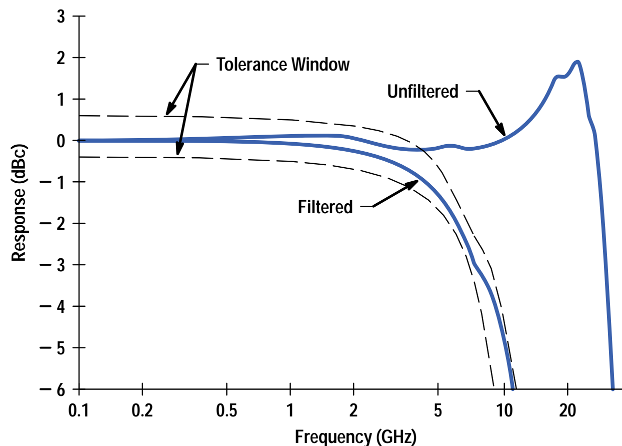


Fig. 5. *Measured frequency responses of the HP 83480 digital communications analyzer with HP 83485B plug-in module.*

Finally, the higher-bandwidth O/E converter and sampler have lower conversion gains, which results in a somewhat lower sensitivity for the HP 83485B plug-in module. Once again, the optical channel sampler bandwidth was adjusted in the filtered measurement mode to give the best possible sensitivity. This results in a 15-μW rms noise level for the 30-GHz bandwidth setting.

HP 83481A Plug-in Module

Compliance testing of 155-Mbit/s and 622-Mbit/s laser transmitters is the intended application for the HP 83481A plug-in module. Whereas laser transmitters for 2.5 Gbits/s and 10 Gbits/s are used primarily for long-haul transmission, transmitters at 155 Mbits/s and 622 Mbits/s are used for shorter-distance transmission and more varied applications. For some applications, the laser output power requirement is lower, which necessitates a measurement receiver with better sensitivity. When this is the case, a low-noise amplifier can be added to the O/E converter, before the sampler, to improve the noise figure of the entire receiver. The HP 83481A was designed with an amplified O/E converter for just such measurement needs.

In addition to an amplified O/E converter, the HP 83481A provides the customer with two switch-selectable Bessel-Thomson filters for compliance measurements at either 155 Mbits/s or 622 Mbits/s. Integration of these two filters within the receiver significantly improves measurement repeatability and reliability. The user can also select an unfiltered mode which bypasses the filters to provide the entire amplifier bandwidth for parametric measurements. The rest of the plug-in module, beginning at the sampler, is leveraged from the HP 83485A.

The amplifier approach chosen for the HP 83481A is a custom transimpedance design realized in a proprietary high-speed silicon bipolar IC process. This IC process is well-suited to delivering the level of performance needed to meet the demanding requirements of the SONET/SDH standards for reference receivers. An amplifier bandwidth of 3 GHz with a conversion gain of 500V/W was achieved. A low noise floor of approximately 1 μW rms results in an overall amplifier dynamic range of nearly 30 dB.

The design of the amplified O/E converter attempts to provide the user with a well-behaved pulse response for the required compliance measurements of laser transmitter waveforms. The amplifier is dc-coupled and the low-frequency gain is carefully controlled to give a stable pulse settling behavior. Pulse overshoot and ringing are damped to a suitable degree but not so much as to seriously degrade rise time. The low output impedance available with this IC technology meant that a series resistance was needed to match to 50 ohms. Consequently, some conversion gain was sacrificed to improve the output

match. Fig. 6 shows the frequency response of the amplified converter used in the HP 83481A. Flatness to the 0.75-bit-rate

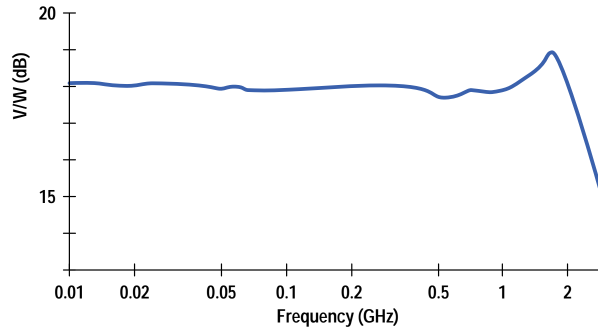


Fig. 6. *Frequency response of the amplified optical converter of the HP 83481A plug-in module.*

characteristic frequency of 466 MHz is better than 0.3 dB peak to peak. The small dip in gain from 500 MHz to 1 GHz is caused by the bond pad capacitance of the transimpedance IC. Peaking, common to all transimpedance amplifiers, is below 1 dB and is well beyond the SONET/SDH specified frequency range. With this level of performance, the amplifier contributes only about half the allowable deviation from the ideal reference receiver transfer function. Shown in Fig. 7 are the measured unfiltered frequency response and the two filtered frequency responses of an HP 83481A-based digital communication analyzer.

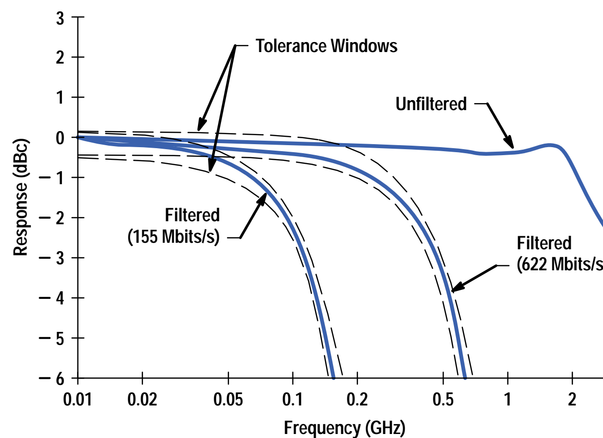


Fig. 7. *Measured frequency responses of the HP 83480 digital communications analyzer with HP 83481A plug-in module.*

Transimpedance Amplifier. The schematic of the complete amplifier is shown in Fig. 8. The transimpedance front end is formed by transistors Q1 to Q3 with resistor R_f (1100 ohms) setting the overall transimpedance gain. A single-stage voltage gain block (Q1, R_1) was chosen for maximum phase margin and has a gain of 17. Transistor Q2 is a buffer and transistor Q3 provides an additional level shift to increase the maximum output voltage swing. Q2 is also part of the dc level temperature compensation. Capacitors C_1 and C_f introduce zeros into the preamplifier frequency response, further increasing the phase margin. The driver amplifier was chosen for its simplicity and is implemented with a double emitter follower capable of driving low-impedance loads. The overall transimpedance of this amplifier is 1000 ohms and its intrinsic bandwidth is 5.2 GHz, resulting in a gain-bandwidth product of 5.2 THz-ohm. The intrinsic bandwidth is the bandwidth with an ideal zero-capacitance photodiode.

To stabilize the dc output level, the IC has passive temperature compensation. Active temperature compensation is capable of achieving essentially zero dc drift but it usually affects the frequency response and ac stability of the amplifier. The passive temperature compensation is obtained by causing the temperature drifts of the components in the signal path to cancel. At the output of the transimpedance preamplifier there is a positive voltage drift equivalent to the drift of the base-emitter junctions of Q1 and Q3 in series. This would ideally be canceled by the negative voltage drifts of base-emitter junctions in the driver. However, because the current densities of these transistors are not identical there is residual temperature drift of 76 μV/°C. This low value is suitable to minimize drift of the signal on the screen of the communications analyzer and allows accurate extinction ratio computations.

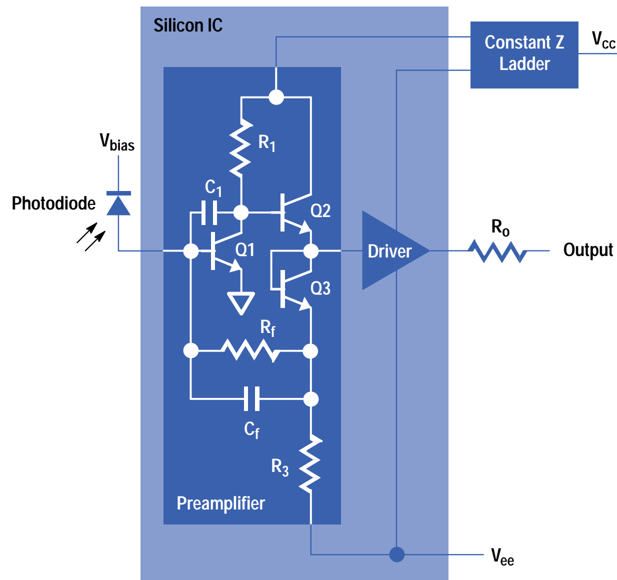


Fig. 8. *HP 83481A amplifier schematic diagram.*

To keep the amplifier response as flat as possible, a special constant-impedance ladder network was designed into the bias circuit of the IC. This circuit is made up of three sections of low-pass impedance transformers. The sections have progressively lower cutoff frequencies as the distance from the IC is increased. The advantage of this approach is that the unavoidable lead inductance is absorbed into the impedance transformer network. A separate network feeds the positive supply lines of the preamplifier and driver sections of the IC. This circuit provides a constant impedance of 7 ohms to the amplifier sections over the entire frequency range. Resonances are suppressed by the resistance of the bias network and a flat frequency response to dc is achieved.

The negative supply is less sensitive to low-frequency resonances but it has a different problem. The negative bias port feeds both the input and the output sections of the IC, and this is an undesired source of feedback. This coupling is made worse with increased bond wire inductance in the ground path. The effect manifests itself in increased frequency response peaking and some droop in the gain. The solution is to minimize the total ground bond length. All bias elements are integrated onto a thin-film substrate and a laser-shaped cutout is used to allow the IC to be mounted in close proximity to the bias circuit and flush with the top of the substrate. Multiple ground bonds of minimal length can be made to the substrate with this approach. The frequency peaking that results is typically less than 1 dB.

Optical Launch and Package. Besides the requirements that the optical launch have low optical reflections and maximum coupling to the photodiode, care must be exercised to ensure that no reflected rays can couple back into the input fiber.

The other design constraint is to minimize aberrations of the illuminated spot on the photodiode that falls predominantly within the 25-μm-diameter active region of the photodiode. This must be maintained over the temperature range of the instrument. To meet these goals an aspheric lens with a magnification of about 2 was chosen. This lens images light from the 9-μm fiber core diameter to a spot size of about 5 μm on the photodiode surface. The lens is coated with an antireflective layer and is mounted at a 20-degree angle to minimize reflections. This angle is enough to keep reflected rays from the photodiode surface from falling within the numerical aperture of the lens.

The lens and amplifier components (see Fig. 9) are mounted inside a hermetic package to protect the photodiode. A single-mode fiber is attached outside of the package. A sapphire window brazed into the lid of the package allows light to pass from the fiber to the lens. Since the window is tilted relative to the optical axis, an angle must be chosen that keeps the polarization dependent loss (PDL) at a reasonable level. An angle of 12 degrees was chosen for the window, which keeps PDL below 1.5%. This design eliminates the need to provide a hermetic seal around the fiber, thus simplifying the mechanical assembly process. The end of the fiber is polished at an angle to minimize reflection from this interface. Using this design a return loss in excess of 55 dB is readily obtained with negligible impact on coupled power.

The package design goal was to create a highly reliable hermetically sealed package to meet all functionality and manufacturability requirements. In selecting material for the package, Kovar (29% Ni, 17% Co + Fe) was chosen for its low thermal expansion, which closely matches that of the thin-film circuits and glass seals inside. Also, Kovar is an excellent material for laser welding and has good corrosion resistance. Fig. 10 shows a picture of the complete converter package.

A new ceramic brazing process called *active metal brazing* was introduced by the manufacturing engineers for attachment of the sapphire window to the Kovar lid. This new process eliminates the necessity of premetallization of the sapphire window edges by using a special active braze alloy of 63% Ag, 34.25% Cu, 1.75% Ti, 1% Sn. During brazing, the oxidization and diffusion of the active alloying element, Ti in this case, creates a chemical bond at the metal-ceramic interface. By means of

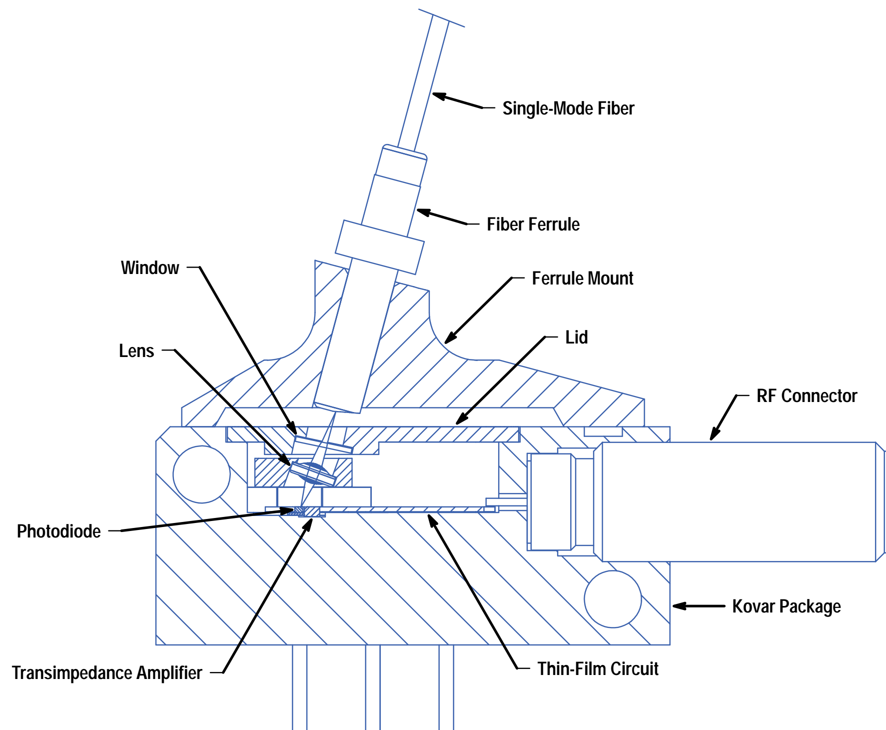


Fig. 9. *HP 83481A amplified converter cross section.*

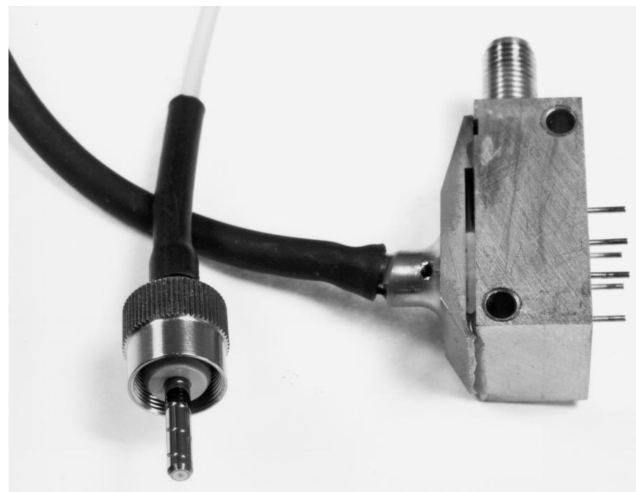


Fig. 10. *HP 83481A O/E converter.*

this relatively simple and robust new process, the cost saving realized in making this window/lid assembly is estimated at 70% compared with the conventional process.

The Kovar lid is hermetically attached to the body by laser welding using a Nd:YAG pulsed laser. The key to a sound weld joint is to keep a tight and consistent fit between the lid and the package body. Gold plating in the weld area has to be removed before welding to prevent cracking. The entire package assembly has withstood strife testing that involved temperature cycling from −50°C to +85°C, random vibration at 8g rms, and mechanical shock to 1800g.

Acknowledgments

The authors would like to thank a number of people who contributed to the design of these optical plug-in modules. Rin Park, Dave Roach, and Jeff Paul contributed their manufacturing and test expertise to the design effort. Kari Salomaa, Mel Duncan, and Mike Powers made significant converter package design contributions. Bernie Hovden provided the design team with his serviceability perspective. Finally, Mike McTigue and Chris Duff of HP's Electronic Measurements Division were extended team members contributing to the hardware and firmware design and were invaluable resources as consultants.

References

1. C.M. Miller, "High-Speed Digital Transmitter Characterization Using Eye Diagram Analysis," *Hewlett-Packard Journal*, Vol. 45, no. 4, August 1994, pp. 29-37.
2. P.O. Anderssen, and K. Akermark, "Accurate Optical Extinction Ratio Measurements," *IEEE Photonics Technology Letters*, Vol. 6, no. 11, November 1994.
3. W.M. Grove, "Sampling for Oscilloscopes and Other RF Systems: Dc Through X-Band," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MMT-4, no. 12, December 1966, pp. 629-635.
4. *High-Bandwidth Oscilloscope Sampling Architectures*, Hewlett-Packard Product Note 54120-3, September 1989.
5. R. King, D.B. Braun, S.W. Hinch, and K. Shubert, "High-Speed Time-Domain Lightwave Detectors," *Hewlett-Packard Journal*, Vol. 44, no. 1, February 1993, pp. 83-86.
6. S.R. Sloan, "Processing and Passivation Techniques for Fabrication of High-Speed InP/InGaAs/InP Mesa Photodetectors," *Hewlett-Packard Journal*, Vol. 40, no. 5, October 1989, pp. 69-75.

Transimpedance Amplifier O/E Converter Design

A block diagram of the amplified O/E converter of the HP 83481A 155/622-Mbit/s optical plug-in module is shown in Fig. 1. The photodiode functions as a detector whose output current reproduces the envelope of the received optical signal. The detected electrical signal is then amplified by a preamplifier which is followed by a postamplifier capable of driving a 50-ohm load. A variety of devices can serve as photodetectors, including p-i-n photodiodes, avalanche photodiodes, phototransistors, and photoconductors. However, for high-speed instrumentation applications, the p-i-n photodiode is favored because of its performance, simplicity, and price.

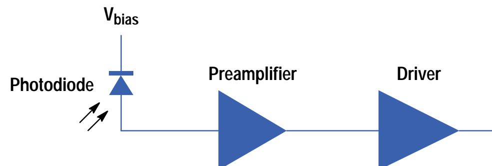


Fig. 1. *Block diagram of amplified O/E converter.*

The sensitivity and dynamic range of the amplifier are largely determined by the design of the preamplifier. Preamplifiers can be classified into two basic categories: voltage amplifiers and transimpedance amplifiers. The generalized schematic of a voltage amplifier is shown in Fig. 2. The photocurrent, i_{pd}, develops a voltage across a load resistance R_L and this voltage is amplified by a voltage amplifier with gain A. To use this converter for instrumentation applications the following relationship must be satisfied:

$$\frac{1}{2\pi R_L C_T} \geq B.$$

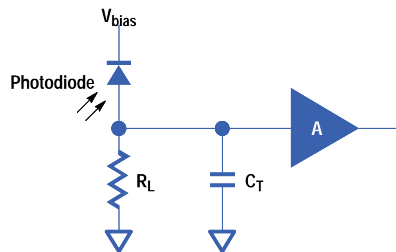


Fig. 2. *Photodiode with voltage amplifier.*

The dominant pole frequency formed by load resistor R_L and capacitance C_T must be larger than or equal to the desired bandwidth B, provided that the preamplifier bandwidth is sufficiently higher. C_T is the total capacitance at the input of the preamplifier, consisting of the diode capacitance, preamplifier input capacitance, and various parasitic capacitances. However, the sensitivity of this converter is relatively low, because only a small voltage signal is developed across a small load resistor R_L, while the Johnson noise current generated is relatively large. A low value for R_L is required to satisfy the above bandwidth relationship. The variance of the equivalent noise current for this front end is:

$$\langle i^2 \rangle = \frac{4kTFB}{R_L},$$

where R_L is the load resistance, F is the amplifier noise figure, B is the bandwidth, k is Boltzmann's constant, and T is the absolute temperature.

From this equation it can be seen that if we can increase the load resistance and somehow maintain the overall bandwidth we can reduce equivalent input noise current and therefore improve the converter sensitivity. This can be achieved by employing the transimpedance design. Here the load resistor R_L is connected as a feedback resistor on an inverting voltage preamplifier with gain –A as shown in Fig. 3. Provided that the preamplifier bandwidth is sufficiently high, the overall bandwidth will be:

$$B = \frac{1 + A}{2\pi R_L C_T}.$$

This equation shows that the bandwidth is increased by a factor of $1 + A$ over that of a voltage amplifier converter with the same R_L and C_T.

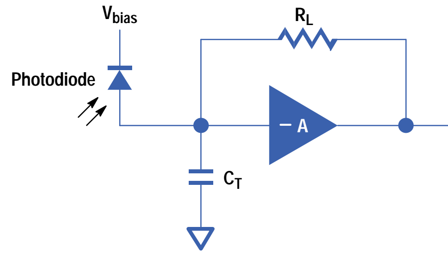


Fig. 3. *Photodiode with transimpedance amplifier.*

Furthermore, the strong negative feedback around the preamplifier decreases its susceptibility to component variation and improves the dynamic range over the voltage amplifier converter by the ratio of the open-loop gain to the closed-loop gain.

The main concern in using a transimpedance amplifier is its ac stability. The output current feedback formed by resistor R_L in conjunction with capacitance C_T results in input voltage feedback and 90 degrees of phase shift at the input of the voltage amplifier. The inverting amplifier contributes a phase shift of 180 degrees. This means that an additional 90-degree shift in the preamplifier would guarantee oscillations. In practice we need to maintain a minimum phase margin of approximately 45 degrees up to the frequency at which the preamplifier open-loop gain becomes less than one. In other words, the pole formed by R_LC_T should be the dominant pole in the circuit to achieve the best stability, and the pole internal to the voltage preamplifier should be substantially higher in frequency so as not to influence the overall frequency response. However, to achieve the highest possible frequency response, C_T is often minimized so that the dominant input pole frequency approaches the preamplifier internal pole frequency. This results in frequency response peaking. The gain increase at high frequencies improves the slew rate of the transimpedance amplifier, but introduces a certain degree of overshoot and undershoot, and moving those two poles even closer together results in additional ringing, ultimately leading to oscillations.

The gain of the transimpedance amplifier is expressed in ohms and its gain-bandwidth product is expressed in ohm-Hz. For a given device technology and circuit design the gain-bandwidth product is constant, so it is possible to trade bandwidth for gain provided that the ac stability is maintained.

The driver amplifier in Fig. 1 provides the final amplification of the signal. It can be as simple as a buffer amplifier to drive a low-impedance load or it can provide an additional gain and signal conditioning function.

Differential Time-Domain Reflectometry Module for a Digital Oscilloscope and Communications Analyzer

The HP 54754A differential TDR plug-in in conjunction with the HP 54750 digital oscilloscope or the HP 83480 digital communications analyzer significantly improves the speed and ease of making critical measurements in today's high-speed systems.

by Michael M. McTigue and Christopher P. Duff

With the advent of higher-speed systems, the issue of signal fidelity has become increasingly important. To maintain signal fidelity at these higher speeds, designers must be able to reduce discontinuities in the signal path and reduce coupling effects between different signal paths. To this end, many newer designs make use of differential transmission lines. An important tool in analyzing signal paths is time-domain reflectometry (TDR). Differential TDR makes this task easier and less time-consuming for differential transmission lines.

To analyze and optimize a differential transmission line system easily, the TDR system should have certain features:

- Fast display throughput. A standard technique in TDR is to observe the TDR trace while probing the line with a small capacitance. This helps locate the various points along the line. If the display throughput is too low, this process is tedious, especially if averaging is used to see small effects.
- Full stimulus/response matrix. To fully characterize a differential transmission line, the TDR system should allow control of both the stimulus and the response. Examples of possible needed measurements are differential response to differential-mode stimulus, differential response to common-mode stimulus, common-mode response to differential-mode stimulus, single-side response to differential-mode stimulus, and so on. Unbalanced differential lines and imperfect terminations can cause one mode to couple into the other, making it difficult to understand signals on differential lines. Being able to look at different corner cases of the stimulus/response matrix can be a big help in understanding these effects.
- Automatic calculation of parameters. The usefulness and ease of use of a TDR system are greatly enhanced if the system combines the various waveforms to get the desired responses. Also, it should be possible to have the response plotted as voltage, reflection coefficient, or ohms.
- Easy calibration of the TDR measurements. This includes establishing the reference planes and measuring the step heights so measurements can be made accurately.

The HP 54754A differential TDR plug-in for the HP 54750 digital oscilloscope and the HP 83480 digital communications analyzer (which share the same mainframe) achieves these goals through the use of optimized acquisition hardware, a rich firmware feature set, and flexible TDR hardware. This TDR system significantly improves the speed and ease of making critical measurements in today's high-speed systems.

Hardware Design

The samplers and TDR step generators for the HP 54754A plug-in are leveraged from the HP 54120 oscilloscope system. The samplers are a microwave design using state-of-the-art GaAs Schottky barrier diodes in a beam-lead package mounted on a thin-film microcircuit. The microcircuit is mounted in a machined cavity using 3.5-mm precision connectors. The step generators use a step-recovery diode to drive an anti-series pair of GaAs Schottky diodes, which switch an 8-mA current into the transmission line leading to the samplers. Like the samplers, the microcircuit is also implemented with beam-lead components on thin film in a machined cavity. This hardware produces a typical system rise time of 35 ps with a step top that is flat within ±5% for the first nanosecond after the edge and within ±1% after that.

The HP 54754A TDR plug-in has two TDR step generators to allow faster and easier differential TDR measurements. One possible hardware architecture would make the rising edges of both step generators coincident in time, as shown in Fig. 1a. Then, to switch from common-mode stimulus (shown) to differential-mode stimulus, the polarity of one of the step generators could be reversed. In practice, this system may have drawbacks. Typically, if the bias and setup of a step-recovery diode TDR step generator are changed, it takes time for the system to settle. This can take many seconds, which could be an issue for the measurement time. Also, the step may settle at a new time position, which would mean that different calibrations would be required for the positive and negative states of the step generator.

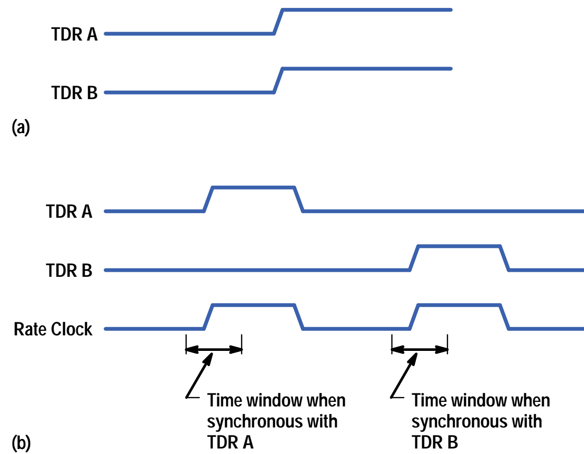


Fig. 1. *(a) One possibility for a dual-step-generator TDR system is to have the rising edges of the two TDR step generators coincident in time. (b) The HP 54754A plug-in pulses the first step generator while the second is quiet, and then pulses the second step generator while the first is quiet.*

To avoid these potential problems, the HP 54754A TDR system uses a *staggered* step generator architecture. This architecture pulses the first step generator while the second one is quiet, and then pulses the second step generator while the first one is quiet, as shown in Fig. 1b. Using this approach, all information needed for either common-mode stimulus or differential-mode stimulus is present without changing the bias or setup of the TDR step generators. The hardware simply selects which edge of the rate clock to send to the oscilloscope's trigger. Each TDR step generator is run at a constant rate, so once it has settled (after initial turn-on), it is stable.

Another requirement for a TDR system is that the trigger-to-TDR-step edge timing have low jitter and be stable. The rate clock signal's rising edge is the source for this timing. This signal has two critical paths. One is through buffer gates to the oscilloscope's main trigger and the other is through buffer gates and a delay line to the TDR step generators. The delay line is necessary to get the step generator's rising edge onscreen. In the HP 54754A system, there is the additional requirement that the system be able to select which TDR generator is synchronous with the trigger. This hardware is shown in Fig. 2. The hardware is set up so that the gates in each signal path are minimized and signals are as fast and as clean as possible to minimize jitter. Also, to minimize the timing drift between the two step generators, the path to the step generators goes through a dual ECL flip-flop. This better matches thermal and bias conditions for these paths. Trigger-to-TDR-step jitter for this system is ~1.4 ps rms.

In most TDR measurements it is desired to define a reference plane at the input to the system under test so that measurements can be made relative to that point. In differential TDR measurements, it is important to have the edges from the two step generators arrive at the reference plane simultaneously. To accomplish this, the step-recovery diode bias in each step generator (bias determines the stored charge on the step-recovery diode and therefore determines when it will fire) is controlled by a DAC to produce a ±400-ps range over which the TDR step can be moved onscreen. This allows connection to the system under test with cables that are as much as 800 ps different in electrical length. The reflected or transmitted signals from the reference plane can be lined up onscreen using the skew adjustment provided for each sampling channel. Having the ability to adjust both the TDR position and the channel skew allows quicker and easier calibration for differential TDR even if the connection cables are not perfectly matched in electrical length.

When the HP 54754A plug-in is not being used for TDR, the two channels can be used as normal high-bandwidth sampling channels. To provide the normal external trigger path needed for non-TDR uses, the trigger path through the plug-in to the main trigger in the mainframe must be preserved. This means that it must be possible to turn off the trigger injected into the trigger path for TDR without affecting the normal external trigger. One way to achieve this would be to use a mechanical microwave switch, but this would add significant cost, take up valuable space, be an additional reliability concern, and add weight. A preferable solution is to use a p-i-n diode to switch in the trigger signal when TDR is on (see Fig. 2). When the p-i-n diode is forward-biased, it has low impedance and drives the trigger path through a coupling capacitor. When the p-i-n diode is reverse-biased, it has very high impedance and low capacitance. This allows the external trigger signal to flow though the trigger path with only a minor discontinuity at the TDR trigger injection point. This injection point is implemented on the printed circuit board using SMA board connectors, a surface mount capacitor, and a p-i-n diode.

The rest of the hardware in the HP 54754A differential TDR plug-in is similar to the other plug-ins for the HP 54750 and HP 83480 systems (see article, page 1). This includes the low-noise front-end amplifiers, temperature-compensated sample strobe generation circuits, and feedthrough compensation circuits. The low-noise amplifiers are optimized to achieve a very low noise floor of ~260 μV rms at 12.4 GHz and ~660 μV rms at 18 GHz. The temperature-compensated sample strobe generation circuits allow good vertical accuracy stability (0.6% for 12.4 GHz and 1.2% for 18 GHz over ±2°C) without using feedback sampling.

Fig. 2. Hardware diagram of the HP 54754A differential TDR plug-in.

Firmware Design

The goal of the HP 54754A firmware was to take full advantage of the power of the hardware and provide an easy-to-use interface to it. Many of the features of the HP 54754A were leveraged from the HP 54120 oscilloscope.[1-4] This product offered single-ended TDR with normalization capabilities and the ability to read out ohms and distance using cursors. The HP 54754A expands upon this foundation, providing full differential characterization in addition to single-ended measurements.

Single-Ended Control. To characterize a single-ended TDR signal, the instrument must be capable of reading out any point along the trace in distance and impedance or reflection coefficient. It also must provide filtering, typically in the form of normalization, which limits the bandwidth through the system so that edges traveling at different speeds can be characterized. Finally, the instrument must be flexible and easy to use.

For the instrument to convert voltage into impedance, a calibration must be performed to determine the signal amplitude and offset at 50 ohms and with a short circuit. The samplers used in the HP 54754A module are subject to drift with changes in temperature. If the temperature has drifted more than $\pm 2°C$, a calibration should be performed before any critical measurements. Determining these offsets allows any signal voltage to be converted into impedance or reflection coefficient. Reflection coefficient is a normalized percentile such that 100% represents an open circuit, -100% represents a short circuit, and 0% represents 50 ohms. Reflection coefficient ρ is determined by:

$$\rho = (V_{in} - V_{50})/V_{amp},$$

where V_{in} is the voltage to be converted to ρ, V_{50} is the voltage of the step with a 50-ohm termination, and V_{amp} is the amplitude of the step into a short circuit.

Voltage can also be translated into impedance using ρ:

$$\text{ohms} = Z_n \times (1 + \rho) / (1 - \rho),$$

where ρ is the reflection coefficient to be converted to ohms and Z_n is the nominal impedance of the system, 50 ohms.

The HP 54754 provides a variety of ways to translate voltage into impedance or reflection coefficient. Waveform markers can be placed on the voltage waveform and set to read out in impedance or ρ. This allows the signal to be viewed in voltage, which is most familiar for many users, but permits any point to be translated to the appropriate units. Additionally, the entire signal can be displayed and controlled in impedance or ρ (Fig. 3). This obviates the need to run the marker down the signal, mapping every point from volts. When the signal is displayed in impedance, the scale and offset controls change to ohms per division and ohms offset.

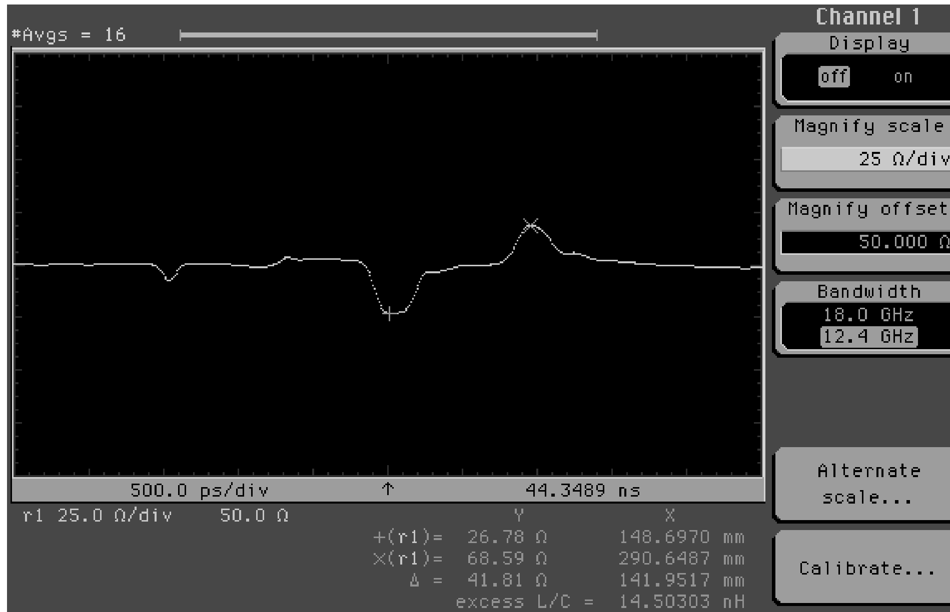


Fig. 3. *Cursors can read out in units of distance in meters or feet along the horizontal axis. Waveforms can be displayed in units of volts, ohms, or reflection coefficient along the vertical axis. Scaling in ohms/division and ohms offset is provided.*

To convert any point along the signal to distance, the reference plane is determined by the calibration. The reference plane is the point at which the TDR pulse is launched into the device under test. This allows any cabling up to the launch point to be eliminated from the distance measurement. The distance to any point after the reference plane is determined by:

$$\text{distance (meters)} = \frac{c}{\sqrt{\epsilon_e}} \frac{(t_{in} - t_{ref})}{2},$$

where c is the speed of light in meters per second, ϵ_e is the effective dielectric constant of the transmission medium, t_{in} is the time to be converted to meters, and t_{ref} is the time at which the TDR pulse is launched at the reference plane.

The constant 2 in the equation above accounts for the round-trip time of the pulse, since only the time from the reference plane to the event is typically desired. The HP 54754A allows the distance to be measured in meters or feet (Fig. 3).

Typically in single-ended TDR the user may see a small impedance variation in the signal and wonder how to flatten the signal. By integrating the reflection coefficient around the variation, the HP 54754A can compute the excess inductance or capacitance[5] causing the impedance variation (Fig. 4). This can lead to trimming the part or adding inductance or capacitance to correct the variation. The waveform markers are used to define a portion of the trace to integrate, time t_0 and t_1, and the following equation is used to compute excess L or C:

$$\tau = \int_{t_0}^{t_1} \rho dt.$$

If $\tau > 0$,

$$L = 2Z_0\tau \frac{(Z_0 + Z_n)^2}{4Z_0Z_n},$$

and if $\tau < 0$,

$$C = -\frac{2\tau}{Z_0} \frac{(Z_0 + Z_n)^2}{4Z_0Z_n},$$

where ρ is the reflection coefficient of the waveform, Z_0 is the impedance at t_0, Z_n is the nominal impedance of the system (50 ohms), L is the excess inductance between t_0 and t_1, and C is the excess capacitance between t_0 and t_1.

Since the impedance of the first point is used as the reference impedance, these equations will work in environments that are not nominally at 50 ohms.

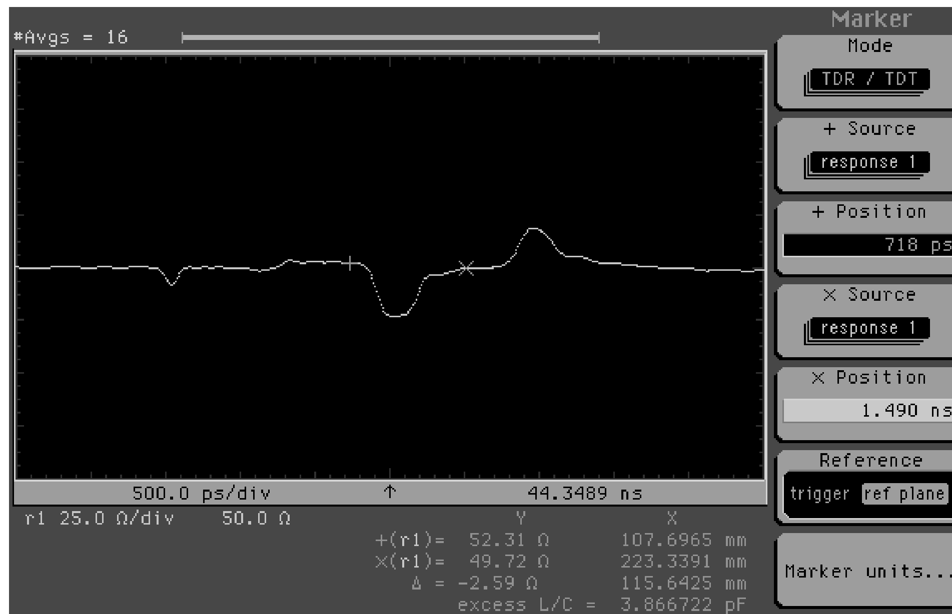


Fig. 4. *The waveform is integrated between the + and × markers to yield an excess capacitance of 3.86 pF. Negative-going variations are capacitive and positive-going variations are inductive.*

The rise time of the step generated by the pulse generators is fixed at 35 ps. Frequently a more realistic stimulus for the device under test is a slower step. Normalization allows the simulation of a slower-speed step through the device under test. If the full accuracy of normalization is not needed, a firmware digital filter can be used. This digital filter does not require calibration and can be applied to any waveform. The step speed can be specified by either rise time or 3-dB bandwidth.

The HP 54120 oscilloscope made measurements one waveform at a time and stored the results in a special waveform memory. The goal of the HP 54750 was to allow normalization and other advanced measurements in real time, updating for every waveform acquisition, without perceptibly slowing down the instrument. The HP 54750/83480 mainframe takes advantage of its dual-computer design and can perform normalization on real-time data. Many excellent articles have been written describing the normalization process, which will not be discussed in this article.[2]

For customers who require a quality step faster than 35 ps, Picosecond Pulse Laboratories makes a product capable of producing steps with 15-ps rise times in conjunction with the HP 54752B 50-GHz plug-in. The HP 54754A provides an external stimulus mode which allows easy control of this step generator.

Differential Control. Since the HP 54754A plug-in module contains two pulse generators that are staggered in time, differential and common-mode measurements can be made. The firmware in the product controls the pulse generators by synchronizing first with one, then with the other for every acquisition so that one TDR is silent while the other is pulsing. Waveforms are acquired for both the active, pulsing TDR and the nonactive, silent TDR. Since the active TDR signal can couple into the nonactive TDR and cause small variations, both the active and nonactive waveforms are acquired for each pulse generator on each acquisition.

If AA is the signal on line A caused by pulsing TDR A, AB is the signal on A caused by pulsing TDR B, BB is the signal on B caused by pulsing TDR B, and BA is the signal on B caused by pulsing TDR A, then for common-mode stimulus, the signal on line A, or signal A, is AA + AB and signal B is BB + BA. For differential-mode stimulus, signal A is AA − AB and signal B is −(BB − BA). The negative sign in the last equation inverts stimulus B for differential TDR to arrive at a more customary differential stimulus consisting of a positive-going step for stimulus A and a negative-going step for stimulus B.

Table I
Derivation of Response Waveforms

Volt Units (A and B in volts)	Differential Response	Common-Mode Response
Differential Stimulus	A − B	(A + B)/2
Common-Mode Stimulus	A − B	(A + B)/2

Ohm Units (A and B in ohms)	Differential Response	Common-Mode Response
Differential Stimulus	A + B	(A − B)/4
Common-Mode Stimulus	A − B	(A + B)/4

ρ Units (A and B in ρ)	Differential Response	Common-Mode Response
Differential Stimulus	(A + B)/2	(A − B)/2
Common-Mode Stimulus	(A − B)/2	(A + B)/2

In additional to controlling the TDR stimulus through the pulse generators, the firmware constructs all of the TDR response waveforms. Differential and common-mode responses are available for both differential and common-mode stimulus. For voltage units, the differential response is the difference of the responses to the two stimulus signals, A and B. The common-mode response is the sum of the responses to the stimulus A and stimulus B waveforms divided by two. Table I lists the derivations of the common-mode and differential response waveforms for units of volts, ohms and ρ.

All of the unit conversion capabilities of the HP 54754A can be applied to differential and common-mode stimulus and response waveforms. The power of the HP 54750/83480 far surpasses any other product of its type. The flexibility with which units can be converted across all stimulus types coupled with the automatic computation of complex differential and common-mode responses provides full-featured characterization of differential lines. The real-time normalization and advanced features like excess L and C computation make the instrument ideal for single-ended TDR as well.

Acknowledgments

Many people were involved in the design, testing, and release of the HP 54750 and HP 83480 mainframes and the HP 54754A plug-in module. John Kerley suggested ideas that led to the staggered TDR implementation. Wayne Helgoth conceived an extraordinary mechanical design that allowed all of the components to fit into the plug-in. Dave Long and Ken Rush developed the original normalization algorithms. Dave Dascher designed and proved the excess L/C measurement.

References

1. *TDR Fundamentals*, Application Note 62, Hewlett-Packard Company, April 1988.
2. *Improving Time-Domain Network Analysis Measurements*, Application Note 62-1, Hewlett-Packard Company, April 1988.
3. *TDR Techniques for Differential Systems*, Application Note 62-2, Hewlett-Packard Company, October 1990.
4. *Advanced TDR Techniques*, Application Note 62-3, Hewlett-Packard Company, May 1990.
5. D.J. Dascher, "Measuring Parasitic Capacitance and Inductance Using TDR," *Hewlett-Packard Journal*, Vol. 47, no. 2, April 1996, pp. 83-96.

Frequency Response Measurement of Digital Communications Analyzer Plug-in Modules

It has been extremely difficult to characterize the SONET/SDH standard receiver with tolerances of ±0.3 dB. This paper describes a method for calibrating photoreceiver frequency response with the low inherent uncertainty of the U.S. National Institute of Standards and Technology Nd:YAG heterodyne system and transferring this calibration to a production test system while maintaining a low uncertainty.

by Rin Park and Paul D. Hale

It is extremely difficult to test the SONET/SDH standard receiver accurately with tolerances of ±0.3 dB as specified by ITU-T standard G.957. Over the years several test methods have been developed, but none has an accuracy good enough to support the 0.3-dB tolerance. There are two main reasons for the high measurement uncertainties: inaccurate knowledge of the optical stimulus and large uncertainties in the microwave power measurement.

To support the SONET/SDH receiver test with an adequate accuracy, we have developed a method for calibrating photoreceiver frequency response with the low inherent uncertainty of the U.S. National Institute of Standards and Technology (NIST) Nd:YAG heterodyne system and transferring this calibration to a production test system while maintaining a low uncertainty. This is achieved by combining a photoreceiver with a microwave power sensor and calibrating the response of the combined unit, eliminating RF calibration and mismatch uncertainties.[1] The calibration expanded uncertainty of the transfer standard is about 0.06 dB with coverage factor of two (2σ).*

The theory for measuring a communications analyzer module using the transfer standard is described below. The production measurements of the communication analyzer modules have given excellent results with very good repeatability and long-term stability.

NIST Nd:YAG Heterodyne Measurement System

NIST uses a Nd:YAG heterodyne system for measuring scalar frequency response because the excitation of the detector can be calculated from first principles. All system calibrations required are well-understood and independent of the frequency response measurement. A schematic of the heterodyne system is shown in Fig. 1. The system uses two commercially available single-mode monolithic-ring Nd:YAG lasers operating at 1.319 μm. The frequency of each laser can be tuned thermally to give beat frequencies from several tens of kilohertz to greater than 50 GHz. The beats have a short-term bandwidth of about 3 kHz. The beat frequency is

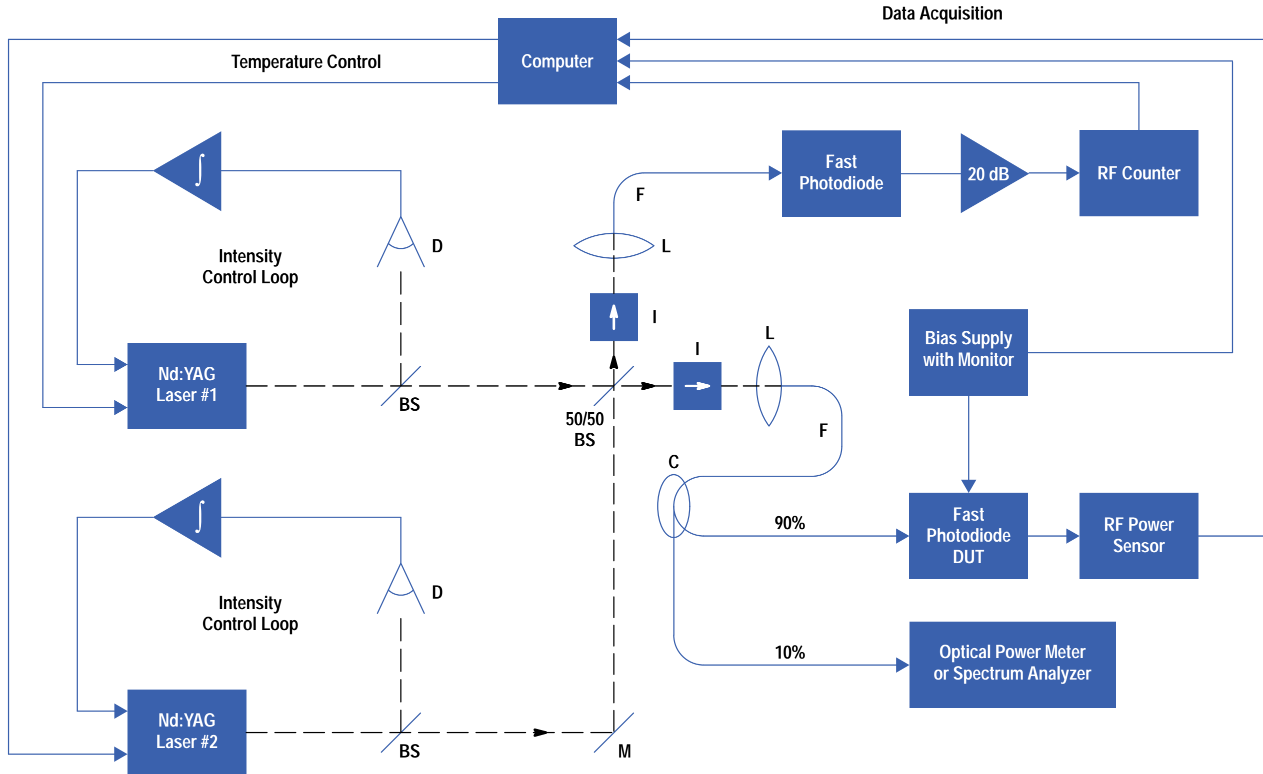
measured with a microwave counter. As the frequency is scanned, data is acquired automatically. The resolution of the system is limited by the scan rate, the frequency jitter, and the time constants of the data acquisition equipment. At present, the highest resolution achievable is about 200 kHz using a swept-frequency technique, although higher resolution can be achieved using a phase-locked loop technique.

The lasers are power stabilized so that nearly equal power from each laser is delivered to the DUT (device under test). The combined laser beams pass through polarizing isolators so that the signal from each laser is in the same polarization state when it reaches the DUT. This ensures nearly 100% modulation depth and eliminates instabilities resulting from feedback. The total optical power incident on the photoreceiver is:

$$P_{total}(t) = (P_{o1} + P_{o2}) + 2\sqrt{P_{o1}P_{o2}}\cos(2\pi ft), \qquad (1)$$

where P_{o1} and P_{o2} are the optical powers delivered to the photoreceiver from the lasers and f is the difference frequency between the two lasers.

* See reference 2 for a complete discussion of expanded uncertainty and coverage factor.



The photocurrent generated by the photodiode is given by:

Fig. 1. *NIST Nd:YAG heterodyne system. Labeled components are the beam splitter (BS), mirror (M), isolator (I), lens(L), single mode fiber (F), large-area detector (D), and integrating amplifier ($\int$).*

$$\begin{aligned} i_p(f,t) &= (P_{o1} + P_{o2})R(0) + 2\sqrt{P_{o1}P_{o2}}\,R(f)\cos(2\pi ft) \\ &= i_{dc} + i_{rf}, \end{aligned} \tag{2}$$

where R(f) is the responsivity of the detector (in A/W) at frequency f. The first term on the right side is the dc photocurrent, i_{dc}, which flows through the bias supply, and the second term is the RF photocurrent, i_{rf}, which flows through the RF load (the microwave power sensor) through a dc blocking capacitor. The mean squared photocurrent generated by the photodiode is:

$$\begin{aligned} \left\langle i_p^2(f)\right\rangle &= (P_{o1} + P_{o2})^2R^2(0) + 2P_{o1}P_{o2}R^2(f) \\ &= \left\langle i_{dc}^2\right\rangle + \left\langle i_{rf}^2\right\rangle. \end{aligned} \tag{3}$$

If P_{o1} is nearly equal to P_{o2}, then $2P_{o1}P_{o2}$ is, to the first order, equal to $0.5(P_{o1} + P_{o2})^2$. The normalized frequency response, $\mathfrak{R}^2(f)$, which is defined as $R^2(f)/R^2(0)$, can then be found by taking the ratio of the RF power to the dc electrical power delivered to a load $R_L = 50\Omega$:

$$\begin{aligned} \frac{2P_{rf}}{\left\langle i_{dc}^2\right\rangle R_L} &= \frac{\left\langle i_{rf}^2\right\rangle R_L}{0.5\left\langle i_{dc}^2\right\rangle R_L} \\ &= \frac{2(P_{o1}P_{o2})R^2(f)R_L}{0.5(P_{o1} + P_{o2})^2R^2(0)R_L} \\ &\approx \frac{R^2(f)}{R^2(0)} \\ &= \mathfrak{R}^2(f). \end{aligned} \tag{4}$$

P_{rf} is a function of frequency, includes corrections for sensor calibration factor and mismatch, and is the power that would be delivered to an ideal load R_L. In an ideal measurement $\langle i_{dc}^2 \rangle$ is constant but in any real measurement system it may vary because of changing optical power coupled to the photodiode. Using the ratio of RF to dc power reduces errors (to the first order) resulting from drifting optical power and eliminates the necessity of monitoring the powers separately. The normalized frequency response can be quoted in decibels as $20\log[\Re(f)]$. The electrical bandwidth of the device is the frequency at which $20\log[\Re(f)]$ has fallen by 3 dB from the low-frequency level.

Arbitrary Modulation Depth

Signals that do not have 100% modulation depth are commonly used in optoelectronic test equipment. An arbitrary modulation depth can be synthesized by varying one or both of the laser powers in the heterodyne system and can be modeled using the formalism presented above. The resulting equations are applicable to any source with an arbitrary modulation depth, such as a laser with direct or external modulation. A notationally simple way to change the model is to let $P_{o1} = \alpha P_o$ and $P_{o2} = (1 - \alpha)P_o$. Then the total optical signal incident on the photodiode is:

$$P_{total}(t) = P_o + 2P_o\sqrt{\alpha(1-\alpha)}\cos(2\pi ft). \quad (5)$$

The average optical power is P_o and the absolute modulation depth is $4P_o[\alpha(1 - \alpha)]^{1/2}$, or in fractional units:

$$M_o = \frac{P_{max} - P_{min}}{2P_o} = 2\sqrt{\alpha(1-\alpha)}. \quad (6)$$

The mean squared currents in the photodiode are:

$$\langle i_{dc}^2 \rangle = P_o^2 R^2(0) \quad (7)$$

and

$$\langle i_{rf}^2 \rangle = 2P_o^2\alpha(1-\alpha)R^2(f), \quad (8)$$

so the ratio of the powers is:

$$\frac{2P_{rf}}{\langle i_{dc}^2 \rangle R_L} = M_o^2\Re^2(f). \quad (9)$$

Hence, the modulation depth of an arbitrary source can be measured with a detector of known normalized response.

Swept-Frequency Ratio Measurements Using the Transfer Standard

The fundamental problem in making photodetector frequency response measurements is getting accurate knowledge of the modulation depth of the source. The modulation depth of the heterodyne beat signal is known from fundamental principles, but the modulation depth of a directly modulated laser diode or Mach-Zehnder modulator is not so simple. Accurate knowledge of the modulation depth of these sources can be obtained using a ratio technique to compare the response of a DUT to the transfer standard. The ratio measurement system, shown in its simplified form in Fig. 2, consists of a modulated light source with unknown modulation depth and a 1-by-2 fiber coupler.

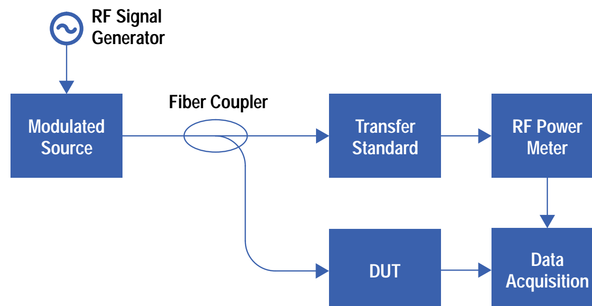


Fig. 2. *Apparatus for measuring the modulation depth of an unknown source to calibrate the DUT.*

The light source consists of a directly modulated laser and a CW laser modulated by an external Mach-Zehnder modulator. The RF signal is fed into the directly modulated laser from 200 kHz to 5 GHz, and at 5 GHz it is switched into the Mach-Zehnder modulator, which covers from 5 GHz to 20 GHz. To cover the frequency band from 200 kHz to 20 GHz, two RF sources are used. The outputs from the directly modulated laser and the Mach-Zehnder modulator are selected by an optical switch, which is connected to the 1-by-2 fiber coupler.

The coupling ratio of the coupler does not need to be specified. The signal from the coupler is delivered to a DUT and to a reference photodiode attached to an RF power sensor. The modulation depth of the source can be calculated using equation 9 and the known frequency response of the transfer standard photodiode. The frequency response of the DUT is found as:

$$\mathcal{R}^2_{DUT}(f) = \frac{1}{M_o^2}\frac{P_{rf_DUT}}{0.5\left\langle i^2_{dc_DUT}\right\rangle R_L}. \qquad (10)$$

Calculation of the DUT frequency response this way includes the calibration uncertainty of the power sensors both at NIST and at the commercial laboratory, giving a total expanded uncertainty of ± 0.2 dB or more. Combined with other factory uncertainties, this may give an unacceptable uncertainty for the intended test system.

One possible alternative is to calibrate the response of the photodiode combined with the RF power sensor. This method totally eliminates uncertainties resulting from power sensor calibration and impedance mismatch. The combined frequency response $\mathbb{R}^2(f)$ measured on the NIST heterodyne measurement system includes the power sensor calibration factor and impedance mismatch, and is given by:

$$\begin{aligned}\mathbb{R}^2(f) &= \frac{\mathcal{R}^2(f)}{C} \\ &= \frac{P_M}{0.5\left\langle i^2_{dc}\right\rangle R_L}.\end{aligned} \qquad (11)$$

In equation 11, P_M is the indication of the power meter after zeroing and calibration against the 50-MHz reference signal (from the power meter) using a calibration factor of 100%. C is given by:

$$C = \frac{1}{k}\left|1 - \Gamma_{pd}\Gamma_{sensor}\right|^2 \qquad (12)$$

where k is the sensor calibration factor and Γ_{pd} and Γ_{sensor} are complex reflection coefficients. However, C need not be known and the power meter reading need not be corrected with a frequency dependent calibration factor. When the standard photodiode/power sensor combination is used in the ratio test system the combined frequency response is used to find the modulation depth of the source:

$$M_o^2 = \frac{1}{\mathbb{R}^2(f)}\frac{P_{M_REF}}{0.5\left\langle i^2_{dc_REF}\right\rangle R_L}. \qquad (13)$$

The normalized response of the DUT is then found using equation 13, the dc photocurrent, and the RF power (including all calibration factors) from the DUT:

$$\mathcal{R}^2_{DUT}(f) = \frac{1}{M_o^2}\frac{P_{rf_DUT}}{0.5\left\langle i^2_{dc_DUT}\right\rangle R_L}. \qquad (14)$$

In some test applications it may be preferable to measure the DUT frequency response in terms of the coupling ratio instead of the bias current or average power. This may be the case when the DUT has poor dc stability or when the frequency response will be normalized to the response at a specific frequency (eliminating the coupling ratio). In this case equations 13 and 14 can be combined to give:

$$\mathcal{R}^2_{DUT}(f) = \frac{P_{rf_DUT}}{P_{M_REF}}\mathbb{R}^2(f)\beta^2, \qquad (15)$$

where β is the ratio of the optical power coupled to the reference arm to the power coupled to the test arm.

Results

When the frequency response of a communication analyzer module is measured, several uncertainties can corrupt the measured data:

- RIN(relative intensity noise) of the lasers
- Wavelength sensitivity of the photodiode receiver
- Harmonic distortions generated by the RF sources, the directly modulated laser, and the Mach-Zehnder modulator
- Jitter of the RF sources and the communication analyzer mainframe
- Finite display resolution of the communication analyzer.

The effect of RIN was negligible, with the detected RF signal about 30 dB above the noise level in the SONET test frequency range. To eliminate the wavelength sensitivity of the photodiode receivers, the wavelength of the lasers used was chosen to be 1.312 μm. To eliminate the harmonic distortion caused by the RF sources, the laser, and the Mach-Zehnder modulator, the amplitude of the RF signal was limited such that the second harmonic was kept below −30 dBc (30 dB below the carrier). The remaining sources of uncertainties were mainly the jitter and the finite display resolution of the communications analyzer.

To verify that a calibration can be transferred with acceptable accuracy, two swept-frequency ratio measurement systems based on Fig. 2 were built using two transfer standards. In the first verification, the frequency response of a transfer standard first calibrated by the Nd:YAG heterodyne system was measured by a swept-frequency ratio measurement system referenced to the second transfer standard. The difference, shown in Fig. 3, is excellent, especially considering that the two systems are working on totally different physical principles. The agreement was within ±0.05 dB to 7 GHz and within ±0.10 dB to 20 GHz. Since in this case the DUT is not a communications analyzer module, the measurement is not affected by the jitter and the finite resolution of the communications analyzer.

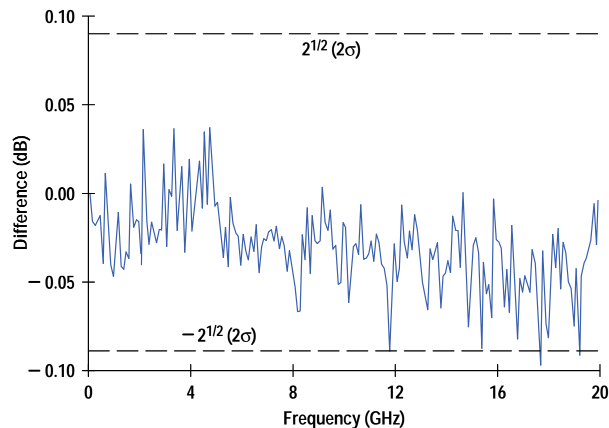


Fig. 3. *Measurement of a transfer standard consisting of a photodiode, a 3-dB attenuator, and an RF power sensor against a second transfer standard of the same type compared with NIST calibration. Uncertainties of the transfer standards alone are also shown.*

In the second verification, a communications analyzer plug-in module was measured on both swept-frequency systems. Each system used a different communications analyzer mainframe, which was triggered by the 10-MHz reference oscillator from the synthesized signal generator. The module's measured response was normalized so that the curves just fit inside the tolerance window specified by ITU-TS G.957 (see Fig. 4). The curves agree well within the G.957 specification. The response curves also agree well within the expected combined uncertainty of the two transfer standards below 2.5 GHz.

Above 2.5 GHz the difference slowly drifts upward, reaching 0.2 dB above 4 GHz. The difference between the two curves also has larger scatter above 2.5 GHz. This discrepancy above 2.5 GHz is attributed to the jitter and the finite display resolution of the communications analyzer mainframe. First, the communications analyzer was triggered directly from the synthesizer's 10-MHz reference oscillator, giving poor timing jitter at high frequencies. The trigger jitter combined with the signal averaging used to reduce the trace noise causes the measured amplitude to be smaller than the actual amplitude. The error caused by the jitter can be reduced by using an external trigger and can be eliminated by taking untriggered data. The finite display resolution (8 bits) caused additional error at high frequencies, where the signal amplitude is only 10% to 20% of the low-frequency amplitude.

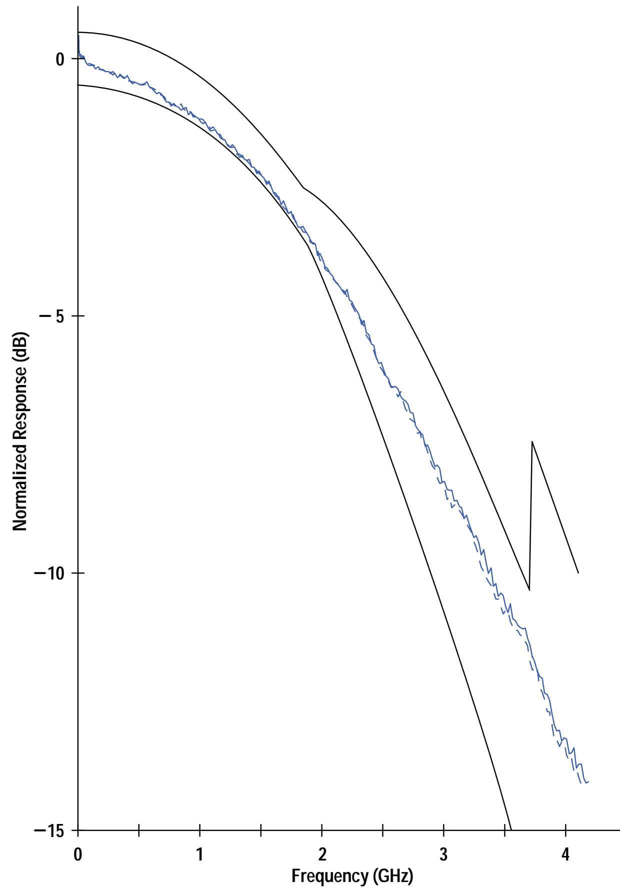


Fig. 4. *Measurement of a digital communications analyzer plug-in module with an STM-16/OC-48 filter on two different ratio measurement systems, using the same two transfer standards as in the experiment shown in Fig. 3, each transfer standard consisting of a photodiode, a 3-dB attenuator, and an RF power sensor. Tolerances given in ITU-TS G.957 are also shown.*

References

1. P.D. Hale, C.M. Wang, R. Park, and W.Y. Lau, "A Transfer Standard for Measuring Photoreceiver Frequency Response," *Journal of Lightwave Technology*, accepted for publication.
2. *Guidelines to the Expression of Uncertainty in Measurement*, ISO, Geneva, 1993.

Radially Staggered Bonding Technology

This new approach to fine-pitch integrated circuit bonding entails a new configuration of bonding pads on the die, dual-loop wire bonding, and a new leadframe design that minimizes wire lengths. The approach bypasses the usual obstacles to fine-pitch bonding that arise with the conventional in-line approach, thus providing appreciable die size and cost reductions with a minimal assembly cost penalty.

by Rajendra D. Pendse, Rita N. Horner, and Fan Kee Loh

Advances in silicon density have made it possible to reduce the core sizes of integrated circuit (IC) devices. However, concomitant reduction of I/O pad ***pitch*** (the pitch is typically defined as the repeat distance between adjacent I/O pads) has been hard to achieve because of packaging limitations. As a result, IC designs that are I/O intensive (so-called ***pad-limited*** designs) tend to have a die size that is significantly greater than the core size, leading to poor utilization of the silicon area. Appreciable savings in the form of greater numbers of die per wafer can be realized merely by reducing the I/O pad pitch and consequently the die size.

From the packaging standpoint, the reduction of I/O pad pitch requires improvements in the technologies used to physically interconnect the I/O pads to the package. While advanced packaging techniques do exist that permit such interconnection, such as ***flip chip*** and ***tape automated bonding*** (TAB), such techniques significantly increase the cost and complexity of the package. Wire bonding is by far the most prevalent interconnection technique in the mainstream industry-standard packages that are used for housing a large majority of ASIC and other IC devices. Quite understandably, therefore, there has been a considerable push to extend wire bonding technology to finer pitches. Some success in wire bonding pitch reduction has been achieved by the use of the so-called ***double-tiered*** package structure. However, once again, this approach increases cost and therefore is unsuitable for mainstream plastic packages that are based on leadframes with a single tier of bonding fingers.

Three broad factors are considered to be obstacles for the reduction of wire bonding pitch. First, the capillary, or the tool that carries the wire, physically occupies the space between adjacent wires, and therefore interferes with the previously bonded wire if the bond pad pitch approaches the capillary wall diameter (Fig. 1). There are physical limits to how much the

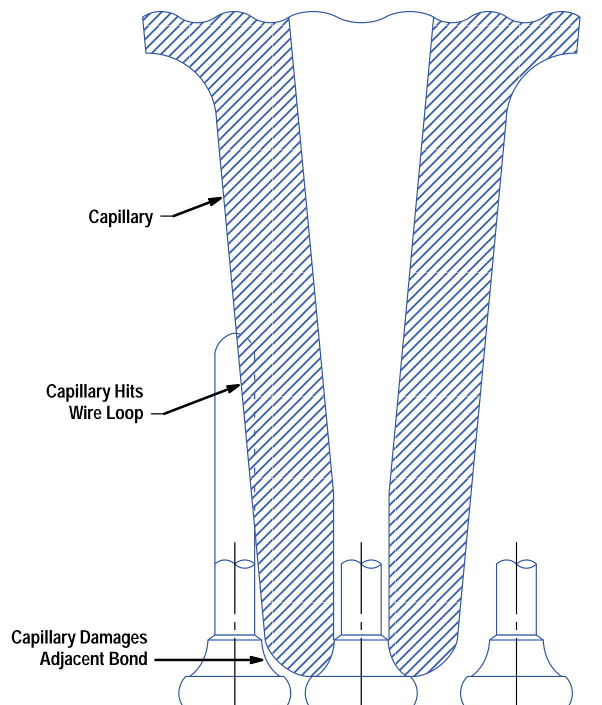


Fig. 1. *Capillary interference.*

capillary wall thickness can be reduced without risking frequent breakage during production runs, and this in turn places a limit on the pad pitch. Secondly, reduction of pitch requires a proportionate reduction in the size of the bond pads and higher bond placement accuracy for the wire bonding machine. This factor leads to significant loss in yield and reliability at finer pitches as a result of bonds that are not completely contained within the bond pad, often referred to as ***off-pad*** bonds. Third is the phenomenon of ***wire sweep*** in plastic packages. Plastic packages are typically fabricated by a process known as transfer molding. During the transfer molding process, liquid resin flowing into the mold cavities at elevated temperature (usually around 180°C) causes the bonding wires to be "swept" in the direction of resin flow (Fig. 2). This phenomenon causes adjacent wires to encroach upon each other or even touch, causing electrical shorts. When the pitch is reduced, two things happen: first, since the die gets smaller, wires become proportionately longer, and second, the wire spacing is reduced. Both of these factors accentuate wire sweep, often making the part unfeasible to manufacture.

Fig. 2. *X-ray micrograph of a molded package showing wire sweep. The curvature in the wire trajectories is induced by resin flow during soldering.*

In this paper, we present a new approach to reducing the effective wire bonding pitch that systematically surmounts the obstacles described above. We also present extensive assembly results on a test chip in a 208-pin PQFP package demonstrating the viability of this approach.

Concept of Radial Staggering

Staggering of bonding pads is a simple way to mitigate the problems of capillary interference and off-pad bond placement. A staggered configuration consists of a set of offset bond pads arranged in two rows as opposed to one row (Fig. 3). Staggering increases the direct distance between any two adjacent bond pads, allowing more room for the capillary to land without impinging on previously made bonds. It also allows the size of the bond pads to be significantly greater than the corresponding size for an inline configuration of the same effective pitch.

A configuration known as orthogonal staggering (Fig. 3b) has been previously employed in conjunction with more sophisticated double-tiered package structures. However, as depicted in Fig. 4, orthogonal staggering does not work with leadframe packages because of the geometric effect of fanout, which leads to severe wire encroachment, or even overlapping, in the corner regions. Fanout is a consequence of the fact that the pitch of bond pads on the die is typically much finer than the pitch of bond fingers on the leadframe.

Radial staggering is a geometric variation of orthogonal staggering in which overlapping of wire trajectories does not occur despite a fanout pattern of wire trajectories characteristic of leadframe packages.[1] This is illustrated in Fig. 3c and in Fig 5. As seen in Fig. 5, adjacent wire trajectories can be placed evenly by virtue of the radial staggering arrangement on the die. In the wire bonding sequence, the outer row of pads is bonded first, followed by the inner row of pads, to prevent any interference of the capillary with the loop portion of the wire trajectory.

Layout Method and Algorithms

A radially staggered bond pad pattern can be designed systematically using either a graphical method or closed-form mathematical expressions incorporated in the layout code. Typically, the layout is done for one eighth of the die, then mirrored about the y-axis to get the layout for one fourth of the die, followed by successive rotations of 90, 180, and 270 degrees about the center. The resulting pad arrangement is uniform and symmetric and can be used as a universal arrangement for standard pin counts. More customized arrangements are possible and can be readily developed based on the concepts illustrated here.

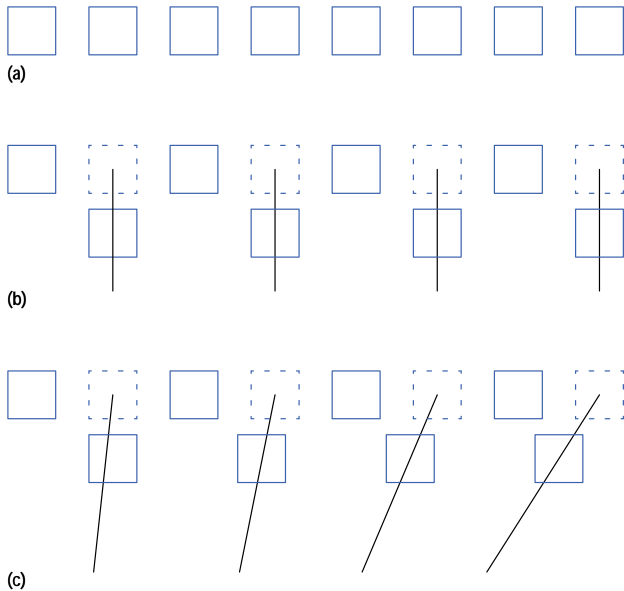


Fig. 3. *Illustration of the concept of staggering. (a) Conventional inline bond pads. (b) Orthogonally staggered bond pads. (c) Radially staggered bond pads.*

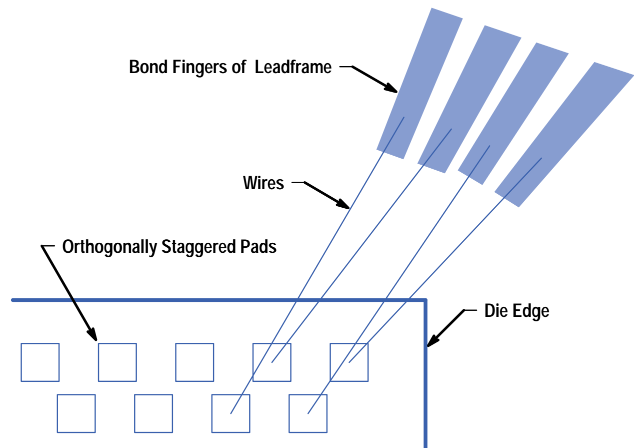


Fig. 4. *Illustration of wire encroachment when orthogonally staggered pads are used with leadframe packages.*

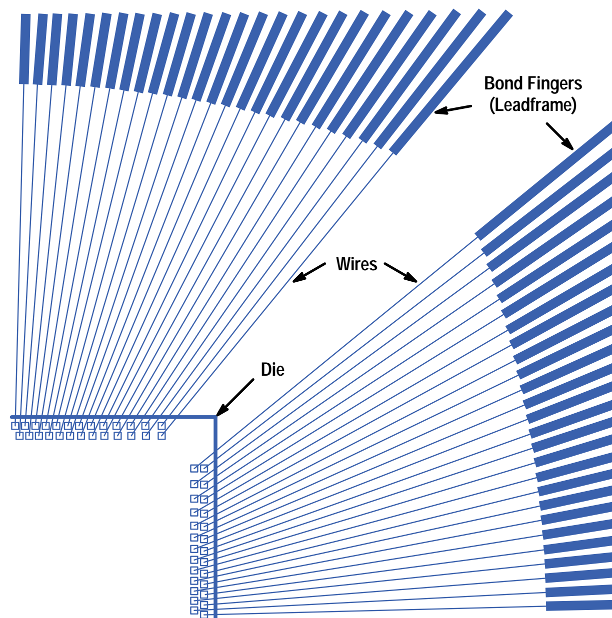


Fig. 5. *Radially staggered bond pad configuration for one quarter of a die.*

Graphical Method. The graphical method is illustrated in Fig. 6. The bond pads are first arranged at the desired effective pitch in a single row; this will be referred to as the *initial inline arrangement*. The effective pitch is usually selected so as to cause the pad ring to "hug" the IC core; see equation 8 below.

Next, a line running parallel to the die edge and representing the second row of bonds is constructed. Then, every other pad is shifted along a radial line emanating from a convergence point until it reaches the point of intersection of the radial line with the line representing the second row of bond pads.

The convergence point can in principle be chosen to be anywhere along the y-axis. In practice, the choice of the convergence point is governed by the resulting *angle of approach* of the bonding wire trajectory to the bonding finger and is a function of the leadframe design (see below).

As the convergence point is moved farther down the negative y-axis, the bond pad pattern approaches the orthogonally staggered configuration. In fact, the orthogonally staggered arrangement is a special case of radial staggering with the convergence point at negative infinity along the y-axis.

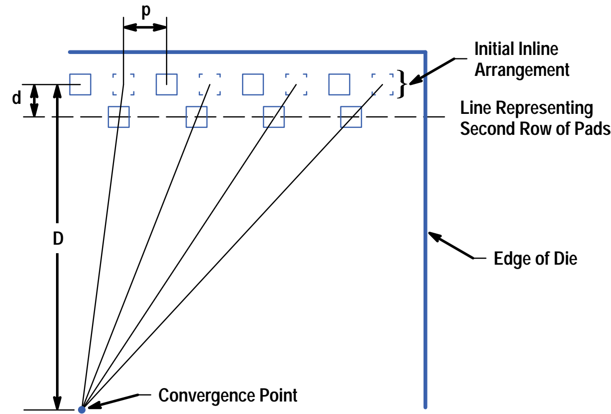


Fig. 6. *Graphical method for radial staggering.*

The value of the offset distance between the two rows of pads is a trade-off between the need to mitigate capillary interference and the need to minimize the consumption of silicon real estate. The offset should be minimized to save silicon real estate while maintaining sufficient clearance for the capillary. Typical values for the parameters described above will be presented later in this paper.

Analytical Method. An alternative to the graphical method for bond pad layout is to place the pads directly at Cartesian coordinate locations calculated mathematically using the convergence point as the origin. The expressions for such placement for one eighth of the die are given below (see Fig. 6). The analysis is divisible into two categories: the first category is for odd values of N/4 and the second is for even values of N/4, where N is the total pad count for the chip.

For odd values of N/4:

$$x(n) = (n-1)p \quad (1a)$$

$$y(n) = D \quad (1b)$$

for odd values of $n = 1, 3, \ldots, (N/4 - 1)/2$, and

$$x(n) = (n-1)(D/d-1)p \quad (2a)$$

$$y(n) = D - d \quad (2b)$$

for even values of $n = 2, 4, \ldots, (N/4 + 1)/2$.

For even values of N/4:

$$x(n) = c/2 + (n-1)p \quad (3a)$$

$$y(n) = D \quad (3b)$$

for odd values of $n = 1, 3, \ldots., N/8$ or $(N/8 - 1)$, and

$$x(n) = [c/2 + (n-1)p](D/d - 1) \quad (4a)$$

$$y(n) = D - d \quad (4b)$$

for even values of $n = 2, 4, \ldots, N/8$ or $(N/8 - 1)$.

In the above equations, n is the bond pad number (starting from the center and increasing towards the corner), p is the effective staggered pitch (repeat distance of the outer row of bond pads), d is the offset distance between the rows of pads, D is defined by $D = d + H + (N/4 + 3)p/2$, where H represents the pad height (to be described later), and c is the spacing between the two center pads on two sides of the y-axis (the value of c is chosen to be equal to the minimum inline pitch capability of the technology).

Once again, as described above for the graphical method, the layout for the entire die is obtained from the the layout for one eighth of the die by a combination of mirroring and rotation about the y-axis and the origin, respectively.

More General Layout Schemes

The layout methodology described above consists of two steps: placement of pads in an initial inline arrangement followed by radial shifting of every other pad towards the convergence point. The initial inline arrangement was taken to be at a uniform effective pitch, p. This need not be the case in general.

It is evident from Fig. 5 that with a uniform initial inline arrangement, the spacing between adjacent wires progressively decreases towards the corner. This geometric effect is particularly detrimental because wire sweep is typically most

pronounced in the corners. This occurs because in typical transfer molds, the resin flow direction is parallel to the diagonal of the chip and is therefore perpendicular to the corner wires.

An appealing choice of the initial inline arrangement would be one in which the pads are placed at a progressively increasing spacing, in a manner that would offset the progressively decreasing wire-to-wire spacing in moving from the center to the corner pads. Such a design would appreciably reduce wire sweep. A geometric progression algorithm has been developed for the initial inline arrangement. The code for layout of progressive pads is complex because closed-form mathematical expressions are not available. This is covered in detail elsewhere.[2]

Angle of Approach

We now briefly examine the criteria for the selection of the convergence point alluded to above. Fig. 7 depicts wire trajectories for two distinct bond pad layouts representing two extreme choices of convergence point. In both cases, the design of the leadframe is held constant. In Fig.7a, the convergence point is selected to be at the geometric center of the die and in Fig. 7b, it is selected to be a large distance away from the geometric center of the die along the negative y-axis. The angle between the wire trajectory and the bonding finger is defined as the *angle of approach*.

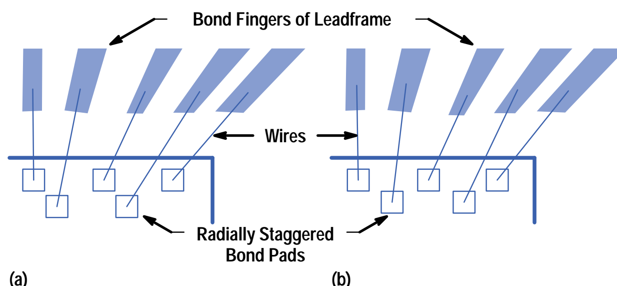


Fig. 7. *Effects on angle of approach of two different bond pad layouts corresponding to two distinct choices of convergence point. The angle of approach is the angle between the wire trajectory and the bonding finger. (a) Convergence point at the geometric center of the die. (b) Convergence point far from the geometric center along the negative y-axis.*

A large value of the angle of approach is likely to induce electrical shorts between a bonding wire and an adjacent bonding finger, leading to lower assembly yields. It is therefore customary to specify a maximum allowable value for the angle of approach. HP's manufacturing specification requires that the angle of approach always be lower than 10 degrees. The approach angles are unacceptably high for at least one wire in Fig. 7b, but are well within the specification in Fig. 7a.

In the more general case, the approach angles are also a function of the leadframe design. As an example, in Fig. 8, a leadframe design is shown that is different from the one shown in Fig. 7, but the choice of convergence point for Figs. 8a and 8b is identical to that in Figs. 7a and 7b, respectively, that is, the convergence point is at the center of the die for Fig. 8a and at a large distance along the negative y-axis for Fig. 8b. However, in contrast to the case shown in Fig. 7, for the case in Fig. 8 the approach angles are significantly lower for the choice of convergence point at a large distance along the negative y-axis than for the choice of convergence point at the center of the die. This is a consequence of the difference in the leadframe design between Figs. 7 and 8, and underscores the role of leadframe design in conjunction with the bond pad layout on the die in determining the approach angle in the general case.

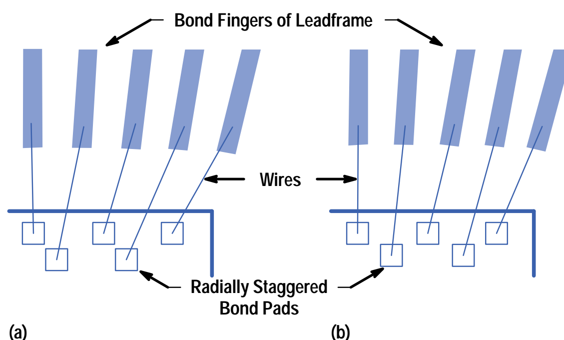


Fig. 8. *The same die pad layouts as in Fig. 7 but with a different leadframe design.*

The designs of typical leadframes used in real plastic packages are very close to the case depicted in Fig. 5, and therefore, low angles of approach are achievable by using the center of the die as the convergence point. This choice was made for the design of a test chip as discussed later.

As a side note, it should be stated that in certain instances, the design of bonding fingers and the corresponding choice of a convergence point for the radial staggering of bond pads on the die can be modified or customized to minimize wire length, and thus wire inductance, when electrical performance is critical. Such custom designs prove valuable in the case of high-performance packages such as ball-grid arrays (BGAs).

It can be shown that the theoretical minimum wire length with a zero angle of approach can be achieved by selecting the convergence point at a location along the y-axis that causes the subtended angle for the corner pad to be the inverse cosine of the ratio of the effective bond pad pitch on the die to the effective bond finger pitch on the package. The theoretical analysis of the general case has been performed,[2] but is beyond the scope of this paper. The author has developed an iterative custom layout algorithm for the coupled layout of radially staggered bond pads on the die and bond fingers on the package that minimizes wire lengths for BGA packages, which typically provide fine-pitch bond finger capability and require high electrical performance.[2]

Implications for Pad Circuitry Layout

The radially staggered bond pad configuration has one important consequence with regard to the layout of I/O pad circuitry, a point that is not initially obvious. The bond pad is typically an integral part of a unit structure known variously as the ***pad cell***, the ***pad buffer***, or the ***I/O pad***. We will refer to it as the pad cell. As illustrated schematically in Fig. 9, the pad cell consists of the I/O circuitry (such as pad drivers), the bond pad, and the ESD protection circuitry. In the conventional inline bond pad design, the pad cell is called from a library and placed automatically along the die perimeter at a repeat distance equal to the bond pad pitch. For an orthogonally staggered design, this procedure is minimally modified to include an x or y offset for every other pad cell, the value of the offset being equal to the spacing between the two rows of bond pads of the staggered configuration. However, if the bond pads have to be in a radially staggered configuration, each bond pad will not line up with the remaining portion of its associated pad cell (Fig. 10). This means that the conventional pad cell placement methodology has to be modified. It also means that an additional routing trace has to be added to connect the bond pad with its associated pad cell. These features are the subject of ***Article 7***.

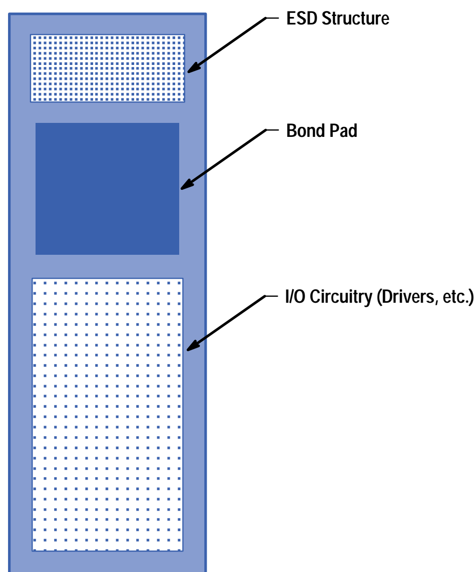


Fig. 9. *Structure of an I/O pad cell.*

Limitations

One limitation of the radially staggered design is that two rows of bond pads are required. The second row of pads uses up space and accrues a penalty in silicon area utilization. It is clear that in designs that are marginally pad-limited, a staggered design may not result in a net reduction of die size. In the following analysis, the important case of a marginally pad-limited design is quantified, and a conditional expression is derived that can be used to determine whether a die size reduction can be achieved by resorting to the radially staggered layout.

Fig. 11 is a schematic illustration of a pad-limited IC design. In this figure, the region defined as white space is unused silicon area resulting from the fact that the inner perimeter of the I/O pad ring falls outside the outer boundary of the core. This condition occurs because the repeat distance for the pad cells (i.e., the pad pitch) is not small enough to pull in the I/O pad ring so that it hugs the core. It can be shown that the size of the white space is independent of the pad height and is expressible as:

$$2W = P_i (N/4) - C \qquad (5)$$

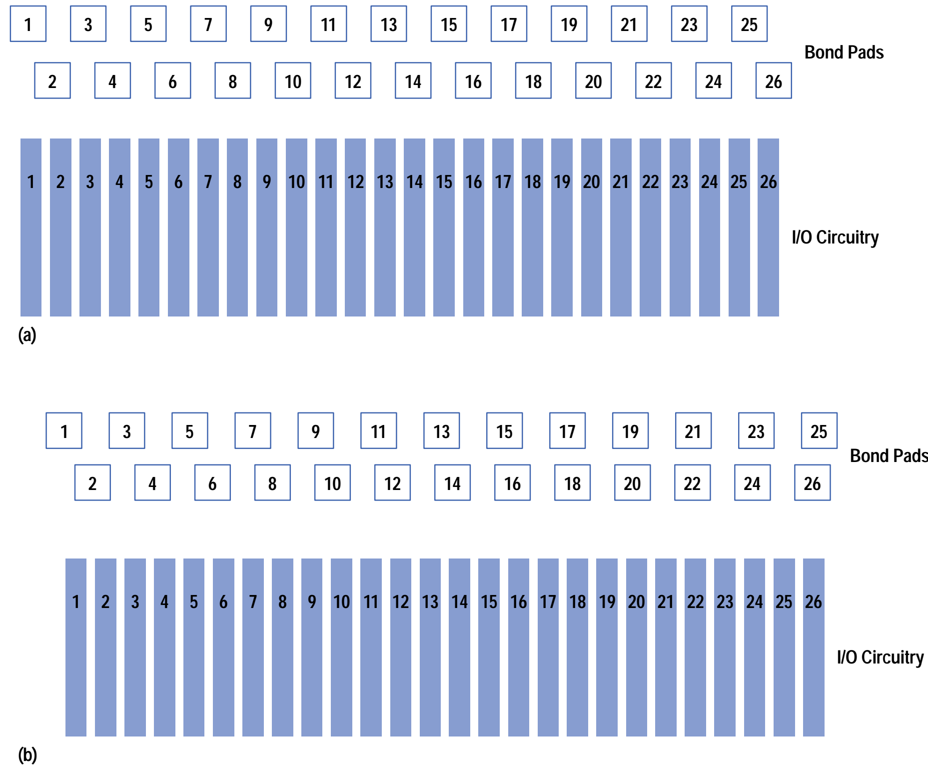


Fig. 10. *Illustration of bond pad shift caused by staggering. As a result of radial staggering, the bond pads do not line up with the associated I/O pad cells.*

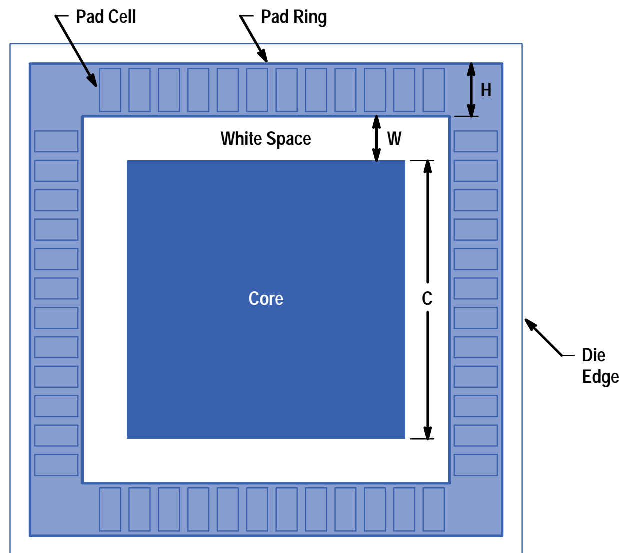


Fig. 11. *The presence of white space identifies a pad-limited IC design.*

where W is the size of the white space, P_i is the lowest qualified pitch (repeat distance of pads) for inline bonding, N is the pin count, and C is the size of the core.

A pad-limited design is defined as a design that has a non-negative white space, that is, $W > 0$, or, using equation 5,

$$P_i(N/4) > C. \quad (6)$$

While the effective repeat distance can be reduced using the radially staggered layout, additional space is required to accommodate the second row of pads, as well as to accommodate a possible increase in pad height necessitated by the narrowing of the pad cell. This additional space is equal to the sum of the row-to-row spacing, d, and the increase in pad height, ΔH, and can be pictured as a broadening of the pad ring. The size of the white space should be greater than the

broadening of the pad ring or the inner perimeter of the broadened pad ring will interfere with the core. This condition can be expressed as follows:

$$2d + 2\Delta H > P_i(N/4) - C. \qquad (7)$$

If equation 7 holds for a given IC design, the design is deemed marginally pad-limited and a radially staggered arrangement will not result in a die size reduction. However, if equation 7 does not hold, then the radially staggered layout will reduce the die size and the layout methodology described above can be used to perform the layout. In performing the radially staggered layout, the effective pitch for the initial inline arrangement should be selected so that the pad ring hugs the core. This condition can be derived from equation 6 by replacing the inequality with an equality:

$$P_{effective} = C(4/N). \qquad (8)$$

If the design is severely pad-limited, the value of effective pitch calculated from equation 8 may turn out to be smaller than the smallest effective pitch supported even by the radially staggered bonding capability. In this case, the smallest supported pitch should be used as the effective pitch.

A further point should be made regarding the implications of the radially staggered design for higher pin counts. In general, the relative value of the penalty in silicon area associated with the second row of pads is diminished at higher pin counts because of the increase in the nominal value of the die size. This pin count effect is contained in the parameter N in equations 6, 7, and 8. On the other hand, the relative area penalty is much greater at lower pin counts. This effect is fortuitous, since an increasing number of ASIC designs in the pin count range well above 208 are pad-limited. The higher-pin-count designs have a significantly higher average selling price and thus provide the greatest potential for cost reduction. Such designs are currently packaged in expensive double-tiered ceramic packages. In addition to the cost reduction resulting from the reduced die size, radial staggering technology provides the interesting additional potential for reducing the cost of the package itself by making it possible to design the ICs into cheaper packages based on a single layer of dense routing as opposed to multitiered structures. This can be achieved by the iterative coupled design procedure alluded to briefly in the foregoing section.[2]

Reduction of Wire Sweep

While radial staggering helps reduce the effective I/O pad pitch and therefore the die size, the length of wires increases proportionately, thus accentuating the wire sweep phenomenon. Table I illustrates typical wire lengths that result from reduction of die pad pitch. As a reference point, today's state of the art for transfer molding technology permits wires no longer than 4.5 mm (0.177 inch) to ensure 6σ manufacturing quality levels.

Table I
Wire Lengths for Different Die Pad Pitches and Leadframe Designs

Effective Die Pad Pitch (mm)	Longest Wire with Conventional Leadframe (mm (inch))	Longest Wire with New Leadframe Design (mm (inch))
0.110	4.3 (0.170)	3.8 (0.150)
0.090	5.1 (0.200)	4.4 (0.173)
0.070	5.5 (0.216)	4.9 (0.193)
0.050	6.0 (0.236)	5.5 (0.216)

The *dual-looping scheme* (Fig. 12) was developed to mitigate wire sweep. In this scheme, wire trajectories are placed in two different planes rather than in the same plane, thus creating a vertical gap in addition to the x-y separation between any two consecutive bond wires. Since wire displacements resulting from sweep typically occur within the same plane, the required displacement of a wire to cause shorting with an adjacent wire is effectively doubled compared to the case without dual looping, thereby considerably reducing the propensity for electrical shorts or leakage.

The separation distance between the two planes in Fig. 12 is nominally 0.125 mm (0.005 inch). Such a design is made possible by virtue of the loop height control features provided by state-of-the-art wire bonders. Loop heights are controllable to within ±0.025 mm (0.001 inch) of the nominal value. Typically, the loop height for the outer row of bond pads is chosen to be lower, and the outer row of pads is bonded first, followed by the inner row, to ensure that there is no interference between the capillary and the loop portion of the previously bonded wire. Since the loop heights and bonding sequence are programmable, the entire bonding operation can be performed in automatic mode without compromising the process throughput. The nominal loop heights used in this work were 0.400 mm (0.016 inch) for the inner row of pads and 0.275 mm (0.011 inch) for the outer row of pads.

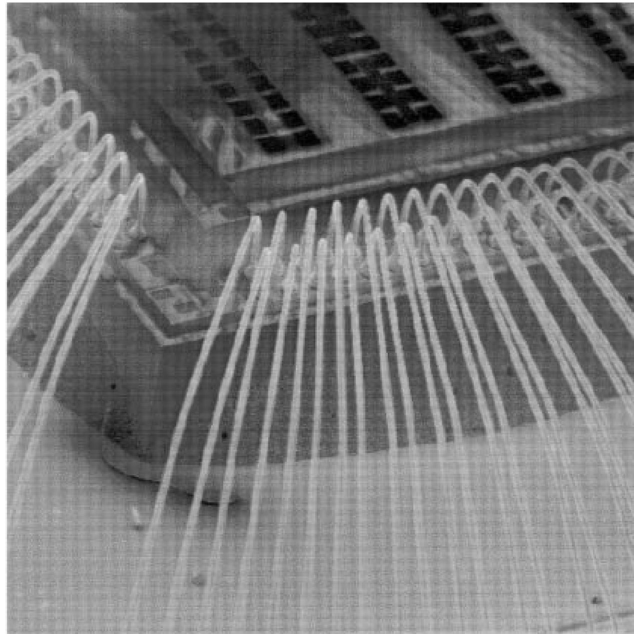


Fig. 12. *An SEM micrograph illustrating the dual-loop wire bonding scheme.*

New Leadframe Design

As is evident from Table I, a reduction of effective pitch to 0.090 mm would lead to a longest wire of approximately 5.1 mm (0.200 inch) if a leadframe with conventional design were used. We have developed a new leadframe design that has reduced the longest wire to 4.4 mm (0.173 inch), which was crucial in reducing wire sweep.

Conventional leadframes are designed using standard CAD tools that have built-in algorithms for bond finger layout. By studying the existing CAD tools, we learned that such tools likely employ a minimization of approach angles algorithm in conjunction with technology limit parameters (such as minimum bond finger width and space) to perform the layout of bonding fingers for the leadframe. The technology limit parameters are derived from the design rules supplied by the leadframe manufacturers. For example, typical design rules for leadframes produced by etching technology are 0.1 mm width and 0.1 mm space, and for leadframes made using stamping technology, they are 0.125 mm width and 0.125 mm space.

We observed that, from a mathematical standpoint, the minimization of wire lengths and the minimization of approach angles (in conjunction with the minimum width and space constraints) can be mutually conflicting conditions. This point is illustrated in Fig. 13. In Fig. 13a, the angles of approach have been minimized (they are zero degrees for all of the bond wires), whereas in Fig. 13b, the wire lengths have been minimized, but the approach angles are nonzero, and are as high as 30 degrees for some wires. While the design in Fig. 13b has appreciably shorter wires, it is unfeasible to manufacture because it violates the specification for the maximum angle of approach of 10 degrees.

After pursuing a theoretical analysis, we arrived at a straightforward (perhaps excruciatingly simple!) graphical method for the layout of bond fingers that would minimize the length of bond wires. The basis for this method is the theoretical condition that the bond fingers can be brought in closest to the center when the minimum widths and spaces reach the design rule limit values simultaneously at the tips of all bond fingers. It can be shown that this condition is attained by laying out the bond fingers so that they subtend equal angles with respect to a convergence point.[2] This latter fact formed the basis for a simple graphical method for performing the bond finger layout. Based on this simple algorithm, a leadframe layout was performed for a 208-pin device and a bonding diagram was generated for a presumed die with radially staggered bond pads at 0.090-mm effective pitch (e.g., see Fig. 5). The approach angles were calculated for this case, and the maximum value was found to be seven degrees, well below the manufacturing specification of 10 degrees. More important, the longest wire could be reduced to 4.4 mm (0.173 inch) compared to approximately 5.1 mm (0.200 inch) for a conventional leadframe.

While the methodology has been illustrated for leadframes, it is more generally applicable to the design of other package substrates, notably high-performance ball-grid arrays. The technology limit design rules for these substrates are more aggressive (e.g., 0.05 mm width and 0.05 mm space), but significantly shorter wires are required to minimize parasitic inductance. In the above analysis, the minimization of wire length algorithm was used in place of the minimization of angle of approach algorithm, and the resulting angles of approach (verified afterwards) fortuitously fell within the 10-degree limit. This fortuitous condition does not always occur, especially in the regime of very short wire lengths and double-tiered package structures. Therefore, a systematic iterative procedure is required that judiciously employs both algorithms and involves a coupled design of the radially staggered layout on the die and the bond finger layout on the package.[2]

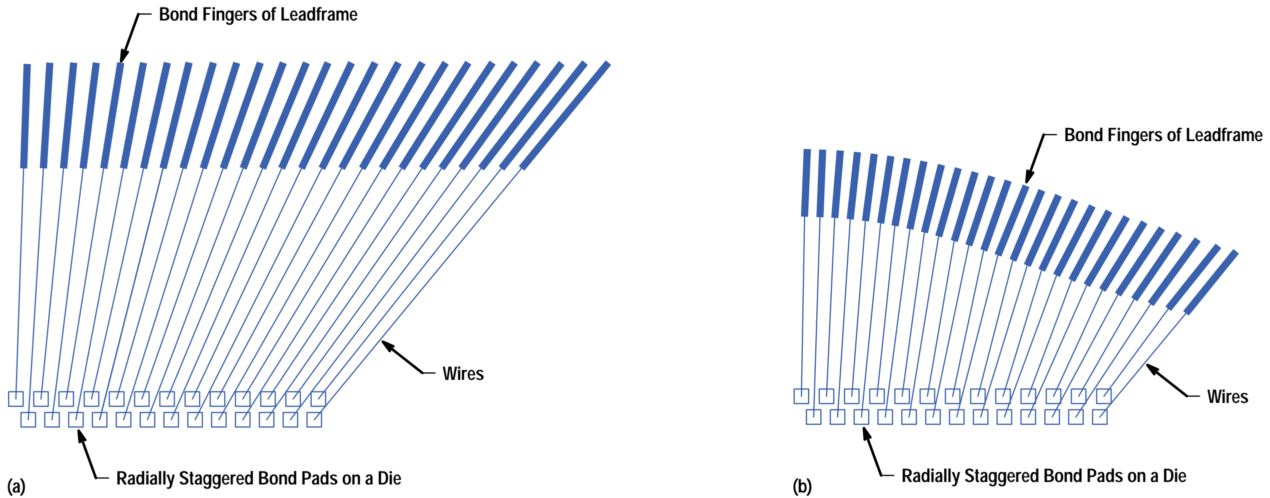


Fig. 13. *Role of leadframe design in angle of approach and wire length reduction. (a) The angles of approach have been minimized. (b) The wire lengths have been minimized. In both cases, the leadframe design rules (minimum width and spacing of leads) are the same.*

Assembly Evaluations

The three key elements of the fine-pitch bonding solution described above are radially staggered bond pads, dual-loop bonding, and a new leadframe design. Extensive assembly and testing were performed to establish the feasibility of these elements. A 208-pin plastic package was used as the test vehicle.

A test chip with 208 pads (N = 208) and having a radially staggered pad configuration was designed and fabricated. The physical structure conformed to standard design rules for the current generation of the CMOS process at HP's Corvallis facility. The effective pitch, p, was 0.0889 mm, the offset distance between the inner and outer row of pads, d, was 0.140 mm, and the pad height, H, was 0.660 mm. Since N/4 is even, equations 3 and 4 were applied with a value of c = 0.110 mm, based on the lowest inline pitch capability available at the time. Using the parameters above, the die size worked out to be 6.02 mm. Details of the chip layout can be found in the ***Article 7***.

Daisy chains were incorporated in the design to check for leakage current between adjacent bond pads. This feature was motivated by the need to test the susceptibility to wire sweep, which is known to manifest itself as an electrical leakage between adjacent bond wires (the leakage at every bond wire can be monitored externally using the package pins).

In addition, a 208-pin leadframe was designed and fabricated by the etching method using the principle of minimization of wire length. The leadframe was fabricated by an external leadframe supplier using standard design rules and production methods.

Assembly and Test Results

The die was assembled in a 208-pin PQFP package at HP's production facility in Singapore. A dual-loop bonding process was developed and the dual looping was performed in automatic mode on production wire bonders. A simple electrical test program was used to perform package testing on a production tester to check for open bonds (continuity) and electrical leakage between adjacent pins. In addition, transmission X-ray analysis was done on finished packages to check for physical evidence of wire sweep. A final test yield exceeding 97% on two independent assembly lots was set as the criterion for acceptability of the technology in the manufacturing environment, based on inputs by the production group.

Results of the assembly evaluations are summarized in Table II. A micrograph of the dual-loop wire bonds taken before the molding operation is shown in Fig. 14. A typical transmission X-ray image of the finished package is presented in Fig. 15. While the results in Table II are considered to be well within the acceptable range, failure analysis was performed on the electrical failures from lot 1 to determine if the failures were in any way related to the fine-pitch bonding design. The details of the analysis are not reported here; it suffices to say that electrical failures were found to be unrelated to the fine-pitch bonding design and are presumed to be defects at the die level (these defects escaped, since no wafer-level testing was done on these lots), implying an effective 100% yield from the standpoint of fine-pitch bonding.

Table II
Assembly Parameters and Yields

	Lot 1	Lot 2
Number of Units Assembled	297	185
Wire Diameter, mm (inch)	0.033 (0.0013)	0.033 (0.0013)
Maximum Wire Length, mm (inch)	4.5 (0.177)	4.5 (0.177)
Nominal Loop Height for Inner/Outer Rows, mm (inch)	0.406 (0.016)/ 0.279 (0.011)	0.406 (0.016)/ 0.279 (0.011)
Final Test Yield, %	99.3	98.4

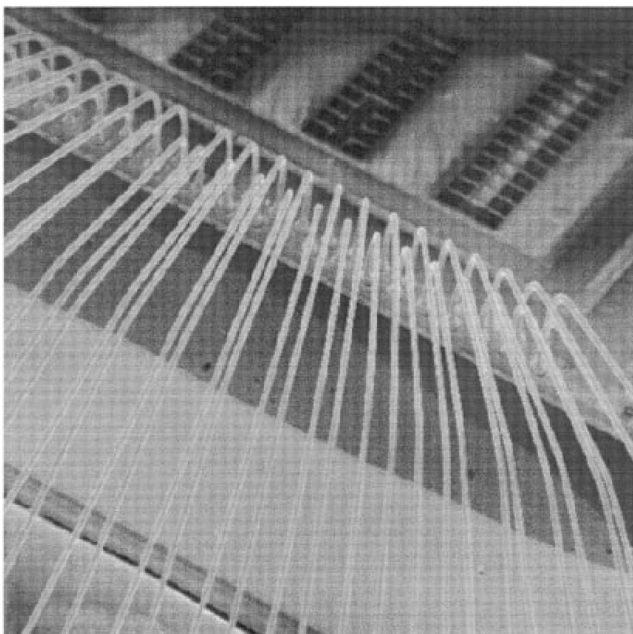


Fig. 14. *Typical dual-loop wire trajectories before molding.*

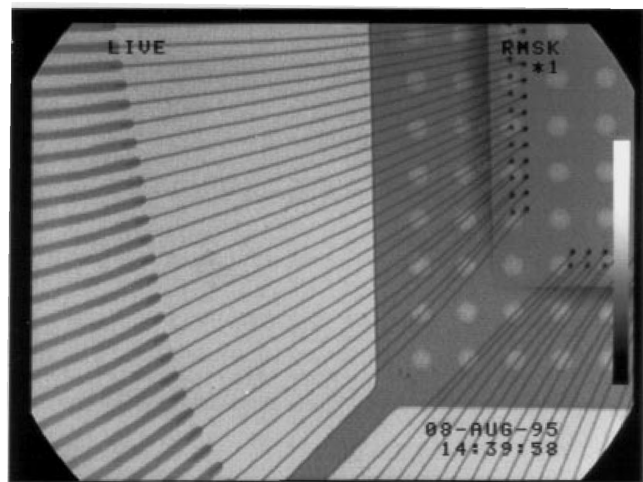


Fig. 15. *X-ray micrograph of a molded package (corner region).*

It should be noted that the wire sweep pattern depicted in Fig. 15 is typical of production packages and shows no evidence of wire encroachment or overlapping. However, in view of the dual-loop structure, it could be argued that there will be no electrical leakage despite any overlapping of adjacent wires, since the wires are in fact spaced in the z direction. This raises the interesting possibility of potentially circumventing the effects of wire sweep altogether, especially when the technology has to be extended to much finer pitches. Therefore, we studied the effects of wire overlapping as seen in transmission X-ray observations on electrical leakage between the overlapping wires. Fig. 16 is an X-ray micrograph of a package in which significant wire movement has been induced, leading to overlapping of adjacent wires in the corner region. Electrical leakage measurements were made on ten sets of overlapping wires from three different packages. In all cases, no leakage was detected up to a limit of 10 megohms. It can be inferred from this preliminary evaluation that the 0.125 mm (0.005 inch) of z separation in the dual-loop design is perhaps adequate to prevent leakage in overlapped wires. A more detailed study is required to characterize this result fully.

Extendability to Finer Pitches

The migration of radially staggered bonding technology to finer pitches (below the nominal 0.090-mm effective pitch demonstrated in this study) is significantly easier with radially staggered bonding than with conventional inline bonding. This point is illustrated in Table III for the case of 0.070-mm effective pitch.

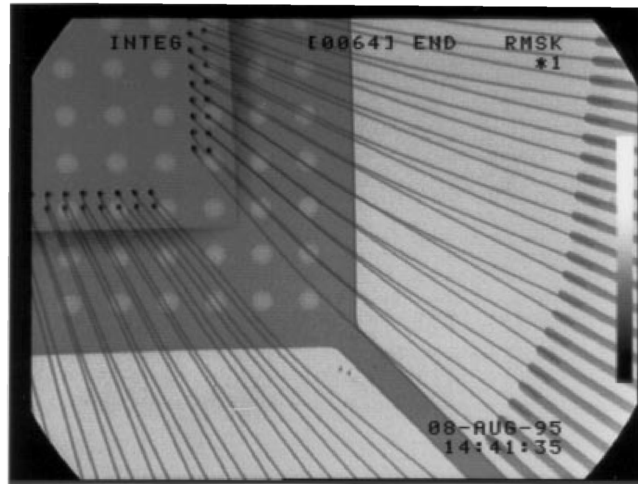


Fig. 16. *X-ray micrograph of a molded package with deliberately induced wire movement leading to the appearance of overlapped wires when viewed from above.*

Table III
Comparison of Wire Bonding Parameters for Radially Staggered and Inline Configurations for 0.070-mm Effective Pitch

	Inline	Radially Staggered
Pad-to-Pad Distance, mm	0.070	0.095
Bond Pad Opening, mm	0.060	0.075
Wire Diameter, mm (inch)	0.025 (0.001)	0.033 (0.0013)
Free-Air Ball Size, mm	0.050	0.070
Capillary Wall Thickness, mm	0.030	0.045

It should be noted that reduction of the bond pad opening to 0.060 mm and the concomitant reduction in free-air ball size* to 0.050 mm required for the inline case translate to ball size control and bond placement accuracy levels that are beyond the capabilities of current wire bonding equipment, while the corresponding values for the radially staggered configuration are well within the range of equipment capabilities. Additionally, it can be shown using finite element modeling that the reduction of wire diameter from 0.033 mm to 0.025 mm translates to an approximately twofold increase in wire sweep. Finally, the reduction in capillary wall thickness will reduce the capillary life by approximately a factor of two. Thus it is clear that, in general, a finer effective pitch is achievable using the radially staggered configuration with considerably relaxed wire bonding design rules compared to the corresponding inline case.

Despite these points, a few important limitations come into play when finer-pitch extensions are considered, even with the radially staggered bonding configuration. First, the reduced die sizes resulting from finer effective pitches lead to proportionately longer bonding wires. This point was illustrated in Table I. As an example, the wire length of 5.5 mm (0.216 inch) resulting from a presumed 0.050-mm effective pitch design is considered unfeasible to manufacture because of wire sweep, based on current process specifications. It is possible that the dual-looping scheme may obviate this problem, but this needs to be evaluated. It is also possible to reduce the bond finger pitch of the leadframe (and thus the wire length) by using interposers,** but the cost and technical feasibility of this approach is so far unproven.

A second limitation is that the routing requirements for interconnection of the bond pads to the I/O circuitry and ESD structures are proportionately tighter at finer pitches and may impose some additional restrictions (see ***Article 7***).

* A typical thermosonic gold wire bond consists of a ball bond at one end and a stitch bond at the other end. "Ball size" refers to the size of the ball that makes up the ball bond, after the ball is attached to the chip. "Free-air ball size" refers to the size of the ball as soon as it is formed, before it is attached to the chip.

** An interposer is a small circuit card that is placed between the die and the leadframe. It contains a pattern of traces that fan out from a fine pitch at the die end to a coarse pitch at the leadframe end.

A third limitation, as illustrated in Table III, is an overall tightening of wire bonding design rules dictated by reductions in pitch. In addition to the burden that this imposes on the capability of wire bonding and molding equipment, it also necessitates the development of new metrology tools and bond quality standards. Bond quality standards can only be set by a careful study of the impact of new wire bonding design rules on reliability. As an example, the smaller ball size, the finer wire diameter, and the presence of off-pad bonds may impact reliability. New specifications for these parameters can be set after the effect on reliability is well-understood. This significant effort has been recently undertaken by SEMATECH, a consortium of semiconductor companies.

Conclusions

A new approach to the reduction of wire bonding pitch is presented that entails the use of two rows of radially staggered bond pads on the die, as opposed to the conventional inline arrangement. We have combined the radial staggering methodology with dual-loop bonding and a new leadframe design, forming an integrated solution for reducing the effective wire bonding pitch and thus the die size of pad-limited IC devices. The approach has been qualified by extensive assembly evaluations of a test device in the production environment. The approach is considered advantageous over the conventional approach because it is implemented with significantly relaxed design rules and therefore with minimal assembly cost penalty.

Acknowledgments

The authors wish to thank Paul Van Loan, Chong Num-Kwee, Chong Chew-Wah, Steve Ratner, and Lynn Roylance for their technical inputs, support, and encouragement throughout the course of this project. We are also indebted to the technicians and operators whose diligence made this work a reality.

References

1. R. Pendse, et al, "Die Size Reduction through Finer-Pitch Bonding," *Proceedings of the HP Design Technology Conference*, 1994, pp. 281-294.
2. R. Pendse, "Algorithms for Optimal Design of High-Pin-Count Packages Using Radially Staggered Configurations," *paper in preparation*.

Implementation of Pad Circuitry for Radially Staggered Bond Pad Arrangements

One approach to pushing the limits of wire bonding pitch in IC packages is to use two rows of radially staggered bond pads. This paper discusses the design of pad circuitry to mesh with the radially staggered bond pad arrangement. A test chip that incorporates suitable test structures was designed, fabricated, packaged and tested to verify the viability of the approach.

by Rita N. Horner, Rajendra D. Pendse, and Fan Kee Loh

Assembly and packaging technologies are major contributors to the success of integrated circuit manufacturing. With increases in silicon density, chip core sizes are shrinking with the minimum transistor size. However, I/O pad circuitry and size are not shrinking relative to core size because of packaging limitations such as capillary interference, long wires, wire sweep problems,† and corner crowding. There are a few workaround techniques such as wedge-wedge bonding and double-tier bonding which are available for the more expensive packages like PGAs. However, these techniques are unsuitable for standard plastic packages. There is need for a better, more reliable, and more cost-effective solution to the bonding pad pitch problem in the near future as more IC chips are becoming pad-limited.

The pad-to-pad repeat distance on a chip, referred to as wirebond pitch, is often the factor that limits both the amount by which the chip size can be reduced and the increase of I/O density, thereby reducing the efficiency of silicon area utilization. Current practice allows minimum straight-line pitches to be in the range of 100 to 125 micrometers. In ***Article 6***, one approach for decreasing the wire pitch in IC packages is discussed. This approach entails the use of two radially staggered rows of pads on the chip periphery as opposed to the more conventional single-row, inline arrangement.[1,2] The bond pads are arranged to ensure no overlapping of bonding wire trajectories, even when conventional leadframes are used for the package. The radially staggered arrangement of bond pads allows uniform placement of wire trajectories despite the geometric fan-out from the die bonding pads to the leadframe bonding fingers. This approach is advantageous because it circumvents the common obstacles to fine-pitch bonding, such as capillary interference, bond placement accuracy, and wire size reduction, which have heretofore proved to be insurmountable.

This paper summarizes the main features of the radially staggered approach and discusses its implications for pad circuitry design. The implementation of radially staggered bond technology presents technical challenges in the area of pad design and layout such as the pad placement scheme, routing from bond pad to I/O circuitry or to power and ground rings, and the influence of such routing on the functionality and performance of the I/O circuitry and the ESD performance.

A test chip that incorporates suitable test structures to address the technical obstacles was designed and fabricated in the HP CMOS14TB process. The chips were packaged and tested to verify the viability of the approach. The verification exercise was performed for the case of 88.9-micrometer effective pitch for a 208-pin PQFP. A paper study was also done for the extension to 70- and 50-micrometer effective pitches. The results of this study will be summarized here.

Bond Pad Arrangement

In a radially staggered bond pad arrangement, every other pad is moved inward in the radial direction to form the second tier, as shown in Fig. 1. The radius used is from the center of the die and is dependent on the height of the I/O circuitry and the total number of pins. The pads are placed in a single row using 88.9-micrometer pitch, and then every other pad is moved inward in the radial direction with respect to the center of the die to form the inner row of pads. In the case of a 208-pin die with an I/O height of 358.9 micrometers, the inner-row bond pad pitch is 84.25 micrometers. Depending on the total number of pad openings and the total I/O height, the inner-row bond pad pitch will vary by a small amount. In the sample bond pad connection shown in Fig. 2, the ESD circuit is placed on the opposite side of the bond pad from the I/O circuitry. The passivation openings are 90 micrometers wide and the metal overlay for the bond pad openings is 4.8 micrometers. The ESD circuitry has an 88.9-micrometer pitch and 72.5-micrometer height. The outer-row-to-inner-row spacing is 140 micrometers. The outer-row bond pad pitch is 177.8 micrometers and the inner-row bond pad pitch is 168.95 micrometers.

† Wire sweep problems are created when liquid resin is forced into the plastic mold cavity. Wires can be moved closer together by the resin, causing possible shorts.

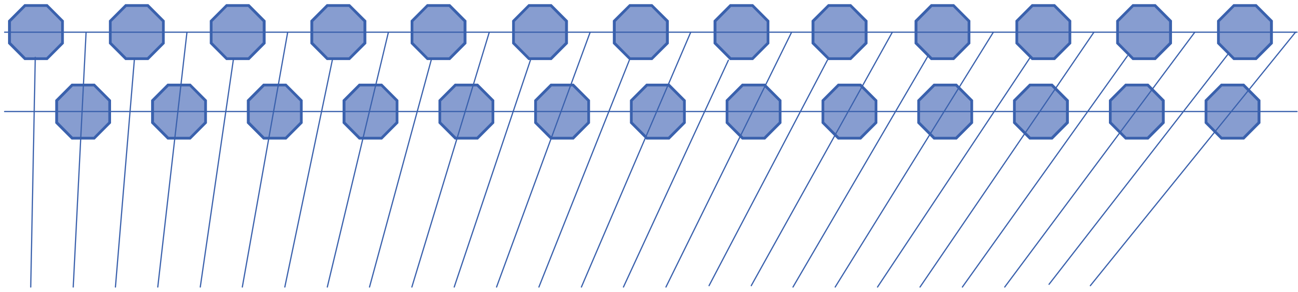


Fig. 1. *Radially staggered bond pad configuration.*

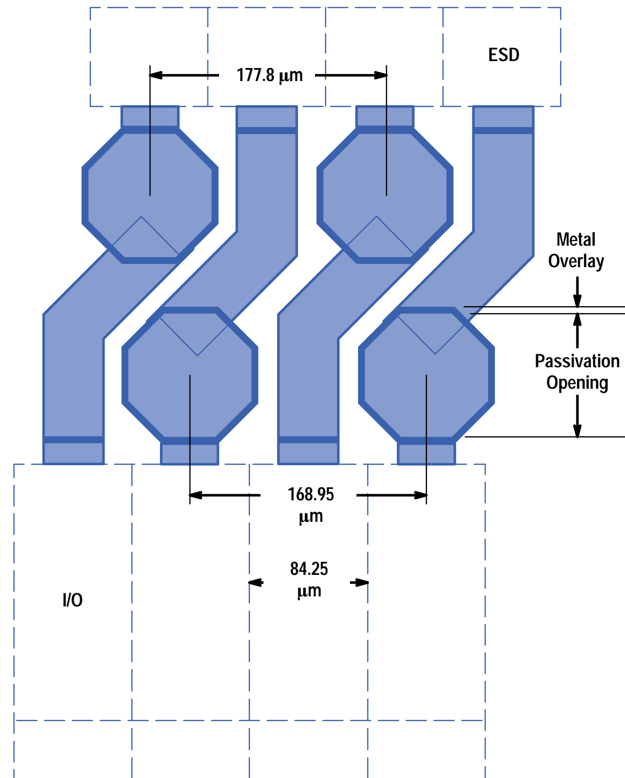


Fig. 2. *Sample bond pad connection.*

Square or rectangular bond pad openings with small amounts of chamfering would limit the maximum bus width for connection to bond pads. Therefore, octagonal bond pad openings with an internal radius of 90 micrometers are used. The octagonal bond pads allow wider buses for the bond pad connections to ESD and I/O circuitry. The maximum metal bus width used for the bond-pad-to-I/O or bond-pad-to-ESD connections for 208-pin die in the CMOS14TB process is 44 micrometers. The maximum interconnect metal bus width is dependent on the process design rules for passivation opening spacing to unrelated metal, minimum metal-to-metal spacing, and total number of pads. As shown in Fig. 3, the spacing between the adjacent interconnect buses gets smaller for the corner pad openings. As the number of pins increases, the inner-row pad pitch increases, while the interconnect metal spacing of the cornermost pads decreases at a faster rate. As a result, 44-micrometer metal width would be too wide in a 240-pin die. In a 100-pin or 196-pin die, the inner row pad pitches are smaller than in a 208-pin die, but because of the absence of the most extreme corner bond pads, three to six micrometer wider metal buses can be used for metal bonding pad connections to ESD or I/O circuitry.

Design Challenges

Routing and Bond Pad Placement. The longest bus connection to the outer-row bond pads from the I/O circuitry is about 197 micrometers and the shortest connection is 155 micrometers. To reduce the interconnect bus resistance to less than 0.1 ohm, all metal layers are used in parallel for the bond pad connection to the I/O or ESD circuitry.

Using the staggered bond pad configuration requires that the I/O and ESD circuitry be laid out in a narrower pitch than the existing inline configuration. It also requires special placement of bond pads and possible manual connection to bond pads. As this configuration becomes more standard, tools may be developed for automatic placement of the bond pads and

automatic bond pad interconnection to I/O and ESD circuitry. A script to perform the bond pad placement of one eighth of the chip was written in the HP ChipBuster layout editor's scripting language. However, since the available autoroute tools do not allow nonorthogonal connections, the connections to the bond pads were made manually.

Long Metal Interconnect Issues. Since all three metal layers are paralleled for the bond pad connection, the maximum interconnect bus resistance to I/O or ESD circuitry is less than 0.1 ohm for the 44-micrometer metal widths in this process. Because of this low resistance, there is no restriction on the choice of the bond pad connection location either for the supply signals or for the different I/O circuits. The 44-micrometer bus is also adequate for the voltage drops during an ESD event. Because the substrate capacitance is large compared to the mutual capacitance of the metal interconnects between neighboring signals, minimum design rules can be used for metal-to-metal spacing of the interconnecting buses without undue concern regarding cross talk.

Placing the ESD circuitry on the die boundary side of the bond pads allows a wider metal connection to the substrate ground, which has less resistance and a smaller voltage drop during an ESD event.

Latchup was found to present no problem as long as the process design guidelines were followed.

Inline versus Two-Tiered Design. In a conventional single-row inline bond pad arrangement, the minimum I/O ring size is a function of the minimum pad pitch, the I/O and ESD circuitry height including the required spacing between them, and the total pin count. The width of one side of a chip can be approximated by the following formula:

$$\text{side width} = (\text{number of pads on the side}) \times (\text{minimum I/O pitch}) + 2 \times (\text{I/O height}). \quad (1)$$

Therefore, for a 208-pin chip with 110-micrometer pad pitch when the I/O and ESD structures heights are about 500 micrometers, the minimum chip size will be 6.72×6.72 mm^2.

In a staggered pad configuration the I/O pitch is reduced to 84.25 micrometers. Because of the two rows of bond pad openings and the narrower I/O pitch, the total block height is increased to 660 micrometers. Therefore, the minimum chip size will be about 5.70×5.70 mm^2. This chip will be approximately 1 mm smaller on each side than a chip with an inline bond pad configuration of 110-micrometer pad pitch even though the pad height has increased by 160 micrometers.

The only differences between the inline bond pad arrangement and the staggered pad arrangement are the increase in the I/O and ESD circuitry heights, the I/O pitch, and the extra 140-micrometer height in the bonding area. Therefore, using equation 1, we can derive the following conditional expression for a square die:

$$[(\text{number of pads})/8] \times [(\text{inline I/O pitch}) - (\text{staggered I/O pitch})] < [(\text{I/O height staggered}) - (\text{I/O height inline}) + 140].$$

If this inequality is satisfied, it is a better area trade-off to use the staggered pad configuration rather than the inline bond pad placement. If the two sides of the inequality are equal, then there is no area benefit from using the staggered bond pad configuration.

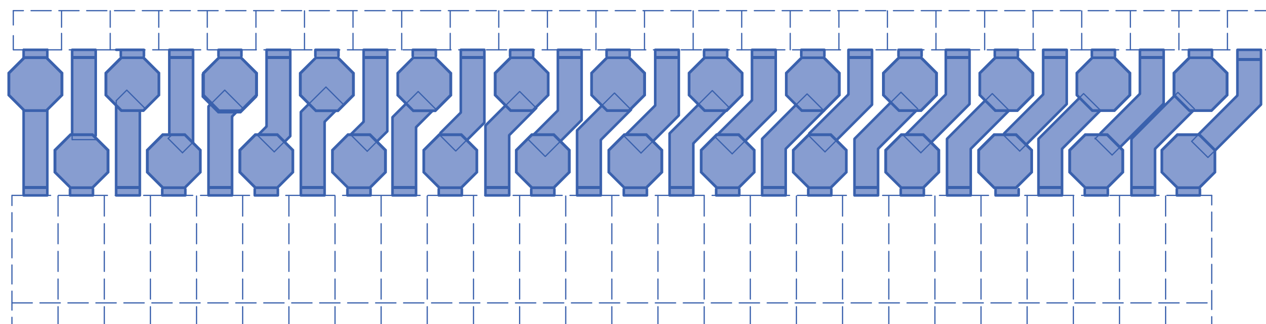


Fig. 3. *One eighth of a 208-pin I/O ring.*

Test Chip Architectures

Two versions of a 208-pin test die were designed and fabricated with suitable test structures. One version of the die was mainly for the reliability qualification of the package. The second version of the die was designed for verification of the functionality and performance of the I/O and ESD circuitry in the new package with the new bond pad architecture.

After considering the voltage drop because of the extra metal bus connections and the pad height, and calculating the optimal I/O pitch and optimal power and ground bus widths, the I/O and ESD circuitries were laid out in 84.25-micrometer and 88.9-micrometer pitches respectively, with small modifications from the inline I/O designs. New support pads such as corner pads and power and ground pads were designed to fit this new I/O ring structure.

As shown in Fig. 4, the bond pad placement on the two halves of each of the die sides is symmetrical around the center of each side. The bond pad placement pitch for each row is constant except in the center of each side. The two center bond pads in the outer row are placed with 110-micrometer pitch because of the minimum inline pad pitch limitations. Effectively, the two center bond pads in the inner rows have 278.9-micrometer pitch.

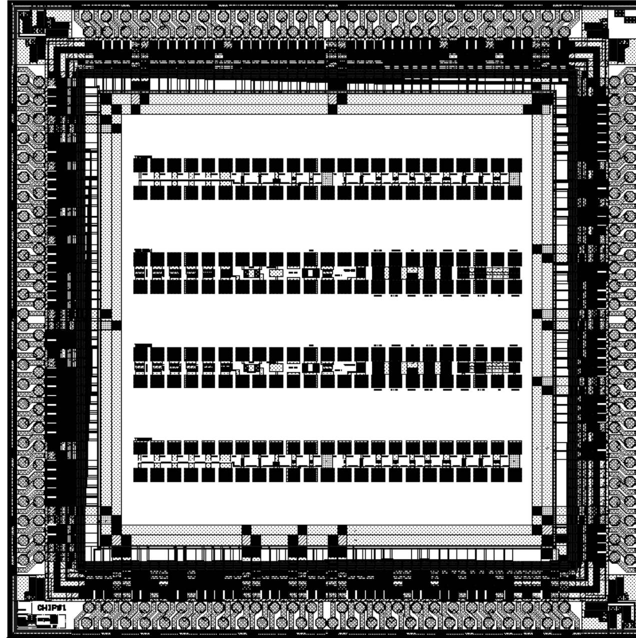


Fig. 4. *Experimental chip fabricated to test the I/O and ESD performance.*

The test die were fabricated and tested for package qualification and I/O and ESD performance. Over 500 parts were packaged and tested. The package qualification yield was over 98% and the parts passed 3500V HBM (human body model) ESD stress.

Extendability Study

As 90-micrometer inline bond pad pitch with minimum passivation opening of 75 micrometers becomes feasible, the radially staggered pad pitch can be further reduced to 70-micrometer and later 50-micrometer effective pitch. These narrower pitches create a new set of issues for consideration. As the bond pad pitches are reduced, the I/O and ESD pitches are reduced as well. The finer pad pitch reduces the interconnect metal bus widths, which would result in higher bus resistance. The I/O circuitry and ESD circuitry would increase in height at a much higher rate than the 110-to-88.9-micrometer conversions, since the smaller I/O width limits the number of metal tracks that can be used for circuitry interconnects within the cell. In the case of the 100-to-84.25-micrometer I/O pitch conversion, the I/O height was only increased by about 20 micrometers. However, in the case of 70-micrometer bond pad pitch, when the I/O pitch will be about 65 micrometers, the I/O height increases from 50 to 100 micrometers depending on the different I/O functionalities. Even with this amount of height increase in I/O, the new 70-micrometer staggered pad die would be about 1 mm smaller on each side than the inline 90-micrometer-pitch die. It can be predicted that the 50-micrometer pitch would require a much larger increase in the I/O height and may face circuit layout limitations for metal interconnect in a three-layer-metal process.

One solution to this problem would be to use processes with four or more metal layers. This will minimize the resistance in the metal interconnect buses from the bond pads to I/O or ESD circuitry, and at the same time will reduce the I/O and ESD circuitry height and possibly minimize the internal cell routing limitations that would be present in the case of 50-micrometer effective pad pitches, at the expense of a more expensive process.

Conclusion

This paper has presented a description of a new methodology for the implementation of radially staggered bonding technology from the standpoint of I/O pad circuit and ESD structure design. The algorithms by which the methodology can be implemented were presented. The issues of layout, placement, and routing for both the present design and the future migration to finer pitch were discussed.

This wire bonding solution was engineered with the goal of achieving die size reduction while minimizing the impact on cost and manufacturability. As such, it is believed that the scheme offers a significant cost reduction opportunity on pad-limited IC designs.

Acknowledgments

The authors wish to thank Alice Aplin and Bob Warren for I/O pads and chip layout, and Ed Chen, Chong Num-Kwee, Steve Ratner, Phil Ritchey, Lynn Roylance, Paul Van Loan, and Lim Chong Yong for their encouragement and support. They are also indebted to Carlos Diaz, Jim Eaton, Larry Lin, many people at the HP California Design Center, and the production staff at Integrated Circuits Singapore for useful discussions and support.

References

1. R. Pendse, et al, "Die Size Reduction Through Finer Pitch Bonding," *Proceedings of the HP Design Technology Conference*, 1994.
2. R. Pendse and R. Horner, *Radially Staggered Bond Pad Arrangements for Integrated Circuits*, U.S. Patent Pending.

A Miniature Surface Mount Reflective Optical Shaft Encoder

The HEDR-8000 Series encoders provide two-channel medium-resolution encoding performance in a very small SO-8 plastic package. Their small size, reflective operation, and low cost enable customers to design them into applications that were impossible for earlier encoders, such as feedback sensing for the miniature motors used in copiers, cameras, vending machines, and card readers.

by Ram S. Krishnan, Thomas J. Lugaresi, and Richard Ruh

Imagine a position servo sensor so small that it can fit almost anywhere. Put it in a surface mount package and give it a price comparable to slot encoders that have much lower resolution. Such a sensor became reality with the introduction of the HP HEDR-8000 Series reflective optical surface mount encoders (Fig. 1).

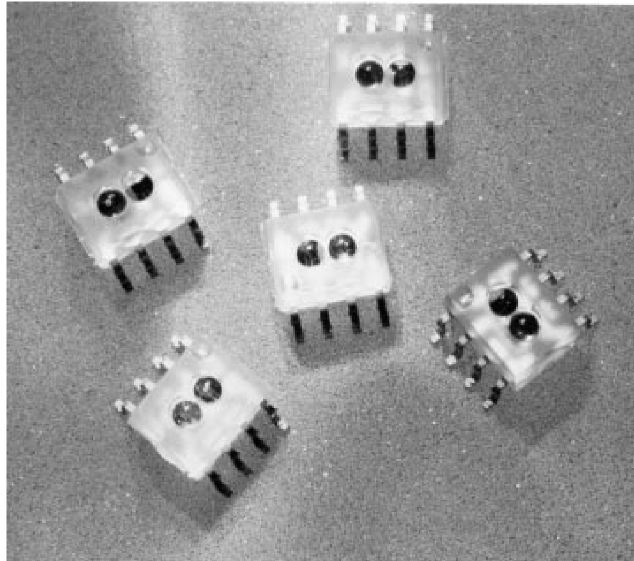


Fig. 1. *HP HEDR-8000 reflective optical surface mount encoder.*

The HEDR-8000 Series encoders deliver HP reflective optical encoder technology in a surface mount package. They provide two-channel medium-resolution (75 and 150 lines per inch) encoding performance in a very small SO-8 (small outline 8-pin) clear plastic package. The reflective technology of the HEDR-8000 Series encoders is inherently different from other HP encoder modules, which use transmissive technology with light passing through a codewheel or codestrip. In the HEDR-8000 Series encoders, light reflects off the codewheel or codestrip.

HP's shaft encoders (sensors that measure the position of a rotating motor shaft) have always been based on optoelectronics, thus providing noncontact measurement and far greater reliability than contact potentiometers. The first HP encoders were the HEDS-5000/6000 complete encoder packages, introduced in 1979. In 1987 HP introduced the low-cost HEDS-9100 encoder module, a small optoelectronic package that must be combined with a coded wheel or strip to function as a position sensor. A powerful feature of this encoder was that it could be assembled with a codewheel without the need for "phasing," or adjusting the position of the encoder to bring the output signals within the desired specification. The HEDS-9100 family quickly became the position sensors of choice in the computer peripherals market. But by 1989, low-cost inkjet printers were introduced and needed a smaller, lower-cost solution. This led to the introduction of the HEDS-9700 family of encoders, which were smaller, less expensive, and ultrareliable, supported high-volume assembly, and offered excellent performance.

The HEDR-8000 Series encoders were developed because new low-end printers required a smaller, less expensive encoder. Customers were redefining high-volume assembly capability to include surface mountability and infrared reflow oven

solderability. The design objectives for the HEDR-8000 Series encoder project were low cost, surface mount capability, and medium resolution.

The design was influenced by a new product introduced for front-panel applications, the HP HRPG family, which proved that reflective sensors were not only feasible and inexpensive, but had the added advantage of stackability, that is, the codewheel or codestrip could be mounted on top of the encoder module and not "through" it, as required by the previous transmissive models. This not only enhances the high-volume assemblability, but also allows the encoder to be placed in space-limited applications.

The small size, reflective operation, and low cost of the HEDR-8000 Series encoders enable customers to design them into applications that were difficult for earlier encoders. One such application is feedback sensing for the miniature motors used in copiers, cameras, vending machines, and card-readers.

Fig. 2 shows the evolution of HP optical encoders in cost and size.

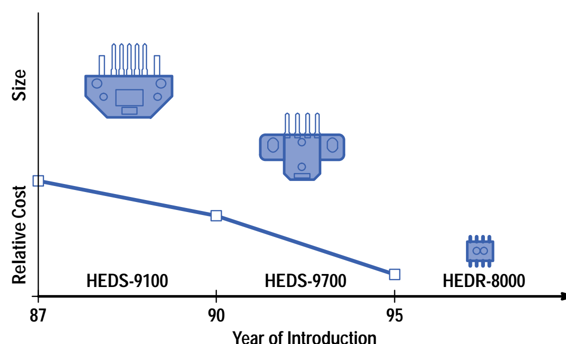


Fig. 2. *Reduction in size and cost of HP encoders.*

Basic Operating Principles

The HEDR-8000 Series encoders combines an emitter and a detector in a single surface mount SO-8 package. As shown in the block diagram, Fig. 3, the HEDR-8000 Series encoders have three key parts: a single LED light source, a photodetector IC, and a pair of lenses molded into the package. The lens over the LED focuses light onto the codewheel, and the image of the codewheel is reflected back through the lens to the photodetector IC.

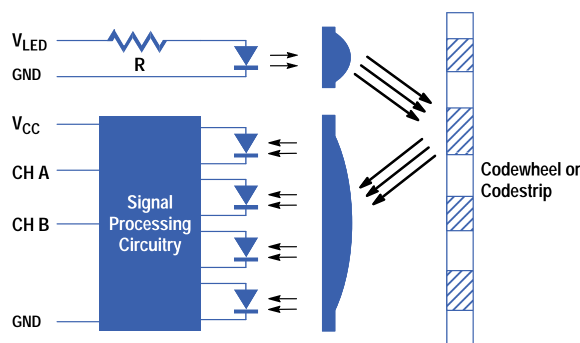


Fig. 3. *HEDR-8000 Series encoder block diagram.*

As the codewheel rotates, an alternating pattern of light and dark corresponding to the pattern on the codewheel falls on the photodiodes. This light pattern is used to produce internal signals A and B and their complements $\overline{A}$ and $\overline{B}$. These signals are fed through comparators to produce the final digital outputs for Channels A and B.

The HEDR-8000 Series encoders' performance is characterized by the quality and consistency of the two encoding signals, Channel A and Channel B. These signals have a quadrature relationship so that, as the codewheel passes in one direction, Channel A leads Channel B, and as the codewheel passes in the other direction, Channel B leads Channel A. Fig. 4 shows the output waveforms. Although the HEDR- 8000 Series encoders match HEDS-9700 resolution, they have generally lower performance than their transmissive predecessors, in part because of smaller size and lower cost.

Encoder Design

The first steps in the design determined the lens sizes required to gather sufficient light, the minimum detector IC size needed, and the minimum number of pins needed. A requirement was to use a standard surface mount package to allow

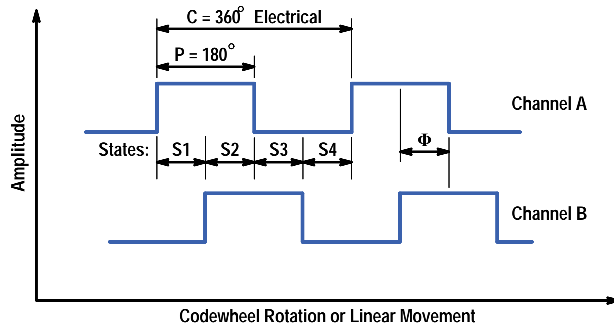


Fig. 4. *Output waveforms of the HEDR-8000 Series encoders.*

customers to use standard assembly equipment such as pick-and-place machines. These requirements were traded off with the customers' need for a small, fit-anywhere size to arrive at the SO-8 package. However, the height was increased beyond the standard to allow for the proper optical focal length of the lenses, after it was determined that this did not interfere with the operation of standard assembly equipment. The low-cost objectives are achieved by using a transfer molding process to form the package and lenses all at the same time. The plastic used to encapsulate the parts needed to be optically clear and transfer moldable. Not many such materials exist, but a moldable material was found that transmits about 90% of the light at the LED's wavelength of 700 nanometers. A book mold was built to mold prototype parts.

Existing ICs were used in early prototypes, and codewheels were made by photocopying a jail-bar pattern onto a reflective mylar sheet and cutting out the disks. Amazingly, they worked!

In its final configuration, an HEDR-8000 Series encoder consists of a clear plastic SO-8 package with two lenses located on the top of the package. A slit-shaped light-emitting diode (LED) that emits red light at 700 nanometers is under one lens and a silicon bipolar detector IC is under the other lens.

Optical Design

The reflectivity of the codewheel surface is an important factor affecting the performance of the HEDR-8000 Series encoders. The ideal codewheel reflective surface is a mirror that reflects almost all the light incident on it. The mirror-like property is specified by a measure called ***specular reflectance***. Specular reflectance (or specularity) is the percentage of the incident light that is reflected back at an angle equal to the angle of incidence. This is the property of a surface ***not*** to scatter light. For example, a shiny surface with a rough finish will reflect most of the light incident upon it, but will also scatter the light, and therefore will have a low specular reflectance. The specular reflectance can be measured with a device called a scatterometer. For proper operation, a minimum of 60% specular reflectance is required in the reflective portions of the codewheel and a maximum of 10% specular reflectance is required in the nonreflective portions. For example, metalized mylar codewheels having 85% specularity and nickel-plated stainless-steel codewheels with typically 65% specularity both perform well with the HEDR-8000 Series encoders.

The HEDR-8000 Series encoder lenses are spherical and have a radius of curvature of about 0.7 mm. Different lens options such as aspherical lenses, cylindrical lenses, no lenses, and others were tried by simulation. Spherical lenses were chosen for their ease and low cost of manufacturing and verification (no null correctors required), as well as their performance (significantly better than no lenses or cylindrical lenses).

The performance of the part was simulated using ASAP, a ray-trace program, with manufacturing tolerances included. The results indicated that the HEDR-8000 Series encoders would work robustly over the normal manufacturing tolerances, but the encoding performance would vary because of these tolerances.

ASAP was used to vary manufacturing tolerances such as die attach locations and lens dimensional variations. Customer assembly tolerances such as codewheel gap were also simulated. These studies were used to determine the optimal lens radius and height.

A critical design parameter is the included angle, defined as the angle subtended by the radii of the lenses. If the lenses were complete hemispheres, the included angle would be 180 degrees. Once again, optical simulations were used to determine the optimum included angle.

Detector IC

The starting point for the the HEDR-8000 Series encoder integrated circuit was the existing two-channel integrated circuit used in HP's high-performance encoders. The challenge was to take this basic design and reduce it in size to meet the low-cost objectives of the HEDR-8000 Series encoders without sacrificing performance.

The encoder circuit contains the following functions:

- Detection, implemented by photodiodes

* Amplification of photocurrents
* Production of stable bias currents
* Current-to-digital-output-voltage conversion with hysteresis
* Provision for testability.

The photodiode area was shrunk by a factor of 10 relative to the existing detector IC and the loss of signal was made up by increasing the lens magnification and the amplifier gain.

The amplifier was reduced to one third of its original number of devices by trading off sensitivity to transistor gain changes resulting from IC process variations. The HEDR-8000 Series encoder amplifier has a very large dynamic range and can work with a wide range of photocurrents. Common-mode rejection is provided by using a differential input configuration.

For the HEDR-8000 Series encoders, a new hysteresis circuit was designed using a geometric offset in the emitter instead of using a resistor as in the previous circuits. This allowed the removal of two large resistors while maintaining the function of the circuit. The hysteresis was found to perform better than the original circuit over temperature and V_{CC} variations, but is more sensitive to variations in the IC fabrication process.

No changes were made to the other functions.

The detector IC is a 5V bipolar device. Fig. 5 shows a partial layout and Fig. 6 shows the equivalent schematic circuit.

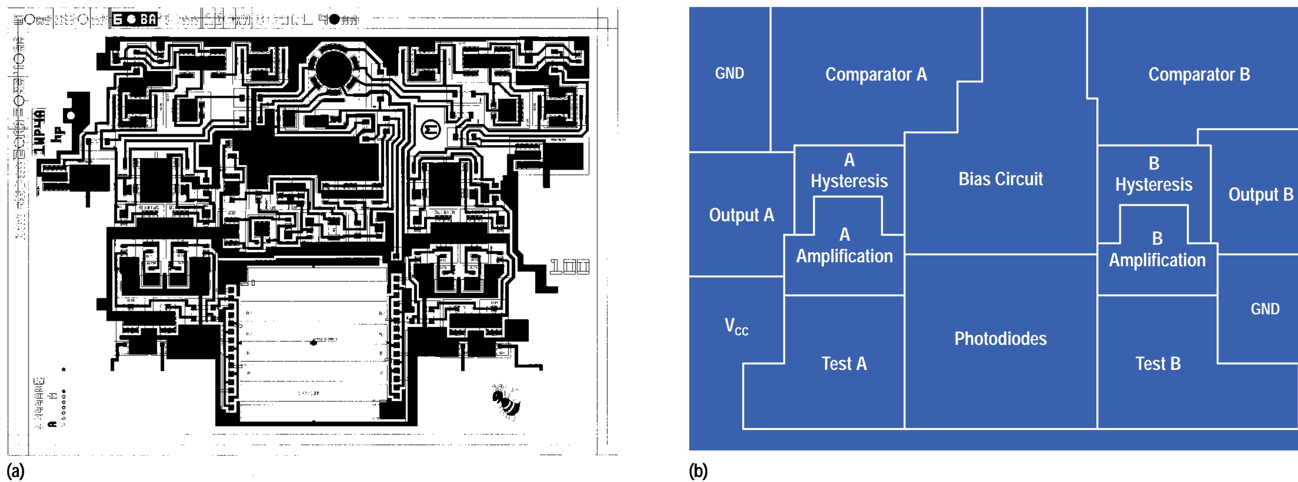


Fig. 5. *Partial layout of the detector IC.*

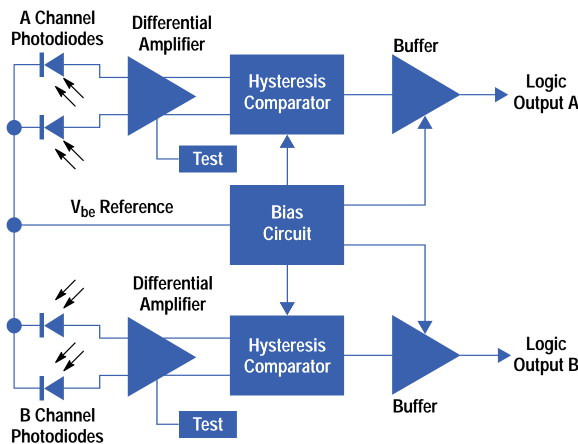


Fig. 6. *Block diagram of the detector IC.*

The circuit uses one or more sets of two photodiodes for each channel. The number of sets depends on the resolution of the particular model. For example, the 75-line-per-inch version has one set (a total of four photodiodes) and the 150-line-per-inch version has two sets (a total of eight photodiodes). The photodiodes in each set are laid out next to each other and labeled A, B, $\overline{A}$, and $\overline{B}$, in that order. If there is more than one set, the pattern repeats. The photodiodes are placed so that the image of

each set spans the pitch of the codewheel, that is, the width of each photodiode is a quarter of the codewheel pitch. The width of the photodiode at the detector IC is reduced by the magnification factor of the lens, which is 4/3. Thus, for the 150-line-per-inch version, the photodiode width is 0.0013 inch. This reduced photodiode size allows a reduction in the size and therefore the cost of the detector IC.

The A and $\overline{A}$ photodiodes together produce the A channel output and the B and $\overline{B}$ photodiodes together produce the B channel output, 90 electrical degrees apart from A. The layout of the photodiodes is such that the A and B photocurrents are 90 electrical degrees apart in phase, the A and $\overline{A}$ photocurrents are 180 electrical degrees apart in phase, and the B and $\overline{B}$ photocurrents are 180 electrical degrees apart in phase. The A and $\overline{A}$ photocurrents are amplified and fed into a comparator. The comparator's output, which mirrors the encoder's output, switches high when the A photocurrent is greater than the $\overline{A}$ photocurrent , and switches low when the A photocurrent is less than the $\overline{A}$ photocurrent. This push-pull operation allows consistent performance despite changing LED performance (which happens over time and varying temperature conditions). If there is more than one set of photodiodes, the photocurrents from all of the A photodiodes are averaged, and similarly for $\overline{A}$, B, and $\overline{B}$, thus minimizing the effect of codewheel irregularity.

Crosstalk between the two output channels is minimized by using an internal capacitor to slow the output transistor fall time to about 100 ns and by separation and shielding of sensitive signals.

The circuit is designed so that one of the photocurrents must be greater than the other by 10% to cause a change of output state. This hysteresis effect is important in preventing oscillations for the case when currents from A and $\overline{A}$ are exactly equal and spurious noise can cause a change in the output state.

The detector IC design is flexible enough to allow resolutions of 68 lines per inch to 200 lines per inch, with a change in the photodiode pitch.

Performance

The HEDR-8000 Series encoders are very tolerant to radial misalignment, tangential misalignment, axial play of the shaft on which the codewheel is mounted, LED current, and codewheel/codestrip gap. Fig. 7 shows typical performance for the 75-line-per-inch version.

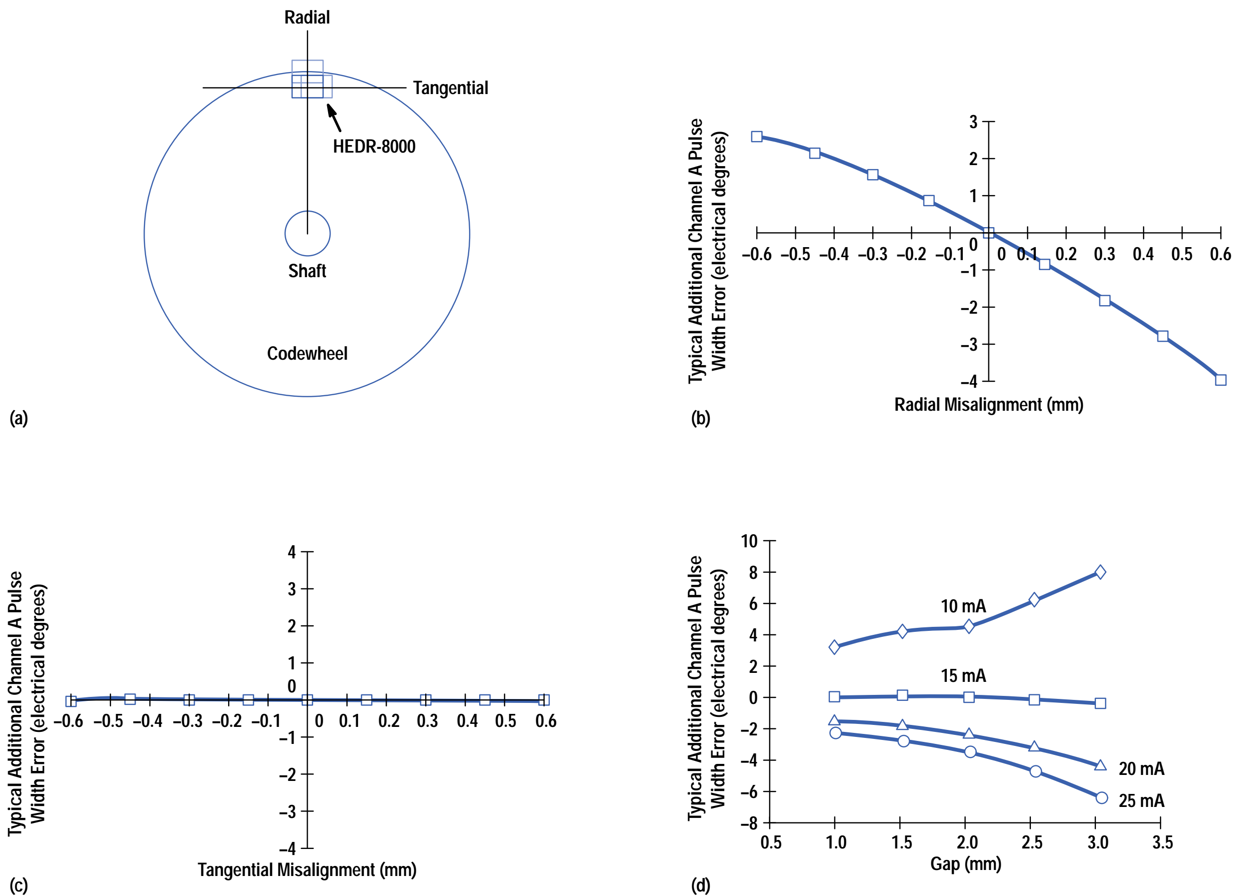


Fig. 7. *Performance curves for the 75-line-per-inch encoder.*

Manufacturing

The HEDR-8000 Series encoders required completely new processes for manufacturing. It was decided at the outset that key processes such as transfer molding of the package would be developed in San Jose, California and then installed in Singapore after the bugs were worked out. However, an important distinction from previous encoder lines was that the entire manufacturing line would be set up in Singapore from the beginning. The line was set up to produce small volumes so that customers could begin to design the HEDR-8000 Series encoders into their applications. The theory was that by the time customers were ready to buy the encoders in volume, a high-capacity line would be in place.

Project Management Issues

Many HEDR-8000 Series encoder customers are in Japan, the manufacturing line is in Singapore, and most of the development staff is in San Jose. Such projects are complicated by their logistics, not to mention the different cultures and time zones. A good many trips were made to Japan and Singapore, and these went a long way to improve personal relationships among team members and move the project along in its difficult times.

The project also experimented with two new concepts. One, called ***iterative product development***, involved building prototypes and presenting them to key customers for evaluation. A firm product definition was avoided until the prototypes got to a stage where the customers were clearly delighted with the concept. At this point, the product specification was finalized and the development began in earnest.

The second experiment involved producing parts on temporary or low-capacity tooling. The intent was to speed up market entry and subsequently develop a high-volume line while the customers evaluated the product in their applications.

Acknowledgments

Bob Steward and Bill Beecher originally identified the market trend that defined the HEDR-8000 Series encoders. For this and for their unwavering support through the many organizational debates, the authors are grateful. John Uebbing co-conceived the optics in the HEDR-8000 and was always a willing sounding board for ideas. Many thanks are due team members from other functions, including Marilyn Wong, Chris Yien, and George Willis, and team members from Singapore, including Richard Gan, Soon Hing Chan, and Wai Kuin Kwok. Many others also helped in support functions and are too numerous to acknowledge individually. Thanks are also due Victor Loewen, Dave Pitou, and Chris Togami for creating the original concepts and the ensuing excitement in the product.

Bibliography

1. ***Optical Encoder with Encapsulated Electrooptics***, U.S. Patent #5,317,149.
2. ***Variable Pitch Position Encoder***, U.S. Patent #5,241,172.
3. ***Reflective Shaft Angle Encoder***, U.S. Patent #4,952,799.

The Global Positioning System and HP SmartClock

The U.S. Department of Defense Global Positioning System has inherent problems that limit its use as a source of timing. HP SmartClock is a collection of software algorithms that solve or greatly minimize these problems.

by John A. Kusters

The Global Positioning System, or GPS, was designed as a ranging system that uses known positions of satellites in space to determine unknown positions on land, on the sea, in the air, and in space. GPS is a passive system in which each satellite transmits its position and the time of the position message. No information about the user or the user's receiver is required for a determination of the user's instantaneous position and velocity (navigational use) or for determining time at the user's receiver (time transfer use).

GPS was initiated by the United States Department of Defense in 1973. The system was recently declared fully operational by the United States Air Force. Twenty-four satellites currently make up the GPS constellation: four satellites in each of six planes spaced sixty degrees apart and inclined at 55° to the equator. Each satellite carries multiple atomic clocks, either cesium or rubidium, for redundancy and reliability. One of the clocks is declared operational for timekeeping purposes. Each satellite is also monitored by several ground reference stations to maintain accuracy. Ultimate timing accuracy is determined by the United States Naval Observatory (USNO) master clock. Fig. 1 shows the three segments of the GPS system: the satellites in space, the monitor and control function, and the user population.

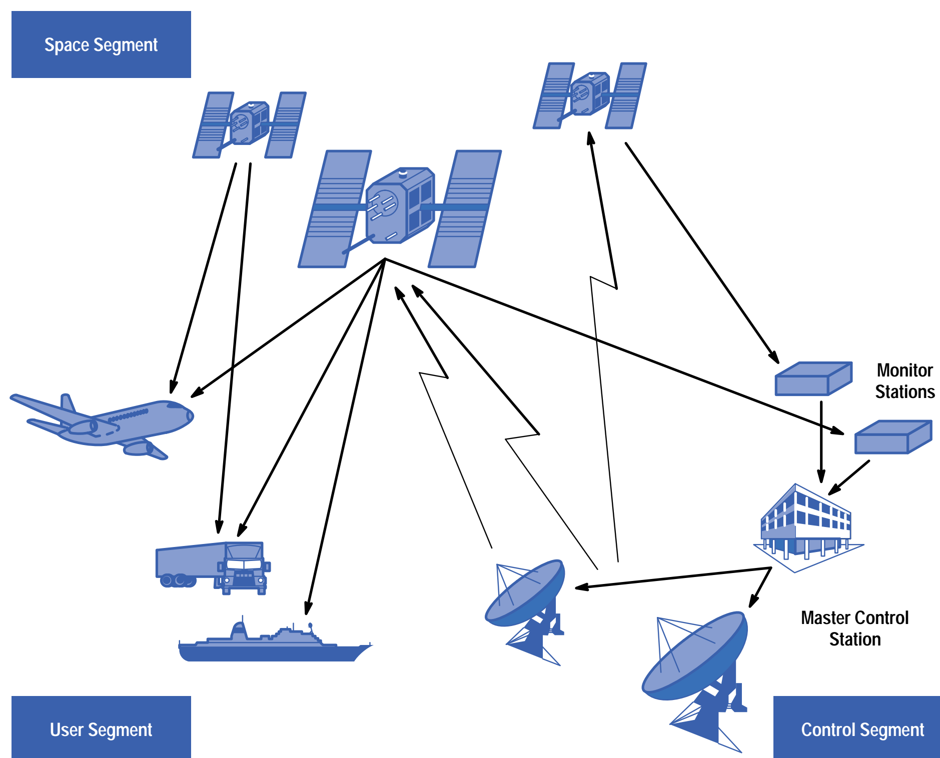


Fig. 1. *The three segments of the GPS system.*

Hewlett-Packard Company has been involved with the GPS program since its beginning. All of the frequency standards at the individual ground reference sites are HP 5061A cesium beam frequency standards. Most of the cesium standards at USNO are the newer HP 5071A primary frequency standard. Many other sites around the world also monitor GPS on a continuing basis. Virtually all of these also use one or both of the HP cesium standard models. HP has actively supported experimental uses of

GPS with equipment and technical expertise. In addition, several former HP scientists were among the first to realize the full commercial utility of the GPS system.

As participants in the GPS program, we were aware of its implications to our primary frequency standard product line. We were also aware that we had technology that complemented the GPS technology in many areas. A natural progression was to explore what we could potentially do with GPS in the areas of time and frequency standards and measurement techniques.

In 1992, one of our customers in the electrical power industry contacted HP about buying one of our cesium standards. At the time, given the nature of the power industry in North America, we were surprised that this utility felt it needed the precision of a cesium clock. How does one correlate needs at 60 Hz with a precision in the cesium standard of parts in 10^{15}? We visited them, and found that their need was not for 60 Hz but to precisely measure a wide variety of commercial GPS receivers to attempt to solve a major problem in their power system. This company generates power at various hydroelectric sites. When power lines fail, as they do, usually in the worst of weather, immediate knowledge of the location of the failure is essential to fix the problem. A key fact is that when the line fails, a traveling wave is generated at the failure and propagates on the power lines to power substations in both directions on the power line. If the utility could precisely time the arrival of the traveling wave at two or more substations, then the failure location could be determined. This requires that each station maintain the same time, and that this time be accurate to about 300 nanoseconds under any weather conditions. The only global timing system that meets these requirements today is GPS.

Looking at the advertisements from many different GPS equipment vendors might give one the feeling that GPS is the answer to any navigational or time transfer need. GPS is viewed by many as the next utility. But, GPS as a utility has problems just as do the power and the telephone utilities.

System Problems

To meet navigational, surveying, and time transfer needs, each of the GPS satellites broadcasts its position and time. However, the message broadcast is not necessarily accurate. The position of a satellite, its ***ephemeris***, is not exactly known and made available to the public until 48 hours after it is broadcast. Much more serious is a characteristic of GPS that allows the U.S. Department of Defense to degrade either the time message or the ephemeris or both. Collectively, the degradation is known as ***selective availability***, or SA. SA is jitter that is deliberately introduced into the system to reduce its overall accuracy for nonmilitary users. Users of the ***standard positioning service*** (SPS) of GPS, therefore, cannot achieve full system accuracy. SPS is specified to provide 100 meters horizontal positioning accuracy, 156 meters vertical accuracy, and 340 nanoseconds time transfer accuracy, 95% of the time.[1] A graphical representation of this is shown in Figs. 2 and 3.

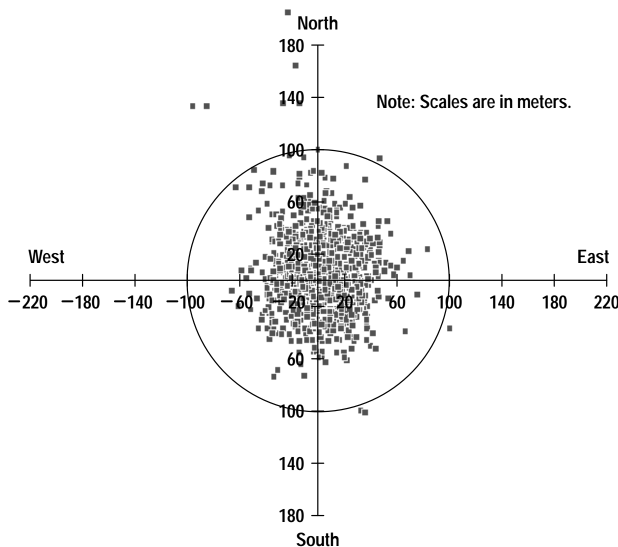


Fig. 2. *GPS horizontal errors over a 24-hour period using SPS (from reference 1).*

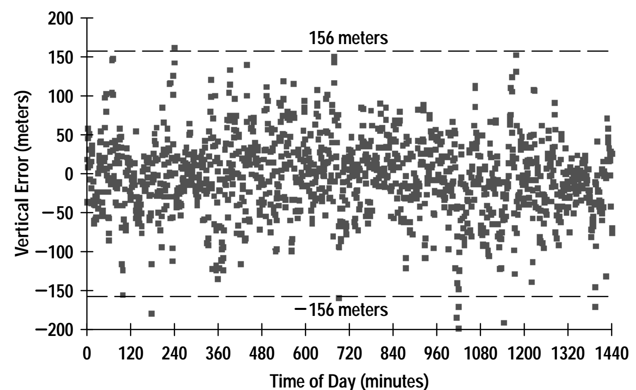


Fig. 3. *GPS vertical errors over a 24-hour period using SPS (from reference 1).*

United States military and other authorized users with the proper security keys can access the ***precise positioning service*** (PPS) of GPS. PPS is specified to provide 16 meters spherical position accuracy and 100 nanoseconds time transfer accuracy.

System errors are a product of the stability of a particular satellite's clock, the predictability of its orbit (ephemeris), and errors in the satellite messages. Under most circumstances the combination introduces a timing uncertainty of 10 to 50 nanoseconds.

A major problem occurs when a satellite malfunctions and is not identified as bad in the satellite message, or when wrong or inaccurate data is sent from the ground control station to the GPS satellites. Effects on the user's data can be significant, with timing errors approaching many milliseconds and positional errors up to several thousand meters. The only protection for the user seeking continuously accurate time and frequency is to use a receiver that has been specifically designed to be a timing reference and provides continued operation without degradation of time if either the GPS system becomes inoperative or bad data is broadcast from the system.

Propagation Problems

GPS satellites are in a half-geosynchronous orbit. They take essentially 12 hours to circumnavigate the earth at an altitude of 10,900 miles. Signals from the satellites propagate through the ionosphere and the earth's troposphere before reaching the user's GPS receiver. Losses and signal delays through the ionosphere can be large when solar activity is high. Other losses and signal delays occur because of localized weather conditions. A standard model for ionospheric time delays is contained in the satellite message, but it is only about 50% accurate. Timing delays caused by ionospheric and tropospheric effects can approach several tens of nanoseconds. Further errors can occur in the receiver because of errors in the processor's calculation of the ionospheric model.

Receiver Problems

Propagation effects and computational errors are also seen in the GPS receiver as timing biases that are a function of the receiver design. These are usually estimated by the GPS designer and proper compensation is provided. However, for critical applications, additional calibration of an individual receiver might be required. If properly designed, receiver bias is usually less than 20 nanoseconds.

User Problems

User installation problems can result in further degradation of positional data and timing data. Most GPS receivers are capable of providing latitude and longitude with sufficient accuracy to obtain good timing. There is little correlation between latitude and longitude errors and timing errors, as long as the positional error is less than 100 meters. The problem is that errors in altitude correlate strongly with timing errors, and altitude errors are usually greater than latitude and longitude errors. Correlation plots are shown in Figs. 4 and 5.[2]

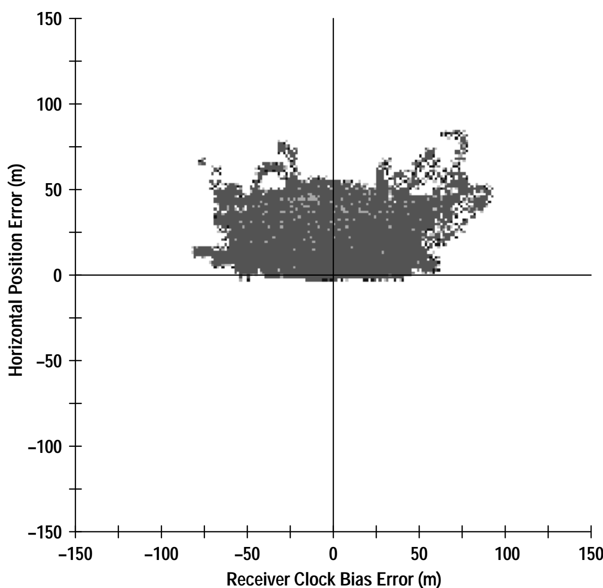


Fig. 4. *GPS measurement, 10-second samples, showing little correlation between horizontal error and time error (from reference 2).*

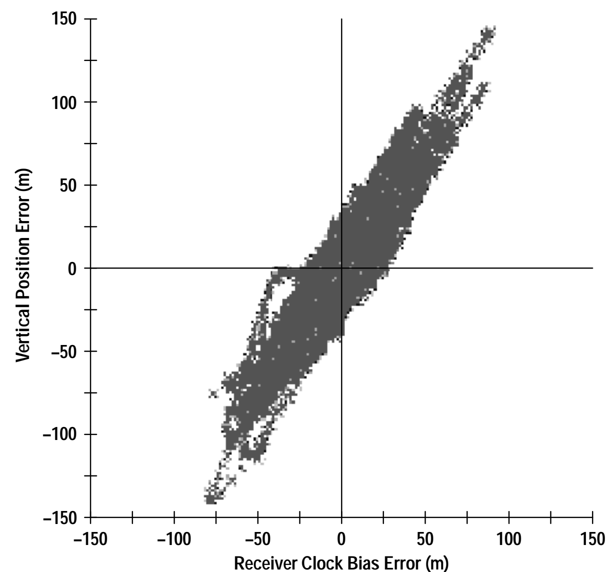


Fig. 5. *GPS measurement, 10-second samples, showing strong correlation between vertical error and time error (from reference 2).*

Another major problem is that GPS determines its position and time at the antenna, not at the receiver. Additional delay must be introduced by the user to account for antenna propagation delay and the delay of any further cabling used to deliver the timing signal to the user. Timing errors are dependent on the length of the antenna cable. The length can be greater than 300 meters, so timing errors greater than 1 microsecond are possible. The most effective method is to measure the actual electrical length of all cables after installation, using time-domain reflectometry techniques.

Antenna siting is another potential problem. If strong radio frequency reflectors are in the vicinity of the antenna, multipath reflection of the GPS signal may occur. Because of multipath, the receiver receives two or more signals from the same satellite, but with different time delays. Without proper consideration of multipath and effective receiver and antenna design to minimize its effects, timing errors up to 50 nanoseconds can be seen.

System and user timing errors are summarized in Table I. The values shown are generally worst-case.

Table I
GPS Timing Errors

Cause	Specification or Expected Error
GPS System	±340 ns (at 95%)
Propagation	
Ionosphere	up to 40 ns
Troposphere	up to 20 ns
Solar Flares	40 ns to system inoperative, dependent on severity
User	
Receiver	<20 ns
Horizontal Position Errors	negligible if self-survey is used
Vertical Position Errors	3 ns per meter of altitude error
Antenna	up to 3 ns per meter of antenna length error
Multipath	up to 50 ns
Environmental	up to 15 ns

Another consideration in antenna siting is that the desired location for the antenna may be in the near-field radiation of another transmitter and antenna. The received strength of a typical GPS signal is about −134 dBm. Many microwave systems, cellular telephone transmitters, and other wireless systems have frequency components and overtones sufficiently near the GPS L1 frequency, 1575.42 MHz, to overload the receiver front end. If the signal strength of the interfering source is sufficient, GPS signals cannot be received. The best solution is to find a site where the GPS antenna is not in the near-field radiation pattern of the interfering source. This may compromise the elimination of multipath. It may also be necessary to filter all out-of-band signals at the antenna or at the receiver's front end.

HP SmartClock

Of all the error sources discussed above, the most serious are the system errors. Selective availability is subject to change at any time. Errors in the satellite message, satellite problems, and other system problems are still observed occasionally even though the system itself is now fully functional. Further problems occur because of the satellite geometry and the antenna location. Because of a marginal location, periods may occur when too few satellites are observed to get the desired positional and timing accuracy. Loss of satellites may also occur because of antenna problems. Antenna leads can be damaged or cut. Snow load on the antenna can reduce its sensitivity. Large birds have been known to perch on the antenna.

In our investigation of the GPS system and its potential use as a source of timing, we have developed a collection of software algorithms that solve or greatly minimize these problems. The overall collection is called HP SmartClock.[3]

HP SmartClock Instruments

HP SmartClock techniques have been used in a wide variety of applications and have resulted in a spectrum of HP products to serve the needs of the general timing population.

The HP 58503A GPS time and frequency reference receiver, Fig. 6, is designed to meet the timing and control needs of small calibration laboratories and the general need for high-precision frequency and timing without buying a cesium standard. It generates precise 10-MHz and 1-pps signals and incorporates an RS-232 or RS-422 port for monitoring and control.

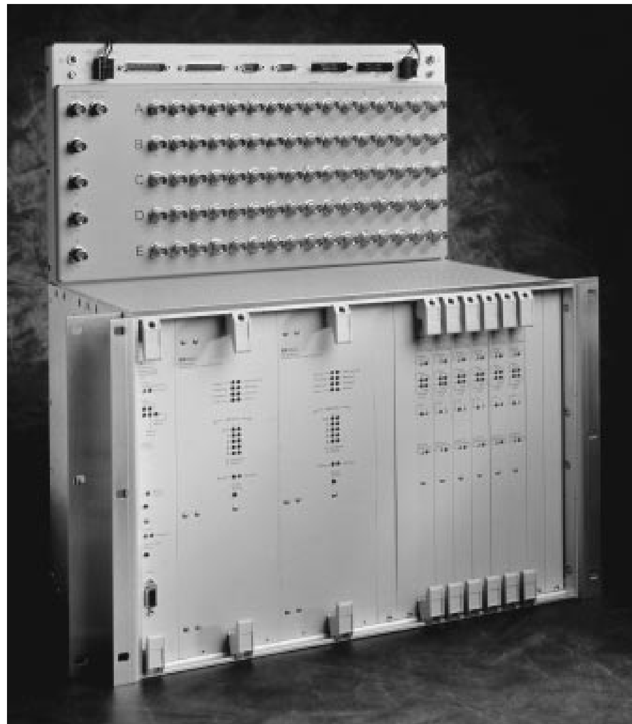
The HP 59551A GPS measurements synchronization module, Fig. 7, is designed to meet specific needs of the power generation and distribution community. It provides a 1-pps signal, IRIG-B, and three channels of high-precision event time tagging.

The HP 55300A GPS telecom primary reference source, Fig. 8a, and the HP 55400A network synchronization unit, Fig. 8b, are designed to meet specific needs of the communications industry.

Fig. 6. *HP 58503A GPS time and frequency reference receiver.*

Fig. 7. *HP 59551A GPS measurements synchronization module.*

(a)



(b)

Fig. 8. *(a) HP 55300A GPS telecom primary reference source. (b) HP 55400A network synchronization unit.*

Supporting these units are a wide variety of special hardware, antennas, line amplifiers, lightning arresters, fiber-optic distribution amplifiers, and special mounting options to meet the needs of specific industries.

Enhanced GPS

When GPS is the reference source for timing, the effects of selective availability (SA) can be greatly minimized. Observations of the spectral characteristics of SA show that it has a correlation peak at about 400 seconds.[4] Thus, any filter that attempts to reduce SA must have time constants that are significantly longer than 400 seconds. *Enhanced GPS* is an HP SmartClock digital filtering technique that exploits the observed correlation peak. When properly designed and matched to an internal frequency reference source, a filter can greatly reduce the effects of SA. The standard specification for SA is 170 nanoseconds rms (340 ns at the 95% level). In a GPS timing receiver using a high-precision quartz oscillator such as the HP 10811D/E, the rms deviation can be reduced to below 30 nanoseconds. With atomic oscillators such as rubidium or the HP 5071A primary frequency standard as a reference, the rms deviation can be further reduced. Experimental results with the HP 5071A have shown an rms deviation of about 2 nanoseconds, an 85-fold reduction in the effect of SA.[5] Fig. 9 shows the effect of the SA filter. The black lines indicate the time instability of the GPS timing signal even after filtering with a 6-channel receiver and averaging over 300 one-second samples. The white line shows the results of using the SA filter to reduce the amount of SA. In this case, the SA filtered data shows a 2.1 nanosecond rms scatter.

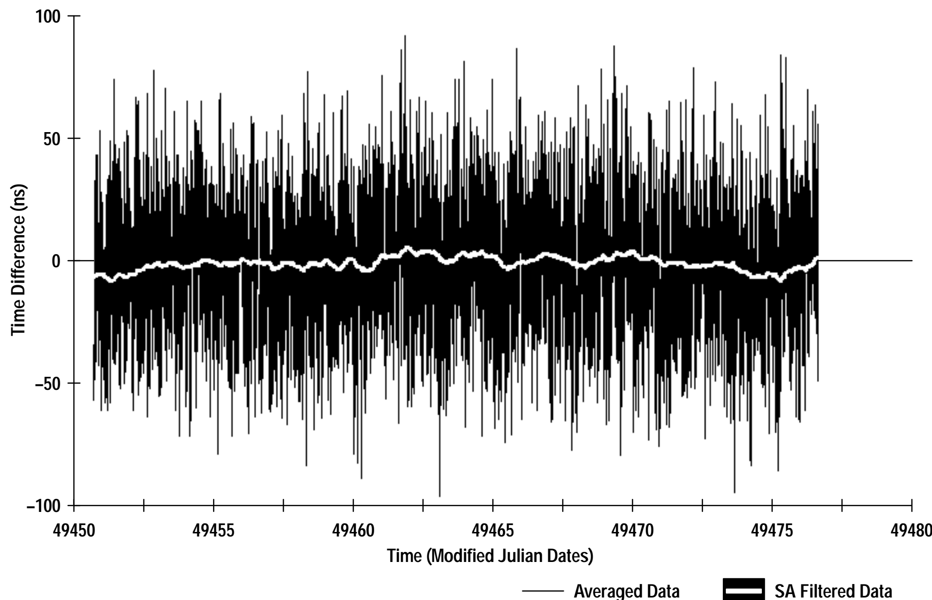


Fig. 9. ***Timing data taken at the U.S. National Institute of Standards and Technology (NIST) using HP SmartClock. The black lines represent the effect of SA (selective availability) after extensive averaging using a 6-channel GPS receiver and 300-second data averaging. The white line in the center represents the output of the SA filter. Observed rms deviation of the SA filtered data is 2.1 ns (from reference 5).***

The key is that the spectral characteristics of SA and the time-domain stability characteristics of the oscillator used must be matched through the types of filters and the loop time constants used in the various control loops. Each oscillator type has a unique filter technique that optimizes the reduction of SA.

Enhanced RAIM

Receiver autonomous integrity monitoring (RAIM) is a series of algorithms that continuously check each satellite against all others under observation. RAIM can take many forms. The GPS engine used in HP timing modules has its own version, T-RAIM, or time-RAIM. The HP timing receivers have an extra layer of RAIM that checks timing information received from the GPS engine against its own timing derived from a precision oscillator. Algorithms monitor the overall health of the timing module, its timing signal, and the signals received from the GPS engine to determine when enhanced RAIM needs to be implemented to preserve the overall timing accuracy.

Enhanced Learning

During normal operation, the internal precision oscillator, usually a quartz oscillator, is phase-locked to the GPS signal by comparing the time difference between the 1-pps (pulse-per-second) signal from the GPS engine to a similar signal derived from the reference source. A block diagram is shown in Fig. 10. While locked to the GPS system, HP SmartClock employs *enhanced learning* to measure the aging and environmental response of the internal reference source. Over a period of time, changes in the oscillator frequency caused by either aging or temperature changes are accurately measured using as a reference the signal from the GPS engine as derived from the enhanced GPS algorithm. Changes caused by humidity or pressure are minimized by using a hermetically sealed oscillator.

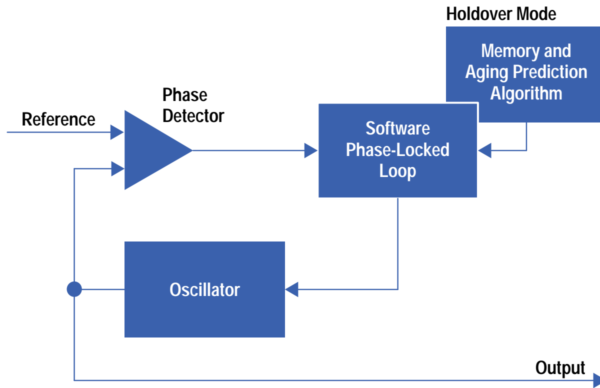


Fig. 10. *Block diagram of the HP SmartClock hardware. Over an extended time period, a signal derived from the oscillator is compared to the reference. The reference can be GPS, or another signal deemed sufficiently accurate to use as a reference. The phase-locked loop locks the oscillator to the reference. Frequency change signals sent to the oscillator are stored and analyzed by the memory and aging prediction algorithm.*

Long-term changes, those occurring over a period of many hours, are related to the aging of the internal oscillator. Frequency changes also occur as a function of temperature. These are measured and stored in internal memory. Constants related to the aging of the oscillator are stored in RAM and are redetermined each time the receiver is turned on. Constants related to temperature performance are stored on EPROM, since temperature performance does not substantially change during periods when the oscillator is not powered.

Normal Operation

Normal operation of the HP timing modules involves initial acquisition of four or more GPS satellites to accurately determine the geographic position of the antenna. Initially, the timing module uses a short time constant to control the oscillator. This facilitates accurate time setting of the module. Following a series of checks of the overall operation of the module, the time constants incrementally increase to their final values. This usually takes from 2 to 18 hours. At this point, the timing module is fully functional and meeting all of its specifications.

While still locked to GPS, HP SmartClock technology in the timing module starts learning the characteristics of the internal precision oscillator. The learning algorithm requires two full days of data to ensure that an adequate determination of the aging can be made. Learning never stops as long as the unit is powered and locked to GPS. Data from the most recent 48 hours is stored in RAM. Older data is discarded.

While locked to GPS, the module shares the long-term stability of GPS. Short-term, the timing module stability is directly controlled by the short-term stability of the oscillator used. A typical stability curve is shown in Fig. 11.

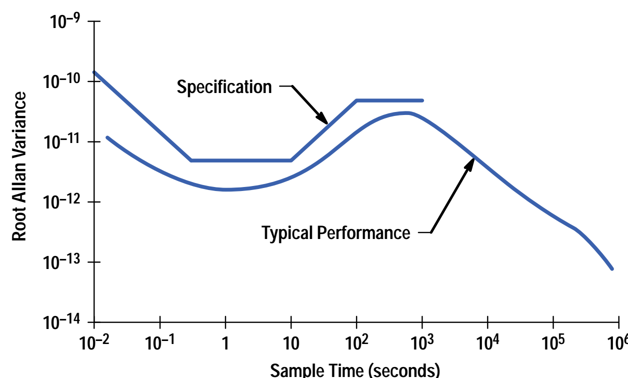


Fig. 11. *Root Allan variance of the SA filtered time difference data measured by an HP 59551A GPS measurements synchronization module.*

Frequency accuracy is essentially independent of most of the errors discussed above. The output of a GPS engine is a 1-pps signal. This is compared directly to a similar 1-pps signal derived by direct division of the oscillator signal. The comparison is made using consecutive 1-pps signals. All of the 1-pps signals from the GPS engine are affected equally by all of the error terms mentioned above. Therefore, to first order, all of the errors discussed above cancel. For averaging times greater than 24 hours (86,400 seconds), the frequency accuracy is better than 1×10^{-12}.

For the same reason, timing stability is essentially independent of the errors discussed. However, timing accuracy is directly affected by the errors discussed previously. Assuming that all of the user-controlled errors (Table I) are negligible, HP SmartClock timing modules with quartz oscillators achieve timing accuracies better than 110 nanoseconds at the 95% level. As an example of this, Fig. 12 shows data taken using an HP 59551A timing module by the United States Naval Observatory. This data was taken using a direct measurement between the USNO master clock and the 1-pps output of the HP 59551A. The peak-to-peak deviation is 100 nanoseconds. The average offset is 20 nanoseconds. The offset is the result of a 10-nanosecond receiver time bias and a known offset of GPS from the USNO master clock of another 10 nanoseconds.

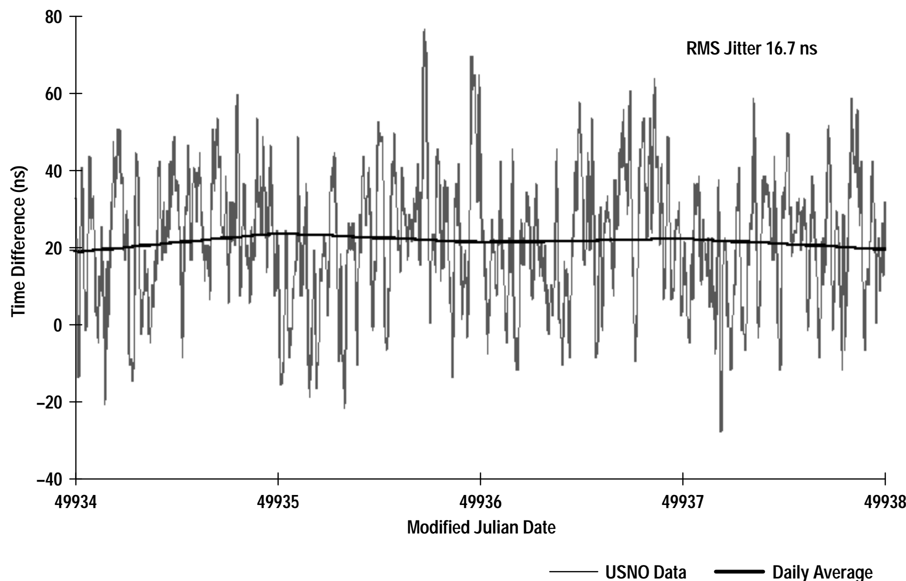


Fig. 12. *Direct measurement of the 1-pps output of the HP 59551A GPS measurements synchronization module against the USNO master clock.*

Neglecting the GPS-UTC (USNO) offset, over this period, the timing module easily met its timing specification of ± 110 nanoseconds. Actual data showed that compared to the master clock, the timing module was within less than ± 70 nanoseconds, with an rms jitter of 16.7 nanoseconds. The maximum deviation of the 1-pps signal was less than 6 nanoseconds over any one-minute period. The standard deviation was less than 1.8 nanoseconds over any one-minute period.

As determined from the 1-pps data, the 24-hour average frequency offset was 4.6×10^{-13}.

Holdover Operation

Occasionally, the GPS reference signal is not available. The antenna may become unusable because of weather, broken or damaged cable, or other causes. The receiver may temporarily lose track of the satellites. The satellite system may receive a bad data upload, or otherwise be unavailable because of military needs. Whatever the cause, during loss of the reference, accurate timing signals must still be generated and used to control customer equipment.

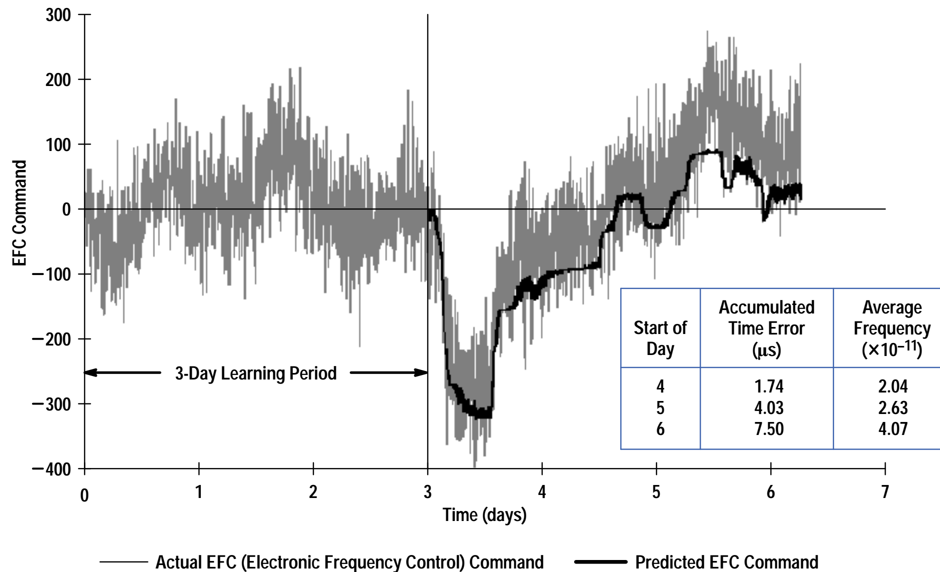
During the loss of the reference, HP SmartClock uses all of the data learned previously about the oscillator to control the oscillator to maintain all timing outputs at essentially the same level of precision as that obtained while locked to the reference. This form of operation is called ***holdover***.

A control loop tracks temperature changes in the module and computes the correct offsets for the oscillator to remove temperature effects. Another loop tracks elapsed time and computes additional offsets for the oscillator to remove any aging effects. Other loops continue to monitor the GPS engine to determine whether normal operation can be resumed.

Normal specification requires that during holdover, the module maintain frequency accuracy to better than 1×10^{-10} and accumulate timing errors no greater than 8.6 microseconds for the first day of holdover, after three days of learning time. Actual performance is highly dependent on the overall length of learning time available before holdover. The longer the learning period, the more stable the oscillator, and the more accurate the prediction.

Fig. 13 illustrates the effects described above. In this case, the value plotted in light gray is the electronic frequency control signal that steers the oscillator.

This unit had previously been operating for several weeks. At the start of this test, we cleared the memory of previously learned data, then started the oscillator relearning. At the end of day 3, we retrieved all of the learned data, including the predicted future performance of the unit.



Start of Day	Accumulated Time Error (μs)	Average Frequency ($\times 10^{-11}$)
4	1.74	2.04
5	4.03	2.63
6	7.50	4.07

Fig. 13. *Example of holdover showing environmental effects.*

During the next three days, we compared actual operation (the light gray curve) to the predicted operation (the dark curve). In our experience, this becomes the most accurate way of determining the quality of the prediction in all circumstances. We could simply disconnect the antenna, then watch what happens. However, it becomes difficult to determine the cause of any unexpected time or frequency error. The data shown in Fig. 13 is a more accurate second-to-second picture of overall performance.

Comparing actual to predicted performance allows an easy determination of both the expected frequency offset and accumulated timing errors. The assumption is that both were perfect at the start of the comparison.

The data shows a diurnal variation caused by changing temperature. The unit was operated in a normal room environment. During the night, the room's climate control was turned off, causing a decrease in room temperature. The large dip in the curve at the start of the experiment marks the start of a weekend, when a much larger temperature change was seen.

Computed values show that at the end of day 4, the first day in simulated holdover, the frequency error was 2.04×10^{-11} and the accumulated time error was 1.74 microseconds. At the end of day 6, the third day in simulated holdover, the frequency change was 4.07×10^{-11} and the time error was 7.5 microseconds.

Acknowledgments

The author wishes to thank the United States Naval Observatory and especially Dr. Mihran Miranian for taking and allowing us to use their data. The basic concept that eventually led to the HP SmartClock was developed by scientists at the National Institute of Standards and Technology (NIST). Special thanks to David Allan, formerly of NIST, for helping us to define many of the concepts that eventually led to the HP SmartClock, and for kindly reviewing this article for technical errors. Finally, none of this work would have been done without the guidance and help of Dr. Leonard S. Cutler and Dr. Robin Giffard of Hewlett-Packard Laboratories. Many key features of the HP SmartClock technology were originally developed and implemented in software by members of the Santa Clara Division Technical Staff, especially Ken Ho, Gary Jacobsen, Robin Poskus, Eric Ingman, Johann Heinzl, Dick Schneider, Terry Nimori, and Doug Nichols.

References

1. U.S. Department of Defense, ***Global Positioning System, Standard Positioning Service, Signal Specification, 2nd Edition***, June 2, 1995.
2. P. Misra, et al, "Adaptive Modeling of Receiver Clock for Meter-Level DGPS Vertical Positioning," ***Proceedings of the ION GPS-95***, The Institute of Navigation, September 1995, pp. 1127-1135.
3. J.A. Kusters, et al, "A No-drift and less than 1×10^{-13} Long-term Stability Quartz Oscillator Using a GPS SA Filter," ***Proceedings of the 1994 IEEE International Frequency Control Symposium***, IEEE Catalog No. 94CH3446-2, June 1994, pp. 572-577.
4. D.W. Allan and W.P. Dewey, "Time-Domain Spectrum of GPS SA," ***Proceedings of the ION GPS-93***, The Institute of Navigation, September 1993, pp. 129-136.

5. J.A. Kusters, et al, “A Globally Efficient Means of Distributing UTC Time and Frequency Through GPS,” ***Proceedings of the 26th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting***, December 1994, pp. 235-254.

Universal Time Coordinated (UTC)

A continuing misconception is that the GPS system presents a timing signal that is always directly related to Universal Coordinated Time or UTC. UTC is a global collection of highly accurate atomic clocks and astronomical observations, coordinated and maintained by the Bureau International Des Poids et Mesures (BIPM) in Paris, under the International Treaty of the Second (Fig. 1). Many GPS receiver specification sheets state that the receiver is accurate within 100 nanoseconds of UTC. The problem is that the timing accuracy of the GPS system, or GPS time, is controlled by the United States Naval Observatory (USNO). USNO is a major contributor to the BIPM time base. The Naval Observatory has the charter to maintain the GPS system to within 1 microsecond of UTC. During the past year, the standard deviation of GPS time with respect to UTC was less than 10 nanoseconds. The observation is that most of the time, GPS time is very near to UTC time. However, this can be changed by the USNO as military needs dictate.

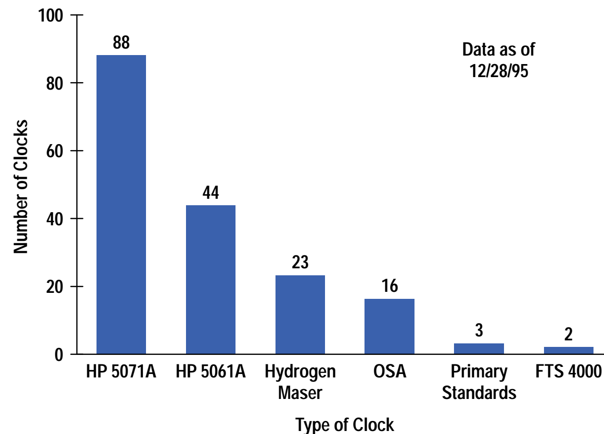


Fig. 1. *Component clocks in the BIPM time base. HP clocks represent 71% of the clocks (OSA and FTS are other cesium standard manufacturers) and over 82% of the weight in defining Atomic Time International (TAI), the size of the second in the UTC time base (from reference 1, 12/28/95.)*

Reference

1. Bureau International des Poids et Mesures, *Circular T Bulletin*, issued every two months, Paris, France.

The Third-Generation HP ATM Tester

Breaking away from the traditional bounds of transmission and protocol analyzers, the HP E5200A broadband service analyzer redefines the way in which the interactions between protocol layers at multiple points in the network are analyzed and presented, leading to the new concept of service analysis.

by Stewart W. Day, Geoffrey H. Nelson, and Thomas F. Cappellari

The HP E5200A broadband service analyzer (Fig. 1) is Hewlett-Packard's third generation of ATM test equipment, focusing in particular on the analysis needs of ATM service deployment.

Fig. 1. *The HP E5200A broadband service analyzer implements the new concept of service analysis, incorporating hardware, software, and usability advances aimed specifically at testing ATM services.*

HP's first-generation tester was called the Series 90. It combined extensive SDH/SONET transmission test capability with the world's first ATM test capability. The addition of higher-layer protocol testing and more complex ATM conformance testing resulted in the second-generation HP tester called the Broadband Series Test System (BSTS). The BSTS has been involved in almost every major ATM field trial around the world.

Now that ATM is progressing from the early field trial phase into the deployment of revenue generating services, the requirements for ATM test equipment have changed. To meet these requirements, HP's Australian Telecommunications Operation has developed the concept of ***service analysis***. The HP E5200A broadband service analyzer implements the service analysis concept, incorporating hardware, software, and usability advances aimed specifically at testing ATM services.

Market Evolution in ATM Testing

HP has been developing test equipment for ATM since this technology first started to gain acceptance in Europe in the mid-1980s (see Fig. 2). It was at this point that the industry started to see the potential for an integrated public broadband network infrastructure, the Broadband Integrated Services Digital Network (B-ISDN).

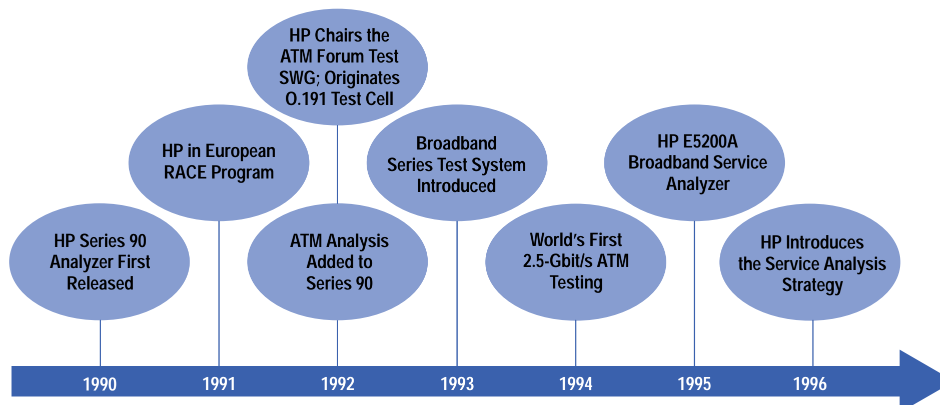


Fig. 2. *HP involvement in broadband and ATM testing.*

HP first became involved in ATM testing in 1988 when a team from the HP Queensferry Telecommunications Division (QTD) in Scotland joined a European Community research program called RACE (Research into Advanced Communications in Europe). The team joined a RACE project called PARASOL, which was a precompetitive collaboration between HP and research teams from a number of other European organizations. These organizations included a cross section of public telephone service providers, network equipment manufacturers, universities, and specialist software houses. The primary goal of the project was to research the test and measurement requirements of B-ISDN equipment and networks being developed by other RACE projects, and develop a test tool to support this work and verify its operation. A great deal of knowledge and experience was gained by all parties and a prototype product fulfilling these needs was developed and used successfully.

By the end of the PARASOL project in 1992, however, it had become obvious that ATM was gaining worldwide acceptance. The ATM Forum was forming and a competitive ATM marketplace was taking shape. This made continued collaboration unfeasible for HP. Therefore, in late 1992, HP's Australian Telecom Operation (ATO) launched the world's first commercial ATM test equipment, the HP 75000 ATM Series 90. This tester was based on the highly successful Series 90 SONET/SDH R&D testers and primarily allowed physical layer and ATM layer testing at OC-3/STM-1 rates, followed soon after by DS3. The Series 90 is essentially a transmission system tester with detailed SDH/SONET overhead test features. Subsequent ATM developments added concatenated payload capability at OC-12c/STM-4c rates and, more recently, OC-48c/STM-16c rates, making the ATM Series 90 the world's only real-time ATM tester at 2.488 Gbits/s.

The ATM Series 90 was a great success, helping HP to build strong global relationships with key ATM customers. It soon became clear, however, that with the speed of technological developments in ATM, there was a strong requirement for features beyond those that the Series 90 could support. HP's Idacom Telecom Operation (ITO) in Canada had been developing WAN testers for the X.25, Frame Relay, and SMDS standards. With ATM also being their next step, it was decided to combine ITO's higher-layer protocol expertise with ATO's expertise in ATM transport and develop a dedicated and extremely powerful ATM tester. The resultant Broadband Series Test System (BSTS) was a success in the R&D ATM test market from late 1993, gaining universal acceptance as the reference tester for ATM developments with almost every telecom operator and network equipment manufacturer throughout the world. Today, the BSTS continues to lead the way in R&D ATM test with support added for conformance test suites and now MPEG-2 video over ATM.

With ATM starting to move out of the R&D lab to early field trials and carriers throughout the world announcing plans for commercial ATM deployment, HP again saw the need to develop test equipment targeted directly at this new phase of the ATM life cycle. ATM features have been added by the HP Queensferry Telecom Operation (QTO) and the HP Cerjac Telecom Operation (CTO) to their transmission test instruments for use in installation and maintenance of the core network infrastructure, and the HP Colorado Communications Operation (CCO) has added ATM to their advisors for LAN/WAN interworking.

To address the speed and complexity of ATM service deployment fully, however, it was felt necessary to break away from the traditional bounds of transmission and protocol analyzers. By redefining the way in which the interactions between protocol layers at multiple points in the network are analyzed and presented, the concept of service analysis was born. It is this initiative that has led to the HP E5200A broadband service analyzer, the third generation of ATM test equipment from the ATO and HP.

Defining the HP Broadband Service Analyzer

As explained in the accompanying articles, the HP E5200A broadband service analyzer has pioneered several technical advances within Hewlett-Packard. It has also pioneered some key strategy and process initiatives, particularly in the area of direct customer involvement in the early stages of product definition. By targeting specific strategically important customers through intimately involving their HP sales engineers in the product's definition and development, it has been possible to build very strong relationships between customers, field, and factory and to ensure that a common understanding of the issues exists throughout. As described in ***Article 14***, graphical user interfaces for ATM service analysis were extensively usability tested on real users, whose feedback helped significantly to shape the design and gave them a feeling of real ownership in the final product.

What Is Service Analysis?

Broadband networks are complex to manage. To provide customers with the quality of service they expect, ATM service providers need to understand the interaction between services, protocol layers, and equipment. Service analysis allows service providers to manage the end-to-end quality of broadband services effectively by showing the important interactions between the elements of the service, such as location dependencies, protocol layer interactions, service interference, and element interoperability.

The need for service analysis is best understood by looking at a few examples.

Example 1: Flow control across protocol layers. Consider an application in which two LANs (TCP/IP) located in different cities are interconnected by a WAN technology (Frame Relay) which is in turn interconnected by a high-speed ATM backbone. In this situation, there will be three completely independent flow control mechanisms operating at the TCP/IP, Frame Relay, and ATM layers (Fig. 3). If the customer were to complain about low throughput, it would be impossible to find the cause of the problem without observing all protocol layers and the way their flow control mechanisms interact.

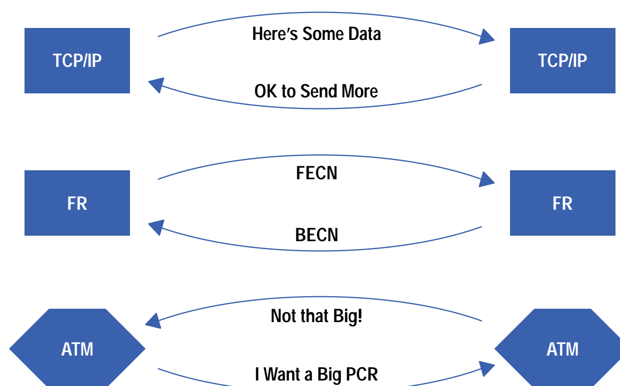


Fig. 3. *Flow control across protocol layers. PCR stands for peak cell rate.*

Example 2: MPEG video and ATM traffic policing. MPEG video over ATM presents special testing problems, particularly when interactions with TCP/IP are involved. Excessively bursty traffic introduced by layer interaction causes PCR (peak cell rate) violations, which when policed at the ATM layer cause video to freeze (Fig. 4).

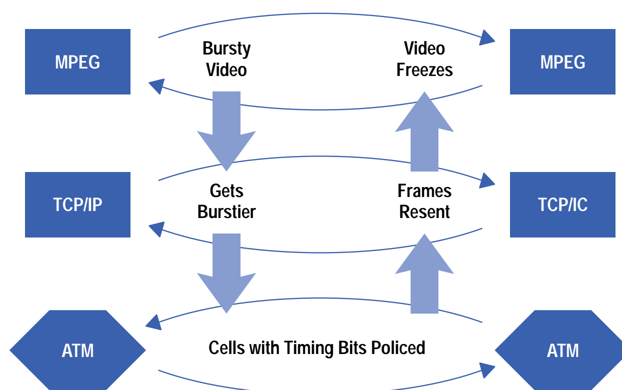


Fig. 4. *MPEG video and ATM traffic policing.*

Example 3: Bit error multiplication and cell loss tyranny. Consider a simple FDDI interconnect service running over an ATM backbone. Because error correction is handled by the higher protocol layers, data services are especially sensitive to cell loss. In this example (Fig. 5), two bit errors in the physical (transport) layer cause an ATM cell to be discarded. This in turn causes an entire 8000-byte FDDI frame to be retransmitted. The net effect is that two bit errors cause an additional 167 cells to be transmitted. In a usage billing scenario, this becomes very expensive. This effect, also known as ***cell loss tyranny***, can result in a downward performance spiral. As frame retransmissions increase, ATM layer congestion may occur, causing more cells to be lost and frame retransmissions to increase even further. To isolate the cause of high frame retransmission rates, testing must be performed across all protocol layers.

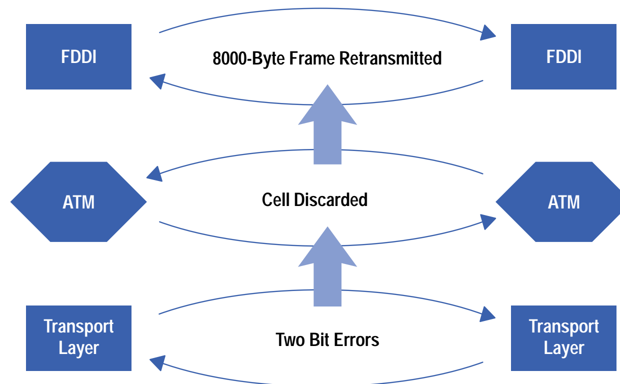


Fig. 5. *Bit error multiplication and cell loss tyranny.*

HP E5200A Broadband Service Analyzer Overview

The HP E5200A broadband service analyzer is a powerful and flexible tool designed to:

- Assist ATM equipment and service installation
- Minimize troubleshooting time
- Assist in meeting service delivery requirements
- Test the interworking of broadband services
- Help manage the performance of broadband networks
- Assist in guaranteeing end-to-end quality of service.

The broadband service analyzer puts users in control by providing service-focused measurements. It allows users to deal with the practical realities of building, operating, and managing broadband networks and to achieve end-to-end service quality. By using the broadband service analyzer, users can be confident that a customer's services are operating reliably and that they can manage the services to maintain a high level of performance and reliability.

The test capabilities provide the information needed to determine the health of a network at all layers of the ATM protocol stack, from the physical layer right through to the AAL (ATM adaptation layer) and above. It can decode LAN protocols running directly over ATM or running over ATM via Frame Relay. It can quickly determine how a customer's services are performing.

System Configuration

The service analyzer consists of four major elements:

- An intelligent base unit that houses the measurement system hardware and software
- A choice of interchangeable line interface modules (interface pods)
- A choice of applications that support analysis of higher-layer protocols and remote testing
- An X display terminal such as a notebook PC or a UNIX® workstation.

The service analyzer can be expanded and updated in the field as new applications are developed. The compact size of the unit means that it can easily be carried onto an aircraft as hand luggage.

Hardware Architecture

The broadband service analyzer's main hardware processing comes from 17 Xilinx XC4013 FPGAs incorporating approximately 221,000 gates. Three i960 RISC processors are used to control the statistics processing, protocol processing, and display graphics and a 500-Mbyte hard disk drive is included. The major functional components are shown in Fig. 6.

The line interface pods contain the line interface circuitry for each interface with capabilities for hot insertion and autoconfiguration. Line interface pods provide physical layer transmit and receive functions and test capability.

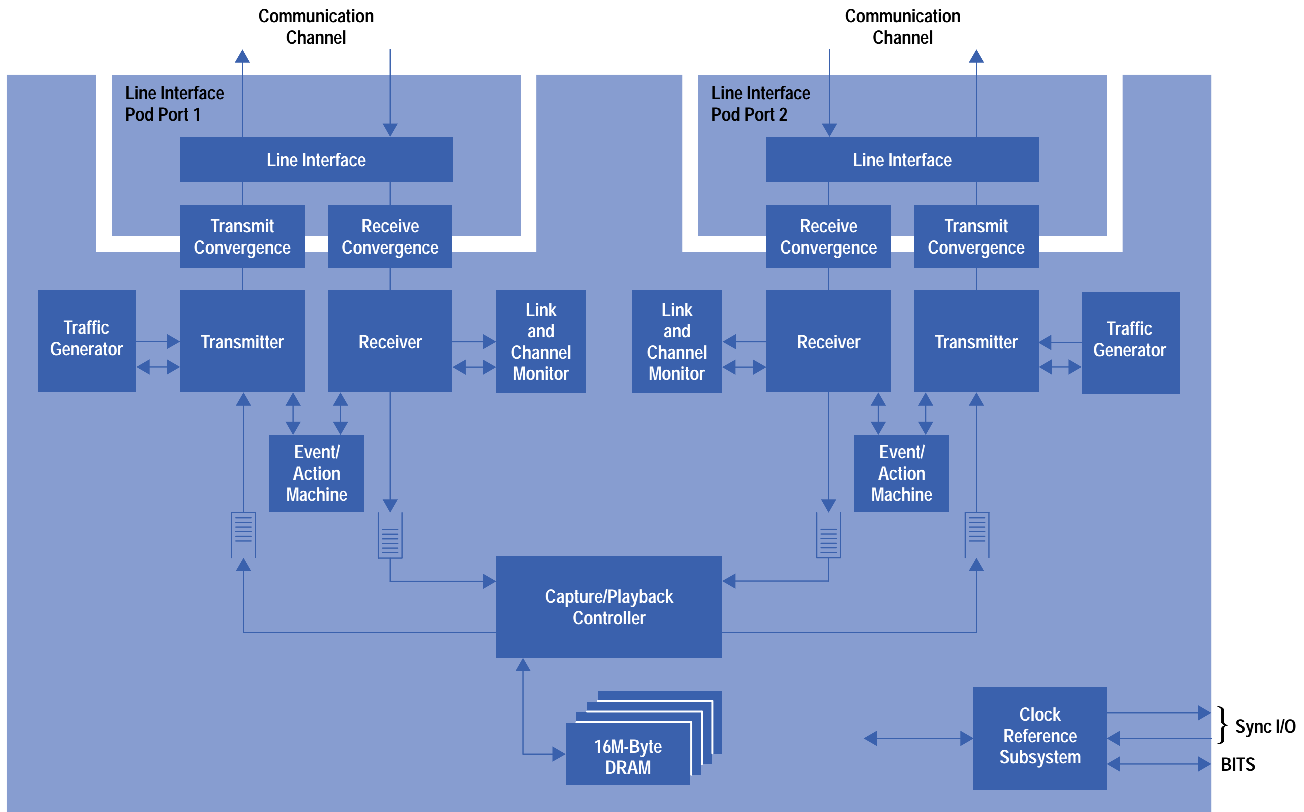


Fig. 6. *Broadband service analyzer hardware architecture.*

The three CPUs are on the main CPU card. The graphics CPU responsibilities include running the X-Windows host via the LAN interface, handling I/O, and interfacing to the hard disk drive. The protocol CPU interfaces with the instrument hardware, controlling the transmit and receive functions for both ports and performing higher-layer processing of the data. The statistics CPU performs the real-time measurements for the link and channel monitors of both ports.

The capture/playback subsystem controls the 16M bytes of DRAM, partitioning it between the two ports as required, adding receive timestamps, and controlling transmit playback.

The traffic generator controls the traffic simulator and the alarm and error simulator functionality. A transmit cell sequence can contain up to 1500 cells.

The link and channel monitor performs real-time processing on up to 1024 different received channels on each port with determination of cell counts, AAL type detection, error statistics, and reassembly statistics.

The event/action machine provides trigger event detection and subsequent analyzer action control.

The clock subsystem provides the clock signals for the different elements of the analyzer. Clock sources that can be used include an internal Stratum 3 reference, a BITS source, or external I/O clock sync.

Software Architecture

The broadband service analyzer's software is designed with object-oriented techniques using managed objects (see ***Article 11***). While this increased the initial development time as new techniques were learned, the use of object-oriented technology will make it possible to add future enhancements and new applications much more efficiently and flexibly than if traditional software design had been used.

The software architecture (Fig. 7) is based around several distinct functional subsections:

- The application presentation subsystem provides the graphical user interface.
- The analyzer subsystem provides SMARTtests, the TCL (Tool Command Language) user scripting environment, and BSTS UPE (user programming environment) compatibility.
- The protocol support subsystem provides protocol analysis and PDU (protocol data unit) segmentation, reassembly, and filtering.
- The measurement and control subsystem provides control of the measurement hardware.

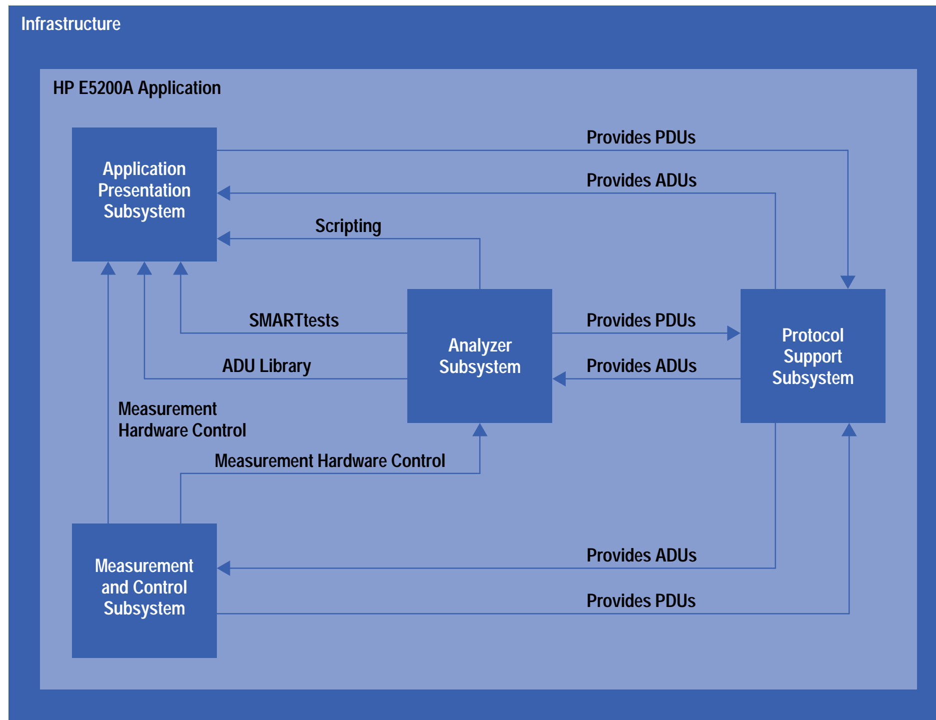


Fig. 7. *Broadband service analyzer software architecture. An ADU is an internal format container of PDUs (protocol data units) plus control information.*

The infrastructure is based on the VxWorks real-time operating system and includes elements to perform basic computing, platform services, operating support, and user interface services. The GUI is based on the X Window System, allowing control via the LAN connection from any X terminal.

Broadband Service Analyzer Features

The analyzer's powerful real-time measurement system allows users to perform extensive real-time measurements and instantly get reports on the status of broadband services. Link and channel information is updated to determine:

- Active VPI/VCI channels
- AAL type and errors
- Network utilization
- Network errors
- Network alarms
- Cell counts.

Measurements can be correlated to identify important interactions between the layers of the broadband protocol stack and any interference between individual services on different ATM channels or on different ports.

Customers are geographically dispersed, and there is a need to be able to guarantee the end-to-end quality of their service. The distributed object measurement technology of the broadband service analyzer makes remote measurements a normal mode of operation. Coupled with the optional HP Broadband Launch Pad software application, access and management of multiple service analyzers distributed throughout the network is possible. This helps reduce the cost of managing broadband networks, improves the speed with which faults are isolated and repaired, and provides a single consistent set of tools for both remote and dispatched applications.

The service analyzer's graphical user interface can be used by people of all skill levels. The measurements made by the service analyzer are displayed so that they are easy to interpret. Users can quickly and easily see what is happening to a customer's service and make an informed decision on the appropriate action to take. Link and channel monitors guide users through the service analyzer's extensive monitoring capabilities, directing the user to the problem areas.

Testing Made Easy

Testing is made easy with the broadband service analyzer, as illustrated in Fig. 8. The user simply connects the service analyzer to a network and the link monitor immediately gives a summary of the health of the network. To look more closely, the user can select a channel monitor to obtain a graphically correlated view of the measurements relevant to a particular

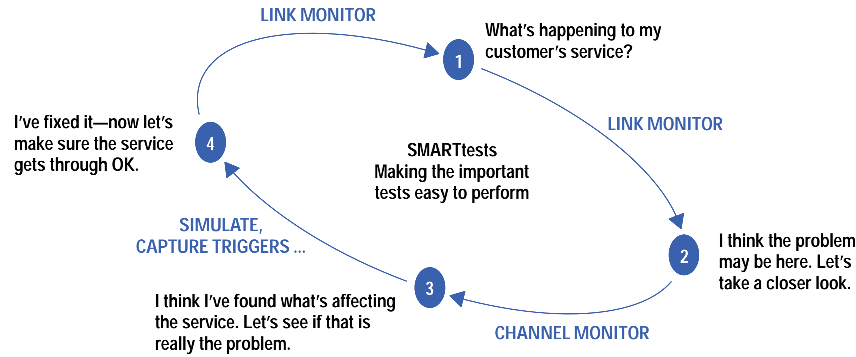


Fig. 8. *The HP E5200A link monitor, channel monitor, and SMARTtests support each other to simplify testing.*

channel. Users can immediately see whether service level problems are related to problems in the other layers of the protocol stack.

A range of predefined SMARTtests reduce the complexity of performing important tests such as cell loss and cell delay. Testing can be started rapidly with one or two clicks of the mouse. Users can also create their own tests by using the macro programming function of the service analyzer. This function provides a record/playback system that creates test programs by automatically recording the results of actions on the service analyzer.

The service analyzer's sophisticated traffic simulation and data capture systems enable users to inject data into the network to exercise the network' s capabilities, simulate specific problems, and then analyze the results as data is captured.

The service analyzer's flexibility is further enhanced by the wide range of interfaces its platform can support, including E1, E3, DS1/DS3, OC-3/STM-1, OC-12/STM-4, 6.3-Mbit/s, and 155-Mbit/s coaxial. Available immediately are OC-3/STM-1 single-mode and multimode optical, STM-1/STS-3c electrical, E3, and DS1/DS3 interfaces.

UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company Limited.

X/Open is a registered trademark and the X device is a trademark of X/Open Company Limited in the UK and other countries.

Glossary

The following are definitions of some of the terms used in this article.

AAL. ATM Adaptation Layer. The AAL translates services from their native format, such as variable-length frames, into fixed-size ATM cells. It returns the service data to its original form at the destination.

ATM. Asynchronous Transfer Mode. A transmission technology that provides high bandwidth, low delay, packet switching, and multiplexing. Usable capacity is segmented into ATM cells that are allocated to services on demand. ATM is designed to handle a variety of service types, such as data, voice, and video.

BECN. Backward Explicit Congestion Notification. A bit in a Frame Relay header that indicates that frames transmitted on this connection may experience congestion (backward notification of congestion).

BITS Source. Building Integrated Timing Source. A reference clock, typically distributed throughout public networks, that is used to synchronize network elements.

Cell. ATM cells have a fixed size of 53 bytes. They consist of a 5-byte header that carries routing information, and a 48-byte payload that carries service data.

FECN. Forward Explicit Congestion Notification. A bit in a Frame Relay header that indicates that the current frame has encountered congestion over the connection (forward notification of congestion).

Frame Relay. A variable-size packet service operating from 56 kbits/s to 2 Mbits/s. Frame relay can efficiently transport high-speed, bursty data, but does not support services that are highly sensitive to transmission delay (such as voice). A number of Frame Relay services can be multiplexed onto a high-speed ATM connection.

LAN. Local Area Network. A short-distance data communications network, typically within a building or campus. A LAN service can be transported directly over an ATM connection or a lower-speed WAN technology such as Frame Relay or SMDS.

Multiplexing. Merging several different signals into one source and separating them at the destination.

PCR. Peak Cell Rate. PCR is one example of a traffic parameter specified in a contract between an ATM network operator and a customer. As long as the customer's service does not exceed the specified PCR, it should not incur any cell loss.

SMDS. Switched Multimegabit Data Service. A high-speed, public, packet-switched data service. SMDS extends LAN capabilities over wider areas. A number of SMDS services can be multiplexed onto a high-speed ATM connection.

Stratum-3 Reference. A highly accurate and stable reference clock, used in network equipment and communications test equipment.

SWG. Sub-Working Group. A group of people within a standards organization who propose new standards or recommendations.

VC. Virtual Circuit. An ATM connection between two endpoints, identified by a VPI/VCI.

VP. Virtual Path. A collection of virtual circuits, grouped together for routing purposes, sharing a common VPI.

VPI/VCI. Virtual Path Indentifier/Virtual Channel Identifier. A field in the ATM cell header that provides routing information.

WAN. Wide Area Network. A network that operates over an extended geographic area. ATM, Frame Relay, and SMDS are examples of WAN technologies.

Managed Objects for Internal Application Control

Managed objects are fundamental to the software architecture of the HP E5200A broadband service analyzer. Typically used to control remote network elements, managed objects are also used internally by the service analyzer's application to control application objects.

by John P. Nakulski

A managed object is a software abstraction that acts as a proxy for a real object. It is used by remote clients, such as objects in other threads of execution, or on remote hosts. Managed objects are often used in network management applications to enable control of remote network elements. For example, a router or switch might offer a managed object interface to enable remote configuration and monitoring from a central network management system.

In the HP E5200A broadband service analyzer, managed objects are used in a different way. The service analyzer's application uses managed objects internally to control application objects. For example, a CellLossTestMO managed object is used to control the service analyzer's CellLossTest object. An interface to the same managed objects is also available to external users.

Besides being fundamental to the service analyzer's software architecture, managed objects have proved useful in other ways, including:

- In the decoupling of the user interface
- As the foundation for a command line interface
- In the development of a macro recording facility
- For backward compatibility with a programmer's interface
- In the unit and regression testing of software components
- For the simultaneous control of several remote service analyzers for distributed testing of broadband networks.

Remote Object Communication

The service analyzer contains three Intel i960 microprocessors running VxWorks—a lightweight, multithreaded operating system. Each processor runs several threads of execution. A layer of software called ***object transport*** provides an interface for streaming ***elemental types*** between threads of execution on the same processor, between processors on the same host, and between hosts on the same network (see Fig. 1). Elemental types include integers, floating-point numbers, characters, strings, and a few classes such as PDU (protocol data unit).

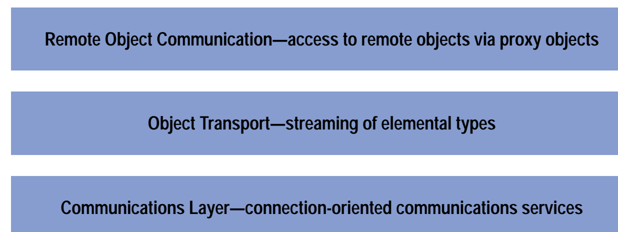


Fig. 1. *Layered communication architecture of the HP E5200A broadband service analyzer.*

Remote object communication is a software layer built upon object transport that makes object services available to remote clients. Using a proxy object, a client object can invoke the member functions of an object in another thread or on another host almost as easily as calling member functions of objects local to its own thread (see Fig. 2). Of course, each argument of the member functions must be an elemental type.

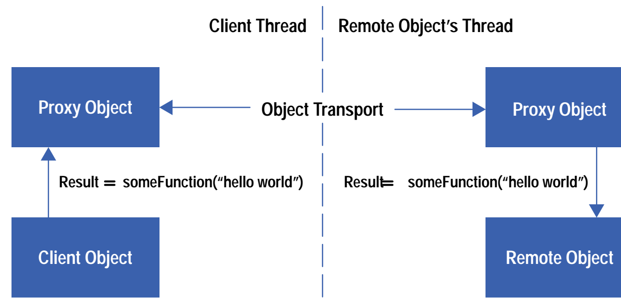


Fig. 2. Remote object communication.

Objects, Attributes, Agents, and Managers

A managed object software layer builds upon the remote object communication layer to provide a dynamically discoverable software interface. The interface consists of named managed objects, each containing a set of named attributes. While it is running, an application can be ***browsed*** through this interface for a list of its managed objects and their attributes, each of which can be accessed by name.

A managed object's executable attributes, which can be invoked by name, enable remote execution of an object's functions. For example, the service analyzer's CellLossTest managed object contains an attribute named start which runs a test that measures loss of ATM (Asynchronous Transfer Mode) cells. A managed object's data attributes, which can be read and written by name, typically mirror the state of the object. For example, the CellLossTest managed object has an attribute named cellLossCountThreshold, which represents the number of cells that can be lost before the cell loss test fails.

Each thread that contains managed objects also contains a ***managed object agent***, which is responsible for dispatching inward communications by calling the desired function on the specified managed object. Each thread that uses managed objects contains a ***managed object manager***, which is responsible for finding and communicating with managed objects in remote threads and on remote hosts. These distributed managers and agents cooperate to provide clients with a single, unified view of managed objects (see Fig. 3).

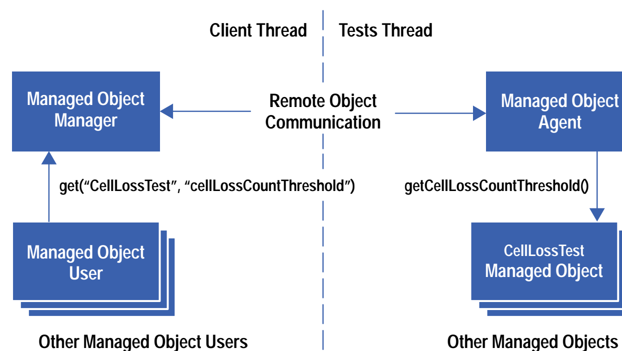


Fig. 3. Managed object communication.

Automatic Code Generation

Each of an application's managed objects is specified in a language-independent ***managed object definition*** (MOD) file. A MOD file specifies and documents the public interface to a managed object. An example of a MOD file is shown in Fig. 4.

After a managed object is specified, the developer uses a customized Interface Description Language (IDL) compiler to parse the MOD file. This produces C++ code for the ***front-end class*** (used by the client of the managed object to encapsulate the behavior of the remote managed object) and skeleton C++ code for the ***back-end class*** (the managed object itself), which the developer must complete.

For example, to use the CellLossTest object from another thread or another host, the client would construct an object of the compiler-generated CellLossTestFE (front-end) class. This offers member functions for invoking executable attributes (such as start() to invoke the start attribute) and member functions for reading and writing data attributes (such as getCellLossCountThreshold() and setCellLossCountThreshold() to read and write the cellLossCountThreshold attribute). The front-end class implements all such functions as operations on the managed object manager.

The thread containing the CellLossTest object itself also contains an instance of the compiler-generated CellLossTestMO (managed object) class. The managed object agent passes the incoming operation to the managed object. The developer needs to ***connect*** the CellLossTestMO class with the CellLossTest class so that incoming requests such as getCellLossCountThreshold() are

```
#include <elementalTypes.idl>

interface CellLossTest {
  attribute AtoUInt16T transmitVpi;
  attribute AtoUInt16T receiveVpi;
  attribute AtoUInt16T transmitVci;
  attribute AtoUInt16T receiveVci;
  attribute AtoUInt32T durationThreshold;
  attribute AtoFloat64T cellLossRatioThreshold;
  attribute AtoUInt64T cellLossCountThreshold;
  AtoStatusT start();
  AtoStatusT stop();
  AtoStatusT getResultSet(
    out CtsPassFailResultE passFailResult,
    out AtoStringT testReason,
    out AtoUInt32T elapsedTime,
    out AtoUInt64T transmitAverageBandwidth,
    out AtoUInt64T receiveAverageBandwidth,
    out AtoUInt64T cellLossCount,
    out AtoFloat64T cellLossRatio);
...
}
```

Fig. 4. *An example of a managed object definition (MOD) file.*

delegated to the CellLossTest object (see Fig. 5). In this simple example, the CellLossTestMO::getCellLossCountThreshold() function body is filled in by the developer with a call to the CellLossTest::getCellLossCountThreshold() function.

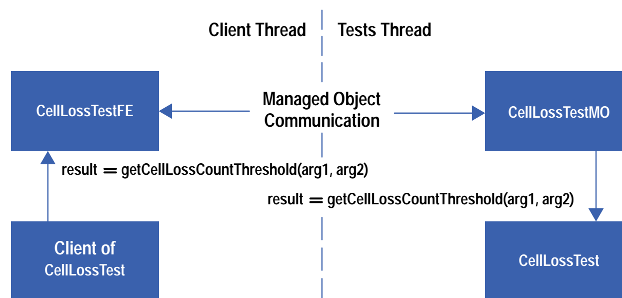


Fig. 5. *Interaction of machine-generated objects. (FE = front end. MO = managed object.)*

Managed object operations are generally ***synchronous***, meaning that the front-end function does not return until the back-end object has completed its execution and has a result available to return. The managed object notification operation, however, is ***asynchronous***. Clients ***subscribe*** to a managed object attribute, supplying a callback function. When the attribute is ***triggered***, a notification message is sent and the supplied callback function is invoked. An attribute is generally triggered when it changes or is executed. For example, the testStatus attribute of the CellLossTest managed object is triggered when the test is started or stopped.

Each managed object data attribute must be an elemental type. Each elemental type corresponds to either a C++ built-in type (such as char) or a low-level class (such as Pdu). Developers of a front-end application using another language need to port only those low-level classes and types that are not native to the language.

Reactive Graphical User Interface

The service analyzer's graphical user interface (GUI) is said to be fully ***decoupled*** from the rest of the service analyzer. The GUI is a separate subsystem (module) and runs in a separate thread, and can even run on a separate host. Other subsystems are not dependent on the GUI; the rest of the service analyzer can run without it.

Further decoupling is achieved through the use of managed objects. Every request from the GUI to another object is a managed object operation. The GUI uses the notification mechanism to learn of and react to the service analyzer's state changes and other events. For example, if an ATM alarm occurs, the GUI is notified and updates the state of its window. This also allows the GUI to react automatically to changes caused by other managed object users (such as the command line interface).

Armed with only the MOD files, a specification of the elemental types, and a specification of the remote object communication and object transport protocols, a developer can write a new user interface in a different language to run on a different host.

Command Line Interface

The design of the service analyzer's command line interface (CLI) exploits the fact that the service analyzer's functionality is available through its managed objects. The CLI ***plugs into*** the service analyzer at the managed object interface. Each command is interpreted as an action on a managed object, enabling CLI users to control the service analyzer without a GUI.

Tool Command Language (TCL), a public-domain script language, is used as the foundation for the CLI. TCL is embedded in the service analyzer and has been extended with a few simple but powerful functions to enable access to managed objects.

The moexec command enables executable attributes to be invoked, while moget and moset allow the user to read and write managed object data attributes. The molst and moattrlst commands retrieve the set of available managed objects and the set of attributes for a specified managed object, enabling the user to browse the service analyzer's interface dynamically.

If access to the GUI is not available, the user can log in to the service analyzer remotely using a telnet session to run the CLI.

Macro Recording and Playback

By tapping the communications between the GUI and the rest of the service analyzer at the managed object interface, it is possible to watch the GUI's interaction with the underlying service analyzer. This is exactly how the service analyzer's macro feature works (see Fig. 6).

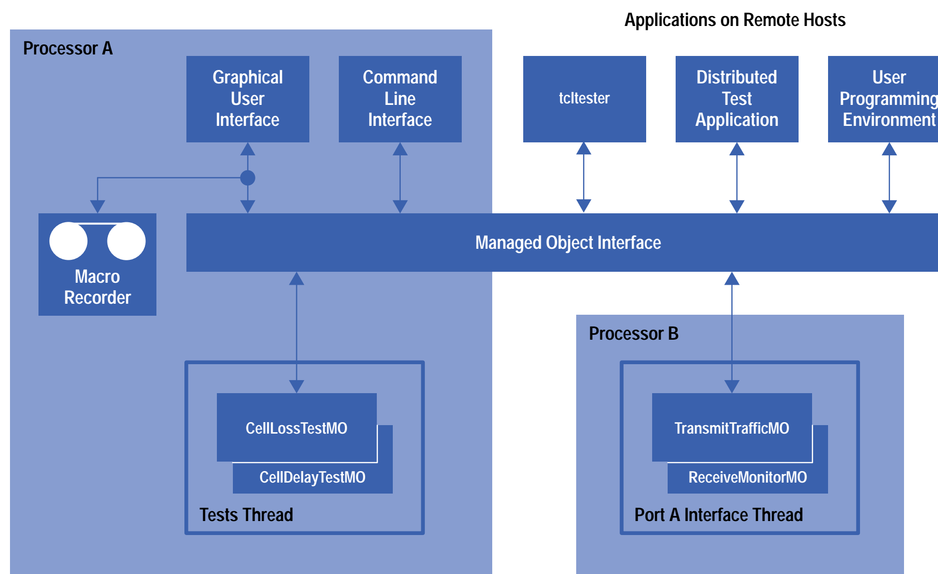


Fig. 6. *Users of the managed object interface.*

When the macro recording facility is used, a subset of the GUI's managed object operations is converted into CLI commands and is recorded into a macro file. For example, a user may want to record a frequently used cellLossTest configuration. When the macro is played back, it is executed in much the same way as a CLI session. If the GUI is active during the playback of the macro, it will respond to service analyzer state changes, allowing the user to watch the effects of the macro as it is running.

Since the macro language is identical to the CLI language, the experienced user can edit and customize a recorded macro or create a new macro. For example, a cell loss test can be executed ten times in succession with increasing transmit bit rates simply by wrapping a for loop around a recorded macro, using a TCL variable to hold the transmit bit rate (see "***Macros***.")

User Programming Environments

Backward compatibility with the HP broadband series test set (BSTS) user programming environment (UPE) is also achieved using managed objects. UPE programs, written in C, can be used to control the service analyzer from an HP-UX* workstation or from a PC by calling functions published in an application programming interface (API).

UPE programs are recompiled on the target host and linked with a new set of supplied libraries. The new libraries implement each API function by invoking one or more of the service analyzer's managed object operations via a managed

object manager. An added bonus for users of the new service analyzer is the ability to control several service analyzers from within a single UPE program.

Testing the Service Analyzer

A further bonus gained from extending TCL with managed object commands was the development of tcltester, a tool used by developers to test the service analyzer. tcltester wraps the CLI into a standalone application that can be executed on a host remote from the service analyzer under test. The developer can enter CLI commands interactively to test and debug the service analyzer or execute whole CLI scripts for regression testing.

Distributed Test Applications

Because the service analyzer's managed object communication layer enables communication not only between threads on the same processor but also between hosts on the same network, it is possible for one service analyzer to control one or both ports of a second, remote service analyzer. A CLI user or a standalone PC application can control several remote service analyzers simultaneously. For example, traffic can be injected into different points in a network, several network nodes can be monitored simultaneously, and performance parameters can be measured and collated across an entire network.

One of HP's goals is to develop useful tests for the installation and commissioning of broadband networks. As we gather experience in using the service analyzer for broadband testing, we expect to develop new tests to add to the test suites.

Although we have been closely involved with users during the development of the service analyzer, it is likely that users will have requirements for distributed tests unimagined by our developers. By providing an open managed object interface, users can develop their own broadband test and measurement applications.

Acknowledgments

I wish to acknowledge the contributions of the many individuals who participated in the design and development of software for the HP E5200A broadband service analyzer, especially Earl Chew, who had the foresight to provide an open software interface, and Chris De Souza, who embedded TCL, developed the macro capability, and fostered the architectural vision in the face of tight deadlines. This investment has already enabled the creation of several new applications. In addition, I would like to thank Dan Smith for helping develop the example macro and Sonja Draeger, Gail Hodgson, and Debbie Augusteyn for their helpful suggestions and thorough review of this paper.

HP-UX 9.* and 10.0 for HP 9000 Series 700 and 800 computers are X/Open Company UNIX 93 branded products.

UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company Limited.

X/Open is a registered trademark and the X device is a trademark of X/Open Company Limited in the UK and other countries.

Macros

The HP E5200A broadband service analyzer uses TCL (Tool Command Language), a public-domain script language, as the foundation for its command line interface and macro recording and playback feature. The following macro configures and runs the cell loss test several times in succession with increasing bandwidths.

```
##########################
#
# This macro runs a series of cell loss tests at
# increasing bandwidths.
#
##########################

### Initialize inputs
set vTransmitVpi 200
set vTransmitVci 88
set vReceiveVpi 200
set vReceiveVci 88
set vTestDuration 15
set vCellLossRatioThreshold 0.001

### Configure input parameters
# Use port A for both transmit and receive.
moset _Test CellLossTest transmitPort _PortA
moset _Test CellLossTest receivePort _PortA
# Configure transmit and receive channel
# parameters
moset _Test CellLossTest transmitVpi
  $vTransmitVpi
moset _Test CellLossTest transmitVci
  $vTransmitVci
moset _Test CellLossTest receiveVpi $vReceiveVpi
moset _Test CellLossTest receiveVci $vReceiveVci # Use a ‘constant’ traffic profile.
moset _Test CellLossTest trafficProfile
  M_CONSTANT
# Configure duration of test
moset _Test CellLossTest durationThreshold
  $vTestDuration
# Threshold on cell loss ratio, not on cell loss
# count.
moset _Test CellLossTest
  cellLossCountThresholdActive M_OFF
moset _Test CellLossTest
  cellLossRatioThresholdActive M_ON
# Configure cell loss ratio threshold.
moset _Test CellLossTest cellLossRatioThreshold
  $vCellLossRatioThreshold

### Fetch the bandwidth limits for the
### ‘constant’ traffic profile.
moexec _Test CellLossTest
  getConstantProfileLimits
  {vMinBandwidth vMaxBandwidth}
### Prompt user for the bandwidth step.
puts ”The bandwidth range is $vMinBandwidth to
  $vMaxBandwidth b/s”
puts ”Enter the bandwidth step.”
set vBandwidthStep [gets stdin]

### Display cell loss test configuration.
puts”
Cell loss test:
===============
```

```
Transmit VPI: $vTransmitVpi
Transmit VCI: $vTransmitVci
Receive VPI: $vReceiveVpi
Receive VCI: $vReceiveVci
Test duration: $vTestDuration
Cell loss ratio threshold:
  $vCellLossRatioThreshold
Min bandwidth: $vMinBandwidth b/s
Max bandwidth: $vMaxBandwidth b/s
Bandwidth step: $vBandwidthStep b/s"

### Run the test several times, using bandwidths
### ranging from minimum to maximum using the
### bandwidth step specified by the user.
set vTestnumber 0
set vCurrentBandwidth $vMinBandwidth
while {$vCurrentBandwidth < $vMaxBandwidth} {
    incr vTestnumber 1
    puts "Running test number $vTestnumber
      - bandwidth $vCurrentBandwidth"
    moexec _Test CellLossTest setConstantProfile
      $vCurrentBandwidth

# Start the test.
    moexec CellLossTest start {}

# Poll the test to see if it is still running.
    while {[moget _Test CellLossTest testStatus]
      == "CTS_RUNNING"} {
# Fetch the interim test results.
        moexec _Test CellLossTest getResultSet
          {vPassFailResult vTestReason
           vExecutionTime
           vAverageTransmitBandwidth
           vAverageReceiveBandwidth
           vCellLossCount
           vCellLossRatio}
    }

### Fetch and display the final cell loss test ### results.
    moexec _Test CellLossTest printResultSet
      {vResults}
    puts"
Test number : $vTestnumber
Bandwidth   : $vCurrentDisplay
$vResults
=============================="

# Calculate the bandwidth to be used for the
# next test.
    set vCurrentBandwidth [expr
      $vCurrentBandwidth + $vBandwidthStep]
}

puts "******* End of macro ********"
## End of macro ##
```

Developing a Design for Manufacturability Focus

The HP Australian Telecommunications Operation has rapidly evolved from a custom test instrument developer to an operation that develops and produces products in higher volumes. Significant cultural and technological hurdles have been overcome during the transition to an operation focused on design for manufacturability.

by John G. Fuller

The need for organizational change can be driven by many factors. In the case of the HP E5200A broadband service analyzer, the HP Australian Telecommunications Organization (ATO) intended to address a different market through a significant step forward in product technology.

ATO's first product was a custom digital performance monitoring system, delivered during the late 1980's. ATO had also recognized an immediate market need for high-performance SONET/SDH (Synchronous Optical Network/Synchronous Digital Hierarchy) test instrumentation, and in 1990, began work on the definition of a VXIbus test system. This product was delivered in the second half of 1991.

These first ATO instruments were largely based upon through-hole manufacturing technology. They used some surface mount technology for the high-speed interface connections to customer equipment. The technologies did not present a difficult manufacturing problem and the testing approach was largely of a functional nature, with some external fault diagnostic tools being developed by ATO manufacturing engineering. These tools consisted of field-programmable gate array (FPGA) configurations that could be used to test the large amount of digital interconnect within the product. Shipment volumes were low enough to be handled by a very small production group which outsourced assembly work wherever possible. It is worth noting that the marketing planning, development, and manufacturing design were done concurrently. This led to a relatively short time to market and provided ATO with a sound cultural basis for future concurrent development of new products.

The next couple of years saw very little change in technologies or volume. Essentially, ATO was making printed circuit assemblies that were easy to build, dealing with product volumes that didn't demand an abnormal amount of effort in design for assembly (DFA) or design for test (DFT), and scrapping the occasional printed circuit assembly that defied all attempts to make it work.

This approach was far from the optimal situation for the manufacturing group, but it was manageable and paid dividends in terms of ATO's profit line and growth. The down-side of the equation was that, while concurrent engineering was being used to some extent during this time, a culture evolved that tended to minimize manufacturing involvement in the design process. This reduced the manufacturing contribution to the design process and limited the interaction between hardware designers and manufacturing engineers. While many reasons for the reduced involvement could be suggested, we feel that the most likely cause was that a lack of product technological change led to less emphasis being placed upon development and manufacturing process improvements. This also led to more emphasis being placed upon manufacturing contributions to functional and quality assurance testing during development. Thus, ATO's manufacturing staff was more involved in the test process than the development process while operating within a low-volume, high-cost, high-mix, and relatively high-profit manufacturing environment.

It is from this cultural atmosphere that ATO's strategic direction shifted away from the R&D market toward the installation market. Associated with the shift was a market requirement to reduce both the size and cost of our product without detracting from its performance. Thus, ATO's manufacturing group was gazing down the development pipeline at a high-technology product that would lead to significant increases in volume, lower selling prices, and eventually, lower profit margins. Clearly, these were very good reasons why change was seen to be necessary.

What Needed to Change?

Having established the need to change, it was necessary to evaluate what processes would have to change to meet the needs of the new marketplace. These were seen to be largely driven by the expected technology leaps in the new product, the volume increases, and the lower profit margins that were expected. Time to market was also seen as a critical factor that had to be minimized in every way possible.

The technological issues were of great significance. ATO was moving from product platforms based largely upon through-hole technology to platforms based largely upon surface mount technology. Worse (or better?) still, the use of surface mount technology was expected to be extreme, with the main printed circuit assembly (400 mm by 230 mm) carrying over 1000 components needing almost 8700 solder joints spread over both sides of the board. The degree of difficulty associated with the solder joints was also of significance: approximately 4500 of them were expected to be extra-fine-pitch (XFP) at 0.020-inch spacing with a further 900 expected to be fine-pitch (FP) at 0.025-inch spacing. A difficult task even for a world-class manufacturer!

Aside from the technology issues, it was also apparent that the small size of the printed circuit board would seriously limit the amount of physical test access, thus requiring an innovative testing approach. With approximately 3000 electrical nodes and access to about only 600 of these because of space limitations, it was clear that diagnostics in the production area (and later in the field) were going to be of paramount importance. Additionally, each printed circuit assembly was expected to have a significantly high component cost. It was never going to be an option to simply scrap the boards that were difficult to repair—a daunting thought when ATO's estimates were predicting around one defect per board after in-circuit test at the manufacturing center!

With the ability to accurately diagnose, test, and assemble the product being so vital to manufacturing success, it was quite apparent that manufacturing engineering teams responsible for materials, test, process, and regulatory functions would need to form and maintain a close alliance with the hardware and software design teams. This would require organizational efficiencies and working relationships, particularly between R&D and manufacturing, on a scale not seen before at ATO. Development cycles would need to shorten still further, placing additional pressure on all aspects of the organization's performance. Effective teamwork and communication skills would have a major impact on the overall success of the product.

In summary, major technological and process improvements would be required, as would significant cultural and behavioral shifts within the organization, to support the technical changes.

Managing the Change

To produce the required technological and process improvements, manufacturing engineering needed to develop a thoroughly different core focus from doing QA testing in parallel with other manufacturing tasks. It was recognized that a significant contribution to the design process could be achieved only by manufacturing engineers becoming more specialized and focused upon particular job functions. Technically, this meant that as manufacturing engineering functions such as test, process, regulatory, and materials became adequately skilled, it was possible to increase the manufacturing contribution to the development process. This ultimately led to a closer involvement with design teams at a much earlier stage in the product life cycle.

Manufacturing targets also became a more important part of the design process, so that measurement of the cost drivers associated with building a product were given more visibility during the early development stages. Manufacturing engineers initiated a sustained effort to review design for manufacturability (DFM) issues, set appropriate assembly, test, and diagnostic targets, and then provide regular feedback into the development process. All the manufacturing engineering functions mentioned previously were involved, an excellent example being the EMC design process (see ***Subarticle 12a***). The manufacturing processes were also subjected to the same kind of scrutiny. Clear targets were assigned for each station in production, and engineering owners were responsible for driving their station toward meeting targets, thus enabling a subsequent handover to the production staff.

With the sharpening of focus on materials, regulatory, process, and BIST (built-in self-test) engineering functions, there was a need for increased contact with the hardware and software designers. This required a significant effort in terms of teamwork and communication skills, and a cultural shift at both the manufacturing and R&D ends of the development process. For example, as manufacturing engineers prepared to contribute to the design process with manufacturing inputs in a way that would benefit the design process and the manufacturing function, designers also had to prepare to take and constructively question the inputs provided by their downstream customers. A cultural shift of this nature does not happen easily and it required an environment of significant trust, teamwork, and communication that was developed over a period of time. The objective was for manufacturing engineers to become design-capable partners for development who can continually improve the DFM process.

Development partnerships were also used to great effect in other areas. ATO has been working successfully with HP's Surface Mount Technology Center in Spokane, Washington for several years. The Surface Mount Technology Center was heavily involved in the surface mount process development and the in-circuit test development for the service analyzer product. Manufacturing partnerships for the plastic components (cases, handles, feet, etc.) and the aluminium die-cast pod casings also made a significant contribution as ATO began to use unfamiliar technologies. As a result of the service analyzer project, a strong emphasis is now placed on the benefits of concurrent engineering involving HP partners inside and outside ATO, and on working closely with partners outside HP.

Many concurrent engineering best practices have been learned during this product development, with major benefits being derived from:

- An emphasis on teamwork and communication skills by everyone involved

- Early and regular involvement throughout the product life cycle by all groups and individuals charged with contributing to that development
- An emphasis on contribution to the design process rather than limitation of it
- Continual improvement of the product life cycle process through critique of the concurrent engineering effort.

Technical Changes

Many technical changes were necessary as a result of ATO's new direction. Broadly speaking, the technologies involved in the service analyzer project drove significant surface mount process changes (see ***Subarticle 12b***), manufacturing engineering process changes (which were addressed by a sharper focus of engineering resources), and major testing challenges.

The testing challenges were posed by the high component and pin counts, the multiple-CPU configuration, the lack of available test points, the high component costs, and the expected degree of difficulty involved in fault diagnosis. It was forecast that these boards would be of a significantly higher volume than anything else that had been manufactured previously by ATO. These volumes were considered relative to the issues mentioned above, and other factors such as head-count limitations, repair and diagnosis times, inventory, test equipment, and expertise required of technicians. It became very clear that an innovative test approach would be necessary to avoid a sustained engineering involvement in the repair and commissioning process.

Boundary scan (IEEE 1149.1) was the chosen test technology, and its extent and use were largely determined by the expected low number of physical test points on the board. It was linked to a detailed built-in self-test strategy that involved an extensive hardware, firmware, and software effort.

With the available physical test access diminishing during the development process (rather than increasing), a large emphasis was placed on getting access through substitution of boundary scannable components wherever possible. This approach eventually covered some 65% of the electrical nodes, with additions being mainly in the area of glue logic, since the ASICs and FPGAs were already scannable. Manufacturing's view of the potential diagnosis issues made the additions well worth the extra cost of the components. The printed circuit board real estate taken up by the extra pins was far less than the space that would have been taken by physical test points, thus providing the printed circuit board designer with more flexibility.

Test Strategy

The overall testing approach can be summarized as follows (see ***Article 13*** for more details):

- In-circuit test was used at the Surface Mount Technology Center after the boards were loaded, with physical test access being supplemented by boundary scan vectors on the Hewlett-Packard 3070 board tester wherever possible. Sufficient test points were placed on the printed circuit board to allow the scan chain to be broken into convenient chunks to aid in debugging and running tests. One of the design trade-offs included a decision to restrict in-circuit test access to some areas of the board. This was primarily driven by a lack of scan capability and layout difficulties in some of the RAM circuits. The rationale behind this potentially risky decision was that inspection at the Surface Mount Technology Center would pick up any obvious manufacturing defects, and built-in firmware test should specifically confirm the functionality of these areas at ATO.
- Extensive scan testing was used at ATO through the IEEE 1149.1 test access port (TAP). This was done before loading any software or firmware on the printed circuit assembly. It uses high-resolution vectors that are generated and inserted by a PC-based system. Automatically generated vectors are coupled with manually generated cluster test vectors to maximize the available scan coverage. This approach was intended to be a good fault resolution and diagnosis tool. However, more work still needs to be done to determine the overlap with scan testing at the Surface Mount Technology Center. When this paper was being written, the PC tools for insertion of the vectors were fully developed and the additional test and fault coverage was considered to be worth the increase in run-time overhead in the manufacturing process.
- Firmware access to the scan chain was made available through a test bus controller IC, which allowed firmware insertion of high-coverage, low-resolution built-in self-test vectors from the service analyzer hard disk. This test ran automatically (and very quickly) at power-up and communicated simple pass/fail results to a four-character alphanumeric display on the printed circuit assembly. A major advantage of this approach is that new vectors can be added to the hard disk or modified when test deficiencies are found during normal production.
- Some very extensive functional testing also runs from the firmware at power-up and includes at-speed tests of various functional blocks within the product. Checkerboard testing of RAMs and microprocessor communications ports, and staged loopback (transmit to receive) tests, are just part of what has been a significant software and firmware attempt to provide built-in, high-resolution diagnosis capability in the service analyzer. All of these tests can be run individually in a postmanufacture (service) situation.
- The final test approach used in production applies mainly to the parametric aspects of the product—for example, the optical and electrical connections to the outside world. A more traditional (minimized) functional

test method is used here with calibrated equipment and gold standards being used to verify that the product meets the published parametric specifications.

Results

From our manufacturing viewpoint, many breakthroughs were achieved during the development of this product. Perhaps the most obvious would be the technological leaps forward. The hardware components that were selected led to a significant extension of the Surface Mount Technology Center manufacturing and test processes. At ATO, the DFM process is considered to be vastly improved and is expected to be a significant contributor to the success of future projects.

In terms of the bottom line, the question should be asked, “Was the result worth the effort?” This is always a difficult question to answer when a specific measurement has not been defined for intangibles such as working relationships. We are of the opinion that the improvements that have been made in working relationships between R&D and manufacturing have made the effort well worthwhile. However, there were also other, more definable results such as:

- A minimal increase in costs for scannable glue-logic parts
- An increase in available printed circuit board real estate as a result of the minimized requirement for physical test access
- Reduced manufacturing test and fault diagnostic time
- Minimized effect of prototype and pilot build times on time to market
- Reduced impact on development work as a result of R&D involvement in resolving manufacturing defects.

When these results are considered along with our ability to make far better use of these tools next time around, the question becomes much easier to answer in the affirmative.

Acknowledgments

Acknowledgment for the significant efforts that have contributed to the manufacturing success of the HP E5200A broadband service analyzer is a big task in itself and identification of every individual would greatly increase the size of this paper. For that reason, the author has classified the contributors as follows, and would like to express the most sincere thanks to all of them.

- ATO manufacturing engineers
- HP E5200A hardware and firmware designers
- HP Surface Mount Technology Center and Fort Collins test engineers
- Surface Mount Technology Center process engineers
- ATO functional management
- The many people who provided extensive review inputs to this paper.

HP E5200A Broadband Service Analyzer EMC Design

Electromagnetic compatibility (EMC) is the ability of equipment to function satisfactorily in an electromagnetic environment and not to introduce intolerable disturbances to the environment or other equipment.

With any new product design there are many areas of risk. One significant area of risk is EMC conformance. Conformance to international EMC standards is fast becoming mandatory in the international marketplace. Failure to meet mandatory EMC standards inevitably leads to being locked out of many large markets. More commonly, EMC conformance, or lack thereof, can often cause delays to project schedules and increase the manufacturing cost of a product. These factors make it very important to consider EMC conformance at all stages in the development life cycle.

At the HP Australian Telecom Operation (ATO), management of regulatory compliance is primarily the responsibility of the EMC specialist within the manufacturing engineering group. The EMC specialist is chartered with the role of ensuring that products meet all required regulatory standards. This is achieved by providing EMC design expertise and managing the product qualification process. However, compliance cannot be achieved by activities within the manufacturing engineering group alone. To drive the compliance process, close cooperation between the R&D and manufacturing engineering groups is required. Responsibility for new product compliance needs to be shared by members of both groups.

With responsibility shared across two functional groups it is very important to coordinate EMC risk-reducing activities with a well-defined process. Fig. 1 shows an overview of the process implemented for the HP E5200A broadband service analyzer. The thrust behind this process was to consider EMC throughout development, by building in compliance mechanisms from the beginning and then continually measuring progress at regular discrete stages.

Product Definition

EMC activities began early in the product definition stage. It was during this stage that many important attributes were decided, such as size, weight, materials, performance, architecture, and technologies. It follows that this stage was one of the most important stages for building in EMC compliance. An informal design and compliance strategy was used to aid the product definition team. This strategy covered the compliance requirements and test plan, risk level trade-offs (i.e., cost versus technologies versus schedule), and EMC design features.

The compliance requirements and test plan were determined from the nature of the product, the target market's legal requirements, and any additional customer requirements. In addition to the hard legal requirements, extra design margin was added. For the service analyzer project, a 10-dB margin was the target for the first prototype, with 6-dB being the target for continuous production.

The risk level was important in estimating and scheduling the amount of resources required for EMC compliance. The estimate of risk accounts for the technologies involved, development time constraints, and cost constraints. For the service analyzer product, risk was higher than ever before.

EMC design features are the mechanisms required to control EMC. Here many decisions were based on experience with EMC design and on research on current techniques and technologies. Credibility of the EMC specialist was of paramount importance. The product development team had to trust and believe in the requirements that the EMC specialist put forward. Many of the requirements clashed with basic product form and function.

Consultation and Review

During the implementation or product development stage, the EMC specialist assumed the role of consultant and reviewer. Designers were making many important decisions that could compromise or enhance EMC conformance. Informal review and consultation occurred regularly. The EMC specialist was required to work as part of the development team to ensure full knowledge transfer and continual feedback and review.

Pretesting and Prototype Evaluation

Later in the development phase, when prototypes became available, pretesting became the main method of risk reduction. Pretesting initially took the form of shielding effectiveness tests for enclosures and conducted emissions tests for power supplies.

Later, as functionality was added, pretesting covered most of the risky EMC tests such as radiated emissions. This testing occurred at regular intervals to catch potential problems at an early stage and to evaluate design changes and tuning.

The number of iterations of pretesting and design optimization depended on factors such as product cost, time to market, level of risk, and technology used. A strategy still commonly observed is to overdesign for EMC and then use iterations to remove unnecessary components. Because time was critical, this was the approach used for the service analyzer product.

Fig. 1. *EMC design process for the HP E5200A broadband service analyzer.*

Qualification Testing

Qualification testing was conducted on a number of typical production samples in calibrated test environments at certified traceable laboratories. This testing was approached with confidence after pretesting.

Conclusion

The risks presented by compliance with EMC standards can be significantly reduced by careful planning and constant feedback and review throughout product development. Acknowledgment must go to the service analyzer hardware design team. Despite pressing schedules and many constraints, they always considered EMC conformance an important priority. As a result, conformance targets were achieved without last-minute complications.

Cary J. Wright
Regulatory Specialist
Australian Telecom Operation

Bibliography

1. H.W. Ott, *Noise Reduction Techniques in Electronic Systems, Second Edition*, ISBN 0-471-85068-3, Wiley Interscience, 1988.
2. M.I. Montrose, *Printed Circuit Board Design Techniques for EMC Compliance*, ISBN 0-7803-1131-0, IEEE Press, 1996.

HP E5200A Broadband Service Analyzer Surface Mount Assembly

Hewlett-Packard's Surface Mount Technology Center located near Spokane, Washington is a high-mix, low-volume facility. It doesn't have the opportunity to fine-tune an assembly over many long runs. The center builds limited runs of prototypes that have extra-fine-pitch components (XFP, with leads on 0.020-inch centers), but before the HP E5200A broadband service analyzer, only two fairly easy designs had reached production status.

By contrast, the service analyzer processor board is one of the most challenging boards that any HP site builds. One image nearly fills our standard 12-by-18-inch panel. This is not unusual, but having 29 XFP components and 15 FP components (fine-pitch, with leads on 0.025-inch centers) on the top side is quite unusual—just these components represent over 6000 individual solder joints.

Loaded on the rest of the dense top side are various 0.050-inch-pitch and passive devices, along with connectors and other through-hole devices representing over 400 through-hole leads. Adding to the challenge is a dense bottom side, with 15 more FP parts (almost 1000 leads), and many other surface mount parts ranging from 0.08-by-0.05-inch passive parts to large SOJ (small outline J-lead) and PLCC (plastic leaded chip carrier) components. One challenge not present is that there are no 0.06-by-0.03-inch or smaller parts on either side, since this design was started before they were our preferred small part. In total, there are over 1000 components and nearly 8700 solder joints on this printed circuit assembly.

Numerous processing problems had to be solved or worked around to build this complex new board successfully. It would be a very expensive printed circuit assembly to scrap, so being able to get good setups and clean starts each time was important. Stenciling the solder paste and placing the XFP components were two key processes that needed to be highly accurate and robust. Tremendous demand on solder paste stencil accuracy and process control highlighted the fact that our existing printers were not designed to handle 0.020-inch pitch requirements. There was also tremendous potential for solder paste drying and plugging the narrow stencil apertures, or slumping and causing widespread bridging between leads on the assembly. A latest-generation printer was implemented to take advantage of an improved vision approach, programmed control of most setup parameters, parameter feedback controls, and built-in stencil cleaning capability. This eliminated paste alignment problems and virtually eliminated variations resulting from operator dependent setups. Setup time and run time were also reduced.

Planarity (flatness) of the raw board and of the hot-air leveled (HAL) solder pads had a noticeable impact on solder paste stenciling and on placing the XFP and FP components. Most of these parts had to be placed on the top side after the bottom side had already been reflowed, which can allow the board to warp. Compounding this was difficulty in finding adequate board support locations because of the density of the bottom side. Further compounding this was having XFP and FP parts at or near the corners on both sides of the board, where warp can be worst. Placement accuracy is also worst in the corners, well outside the machine's optimum area. A customized tooling approach helped reduce this problem. An extra complication was that some parts were larger than the normal field of view on the pick-and-place machines, and required a camera modification.

Because of the number of components and the preroutes for the many connectors, the panel tends to sag under its own weight. Proper board support was useful not only in the assembly equipment, but also during furnace reflow and in wave soldering. Additionally, it was no small trick to keep the wave solder from bridging leads on the bottom side of the board. Once the board had progressed that far, the many components on the bottom side, some of which were fairly tall, made even the depanel process harder than usual. A revised approach to fixturing solved this challenge.

The initial prototype build was time-consuming and had many assembly process problems. The defect rate was about 30,000 ppm (parts per million) on XFP solder joints and placement accuracy. Numerous process changes (many discussed above) and some design changes were realized by closely monitoring the prototype build process and keeping an open dialogue with customer support engineers at ATO on design for manufacturability (DFM) issues. As a result, defect rates on the first production boards were brought down to a level comparable to simpler XFP boards in high-volume production at HP, despite being built on second and third shifts. This milestone represents a tremendous accomplishment for the Surface Mount Technology Center. Ongoing improvements will bring defect rates down further as production continues.

Wyatt Luce
Process Engineer
Hewlett Packard Surface Mount Technology Center

Production Test Strategy for the HP E5200A Broadband Service Analyzer

Boundary scan and built-in self-test are supplemented by conventional testing techniques. Eight discrete levels of testing were implemented.

by Cary J. Wright

The HP E5200A broadband service analyzer is a highly complex digital system packed into a field-portable case. During its design, it was essential to achieve the highest densities possible by using some of the latest technologies.

To achieve the functionality required, the main processor printed circuit assembly for the product had to be extremely densely populated. This printed circuit assembly contains over 1000 components, including 29 XFP parts, 30 FP parts, and almost 8700 solder joints (see "***HP E5200A Broadband Service Analyzer Surface Mount Assembly***" for further details). Because of the sheer density of the printed circuit assembly, testing by conventional bed-of-nails in-circuit techniques alone was not feasible.

In addition, normal defect rates specified by surface mount placement vendors were in the order of 200 ppm. From this data, each printed circuit assembly was predicted to have at least one defective solder joint after the placement process. Because of the high cost of this printed circuit assembly, it was very important to be able to accurately and quickly identify defects with high success rates. The target yield for the assembly and commissioning process was set to 100%.

Overall Test Strategy

To overcome the challenges presented, it was necessary to implement a comprehensive test strategy that included boundary scan and built-in self-test, supplemented by conventional testing techniques.

With this complex printed circuit assembly it was decided that a multitiered testing strategy would be used. Such a strategy would increase yield by identifying faults early and by verifying functionality in small discrete stages. Another advantage was that the tests could be reused for other purposes such as service, calibration, or confidence tests.

Fig. 1 shows the production test strategy that was implemented for the service analyzer project. Eight discrete levels of testing were implemented. Each level verifies basic functionality before proceeding to the next stage of assembly or testing. The chance of catastrophic and difficult-to-diagnose failures is minimized.

Boundary Scan Technology

Boundary scan technology forms the key to the production test strategy. Boundary scan is often referred to as JTAG (Joint Test Action Group) and is defined by the IEEE 1149.1 standard. Fig. 2 shows a typical IEEE 1149.1-compliant device.

Boundary scan functions of a device are stimulated via the TDI (test data input), TCK (test clock), TMS (test mode select), and TRST (test mode reset) pins. Responses are received from TDO (test data output). For a printed circuit assembly or system, TDI and TDO pins of each device or submodule are connected in a chain, with the TDO of one device or submodule connected to the TDI of another. The TCK, TMS, and optional TRST pins of devices are connected in parallel (see Fig. 3).

The TAP (test access port) controller is a simple state machine with 16 states. Transitions between states are controlled by TMS and TCK. The TAP controller states determine how data is shifted into or out of the various device registers. Boundary, instruction, and bypass registers are the most important registers of a boundary scan device. The instruction register allows control of the various test modes. Boundary registers capture and control device pin states. The bypass register provides the ability to bypass a device in the chain.

With these three registers and knowledge of the printed circuit board netlist, it is possible to generate and analyze long strings of serial data (test vectors) to detect pins and traces that are shorted or open. Special tests can also be devised for testing nonscannable devices that are surrounded by scannable devices, (i.e., cluster testing). U2 in Fig. 3 is an example of such a device. In addition, the internal logic of a device can also be exercised.

The loaded HP E5200A printed circuit assemblies are first tested using the HP 3070 in-circuit tester. The HP 3070 allows combined testing (boundary scan and test points) for the greatest coverage and resolution. Tests are performed extremely quickly, avoiding component damage.

Test vectors for the service analyzer were also generated using ASSET™ scan software. These vectors are applied using both a PC with an ASSET hardware interface and an embedded scan engine. Greater test coverage is achieved when the PC

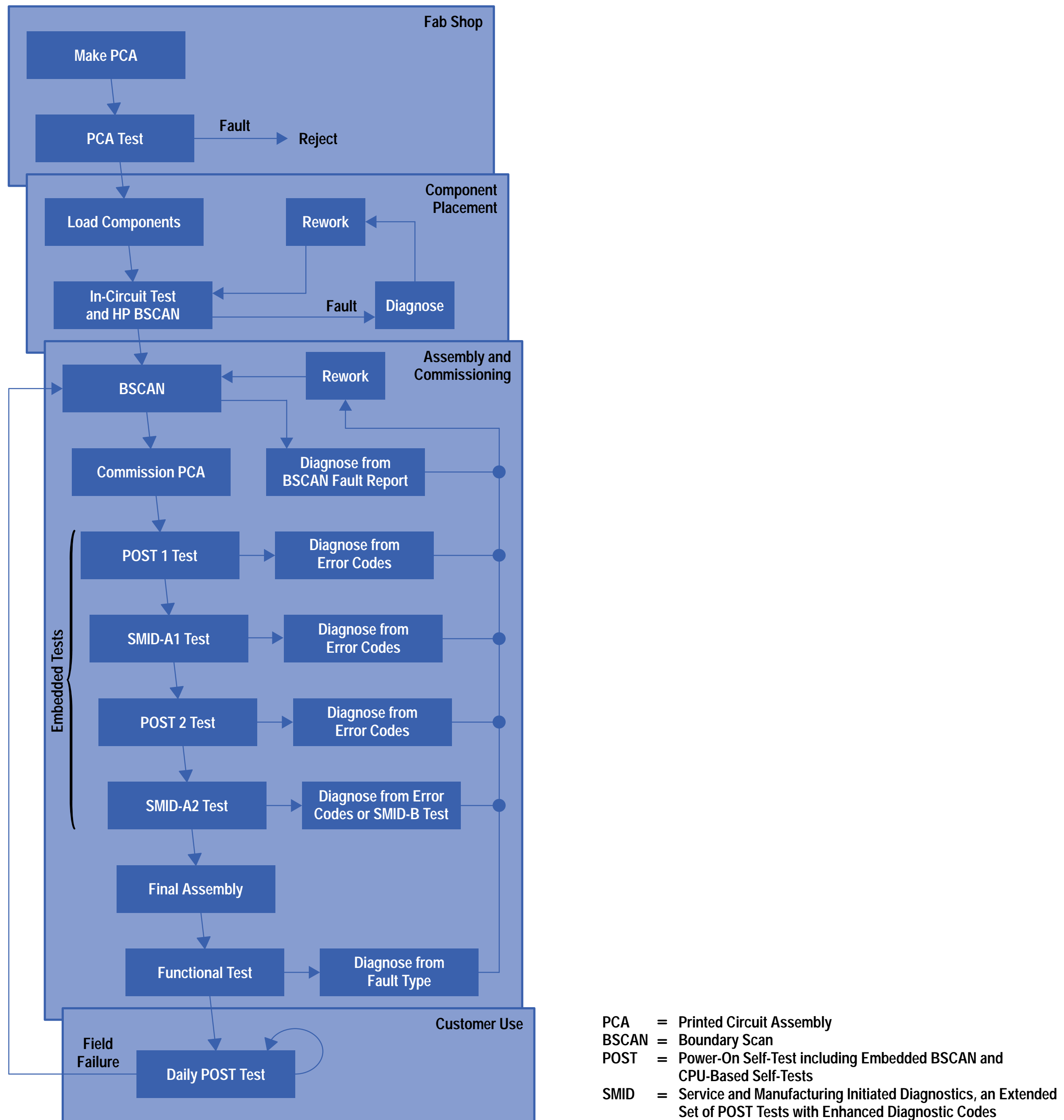


Fig. 1. *Production test strategy for the HP E5200A broadband service analyzer.*

and ASSET hardware interface are used. The embedded scan engine provides less coverage, but is useful as a power-up self-test for the user-initiated or service-initiated diagnostic. Coverage and speed of ASSET tests are not as great as the HP 3070, but setup costs and portability are much more attractive for a service and diagnostic environment.

Boundary Scan Limitations

Boundary scan cannot be used for testing all circuits. Analog devices, nonscannable parts, and passive parts require the use of conventional testing techniques. At-speed testing cannot be performed because of boundary scan's serial architecture. Boundary scan is essentially a static (dc) test. Component costs are increased, although this is offset by savings in real estate, increased production yield, and reduced repair diagnosis time.

Issues that arose during the implementation of boundary scan included difficulties achieving clean TCK waveforms at all nodes, incorrectly supplied component description files, and failures caused by scan chain data corruption.

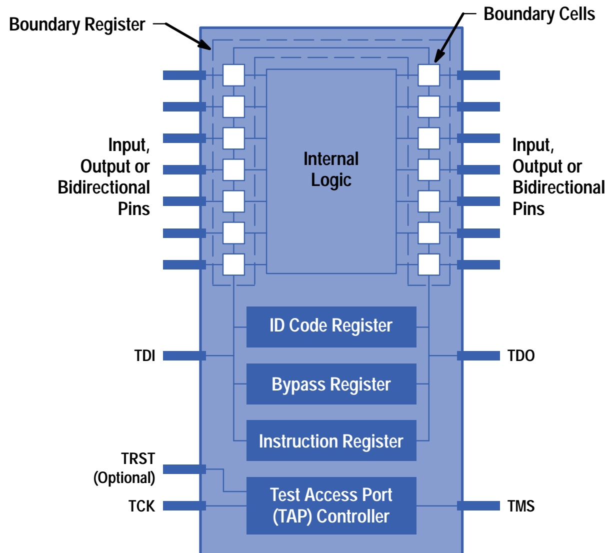


Fig. 2. Typical IEEE 1149.1-compliant device.

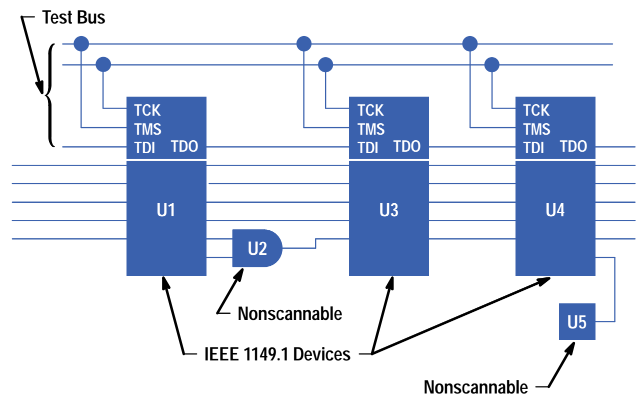


Fig. 3. Simple boundary scan chain.

Boundary Scan Benefits

Many benefits were gained by adopting boundary scan for the service analyzer project. The use of boundary scan helped to achieve size requirements for the product, and was invaluable for getting early prototypes running. Embedded boundary scan now provides the user with high confidence in the product, and allows for quick diagnosis at repair centers without expensive tools and fixtures. Production repair and diagnosis time are down and yield is improved. Overall, boundary scan has proved to be a flexible and powerful tool.

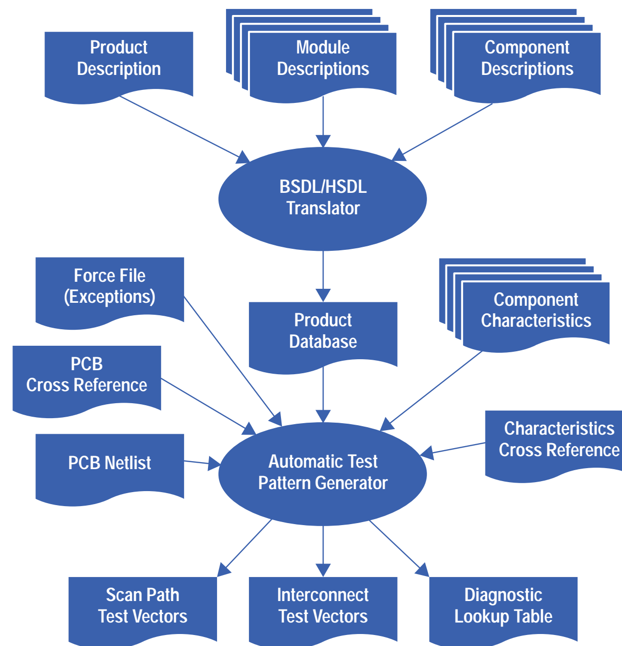


Fig. 4. Test vector generation.

Acknowledgments

Design for testability is often an area that suffers neglect during projects that have severe time and resource constraints. It is important to acknowledge the efforts of the service analyzer product design team, which despite these constraints implemented a valuable test strategy.

Bibliography

1. *IEEE Standard 1149.1-1990*, ISBN 1-55937-350-4, IEEE Computer Society, October 23, 1993.
2. K.P. Parker, *The Boundary Scan Handbook*, ISBN 0-7923-9270-1, Kluwer Academic Publishers, 1992.

Usable Usability

Usability engineering aims to improve a product's ease of use by focusing on user needs. "Usable usability" also considers the needs of the product developers.

by Peter G. Tighe

Misguided usability engineering effort is a waste. So how do you get maximum benefit from your investment in usability engineering? This article discusses some of the methods HP's Australian Telecom Operation (ATO) used while developing the new HP E5200A broadband service analyzer (Fig. 1). The service analyzer is designed for installers and maintainers of large telecommunications networks that employ BISDN (broadband ISDN) technology, such as the information superhighway networks.

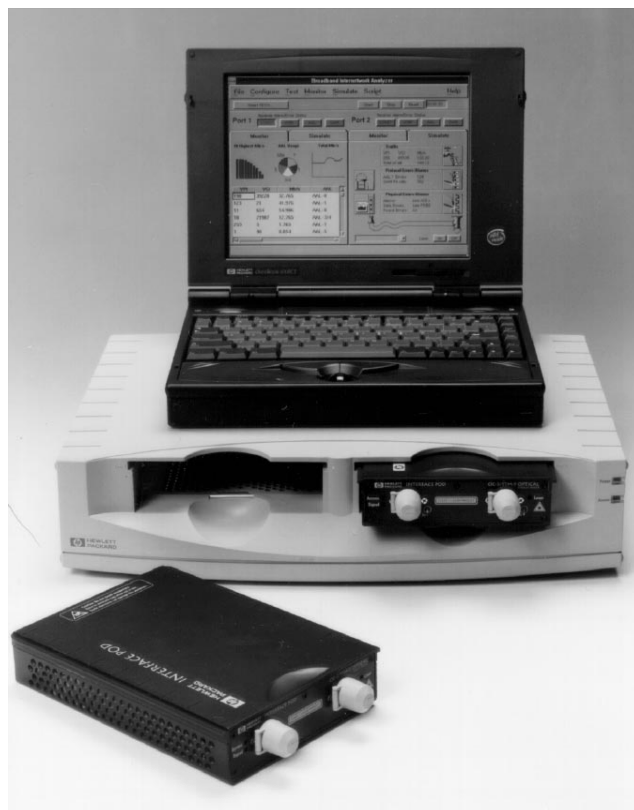


Fig. 1. *HP E5200A broadband service analyzer.*

Traditionally, usability engineering aims to improve a product's ease of use by focusing on the users' needs, behavior, environment, skills, and so forth. "Usable usability" takes this one step further by also placing high focus on the needs of the product developers—in this case, the ATO staff.

The usability engineering motto, "Know your user," can be easily extended to include "Know your organization." This can be as simple as understanding that the development engineers frequently don't read paper reports, that they prefer to read e-mail, that they are professional but are always in the mood for a laugh, or that when they sarcastically criticize someone they are actually complementing that person. So the first step was to align usability efforts with ATO's business and staff needs. The remainder of this article explains some of the results.

Understanding the User

One of the first and most essential steps quoted by almost every textbook in usability engineering is user analysis. Understand, live with, and love your users so much that you feel that you can predict their behavior.

We lived with the customers, we stood over their shoulders while they did their jobs, we asked questions when we could, we helped them do their jobs, we ate lunch with them, and we interviewed some of them intensely. We gathered information on what users looked like, on what they did, on what their environments were like, on what skills they had, on their education, on how they preferred to learn, on their tolerance limits, and on what they liked or disliked.

Being in Australia, we couldn't visit a customer in the U.S.A. whenever it suited us. We needed to condense all of our visits into one or two trips. The sheer quantity of quality information gathered created challenges, too. To meet these challenges, four techniques evolved and proved very successful:

- Onsite experts
- Meet-the-user e-mail
- User nicknames
- User artifacts.

Onsite Experts. We found that giving a customer access to a telecommunications engineer from ATO was a very effective way of gaining the customer's attention. It's a win-win situation. Both the customer and ATO engineers have the challenge of keeping up with the rapid changes in telecommunications technology. The ATO engineers get access to real-life situations in which ATO products are being used. The customer gets some free advice from the ATO engineer, has a chance to influence the direction of ATO products, and sometimes gets to see a prerelease product. Meanwhile, a usability engineer gathers the information required to make ATO's service analyzer easy to use.

Meet-the-User E-Mail. An example of meet-the-user e-mail is "Meet Billy B. Bob," one of the trip reports produced from our customer visits. An outsider might view this trip report as unprofessional or brash. Yet, to the ATO staff, it was essential, entertaining, memorable, informative reading.

Billy was not ATO's target user, but he works with and shares our equipment with our target users. The lessons ATO learned from Billy were explained explicitly in a summary trip report and presentation. For example, Billy only reads the manual for the exceptions to the rule. He actually expects to be able to use the service analyzer without training, relying on his industry knowledge. Billy also states he is new to BISDN technology and is very willing to learn. This affected ATO's documentation team, resulting in an online help system that not only explains how to use the service analyzer, but also allows users like Billy to probe (using hyperlinks) into details about BISDN technology (see Fig. 2).

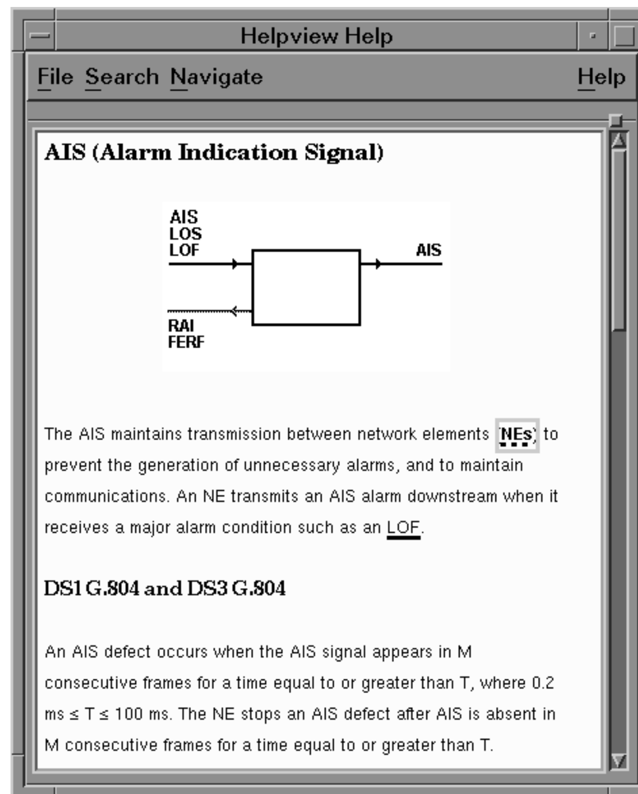


***Fig. 2.** HP E5200A broadband service analyzer online help also teaches users about broadband ISDN technology.*

Another deliberate approach was to include very personal details about Billy in the meet-the-user e-mail. This was a very effective method of influencing product design decisions. The product developers became very conscious of the user, which made a big difference in their attitudes—they became more sensitive and understanding to users' needs, resulting in better design decisions.

User Nicknames. The results of the user analysis revealed that most of ATO's users fall into three distinct categories. Each category has different skills, education, working environments, aspirations, and day-to-day tasks. To characterize these types of users, ATO marketing invented "Telecom Tom," "Protocol Pete," and "Gigabit Guru." These names are now part of the vocabulary of every ATO staff member. The nicknames are referred to in situations ranging from manufacturing problems to design opportunities, from order processing requirements to marketing strategies, and from lunch-table discussions to high-level business plans.

The real success of the nicknames is not just that they summarize a user's characteristics swiftly and succinctly. They also summarize the use scenarios of ATO products, that is, they represent complete descriptions of the situations in which ATO products are used. Bruce Tognazzini[1] describes how important scenarios are to user analysis: "The scenarios were not dwelled upon during the design process, but were always in the background, ready to be referred to when pressing an argument or forming a new idea." This is much like what happened at ATO, thanks to the user nicknames. For example, people would say, "This is mainly used by Gigabit Guru, so in this case ease of doing is probably more important than ease of learning." "Yes, but might Telecom Tom need this information?"

For example, the most popular use scenario for Telecom Tom can be summarized by one particular users comment during a customer visit: "I don't want to understand the technology, like how cell loss is measured, I just want a button that says Test Cell Loss, and when I click it, it tells me if cell loss is good or bad."

The result was the cell loss SMARTtest, as shown in Fig. 3. Telecom Toms use at least five other network analyzers. They have to support high-profile companies, such as airlines and banks, with mission-critical network connections. Every problem is important and urgent. They also install and maintain at least five different types of network elements (switches, routers, bridges). They simply don't expect to remember technology details—they expect help from the technology itself, or from an expert in that technology. The cell loss SMARTtest is an example of how the service analyzer directly addresses this concern.

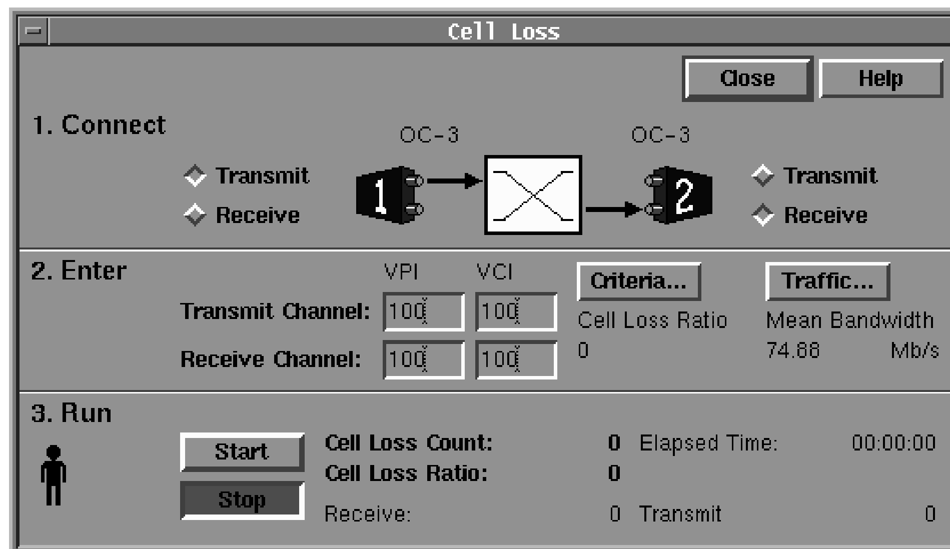


***Fig. 3.** The cell loss SMARTtest.*

User Artifacts. Meet-the-user e-mail messages and user nicknames, although very efficient forms of communication, are both secondary to hearing it "from the horse's mouth." The reader must place a lot of trust in the storyteller. Here, videos, photos, documents, broken equipment, or even just a user's notepad can be extremely valuable.

Usability Testing

As mentioned earlier, the purpose of user profiling is to understand, live with, and love your users so much that you feel that you can predict their behavior. Unfortunately, no matter how well you know your users, you can never fully predict their behavior. This is why usability testing is so important. Says Donald Norman,[2] "... even the best-trained and best-motivated designers can go wrong when they listen to their instincts instead of testing their ideas on actual users."

For those not familiar with usability testing, it's a very formal process of observing users attempt to learn and operate a product in real situations. A good reference for more information is Rubin.[3]

Usability testing can be tackled in a number of ways. For example, there are expensive ways such as using laboratories with one-way mirrors and complex data recording equipment. Or there are cheap ways such as visiting a customer with a prototype computer and a portable video camera.

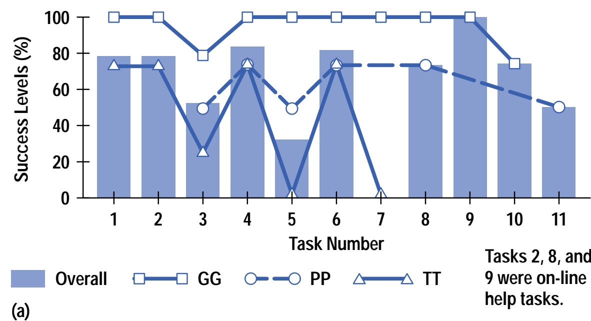
Fortunately for ATO, the cheaper option was the most suitable. ATO's customers are very busy. It was much easier to get a block of a customer's time if ATO engineers traveled to the customer's site. This gave HP a chance to see the users in their real environments and to meet other users in the organization.

One of the disadvantages of this cheap form of usability testing is that it is more difficult to control. Many external factors can intimidate the user and corrupt the test. For HP, this proved to be a minor problem. Low-skill users showed intense concentration while trying to grasp BISDN technology, and soon lost signs of intimidation. High-skill users loved the chance to show off what they knew and talked about the technology. Regularly they would stay after the usability testing and drive very motivating discussions about the technology and how ATO could benefit from it.

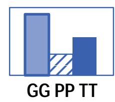
Usability testing was a very motivating and creative experience for everyone involved. The sales force was completely motivated after seeing ATO staff listen to and watch the user. They could also see the excitement the customer showed.

User "Hot Spots"

User hot spots is a technique that focuses on minimizing the development team's information burden. No matter how much information was gathered from a usability test, we always reduced it to approximately five of the most important items (see Fig. 4). Anyone who has ever done a usability test will realize how difficult this can be—so much information is gathered in one test.

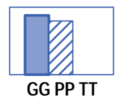


(a)



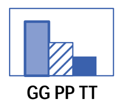
Mental Model Clash

- Previous experience with BSTS and other analyzers is affecting ease of use
- Users did not expect autoconfigure and an automatic link summary
- Wanted to override with a manual configure anyway
- Solve: Bootup and hot swap messaging
- Loopback expected inside configure.



"What does Commission Mean?"

- Commission confused with configure.
- "Equipment" and "Service" grouping not understood.
- Solve: Put tests on front panel "Cell Loss & Cell Delay," remove test grouping for release 1, more effectively, communicate "predefined" tests as being a separate environment.



Labeling and Messaging

- Monitor windows and simulate windows confused.
- Not enough feedback when errors/traffic is simulated
- Four alarm buttons still visible when simulate tab is active.

(b)

Fig. 4. *Usability test reports. These condensed weeks of data gathering into approximately five basic critical issues, called user hot spots.*

Testing Enhances Creativity and Motivation

Putting it simply, de Bono[4] explains why usability testing actually enhanced ATO's creativity. "Having to be correct at every step makes creativity virtually impossible." Knowing there would be second chances, ATO engineers were able to risk all sorts of new design ideas. They exploited this opportunity to do some creative thinking. The de Bono technique called "Six Hats Thinking" was used.[5] When "wearing" the green (creative) hat, we used the "Provocation" and "Random Word" techniques to help generate alternative, unorthodox ideas. De Bono[5] describes a provocation as crazy idea that is designed

to take people out of their normal perceptual patterns. It challenges assumptions and impressions that everybody seems to take for granted.

For example, one provocation ATO engineering used was: "What if the user could do the most frequent task without having to press a single button?" After overcoming the apparent absurdity of this idea, the engineers soon realized that it was going to be possible. From this, the link monitor was created (see Fig. 5).

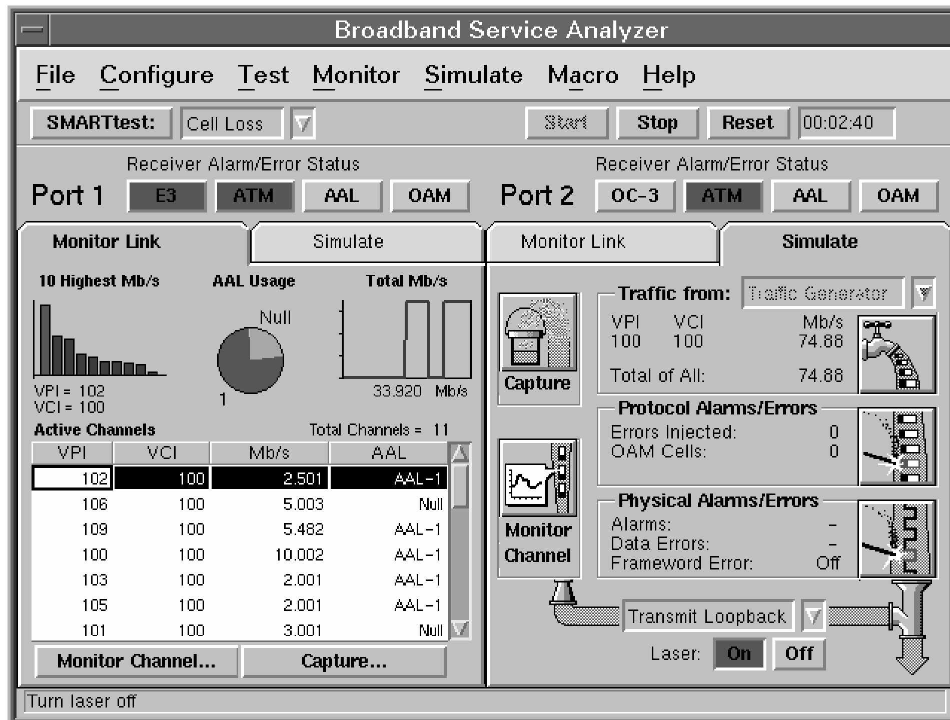


Fig. 5. *The link monitor tells the user what's happening on the network without any button clicks.*

The link monitor is a new concept in network analyzer design. Without any button clicks, it performs the most frequent task—discovering what's happening on the network. This task is intuitively basic and not immediately obvious. Ask any group of network troubleshooters what they do, and only a few will mention this task. The rest will assume they already know what's happening on their network. It's similar to typing using a word processor. Imagine if you had to select Edit/Type from the pull-down menu every time you wanted to start typing.

The second new idea in the link monitor can be compared to tools such as automatic spell-checking and grammar checking in a word processor. For example, as you type "hte" it changes it to "the." Similarly, the link monitor tells its users of network problems without their having to click any buttons. So when do users click buttons? Only when they want more details.

Task-Based Design

A rule of thumb for usability engineering is, "Design for your user's tasks." The HP E5200A broadband service analyzer project was no exception. The online help describes realistic tasks to users. The user interface is structured to follow a typical user's task flow (see Fig. 6).

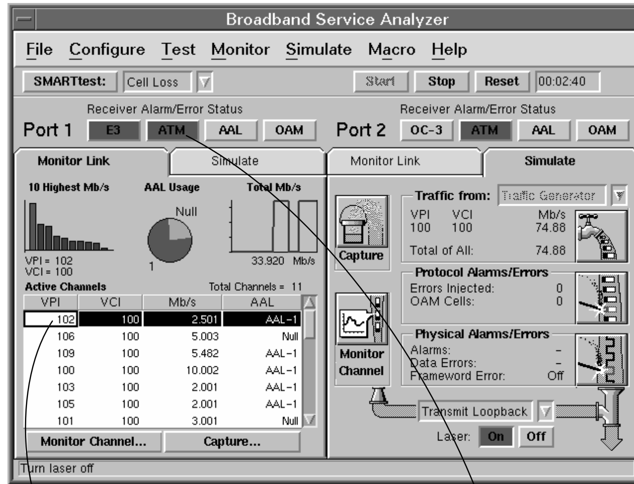
The end result seems quite simple, but it's a challenge that ATO worked hard to achieve. Why? Telecommunications technology is full of structured detail about network protocols and standards. It's very easy to get involved in these technology details and lose focus on the user's task. ATO avoided these complications by using the user nicknames (mentioned earlier) and regularly repeating the question, "What's the user trying to do with this information?"

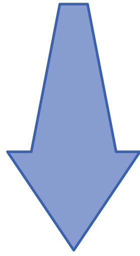
Rapid Prototyping

Rapid prototyping proved to be an ideal strategy for this project. ATO chose Microsoft® Visual BASIC as the prototyping tool. Visual BASIC was not suitable for ATO's application development environment. Its major advantages were the speed and flexibility with which screens and screen behavior could be prototyped. These advantages suited ATO's dynamic and creative attitude to GUI prototyping. In fact, it gave ATO much more freedom than expected:

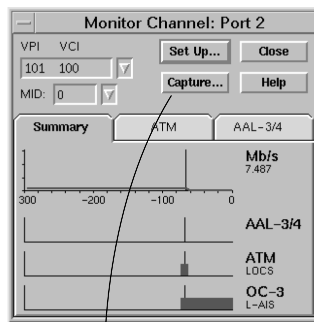
- Initially, the prototype was way ahead of the user interface specification. It actually replaced the need for a GUI specification in the early stages of product definition.

1. What's happening on my network/network element?

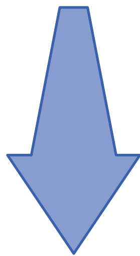


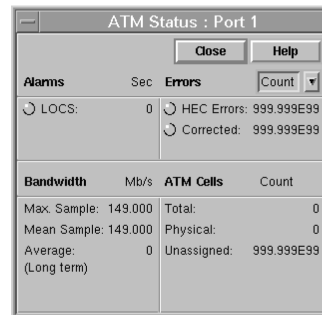


2. I think the problem may be here, let's take a closer look.



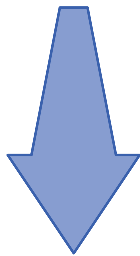
Channel monitored in real time.

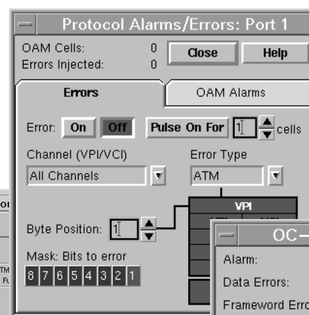




Alarms monitored in real time.

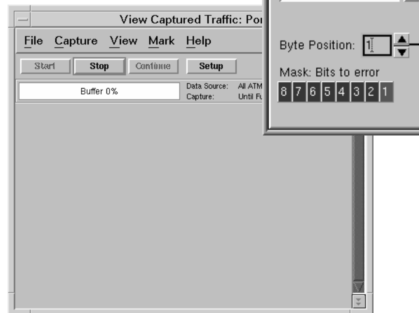
3. I think I've found it, let's see if that really is the problem.





Simulate network behavior such as artificial load, alarms, and errors.

ASCII, hex, or binary representations of network traffic.



4. I've fixed it—let's make sure everything is OK now.

Fig. 6. *A user's typical troubleshooting tasks, and how the user interface makes these tasks easy. The more detailed and less frequent the task, the deeper the user delves into the user interface.*

- The prototype was used as a selling tool. Sales force and customers were able to see and interact with a prototype product long before it was even on our price list.
- Early in the project, everyone felt they were talking about the same thing.
- Early prototyping eliminated major product changes from the middle to the end of the project. The "It's too hard" excuse or the "programmer's attachment to work" did not hinder conceptual design.

Conclusion

This article has conveyed some ways of maximizing the benefit from an investment in usability engineering. The staff at ATO believe some of this work was breakthrough and an inspiration for other usability projects. But remember, most of the methods described in this article came from first analyzing the organization's needs. These methods may not apply to other organizations, especially if they have different needs and a different culture from ATO.

There is one factor that is important to secure before trying to apply these methods in other organizations—that is, management's support to experiment with how usability engineering is applied. This support is all a motivated team needs to succeed with usability engineering.

Acknowledgments

Most of the best practices described in this article have been successful because of the advice and direction given by the intelligent product and media group at Hewlett-Packard Corporate Engineering, Palo Alto, California.

References

1. B. Tognazzini, *TOG on Interface*, Addison Wesley, 1992.
2. D. Norman, *The Design of Everyday Things*, Doubleday Currency, 1988.
3. J. Rubin, *Handbook of Usability Testing*, Wiley, 1994.
4. E. de Bono, *Parallel Thinking*, Viking, 1993.
5. E. de Bono, *Serious Creativity*, HarperColliins, 1993.

Bibliography

1. D. Hix and R. Hartson, *Developing User Interfaces—Ensuring Usability through Product and Process*, Wiley, 1993.