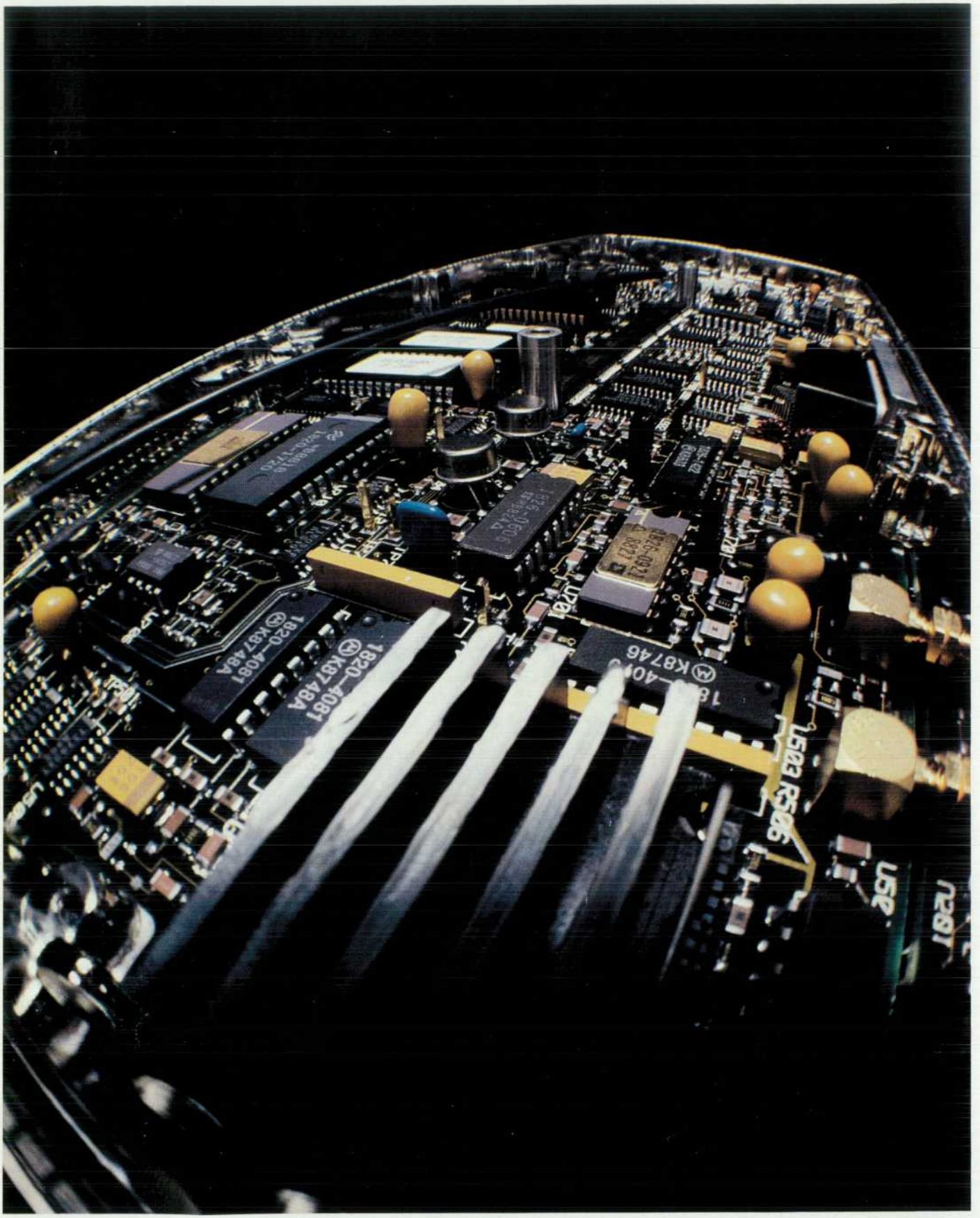
HEWLETT-PACKARD JOURNAL

OCTOBER 1989



HEWLETT-PACKARD JOURNAL

October 1989 Volume 40 • Number 5

Articles

Departments

In this Issue

The first issue of the *HP Journal* was published in September 1949, so this issue, October 1989, is the first of the *HP Journal*'s forty-first year of publication. (However, because of a 1978 change, Volume 41 doesn't begin until February 1990.) Coincidentally, the Hewlett-Packard Company is celebrating its fiftieth anniversary this year. To Associate Editor Chuck Leath, this double milestone—40 years of the *HP Journal* and 50 years of HP—fairly cried out for recognition, so he has assembled a special commemorative article to mark the occasion. The article, which begins on page 6, traces the development of HP technology as reported in the pages of the *HP Journal*.

As signal generator designers at HP's Spokane Division saw things, general-purpose signal generators were delivering a lot of capability to the customer—so much, in fact, that few users needed all of it and were therefore paying for features they never used. On the other hand, offering a different product for each application could mean the loss of manufacturing economies and therefore higher prices. Their solution is a new family of internally modular signal generators. While options allow customers to buy only as much capability as they need, production economies are retained by using a lot of the same hardware and firmware in all of the members of the family. Called the Performance Signal Generators (PSG), the family includes the HP 8644A 1-GHz or 2-GHz Synthesized Signal Generator, the HP 8645A 1-GHz or 2-GHz Agile Signal Generator, and the HP 8665A 4.2-GHz Synthesized Signal Generator. The HP 8644A is designed with traditional out-of-channel transceiver test applications in mind, and the HP 8645A is focused on frequency agile transceiver testing. The HP 8665A is designed for radar, telemetry, spurious testing of UHF transceivers, and similar applications. On page 14, two PSG R&D project managers give an overview of the family and the basic design, which differs from previous microwave synthesized signal generators in that it uses only a single phase-locked loop instead of the multiple loops common in earlier designs. An advanced fractional divider makes the simplification possible. Supplementing the single-loop design are a new method of introducing frequency modulation and one or two optional frequency-locked loops for additional phase noise reduction. Details of the synthesis, FM, and noise reduction techniques are in the article on page 27. How fast frequency hopping, a capability of the HP 8645A, influenced its design is explained in the article on page 34. The articles on pages 37, 42, and 51 cover the design of the HP 8665A synthesis, output, and pulse modulation systems. Firmware design for the three instruments is treated in the article on page 20, and RFI (radio frequency interference) reduction is the subject of the article on page 59. A particularly simple and quick RFI test turns out to be one of the most difficult to pass—it consists of placing a pager next to the signal generator and noting whether it detects anything.

The **Hewlett-Packard Journal** is published bimonthly by the Hewlett-Packard Company to recognize technical contributions made by Hewlett-Packard (HP) personnel. While the information found in this publication is believed to be accurate, the Hewlett-Packard Company makes no warranties, express or implied, as to the accuracy or reliability of such information. The Hewlett-Packard Company disclaims all warranties of merchantability and fitness for a particular purpose and all obligations and liabilities for damages, including but not limited to indirect, special, or consequential damages, attorney's and expert's fees, and court costs, arising out of or in connection with this publication.

Subscriptions: The Hewlett-Packard Journal is distributed free of charge to HP research, design, and manufacturing engineering personnel, as well as to qualified non-HP individuals, libraries, and educational institutions. Please address subscription or change of address requests on printed letterhead (or include a business card) to the HP address on the back cover that is closest to you. When submitting a change of address, please include your zip or postal code and a copy of your old label.

Submissions: Although articles in the Hewlett-Packard Journal are primarily authored by HP employees, articles from non-HP authors dealing with HP-related research or solutions to technical problems made possible by using HP equipment are also considered for publication. Please contact the Editor before submitting such articles. Also, the Hewlett-Packard Journal encourages technical discussions of the topics presented in recent articles and may publish letters expected to be of interest to readers. Letters should be brief, and are subject to editing by HP.

Please address inquiries, submissions, and requests to: Editor, Hewlett-Packard Journal, 3200 Hillview Avenue, Palo Alto, CA 94304, U.S.A.

Hewlett-Packard's first Technical Women's Conference was held in October 1988. The conference was organized to showcase the achievements and contributions of technical women at HP and to promote their professional development into leadership roles. Approximately 400 HP women engineers and scientists and their managers attended the conference. Many section and Division managers participated as speakers and panel discussion leaders. Topics of the technical presentations included hardware, software, computers, instruments, manufacturing, components, and other subjects. Papers based on six of the presentations appear in this issue. On page 69, Susan Sloan of the Microwave Technology Division reports on work done to determine the best method of surface preparation and passivation for low-dark-current photodetectors for use in HP lightwave receivers. Among the requirements for these photodiodes are maximum dark currents of a few nanoamperes and good response to modulation frequencies beyond 22 GHz. On page 76, Eve Tanner of the Personal Computer group tells how a useful symbolic debugging capability can be given to programmers developing drivers to run under the HP-UX operating system when the HP-UX source code is not available. The technique takes advantage of available compiler information to insert HP-UX global data records into the user's file. On page 81, Cathy Keely of HP Laboratories describes an experimental laser-based method of finding bad solder joints on surface mount components. Vibrations induced in the leads by an air jet are detected and analyzed by laser Doppler vibrometry. On page 86, Stacy Martelli of the General Systems Division discusses a model for HP-UX shared libraries that was developed to provide shared library capability to users of a particular software package in the absence of operating system support for them. The model required only minor changes to the HP-UX linker. On page 90, Lucy Berlin of HP Laboratories describes a methodology for systematically acquiring and applying user information in the definition of software applications. Interviewing, task analysis, and storyboarding are key elements. On page 98, Carolyn Jones of the Optoelectronics Division discusses the design of light guides for an array of light-emitting diodes used for selective erasing in electrophotographic copiers. Well-defined spots of light have to be formed some distance away from the LEDs, which are one millimeter apart and emit light in all directions.

R.P. Dolan
Editor

Cover

Not a space station from a science fiction film, but the fractional-N module from HP's Performance Signal Generator family. For a more conventional view, see Fig. 8 on page 19.

What's Ahead

The HP Starbase Graphics Library is a high-performance 2D and 3D graphics library that runs on HP 9000 Computers under the HP-UX operating system. In the December issue, six articles will describe the Starbase/X11 Merge System, which enables Starbase applications and X Window System™ applications to coexist in the same window system. There will be two papers on aspects of the HP 9000 Series 300/800 Turbo SRX graphics subsystem—one on the radiosity method of global illumination modeling and one on the custom VLSI chips used in the graphics pipeline. HP Source Reader, a CD-ROM system that gives HP support engineers fast access to HP 3000 Computer operating system source code, will be described. The December issue will also contain the 1989 Index.

Fifty Years of Looking to the Future

40 Years of Chronicling Technical Achievement

Over the last 40 years the HP Journal *has created a record of HP's technical achievements by communicating technical information to professional people in all fields served by HP. With Hewlett-Packard celebrating its 50th anniversary it seems appropriate to take a look at the* HP Journal, *past and present, and some of the technological history of Hewlett-Packard it has chronicled.*

by Charles L. Leath
Associate Editor

THIS YEAR HEWLETT-PACKARD is celebrating its first fifty years as a company and coincidentally the *HP Journal* is in its 40th year of publication.

In September 1949, when the first issue of the *HP Journal* was published, HP had 150 employees, 70 electronic measurement instruments in the product catalog, and 2¼ million dollars in sales. Other significant technological events during this period were the invention of the transistor at Bell Laboratories in 1947 and the operation of the first supercomputer, the ENIAC, in 1946. Today HP has over 90,000 employees, over 10 billion dollars in sales, and over 10,000 products that appear or are alluded to in the HP product catalog. The HP product line today includes not only electronic measuring instruments, but also computer hardware and software systems, peripherals, medical electronic equipment, analytic instruments for chemical analysis, electronic components, and a host of other services and products.

The initial idea for the *HP Journal* came from a periodical called the *General Radio Experimenter*. General Radio, now known as GenRad, produced primarily electronic measurement instrumentation, and the *Experimenter* was a monthly publication devoted to supplying "unbiased information pertaining to radio apparatus design and application."[1] The *Experimenter* started publishing in 1926 (11 years after the founding of General Radio) and was the first periodical of its type in the radio industry. It was distributed free of charge to qualified experimenters and it was a popular periodical in the industry. Many HP engineers were devoted readers of the *Experimenter*. One of those readers, David Packard, was impressed with the quality of the articles and thought that HP should have something similar. This idea turned into reality when Frank Burkhard, the first *HP Journal* editor, was assigned the task of creating HP's version of the *Experimenter*. There was some consideration given to calling the *Journal* the *HP Experimenter*. However, after examining periodicals from other companies (e.g., the Bell System Technical Journal) the name *HP Journal* stuck.

Although one of the initial objectives for the *Journal* was to produce a periodical similar to the *General Radio Experimenter*, other objectives included telling customers about HP products and their applications, showing the quality and workmanship that went into each product, and giving credit to the product designers. The *Experimenter* also included information about the product designers. Like HP's corporate objectives, the *Journal*'s original objectives have been expanded and clarified. Today the *HP Journal* is a periodical for the communication of technical information from all of HP's R&D, manufacturing, and quality organizations to professional people in all fields served by HP.

When this issue (Volume 40 Number 5) is printed there will have been 466 issues of the *HP Journal*. Since the late 1950s the *HP Journal* has been a monthly publication, and in December 1987 it became bimonthly. When the *HP Journal* started, it was circulated only in the continental U.S.A. Today the HP Journal's circulation includes subscribers all over the world. Up until the late 1950s the *HP Journal* averaged four to six pages per issue, and today it averages around 90 pages per issue.

First Journal

The first *HP Journal* featured the HP 460A, a wideband noninverting amplifier whose main features included distortionless pulse amplification with a very short rise time (3 ns) and almost zero overshoot, a frequency range of 3 kHz to 140 kHz, a gain of 20 dB into a 200Ω load, and a weight of 10 pounds. The rise time was pretty good even by today's standards. The HP 460A was used in situations where it was necessary to amplify pulses faster than 0.01 μs. Application areas included nuclear research and television, VHF, UHF, and SHF (superhigh-frequency or microwave) research. It was also used as a preamplifier for oscilloscopes and to increase the sensitivity of voltmeters. The original cost of the HP 460A was $185.00, and when it was discontinued in 1967 it cost $285.00. Fig. 1 shows a reduced reprint of this first *HP Journal* issue.

In 1964, the year HP celebrated its 25th anniversary and the *HP Journal* was up to eight pages per issue, two broadband, solid-state versions of the HP 460A were introduced, the HP 461A and HP 462A (March 1964).* The HP 461A was optimized for a flat frequency response over a wide bandwidth, and the HP 462A was similar to the HP 461A, but had a smaller bandwidth and was optimized for a fast rise time (less than 4 ns) with an overshoot of less than 5%. These amplifiers could be cascaded with the HP 460 Series broadband amplifiers for high-voltage applications by using a 50Ω-to-200Ω matching transformer. The HP 461A had a frequency range of 1 kHz to 150 kHz, a maximum gain of 40 dB operating into a 50Ω load, and a weight of four pounds. It cost $325.00.

Today the performance characteristics of the HP 460 and 461 Series amplifiers are achieved and exceeded in thick-film hybrid circuits and HP amplifiers. For example, the HP 8449A Preamplifier, which is used in RF and microwave applications, is programmable and has up to 28 dB of gain at 22 GHz. There are also pulse

*The month and year in parentheses indicate the Journal issue in which the referenced article or topic appears.

Generator, which can produce pulses with widths less than 0.5 ns, rise and fall times less than 200 ps, and frequencies to 500 MHz.

After Vol. 1 No. 1

HP Journal articles have covered everything from products and applications to tutorials and research topics. Many of the application articles showed the practical and diverse uses of HP products. For example, the October 1957 issue included a supplement that described the use of an HP counter and digital recorder to capture transmissions from the Russian earth satellite Sputnik. One of the more international applications appeared in the July 1964 issue which described the "Flying Clock" experiment, in which the HP 5060A Cesium-Beam Frequency Standard was used to compare time standards in the U.S.A. and Europe and the results were used by the official agencies of the two continents to improve their time synchronization. This same product was later used in a test of Einstein's relativity theory. Although applications, tutorials, and other items populated the *Journal* pages, new product designs dominated the *Journal* articles.

In 1983 reprints of articles from 32 HP Journal issues were bound together in a book called *Inventions of Opportunity*.[2] The articles included in the book were chosen because they described products or innovations that represented important milestones in the technological history of the Hewlett-Packard Company. The selec-

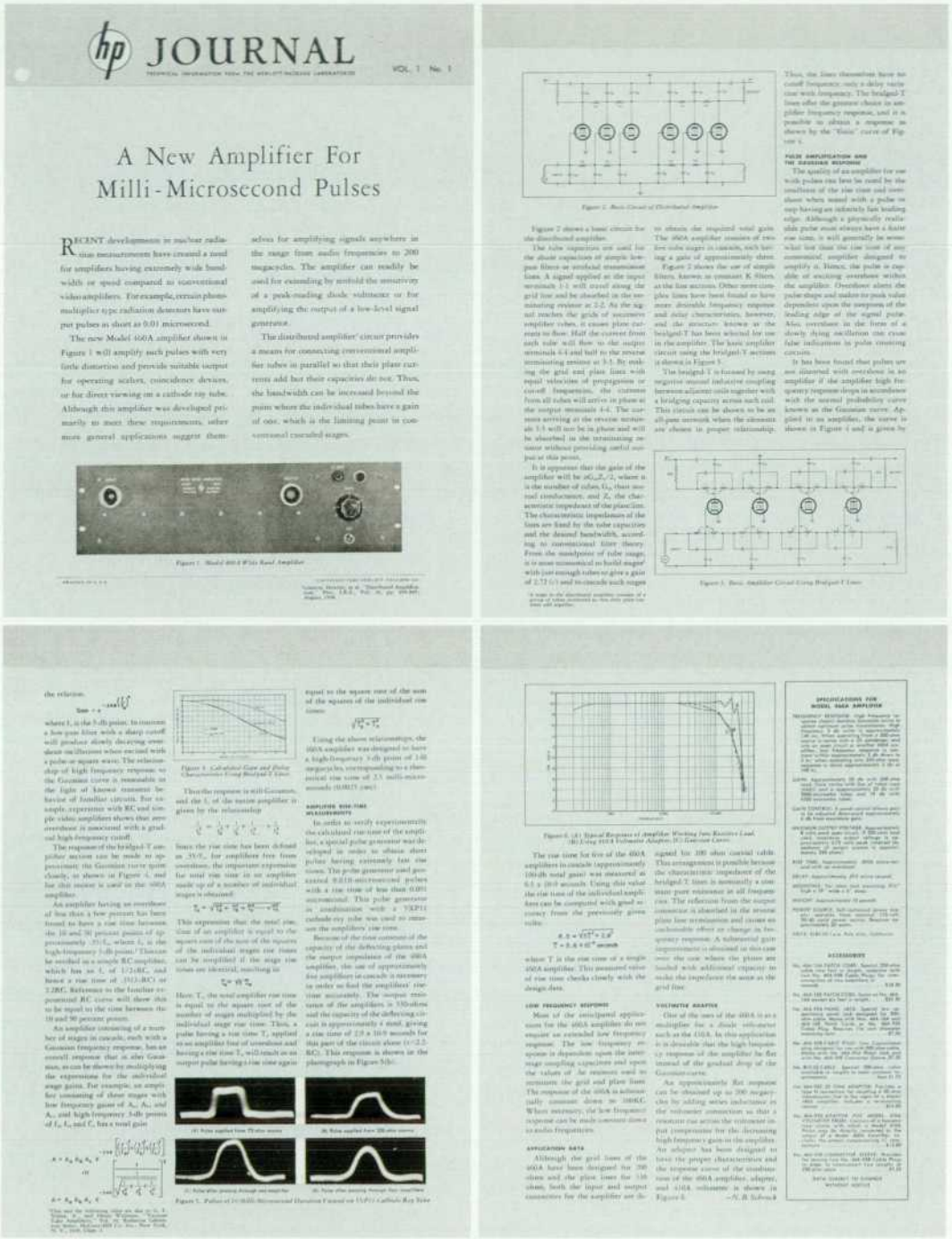


Fig. 1. *The first issue of the* HP Journal.

Fig. 2. *(a) Nov. 1949. This issue is the first of a two-part series describing design issues related to the resistance-capacity oscillator. (b) June 1951. The* Journal *issue describing the HP 202A Low-Frequency Function Generator. This instrument had a frequency range of 0.01 Hz to 1 kHz. (c) Nov. 1965. This issue describes a successor to the HP 202A, the HP 3300A Low-Frequency Function Generator. The basic instrument provided sine, square, and triangular waves throughout a frequency range of 0.01 Hz to 100 kHz. The* Journal *cover shows some of the waveforms that could be generated with this instrument.*

tion committee consisted of 20 or more individuals who had been with HP for a long time and who had a broad perspective on the technological history of HP. The collection begins with the resistance-capacity oscillator (November 1949) and ends with an article on a new plotting technology (October 1981). Bill Hewlett wrote the introduction, describing the history behind each of the products included in the book.

What follows is a synopsis of some of the Journal issues and technical areas covered in *Inventions of Opportunity*, and of some significant *Journal* issues published since October 1981. The discussion also shows what has happened to a particular product line since it first appeared in the pages of the *Journal*. Because of space limitations we can only show the front covers of the *Journal* issues being discussed, but these covers are interesting in themselves, because in addition to showing the product or something about it, they also show the changes in the style of the *Journal* over the years.

Signal Sources

Many of the *HP Journal* issues in 1949 described products that had been developed years earlier. The most famous of these products was the resistance-capacity oscillator, HP's first product. The RC oscillator had been introduced ten years before the *Journal* article. These two issues discussed the problems that had to be overcome in designing these instruments, such as extending the low and high frequency ranges with the components available at the time. Fig. 2a shows the first of these two issues.

Although HP had oscillators that operated as low as 1 Hz, there were needs for much lower-frequency sources. A technical paper from a student contest sponsored by the the San Francisco section of the IRE (predecessor to the IEEE) in 1950 described a technique for extending the range of an oscillator to very low frequencies. The rights to the technique were acquired from the student and his professor and very shortly HP developed the HP 202A Low Frequency Function Generator. This generator also had the ability to produce triangular and square waveform shapes—a big deal at the time. See Fig. 2b.

Today, signal sources are usually frequency synthesizers. The first synthesizer produced by HP, the HP 5100A/5110A Frequency Synthesizer (see Fig. 3), could generate high-stability signals from 0.01 Hz to 50 MHz. It used a direct synthesis approach. Direct synthesis involves performing a series of arithmetic operations on a signal from a frequency standard to achieve the desired output frequency. In today's synthesizers, indirect synthesis, which involves deriving the signal from one or more oscillators phase-locked to a reference source, is used. The development of the HP 5100A/5110A instrument was the first time HP had made such a major commitment to a new technology—the project required almost 40-engineer-years of development effort spread over a period of about three years.[3]

Microwave Equipment

During World War II, development of microwave systems was quite intensive. However, after the war many companies dropped much of their microwave work. HP continued to develop microwave equipment including such items as detectors, standing-wave indicators, directional couplers, waveguides, and so on, "all with the aim of simplifying measurement in this very difficult field."[4] Microwave products were so pervasive in the early days that out of 24 *HP Journal* articles between September 1950 and September 1952, half were devoted to microwave equipment. The *Journal* cover shown in Fig. 4a does not represent a significant product, but it is representative of the type of *Journal* issues published at the time that covered microwave equipment.

Today with new technologies, HP continues to develop microwave components, signal sources, and measurement instruments.

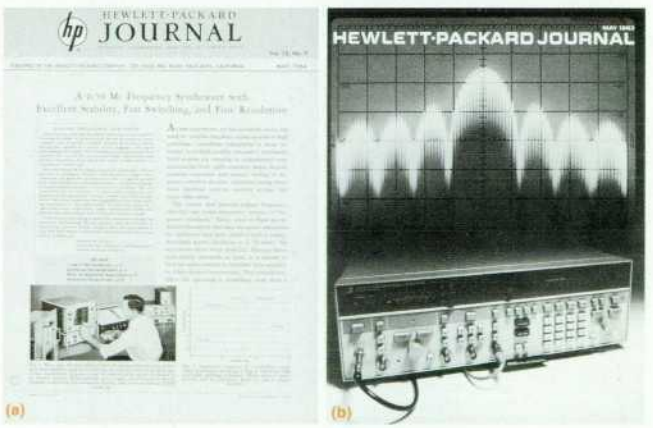


Fig. 3. *(a) May 1964. This issue described the HP 5100A/5110A Frequency Synthesizer. The unit with all the buttons (top unit) is the HP 5100A Synthesizer and the unit on the bottom is the HP 5110A Synthesizer Driver. The driver unit contains the frequency standard. (b) May 1983. This issue describes a product that shows where we were in 1983 with regard to frequency synthesis techniques. The HP 8673A Synthesized Signal Generator, which has a frequency range of 2 to 26.6 GHz, is featured.*

Fig. 4. *(a) Feb. 1951. This issue explained HP's program in 1951 regarding microwave waveguide equipment, and it is typical of the many* Journal *issues at the time that featured microwave equipment. Among the waveguides discussed in this issue, the one with highest range went to 18 GHz, or 18 kilomegacycles in the vernacular of the time. (b) Nov. 1986. This* Journal *cover shows four HP 11970 Series waveguide harmonic mixers. Three of them have their own horn antennas which come in different sizes for operation in different frequency bands.*

One of the best examples of these efforts is the HP 11970 Series of waveguide harmonic mixers described in the *HP Journal* issue shown Fig. 4b. This issue describes the efforts by HP Laboratories and the HP Microwave Technology Division to develop small-scale integrated and hybrid circuits that operate at 100 GHz and beyond. Electromagnetic energy in the frequency range of 26 to 300 gigahertz is called millimeter waves.

Counters

The first high-speed frequency counter, the HP 524A, was developed in 1951. It revolutionized the measurement of unknown frequencies. At the time, frequency measurement required three separate pieces of equipment—a frequency standard, an interpolating system, and a detector. The technology for the HP 524A came from an instrument called a scaler, a frequency-dividing instrument capable of dividing quantities of pulses by an arbitrary factor, usually 10^n or 2^n. The HP 520A, a 10-MHz scaler that provided the technology for the HP 524A, consisted of two decades, giving a scaling factor of one hundred, so that for every 100 pulses applied to the scaler a single output pulse was generated. This scaler was intended for the nuclear instrumentation market. Conventional scalers at the time operated to about 100,000 counts per second. The HP 520A could be used as a prescaler that could accept count rates as high as 10 million counts per second and scale them down by a factor of 100 to be fed into conventional scalers. The HP 520A was successfully developed, but the need for it never materialized.[5]

The technology of the HP 520A was combined with a very accurate time gate circuit to produce the HP 524A. The complete counting circuit for the HP 524A consisted of eight cascaded scalers. All the scalers generated one output pulse for every ten counts. The first two scalers produced their counts on a panel meter and the remaining six displayed their counts in neon lights.

With more functionality and greater capabilities provided by firmware and new circuit technologies, counter technology has come a long way since the HP 524A as evidenced by the descendants described in the *HP Journal* issues shown in Fig. 5.

Oscilloscopes

Technical innovation is not always the result of a major technological breakthrough. Sometimes it is simply the application of old or existing concepts in different ways and possibly with new technologies. The HP 185A/187A Sampling Oscilloscope was just such an example of using an existing concept. The concept was waveform sampling, a technique that had been used earlier in the power generation field to measure the waveform of an alternator. The technique had proven to be applicable to very high-speed signals, and the HP 185A/187A was the first practical application of this concept to measuring periodically recurring waveforms. In the instrument, the sampling technique used a stroboscopic approach to reconstruct the input waveform. In reconstructing the waveform, the sampling circuit was turned on for a very short interval. The waveform voltage at that instant was measured and the CRT spot was positioned to correspond to the sampled amplitude. This process was repeated on the next cycle of the waveform except at a later point than the first sample. Continuing in this way the sampling scope reconstructed the entire

Fig. 5. *(a) Jan. 1951. This issue describes the HP 524A High-Speed Frequency Counter. The counter provided frequency measurement up to 10 MHz and five selectable sampling periods: 0.001, 0.01, 0.1, 1.0, and 10.0 seconds. (b) July 1975. The HP 5328A Universal Counter has a direct count frequency range of 100 MHz (512 MHz with a special option), a period resolution of 100 ns (10 ns with a special module), and other functions such as time interval averaging, totalize, ratio, and so on. The* HP Journal *cover shows the counter's optional modules superimposed on a background showing the counter's interior. (c) Feb. 1989. The latest contribution to frequency and time measurement, the HP 5371A Frequency and Time Interval Analyzer. The cover shows a sculpture in plastic of the circuit diagram of a zero-dead-time counter, a key component in the HP 5371A. The HP 5371A uses a method known as continuous measurement technology. The zero-dead-time counters implemented with high-speed integrated circuits make one measurement after another without stopping. Like the HP 524A, this instrument has also made it easier to make certain frequency and timing measurements that were formerly difficult or impossible.*

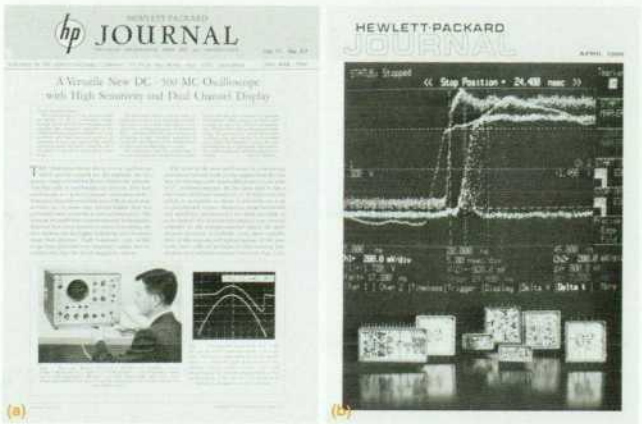


Fig. 6. *(a) March 1960. This* Journal *issue describes the HP 185A/187A Oscilloscope. Also in this issue is an article describing the phenomenon of reverse conducting diodes and how the new sampling scope was used to observe this phenomenon. (b) April 1986. This* Journal *issue describes the HP 54100A/D and HP 54111D Digitizing Oscilloscopes. The cover shows the display of the HP 54110D, and in front of the display are the seven hybrid circuits used in each instrument.*

input waveform.

At the time the sampling scope was produced, most scopes had a top frequency of 50 to 100 MHz. The sampling technology enabled the HP 185A/187A to provide a 500-MHz upper limit. Sampling technology was also applied to other instruments such as the HP 8405A RF Vector Voltmeter (May 1966), the HP 3406A Voltmeter (July 1966), and the sampling plug-ins for the HP 140A and 141A oscilloscopes (October 1966). The latter had a range of dc to 12.4 GHz.

Today we have digitizing oscilloscopes that still employ the waveform sampling concept but use more advanced technologies and digital techniques to operate on the signal as opposed to pure analog signal processing in the older scopes. The HP 54100A and 54110D Digitizing Oscilloscopes (Fig. 6b) are the best examples of these new oscilloscopes. These scopes have capabilities such as a 1-GHz range and a 350-picosecond rise time (4 nanoseconds for the the HP 185A/187A oscilloscopes). These digitizing scopes use a sampling technique that reconstructs the input signal from a series of samples taken at random points on the waveform.

Computation

It is probably a gross understatement to say that the field of computation, which includes calculators, computers, peripherals, and associated technologies, has gone through some dramatic changes since the days of the ENIAC. The ENIAC computer, which began operation in 1946, weighed 30 tons, consumed 150 kW, contained 18,000 vacuum tubes, and could perform 400 multiplications or 5,000 additions per second. Contrast this with one HP Precision Architecture NMOS-III single-chip CPU, which contains 115,000 transistors packed onto a square die measuring 8.4 mm on a side, weighing just a few ounces and consuming less than 12 watts in a pin-grid array package, and capable of executing 7 MIPS (million instructions per second). This may seem like comparing an elephant to an ant. However, it does illustrate how far this field has come in a little over 40 years.

Before the introduction of the HP 3000 Computer in 1973, HP was regarded as strictly an instrumentation company. Today computation products make up over 60% of HP's sales and orders. This change has been reflected in the pages of the *HP Journal*. For instance, in 1970, out of 25 articles, only five articles or 20% featured something about computation products, whereas in 1980, out of 83 articles, 31 articles or 37% featured computation products.

Although HP's computer products also include all the peripheral items that go with a computer, such as terminals, printers, disk drives, and different types of input devices, this section will focus only on computers and calculators. Since computers are worthless without software, a section on software is also included.

Calculators

The story of HP's entry into the calculator business started in 1965 with the development of the HP 9100A Calculator. Four ingredients helped to pull this product together. The first ingredient was the product of a young inventor who had perfected a simple calculator using reverse Polish notation. At about the same time another inventor brought to the attention of HP management another prototype calculator that could calculate transcendental functions—this was the second ingredient. The third ingredient was the recognition that these two inventions could be brought together to build a powerful calculator. The last important ingredient was the availability of a project already under way at HP to develop a read-only memory (ROM). This ROM, which was composed of 32,267 bits, was eventually used to store all the calculating and display routines for the HP 9100A. Two major problems had to be overcome by the HP 9100A development team. The first was packing eight circuit boards (remember the read/write memory was magnetic core and the control logic diode-resistor logic), a power supply, a CRT, a magnetic card reader, and a keyboard into a package 8¼ inches high, 16 inches wide, and 19 inches deep. For the technology available at the time this was quite a challenge. The second major challenge was compressing 27 floating-point arithmetic operations and functional computations into 4K bytes of ROM. This is quite an accomplishment when you consider that software engineers today consider it a challenge to fit a simple application into 1 Mbyte of memory.

The HP 9100A was a very easy to use, programmable calculator that had 196 program steps, 27 single-keystroke mathematical operations including trigonometric and logarithmic operations, and instructions and constants for repetitive operations. Single-keystroke computation is not a big deal today; in fact, we take it for granted, but before the HP 9100A, to do anything more than a simple addition or subtraction was a multistep, error-prone process.

The successor to the HP 9100A, the HP-35, the first handheld scientific calculator, turned out to be one of the most successful products ever introduced by HP. Fig. 7b shows the *HP Journal*

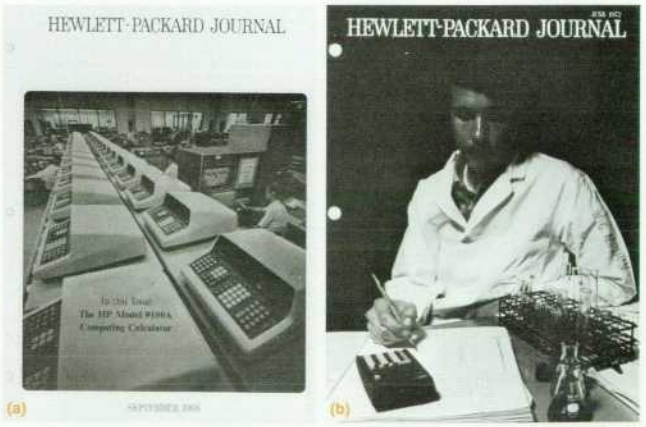


Fig. 7. *(a) Sept. 1968. This entire issue was devoted to the HP 9100A Calculator. The cover shows the final test area for the calculator. The price for the HP 9100A was $4900.00. (b) June 1972. The* HP Journal *issue describing the HP-35. The HP-35 cost $395 at introduction.*

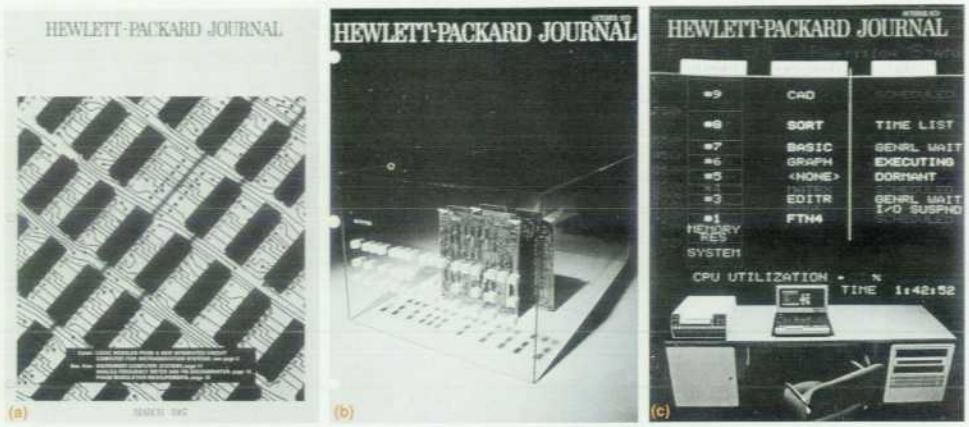


Fig. 8. *(a) March 1967. The issue describing the HP 2116A Computer. The cover shows some of the integrated circuits used in the computer. The machine could handle up to 32K bytes of magnetic core memory with a 1.6-μs cycle time. The HP 2116A also came with a software package that included a Fortran compiler, an assembler, a symbolic editor, and a basic control system (a very primitive operating system). (b) Oct. 1971. The cover shows a mockup of the front-panel buttons of the HP 2100A Computer, and inside the plastic box are two boards containing the microprocessor, which was the heart of the machine. The machine contained a special emulator to emulate the earlier HP 2116A Computer. The instruction set for the HP 2100A was decoded and executed partly in hardware and partly in firmware. This meant that the basic instruction set could be expanded without new hardware. The machine still used magnetic core memory that had a 980 ns cycle time. (c) Oct. 1978. The HP 1000 system, which was introduced in 1976, was based on a successor to the HP 2116A, the HP 21MX Computer. The HP 1000 provided a complete package—computer system, a set of standard computer peripherals, the RTE-IV operating system, and other software. The display shows the output of a utility that monitored the activities of the RTE-IV operating system. The HP 21MX F-Series machine had a system cycle time of 420 ns and a maximum memory size of 1280K bytes in the mainframe. The memory technology was semiconductor dynamic RAM.*

issue that described the HP-35. The HP-35 provided 17 arithmetic, trigonometric, and logarithmic functions, and functions for data storage and positioning. It was battery-powered, weighed nine ounces, was small enough to fit in a shirt pocket, and computed answers to 10 significant digits. At the time the development of the HP-35 began, it was estimated that if 10,000 units could be sold in the first year the program would be a success. HP sold over 100,000 units the first year. The HP-35 completely supplanted the major computational device for engineers and scientists, the slide rule.

Computers

HP's involvement in the computer business also started in 1965 with a minicomputer project at HP Laboratories and a young engineer with a PhD in computer science (a rare degree in 1965). HP Laboratories carried the research far enough to demonstrate the feasibility of the project, and because of customers' increasing demands for automated testing, the project was given the go-ahead. At the time there was a separate HP organization called Dymec that was responsible for doing work related to automated testing, so the responsibility for computer development was transferred from HP Laboratories to Dymec.

HP 2116A. The minicomputer project at Dymec produced HP's first computer, the HP 2116A, in November 1966 (see Fig. 8). Emphasis was placed on the fact that HP was not in the computer business, but that the HP 2116A was intended for instrument automation. Since the machine was intended for instrument automation it was ruggedized to operate under the same environmental conditions as HP's instruments. This gave the HP 2116A a good reputation for reliability.

One of these early automated system applications was the HP 8542A Automatic Network Analyzer (February 1970). The HP 8542A was a combination of microwave test equipment and a computer that was used for characterizing RF and microwave devices. The system was configured either with an HP 2114B or an HP 2116B, two descendants of the HP 2116A. This system demonstrated the advantages of connecting a collection of instruments to a computer specially tailored for instruments.

Despite the emphasis on using the HP 2116A for instrumentation systems, one of its early successors, the HP 2116B, was used to build a time-sharing system. The system was called the Hewlett-Packard Model 2000A Time-Shared BASIC System (July 1968), and was capable of handling 16 users at once. The terminals were teletypewriters and a special teleprinter multiplex card occupied one I/O slot in the host computer to service all sixteen terminals. For direct connection, the system could handle terminals up to one mile away, and for longer distances voice-grade telephone coupling equipment was used. One interesting application of this system was computer-assisted instruction (CAI) (February 1971). The HP/CAI system used an HP 2000B system that could handle up to 32 users. The system provided mathematics drills and practice programs for elementary school children.

The HP 2116A eventually evolved into the HP 2100A (see Fig. 8b), the 21MX, the HP 1000 (see Fig. 8c), and today the HP 1000 A-Series machines (February 1984). These systems continue to be used for real-time applications, such as automatic test systems and data acquisition and control systems.

HP 3000. In 1972 HP introduced its first small general-purpose computer system, the HP 3000 (see Fig. 9). The HP 3000 was a significant milestone because it was HP's first entry into the busi-

Fig. 9. *Jan. 1973. This is not an organist in concert, but a system administrator sitting at the system console for the HP 3000 Computer System. On the table on each side are control panels used for system maintenance and system checkout. In the background is the HP 3000 and some of the peripherals typically connected to the system (card reader, cartridge disk, paper tape reader and punch, and so on).*

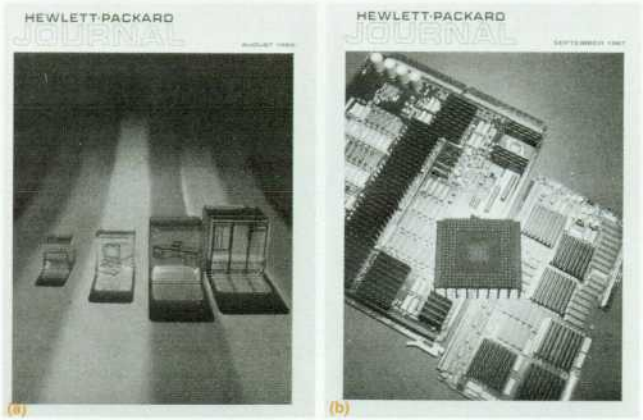


Fig. 10. *(a) Aug. 1985. This was the first of a series of* Journal *issues covering HP Precision Architecture. This particular issue introduced the major design principles associated with HP Precision Architecture. There is a discussion of the genealogy of computer architectures from von Neumann to HP Precision Architecture. (b) Sept. 1987. This issue describes the NMOS-III implementation of HP Precision Architecture computers. The cover shows the processor boards for the HP 9000 Model 825 and the HP 9000 Model 850S/HP 3000 Series 950, and an unmounted pin-grid array package housing an NMOS-III VLSI chip.*

ness computer market, and it was a major change from HP's traditional electronic test and measurement instrumentation business. The HP 3000 was also HP's first truly integrated software and hardware design. Unlike many computers of the past, it was not built by hardware engineers and then turned over to the software engineers to see what they they could do with it—it had the total involvement of both teams in the design. The HP 3000 had features that in the early 1970s were only found in large mainframes, including:

- A hardware stack architecture and virtual memory
- Separation of code and data
- Automatic relocation of programs and data
- Reentrancy to permit code sharing
- Recursion to allow a routine to call itself.

The operating system for this machine, the HP 3000 MPE (Multiprogramming Executive) provided time-sharing and batch processing concurrently. The MPE operating system has evolved into MPE XL, which runs on the new HP Precision Architecture machines.

HP Precision Architecture. HP's Precision Architecture is still technological history in the making. Research on the architecture started in 1981 and was known within HP as the Spectrum program (see Fig. 10a). It was the largest development program ever undertaken by HP. HP Precision Architecture fits loosely within the class known as reduced instruction set computers, or RISCs, and it takes full advantage of VLSI and new software technologies. It does not, however, depend on any particular circuit technology, so instead of being rendered obsolete by the inevitable development of new circuit technologies in the future, it allows designers to exploit new technologies for increased performance gains.

The first systems with the new architecture were introduced in 1986 and included the HP 9000 Model 840 technical computer and the HP 3000 Series 930 commercial system. The Model 840 ran the HP-UX operating system and the Series 930 ran the MPE XL operating system. Both systems used the same processor which is based on a relatively old-fashioned integrated circuit technology, TTL, and yet achieves an average performance rating of 4.5 MIPS, about four times that of the fastest of HP's previous-generation machines. In 1987 two more HP Precision Architecture systems (the HP 3000 Series 950 and HP 9000 Model 850S) were introduced, using HP's proprietary NMOS-III integrated circuit technology (see Fig. 10b). The system processing units, or SPUs, in both systems are the same. These systems verified the point that HP Precision Architecture is portable over different integrated circuit technologies. The VLSI development involved 12 chips with up to 150 thousand transistors per chip and was the largest multichip VLSI project HP has undertaken to date.

The saga of HP Precision Architecture computers continues with higher performance being achieved with faster VLSI parts, bigger caches, floating-point coprocessors, and other enhancements. For instance, one of the latest versions, the system processing unit for the HP 3000 Series 935 and the HP 9000 Model 835, has been benchmarked at 14 MIPS (June 1989).

Software

Software today is like the knobs and dials on instruments. It is expected to be there and provide the human interface to the machine. At the time HP started to develop and produce computers (1966), most work in computers at the time focused on commercial and business applications and special-purpose machines for military applications. There were, however, many innovative concepts in use and being developed. For example, the technique of virtual memory was being developed to get around the problem of limited core memory space, and high-level languages such as Fortran, BASIC, and Algol were beginning to be accepted as standards for interfacing to computers.

Since the HP 2116A was strictly intended as an instrumentation computer all of the software development was focused on making it easy for the machine to communicate effectively with instruments. As a result, many of the early *Journal* articles on software were about software for automatic test and measurement systems (see Fig. 11a). With the introduction of the HP 3000 Computer

Fig. 11. *(a) Nov. 1968. The first* HP Journal *article in which a software product made the front cover and where a software product filled half of the issue. HP's motivation for developing a BASIC (Beginner's All-purpose Symbolic Instruction Code) compiler grew out of the need to provide HP 2116A users with a simple interface to the machine for writing programs for automatic testing. The project started in May 1967 with the goal of developing a single-terminal BASIC compiler to run on any HP computer having at least 8K bytes of core memory and an ASR-33 teletypewriter. (b) July 1974. This issue contained an article on the HP Image/3000 Data Base Management System. The images on the head are supposed to represent information that must be organized, made readily available to those who need it, and protected from those who don't—capabilities inherent in Image.*

System, HP was officially in the computer business and therefore in the software business (see Fig. 11b). One software product, the HP Image/3000 Data Base Management System, was rated the best in its class for three consecutive years by Datapro International (see Fig. 11c).

Today at HP, software design and development occupy the same level of importance as many hardware developments of the past. One area where this is noticeable is in instrument design and development. Since the introduction of microprocessors into instruments, firmware routines have added a large amount of functionality to instruments and in some cases replaced some functions formerly performed in hardware. As a result, almost every new instrument development has a complement of hardware and software design engineers working in parallel to develop an instrument. The synergism between hardware and software engineers was evident in the development of HP Precision Architecture discussed earlier.

Journal pages for the last five or six years have reflected the increasing importance of software in HP. In some cases over half an issue has been devoted to software products and research efforts.

Components

In the mid-1950s HP established a semiconductor operation to design and produce special-purpose diodes for use in instruments. Commercial diodes, although less expensive, did not have the performance required for HP instruments. To gain a foothold in the rapidly emerging field of semiconductors, a subsidiary called HP Associates was founded in 1961. The purpose of this organization was to perform research, development, and manufacturing in the semiconductor field (see Fig. 12). Today there are many entities within HP doing this type of work. The components produced by these organizations form the cornerstone of many HP products—from instruments to computers (see Fig. 13).

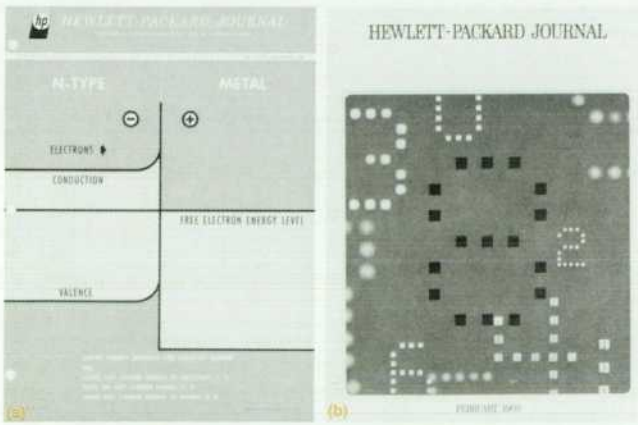


Fig. 12. *(a) Dec. 1965. The cover shows the energy diagram for a Schottky barrier. This issue describes the hot-carrier diode, a device that was used as a very fast detector, as a microwave mixer, and as a switch. Hot-carrier diodes were distinguished from conventional semiconductors at the time in that the junction consisted of a metal and a semiconductor rather than two semiconductors. (b) Feb. 1969. The cover shows the HP Model 5082-7000 Numeric Indicator, a small, low-power, all-semiconductor module, which used four-line binary-coded-decimal input signals to display digits 0 through 9 as an array of brightly glowing red dots. Each red dot is a GaAsP light-emitting diode. Today one can get LEDs in other colors besides red, and the new red AlGaAs LEDs (August 1988) are bright enough to be considered for such applications as light bulbs in automotive tail lights, airport markers, and traffic signals.*

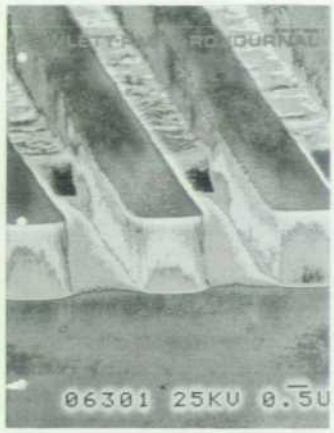


Fig. 13. *Aug. 1982. The cover shows a solarized version of the profile of a trilayer process after reactive ion etching of the bottom polymer layer. This issue describes HP current and future efforts in integrated circuit process technology, particularly in the area of very large-scale integration (VLSI).*

In instruments, special thin-film and thick-film hybrid packages containing such things as amplifiers and samplers are used to lower costs and improve performance. Likewise, in computer products, VLSI (very large-scale integration) circuits implement fast floating-point processors, ALUs, and other circuits. VLSI circuits are also being incorporated into instruments to supplant functions formerly performed by discrete hardware or by software.

Not to be Forgotten

There are many more product innovations that appeared in the *HP Journal* that can be considered to have made significant contributions to the technological history of Hewlett-Packard, but there is not enough room in this article to discuss them. Therefore, to ensure that these products are not forgotten the following is a list of these products and their associated *HP Journal* references.

- HP 8551A/851A Microwave Spectrum Analyzer (August 1964)
- HP 5000A Logic Analyzer (October 1973)
- Time-Domain Reflectometry
  - HP 1415A Time-Domain Reflectometer Plug-in (February 1964)
  - HP 8145A Optical Time-Domain Reflectometer (December 1988)
- HP 5525A Laser Interferometer (August 1970)
- HP 3406A Broadband Sampling Voltmeter (July 1966)
- HP-IB (IEEE 488 and IEC 625)(October 1972, January 1975)
- Chemical Analysis Instruments
  - HP 8450A Spectrophotometer (February 1980)
  - HP 1090 Liquid Chromatograph (April 1984)
- Medical Instrumentation and Systems
  - HP Model 47210A Capnometer (CO_2 Analyzer) (September 1981)
  - HP 77020A Ultrasound Imaging System (October 1983)
- Printers
  - HP ThinkJet (thermal inkjet printer)(May 1985)
  - HP DeskJet and HP PaintJet printers (September 1989, October 1989)
- Plotters
  - HP's low-mass, low-inertia plotting technology (October 1981)
  - HP 7580A Drafting Plotter (November 1981)

References

1. A. Thiesen, *A History of the General Radio Company*, General Radio, 1965, p. 19.
2. *Inventions of Opportunity: Matching Technology with Market Needs*, Hewlett-Packard Company, 1983.
3. *Ibid.*, p. x.
4. *Ibid.*, p. xi.
5. *Ibid.*, p. viii.

A Modular Family of High-Performance Signal Generators

Three signal generators, each designed for a particular type of application and each offering several options, let the user choose and pay for exactly the capability required.

by Michael D. McNamee and David L. Platt

ADVANCES IN DIGITAL AND RF TECHNOLOGY over the last decade have made the RF signal generator a common tool on the workbenches of design engineers. High-quality commercially available generators offer excellent frequency stability, wide tuning range, reasonable switching speed, and a broad array of modulation types, including AM, FM, ΦM, and pulse. However, while LSI and design for manufacturability have brought down the price of the typical generator, most of these signal generators are general-purpose in nature and users are often forced to pay for performance that they don't need.

As test system complexity grows, the need to reduce both the price and the rack space of each piece of test equipment becomes critical. This requirement is the driving force behind such externally modular schemes as the VXIbus and the HP Modular Measurement System. Modularity offers the flexibility to tailor a system more specifically to a given application. With this in mind, HP has developed a new family of internally modular signal generators called the Performance Signal Generators (PSG). The three members of this family are the HP 8644A 1-GHz or 2-GHz Synthesized Signal Generator (Fig. 1), the HP 8645A 1-GHz or 2-GHz Agile Signal Generator (Fig. 2), and the HP 8665A 4.2-GHz Synthesized Signal Generator (Fig. 3).

These three signal generators are designed for three specific application segments. The HP 8644A is for the traditional out-of-channel transceiver test applications. The HP 8645A Agile Signal Generator is focused on frequency agile transceiver testing. The HP 8665A is designed for high-performance applications up to 4.2 GHz, particularly radar, telemetry, and spurious testing of UHF transceivers.

The internal modularity of the three PSGs allows a user to select among several options for each that provide a wider range of performance at a lower price than a single high-performance signal generator can provide. The user pays only for the performance required. Optional capabilities include phase noise enhancement, extended frequency range, numerically synthesized internal modulation, high-performance pulse modulation, and improved RFI leakage.

User-Selectable Performance Trade-Offs

The most unusual feature of all three PSGs is the ability to make trade-offs among phase noise performance, maximum FM deviation, and switching speed from the front panel. This is done manually by selecting the desired mode of operation or automatically by selecting auto mode. When the PSG is in auto mode, the instrument automatically selects the mode that allows the lowest phase noise for a given carrier frequency and selected deviation.

Both the HP 8644A and the HP 8645A offer one standard and one optional enhanced mode. In the standard enhanced mode, the typical phase noise at a 20-kHz offset on a 1-GHz

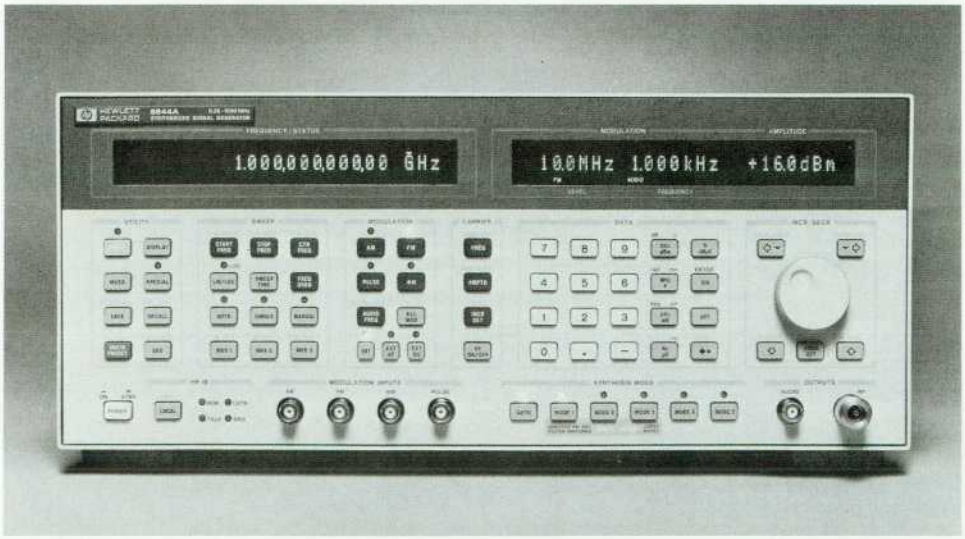


Fig. 1. *The HP 8644A Synthesized Signal Generator comes in 1-GHz and 2-GHz versions and is designed for out-of-channel transceiver testing.*

Fig. 2. *The HP 8645A Agile Signal Generator is available in 1-GHz and 2-GHz versions. It is designed for testing frequency agile transceivers.*

carrier is around −134 dBc/Hz. The optional enhanced mode improves this to −142 dBc/Hz. The HP 8665A offers one optional enhanced mode which improves the typical 20-kHz noise on a 1-GHz carrier from −124 dBc/Hz to −142 dBc/Hz. In an enhanced mode, a frequency discriminator detects the FM noise on the internal oscillator and cancels it in a feedback loop. This lowers the phase noise substantially over that of the phase-locked oscillator alone. However, a trade-off is required. The maximum available FM deviation is lower and frequency switching time is longer than in the unenhanced mode. Fig. 4 shows the phase noise obtained with the three available modes of the HP 8644A.

Surface Mount Technology

For many years, surface mount technology (SMT) has been used in critical RF circuitry because of its improved and consistent RF performance. Only recently has it been recognized that SMT offers improved reliability and lower-cost parts placement in addition to smaller size. The PSGs are the first signal generators from HP using SMT extensively, including the power supply and controller. An extensive SMT process center was established specifically for automatic loading and soldering of the PSG surface mount assemblies. Electrical designers were asked to converge on a minimum set of "core" SMT components. As a result, approximately 80% of the components in the PSGs represent less than 400 SMT components. This approach reduces the cost traditionally associated with purchasing and maintaining large numbers of leaded components.

Automatic Calibration

Manual calibration adjustments are time-consuming and costly. Moreover, the components that allow these adjustments, trimmer potentiometers and capacitors, have higher failure rates than fixed-value components. For these reasons, the PSG designers replaced manual adjustments wherever possible using either precision fixed-value com-

Fig. 3. *The HP 8665A Synthesized Signal Generator is a 4.2-GHz instrument designed for radar, telemetry, and UHF transceiver testing.*

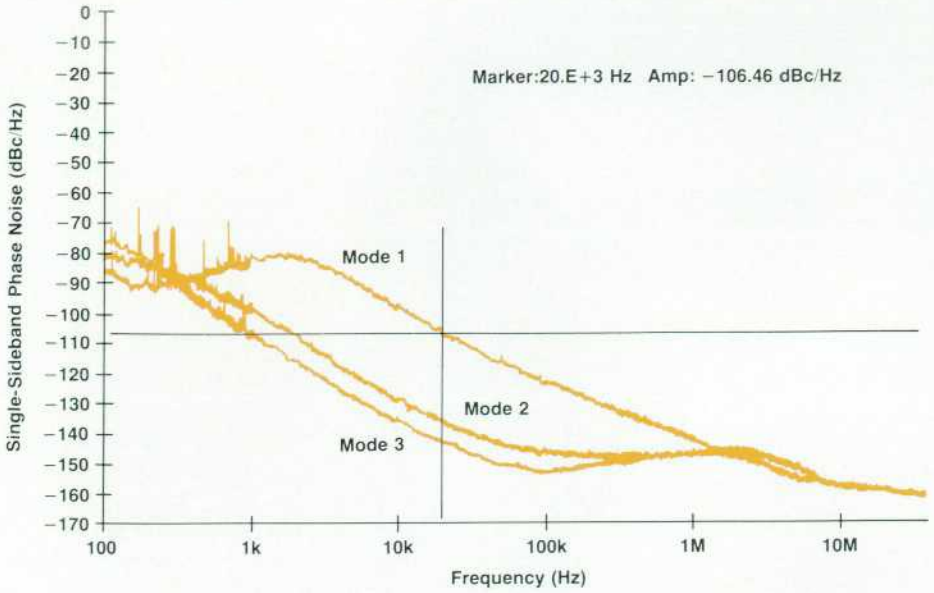


Fig. 4. *HP 8644A phase noise in the normal and two enhanced modes.*

ponents or digital-to-analog converters (DACs) controlled by instrument firmware. As a result, the HP 8644A and HP 8645A have no manual adjustments. All calibration is done automatically by the microprocessor controller using precision internal sources and calibration data stored in EEPROM. Even with its complex YIG circuitry, the HP 8665A only requires a few manual adjustments for calibration.

Single-Loop Design Concept

All three PSGs are based on the same fundamental frequency synthesis scheme, shown in Fig. 5. This scheme uses a single phase-locked loop (PLL), which is a fractional-N loop of the type used in many earlier HP signal generators. This loop provides the required frequency resolution as well as highly accurate digital FM at rates within the loop's bandwidth. The single-synthesis-loop approach has the advantages of better spurious performance and lower overall cost, parts count, and mechanical complexity. The single-loop approach, with its low reference frequency, requires a relatively narrow loop bandwidth to prevent multiplied reference phase noise from dominating the output phase noise. This impacts both switching speed and power line spurious cleanup. The PLL bandwidth of the PSGs varies between 150 Hz and 3 kHz.

Single-loop synthesizers have not been widely used in general-purpose signal generators. Typically, multiple loops are used—one loop providing frequency modulation, another providing high-resolution frequency control, and finally a sum loop, in which the VCO output is mixed with the outputs of the other loops and fixed reference frequencies to produce the signal that is fed to the phase detector. The reason the single-loop approach has not been used is that it has been difficult to achieve both fine frequency resolution and good spurious performance close to the carrier at the same time with a single phase-locked loop. The high-resolution fractional divider developed for the PSGs overcomes these problems, allowing 0.01-Hz frequency resolution and spur levels typically less than −48 dBc at less than a 1-kHz offset from a 3-GHz carrier.

This synthesis scheme results in an unusual approach to frequency modulation (FM). Traditionally, FM is done in a phase-locked loop by injecting the modulating signal in two places—directly into the VCO to generate the FM and also into the loop integrator (after integration) to prevent the loop from trying to cancel the FM that was generated within its bandwidth. In the PSG approach, the FM that occurs within the loop bandwidth is not injected in an analog fashion, but digitally, by digitizing the FM input and adding it to the instantaneous frequency data that is fed into the fractional-N control chip. This ensures that the loop does not try to remove the FM, since it is programmed to generate the correct instantaneous frequency.

With this scheme, FM occuring at rates well within the loop bandwidth is done digitally and FM at rates well above the loop bandwidth is done in the traditional analog

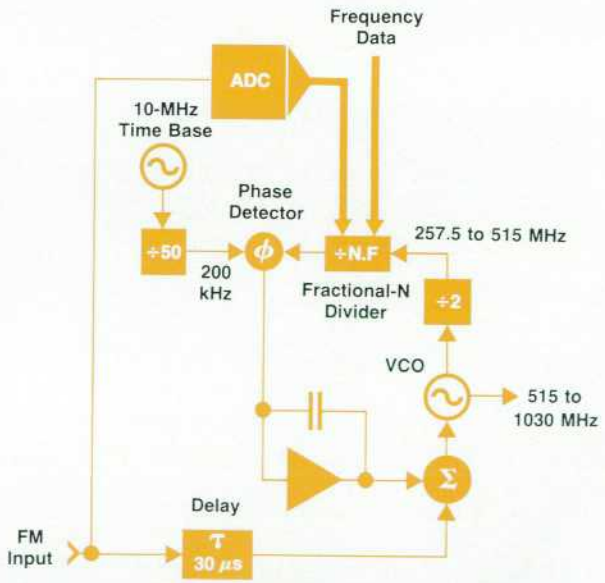


Fig. 5. *The same fractional-N frequency synthesis system is used in all three generators. However, in the HP 8665A, the voltage-controlled oscillator (VCO) runs at 3 to 6 GHz.*

fashion. A major drawback is that the digital FM is delayed by 30 microseconds, the time required to digitize it and clock it into the fractional-N control chip. This results in a large group delay variation, from 30 microseconds at low rates to less than a microsecond at high rates. For the loop bandwidth of 3 kHz, this would cause a large FM inaccuracy at rates around 3 kHz. Moreover, such a high group delay variation would cause excessive FM distortion for complex modulating signals. To correct this problem, a 30-microsecond delay is added into the FM path to the voltage-controlled oscillator (VCO). This causes all FM to experience a 30-microsecond delay, and the variation in group delay is reduced to only a few microseconds.

HP 8645A Concept

The first of the three PSGs to exit the investigation phase was the HP 8645A Agile Signal Generator. A basic block diagram of the HP 8645A is shown in Fig. 6. In addition to the single PLL described previously, the HP 8645A also contains one or two delay line frequency discriminators to detect and feed back FM noise to the VCO for cancellation. When operated in mode 1, the HP 8645A uses only the PLL for frequency stability. This mode offers the fastest synthesized frequency switching speed and greatest FM deviation. Mode 2 adds a frequency-locked loop (FLL) containing a 70-nanosecond discriminator, which improves the phase noise at the expense of reduced deviation and switching speed. Mode 3 adds an optional 140-nanosecond discriminator which extends this trade-off even further, reducing the phase noise another 8 dB.

The HP 8645A also offers a fast frequency hopping mode (mode 5). In this mode, the PLL is opened but the 70-nanosecond FLL remains connected. Thus, the VCO remains frequency locked but not phase locked. Fast hopping is done using several DACs to set control and pretune voltages in the FLL to force the VCO frequency to change very rapidly. The exact control voltages are determined by the microprocessor controller during a "learn" process prior to fast hopping.

In the fast hopping mode the frequency can switch in less than 15 microseconds. The frequency accuracy, which is a function of the drift of several critical components in the FLL, is typically better than ±1 part per million. Frequency hopping of the HP 8645A is discussed further in the article on page 34.

The VCO output of 515 to 1030 MHz is digitally divided to the desired output frequency (doubling is also available with Option 002). A switched, half-octave low-pass filter follows the divider to reduce harmonics. The signal is amplified and leveled in an automatic level control (ALC) loop. AM is performed in the traditional manner within the ALC loop bandwidth. The loop bandwidth can also be narrowed greatly to improve intermodulation performance to better than −50 dBc. The output attenuator is a set of two mechanical attenuators from the HP 33320 Series. This is followed by a reverse power protection relay that protects the output from an accidental application of up to 50 watts of RF.

HP 8644A Concept

It was recognized that the single-loop synthesizer with discriminator-enhanced spectral purity is ideal for out-of-channel transceiver testing because of its ability to achieve −100 dBc spurious and −134 dBc/Hz phase noise at a reasonable cost. This led to the concept of the HP 8644A as a cost-reduced version of the HP 8645A.

During the investigation phase of the HP 8644A it was tempting to scale down some of the common modules such as the power supply and the microprocessor controller, since that hardware had much more capability than we actually needed. While this redesign would have allowed a lower HP 8644A cost, it would not have allowed some of the reuse advantages discussed later in this article. Therefore, we focused on changes in the VCO and modulation circuitry.

The HP 8644A block diagram is similar to that shown for the HP 8645A except that the fast hop controller is not present. The approach taken was to remove all the circuitry associated with fast frequency hopping as well as the high-rate FM capability. This allowed us to combine the VCO, standard discriminator, and phase shifter functions into a single module. The modulation drive circuitry was greatly simplified and standard modulation features, such as FM preemphasis and internal 300-Hz, 400-Hz, 1-kHz, and 3-kHz modulation oscillators were added.

The HP 8644A output section is identical to the HP

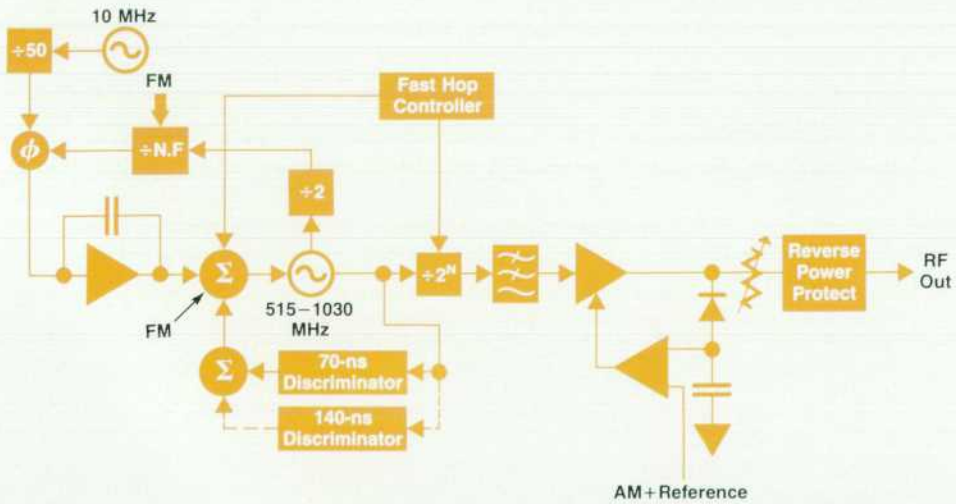


Fig. 6. *Block diagram of the HP 8645A Agile Signal Generator.*

8645A output section except that an optional high-reliability 1-GHz electronic attenuator (Option 005) is also available. This attenuator uses pin diodes instead of mechanical switches and is identical to the attenuator used in the HP 8657A Synthesized Signal Generator.

HP 8665A Concept

The HP 8665A concept was driven by two factors. First, users find it increasingly necessary to push RF signal generator performance into the low microwave region. For example, 900-MHz radio manufacturers need spurious testing capability to at least 4 GHz. Spurious testing requires precise level control to very low signal levels and low adjacent channel spurious content. The HP 8665A's −100-dBc spurious performance and ±1-dB level accuracy at 3 GHz allow these high-performance measurements.

The second driving factor was the desire for high reuse of the single-loop, discriminator-enhanced PSG block diagram. By using whole modules from the other PSG projects along with a few special modules, the HP 8665A development required far less in engineering resources than would have been required if the product had been designed "from scratch."

The HP 8665A block diagram, Fig. 7, is very straightforward. The synthesis is accomplished by a single phase-locked loop. A phase-noise-selected YIG oscillator running from 3 to 6 GHz is prescaled by four with a specially developed GaAs divider to 750 to 1500 MHz. A silicon ECL divider further divides the signal to the phase detector module input frequency of 257.5 to 515 MHz.

The phase detector module is the same fractional-N module used by the HP 8644A and HP 8645A. This module is highly complex and handles much of the FM function as well as the loop synthesis.

FM is added to the YIG's output via two paths. For dc and low frequencies the modulation signal is applied to the fractional-N divider by digitizing the signal. At higher modulation frequencies, the modulation is applied outside the loop bandwidth through the YIG driver. A detailed discussion of the synthesis section can be found in the article on page 37.

The spectral purity of the HP 8665A can be enhanced with the use of a discriminator attached to the input to the phase detector module. The error signal generated by the discriminator is an indication of the phase noise content of the synthesis loop and is applied through the YIG driver to reduce the noise of the output. The discriminator module is another of the common assemblies developed for the PSG family.

To allow phase noise to be scaled with frequency, the output section is constructed around a hybrid approach found in most of HP's high-performance RF generators. For frequencies from 187.5 MHz to 4200 MHz, the signals are derived by frequency division. For the highest octaves, the frequency division is done by the GaAs divider in the synthesis loop. To preserve the FM deviation at lower frequencies, a heterodyne approach is used below 187.5 MHz. A 1-GHz reference is developed from the 10-MHz time base and this signal is mixed with a 1000-to-1187.5-MHz frequency range produced by a divider in the output section to form the output signal. The HP 8665A output system is discussed in the article on page 42.

Common Hardware

The following sections describe hardware that is used in all three PSG products.

Fractional-N Module

The PSG fractional-N module performs the fractional division, phase detection, and integration functions of the PSG phase-locked loop. It also digitizes the low-rate FM and sums it with the desired frequency information. The RF input is 257.5 to 515 MHz and the reference input is a 200-kHz pulse train supplied from the reference module by dividing down the 10-MHz time-base. The fractional-N board, shown in Fig. 8, is a six-layer Class III printed circuit board. This board is extremely complex, and because of the relatively high reference frequency, required several iterations to reduce the fractional-N spurious below −60 dBc.

Two custom integrated circuits, which can be seen in Fig. 8, make this level of complex synthesis circuitry possible. The first is a four-watt ECL gate array, which performs the fractional division (÷N.F). The second is the fractional-N control chip, which combines the FM input with the desired carrier frequency and controls the fractional division.

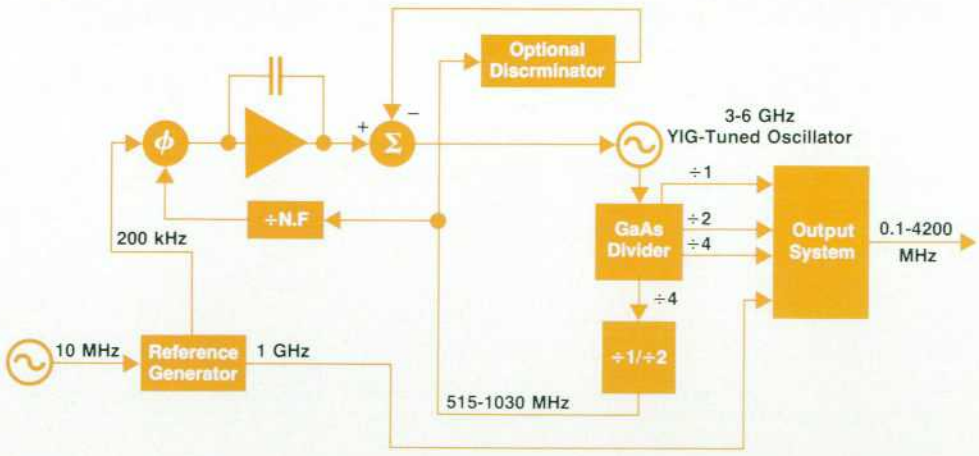


Fig. 7. *Block diagram of the HP 8665A Synthesized Signal Generator.*

Fractional frequency division generates spurious signals, as already indicated. A technique known as automatic phase interpolation (API) allows these spurs to be largely cancelled. This cancellation requires extremely precise control of very small currents, and is only effective to −50 to −60 dBc. The API correction in the PSG uses an automatic spur nulling technique in which the spurs are synchronously detected and the API currents are adjusted to minimize the spur levels. For more information, see the box on page 28.

PSG Modulation Oscillator

Early in the PSG development we realized that simulating complex modulation signals is becoming difficult for our customers. With the advent of integrated implementations of numerical synthesis techniques, a powerful solution exists. The key synthesis IC developed for the HP 8904A Multifunction Synthesizer[1] was used to build the PSG's compact and versatile modulation assembly. This assembly consumes little more than 15 square inches of board space yet allows synthesized modulation signals from dc to 400 kHz. Square waves and Gaussian noise can be generated in addition to sine waves. The oscillator can produce two simultaneous waveforms to emulate many signaling tones and it can internally modulate itself. Modulation waveforms such as those used in aircraft VOR and ILS applications can now be generated totally inside the signal generator.

PSG Front Panel

The PSG front panel assembly deviates from those found on previous generators in two ways. An in-depth reliability study showed that keyswitch problems accounted for a disproportionately high number of failures. This drove our choice of rubber keyswitches for the front panel. Rubber keyswitches have been shown to have a very low failure rate compared with their mechanical counterparts.

The other area of difference is the use of vacuum fluorescent displays. These devices are rugged and reliable, and have very high readability. Special consideration was given to their driving waveshapes to minimize RFI emissions from the generator and RFI effects on the low-spurious circuitry inside the product.

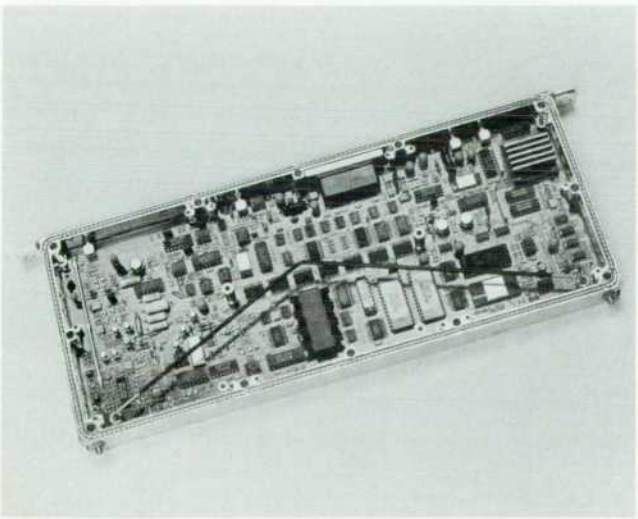


Fig. 8. *Two custom ICs, an ECL gate array and a control chip, make the small size of the fractional-N module possible.*

PSG Microprocessor Controller

The PSG digital control unit controls all of the PSG hardware. The design is based on an 8-MHz Motorola 68000 microprocessor. The control board holds 512K bytes of ROM, which contains all the instrument control firmware. This board also contains the HP-IB interface, which uses a TMS9914. Interrupt handling, timers, and serial communication to the front panel are provided by a Motorola 68901 multifunction peripheral chip. For setting up the RF hardware and performing diagnostics, the board includes voltmeter circuitry to measure dc or rms voltages throughout the instrument. The voltmeter and a simple RF detector provide basic power measurement capability for RF hardware diagnostics. The processor execution stack and instrument state reside in 64K bytes of static RAM. Store/recall registers and the instrument state are maintained by a NiCad-battery-backed RAM when the instrument is powered down. The battery is charged during normal operation by an on-board regulator. The processor communicates with the rest of the instrument through memory-mapped parallel I/O ports.

PSG Power Supply

The PSG power supply is a linear supply capable of 180 watts distributed among several supply voltages. The supply is designed for extremely low levels of conducted and radiated line-related spurious outputs. Careful attention was paid to the layout to minimize magnetic loops and ensure that sensitive circuitry and noisy rectifier circuits did not have common ground return paths.

The PSG power supply contains circuitry to shut down the supply when the internal temperature exceeds safe limits. This can occur, for example, if the airflow from the fan is interrupted. Circuitry also exists to vary the fan speed with temperature to improve instrument reliability at high temperatures while maintaining a quiet fan at room temperature. Precision reference ICs are used in the regulators to avoid manual adjustments of the supply voltages.

Modularity Effects in Development

An important benefit of internal modularity lies in the resulting potential for reuse of circuitry among instruments. One of the primary PSG design goals was to make as much hardware and firmware as possible common to all three PSGs. Using this approach, the development of these three high-performance products took place almost as quickly as a single product, and used only slightly more resources.

Probably the most difficult phase in a product's development is the production buildup. Unforseen yield and process problems show up, along with design tolerance problems in mechanical parts as a greater number of parts are processed. Debugging and streamlining of test and assembly procedures is required. This phase takes a lot of time and resources and anything that can be done to shorten this critical learning process will help.

The modularity of the PSG family was an important ad-

vantage in this phase. Because of the high degree of reuse, we are able to build all three signal generators on the same production line using the same assembly processes and test equipment. Even in the product-specific modules, designers were constrained to use a common pool of surface mount parts and design techniques. The result was that, while it took more effort to handle all three products initially, the overall time to get through the learning process was much shorter than it would have been with three very different products produced on their own production lines.

The development effort for the PSG family was organized into three separate concurrent projects. In addition to product-specific modules, each project developed one or more of the modules used in all three signal generators. The HP 8644A project developed the fractional-N synthesis circuitry. The HP 8645A project developed the power supply, microprocessor controller, and low-noise frequency discriminator. The HP 8665A project developed the front panel and numerical-synthesis modulation source. This concurrent development required a high degree of discipline on the part of the project designers to reuse hardware and firmware and to resist the urge to optimize by redesigning.

Another key component of this concurrent development was interproject communications and management. The magnitude of the overall management task was much greater than originally thought. It was necessary to maintain rigid, detailed I/O specifications on each assembly. In addition, all three project managers met regularly with their section manager and with manufacturing, marketing, QA, and finance people.

Acknowledgments

The authors would like to thank Ray Fried for his fundamental contributions to the PSG concept and his support throughout the development. We would also like to thank Dave Molinari for his contribution as HP 8645A project manager. Finally, we would like to acknowledge the tremendous effort on the part of numerous R&D, production, and assembly and test engineers who worked on the PSGs during the initial production phase.

Reference

1. F.H. Ives, "Multifunction Synthesizer for Building Complex Waveforms," *Hewlett-Packard Journal*, Vol. 40, no. 1, April 1989, pp. 52-57.

Firmware Development for Modular Instrumentation

Of three major subsystems in the Performance Signal Generator control firmware, only one contains instrument-specific code. Additional hardware and firmware for calibration and diagnostic purposes provide important customer and production benefits.

by Kerwin D. Kanago, Mark A. Stambaugh, and Brian D. Watkins

THE PERFORMANCE SIGNAL GENERATOR (PSG) product line described in this issue represents an internally modular platform approach to the development of signal generators. There are currently three full-function signal generators in this family, each with a number of options. The similarities between these signal generators, including a large amount of shared RF hardware, suggests that the firmware to control them can be shared or highly leveraged between versions. This is in fact the case.

The major design objectives for the control firmware were:

- Reuse as much firmware as possible to reduce development and test time relative to developing three different sets of signal generator firmware.
- Structure the firmware so that it can be extended to allow for future versions and options.
- Give the three instruments a common look and feel from the front panel.
- Use consistent symbolic HP-IB control language and formats for all three signal generators that could also be used in future signal generators.

Since reuse was a key issue in PSG control firmware development, the majority of the control firmware was written in a high-level language. Only time-critical routines, code that directly interfaces with hardware, and certain library routines were written in assembly code. The control firmware for all three PSG instruments consists of 100 KNCSS (thousands of noncomment source statements) of Pascal source code and 8 KNCSS of assembly language

source code. The firmware executes on a 68000 microprocessor with 512K bytes of ROM and 64K bytes of RAM.

Control Firmware Architecture

The control firmware for the PSG family is partitioned into three major subsystems (Fig. 1). They are the KRNL (operating system kernel), the UI (user interface), and the RFHI (RF hardware interface). The KRNL contains the operating system, support libraries, and low-level I/O drivers. It passes input data from the front panel and HP-IB to the UI. The UI takes character data from the KRNL and converts it to internal instrument commands with parameters which are passed to the RFHI. The UI also formats data to be output via the HP-IB (IEEE 488, IEC 625) or displayed on the front panel. The RFHI section takes commands from the UI, range checks the parameters, and configures the hardware appropriately. The RF hardware interface also stores the current instrument configuration.

To keep the interface between subsystems small and well-defined, the KRNL, UI, and RFHI communicate through a command interface. Each major subsystem has a do_command procedure, which has two parameters. The first parameter is an element of an enumerated type which defines internal instrument commands. The second parameter is a pointer to any data needed with this command. For example, the KRNL may pass a character to the UI by calling the procedure UI_do_command, with the command Parse_HP-IB_character, and a pointer to the character to parse. The type of data pointed to by the second parameter depends on the command. It may point to a real, integer, Boolean, enumerated, or structured type. It may also be a nil pointer with commands like Preset_the_instrument_state, which need no data.

The KRNL is identical for all three PSG instruments, since they all share a common microprocessor board and operating system. Although there are some differences in the operation of the three instruments, they also share a common UI. Where the three instruments have different ranges for the same setting, such as the allowable frequency range, the user interface accepts any value for frequency and passes it to the RFHI. The RFHI determines whether that value is acceptable for that instrument, and either sets the hardware to that value or signals the user interface to show an error message. Each instrument has one or more functions not found in the others. The user interface will always accept the commands to operate these functions, but the RFHI will signal an error if the function is not available.

Differences in the RF hardware of the three signal generators required three different RFHI firmware sections, which share some of their source code. The amount of unique source code was minimized to save development and test time. Since the HP 8645A was developed first, its firmware was used as a base. The HP 8644A shares much of the HP 8645A hardware and firmware. It required about 5K lines of HP 8644A-specific firmware. The HP 8665A shares less hardware and required about 14K lines of HP 8665A-specific firmware.

Operating System Kernel

The KRNL contains all of the operating system, run-time libraries, exception processing, and support for the hardware on the microprocessor board. The code here is specific to the processor hardware and the execution environment. It is not dependent on the RF hardware being controlled or how the user controls it.

The KRNL includes low-level I/O drivers for the front panel and the HP-IB. These drivers implement HP-IB and front-panel protocols. They pass character data back and forth between the external world and the user interface firmware without any knowledge of what the data represents.

User Interface

The UI manages all interaction with the user. It contains all of the firmware to parse incoming characters and format messages to be sent to the front panel or the HP-IB. All information used to parse input and format output is contained in tabular format to allow easy modifications or addition of features as new options are developed. The user interace generates commands and parameters for the RF hardware control firmware.

There are five major sections in the UI (Fig. 2). They are the keyword scanner, the command parser, the HP-IB output formatter, the keyboard parser, and the display control. The scanners and parsers process incoming data as it comes in, rather than after whole messages are received, to give the fastest possible response. The display section manages the display, which is accessed via a serial bus, and attempts to minimize the number of characters sent to the display.

HP-IB Keyword Scanner. Characters from the HP-IB are passed to a keyword scanner. The scanner converts the alphanumeric characters to tokens representing keywords. The scanner is implemented as a state machine, with a tree of characters in allowable keywords described by a table in ROM (Fig. 3). The scanner walks through this tree structure one character at a time. When the end of a keyword is encountered, the position on the scanner tree is checked. If the current position represents a valid keyword, then a

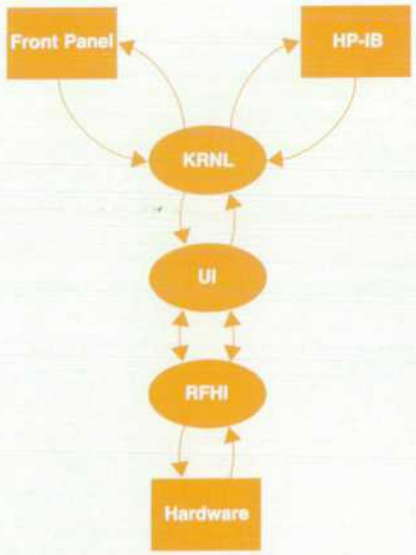


Fig. 1. *Performance Signal Generator (PSG) control firmware subsystems.*

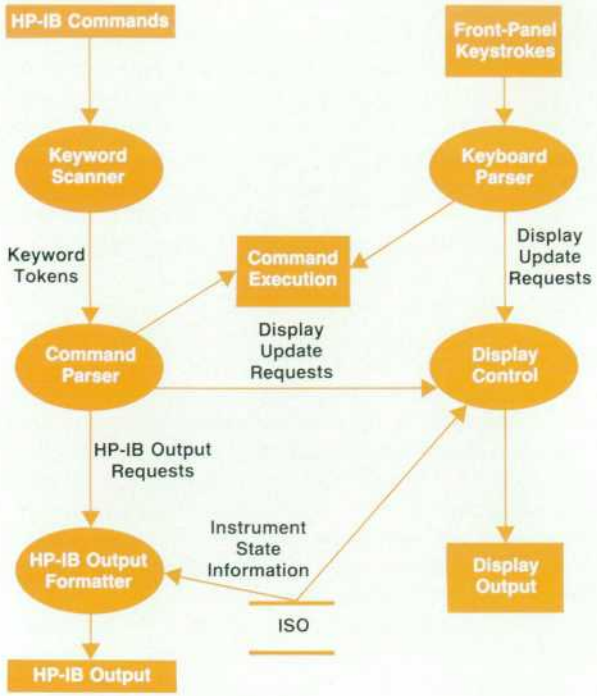


Fig. 2. *Structure of the user interface module (UI) of the PSG control firmware. The ISO is the instrument state object, a central data base.*

token representing that keyword is output to the HP-IB command parser. If a character is sent to the scanner that does not match an allowable keyword, the scanner will generate an invalid keyword error. If the characters coming in from the HP-IB represent a number, they are converted from ASCII characters to a binary representation of that number. The number is saved and a token indicating a number was received is sent to the command parser. Generation of the keyword parsing state table was simplified by the development of a utility program. This program reads in a list of all acceptable keywords and the tokens to output when they are encountered. From this input, the program generates a Pascal-compilable lookup table for the scanner.

HP-IB Command Parser. The command parser operates much like the scanner. However, it accepts tokens rather than ASCII characters. It walks through a list of valid token syntax until it encounters invalid syntax or the end of a command. At that point, if a valid command has been processed, it outputs an internal instrument command to the RFHI for processing. If the command contains a numeric parameter, such as a frequency to which to set the signal generator, the value is passed with the command as its parameter. The command parser can also generate requests to the display control to update part or all of the display and requests to output data to the HP-IB as needed. The syntax information for the command parser exists in a lookup table. This table is generated by a utility program, which reads in a list of allowable combinations of keywords and the commands to execute when they are encountered.

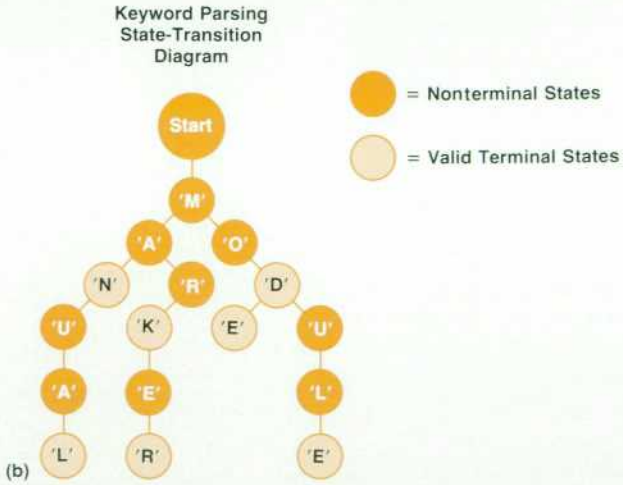
HP-IB Output Formatter. The HP-IB output formatter accepts data output requests from the scanner and formats the data into ASCII character strings. These strings are passed to the low-level HP-IB driver in the KRNL. Requests for data output are accompanied by the name of the variable to be output. The formatter has no implicit knowledge about the variable to be output. All information about a variable to be output is obtained from a data base inside the RFHI. In addition to the current value of the variable, the data base also tells the formatter the type (real, integer, Boolean, enumerated, etc.) and units of the variable. The data base is called the ISO, for instrument state object, and is described in detail later.

Keyboard Parser. The keyboard parser operates like the keyword scanner or command parser, but accepts front-panel keystrokes. It keeps track of the active function and uses that information to interpret the keystrokes passed to it. It has its own numeric parser which is coupled to the display control section so that the numbers appear on the display as they are entered. When numbers and terminators are entered or immediate-action keys are pressed, the keyboard parser generates internal instrument commands to the RFHI. As with the command parser, the numbers entered are passed as values to the commands. The keyboard parser also generates requests to change or update the contents of the display. Like the token scanner and command parser, the keyboard parser's operation is based on a lookup table. This table is generated by a utility pro-

Example Input to Keyword Table Generator

Keywords	Token to Output
'MAN'	TOKEN_MANUAL
'MANUAL'	TOKEN_MANUAL
'MARK'	TOKEN_MARKER
'MARKER'	TOKEN_MARKER
'MODE'	TOKEN_MODE
'MOD'	TOKEN_MODULE
'MODULE'	TOKEN_MODULE

(a)



(b)

Fig. 3. *Keyword scanning in the user interface module. (a) Example input to the keyword table generator. (b) Keyword parsing state transition diagram.*

gram which is given a list of tokens that change the active function, the keystrokes allowed for each active function, and the commands to send to the RFHI.

Display Control. The display control firmware is responsible for maintaining and updating the state of the front-panel display. The PSG front panel has its own 6803 microprocessor, which exchanges data with the main 68000 serially at 9600 baud. Since this serial communication is relatively slow, the display control module maintains a copy of what is currently being shown on the display and sends update information only when it is required. The display control module operates on a list of display segments stored in ROM. Each entry in the list contains the position and length of the display segment, the name of the variable to show in that segment, and any fixed character strings to show in that display segment. When a request for display update comes into the display control module, it looks up the display segment information in its table and requests information from the ISO about the variable to be displayed.

RF Hardware Interface

The RFHI subsystem is split into four layers: command, control, driver, and library (Fig. 4).

Command Level. Command level modules take commands from other subsystems and begin their execution. The first function of a command module is to consolidate commands that perform the same function. The instrument frequency can be set by entering a value, incrementing it by a programmable step, or turning the knob on the front panel. Each of these actions causes a different command to come into the RFHI, but they all perform the same function. For each of these commands, the command level firmware calculates the frequency and calls a single set-frequency procedure. This procedure compares the frequency with the limits stored in the ISO. If the value is out of range, an error is signaled. If the value is acceptable, the procedure initiates execution of the function in lower-level modules. The final function of the command module is to update the ISO. When a setting is changed, the command level firmware stores the new setting in the ISO, then recalculates limits on other parameters that may be affected by this new value, storing the new limits in the ISO.

There are ten to twelve command level modules, depending on the instrument. One of the command modules, RFHI_COMMAND, acts as the interface to the RFHI from the other subsystems. The remaining command level modules are divided by functional grouping, with all frequency related commands in one module, all amplitude commands in another, and so on. RFHI_COMMAND dispatches all commands processed by the RFHI to the appropriate functional command module. Command modules know how to interpret commands, but do not contain information about how the command is executed. They derive most of what they know about limits by asking the lower-level routines. If a command is not allowed in one of the instruments, the procedures to execute it are omitted and replaced with code to generate an insufficient-capability error. Using these techniques, a single set of command level modules meets the needs of all three instruments.

Control Level. Control modules are called by the command modules to execute intermediate-level instrument functions, such as setting the frequency of the synthesis loop. Most of the knowledge of the instrument block diagram, but none of the hardware details, is contained in the control modules. Control modules know what each hardware module in the instrument must do to implement a command, but they do not know how to program the hardware to do it. Although the block diagram of each instrument is different, many sections of it are shared by all three instruments. Sections that are not shared are handled by compiler directives. Optional modules are handled by calling a driver level function to determine if the option is present.

Driver Level. Driver modules are the lowest level of the RFHI. These modules directly manipulate the RF hardware to implement very simple functions, such as setting the VCO pretune voltage for a specified VCO frequency. There is one and only one driver module for each hardware module. Strictly defining the driver module interface makes it possible to hide knowledge of specific hardware from higher levels, allowing the use of different hardware by merely substituting a new driver module. Knowledge of the hardware's control bit mapping and dynamic response is contained in the driver levels.

Driver modules provide a method of backwards compati-

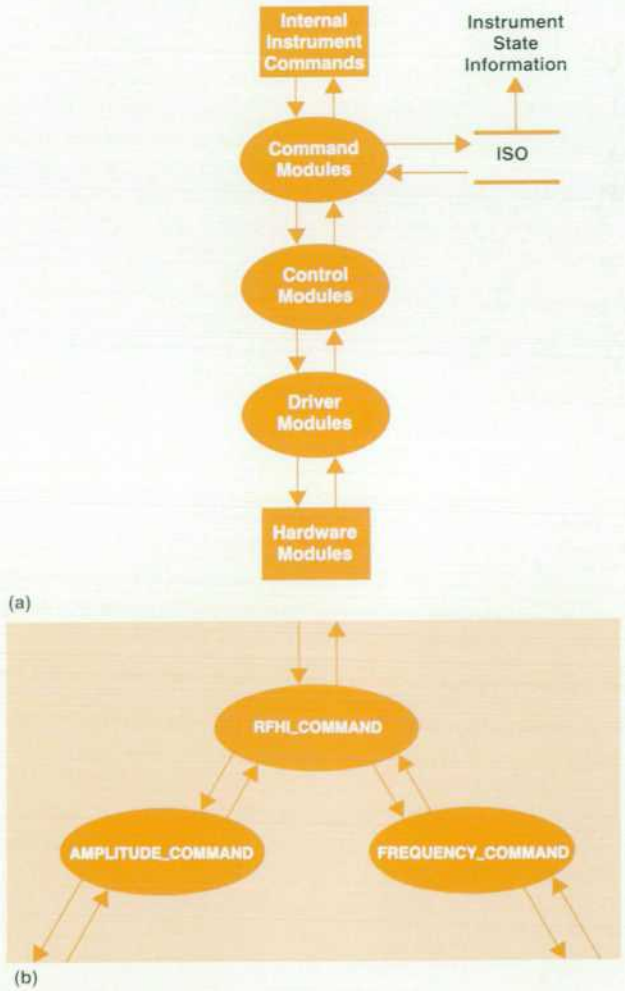


Fig. 4. *(a) Structure of the RF hardware interface module (RFHI) of the PSG control firmware. (b) Structure of a command module.*

bility for hardware modules. All hardware module updates that must be treated differently by the firmware provide a mechanism by which the firmware can determine which revision is installed. This is often an analog multiplexer point that is tied to a unique dc voltage. On initialization, the driver can read that point to determine which revision is installed. This saves time, since older modules do not have to be reworked. It also saves materials, since older modules do not have to be scrapped if modifications are not possible.

Some of the hardware modules—for example, the three instruments' VCO sections—share identical block diagrams but are implemented with entirely different hardware in each instrument. In this case, three separate VCO driver modules exist in the firmware. All VCO modules have identical external procedure definitions to keep the interface to the control level consistent.

Library Modules. Library modules are used to perform common RFHI functions. The ISO and I/O board driver are considered to be library modules.

The ISO (instrument state object) is a large data structure that provides data encapsulation, restricted access, reference by name, and knowledge of what data type each ISO variable represents, as well as the unit, value, and limits of each variable. ISO functions provide access to each variable's type, unit, value, and limits, requiring only a name in the form of an enumerated type to identify the variable of interest. The ISO as a whole stores the entire state of the instrument, providing a convenient store/recall structure. Access is restricted to the ISO simply by not importing the ISO module into the modules where access is denied.

Modularity of the driver level firmware was enhanced by making the serial interface to each hardware module identical. With this scheme, a single set of I/O functions could be written to communicate with all hardware modules. The routines require only a port number, the length of the control word, and a pointer to the control bit structure. The I/O board driver exports constants associating each hardware module with a port number, further isolating the driver level firmware from the module drivers.

Calibration and Diagnostic Firmware

Instruments in the PSG family contain a large amount of firmware to perform calibration and diagnostic operations. The calibration firmware adjusts the instrument hardware so that it will meet the operating specifications. The diagnostic firmware is used to detect a hardware failure and indicate to a high degree of certainty which hardware module should be replaced.

Calibration

The PSG instrument hardware is designed to allow firmware control of the adjustments required to ensure proper operation. As a result, the HP 8644A and 8645A contain no manual calibration adjustments and the HP 8665A requires only four or five.

Calibrating the instrument under firmware control means that the production line does not require technicians and test equipment to calibrate the instrument, resulting in lower production costs and assembly time. The customer can recalibrate the instrument to compensate for aging components or extreme operating conditions without sending the unit back to the factory or requiring specialized test equipment. If a hardware module fails, the customer can swap in a new module and recalibrate the instrument to ensure that it will still meet the operating specifications.

There are three types of calibrations: power-up calibration, run-time calibration, and external calibration.

Power-up Calibration. Power-up calibrations are performed when power is applied to the instrument for the first time or when requested by the user. These calibrations require several minutes to complete, and the calibration factors are stored in RAM. Because the RAM is powered by a rechargeable battery, the stored calibration factors are valid each time the instrument is turned on.

A good example of a power-up calibration is the VCO pretune calibration. The VCO pretune voltage is controlled by a DAC (digital-to-analog converter) that must be programmed to provide a frequency within the capture range of the phase-locked loop (PLL). The PLL frequency is determined by the setting of the fractional-N divider circuit (N.F). The calibration consists of finding the pretune DAC setting that results in the minimum PLL error voltage for a given frequency. The N.F divider is programmed for the desired frequency and the pretune DAC is swept until a zero crossing is detected in the PLL error voltage. The two pretune DAC settings that bracket the zero crossing are examined to see which setting is closest to zero. This setting is the pretune DAC calibration factor for the desired frequency.

Run-Time Calibration. There are times when it is not possible to have enough calibration factors to cover all operating conditions. In this case a calibration operation must be performed during normal instrument operation to determine the required value.

For example, without additional calibration, the power-up VCO discriminator gain calibration in the HP 8645A would not meet the FM accuracy specification for frequencies between the calibration points. When FM is turned on and the VCO discriminator is being used, a run-time cali-

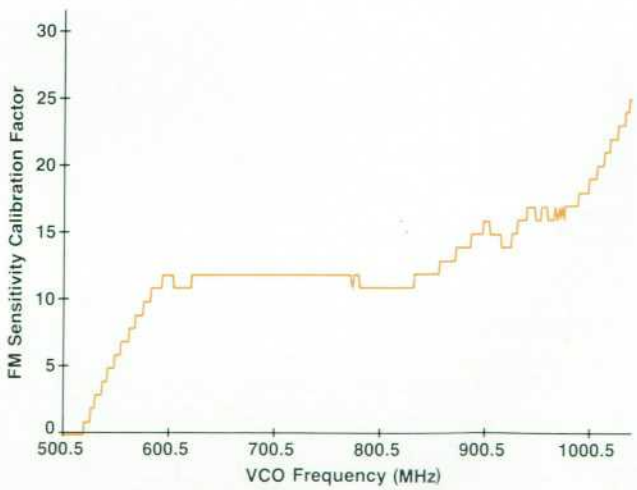


Fig. 5. *Typical HP 8645A VCO FM sensitivity calibration factors.*

bration must be performed whenever the frequency is set.

Run-time calibrations are optimized to reduce their execution time.

External Calibration. There are several calibrations that require external test equipment to measure power levels or timing intervals. These calibrations are performed in the factory and the calibration data is stored in EEPROM located on the hardware module that uses the data. The user only needs to perform these calibrations when making component level repairs on a module containing external calibration data.

The external calibration firmware is designed so that it requires no knowledge of the external test equipment. An external HP-IB controller commands the instrument to start a specific external calibration. The instrument performs some initialization operations, sets itself to the state corresponding to the first calibration point, and then waits for the controller to relay a measurement from the test equipment. If the test equipment needs to know the current state of the instrument, the controller can query the instrument using HP-IB commands. When the instrument receives the measurement from the controller, it uses the measurement to generate calibration factors and configure itself for the next calibration point. These steps are performed in a loop until the instrument has generated calibration factors for all of the calibration points. The controller determines from the instrument state that there are no more calibration points and sends a command to the instrument to store the calibration factors in EEPROM.

Calibration of some HP 8665A hardware modules requires the service technician to adjust components to achieve certain results on the front-panel display. These calibrations are also classified as external calibrations since they store the calibration factors in EEPROM.

Calibration Errors. When a calibration step fails the firmware usually supplies a default calibration factor and proceeds to the next calibration step. Some failures are so severe that the rest of the calibration procedure is terminated.

After the execution of an instrument calibration, the results of the calibration are displayed on the front panel as a numeric result code. This code number identifies the module that was being calibrated, the hardware configuration of the instrument, and a number indicating the calibration step that failed. The instrument diagnostics must be executed to determine which hardware module caused the calibration failure.

Accessing Calibration Data. A special mechanism allows an external HP-IB controller to read any arbitrary calibration factor. A program was written to display the calibration data graphically. The ability to inspect the calibration data visually was particularly useful in development when the RF engineers were turning on new revisions of their hardware modules. An engineer could usually tell at a glance whether the module was working correctly. The PSG production line developed a version of the same program that lets them examine the calibration data as soon as an instrument is assembled. The production line technicians can then compare the calibration data graphs to the graphs produced by a known-good instrument.

Fig. 5 shows an example of a calibration data graph for the HP 8645A VCO FM sensitivity calibration.

Diagnostics

The service strategy for PSG instruments is centered around a concept called module swap. If a hardware failure is suspected, the diagnostic firmware is executed to verify the failure and identify the faulty module. The customer will replace the indicated module with a spare and send the faulty module to the factory for repair. The repaired instrument is recalibrated to compensate for the new hardware and is returned to the customer's application. To make module swap a success, the goal of the diagnostic firmware was to detect greater than 90% of the hardware failures and then correctly identify the faulty module greater than 99% of the time.

Diagnostic Circuitry. Most PSG hardware modules contain an analog multiplexer (MUX) that selects various internal analog signals and routes them to a voltmeter circuit on the digital controller board. These MUX points are located at the inputs and outputs of the module and along various

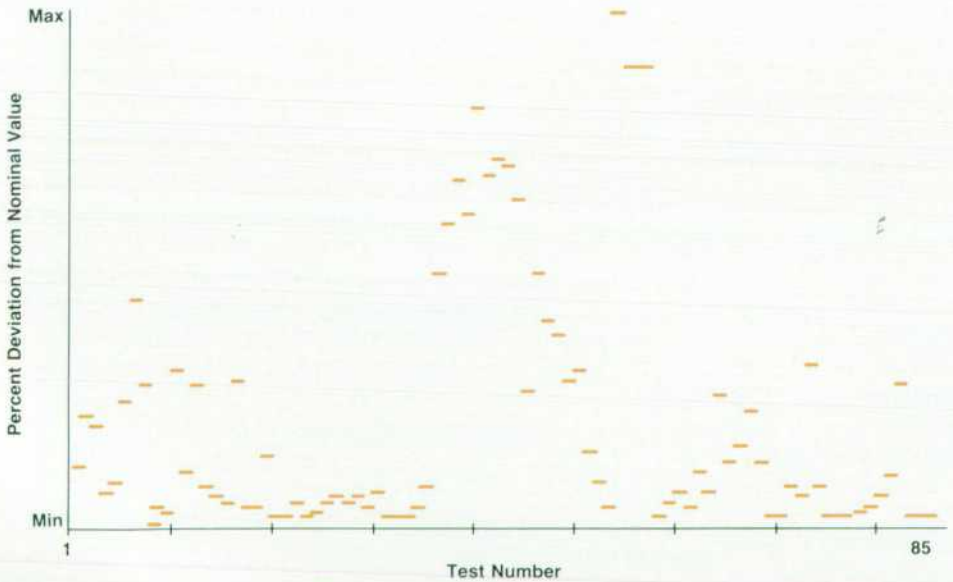


Fig. 6. *Typical test margins for the HP 8645A VCO module diagnostic tests.*

internal signal paths. Various hardware modules can be operated in special modes to apply calibrated signals to other modules.

Order of Module Testing. The hardware modules must be tested in a certain order. Modules that rely on the proper operation of other modules must be tested after the other modules have been tested. The result is a bottoms-up order of module testing.

Interactive Tests. If a test measurement indicates that a module input is not within the expected range, there are three possible explanations: (1) the output of the module supplying the signal is bad, (2) the cable connecting the modules is bad, or (3) the electrical path between the module input and the MUX point is bad. To isolate the error, there is a mechanism to test the connecting cable. This mechanism is called an interactive test because it requires interaction with the user.

When the diagnostic firmware determines that an interactive test must be performed, the front panel indicates that the user must connect the output of the supplying module to a specified connector on the controller board through a known-good RF cable and instruct the instrument to continue the test. If the output of the supplying module is good, the user must connect the output of the cable being tested to the controller board connector. If this measurement is bad, the cable is faulty and must be replaced. If this measurement is good then the input of the module being tested is bad.

Some tests of a module's output require an interactive test, since a short in the output cable or the input of the next module could result in a failure of the output signal test.

Test Margins. Most of the test limits used in the diagnostics were determined mathematically by analyzing the design of the hardware. A method was required for validating these limits.

The vast majority of the diagnostic measurements use a call to the same voltmeter function. Special code was inserted into the function to calculate the percentage deviation from the expected value. The deviation, called a test margin, is then written to a memory address that is a function of the module being tested and the test being performed. This address is in an unused portion of the microprocessor address space. The microprocessor emulator used to develop the firmware can trace these memory writes and record the test margins. A logic analyzer can also be connected to the microprocessor to record the test margins. An example of a test margin record for the HP 8645A VCO module diagnostic tests is shown in Fig. 6.

Diagnostic Errors. Diagnostic errors are reported using the same numeric result code format as calibration errors. Diagnostic errors also contain additional information. If the result code is a positive number, the diagnostics have determined that the indicated hardware module has failed and should be replaced. If the result code is a negative number, the user must look up the result code in the diagnostic documentation and perform the specified action, which may be an interactive test or a test that cannot be performed by the instrument firmware.

The diagnostics are designed to detect single hardware failures because the probability of multiple failures in the field is relatively low and interactions between failures could make isolation of the failures virtually impossible. Although the diagnostics only report the first failure, the remaining diagnostic tests are usually performed so that their test margins can be monitored using a microprocessor emulator or a logic analyzer. A variation of the instrument diagnostic command aborts all remaining tests when a failure is detected, leaving the instrument hardware and software in the same state as for the failed test. This diagnostic command allows a service technician to use some low-level commands to attempt to service the hardware module to the component level.

The diagnostic firmware is designed to spot catastrophic failures. Failures that cause subtle errors in the operating parameters (such as increased phase noise) may not be detected.

Conclusions

The calibration and diagnostic firmware have been used extensively on the PSG production line. The test equipment used to verify the instrument operation is very expensive and the verification tests are lengthy. Therefore, the ability to diagnose problems before the instruments reach the test station results in a great savings of time and money.

The PSG production line assembles the hardware modules, installs them in a chassis, and executes the instrument diagnostics. Modules that fail are repaired and the test is performed again until there are no failures. Once all the catastrophic errors have been found, the instrument is sent to the module pretest station where the RF performance of the individual hardware modules is verified. When the instrument comes back from module pretest the instrument diagnostics are executed to verify that all cables that were disconnected were reinstalled correctly. The instrument is calibrated and sent to the final test station, where the instrument operating specifications are verified. Using the diagnostic and calibration firmware has resulted in a drastic decrease in the number of faulty modules detected by the test stations.

Adding the calibration and diagnostic capabilities to the PSG instruments was costly in terms of extra hardware components and development time, but the reduction of production expenses and the added benefits for the customer have made these features well worth the cost.

Acknowledgments

The authors would like to thank Kevin Kelley and Troy Beukema, who contributed to the structure and development of the PSG control firmware. Gary Sprader and Brooks Holland were the service engineers who developed most of the diagnostic tests. We would like to give special thanks to the PSG production line technicians, production engineers, and technical writers for their efforts to test and document the calibration and diagnostic firmware.

RF Signal Generator Single-Loop Frequency Synthesis, Phase Noise Reduction, and Frequency Modulation

This signal generator design uses only a single phase-locked loop for frequency synthesis and one or more frequency-locked loops for phase noise reduction. The frequency-locked loops are based on delay line discriminators. Frequency modulation is introduced into all loops.

by Brad E. Andersen and Earl C. Herleikson

THE THREE PERFORMANCE SIGNAL GENERATORS (PSG) share a common synthesis block diagram, as discussed in the article on page 6. This method of synthesis is quite different from many signal generators designed for low noise. For example, the HP 8662A uses seven phase-locked loops (PLLs) and seven voltage-controlled oscillators (VCOs) for frequency synthesis, and the HP 8642A uses six PLLs and 12 VCOs. In contrast, the HP 8644A, 8645A, and 8665A PSGs use one PLL, one VCO, and up to two frequency-locked loops (FLLs). The simplicity of this new design results in lower parts count, higher reliability, and no spurious mixing outputs. The FLLs are added when lower phase noise is desired and each instrument uses the FLLs whenever possible.

In this article, operation of each loop and its effect on noise performance will be discussed. The frequency modulation scheme will also be explained, including loop crossovers and the various operating modes.

Frequency Synthesis

The voltage-controlled oscillator used in the HP 8644A and 8645A covers the 515-to-1030-MHz octave, while the HP 8665A VCO covers the 3-to-6-GHz octave. A PLL locks the VCO to a reference frequency with high accuracy and fine resolution. A pretune DAC is used to tune the VCO over this octave with enough resolution for the PLL to acquire lock easily. Without the PLL, the frequency accuracy is on the order of several hundred kilohertz. This is generally inadequate for most applications. The phase noise without a PLL is just that of the free-running VCO. Fig. 1 illustrates a free running VCO with pretune circuitry for tuning, and shows an idealized phase-noise plot.

The phase-locked loop is placed around the VCO to lock its frequency to a stable reference frequency as shown in Fig. 2. A fractional-N technique (see box, page 28) is used to achieve a frequency resolution of 0.01 Hz. This particular fractional-N PLL operates from 257.5 to 515 MHz (each signal generator's VCO output frequency is divided down to meet this requirement) and achieves excellent spurious performance, typically lower than −60 dBc.

An operating characteristic of a PLL is that the phase noise of the loop's RF output will either be that of the reference input or the noise of the PLL circuitry. In this case the noise rolls down with the reference and then with the 1/f characteristic of the fractional-N PLL until it gets to the noise floor of the PLL at −85 to −90 dBc/Hz. The noise then stays at this level throughout the bandwidth of the PLL, even beyond where the raw VCO noise is less than the PLL's. To optimize the noise performance, the PLL bandwidth is chosen to be equal to the frequency at which the raw VCO noise intersects the noise floor of the PLL. This turns out to be about 3 kHz. The resulting phase noise plot follows the reference to the PLL floor, the PLL floor to the PLL bandwidth, and then the VCO noise itself.

Phase Noise Reduction

To reduce the VCO phase noise, one or more frequency-locked loops can be placed around the VCO. A user-selectable FLL based on a 70-ns delay line discriminator is standard in the HP 8644A and 8645A. The design of this dis-

(continued on page 29)

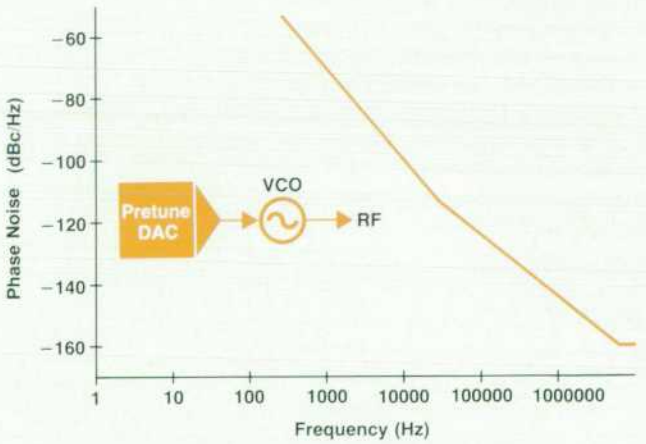


Fig. 1. *A voltage-controlled oscillator (VCO) and an asymptotic phase noise plot based on typical measured data for the HP 8644A/45A VCO.*

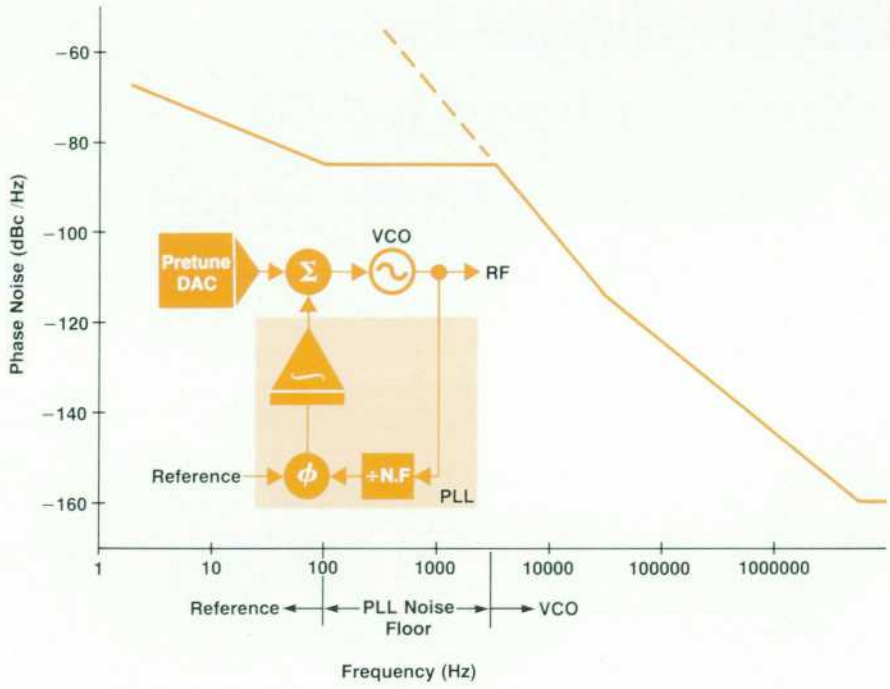


Fig. 2. *Asymptotic phase noise for the HP 8644A/45A VCO stabilized by phase-locking it to the reference with the fractional-N loop.*

Fractional-N Synthesis Module

Fractional-N frequency synthesis provides vastly improved phase noise performance over any other single-loop synthesis technique with similar frequency resolution.[1-3] In a fractional-N phase-locked loop, the output frequency is N.F times the input frequency, where N is an integer and F is some fraction. The fractional division (÷N.F) is implemented by switching between two integer divisors with varying duty cycles, depending on the fraction desired.

Since the divider output in the fractional-N loop is not an integer multiple of the reference frequency, the loop phase detector output contains a phase ramp with a period proportional to the fractional frequency offset. This ramping phase, if uncorrected, would result in very large phase modulation sidebands on the output frequency. To correct these spurious outputs, a digital phase accumulator in the fractional-N control chip (a special integrated circuit) feeds a multiplying digital-to-analog converter (DAC), which provides a current ramp to the loop integrator to cancel the effect of ramping phase at the phase detector output. This is called automatic phase interpolation (API).

Phase correction is limited to 360°. When one complete cycle of extra phase is accumulated, the fractional-N control chip instructs the divider to "swallow" one pulse of the VCO output. This gives the signal applied to the phase detector an average frequency equal to the reference frequency.

PSG Fractional-N Module

The fractional-N synthesis module used in the three Performance Signal Generators (PSGs) described in this issue makes three contributions to the state of the art:

- Pulse swallowing at the 257.5-to-515-MHz octave. This is a higher frequency than previous loops and results in much lower fractional-frequency spurious content.
- Self-calibration of fractional spurs to reduce their level.
- Digitally prescaled, digitized frequency modulation capability for accurate dc-coupled FM with low distortion.

The key to the higher-frequency operation is a dual-modulus ECL divider capable of swallowing cycles at the 257.5-to-515-MHz octave.

Fractional spurs are further reduced by self-calibration. This is implemented by spur-detection and spur-nulling circuits. The spur-detector circuit coherently detects the API spur residue (the phase ramps do not cancel perfectly) on the tune line of the fractional-N module. The coherent detection measures both the magnitude and the sign of the spur residue. The spur-nulling circuit causes a digitally controlled gain adjustment of the API DAC that attempts to reduce the spur-detector error signal to zero.

Digitized FM

Highly accurate dcFM is implemented by passing the FM signal through an analog-to-digital converter (ADC) and using the result to control the programmed frequency of the synthesizer. The FM is scaled for different deviations after it passes through the ADC. In this way, the FM signal uses the full range of the ADC, ensuring maximum accuracy and linearity.

Acknowledgments

Many thanks to Alan Hedge, who developed these concepts into a working module.

References

1. D.D. Danielson and S.E. Froseth, "A Synthesized Signal Source with Function Generator Capabilities," *Hewlett-Packard Journal*, Vol. 30, no. 1, January 1979, pp. 18-26.
2. M.B. Aken and W.M. Spaulding, "Development of a Two-Channel Frequency Synthesizer," *Hewlett-Packard Journal*, Vol. 36, no. 8, August 1985, pp. 11-18.
3. T.R. Faulkner, et al, "Signal Generator Frequency Synthesizer Design," *Hewlett-Packard Journal*, Vol. 36, no. 12, December 1985, pp. 24-31.

Barton L. McJunkin
Development Engineer
Spokane Division

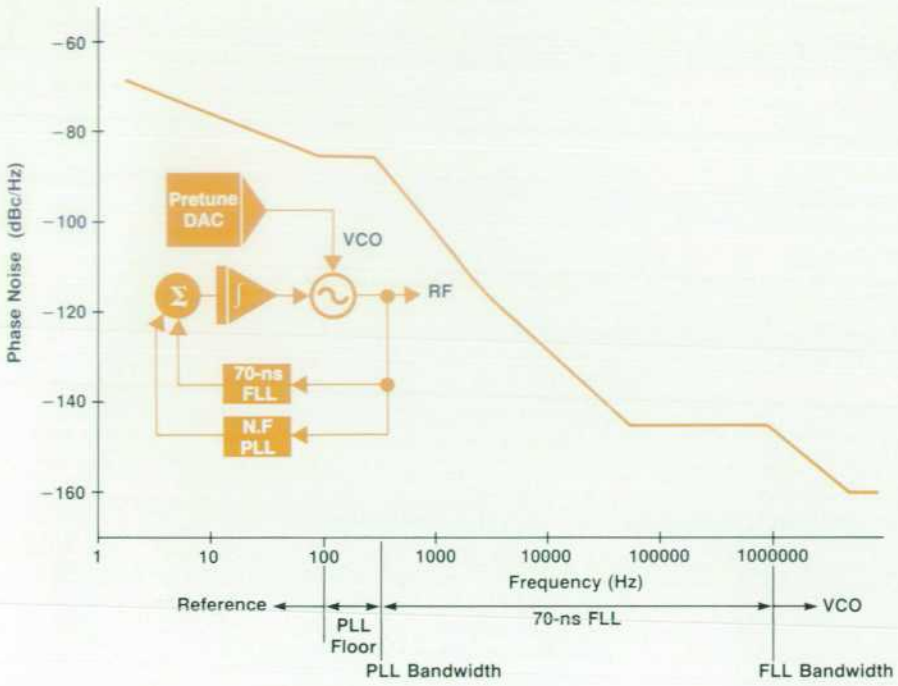


Fig. 3. *Reduced phase noise resulting from the addition of a frequency-locked loop based on a delay line discriminator.*

criminator is described in the article on page 34.

Fig. 3 shows a block diagram and a phase noise plot of this circuit arrangement. When this loop is selected, the PLL sees a lower-noise VCO, so the PLL bandwidth is reduced to coincide with the new crossing of the PLL noise floor and the cleaner VCO noise. The new PLL bandwidth is about 350 Hz. The FLL bandwidth determines the noise floor for offsets below the FLL bandwidth frequency. Extending the bandwidth further to get a lower floor can produce peaking of the phase noise beyond the bandwidth point because of phase and gain margin problems. The FLL bandwidth of about 1 MHz provides an adequate noise floor and avoids peaking of the noise beyond the bandwidth frequency. See page 30 for more information on the use of delay line discriminators for noise reduction.

The noise characteristic of the FLL discriminator can be shown asymptotically by calculating the noise floor and the delay line corner frequency (see Fig. 4). The noise goes up with a 20-dB/decade slope for frequencies less than this corner frequency and then up with a 30-dB/decade slope when the 1/f corner of the discriminator (usually the phase detector) is reached. The FLL discriminator noise floor is lower than the VCO's, so the resulting phase noise beyond the FLL bandwidth is that of the VCO.

The overall phase noise graph follows the FLL noise down at 30 dB/decade until the 1/f point, where the slope becomes 20 dB/decade. When the phase noise reaches the VCO noise level at the FLL bandwidth, the slope becomes zero. This becomes the noise pedestal of the FLL-stabilized VCO. The phase noise plot remains flat out to the FLL bandwidth, and then follows the VCO phase noise characteristic for higher frequency offsets from the carrier.

Additional FLLs with longer delays can be added to decrease the noise of the VCO even more. The noise keeps getting lower between the PLL bandwidth and the FLL bandwidth frequencies as the FLLs are nested one on top of another. For each additional FLL added, the PLL bandwidth is reduced to correspond to the cleaner VCO noise characteristic. One additional FLL based on a 140-ns delay line discriminator is currently available as an option for the HP 8644A and 8645A. It is also optional in the HP 8665A. The design of this discriminator is presented later in this article.

Frequency Modulation

Fig. 5 shows a block diagram of the synthesis loops with the PLL, two FLLs, and the FM connections. FM is summed into each loop in addition to being connected to the VCO's FM input.

Within the bandwidth of each loop that is connected around the VCO, the FM deviation is controlled by the loop with the lowest bandwidth. For example, for FM rates below the PLL bandwidth, the PLL controls the deviation.

(continued on page 31)

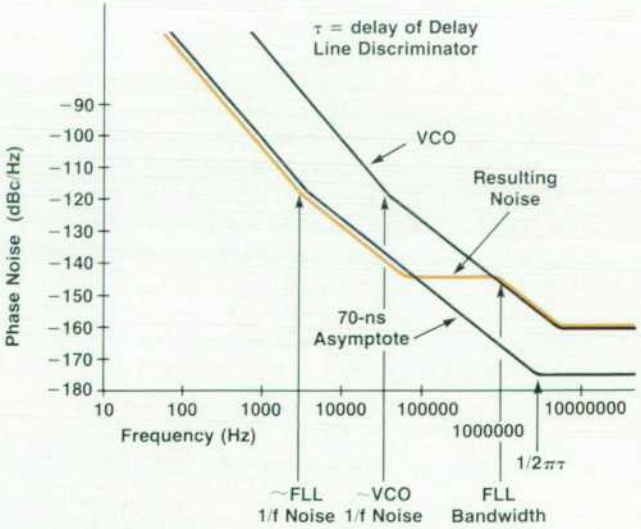


Fig. 4. *Contributions of the phase-locked VCO and the frequency-locked loop to the phase noise of Fig. 3.*

Delay Line Discriminators and Frequency-Locked Loops

Delay line discriminators are a well-known type of FM discriminator (frequency-to-voltage converter) characterized by high sensitivity, low noise, and wide bandwidth. This type of discriminator consists of an RF power amplifier, a power splitter, a delay line, a phase shifter, and a phase detector, as shown in Fig. 1. The purpose of the power amplifier is to provide adequate signal levels at the phase detector. It compensates for all of the loss of all the passive components between the power amplifier and the phase detector.

The power splitter produces two signal paths. By delaying one signal path and not the other, a phase shift proportional to input frequency is produced and is detected by comparing the two paths at a phase detector. Most RF phase detectors are mixers that have an output voltage proportional to the cosine of the phase difference between their inputs. To make it possible to operate the discriminator at any carrier frequency within a band of operation, the phase shifter provides a phase offset such that the cosine of the phase difference is near zero at the frequency of interest. It does not matter which path the phase shifter is placed in, but it is typically placed in the nondelayed path to minimize the loss in the delayed path. The phase shifter changes the phase detector response to a sine function of the form

$$V_{out} = V_{peak} \sin((\omega - \omega_o)\tau)$$

where V_{out} is the output voltage of the phase detector, V_{peak} is the peak value of V_{out}, ω angular frequency, and τ is the length of the delay. In practice, discriminator operation is limited to a small range of frequencies around ω_o where the response is relatively linear.

For *low-rate FM*, computing the sensitivity of the delay line discriminator is quite simple. Group delay of any device is simply the derivative of relative phase with respect to radian frequency, where relative phase is the phase at the output relative to the input. Since the discriminator hardware measures the relative phase with a phase detector operating in its linear range, the voltage at the output is $K_0\tau$ times the radian frequency at the input, where K_0 is the phase detector constant (volts/radian) and τ is the length of the delay (seconds). The frequency-to-voltage sensitivity is therefore $2\pi K_0\tau$. The sensitivity of the discriminator can be increased by increasing either K_0 or τ. K_0 is typically a function of the RF power at the inputs of the phase detector and of the efficiency or conversion loss of the phase detector. Increasing delay length will decrease RF power at the phase detector. This forces a trade-off between K_0 and τ, which can be overcome with a lower conversion loss in the phase detector, higher power from the power amplifier, or a lower-loss delay line. Higher power increases heat dissipation and cost. Lower-loss delay lines increase cost and increase weight because they use larger-diameter coaxial components. Therefore, a low-conversion-loss phase detector is extremely important.

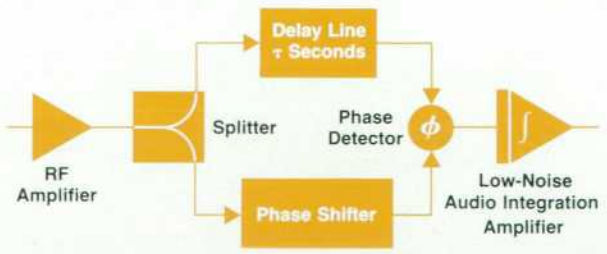


Fig. 1. *Delay line discriminator.*

The transfer function from frequency modulated RF at the input to demodulated audio at the output is almost constant up to FM rates of $0.1/\tau$ Hz, assuming a sensitivity of $2\pi K_0\tau$ as described above. The delay from input to output is constant for all FM rates and is simply $\tau/2$.

Since the phase detector cannot detect the difference between a nondelayed FM signal and an FM signal delayed by 360 degrees, the transfer function drops to zero at an FM rate of $1/\tau$. This *high-rate FM* response of the transfer function is ideally described by

$$2\pi K_0\tau \sin(\pi f_m\tau)/(\pi f_m\tau),$$

where f_m is the FM modulation rate (Hz). The multiplier, $\sin(\pi f_m\tau)/(\pi f_m\tau)$, is equal to one when f_m is small compared to $1/\tau$, and is equal to zero when $f_m = 1/\tau$.

Sine wave phase modulation is related to frequency modulation by the equation

$$\theta_{pk} = \Delta f_{pk}/f_m,$$

where θ_{pk} is the peak phase deviation, Δf_{pk} is the peak frequency deviation, and f_m is the modulation frequency. Therefore, the sensitivity to phase modulation will decrease with decreasing modulation rates. The transfer function from phase modulation at the input to voltage at the output is:

$$2K_0 \sin(\pi f_m\tau).$$

The low-noise characteristic of the delay line discriminator is achieved by careful consideration of all noise mechanisms. Since the phase of the delayed path relative to the nondelayed path is proportional to frequency at the input, thermal phase noise floors or additive phase noise anywhere after the power splitter will appear as FM noise at the input. Because there are no active devices in either path after the power splitter, additive noise is minimal and thermal noise is the limiting factor. Therefore, increasing the signal level with the power amplifier will increase the signal-to-noise ratio (SNR) at the phase detector inputs. The conversion loss of the phase detector then lowers the signal level and decreases the SNR. Further degradation of the SNR occurs with the additive noise of diodes in the phase detector and the equivalent input noise of the audio amplifier at the output of the phase detector. The additive noise from the mixer diodes is usually flicker (1/f) noise and the dominant source of noise for rates less than 1 kHz.

Frequency-Locked Loops

The low noise, high sensitivity, and wide bandwidth of the delay line discriminator, which make it useful for measuring low-noise oscillators, also make it useful for lowering the noise of an oscillator. Measuring the noise of the oscillator with the discriminator, amplifying it, and feeding it back to the FM port of the oscillator, as shown in Fig. 2, will cancel the noise at the output of the oscillator. The amount of cancellation is dependent on the gain of the feedback loop and limited by the FM noise floor of the discriminator. This type of loop is called a frequency-locked loop (FLL) because it is the frequency of the oscillator that is detected and fed back to the frequency tuning port.

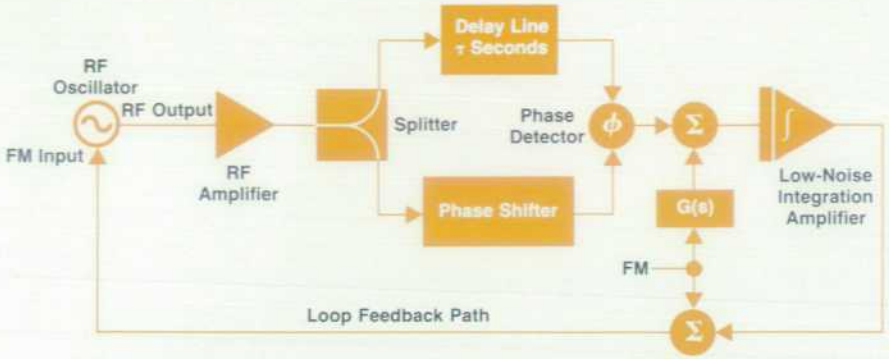


Fig. 2. *Frequency-locked loop based on a delay line discriminator.*

A first-order frequency-locked loop cannot be achieved without adding an integration amplifier between the discriminator output and the oscillator input. Integration is necessary because the gain from voltage input at the oscillator to voltage output at the discriminator is not a function of the FM rate. The oscillator converts voltage to frequency with a constant of K_v Hz/V and the discriminator converts frequency to voltage with a sensitivity of $2\pi K_0\tau$ V/Hz that is constant for FM rates less than $0.1/\tau$.

An integration amplifier with a gain of $G/(f_p + jf_m)$ has a gain of G/f_p when f_m is zero, and the gain decreases with a single-pole roll-off for f_m greater than f_p. With such an amplifier, the total loop gain becomes

$$K_v(2\pi K_0\tau)G/(f_p + jf_m).$$

The loop bandwith is defined by the frequency at which the magnitude of the loop gain equals unity. In this case the loop bandwidth is

$$K_v(2\pi K_0\tau)G$$

for $f_m >> f_p$.

Any FM applied at the VCO will be attenuated by the FLL loop gain. Therefore, FM must be applied before the integration amplifier. A voltage summed at the phase detector output will be integrated and applied to the VCO, tuning the frequency until the phase detector output voltage cancels it. Frequency modulation by means of a summation voltage at the phase detector output forces the phase detector operating voltage away from zero. FM sensitivity at this point is simply the inverse of the discriminator gain or $1/(2\pi K_0\tau)$ Hz/V. Therefore, FM accuracy is determined by the phase detector gain constant and the delay length.

Earl C. Herleikson
Development Engineer
Spokane Division

As the rate increases, the FLL will take over control of the deviation if it is connected.

FM on the VCO output is primarily caused by the FM signal that is fed to the VCO directly at point A in Fig. 5. However, each loop must add in the FM signal to cancel the FM that is detected by that loop. Any difference between the VCO FM and what any particular loop determines the FM should be results in a small correction signal to the VCO from the loop. For example, the signals at points B and C sum together (actually the difference is found) and if there is a difference in amplitude and/or phase, a correction signal is produced. The correction signals tend to modify the resulting deviation to satisfy each loop's determination of what the deviation ought to be. If each loop is calibrated perfectly and in agreement in amplitude and phase with the FM calibration of the VCO's FM input, then the error signals will all be zero. If the calibrations are not exactly the same, then the FM deviation response will not be flat as the FM rate is swept through the different loops' control bandwidths. This is well illustrated by Fig. 6, which shows the areas of control by the various loops and the levels that must be calibrated. Phase differences affecting FM flatness will become noticeable near the loop crossover frequencies when the adjacent loops contend for control of the FM.

It should also be mentioned that the FM summing point within each loop can contribute to the noise of that loop. As the FM deviation is increased, the noise in the loop tends to increase once the deviation exceeds the threshold defined by the residual noise of the FM circuits in that particular loop.

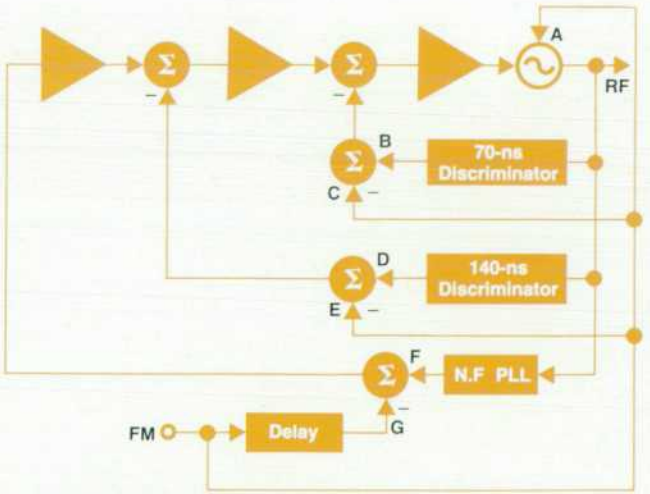


Fig. 5. *Frequency modulation is summed into each loop in addition to being connected to the VCO's FM input.*

Digitized FM Effects

The FM signal is digitized in the fractional-N PLL circuit and summed digitally. All this takes time (approximately 30 μs), which appears as an increasing phase shift of the FM signal with increasing frequency. This delay is shown as the delay block in the FM path of the PLL in Fig. 5. If the FM path to the VCO (and other FLLs) does not experience this delay, a large ripple in FM flatness can occur near the PLL bandwidth. It is especially bad with a higher PLL bandwidth because the correction signals from the PLL summing point (of F and G) and the FLL summing point (e.g., B and C, or D and E) will not be of the same phase, resulting in a lack of flatness near the crossover frequency. Since the PLL controls the final FM of the VCO's RF output within its bandwidth, this delay is present from the external FM input of the instrument any time digitized FM is used from dc to the PLL bandwidth.

One solution is to delay the FM signal to the VCO (and the FLLs) so that all FM signals experience the same delay as the PLL circuit. This helps the flatness problem but still presents a time delay for external FM applied to the instrument. The HP 8644A uses this approach by implementing an active all-pass network that has a flat 30-μs group delay and a 100-kHz bandwidth. The HP 8645A has a 10-MHz FM bandwidth and it is very difficult to provide the 30 μs of delay with that much bandwidth. Therefore, a passive all-pass network is used that provides the correct group delay throughout the PLL bandwidth and then becomes nonlinear well beyond that point. This does not affect the FM flatness beyond the PLL bandwidth, but does provide the necessary delay to flatten the crossover from the PLL to the VCO or FLL. For mode 3 operation in the HP 8645A (140-ns delay line discriminator), the bandwidth of the PLL is low enough that FM delay compensation is not used for reasonably flat FM. In this case, the external FM port experiences a delay within the PLL bandwidth and a negligible amount of delay beyond that point.

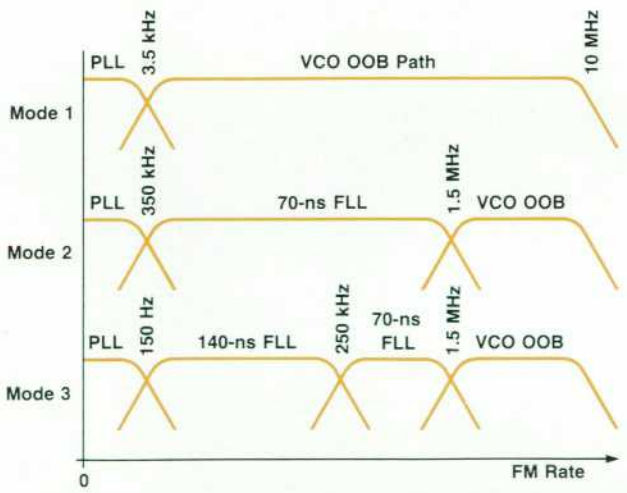


Fig. 6. *FM rates at which various loops dominate the FM response in various operating modes. VCO OOB (out of band) refers to VCO signals at frequencies beyond the phase-locked loop's bandwidth.*

Linear (Nondigitized) FM

Two types of linear FM are available to the user: linear dcFM and linear acFM. The linear dcFM mode of operation disconnects the PLL, resulting in either a raw VCO or a VCO cleaned up by an FLL. The resulting noise keeps climbing instead of flattening at what was the PLL bandwidth. However, the FM now has little time delay and phase shift, since the delay compensation is removed.

With digitized FM and its FM path delay, connecting the signal generator in a phase-locked loop situation could result in loop instability. Linear dcFM is a solution since there is sufficient center frequency accuracy in the FLL modes to allow use of this type of FM in phase-locked loop applications requiring low additional phase shift.

Linear acFM is an interesting mode. The PLL is used to control the dc frequency but does not have FM summed into its loop. This results in the PLL's rolling off the FM of the VCO within the PLL bandwidth, leading to the name acFM. With this mode of operation, the center frequency accuracy is very good but the FM rate must be higher than the PLL bandwidth.

140-ns Discriminator Design

The key to the low-noise capability of the 140-ns discriminator module is its new phase detector design. This phase detector has 5 to 8 dB better SNR than a double balanced mixer. The VCO to be cleaned up can be either a free-running oscillator as in the case of the HP 8665A's YIG oscillator, or an oscillator that already has a discriminator loop locked around it as in the case of the HP 8644A and 8645A. The only criteria are that the output of the VCO be in the frequency range of 500 to 1040 MHz and that the VCO have a broadband FM port with a bandwidth greater than 2 MHz.

The phase detector used in the 140-ns discriminator is made up of a 3-dB coupler and two low-noise AM detectors (Fig. 7). The coupler combines the RF signals (applied at the isolated ports) at the through and coupled ports. By amplitude-detecting the combined RF signals at these ports and subtracting, a voltage proportional to the relative phase

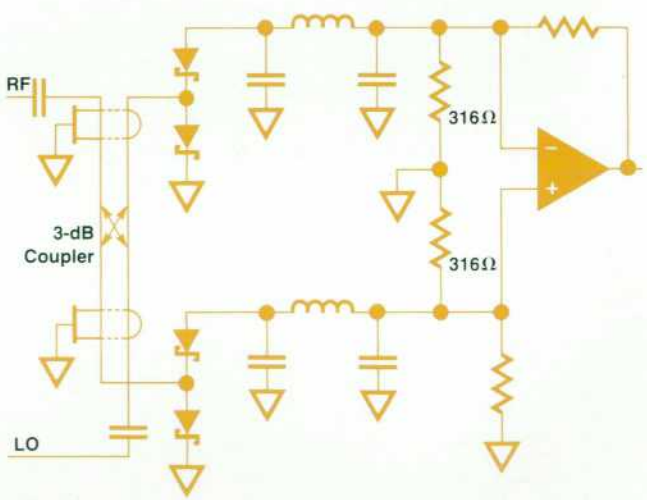


Fig. 7. *The phase detector is the key to the low-noise capability of the 140-ns delay line discriminator.*

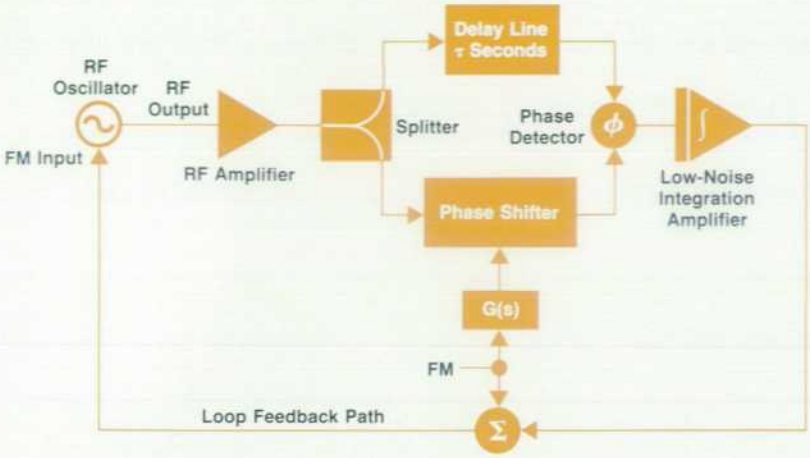


Fig. 8. *FM is applied at the phase shifter in the 140-ns frequency-locked loop.*

of the two RF ports can be obtained. Each amplitude detector provides a 50Ω termination at the coupled and through ports of the coupler for amplitudes large enough to switch the Schottky diodes. This 50Ω match is set by the dc bias point of the diodes, which is set by the 316Ω resistors to ground. The dc source impedance of the detector is simply 2Z/d, where Z is the RF impedance and d is the duty cycle, or the fraction of the period that each diode is on. It can be shown that the SNR is maximized when each diode is on for approximately one third of a period. This gives a 300Ω source impedance for each amplitude detector. Just as one would expect for optimum SNR, the RF is properly matched at 50Ω and the dc is properly matched at 300Ω. The total source impedance of the two amplitude detectors is 600Ω.

The advantages of this phase detector are a low conversion loss and a 600Ω source impedance. If one phase detector requires 5 dB more RF power than another to achieve the same signal-to-noise ratio, then 5 dB greater power will be required from the power amplifier to achieve the same discriminator noise floor. Since every additional dB of power from the power amplifier is increasingly expensive, a low-loss phase detector is very important.

For the same signal-to-noise ratio, the voltage from a 600Ω source would be 3.4 times the voltage from a 50Ω source impedance. The larger voltage output of the 600Ω phase detector means that the noise voltage of the integrator has an insignificant effect on the noise floor of the discriminator. The lower conversion loss and higher source impedance give a combined phase noise performance improvement over a standard double balanced mixer of about 7 dB.

Applying FM at the output of the phase detector has several disadvantages. Maximum phase detection and minimum amplitude detection occur when the detector output is near zero volts. FM applied after the phase detector will cause the detector to operate at nonzero voltages. Also, the FM sensitivity is a function of the phase slope, which is a function of the power levels at the phase detector. This means that as temperature changes, the phase slope will change, and FM sensitivity will change. In the 140-ns delay line discriminator, these problems are eliminated by applying the FM signal to the phase shifter instead of the phase detector, as shown in Fig. 8. The phase detector always operates with zero volts at the output and the FM sensitivity is independent of the phase slope of the phase detector.

Acknowledgments

The authors wish to acknowledge the other PSG engineers' work and contributions to the understanding of the operation and interactions of the various loops. Also, the production engineers and technicians made invaluable contributions in transferring the instruments to production.

Design Considerations in a Fast Hopping Voltage-Controlled Oscillator

The fast hopping requirement affected the design of the discriminator power amplifier, phase shifter, and delay line, the wideband feedback loop, and the VCO pretune circuit.

By Barton L. McJunkin and David M. Hoover

IN ITS FAST HOPPING MODE of operation, the HP 8645A Agile Signal Generator can switch to a new frequency in less than 15 microseconds with an accuracy of one part per million or better. This article describes the fast hopping VCO block diagram and the five major technical challenges that had to be met to build the fast hopping VCO.

System Architecture

The HP 8645A uses a 515-to-1030-MHz oscillator whose stability is controlled by a fractional-N phase-locked loop (PLL), by a frequency-locked loop (FLL) based on a delay line discriminator, or by both the PLL and the FLL. Fig. 1 is a block diagram of the system.

The frequency-to-voltage conversion characteristic of the discriminator is inverted by placing the discriminator in a feedback path around the VCO. The stabilized VCO now has the voltage-to-frequency characteristics of the inverted discriminator. For fast switching speed, the feedback loop must be wideband and high-gain without degrading the stability of the discriminator.

The ideal model for the FLL loop gain is a single integrator with a unity-gain crossover frequency of 1 MHz. Because of the relatively flat gain and excess phase shift of the discriminator, the gain margin of the loop required more attention than the phase margin.

As explained in the box on page 30, the frequency-to-voltage conversion characteristic of the delay line discriminator is a repetitive function of the form:

$$V_{out} = V_{peak}\sin((\omega - \omega_o)\tau),$$

where V_{peak} is the peak voltage from the phase detector, ω is the input angular frequency (radians/second), and τ is the time delay of the delay line (seconds). Although each input frequency results in only one output voltage, the inverse is not true. Each output voltage can indicate a multitude of frequencies. The inverted response of the discriminator-stabilized VCO is single-valued only over a small range of $(\omega - \omega_o)\tau < \pm 90$ degrees at the FLL phase detector. This means that the VCO must be pretuned to within $1/4\tau$ Hz of the final frequency before the discriminator will stabilize the VCO to the correct frequency. To ensure that no transient condition can result in the FLL's locking the VCO to the wrong frequency, the FLL is designed to have a pulling range less than $1/4\tau$ Hz. The pulling range is limited to this value by a diode limiter on the FLL VCO control line.

The standard signal generator operation of the instrument uses the fractional-N loop to phase-lock the VCO to the reference signal and the discriminator to reduce the phase noise of the oscillator.

Learn Sequence

Before the system enters the fast hopping mode, it must learn the correct values for the pretune DAC (digital-to-analog converter), phase shifter, FLL DAC, and PLL DAC. The learn sequence is as follows.

The fractional-N loop phase-locks the VCO to the desired frequency. The tune line offset is used to optimize the pretune value. The phase shifter is used to null the FLL phase detector within ±6 degrees of quadrature. The FLL DAC is used to cancel the phase detector offset at the input of the low-noise amplifier. The PLL DAC is used to replace the PLL tune line with an equivalent dc voltage. This procedure is followed for each frequency that is hopped to. The DAC values are then saved in internal memory to be called back when the instrument hops to that frequency.

In the fast hopping mode, the PLL is opened and the discriminator controls the frequency accuracy of the signal generator in addition to reducing the phase noise of the oscillator. Since the discriminator controls the frequency accuracy of the signal generator, it is necessary to optimize the stability of the discriminator for best performance. The three components of the discriminator that have the greatest effect on frequency stability are the power amplifier, the phase shifter, and the delay line. Besides the discriminator, two other components can cause considerable frequency error: the feedback loop and the VCO pretune circuit.

Power Amplifier

The power amplifier drives the input of the discriminator with a leveled RF signal. Automatic level control (ALC) maintains the power amplifier output at six volts peak into the backmatch resistor. This high power level is necessary for low-noise, high-sensitivity operation.

Absolute accuracy of the power amplifier ALC loop is not as important as switching speed and stability. The ALC loop has a minimum bandwith of 500 kHz and must settle to within 0.012% of the final level within 15 microseconds. This is the kind of stability that is necessary to keep the AM-to-ΦM conversion of the phase detector to acceptable levels. Since the AM-to-ΦM rejection of the phase detector

is only 10 dB worst-case, 0.012% AM can cause 0.1 ppm of frequency error in the FLL.

The extreme stability specification for the AM level requires a very accurate level detector. The detector is a peak-to-peak Schottky diode detector with an output level of −12V dc. Because of the high output level, temperature compensation of the Schottky diodes was not necessary.

The detector output is compared to a stable reference voltage and the error signal is integrated and applied to the modulator at the input to the power amplifier. The modulator is a limiter amplifier with a variable bias. The modulator has 15 dB of linear range with the output varying from 5 to 20 dBm.

The final ALC loop bandwidth varies from 500 kHz to 2 MHz across the 500-MHz-to-1040-MHz frequency band. Since the bias loop of the limiter is used to control the modulator, and since the power amplifier is near limiting, the bias loops for both the limiter amplifier and the power amplifier are wideband loops. To reduce the overall detector error signal, the power amplifier bias circuit uses an integrator in the bias loop for high dc gain.

Phase Shifter

The main design goal for the phase shifter was to produce no more than 0.2 ppm of error in the stabilized VCO from t = 5 microseconds to t = 10 minutes. This requires phase stability within <0.0012 degree. Other design goals were low loss, good input/output match, low phase noise, good phase stability with temperature, monotonicity of phase shift with frequency, fast switching (<5 μs for <0.0012° stability), and excellent long-term phase stability.

Two specifications drove the choice of implementation: phase noise and switching speed. A varactor phase shifter could not satisfy both conditions. With a switching speed specification of 5 microseconds, sufficient filtering for phase noise was not possible. The topology that was chosen is a transmission line structure with double-pole double-throw pin switches. Five binary-weighted phase shift elements have phase shifts of 90°, 45°, 22.5°, 11.25°, and 5.6° at 500 MHz. Because of the transmission line design, the minimum phase resolution at 1000 MHz is 11.25°.

The initial switch design used discrete pin diodes and simple driver circuits. A large phase drift occurred when the current was switched from one diode to another. The drift was caused by the heating of the pin diodes, which affected the RF resistance and forward bias voltage. The intial phase drift was approximately 0.5° from t = 5 μs to t = 10 s.

Dual monolithic pin diodes are now used in the DPDT switches. Since one or the other of the diodes is forward biased, the total power dissipation in the diode chip is constant. Also, the driver circuits were replaced with stable current sources. The constant bias improves the long-term stability. These changes reduced the phase drift from t = 5 μs to several minutes to something on the order of 0.02°.

The remaining phase drift was difficult to track down. It was finally traced to shifts in the dielectric constant of the printed circuit board material caused by localized heating. The heating was caused by both bias and RF power dissipation. The shifts in dielectric constant caused changes in the group delay of the transmission lines. Placing the design onto a Teflon board allowed the phase shifter to meet its short-term phase stability specification.

Even with Teflon as a dielectric material, a few degrees change in ambient temperature caused a long-term phase stability error. The phase shifter was split into two parts to improve the stability. The two most-significant bits were placed in the discriminator leg with the delay line, and the three least-significant bits were placed in the other leg. This improved the long-term stability by a factor of 20 to 50.

Delay Line

Two specifications affected the choice of the delay line length of 70 nanoseconds: phase noise pedestal and FM deviation. A 1-MHz loop bandwidth is necessary to maintain a typical phase noise pedestal of −145 dBc. At 1 MHz, the excess phase shift added to the open-loop gain of the discriminator by the delay line has to be small enough not to reduce the loop phase margin unacceptably. Also, a peak FM deviation of 1.76 MHz was required. To ensure linear

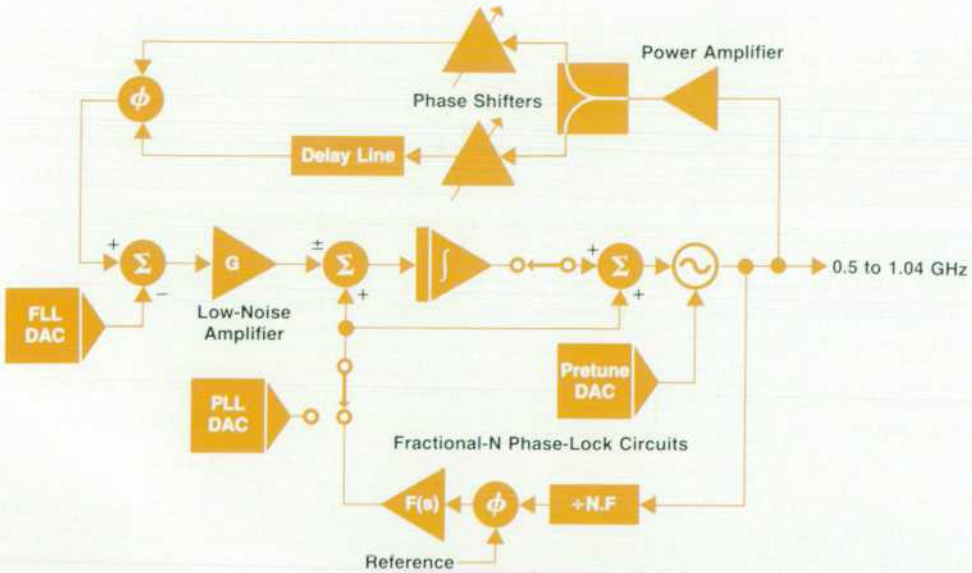


Fig. 1. *Block Diagram of the voltage-controlled osciilator used in the HP 8645A Agile Signal Generator, showing the fractional-N phase-locked loop and the frequency-locked loop based on a 70-nanosecond delay line discriminator.*

operation, this needs to be less than half the pulling range of the FLL. A 70-ns delay line meets both requirements. It adds only 25° of excess phase shift at 1 MHz, and it results in a pulling range of $1/(4 \times 70\ \text{ns})$ Hz, which is more than twice the required FM deviation. The instantaneous gain of the FLL changes only 3 dB over the FM range.

The coaxial delay line is ovenized to shield the coax from temperature fluctuations resulting from ambient temperature changes and instrument turn-on. The electrical length of solid-Teflon-dielectric semirigid coaxial cable is very sensitive to temperatures below 25°C. At 70°C, the sensitivity reaches a value of zero. At first, the plan was to raise the temperature of the delay line to 70°C. However, heat loss calculations showed that continous power consumption in the coax heater could be as high as 20 watts at low ambient temperatures. A temperature of 45°C was decided upon as a good compromise. At ambient temperatures below 45°C, the coax temperature is held at 45°C. As the ambient temperature goes above 45°C, the heater shuts off. Thus, at high ambient temperatures, the stability of the delay line is dependent on the insulation of the oven and the reduced sensitivity of the delay line to temperatures above 45°C.

Wideband Feedback Loop

The wideband feedback loop includes the low-noise amplifier, the summing amplifier, and the loop integrator. The design of all three blocks involved low-noise, wideband techniques, but special attention was paid to temperature stability and saturation recovery time. Because of the varied specifications, all of the amplifiers in the feedback loop use discrete operational amplifiers.

The best amplifier design for temperature stability uses a monolithic differential pair with no emitter degeneration as the input stage. Even small values of emitter resistance can degrade the temperature coefficient of the input offset voltage by a factor of 10. Because of the low-noise requirements of the low-noise amplifier, the differential pair in this circuit uses two discrete transistors that are thermally coupled.

During a switching transient, the high-gain secondary stages in the amplifiers are likely to saturate or enter cutoff. Reverse-biased Schottky diodes are placed across the collector-to-base junctions of those transistors to reduce saturation recovery time. This technique works well with fast switching transistors that have a low V_{ce} saturation voltage. Many RF and microwave transistors with high V_{ce} saturation voltages come out of saturation with large thermal transients, so they are not recommended for fast switching amplifiers.

Many switches are used in the wideband feedback circuitry to protect the amplifiers from saturating when the FLL is not in use. Also, since each amplifier stage has high dc gain, the stages are separated from each other to keep offset voltages from one stage from saturating another. These circuits can be switched on and off without introducing errors caused by saturation transients.

VCO Pretune Circuit

The VCO pretune circuit must tune the oscillator to within 1 MHz of the final frequency in 3.5 microseconds. After that time, no switching transients from the pretune circuit should be large enough to affect the center frequency accuracy of the stabilized VCO. The pretune circuit also needs sufficient filtering to avoid degrading the noise of the FLL.

The pretune filter is a simple RC low-pass filter with a speed-up switch to short the resistor during the 3.5-μs switching time. After the 3.5-μs switching time, the switch is open. The time constant of the RC filter is one millisecond. The reason for the extremely large time constant compared to the instrument switching speed of 15 microseconds is the shape of the loop gain of the FLL. The one-pole design of the loop means that the stabilized VCO is more senstive to the rate of change or slope of the VCO pretune transients than to the peak value of the transients. The loop has sufficient gain at low frequencies to cancel a large VCO pretune transient with a time constant of one millisecond.

The design approach for the fast pretune circuit was to ensure that the only transients generated have time constants much larger or much smaller than the 15-μs instrument switching speed. Fast-time-constant transients settle out before $t = 15\ \mu s$. Slow-time-constant transients change so slowly that the loop gain of the FLL cancels the errors without generating a large error frequency. Transients with time constants between 12 μs and 15 μs cause considerable frequency error at $t = 15\ \mu s$. Dielectric absorption in a polypropylene pretune capacitor had considerable transient energy with a time constant of 12 to 15 μs, so a specially designed Teflon-dielectric pretune capacitor was chosen instead. The dielectric absorption at 12 μs is specified at 0.004%. This very tight dielectric absorption specification guarantees less than 0.2 ppm center frequency error as a result of VCO pretune transients.

Acknowledgments

The authors would like to recognize Don Borowski for contributions to all facets of the fast hopping VCO design, and Brian Watkins, the firmware designer for the fast hopping VCO. Many thanks to others that listened to some real blue-sky theories without batting an eye.

High-Spectral-Purity Frequency Synthesis in a Microwave Signal Generator

A low-noise YIG-tuned fundamental oscillator and a GaAs divider contribute to the spectral purity of the HP 8665A 4.2-GHz Synthesized Signal Generator.

by James B. Summers and Douglas R. Snook

THE HP 8665A IS THE HIGH-FREQUENCY MEMBER of the Performance Signal Generator family. Like the other PSG instruments, the HP 8644A and the HP 8645A, it employs a simplified single-loop frequency synthesis scheme. The HP 8665A differs, however, in that its fundamental oscillator operates in the octave from 3 to 6 GHz. The fundamental oscillator is a YIG-tuned oscillator that provides low phase noise and low-distortion FM. A low-noise GaAs divider IC was developed to allow division of the 3-to-6-GHz octave into the frequency range of the PSG fractional-N synthesis hardware and the optional frequency discriminator noise reduction hardware. This allows the entire performance signal generator family to share common synthesis hardware, reducing development time and production costs.

Synthesizer Block Diagram

The block diagram of the HP 8665A synthesis section is shown in Fig. 1. The YIG-tuned oscillator (YTO) frequency is set by current flowing in two coils—the main coil and the FM coil. The coil currents determine the magnetic field on the YIG (yttrium-iron-garnet) sphere that serves as the resonator for the oscillator. The main coil is used to tune the oscillator over the full 3-to-6-GHz frequency range. The FM coil provides a fine-tune capability of ±30 MHz which is used to lock the phase-locked loop (PLL) and the optional frequency-locked loop (FLL), as well as for frequency modulation as the name implies. The pretune circuitry sets the main coil current to place the YTO frequency near the desired operating frequency so the PLL can acquire lock.

The YTO output drives the GaAs divider microcircuit (described later) which contains a splitting amplifier and two binary dividers. The divider microcircuit provides output frequencies divided by 1, 2, and 4 for use in the output section of the HP 8665A and a divided-by-4 output for the synthesis section.

Since the 3-to-6-GHz octave does not divide simply into the 515-to-1030-MHz octave used in the HP 8644A and 8645A, additional divider hardware is included to allow reuse of the common PSG fractional-N and frequency discriminator hardware.

The PLL is formed as follows. The YTO signal divided into the 257.5-to-515-MHz octave drives the fractional divider circuitry in the fractional-N module. The signal is further divided down to 200 kHz and compared in a phase detector to a 200-kHz reference signal derived from the instrument's 10-MHz reference. The phase detector output is integrated and then scaled before leaving the fractional-N module. This scaling provides a constant loop bandwidth as the loop frequency (and therefore the loop divide number) is changed.

For PLL-only operation (no FLL), the fractional-N output signal must be fed back to the oscillator to complete the PLL. Since the HP 8665A oscillator has different characteristics from those of the HP 8644A and HP 8645A, further loop gain shaping is required. This is provided by a variable lag-lead network. A low-pass filter is also provided to reduce the levels of spurious products that result from the fractional division process. These fractional spurs are typically less than −48 dBc at 3 GHz within the PLL bandwidth and less than −90 dBc at offsets greater than 10 kHz. The output signal from the low-pass filter controls the YTO frequency by controlling the current in the FM coil.

For PLL + FLL operation, the fractional-N module phase detector output signal is fed to the frequency discriminator module instead of the YTO. The FLL is a wideband loop that senses the FM noise of the oscillator and feeds back a control signal to the YTO to reduce this noise. Any attempt by the fractional-N module output to change the YTO frequency by changing the FM coil current would be seen by the frequency discriminator as a frequency error, and since there is more loop gain in the FLL than in the PLL, the FLL would prevail. To avoid this problem the PLL signal is fed to the discriminator module, where it changes the phase shift in one leg of the delay line discriminator to control the frequency of the FLL. The FLL then appears to be a composite VCO for the PLL. (See the article on page 27 for more information on discriminator module operation.)

The frequency-locked loop is formed by feeding the 515-to-1030-MHz divided YTO signal to the delay line discriminator module. In this module, the RF signal is split into two paths. One path is fed through a long piece of coaxial transmission line to delay it 140 nanoseconds before it is applied to a phase detector. The other path is fed through variable phase shifters to the other input of the phase detector. A frequency discriminator is formed when the variable phase shifters are set so that, at the desired operating frequency, the two signals at the phase shifter are in phase quadrature. As the input frequency deviates from the operating frequency, the phase of the delayed signal changes much more rapidly than that of the non-delayed signal, causing an output from the phase detector that is proportional to the frequency offset. The output of

the phase detector is integrated to produce the output of the discriminator module. The frequency-locked loop is completed by connecting the discriminator module output to the FM coil of the YTO through a lag-lead network. This lag-lead network sets the bandwidth of the FLL and provides high gain at dc for YTO drift tracking.

Frequency Modulation

For PLL-only operation, the synthesis loop is frequency modulated by driving the YTO FM coil and by changing the fractional divide number in the fractional-N module. The modulation signal must be applied in both places because without the fractional-N path the PLL will see the FM as a dynamic phase error and will act to suppress the modulation if the modulation rates are within the PLL bandwidth.

Changing the fractional divide number to generate FM has some advantages. First, it allows true dc-coupled FM without unlocking the PLL. This provides much better carrier frequency accuracy and less frequency drift than the conventional method of providing dc-coupled FM by unlocking a PLL and relying on the drift characteristics of a free-running VCO. Second, the modulation index limitation of the conventional ac-coupled FM system is eliminated.* In a conventional acFM system, if the modulation

*Modulation index = $\beta = \Delta f/f_m$ = peak phase deviation.

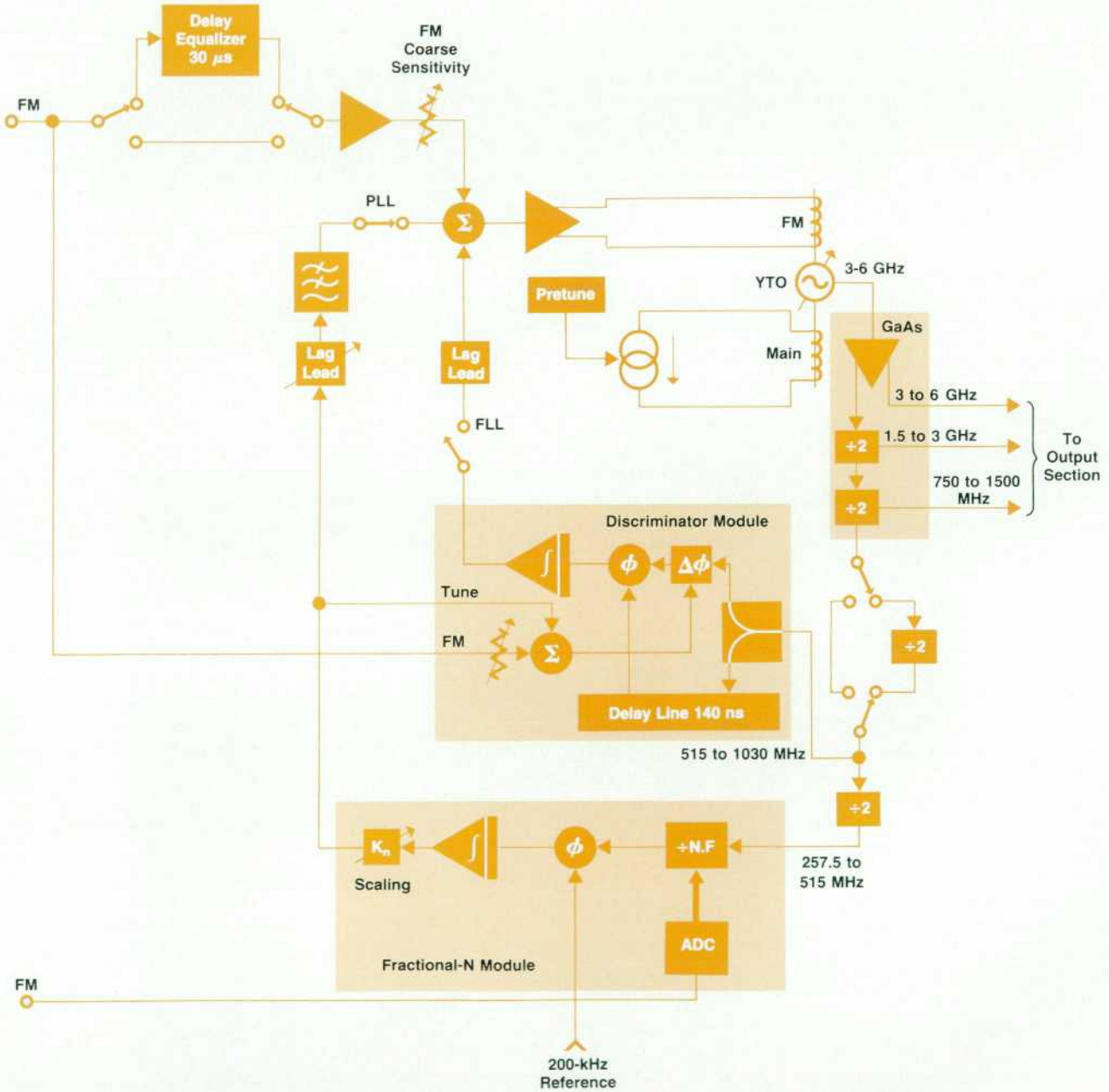


Fig. 1. *Block diagram of the frequency synthesis section of the HP 8665A Synthesized Signal Generator.*

index of the modulated signal at the PLL phase detector exceeds the range of the phase detector the loop will lose lock. The maximum modulation index possible is the peak range of the phase detector multiplied by the divide number of the PLL. Changing the divide number in the PLL as the VCO is frequency modulated produces a constant instantaneous frequency at the output of the loop divider and the phase detector sees no modulation. There is therefore no limitation to the modulation index of the HP 8665A FM system.

There are two disadvantages associated with changing the fractional divide number in some applications. The first is that for very low FM rates the modulation occurs in discrete steps. This is only a problem when an external FM input is used and the peak input levels are in the millivolt range. The quantization step size at the front panel will be 0.5 to 1 millivolt. The second problem results from the time it takes to digitize the FM signal, convert it into the new divide number, and finally, have that new divide number affect the frequency at the phase detector. Altogether, the time delay is approximately 30 microseconds. This time delay causes a phase shift between the YTO and fractional-N FM paths, which results in unflatness of the FM frequency response at modulation rates near the PLL bandwidth. To provide a flat FM response, the FM signal coupled directly to the YTO must be delayed a similar amount or a very narrow PLL bandwidth must be used (since a fixed time delay represents less phase shift at lower frequencies). To provide this delay, a delay equalizer is used. The delay equalizer maintains its delay out to a modulation rate of 15 kHz, after which the delay falls to near zero. When a complex modulation waveform is passed through this nonconstant delay, some waveform distortion will result because the time delays experienced by the different harmonic components of the waveform are unequal. A special function can be used to bypass the delay equalizer and narrow the PLL bandwidth if this is a problem in a particular application.

FM during FLL + PLL operation is similar to PLL-only FM except that the FM signal must also be fed to the discriminator module. Without this path the FLL would act to reduce the frequency modulation at rates within its loop bandwidth. The delay equalizer in the YTO FM path is not required when the FLL is in operation because the FLL reduces the YTO phase noise, allowing a narrow PLL bandwidth to be used at all times.

Single-Loop Synthesizer Design

For the HP 8665A, the PSG single-loop synthesis approach has some real advantages. Since there is no heterodyning in the synthesizer, spurious products are exceptionally good. The nonharmonic spurious performance of the HP 8665A is specified to be better than −90 dBc for frequency offsets greater than 10 kHz from the carrier in the 3-to-6-GHz octave (output section coverage is only to 4.2 GHz.) Another advantage is lower complexity. Since there is only one VCO and one PLL the part count is lower, resulting in lower cost and higher reliability. Another less obvious advantage is lower mechanical complexity. A multiple-loop synthesizer must have high levels of internal shielding to avoid spurious signals resulting from couplings between adjacent circuits.

A disadvantage of the single-loop approach is higher phase noise close to the carrier. Since a relatively low reference frequency is used (200 kHz), the divide number is quite large (30,000 at 6 GHz). This large divide number multiplies the noise floor of the reference and phase detector, raising it as much as 90 dB. To achieve reasonable noise performance, the PLL bandwidth must be limited so that the multiplied reference noise does not degrade the noise of the YTO. A narrow loop bandwidth can adversely affect the switching speed of the synthesizer as well as immunity to power line related spurious and microphonics. A trade-off has to be made between these characteristics. The HP 8665A PLL bandwidth is set at 1 kHz in the PLL-only mode of operation. This provides typical switching speeds of 30 ms (50 ms specification) and phase noise better than the YTO for offsets less than 1 kHz. The YTO noise is degraded slightly between 1 and 10 kHz, but beyond 10 kHz the noise is that of the YTO alone.

While the PLL cannot reduce the phase noise at large frequency offsets from the carrier, the frequency discriminator can. Since the discriminator used in the HP 8665A operates at frequencies between 515 and 1030 MHz, the multiplication of its noise floor is much less (a factor of 8 at 6 GHz). The discriminator frequency-locked loop improves the YTO phase noise at offsets out to 125 kHz from the carrier at 6 GHz. It also provides improved immunity to YTO microphonics because of its relatively wide loop bandwidth. The phase noise improvement provided by the discriminator FLL is typically 17 dB at a 20-kHz offset and 27 dB at a 1-kHz offset from the carrier. Fig. 2 compares PLL-only and FLL-enhanced phase noise performance in the HP 8665A measured at an output carrier frequency of 1 GHz.

YIG-Tuned Oscillator

The selection of an oscillator for the HP 8665A was primarily based on two factors. First, the oscillator phase noise performance had to be very good since the oscillator noise cannot be reduced by the PLL at adjacent channel offsets. (The discriminator can reduce this noise, but it is

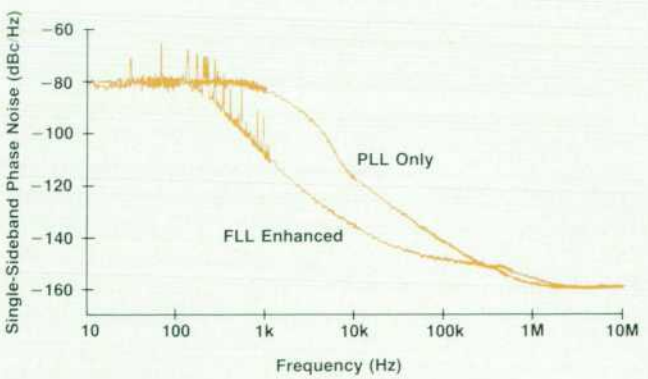


Fig. 2. *HP 8665A phase noise performance at a 1-GHz carrier frequency.*

optional.) Second, for rapid frequency switching it is important that the oscillator be accurately pretuned to allow loop acquisition. The YIG-tuned oscillator was chosen because it satisfies these requirements, having both low phase noise and linear tuning characteristics.

A YIG-tuned oscillator uses a sphere composed of yttrium-iron-garnet as the resonator for a negative-resistance oscillator. This sphere is placed between the poles of a magnet structure which provides a uniform magnetic field in the vicinity of the sphere. The strength of the magnetic field is determined by the current flowing in coils that surround the pole pieces of the magnet structure. A coupling loop is placed around the sphere in a plane parallel to the magnetic field to form the RF resonator. The magnetic dipoles in the YIG sphere have precessional frequencies that are controlled by the bias field set up by the magnet structure. The RF magnetic field of the coupling loop is perpendicular to the bias field and will couple energy to the precessing dipoles at the precessional frequency, thus forming the resonator. The resonant frequency of the sphere is a linear function of the bias field strength, which in turn is a linear function of the current in the main and FM coils. The oscillator tuning characteristics provided by the YIG resonator are therefore very linear, allowing accurate pretuning as well as low-distortion FM. The FM distortion of the HP 8665A in the PLL-only mode of operation is typically less than 0.2% for deviations up to 20 MHz in the 3-to-6-GHz band. This performance is not limited by the YTO linearity, but rather by the audio amplifiers in the FM signal path.

The high-Q YIG resonator and a low-noise bipolar transistor form the low-phase-noise oscillator required for the HP 8665A application. To prevent degradation of the oscillator phase noise, a low-noise main coil driver was developed. At a 20-kHz offset, the phase noise specification in the oscillator band is −105 dBc/Hz.

GaAs Frequency Divider

To allow the common PSG fractional-N divider and frequency discriminator modules to be used with the 3-to-6-GHz YTO, a proprietary gallium arsenide divider IC was developed.

The divider exhibits low residual phase noise, having a noise floor typically less than −160 dBc/Hz. The residual noise of the divider must be better than the divided phase noise of the YTO to prevent the frequency discriminator noise reduction circuitry from acting upon the divider noise by modulating the YTO to cancel the divider noise. This would degrade the phase noise performance in the YTO octave.

Traditionally, a 3-to-6-GHz synthesizer is locked to the reference oscillator by sampling the output signal down to a frequency range that can then be phase locked to a signal derived from the reference oscillator. With the advent of gallium arsenide (GaAs) IC technology, it was found that this function could be accomplished by designing a frequency divider that worked at the output frequency of the main oscillator. A divider chain could then be used instead of a sampler, reducing the complexity and improving the spurious performance of the synthesizer.

Divider Block Diagram

For the HP 8665A, a single hybrid microcircuit (see Fig. 3) was designed to take the output of the 3-to-6-GHz YIG-tuned oscillator and provide the output frequency of 750 to 1500 MHz, which is then further divided by silicon ECL frequency dividers. This microcircuit also provides the drive signals for the output section. A block diagram of the microwave divider microcircuit is included in the synthesis block diagram, Fig. 1.

The HP 8665A microwave divider microcircuit contains three GaAs ICs and associated thick-film substrates. The GaAs ICs are as follows:

- Dual Limiting Amplifier: Amplifies, limits, and splits the 3-to-6-GHz input signal to provide a 3-to-6-GHz output and a drive signal for the first divider.
- 3-to-6-GHz Divider: Takes the 3-to-6-GHz signal from the amplifier and divides the frequency by two. This divided signal is used to drive the next divider and to provide a 1.5-to-3-GHz output.
- 1.5-to-3-GHz Divider: Takes the 1.5-to-3-GHz signal from the first divider and divides the frequency by two, providing two 0.75-to-1.5-GHz outputs, one for the synthesis loop and one for the output section.

The same design is used for both of the dividers in the microcircuit.

The thick-film substrates provide the RF paths to and from the GaAs ICs, the dc bias paths from the bias board to the ICs, and proper terminations for the divider's internal division block.

The microcircuit plugs into a fairly simple printed circuit

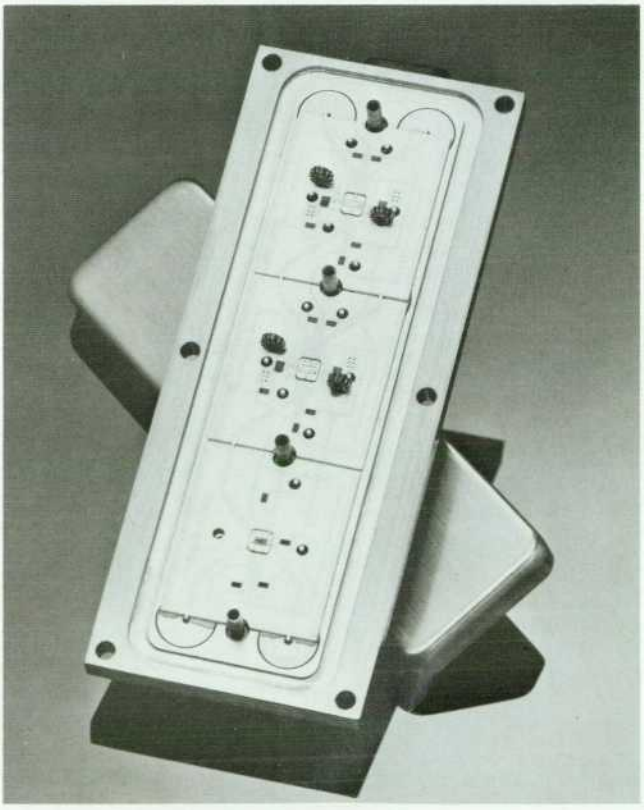


Fig. 3. *GaAs microwave divider.*

board that provides the frequency dependent pretune voltages for the dividers, the adjustments to set the desired power levels for the output section, and the required bias regulation.

GaAs IC Design

The design philosophy for the GaAs ICs called for integrating as much circuitry as possible onto the ICs. This includes matching, bias decoupling, and output stages. The ability to add on-chip matching and output stages shows an advantage of GaAs IC design, that active devices are easy and cheap to integrate. In a typical hybrid microcircuit design, one tries to minimize the number of active components and do as much as possible with passive elements that can be made part of the thick-film or thin-film substrate. With a GaAs IC, it is actually less costly (that is, less space-consuming) to add other active devices, such as FETs, than to use passive components, such as transmission lines, spiral inductors, or capacitors.

Dual Limiting Amplifier. The dual limiting amplifier is a three-stage GaAs IC built using a proprietary HP Microwave Technology Division process.

The first stage of the amplifier is an active lossy match design that provides a 50Ω match to the YTO. The second stage provides the gain and the power-splitting capability. The third stage, the output stage, provides the final limiting and a reasonable match to the output load. The output FET is designed to run out of signal swing at the desired maximum output power. The gate bias of this FET can be set by an external voltage to provide an adjustment range for the desired output power. The second and third stages are replicated to provide two outputs from this IC.

1.5-to-6-GHz Frequency Divider. The second and third ICs in the microcircuit are the frequency dividers. The same type of divider is used in both applications. This device is also an HP Microwave Technology Division IC.

Important design considerations were as follows:

- Frequency range. The 1.5-to-6-GHz input frequency range had to be covered over the instrument temperature extremes.
- Low phase noise. The residual noise of the divider must be lower than the worst other noise contributor in the synthesis loop at the offset of interest. If this is not the case, the divider performance sets the phase noise level of the synthesis loop. It turns out that the offset where many dividers have contributing noise is in the 10-to-100-kHz range, which is significant for many test applications.
- Output power. The output level of the dividers was part of the overall gain budget that set the 8665A output power specification.

The standard design for frequency dividers uses a static or master-slave approach. A flip-flop circuit with feedback generates the divided-by-two signal. This topology has the advantage of not needing frequency bias adjustments for multioctave operation. However, bandwidths up to only about 3 GHz were achievable using this approach with the available GaAs IC process. Another problem was that the phase noise was too high.

The frequency dividers in the HP 8665A use the concept of an injection-locked oscillator. The output is locked to one half of the input frequency. The oscillator is pretuned to a value that is close to the desired output frequency to allow locking by the proper divide number. The ability to design the oscillator to work from 750 MHz to 3 GHz allows the same divider IC to be used in both octaves of operation. Advantages of the injection-locked divider approach are the frequency range and the phase noise performance. Disadvantages are power dissipation and the requirement for a frequency dependent bias adjustment to provide the oscillator pretuning.

All of the divider is integrated onto the IC with the exception of an off-chip bias tee and termination. Also included on the IC are an input matching stage and an output stage. The input stage is an active lossy match design to provide a reasonable 50-ohm load to the input source. A passive match could not provide a good enough return loss without severely attenuating the input signal. The divider output stage provides a buffer between the divider cell and the output load and increases the voltage swing to meet the desired output power requirements. The final part of the output stage includes an adjustment so that the output level can be set for a given device.

Package Design. The three GaAs ICs making up the divider microcircuit are packaged using a steel baseplate, a molybdenum pedestal, and thick-film substrates. The package is made relatively hermetic by using epoxy preform seals for the RF connectors, dc feedthroughs, and lid. The baseplate and pedestal are designed to dissipate the heat generated in the ICs without requiring a separate carrier, which would be required if the baseplate were fabricated from aluminum. The thick-film substrates provide the paths for dc bias and RF signals, along with the bias tee and RF termination for the dividers.

Acknowledgments

We would like to thank technicians Jack Wagner and Matt Dixon for many hours of work and troubleshooting during the development and environmental testing of the HP 8665A. In addition, We wish to thank the printed circuit designers and the R&D procurement organization for their assistance in producing hardware from schematics, and the assembly and test engineers for their work on the test systems that made the developmental and environmental testing of the HP 8665A design possible.

Fred Ives did the early work on the dual limiting amplifier IC and provided technical leadership for the design of the divider IC. Key contributions to the divider design were made by Rory Van Tuyl, Val Peterson, and Don Estreich. The mechanical design of the microcircuit packaging was done by Bruce Roeder and Ben Helmso. Many thanks are due the fabrication and test groups at the HP Microwave Technology Division. Thanks are also in order to Dick Waite, Todd Wendle, Al Reuter, Jodi Zellmer, Don Roberson, and Lugene Arnold for work on process development, assembly, and testing of the microcircuit.

Microwave Signal Generator Output System Design

Noise performance and level accuracy were major design concerns. Thick-film microcircuits, some "packageless," are used extensively.

by Steve R. Fried, Keith L. Fries, and John M. Sims

THE OUTPUT SYSTEM OF THE HP 8665A Synthesized Signal Generator takes the synthesized and divided signals from the GaAs divider described in the preceding article and produces an output signal in the range of 0.1 MHz to 4.2 GHz. It also provides automatic level control (ALC), amplitude modulation, and reverse power protection.

The main output section, which provides output frequencies from 0.1 MHz to 3 GHz, consists of a divided output section for frequencies from 0.1875 to 3 GHz and a heterodyne output section for frequencies from 0.1 to 187.5 MHz. There is a separate microwave extender output section for frequencies from 3 to 4.2 GHz. Fig. 1 is a simplified block diagram of the HP 8665A output system.

Divided Output Section

The divided output section takes two of the octaves generated by the synthesis block and provides level control and AM capability for output frequencies from 187.5 MHz to 3.0 GHz. The upper three octaves of this range are covered by the high-frequency driver and the last octave is covered by the low-frequency driver, along with the heterodyne band, 0.1 to 187.5 MHz. The two drivers supply the inputs for the two sections of the main output amplifier.

High-Frequency Driver. The HF driver accepts two signals from the GaAs divider: 1.5 to 3.0 GHz (DIV2) and 0.75 to 1.5 GHz (DIV4). These signals have high harmonic content and their output levels vary with frequency. The HF driver filters the harmonics and levels the signals to provide a constant level to the high-band modulator. There is also a silicon binary divider on the HF driver which generates the 0.375-to-0.750-GHz octave. The block diagram of the HF driver is shown in Fig. 2. Because of the frequencies involved, the majority of the RF signal handling is done with microcircuits.

The desired frequency octave is selected by the input multiplexer on the premodulator microcircuit. If the 0.375-to-0.75-GHz band is selected, the DIV4 signal from the GaAs dividers is directed to the DIV8 silicon divider. The DIV8 output then returns to the premodulator microcircuit for further processing.

The next circuit on the microcircuit, the premodulator, acts like a variable attenuator controlled by the preleveling loop. If the drive to the premodulator increases by 2 dB then the premodulator attenuation will increase by 2 dB to keep the output level constant. Similarly, if a following stage rolls off by 1 dB the premodulator attenuation will decrease by 1 dB to ensure that the level at the input of the high-band modulator remains constant. While the premodulator performs a function similar to that of the high-band modulator, the performance requirements are much lower, so the design is simpler. The premodulator does not have to do amplitude modulation, so the dynamic range required is smaller. Also, harmonics are not very critical since the filter microcircuit follows the premodulator.

The premodulator circuit is followed by a two-stage GaAs amplifier. The lowest signal level in the HF driver signal path occurs at the input to this GaAs amplifier, so the noise floor is set at this point. Care must be taken to ensure that the instrument noise floor is not degraded by the HF driver. This is accomplished by carefully selecting the gain and flatness of the following stages, and by designing the amplifier for a low noise figure. Harmonics are not a major concern because the filter microcircuit attenuates them.

The filter microcircuit reduces the signal harmonics to the point where the harmonics at the output of the instrument are not dominated by the HF driver. The microcircuit contains two lumped-element low-pass filters and four transmission line low-pass filters. At high frequencies the transmission line filters no longer provide attenuation. This occurs at frequencies where the length of the filter elements is approximately one half wavelength. The lower the filter corner frequency, and therefore the larger the filter elements, the lower the frequency at which the filter no longer attenuates. To ensure that the filter microcircuit rejects very high-frequency harmonics, the 3-GHz filter is always left in the signal path. Being the highest-frequency filter on the microcircuit, it provides isolation to the highest frequencies. An additional output is included to provide a 1000.1-MHz-to-1187.5-MHz signal to the down-converter section.

Following the filter is the high-band modulator microcircuit. This microcircuit includes the peak detector for the preleveling loop and the AM modulator for the 0.375-to-3.0-GHz band. Placing the preleveling loop peak detector inside this microcircuit ensures that the level into the modulator will remain constant. The modulator is a variable attenuator that provides vernier level control and AM and compensates for unflatness in any of the following stages. Since there are no filters after the modulator, it must have very good harmonic performance. It is an absorptive modulator that is constructed as two cascaded pi attenuators. The diodes in each leg of the modulator are chosen based on trade-offs between harmonic performance, modulator

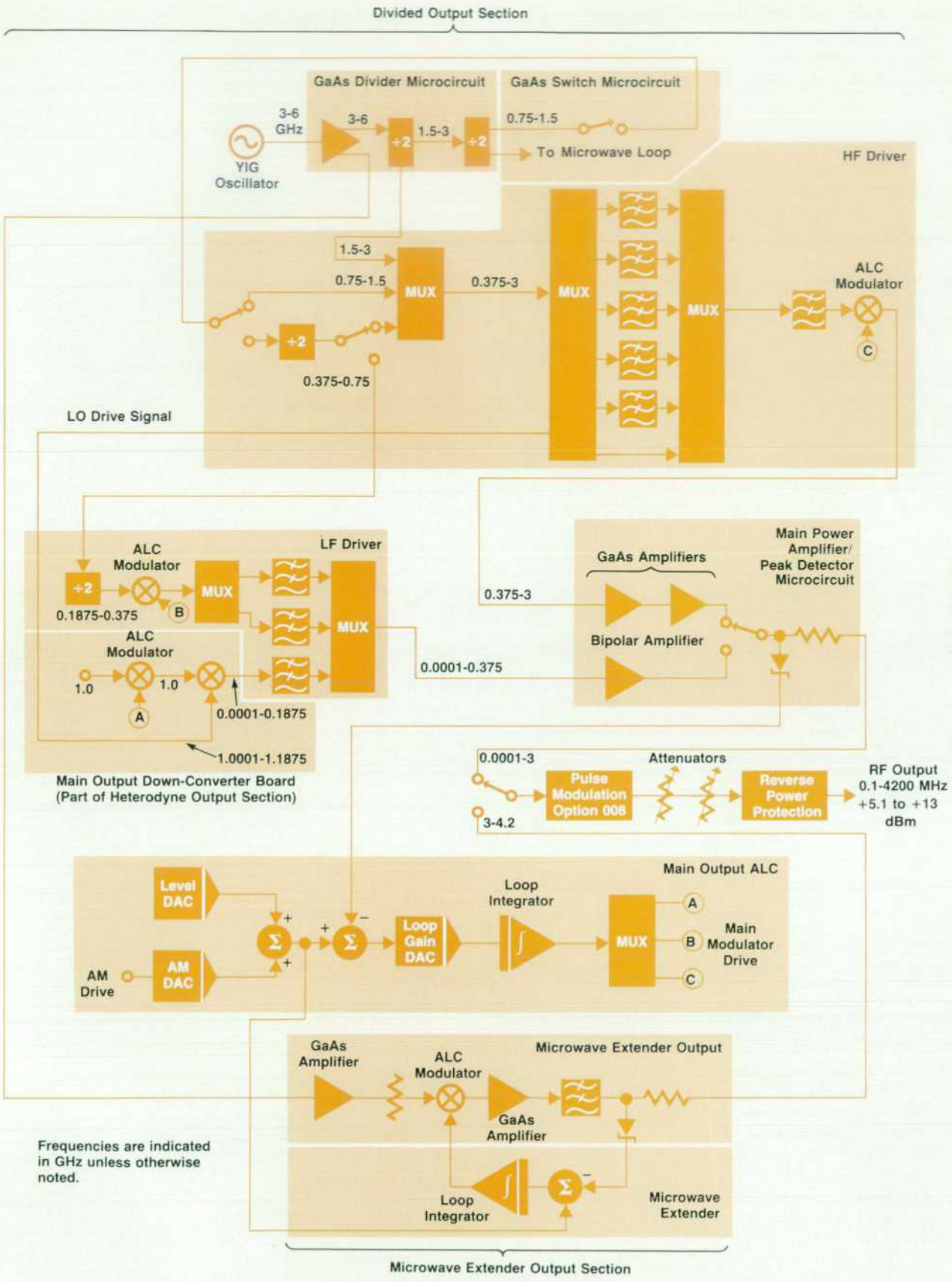


Fig. 1. *Block diagram of the output system of the HP 8665A Synthesized Signal Generator.*

"Packageless" Microcircuits

The "packageless" microcircuit concept was developed in an effort to save on the number of separate microcircuits needed and thereby decrease the cost of packaging these circuits. In the example illustrated in Fig. 1, six low-pass filters are contained on a single large thick-film hybrid on a ceramic substrate. The substrate is suspended in a cutout in the printed circuit assembly (PCA). It is retained by the contact holder and spring fingers from the clamp. Electrical contact between the substrate and the PCA is made by bridging the gap with a conductive rubber contact. Details of this can be seen in Fig. 2.

Both dc and RF connections are made between the substrate and the PCA. The dc connections use a standard pad on both the substrate and the PCA. The RF connections have to accept 50-ohm microstrip transmission lines. Since the transmission line widths are so different between the PCA (0.115 inch wide on epoxy glass, $\epsilon_r = 4.7$) and the substrate (0.025 inch wide on ceramic, $\epsilon_r = 9.6$), special transitions are used coming to the edge of each. The RF performance of the connection between the PCA and the substrate is important to overall circuit performance. The transitions used are designed to have a minimum of 15 dB return loss (VSWR 1.44:1) to 3 GHz. Network analysis and time-domain reflectometry techniques were used to evaluate the transitions during development.

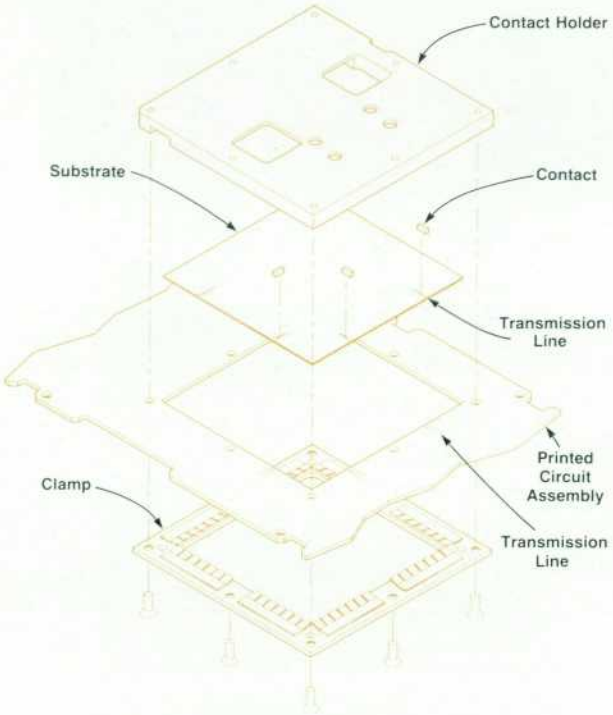


Fig. 1. *"Packageless" microcircuit concept.*

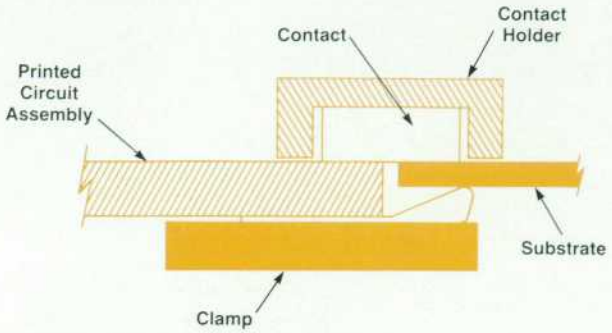


Fig. 2. *Contact detail between the microcircuit substrate and the printed circuit assembly (PCA).*

The gap between the PCA and the substrate is bridged with a conductive silicone rubber contact. This contact is held in a pocket in the contact holder. The pocket is sized so that the contact can expand into it as it is compressed by the clamping force of the assembly. When completely clamped the contact is compressed 20 to 30 percent from its free state.

This system has been quite successful both for electrical performance and for packaging flexibility. Nearly the entire perimeter of the substrate is available for contact sites. This design also makes it unnecessary to package individual circuits in separate microcircuit packages.

Bennie E. Helmso
Development Engineer
Spokane Division

linearity, and dynamic range. The design evolved from a modulator used in the HP 8663A, with the frequency range extended both higher and lower for use in the HP 8665A. The modulator output drives the high-frequency section of the main output amplifier.

Microcircuits are typically very costly. An analysis of previous microcircuit designs showed that a significant portion of the cost could be attributed to the microcircuit package itself. To reduce the cost of the instrument, the premodulator and filter microcircuits were designed as "packageless" microcircuits (see box above).

The design of the HF driver also required careful attention to grounding, layout, bypassing, and cavity resonance suppression to achieve the required level of performance.

Low-Frequency Driver. The low-frequency driver of the divided output section covers 187.5 to 375 MHz. The block diagram is shown in Fig. 2. A 375-to-750-MHz signal from the high-frequency driver drives a silicon binary divider to generate 187.5 to 375 MHz. A separate AM modulator, a reflective pin diode modulator, is used for this octave. Two half-octave filters are used to filter off any harmonics generated by the divider or the buffer amplifiers. The multiplexer on the output of the filters also controls the 0.1-to-187.5-MHz signal from the heterodyne output section. The signal then goes to the low-frequency section of the main output amplifier.

Heterodyne Output Section

The HP 8665A uses frequency division to obtain output frequencies from 187.5 to 4200 MHz. Below 187.5 MHz, a heterodyne section is employed. With this scheme, FM deviation capability is greater than it would be if frequency

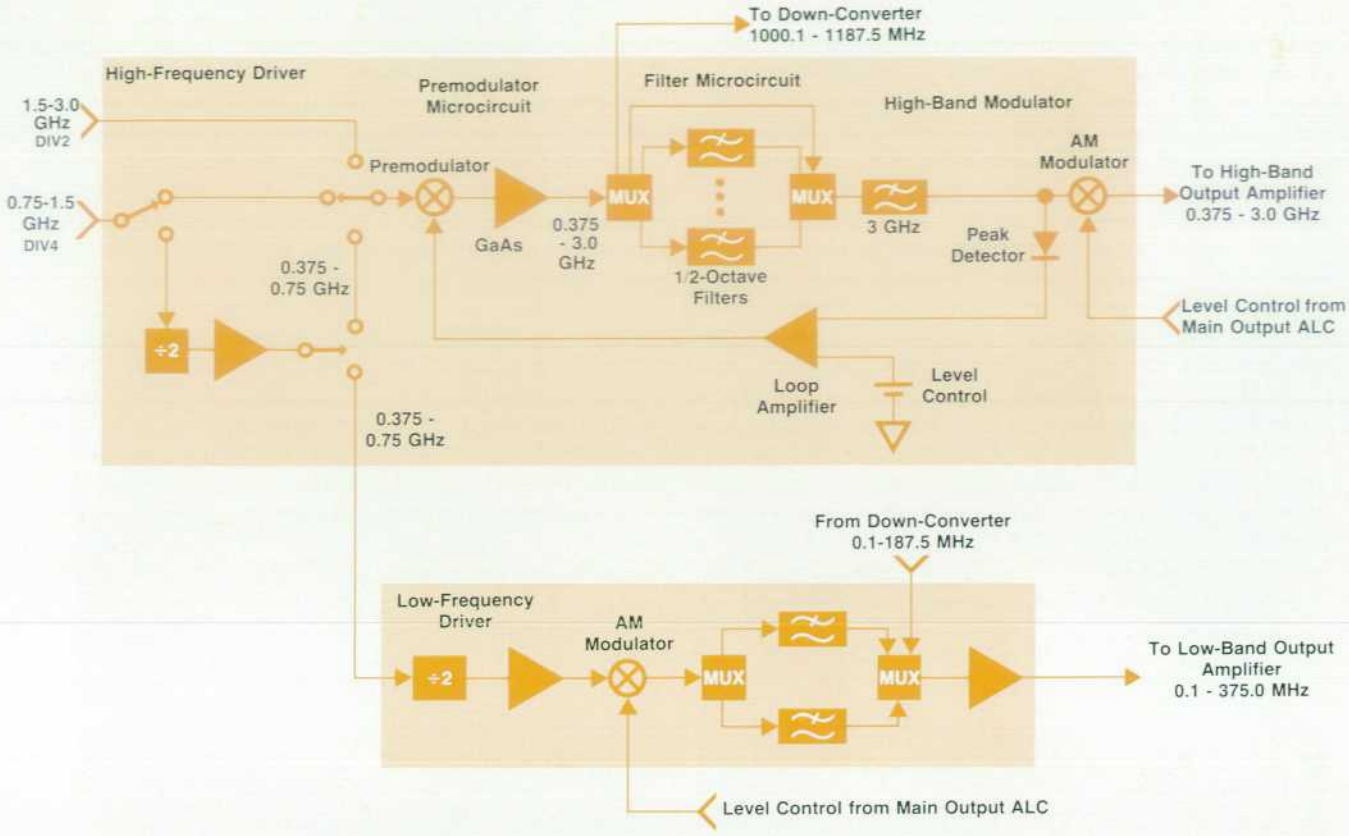


Fig. 2. *Block diagram of the divided output section, part of the main output section.*

division were used. While delivering better FM deviation, the heterodyne process brings with it its own problems in the form of intermodulation products and other mixing spurs, which must be 90 dB below the carrier amplitude. To obtain the 0.1-to-187.5-MHz signals, the heterodyne output section mixes an RF signal at 1.00 GHz with a local oscillator (LO) signal ranging from 1.0001 to 1.1875 GHz.

250-MHz Phase-Locked Loop (PLL). The 1-GHz RF signal is derived from the 10-MHz instrument reference, which is also used in the synthesis portion of the HP 8665A. Instead of phase locking the 250-MHz signal shown in Fig. 3 to the 10-MHz reference by dividing 250 MHz down to 10 MHz, a different method is used. First, the 10-MHz reference signal is multiplied by five to obtain the PLL reference signal. To get down to the 50-MHz signal required for phase detection, the 250-MHz signal is divided by five. The two 50-MHz signals are then phase-locked together, thus providing a 250-MHz signal with the same stability as the 10-MHz reference signal. The 250-MHz signal is then multiplied by four to get the 1-GHz RF signal needed at the mixer.

When operating the HP 8665A, the user can either supply an external 10-MHz reference signal or use the internal low-noise reference. The internal reference is an ovenized voltage-controlled crystal oscillator (VCXO) whose frequency is controlled by a digital-to-analog converter (DAC). An optional high-stability time base can be used as the external reference when better stability is desired. Whichever 10-MHz reference is selected provides to the rear panel a 10-MHz signal phase-locked to the output signal of the HP 8665A.

The 250-MHz signal needed in the PLL is generated by doubling the frequency of the output of another VCXO, which generates a low-noise 125-MHz signal. A varactor in series with the crystal allows tuning of the signal frequency by several kilohertz. A resonant circuit extracts the crystal current. The 125-MHz signal is amplified and doubled using the familiar full-wave rectifier.

To provide the 250-MHz output signal as well as the signal that ultimately gets fed back to the phase detector, an asymmetrical resistive splitter is used. The output with the smaller level is first sent through an isolation amplifier and then to an ECL divide-by-5 circuit. The 50-MHz signal is filtered, buffered, and sent to the phase detector.

A double balanced mixer is used as the PLL phase detector. The phase slope of the detector is 0.2V/rad. The IF output is diplex filtered to provide a 50-ohm termination to the sum frequency and to low-pass filter the difference frequency at dc. The low-pass filtering removes any unwanted signals that might modulate the VCXO via the tune line. The loop integrator, which follows and makes the PLL a second-order loop, has the capability of providing PLL bandwidths of either 100 Hz or 1 kHz. This choice allows optimization of noise either close to the carrier or farther out. The output of the integrator is used to tune the 125-MHz crystal oscillator.

The larger output of the 250-MHz asymmetrical resistive splitter is applied to a double-tuned bandpass filter. This filter rejects subharmonically related signals at multiples of 125 MHz coming from the doubler as well as other spuri-

ous signals generated in the ×5 section. These signals include 10-MHz and 50-MHz offset and feedthrough signals.

1-GHz Generation. Once the 250-MHz signal is phase locked to the 10-MHz reference, all that is needed is to multiply it by four to get the 1-GHz signal. Although this sounds easy, because of concern over spurious signals entering the mixer it is not. The 250-MHz signal is first doubled using a doubler similar to that used in the PLL for the 125-MHz signal. Amplification and filtering are performed twice on the resulting 500-MHz signal. The filtering is accomplished with tunable, helical bandpass filters. At this point, the signal is split. One signal is routed to the rear panel where it can be used for various applications, and the other signal is sent to another doubler.

The 1-GHz signal that exits the doubler has spurious signals that occur every 250 MHz, even though some are relatively low in amplitude because of the 500-MHz filters These signals must be filtered well before reaching the mixer. To accomplish this, a 1-GHz helical bandpass filter similar to the 500-MHz filters is used. Although this filter provides reasonable stop-band performance close-in, above 3 GHz it has very little attenuation. For this reason, a 1-GHz low-pass filter follows the bandpass filter. Along with all the filtering, several shield cans are used to separate the 250-MHz-to-1-GHz circuitry. The shield cans effectively stop on-board coupling from one circuit to another.

A pin diode reflective modulator controls the vernier level, amplitude modulation, and any other unflatness in the ALC loop. Because the gain of the modulator, defined as the change in the RF voltage divided by the change in the modulator control voltage, is not a linear function of the modulator control voltage, a shaper was designed to linearize the gain and effectively reduce the variation in the ALC loop gain and bandwidth. An isolation amplifier between the modulator and the mixer decouples the LO feedthrough signal from the modulator.

LO Drive Signal. The LO signal comes from the high-frequency driver (see Fig. 1). Its frequency range is from 1.0001 to 1.1875 GHz. Since the LO signal at the mixer must be between 14 and 21 dBm, the signal from the HF driver is amplified by two transistor amplifiers that use high-pass structures to boost the gain at 1187.5 MHz. A nine-pole

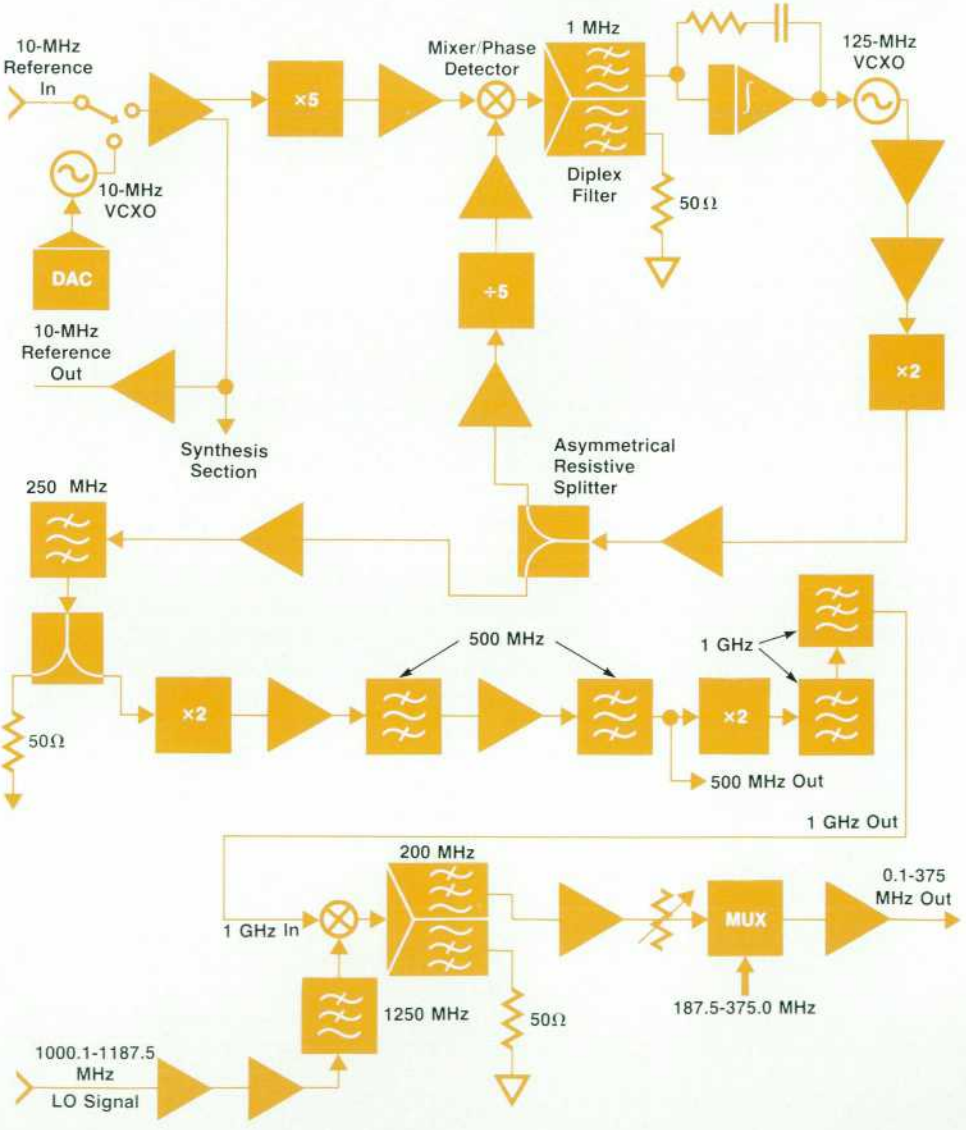


Fig. 3. *Block diagram of the heterodyne output section, part of the main output section.*

low-pass filter made with thick-film technology removes harmonics of the LO signal before it enters the mixer.

IF Circuitry. The mixer used in the heterodyne output section is a double balanced mixer specifically selected for low intermodulation distortion. Two internally generated distortion products of concern are the $5f_{LO} - 6f_{RF}$ and the $4f_{LO} - 5f_{RF}$ products. Each mixer is tested to make sure that these two distortion products are low enough to meet the instrument specification.

As shown in Fig. 3, the IF (output) port of the mixer is terminated in a diplexed filter. This filter performs several functions. First, it provides a 50Ω termination to the sum frequency coming out of the mixer. Second, it attenuates some harmonics and intermodulation distortion products, especially $4f_{LO} - 5f_{RF}$, which almost comes into the IF frequency range. Lastly, the filter attenuates the LO and RF feedthrough signals. Because of leakage to other parts of the board, the diplex filter is designed with no leads protruding through to the bottom side of the board, and it is covered by a shield can.

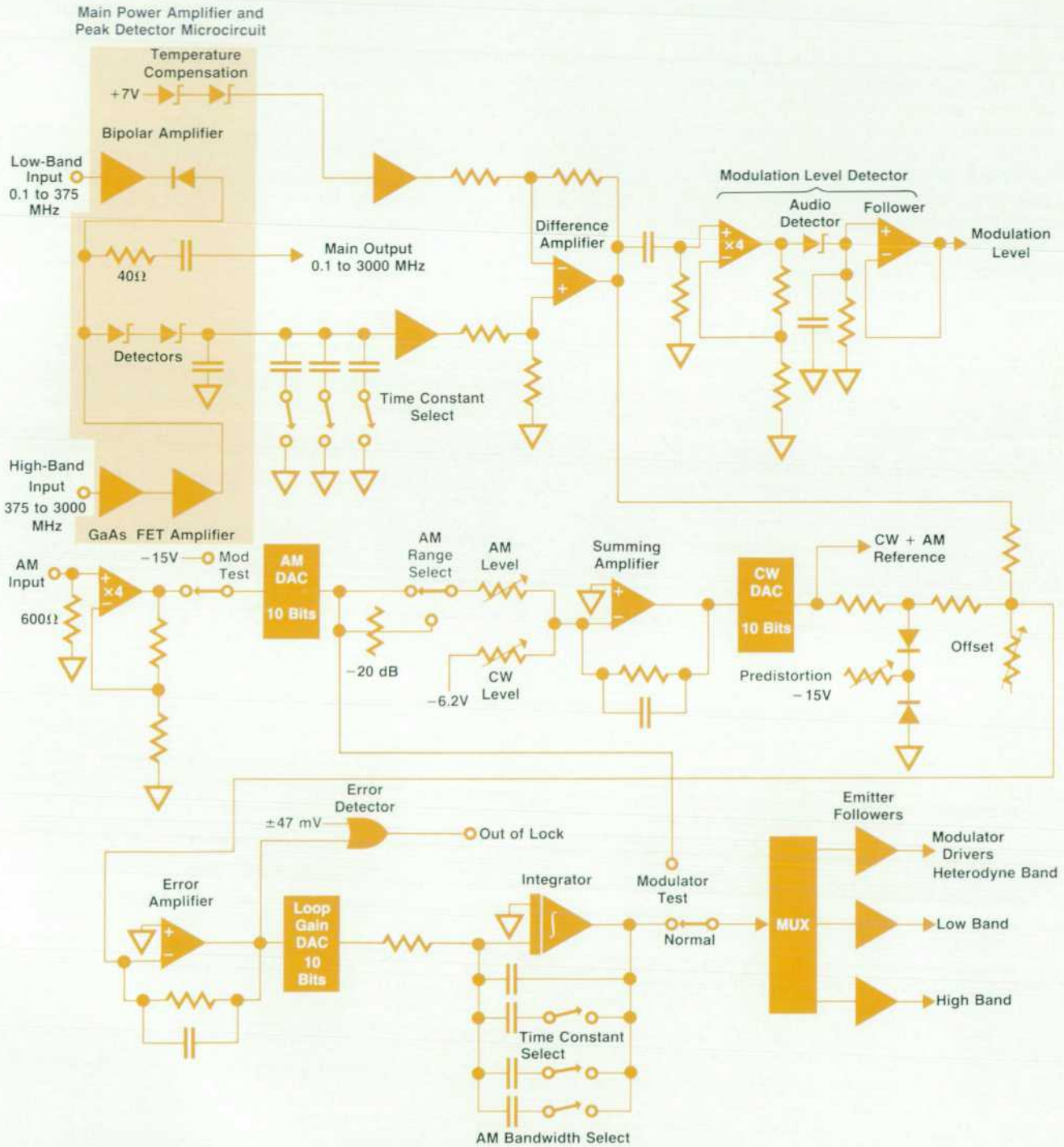


Fig. 4. *Main power amplifier and main ALC loop.*

Following the diplexed filter is the IF amplifier stage. This stage provides good harmonic performance. This is important because there is no IF filtering afterward. Besides being very linear, these amplifiers also have a low noise figure so that they do not degrade the noise floor, which is primarily set by the level out of the mixer. Since the level out of the mixer sets the noise floor, switchable attenuators are included after the amplifiers in an attempt to keep the output level of the mixer within a 2-dB window as the amplitude is varied. As the amplitude approaches the bottom of the vernier, attenuators are switched in, causing the level out of the mixer to be higher than it would be without the attenuators. The multiplexer at the output of the attenuators switches between the 187.5-to-375-MHz signal, which comes from the last divider section in the instrument, and the heterodyned signal. Both divider and heterodyned signals are sent to the main output amplifier.

Main Power Amplifier and Peak Detector Microcircuit

The main power amplifier microcircuit (see Fig. 4) is designed to provide signal switching and gain to two independent RF sources within the main output section while maintaining high spectral purity. The microcircuit selects between the LF driver (0.1 to 375 MHz) and HF driver (375 to 3000 MHz) signals, depending on the desired carrier frequency. The circuitry consists of a single bipolar amplifier for the LF band and a dual-stage GaAsFET amplifier for the HF band. To conserve power and provide increased signal isolation, only one of the amplifiers is switched on at a time. A pin diode performs the RF switching between the amplifier outputs. The LF amplifier provides 5 to 7 dB of gain to the LF driver output signal, which is prefiltered to at most −38 dBc harmonics. The HF amplifier adds 15 to 20 dB of gain to the HF driver output signal, which has better than −35 dBc harmonics. Each power amplifier stage provides greater than +17 dBm saturated output power with at most −30 dBc harmonics for instrument levels up to +10 dBm.

A peak detector samples the output RF envelope voltage for power leveling purposes. Two low-barrier Schottky diodes are connected in series to provide low junction capacitance and high breakdown voltage, which are very important parameters in this application. A second matched diode pair is placed within the microcircuit to compensate for the temperature drift characteristics of the main detector diodes. A backmatch resistor is inserted between the detection node and the output connector to provide the 50Ω source impedance under leveled conditions. The result is a highly accurate, stable RF output level from 0°C to +55°C.

Main ALC Loop

The main RF output signal (0.1 to 3000 MHz) is leveled by comparing the detected RF envelope voltage to the de-

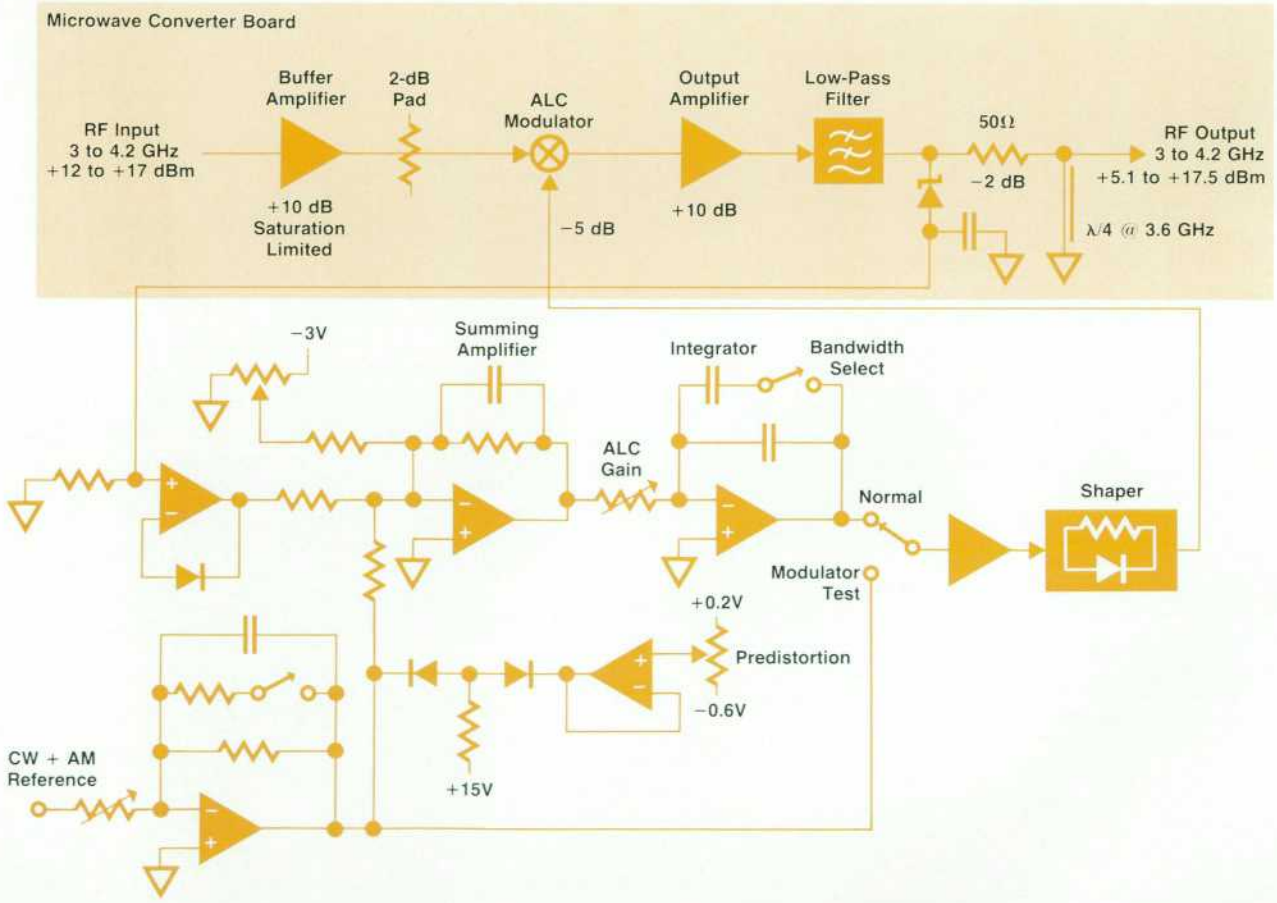


Fig. 5. *Block diagram of the microwave extender output section and ALC loop.*

sired reference voltage. The reference signal contains both CW level and AM information, and is corrected to include the effects of cable attenuation and the square-law response of the peak detector. Any difference between the detected and reference voltages is integrated to provide a control voltage to the ALC modulator. The results are a highly accurate RF output level over temperature and accurate, low-distortion AM. The main ALC loop is shown in Fig. 4.

Switchable integrator time constants and an adjustable-gain DAC provide a stable ALC loop bandwidth as the carrier frequency and level are changed. In addition, the detector time constants are switched to allow proper tracking of the sampled RF envelope voltage. This produces very low-distortion AM and permits fast level recovery after the carrier frequency is changed. A self-calibration routine determines the optimum gain DAC settings for constant ALC bandwidth so that manual adjustments are not required. The ALC integrator time constants are varied to compensate for large loop gain variations and to reduce the possibility of intermodulation distortion. The ALC loop bandwidth is approximately 100 kHz with AM switched on and 100 Hz with AM switched off. By reducing the bandwidth in CW mode, the possibility for third-order intermodulation distortion is reduced, since the ALC loop responds less to external signal sources.

An AM predistortion circuit compensates for the linear-to-square-law transition of the main leveling detector. By shaping the reference AM waveform at levels below a set threshold, the reference signal closely approximates the actual response of the detector. The result is low-distortion AM, independent of carrier level and AM depth.

Microwave Extender Output Section

The purpose of the microwave extender module (Fig. 5) is to amplify, level, modulate, and filter the 3-to-4.2-GHz signal from the GaAs divider microcircuit. The microwave extender output signal is fed to the transfer switch, attenuators, reverse power protection system, and finally to the RF output connector. It bypasses the main power amplifier.

The GaAs divider signal is supplied at a level from +12 to +18 dBm and may contain harmonics as high as −6 dBc. This signal must be amplified or attenuated sufficiently to obtain the desired instrument output level, including the effects of AM. Additionally, the harmonics must be attenuated to a maximum level of −30 dBc.

Because of the narrow frequency band requirements of this module, Teflon printed circuit board was chosen as the substrate medium. This allowed easier integration into the module frame without the need for an expensive microcircuit package design. This choice provided some design challenges because of the effects of component parasitic elements at microwave frequencies. Packaged GaAsFETs, leaded ultralow-capacitance pin diodes, and sapphire variable capacitors are among the special components that were selected to meet these challenges.

The 3-to-4.2-GHz input signal is buffered by a GaAsFET amplifier and a 2-dB pad to provide a fixed +21-dBm level to the ALC modulator. This stage also isolates the GaAs dividers from impedance variations of the modulator across its operating range. The modulator serves as the control element in the ALC loop by providing variable attenuation from 5 to 55 dB. A second GaAsFET stage amplifies the modulator output signal to as high as +23 dBm, depending on the modulator attenuation. The low-pass matching structure of the two GaAsFET amplifiers attenuates harmonics to about −20 dBc. The output of the second amplifier drives a nine-section microstrip low-pass filter, which cleans up the harmonics by an additional 20 dB. The result is a controllable output level from −30 dBm to +18 dBm with harmonics typically better than −40 dBc in the extender band.

The detector portion of the microwave extender is similar to that used in the main output section. Direct peak detection using a dual series low-barrier Schottky diode is employed to maximize the signal-to-noise ratio and detector linearity and minimize drift over temperature. As a result of the high operating level at the detector, AM distortion

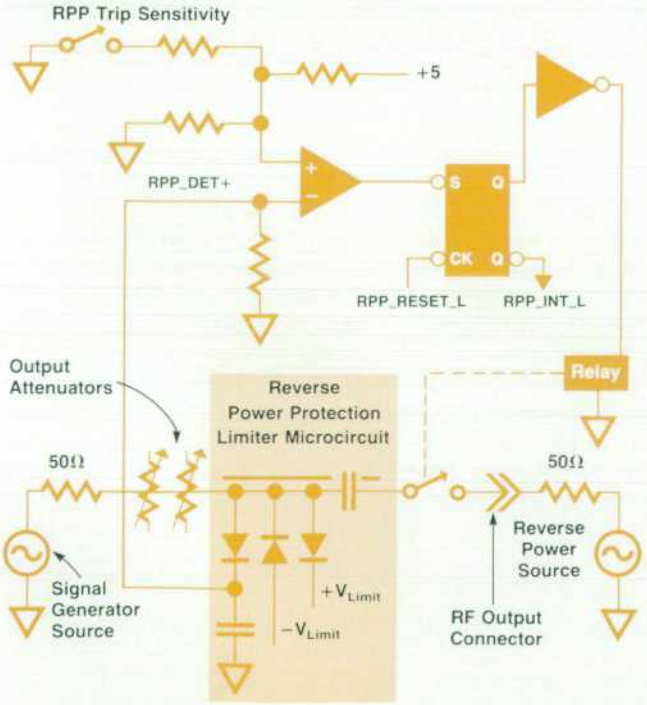


Fig. 6. *Reverse power protection system.*

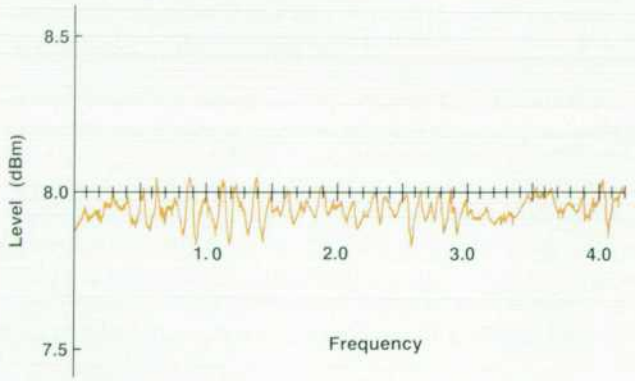


Fig. 7. *Typical HP 8665A level accuracy at +8-dBm output power at 25°C.*

is typically better than 2% at 90% AM depth. A second matched diode pair compensates for the residual temperature drift of the detector. The 50Ω module source impedance is achieved by placing a backmatch resistor between the detection node and the microwave output connector. A quarter-wave shorted stub across the output protects the module circuitry from electrostatic discharge transients.

The microwave extender ALC system sets the output power at a specified value within the resolution of the 10-bit CW level DAC. The ALC system is also responsible for amplitude modulating the output signal in response to either internal or external audio sources. The ALC loop functions by comparing the detected microwave envelope voltage to a desired reference voltage. The resultant error signal is integrated to drive the modulator to the correct attenuation, thereby achieving leveled output power. Because the reference voltage contains both CW level and AM signal information, the output microwave envelope is leveled and amplitude modulated in accordance with the CW and AM DAC settings. With AM switched on, the closed-loop bandwidth exceeds 100 kHz, producing minimal AM distortion for rates up to 1 kHz. With AM switched off, the integrator time constant is increased to produce a loop bandwidth of 100 Hz.

Reverse Power Protection System

A diagram of the reverse power protection system is shown in Fig. 6. The purpose of reverse power protection is to protect the output amplifiers, detectors, and attenuators from overload in case of accidental reverse power transmission, such as when testing RF transceivers, and from electrostatic discharge energy. The reverse power protection relay is designed to interrupt the flow of reverse power into the instrument output section in less than 12 microseconds from initial application. Short-term overloads are suppressed by the reverse power protection limiter. The maximum power capability is limited to 25 watts by the amount of energy that the limiter pin diodes can absorb in the brief 12-microsecond period before the relay opens. The reverse power protection system is designed to handle rated power up to 2060 MHz, where the majority of high-power transceivers operate. Above 2060 MHz, the instrument is designed to handle up to one watt of reverse power. As an extra safety measure, the output attenuators are switched to maximum insertion loss settings whenever reverse power is sensed. When the reverse power protection system activates, its status is indicated by a front-panel display message.

Although the reverse power protection detector is designed to respond only to sources up to 2060 MHz, the reverse power protection limiter and relay must pass signals in the entire 0.1-to-4200-MHz range with low insertion loss and minimal degradation of the output standing wave ratio. This is accomplished by the use of a thick-film limiter microcircuit and a coaxial reed relay. The limiter microcircuit contains the reverse power protection detector, positive and negative diode clippers, and a series dc-blocking capacitor. The reverse power protection relay consists of a reed module with SMA connectors attached to each end, surrounded by conductive epoxy. Because of the coaxial structure of this design, the relay approximates a pure 50Ω transmission line with insertion loss less than 0.7 dB and return loss greater than 16 dB up to 4200 MHz. The output of the reed relay feeds the instrument RF output connector directly.

Level Accuracy Enhancement

The HP 8665A uses correction algorithms to control the RF output power within ±1 dB for levels of −119.9 dBm to +13 dBm. Highly accurate production test systems measure the actual instrument output level to determine the cable, attenuator, and reverse power protection system losses. The level error data is transferred to instrument EEPROM where it is used to compensate the output level into a 50Ω load. Additional power measurements are made for the main output and microwave extender modules to correct for their individual output level errors. This allows modules to be swapped without recalibrating should the necessary test equipment not be available, although there will be some accuracy degradation because of mismatch uncertainty between modules. The module correction data resides in EEPROM inside the affected unit.

To correct for the cable, attenuator, and reverse power protection losses, power measurements are made every 50 MHz at high levels and every 100 MHz at low levels. Linear interpolation is used to correct for frequencies between the sample test points. Measurements are repeated for every attenuator setting to correct for all possible instrument settings. The correction resolution is 0.01 dB, while the CW DAC resolution is 0.05 dB or better. The result is a highly accurate corrected output level, typically within ±0.2 dB of the desired value across the entire frequency range at room temperature. Module swapping without recalibration will degrade this typical accuracy somewhat.

Fig. 7 is a plot of typical level accuracy at +8-dBm output power at a temperature of 25°C.

Acknowledgments

We wish to thank Tom Yeager for his contributions to the design of the heterodyne section and the main output section, which includes the power amplifier, ALC loop, and AM circuitry. Steve Curtis also made significant contributions in his design of the reference hardware in the HP 8665A. Technicians Jack Wagner and Matt Dixon deserve much credit for their long hours of product support and testing. Assemblers Jan Neddo and Marsha Suman are to be commended for their dedication to assembling quality boards, no matter how difficult the processes involved. The assembly and test group did a fine job of developing automated tests, the manufacturing engineers in setting test-line specifications and in coordinating last-minute design changes, and the microcircuits group in ensuring producible designs. Special thanks go to project manager Dave Platt for his strong support to the entire design team.

Design of a High-Performance Pulse Modulation System

The pulse modulation option for the HP 8665A Synthesized Signal Generator adds a pulse modulator and an internal pulse generator. The pulse modulator uses gallium arsenide field-effect transistor switches on microwave monolithic integrated circuits.

by Douglas R. Snook and G. Stephen Curtis

PULSE MODULATION IS AN IMPORTANT modulation format for radar testing and calibration, design and evaluation of avionics systems such as DME and SSR, IF filter characterization, and EW or ECM work. Pulse modulation is available as as an option in the HP 8665A Synthesized Signal Generator.

Key contributions of the HP 8665A pulse modulation are:

- Internal pulse generator with programmable delay from 50 ns to 1 s, width from 10 ns to 1 s, and pulse repetition frequency (PRF) from 0.1 Hz to 400 kHz (external PRFs from dc to 10 MHz are allowed)
- 5-ns rise and fall times, with less than 2-ns typical
- Greater than 80-dB on/off ratio
- No change in the level accuracy performance from the standard HP 8665A. The maximum output power for the pulse modulation option is +9 dBm.

Internal Pulse Generator

The HP 8665A pulse option includes an internal pulse generator, which provides the video signal that drives the pulse modulator. Including this function in the HP 8665A eliminates the need for an external instrument to set the pulse parameters of PRF, delay, and width. This integration is a first for a Hewlett-Packard synthesizer.

The internal pulse generator is controlled by either the internal modulation source or an external signal. These signals initiate the generation of the video signal in one of two ways: direct pulse control, in which the video characteristics are set by the control signal's width and PRF, and internal pulse generator, in which the control signal triggers the delay and width generators. The positive edge, the negative edge, or both edges of the control signal can be used as the trigger event for the internal pulse generator mode. In the case of the direct control mode with the internal modulation source as the control signal, the resulting video signal is a 50%-duty-cycle square wave at the frequency of the modulation source. Fig. 1 shows timing waveforms for the two control modes.

The internal pulse generator provides programmable delay and width, triggered by either the internal modulation oscillator or an external stimulus. Three digits of resolution are programmable, although the actual resolution is specified as 0.2% of full scale for a given range, which turns out to be less than three digits for some settings. The

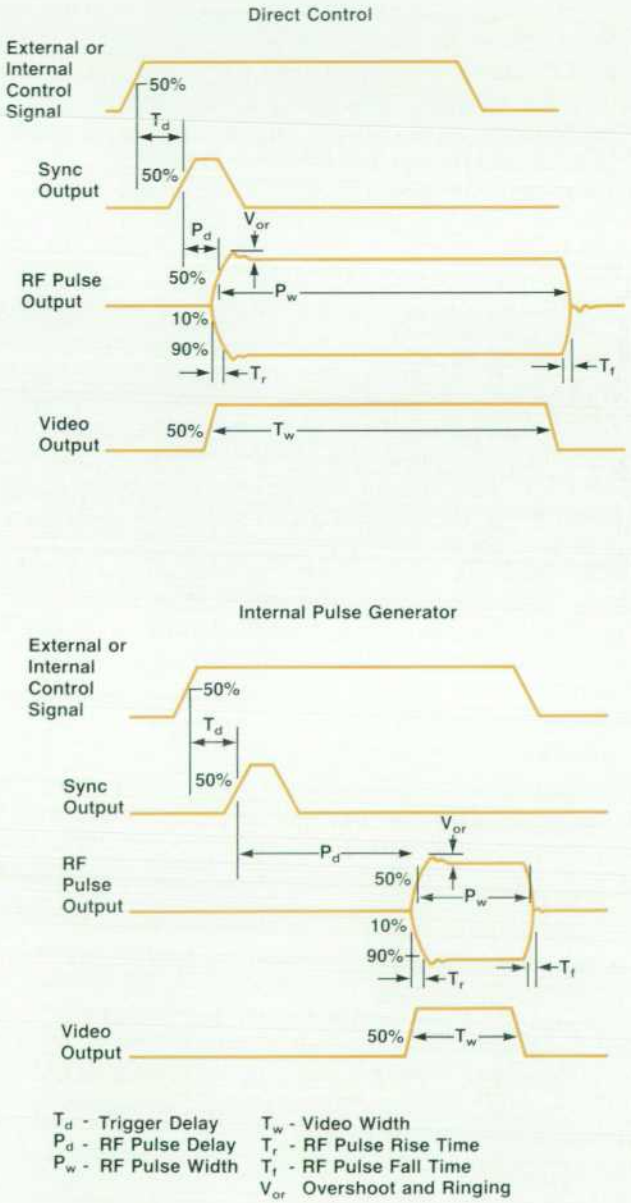


Fig. 1. *Timing signals for the two control modes of the internal pulse generator of the HP 8665A Synthesized Signal Generator with pulse modulation option.*

pulse delay can be set from 50 ns to 1 s in eight decade ranges. The lowest range is 50 ns to 99.9 ns. The accuracy is ±5% of the programmed setting ±2 ns. The pulse width can be set from 10 ns to 1 s in eight decade ranges. The accuracy is ±5% of the programmed setting ±2 ns for widths greater than 50 ns and typically ±5% ±2 ns for widths between 10 ns and 50 ns.

As already mentioned, internally generated PRFs range from 0.1 Hz to 400 kHz, and external PRFs from dc to 10 MHz are accepted. The external input can be set to a high (TTL gate) impedance or to 50Ω.

Two outputs are generated: a 50-ns-wide sync pulse and a video pulse that is representative of the RF pulse in both width and delay. The rising edge of the sync pulse is the reference point for the RF pulse, that is, the programmed delay is the time between the rising edge of the sync pulse and the rising edge of the RF pulse. The video and sync outputs are designed to drive 50Ω loads.

The pulse modulator consists of two HP-designed GaAs ICs in a custom microcircuit assembly. This modulator provides several key performance advantages:

- Low insertion loss with good flatness across the instrument's 100-kHz-to-4.2-GHz frequency range
- Good return loss characteristics in both on and off states
- Excellent rise and fall time performance, typically less than 2 ns
- On/off ratio of better than 80 dB
- Minimum pulse widths of 10 ns.

Pulse System Design

There are two common ways of including pulse modulation in a signal generator block diagram: either within the automatic level control (ALC) loop or external to the ALC loop. These two options trade off the complexity of the ALC system for the required performance of the pulse modulator.

Inside the Loop. Pulse modulation within the ALC loop can be implemented in two ways. The first option is for the ALC loop to use a sample-and-hold circuit to set the proper voltage in the loop when the RF is turned off, maintaining leveled output power.[1-4] The second option is to unlock the ALC loop and set the output level by driving the vernier with stored calibration data.[5]

Pulse modulation within the ALC loop has several advantages. First, modulator unflatness can be compensated by the ALC loop. Second, the pulse can be translated to other frequencies, allowing a lower-bandwidth pulse modulator. Third, video feedthrough can be suppressed if there is a high-pass filter following the pulse modulator, as is the case in a YIG-tuned multiplier. Fourth, mismatch related transient problems are minimized.

There are also several disadvantages of this approach. First, the ALC loop must be sampled or broken to set the output level, adding to the complexity of the ALC design. Second, when the ALC loop is locked, its operation sets the pulse performance in terms of minimum pulse width and duty cycle. This in turn sets the range of possible PRFs. Third, the pulse waveform may be degraded by the circuitry following the modulator, especially in the case of frequency translation. Fourth, the video feedthrough coming out of the pulse modulator may adversely affect the bias of the following circuitry.

Outside the Loop. Pulse modulation external to the ALC loop places the modulator after the power amplifier but before the output attenuators. A major advantage of pulse modulation external to the ALC loop is that the ALC loop remains locked during pulse modulation, so the output level remains constant independent of pulse width or PRF. Note that is this is only true if the match presented to the power amplifier detector is the same during both the on state and the off state of the pulse modulator. Otherwise the difference in load will change the drive required to set the desired detector voltage, causing the output power to vary.

Pulse modulation outside the ALC loop is easy to include as an option. No changes are required in the ALC circuitry, except that enough vernier range must be in place to account for pulse modulator unflatness. The insertion loss and unflatness of the pulse modulator can be calibrated out in the same way as for general level accuracy calibration.

In the HP 8665A, the use of the GaAs IC pulse modulator

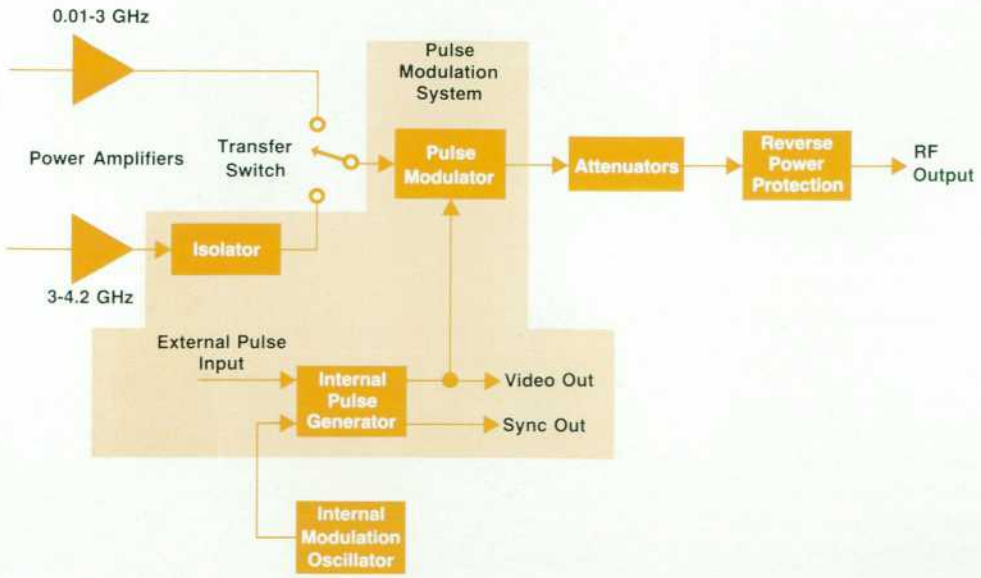


Fig. 2. *Block diagram of the HP 8665A pulse modulation system. The pulse modulator is outside the ALC loop.*

microcircuit makes it possible to implement the less expensive, more modular option by putting the pulse modulator external to the ALC loop. Fig. 2 is a block diagram of the HP 8665A output system with the pulse modulation option. For the HP 8665A implementation, no changes were required between the standard instrument and the pulse option, other than adding the pulse-specific hardware.

Putting the pulse modulator outside the ALC loop does have some disadvantages. First, a high-performance pulse modulator is required in terms of bandwidth, insertion loss, match, on/off ratio, rise and fall times, and video feedthrough. All of these parameters are passed on directly to the output. Second, the insertion loss must be relatively constant over temperature because level calibration is done at only one temperature. Third, the return loss of the power amplifier driving the pulse modulator must be well-controlled. If the mismatch between the amplifier and the modulator is significant, the resulting multiple reflections during the turn-on and turn-off transients can degrade the system rise or fall time. In the HP 8665A, this problem was minimized in the 3-to-4.2-GHz band by adding an isolator between the power amplifier and the pulse modulator. Performance below 3 GHz was adequate without additional components.

User Interface

The user has access to the pulse modulation features through a mixture of front-panel keys and special functions. Pulse modulation is enabled or disabled by pressing the **PULSE** key on the front panel and then pressing **ON** or **OFF**. The choice of using the internal modulation oscillator or an external signal for the control signal is made by pressing either **INT** or **EXT DC** on the front panel. External ac coupling is not a option for pulse modulation.

Following an instrument preset, the default pulse modulation mode is the direct control option, with **EXT DC** selected as the modulation source. To gain access to the other pulse modulation features, the following special functions are used:

- Special 210: 50-Ohm Pulse. Press **ON** or **OFF** to set the input impedance of the external dc-coupled input port. **ON** is 50 ohms, **OFF** is a high impedance (the input to a Schottky TTL Schmitt trigger).
- Special 211: Pulse Control. This function is used to select between the direct control mode and the internal pulse generator mode. To view the current control mode, press **ON**. To select between the two choices, press the up or down keys or rotate the knob.
- Special 212: Pulse Delay. This function sets the delay from the sync pulse to the RF pulse in internal pulse generator mode.
- Special 213: Pulse Width. This function sets the width of the RF pulse in the internal pulse generator mode.
- Special 214: Pulse Trigger Edge. This function selects the trigger event that initiates the internal pulse generator timing sequence. To view the current control mode, press **ON**. To select between the three choices of positive, negative, or both, press the up or down keys or rotate the knob.

Pulse Generator Design

The block diagram of the internal pulse generator is shown in Fig. 3. The heart of this circuitry is the timing IC used to develop the delay and width timing information. Two of these ICs are used. The first timing IC generates the delay signal and triggers the second timing IC, which generates the width signal. This results in the video signal which is sent to the rear panel and the pulse modulator.

The timing IC was developed for use in the HP 8111A, 8112A, and 8116A Pulse/Function Generators.[6,7] The IC consists of a current-controlled oscillator (CCO) that can

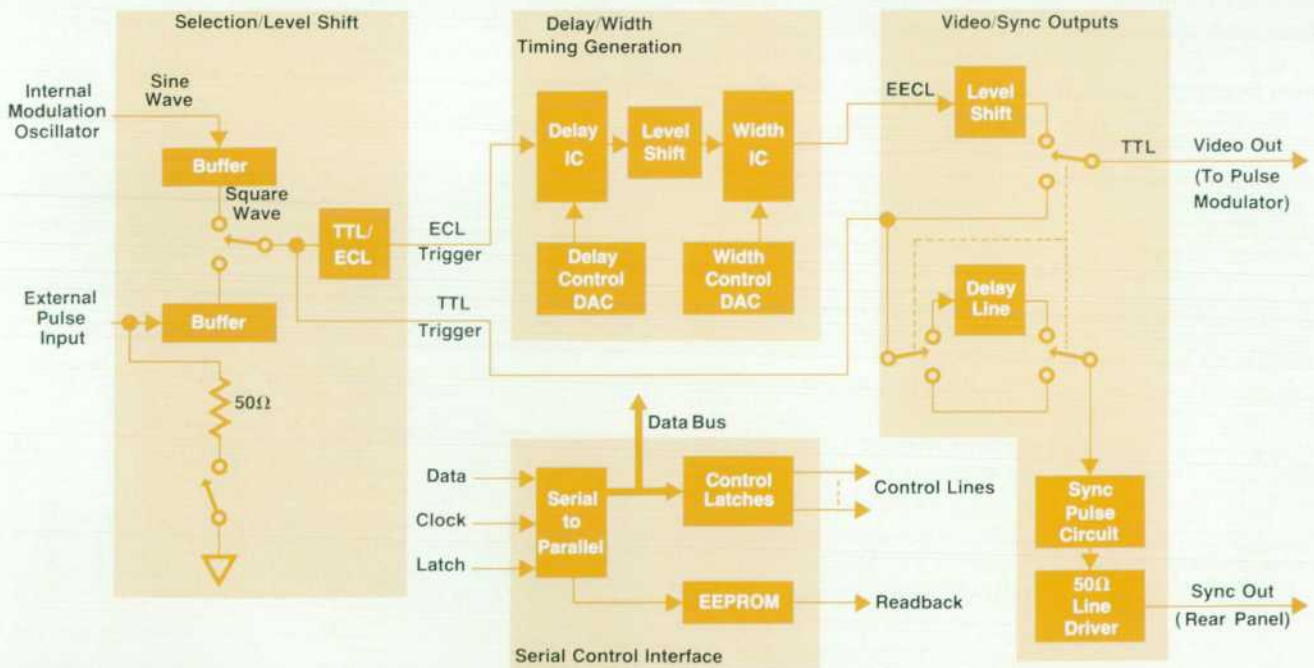


Fig. 3. *Block diagram of the HP 8665A internal pulse generator.*

be tuned from 10 MHz to 100 MHz and 2 MHz to 20 MHz. This corresponds to timing signals of 10 ns to 500 ns. Longer time signals are obtained by an internal flip-flop and decade counters. No external capacitor switching is required. The various decades of operation and the desired trigger mode are set by TTL-compatible lines driven by the instrument microprocessor. Fine control within the selected decade is provided by a DAC that supplies the tuning current for the CCO. This DAC is used to calibrate the video signal so that the RF signal has the proper delay and width characteristics relative to the sync output from the rear panel. This calibration compensates for variations in the timing ICs and the RF path, including differences in delay because of varying cable lengths and in width because of differences in the RF pulse compression characteristics of the pulse modulator. The DAC circuitry also includes temperature compensation, which corrects for temperature drift in the timing ICs.

The sync signal is generated by the event that triggers the delay and width timing generation, or for the direct control mode, the rising edge of the control signal. For the internal pulse generator mode, a delay line is included in the sync pulse path to compensate for setup times and propagation delays through the timing circuitry.

The input circuitry chooses whether the controlling signal comes from the front panel or from the internal modulation oscillator. Signal conditioning circuits shape the wave properly, set the voltage levels for TTL or ECL, switch the trigger signal to the appropriate path, and provide the desired input impedance for the external input—either 50 ohms or high impedance (the input to a Schottky TTL Schmitt trigger).

The internal pulse generator is controlled by the instrument microprocessor via the standard serial-to-parallel bus design used throughout the PSG family. The DAC calibration data is stored in an EEPROM on the pulse generator board and is loaded into the instrument main RAM during instrument calibration.

Pulse Modulator Design

The pulse modulator microcircuit consists of two identical proprietary GaAs ICs in a package designed for high RF isolation in a small area. In the off state the input and output connections are terminated in 50Ω to prevent unwanted reflections. The isolation in the off state is >90 dB, achieved by using cascaded series and shunt FET switches. In the on state insertion loss is less than 4 dB to 4.2 GHz, primarily a result of resistive loss in the series switches and mismatch loss caused by the capacitance of the shunt switches. The low capacitance of the FET switches means that the rise and fall times of the pulse modulator are very fast, on the order of 1 ns. The modulator ICs have built-in buffer stages to speed up the edges of the clock input to the IC, allowing it to be driven by relatively slow logic gates without degradation of switching time.

The frequency range of the HP 8665A requires that the pulse modulator design work from frequencies below 100 kHz to over 4 GHz. Gallium arsenide FETs are among the few devices that can act as a switch over this frequency range and simultaneously exhibit low loss, high isolation, and low distortion. FET switches also generate little video feedthrough (i.e., the switch control signal couples onto the output signal). Generally, microwave signal generators employ pin diodes because of their ability to act as variable resistors for RF signals. A pin diode modulator would not have been able to cover the heterodyne band of the HP 8665A because of its inherent distortion at low frequencies. If pin diodes had been used, a second modulator for the heterodyne band would have been needed.

Using GaAs FETs in modulators is not a new idea, but in the past the cost of assembling a high-isolation modulator out of discrete FETs would have been prohibitive. The availability of GaAs monolithic microwave ICs made it possible to integrate enough FETs to achieve the desired performance and lower the cost to a reasonable amount. The HP Microwave Technology Division supplies the pulse IC used in the HP 8665A. Integration of additional functions such as driver amplifiers into the IC simplified the design of the microcircuit and pulse level shifting circuits.

A feature of this design is its near 50Ω impedance in both on and off states. Because of this, the modulator is always matched to the transmission lines at its input and output, which minimizes signal reflections between the pulse modulator and the power amplifier and between the load connected to the signal generator and the pulse modulator. These reflections can cause distortion of the leading and trailing pulse edges, as well as level inaccuracy.

Package and Microcircuit Design

The pulse modulator microcircuit, shown in Fig. 4, uses two identical GaAs ICs connected in series for switching the signal on and off. Two ICs are necessary to get 90 dB of isolation at 4.2 GHz because of undesired signal coupling between the device input and output. Three thin-film-on-sapphire substrates are used to make RF and bias connections to the ICs. The gold-plated aluminum package has two cavities, each of which holds an IC and a substrate. In a narrow slot connecting the two cavities is a thin-film substrate with a microstrip transmission line that carries the RF signal from one IC to the other. The slot forms a waveguide below cutoff, which attenuates any signal

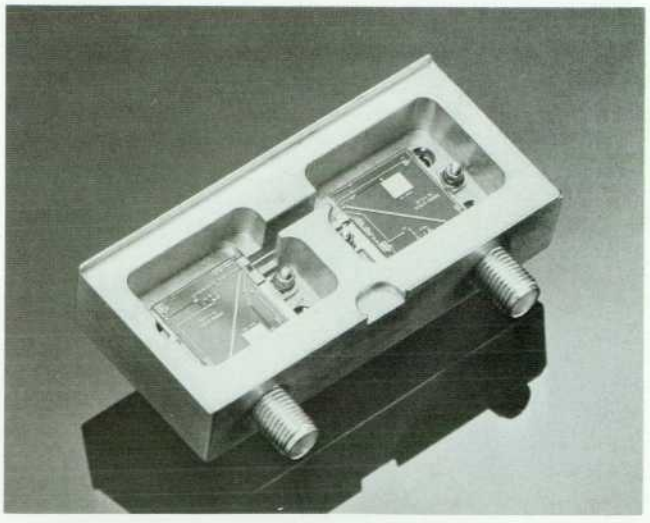


Fig. 4. *Pulse modulator microcircuit.*

radiating from the first cavity to the second, preventing unwanted coupling from the input of one IC to the output of the other, which would degrade isolation.

Each IC is mounted in a recess in a molybdenum pedestal. The pedestals raise the IC so that it is flush with the thin-film substrates used for signal transmission and bias connections. The top of the pedestal is flush with the IC top surface, allowing short ground bonds from the IC edge to the pedestal. The gold-plated pedestals are machined using wire EDM. Molybdenum has a thermal coefficient of expansion similar to GaAs and has good thermal conductivity, so heat from the IC is conducted out to the package.

The RF signal enters the package through a coaxial connector. A ribbon bond is used to attach the connector center conductor to a microstrip transmission line on a 0.010-inch-thick sapphire substrate. The substrate is thin to reduce the area of signal current loops and improve isolation. On this substrate, a 50Ω microstrip transmission line is about 0.010 inch wide. To maintain a repeatable transmission line width (which determines the impedance), the more precise gold thin-film-on-sapphire technology was chosen over a thick-film process. An added benefit is that the signal transmission loss of thin-film conductors is less than that of thick-film. The thin-film process chosen also made available thin-film resistors, which are used on the substrate as damping resistors on the bias lines and for a resistive divider network, which can be used for bias adjustment. The bias and clock lines are relatively high-impedance traces with capacitive terminations at one end—a 0.015-μF feedthrough capacitor in the package base for the V_h and V_l power supply lines, 50 pF for the clock line. In addition, the V_h line has 20 pF of filtering capacitance on the IC. At around 3 GHz, the bias lines resonate with the feedthrough capacitors, causing a sharp increase in the insertion loss of the pulse modulator. This problem was eliminated by inserting 5Ω and 10Ω damping resistors near the low-impedance point in the line (near the feedthrough capacitors).

The substrate and pedestal are attached to a gold-plated Kovar carrier. The thermal coefficient of expansion of Kovar is similar to that of sapphire. If the carrier were not used and the substrate attached directly to the aluminum base, the substrate would probably pop off or break at temperature extremes because of the difference in expansion. Each carrier is held down in the package with nuts on two studs protruding from the package base. The difference in thermal expansion is accommodated by allowing the carrier to slip on the base. Using a carrier also simplifies rework of the microcircuit, since the assembly can be lifted out after removing a few bond wires and two retaining nuts. Bias and the pulse clock are brought through the package base using feedthrough capacitors. These reduce RF leakage from the package and prevent isolation degradation that might otherwise result from signals leaking out one side of the modulator and back into the other side via the bias or clock lines. The bias lines, clock line, and signal trace all converge next to the IC. Small bonding pads are provided to simplify wire-bonding to the IC.

Additional pads on all traces are provided for ground bonds, which are used for electrostatic discharge protection during the assembly process. Although the same IC is used on both sides of the microcircuit, the two substrates are not mirror images, since one IC is rotated 180 degrees relative to the other to place the switchable 50-ohm termination on the proper side.

Probably the most challenging part of this design was understanding the factors that determine isolation across an IC and a package. Even when the signal path is brought right to the side of an IC and bond wires to the IC are kept short, large loops can be created by circuitous ground paths. As depicted in Fig. 5, when the IC is in the off state, signal current flows from the microstrip line on the substrate across the bond wire to the IC input, down through a shunt FET switch and a 50Ω terminating resistor, through ground bond wires to the pedestal, down the side of the pedestal, and onto the carrier, then returns to the source on the ground plane plated onto the bottom side of the sapphire substrate. On the other side of the IC (right side of Fig. 5) a current loop is formed by the bond wire from the microstrip line on the center substrate to the IC, then through a shunt FET switch, through ground bonds to the pedestal, down the side of the pedestal to the carrier, down the side of the carrier to where it contacts the package base, and then up the package wall to where it contacts the ground plane of the center substrate. The mechanical arrangement was chosen to make the circuit manufacturable, with only the center substrate being difficult to remove from the package (it is attached with conductive epoxy). The drawback of this arrangement is that the carrier contributes a significant proportion of the loop size on the side of the IC attached to the center substrate. Substrate thickness is important because it determines the minimum height of the loop. Originally a 0.025-inch ceramic substrate was used before switching to 0.010-inch to meet isolation specifications. The other critical factor determining loop area is how close together the substrates, pedestal, carrier, and package wall are. The assembly process minimizes gaps by pushing the pedestal to the edge of the carrier, then pushing the substrate against it. When the carrier is installed, it is pushed up against the package wall.

An ideal arrangement to get high isolation would have the substrates butting right up to the IC with a common, continuous ground plane. Because this would require cutting a hole or slot into a substrate, adding a substrate, or

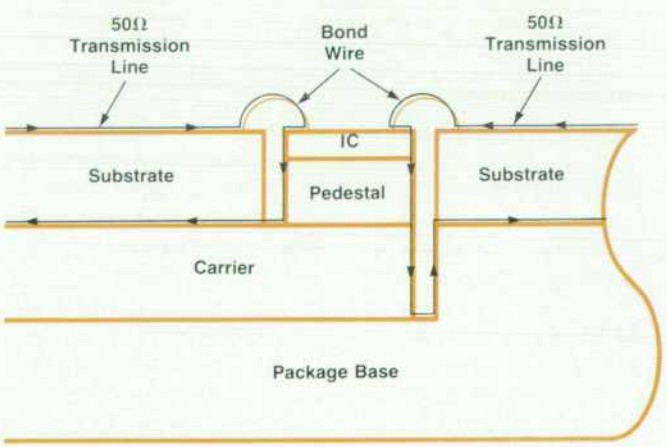


Fig. 5. *Coupling loops in an IC in the off state.*

getting rid of the carriers, all of which increase cost or reduce manufacturability, this alternative was not chosen for this design.

One assumption made in the design was that the bottoms of the carriers and the package base are flat and make contact or are very close to each other everywhere. If this is not true, the size of the current coupling loop can increase, degrading isolation. This has been observed in some instances.

Pulse IC Design

Designing a pulse modulator requires trading off insertion loss for isolation. For this design, as little insertion loss as possible was desired, but it was decided that less than 4 dB to 4.2 GHz was adequate. Isolation had to be good enough to meet the industry standard for microwave generators, an 80-dB on/off ratio. Since it was improbable that this could be done with one IC, the goal was to do this with just two ICs, with a specification of 45-dB isolation and 1.5-dB insertion loss for each IC. For this application, the modulator only needs to switch between on and off, functioning as a single-pole, single-throw (SPST) switch. Commercially available ICs have been designed to perform this function, but only in the last two years have they come close to the needed isolation. A FET used as an SPST switch can be connected as a shunt switch (shorting the signal to ground), or as a series switch (opening the signal path). The two types can be combined for even more isolation.

FET Modeling. A simple model for a FET in the pinched off state is a capacitor from drain to source, C_{ds}, a capacitor from gate to drain, C_{dg}, and one from gate to source, C_{gs}. The resistances at pinchoff are large enough to be considered infinite. The on state model is simply a resistor, $R_{ds(on)}$, connected from drain to source, with equal value capacitors, C_{dg} and C_{gs}, connected from gate to drain and gate to source. Depending on the accuracy needed, the resistance of the ohmic contacts must also be considered. In many cases it is sufficiently accurate to combine the contact resistance with R_{ds}. In the on state, the drain-to-source capacitance, C_{ds}, can be ignored, because a much lower impedance, $R_{ds(on)}$, is in parallel with it.

A shunt switch in the off state uses the FET in the on (low-resistance) state. Fig. 6 shows this configuration along with parasitic elements. The amount of signal attenuation achieved depends on the ratio of $R_{ds(on)}$ to the source impedance seen at the switch input terminals. A larger FET will have lower on resistance and provide more signal attenuation. At high frequencies, the inductance of the wire bonds connecting the FET to ground causes the total impedance to ground to increase proportionally to frequency, degrading the switch isolation. Ground bonds need to be kept as short as possible. Several can be placed in parallel to reduce inductance. This design uses three very short bonds in parallel for about 0.1 nH of inductance.

In the FET shunt switch in the on state (see Fig. 6), the off FET parasitic capacitances mentioned earlier cause a mismatch on the signal trace, which must be considered as a transmission line at high frequencies. This mismatch will cause some of the input power to be reflected back to the source instead of all going to the load. This loss can be about 2 dB at 4.2 GHz unless it is reduced by adding inductive matching elements in series with the switch. Another problem that can occur, and had to be fixed in this design, is that if a low resistance is connected in series with C_{dg}, then at higher frequencies, some of the signal that is supposed to be passing through the switch to its output will be capacitively coupled to the resistance, causing signal attenuation. The solution used here is to place a moderately large resistance in series with the gate to minimize signal loading. The value of resistance is not so large that it degrades the switching time of the FET.

For a series switch in the off state (see Fig. 6), the FET is also off. When a FET channel is pinched off, the resistance between between drain and source is so large that any signal leakage through the switch can be attributed to capacitive coupling through C_{ds} or the series combination of C_{dg} and C_{gs}, even at low frequencies. FET size can be reduced to decrease feedaround, but this will increase the on resistance of the FET, resulting in higher insertion loss when the switch is on. The sizes of FETs in series switches are generally chosen based on the desired insertion loss and distortion characteristics. As was the case with the shunt switch, a low resistance connected to the gate can cause excess insertion loss. This design avoids that problem because the series FET gate is connected to the drain of a driver FET, which is pinched off when the series FET is on. The gate of the series FET then sees a very high impedance, minimizing any loading of the signal via the source-to-gate or drain-to-gate capacitance.

Matching. To increase isolation, the IC design uses three shunt switches alternating with two series switches (see Fig. 7). This increases the isolation to over 50 dB at 4.2 GHz while keeping insertion loss to around 1.5 dB at the same frequency. One of the shunt switches is used to connect a 50Ω resistor from ground to the signal trace, thus terminating the input of the IC with a good match to the transmission line. This is necessary because the series switch that follows is an open circuit when off, and would cause undesirable reflections of the incoming signal back

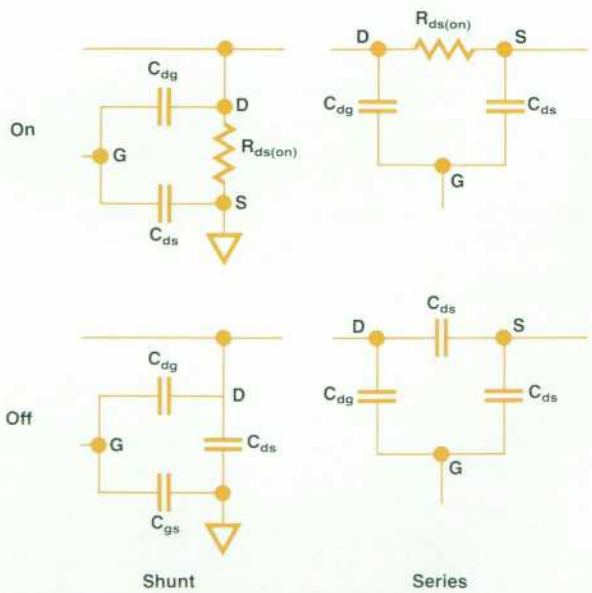


Fig. 6. *Models for field-effect transistor switches.*

to the signal source. The switchable resistor is located on the left side of the IC (also called the "out" side). The out side is connected to the transmission line that leads to the RF connector on the side of the package. One IC is rotated 180 degrees to connect the nonterminated sides (the "in" sides) together. Because the series switch that follows the 50Ω load isolates the input from the capacitance of the shunt FETs that follow, the match is very good in the off state, typically better than 20 dB return loss, even at 4.2 GHz.

In the on state, matching is mainly dependent on the approximately 0.6-nH inductance of the bond wires connecting to the IC. The inductance cancels some of the capacitance to ground of the shunt FETs. The resistance of the series FETs also affects the match, raising the input impedance. The match is not nearly as good in the on case, but return loss is usually better than 12 dB.

Distortion. In a series switch, all the signal current must flow through the FET. Even with the gate-to-source voltage equal to zero, sufficient current through the channel will cause it to saturate and act as a current source instead of a resistance. This means that for small voltages, the current through the FET is linearly proportional to the input voltage, but at high voltages, the FET cannot supply more than the saturation current I_{dss}, which is determined by its size. Depending on the distortion level required, the signal current must be kept to some fraction of I_{dss}.

Another important element in the design of a series switch is how the gate-to-source voltage is held to zero, which allows operation at minimum resistance and maximum I_{dss}. In many series FET switches, a voltage source is used to change the gate voltage. When the switch is on, the gate voltage is held at zero, and V_{gs} is approximately zero when the signal voltage is small. For a low-loss switch, most of the signal voltage at the switch input appears at its output, across the load. A voltage impinging on the switch will cause the source and drain voltages to rise relative to the gate voltage. As this voltage differential increases, the conductance of the channel decreases, and the channel will eventually pinch off if the signal level is large enough. This can cause severe signal distortion. One method to reduce this problem that works at higher frequencies is to place a resistor in series with the FET gate. If the resistance is larger than the impedance of the gate-to-source or gate-to-drain capacitance, most of any signal voltage on the source or drain will couple onto the gate, causing the gate voltage to track this signal voltage, keeping V_{gs} at zero and distortion low. This technique fails at low frequencies because the capacitive impedance from signal trace to gate becomes large compared to a practical resistor value. On an IC, FETs are readily available, so a 100-μm device is used to drive the gate of the series FET. A 50-μm FET is used as an active pull-up for the driver. The source of the driver is connected to the negative voltage supply (V_h) used for reverse-biasing the gates of the switch FETs. The drain is connected to the gate of the switch FET. The gate and source of the active pull-up are connected to the driver drain, and its drain is connected to the signal trace. When the driver FET is off, the pull-up FET shorts the gate of the switch to the signal trace. The beauty of this approach is that almost all the source voltage at the switch source or drain also appears at its gate because the gate is essentially open-circuited (the driver FET loads the gate of the switch FET very little). Distortion is very low at all frequencies. Placing an open circuit in series with the gate also makes the capacitive loading of the signal trace by C_{dg} or C_{gs}

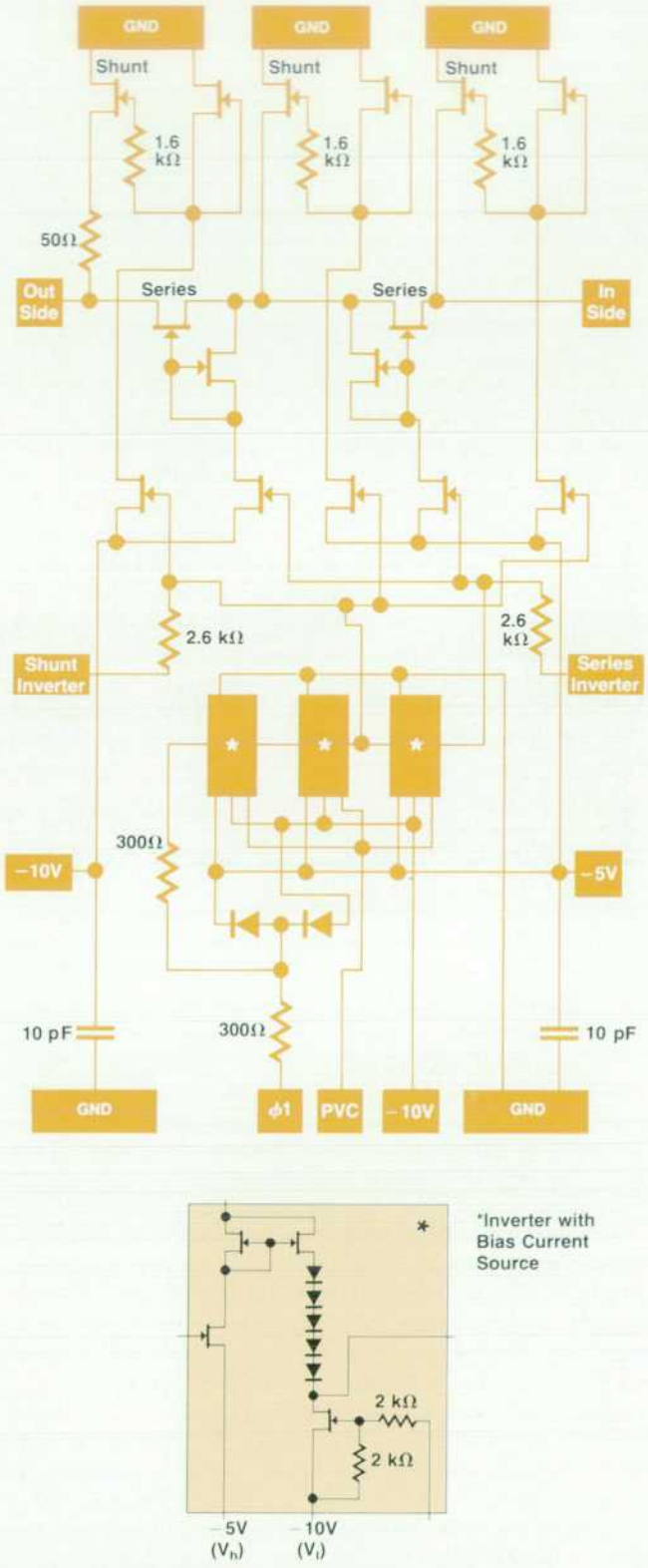


Fig. 7. *Schematic diagram of the GaAs pulse modulator IC.*

negligible.

The shunt FET switches can also cause distortion. To switch the signal off, the shunt switch FET gates are held at some voltage below the pinchoff voltage. Large negative signal swings can reduce the gate-to-drain voltage enough that conduction occurs either through the channel or by capacitance modulation. This type of distortion can be remedied by making the gate-to-source voltage more negative.

Video Feedthrough. Video feedthrough in FET switches is fairly small because they are voltage-controlled devices with low capacitance. High-speed switching can pull current through the gate capacitance in the form of narrow spikes. In this design, a larger effect masks this phenomenon. During switching, the bias for the series FET drivers is pulled out of the RF trace. The current normally comes through the center shunt FETs, which are on when the series drivers pull current. The result is a negative-going, 1-ns-wide spike at the leading and trailing edges of the pulse. The energy in these narrow spikes is small enough that they are not a problem in most applications.

Switching. As mentioned elsewhere, the pulse modulator IC uses three shunt switches and two series switches. A complementary drive signal must be applied to the gates of the series and shunt switches. An integrated inverter is used to do this job, with its output connected to the shunt switches and its input connected to the series switches. Two more inverters are cascaded in front of the control line for the series switches. These inverters serve to sharpen the edge of the external clock signal, making the rise and fall times of the IC insensitive to the clock rise and fall times. Both inverter outputs are brought out to a pad at the chip edge to make probing their voltage possible. The inverters all operate with two negative bias voltages, V_h and V_l. The outputs of the inverters swing approximately between these two levels. The threshold voltage is the level where the input FET in the inverter operates at one half its saturation current, I_{dss}, which is about a volt below V_h, the less negative of the two supplies. The FET current sources used for biasing the voltage follower in the inverter output stages have a resistor network tied to their gates and brought out to a bonding pad. Decreasing the voltage on this pad will shift the output voltage swing of the inverters downward. Although it has not been needed, this pad was added for process variation control (PVC in Fig. 7). The external clock input to the first inverter has ESD protection added in the form of a 300Ω series resistor and clamp diodes to V_h and V_l.

The shunt and series inverter outputs do not drive the shunt and series FET gates directly. As mentioned elsewhere, a driver with an active pull-up is used to control the gates. The sources of the driver FETs are tied to V_h, so that the switching threshold is the same as for the inverters. One advantage of using driver FETs for each switch is that isolation is improved. Signal leakage through a series FET to the gate of the next series FET is reduced by the isolation of two driver FETs. The driver FETs also provide an extra stage of amplification, further sharpening the edges of the external clock. The switching time of the pulse modulator is about 1 ns, even using a clock signal with a 14-ns rise time.

Several factors can cause the rise and fall times of the modulator to be longer than 1 ns in practice. When the pulse modulator is off, the power amplifier still generates a signal. This signal reflects from the modulator and then from the power amplifier again. It takes twice the time delay between the power amplifier and the modulator for the first reflection from the modulator to return to it again. This round-trip delay time is about 5 ns in the HP 8665A. When the pulse modulator switches on, the signal previously reflected from the modulator adds to or subtracts from the amplifier signal and passes through the modulator. Five nanoseconds later, the signal reflected from the modulator after it turned on adds to the amplifier signal (this analysis assumes a slow ALC loop time constant relative to the pulse switching time). The new doubly reflected signal may add to the amplifier signal with a phase and magnitude different from the previous one, causing a change in level. This new level will be the steady-state level. The result of the reflected signals adding to the transmitted signal may be a shelf lower than the steady-state voltage or a pedestal higher than the steady-state voltage. The shelf or pedestal will be about 5 ns in length, starting at the leading edge of the pulse. If the shelf is less than 90% of the final amplitude, the 10%-to-90% rise time is degraded. To avoid this problem in the HP 8665A, the amplifier match and the pulse modulator match are made good enough that the doubly reflected signal is less than 10% of the transmitted signal. Another method can also be used, that is, to design the modulator so that the reflection coefficient is the same in both on and off modes. However, this will not reduce problems caused by mismatches beyond the pulse modulator.

A similar set of reflections can occur on the output side of the pulse modulator, and will affect the trailing edge of the pulse by causing a pulse that does not die out until after twice the delay between the pulse modulator and the load. Again, both the pulse modulator match and the load match should be made as good as possible.

During the pulse rise time, the pulse modulator match is not controlled, and large reflections can result during this 1-ns period. These reflections can show up about 5 ns later as a "blip" or "dimple" on the pulse envelope, about 1 ns wide. Whether the envelope is increased or decreased depends on how the reflection adds to the carrier, which is determined by frequency, reflection coefficient, and cable length.

It is desirable for these waveform distortions to occur near the pulse edge, if they have to occur at all. This implies that pulse modulator systems such as this should be as compact as possible, with minimum cable lengths between components. In the HP 8665A, the cables are longer than optimum because of the block diagram, which uses three different power amplifiers and only one pulse modulator. However, this is not a major problem, because the power amplifier match is reasonably good.

Acknowledgments

Fred Ives did the first pulse IC design. Tom Higgins did significant further development and modeling of the IC before handing it off. Rob Dickerson also contributed to the pulse IC design. At the Microwave Technology Division, Val Peterson, Scott Trosper, Don Montgomery, and

Andy Teetzel put in many hours helping with the development and testing of the IC. Pat Herron did the mechanical design of the microcircuit package. Keith Fries designed much of the internal pulse generator. Mark Stambaugh was responsible for the firmware control and calibration algorithm of the internal pulse generator. Joyce Smith took sketchy documentation and turned a handful of parts into assembled microcircuits in record time. Dick Waite, Todd Wendle, Kimball McKeehan, and Joyce all provided support in setting up the production process and provided valuable feedback on the design. Marc Brodeur developed the production test system for the pulse modulation option. The lab procurement group was instrumental in keeping the schedule on track. The quick turnaround time from the Microwave Technology Division thin-film group also helped greatly. Special thanks are in order to Tim Halloran for marketing support, Matt Dixon for long hours of testing and troubleshooting, and Dave Platt for overall project coordination and support.

References

1. W.W. Heinz and P.A. Zander, "2-to-26.5-GHz Synthesized Signal Generator Has Internally Leveled Pulse Modulation," *Hewlett-Packard Journal*, Vol. 34, no. 5, May 1983.
2. R.K. Larson and L.A. Stark, "A Wideband YIG-Tuned Multiplier and Pulsed Signal Generation System," *Hewlett-Packard Journal*, Vol. 34, no. 5, May 1983.
3. D.R. Chambers and S.N. Sanders, "Solid-State Microwave Signal Generators for Today's Exacting Requirements," *Hewlett-Packard Journal*, Vol. 33, no. 7, July 1982.
4. J.F. Catlin, "A Wide-Dynamic-Range Pulse Leveling Scheme," *Hewlett-Packard Journal*, Vol. 33, no. 7, July 1982.
5. R.B. Collison, J.B. Summers, M.W. Wagner, and B.D. Ratliff, "Wide-Frequency-Range Signal Generator Output Section Design," *Hewlett-Packard Journal*, Vol. 36, no. 12, December 1985.
6. M. Fleischer, H. Rossnes, and U. Neumann, "A New Family of Pulse and Pulse/Function Generators," *Hewlett-Packard Journal*, Vol. 34, no. 6, June 1983.
7. C. Hentschel, A. Leiter, S. Traub, H. Schweikardt, and V. Eberle, "Designing Bipolar Integrated Circuits for a Pulse/Function Generator Family," *Hewlett-Packard Journal*, Vol. 34, no. 6, June 1983.

Reducing Radiated Emissions in the Performance Signal Generator Family

Two levels of radiated emissions are offered: one standard and one optional. The optional level, −133 dBm into a two-turn loop one inch away from any surface, is 26 dB lower than the standard specification.

by Larry R. Wright and Donald T. Borowski

RADIATED INTERFERENCE IS A COMMON PROBLEM in signal generators. The Performance Signal Generator product line described in this issue offers the user a choice of two levels of radiated emissions. The lower level, Option 010, is for extremely sensitive applications, such as testing pagers and transceivers, for which the standard level is not acceptable. The goals of the Option 010 project were:

- Maximum emissions level of 0.05 μV (−133 dBm) into a two-turn loop, one inch away from any surface (the standard level is 1 μV)
- Frequency range 0 to 1 GHz
- Leakage specified up to 0-dBm output power
- Installable on all three PSG instruments
- Field retrofitable at service centers.

It is important for HP to provide our customers with quiet, quality instruments. Some customers perform a quick, simple test using a pager to determine the shielding effectiveness of any signal generator. This test involves tuning the signal generator to the pager frequency and terminating the output connector with a suitable termination. The pager is then placed directly on every outside surface of the instrument to see if it will unsquelch. While this test may not be a realistic quantifier of overall instrument performance or quality, it is done on a regular basis. In the past, our signal generators did not consistently meet our expectations.

The original goal for the HP 8645A Agile Signal Generator was to have an RF leakage specification of 0.5 μV measured into a two-turn loop, one inch away from any surface. This is equal to −113 dBm. The first pilot unit measured approximately −115 dBm. However, the first several production units measured −100 to −115 dBm. This is too high for testing pagers and transceivers, since some transceivers have sensitivities as low as −131 dBm.

Measurement Technique

It became apparent early in the project the technique used to measure emissions was not part of the solution, but part of the problem. It is disturbing to implement a

modification to the instrument to reduce RFI and have a measurement show a contradicting result. The two-turn loop and its separate coaxial cable are subject to such things as the location of the user's hand while holding the body of the loop, the perpendicularity of the loop to the surface measured, fields affecting the side of the loop, and reflections from the user's body and other surfaces close to the loop.

To work around this problem somewhat, initially pagers were relied on for performance levels and a near-field probe (HP 11940A) was used for locating sources of leakage. The pagers ranged in frequency from 154 MHz to 932 MHz. It turns out that to maintain shielding the higher frequencies require narrower slots or openings in the instrument. As the leakage was reduced, the pagers were replaced by transceivers. The pager sensitivities ranged from −125 dBm to −128 dBm. The transceivers went down to −131 dBm.

We realized that a new measurement technique would be required not only for ourselves, but service centers and customers as well. As a result, we now are using a tuned dipole antenna system to make measurements outside of the instrument. The theory and experiments leading to this choice are discussed later in this article.

Cabinet Modifications

The results of early tests demonstrated several problems with the HP System II enclosure. Some of the trouble spots were the interfaces between the side cover and the top, bottom, and front frames, and the fit of the top and bottom covers into the front frame.

The original System II cabinet provided from 2 to 15 dB of shielding. There was not enough room to place a bulkhead with connectors between the instrument and the front panel, but by increasing the length of the instrument, room for the bulkhead was created. This also allows the cover to be attached directly to the bulkhead rather than to the front frame. This is shown in Fig. 1.

Our approach to the design of a reduced-leakage cabinet was to create a Faraday cage using a seamless outside cover, a front bulkhead, and the rear casting. All input and output outer conductors are grounded to the interfaces and filtered connectors are used where necessary.

The reduced-leakage configuration has three aspects: cover, bulkhead, and instrument extension.

The cover is a one-piece aluminum wraparound secured at the seam by two rows of staggered spotwelds. The lap width, spacing, and staggering of the spotwelds were selected to maintain high shielding up to 4 GHz. The spacing is kept within approximately 1/30 of the wavelength being generated. The one-piece cover eliminates four seams in the System II design.

The interface between the cover and the rear casting requires a clearance between the folded edge and the casting. This significantly reduces friction during installation and removal. Twisted strip RFI gasketing is installed at the inside rear of the cover to make contact with the casting.

An expanded aluminum mesh is required in the ventilation perforation pattern to meet the leakage objectives. The size of the mesh openings is approximately 2.5 mm by 2.5 mm. The mesh is held in place with a 1.6-mm-thick frame that is secured with rivets to the cover. The rivet spacing is not critical. The cover is spotfaced to remove the paint for positive grounding of the rivets. The mesh selected is readily available and worked the first time it was installed.

Experiments with cover attachment hole spacing led to the present spacing, which meets the leakage objectives. Hole spacing at 50.0 mm was not sufficient. The current spacing is 25 mm.

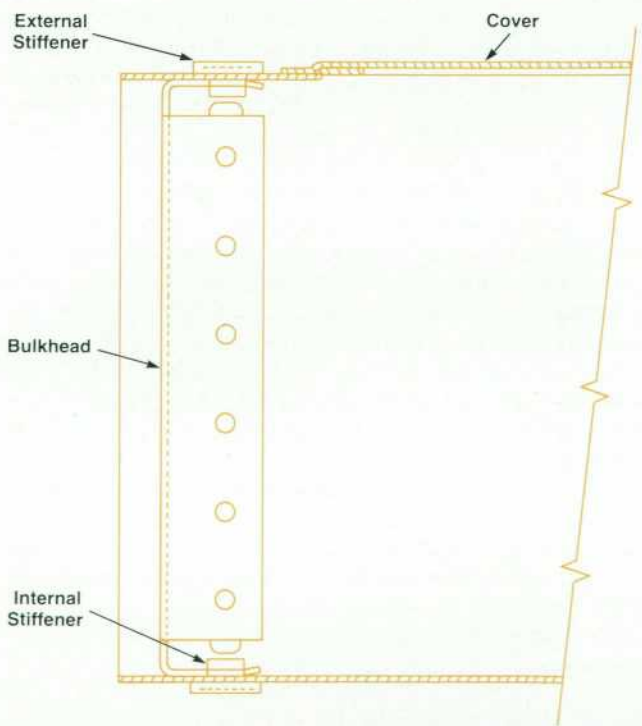


Fig. 1. *For Option 010 (very low RFI emissions), a front bulkhead was added to the PSG instruments by increasing their length.*

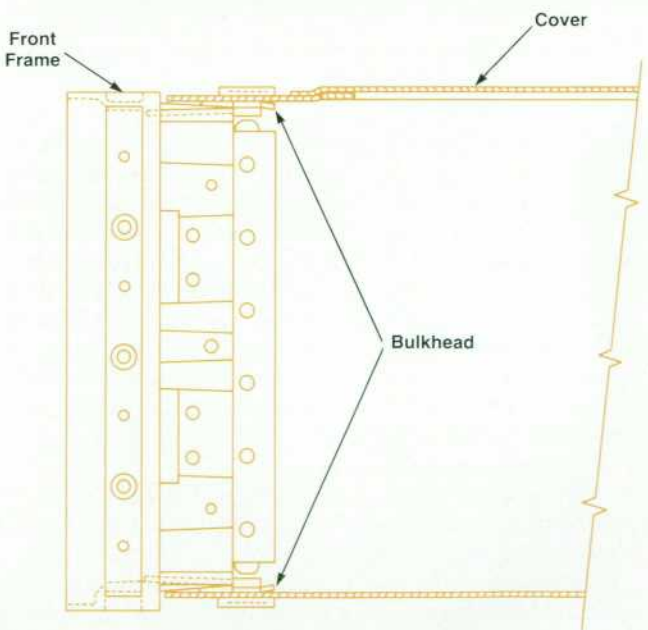


Fig. 2. *New Option 010 interface between the cover and the front casting.*

The cover does not engage the front casting groove as is normal in System II cabinets. This is for two reasons. First, the cover has to be large enough to clear the rear casting during installation. This led to an increase in size, so that the cover is now larger than the groove opening. Second, the original interface between the cover and the front casting proved unreliable. The new cover interface is shown in Fig. 2. The bulkhead makes contact with the top and bottom of the wraparound cover and with the inside surface of each extender plate. The top and bottom of the bulkhead have internal tapped stiffeners that act in the same way as the external stiffeners. These stiffeners increase the pressure area when clamping two thick sheet-metal pieces together. The sides of the bulkhead are attached to the extender plates with screws every 25 mm to match the spacing found necessary on the top and bottom of the cover.

The five flexible coaxial cables to the front panel use SMC bulkhead feedthrough connectors. Only one cable was particularly noisy, measuring approximately −120 dBm, but all are grounded. This design allows the existing cables to be used from the front panel to the bulkhead.

The ribbon cable to the front panel radiated at a level of −125 dBm. To reduce this, a filtered feedthrough "D" subminiature connector is used. A 400-to-600-pF pi filter was selected so as not to interfere with the signal rate required for the front panel assembly.

The semirigid output cable also requires grounding at the bulkhead. The two problems here were the size of the hole in the bulkhead to permit passage of the SMA connector and the uncertainty of the location of the bulkhead for a rigid grounding point. A plate with a small clearance hole is assembled onto the cable with flexible copper braid soldered to the plate and cable. This braid allows the plate to move up against the bulkhead where it is securely grounded with three screws. The plate and cable interface is shown in Fig. 3.

The same twisted strip RFI gasketing is used in the corners on the top and bottom of the bulkhead to ensure proper and continuous grounding along those surfaces. Because the cover does not make contact to the bulkhead in each corner, gasketing is required. Fig. 4 shows an example of cover clearances in the corners.

The length of the cabinet was increased primarily to allow room for the bulkhead and feedthrough connectors. The extender plate, which does this, also provides the mounting surfaces to attach the sides of the wraparound cover. The grounding of the extender plate to the System II frame is not required to be an RFI-tight fit. The RFI seal is between the cover, the extender, and the bulkhead. This is shown in Fig. 5.

It turned out that the rear of the instrument radiated at about −125 dBm. The main contributors were the line cord module, the HP-IB and RS-232-D connectors, and the fan grill. The easiest to fix was the fan grill opening, which is part of the rear casting. Since the instruments are almost always mounted in a rack and testing is performed in front of the instrument, we reduced the radiated emissions specification for the rear. The benefit of fixing the fan grill and the HP-IB and RS-232-D connectors was not worth the investment.

One-Microvolt Project

Once we had a handle on the very low-leakage Option 010 instruments, our focus changed to the standard production instruments. The specification for production instruments was −113 dBm, but the first production units measured between −100 and −115 dBm. This prompted a serious review of the performance requirements. It was found that most applications could tolerate more leakage, so the radiated RFI specification for the HP 8644A and 8665A signal generators was changed to 1 μV (−107 dBm). The HP 8645A specification was reduced to −97 dBm.

The 1-μV project concentrated more on producibility and performance margin. We knew from the Option 010 project that the System II covers would have to be replaced. A two-piece clamshell cover eliminates the cover side seams. The two pieces are connected together by five screws on each side under the side handle straps. The front sides of the covers are secured to two small plates attached and grounded to the front casting. These small plates are required because of the painted irregular surface of the cast front frame. These are shown in Fig. 6. The bottom cover attaches directly to the bulkhead, like the Option 010 ver-

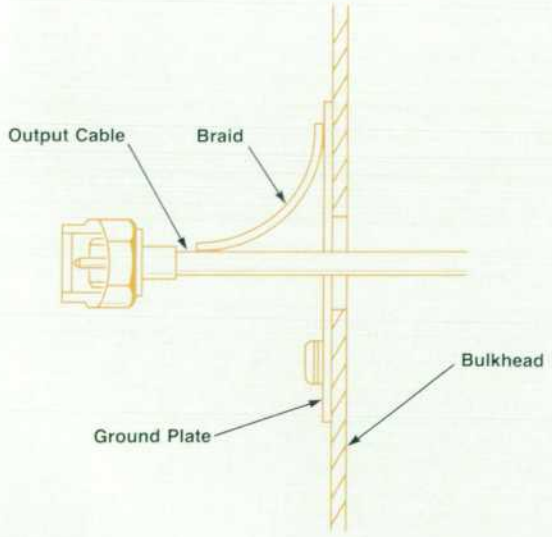


Fig. 3. *Coaxial cables to the front panel are grounded to the bulkhead for Option 010.*

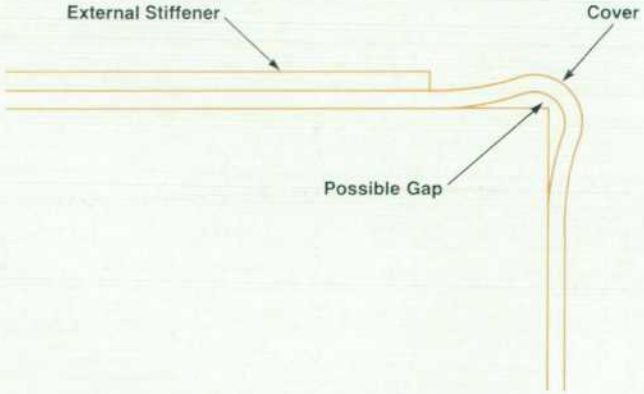


Fig. 4. *Gasketing is placed in each corner where the cover would otherwise not contact the bulkhead.*

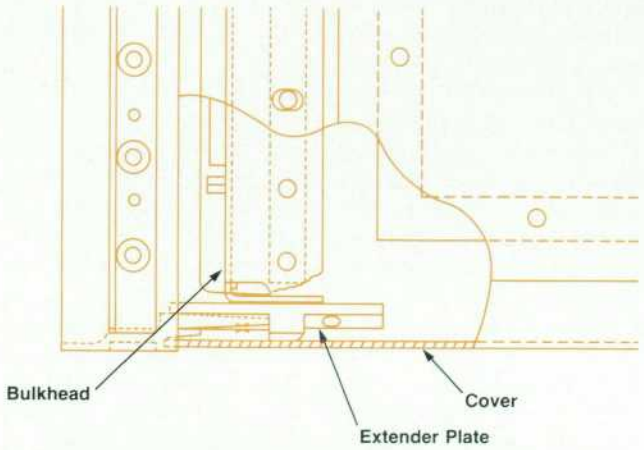


Fig. 5. *Detail of the interface between the cover, extender plate, and bulkhead in the Option 010 instruments.*

sion, but without the stiffeners and close screw spacing. The top cover engages the front frame as usual with a strip of round mesh RFI gasket inserted into the casting. The rear feet drive the cover into the gasket before the side and bottom screws are tightened. A cross-sectional view showing the bulkhead, covers, and front frame is shown in Fig. 7.

The bulkhead turned out to be a simpler version of the Option 010 bulkhead. The flexible coaxial cables go through 12-mm-clearance holes and the semirigid output cable is grounded directly to the bulkhead using the same copper braid. A filtered connector is not required; instead, the ribbon cable is secured between a small flat plate and the bulkhead. The close proximity to the bulkhead drains off some of the conducted noise from the instrument interior. It might have been possible to eliminate the bulkhead if the front panel had been sufficiently grounded on all four sides. However, this would have required a large revision to relocate the inputs and outputs and revise the two rubber keypads. The main problem was that both the plastic front panel and the keymats were already hard-tooled.

During this redesign effort, it was found that the internal instrument radiation level was increasing. The specified −100-dBm level was drifting towards −90 dBm on a few instruments that had been around for a while. The reason turned out to be the corrosion of the interface of the metal RF gasket and the aluminum internal module casting and aluminum covers. While the gaskets were tin-plated beryllium and should have been galvanically inactive, they were not. The gaskets are tin-plated and then slit to width, leaving the bare copper edges exposed. The result is galvanic corrosion. Plating the gaskets after slitting turned out to be expensive, so the casting and covers are both tin-plated. Special attention was given to the possibility of tin whisker growth to assure ourselves that problems would not appear in the future.

Tuned Dipole Antenna System

As explained earlier in this article, the testing of mobile radios and pagers places stringent limits on radiated emissions from signal generators. It was found that the two-turn loop antenna formerly used to measure emissions close to the generator gave inconsistent results. These measurements are now made with a tuned dipole antenna system.

Background

The radio regulatory agencies in various countries, for example the Federal Communications Commission in the United States, have set limits on the electric field emissions from electronic equipment. HP has generated a field strength specification that ensures compliance in practically any country by combining the most stringent requirements of all these countries into one field strength curve.

Many mobile two-way radios and pagers operate in the frequency range of 100 MHz to 1000 MHz. In Fig. 8, the

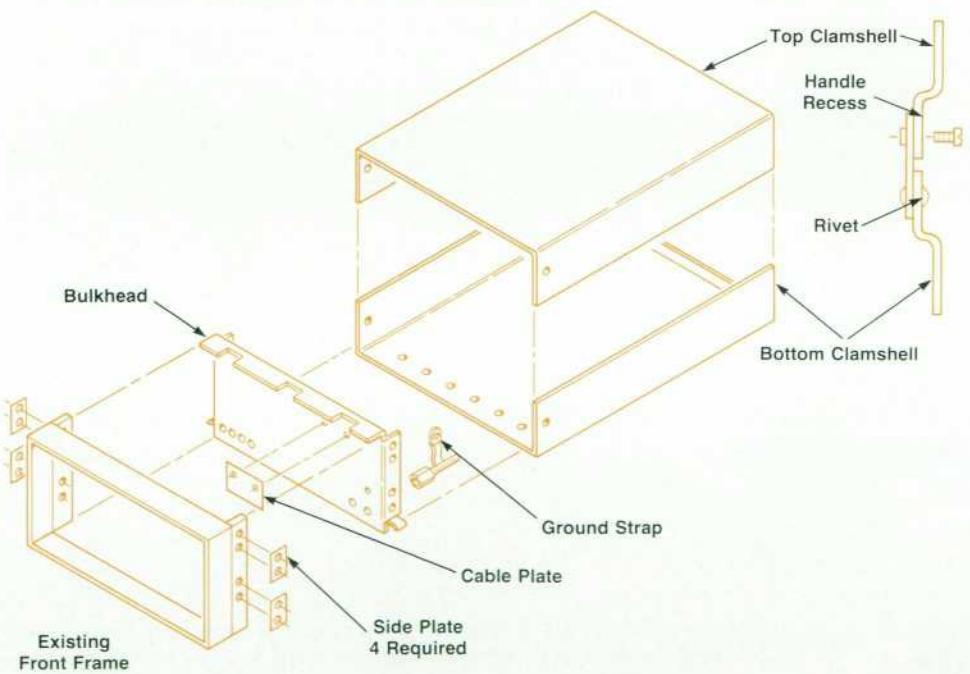


Fig. 6. *New cabinet design for the standard PSG instruments.*

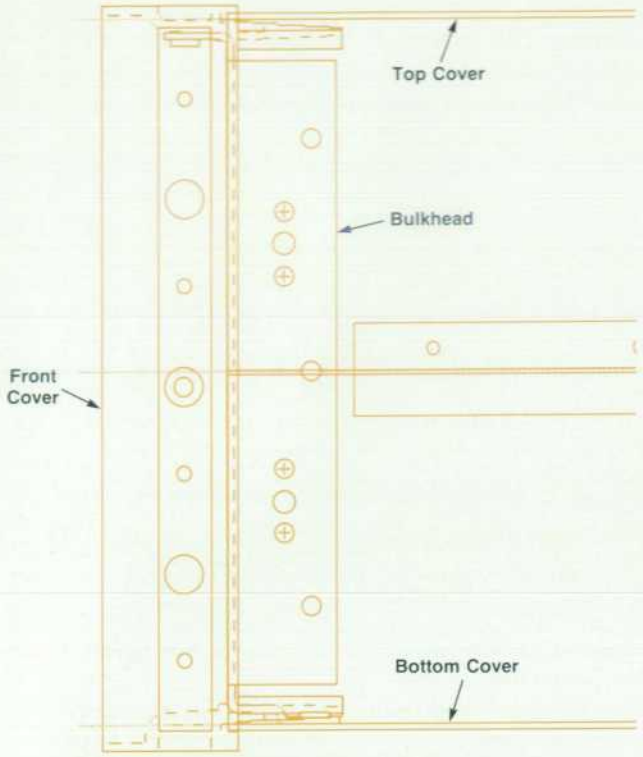


Fig. 7. *Side view of the standard PSG bulkhead, covers, and front frame.*

portion of the HP radiated emissions field strength curve for 100 MHz to 1000 MHz is shown. The measurement distance is 10 meters away from the electronic product. Fig. 8 also shows typical sensitivities (0.25 μV into 73 Ω) of mobile radios in this frequency range, assuming a dipole antenna, or equivalently, a monopole or whip antenna.

Fig. 8 shows that even at 10 meters away, emissions from equipment passing the regulatory requirements for RF leakage can be detected by these radios quite easily. Radio tests are often carried out at distances much closer to the signal generator, usually less than one meter away. This further compounds the problem.

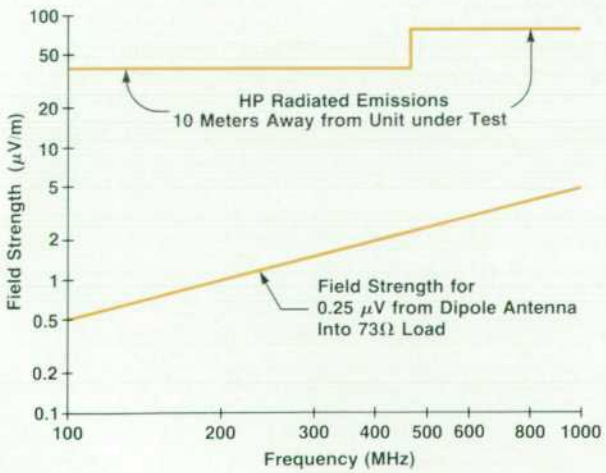


Fig. 8. *HP radiated emissions specification for regulatory requirements compared with typical sensitivity of a mobile radio.*

The signals leaking (radiating) from a signal generator are of two types. The first type, familiar to many home computer users, is the broadband spectrum generated by the clock and switching pulses of the digital circuitry. These are generated by the computerlike controller built into modern signal generators. Similar spectra are caused by switching power supplies and scanned displays. This broadband noise can cause continuous interference to radio tests at all frequencies of interest. The second type of leakage is that of the internally generated signal used in the radio test, which oftentimes is the worst offender because of its high amplitude inside the signal generator. This signal should leave the signal generator by the RF output connector only, but it can also escape by a radiated leakage path. If the radio under test is not well-shielded, or if the radio's covers are off for test purposes, the leakage signal can enter the radio along with the desired signal from the RF output connector. The two signals can add either constructively or destructively, resulting in a total signal of unknown amplitude. This reduces the accuracy of the test.

Signal generators, like other electronic products, are tested to the HP radiated emissions specifications. Since this level of emissions can be much higher than tolerable for radio tests, leakage measurements are made closer to the unit to gain additional sensitivity. The antennas used for regulatory emissions tests are usually quite large and heavy, making them unsuitable for close, handheld measure-

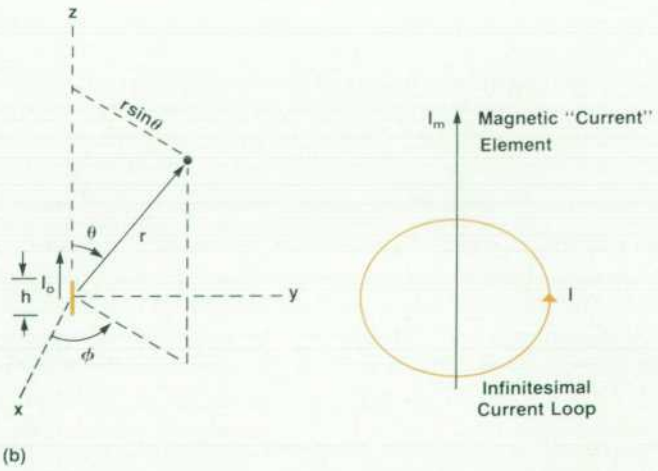
$(I_m \leftrightarrow I_o)$

$(-E_\phi)$ $$H_\phi = \frac{I_o h}{4\pi} e^{-jkr}\left(\frac{jk}{r} + \frac{1}{r^2}\right)\sin\theta$$

(H_r) $$E_r = \frac{I_o h}{4\pi} e^{-jkr}\left(\frac{2\eta}{r^2} + \frac{2}{j\omega\epsilon r^3}\right)\cos\theta$$

(H_θ) $$E_\theta = \frac{I_o h}{4\pi} e^{-jkr}\left(\frac{j\omega\mu}{r} + \frac{1}{j\omega\epsilon r^3} + \frac{\eta}{r^2}\right)\sin\theta.$$

(a)



(b)

Fig. 9. *(a) Equations for the field components of an electric field point source (infinitesmal dipole antenna). Field components for a magnetic field point source (infinitesmal loop antenna), shown in parentheses, are found by substituting I_m for I_o in these equations. (b) Definitions of I_o and I_m.*

ments. Traditionally, we have measured radiated leakage from our signal generators using an RF "sniffer" antenna—a two-turn loop, 25 mm in diameter, constructed with a 25-mm spacer to set the measurement distance from the signal generator cabinet. As the sensitivity of the radios under test has improved and their frequency of operation gone up, the loop antenna has become inadequate. For the same reasons, the conventional equipment cabinet has become inadequate, as explained earlier. A new approach was needed to measure the leakage from signal generators.

Investigation and Experiments

An investigation into various antennas for RF "sniffing" was made. These antennas are designed to be broadband, which is to their advantage. However, they suffer from low sensitivity. They are designed to find leakage that could cause the unit under test to fail the regulatory emissions standard. As mentioned above, this level of leakage is orders of magnitude higher than our requirements. Both electric field and magnetic field antennas are available, and measurements may have to be done twice at each frequency because of the close distance.

We decided to retain our traditional 25-mm measurement distance for RF leakage. This decision presents the problems of a near-field source. At a distance beyond several wavelengths from a radiating source in free space, that is, in the far field, the magnitudes of the E field and the H field have a constant ratio. That ratio is 377 ohms, the impedance of free space. Close in, in the near field, their ratio is not constant, but depends on the distance and the nature of the radiating source.

The equations of the field components for an electric field point source are shown in Fig. 9a. An infinitesimal transmitting dipole antenna is an electric field point source. If the terms in parentheses are substituted into the equations, and I_m substituted for I_o, the field components of a magnetic field point source are described by the equations. An infinitesimal transmitting loop antenna is a magnetic field point source. The I_m term is a magnetic current element analogous to the electric current element. It is oriented with respect to the infinitesimal current loop as shown in Fig. 9b.

The terms that remain significant in the far field are those that vary as 1/r. The terms that vary as $1/r^2$ and $1/r^3$ die out rapidly. Thus the radial terms E_r and H_r are not present in the far field.

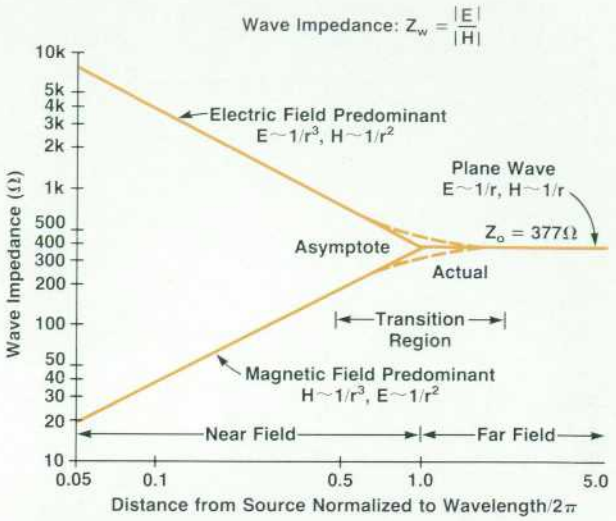


Fig. 10. *Wave impedance as a function of distance and type of source, showing near-field and far-field regions.*

In Fig. 10, the wave impedance, that is, the ratio E/H, is shown as a function of distance and type of source, neglecting the radial field components. For point sources, the transition between near field and far field can be put at a distance equal to $\lambda/2\pi$, where λ is the wavelength. For a distance of 25 mm, measurements below 1.9 GHz are in the near field.

In the far field, the receiving sensitivity of an antenna is not a function of the nature of the source, since the ratio of E to H is a constant. In the near field, however, this is not the case. A loop and a dipole with equal sensitivities in the far field will respond differently in the near field. The loop will be more sensitive to magnetic sources, while the dipole will be more sensitive to electric sources. This presents some practical problems. Leakage measurements must be done twice, once with a dipole antenna and again with a loop.

Ideally, we would like to have just one antenna to make a measurement at any given frequency. Therefore, we did several experiments to see if a resonant dipole would serve.

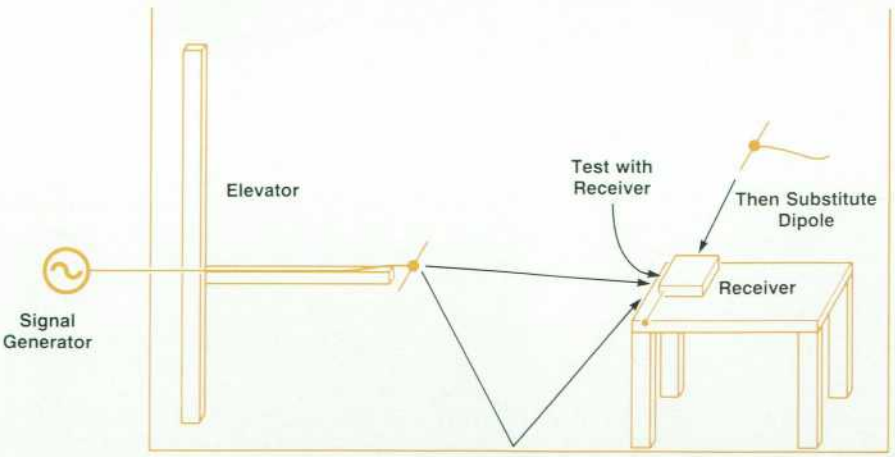


Fig. 11. *Setup for antenna tests.*

The tests were performed in a semianechoic chamber. The chamber has RF absorbers on all four walls and the ceiling. The floor is a reflecting surface. The purpose is to simulate, in an enclosed area, free space above a ground plane such as might be found outdoors.

A 500-MHz radio with a monopole antenna was placed on a wooden table about one meter above the floor (see Fig.11). A 500-MHz dipole to transmit the signal was placed on an antenna elevator at a horizontal spacing of about 1.5 meters. This dipole was fed by a signal generator. The vertical position of the dipole was adjusted with the elevator to maximize the signal received by the radio. The signal strength was then reduced until the signal received by the radio was at the threshold of audibility. A 500-MHz dipole to receive the signal was then substituted at the position of the radio. The strength of the signal received by the dipole was measured with a spectrum analyzer.

Next, the signal generator was connected to a 500-MHz circuit-fed resonant slot antenna. A slot antenna is the magnetic field radiator equivalent of a dipole antenna, which is an electric field radiator. The slot antenna simulates a cabinet seam leaking RF energy. The position of the radio and its antenna relative to the the slot was adjusted for maximum signal, maintaining the 25-mm distance with a spacer from the plane of the slot antenna. When the maximum was found, the signal generator output was reduced until the signal received by the radio was at the threshold of audibility. A 500-MHz dipole was then scanned over the surface of the slot antenna, also maintaining a distance of 25 mm with disk-shaped spacers placed on the two elements of the antenna. A spectrum analyzer was used to monitor the amplitude of the received signal. The amplitudes of the signals received on the spectrum analyzer for the first and second parts of the test were compared and were found to be within about 1 dB.

This test shows that a resonant dipole antenna can be calibrated in the far field to match the sensitivity of the radio. This calibration holds in the near field. Thus leakage measurements made with a dipole will ensure that radios can be tested successfully without loss of accuracy from RF leakage.

One additional experiment was performed with a resonant dipole in the near field. The 500-MHz slot antenna was fed with a 0-dBm signal at 500 MHz. The 500-MHz dipole with the spacers was scanned across the slot antenna. The signal peaked with the dipole perpendicular to the slot, centered along the length of the slot. The measured signal level was −3 dBm, only 3 dB lower than the signal fed to the slot antenna. Coupling efficiency this high was not expected. We believe the mechanism is capacitive coupling. A potential difference exists across the slot of the slot antenna, and is maximum at the center along the length of the slot. With the dipole positioned as previously described, the capacitive coupling of this potential difference is maximized. A radio with antenna placed across this slot would also experience the same sort of capacitive coupling. This is consistent with the radio test described above.

Acknowledgments

A tip of the hat to Bill Gaines for helping with the one-microvolt clamshell design.

Authors

October 1989

14 High-Performance Signal Generators

Michael D. McNamee

Mike McNamee joined HP's Stanford Park Division in 1978, after receiving his BSEE degree from the University of Illinois. He worked as a marketing and production engineer on the HP 8662A Synthesized Signal Generator while earning an MS degree in engineering management from Stanford University through the HP Honors Co-op program. In 1981, he became a development engineer at the Spokane Division, where he contributed to the design of the HP 8901B Modulation Analyzer, the HP 8902A Measuring Receiver, and the HP 8644A Synthesized Signal Generator, for which he was also the project manager. He has authored trade journal articles on phase noise measurement and low-level signal detection, and is named coinventor in a patent on a digital frequency synthesis method used in the HP 8904A. Mike grew up in Chicago and lives in Spokane, Washington with his wife and two-year-old son. He plays first violin with a local symphony orchestra. He is also a jazz drummer, an amateur radio operator, and an avid fisherman.

David L. Platt

Dave Platt's career at HP began in 1972, after he received his MSEE degree from the University of Wisconsin. He contributed to the design of synthesizer and network analyzer products and served as the R&D project manager for the HP 8663A Synthesized Signal Generator. Dave developed S-NODE, a widely used circuit analysis program, and he authored an article in the February 1981 issue of the HP Journal describing the output section of the HP 8662A. He is now project manager for a similar project. Dave, who is a Corvette enthusiast and enjoys golf, is married, has two sons, and lives in Veradale, Washington. He holds a professional engineering license in California.

20 Signal Generator Firmware

Brian D. Watkins

Brian Watkins has been a development engineer for HP since he received his BSEE degree from Colorado State University in 1984. As a firmware writer on the Performance Signal Generator project, he developed firmware for the HP 8645A Agile Signal Generator, including the fast-hop level control, synthesis loop, instrument state object (ISO), and calibration and diagnostics. Although his specialty is microprocessor-based design, Brian is also interested in embedded systems, structured analysis and design, and signal processing. Currently, he is designing controller boards for a modular measurement system. He divides his free time among golf, scuba diving, and skiing. He is single and lives in Spokane, Washington.

Kerwin D. Kanago

For two years, Kerwin Kanago was chief engineer at radio station KICD in his hometown of Spencer, Iowa. He developed software that allows radio stations to use personal computers to monitor, store, and edit information from news and weather wire services and draw weather maps from the data. He came to HP in 1984 and later joined the Performance Signal Generator project as firmware project leader. Kerwin received his BSEE degree from Iowa State University in 1982. He is married, has two young daughters, and lives in Veradale, Washington. In addition to his hobbies of ham radio, bicycling, and bowling, he has acquired a large collection of old comic books, including over 600 Spider Man volumes. He is currently project manager for the source section of a new product.

Mark A. Stambaugh

After receiving his BSEE degree from the University of Cincinnati in 1981 and his MEE degree in computer engineering from Rice University in 1987, Mark Stambaugh joined the HP team developing the HP 8665A Synthesized Signal Generator. As a firmware engineer, he was responsible for the diagnostic, calibration, and output sections. Before coming to HP, Mark worked for six years in microprocessor design at Texas Instruments, where he developed semiconductor circuits and architectures described in five patents; a sixth patent is pending. His paper titled "A VLSI Communication Processor Designed For Testability" was published in the International Solid-State Circuits Conference proceedings in 1984. He is married, lives in Veradale, Washington, and spends much of his spare time fishing, backpacking, skiing (water and snow), and bicycling.

27 Single-Loop Frequency Synthesis

Earl C. Herleikson

Until May of this year, Earl Herleikson was a development engineer in HP's Spokane Division, working on the HP 8644A, HP 8645A, and HP 8665A family of signal generators. A transfer took him to McMinnville, Oregon, where he holds a similar position in the development of cardiology equipment. Earl has written for the HP Journal in the past, describing the HP 8642A Synthesized Signal Generator in 1985. He has been employed at HP since he received his BSEE degree from the University of California at Davis in 1980. He recently received notification that his application for a patent on a calibrated high-bandwidth white-noise reference has been approved. He is married, has five children, and lives on a small ranch in Yamhill, Oregon.

Brad E. Andersen

Camping and bicycling enthusiast Brad Andersen was responsible for the voltage-controlled oscillator and electronic attenuator of the HP 8644A Synthesized Signal Generator and the HP 8645A's fast-hop level control. He came to HP in 1984 after spending over five years at AT&T Bell Laboratories, where he developed a control for high-power DC-DC converters. He is currently working on the development of transceiver test products at the Spokane Division. Brad received

his BSEE degree in 1977 and MSEE degree in 1978 from Washington State University. He is coauthor of an IEEE conference paper on DC-DC converter operation, and is named coinventor of a patent on a current-limit technique for switching power supplies. He grew up in Oregon and southeastern Washington and makes his home in Spokane, Washington. He is married, has two sons, and is active in his church. He spends some of his spare time restoring and driving a 1969 MGB-GT sportscar. He is a member of the IEEE.

34 Fast Hopping VCO

Barton L. McJunkin

So enthusiastic is Bart McJunkin about horse racing that he traveled to Australia last year to attend the Melbourne Cup race. Born in Billings, Montana, he joined HP as an R&D engineer immediately after receiving his BSEE degree from Montana State University in 1981. He has devoted much of the past eight years to working on the fractional-N module and fast voltage-controlled oscillator for the performance signal generator family. Bart is married, lives in Spokane, Washington, and spends his spare time playing golf, bowling, reading a wide variety of books, and, of course, going to horse races.

David M. Hoover

Development of the synthesis section, phase shifter, and digital control of the HP 8645A Agile Signal Generator was Dave Hoover's major contribution to the Performance Signal Generator project. Born in Seattle, Washington, he began working for HP after receiving his BSEE degree (1980) and his MSEE degree (1981) from the University of Washington. As an R&D engineer, his expertise in radio frequency and analog design has yielded a patent on a concept of fast-hop signal generators using frequency-locked loops. Dave is a scuba diving, skiing, golfing, and backpacking enthusiast. He and his wife live in Newman Lake, Washington in a house he built himself. They are expecting their first child in October.

37 Microwave Frequency Synthesis

Douglas R. Snook

As a development engineer, Doug Snook has been contributing to the development of synthesized signal generators since he began his HP career in 1981. His most recent work has been on the GaAs frequency divider and internal pulse generator of the HP 8665A Synthesized Signal Generator. Before that, he was responsible for test system development for the HP 8663A Synthesized Signal Generator. A graduate of Washington State University, he has a BSEE degree (1980) and an MSEE degree (1981). Doug spends much of his free time bicycling, skiing (downhill, cross-country, and water), playing piano and guitar, and singing. He is married to a former HP engineer, lives in Veradale, Washington, and has two children.

James B. Summers

Jim Summers has worked for HP since he received his MSEE degree from the University of Wisconsin at Madison in 1979. Analog and radio frequency circuitry are his professional specialties. He originated FILSAP, a filter synthesis/analysis computer program. He is also the coauthor of an article on the HP 8642A low-noise Synthesized Signal Generator, which appeared in the December 1985 issue of the HP Journal. A native of Michigan City, Indiana, Jim is an amateur radio operator who enjoys fishing, golf, and woodworking in his spare time. As development engineer on the HP 8665A Synthesized Signal Generator, he was responsible for the synthesis section. He is single and lives in Spokane, Washington.

42 Output System Design

Steve R. Fried

As an R&D electrical engineer at HP's Spokane Division, Steve Fried has worked on various aspects of synthesized signal generator design. His contribution to the HP 8665A Synthesized Signal Generator was primarily in the heterodyne output section. Steve was a SEED student at HP for two summers, while he earned his BSEE degree from Montana State University. He graduated in 1986 and joined HP full-time in January 1987. Born in Lewistown, Montana, he is married and makes his home in Spokane, Washington. His outside interests include fishing, playing racquetball and softball, and teaching a church youth group.

Keith L. Fries

Working on the HP 8665A Synthesized Signal Generator since its inception, Keith Fries' contributions have been in the development of the high-frequency driver and the output gain budget. Before that, he contributed to the design of the HP 8663A Synthesized Signal Generator. Keith is a graduate of the University of Illinois; he holds a BSEE degree (1978) and an MSEE degree (1980). He joined HP in 1979 and spent a year at the Stanford Park Division working on the HP 8901B Modulation Analyzer before moving to Spokane, where he now lives. He is a member of the IEEE and the National Society of Professional Engineers. A native of Chicago, he enjoys photography and hiking and is an amateur radio operator.

John M. Sims

John Sims's eleven-year career at HP has included work on numerous synthesized signal generator projects, including the HP 8665A. His contributions to this project were in the development of the output module, the microwave extender module, and the reverse power protection system. John received a BSEE degree from the University of California at Davis in 1978 and an MSEE degree from Stanford University in 1984. He was born in Mercedes, Texas, grew up in California, and lives in Liberty Lake, Washington. His interests outside of work include boardsailing, aerobics, running, and downhill and cross-country skiing. He is a recently qualified scuba diver and occasionally pursues his interest as an amateur radio operator (N6AB).

51 Pulse Modulation System

Douglas R. Snook

Author's biography appears elsewhere in this section.

G. Stephen Curtis

When he is not fixing up an old farmhouse he recently purchased, Steve Curtis is a manufacturing development engineer at the Vancouver Division, supporting the HP RuggedWriter 480 printer. Steve is a native of Bradenton, Florida, and attended Florida Atlantic University, where he earned a BSEE degree in 1981. He applied his expertise in low-noise design and frequency synthesis to designing the low-phase-noise frequency reference in the HP 8665A Synthesized Signal Generator and contributed to the pulse modulator integrated circuit and microcircuit. He has also worked on the optional discriminator design used in the Performance Signal Generator project. Since joining HP in 1981, Steve has written a symposium paper on the relationship between, and measurement of, resonator and oscillator phase noise and is named as inventor in a patent for FM calibration in a phase-locked loop. Another patent is pending for his concept of a circuit to reduce phase noise in frequency synthesizers. He lives in Vancouver, Washington, and enjoys backpacking, running, and bicycling.

59 Radiated Emissions

Larry R. Wright

After spending 14 years in the aerospace industry and attending college part-time, Larry Wright earned his BSME degree from the University of Colorado and joined HP's Loveland Instrument Division in 1974. He is the author of a technical paper on the properties of cesium iodide and coauthored an article in the February 1989 issue of the HP Journal describing the HP 8904A Multichannel Synthesizer. He was a product designer on the HP 8665A Synthesized Signal Generator. Born in Barberton, Ohio, Larry is married, has five children, and lives in Post Falls, Idaho. He spends his spare time building, driving, and enjoying a 1931 Ford coupe, which he calls a "chopped street rod" and which he drives to work every day, weather permitting.

Donald T. Borowski

Don Borowski bicycles over twelve miles every day (weather permitting) to his job as a quality and reliability engineer at the Spokane Division of HP, where he works on low-level radio frequency leakage measurements and other projects. Previously, as a lab engineer, he contributed to the design of the HP 8662A, 8663A, and 8645A Synthesized Signal Generators. He began his career at HP's Stanford Park Division in 1975 and still visits friends every year in Palo Alto, where he enjoys performing Gregorian chant at a local church during the Easter season. A member of the IEEE and the National Society of Professional Engineers, Don wrote an article in the February 1981 edition of the HP Journal, describing his work on the frequency doubler for the HP 8662A. He was born in Pulaski, Wisconsin, and received his BSEE degree in 1973 from the University of Wisconsin. In 1975, he earned his MSEE degree at the same school after spending two months in Antarctica testing his research project, a 150-MHz radar, which is still used by the university to measure the thickness of glaciers. An amateur radio operator and collector of vintage electronics, Don also is his own auto mechanic.

69 Mesa Photodetectors

Susan R. Sloan

Susan Sloan is an R&D engineer at the Microwave Technology Division, where she has worked since joining HP in 1985. She is currently working on process development, testing, and manufacturing of photodetectors. She coauthored an earlier article for the HP Journal on the topic of diode integrated circuits and coauthored an article for Applied Physics Letters describing planar doped barrier devices which act as field-effect transistors. A native of Madison, Wisconsin, Susan resides in Petaluma, California. Her outside interests include skiing, hiking, reading, and writing fiction which she hopes to publish someday. She has a BS degree in physics and English from the University of Wisconsin (1981), and an MS degree in physics from the University of Colorado at Boulder (1984).

76 Driver Debug Technique

Eve M. Tanner

Pennsylvania was originally home for Eve Tanner, and she stayed there long enough to earn her BA degree (1970) and MEd degree (1973) from Pennsylvania State University. She joined HP in 1986 and is now a development engineer at the Personal Computer Group. In this issue of the HP Journal, she has written about her work on HP-UX software tools. Previously, she worked on documenting the developer version of the HP NewWave environment. Eve's primary interests outside of work are her two children, but she still finds time for her hobbies of N-scale trains and hiking in California.

81 Solder Joint Inspection

Catherine A. Keely

As a development engineer at HP Laboratories since 1985, Cathy Keely has applied her expertise in the field of optics, sensors, and lighting to the computer vision project. She earned her BS degree in applied mechanics at the University of California at San Diego in 1981 and her MS degree in mechanical engineering at the University of California at Berkeley in 1985. Before joining HP, she worked at McDonnell Douglas, where she specialized in videodisc research. She has contributed two papers on optical techniques for inspection at professional conferences and currently has two patent applications being considered, one on optical techniques and one on the concept described in her article in this issue of the HP Journal. A resident of Cupertino, she is an outdoorswoman with interests in mountain activities, skiing, and soccer.

86 HP-UX Shared Libraries

Anastasia M. Martelli

Since joining HP in 1983, Stacy Martelli's responsibilities have included the PORT/HP-UX emulation package, mainly in file system emulation. Process management and virtual-memory management for the HP-UX operating system have been the most recent focus of her work. In an earlier position at Standard Oil Company of California, she worked in computer services support. A native of Seattle, Washington, she received her BS degree in computer science from Washington State University in 1983. She currently lives in Santa Clara, California. She enjoys sailing, camping, needlework, and singing with the HP choir.

90 User-Centered Application Definition

Lucy M. Berlin

Lucy Berlin has been involved in improving programmer productivity since she joined the HP Laboratories as an R&D engineer in 1984. She is part of the team that defined and carried out the user-centered application definition process described in this issue of the HP Journal, and later participated in the development of the hypertext prototype. Her earlier responsibilities included the user interface for MicroScope, an experimental program analysis toolset, and the execution monitor tool of the HP Common Lisp Development Environment. Lucy has a BA degree in physics and computer science from Queen's College, New York (1981) and an MA degree in computer science from Stanford University (1983). She is a member of IEEE and ACM/SIGCHI. She was born in Prague, Czechoslovakia, and lives in Sunnyvale, California. For recreation, she enjoys working on her house and choral singing.

98 Reflective Light Guides

Carolyn F. Jones

Carolyn Jones is a technical advisor in the R&D product development department at HP's Optoelectronics Division. In this position, she is responsible for new product development for electrophotographic applications, including the evaluation of opportunities and recommendations for direction, with specific emphasis on packaging materials, interconnections, and optical interfaces. She earned an AB degree in physics from the University of California at Berkeley in 1964 and an MS degree in solid-state physics from Tufts University in 1966. Before starting her career at HP in 1975, she had ten years of experience in hybrid optoelectronics, most recently at Monsanto's Electronic Special Products Division. She originated a patent on the optical design concept described in this edition of the HP Journal. She is active in several professional organizations, including the International Society for Hybrid Microelectronics, where she served two terms on the national executive council. She is a member of the IEEE and the Society of Women Engineers. Carolyn is married, has a two-year old son, and lives in Menlo Park, California. Her spare time is devoted to various creative endeavors ranging from handicrafts to fine arts. Her oil paintings and drawings have been exhibited at the HP corporate headquarters.

Processing and Passivation Techniques for Fabrication of High-Speed InP/InGaAs/InP Mesa Photodetectors

Proper surface preparation and a conformal mesa passivation covering are critical to the production of low-dark-current photodiodes. The best results have been obtained with a wet chemical etch followed by double-layer polyimide passivation.

by Susan R. Sloan

FIBER OPTIC COMMUNICATION SYSTEMS are rapidly growing. Optical fibers provide a low-loss, low-dispersion medium for transmitting data at high speeds (>1 Gbit/s). The lowest optical losses occur for 1.3 and 1.55 μm wavelengths, and these have become the wavelengths of choice for fiber optic systems. High-speed InP/InGaAs/InP photodetectors optimized for these wavelengths have helped HP make a contribution in the lightwave instrument field. These pin detectors are a key element in HP's new lightwave receivers.

The pin photodetectors are designed for front-side illumination with light at wavelengths between 1.2 and 1.6 μm. Their low-capacitance mesa structure provides good response to modulation frequencies beyond 22 GHz. The 25-μm-diameter devices show responsivities greater than 0.9A/W at 1.3 μm and 1.55 μm, capacitance values of 0.07 pF (5×10^{-9} F/cm^2), and low optical reflection (less than 2% at both wavelengths). Diode lifetimes are calculated using high-temperature operating life (HTOL) data and extrapolating for ambient instrument temperatures. Lifetimes of 3.5×10^5 hours at 55°C are typical.

Low detector dark current is important for low noise in lightwave receiver applications. Photodetector dark currents at -5V of 0.15 nA (1.1×10^{-5} A/cm^2) after mesa etch, and 1 nA (5×10^{-5} A/cm^2) after fabrication is complete, have been achieved. Dark current is a measure of the leakage current, or reverse current of the diode at a given bias without illumination (see Fig. 1). The surface preparation and passivation of the photodetector mesa walls determine the dark current of the device. This study will examine different methods of surface preparation and surface passivation aimed at achieving low-dark-current devices.

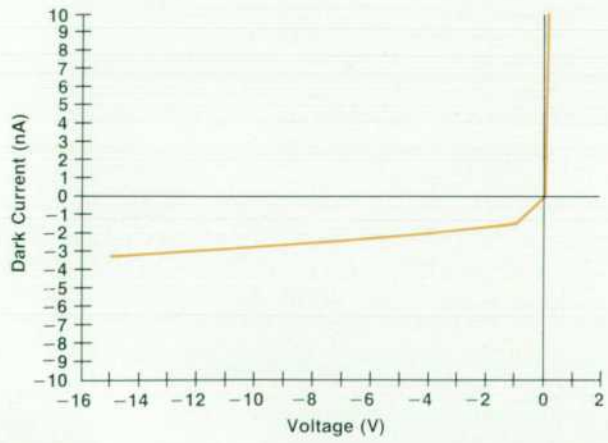


Fig. 1. *Typical I_d-V characteristics of a 25-μm-diameter photodiode.*

Operation and Fabrication

The epitaxial material for these photodetectors is grown by organometallic vapor phase epitaxy (OMVPE). The epitaxial structure is shown in Fig. 2. The pin structure is made of a p-type InP zinc-doped window layer and an InGaAs intrinsic region active layer on an n+ substrate. An InP buffer region is grown to facilitate the growing of the InGaAs active layer. This $In_{0.53}Ga_{0.47}As$ layer is lattice matched to the InP and must have low background doping ($\approx1\times10^{15}$/cm^3). The InP cap layer is transparent to the wavelengths of interest, 1.3 and 1.55 μm. InP will not absorb for $\lambda >$ 0.92 μm, that is, for

$$\lambda > \frac{hc}{E_g} = \frac{1.24\ \text{eV-}\mu\text{m}}{E_g},$$

Fig. 2. *Photodetector epitaxial structure (not to scale).*

where E_g = 1.35 eV is the bandgap energy for InP, h is Planck's constant, and c is the speed of light. InGaAs will absorb wavelengths up to ≈1.7 μm and is therefore an excellent material for the active region.

The pin detector works by converting incoming light energy into an electrical current. A photon of wavelenth 1.2 to 1.6 μm incident upon the top InP cap region moves through without absorption. The photon is absorbed in the InGaAs active region, creating an electron/hole pair. When the device is operated in reverse bias, an electric field is present and the carriers are swept out, creating a current (see Fig. 3). These detectors are designed for high-speed operaton. The frequency response is maximized by minimizing the detector area and thus its capacitance, and optimizing the i layer thickness for transit time considerations (Fig. 4).

The high-speed detector is fabricated on a conductive n-type InP substrate. Fig. 5 shows top and cross-sectional views of the photodetector. The wafer is backlapped, and a backside ohmic contact of AuGe is deposited. The frontside AuZn ohmic contact is patterned with a liftoff process, and a Si_xN_y antireflection coating is deposited by plasma enhanced chemical vapor deposition (PECVD). The mesa of the pin structure is formed next by a wet chemical etch in a HBr-based etchant. The device is now functional; however, passivation is applied to seal the mesa walls, to give a stable, low dark current for reliable operation, and to form an insulating layer upon which to plate a bonding pad. The passivation layer is then etched from the active area, revealing the antireflection coating. The devices are separated by sawing with a diamond-embedded resin blade.

The critical steps for ensuring low-dark-current devices are the mesa etch and the subsequent mesa passivation. While the mesa design provides a low-capacitance structure for high-frequency performance, it has the disadvantage of leaving the InGaAs active region exposed. An exposed pn junction in InGaAs is known to be a potential source of high dark current.[1] This exposed surface gives rise to a leakage path for surface leakage current which contributes to the dark current of the device. The dark

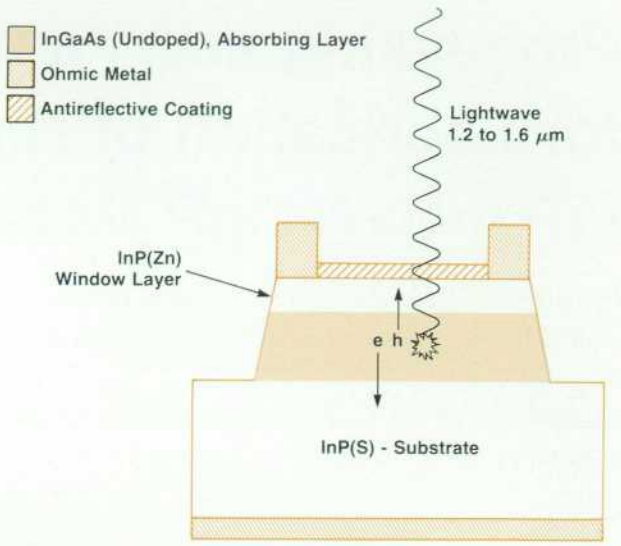


Fig. 3. *The photodetector converts incoming light into electron-hole pairs, resulting in an electrical current.*

current of the photodetector is given by the following equation:[2]

$$I_d = I_{dG-R} + I_{dDIFFUSION} + I_{dSURFACE},$$

where I_{dG-R} is the dark current resulting from generation and recombination, $I_{dDIFFUSION}$ is the dark current resulting from diffusion, and $I_{dSURFACE}$ is the dark current resulting from surface leakage.

Types of Passivation

Three basic types of surface passivation for InP-type devices have been examined and tested. The first is oxides. These oxides can be native oxides formed when the InP or InGaAs surface is exposed to air, or they can be purposely formed on a surface using anodization or an oxygen atmosphere. The second type of passivation uses a column-five (V) element from the periodic table, such as phosphorus

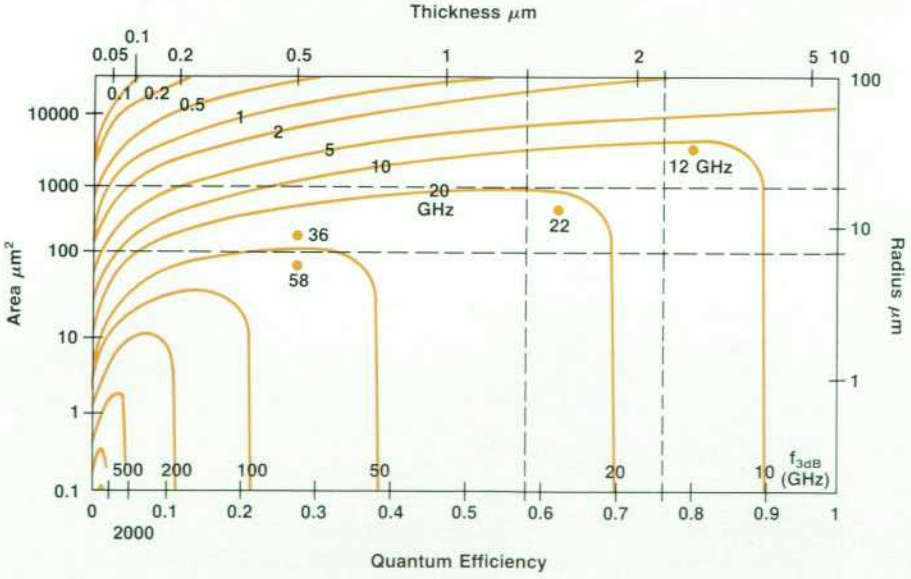


Fig. 4. *InGaAs layer thickness can be determined from this plot, which is from J.E. Bowers, et al, "Millimetre-Waveguide-Mounted InGaAs Photodetectors," Electronic Lettres, Vol. 22, 1986, p. 633. The dashed lines show the target ranges for i layer thickness and detector area.*

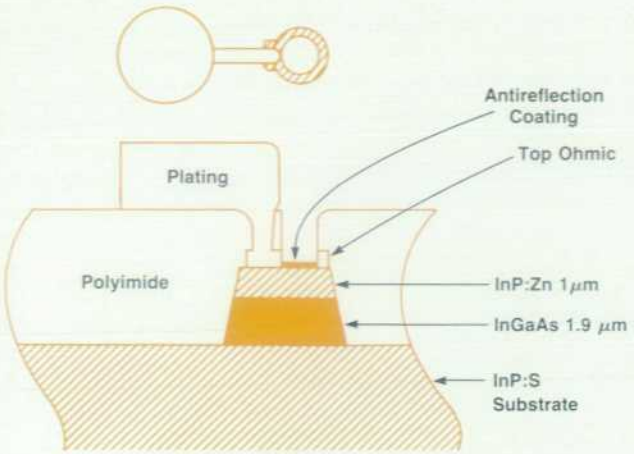
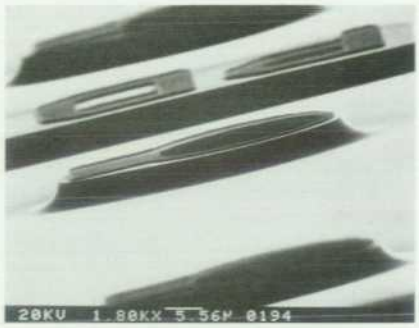


Fig. 5. *Photodetector diode top view and cross sectional view.*



Photodetector Mesa Surface
1:1:1 (IN)$K_2Cr_2O_7$:CH_3COOH:HBr

Surface Studies - ESCA

InGaAs	InP
As_2O_3	$In(OH)_3$
$In(OH)_3$	PO_x
*In_2O_3	InP
Ga	
GaAs	

*Conductive

Fig. 6. *Scanning electron micrograph of a typical photodiode after mesa etch. The list of native oxides formed on the InGaAs and InP after etching in 1:1:1 (1N)$K_2Cr_2O_7$:CH_3COOH:HBr was determined by ESCA (electron spectroscopy for chemical analysis).*

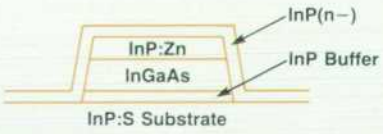
or arsenic. This has been done using arsenic partial pressure in a molecular beam epitaxy system,[3] or by cracking and depositing phosphorus.[4] The methods examined here include using elemental arsenic or regrowing epitaxial layers over existing layers in the OMVPE system. The third and most common type of mesa passivation is dielectrics. It is critical to have a clean, newly etched mesa surface before passivation.

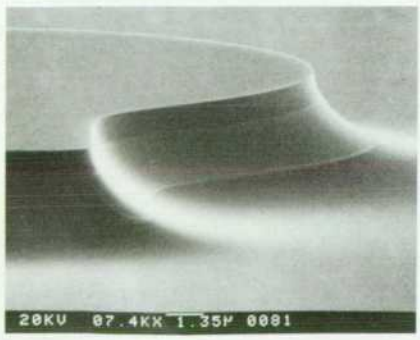
Oxides

The condition of the mesa wall surface after etch determines the device dark current. The use of a bromine-based etch leaves surface oxides on the mesa walls.[5] Chemical analysis of the InP and InGaAs surfaces by electron spectroscopy shows that the oxides listed in Fig. 6 are found after the use of (1N)$K_2Cr_2O_7$:CH_3COOH:HBr mesa etch. Fig. 6 also shows a scanning electron micrograph of the device after mesa etch. The most critcial oxide is the In_2O_3 formed on the InGaAs layer. This oxide is known to be conductive, and thus would contribute to the surface leakage current of the detectors. Increasing the thickness of this oxide layer by using an ozone stripper as a source of oxygen causes a further increase in dark current. Stripping this oxide in a wet etch can clean the mesa surface and leave a low-dark-current device that can then be passivated with a dielectric. An H_2SO_4-based etch is another way to remove the In_2O_3 from the mesa surface, leaving an arsenic-rich surface.[6] This cleanup etch is a good way to prepare the device for regrowth or dielectic application. These ozone-treated devices perform well under high-temperature operating life conditions. Devices are placed at an elevated temperature, 175°C, and under operating bias, −5V. This accelerates failures and allows the calibration and extrapolation of an instrument ambient lifetime for the devices. Lifetimes of 1×10^6 hours or 114 years at 55°C have been found for these ozone treated chips. This again points to passivation as a key to low dark currents and highly reliable devices.

Group-V Element Passivation

OMVPE regrowth is a method of covering the exposed InGaAs pn junction, thus suppressing or avoiding the $I_{dSURFACE}$ term in the dark current expression. Two regrowths will be examined. Fig. 7 shows a mesa before and





PIN Mesa

Mesa + Regrowth

Fig. 7. *OMVPE regrowth of InP (n−) over an etched mesa. (Left) Before regrowth. (Right) After regrowth.*

after regrowth. Here, a 500Å layer of InP is grown over the standard pin etched mesa. The InP is nominally undoped, but is actually slightly n-type ($1 \times 10^{15}/cm^3$). This n region will place the pn junction in the InP instead of the InGaAs, creating a more stable surface to reduce the surface leakage. The coverage is, however, incomplete. The device fabrication continues, adding backside metallization and frontside ohmic contacts on top of the mesa structure. The devices formed exhibited high dark currents, on the order of several hundred nanoamperes.

Fig. 8 shows an InP:Zn layer grown over an etched InGaAs mesa forming the p layer of the pin structure. The conformal p layer forces the pn junction to be buried in the bulk of the mesa leaving only InP exposed. This 1-μm p layer is, again, not continuous and thus leaves some areas of exposed InGaAs. A second etch is performed to etch away the newly grown p region between devices to separate them electrically. Again, backside and frontside ohmic contacts are deposited to allow electrical probing and testing of the devices. These devices were leaky, having dark currents of several hundred nanoamperes at $-5V$.

Two reasons for the high dark current are the incomplete coverage of the mesa surface and uncontrolled surface conditions before reentry into the OMVPE growth tube. The surface needs careful attention such as immediate etching and oxide removal before entering the chamber for regrowth. This processing was not possible because processing and regrowth were done at separate facilities. It is interesting that the incomplete coverage occurs in the same regions on both wafers and appears connected to the crystal orientation. The sides parallel to the major flat of the wafer, the $[0\bar{1}1]$ crystal direction, are completely covered. The sides perpendicular to the flat, [011], are not (see Fig. 9). It can be speculated that the different crystal planes preferentially form different native oxides, and that some oxides are less conducive to regrowth of InP material. Leaving exposed planes of In_2O_3 would result in high-dark-current devices.[7]

Another way of placing the V element of the III-V compound semiconductor on the exposed InGaAs mesa walls is by arsenic evaporation. Samples are prepared through mesa etch and placed in a special evaporation chamber for a thin (<100Å) coating of arsenic. Careful handling of the source material and wafers is necessary because of the toxicity of arsenic. The wafers are subsequently coated with polyimide, an insulating material, and follow standard processing. The devices did exhibit low dark currents, a few nanoamperes, at $-5V$. One adverse effect of this process is a dark current variation over time. Variations of about two nanoamperes on a dark current of the same magnitude occur over time in an irregular fashion. This dark current variation contributes to noise in the lightwave receivers, and therefore these devices are not usable for instrument applications.

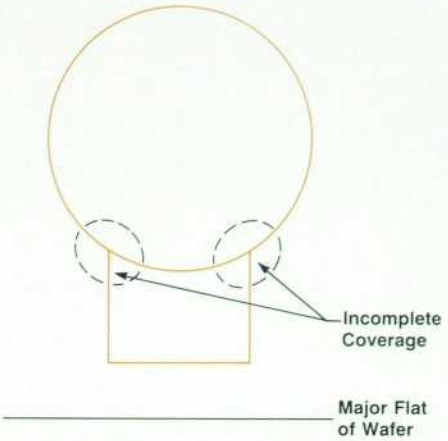
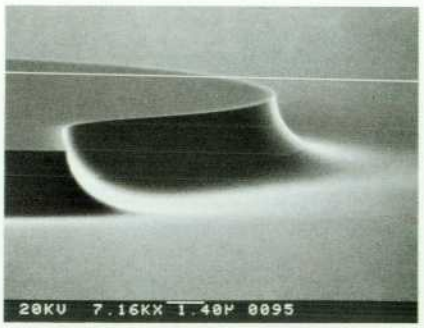


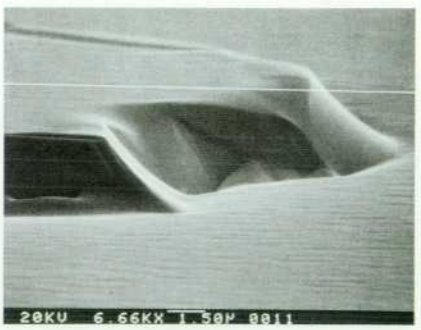
Fig. 9. *Incomplete coverage of regrown InP along the [011] direction.*

Dielectric Passivation

Dielectric coatings provide the third major method of passivating mesa devices. They are widely used and involve simple process techniques. Dielectrics that have been evaluated include: silicon monoxide, silicon dioxides, silicon nitrides, spin-on glasses, sodium hydroxide, and polyimides. Fig. 10 shows three such coatings. The left micrograph shows a mesa device after a typical SiO or SiO_2 dielectric deposition. SiO is deposited by thermal evaporation of a solid source. SiO_2 is deposited by chemical vapor



Mesa



Mesa + Regrowth

Fig. 8. *OMVPE regrowth of InP:Zn over an InGaAs mesa. (Left) Before regrowth. (Right) After regrowth.*

Fig. 10. *Dielectric passivations. (Left) SiO/SiO_2:P. (Center and Right) Spin-on SiO_2.*

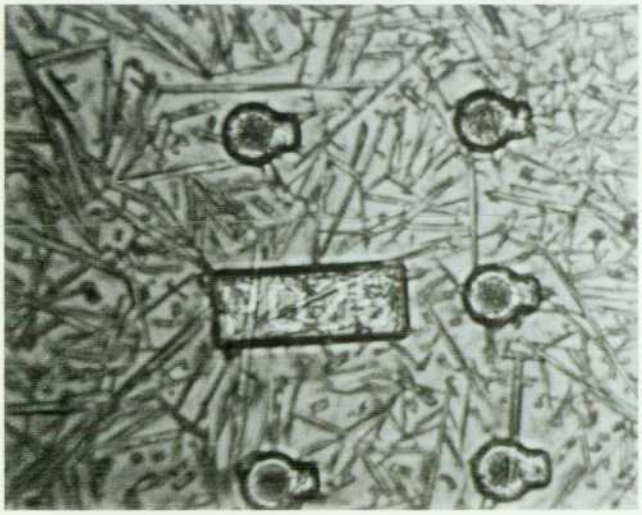


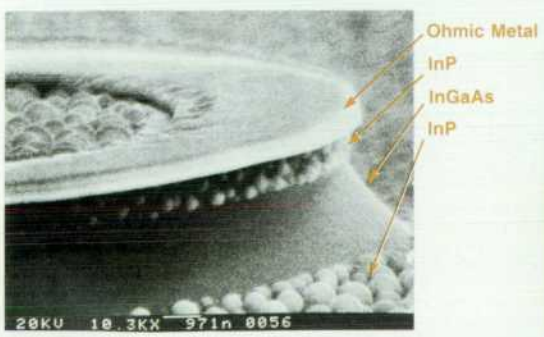
Fig. 11. *NaOH passivation.*

deposition (CVD). As shown in this photograph, the coating does not completely cover the mesa walls. These dielectric coatings have not yielded low-dark-current devices. One dielectric passivation attempts to incorporate the insulating dielectric with the V element passivation discussed earlier. This is a plasma CVD SiO_2 deposition with the inclusion of phosphorus in the form of phosphine gas. The coating did not result in complete sidewall coverage, and the devices had high dark currents.

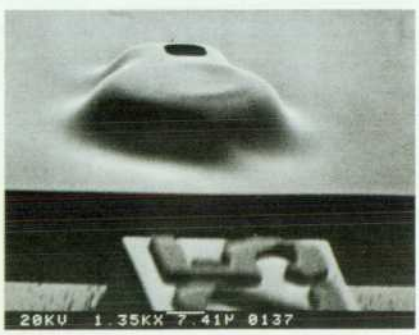
Silicon nitride does provide a very good conformal coating of the mesa walls. It is deposited by PECVD. Gaseous mixtures of silane (SiH_4), ammonia (NH_3), and nitrogen (N_2) fill a chamber and a plasma is ignited. The index of refraction of the mixture can be adjusted by controlling the relative amounts of silicon and nitrogen in the mixture. The nitride can have the dual purpose of passivating and serving as the device antireflection coating. Various stoichiometries of Si_xN_y were used as passivation. Although low-dark-current devices were produced, they were not consistently obtained.

Another type of silicon-based passivation is spin-on glasses. These coatings have the benefit of being easily doped with a variety of elements. They are also easy to process. The coating is dispensed and spun on like a conventional photoresist. It is then cured in a furnace to form a hard glass coating. Silicon-based liquids doped with the V elements arsenic and phosphorus were obtained as well as coatings with iron. Iron is added to InP as a dopant to create semi-insulating InP. The idea is to form an insulating layer at the mesa walls that will not be conducive to surface currents. Fig. 10 shows the results of two of these spin-on glasses. The center photograph is SiO_2:Fe. After curing, the coating is porous. The right photograph shows SiO_2:P. Here, the coating pulled away from the sidewalls and cracked. These coatings leave gaps through which water vapor and other substances can come into contact with the InGaAs active layer. This will lead to reliability problems over time. For this reason, and because of high dark currents, these spin-on coatings are unacceptable.

Lowering surface recombination rates of carriers at the mesa surface is a way of lowering device dark current. NaOH and KOH are said to have the effect of lowering surface recombination velocities in InGaAs.[8] A wafer processed through mesa etch is placed in a beaker of concentrated NaOH. The result is a very crystalline covering of NaOH over the wafer surface(see Fig. 11). This does not



Ammonia Plasma and Si_xN_y



Polyimide

Fig. 12. *(Left) Ammonia pretreatment followed by Si_xN_y deposition. The bumpy InP surface is caused by residual In. (Right) Polyimide passivation.*

provide a continuous covering of the mesa and is difficult to probe through to test device characteristics. This idea did not provide a practical or usable fabrication process.

Proper surface preparation coupled with a conformal mesa covering is the best way to achieve and maintain low-dark-current devices. One approach to achieving this involves using the ammonia gas present in the silicon nitride system as a pretreatment to the normal silicon nitride deposition. This in-situ process is performed without breaking vacuum, thereby maintaining a clean and controlled environment.[9] The ammonia plasma is meant to stabilize the mesa surface. The Si_xN_y is then deposited over that surface, sealing in the clean low-surface-current active layer. Fig. 12 shows a scanning electron micrograph view of part of the mesa area (magnified >10,000 times). The InP surfaces are bumpy while the InGaAs surface remains smooth. The ammonia plasma has knocked off the lighter phosphorous atoms (atomic weight 30.97) leaving behind a surface of In (atomic weight 114.82) bumps. This has been confirmed by electron dispersive x-ray spectroscopy (EDS) of the surface composition performed with the scanning electron microscope. The Si_xN_y film is then deposited covering the mesa and its bumpy surface. These devices did not have dark currents that met specification. The center of the ring where the fiber makes contact is also bumpy, leaving in doubt the optical reflection specification of less than 2%.

The present detector process uses a wet chemical etch followed by a double-layer passivation of polyimide cured at a low temperature (see Fig. 12).[10] The polyamic acid cures by driving out solvents and water vapor to provide a cured polyimide of about 2 μm per layer. It is important that the polyimide form a good seal around the mesa, be curable at low temperatures, and planarize, providing a surface upon which to plate the device bond pad.

Double-layer coatings provide increased planarity. Other parameters affecting polyimide planarity include viscosity and the amount of dissolved solids. Fig. 13 shows how planarity of a coating is examined. The device mesa is typically 3 to 4 μm high. This coating must cover and seal the mesa, planarize enough to accommodate the device bond pad, and be capable of being dry etched to create a via-hole connection of plated metal to the p-side ohmic contact. The polyimide must also be cleared completely from inside the ring contact to expose the antireflection coating.

Low-temperature processing must ensure a cured coating for rugged, reliable devices. The tall mesa requires a thick coating (4 μm) and requires cutting a deep hole through

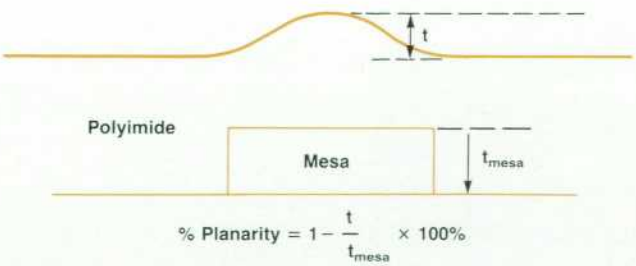


Fig. 13. *Polyimide planarity.*

Dielectric	Results
SiO	Poor Coverage, High I_d
SiO_2:P CVD	Poor Coverage, High I_d
Si_xN_y	Good Coverage, High I_d
SiO_2:Fe	Porous, High I_d
SiO_2:P	Cracking, High I_d
NaOH	Crystalline, High I_d
Ammonia CVD	Bumpy InP surface, High I_d
Polymide	Good Coverage, Low I_d

Fig. 14. *Dielectric Passivation Summary*

the polyimide. This is done using an oxygen plasma and reactive ion etching (RIE). The RIE cuts a deep hole with vertical walls (see Fig. 12). To plate the bond pad, metal must plate up through this hole as well as on top of the polyimide. This requires a base coating of metal covering all sides of the deep via hole. A special low-power, low-temperature sputter process is used to do this.

The low-curing, planar polyimide coating provides good mesa coverage and yields low-dark-current devices. 25-μm-diameter devices with dark currents of 1 nA (5×10^{-5} A/cm^2) at -5V have been fabricated. Typical dark currents are 1 to 5 nA at -5V. Devices are found to have 55°C instrument ambient lifetimes of 3.5×10^5 hours or 40 years.

Fig. 14 summarizes the dielectrics used for passivation. It includes the present polymide process.

A photograph of the completed photodetector chip can be seen in Fig. 15. This chip contains five 25-μm-diameter devices. These devices operate beyond 22 GHz.[11] Detectors of several different sizes for a variety of applications have been designed and fabricated.

HP Lightwave Products

Currently three different photodetector chips are used in HP products. One family of such products includes the HP 8702A Lightwave Component Analyzer and the HP 83400 series of lightwave sources and receivers (Fig. 16).[12,13] The 3-GHz and 6-GHz lightwave receivers contain photodetectors designed especially for these applications.

The PD25 chip from Fig. 15 is used in the HP 71400A Lightwave Signal Analyzer (see Fig. 17). This system has

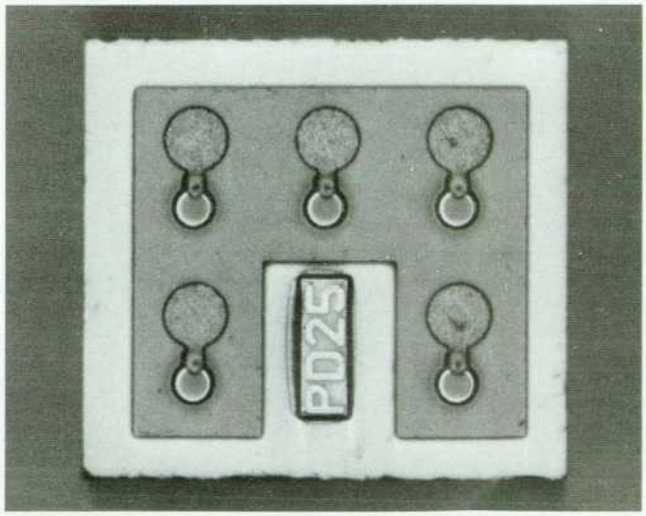


Fig. 15. *PD25 photodetector chip.*

Fig. 16. *HP 8702A Lightwave Component Analyzer system.*

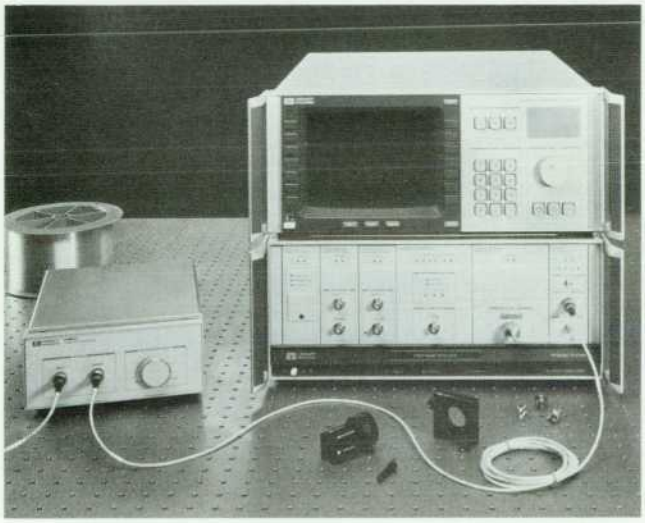


Fig. 17. *HP 71400A Lightwave Signal Analyzer.*

a receiver bandwidth of 100 kHz to 22 GHz and a sensitivity of −60 dBm (optical). This system can be used for laser, modulator, and transmitter characterization, and is calibrated so that it can be used as a reference receiver for lightwave detector and receiver testing.

Acknowledgments

Getting these lightwave products to market involved multidivisional efforts. Contributors to the photodetector project include: David Braun, Bob Bray, Eric Ehlers, Scott Elliott, Forrest Kellert, Nina Maddox, Monique Pearl, Mimi Planting, Jimmie Russell, Karl Shubert, Mary Stone, and Mark Zurakowski, with assistance from Ted Shimkowski, Mark Lightner, and Bill Anklam from the HP Microwave Technology Division and Kent Carey, James Chang, Gary Trott, and S.Y. Wang from HP Laboratories. Special thanks to Mimi Planting for her work during process development and shipments.

References

1. K. Ohnada, et al, "A Low Dark Current InGaAs/InP p-i-n Photodiode with Covered Mesa Structure," *IEEE Transactions on Electron Devices*, Vol. ED-34, no. 2, 1987, pp. 199-204.
2. J. Muller, "Photodiodes for Optical Communication," *Advances in Electronics and Electron Physics*, L. Maron and C. Marton, editors, Vol. 55, Academic Press, 1981, p. 212.
3. J. Chave, et al, "Arsenic Passivation of the InP Surface for Metal-Insulator-Semiconductor Devices Based on Both Ultra-High Vacuum Technique and Chemical Procedure," *Journal of Applied Physics*, Vol. 61, no. 1, 1987, pp. 257-260.
4. R. Schachter, et al, "Summary Abstract: Passivation of InP by Plasma Deposited Phosphorus: Effects of Surface Treatment," *Journal of Vacuum Science and Technology B*, Vol. 4, no. 4, 1986, pp. 1128-1129.
5. H.J. Stocker and D.E. Aspnes, "Surface Chemical Reactions on $In_{0.53}Ga_{0.47}As$," *Applied Physics Letters*, Vol. 42, no. 1, 1983, pp. 85-87.
6. *Ibid.*
7. K. Carey, et al, "Leakage Current in GaInAs/InP Photodiodes Grown by OMVPE," to be published in *Journal of Crystal Growth*, 1989.
8. E. Yablonovitch, et al, "Nearly Ideal Electronic Surfaces on Naked $In_{0.53}Ga_{0.47}As$ Quantum Wells," *Applied Physics Letters*, Vol. 52, no. 12, 1988, pp. 1002-1004.
9. F. Capasso and G.F. Williams, "A Propsed Hydrogenation/Nitridization Passivation Mechanism for GaAs and Other III-V Semiconductor Devices, Including InGaAs Long Wavelength Photodetectors," *Journal of the Electrochemical Society: Solid-State Science and Technology*, Vol. 129, no. 4, 1982, pp. 821-824.
10. R. Yeats and K. Von Dessonneck, "Polyimide Passivation of $In_{0.53}Ga_{0.47}As$/InP p-n Junction Structures," *Applied Physics Letters*, Vol. 44, no. 1, 1984, pp. 145-147.
11. E.R. Ehlers, et al, "Comparison of Frequency-Response Calibration Techniques for Wide-Bandwidth Photodetectors," *Digest of Technical Papers, 1988 Optical Fiber Communication Conference*, WQ22.
12. R.W. Wong, et al, "High-Speed Lightwave Component Analysis," *Hewlett-Packard Journal*, Vol. 40, no. 3, June 1989.
13. D. Albin, et al, "Design and Operation of High-Frequency Lightwave Sources and Receivers," *ibid.*

Providing Programmers with a Driver Debug Technique

Symbolic debugging is difficult for programmers who are developing drivers to run under the HP-UX operating system but do not have HP-UX source licenses. A technique is described to use available compiler information to provide access to certain HP-UX debug records.

by Eve M. Tanner

SOFTWARE DEVELOPMENT PROCEEDS most efficiently when programmers can request information on memory locations by typing in an ASCII name (symbol) rather than the address (hexadecimal/octal/decimal) of code being debugged. Symbolic debugging has revolutionized code development, placing the burden of numeric translation on the computer, allowing programmers to concentrate on the higher-level design issues of their code.

Running a symbolic debugger with code requires that the program source code first be compiled with a debug option. The compiler builds numerous useful information records into the compiled code, which, upon request by the programmer, are later used by a symbolic debugger to display lines of source code and values of data structures. Two important facts regarding compiler action are that debug information can be generated only for source code, not for the executable object code, and that HP-UX compilers currently generate debug information only on unoptimized source code—the compiler's optimizer cannot be invoked in the same compile time.

Programmers who are developing drivers to run under HP's HP-UX operating system need to include their code with the HP-UX code, and the two programs need to be debugged together. What happens when programmers do not own the HP-UX source code? The driver developer loses the ability to display HP-UX source code and data structures. Not owning source code is not an unusual occurrence, especially when dealing with operating system code, which tends to be very large and costly. Additionally, object code for operating systems tends to be optimized for system performance reasons. When only optimized HP-UX object code is available, the programmer's most potent weapon in the "debugging war zone," symbolic debugging of source code and data, is severely impaired, because the operating system code and data cannot be displayed.

Programmers in this situation needed to be supported in their efforts to develop drivers under HP-UX. Given that operating system code will not become generally available, and given that programmers will continue to demand symbolic debugging tools, a search was instituted to find a sensible solution to the problem. "Sensible" implies using information already available and making changes to files in a way that is transparent to the programmer and requires no changes to the debugging process.

The technique developed as a solution is based on the fact that kernel data structures are a primary point of interaction between drivers and the operating system. While the technique does not allow source code to be displayed, symbolic display of HP-UX data structures is restored to the driver developers.

The technical problem was to find a technique to strip data records from an unoptimized HP-UX kernel and insert them into users' programs that contain user source code and optimized HP-UX object code, all the while maintaining a valid file that appears untouched to the programmer.

This article describes the internal HP technology underlying this method. It is not intended as a description of any HP product.

Optimal Driver Development

Driver writers have special programming tool needs. They are inserting their own code into one of the lower layers of operating system code. In the case of HP-UX, driver modules reside between operating system kernel code and the hardware. Fig. 1 shows the position of device drivers in HP-UX.

Programmers who are debugging driver code need to test their driver routines with the kernel code, looking for any

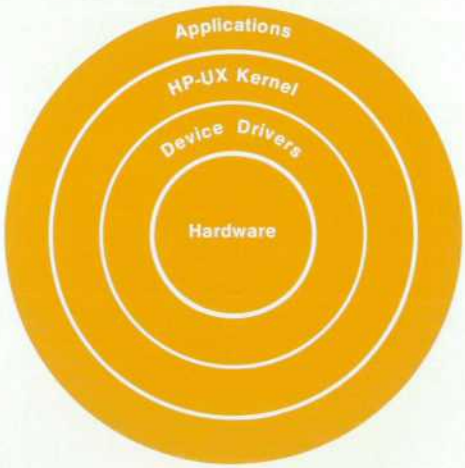


Fig. 1. *Position of drivers in HP-UX.*

effects their code might have on the operating system code, and vice versa. They need to examine the contents of data structures in any code for which they have the source. To do this, they need to have access to the HP-UX operating system source modules.

For symbolic debugging, the programmer submits the driver code and HP-UX code to the compiler, invoking the compiler with the debug option, which instructs it to add debug information into the final compiled object module. A symbolic debugger that knows how to access these special records is then run on the compiled code. When the programmer requests information on the HP-UX symbol proc, for example, the symbolic debugger will make the link between the symbolic name and the memory location where the contents of proc are stored, and will display the value of proc.

Partial Source Leads to Partial Debugging

Driver writers who do not have the source code for HP-UX are handicapped by an incomplete process, as illustrated by the debugging scenario in Fig. 2. A few quick command lines at the terminal allow a compilation of all of the source modules into a single, debuggable file, but since this programmer does not have access to the source files for the operating system, the compiler can generate debug records only for the driver module, not for the operating system modules. When the programmer hands off the compiled file's name to the symbolic debugger and starts to dig into the driver code, the debug section at the end of the compiled module contains pointers to the addresses of the driver code and data structures only. Suppose that the program runs only a short time before hanging the system. Stepping methodically through the driver code up to the hangup point and displaying all of its symbols does not find the bug. Perhaps a change in a certain little section of code will do the trick. This means shutting down the debugger, rewriting some lines of driver code, recompiling the source, and restarting the debug process again.

This is not a reasonable way to debug a program. It is not only a frustrating process, but also an expensive one in terms of resources, programmer time, and machine use. The programmer in Fig. 2 is wearing a wizard hat because some magic may help here. Important HP-UX data structures cannot be observed during driver debugging. The programmer cannot watch the interaction between the driver code and the operating system, so educated guesses must be substituted for on-line observations. The result is that many iterations of compilation and fixing errors must occur to get a driver program that works properly. "Working properly" means not only doing its appointed tasks, but refraining from creating havoc with the operating system.

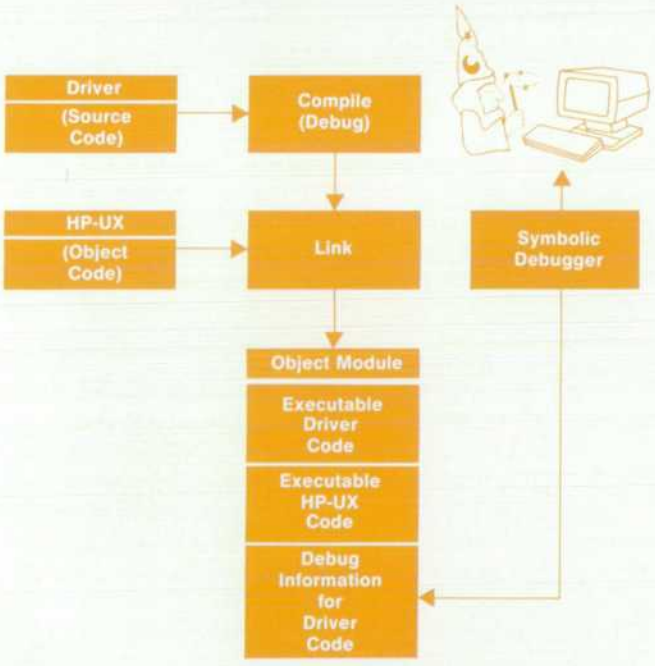


Fig. 2. *Information flow in the symbolic debugging process when HP-UX source code is not available. The programmer must use educated guesses or "magic" to find many bugs.*

Strip Out and Merge In

The solution to this unappealing debugging scenario is to provide a method that strips HP-UX operating system data structure records out of the HP-UX source code compiled with the debug option and merges them into a programmer's source file that was compiled under HP-UX. The resulting object module must look as if the programmer had actually compiled the HP-UX source code with the driver code. It must be compiler correct, and be usable in any plausible programming situation.

Implementing these concepts led to the development of two program files and one data file, as shown in Fig. 3.

Strip Program. This program strips global data debug records from HP-UX source compiled with the debug option (-g). This program is typically used on a one-time basis to create the strip file. The strip program is sometimes used by developers who have access to HP-UX source code but wish to do symbolic debugging in the most efficient manner; another entire HP-UX kernel need not be loaded in, only the stripped-out debug records.

Strip File. This file is created by the strip program. It contains the global debug records needed for symbolic debugging. Programmers submit this file to the merge program along with their own compiled driver code.

Merge Program. Programmers use this code to merge any existing code and data debug records contained in their own driver source code with the HP-UX data debug records in the strip file. A merged file is created, which can then be submitted to a symbolic debugger that knows the structure of code compiled under HP-UX.

Fig. 3. *Using available compiler information, two programs and a data file were developed to provide useful symbolic debugging capability to driver writers who do not have HP-UX source licenses.*

HP-UX Object Module Structure

The HP-UX object module can contain up to five distinct sections (Fig. 1). Four sections—the header, tables and dictionaries, code, and data sections—are always present. The debug section is generated only upon request, through the setting of an optional flag (−g) which asks the compiler to build the debug section along with any accompanying dictionary records and other pointers.

Four sections of the HP-UX object file are of particular interest in this project. The header section, the tables and dictionaries, and the debug section contain records that require updating, while the data section contains the address information, which must be duplicated in the newly merged debug section.

The header section contains typical header information, such as pointers to dictionaries containing information on the various sections and their subsections (into which each section is divided). It also contains the length of the entire compiled object module.

The tables and dictionaries contain a myriad of pointers and lengths that must be updated when merging the debug sections of the HP-UX file and the developer's driver. Included here is the symbol dictionary, which contains information on each code or data symbol that is defined or referenced in the compiled file. Among other information, each symbol record contains a pointer to the section (code or data) where the symbol is defined or referenced. Also included here are the start points and lengths of the debug section and its subsections and pointers to their ASCII names.

The debug section is built by the compiler when a debug request flag is set at compile time. This section is primarily used by a symbolic debugger to reconstruct information about the program. It contains all pointers needed to display various code and data information. Symbolic debuggers reconstruct program information by forming three important links:

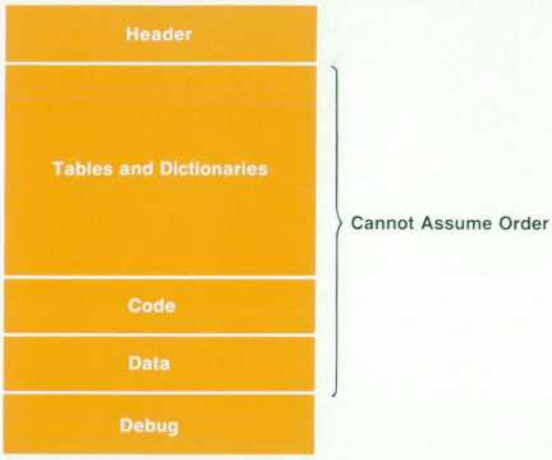


Fig. 1. *HP-UX object module structure.*

- The link between the debug section's symbolic data names and their addresses in the data section.
- The link between the debug section's procedure and label names and the code section.
- The link between the debug section's source line numbers and addresses in the code section.

The technique described in the accompanying article is only concerned with reconstructing the first of these three links.

As of this writing, only two major sections of an object module compiled under HP-UX are location dependent. The header section is always first and the debug section is always last (when it is present). The locations of all other portions of the object module are set at compile time, and can be traced by studying the header section.

Improved Driver Debugging

Using the strip file and the merge program described above, driver writers can debug their driver code using symbolic names for HP-UX global data structures. The debugging scenario now proceeds as follows.

The programmer compiles the driver source under HP-UX and is ready to begin debugging. Typing the <merge program> <strip file> <driver name> command line at the terminal produces a merged file. The symbolic debugger is invoked, with the merged file name in the command line along with any desired option flags. Symbolic debugging can now begin.

The strip and merge process does not restore all of the debugging possibilities that are lost when access to the HP-UX source code is not available. The only debug information restored is for what is known as global data. Thus, information can be retrieved for static variables and constants—data that never changes memory location. Debug information for local data is not involved in the strip and merge process, since local (dynamic) data can change memory locations. Also, source line information, which is also a typical compiler output, is not retrieved in the strip and merge process. Providing source line debugging, a common method in which the programmer can view lines of code as they appear in an ASCII code listing, was not included in the scope of this effort.

Challenges

As the name of the strip and merge process indicates, the HP-UX operating system debug records cannot be simply appended to the compiled driver code module. The challenge here was to be a good detective and find all of the lengths and pointers that needed updating. The act of appending or updating data in one section sometimes affects file areas considered safe from change.

The box above describes the structure of a file compiled under HP-UX, including the links between the debug information and the rest of the file.

Different HP-UX Versions

Disaster will strike the programmer who tries to merge debug records of one HP-UX version with an HP-UX file of another version. Code changes between the two versions may result in different data records in the debug sections of the two HP-UX files. Matching data tables in the debug sections are essential to the success of this technique.

For this reason, version checking became one of the first processing tasks of the merge program. It was decided that the product number would not be changing within an HP-UX version. The merge program ensures matching HP-UX versions by identifying the HP-UX what strings in each of the two files, comparing the product number and version information in each string.

Identifying Useful HP-UX Debug Records

The strip and merge process described in the accompanying article deals only with restoring global data information to symbolic debuggers. It is not involved in providing the ability to view lines of source code or local data structures.

Any information that can be used to reconstruct a source listing complete with labeled procedures and entry points may not be legally shared with nonpurchasers of HP-UX source licenses. Simply put, any debug records that point into the code section may not be shared.

Careful examination of the object module's debug section reveals three useful types of information for use in the strip and merge process. Fig. 1 shows the contents of the debug section, which is located at the end of the compiled file. The section's records are physically partitioned into five subsections by type of information.

The three asterisks indicate debug subsections that contain legal and useful information for retrieving global data names and values. The header subsection contains pointers to the remaining subsections. The global data subsection contains information about global data, and thus is limited to variables, types, and constants. The ASCII names section contains the ASCII string names of the global data items.

The local data records contain information about local data, such as entry point names and procedure names. The source line records contain information relating source (listing) lines to code addresses (code section). This is the linking that allows reconstruction of the object code to a typical source listing, and is not part of the effort described in this article.

Fig. 1. *Debug section of an HP-UX object module.*

Different Types of User Files

We were aware of the possibility that programmers might try to merge three different types of driver code modules with the strip file. Their driver modules might contain varied amounts of debug information records, such as:

- No debug information at all
- Debug records for only the driver code under development
- Debug records (a complete or partial set) for the HP-UX code.

In the first instance, the programmer did not invoke the debug option when compiling the driver code. This might seem unlikely, but it is a possibility that the merge program must and does handle. In this case, the merge program can begin in a simple way, appending the HP-UX data debug records to the end of the programmer's compiled file. A complicated cut-and-paste job must follow, however. As the box on the left shows, each distinct section (type of information) in the compiled file must have its accompanying pointers and lengths. When a compiler builds a section of information into a file, it also generates many pointers to the information. Each major section of the file has its own dictionary records, which include things like its length, pointers to its start, its ASCII string names, and so on. Since the merge program is now acting like a compiler, whenever it adds a new section of information to the file, it must also build all of the necessary new dictionary records, placing them correctly in the file. For example, dictionary records for the debug section and its subsections must be built and added into the tables and dictionaries section. ASCII strings for the new debug section and subsection names must be added into a strings table. Pointers and lengths of new and updated records must be added into the header section. If the cut-and-paste operation disturbs pointers to existing pointers and lengths, they need to be fixed.

The second possibility above, debug records for only the driver code, is probably the typical use of this process. The programmer compiled the driver source code along with the HP-UX object code, requesting that the compiler build debug records for all source modules.

The third possibility, an unknown amount of HP-UX debug records, is not a predicted use of this product, but it is a possible scenario that the merge program does handle. In this case, the programmer had source code for the driver and some (or possibly all) source code for the HP-UX operating system. The compiler was invoked with the debug request, and the resulting compiled module has a complete set of driver debug information and a set (the extent of its completeness is unknown) of HP-UX debug information records.

In the second and third cases above, the merge program merges the global information contained in the strip file into the debug section of the user's compiled file. The debug section of the file is changed, not newly created as for the first case. New dictionary records do not have to be built, but the existing ones do have to be updated with their new start positions and lengths. As can be expected, any other section records whose pointers are disturbed by the expansion of the debug section must also be updated.

Getting Those Tricky Pointers

In all of the above cases, the most interesting pointers to update are the pointers to the absolute addresses in the data section for each global data item in the HP-UX debug section. Fig. 4 shows these critical pointers.

Envisioning the compiled object file as a book, one can think of the symbol dictionary as the index of the book. It contains the pointers into the code and data sections for each code and data symbol in the compiled code. When a debug section exists, it too contains pointers to the code and data sections, so there is a second index. Recall that

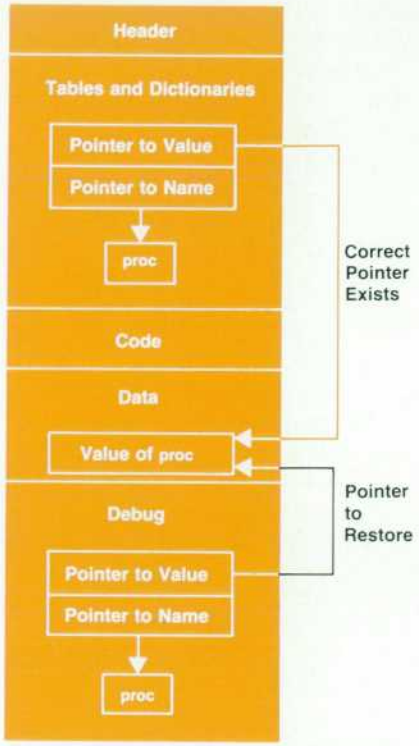


Fig. 4. *Restoring pointers into the data section.*

while the strip and merge process builds a new debug section for global data, the symbol dictionary entries for that global data already exist. Recall also that the newly added debug entries contain pointers to the data section of a different HP-UX file, the strip file. These pointers are now invalid. It is imperative that the newly added debug entries point to the same addresses in the data section as do the existing symbol dictionary entries. In other words, the two indexes (i.e., symbol dictionary and debug section) must have matching pointers to their shared index items.

The merge program's critical task is to set up a giant processing loop to recreate the link between the current file's data section and the global HP-UX data debug records. The global data records are examined and, for each record that contains a constant or variable, the following processing occurs:

- Retrieve the pointer to the ASCII name of the record.
- Use this ASCII name to index into the symbol dictionary.
- Retrieve the symbol's address into the data section and update the global data debug record to point to the new data section location. This updated pointer is shown in Fig. 4 (black arrow).

The stdio function nlist.c is used to help make this link between the strip file's global data debug records, the symbol dictionary, and the data section. The merge program passes the ASCII global data name to nlist.c, which returns a structure containing, among other things, the pointer to the data section.

Conclusions

The goal was to find and implement a technique that allows driver developers who do not have HP-UX source code to display global HP-UX data constants and variables symbolically while debugging their own driver code.

The HP internal technique described here is one that has possibilities for other programming tasks. It points out the great amount of free information an HP-UX compiler generates, and provides a working example of how this information can be used.

Acknowledgments

I would not have been able to implement this technique without the help of many outstanding people in the Workstation Technology Division's R&D, marketing, and manufacturing groups. I extend my thanks to all. In particular, I would like to recognize Ray Spear, who brainstormed over many of the technical issues in this effort, and Kevin Morgan, who had the task of encouraging me to turn concepts into working code.

Solder Joint Inspection Using Laser Doppler Vibrometry

Good solder joints can be distinguished from bad joints by their vibration spectra. Vibration frequencies for bad joints are consistent for a given lead type.

by Catherine A. Keely

LIKE MOST ELECTRONICS COMPANIES, Hewlett-Packard has been striving to fit more functionality into smaller packages. To achieve this, many of HP's products now use surface mount technology (SMT) and other technologies in their printed circuit assembly production. SMT involves the use of packages significantly smaller than standard through-hole packages. The package leads are soldered directly to pads on the surface of the board instead of passing through the board. SMT components have lead widths of 0.025 inch (0.625 mm) or less, with lead pitch double the lead width. As one can imagine, when a printed circuit board is packed with SMT components, inspection of the solder joints by visual means becomes a difficult task. This is especially true with certain lead shapes, such as J leads, where most of the solder joint is under the component. Sophisticated electrical testers can catch many solder joint errors, such as shorts or solder bridges. They can sometimes find unsoldered lead problems, but frequently these faults go unnoticed because the tester forces mechanical contact between the lead and the pad, making the missing joint undetectable.

To address this problem, the computer vision project of HP Laboratories has investigated alternative methods of inspecting SMT solder joints. The subject of this paper is one method that has proven successful at detecting one major joint error type: the unsoldered joint. The method, which does not actually involve computer vision although it is an optical method, is based on the vibration characteristics of an unsoldered lead. The idea is that an unsoldered lead, when stimulated, will vibrate at its resonant frequencies, which depend on its material and geometry. A soldered lead, under the same stimulation, will have different vibrational characteristics because it has different geometrical constraints. A laser Doppler vibrometer or velocimeter (LDV, see box, page 82) is used to measure the velocity of a vibrating lead, and as this paper describes, the peaks in the frequency spectrum of the vibration indicate whether the joint is soldered or not.

The SMT Application

The most common surface mount components fall into three general categories (see Fig. 1): J leads, which are found on PLCCs (plastic leaded chip carriers), gull wings, found on SOICs (small-outline integrated circuits) and PQFPs (plastic quad flat packs), and brick-shaped passive components. The gull wings can be further broken down into different shape standards: narrow SOICs, wide SOICs, and fine-pitch PQFPs (0.025 in./0.625 mm pitch or less). The passives can also be broken down into the many available shapes and sizes, since the geometry determines the natural frequency, but this paper does not distinguish between the different passives.

There are many common joint error types, including shorts or bridges, solder balls or splashes, cold joints, insufficient solder, excess solder, dewetting or wicking, nonwetting, and lifted leads. In the first five error types, some sort of joint is formed, so examining the vibration should only indicate that there is a joint. For the last three, which form a significant portion of the errors, there is no mechanical joint, and thus they are considered open or unsoldered joints for the purposes of vibration analysis. Cracked joints fall somewhere in between, depending on how freely the lead can vibrate.

Before building the LDV system to measure the vibration, finite element models of the various leads were made and analyzed for their vibration characteristics. This was done to determine analytically the expected resonant frequencies of soldered and unsoldered leads, to verify that the expected resonant frequencies for good and bad joints are distinct, and to make sure that the frequencies are in the range detectable by the proposed LDV setup, that is, below 200 kHz. The results of the finite element analysis gave the following resonances under 100 kHz for unmounted leads: about 29 kHz for wide SOICs, 51 kHz for narrow SOICs, 14 kHz (first mode) and 40 kHz (second mode) for J leads (PLCCs), and 20 kHz for fine-pitch PQFPs.[1] No modeling was done for the leadless passive components. First-mode resonant frequencies of properly soldered leads are at least five times those of unsoldered leads, and are always above 85 kHz, with the specific frequencies depending on the lead type and thickness. The frequencies increase slowly as the lead thicknesses increase, with slopes of 50 to 400 Hz/μm. These results indicate that in theory the vibration characteristics should distinguish soldered leads from unsoldered leads.

To measure the vibration, the leads first have to be stimulated to vibrate. This can be done in several ways, such as using an impulse force, shaking the board, or sweeping a stimulus frequency, but the method we implemented is an air jet (at a pressure of about 35 kPa or 5 psig) aimed near the joint, as shown in Fig. 2.[2] The air jet is a source of acoustic white noise, which sets up resonance vibrations in the leads. It is simple and flexible, and it works.

The LDV measuring beam is aimed at the lead shoulder

(continued on page 83)

Laser Doppler Vibrometry

Laser Doppler vibrometry is a widely used noncontact technique for measuring velocity or vibration. It is an extension of laser Doppler anemometry, which was developed in the 1960s and 1970s for fluid flow measurements.[1-3] It is based on the fact that light reflected from a moving object will have a Doppler frequency shift proportional to the object's velocity. The laser Doppler vibrometer (LDV) system provides a means of measuring the frequency shift caused by the vibrating object. Commercial LDVs are available, but because of the bandwidth and flexibility requirements of this feasibility investigation, a lab system was built.

Fig. 1 is a schematic diagram of the optics in the lab system. Laser light with frequency f_0 is incident on a target, which is moving in a sinusoidal fashion. The lead displacement x is given by:

$$x = C\sin(2\pi f_v t)$$

and the lead velocity v is:

$$v = \dot{x} = 2\pi f_v C\cos(2\pi f_v t),$$

where f_v is the target vibrational frequency in hertz, C is a constant representing the displacement amplitude, and t is time. The laser light E, which has the form:

$$E = A\cos(2\pi f_0 t),$$

where A is the amplitude and f_0 is the frequency of the laser light, will be frequency shifted after reflecting off the target by an amount

$$\Delta f = 2v\cos(\phi)/\lambda_0,$$

where ϕ is the angle between the target velocity vector and the incident light direction, and λ_0 is the wavelength of the laser light.

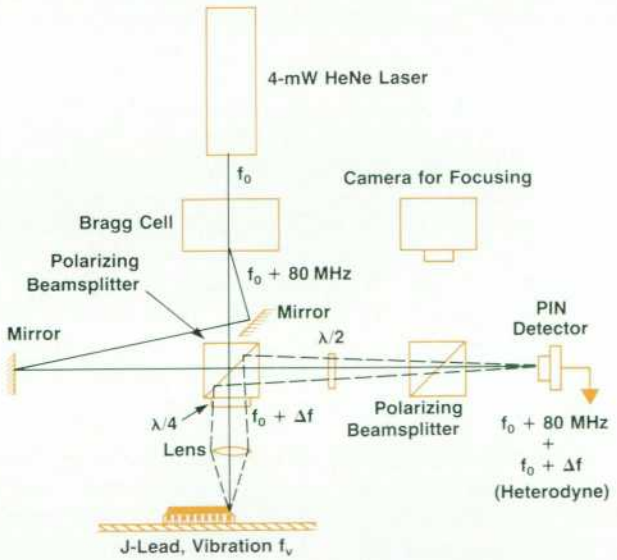


Fig. 1. *Optics for Laser Doppler vibrometry. The direct laser light is represented by solid black lines and the laser light reflected from the subject is represented by dashed black lines.*

(For the HeNe laser used, λ_0 = 632.8 nm.) Thus the reflected light E_v can be written as:

$$E_v = A\cos(2\pi(f_0 + \Delta f)t).$$

To separate the frequency shift from the incident frequency, the reflected beam is recombined with a reference beam. The reference beam is a beam from the same laser that is shifted by a known amount. In this case, the shift is generated by a Bragg cell, or acoustooptic modulator. A Bragg cell is basically a section of glass with a acoustic wave traveling through it, generated by a piezoelectric crystal. When light passes through the glass, part of it is diffracted at the acoustic wavefronts, which have a different refractive index. The light diffracted is essentially Doppler shifted by the frequency of the traveling acoustic wave. In this application, the Bragg cell causes a shift of f_b = 80 MHz in the light frequency, so the reference beam E_b is of the form:

$$E_b = B\cos(2\pi(f_0 + f_b)t),$$

where B is the amplitude of the light. When the reference beam is aligned with the measuring beam, the two beams interfere. The combined intensity is detected by a pin diode:

$$I(t) = (E_b + E_v)^2$$

$$= \frac{A^2}{2} + \frac{B^2}{2} + AB\cos(2\pi(f_b - \Delta f)t)$$

$$+ \text{higher-frequency terms.}$$

The first two terms are dc terms and are lost in the detector amplifier. The third term is the beat frequency between the two beams. Further terms have a frequency of order f_0 or higher and cannot be detected. The shift frequency Δf can be extracted from the signal by sending it to a frequency demodulator (FM receiver), and tuning the demodulator to the carrier frequency f_b. The demodulator output will be proportional to the frequency shift Δf, which is proportional to the target velocity. This signal is then sent to a spectrum analyzer so the vibration frequency content can be examined. Instead of frequency demodulation, the signal could be phase demodulated, and then the output would be proportional to the target displacement without the dc (or offset) terms. Either way, the frequency content is the same.

The vibration frequencies that can be detected are mostly a function of the demodulator. The velocity of the target causes the frequency shift, or deviation, and the change in velocity at the vibration frequency causes a rate of change in the frequency shift. In this application, the vibration is natural, so the vibration frequency and the rate of change are functions of the lead geometry. This cannot be changed, but the velocity is determined by the vibration frequency and the displacement, and the displacement can be changed by limiting the forcing function, which in this case is an air jet, as explained in the accompanying article.

The demodulator used can analyze an 80-MHz signal with frequency modulation rates f_v of 20 Hz to 200 kHz and deviations Δf (frequency shifts) up to 400 kHz with 100-Hz resolution. This means the target vibration frequencies can be up to 200 kHz and the velocity can be such that the deviation is up to 400 kHz. In practice, the demodulator deviation range is set for up to 40 kHz of frequency shift Δf and the resolution is 10 Hz. This setting corresponds to a target velocity $\dot{x}$ range of 3.2 μm/s to 12 mm/s. This corresponds to a displacement C range of, for example,

0.02 nm to 80 nm if f_v is 25 kHz.

A limit on the system performance is the power of the detected signal, or the amplitudes A and B of the light waves. This is primarily determined by the amount of light reflected from the target, but is also affected by the sensitivity of the detector, the amplifier, and the signal level required by the demodulator. To get a good signal, the measuring beam must be aimed at a surface that will reflect light back into the system.

References 1 to 3 provide more details on laser Doppler vibrometry.

References

1. N.A. Halliwell and R. Jones, *Vibration Measurement Using Laser Technology*, Course Notes, SPIE/ESD Conference, Dearborn, Michigan, June 1988.
2. P. Cielo, *Optical Techniques for Industrial Inspection*, Academic Press, 1988.
3. B.M. Watrasiewicz and M.J. Rudd, *Laser Doppler Measurements*, Butterworths Ltd., 1976.

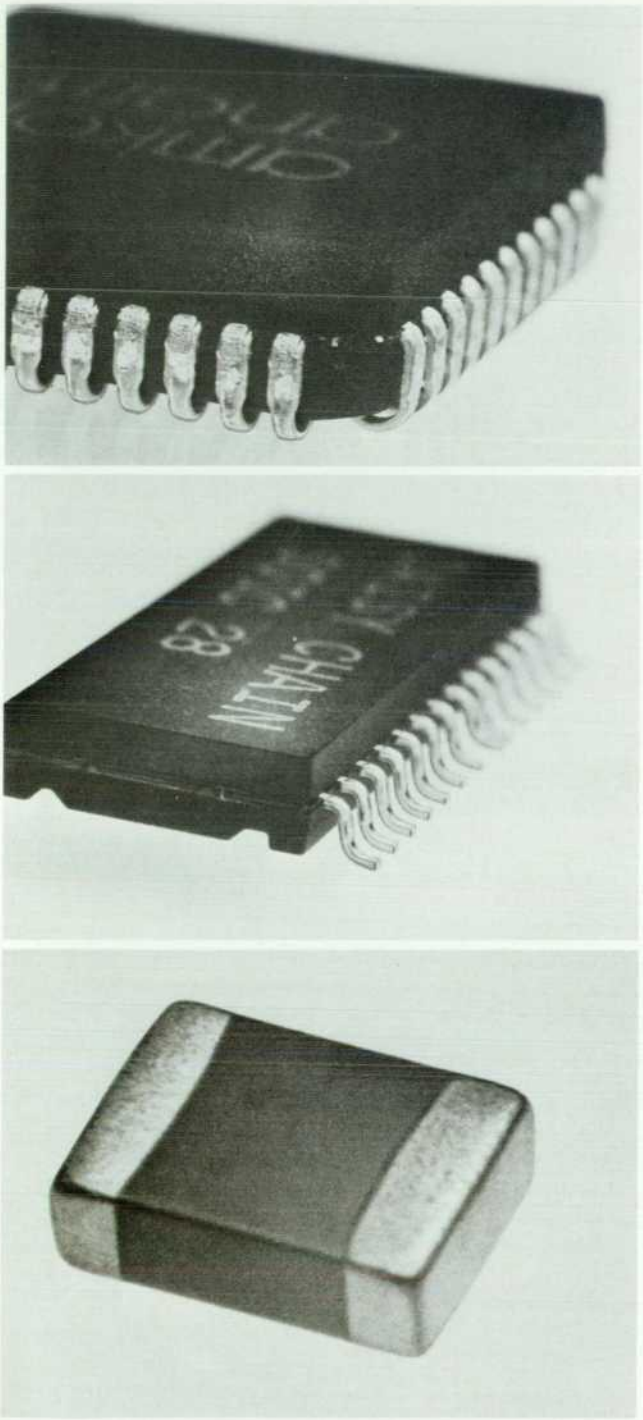


Fig. 1. *The three general lead types considered in this study. (top) J lead. (middle) Gull wing. (bottom) Passive.*

because this area provides the best light reflection and is easily accessible on all lead types (except the passives which have no shoulder). The shoulder deflection during vibration is more than sufficient to generate a detectable Doppler shift. For the passives, the beam is aimed at any point that reflects well near the end of the component.

As the lab system is presently set up, one lead is measured at a time, and then the printed circuit board is moved by an x-y table to access another lead. The time required to make each measurement is very short, depending primarily on the filter bandwidth and the averaging necessary to reduce the noise. Methods of automating the system to access leads quickly are being explored to increase the overall speed of the system.

Measurement Results

The results of the vibration measurements are in the form of vibration frequency spectrum plots. As explained earlier, the frequency peaks in the vibration spectra should (according to theory) be indicative of the state of the solder joint. The theory predicts specific frequencies related to the lead geometry and boundary conditions (i.e., soldered or unsoldered). In practice, other frequencies may be present because of the complete component geometry, but these are in addition to the lead vibration frequencies. The frequencies of the soldered lead are in general too high to be detected. The practical limit is about 100 kHz because of the small deflection and the amount of noise, even though in principle the system limit is 200 kHz.

Fig. 3 shows typical vibration spectra for unsoldered leads (e.g. no solder, solder wicked, etc.) and soldered leads for the five component types. Each plot contains a trace for the unsoldered lead and a trace for a nearby soldered lead on the same component, so the differences between the traces can be easily compared. For most components, the spectra for good, soldered joints are fairly flat unless

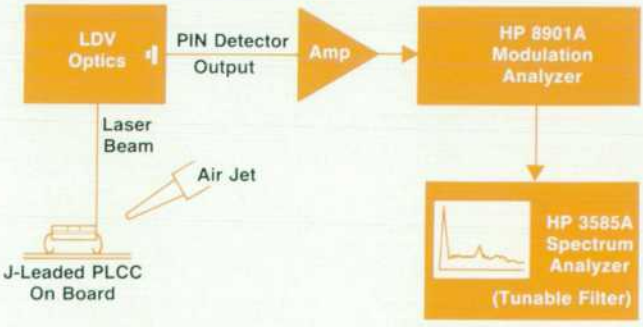


Fig. 2. Processing electronics for the laboratory laser Doppler vibrometry system.

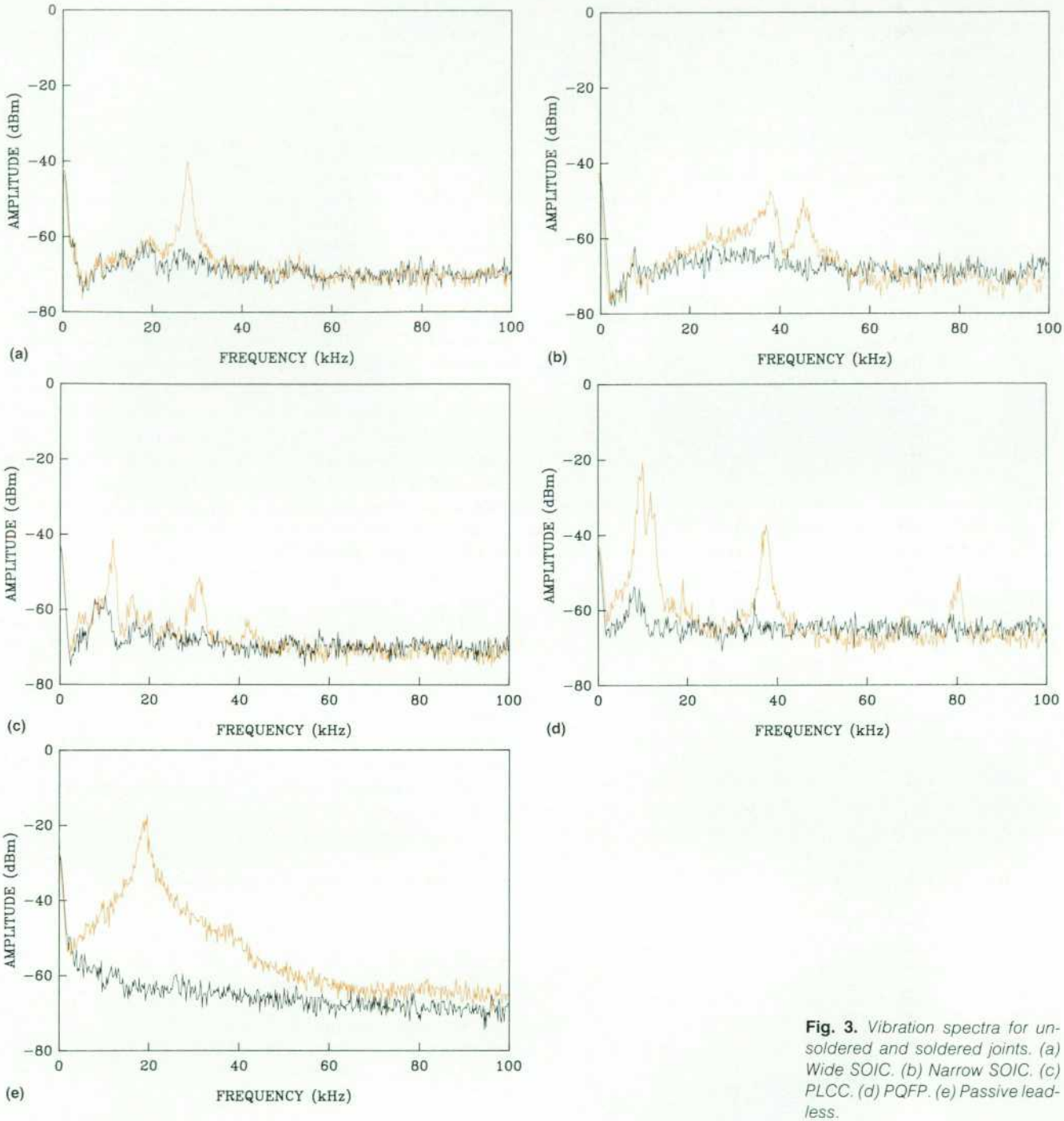


Fig. 3. *Vibration spectra for unsoldered and soldered joints. (a) Wide SOIC. (b) Narrow SOIC. (c) PLCC. (d) PQFP. (e) Passive leadless.*

there are many unsoldered leads so that the entire component can vibrate. In these plots the good joints' spectra are indeed reasonably flat and the unsoldered leads' spectra have resonance peaks that stand out, making them easily detectable.

Within a component type, the measured resonance frequencies are very consistent as shown in Fig. 4 by the two overlaid spectra from the same lead type (J-lead, PLCC) but different components. As can be seen from the plots for the leaded components, the peaks correspond well to the natural frequencies predicted by the finite element analysis given above.

The most important result is the distinction between soldered and unsoldered joints' vibration amplitudes at specific frequencies, depending on the lead type. The large difference in the vibration amplitudes at the resonance frequencies for good and bad joints allows this to be used as an inspection criterion.

To quantify the difference in vibration, the signal-to-noise ratio can be used. The resonant-frequency response of an unsoldered lead is the signal and the response at that same frequency of a soldered lead is the noise. At a typical peak in the spectrum of a bad joint, this signal-to-noise ratio, or separation (readable from the plots), is about 20

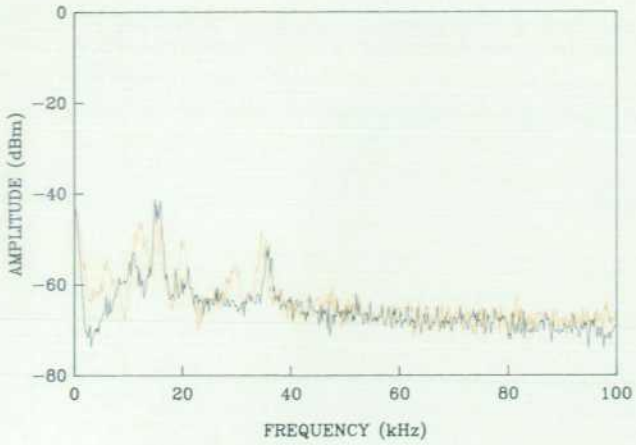


Fig. 4. *Two bad-joint spectra from different components of the same type (PLCC) are very similar, showing that responses are consistent for a given lead type.*

to 40 dB with the resolution bandwidth of the spectrum analyzer set at 1 kHz.

Conclusions

The vibration spectrum of surface mount leads is indicative of the state of the joint, that is, soldered or unsoldered. The use of the laser Doppler vibrometer to measure lead vibration after the solder process, followed by spectral analysis of the vibration, is a promising technique for detecting unsoldered leads in surface mount solder joint inspection. With the feasibility of the technique proven, work is now being continued to develop and assess the practicality of the system.

Acknowledgments

Many thanks to John Lau for the finite element analysis, and to Bill Gong, Dick Henze, Henrique Martins, Chuck Morehouse, Terry Pierce, and Wayne Sorin for their technical contributions and support of the LDV system.

References

1. J.H. Lau and C.A. Keely, "Dynamic Characterization of Surface Mount Component Leads for Solder Joint Inspection," *Proceedings of the 39th IEEE Electronic Components Conference*, Houston, Texas, May 1989, pp. 603-615.

2. T.H.T. Ninomiva and Y. Nakagawa, *Inspection method and apparatus for joint junction states*, U.S. Patent no. 4,641,527, assigned to Hitachi KK, WPI Acc. No. 87-055982/08.

Correction

In the August 1989 issue, the caption of Fig. 5 on page 89 is incorrect. The correct caption is as follows:

Fig. 5. *Simple inheritance.* object_b *inherits* method_1 *from* object_a.

A Model for HP-UX Shared Libraries Using Shared Memory on HP Precision Architecture Computers

To meet the needs of the PORT/HP-UX product, a special model for shared libraries was developed and implemented on HP 9000 Series 800 Computers.

by Anastasia M. Martelli

CURRENTLY, WHEN AN HP-UX USER PROGRAM includes calls to a library, a copy of the library code is added to the executable file at link time. This means that each executable file that uses a routine has a copy of that routine on disk and in memory when it is running. This type of library technology is called archived libraries (Fig. 1a).

Shared libraries allow multiple processes to share a single copy of library code. This is accomplished by loading shared library code into the system separately from any one program. When a program calls a routine in the shared library, it branches into and out of the shared library at run time (Fig. 1b). Thus an executable file that uses shared libraries does not hold library routines and initialized data. This results in smaller files and lower disk space requirements. Memory use may be reduced because only one copy of the library routines exists in memory.

After investigation, the HP-UX product team proposed that shared libraries not be supported for Release 2.0. However, the PORT/HP-UX project was receiving strong user feedback to reduce the size of executable program files that use their emulation library.* Therefore, a solution had to be provided. This paper describes a model that was designed and implemented to help meet this customer need. No kernel support was necessary, but minor changes were made to the linker.

Goals of the Implementation

Besides supplying the functionality required by the PORT/HP-UX product, there were two major design goals. The first was to avoid modification of the HP-UX core product, that is, the kernel, compilers, and linker, to support this model of shared libraries. The second was to avoid hard-coding the shared library text and data addresses. Hard-coding the addresses requires the user process to attach shared library code and data at predetermined addresses within their virtual address space, which is a well-known problem with some other UNIX** implementations of shared libraries.

*PORT/HP-UX (now renamed PORT/RX) is a software product that aids the migration of applications from the HP 1000 Computer with the RTE operating system to the HP-UX operating system on HP 9000 Series 800 Computers. The largest part of PORT/HP-UX is an emulation library that includes most of the system entry points available in RTE.

**UNIX is a registered trademark of AT&T in the U.S.A. and other countries.

Mapping the Shared Library

The design of the shared library began with the decision of where in the process's address space the shared library would reside. A shared library consists of two parts: the text (or code) and the data. Although there is only one copy of the text for all processes that attach to the shared library, each process must have its own private copy of the data. Therefore, the text and data must be handled separately.

Consider the HP-UX use of a process's address space on the HP 9000 Series 800 (Fig. 2). The address space of a process is divided into four areas. The first area is used for the text of the process. When there are multiple instances of the same program running, this text is shared across all instances, but the user process may not write to this area. The second quadrant is used for process private data, heap, and stack. The user process can write to this data area and

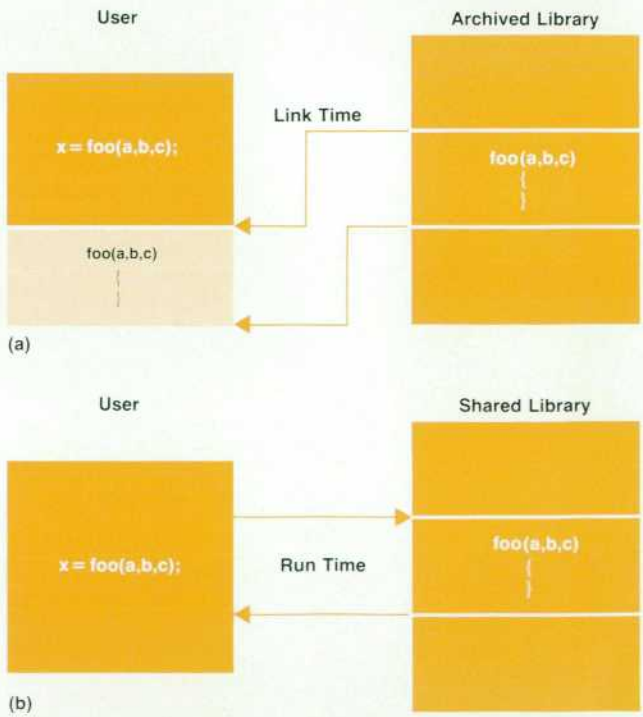


Fig. 1. *(a) Current library use on HP-UX. (b) Shared library use.*

it can be dynamically expanded to a system-imposed limit. The kernel stack resides in the third area. This is the stack information of the process while it is executing in the kernel and the process is not allowed to write on it. The fourth area is used for things that may need to be accessible to all processes on the system. Besides some miscellaneous kernel information, this area includes shared memory segments created by user processes. The user can write to this area by creating and attaching a shared memory segment.

At first glance, it may seem that the first area, which contains sharable text, would be a good place to put the shared library text. This is not the case, for two reasons. First, the user cannot place text in this area of the user process space except through the exec system call, which completely overlays what is already there; thus, the shared library text would be there, but the user code would have disappeared. Second, for architectural reasons the area that contains the shared library must be the same for all processes; thus, the user would have to keep track of two areas with the same short address—the one containing user text and the one containing the shared library text. Besides being very complex, this would also degrade performance.

The only other area of the process's address space that a user shares with other processes is the fourth area. The user can set up a shared memory segment and any other process can attach to it as long as the proper key is provided. A shared library can be set up in a shared memory segment with a well-known key.

Unlike text, each process has its own private copy of shared library data. Since a single set of shared library instructions must be able to access the data areas attached to multiple processes, some implementations have had to reserve specific portions of a process's virtual address space for shared library data areas, and the data of the shared library must be mapped to the same virtual address in each process that uses it. Additionally, that virtual address must be decided upon when the shared library is built.

Our implementation takes advantage of a feature of HP Precision Architecture to get around the inflexibility and complexity of the method just described. HP Precision Architecture instructions reference data relative to a register called the data pointer or dp. Using this feature, shared library data can reside at different addresses for each process. During most of the life of a process the data pointer is the address of the first word of the user's data area. If the data pointer is changed to the address of the shared library data area when entering a shared library routine, the shared library can access its data. The data pointer must be changed back to the valid value for the user data when exiting the shared library. Any data that is not dp-relative (such as initialized pointers) must be fixed when the data is relocated in this way. This allows the data for the shared library to be placed anywhere the user has write access. This could be at the end of the process private data, or in a private shared memory segment. Since the total amount of shared memory in the system is limited, it is wasteful to use a shared memory segment for process private data, so PORT/HP-UX chose to put the data at the end of the process private data.

Since the data pointer is changed between the user code and the shared library code, the user process cannot access the data of the shared library and the shared library cannot access the data of the user process except through passing parameters and return values.

Fig. 3 shows how the process's address space is used in this implementation of shared libraries.

Loading and Attaching the Shared Library

Theoretically, the shared library text only needs to be

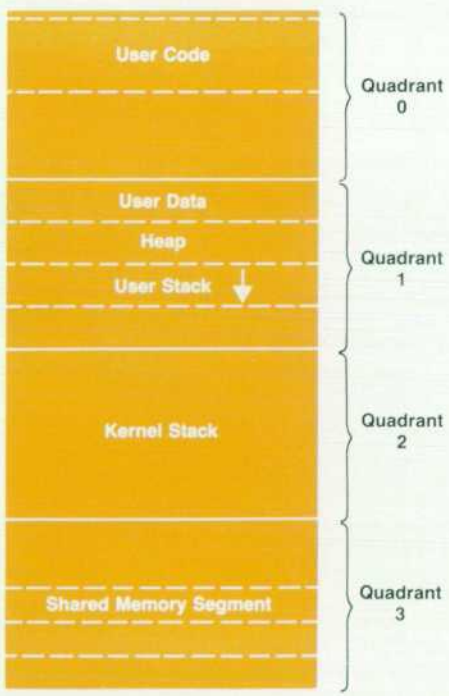


Fig. 2. *HP 9000 Series 800 Computer address space use.*

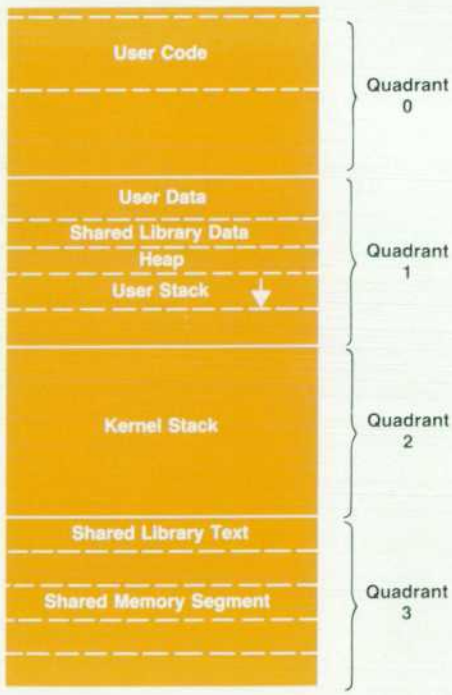


Fig. 3. *Address space use in this implementation of shared libraries.*

loaded into the system once. Actually, it may either be loaded into the system before running any processes that need it, or it may be loaded when the first process attaches to it and unloaded when the last process detaches from it. While the second approach is more efficient in the use of shared memory, the first approach is less complex and more efficient from a performance standpoint. The PORT/HP-UX project uses the first approach with their shared library, since an initialization process has to be executed before executing any other processes anyway.

To load the shared library, a shared memory segment is created and attached to the process that is loading the library. The system call to attach the shared memory segment returns the address of the segment. Since the HP 9000 Series 800 does not let the user specify the attach address of a shared memory segment, the shared library must be relinked at this point, specifying the shared memory attach address as the beginning of user text. Once the link is complete, the shared text is read into the shared memory segment and is scanned for the address of a single shared library entry point, which will be used later to branch into the shared library. Then, when a process starts up that attaches the shared library, it must attach to the shared memory segment. The shared library entry point may be stored in another shared memory segment, or each process may search for it when it attaches to the shared memory segment.

Since a copy of shared library data is kept for each process that is attached to it, the data is mapped at each process start-up. A section of memory is allocated from the process's heap and the shared library data is read into that area. The address of this allocated area is kept to be loaded into the data pointer when the shared library is entered, as discussed above, and fixups are performed on initialized global data.

Calling the Shared Library

Once the shared library is attached, a process can call the routines that reside in it. Because the data pointer must be changed between the shared library and user code, and a special branch instruction must be made to branch between the separate areas described above, two types of stubs have been added to user and shared library code. A stub can be thought of as an interceptor of a subroutine call.

Fig. 4 shows the flow of a shared library call. There is an assembly language branch stub for every routine in the shared library that may be called by code that resides outside the shared library. These stubs have the same names as the routines in the library and are linked with the user-supplied code. When a call to the shared library is made, the corresponding branch stub in the user's text area is entered. Satisfying the external reference to the shared library routine, this stub serves many purposes. It first saves the return pointer and the user data pointer on the stack. It then changes the data pointer to the area that was allocated for the shared library data area. Finally, it performs a branch to the shared library entry point stub. Control never returns to the branch stub; rather, control returns directly to the return point stored on the stack through the shared library entry stub.

The shared library is always entered via a single entry point—the shared library entry stub. Individual routines are then entered via a jump table which is indexed by a number passed to the library by the branch stub. Entering the library through this single stub requires that only one address in the shared library be known to the user process. If it were to enter at each routine, each process would have the additional overhead of setting up the address of each entry point.

The exit to user code is also done through this entry point. In this way each routine in the shared library does not have to be specially written to restore the data pointer and jump back into the user's text space via a long branch.

Signals

Signal handling was one of the most interesting issues in this design. A signal is a software interrupt sent to a process to inform it of a special situation or event. Processes that are expecting signals generally install signal handling routines which are invoked immediately upon receipt of the signal regardless of where the process is executing.

A signal can be delivered to a process while it is executing in either user text or shared library text and a signal handler can reside in either the users' text space or the shared library. Because the state changes between the user code and the shared library code (namely the data pointer), special signal handling code is added to this implementation of shared libraries. If a signal is received while the process

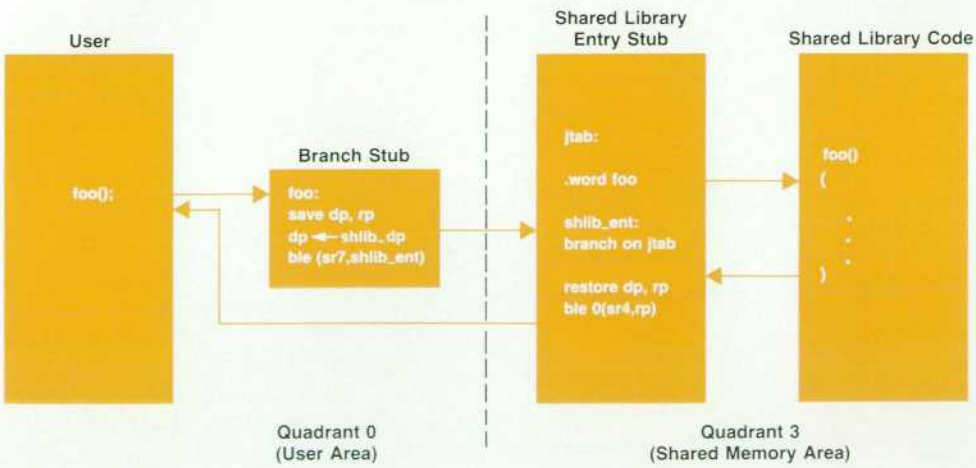


Fig. 4. *Control flow for a shared library call.*

is executing in shared library code and the user has previously set up a signal handler in the user code space, the data pointer must be set back to the user space data pointer while in the signal handler, and it must be set back to the data addresses of the shared library data on return from the signal handler. This implies that a state handling routine must get control of the process before and after the signal handler.

This is accomplished by providing a modified version of the sigvector stub. Sigvector is the system call that is made to install user-defined signal handlers. In the core HP-UX product, the sigvector stub just invokes the system call, passing along the supplied address of the handler to be invoked for a particular signal. The modified sigvector stub passes the address of a state handling stub to the kernel instead. The address of the actual handler is stored in a table along with the data pointer that should be in effect for the handler. The other system call for installing signal handlers, signal, will work this way as well, since it is implemented on top of sigvector.

When the signal is delivered, the state handling stub is invoked to branch to the real handler with the correct data pointer. The state handling stub and the signal information table used for storing information about the real user handlers must reside in user space. The reason for this is that being able to access a well-known data pointer is necessary for the state handling stub. Since receipt of a signal can occur at any time, it is difficult at best to tell what state the process was in at the time. The HP-UX linker sets a constant at link time, $global$, which is the first word in the data space. Since the shared library data space is meant to be dynamic but the user data space is not, the process can load the data pointer with the link-time value of $global$ and be pointing to the first word of the user's data. Therefore, as soon as the state handling stub is entered, it saves the current data pointer onto the stack and loads dp with $global$. Then the information about the real signal handler is retrieved and dp is set to the correct value for the handler. The real signal handler can now be invoked.

After handling the signal, the data pointer must be restored to its value when the signal occurred. This can be done in this same state handling stub for handlers that reside in the user's text space. However, so that handlers in the shared library don't have to be specially written to do a long branch back to this stub, an additional stub is added to the shared library that restores the data pointer and returns to the kernel, which in turn restores control to the code stream that was executing when the signal occurred. The address of this stub is placed in the return pointer by the state handling stub before the handler is invoked.

The modified sigvector stub resides in the shared library, so both the user text and the shared library text use the same entry to the sigvector system call. This implies that the shared library must find out from the user process what the addresses of the state handling stub and signal information table are. This information is passed to the shared library when the process attaches to it.

Fig. 5 shows the flow of control for handling a signal.

Setjmp and Longjmp

Setjmp and longjmp are routines that save and restore the state of a process. Special versions of these routines must be provided with the shared library since they now must save and restore the data pointer as well.

Limitations of the Model

Two major limitations of this model are caused by changing the data pointer between user code and shared library code. First, this version of shared libraries is not portable to other architectures that do not access data indirectly through a data pointer, specifically the HP 9000 Series 300.

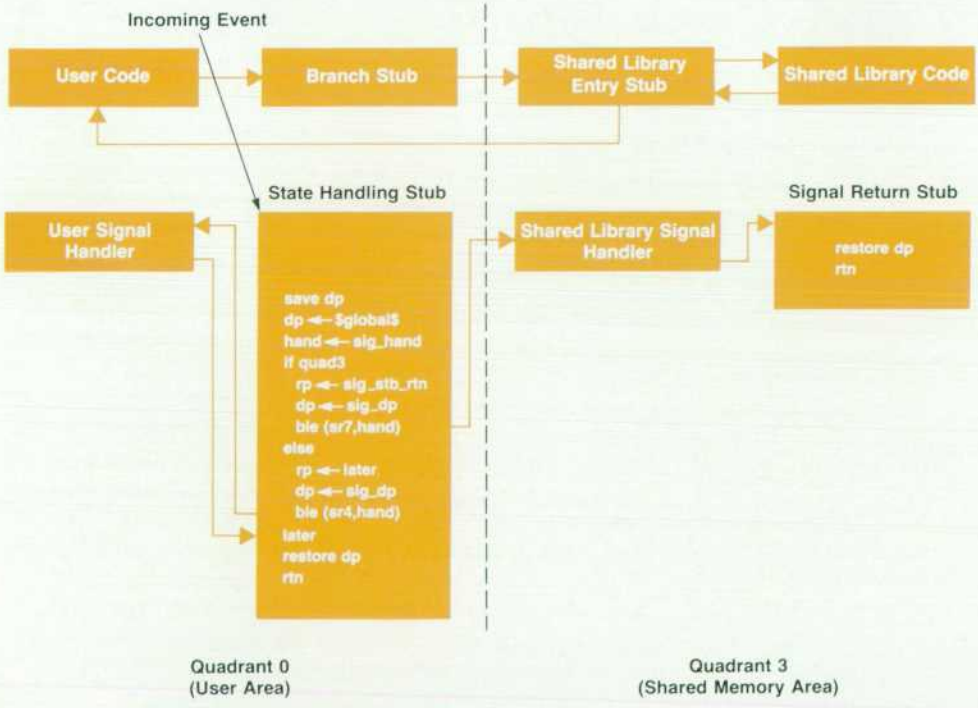


Fig. 5. *(Top) Normal flow of control for a process. (Bottom) Flow of control when a signal is delivered to a process in which a user has set up custom signal handlers.*

Secondly, it is not possible to share data between the shared library and the user code except through passed parameters and return values. The most well-known example of data that is shared directly between user and library code on HP-UX is the variable errno. When an error occurs in a library routine or system call, errno is set to signify a specific error and −1 is returned to the user. When the user is returned −1, the value of errno is checked to determine the error. This behavior is not supported in this model. The PORT/HP-UX project did not need this type of data sharing in the general case, but since their code includes calls to the archived C library (a library that holds standard routines that are released with the core HP-UX product) that might also be made in the user's code, they ran into this problem. For their application this is solved by copying specific pieces of data from user space to shared library data space when the shared library is attached and placing the code from libc into their library so the user gets the same copy as the shared library.

This model does not include the feature of calling a user routine from the shared library, since this feature was not required by the PORT/HP-UX project.

Conclusions

Once again, the goals for this project were to avoid hard-coding the text and data addresses of the shared library and to avoid modifying the HP-UX core product. Both of these goals were met with minor exceptions. First, the data address for the shared library can be different for each process that is attached to it. On the other hand, the text address of the shared library entry point is the same for every process. This was an acceptable solution since on the HP 9000 Series 800 all shared memory segments must be attached at the same address for all processes.

The HP-UX linker needed to change to support the model. The HP-UX linker did not generate the loader fixup information that was necessary to fix the initialized global pointers when moving the data pointer. This functionality was already included in the MPE-XL version of the linker and was not difficult to add.

There was a restricting software convention in the HP Precision Architecture control document that caused problems for this implementation. This convention was deemed unnecessary for the HP-UX product, and the code generation in the linker that depended on it was changed.

This shared library work was continued at the HP Data Systems Division and released as part of the PORT/HP-UX product with Release 2.0. The result is a 40-60% savings in the length of a typical user's program.

Acknowledgments

I would like to thank Cary Coutant of the HP-UX linker project for his invaluable assistance throughout the design and implementation of this project.

User-Centered Application Definition: A Methodology and Case Study

This paper presents a practical user-centered methodology for application definition. The methodology encompasses interviewing strategies, task analysis, and storyboarding techniques. The need for systematic user analysis is demonstrated, and the methodology is illustrated by a case study.

by Lucy M. Berlin

GOOD PRODUCTS MEET USERS' NEEDS, present their interactions in terms of their users' model of operations, and are easy to use. They accomplish this by embodying knowledge about the user's tasks and sequencing of operations, and about the relationship of the tasks to the overall work context. For example, VCRs can be difficult to use because they have an obscure tape counter, a control that reflects the internal structure of the machine, rather than a "minutes of playing time" indicator, which would reflect the user's view of the information. In contrast, a washing machine is easy to use, since the user simply indicates the type of clothes and the size of load, and is not required to understand the machine's internals or to specify the hot water flow rate, gallons of rinse water, and spin rate of the spindle. Good products present user interfaces that are consistent with the user's model of their operation rather than mapping too directly to the product's internal organization.

Software applications, and in particular their user interfaces, also need to be consistent with their users' model of operations. Dramatic advances in hardware integration and an exponential decline in costs have expanded computer use beyond the original computer science and engineering domains into many new contexts. As a result, software users can no longer be assumed to be computer experts who are comfortable with the traditionally terse user interfaces. These noncomputer specialists are too often required to translate between their professional domain and the restrictive terms of the computer. For them, commands such as find . -type f -mtime -7 -exec grep -l 'Hello' {} \; are viewed as magical incantations, not meaningful communication.* They must try to model their needs into the system's internal structures and concepts (similar to manipulating tape counters) instead of communicating with the system in the terms of their domain and tasks. To provide good products for these new users, developers must actively seek out an understanding of their users' domain and tasks. It is this need for developers to do more to understand users and their needs that we call user-centered design.

In the project covered in this paper, we set an explicit goal to produce a user interface design that is focused on user tasks. We developed the user-centered design methodology described here to help us gain this crucial understanding of our users' needs, and now propose it as a more generally useful process. Using this approach, other developers may learn, for example, that seemingly similar tasks in different domains or involving different users may actually have very different, domain-specific requirements for functionality and user interface. For example, computer-aided design applications for VLSI circuits and for mechanical engineering appear similar at first glance. Both systems use hierarchical, multiple-view design representation data structures and graphical editing. However, these two domains use very different design verification methodologies, each with its own optimal user interface. Each system is most useful if the users can interact using their own domain concepts, if common tasks are easy to perform, and if the systems fit into the specific work context. To be done well, both applications require that the developers go beyond their own experience and acquire an understanding of the problem space from their users' point of view. The methodology described here may help developers acquire this understanding and provide systems and user interfaces that more closely match their users' task models.

Since we use our project's hypertext platform definition process as a case study of the application of the user-centered methodology, we briefly describe the project's background and goals in the next section. The following section defines the user-centered methodology we developed and how it relates to other work. Each succeeding section details a single step of the methodology, and uses our project as a case study of how the step was applied in practice. We end with a discussion of the project's current status and a retrospective view of the contributions and general applicability of the methodology.

*The incantation means "Find any file in the current directory or its subdirectories that has been modified less than seven days ago, and if the file contains the string 'Hello' then print the file's name."

Project Background and Goals

The information interfaces project is in the Human-Computer Interaction Department of the Software Technology Laboratory at Hewlett-Packard Laboratories. The Software Technology Laboratory develops environments, tools, and methodologies to improve software productivity.

Our project's charter is to explore information management issues and technologies. We hypothesized that hypertext is a useful enabling technology in many applications that access complex information. Our approach was to demonstrate the power of hypertext by defining and prototyping a hypertext-based software platform. A software platform is a common toolkit of concepts and capabilities for use by developers of many applications. The hypertext capability makes available nonlinear text with connections (called links) among related items.

To define a platform to meet the applications' needs, we needed to understand applications and their uses in detail. This is a common problem of platform developers, yet there was no set of guidelines available. Our lab's focus on software productivity led us to develop a methodology for software developers, to apply it within our project, and to introspect about the methodology's effectiveness and impact on our success.

User-Centered Methodology

The user-centered software definition methodology we are using in this project is a six-point process. Its compo-

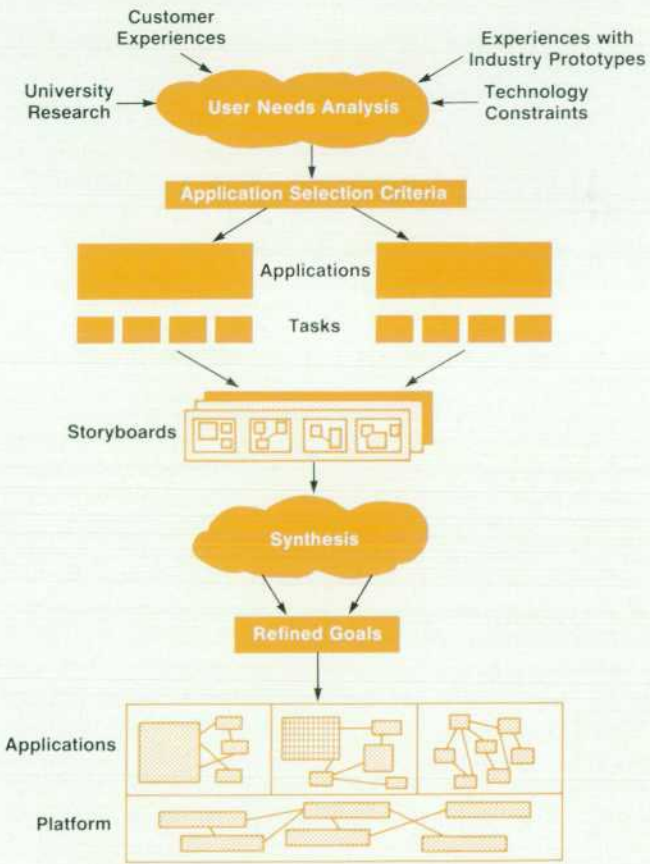


Fig. 1. *Methodology for doing user-centered application definition.*

Interviewing Techniques

The crux of the user-centered design methodology is making contact with users during the design process. A crucial step in this methodology is for designers to interview users to develop a better understanding of their needs, tasks, and goals. However, interviewing users means more than just having informal question-and-answer sessions. Properly done, interviews can provide engineers with a wealth of relevant information; improperly done, they do little more than buttress the biases the engineer brings into them.

Human factors engineers are good resources to involve early in the interviewing process. Most human factors engineers have had formal training in interviewing techniques and can assist in everything from determining who to interview and setting up an initial set of questions to analyzing the data collected. Their specialized knowledge will be especially valuable where sensitivity to subtle aspects of the situations are required. This article covers basic, straightforward techniques for those without access to more specialized help. For further information, see the references.

First, begin by choosing the right people to talk to. Know your customer environment well enough to talk to all categories of people who will be affected by the new application or have experience with a similar one. For example, when defining a project planning tool, talk to the managers who do or will use it, as well as the engineers whose managers currently use project planning tools. The engineers, even though they are not themselves direct users of the tools, may reveal issues that the current tools do not address.

Second, pick a good time. Arrange a meeting when the customer has time to focus on your questions and won't be interrupted or preoccupied.

Third, capture what was said. Either take good notes or audiotape the meeting so that both the specific problems and the overall situation will be recorded. Be aware that each hour of tape requires several hours of analysis; this can be somewhat improved if you timestamp your notes during the interview.

Fourth, studiously avoid biasing the results. Make sure your questions do not presume where the problems are, or what the right solution is. This is a common problem with nonprofessional interviewers, and an easy mistake to make, especially when the right answer seems "intuitively obvious." However, the answer that seems obvious from a designer's perspective may miss a crucial factor, and that is exactly the factor the interview needs to expose. So, ask questions about the flow of and problems with the overall task, not just the specific problems addressed by the software. You may learn that the major bottlenecks are in a completely different area.

Use good listening skills. Ask open-ended questions first, such as "What do you think of...?" and "When do you need...?". This helps to identify the breadth of issues, and lets the customer set the priorities based on perceived needs. Questions like "How would you change this?" or "What do you find most and least useful?" may identify new issues and may expose problems that can keep a product from being accepted. Rephrase important points to check that you have understood, and ask for clarification if you've misunderstood.

To understand how the application fits into the work flow, ask to see what's currently done and how it interacts with other people and tools. The customer may point out additional problems while demonstrating the use of the application. You may notice obvious places where information has to be manually computed or transcribed, although it could be handled by the computer, or where information is organized in an illogical order. You may find sequences of actions that are error-prone or could be eliminated.

Be sensitive to the social aspects of the work situation. Consider whether proposed changes in the work process would, for example, constrain or isolate people who now feel empowered and connected. A surprisingly large fraction of systems that offer excellent technical functionality fail in use because they disrupt the normal work flow or social structure of the work group. In many cases, relatively small design changes would have made these systems fit in with the preexisting social and work structures.

Iterate if possible. A few pilot interviews will give preliminary feedback and let you focus a section of the subsequent interviews on the presumed problem areas. The pilot interviews may allow you to develop a preliminary solution; you can then create a storyboard and present it to the customer for a reaction. People are much better at spotting flaws in a concrete example than generating correct solutions a priori; thus, having a shared straw man, based on the specifics of their particular situation, helps refine the proposed application. However, be careful to present the presumed problems and hypothesized solutions only late in an interview or you will sidetrack the user and never learn of more fundamental issues and problems.

References

1. G.A. Bassett, *Practical Interviewing*, American Management Association, 1965.
2. R.L. Gordon, *Interviewing: Strategy and Techniques*, Dorsett Press, 1969.

nents, illustrated in Fig. 1, are:

- Understand customer needs and available technologies.
- Select multiple contrasting tasks or applications.
- Analyze common user tasks in the applications.
- Storyboard scenarios to explore how the tasks could be better performed.
- Use the scenarios to synthesize platform goals and requirements.
- Prioritize features to implement.

This methodology is synthesized from ideas in software engineering, systems analysis, market research, and contextual field research. Traditionally, software methodologies[1,2] do emphasize "building software to meet users' needs," but do not teach how to identify the needs. They cover the transitions from requirements specifications to design and code, but give little detail on the requirements specification itself.

Systems analysis, for example Fitzgerald's text,[3] teaches the analysis of a system's inputs, operations, and outputs, but is oriented to streamlining a formalized business system, rather than unstructured information management tasks. Market research[4,5] teaches selection and analysis of potential users to identify users' general problems and potential solutions, but is mostly oriented to tangible consumer goods, not processes or complex systems. Understanding users is seen as essential, but software poses problems not found in consumer goods needs analysis. End-user customers of software often do not have enough technical computer background to know what is feasible and what they can request. Focus groups of unrelated customers often

do not provide enough detail of tasks to specify platform capabilities. However, contextual field research [6,7] provides a framework for identifying useful solutions. This technique focuses on users' problems in the context of their overall job and organizational structure, using observation and field interviews to learn what users really do and what problems they have.

Our methodology combines the process focus of systems analysis and market research, the user focus of the human-machine interaction community,[8] and the tools of contextual field research and traditional human factors engineering, and applies them to the domain of application definition in software development. Thus, we aim to provide a practical path for software developers truly to "build software to meet users' needs."

Understanding User Needs and Technologies

Studying user needs is a first step to any solution, along with gaining an understanding of available technologies and existing tools. These two tasks interact. Without an understanding of technologies one may aim for the impossible, and without an understanding of needs, one may solve the wrong problem.

A software project often begins with a general statement of need or market opportunity. The problem statement is likely to be broad—for example, "an online library catalog and inventory system," a "blood bank control and information system," or a "general multiuser information management platform." In current practice, a software engineer is typically expected to be able to build systems across this wide range of domains. To succeed, the software engineer's knowledge of algorithms and data structures must be supplemented with an understanding of the users' domain.

As mentioned earlier, it is difficult to build systems to solve needs that are not well-understood, or to build systems for users who are different from the developer. A deep understanding requires "walking a mile in the user's shoes." This can be best achieved by spending a few hours interviewing selected customers, discussing their jobs and their interactions with other people. These interviews give a richer understanding of the problem, may expose misunderstandings and hidden constraints, and can point out novel solutions. See the box on page 92 for the interview guidelines we followed.

In addition to exploring user needs, the software engineers must actively learn about the available technologies and state-of-the-art products by comparing product goals and successes and learning how these technologies interact. The goal is to identify technologies with the potential to help meet the identified customer needs. The result of this approach should be a well-understood set of user needs and knowledge about relevant technologies and potential solutions.

Since we are a research group, our focus is necessarily somewhat longer-term than that of a product development team. We are interested in research systems as well as products already in the market. When we approached the task of understanding the available technologies and customer needs, we compiled a bibliography of papers on hypermedia research and discussed the state of the technology with our potential user community. We organized a study group in which we distributed and discussed selected papers and analyzed the goals and capabilitities of existing systems. We evaluated their usefulness for different tasks and what was required to implement them. We also invited speakers to give their perspectives on the field and on their needs. These contacts exposed us to a diverse range of applications and customer needs and helped us gain a better understanding of their functionality requirements.

Selecting Multiple Applications

After selecting a general area to be supported by a software platform, detailed requirements must be identified. Platform designers need diverse, representative, example applications to help determine and validate the platform's specific capabilities. A good validation application is one that could clearly benefit from the proposed solutions, is based on technology available to the customer, is doable, and will exercise many features of the platform (otherwise it would not help define and constrain the platform). We found it useful to define a table of technical and pragmatic criteria for use in selecting an application.

Examining multiple applications helps to avoid false optimizations, identify the crucial requirements, prioritize features, and decide among trade-offs, and may define the boundaries of the problem space. The applications we chose were a multiuser, annotatable, hypertext-structured, on-line manual and a community calendar interface to group information. Each application appeared doable, could be important to HP, and could clearly benefit from hypertext capabilities. However, the two applications emphasized very different mixtures of platform capabilities.

When designing an application instead of a platform, the requirements definitions must be more specific and should cover the full range of user tasks rather than just representative ones. For example, application designers might consider both data generation and the specific requirements for producing a variety of graph types on multiple output devices rather than simply looking at general graphing functionality requirements. To ensure coverage of the full range of tasks, developers should consider the application users' goals and their associated tasks.

Analyzing Common Tasks

The goal of this step is to experience a user's tasks and problems and the currently available applications and tools first-hand, thereby building a detailed understanding of the issues involved. For each chosen application, the common tasks should be listed and characterized, preferably with the help of a "customer." Then, each project member should select a specific task in the list. This should be a concrete, nontrivial task—for example, "Find the last eight quarters' earnings and expenses and present the result as a slide containing a bar graph."

Using the existing tools, each engineer then records the specific steps required to accomplish the goal, frustrations encountered, sources of problems or mistakes, and ideas for new tools. Each engineer also interviews current users of similar products, or potential users, to learn how they would go about solving the problem. In addition to observing current work styles and tool use, the interviewer also asks open-ended questions about the user's needs for com-

puter support in the area, and requests prioritized suggestions for improvements.

The result of this process is an insider's understanding of the tasks and the aspects that make the current solution frustrating or time-consuming, plus a developing vision of a better solution.

After discussing the goals and expected results of the task analysis, each engineer in our project selected an example of a common, nontrivial task in the target application. We conducted interviews with users and gained a number of surprising insights into what users really do, how they use or why they avoid a system, and what they would really like to have help with. We confirmed that even a few hours of interviews can be very valuable, even when conducted by engineers who have had only brief training in interviewing and user needs analysis. We were fortunate to have a couple of project members who had been formally trained in interviewing techniques and were able to brief the rest. An overview of the guidelines we use is presented in the box on page 92.

One sample task involving the manual for a document preparation system was to discover how to print labels for three figures on a page: two for figures in the upper quadrants and one across the lower half of the page, as shown in Fig. 2. In analyzing the way in which the current paper manual is used, it became clear that there are some strong advantages of a paper document that we would want to preserve, and some glaring problems where we could make a contribution. A physical book can be skimmed, highlighted, dog-eared, and written on. On the other hand, a good solution using this particular system requires understanding of the concepts and syntax of "floating object," "figure," "caption," and "fragile command," and it was difficult to identify all the correct names for the concepts and then flip among the scattered descriptions to understand how they fit together.

In solving this problem using the manual, the project engineer would skip back and forth among about eight groups of pages that described the concepts, a tutorial section, and the index, using pieces of paper, pens, and fingers holding pages open as bookmarks to facilitate moving among the eight sections of text. It was frustrating not to be able to differentiate index entries for the main definition of a concept, an example, or a mere mention in the tutorial. Other problems included the facts that electronic mail concerning use suggestions, questions, and bugs in the system was kept elsewhere and not indexed in the manual, that there was no good way to add new entries to the index, and that once the problem was solved there was no easy way to annotate the book with the solution and links to the relevant entries or to share the solution with colleagues.

We found that working through this example task using the current system was quite revealing. In particular, being exposed to the low-level interaction details brought out user issues and requirements that might otherwise have been overlooked.

Storyboarding Scenarios

Based on the previous step's learning process, analysis, interviews, and suggestions, each engineer prepared a report summarizing the user problems to be solved and ideas for an improved application. The engineer's presentation of this report to the group tended to enhance the organization of the design process and to expose the proposed solution to the different mental models, approaches, and domain knowledge of the other group members.

Part of the report is a sample scenario of a user's path through the improved system, showing how the user solves the selected problem. This scenario is presented in the form of a storyboard (see box, page 95). It may be a mostly textual, sequential description of the low-level steps the user goes through, including a description of what tools and information are visible at different times, how the user navigates through the system and uses the help facility, and how data is moved from one tool to another. The scenario may also be a graphical sequence of screen images, with captions that describe the transitions and user actions.

The goal of the scenario is to clarify assumptions about the user's goals and context and to expose missing components, awkward links between steps, and constraints on the technology. The scenario presentation includes a list of technical capabilities it assumes, which represents functionality the team must either get or build. At this point, the scenarios communicate individuals' visions to the group and do not represent a group vision; in fact, diversity of perspectives is an advantage.

After each scenario is presented, the group discusses any omissions and brainstorms improvements. The resulting requirements are categorized as either resources to be provided by other groups, capabilities to be built by the project team, or paradigms and tools that are useful additions to be supported but not built. These lists form the starting points of the synthesis phase.

We found that succeeding scenarios built on previous ones, but also had very different foci—multiuser interactions, help and error recovery, performance, or navigation in complex information. This diversity was especially

(continued on page 96)

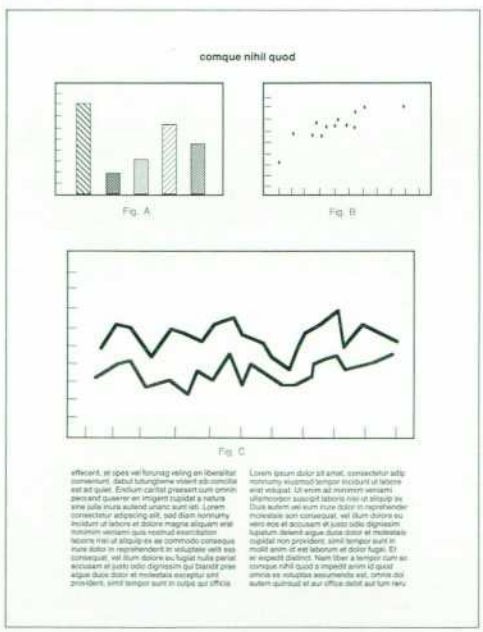


Fig. 2. *A sample task for a document preparation system involved printing labels for three figures on a page.*

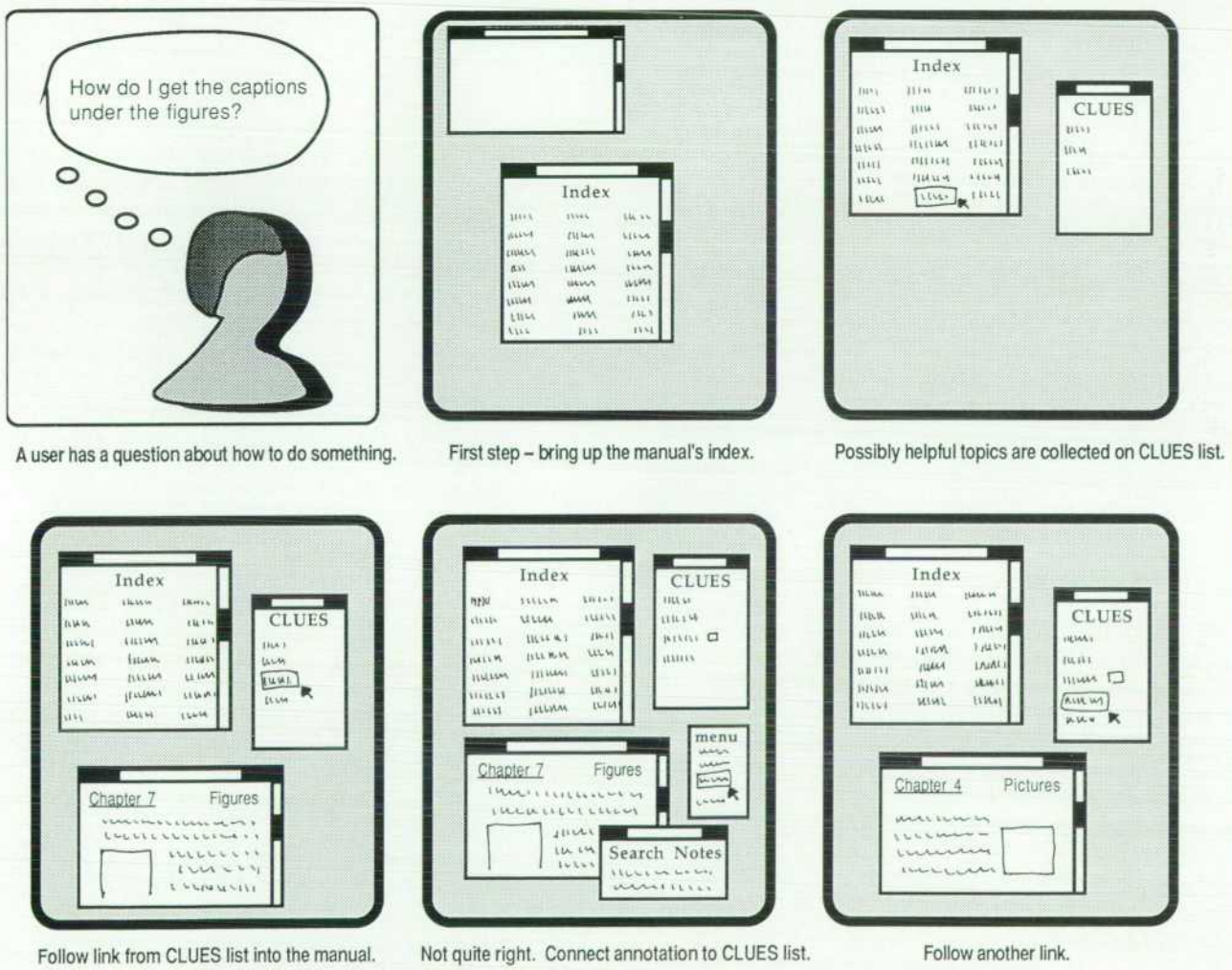
Storyboarding Techniques

To most people, the word "storyboard" probably means animators sketching scenes for a cartoon or movie makers planning a filming sequence. But the storyboarding technique can be used to organize information in many other situations. In particular, it is a fine tool for the design phase of a software project. The storyboard is also a powerful process tool for keeping design meetings focused and for systematically exploring whether alternative designs meet all requirements. After a group storyboard has been developed, it provides an explicit, permanent, and shared record of design issues and decisions. This record facilitates intergroup and intragroup communication and serves as a planning board with which teams can track implementation progress. In our design phase, because we were not making design decisions for a specific implementation, we modified the usual approach. Our individual storyboards served as multimedia tools for communicating with one another how our user scenarios would unfold (Fig. 1).

In the approach we describe here, the elements of the storyboard are organized in a two-dimensional matrix of categories and details. This structure helps organize and point out gaps in a scenario or a design. It also brings all the design information together in one place, and in so doing helps team members avoid the tendency to remember only information relevant to their own concerns. Spatial structure, colors, and symbols are memory aids, and can be used to help organize the information on a storyboard.

The basic storyboarding technique is simple. In design meetings, ideally guided by a skilled meeting facilitator, the project team alternates in short cycles between generating ideas in a brainstorming phase and then evaluating them. Simple materials suffice: notecards and a corkboard are better than sheets of paper or flip charts. Spreadsheet software (treating cells as notecards) can be used to create a more portable record.

In the brainstorming phase, participants should generate ideas



A user has a question about how to do something.

First step – bring up the manual's index.

Possibly helpful topics are collected on CLUES list.

Follow link from CLUES list into the manual.

Not quite right. Connect annotation to CLUES list.

Follow another link.

Fig. 1. *Part of a storyboard for a user scenario. This example is untypical because a computer was used to prepare it (most are only sketched).*

freely, without evaluating them. Ideas are recorded on notecards and presented to the group. Team members should try to build on previous ideas—creative ideas come as often from rearranging existing concepts as from original thinking. The flow of ideas in this phase should feel like popping popcorn, and should take just about as long.

In the evaluation phase, the team critically examines the ideas and filters out inappropriate ideas and duplicates. They categorize, sequence, and prioritize the accepted ideas, and flag items that require further investigation. If an issue or view has a champion but not wide acceptance, the team can defer conflict over the item by assigning the card to a "pending" category. The champion should then attempt to win the support of other team members in off-line discussions. In subsequent design sessions, the issue can be reevaluated.

The storyboarding technique has several benefits. Brainstorming is a good method for tapping creativity. Writing ideas on cards captures ideas that might otherwise get lost, and permits even the quieter team members to have an equal voice. It also avoids the slowness and inherent filtering that happen when there is only one person recording ideas on a flip chart. Because deferred evaluation permits a more rapid and uninterrupted flow of thoughts, new ideas can more easily build on previous ones.

Evaluating groups of ideas together, rather than individually, defuses individual ownership of ideas and permits a more thorough exploration of design alternatives. It also ensures that all issues are examined and negotiated explicitly. The process facilitates consensus building and team ownership of the design; the publicly accepted and recorded results become "community property."

The resulting storyboard is a superb tool for communicating design decisions and checking the validity of design assumptions with a team's customers. It can be quite effective in helping customers feel that they own the decisions along with the designers. Involving customers in the design stage can help identify ineffective design choices early, thus saving the expense of late-stage redesign.

Cathy Fletcher
Member, Technical Staff
HP Laboratories

strong because of the variety of backgrounds and interests in our project, and greatly enhanced the quality and versatility of the resulting platform. However, even in a homogeneous group, we feel that members should try for a diversity of approaches because it leads to insights and exposes issues.

Synthesizing Goals and Requirements

The first step in synthesis is to reexamine the original problems to be solved and refine them according to the user feedback and task analyses of the above steps. Then, use the project goals to combine and filter the lists of requirements from the storyboarding step.

For a software platform this results in three lists: the platform requirements and programmatic interface, the set of capabilities to be built by the project, and the functionality expected to be imported from other projects or products. For an end-user application, the lists would be: user-level functional requirements including extensibility and customization hooks, capabilities to be built by the project, and capabilities expected to be imported.

After we discussed each task and collected the lists, we began the synthesis phase. We reexamined and agreed on the results we wanted to achieve. Our goals were to understand the interaction of programmatic and user interface issues in the proposed group environment for information management and to build a prototype hypertext platform and application prototypes to validate it. When we agreed on the goals, we were able to combine the various lists of requirements into a platform definition.

Prioritizing Features

The last step is to set up a plan of attack for the project. The goal is to define a phased process, so that critical issues can be resolved quickly and problems identified, and yet provide a vision of the full system.

Especially for a new domain, or a combination of new technologies, it is critical to constrain the size of the first prototype. It is common to underestimate the problems in a new area, so one needs to make sure that the first pass is doable. In building an application, it is often possible to plan a vertical slice of useful functionality and a user interface, which may then be given to users for early feedback. A platform may not have a user interface and may not solve any application's complete problem. This makes getting early user feedback difficult and makes it even more important for the developers to match the application requirements carefully to the platform capabilities and to build and test the critical capabilities first.

Defining a core is not enough; it is also necessary to define the full system, to consider the dimensions in which the system should be extensible, and to understand how such extensibility will be designed in. These longer-term plans need not be as detailed as the definition of the core features, but must be detailed enough that the full functionality and extensibility requirements can be anticipated and incorporated into the initial design.

We reorganized the requirements lists into three groups, corresponding to three implementation phases. In the first layer were items whose design had fundamental impact on our architecture and that would provide a core slice of multiuser functionality. The second layer defined the higher capabilities needed to provide a programmatic interface for hypertext application developers. Our top layer's list included the new tools and organizing concepts we felt would be useful in various end-user applications but were not part of the core platform.

Our goal was to define an extensible platform architecture with which we could validate our theories or expose problems early. In our first-pass prototype our plan was to design and implement core data structures, algorithms, and communication strategies. We designed and implemented a simple, extensible user interface, which was adequate for our initial applications.

For our second pass, we planned to revise the basic architecture based on lessons learned in the first prototype, and to have a couple of applications that would be sufficiently complete and efficient to allow us to evaluate daily

use by a few dozen users. We even planned a third-pass system, which had much richer functionality. We didn't design that fully, but checked our fundamental decisions to see if they would meet the third pass's functionality requirements.

Current Status and Evaluation

Our methodology greatly influenced our design of the platform for hypertext applications. In design meetings we often referred back to a scenario to identify the user model we wished to support or to check the impact of a design decision. This simplified the design discussions.

We have now built the first-pass prototype platform and are analyzing how it was used to build a computer-conferencing application. This application was designed and built as the thesis project of an MIT Master's student intern. It has been a fine validation system for our platform, since the student did not participate in our definition phase and this system was not one of the applications we had analyzed in detail. Our platform matched the conferencing application's requirements quite well. Only one missing concept was identified, which required only minor redesign. On the whole, the platform design was quite robust; our planned versatility paid off well.

We expect to use storyboarding again before we prototype one of our validation applications. This time we would use it as a group tool to build consensus and to define and document a shared view of the application's goals, constraints, and tasks.

Contributions and Applicability

This paper has tried to show that to develop applications that truly help users, developers must understand their users' needs and constraints. While this may seem obvious, it is tempting to begin immediately to "solve" quickly, rather than taking the time to probe and listen with an open mind. A principled, objective study of users and their overall goals will elicit their real obstacles and needs and simplify the selection and design of the appropriate technical platform application solutions.

Our user-centered methodology contributed to our goals in many ways. We found that the analysis of how existing systems are actually used, the focus on two diverse applications, and the analysis of multiple scenarios of user tasks were very revealing. By focusing on representative tasks and developing realistic scenarios, we identified how people actually do tasks, what their concerns are, and where their current methods fail to meet their needs. In so doing, we identified the essential capabilities of the hypertext platform and a set of desirable extensions, the crucial features for initial test applications, and evaluation questions for the platform prototype.

The selection of criteria for user applications and analysis of user needs took less than two weeks, including the synthesis of platform requirements. Our results let us prioritize the capabilities to provide, and also suggested new tools for organizing the richly interconnected information. The synthesis phase resulted in a shared, documented vision, which focused and guided decisions throughout the design and implementation phases. The methodology also provided validation metrics for the resulting platform via two test applications which represent diverse uses for the platform.

We believe that our strategy was a successful implementation of a user-centered analysis of a new technology, and that this simple and powerful methodology is widely applicable to application design both within and outside Hewlett-Packard.

Acknowledgments

We wish to thank Nancy Kendzierski (our early project manager and the current department manager) and Robin Jeffries for leading this project and for their insightful comments on the HP Technical Women's Conference talk and on drafts of this paper. Cathy Fletcher designed all the illustrations and ably helped revise many sections. Bob Leichner, Jarrett Rosenberg, and Vicki O'Day participated in the definition and implementation of our system and gave helpful comments on this paper. Michael Creech, Cathy Fletcher, Dennis Freeze, Andreas Paepcke, and Warren Harris generalized and adapted their experimental technologies to meet our platform's needs. Shari Jackson, our MIT student intern, defined and built the conferencing application which helped us test and refine our platform. Jean Gascon (then of HP Laboratories), Merijane Lee (then of the HP Application Support Division), and many others provided insight through discussions of their real information management tasks. Will Willams and Bryan Hoover of the HP Systems Technology Division discussed the central role that storyboarding played in their project and analyzed their use of the storyboarding methodology in requirements analysis and design.

References

1. M.A. Jackson, *Principles of Program Design*, Academic Press, 1975.
2. M. Page-Jones, *The Practical Guide to Structured Systems Design*, Second Edition, Yourdon Press, 1988.
3. J. Fitzgerald, A. Fitzgerald, and W.D. Stallings, Jr., *Fundamentals of Systems Analysis*, Second Edition, Wiley, 1981.
4. J.L. Pope, *Practical Marketing Research*, AMACOM, 1981.
5. G.A. Churchill, Jr., *Marketing Research*, Fourth Edition, Dryden Press, 1987.
6. J. Whiteside, J. Bennett, and K. Holtblatt, "Usability Engineering: Our Experiences and Evolution," in M. Helander, editor, *Handbook of Human-Computer Interaction*, Elsevier, 1988, pp. 791-817.
7. M. Good, chair, "Experience with Contextual Field Research," *CHI'89 Conference Proceedings*, 1989, pp. 21-24.
8. J.D. Gould and C. Lewis, "Designing for Usability: Key Principles and What Designers Think," in B. Shackel, editor, *INTERACT '84: First Conference on Human-Computer Interaction*, North-Holland, 1985.

Partially Reflective Light Guides for Optoelectronics Applications

The guides control the light from an array of light-emitting diodes in a high-performance, low-cost erase bar for electrophotographic copiers.

by Carolyn F. Jones

THE OPTICAL DESIGN for HP's programmable erase bars is an elegant application of the optical properties of dielectric materials. In newer generation electrophotographic copiers a selective erase function is used for a variety of editing functions, from erasing areas as small as single characters to adding picture frames. In spot color copiers it is used to transform selected areas of a black-and-white original into a specific color, giving colored words, logos, frames, and background blocks. For the erase function, an array of light-emitting diodes (LEDs) offers many potential advantages over other technologies.

A high-performance, low-cost optical design was the key to making the LED erase bar a viable reality. The application requires that discrete, well-defined spots of light be projected onto a photoreceptor located some distance away from an array of LED light sources spaced on about one-millimeter centers. LEDs emit light in all directions and the diverging light rays can be difficult to control. The background light level must be low, meaning that light rays not forming part of the desired signal must be eliminated. Traditional optical designs were not satisfactory because of the tight array spacing, the large number of LEDs, and cost constraints.

The solution developed for this problem combines a simple design with proper choice of materials. The reflective and refractive properties of opaque dielectrics are used to enhance the signal and extinguish the noise simultaneously. The desired signal is channeled and focused by reflection under conditions where the reflectivity of the dielectric is high. The unwanted light is extinguished in two ways: by multiple reflection under conditions where the reflectivity of the dielectric is low, and by absorption in the opaque dielectric.

The main element of the design is an optical baffle which can be positioned some distance away from the light sources without loss of performance. This flexibility makes the erase bars easier to produce and assemble into a product.

The partially reflective optical light guides operate with the refractive index of the guide lower than the refractive index of the containing medium.[1] This is the converse of optical light guides that use the principle of total internal reflection. This paper presents the theory and principles governing the design and shows an example of implementation.

Theory

The partially reflective light guides are based on Fresnel reflection of light from dielectric surfaces. This theory can be found in any textbook on classical optics.[2]

We were interested in the case where the refractive index of the reflective surface is larger than the medium containing the incident beam, specifically in the case of light reflecting in air from a plastic surface (see Fig. 1).

The reflectivity $R(\theta)$ is the fraction of light reflected from the surface, or the ratio of the reflected intensity to the incident intensity. It is also related to the reflection coefficient $r(\theta)$:

$$R(\theta) = \frac{\text{reflected intensity}\,(\theta)}{\text{incident intensity}\,(\theta)} = r^2(\theta) \qquad (1)$$

The manner in which the light is reflected depends on the refractive indexes of the materials (n_1 and n_2), the angle of the incident beam (θ_1), and the polarization of the beam. The reflection coefficients are generally described separately for two polarization components depending on the orientation of the electric field: parallel and perpendicular to the plane of reflection. These can be expressed most simply in terms of the associated angle of refraction θ_2,

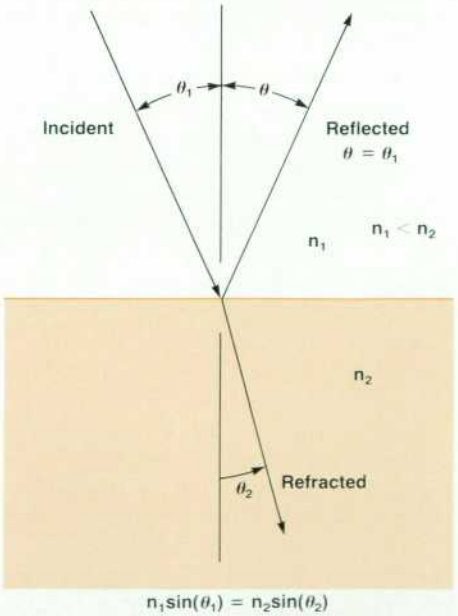


Fig. 1. *Reflection and refraction at a dielectric boundary. The refractive indexes of the two materials are n_1 and n_2.*

defined by Snell's Law:

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \quad (2)$$

$$r(\theta)\,\text{parallel} = \frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)} \quad (3)$$

$$r(\theta)\,\text{perpendicular} = \frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)} \quad (4)$$

Brewster's angle θ_B is the angle at which the reflectivity for the perpendicular component goes to zero, or $(\theta_1 + \theta_2) = 90°$. Using Snell's law this condition can also be expressed as:

$$\tan(\theta_B) = n_2/n_1. \quad (5)$$

Fig. 2 shows typical reflection coefficients and reflectivities for the two polarizations as a function of incident angle (using an air-dielectric system where $n_1 = 1.0$ and $n_2 = 1.8$). At normal incidence ($\theta_1 = 0$) the reflectivity is low. For angles up to about Brewster's angle, the average reflectivity remains low and is not significantly higher than the reflectivity at normal incidence. For angles greater than Brewster's angle, the reflectivity climbs rapidly to 100% at $\theta_1 = 90°$.

Reflectivity of Opaque Dielectrics

For our purposes, we were interested in looking at reflections from filled dielectrics, especially molded plastics, and specifically on the impact of the filler on the light reflected from the surfaces of these materials.

Going back to theory, the surface reflections from a dielectric depend on the refractive index of the material within a few wavelengths of the surface. Inclusions, boundaries, and changes in the material at depths below the surface should have no effect on the surface reflections (provided that these discontinuities are not regular and do not contribute to interference phenomena). The material below the surface may add to the total reflected light and increase the measured reflectivity, but it will not subtract from it. The only real requirement is that the material have a specular surface.

For molded plastics, the skin on the surface is resin-rich and the filler particles do not intersect the surface. Therefore, we believed that surface reflections from a filled molded plastic would depend only on the refractive index of the matrix resin and not on the filler.

This was demonstrated by experiment. The reflection coefficients were measured on three samples of molded polysulfone: clear (no filler), white (titanium dioxide filler), and black (carbon filler). Within experimental error, there was no measurable difference from one sample to another

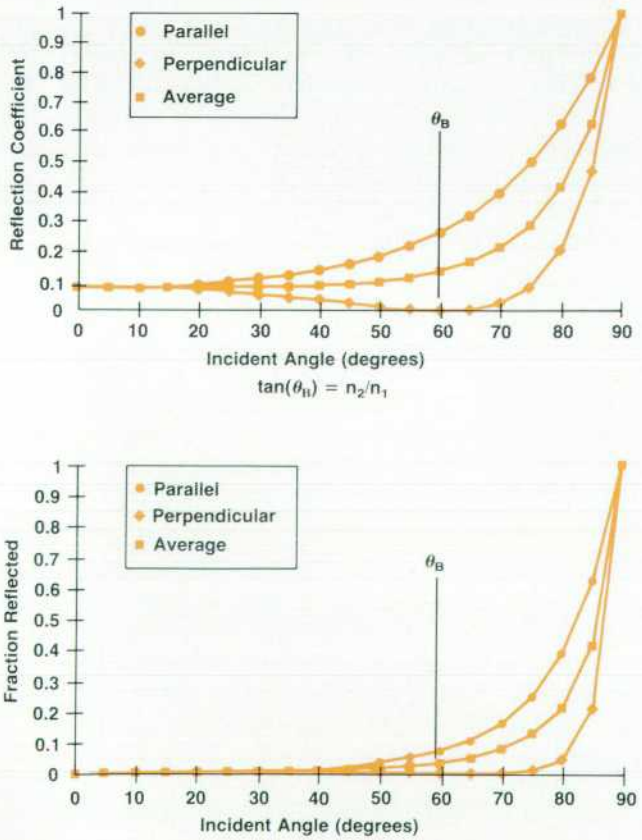


Fig. 2. *(a) Reflection coefficient as a function of angle of incidence for an air-dielectric boundary where $n_1 = 1.0$ and $n_2 = 1.8$. (b) Reflectivity for the same conditions.*

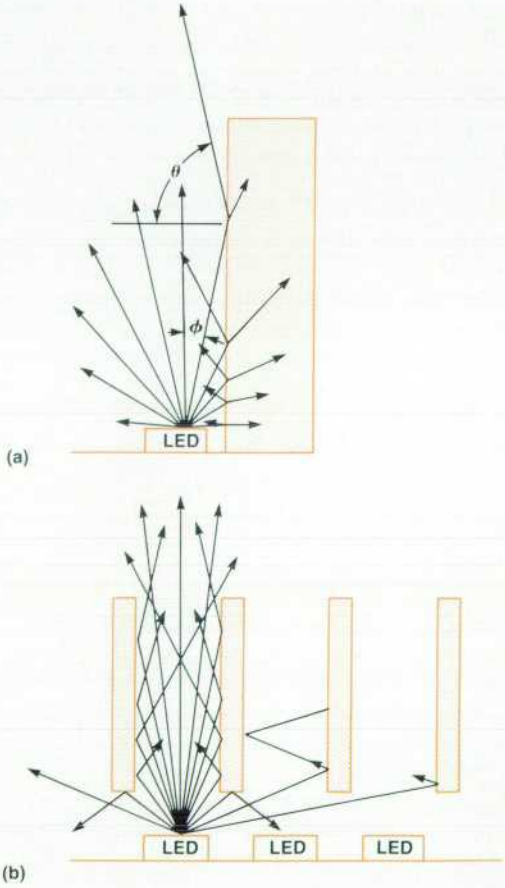


Fig. 3. *(a) Effect of placing a wall of opaque dielectric material next to a light-emitting diode. (b) Effect of adding a second wall, forming a well.*

and all agreed with values calculated from equations 3 and 4 using the refractive index of the polymer material.

Certainly light refracted into the material will be scattered by the filler particles. Some of this light may reemerge. Although the optical scattering and absorption characteristics were not explicitly measured, the fact that heavily filled, opaque dielectric materials have high optical absorption coefficients indicates that any light refracted into the bulk material will have a high probability of absorption by the material. Therefore, the refracted light should not tend to reemerge from the front surface as scattered reflection and should not tend to be transmitted through the material to emerge at the back surface.

Since refracted light could be a potential problem in a reflective optics design, and since filled, optically opaque dielectrics will tend to absorb the refracted component without degradation of the surface reflections, optically opaque dielectrics would appear useful in reflective optics designs.

Light Pipe Principles

In this particular optical problem, we needed to project well-defined spots of light from an array of light sources emitting in all directions. Light rays emerging from an LED at angles far from the optic axis are troublesome because they are difficult to direct into the desired signal beam. This application required the background light level to be low, meaning that light rays not forming part of the signal beam had to be eliminated.

Looking at the effect of placing a wall of opaque dielectric material next to the LED shows how the reflective characteristics can be used to advantage (see Fig. 3a). Light emerging from the LED at angles close to the optic axis will hit the dielectric at angles close to the grazing angle. This is the condition where the reflectivity of the dielectric is high. Light emerging from the LED at higher angles will hit the dielectric at angles where the reflectivity is lower. At these angles, the majority of the light will be refracted into the dielectric and tend to be absorbed.

If a second wall of dielectric is added, forming a well, multiple reflections will take place (see Fig. 3b). If the height of the well is finite, light reflected at grazing angles will tend to be propagated because the reflectivity of the walls is high, while light reflected at angles closer to the normal of the dielectic surface will undergo multiple reflections, losing intensity with each bounce, and will tend to be extinguished.

For this particular application, straight walls parallel to the optical axis of the LED gave adequate performance. It would appear that the same principles of partial reflection at close to grazing angles could also be applied to reflective optics designs other than the simple case presented here.

Design Considerations

Let us define some angles in terms of the physical dimensions of the well, as shown in Fig. 4. The dimensions are:

a = half width of light guide
b = height of light guide measured from source plane
c = center-to-center spacing of light sources ($c = 2a + w$)
d = half width of primary beam at image plane
h = distance of image plane from top of light guide (working distance)
s = cutoff at bottom of light guide
w = light guide wall thickness.

We define the following angles:

$\tan(\phi_1) = a/b = d/(b+h)$ cutoff for primary beam (6)

$\tan(\phi_2) = 3a/b$ cutoff for propagation with one reflection (7)

$\tan(\phi_3) = a/s$ cutoff for reflection into cavity (8)

$\tan(\phi_4) = (c+3a)/b$ cutoff for one reflection into adjacent cavity (9)

$\tan(\phi_5) = (c-a)/b$ cutoff for cross talk into adjacent cavity (10)

$\tan(\phi_B) = n_1/n_2$ reflection from walls at θ_B. (11)

From simple arguments based on geometry and reflection characteristics, restrictions can be placed on the design of the well for implementation with a point light source.

For signal enhancement, the light should be propagated under conditions where the reflectivity of the material is high, requiring:

$$\phi_2 < \phi_B \quad (12)$$

$$3a/b < n_1/n_2 \quad (13)$$

There is no point in trying to capture light that cannot be efficiently propagated either because of low reflectivity or because of multiple reflections. Up to some height, the bottom section of the well contributes virtually nothing to the signal.

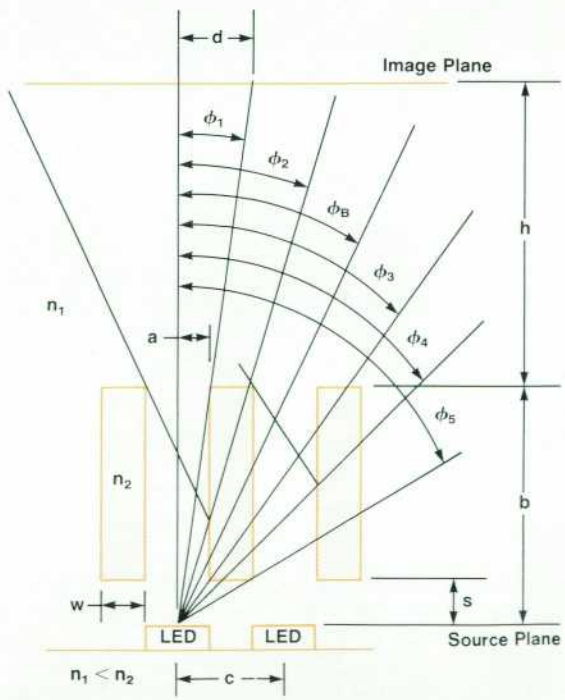


Fig. 4. *Angles and dimensions for design calculations.*

Up to a height s_B associated with Brewster's angle, the walls have low reflectivity:

$$\phi_3 = \phi_B \quad (14)$$
$$s_B = an_2/n_1 \quad (15)$$

Up to a height s_R the walls only contribute to multiple reflections:

$$\phi_3 = \phi_2 \quad (16)$$
$$s_R = b/3 \quad (17)$$

To minimize cross talk in an array, light entering an adjacent cavity must be baffled out by multiple reflections. This leads to the requirement:

$$\phi_5 > \phi_4 \quad (18)$$
$$s < b(c-a)/(c+3a) \quad (19)$$

The highest and most uniform beam intensity occurs when the working distance is as small as possible. This also gives the smallest spot size in the image plane. In the limit,

$$h = 0 \text{ and } a = d. \quad (20)$$

At larger working distances, the condition where the two reflected beams just meet in the center, but do not overlap, also gives a reasonably uniform beam. This condition occurs when:

$$h = b \text{ and } a = d/2. \quad (21)$$

For working distances beyond overlap, the beams begin to diverge with loss of definition and intensity.

The range of reasonable operating conditions for intensity and uniformity would appear to be bounded by equations 20 and 21. For our purposes the optimum conditions are probably close to the conditions outlined in equation 21:

$$h \leq b \text{ and } d/2 \leq a \quad (22)$$

Although the LED is not a point source, the key features and design requirements are evident using simple point-source considerations. A comparison of predicted optical characteristics as a function of working distance for a point source and a finite source is shown in Fig. 5. In concept, a finite source can be viewed as an array of point sources distributed over the width of the LED. This has the effect of smearing each of the cutoff angles over a range $\Delta\phi$ described by the included angle of the emitting area. Closed-form analysis of a two-dimensional distributed source is difficult since the axial rays are the only ones that are simple to describe.

The optimum dimensions for a light guide can be calculated from these design considerations. For our particular application the array spacing and working distance were given and we chose to design using the conditions outlined in equation 22. By using the values calculated for a point source and designing for a primary beam width equal to the array spacing, we were essentially using the physical size of the LED to provide the required image overlap.

Table I shows a logical sequence for developing the optimum light guide dimensions using the above design considerations and working from the given specifications. For

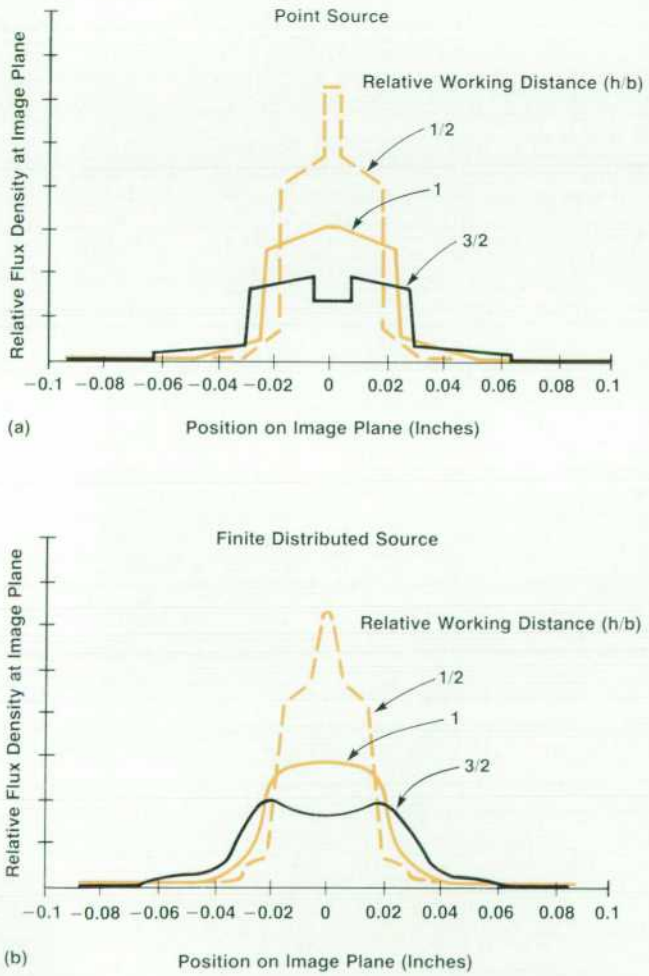


Fig. 5. *A comparison of predicted optical characteristics as a function of relative working distance for a point source and a finite source.*

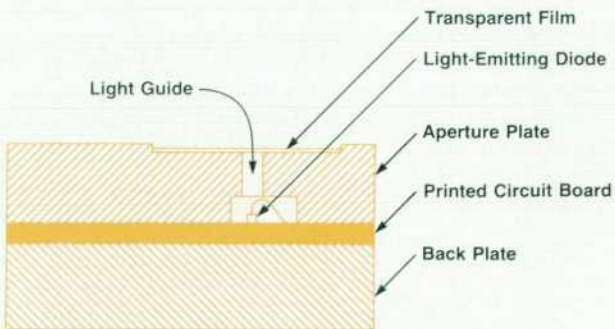


Fig. 6. *Cross section of a typical light guide.*

comparison, the actual values used in the design are also shown.

Table I

Attribute	Symbol	Equation Used	Given	Calculated Value (in.)	Actual Value (in.)
(23) Spot Size	2d	2d = c	LED spacing c = 0.040 in. No overlap of primary beam.	0.040	0.040
(24) Light Guide Width	2a	22	d from (23)	≥0.020	0.022
(25) Light Guide Height Minimum	b	22	Working distance specification h = 0.070 in.	≥0.070	0.071
(26) Light Guide Height Minimum	b	13	2a from (24) $n_1 = 1.0$ $n_2 = 1.8$	>0.065	0.071
(27) Unnecessary Height at Bottom	s_B	15	same as (26)	0.013	0.022
(28) Unnecessary Height at Bottom	s_R	17	b from (25)	0.023	0.022
(29) Maximum Cutoff at Bottom	s	19	b from (22)	<0.030	0.022

Other Design Aspects

There are other aspects of the design that impact implementation.

Short Working Distance. This optical design is not intended to transmit light making more than one reflection from the light guide walls and is limited to short working distances. The design considerations place restrictions on the light guide dimensions.

Other constraints come from implementation requirements. Flux density requirements can limit the working distance because the flux density at the image plane decreases roughly as the square of the distance from the light sources. There are also fabrication limitations associated with closely spaced holes that have a large aspect ratio—this can impact the maximum baffle height. For our application these constraints were not a problem because the specified array density and working distance allowed operating with an optimum design and were consistent with available fabrication technology.

Clearance at Light Source. As noted earlier, the containing walls do not have to extend all the way to the light source plane. For implementation with a hybrid circuit, this allows clearance for semiconductor chips and wire bonds, which sit above the light source plane. (The bond wires making contact to the LED chips do protrude above the top surface of the LEDs.) Reflective cavities with the walls surrounding the LED, or at least in close proximity, are difficult to implement because any contact with the LED or its bond wire can cause mechanical damage and lead to device failure.

Since the bottom section of the cavity is unnecessary for optical performance, it can be eliminated, thus avoiding all of the problems associated with close proximity as well. There is ample clearance for bond wires and other components, allowing x-y freedom during optical alignment without risk of damaging the assembly. The design is relatively insensitive to LED placement within normal process control limits. The piece parts can be fabricated without ultratight tolerances and assembled using simple techniques.

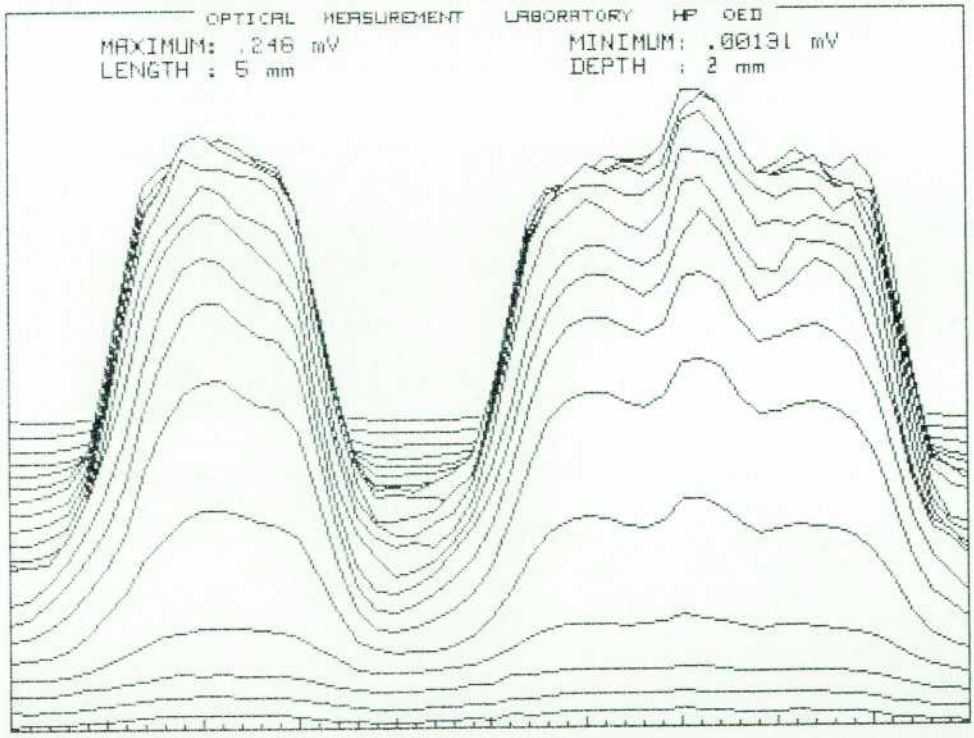


Fig. 7. *Optical performance of the LED array for a pattern of one LED on, one off, and two on.*

Air Cavity. The requirement that the refractive index of the optical cavity be less than that of the containing walls allows the optical cavity to be air. Many other optical design concepts used in optoelectronics are based on fiber optics or immersion optics where the LED must be encapsulated in a medium of higher refractive index. These designs have serious potential problems in device reliability, resulting from a variety of causes.

There is a limited choice of optically clear materials. The inherent thermal expansion coefficient mismatches in the materials system can increase the probability of device failure during normal operation.

If thermosetting plastics are used, any thermal expansion coefficient mismatches are aggravated by the built-in strain developed as a result of shrinkage during the polymerization process. These problems are compounded as the physical size (or complexity) of the assembly increases.

Any coating on the LED also has the potential problem of contamination by entrapping, providing a transport vehicle or even being a source for undesirable ions. The presence of mobile ions can impact device reliability by causing an abnormally high degradation of the LED light output with time. The mechanisms involved in this degradation are not well-understood, are difficult to predict, and can be sensitive to subtle variations in process and materials. The degradation is seen only when the LEDs are stressed in operation, making it necessary to perform extensive reliability studies.

The ability to implement this optical design without having to encapsulate the LEDs is a benefit because all of the associated reliability concerns are avoided. The absence of encapsulant also allows flexibility for repair.

Application

The erase bar is used for a selective erase function in newer-generation electrophotographic copiers. Since the LEDs are individually addressable and spaced about one millimeter apart, areas as small as single characters can be erased from the photoreceptor before printing the copy.

This capability allows a multitude of editing functions such as cleaning up the edges, deleting sections of the original, and adding picture frames. In color copiers with spot color features, the erase capability can select areas for transformation into color by erasing all but the part of the image to receive a specific toner color. For example, colored words, frames, logos, or background blocks can be generated from a black-and-white original.

Performing the erase function imposes severe requirements on the light intensity profiles at the image plane. The profiles of individual LEDs must be uniform, have controlled and well-defined light-to-dark transitions, completely fill a specific area, not fill beyond that area, and not contribute to background scattering in other areas. Adjacent LEDs must also give a uniform profile, without gaps or overexposure.

The partially reflective light guides described in this paper were first used in HP's HEXP-GM01 LED erase bar, a 14-inch array of 352 individually addressable LEDs with drive electronics and optics in a compact assembly.

Implementation

The first-generation light guides were fabricated from black anodized aluminum. Working models could not be machined from filled plastic since the optical performance depends on the presence of a dielectric skin, which might have been destroyed by the machining process. Plastic versions would have to be tooled and molded.

Aluminum was chosen for prototyping purposes because it was easy to machine and the sapphire skin grown during the anodizing process provided the necessary dielectric interface for surface reflections. The black dye served to absorb the light refracted into the surface so that it did not emerge as reflections from the underlying aluminum. The reflectivity of the black anodized aluminum was measured on an optical bench and exhibited the surface reflectivity characteristics predicted based on the refractive index of sapphire. Machined versions also offered the advantage of quick turnaround and flexibility for design modifications without the capital investment of hard tooling.

The optical baffle was fabricated in two sections with a series of shallow saw cuts forming a row of teeth, which interdigitate when mated, creating an array of rectangular apertures. Even using a machined optical baffle, the optics had better performance and were less expensive than other alternatives, specifically self-focusing lens arrays. The combination of the optics, the LED array, and the hybrid circuit gave HP a cost-effective solution with performance superior to other technologies for this application.

The first-generation product was developed under a tight schedule and designed against changing specifications. As a result, some of the dimensions used were empirical and the design was frozen once satisfactory performance was achieved. Although the design might have been adjusted for ultimate performance and realized in molded plastic instead of anodized aluminum, the improvements did not justify the time, risk, and other costs.

A cross section of a typical light guide cavity is shown in Fig. 6. The light guides were constructed using the dimensions given in Table I. A flat cover of clear plastic was added to protect the LEDs and the hybrid circuit from dust

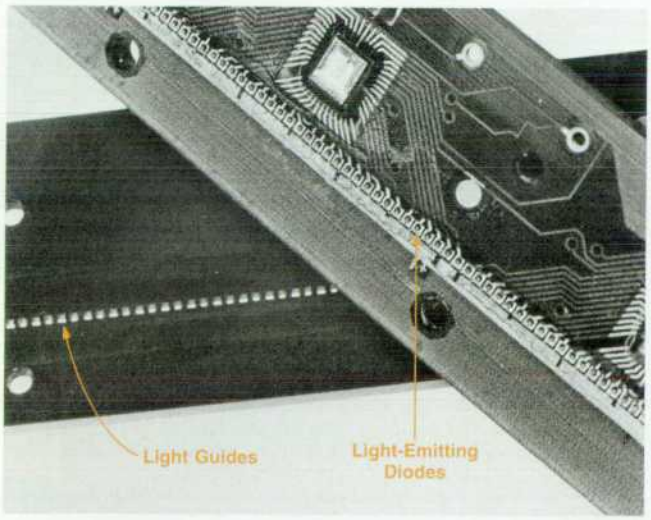


Fig. 8. *Erase bar assembly. The cover plate containing the light guides fits on top of the circuit board containing the LEDs.*

and other particulate matter present in the electrophotographic copier environment. To a first approximation, the flat cover does not contribute to the shape of the light beam. The main impact on the optical characteristics is to reduce the intensity by Fresnel reflection losses at two surfaces (or by about 8%).

The optical performance of the aperture LED array is shown in Fig. 7. The intensity profiles were measured at the image plane with one LED on, one off, and two on. The characteristics of adjacent pixels overlap smoothly without gaps or overexposure. The transition edges are abrupt. The background level is low, showing an absence of cross talk, scattering, noise, and other unwanted signals.

Fig. 8 shows the erase bar assembly with its small size and compact design.

Conclusions

The partially reflective light guide design makes a major contribution to the LED erase bar. It is the optical baffle with its efficient performance, simple design, and low manufacturing costs that allows the LED erase bar to be a viable alternative for the erase function in electrophotographic copiers.

The light guides are specifically tailored for use with an LED array in this application, simply to project well-defined, narrow beams of light. The resulting performance is superior to self-focusing lens arrays, which compromise other performance attributes in favor of projecting a high-resolution image. Compared to these, the light guides provide a larger range of working distances (depth of field), more controlled and reproducible light-to-dark transitions (image sharpness and contrast), and higher optical efficiency (numerical aperture).

The combination of the optics, high-sterance* LEDs, and the hybrid circuit gives a high-performance, low-cost, solid-state erase bar that is superior in performance to other technologies providing the same function, such as incandescent bulbs, vacuum fluorescent bars, and gas plasma lamps.

*Sterance = flux per unit solid angle per unit area of the source.

In newer-generation copiers, where the erase function is more sophisticated, requiring features such as edit, zoom and selective color, the intelligent LED erase bar with its individually addressable LEDs offers significant advantages:

- Visible wavelength compatible with copier photoreceptors
- Low power consumption
- Predetermined pixel shape
- Simplified system timing and clocks
- Solid-state construction and reliability
- Low cost for equivalent function.

Acknowledgments

The author wishes to thank the following persons for their invaluable support: Ken Brownlee, Don Lapray, Dave Collins, Ed Aoki, and other members of the HP Optoelectronics Division (OED) erase bar team for their contributions in implementing the optical design in HP's first LED erase bar and for providing background information for this article, Al Petrucello of the HP OED optics lab for making bootlegged optical measurements, Bill Franks and the HP OED machine shop for fabrication of prototypes and practical design recommendations, Hugh Williams of the HP plating shop for processing prototypes and providing information on anodized aluminum, and Stan Gage of HP OED product development for having faith in the idea before it was proven.

References

1. C.F. Jones, *Array of Limited Internally Reflective Light Guides*, U.S. Patent no. 4,759,603, July 26, 1988.
2. J.R. Meyer-Arendt, *Introduction to Classical and Modern Optics*, Prentice-Hall, 1972.

HEWLETT-PACKARD JOURNAL

October 1989 Volume 40 • Number 5

Technical Information from the Laboratories of Hewlett-Packard Company
Hewlett-Packard Company, 3200 Hillview Avenue
Palo Alto, California 94304 U.S.A.
Hewlett-Packard Marcom Operations Europe
P.O. Box 529
1180 AM Amstelveen, The Netherlands
Yokogawa-Hewlett-Packard Ltd., Suginami-Ku Tokyo 168 Japan
Hewlett-Packard (Canada) Ltd.
6877 Goreway Drive, Mississauga, Ontario L4V 1M8 Canada

5953-8577