JANUARY 1984

HEWLETT-PACKARD JOURNAL

HEWLETT-PACKARD JOURNAL
Technical Information from the Laboratories of Hewlett-Packard Company

JANUARY 1984 Volume 35 • Number 1

Contents:

In this Issue:

Here in our *Hewlett-Packard Journal* offices, we recently took delivery on a new phototypesetting system. Our old system, acquired more than ten years ago, was a noisy nightmare of complicated electromechanical design. It had whirling drums, snapping relays, rotating mirrors, traveling lenses, and assorted shutters and rollers. Usually it worked, and it served us very well for many years. Our new system has only four moving assemblies, reflecting the universal trend towards doing more things electronically and fewer things mechanically, thereby gaining reliability. Significantly, two of our new typesetter's four moving mechanisms are disc drives—"floppies", like the ones you can see in nearly every personal computer system. From microfloppies to large multimegabyte drives, disc drives are an exception to the trend to more-electronic systems. Spinning platters coated with magnetic oxide are still the preferred media for storing massive amounts of data so a computer can have rapid access to any piece of it. Bubble memory, the only serious contender for the disc drive's role, has so far made inroads only where severe environments would knock out a disc drive's finely tuned mechanisms. A cost-saving feature of some disc drive designs is media removability, which lets users access more data with each drive by changing discs. It's taken for granted in a floppy drive, but becomes less cost-effective as drive capacity increases.

Our cover subject this month is the industry's largest removable disc drive media module. Its seven discs hold 404 megabytes, or over four hundred million characters, roughly the amount of information in all of the books in the photograph. It works in the HP 7935 Disc Drive, which is also distinguished by an unusually comprehensive set of built-in diagnostics. On pages 3 through 26, you can read about the design of this high-capacity drive and its companion fixed-media drive, the HP 7933.

Two speech output modules for HP computers are described in the articles on pages 29 and 34, one for HP Series 80 Personal Computers, and one for HP 1000 and HP 3000 Computers. The latter is compatible with the industrywide RS-232-C/V.24 interconnection protocol, so it can also be used with other manufacturers' computers if you can get along without the special software that makes the modules easier to use with HP computers. If you can't always be looking at the computer, it does help if the computer can talk.

-R.P. Dolan

Two High-Capacity Disc Drives

One of these 404-megabyte drives is the current industry leader in removable disc pack capacity. The other is a lower-cost nonremovable drive.

by Kent Wilken

THE DESIGN of a high-capacity disc drive requires expert attention in a variety of areas. Magnetic media and magnetic read/write heads of exacting specifications must be obtained. Mechanics that can precisely and repeatably locate data on the magnetic surface must be employed. Low-noise electronics and sophisticated modulation techniques must be used to achieve high bit packing densities and retrieve the data reliably. Real-time signal processing techniques must be used to ensure that the disc heads are constantly in the correct position over the data. Error correcting codes must be designed to detect errors in reading the data and to correct certain types of those errors. And a microprocessor running what amounts to a real-time operating system must be employed to orchestrate the entire operation.

The HP 7933 and HP 7935 Disc Drives (Fig. 1) provide technical advancements over previous HP disc products in these as well as many other important areas. This pair of 404-megabyte disc drives includes the industry's highest-capacity removable-pack disc drive, the 7935. Its sister product, the 7933, is a less expensive nonremovable design.

Automatic Head Alignment

The development of such large-capacity disc products including the requirement of removability presented a number of challenges. Head positioning accuracy needed special attention. A data pack that is written on one drive must be readable on another drive. The slight differences in relative head position from drive to drive can make this difficult. Similarily, a pack written at one temperature ex-

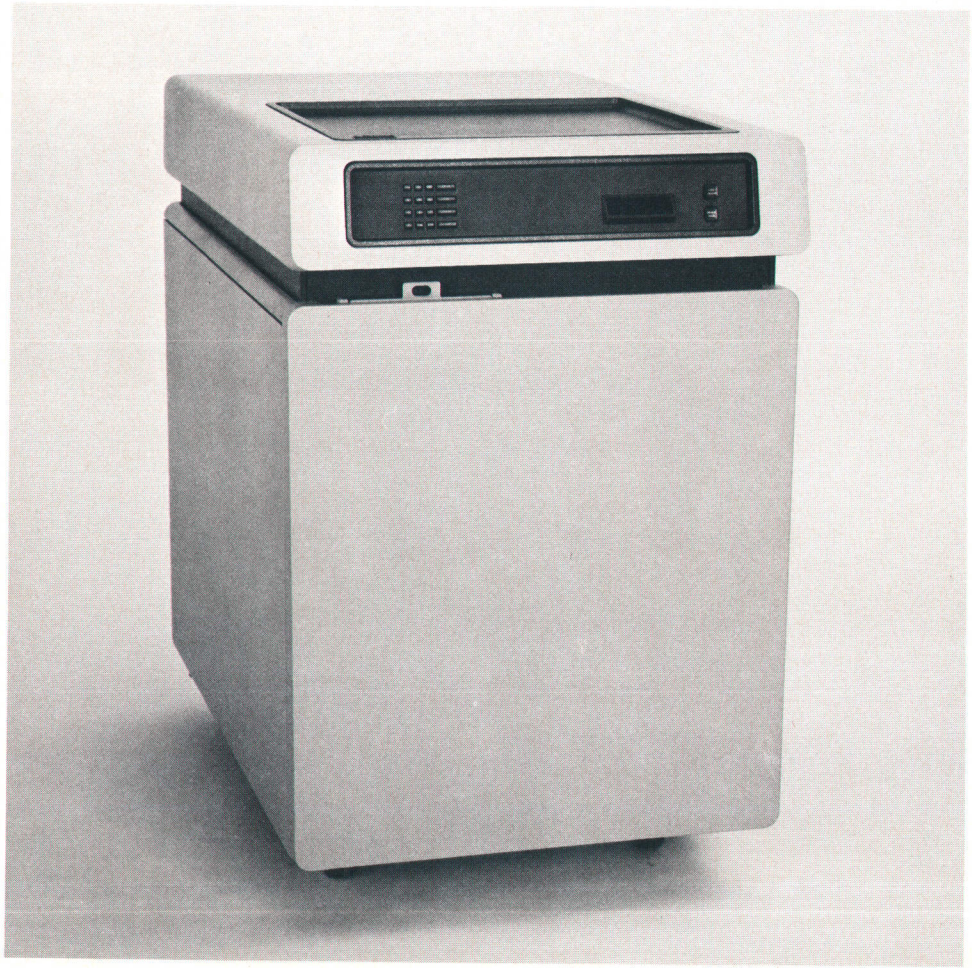


Fig. 1. *The HP 7933 and 7935 Disc Drives are 404-megabyte drives that feature data error correction and extensive diagnostics. Externally, the drives are identical. The difference is that the 7935's disc pack can be removed and replaced by the user.*

treme must be readable at the other extreme. Because of thermal expansion, the head position relative to the data will vary. To overcome these and related position variations, a technique dubbed automatic head alignment (AHA) is used. Each data surface in the disc pack has a set of calibration bands. When a pack is first inserted in the drive, and periodically thereafter based on time and temperature changes, the built-in microprocessor makes a calibration reading on each of the bands on each of the surfaces. This reading tells the processor how far that band is from the positioning bands on the reference (servo) surface. The difference is stored in a microprocessor RAM table. Later, when a head positioning (seek) operation occurs, an interpolation is done based on the tabular data to determine the offset required for the target location.

The main benefit of AHA is higher capacity as a result of being able to locate the target data more precisely. There is also a side benefit. Previous removable disc products required periodic fine tuning of the head position by qualified service personnel. AHA virtually eliminates this manual adjustment, thus reducing maintenance costs and system down time.

Sector Sparing

Another area that required a lot of design attention was ensuring defect-free media in the area where user data is stored. It is a virtual impossibility to create a large magnetic surface for high-density recording that is defect-free. Higher-capacity disc products such as the 7933/35 tend to see more defects because the area that they use to store a bit of data is smaller, which means that smaller defects become visible. Previous products use a method called track sparing. When a defect is encountered, that track of data (one concentric ring) is flagged as being defective and that logical track is assigned to a new physical location taken from a spare track pool.

Anticipating an increase in the number of observed defects, a new method called sector sparing is employed in the 7933/35. Each track is allotted one extra data block (sector). If a sector should contain a defect, the extra sector is migrated to the defective area by moving the intervening sectors one position. The drive microprocessor handles

Fig. 2. *The HP 97935A Data Pack for the 7935 Disc Drive provides contamination resistance similar to a sealed chamber. When the pack is inserted in the filtered air of the drive, a trap door opens to allow the heads access to the media.*

this data movement as well as the subsequent skipping of the defects in a manner that is transparent to the host computer.

Sector sparing greatly increases the number of media defects that can be compensated while still presenting 404 megabytes of defect-free data space to the user without increasing the fraction of disc space that is reserved for sparing and unavailable to the user.

Controlling Contamination

Controlling particle contamination is another challenge in designing a high-capacity removable disc product. Just as there is a greater sensitivity to media defects, contamination becomes a greater concern. The HP 97935 Data Pack (Fig. 2) incorporates a new design concept that has contamination resistance similar to a sealed chamber while providing the necessary removability. When not in use, the disc platters are completely enclosed in an outer set of plastic covers that prevent dust particles, fingerprints, and other pollutants from contaminating the media. When the pack is inserted in the filtered air of the drive chamber, a trap door opens that allows the heads to access the media. Previous designs exposed disc platters directly to dirty air and fingerprints during insertion and removal; to aid in avoid-

A Command Language for Improved Disc Protocol

by Douglas L. Voigt

By itself, a disc drive is of little use. It must be connected to a system. The system designer will probably want to support a variety of disc products, but would like to minimize the effort in designing the supporting software. Command Set 80 (CS-80) was created as an adjunct to the 7933/35 Disc Drive project to provide a flexible and forward-looking method of communicating between the disc and a host computer.

One CS-80 feature that eases the software support burden is the describe command. When sent by the host, the describe command will cause the drive to send back a list of parameters detailing its characteristics. This includes the amount of data the device can store, how fast it transfers data, how much RAM buffering it has for I/O—all of the information needed to connect a drive to a system. This makes it possible for the host computer to configure a disc drive into its system without prior knowledge of the type of drive or its characteristics.

CS-80 also minimizes communication overhead between the host computer and the disc. Previous communication schemes had the computer tell the disc drive to seek to a specific address. Upon arrival, the computer would, in a separate command stream, declare whether the operation was to be a read or a write. CS-80 has a locate-and-read (write) command that sends all of this information in the initial command stream.

CS-80 is supported on all of the newer HP disc drives ranging from the 16-megabyte HP 7908 to the 404-megabyte HP 7933/35. It will also be supported on future drives. This consistent support of CS-80 makes it easier to provide a broad range of mass memory support on HP's computer systems.

Transactions in CS-80

CS-80 states how to conduct transactions, the basic unit of work in this protocol, and it tells how to specify the operations to be performed.

A transaction consists of three phases, as illustrated in Fig. 1. In the command phase the CPU sends a command message describing an operation to be performed by the disc. The command is decoded, and the drive enters the execution phase, during which the specified activity takes place. If execution of the command involves the exchange of data, the disc controller uses the HP-IB parallel poll function to synchronize an execution message containing the data with internal events; this minimizes the time during which the bus is dedicated to one transaction.

When execution is complete, the reporting phase is entered, and parallel poll is again used to initiate a reporting message that signals the end of the transaction. The reporting message contains an indication of success or failure of the transaction. This can be interpreted by the CPU, allowing the next transaction to begin as soon as possible.

In error situations the disc aborts directly to the reporting phase and waits for the CPU to reach that point. Between transactions, the disc controller monitors the vital signs of the disc and performs limited diagnostics. If anything happens that requires CPU intervention during this time, an unsolicited report is entered to get the CPU's attention. This reflects a trend towards equality of communication between CPUs and intelligent peripherals, which can initiate transactions as a result of events detected outside the CPU.

The formalization of a transaction allows a CPU to support the wide range of features of the 7933/35 Disc Drives using only three transaction templates. The transaction is also used as a level of modularity in 7933/35 firmware. The firmware impact involved in converting the 7933/35 to a new CPU interface would

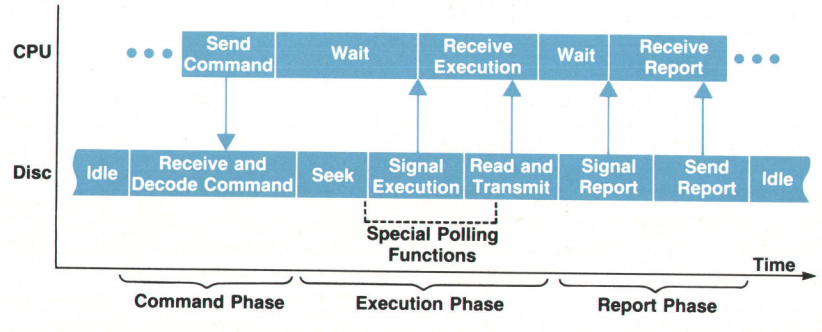


Fig. 1. *Command Set 80 organizes CPU and disc activities into three-phase transactions. All disc functions are executed in a command phase, an execution phase, and a report phase. Shown here is a locate-and-read transaction.*

be limited to the changes necessary to implement transactions on the new interface. Control of the 7933/35 by a CPU can be viewed as transaction-oriented communication between a process running inside the CPU and a process (called the logical machine) running in the disc controller to control the disc hardware.

Command Efficiency

CS-80 commands are designed for flexibility and communication efficiency. The disc controller maintains a logical machine state, which contains the variables used in making the decisions involved in each disc operation. These variables are set by the CPU through the use of complementary opcodes. A command message contains a series of optional complementary opcodes, which may be followed by one opcode specifying an operation to be performed. Complementary opcodes used in a command that contains an opcode specifying an operation are temporary, and apply only to the current transaction. Complementary opcodes specified in a command that does not call for an operation retain their new values permanently. The permanent value of each complementary opcode is used during any operation that does not temporarily override it.

This structure allows a CPU to generate commands containing only the new information necessary to do an operation. Parameters that do not change often need not be retransmitted with each command, since their most frequently used values can be retained in the disc controller. This scheme also offers a CPU several ways of expressing the new value of a single parameter. This capability is used to allow disc addresses (locations of data to be read or written) to be specified with cylinder, head, and sector numbers corresponding to the disc's physical organization, or as a single block number forming a linear contiguous address space including all of the sectors on a disc, which is easier for a CPU to deal with. The 7933/35 firmware is optimized to decode a few frequently used opcodes an order of magnitude faster than the rest, thus retaining command set power while offering speed where it is needed.

Status Information

CS-80 also specifies a new format for status information returned to the CPU. Error and status conditions are divided into four classes: protocol reject errors, hardware fault errors, procedural access errors, and nonfatal information errors. A six-byte parameter field is provided to be used in conjunction with specific status bits, or for current disc address information. Any combination of status bits outside of the fault field can be suppressed as they occur using the status mask complementary command. The reporting message mentioned above gives a quick indication as to whether or not there is any status information available upon completion of each transaction.

CS-80 allows the 7933/35 disc controller to take advantage of microprocessor technology to distribute the intelligence and increase the amount of parallel activity available in HP systems today. In the future we hope to use it to help provide HP customers with a consistent growth path involving a wide range of system building blocks.

ing contamination, a dummy platter was put on the top and bottom. In the 7935, these cover discs are not required, meaning that for a given number of platters, two more are available for data, thus increasing capacity.

Troubleshooting and Repair

Diagnostics and serviceability are fundamental parts of the design of the 7933/35. The most visible part of the diagnostic design is the diagnostic front panel. This consists of a set of keys and an alphanumeric display. With these, a trained service person can access data and test routines internal to the drive.

The internal diagnostics and utilities are a major improvement. Previously, most disc diagnostics were written to run on the host computer. This has several disadvantages. First, a product such as the 7933/35 is generally supported on more than one of HP's computer systems. Having to create a system-specific version of a diagnostic is a duplication of design effort. The internal diagnostics of the 7933/35 needed to be implemented only once. Second, previous diagnostics have been written by system software designers not necessarily extremely knowledgeable about the details of disc drive operation. The diagnostics on the 7933/35 were written by members of the design team in close cooperation with the hardware designers.

Possibly the biggest advantage of the internal diagnostics on the 7933/35 is the array of test hooks designed into the hardware. External diagnostics must treat the disc drive as a black box. The internal diagnostic has more direct access to the hardware. It can use such facilities as an analog-to-digital converter to test power supply levels and other critical voltage settings. The internal diagnostic can disable certain sections of electronics while testing a neighboring board to help isolate the problem to a particular board. Extensive fault isolation in the diagnostics virtually eliminates guesswork in repairing consistent failures in the drive.

The most difficult problems to diagnose in a disc drive, however, are not the consistent failures. The intermittent failures are more difficult. The diagnostic design team recognized this and felt the best way to assist the service personnel was by keeping a detailed record of each failure the drive encountered in the course of operation. To that end, certain tracks inside the disc are reserved for storing this fault history. This is in addition to any fault logging a host system might have for drive errors. These maintenance tracks are outside the normal disc address space and are accessible only by diagnostic commands from the I/O channel or through the diagnostic front panel. Besides preserving the fault history, these tracks contain a history of data error occurrence, test patterns for verifying proper operation of the read data path, an area for doing write-then-read testing, a table of the spared areas of the disc, and other information.

In addition, the drive contains internal utilities that do error rate testing, test the performance of the positioning system, and perform other tests. The diagnostics, maintenance information, and internal utilities can be accessed through the diagnostic front panel or through the HP-IB (IEEE-488) I/O port using a special set of commands. Through use of the diagnostic front panel a service person can, for example, disconnect and service the failed drive while the rest of the computer system remains in operation. By passing diagnostic results over the I/O channel to the host computer, it is possible to display the results on the system console or even on a remote terminal.

Second-Generation Disc Read/Write Electronics

by Robert M. Batey and James D. Becker

THE FIRST HP DISC PRODUCTS brought high performance to Hewlett-Packard computational systems. However, by 1978, the capacities of these drives were becoming inadequate because of their use of traditional methods of channel encoding and decoding, and their low recorded track densities (on the order of 300 to 400 tracks per inch).

By increasing track densities and by clever use of information-packing coding techniques, the new 7933 and 7935 disc products deliver slightly over three times the capacity of previous products (7925 family) at about the same cost. The effect is a reduced cost per megabyte.

Coding Schemes

Coding schemes used by earlier HP disc products include the MFM (modified frequency modulation) code and the FM or Manchester code. Their simplicity makes system design easy. It is significant that these codes use little of the channel's available bandwidth. This proves to be a disadvantage in terms of information content, but an advantage in terms of simplicity of signal processing.

To be more specific, consider MFM. The code construction map is as follows:

Data Bit	Code Sequence
0	X 0
1	0 1

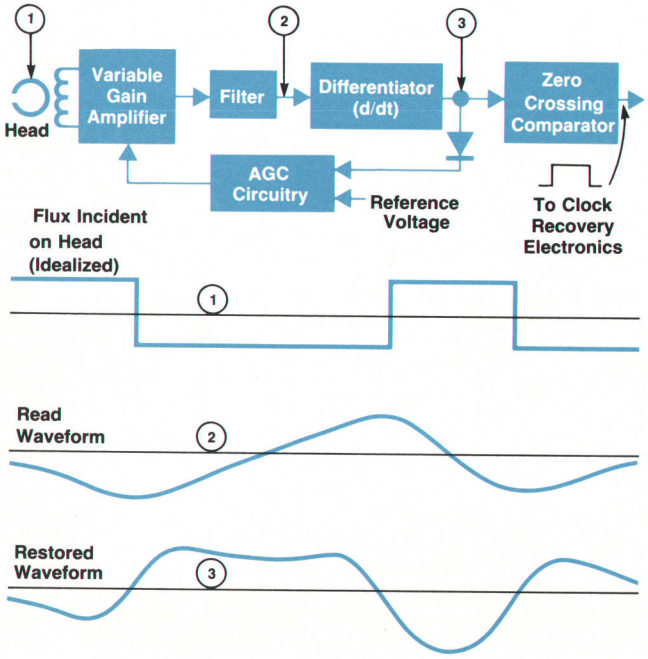


Fig. 1. *Block diagram and waveforms for simple read chain electronics for the MFM code.*

The coding rule requires that a 1 stand for a flux transition and that a 0 represent no flux transition. The X stands for a possible flux transition depending on the last state of the previous code sequence transmitted. The X is simply replaced by the complement of the previous state, that is, if the previous state was a 1 (flux transition), replace X with 0 (no flux transition).

It is trivial to build a state machine to implement an encoder and decoder for this code, and processing the raw analog data from a read head is also very simple. We will examine why.

The voltage output of an inductive head is proportional to the derivative of the flux, that is, it is proportional to the rate of change of the flux incident on the head transducer. Consider a waveform in which the flux reverses arbitrarily with time. The flux waveform steps between $+p_h$ and $-p_h$. Head output, the derivative of this waveform, consists of alternating impulses of value $+Kp_h$ and $-Kp_h$. Since the read system is band-limited we can add a low-pass filter to the model. The output of the model will now be a series of alternating pulses of finite width and value. In reality, the low-pass function is determined by the rise time of the transitions (transition width) and the head response. These factors determine the Nyquist signaling rate for the system, that is, the maximum rate at which independent symbols can be sent over a channel without intersymbol interference.[1]

Earlier coding techniques took advantage of the signaling limit to achieve an easy implementation of a pulse position detector. Consider a simple signal processing system consisting of a read head, a differentiator, and a comparator to detect zero crossings. If we control the spacing of the pulses we introduce enough intersymbol interference so that the resulting waveform's only regions of zero slope are at the peaks. A coding scheme that controls the maximum distance between pulses can keep the resultant inflection point between adjacent pulses from reaching zero slope. The minimum distance is also controlled to keep the pulse position from being shifted out of a timing window. The result is that if the waveform is differentiated, the zero crossings will correspond to pulse peaks, and the comparator output will be a reasonable reconstruction of the flux waveform.

Conditions necessary for this simple analog processing technique to be feasible seem to be:

1. An "appropriately" band-limited channel
2. Tight constraints on the maximum and minimum signaling rates (use of a proper coding technique)
3. About 25 dB signal-to-noise ratio for a reasonable error rate.

MFM and FM produce recording densities of 1.0 and 0.5

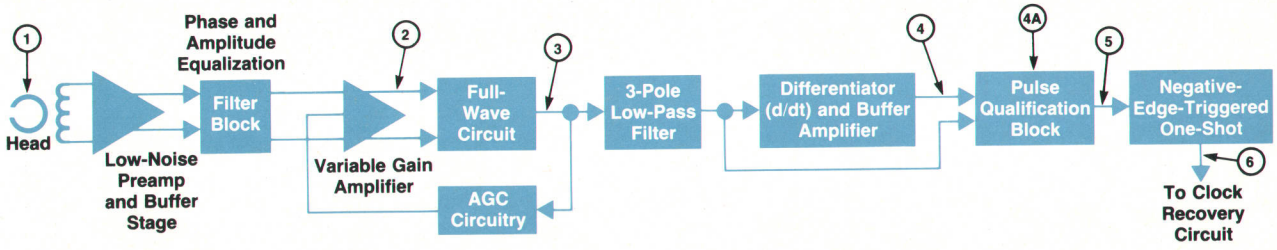


Fig. 2. *Block diagram of 7933/35 read chain electronics for the VLFM code. The numbers in circles refer to waveforms in Fig. 3.*

respectively, with MFM being the better of the two codes at about 1.0 density (i.e., about 1 data bit per recorded transition).[2] It is evident that the data rate is only equal to the rate of recorded transitions on the disc at best. Unless we can improve the data-to-transition ratio, the only way to increase capacity per track is to increase the number of flux reversals (transitions) per track, and this means great improvements in the heads or media must occur.

The goal of the 7933/7935 project was to gain 33% in recorded density by means other than improved heads and media. The approach was to deal with the channel modulation code, not only as a vehicle to simplify record recovery, but as a means of data packing. For every three transitions written on the disc, on the average we wanted to get four data bits of information.

In surveying the literature, we found a code that gave us reasonable results.[3] We chose to call the code VLFM in light of its predecessors. Horiguchi and Morita described the code with the following coding table. The ones and zeros on the left represent bit positions in a data string, and the ones and zeros on the right represent potential positions for recorded transitions on the disc (with the ones being the actual positions of recorded transitions).

VLFM ENCODING TABLE

Data Sequence	Code Sequence
00	X01
01	010
10	X00
1100	010001
1101	X00000
1110	X00001
1111	010000

Notice that some of the code strings are preceded by an X. This shorthand means that to form this code string the first bit must be assigned the complement of the trailing bit of the previously transmitted code string. This is done to restrict the minimum distance between transitions. For VLFM and also for MFM, the minimum distance between coded transitions is a distance of one transition cell or window. We will call this distance the d constraint for transition placement.

Similarly, there is for many codes a maximum recorded distance between transitions. For MFM it is three windows. This constraint we will call the k constraint. For VLFM we can observe a maximum of seven windows for the k constraint. (Observe this by encoding hexadecimal 4D4D4D. 1000000010100000001 will appear in the binary code sequence.)

Since the d and k constraints determine the minimum and maximum number of data windows between recorded transitions, they determine the upper and lower require-

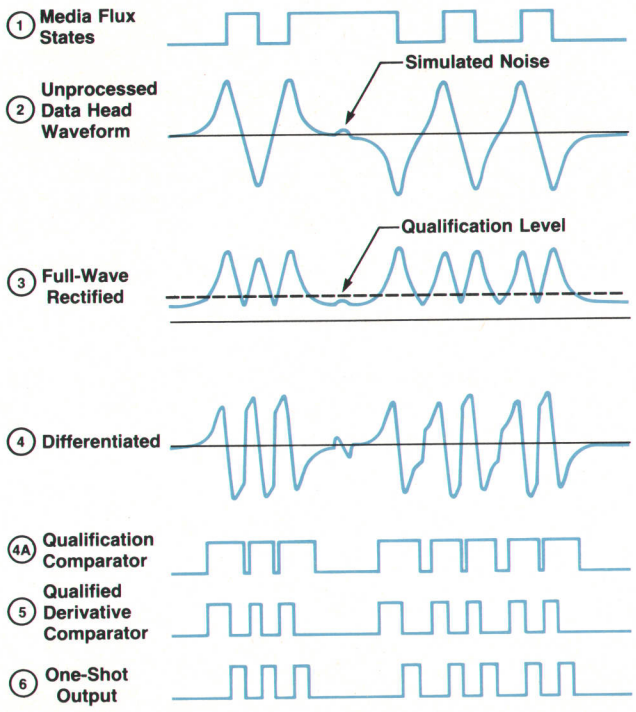


Fig. 3. *Waveforms in the circuit of Fig. 2.*

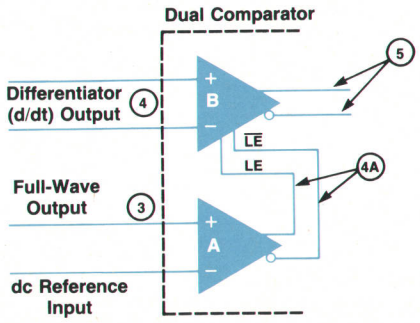


Fig. 4. *Pulse qualifying circuit gates out noise pulses as shown in Fig. 3.*

ments for the data channel bandwidth (when the clocking rate is specified). Both MFM and VLFM have d constraints of one, and thus they have identical upper-frequency requirements. Because VLFM has a k of seven compared to the k of 3 for MFM, VLFM coding allows transitions to be more than twice as far apart in time. Therefore VLFM results in a much larger signal energy at lower frequencies. The result is a VLFM requirement of approximately twice the bandwidth required for MFM. Note that this bandwidth is achieved by decreasing the lower band limit rather than by increasing the upper band limit, which is the more commonly encountered technique.

Analog Signal Processing

Processing codes that require more channel bandwidth than MFM make signal processing more difficult.

Early schemes require signal amplification followed by a differentiator (d/dt) circuit. The output of the differentiator circuit is then input to a comparator whose reference input is at ground potential. The comparator output is a digital waveform whose edges coincide with the flux reversals applied by the rotating disc to the read head. Fig. 1 illustrates how the waveforms look at different stages, and the block diagram for this configuration. Note the simplicity of this scheme.

In the 7933/35 product the waveforms look a bit different. Fig. 2 is a block diagram of the system and Fig. 3 shows typical waveforms in the electronics. For VLFM encoding the pulses are spaced twice as far apart as for MFM, worst case. Since the head voltage waveform attempts to approximate the derivative of the flux, the system readback waveform settles to the zero baseline (almost) between some pulses when they are spaced far apart.

Differentiating this waveform would result in the derivative's settling to near zero between pulses. Add a little noise to the system and a false zero crossing is easily generated. This results in a data error, something not easily

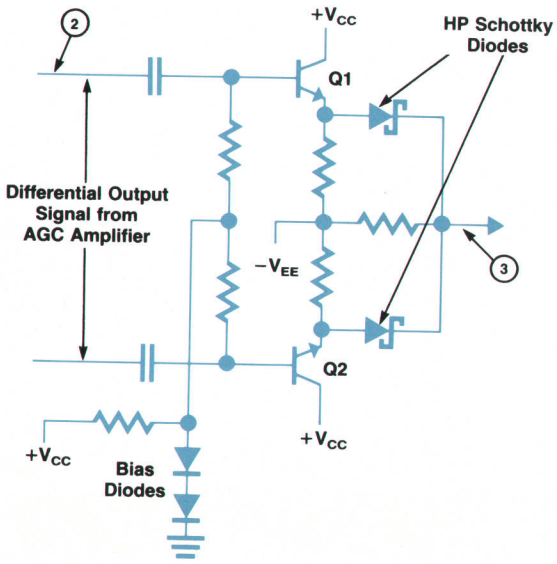


Fig. 5. *Schematic diagram of the full-wave circuit shown in Fig. 2. Q1 and Q2 are a matched pair of 500-MHz transistors.*

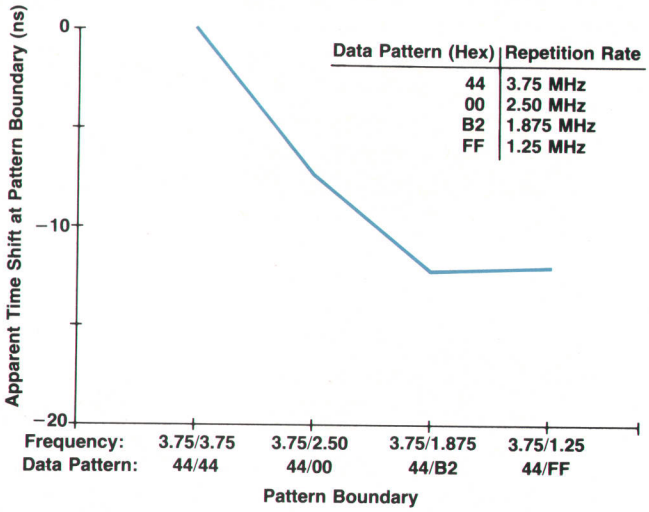


Fig. 6. *Pattern phase sensitivity was measured by writing all sectors with 32 bytes of data pattern 44 alternating with 32 bytes of another data pattern, then measuring the offset of the pattern boundary in the read waveform.*

tolerated in the disc business. We needed to design a system that could tolerate this condition without generating false zero crossings in these shouldering regions.

Fig. 2 illustrates the approach. The read signal is first full-wave rectified. An active circuit using two HP Schottky diodes and a differential pair of transistors is used for this purpose. The speed of the circuit is important; so is minimizing the dead zone at crossover.

After low-pass filtering, the full-wave-rectified signal is differentiated. This causes all negative-going transitions to track valid flux reversals and all positive-going transitions to track approximate midpoints between flux reversals. This scheme generates waveforms approaching our requirement, except that the midpoint between pulses may have a slope that approaches zero. A small noise pulse will almost certainly generate a false zero crossing in the differentiated signal.

It was necessary to provide some noise immunity in these regions between pulses. We use a level comparator set about 15% above the lowest level of the full-wave signal. Any signal above the qualification threshold gets gated through, and any signal below the threshold is ignored. We will explain this gating circuit in detail, shortly.

Triggering a negative-edge-triggered one-shot from the resultant output of the level qualifier results in a pulse train whose leading edges are timed synchronously with the read pulses coming from the head. Small noise pulses between data pulses are ignored. Error rates less than one part in 10^{11} have been obtained using this scheme.

Level Qualifier

The circuit that gates out noise pulses during waveform shouldering periods bears a bit more explanation. Basically, the circuit is made up of a dual comparator circuit. Each comparator within the circuit has a track-and-hold line. In one state the output of the comparator follows the polarity of the inputs as any comparator would, but in the

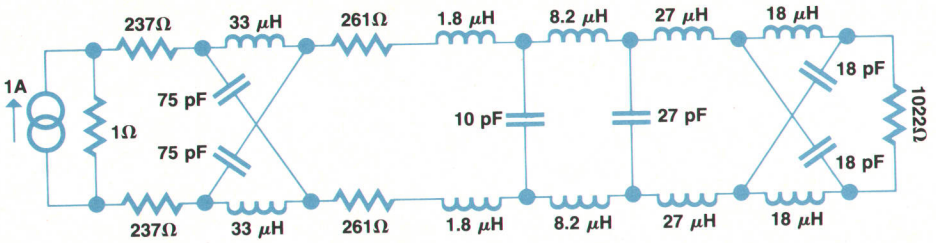


Fig. 7. *This filter compensates for the pattern phase sensitivity shown in Fig. 6 and improves the drive error rate by over two orders of magnitude.*

second state the comparator holds the last output state. This feature is used to implement a pulse qualifying circuit. It works in the following way (see Fig. 4).

The full-wave-rectified waveform is input to the positive input pin of comparator A. When the signal is greater than the reference, comparator B is allowed to follow the differentiator output polarity. When the full-wave signal is less than the reference, comparator B is latched in its previous state, effectively gating out any noise pulses that may occur during the shouldering period.

The full-wave circuit is shown in Fig. 5. The two input transistors form a buffer stage so that clamping at the input capacitors is minimized. Bias diodes are included to compensate the circuit thermally, thus minimizing dc drift. Since the circuit is basically a differential switch, the Schottky diodes are switched without the dc crossover problems that normally plague such circuits. Some signal compression occurs around the zero baseline, but this is generally tolerable because pulses are not qualified by level when the signal level is in this region.

Clock Recovery Techniques

This section deals with synchronizing and decoding the serial pulse streams delivered by the analog processor in the read path. The task is more complex than it might seem since we would like to use a synchronous state machine to decode the encoded bit streams. The task requires a clock. The reliability of the clock recovery electronics must be good enough to deliver an error rate less than one part in 10^{11}, like the rest of the electronics.

It might seem that recovery of the timing information depends solely on the quality of the phase-locked loop circuitry. Although this is almost true, we found that even after we had constructed an adequate phase-locked loop, our error rate was not as good as we needed. A little further investigation revealed that the timing information coming off the disc was dependent upon the data pattern. We proved that this effect was independent of the phase-locked loop by the following test.

First, we used a data generator to generate the patterns that were troublesome. When we locked the phase-locked loop to these patterns the loop had no trouble recovering them. This indicated that the trouble was either in the recorded patterns on the disc or in the analog processing electronics.

The phase nonlinearities in the system were found to be caused by:

1. The electrical resonance of the data heads
2. Nonlinearities in the write process
3. The phase/frequency characteristic of the read amplifier chain.

It was felt that even though we could minimize these effects to some extent, we would still need to compensate for some of the phase nonlinearities. Since effects 1 and 3 were controlled by component specifications, and since these two effects were the major contributors to phase nonlinearities, we felt that a filter would provide enough correction.

Our biggest difficulty was to measure the delay characteristic of the channel. Because the channel includes the write process, the magnetic media, and the head, we could not simply use a network analyzer to measure the transfer function. Our approach was to vary the pattern repetition rate and measure the step offset in the phase detector output of the phase-locked loop.

Fig. 6 shows the results of the test. The high-repetition-rate pattern (44 hexadecimal at 3.75 MHz) was chosen as a reference. All sectors were written with 32 bytes of the

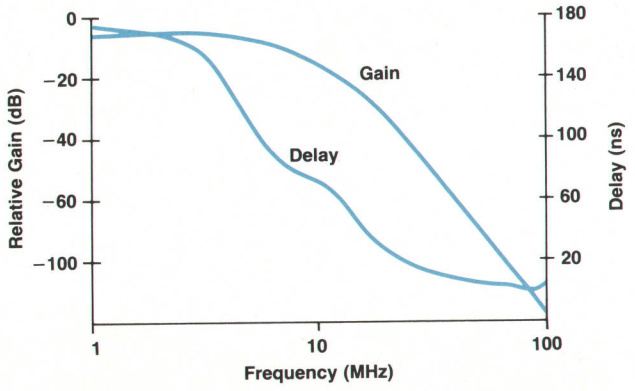


Fig. 8. *Calculated frequency response of the filter of Fig. 7.*

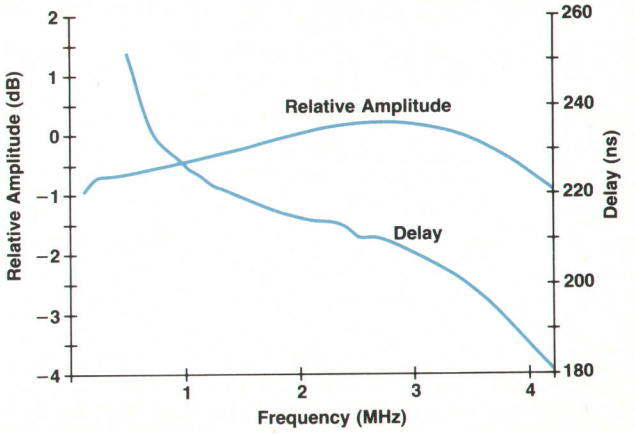


Fig. 9. *Overall frequency response with the filter of Fig. 7 in the read path.*

reference pattern alternated with 32 bytes of the pattern under test. All patterns chosen had 50% duty cycle. The phase step was measured for each case and plotted as in Fig. 6.

This led us to design a filter with a group delay characteristic compensating the pattern phase sensitivity shown in Fig. 6. The filter network is shown in Fig. 7 and has the calculated response shown in Fig. 8. When measured as part of the read path electronics, the characteristics are as shown in Fig. 9. Note the deviation from the filter's own characteristics below 1 MHz. This is largely caused by the ac coupling capacitors between amplifier stages. The filter improves the error rate by over two orders of magnitude. Fig. 10 shows a distribution of soft error rate for the product in the early days of production. 181 units were sampled (2,353 head and media combinations), showing that the average error rate was about 8×10^{-11}. With error correction we would expect over an order of magnitude improvement, or an average error rate for the user less than 8×10^{-12}.

The clock recovery phase-locked loop is shown in Fig. 11a. The one-shot pulse width is set to single-window width. This means that a ..101010.. code pattern (44 hexadecimal data pattern) will generate a square wave out of the one-shot. One window is therefore the minimum distance between pulses.

Below the block diagram, a state machine is shown for the phase detector (Fig. 11b). The waveforms to the right are for a locked condition. To put things in perspective, the pulse width out of the one-shot is 66.7 nanoseconds. Signals P1 and P2 are the sampled outputs of the phase detector. If the one-shot data is either early or late, the P1 and P2 pulse widths are altered. For instance, if the one-shot data is late, the P1 pulse width is longer and P2 is shorter. The opposite is true if the one-shot data is early. Notice also that one-shot data cannot be later than sync data because of the J-K state machine. This phase detector will not operate with more than a ±90-degree phase ambiguity. Only a very small unstable region exists, because of logic propagation delays. If no pulse occurs during any cycle of the divide-by-two VCO clock, no sample output

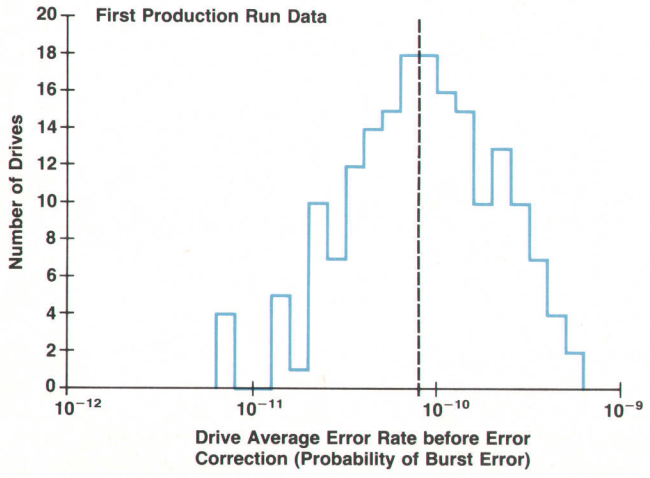


Fig. 10. *Distribution of soft error rate for early production 7933/35 Disc Drives without error correction. Error correction improves the error rate by another order of magnitude.*

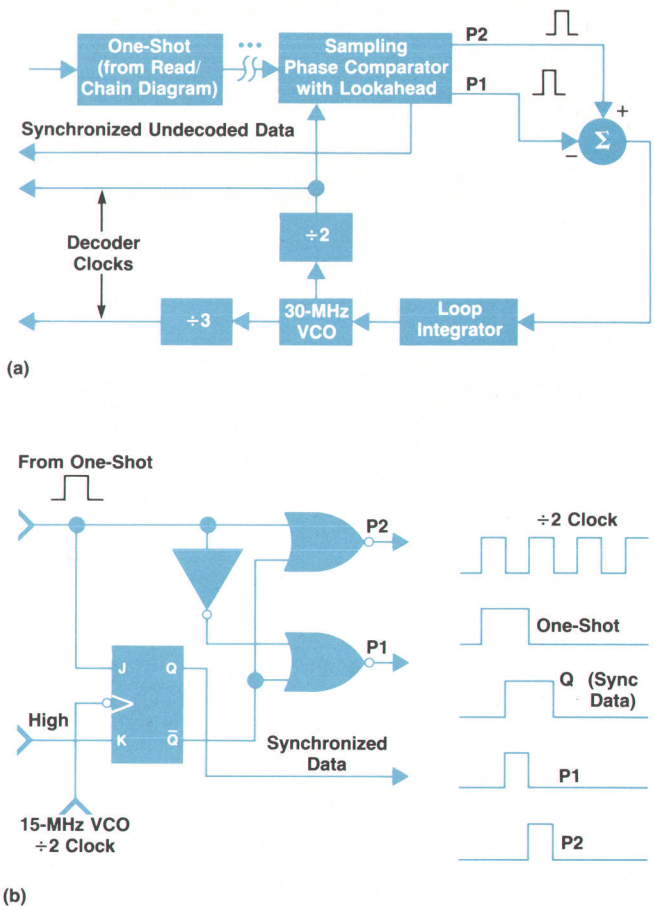


Fig. 11. *(a) Clock recovery phase-locked loop. (b) Phase comparator state machine and waveforms for a locked condition.*

is generated. This phase comparator will not sample during an absence of data pulses. It operates much like a sample-and-hold, with the loop integrator performing the hold operation.

This phase-locked loop can remain locked to a string of pseudorandom pulses, evenly clocked. The phase detector sample rate is determined by the data pattern since no phase comparator output is generated when there are no pulses. The maximum loop bandwidth is realized when a ...1010101.. field is presented to the phase detector input. For this reason the beginning of each sector is written with about 16 microseconds of this pattern to assure proper phase coherence of the system to the sector's data field. The loop bandwidth is set so that 90% settling occurs in about eight microseconds.

Acknowledgments

The authors would like to acknowledge the following individuals for their contributions: Pete Petroski for his low-noise preamp design, Richard Wilson for media characterization, Mary Carter for read head characterization, and Russ Mendenhall and Wally Overton for guidance from their previous experiences in designing this kind of hardware. Also, we thank the many people in the production area, the materials area, and other divisional support

areas for their endless hours in making the 7933/35 data path a success.

References

1. A.B. Carlson, *Communication Systems*, McGraw-Hill, 1975, p. 160.
2. J.C. Mallinson and J.W. Miller, "Optimal Codes for Digital Magnetic Recording," *The Radio and Electronic Engineer*, Vol. 47, no. 4, April 1977, pp. 172-176.
3. T. Horiguchi and K. Morita, "An Optimization of Modulation Codes in Digital Recording," *IEEE Transactions on Magnetics*, Vol. MAG-12, no. 6, November 1976, pp. 740-742.

Disc Drive Error Detection and Correction Using VLSI

by Peter M. Galen

A DISC DRIVE is subject to errors when reading data from the disc. The errors can be induced by random noise, correlated noise, media defects, mechanical nonlinearities, and other causes. The purpose of error correction is to improve the integrity and recoverability of data.

There are two elements to error correction. The first is the detection of errors and the second is the correction of the erroneous data. Placing error correction entirely within the disc drive eliminates the need for delay in the case of most errors and thereby improves the system performance.

Error Correction

Errors in disc drives are best described as bursty. This means that errors tend to occur in groups. This fact led to the choice of a burst-correcting code, called a Fire code,[1] for the 7933/35. The characteristic polynomial of this code is

$$(X^{23}+1)(X^{12}+X^{10}+X^9+X^7+X^6+X^4+1)$$

This code allows the correction of a single 12-bit burst in a sector on the disc (2147 bits).

The first step in the error correction process occurs during a write of data to the disc. As the data is written a division modulo the above polynomial occurs. At the end of the data stream the remainder is appended to the data and written on the disc. 35 bits are added in this encoding process.

During a read the data stream with the appended error correction information is again divided modulo the chosen polynomial. If the remainder, or syndrome, is zero, the data is correct. If the syndrome is not zero, then an error has occurred. There are two error types: correctable and uncorrectable. Correctable errors are errors 12 bits in length or less, while uncorrectable errors are greater than 12 bits long. For a correctable error the syndrome contains information about the error pattern and the error location.

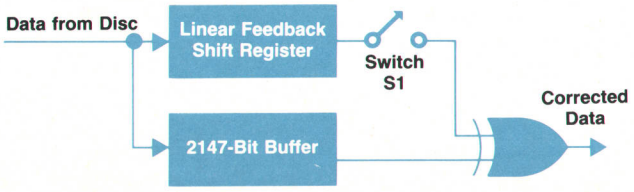


Fig. 1. *Circuit for error correction using a Fire code.*

Error Detection

Error detection is critical to ensuring the integrity of customer data. The Fire code chosen can detect most uncorrectable errors. There is a chance that an uncorrectable error will be mistakenly found to be correctable and correction attempted. To prevent this, a 16-bit CRC (cyclic redundancy check) is used. This is appended to the data on a write before encoding and is used for error detection after error correction has occurred. This decreases the probabil-

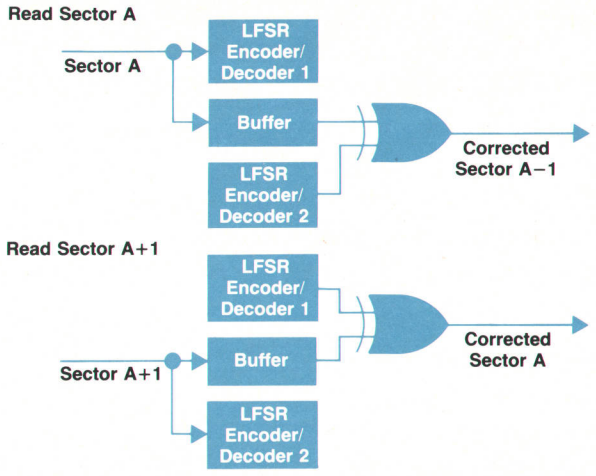


Fig. 2. *In the 7933/35 Disc Drives, two linear feedback shift registers are used so that the syndrome for one sector can be generated while the preceding sector is being corrected.*

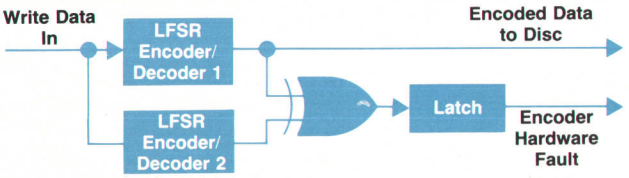


Fig. 3. *The same linear feedback shift registers are used for both encoding and decoding. In encoding data, having redundant encoders helps prevent the writing of bad data because of an encoder hardware fault.*

ity of undetected errors to less than one occurrence in 10^{15} bits transferred.

Real-Time Correction

Correction can be accomplished by several methods using a Fire code. The method chosen is outlined in Fig. 1. The syndrome is contained within a 35-bit linear feedback shift register (LFSR). The data stream is contained in a 2147-bit buffer. Each shift of the LFSR moves the beginning of the error one bit in the data stream. The error mask is found when only the last 12 bits of the LFSR are nonzero. This enables the data from the buffer to be clocked out on a one-for-one basis with clocking of the LFSR. When the error mask is found, switch S1 is closed and the mask is EXCLUSIVE-ORed with the data, thereby correcting it.

To allow generation of the syndrome for one sector while the previous sector is being corrected, two LFSRs are required. Fig. 2 shows how these registers are configured. This allows correction of errors in consecutive sectors without inducing additional delay.

Fault Detection on Writes

An added advantage of the use of two LFSRs is their use in redundant encoding. The LFSR used for encoding is nearly identical to the one used for syndrome generation and correction. This allows the same LFSR to be used for both operations. Fig. 3 shows the configuration used to detect hardware failures in the encoder. This helps prevent a customer from storing bad data without knowing it.

VLSI Implementation

The error correction function was integrated into a 28-pin IC fabricated using HP's SOS (silicon on sapphire) process. The design contains approximately 31,000 active devices. The IC is designed to be compatible with several disc drives and is currently in use in the 7911, 7912, 7914, 7933, and 7935 Drives.

SOS was chosen for two reasons; these were the need for a fairly dense process and the speed requirements. This chip can accept data at a rate greater than 11 MHz. SOS is a CMOS process, but the substrate used is sapphire instead of the more traditional silicon.

Designing Testability Into the IC

Testing a VLSI chip can be difficult. One design goal for the 7933/35 project was that a drive be able to test itself functionally and detect and locate failing assemblies. To achieve this goal, the error correction chip was designed so that any input to it can be driven from a microprocessor-accessible register. All output lines can be read from microprocessor-accessible registers on the chip. This concept is shown in Fig. 4. This method will not find speed-related problems, but can find most other failure types.

Acknowledgment

Credit for the CMOS design belongs to Bob Ko.

Reference

1. P. Fire, "A Class of Multiple-Error-Correcting Codes for Non-Independent Errors," Sylvania Reconnaissance Systems Laboratory Report RSL-E-2, 1959.

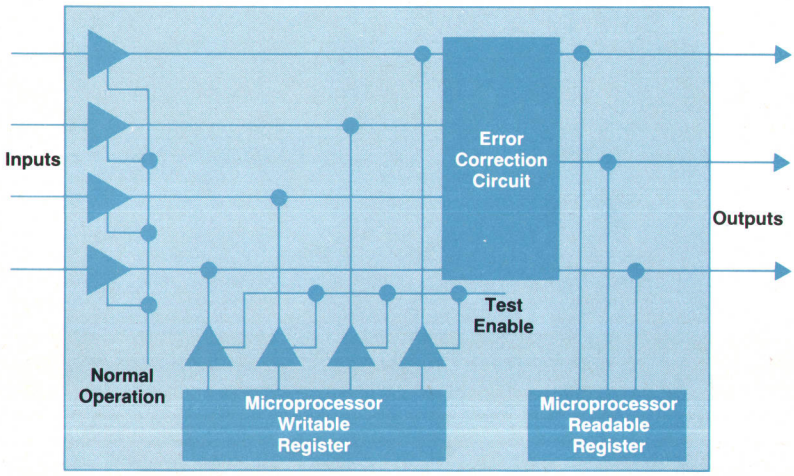


Fig. 4. *VLSI error correction chip is designed so that any input can be set by the drive microprocessor and any output can be read by the microprocessor for testing the error correction circuitry.*

Head Positioning in a Large Disc Drive

by R. Frank Bell, Eric W. Johnson, R. Keith Whitaker, and Roger V. Wilcox

THE HEAD POSITIONING SYSTEM in the HP 7933/35 Disc Drives has two major purposes. The system controls the position of the data heads to follow the data tracks within approximately one tenth of the track width. It also moves the heads from one cylinder to another in a minimum amount of time.

The high track density of the 7933/35 (625 tpi) requires a new approach to data head alignment techniques. To minimize the effects of thermal drifts and other low-frequency changes, the drive automatically aligns the data heads to prewritten alignment tracks existing on every data pack. The implementation of automatic head alignment in the drive requires special electronics and interaction with the drive microprocessor.

The large capacity of the 7933/35 was achieved in part by having 13 data surfaces on seven discs. Including the servo head, there are 14 heads. The resulting high mass requires high power levels to move the heads rapidly from track to track at performance levels similar to previous HP disc drives (7920/25). This requires a special design for the actuator amplifier, and a linear actuator capable of moving this large mass at high accelerations.

The head-carriage assembly in the 7933/35 is controlled by the microprocessor and dedicated electronics. The dedicated electronics provide the necessary speed and a relatively simple and reliable compensation technique for the positioning system.

Head Positioning Control Systems

Two separate control systems are used to control the position of the heads. The linear position control system, or fine position servo, keeps the servo head at the center of the servo track, and is used when reading and writing data. A nonlinear position control system is used when the heads are moved from one track to another (a seek). This nonlinear system is designed to move the heads relatively long distances (0.0016 inch for a single-track seek, and 2.1 inches for the longest seek) in the shortest possible time.

The fine position servo system is designed to keep the servo head as close as possible to the center of the servo track. In addition, the servo head must quickly settle on track at the end of a seek. For the 7933/35 the specific goals were to have a servo head tracking error of less than 75 microinches and a settling time of 2.5 milliseconds.

The largest part of the tracking error comes from the radial motion of the disc pack. The disc pack spins at 45 Hz (2694 r/min), and the design of the spindle keeps the once-around runout component of the disc motion within 300 microinches. This component needs to be reduced by at least a factor of 10 to keep the servo head tracking error within the desired limits. This implies that the open-loop servo gain should be approximately 20 dB at the once-around frequency of 45 Hz.

The presence of nonlinearities requires the system to have a higher gain to ensure that the runout is reduced. Some of the nonlinearities include deadband in the power amplifier and the static friction of the head assembly carriage bearings. An open-loop gain of 30 dB overcomes the effect of these nonlinearities and still reduces the 45-Hz component by at least a factor of 10.

To achieve fast settling times, the bandwidth of the fine position system needs to be as wide as possible and the transient response must be good. The bandwidth is limited by the presence of a mechanical resonance at 3400 Hz. The compensation technique that is used includes a lead/lag network and a low-pass filter (see Fig. 1). The final phase margin is 35 degrees and the open-loop 0-dB crossover frequency is 500 Hz. The low-pass filter controls the magnitude and phase of the resonance so that stability is not a problem.

A block diagram of the system used to control seeks is shown in Fig. 2. This nonlinear position control loop is used to optimize the time for long moves. The move lengths are always known ahead of the seek.

The first summing junction shown in Fig. 2 is a digital counter, which keeps count of the present distance error (number of tracks remaining to the target). This error signal is fed to a nonlinear velocity profile generator, which supplies the input to the velocity control loop. The position signal is fed back to the digital counter, which is decremented by 1 for every track crossing.

The nonlinear velocity profile generator contains a programmable read-only memory (PROM) and a companding digital-to-analog converter (C-DAC). Using a PROM allows any function to be programmed for the velocity profile. The companding DAC has a wide dynamic range. It requires only seven input bits to achieve an output ratio of 150:1

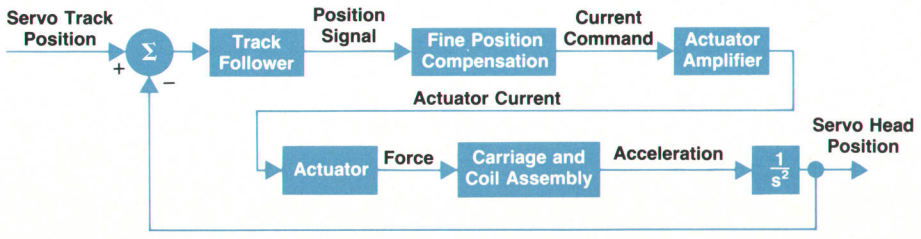


Fig. 1. *7933/35 Disc Drive position control loop block diagram.*

with 6% accuracy for signals near 0 volts. A linear DAC with a similar output ratio and low-end accuracy would require 11 bits.

The velocity profiles stored in the PROM are designed to provide the maximum possible current to the actuator, within the limits of the available supply voltage and the current rating of the amplifier. During the acceleration portion of a seek, the large distance error creates a large input to the velocity loop. The high gain of the velocity loop causes the power amplifier to saturate until the head velocity equals the command velocity. This accelerating current level is limited by the power supply voltage, the actuator coil resistance, and the back EMF of the actuator. Once the head velocity equals the command velocity, the heads start to decelerate. The deceleration velocity profile is designed to provide the maximum current possible for the given supply voltage, actuator coil resistance, back EMF, and power amplifier characteristics. Fig. 3 shows the basic shapes of the current and velocity waveforms during a track-to-track seek.

When the actuator heats up, the actuator coil resistance and back EMF constant change. A profile that works for a cool actuator would cause the power amplifier to saturate as the actuator heated up. To provide optimal seek times over a range of coil temperatures, four different profiles are stored in the PROM, each one designed to give optimum performance over a given range of actuator temperatures. At the start of each seek, the actuator coil temperature is checked by the microprocessor, which then selects the proper profile for that temperature.

This system gives seek time performance of 5 ms for single-track seeks, 24 ms for the average random seek, and 42 ms for the longest seek.

Actuator Driver Amplifier

One of the primary goals in the design of the 7933/35 Disc Drives was to keep the seek time as short as possible. As shown in Table I, the mechanical power required to move a fixed mass a fixed distance is inversely proportional to the third power of time, and the electrical power required is inversely proportional to the fourth power of time. A small reduction in seek time therefore requires a large increase in drive power. In the 7933, which has a moving mass of nearly 1.7 lb (0.765 kg) and a maximum seek distance of 2.1 inches, a nominal average seek (motion only, without overhead) takes 19.5 milliseconds, and a maximum-distance seek under nominal conditions takes 35 milliseconds. The power required to achieve this performance reaches peaks of 1000 watts, and the maximum acceleration is approximately 900 ft/s/s (28 times gravitational acceleration). On the other hand, when the drive is maintaining head position under servo control, the power requirements

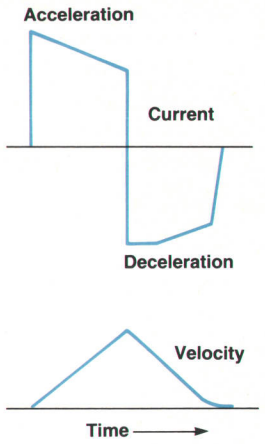


Fig. 3. *Current and velocity during a seek.*

Table I
Seek Power Derivation

Mechanical Power:

P_M = Mechanical Power
x = Distance t = Time
a = Acceleration v = Velocity
f = Force m = Moving Mass

$$x = \tfrac{1}{2}at^2$$

$$a = \frac{2x}{t^2}$$

$$f = ma = \frac{2mx}{t^2}$$

$$v = \frac{x}{t}$$

$$P_M = fv = \frac{2mx^2}{t^3}$$

Electrical Power:

K_F = Motor Force Constant
R = Electrical Resistance of Motor
I = Current through Motor
P_E = Electrical Power

$$f = ma = K_F I$$

$$I = \frac{ma}{K_F} = \frac{2mx}{K_F t^2}$$

$$P_E = I^2R = \frac{4m^2x^2R}{K_F{}^2t^4}$$

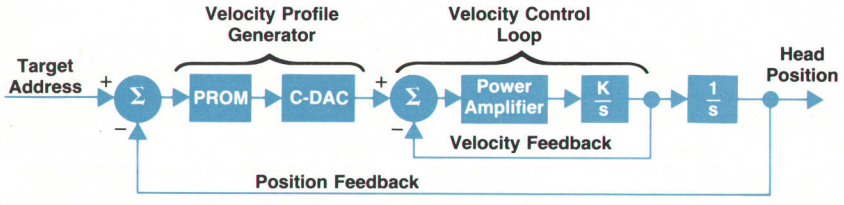


Fig. 2. *7933/35 seek control system.*

are on the order of 0.025 watts. A conventional linear amplifier can accommodate the required dynamic range, but is inefficient unless it is operating at low output currents or high output voltages compared to its design limits. A switching amplifier is efficient, but is impractical over the required dynamic range.

The 7933/35 actuator driver amplifier successfully combines the dynamic range of the linear amplifier with the efficiency of the switching amplifier. This was done by designing a dual-mode circuit that operates as a linear amplifier when the required output is small and as a switching amplifier at higher outputs. Although this is simple in concept, the design requirements of the two amplifier types often conflict. The most obvious conflict is the wide bandwidth required by an efficient switching amplifier and the relatively narrow bandwidth required to compensate the feedback loop in a multistage linear amplifier.

A simplified block diagram of the power amplifier is shown in Fig. 4. The output stage is a differential or bridge type, which is driven by one of two input stages depending on which mode (linear or switching) is selected. The mode selection is automatic and is determined by the amplitude of the input signal. Overall feedback is provided in the form of a voltage proportional to load current. This makes the circuit operate as a transconductance amplifier, that is, the output is a current in response to a voltage input. The linear mode input stage is a standard differential operational amplifier circuit in which the feedback signal is compared to the input signal. The linear loop compensation is also included here. A similar differential circuit is used in the switching mode input stage, but the gain and compensation are different. A pulse width modulator is included to convert the analog error signal (difference between actual and desired output) to a constant-amplitude pulse width modulated signal. The average value of the pulse width modulated signal is proportional to the switching mode error signal and the pulses may be of either polarity depending on the polarity of the error signal. The amplitude of the pulses is controlled and the power stage is designed so that only the output transistors saturate. This avoids latch-up and saturation recovery problems, which can occur if the earlier stages saturate. Also, since the compensation is placed in the first stage where the signal is still linear, the driver and output stages can be designed with very wide bandwidth for fast switching times. Notice that the linear mode compensation is independent of the switching mode compensation, allowing the two loops to be compensated separately. Mode control is accomplished by a pair of analog voltage comparators connected as a window detector.

Whenever the input signal exceeds a preset threshold of either polarity, the mode control circuit disconnects the linear input stage and connects the pulse width modulator to the power driver. A fast-acting electronic switch is used to achieve the desired switching time.

The power output stage is a pair of fully complementary linear power amplifiers which are driven differentially. Local feedback is used to control the stage gain and ensure stability. Complementary Darlington transistors are used for the predriver and output sections of the stage to achieve the required gain with a minimum number of parts. The overall feedback signal representing the load current is obtained from the voltage across a small current sensing resistor in series with the linear motor. Since this resistor floats with respect to ground, a differential amplifier is required to convert the feedback signal to a voltage referenced to circuit ground.

Track Follower

The basis for track-to-track seeking and on-track servoing in HP drives is sensing track position by monitoring magnetically recorded servo information. The circuitry that gives the track position information is called the track follower.

A single surface in the center of the multidisc pack is dedicated to track position information. The information recorded is called servo code. The servo surface is composed of servo bands written in concentric circles across the surface of the disc. The servo bands alternate between A and B bands. Data track centers occur when the servo head is positioned over equal portions of A and B tracks. As the servo head moves off track center, it encounters more of one and less of the other, producing a positive or negative signal corresponding to the direction of motion. Using this position error signal, the servo system adjusts the position back to track center.

The servo code is marked at intervals with sector timing information. The track follower circuit decodes these timing marks and provides start-of-sector information to the read/write and drive control circuitry.

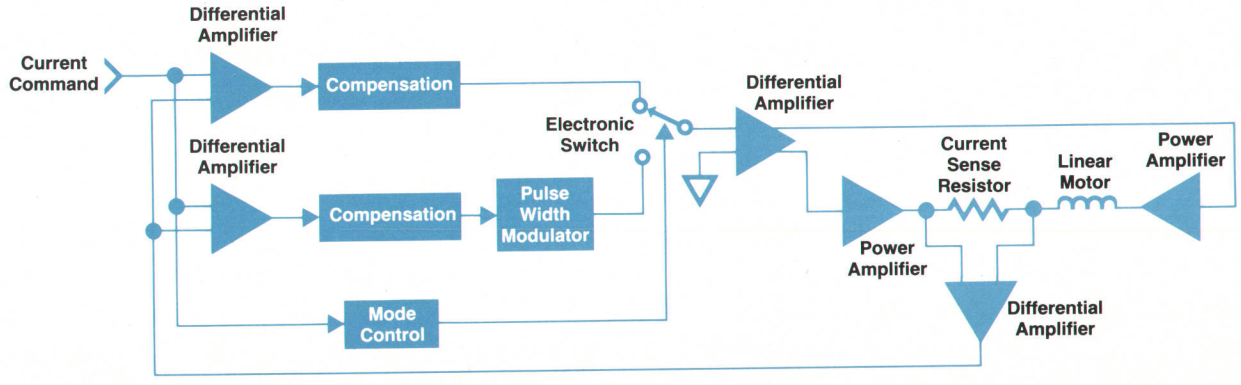


Fig. 4. *Actuator driver amplifier block diagram.*

Servo Code Format

As mentioned above, the servo surface consists of alternating bands of A and B servo information. Fig. 5 describes the format of the servo code. Each field consists of pairs of closely-spaced magnetic transitions. The electrical response to each pair of closely spaced transitions is called a di-bit.

The A and B bands of servo information are similar, differing only in the position in time of the closely spaced transition pairs. The electrical response of a head positioned over equal portions of A and B bands is a series of di-bits. Di-bits from the A band alternate with those from the B band.

The sector timing marks are sequences of extra di-bits that are placed between the A and B servo di-bits. They contribute nothing to the position sensing system, but are detected as a sequence to generate sector timing information. The timing information is also used to identify the A and B di-bit information.

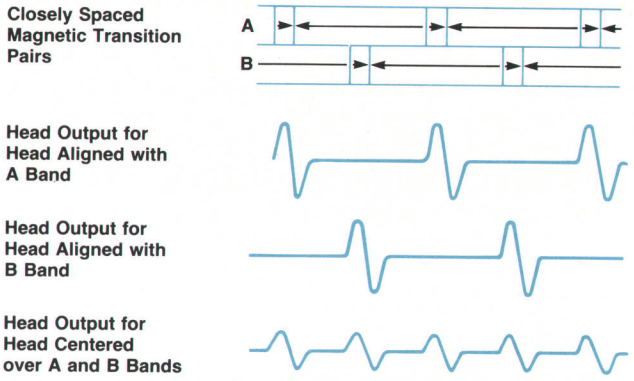


Fig. 5. *Format of the servo surface, a surface in the center of the multidisc pack dedicated to track position information.*

Circuit Functional Description

Fig. 6 is a block diagram of the track follower. The servo head signal is amplified by the AGC amplifier chain. The resulting signal is fed to the phase-locked loop, the position discriminator, and the level detecting circuitry. The phase-locked loop locks to the servo code waveform and produces timing signals that are used by the amplitude detectors, sector counters, and position discriminator. The position discriminator uses the incoming servo code and the phase-locked timing reference to demodulate the A and B amplitudes. The A and B amplitudes are subtracted in a differential amplifier to produce the position error signal.

The amplitude detection circuit uses phase-locked gating signals to detect the sector timing sequences encoded in the A and B bands. The detector outputs are used to synchronize the sector counters and initialize the start-of-sector timing pulse.

The AGC reference is derived by summing the A and B amplitudes. The resulting signal is fed to the AGC integrator. The integrated output is fed into the AGC amplifier chain to keep the sum of A and B constant.

Di-Bit Integrator

The preceding functional description holds for many HP disc products. In the 7933/35, the di-bit integrator deserves special attention. Fig. 7 is a simplified schematic. Usually, the di-bit amplitude is peak detected by A and B peak detectors, which are carefully matched. The 7933/35 uses a dumped integrator followed by matched sample-and-hold amplifiers to separate the A and B channels. The process of integration has superior noise immunity compared to the process of peak detection, if properly implemented.

Our implementation, shown in Fig. 7, avoids the difficult task of matching dual integrators by using the same integrator to process both A and B di-bits. After the integration cycle, the integrator output is sampled by either the A sample gate or the B sample gate. Then it is reset to zero by the dump pulse to initialize it for the next integration cycle. This implementation requires only careful matching of the sample-and-hold circuits rather than matching integrator circuits.

The implementation consists of matched transistors Q1 and Q2, which are arranged in a circuit similar to the Widlar current source.[1] R1 biases Q1 with a very small collector current. Because of the matched transistors, the quiescent current in the collector of Q2 should be equal to that of Q1. However, because of the presence of emitter resistor R3 and base connecting resistor R2, the quiescent current in Q2 is less than that in Q1.

R1 sets the collector current of Q2 to a level sufficiently small that there is no significant voltage change in capacitor C2 in the quiescent state. The inverted servo signal is ac-coupled through C1 to the base of Q2. The negative swing of a servo di-bit pulls Q2 into full conduction. A bias voltage very near to that of the di-bit is imposed across the emitter resistor. The emitter current set by resistor R3 is proportional to the di-bit voltage waveform and flows through Q2 into capacitor C2. The voltage level at C2 is proportional to the integral of the di-bit waveform. The resulting circuit operation produces a smooth and accurate position error signal.

Automatic Head Alignment

Previous HP disc products have used a special electronic tool and procedure to allow the HP Customer Engineers who service the drives to adjust the alignment of the data heads to the center of the data tracks. A special pack (CE pack) is placed in the drive along with the electronic tool. The electronic tool indicates the amount of adjustment required to bring the head to track center. An alignment procedure is performed if the head is out of specification. This is an expensive and difficult process. First, each CE pack must be very closely matched to other CE packs. Second, each CE must carry a bulky and expensive CE pack to the customer's site. Third, very tight control of electrical and mechanical specifications is required to keep environmental factors such as temperature from causing large head position shifts and data errors. Fourth, higher track densities cause these error effects to become so severe that a higher-density drive could not be built if this traditional alignment method had to be used.

The 7933/35 adapted the traditional approach to head alignment by including the special electronic tool in each drive and special reference tracks in every production pack.

Supplied with special controller firmware, the drive is able to correct errors up to ±1/4 track. This adjustment range allows the drive to operate at a significantly higher track density without increasing mechanical tolerances.

Each read/write board includes a circuit that duplicates the position sensing and sector timing circuits of the track follower board. Autoalignment bands, which duplicate a pair of servo bands, are included on each data surface. Alignment bands are located at the outside, near the center, and near the inside of each data surface. Periodically, the drive controller requests the servo system to position the data heads over an alignment band. When the alignment circuitry is then enabled, the drive controller measures the resulting position error signal from the selected data surface. A command is issued to the servo system to offset a corresponding amount in the opposite direction. The process is repeated up to four times at each band or until the error is small enough to be insignificant. This entire alignment process is repeated on each surface and at all three bands. The offset value required to null the error is recorded for each surface at all three bands. When any track on any surface is then accessed, the offset values for the alignment bands on either side are recalled and an offset is calculated by interpolating between the two alignment values.

The autoalignment approach limits drive head alignment errors to less than 3% of track width (50 μin) over widely varying temperature conditions (10 to 40°C ambient).

Automatic Skew Correction

The autoalignment circuitry also contains adjustments for correcting the circumferential alignment error of the heads with respect to the servo head. In previous drives a wide intersector gap was created to compensate for misaligned heads. The automatic skew correction circuit allows the intersector gap to be decreased by 80%. Virtually all the remaining gap is required for reasons other than physical misalignment.

The autoskew measurements are made at the same time as the autoalignment measurements. The alignment bands on each surface contain sector marks identical to the servo surface. The difference in time between the sector marks on each band and the servo surface is measured by the skew correction circuit on the track follower board and then stored by the drive controller along with the autoalignment information. Any required skew correction is computed in the same manner as the autoalignment correction and fed back to the autoskew circuit, which can advance or delay the sector timing information by the required amount.

The correction capability of the circuit is ±1/6 sector. This greatly reduces the precision required for mounting heads in the drive assembly.

Writing Servo Code for Automatic Head Alignment

The autoalignment functions require very accurate and repeatable location of the alignment bands from band to band on each data pack. In the case of the 7935 data packs, they also require good repeatability from pack to pack, since the packs will be interchanged between different drives. This means that each data pack must be produced and formatted to an accuracy equal to that of a CE pack.

It is the job of the servo writer to format the servo surface at the factory with position and timing information and to write the head alignment bands on the data surfaces. The alignment bands consist of servo information located in three bands on each of the data surfaces. Each band contains one track center per data surface, and each of these must be accurately located above the corresponding track center on the servo surface.

In a basic sense, the servo writer is very similar to a disc drive. It must seek and position a carriage that holds all of the data heads and it must read and write to all surfaces on the data pack. In fact, much of the servo writer electronics consists of modified 7933 boards. However, most of the mechanics consists of specialized parts which provide the best possible system repeatability and freedom from vibration. The major differences are that the servo writer must position the data heads very accurately and repeatably without servoing on a servo surface and it must provide an external source of accurate circumferential timing.

The major technical challenges in designing the 7933/35 servo writer were:

- To provide accurate, repeatable positioning and circumferential timing
- To remove or compensate for any thermal or mechanical offsets and vibration
- To design an automated system for repeatability, accuracy, and data collection
- To overcome long thermal stabilization times and achieve less than a 20-minute process time per pack
- To provide a means of calibrating and monitoring servo writer performance on a day-to-day basis
- To be able to produce multiple servo writers with identical performance.

Circumferential timing is established by phase locking to an optical encoder mounted on the spindle shaft. To provide accurate positioning, we use an air-slide carriage

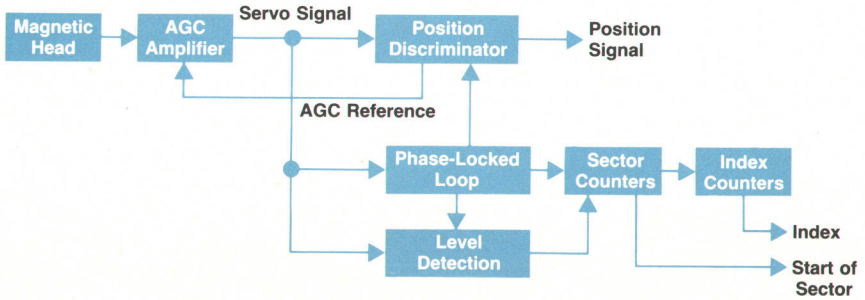


Fig. 6. *Simplified track follower block diagram.*

bearing driven by a linear motor and a positioning system based on the HP 5501A laser and optics. Position is measured by reflecting the laser beam off a retroreflector, which is placed just behind the head mounting point on the carriage. The laser information is then converted to track position information and fed into a normal disc drive servo board and actuator amplifier electronics. The carriage can be offset in quarter-wavelength (6.25 μin) increments over its entire length of travel for fine adjustments or it can seek a given number of tracks (1600 μin) for larger displacements. The initial position reference is established by pulling the carriage back against its end-of-travel stops.

The accuracy with which the servo information can be written is limited by the amount of vibration and thermal and mechanical offsets in the system. The vibration problem was solved mechanically by using an air bearing spindle with an integrated optical encoder and by building the servo writer on an air-shock-isolated table. Long-term thermal drifts caused by the environment were solved by operating the servo writer in a temperature controlled room (±0.5°C). The mechanical offsets are removed in the calibration procedure.

It is important that this machine be automated to ensure that the operations are performed and monitored exactly the same way for each pack and to free the operator from tedious machine adjustments, instrument reading, and data logging. Automation allowed us to develop format, verify, and calibrate routines that iterate their procedures, adjusting the location of each alignment band until it is within specification; this is done very rapidly. The servo writer is a Z80-based system using the 7933 processor board. All of the production programs and maintenance subroutines are stored in EPROM within the servo writer and all of the data results and servo writer calibration information are sent to the production area data base system via an RS-232-C link. Using the processor-based system allowed us to incorporate many automatic checks on system calibration, operating conditions, and data results which increase the integrity and observability of the system.

The long periods of waiting for thermal stability normally encountered in writing CE packs were minimized by paying close attention to the format, verify, and calibrate routines and to the amount of down time allowed for the servo writer. While the heads are loaded on the discs, friction is generated, and the heads and discs heat up and begin to expand, although not all of them equally. The amount of expansion depends upon the particular disc (middle, top, or bottom), the location of the heads, the time spent there, and their previous thermal history. Our problems were largely overcome by designing the format, verify, and calibrate routines to locate the data heads over the same portions of the disc for the same amount of time and to limit the amount of down time allowed between packs. In this way, the machine is warmed up and achieves a state of process equilibrium from which only short periods of time are required to stabilize the position of the heads. If the system is down for an extended period of time, warmup (simulated verify) operations are performed until the system is again ready for use.

The servo writer uses a set of standard packs and corresponding correction factors to calibrate itself for daily operation. These packs have been carefully selected and tested many times to determine the position inaccuracy (correction factor) for each head and data band. The servo writer measures the difference between where it thinks the corrected position of the alignment band should be and where it iterates to find it. This difference is saved in the servo writer and becomes the system calibration factor for that head and alignment band. Then, for each pack that it writes afterwards, it will offset its position by an amount equal to the calibration factor, and thus each data pack is written to look exactly like a standard pack.

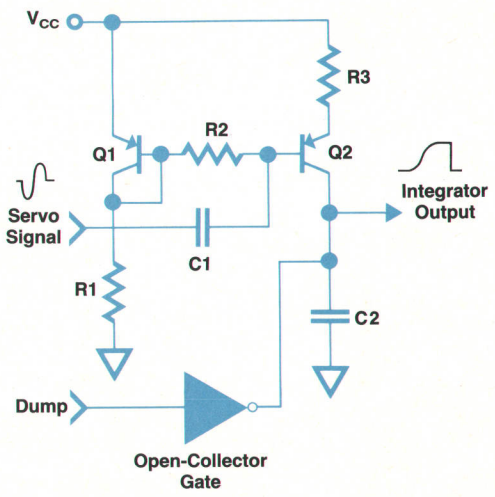


Fig. 7. *Simplified schematic of the di-bit integrator, part of the position discriminator in the track follower circuit.*

The servo writer calibration factors are sent to the production data base system and can be monitored to determine the long-term stability of the servo writer and the condition of the standards packs. There should be no sudden step changes in the values of the calibration factors, so limits are set in the servo writer. If these limits are exceeded, the system will report a calibration error, will request service, and will not format any more packs.

The volume of packs required for the 7933/35 products meant that several servo writers would be needed. To duplicate the servo writer and to prove that a single servo writer could produce packs repeatably within tolerance, a large amount of testing was done to determine the system repeatability. Error components were identified and measured and an error budget was drawn up. Each succeeding servo writer is now required to go through a certification testing program in which its accuracy and repeatability are checked and verified before it is released to production.

Acknowledgments

The servo writer design team included Dale Wolin and Paul Mui, with assistance from Keith Grethen. Thanks also to the technicians and assemblers who helped with the servo writer development. Frank Klekner, Sandy Greeley, Dan Makabe, Ron Miller, and Rich Priest were of great help from the early phases of the project. Dave McIntyre wrote much of the firmware associated with the servo electronics. Production engineering support for this area was provided by Sam Gailbreath. Mike Rusnack was the reliability en-

gineer. Mike Byce and Tom Moore provided suggestions on field serviceability from Customer Engineers. The final test system was handled by Warren Greaves, Chick Slutz, and Terry Smith. The board tests were designed by several people in Bill Horner's group. Art Beale was involved in many aspects of the drive design, and his contributions to the final design are greatly appreciated. Jim Brezinsky designed the spindle motor amplifier. The mechanical engineering section was faced with several difficult problems in the servo area. Thanks to Larry Albrecht, Art Beale, Rick Connolly, Pat Donnelly, Steve Edwards, Paul Mui, Jim Smith, Ed Walsh, and Charlie Woodard for their efforts in solving these problems and for the mechanical design of the product.

Reference

1. R.J. Widlar, "Some Circuit Design Techniques for Linear Integrated Circuits," *IEEE Transactions on Circuit Theory*, Vol CT-12, 1965, pp 586-590.

Mechanical Design of a Large Disc Drive

by James H. Smith

ALTHOUGH THE HP 7933/35 DISC DRIVES can be considered higher-performance replacements for the earlier HP 7925 (see Table I), they bear little resemblance mechanically to the older drive. A new mechanical design approach was taken, with high-volume manufacturing the goal. Many innovative concepts were evolved, both for the drive and for the media module.

Because of the projected volumes for the 7933/35, we were able to show that payback periods of three months or shorter could be obtained by investing in parts tooling. We therefore investigated high-volume parts tooling on the front end of the project. We also looked at various ways that we could combine the functions of various parts and decrease the overall parts count.

The approach that we finally decided upon is shown in Fig. 1. This design is based on a molded cabinet with a maximum amount of molded-in details and modularity of design. There are four major assemblies or modules: the actuator spindle base, the power supply, the card cage, and the cabinet itself.

Cabinet Design

The cabinet is a two-part polycarbonate foam molded design consisting of the cabinet or outer shell and the component mounting frame. The two units are bonded together by an induction heating process, using a compounded polycarbonate material (with metal flakes) as bonding material, to form one structural cabinet part. All of the cabinet parts are molded using straight pulls to eliminate slides, which would be costly and impractical because of the size of the parts involved. Each of the molded cabinet parts weighs approximately 32 pounds, making this the largest molded cabinet of its kind in the computer industry. The component mounting frame provides for mounting two of the three subassemblies, the actuator spindle base and the card cage. In addition, there are details for mounting the filter, the prefilter, and ducting to direct cooling air to the card cage and the power supply. The power supply is mounted in the cabinet portion of the assembly and is accessible through the rear of the cabinet.

The molded cabinet probably presented more problems to the industrial designer than to the design engineers. To obtain the advantages of simple tooling and consequently a straight pull from the mold, we had to add one degree of draft to the cabinet sides. This was considered a poor industrial design practice, but once models were made that demonstrated the function of the parts as well as the draft, it did not appear to be very important. When the advantages of the design were weighed against the potential aesthetic disadvantage of one degree of slope to the sides, the practicality of the design won out.

Table I

Performance Comparison
7933/35 versus 7925

	7925	7933/7935
Storage	120M Bytes	404M Bytes
TPI	384	625
Track Width	2600 μin	1600 μin
Recording Surfaces	9	13
Access Time	25 ms	24 ms
Maximum Off-Track Error	500 μin	260 μin
Peak Force to Seek		
Acceleration	110.0 Newtons	208 Newtons
Deceleration	70.0 Newtons	175 Newtons
Kinetic Energy Required for an Average Seek	0.5 Joules	1.5 Joules

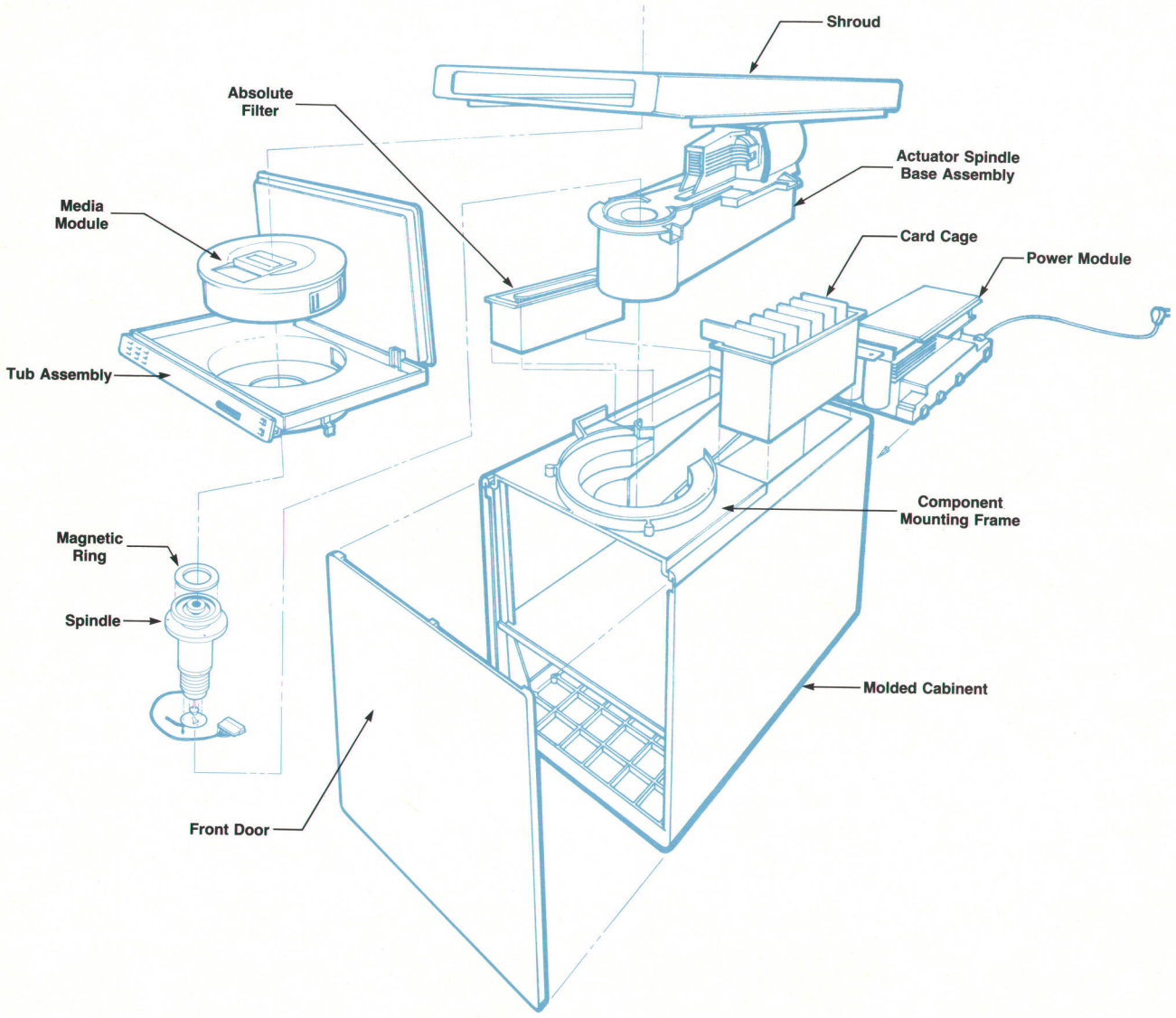


Fig. 1. *Exploded view of the 7933/35 Disc Drives.*

Modularity

The modularity of the system is made possible by separating the mechanical and electrical assemblies. We were able to have a modular card cage by standardizing the printed circuit board size. There are eight boards in the card cage and all but one are the same size. The read/write board was altered to make room for the preamp circuitry, but it still fits within the card cage. The other boards—the track follower, formatter/separator, DMA, microprocessor, servo, actuator driver, and power regulator—are all on standard HP boards. The card cage is a die casting and its cover is a die casting with a stamped and punched sheet-metal insert. The card cage is completely tested as an assembly before it is put into the cabinet.

The actuator spindle assembly is a die casting. In it are mounted the spindle motor, linear motor, and actuator assembly. This assembly is completely built and aligned and checked out before it is installed. This is a major mechanical assembly of the drive and presented the most problems during the design (see article, page 23).

The power supply is built on a die-cast base and is also completely assembled and tested before assembly.

7935 versus 7925

Comparing the total number of cast, molded, sheet-metal, and machined parts for the 7935 versus the 7925 shows that the mechanical parts have been reduced by 26% from 509 parts to 376 parts. The sheet-metal parts count is 59% lower and the machined parts count shows a reduction of 42%. For molded and cast parts, the amount of tooling has increased by 137%, from 72 parts for the 7925 to 171 for the 7935.

The assembly area required for the 7935 is about 50% of that required for the 7925. The reasons for this are the modularity of the design, the fact that the drive cabinet itself is the workstation, and the new material flow con-

cepts, which dictate that a minimum amount of material should be in the assembly area at any time. By having completed and tested modules that can be dropped into the cabinet, the costly build stands that are required for the 7925 are eliminated. These stands were not only expensive, but took up valuable space.

Because all of the modules are tested at the subassembly level and all of the mechanical components are tested before assembly, the 7935 was able to go to a 24-hour test cycle which includes burn-in. The 7925 requires a 48-hour burn-in before testing, because the 7925 components cannot be tested at the subassembly level; consequently, it is necessary to wait until they are all assembled before they can be burned in and tested. This has allowed the 7935 to use a five-step final build and test operation instead of the eight-step operation of the 7925. The 7925 goes from top-level build to initial turn-on to burn-in to put-up to diagnostic testing to button-up, integration, and final test. The 7935 combines the initial turn-on, burn-in, and put-up steps into one step and does not do the integration. This reduction in steps has not only reduced assembly labor but floor space as well.

The manner in which the 7935 is built and tested compared to the 7925 is probably a good proof for the statement made by W. Edwards Deming, "The later you wait in a process to test, the more it costs you and the harder it is to find a problem."

The 7935 was able to build up to mature production in about four months versus twelve for the 7925. The quantity at mature production is about 50% higher than for the 7925. Part of the reason for this is that mechanical parts do not restrict production. 171 parts are tooled for high-volume production. All of these were tested at the production prototype and pilot-run stages. Once production had started, vendors were able to supply very large quantities of parts in very short times. The materials engineers were involved very early in the design stages of the project and were able to design in the material flow from the vendors along with the parts and tooling design. There was a very large commitment early in the project to high-volume production in a very short period of time. Critical purchased parts were negotiated with the vendors long before the release of the project. The material flow of the critical path was worked out well in advance.

Media Module

A major design feature of the 7935 Disc Drive is the media module.

Handling and cleanliness of disc media has always been approached with a certain fear and mystique. Media damage and subsequent loss of data are usually blamed on contamination and/or handling-induced defects. Typically, drive manufacturers have used existing media cartridge and pack designs for their drives. When we investigated these existing designs, we found that they did not satisfy our design requirements. Through the investigation of these designs and extensive testing we were able to come up with a design that minimizes the effects of contamination and handling, the two major factors in media longevity.

The media module consists of seven 14-inch oxide-coated discs assembled onto a hub, which contains a precision-ground armature plate at the base. This armature plate completes the magnetic circuit created by a magnetic ring mounted to the spindle hub. When the module is inserted into the spindle, this circuit secures the disc stack to the spindle with a clamping force of approximately 45 lb. The disc assembly is located on the precision-ground nose of the spindle in the radial direction by a three-pronged flexure which is part of the disc hub assembly. This arrangement allows the pack to be removed from and put back onto the spindle with a repeatability of eight microinches. The handle assembly acts as a cam mechanism; when lifted, it separates the hub from the spindle, allowing for pack removal.

The pack assembly is contained in a two-piece polycarbonate injection-molded case. Upon insertion into the pack chamber and onto the spindle hub, the pack is rotated to open the sliding door, which allows head access and a flow of clean filtered air onto the discs. Thus the door to the discs is only opened in the clean environment of the pack chamber to minimize the effects of contamination. In addition, a purge and cleaning cycle is used before data is written onto or retrieved from the media.

Specifications call for the module to maintain protection after a 30g impact with a 12-ms duration. Typically, the module will withstand 75g and still provide protection for the media.

Acknowledgments

I would like to acknowledge all the members of the mechanical design team for their contribution to the mechanical design of the 7933 and 7935 products. This list includes: Larry Albrecht for the design of the media module mechanical components, Art Beale for the initial spindle design and the data acquisition and retrieval system for the servo system testing, Rick Connolly for the front-panel and acoustical design and for his contamination/removability work, Pat Donnelly for the power module and the thermal and RFI design, Steve Edwards for the actuator design, Paul Mui for the media module design and the mechanical design of the servo writer, Dave Prouty for the spindle motor and RFI design, Ed Walsh for the spindle mounting base design and servo analysis, and Charlie Woodard for the overall product design (including the cabinet) and mechanical servo analysis. We are all grateful to Ross Casey and Gene DeShon for their valuable assistance and advice during the parts and tooling design and for their patience in helping the engineers design parts for high-volume manufacturing. Finally, thanks to Mike Okamura, who drew Fig. 1 for us on very short notice.

CORRECTION

In the November 1983 issue, the two captions at the bottom of the box on page 8 are reversed. The line chart is on the right. The slide is on the left.

High-Capacity Disc Drive Servomechanism Design

by Stephen A. Edwards

THE ACTUATOR SPINDLE BASE ASSEMBLY (ASB) was the major mechanical design project for the 7933/35 Disc Drives. Two of the fundamental design considerations for the ASB were modularity and primary functionality. In the past, HP products have tended to use the ASB as part of the structural cabinetry, thus subjecting critical components to external stresses and vibrations which could adversely affect their rigidity and precision alignment. By defining the cabinet/ASB interface early in the 7933/35 project, we were able to separate all the mechanically critical components, specifically the spindle, linear actuator, carriage/rail systems, and the base which supports them, from structural constraints and concentrate on the complicated servo performance issues. By eliminating many of the smaller extraneous subassemblies from the main base, several potentially annoying vibrational modes were avoided. As an added benefit, a single high-precision assembly emerged which requires fewer tools and less assembly space and is fairly simple to monitor and troubleshoot. Subassemblies are pretested before reaching the final assembly station, resulting in higher yields.

Fig. 1 shows the components of the actuator and base.

Spindle

The spindle assembly consists of a precision shaft ground on centers, supported by two ABEC-7 class ball bearings, and an internal preload spring all encased in a steel permanent mold casting. Extremely tight runout requirements dictate stringent quality control at all phases of manufacture and assembly. Final grinding is done in-house, on the spindle's own bearings, to obtain total runouts less than 50 μin peak to peak with nonrepeatable components representing less than 15 μin p-p of the total.

The driving torque is supplied by an integral dc motor with a latching Hall-effect sensor used for shaft position and velocity feedback to the spindle driver electronics. Rotational speed of 2694 r/min is allowed to vary only $\pm$0.5% because of the extremely narrow bandwidth of the read/write phase-locked loop.

Linear Actuator

To supply the necessary radial access capability, a linear motor actuator system is employed. The linear motor housing is a 1010-steel shell mold casting containing four 76-degree sections of M7 grade ceramic magnets, radially magnetized to achieve a gap flux density of 2700 gauss. The armature is a three-piece bonded assembly with a 200-turn rectangular aluminum wire coil, developing a nominal force constant of 12.4 newtons/ampere throughout the 54.8-mm stroke.

It is critical to the proper operation of the drive to establish and maintain a single degree of freedom for the head stack (i.e., radial to the disc center). Radial location is decoded by the dedicated servo head in the center of the seven-platter disc stack. The remaining thirteen heads are used for data manipulation based on the perceived servo head location. Sector timing information is also determined through the servo code. Any nonrepeatable motion of data relative to the servo heads results in misplaced data with respect to the anticipated cylinder and/or sector. This misplaced data, upon subsequent readback, results in unpredictable and typically uncorrectable errors.

The dynamic components of the actuator consist of the armature, the carriage assembly, the tach rod, and fourteen magnetic read/write heads. The maximum acceleration seen by the nominal 765 grams of the carriage system's moving mass is 28g.

Structural support for the carriage system relies upon three precision dual-row miniature ball bearings, preloaded between two tungsten carbide gauge rods. On the outer surface of the bearings is ground a gothic arch which,when mated with the rod, creates two-point contact. This suspension system results in compact single-degree-of-freedom motion as long as bearing/rail contact is maintained. Attached to the carriage and suspended inside a coil of 44 AWG wire are the tach rod and magnet which supply velocity feedback (nominally 2 V/m/s) to the servo system during long seeks and head loading.

By removing the heads from the disc chamber, it is possible to remove the media in a module that is both easy to handle and inexpensive. Because of the nature of the head/oxide media interface, it is necessary to load and unload the heads from the media while the discs are spinning. This is achieved by means of a head ramp and separation device called a cam tower.

Base

The base supports the linear actuator assembly and the spindle assembly. Initially, the base was envisioned as being an infinitely stiff beam which would support and maintain the two assemblies in their required orientation. Soon after the first unit was assembled and seek testing began, we found how wrong this assumption was. As shown in Figs. 2 and 3, the open-loop servo transfer function indicated a phase margin of approximately eight degrees at the gain crossover frequency of 353 Hz. Fairly minor mechanical variations can degrade this margin, resulting in an unstable condition. After several iterations of beam models and prototypes, it became apparent that a 16-in dumbbell with 30 lb on one end and a 35-lb random moment generator on the other was a very uncooperative device. The moment generated by the linear motor, when

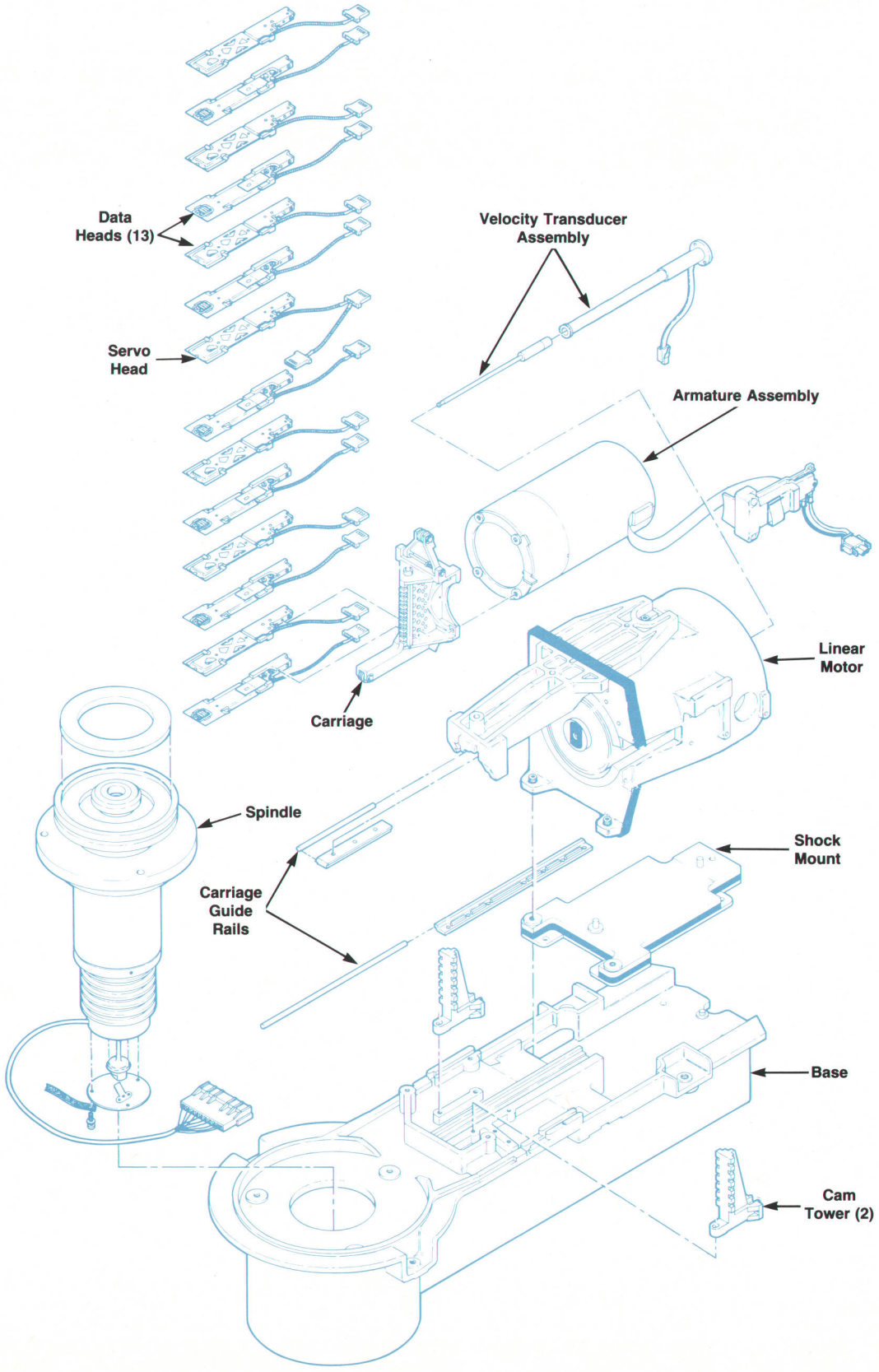


Fig. 1. *Elements of the actuator spindle base assembly, the major mechanical component of the 7533/35 Disc Drives.*

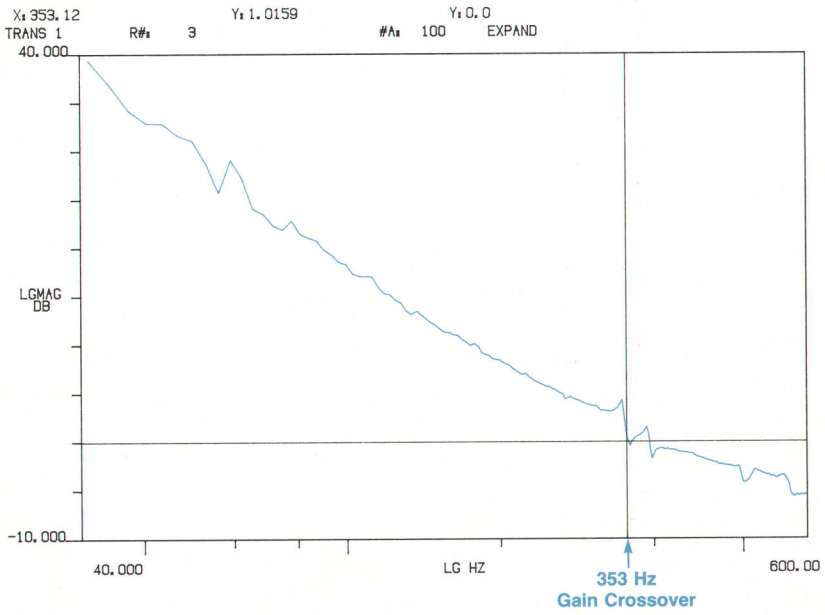


Fig. 2. *7933/35 servomechanism open-loop transfer function without shock mount. Open-loop gain crossover is at approximately 353 Hz, and the gain drops sharply there.*

coupled to the base, drove a 400-Hz bending mode of the base assembly. This was so close to the crossover frequency that there was no servo gain to correct for position errors (see Fig. 2). The only solution seemed to be to decouple the source of random energy input from the rest of the system. What we needed was a thermally and dimensionally stable composite that would absorb vibrational energy at a low frequency (approximately 60 Hz), where the servo system has 20 dB of gain to correct for position errors.

The result of these findings is a shock mount. This is a vulcanized sandwich of two aluminum plates with a specially formulated butyl rubber between them. Its effect is to reduce the moments applied to the end of the base by the reactive load on the actuator caused by accelerating 1.7 lb at 28g. The change in servo performance is illustrated by the open-loop transfer functions in Figs. 4 and 5. The phase and gain drops at crossover have been eliminated and the effective crossover frequency has increased to approximately 486 Hz. Design tradeoffs involved larger assembly tolerances, seek-length-dependent servo performance caused by relative velocity between the velocity transducer and the base, and the requirement to maintain

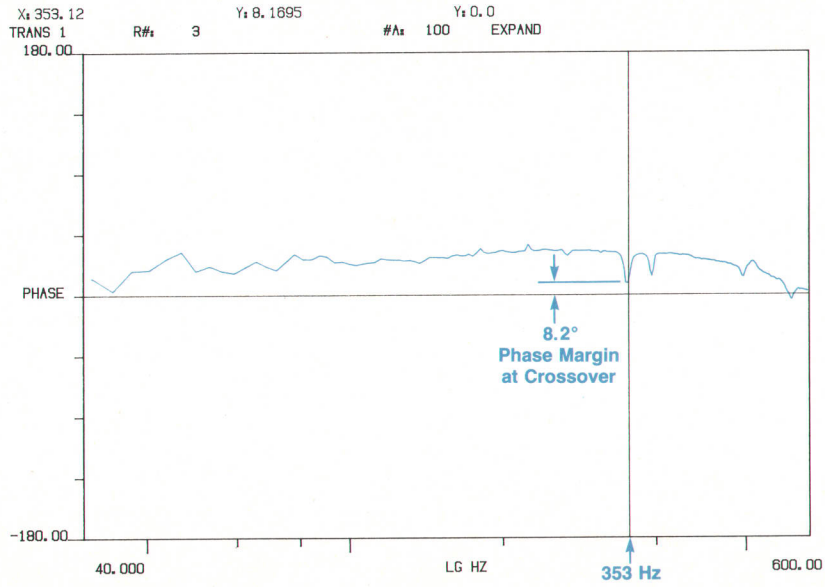


Fig. 3. *7933/35 servomechanism open-loop transfer function without shock mount. Phase margin at gain crossover (353 Hz) dips to only 8.2°.*

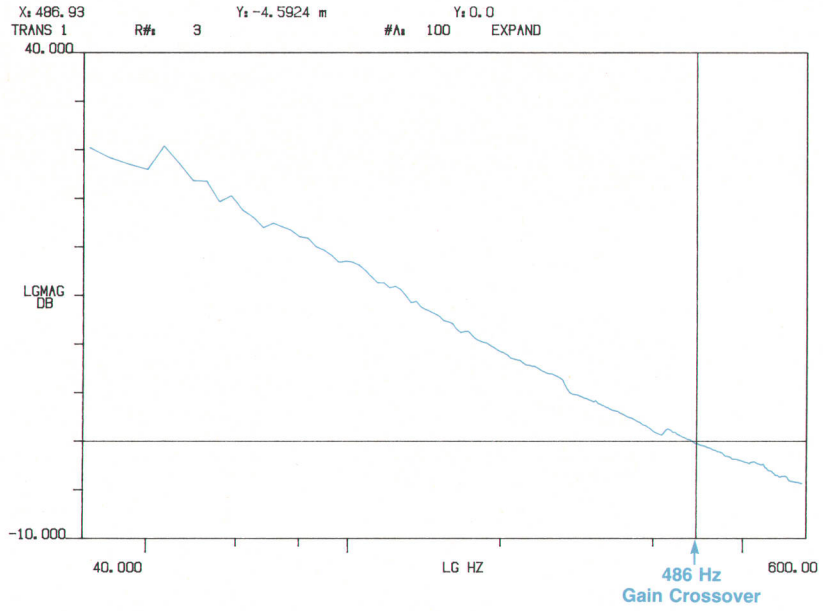


Fig. 4. *7933/35 servomechanism open-loop transfer function with shock mount. Gain crossover frequency has been increased to 486 Hz and there is no sharp drop.*

critical composition tolerances on an elastomeric material.

Drive Performance Issues

There are two major factors to consider when discussing overall drive performance. The first is whether the servo system can seek and accurately locate the servo head where desired within a fixed time (preferably very short). The second is whether the data heads follow the servo head.

Servo head settling time is the period required for the carriage assembly to move and settle within a quarter track width of the target track center. When within a quarter track (400 μin), the track follower circuitry homes in, and at a suitable time, enables the read/write circuit to use one of the other 13 heads. If the servo system is unable to position the head in the track follower range within a software-specified time limit, a seek error is flagged and the drive does a recalibration and reattempts the seek.

To measure accuracy and settling time performance, we determined the open- and closed-loop system transfer functions and settling characteristics using HP Digital Signal Analyzers. Initially, the error rate observed was quite poor, even when the servo position was extremely accurate. This

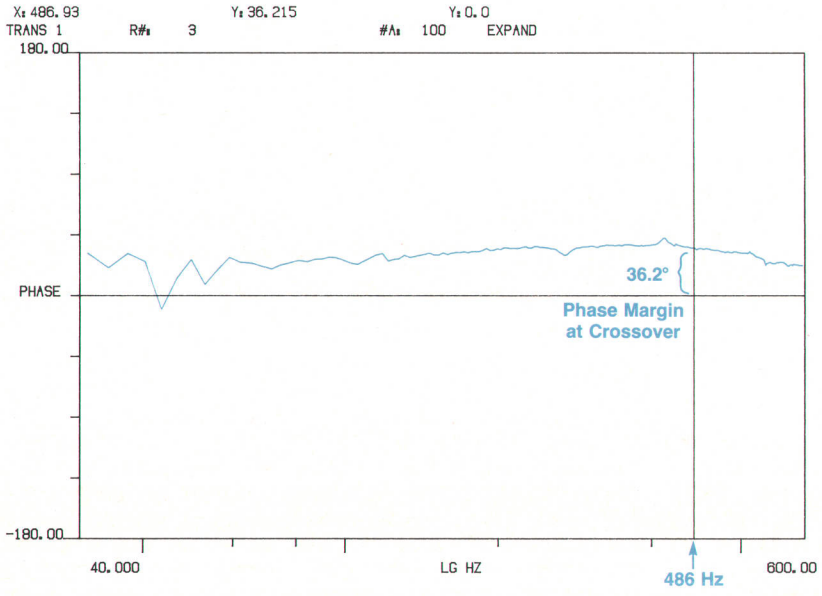


Fig. 5. *7933/35 servomechanism open-loop transfer function with shock mount. Phase margin at gain crossover is a comfortable 36.2°.*

was a result of the data heads' going astray with respect to the servo head. The major factor allowing this to occur was that the carriage bearings were not maintaining contact with the rail. As the carriage tilts, the heads above or below it are no longer in a vertical line perpendicular to the disc surface, that is, the heads are not in the same cylinder. This causes data to be miswritten with respect to the servo reference and therefore causes errors when the data is retrieved.

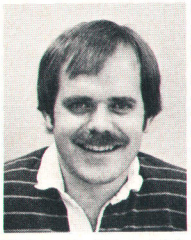
One major cause of this behavior was a nonsymmetric magnetic field in the linear motor, which applied a horizontal force to the coil upon excitation. This resulted in a moment in the carriage system, forcing the bearings to slide sideways with respect to the rail, thereby tilting the carriage up or down dependent upon the polarity of the coil current. By balancing the magnet volume about the horizontal axis of the linear motor, this effect was eliminated. This phenomenon was monitored using an instrumentation pack (servo code written on all surfaces) and several sets of track follower electronics. It was thereby possible to see relative settling between servo and data heads.

Authors

January 1984

3

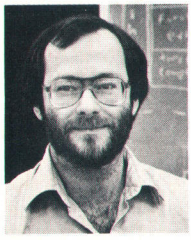
Kent Wilken



Kent Wilken received his BSEE and MSEE degrees in 1976 and 1977 from Stanford University. He joined HP in 1977 as a development engineer doing digital circuit design. He's now a project manager for disc controller design at HP's Disc Memory Division in Boise, Idaho. Kent was born in New Haven, Connecticut and now lives in Boise. He enjoys tennis, skiing, and other outdoor sports.

5

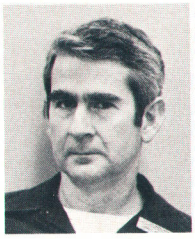
Douglas L. Voigt



Doug Voigt received his BSEE degree from Cornell University in 1978 and joined HP the same year. He participated in the specification and implementation of Command Set 80 and later served as project leader for CS 80 development and enhancements. A native of Pennsylvania, Doug now lives in Boise, Idaho. He plays keyboard musical instruments, including performing and composing music for them.

7

James D. Becker



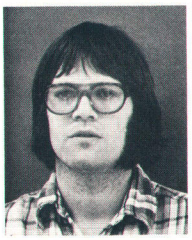
James Becker received his BSEE degree in 1967, his MSEE in 1968, and his PhDEE in 1980, all from the University of Texas. He joined HP's Disc Memory Division development lab in 1980, contributed to the design of the 7933/35 Disc Drives, then moved to manufacturing engineering to support the 7933/35. He's now doing test equipment design. Before joining HP, he designed analog and digital circuits, wrote BASIC, FORTRAN, and assembly language software, and served as a project manager in an office products development lab. He also put in 3½ years of active duty with the U.S. Air Force and is now a major in the Air Force Reserve with a total of 15 years' service. Born in Columbus, Texas, James is married and the father of twins. He lives in Meridian, Idaho, and enjoys photography, hunting, and gardening.

Robert M. Batey

Bob Batey is a development engineer with HP's Disc Memory Division in Boise, Idaho. With HP since 1975, he has contributed to the design of the read/write electronics of several disc drives including the 7933/35. Born in Gainesville, Florida, he attended the University of Florida for a pair of BS degrees, one in education (1972) and one in electrical engineering (1975). He received his MSEE degree from Stanford University in 1978. He's a member of the IEEE and a former member of the U.S. Navy Reserve. Bob is married, has a daughter, and lives in Boise. His interests include personal computing and outdoor activities like camping, fishing, hiking, and canoeing.

12

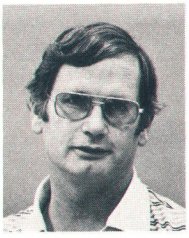
Peter M. Galen



Peter Galen received a BS degree in biomedical engineering from Case Western Reserve University in 1976 and an MSEE from the University of Washington in 1978. Since joining HP's Disc Memory Division in 1978, he has worked on error detection and correction for disc drives, and is now a project manager for a disc controller. A native of New York City, Peter is married, has one child, and lives on a five-acre farm in Eagle, Idaho, where he raises steers and chickens. He's a member of Amnesty International and his interests include reading, politics, camping, and travel.

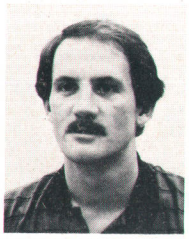
14

Roger V. Wilcox



Roger Wilcox returned to his native Idaho to join HP in Boise in 1979. Born in Montpelier, he studied electrical engineering at Brigham Young University, graduating with a BS degree in 1969, then spent ten years in California working on radar signal processing for the U.S. Navy. His work there resulted in one patent. With HP's Disc Memory Division in Boise, he designed the track follower and served as project leader for the 7933/35 Disc Drives. Roger is married, has six children, lives in Boise, and teaches digital electronics at Boise State University. His other interests include woodworking, backpacking, and coaching youth sports.

Eric W. Johnson

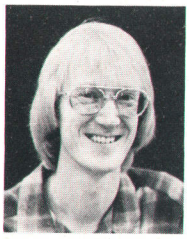


Eric Johnson graduated from Washington State University with a BSEE degree in 1979 and joined HP the same year as a development engineer. He contributed to the design of the 7933/35 Disc Drive servo electronics and servo writer. In 1982 he received his MSEE degree from Stanford University. A native of Kelso, Washington, Eric lives in Boise, Idaho, and enjoys sports, especially softball, basketball, and skiing.

R. Frank Bell

Frank Bell joined HP in 1975 with experience in the design of instrumentation, security systems, and sonar. At HP's Disc Memory Division, he's done analog and interface circuit design for disc drives, including the actuator driver amplifier for the 7933/35, and he acts as a tutor for videotaped circuit theory courses from Stanford University and the University of Idaho. His work has resulted in one paper and two patents on intrusion detection systems and one pending patent on a dual-mode amplifier. Born in Nashville, Tennessee, Frank received his BSEE degree from Lehigh University in 1962 and his MSEE degree from the University of Santa Clara in 1973. He's married, has two teenagers, lives in Meridian, Idaho, and enjoys skiing, camping, sailing, and home computers.

R. Keith Whitaker



Keith Whitaker attended Weber State College and the University of Utah, graduating from the latter with a BSEE degree in 1977 and an MSEE in 1978. With HP since 1978, he has contributed to the design of the 7925 Disc Drive servo system and the 7933/35 Disc Drive position servo and seek control, and has coauthored an IEEE correspondence on phase-locked loops. Keith's hobbies are skiing, photography, backpacking, and scuba diving. He was born in Portland, Oregon, grew up in Ogden, Utah, and now lives in Boise, Idaho.

20

James H. Smith

Jim Smith was mechanical design project manager for the 7933/35 Disc Drives. With HP since 1977, he's a graduate of California State University at San Jose (BSME 1959) and the University of Santa Clara (MBA 1974). His work has resulted in a pending patent on the 7933/35 media module. Jim was born in Eureka, California. He's married, has two children, and is a founder and board member of the Idaho Youth Soccer Association. He also serves as a soccer coach and as a board member of the Boise City Parks and Recreation Department. He enjoys hunting and fishing.

23

Stephen A. Edwards

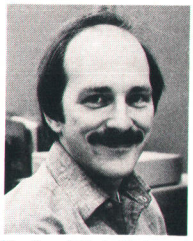
Steve Edwards was responsible for the 7933/35 Disc Drive actuator design. A 1978 BSME graduate of the University of Texas at Austin, he's been with HP since 1979 as a development engineer. Steve's interests include horses (he and his wife—another HP engineer—have an Arabian and a quarter horse), outdoor sports, sailing, and "warm sunny vacations." Originally from Tyler, Texas, he now lives in Boise, Idaho.

29

Timothy C. Mackey

Tim Mackey graduated from Portland State University with degrees in business adminstration and computer science in 1974. He was a product manager for communications related products at HP's division in Corvallis, Oregon before leaving the company recently. Tim is married and now lives in Upton, Massachusetts. His interests include music and motor sports (cars, boats, motorcycles, airplanes).

Loren Koehler



Loren M. Koehler joined HP in 1979 and has worked on several different I/O products for HP Series 80 Computers, including the Serial and BCD I/O Modules. He was the project leader for the HP 82967A Speech Synthesis Module. A native Oregonian, Loren attended Portland State University, receiving a BSEE degree in 1979. He is married, has three children, and lives in Corvallis, Oregon. His outside interests include woodworking, skiing, music, and involvement in Full Gospel Business Men's Fellowship International.

34

Jeffrey R. Murphy

Jeff Murphy joined HP in 1979 after receiving a BSEE degree from Cornell University. He worked on the 12060A ADC for the A-Series Computers, the 27201A Speech Module, and the 27203A Speech Library. He now is working on broadband local area networks. Born in Williamsport, Pennsylvania, Jeff now lives in Rocklin, California. He is a member of the IEEE and enjoys bicycling and skiing when not practicing classical piano (he has taken lessons for 12 years).

Elizabeth R. Hueftle

Born in Kearney, Nebraska, Beth Hueftle studied mathematics at the University of Nevada at Reno (BS 1976) and computer science at California State University at Chico (MS 1978). She then joined HP and has contributed to the FORTRAN/77 compiler for the HP 1000, the design of process control systems, and the 27203A Speech Library. Beth currently is working on local area networks for personal computers. She is a member of the Society for Women Engineers and lives in Loomis, California. Outside of work, she has many interests—skiing, soccer, bicycling, rock climbing, photography, and collecting antiques.

Speech Output for HP Series 80 Personal Computers

This module allows a computer to provide informative prompts and alarms, freeing the user from frequent attention to a display.

by Loren M. Koehler and Timothy C. Mackey

THE HP 82967A SPEECH SYNTHESIS MODULE (Fig. 1) for Hewlett-Packard's Series 80 Personal Computers allows these machines to output audible information to prompt operators, sound alarms, indicate error conditions, or request service. The 82967A is accompanied by enough vocabulary and software to provide a variety of tools for using speech. A similar module, the HP 27201A, is available for computers with an RS-232-C/V.24 interface. A powerful supporting software package makes this module easy to use on the HP 1000 and HP 3000 Computer Systems (see article on page 34).

Computer-generated speech output can improve the efficiency of some operations, and in many cases, provide new capabilities. Some possible applications are:

- Test/measurement. A typical test station might use a voltmeter controlled by a Series 80 Computer via the HP-IB (IEEE 488). With speech output, the computer can tell a technician when to move a test probe to a new circuit node without the technician's having to divert attention away from the circuit board to obtain the next test instruction. The computer also can be programmed to warn the technician verbally when a high voltage is encountered.
- Data entry. An accounting clerk who enters data into ledgers all day long gets very good at using the 10-key numeric pad on the computer's keyboard. This user typically does not look at the screen during every data entry. Accuracy can be increased greatly by either having audible feedback of entered data or verbal warnings of "out-of-balance" account situations to such users, thus getting their attention and causing them to check their entries.
- Monitoring/process control. With speech output, a computer monitoring peripheral status or controlling a process can provide information with an audible alarm, eliminating the time required for an operator to locate the appropriate display and read the alarm message. For

Fig. 1. *The HP Model 82967A Speech Synthesis Module provides plug-in speech output capability for HP Series 80 Computers. The speech output can be supplied to headphones, an external audio system, or the speaker contained in the video monitors for HP-86 Computers. Included with the 82967A is a disc containing a 1500-word speech vocabulary and binary routines to simplify the application of speech output to programs.*

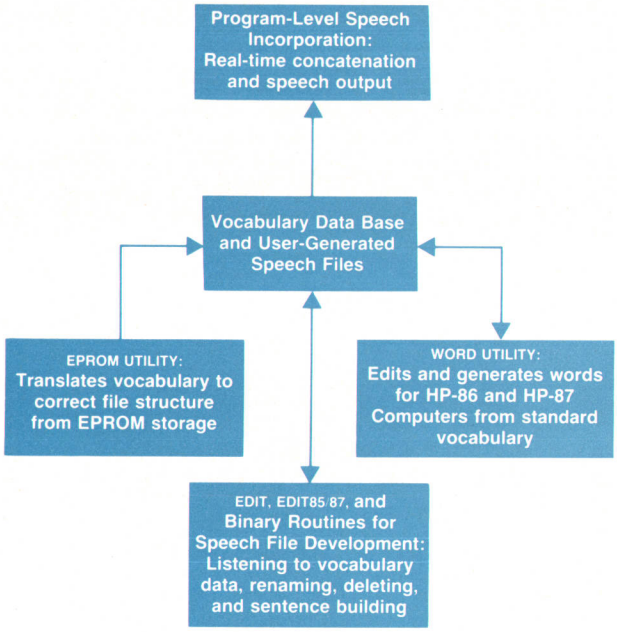


Fig. 2. *Outline of software package for implementing speech output in Series 80 Computers. All portions are included with the 82967A except the EPROM UTILITY and the WORD UTILITY, which are available from the HP User's Library.*

example, a computer can announce, "Printer 6 is out of paper," or "Furnace temperature is too high," to indicate not only the nature of the problem, but also the device affected.

Speech Output Technology

A computer system can use several techniques to generate speech output.[1] Some of the most common methods are converting text to speech, reconstructing speech using digitized samples or recordings of actual speech, and synthesizing speech using linear predictive coding (LPC, see box on page 32). The LPC technique is used by the 82967A and its companion product, the 27201A Speech Output Module.

Text-to-Speech Conversion. Translating text in ASCII characters into verbal output is done by examining the text in sequence and trying to figure out how to pronounce the syllables observed. Because most languages, in particular English, contain words or combinations of letters that look the same but sound different, or sound the same but look different, the set of translation rules can become quite complicated. For example, consider the italicized words in the following sentences:

Your assignment is to *read* the chapter about *reed* boats.
After you have *read* the instruction, pick up the *red* envelope.
After the archers picked up their *bows*, they faced the spectators and *bowed*.

To handle these problems, a large exception dictionary and an intricate set of context-dependent rules are required.

The advantage of translating text is its flexibility. However, the quality of the speech is generally very poor. The listener must pay close attention, because the speech output is usually flat and unemotional, and sounds mechanical. In some cases, the listener must have some knowledge of the general context of the spoken message to understand it correctly. Some of these deficiencies can be corrected, but then the benefit of flexibility is lost.

Recorded Speech. Actual speech can be digitized and stored in the computer's memory for reproduction when desired. The benefit is natural-sounding speech, but the memory requirements are large even for brief speech outputs, typically 12K bytes for speech lasting one second. An alternative approach is to use the computer to search for a desired phrase recorded on a tape recorder and direct the recorder to play back the phrase. The disadvantage of this method is the long access time and the large number of phrases required to verbalize a full range of numbers or conditions.

Synthesized Speech. Speech synthesis can generate high-quality speech using a mathematical model of the human vocal tract and digital data based on the recording of spoken words and sounds. Synthesized speech output has some clear advantages over more conventional methods of sound reproduction such as the tape recorder. Synthesized speech has solid-state reliability, allows real-time random selection and concatenation of words, phrases, and sounds, and has significantly lower data requirements than digitized speech. The LPC technique used in the 82967A and 27201A Modules is based on the assumption that the sound a person makes at one instant is a continuation of the sound made in the previous instant. LPC removes natural redundancies in speech and reduces the number of bits required to reproduce one second of speech by as much as 98.5% compared to purely digitized speech. Hence, the benefit of the LPC technique is that it greatly reduces memory requirements

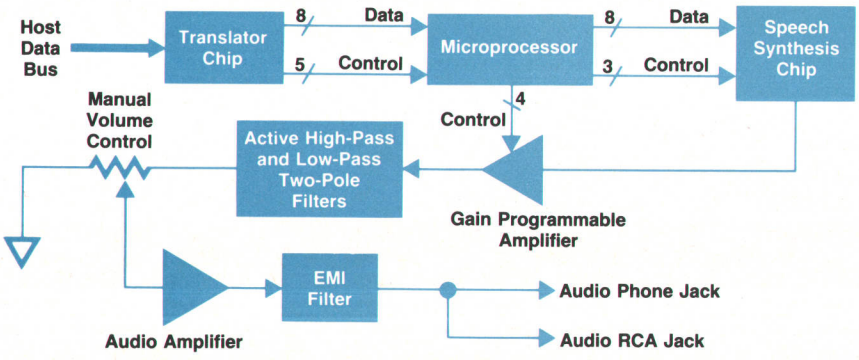


Fig. 3. *Block diagram of the 82967A Speech Synthesis Module's hardware system.*

while retaining a high-quality speech output.

System Overview

The design objectives for the 82967A were to provide speech output with as high a degree of quality as possible, and to provide everything needed to allow its use as a viable computing tool. The 82967A hardware is closely knit with software and vocabulary (refer to Fig. 2).

The EDIT program generates speech files. A speech file is a subset of the vocabulary data base, with words, phrases, and sounds put into user-defined combinations and renamed to fit the given application. A speech data file, as with any data file, requires a certain amount of time to load from mass storage into the computer and requires a certain amount of memory space. Speech file construction allows the user to make tradeoffs between vocabulary residing in computer memory and the amount of load time from mass storage.

The EPROM UTILITY program converts additional vocabulary words, usually residing on an EPROM (electrically programmable read-only memory), to the correct data file structure. EPROMs programmed with additional vocabularies selected by the user are available from the vendor for the speech synthesis integrated circuit used in the 82967A Speech Module. The EPROM UTILITY program is available from the HP User's Library service. It works in conjunction with the HP 82929A Programmable ROM Module for Series 80 Computers. The 82929A holds up to two 64K-bit EPROMs and fits directly into the Series 80 I/O backplane.

The WORD UTILITY program can be used to develop new words from the existing vocabulary. New words are constructed by deriving the required syllables and phonetic sounds from existing words in the vocabulary data base and patching them together. The WORD UTILITY program also allows a word, phrase, or sound to be altered by varying the digital parameters that define it (refer to the box on page 32).

Hardware Design

The 82967A hardware system (Fig. 3) is centered around an 8049 microprocessor and a TMS5220 speech synthesis chip. In addition, the module contains a programmable amplifier for attenuation of the volume under software control, active low-pass and high-pass filters, and a final audio amplifier.

The 8049 microprocessor services the speech chip, controls the gain programmable amplifier, and handles the I/O protocol between the 82967A and the host computer via a translator chip.

The IMB5 translator chip provides the interface between the module and the bus to the host computer. To service the speech chip, the microprocessor provides an internal 90-byte first-in, first-out (FIFO) buffer between the host computer and the speech synthesis chip. The computer invokes verbal execution of an utterance (a sound, a word, or a group of sounds and/or words) by passing to the host computer a command and a 16-bit value representing the total number of bytes in the speech file containing the data for the utterance. The computer then transmits the speech data for the utterance to the microprocessor's FIFO buffer. After it fills its FIFO buffer or receives all of the bytes in the utterance (if less then 90 bytes), the microprocessor sends a speak command and 16 bytes of speech data to the TMS5220 speech chip. The microprocessor then passes additional eight-byte blocks of speech data on a request basis to the speech chip until the utterance is finished. As data is moved from the microprocessor's FIFO buffer to the speech chip, more data is sent from the host computer until all of the requested speech data has been transferred from the computer to the speech module.

The data transfer used between the Series 80 Computer and the 82967A card is known as an OUTPUT data transfer. The OUTPUT data transfer was selected as the only means of transferring speech data from the computer to the speech module because this type of data transfer cannot be suspended by another device. Therefore, it can guarantee that all speech data in an utterance arrives at the 82967A without interruption. The OUTPUT data transfer approach reduces the memory requirements on the speech module, which keeps the hardware cost down. After completion of the OUTPUT speech data transfer, the other I/O operations of the computer can resume or begin while the speech module is still speaking, using the speech data remaining in the FIFO buffer.

The TMS5220 speech chip produces synthesized speech by taking the encoded LPC parameters, decoding the information, and placing it in an internal digital lattice filter. The ten most-significant bits of the filter output are supplied to an on-chip digital-to-analog converter (DAC) every 125 microseconds, yielding an analog signal representing the appropriate utterance at the output of the speech chip.

Programmable volume control is accomplished by using a gain programmable amplifier (GPA). This amplifier is actually a digital-to-analog converter. The analog signal from the speech chip is fed into the GPA's reference node and attenuated to one of sixteen possible levels, depending on the setting of the four binary inputs to the GPA by the microprocessor. The GPA setting is controlled from the keyboard or program by using the binary keyword SVOL, or alternately, the CONTROL command from the Series 80 I/O ROM. For example, the command SVOL 10,12 sets the output to level 12. The value 10 is the factory setting for the 82967A's select code (hardware address). Writing a value of 15 to the GPA sets the maximum volume level; a value of 1 yields a minimum volume level. Writing a value of 0 to the GPA turns off the output. At power-on or after a system reset, the speech module sets a default value of 15 for the GPA.

Two-pole high-pass and low-pass active filters are used to modify the analog signal before final amplification on the card. The −3-dB point of the high-pass filter is 100 Hz, and the −3-dB point of the low-pass filter is 3 kHz. These points were picked based on the bandwidth of human speech and background system noise. An operation called deemphasis, which attenuates the high frequencies of the synthesized speech, is performed directly by the TMS5220 speech synthesis chip. After the analog speech signal is filtered, it passes through a manual volume control and into an audio amplifier. The output of the audio amplifier is 0.22 watt into an eight-ohm load, which is more than

Linear Predictive Coding

Linear predictive coding (LPC) is a speech analysis/synthesis technique which reduces the amount of stored information required to reconstruct an utterance. If actual speech is digitized for storage, 96,000 bits of memory are required to reproduce one second of speech. By storing only the data required by LPC to synthesize the same one second of speech with high quality, the memory required is reduced to less than 1200 bits.

LPC analysis begins with a recording of the actual sound, word, or combination of words and/or sounds. This recording is then converted into digital data by first sampling the recorded waveform at a fixed rate. This data is then compressed to extract amplitude, source, and filter information to reconstruct the utterance based on a mathematical model of the human vocal tract. The amplitude, or energy, is simply the loudness of the utterance. The source information specifies whether or not the vocal cords are vibrating, and if so, at what frequency (pitch). The filter parameters describe the relative positioning of the tongue, lips, and teeth in the vocal tract model.

Speech Analysis

The filter parameters are derived based on the assumption that whatever is spoken now is to a large degree a linear continuation of what was spoken an instant earlier. That is, the filter parameters K_i are determined by minimizing the mean square error between the actual value v_n of a sample of the utterance and the value v_{ne} estimated from a weighted sum of a number of previous actual values. The relation between v_{ne}, the K_i coefficients, and the previous sample values is given by

$$v_{ne} = \sum_{i=1}^{j} K_i v_{n-i} \qquad (1)$$

where the analysis period, i = 1 to j, is a function of the sampling rate and the value chosen for j. For the LPC technique used in the 82967A and 27201A Modules, the sample rate is 8 kHz and j = 10.

Matrix algebra is used to solve for the K_i coefficients for each analysis period. The amplitude value is derived from the rms value of the speech waveform. The pitch value is derived from the periodic variation of the sample values.

Phonetics

Before discussing how the LPC parameters are combined to form a frame of speech data and how that frame is used to synthesize speech, the following information about basic phonetics will be useful.

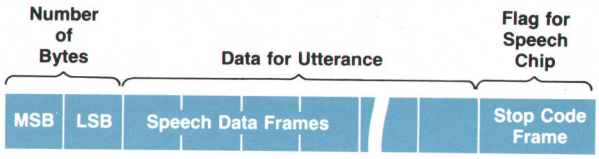


Fig. 2. *An utterance is coded as a sequence of speech frames with a header that indicates the total number of bytes in the sequence and a final stop speech frame.*

Speech sounds can be divided roughly into two categories: unvoiced and voiced. Unvoiced sounds are random noise generated by a constriction somewhere in the vocal tract. Examples of unvoiced sounds are the letters f and s. Voiced sounds involve the vocal cords vibrating at a certain pitch. All vowels are voiced sounds as well as the letters l, m, n, r, w, and y. Some letters such as v and z are combinations of unvoiced and voiced sounds.

Some letters have other characteristics that must be considered for speech synthesis. All of the consonant letters except h involve a constriction in the vocal tract that is relaxed rather quickly. For the stop letters b, d, g, k, p, and t, the constriction completely blocks the flow of air through the vocal tract. Hence, stop letters are characterized by a short silence lasting 10 to 100 milliseconds.

All letters characterized by either a partial or a complete vocal tract constriction are also characterized by rapid spectral (frequency) changes before and after the occurrence of the constriction. This is caused by the rapid changes in the structure of the vocal tract as the constriction is formed and then relaxed.

Speech Frames

The LPC parameters for each analysis period are stored with a repeat bit in speech frames of different lengths as shown in Fig. 1. Each word or sound is stored in the vocabulary as a sequence of these speech frames headed by a digital value that indicates the number of bytes in the sequence (Fig. 2).

The speech frames can be classified into five basic types: voiced, unvoiced, repeat, zero energy, and a stop code frame. The 50-bit voiced (V) frames specify energy, pitch, and all ten filter coefficients. The 29-bit unvoiced (U) frames specify energy, zero pitch, and only the first four filter coefficients. The 11-bit repeat (R) frames specify only energy and pitch, and have their

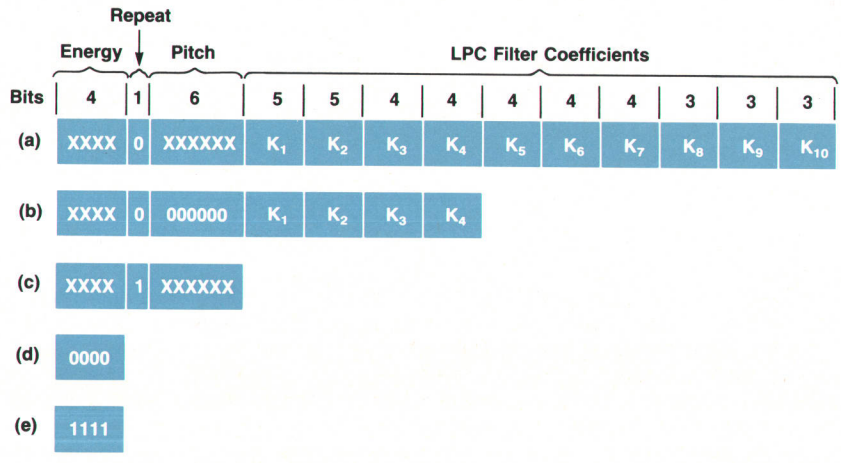


Fig. 1. *LPC speech synthesis parameters are stored into one of five speech frame formats: (a) voiced, (b) unvoiced, (c) repeat, (d) zero energy, and (e) stop frames (see text).*

Frame		Energy	Repeat	Pitch	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_{10}
V	-	12	0	43	22	11	6	3	10	11	10	6	3	3
V	-	13	0	39	25	13	9	2	4	11	8	6	4	4
V	-	13	0	38	25	13	9	2	3	11	9	6	4	4
V	-	13	0	38	24	16	8	1	4	11	9	6	3	3
V	-	13	0	41	25	17	8	0	3	11	9	6	3	4
V	-	11	0	45	24	14	6	0	5	13	10	6	3	2
V	-	7	0	49	21	12	2	3	12	13	12	6	2	3
V	-	1	0	1	25	8	10	5	5	4	2	5	3	0
Z	-	0	-	-	-	-	-	-	-	-	-	-	-	-
Z	-	0	-	-	-	-	-	-	-	-	-	-	-	-
Z	-	0	-	-	-	-	-	-	-	-	-	-	-	-
Z	-	0	-	-	-	-	-	-	-	-	-	-	-	-
U	-	1	0	0	21	6	6	4	-	-	-	-	-	-
U	-	1	0	0	24	0	6	8	-	-	-	-	-	-
U	-	7	0	0	31	15	1	2	-	-	-	-	-	-
U	-	7	0	0	27	1	3	9	-	-	-	-	-	-
U	-	7	0	0	31	10	4	5	-	-	-	-	-	-
UR	-	6	1	0	-	-	-	-	-	-	-	-	-	-
UR	-	4	1	0	-	-	-	-	-	-	-	-	-	-
S	-	15	-	-	-	-	-	-	-	-	-	-	-	-

Fig. 3. *Listing of the twenty speech frames and LPC parameter values in the sequence for the word "eights." The utterance begins at the top of the list and concludes with the stop frame at the bottom.*

repeat bit set to one to indicate to the speech synthesis IC that the filter coefficients for the previous frame are to be retained. The 4-bit zero energy (Z) frames specify zero energy and have no other values. The 4-bit stop (S) frame specifies an energy value of 15 and has no other values. The shorter length of the unvoiced, repeat, and zero energy frames allow the data rate and storage requirements for speech synthesis to be reduced considerably.

In each frame, the energy value can range from 0 to 15, and the pitch value can range from 0 to 63. The repeat bit is either zero or one. Note in Fig. 1 that the number of bits allocated for each filter coefficient varies depending on the influence of that parameter on speech quality.

Fig. 3 shows a listing of the values in each of the 20 frames in the 74-byte sequence of values required to synthesize the word "eights." The Z frames describe the momentary silence before the "t" and the U and UR frames describe the "ts" sound.

Speech Synthesis

Fig. 4 shows a block diagram of the basic speech synthesis process. The excitation sources are a variable-frequency pulse generator for voiced frames and a random noise generator for unvoiced frames. The amplitude of the selected source is modulated by the desired amplitude or energy level and then is applied to the recursive filter, which is programmed by the filter coefficients for the speech frame. The transmission function of the filter is

$$T(z) = 1/(1-H(z)) \quad (2)$$

where

$$H(z) = \sum_{i=1}^{10} K_i z^{-i} \quad (3)$$

Therefore, if e_n is the nth sample of the selected excitation, the synthesized output sample v_n' is

$$v_n' = e_n + \sum_{i=1}^{10} K_i v_{n-i}' \quad (4)$$

This synthesized output is then sent to additional circuitry to be shaped by high-pass and low-pass filters and amplified for output as audible speech.

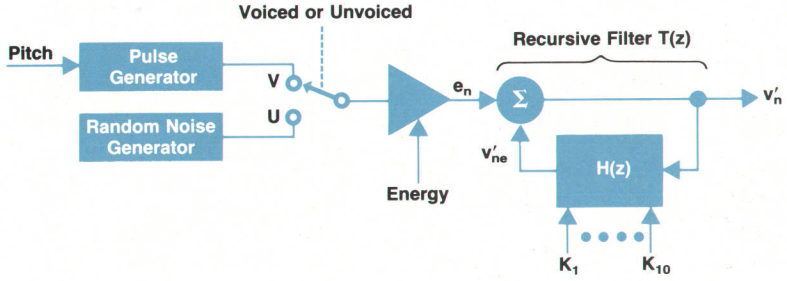


Fig. 4. *Block diagram of LPC speech synthesis process.*

adequate for most needs.

The 82967A comes with a phono jack and an RCA-type speaker jack connected in parallel. These two jacks enable the user to connect the speech module to an HP-86 Computer's video monitor (whose audio input also uses an RCA connector), to stereo headphones, or to whatever else meets the user's needs, without the need for an adapter. The manual volume control is designed so that by turning it all the way down, the correct volume setting is obtained for use with headphones.

High-Level Speech Incorporation

Incorporation of speech output in an HP BASIC program is a simple task. Special keywords provided by the "speak" binary program included with the 82967A's software allow quick implementation of speech output in programs. The software also contains easy-to-use routines for generating speech files that contain the desired utterances.

Retrieving Speech Data. To retrieve a speech file from mass storage, the procedure is the same as that required to read any data file into memory from mass storage with one exception—the keyword is different. For example,

```
10 ASSIGN #1 TO"HP" !Open the file
20 DLOAD;A$,P$ !A$ = name and file information, P$ = speech data
30 ASSIGN #1 * !Close the file
```

The above three HP BASIC statements load a speech file into computer memory. The actual speech data resides in the variable P$. Information related to the speech data such

Speech Output for HP 1000 and HP 3000 Computer Systems

by Elizabeth R. Hueftle and Jeffrey R. Murphy

The efficiency of some users of a large computer system can be notably improved by the addition of local speech output, which can be used to request service for peripherals such as plotters and printers, prompt operators for inputs, announce the occurrence of input or output error conditions, and sound alarms. The HP 27201A Speech Output Module in conjunction with the appropriate HP 27203A or HP 27205A Speech Library provides this capability for HP 1000 and HP 3000 Computer System users. The 27201A can also be used with other RS-232-C/V.24 computer systems, but the user will have to develop the speech files directly without the assistance of the Speech Library software.

The 27201A is a microprocessor-based peripheral whose appearance and internal speech synthesis circuitry are very similar to the 82967A Speech Synthesis Module for HP Series 80 Computers discussed in the accompanying article. The differences are the interfacing method and the addition of limited internal speech data storage.

Interface

The 27201A and its interconnecting cable implement a standard three-wire (for transmit data, receive data, and ground signals) RS-232-C/V.24 data communications interface. Because no other signals are provided, the module does not support connection to long-haul modems. However, short-haul modems using three-wire data communications can be used.

This interface scheme allows the 27201A to be connected directly to the host computer or to be inserted in series with any RS-232-C/V.24 peripheral such as a terminal, printer, or plotter. The serial configuration (Fig. 1) eliminates the need for another port when adding speech capability and allows the 27201A to be connected easily to a workstation for interactive speech output program development.

To coexist with an in-line peripheral on the same data communications line, the 27201A operates in an eavesdrop mode, monitoring all data on the line while passing it through to the in-line peripheral. The 27201A's speech output functions are invoked by a special escape sequence (Esc&yS<command>Esc&yU), which triggers command recognition by the speech synthesis circuitry. Table I lists the available commands. When the module must communicate with the host computer, it intercepts the peripheral's CTS (clear-to-send) line to suspend data flow from the peripheral during this time. The 27201A implements the XON/XOFF handshake, which means that data flow control resides with the 27201A, even when the peripheral initiates the handshake.

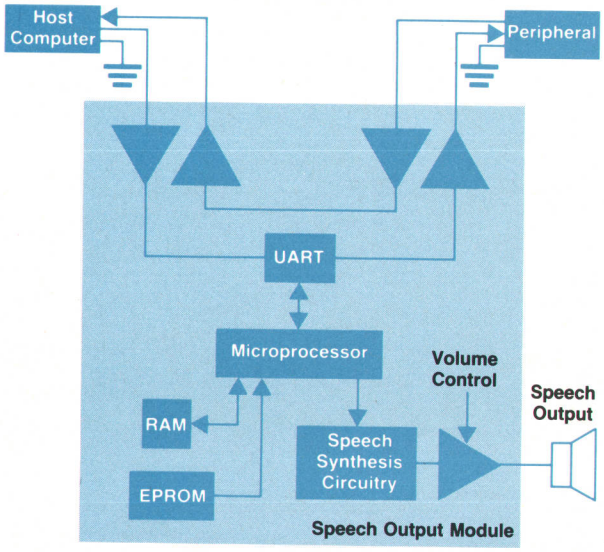


Fig. 1. *Serial configuration for HP 27201A Speech Output Module. The module is inserted in the RS-232-C/V.24 line connecting the peripheral to the host computer. During nonspeech operation, the module passes all signals between the peripheral and the host computer, but monitors the signals to detect the special escape command sequence that addresses the module.*

Table I

HP 27201A Speech Output Module Commands

Command	Action
CLEar	Deletes a group* of word data in the 27201A
DOWnload	Transfers word data from host computer to a group in the 27201A
PITch	Varies pitch of speech
RESet	Clears 27201A and runs self-test
SPEak	Causes 27201A to speak using word data in a group
STAtus	Identifies buffer, register, and memory status of 27201A and conveys error messages
TRAnsparent	Passes all data through 27201A without recognizing commands
UPLoad	Transfers word data from a group in the 27201A to the host computer

*A group in the 27201A represents one RAM or EPROM component.

Internal Memory

The Speech Output Module can store the data for synthesizing up to 100 words (50 seconds of speech) by downloading the data from the speech library in the host computer to the module's internal RAM, or can store the data for up to 200 words (100 seconds of speech) using EPROMs installed by the user. In the latter case, users can store the vocabulary most appropriate for their application and eliminate the time required to obtain speech data from the host system.

Speech Library Software

Fig. 2 illustrates the speech software package used by HP 1000 and HP 3000 Computers to generate speech output with the 27201A Speech Output Module. This package provides several useful features. First, it provides a vocabulary of over 1700 commonly used words and sounds in a form ready to be downloaded to a module and spoken. Second, the package provides easy access to these words and sounds. Third, the speech software provides a data base structure for the speech data and a program for managing that data base. Fourth, the software makes it easy for a user to include speech output in the user's application programs.

The VX (voice exerciser) program lets the user control the

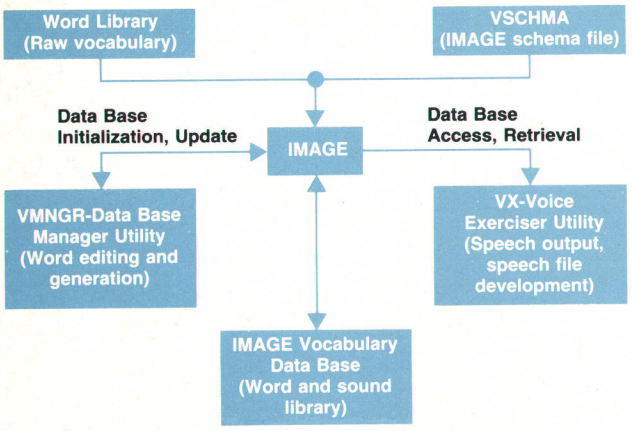


Fig. 2. *Outline of the Speech Library software package for HP 1000 and HP 3000 Computers.*

27201A without having to code module commands and speech data in a program. The interactive access provided by VX lets the user clear the module, download selected speech data to it, and cause it to speak by using only a few keystrokes. Thus, a user can listen to different outputs quickly to choose the best sounding one for the user's application.

The IMAGE schema file VSCHMA contains instructions used during the installation of the speech software to construct the data base, which is based on HP's IMAGE data base management system.

The VMNGR program manages the speech data base. It allows the user to:

- Install speech data into the data base
- Make other data bases out of subsets or supersets of the standard speech data set. For example, if an application requires a data base that takes up only a small amount of storage, VMNGR can be used to create a small data base containing only the words needed.
- Make up new words by combining parts of existing words. Instructions and hints about how to do this are included in the user manual.
- Add custom words to the data base. If a user needs words that are not in the standard vocabulary and cannot make them by modifying existing words, the user can purchase additional words from the vendor of the speech synthesis circuit used in the 27201A. These custom words are provided on an EPROM that the user can install in the speech module. VMNGR can then be used to access those words for installation in the central data base.
- Create a burn file for an EPROM. VMNGR can be used to format selected speech data into the proper file for programming an EPROM. The user can then program EPROMs for insertion in the memory sockets of the 27201A, which provides nonvolatile storage of speech data for fast speech output.

Acknowledgments

Many people contributed to the development of the speech products. Two of the individuals we would especially like to thank are Mike Thompson who developed the software for the HP 3000, and Dave Kenyon, our project manager, who was a source of continual encouragement and incredible enthusiasm. We would also like to thank the marketing team, especially Duncan Campbell and Tom Nakamura, for their contributions.

as the names of the words in the speech file and their location are contained in the variable A$. The binary keyword DLOAD provided by the binary speech software is used instead of the normal LOAD command because the vocabulary data base and user speech files have unique data structures.

Real-Time Concatenation. Combining words, phrases, and sounds residing in the host computer memory is an easy task for the applications program. One useful example is the generation of a number to be verbalized by a speech module in an HP-86 or HP-87 Computer. The statements

```
60 LET X$="20"
70 NUMBER$=PAR$("100",A$,P$)&PAR$(X$,A$,P$)
```

put the speech parameters for the word 120 in the string variable NUMBER$. This procedure enables any number between zero and 999 million to be constructed and verbalized in real time using the existing English vocabulary. To generate the given range of numbers, the speech file must contain speech data for the numbers zero through 20, the multiples of ten (30, 40, 50, etc.) through 100, and the numbers "thousand" and "million." The keyword PAR$( ) is part of the "speak" binary routine.

Speech Generation. To have a program speak an utterance is the most straightforward task of all. For example, to speak the number 120 generated above, requires only the program statement:

```
80 SPEAK 10; NUMBER$ !Where 10 is the 82967A's select code
```

A phrase of speech stored in a user's speech file can be as brief as one sound or word, or as long as thirty minutes of continuous speech. The phrases in a speech file can be named using a total of up to 195 characters on the HP-86/87 Computers and up to 95 characters on the HP-83/85 Computers by using the EDIT program. For example, in the statement

```
100 SPEAK 10; PAR$("DEMO",A$,P$)
```

DEMO is the label for the phrase "GOOD DAY PAUSE10 THIS IS THE HEWLETT PACKARD 8 2 9 6 7 A SPEECH SYNTHESIS MODULE." This phrase could have been constructed using the EDIT or the WORD UTILITY programs, or all the separate words could have been placed in separate phrases in a speech file and concatenated in real time to produce the same result.

Finding Words. The first step is to look up the desired words on the vocabulary disc that comes with the 82967A. This is done by running the EDIT program and pressing the VIEW softkey. The Series 80 host computer then asks the user which dictionary to look in. If the word wanted is "hello," the user enters H, the first letter in hello, because the standard vocabulary is arranged into 26 alphabetic dictionaries (one for each letter in the alphabet). A complete listing of the standard vocabulary on a removable placard is also included in the manual for the 82967A. User dictionaries are also allowed, and any name of up to eight characters can be assigned to a user-created dictionary.

Listening to Words or Phases. Once the user looks through

the displayed words to confirm the existence of the desired words in the vocabulary, the user can listen to different sentence combinations by pressing the LISTEN softkey. The host computer then prompts the user by asking Word/Phrase to Speak? The user then enters each desired word separated by underscore marks. For example,

Word/Phrase to Speak?
HELLO_I_AM_THE_HEWLETT_PACKARD_EIGHTY_SIX_
PERSONAL_COMPUTER

The words are read from the vocabulary disc and the 82967A promptly "speaks" them. Next, a set of three softkeys (AGAIN, INCLUDE, CONTINUE) is displayed by the computer. The AGAIN softkey causes the phrase to be spoken again. The INCLUDE softkey causes the entire phrase to be brought into the dictionary residing in the host computer's memory. Pressing the CONTINUE softkey causes the original set of softkeys (VIEW, LISTEN, FET, SAVE, INCLUDE, DELET, RENAME) to be redisplayed.

Saving Results. Once the desired results are in the host computer's memory, the user can save this resident dictionary on disc by simply pressing the SAVE softkey. The computer then prompts the user for a name for the dictionary file this information will be stored in. Once the phrases are created and stored on disc, the user can write a BASIC application program to use them under program control with the help of the SPEAK command provided by the "speak" binary program that comes with the 82967A.

Packaging

The packaging scheme for the 82967A is designed so that the user does not have to assemble or disassemble anything to install the unit. The package is a plug-in module (Fig. 1), which in external appearance looks similar to the Series 80 HP 82950A Modem. A volume control and two audio jacks are the only clues that this is not an 82950A. Plugging the module into any of the four I/O slots in the Series 80 Computer backplane is the only installation needed (Fig. 4). Power is drawn from the mainframe, thus eliminating the need for any external power supplies and/or cords.

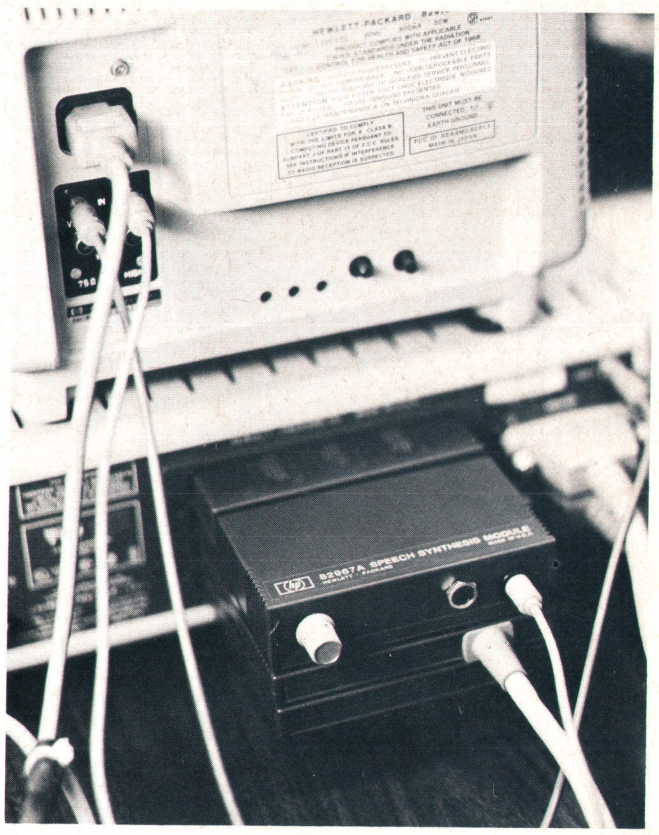


Fig. 4. *Installing the 82967A in a Series 80 Computer is a simple process. The module is plugged into one of the four I/O slots in the backplane of the computer as shown and derives its power from the mainframe.*

Acknowledgments

Homer Russell contributed to the definition of the speech system software and helped implement the 82967A for Series 80 Computers.

Reference

1. L. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.

Hewlett-Packard Company, 3000 Hanover Street, Palo Alto, California 94304

Bulk Rate
U.S. Postage
Paid
Hewlett-Packard
Company

HEWLETT-PACKARD JOURNAL

JANUARY 1984 Volume 35 • Number 1

Technical Information from the Laboratories of
Hewlett-Packard Company
Hewlett-Packard Company, 3000 Hanover Street
Palo Alto, California 94304 U.S.A.
Hewlett-Packard Central Mailing Department
Van Heuven Goedhartlaan 121
1181 KK Amstelveen, The Netherlands
Yokogawa-Hewlett-Packard Ltd., Suginami-Ku Tokyo 168 Japan
Hewlett-Packard (Canada) Ltd.
6877 Goreway Drive, Mississauga, Ontario L4V 1M8 Canada

CHANGE OF ADDRESS: To change your address or delete your name from our mailing list please send us your old address label. Send changes to Hewlett-Packard Journal, 3000 Hanover Street, Palo Alto, California 94304 U.S.A. Allow 60 days.

5953-8519