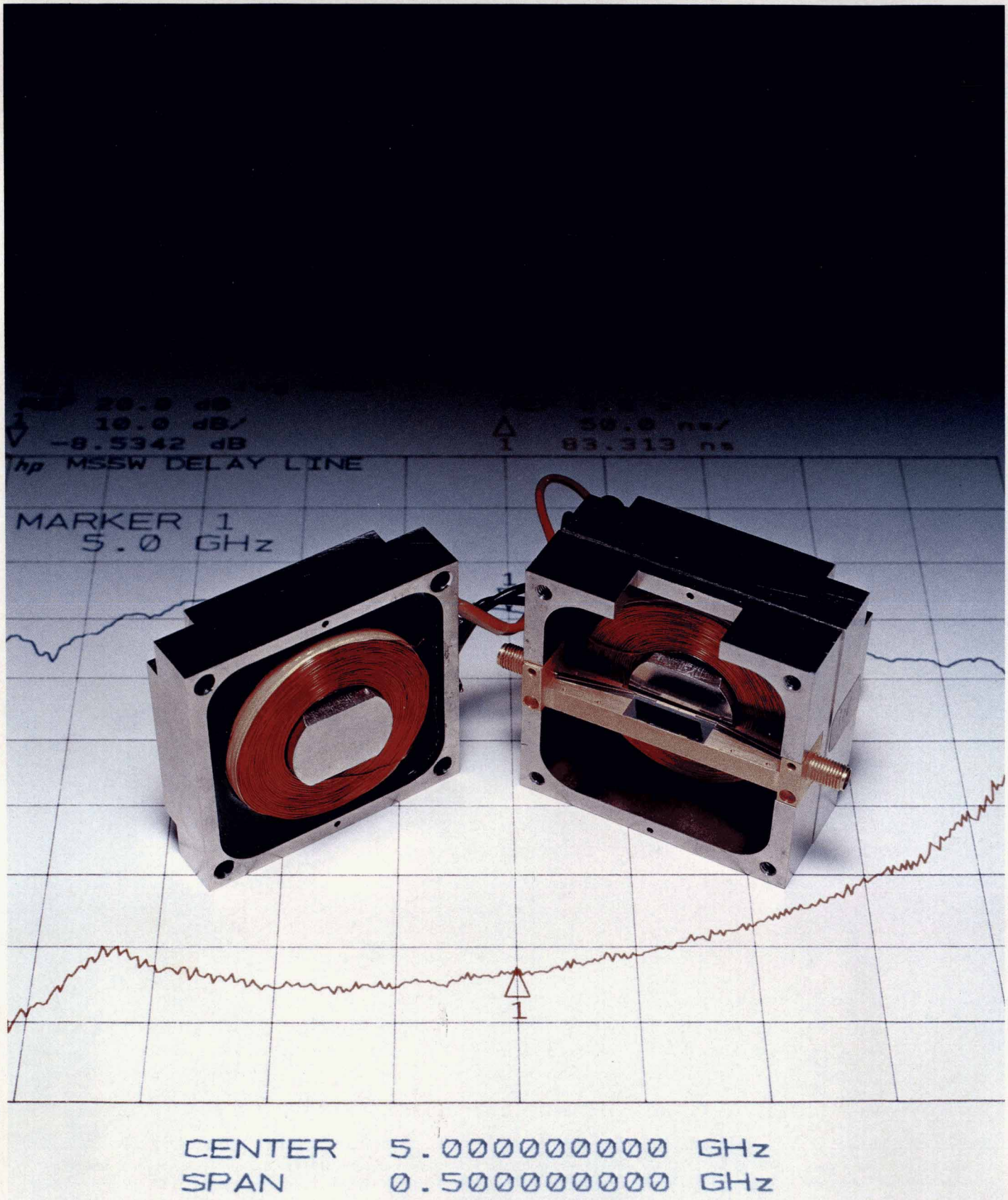
HEWLETT·PACKARD
JOURNAL

FEBRUARY 1985



HEWLETT-PACKARD JOURNAL

February 1985 Volume 36 • Number 2

Articles

4 **HP TechWriter: Illustrated Documents for Engineers,** *by Elaine C. Regelson and Roy E. Anderson* *With an HP 9000 Series 200 Computer and an appropriate HP graphics printer, engineers can produce and print many documents themselves.*

8 **HP TechWriter Security**

40 **Authors**

Research Reports

10 **Magnetostatic-Wave Devices for Microwave Signal Processing,** *by Waguih S. Ishak and Kok-Wai Chang* *They're useful for tunable delay lines, couplers, signal-to-noise enhancers, filters, oscillators, and frequency multipliers.*

14 **Magnetic Resonance and YIG-Sphere Devices**

16 **Spin Waves and Magnetostatic Waves**

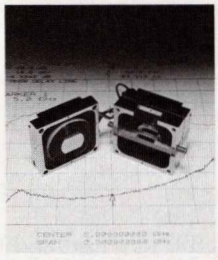
21 **Disc Caching in the System Processing Units of the HP 3000 Family of Computers,** *by John R. Busch and Alan J. Kondoff* *Excess main memory and processor capacity are used to eliminate disc access delays and improve processor utilization.*

21 **Glossary**

23 **Disc Cache Performance Tools**

25 **The MPE-IV Kernel**

In this Issue



A basic computer system has a central processor, some main memory, and some mass memory. The mass memory is slower than main memory but has much more capacity. Recent trends are for processors to get faster, and for semiconductor main memory to get not only faster but also cheaper, so computer systems have more of it. Disc memory mass storage systems have grown in capacity, but disc access times haven't changed much. The result is that in many systems, fast processors and main memories are underemployed, slowed down by the disc drives they have to work with. On page 21, two HP computer scientists report on their research into this problem, the various possibilities for alleviating it, and the solution they chose for the HP 3000 family of computers. Their answer is disc caching—temporarily storing in excess main memory the disc files that are most likely to be needed soon. This way the processor needs to access the slow discs much less frequently. It's the same principle that has led to the use of very fast semiconductor cache memories between the processors and the main memories in many high-performance computer systems. And it does work—an HP 3000 with disc caching completed a benchmark program faster than another computer that had a processor five times faster but no disc caching.

On our cover this month is a photograph of a magnetostatic-wave delay-line filter. Magnetostatic-wave devices are being studied by HP's central research laboratories for applications in delay lines, couplers, signal-to-noise enhancers, filters, and oscillators for the microwave frequency range, roughly 0.5 to 26.5 gigahertz. These devices depend on magnetostatic waves propagating in thin films of ferrimagnetic materials. Although they've been studied since the 1950s, they didn't begin to show attractive performance until better methods for growing high-quality thin films were discovered in the 1970s. On page 10, two HP Laboratories scientists review the basic principles of these devices and report on recent results indicating that magnetostatic wave devices may soon be ready for large-scale commercial application.

HP 9000 Series 200 Computers are widely used as personal workstations for engineers. The article on page 4 describes a document editing software package for these computers that makes it much easier for engineers to produce such documents as engineering notes, product change descriptions, and design specifications. HP TechWriter is compatible with the Pascal 3.0 operating system and with a wide range of printers, including HP's LaserJet printer. Text and graphics are electronically merged and appear on the computer screen as they will in the printed document.

-R.P. Dolan

What's Ahead

In March, we'll have articles on two tape-drive designs. One drive is a ¼-inch cartridge tape unit that can back up a 60-megabyte Winchester disc in 30 minutes. The other is a high-performance, low-cost, dual-density ½-inch unit for backing up systems that have 400 megabytes or more of storage. We'll also have an article on HP Maintenance Management, an equipment-maintenance software package for HP 3000 Computers.

HP TechWriter: Illustrated Documents for Engineers

This document editing software package for HP 9000 Series 200 Computers electronically merges text with pictures from many HP graphics software packages. Text and graphics appear on the screen as they will in the printed document.

by Elaine C. Regelson and Roy E. Anderson

HP TECHWRITER is a document editing software product that assists engineers in producing illustrated documents, as shown in Fig. 1. It runs on HP 9000 Series 200 Computers in the Pascal 3.0 operating system. In addition to providing illustrations, HP TechWriter offers significantly enhanced document support commands and numerous hard-copy options. It was recently enhanced to support larger pictures and more hardware, including the HP LaserJet printer.

HP TechWriter consists of three original programs (Fig. 2). The Editor is used for text entry and editing and producing listings with draft-quality pictures. The Lister produces listings of multiple files with final-quality pictures and generates tables of contents. The Picture Processor converts plot files produced by other programs into picture files suitable for illustrating documents.

Engineers, scientists, and first-level technical managers are the target users for HP TechWriter. These people need the ability to include drawings, charts, and sketches in their documents. Our analysis of these users indicates that they are willing to spend some time learning this type of product, but they insist that it be very responsive (fast), and that it be easy to use once they learn it.

There was a text editor already available on HP's Pascal workstations that had some of these qualities, but not the needed functionality. It proved possible to start the development of the HP TechWriter Editor by leveraging this editor. The Pascal workstation editor is very fast at text display and scrolling. Most engineers master it in about a day of intensive use, and experienced users find it quite easy to use, with little overhead in the commands. This made it a logical foundation for HP TechWriter.

HP TechWriter allows users to produce most internal publication-quality documents, such as engineering notes, product change descriptions, and design specifications. It also allows users to take advantage of any capabilities offered by their printers, such as math, bold, or italic fonts. The new features required major enhancements to the Pascal workstation editor.

HP TechWriter Features

The HP TechWriter Editor retains the basic qualities and speed of the standard Pascal workstation editor, and the data files of the two products are fully compatible. The HP TechWriter Editor offers an extended set of features, allowing the user to do basic editing, to set and use the global editing environments, to specify illustrations, to tailor hard copy, and to format documents. There are two kinds of command interfaces. Global commands are entered from the keyboard and take effect immediately. Document, or embedded, commands are inserted into the file, and affect the text following them.

HP TechWriter provides basic editing functions including insert, delete, change, copy, find, and replace text, produce a hard-copy listing of the document, and save, retrieve, and copy from text files. The user can scan and work through the entire text file quickly by using arrow keys, or a knob if it is available. The user can also move by screenfuls of text, or to specific marked locations in the file. In addition, HP TechWriter offers several features useful for document text preparation, such as paragraph filling and word wrap, ragged and straight right-edge justification, flexible line adjustment (center, right, anywhere), and the ability to specify and change margins easily and override global margins whenever desirable. The Editor stores the most recently deleted text to allow the user to move text or recover from an unintentional deletion.

Embedded commands let the user force new pages, specify headers and footers, mark table of contents entries, set spacing, and turn off listing of particular areas of the document. A reminder-level help facility lists all command names and options and states very briefly how they work. This is generally all the prompting that is needed to support full use of all features.

There is a global environment for each document that defines standard margins and the text entry mode. The document mode results in word wrap, while program mode preserves all entries as typed. The global environment also defines special characters including the command character, which is used to mark command lines embedded in the file. The default command character, ^, is used in all examples and illustrations in this article. In addition to the global environment, both the Editor and the Lister use a list environment, which defines the printer type and the print destination and includes parameters specifying the position of the print on a page.

The user can tailor HP TechWriter by setting default values for all parameters in either environment in a special data file called the configuration file. This file also allows the user to enter all text that appears in any hard copy (the

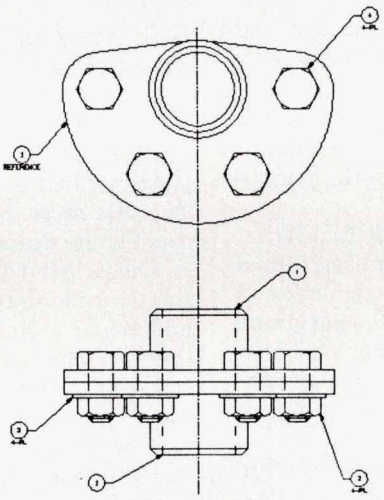
default footer, table of contents column headings, etc.). The user can also specify the format for the date (month/day/year, day-month-year, etc.) printed in headers or footers.

Real-Time Margination

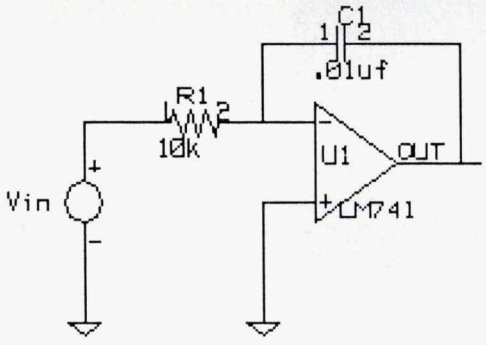
HP TechWriter recognizes paragraphs when it is in document mode, and formats them according to their margins, either when their entry is completed or by explicit command. Margins are specified in the global environment and may also be specified by local commands. The user can specify ragged or straight right margins. Document commands can also be used to turn off margining for a region

HP TechWriter is an engineering documentation product that provides Series 200 customers with a version of the familiar Pascal text editor that has been enhanced to allow customers to create and print illustrated documents. The illustrations are contained in plot files produced by graphics editors such as HP EGS/200 or by customer programs.

HP EGS/200 drawing:



Pictures are drawn scaled to a size specified by the picture command. If you do not need to see the pictures, HP TechWriter allows you to turn off the display of your figures if you wish to. The editor also provides a number of convenient features for developing documents, such as right justification, multi-line headers and footers, and automatic table of contents generation.



Side by side text and graphics allows you to produce annotated pictures on any printer which can support this feature. Local margin commands allow you to use special margins for quotation text or the text describing your picture, or to turn margins off completely for tables.

This sample was produced on a LaserJet printer (HP 2686A) showing its variable resolution graphics capability - the top picture is at 150 dots per inch while the lower picture is at 75 dots per inch.

Fig. 1. *A sample of an HP TechWriter document. The user sees the document on the display as it appears in this hard copy.*

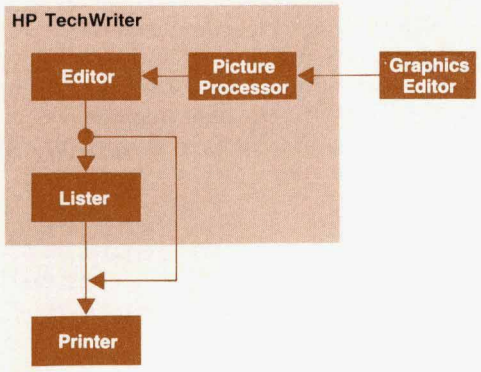


Fig. 2. *HP TechWriter consists of three programs. It accepts picture files from many HP graphics software products and works with most HP printers that have graphics capability.*

of the document. Turning off margination with a margin document command is a feature that is used to protect tables and similar text from being rearranged by the margining operation.

When margining text into a paragraph, HP TechWriter first checks to ensure that margining has not been turned off for that paragraph. After verifying that margination of the current text is allowed, the text ahead of the paragraph is scanned for a document command specifying local margins. If an applicable margin command exists, margin values are obtained from it. Otherwise, the global environment margins are used.

Each line of the new paragraph is then constructed word by word, with single spaces separating words regardless of how many spaces were originally typed. When a word is encountered that exceeds the right margin, a new line is started. When the last word of the paragraph is reached, the margining operation is finished if a ragged right margin was specified. If a straight right margin is indicated, spaces are added between the words of each line until the ends of the lines align with the right margin. To avoid having paragraphs that look more dense on one side than the other, HP TechWriter inserts the needed blanks from alternating ends of the lines, distributing them across the line, in an "even" manner, but with priority given to words that end with a punctuation mark. Margining is very fast. An average paragraph is remargined in about 200 milliseconds.

Illustrating Documents

Illustrations used in documents may, at the user's option, be displayed on the screen while editing the document. Using this, the user actually views the picture while writing about it and can see how the picture will look next to the text before producing the hard copy (Fig. 3).

To illustrate documents, the user must first produce a plot file of a picture from a graphics editor such as HP EGS/200 or HP Graphics Presentations/200, or from programs written by the user in BASIC 3.0 or Pascal 2.1/3.0. This file must be processed using the Picture Processor. The Picture Processor reads the Hewlett-Packard Graphics Language (HP-GL) commands in the plot file, finds the limits of the drawing for ease in scaling the illustration, and writes a binary file containing the move, draw, and linetype data from the plot file. A binary file is used because the Editor can read it much faster than it can read the original ASCII file.

HP TechWriter is one of very few products that displays text and pictures simultaneously for the user. High-speed picture display is vital to making this feature useful. To include a processed picture in the document, the user enters a drawing command as shown in Fig. 4.

The drawing command line contains the file name of the

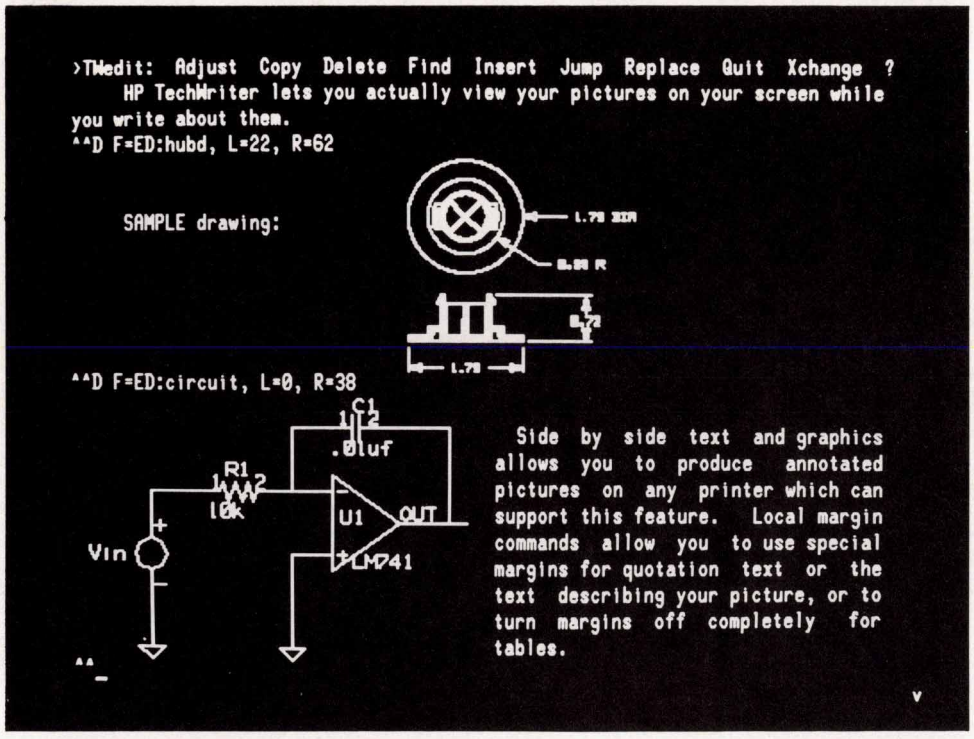


Fig. 3. *An example of an HP TechWriter screen.*

processed picture file and a specification for the left and right picture edges in columns. The drawing is completed by the next command line that appears in the file. The picture is drawn between the drawing command line and its terminator (the next command line) and between the right and left margins. The picture in Fig. 4, for example, is eight lines high and is drawn between columns 0 and 15.

Plot Files

HP TechWriter was originally conceived as a companion product to the HP Engineering Graphics System (HP EGS/200) to allow users of that system to produce documents illustrated with their drawings. Using the HP EGS/200 file structures, however, would have required a special interpreter to read those files, and would have limited HP TechWriter to illustrations produced by HP EGS/200. HP-GL plot files were ultimately chosen as the picture input file for HP TechWriter illustrations because they can be produced easily by any program that drives a plotter. The standard BASIC 3.0 and Pascal 2.1/3.0 (DGL) operating systems for HP 9000 Series 200 Computers produce these plot files. This allows the user to choose from a wide variety of currently marketed sources for document illustrations including, in addition to HP EGS/200, HP Graphics Presentations/ 200, HP Statistics Library/200, Graphics Editor/200, and Data Grapher/200.

HP TechWriter interprets only the limited subset of HP-GL commands produced by the standard BASIC and Pascal systems. These commands (PU, PD, LT, SP, and PA) produce consistent results on all HP plotters.

Text and Graphics Display

To maximize the size of a document that can be handled by the Editor, a drawing is read in from the disc and redrawn each time it appears on the screen. The process of reading and drawing the picture is relatively slow compared to the rest of the editing process. HP TechWriter minimizes the number of times the picture is redrawn by scrolling the picture when possible and keeping track of its location, redrawing it only when necessary. If there is no need to see the drawing at a particular time, the user can turn the picture draw option off, eliminating the time needed for redrawing.

Most Series 200 Computers have screens with two planes for display, alpha for the text and graphics for pictures. Either plane can be turned on or off independently. To display environment or help screens and then return to the main text display, the Editor turns off the graphics plane and redraws the alpha plane to show the desired display.

The two independent planes present a significant problem when trying to scroll pictures. Since the information for drawing a picture is not stored, pictures scrolled off the screen could lose tops or bottoms when scrolled back on. HP TechWriter solves the problem by scanning the displayed section of text and displaying only pictures that completely fit on the screen. When scrolling, the program first shifts the alpha plane one line, then shifts the graphics plane one alpha line (8-12 graphics lines). Assembly language procedures were written to shift graphics display lines quickly up or down a specified number of lines, and to erase only selected graphics lines. These special routines result in smooth and fairly fast scrolling for all Series 200 Computers, although moving the graphics plane of the Model 236C's color display is slow enough that a mild "ripple" effect is visible while scrolling pictures.

The HP 9000 Model 237 display is a high-resolution bit-mapped display that represents a departure from the separate alpha and graphic planes architecture, since it makes no distinction between alpha and graphics display information (it is all graphics). As a result, it represents both a complication and a simplification for HP TechWriter to support. On one hand, scrolling becomes a single operation with no distinction between scrolling alpha and scrolling graphics, but on the other hand, displaying text over a picture and switching to alternate HP TechWriter displays is more difficult.

Instead of turning off the graphics display, as is done on the other Series 200 Computers, HP TechWriter erases the Model 237 display. However, the Model 237's bit-mover hardware makes this operation very fast. The unavoidable result is that the equivalent of "turning on" the display now involves redrawing the picture from disc.

Displaying text over a picture on most Series 200 Computers is simple, since there are two display planes. For the Model 237, the replacement rule register is used to cause figures to be ORed as they are drawn with any existing text displayed on the screen. Again, the result is a high-performance integration of alpha and graphics information displayed by the Editor.

Printing Side-by-Side Text and Graphics

HP TechWriter is exceptional in its ability to print out pictures with, next to, or even on top of the text in the

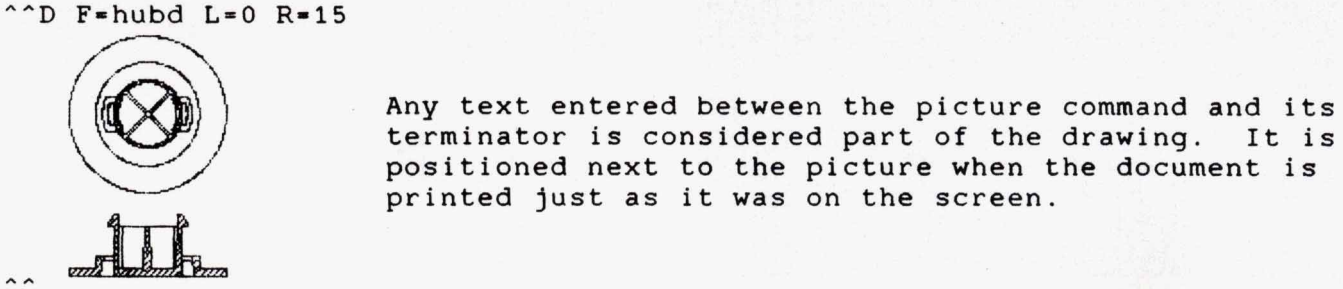


Fig. 4. *To include a drawing in the document, the user enters a drawing command as shown on the top line in this hard copy.*

HP TechWriter Security

HP TechWriter uses a software security mechanism developed for HP EGS/200. The mechanism is designed to allow program copying, but to limit full use of the program to those who have purchased the software. The mechanism also permits HP sales personnel to demonstrate the program on any available machine.

The HP TechWriter package includes a right-to-use certificate, which has a serial number and instructions for contacting Hewlett-Packard to obtain a code word. The code word provided is based on the certificate number and the model and serial numbers contained in the ID PROM of the user's HP 9000 Series 200 Computer, and is valid only for that machine.

When HP TechWriter is executed, the program looks for a code word in the configuration file. If no code word is found or the code word does not agree with what the ID PROM model and serial numbers indicate it should be, then the program will not allow disc write operations. By allowing HP TechWriter to execute in either case, the functionality of the program, including reading, modifying, and listing existing disc files, can be easily demonstrated. Some example text and pictures files are provided with the product for this purpose. However, it is not possible to save the results of the Editor on disc, so it is difficult to do useful work when the correct code word is not present.

HP TechWriter allows one configuration file to be used by more than one computer by allowing multiple code word entries in the configuration file. This means that a shared resource disc or a removable disc volume may be used by more than one computer using HP TechWriter by including a code word for each computer.

document on a wide range of printer types. Several technical problems were solved to make this possible. Different printers have a variety of resolutions, features, and interfaces to those features. The HP Printer Command Language, HP-PCL, is now standardizing feature interfaces, but HP TechWriter has to work with existing printers. The variety of printers results from advancing technology and the wide differences in cost and uses for printers. We chose to support current printers to make the product useful for our present as well as future users, and to offer users a range of options in terms of printer speed, price, quality, and type. Supporting any one printer would have been much easier, but would not have offered users this range of options. The few printers qualified on the HP 9000 Series 200 that are not supported by HP TechWriter generally do not offer graphics and have feature interfaces that are radically different from the HP-PCL definition.

To produce a rough hard copy of a picture, HP TechWriter uses the Pascal library Device-independent Graphics Language (DGL) to plot the picture to graphics memory. This takes advantage of that product's fast vector-to-raster conversion. HP TechWriter then reads the screen memory and dumps raster rows to the printer. Since there are several ways in which graphics memory is organized among the various HP 9000 Series 200 Computers (ranging from a bit per pixel to a byte per pixel), assembly language procedures are used to read whichever graphics memory is in use and construct the one-bit-per-pixel representation required by all supported printers.

To produce side-by-side pictures and text, HP TechWriter writes the alpha line without advancing the paper, then writes one alpha line equivalent of graphics rows to the printer. Side-by-side text and graphics requires a printer that can write alpha lines without advancing the paper.

Pictures are positioned and sized in the document based on screen text columns and lines. The HP 9000 Series 200 Computer display screens vary in their graphics-dots-to-alpha-cell ratio, and no printers match any of the graphics-to-alpha dot ratios of the screens. Therefore, pictures produced this way tend to look different on the screen and in hard copy, as shown in Fig. 5.

In general, the printers have narrower and taller character cells than the HP 9000 Series 200 displays. To maintain the aspect ratio of the picture as specified in the document, pictures generally print short in the X direction and long in the Y direction with respect to the surrounding text. The exact discrepancy, of course, depends on which printer and display are in use. Some match the height of the cell in dots exactly, and others may be short in the Y direction.

HP TechWriter dealt with this problem initially by com-

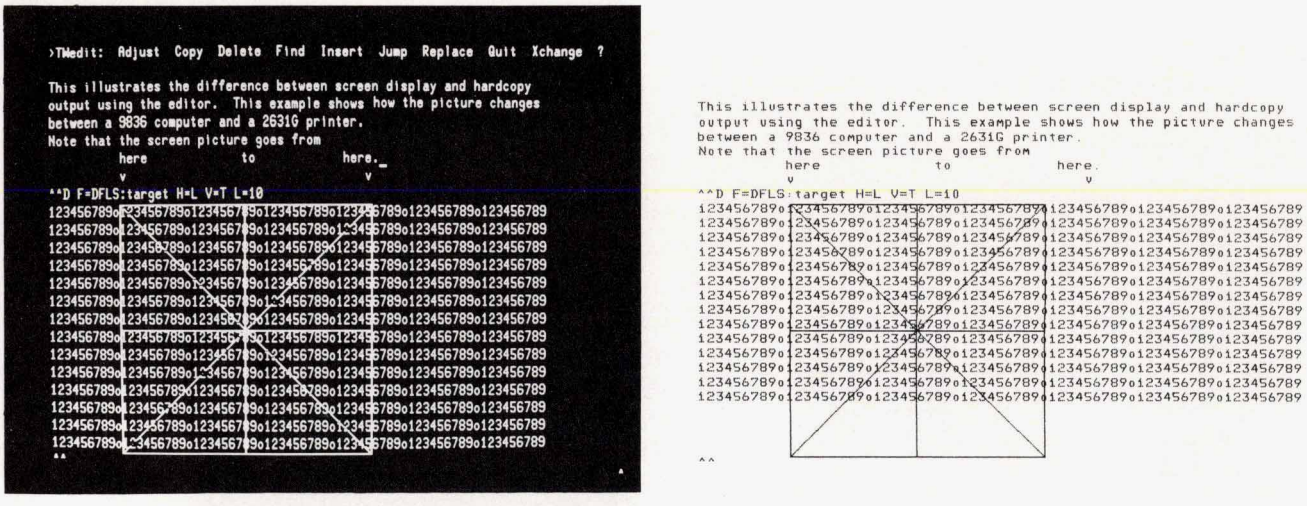


Fig. 5. *Because the ratio of graphics dots to alpha dots varies from display to display and from printer to printer, text may look different on the screen (left) and in hard copy (right).*

```
~027)0M~015

      Mean time to failure for the Weibull model is
^^m off on
^^s n=0
                                -1/~014b~015        1                ~027=~014
                    m~015 =~014 a          +~015(1+-)             ~027=~014
                                           b                          ~015


^^s
```

(a)

Mean time to failure for the Weibull model is

(b) $$\mu = \alpha^{-1/\beta}\,\Gamma\left(1+\frac{1}{\beta}\right)$$

Fig. 6. *Using an escape character (~ in this case), the user can take advantage of all of the printer's features. Here the input shown in (a) produces the formula shown in (b).*

puting the difference in resolution between the printer and the screen in each dimension, then allowing the user to specify what to do with the difference. The Editor still produces pictures in this way, providing draft-quality picture hard copy. The recently upgraded version of the Lister goes one step farther; it plots the picture using the correct horizontal and vertical character cell size in graphic dots for the particular printer in use, then prints that information. Pictures printed from the Lister use all of the graphics resolution available from the printer, and can be larger in both dimensions than pictures printed by the Editor. These larger, printer-density pictures are called "final-quality," and look as good as the printer can produce.

User Ingenuity

Every printer is different and every engineer has different specifications for documents. HP TechWriter does not attempt to address all combinations. It provides a minimum functionality, then gives the user the ability to program extended functionality into the document.

For instance, the spacing command was introduced to allow production of double-spaced documents. The spacing command lets the user set any spacing desired, including zero spacing (i.e., overprint). By using this feature, the user can produce pseudobold output or special combination characters.

The most powerful feature of this kind is the escape character. The escape character packs the three digits immediately following it into one byte to send to the printer. Using the individual printer's manual, the user can take advantage of all features provided by the printer. For example, if the escape character is a ~, the sequence shown in Fig. 6a causes an HP 2934A Printer to print the formula shown in Fig. 6b.

Briefly, the printer is instructed to select a math character set as the alternate character set by the command ~027)0M. The ~015 directs use of the normal character set, and ~014 specifies the alternate character set. The ~027= forces a half-line line feed. Note that this example uses half-line spacing since the HP 2934A Printer does not have subscript or superscript commands. Half-line spacing moves the paper up while spacing is turned off (^^s n=0), thereby causing a mismatch between the line-on-page count and the actual line on the page. This will cause the footer to be misplaced on the page. To reposition the footer correctly, either the page could be half-line spaced up again, or the number of lines per page for the printer could be doubled for one line and a blank line printed. Printing two lines at twice the lines per page in this way would result in one blank line on the output and would correctly reset the line-on-page count.

Acknowledgments

Many people contributed to HP TechWriter. Product managers Kathy Gillich-Adams and Doug Blackwood defined a sufficient but not overly ambitious set of document functionality. Project manager Danny Darr provided leadership and support to help get the project done correctly and on time. HP TechWriter was vigorously tested and improved by the efforts of our 66 alpha sites, who represented our target users and who supported us in our evaluation of necessary features and the general usefulness of the product.

Magnetostatic-Wave Devices for Microwave Signal Processing

By locally perturbing the magnetic dipoles formed by spinning electrons in thin ferrimagnetic films, a propagating wave can be initiated. Devices based on this principle can be used to process microwave signals.

by Waguih S. Ishak and Kok-Wai Chang

MAGNETOSTATIC WAVES propagating in a thin film of ferrimagnetic material provide an attractive means for processing signals in the 0.5-to-26.5-GHz frequency range. In addition to their application as tunable delay lines, couplers, and signal-to-noise enhancers, the simple fabrication steps and the excellent performance of magnetostatic-wave (MSW) devices makes them very competitive with other microwave device technologies such as yttrium-iron-garnet (YIG) spheres in such applications as filters and oscillators. This report reviews some of the basic principles of MSW device technology and describes the current state of the art and some of its applications.

History and Current Status

Early experiments with magnetostatic waves in bulk YIG material were begun in the late 1950s to demonstrate tunable microwave delay lines for use in pulse compression, frequency translation, and parametric amplifier circuits. However, these devices did not reach product engineering status because of basic material problems such as the nonuniform internal magnetic field inherent in all nonellipsoidal ferrimagnetic shapes[1] such as rods and bars. The bulk delay lines demonstrated at that time had more than 30 dB of insertion loss and exhibited very limited dynamic range.

With the advent of liquid-phase epitaxy techniques for growing high-quality, single-crystal YIG films (initially developed for the fabrication of magnetic bubble memory devices), a renewed interest in MSW devices started in the mid 1970s. Since thin (1 to 100 μm) YIG films can be grown now with less than one defect per square inch, and because these films exhibit a uniform internal magnetic field throughout most of their cross sections, which reduces losses, magnetostatic-wave delay lines are presently being built with less than 5 dB of insertion loss at 10 GHz. Furthermore, the thin-film geometry lends itself to the use of integrated circuit manufacturing techniques, resulting in high device yield and excellent repeatability of performance.[2-5]

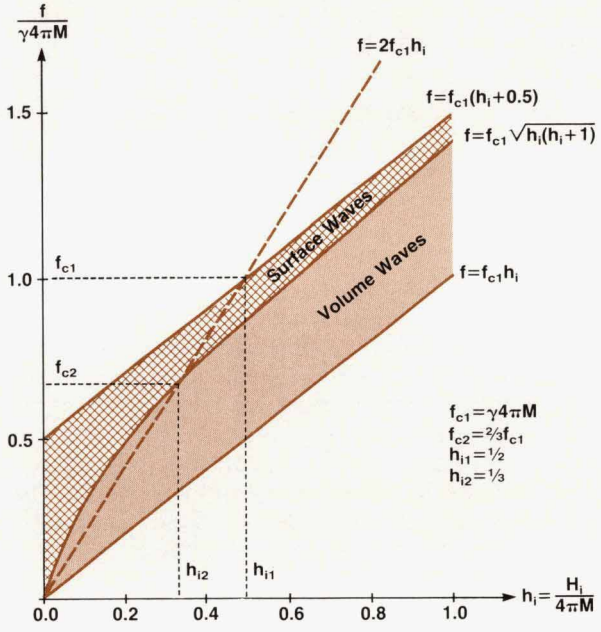


Fig. 2. *Graph of areas of operation for propagation of magnetostatic volume and surface waves in thin YIG films.*

MSW Propagation

Magnetostatic-wave propagation in thin ferrimagnetic films has been extensively considered. Three modes—mag-

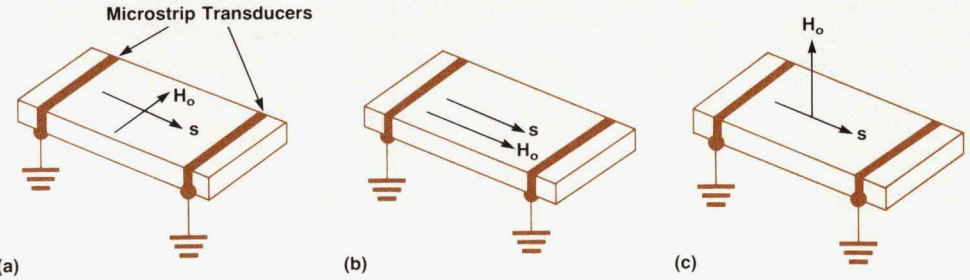


Fig. 1. *Three magnetostatic wave modes as determined by the orientation of the propagation vector* **s** *and the direction of the external magnetic bias field* H_o.[3] *(a) Surface wave mode. (b) Backward volume wave mode. (c) Forward volume wave mode.*

netostatic surface waves, backward volume waves, and forward volume waves—are of interest for device applications. These modes are determined by the orientations of the external magnetic bias field $\mathbf{H}_o$ and the propagation direction **s**, as shown in Fig. 1. These modes are dispersive and characterized by a narrow propagation passband and limited magnetic bias field tunability. The latter property can be used to choose a particular wavelength at any particular frequency in the passband. Table I summarizes the important dispersion characteristics of these three modes,[3] where:

β = wave number = 2π/wavelength = $2\pi/\lambda$
$4\pi\mathbf{M}$ = YIG saturation magnetization
γ = gyromagnetic ratio = 2.8 MHz/oersted
$\mathbf{H}_i$ = internal field = $\mathbf{H}_o + \mathbf{H}_{anisotropy}$

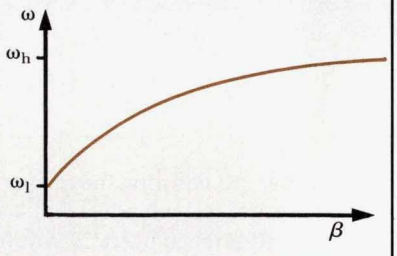
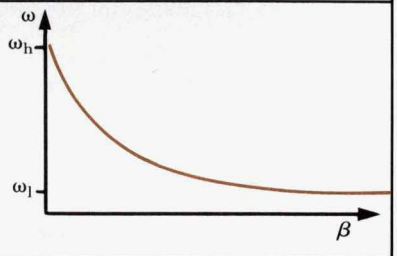
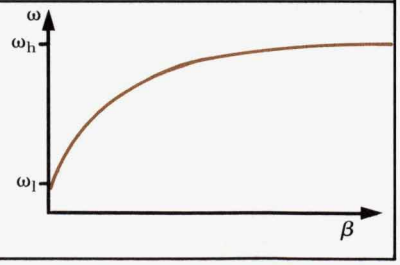
The potential bandwidth of these MSW devices is given by $\omega_h - \omega_l$, where ω_h and ω_l are as defined in Table I. For example, these expressions predict a bandwidth of more than 2.2 GHz at a center frequency of 10 GHz for a magnetostatic forward volume wave device.

As seen from the ω-β dispersion diagrams in Table I, the group-delay-versus-frequency characteristics of these three MSW modes are highly nonlinear. Fortunately, it is possible to modify these dispersion characteristics by introducing an external conductive plane in the vicinity of the YIG film or by growing a multilayer structure of a YIG film on a gadolinium gallium garnet (GGG) substrate. The separation between the YIG film and the conductive plane and/or the different thicknesses of the layers in the multilayer structure are parameters that can be controlled to determine the dispersion properties of these devices.

Propagation losses of MSW devices are proportional to the ferrimagnetic resonance linewidth ΔH. These losses (in dB/μs) are given approximately by 76.4 ΔH, where ΔH is in oersteds. Because ΔH increases approximately linearly with frequency, the propagation loss increases in a like fashion with frequency. Presently, pure YIG films are grown with ΔH less than 0.5 oersted at 10 GHz, resulting in a propagation loss of less than 36 dB/μs for X-band delay lines.

The various properties discussed above can be sum-

Table I
Magnetostatic Waves in Thin YIG Films

Mode	Dispersion Relations	Dispersion Diagram
Surface Wave	$\exp(2\beta\mathbf{s}) = \dfrac{(2\pi\mathbf{M})^2}{(\mathbf{H}_i + 2\pi\mathbf{M})^2 - (\omega/\gamma)^2}$ $\omega_l = \gamma\sqrt{\mathbf{H}_i(\mathbf{H}_i + 4\pi\mathbf{M})}$ $\omega_h = \gamma(\mathbf{H}_i + 2\pi\mathbf{M})$	
Backward Volume Wave	$2\cot(\alpha\beta\mathbf{s}) = \alpha - \alpha^{-1}$ $\omega_l = \gamma\mathbf{H}_i$ $\omega_h = \gamma\sqrt{\mathbf{H}_i(\mathbf{H}_i + 4\pi\mathbf{M})}$	
Forward Volume Wave	$\tan(\beta\mathbf{s}/2\alpha) = \alpha$ $\omega_l = \gamma\mathbf{H}_i$ $\omega_h = \gamma\sqrt{\mathbf{H}_i(\mathbf{H}_i + 4\pi\mathbf{M})}$	

$$\alpha = [(\omega/\gamma)^2 - \mathbf{H}_i^2] \times [\mathbf{H}_i(\mathbf{H}_i + 4\pi\mathbf{M}) - (\omega/\gamma)^2]^{-1}$$

Magnetic Resonance and YIG-Sphere Devices

Magnetic resonance is a phenomenon found in magnetic systems that possess both magnetic moments and angular momentum. According to classical theory, an atomic magnetic moment behaves as a gyroscope in the presence of an external static magnetic field. The term resonance implies that the atomic magnetic moment is in tune with a natural frequency of the magnetic system. In this case, the frequency corresponds to the rate of gyroscopic precession of the atomic moment about the direction of the external field.

Resonance occurring in ferrimagnetic materials such as single-crystal yttrium iron garnet (YIG) was first studied by C. Kittel in 1948. Since then it has been used to determine several material parameters such as relaxation time, crystal anisotropy, and magnetostriction. In 1957, Dillon observed very low ferrimagnetic resonance linewidth values for single-crystal YIG spheres and found that impurities, especially relaxing ions, and surface roughness increase linewidth considerably because of the excitation of spin waves.

Fig. 1 shows an atom with its associated magnetic moment created by a spinning electron. The magnetic moments of the atoms in each molecule of a YIG crystal combine to form a strong magnetic dipole. Application of an external magnetic field H_o causes these magnetic dipoles to align themselves in the direction of the field. This produces a strong net magnetization **M**. Any magnetic force acting at right angles to the bias field H_o results in precession of the dipoles around the field at a rate that depends on the strength of the field and, to a limited degree, on the basic properties of the material. Such lateral forces may result from RF magnetic fields orthogonal to H_o. If the frequency of the RF field coincides with the natural precession frequency of the magnetic dipoles in the material, a strong interaction results.

For spherical YIG resonators, the resonance frequency is given by:

$$f_r = \gamma(H_o + H_a)\ \text{MHz}$$

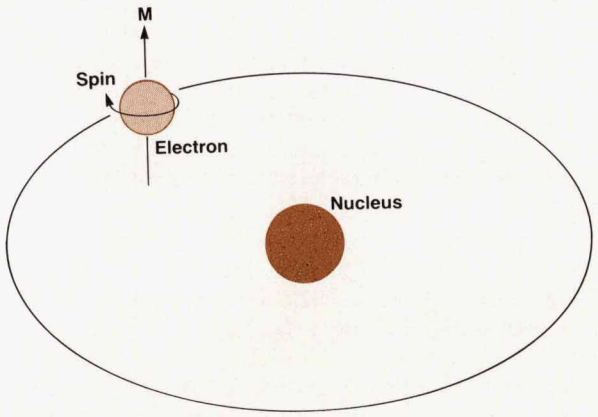


Fig. 1. *In a crystal of ferrimagnetic material, the spinning electrons circulating around the nuclei of the atoms forming the crystal lattice generate a strong magnetic moment* ***M*** *that precesses around the direction of an external magnetic field.*

where H_o is in oersteds, H_a is the anistropy field in oersteds, and γ is the gyromagnetic ratio (charge-to-mass ratio of an electron), which has a value of 2.8 MHz/oersted. For more than two decades, YIG spheres have found applications as practical microwave components such as bandpass and bandstop filters and resonators for oscillators. Because the resonance frequency can be changed by varying the strength of the bias field, YIG-sphere devices are widely used in such applications as swept receivers, tunable signal sources, test equipment, and electronic countermeasure systems operating at microwave frequencies over a bandwidth of several octaves. The interested reader desiring more details can consult the reference listed below.

Reference

N.K. Osbrink, "YIG-Tuned Oscillator Fundamentals," *Microwave Systems News*, November 1983, pp. 207-225.

marized in a graph defining the area of operation for different magnetostatic wave modes. Fig. 2 shows the frequency range of YIG-film magnetostatic volume and surface wave mode devices as a function of the internal field $\mathbf{H_i}$.

MSW Delay Lines: How They Work

Fig. 3 shows a YIG film grown on a GGG substrate and placed in an external magnetic bias field $\mathbf{H_o}$. The magnetic moments created by the spins of the bound electrons in the YIG film precess around the direction of the bias field in a manner analogous to the behavior of a gyroscope in a gravitational field. The gyroscope corresponds to the spinning electron and the force of gravity corresponds to the external bias field. Because these magnetic moments are coupled, a local perturbation caused by a current-carrying conductor (input transducer) creates a disturbance (wave) that propagates through the YIG film in a manner similar to that shown in Fig. 3b. When this magnetic disturbance reaches another conductor (output transducer) sometime later, it induces a current in the conductor similar to that initiating the disturbance, thus forming a simple delay line.

The generation and propagation of the magnetic disturbance can be likened to plucking a guitar string, which

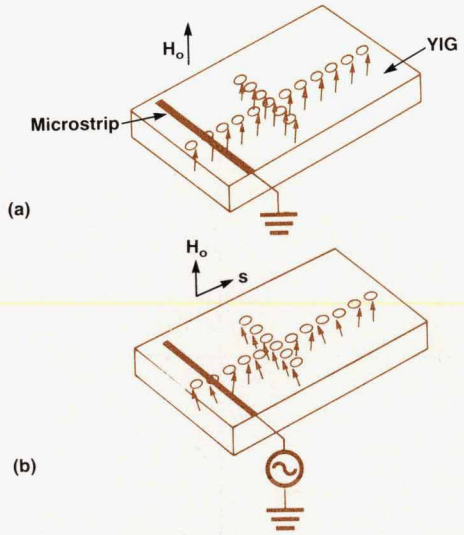


Fig. 3. *(a) Under uniform field conditions, the magnetic moments in a thin ferrimagnetic film precess at the same rate and have the same phase. (b) When an RF field is applied to the microstrip transducer, a local perturbation of the magnetic field occurs, which then propagates through the film via the coupling of adjacent magnetic moments.*

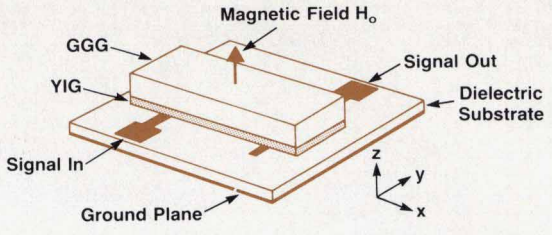


Fig. 4. *Basic configuration of a magnetostatic-wave delay line.*

causes a vibration to travel along the string. If the tension of the string is increased, it vibrates at a higher frequency. In a like manner, if the strength of the magnetic bias field is increased, the MSW delay line can be tuned for a higher frequency.

Fig. 4 shows a basic MSW delay line configuration. The input and output transducers, usually part of a microstrip circuit, are deposited on a dielectric substrate such as alumina or sapphire with a ground plane on the bottom surface and then brought close to the YIG/GGG structure. The thickness of the dielectric substrate determines the distance between the YIG film and the ground plane, and therefore, determines the dispersion characteristics of the delay line. This dispersion relation is plotted in Fig. 5a for a magnetostatic-surface-wave delay line using a YIG film 20 μm thick placed 254 μm away from the ground plane and an applied bias field of 1000 oersteds.

A more important parameter for delay lines is the group delay τ_g, which can be obtained by differentiating β with respect to ω ($2\pi f$). Fig. 5b shows τ_g as a function of frequency for the delay line described above. Notice the relatively flat delay region from 4.85 GHz to 5.05 GHz followed by a dispersive delay region above 5.05 GHz. Because MSW devices can be tuned by varying the bias field, these delay lines can be tuned electronically to have either a constant or a dispersive delay as shown in Fig. 6. In Fig. 6a, the device, which uses a doped, 18.7-μm-thick YIG film, is tuned to a center frequency of 840 MHz using an applied bias field of 85 oersteds, resulting in a nearly constant delay of 150 ns over a 120-MHz bandwidth. Without changing the bias field, the same device can be used as a dispersive delay line at a center frequency of 940 MHz with a dispersion of about 1.2 ns/MHz over a 120-MHz bandwidth as shown in Fig. 6b. By reducing the bias field to 70 oersteds, the dispersion can be recentered at 840 MHz as shown in Fig. 6c. This process can be realized over the tuning range of the MSW delay line (for this specific device, the tuning range is 640 MHz to 2.4 GHz).

Various methods have been suggested to increase the bandwidth of linearly dispersive MSW delay lines. In one method, the linearization of the dispersion characteristic is achieved by varying the ground plane spacing in a YIG-dielectric-ground-plane structure,[6] as shown by the cross section of the device in Fig. 7. The frequency response of this device is plotted in Fig. 8, which shows a linear dispersion over a bandwidth of more than 500 MHz at a center frequency of 3 GHz.

A nondispersive, electronically variable time-delay unit[7] can be constructed by cascading a dispersive magnetostatic-surface-wave delay line with a backward-volume-wave delay line as shown in Fig. 9a. When the individual delay lines are properly adjusted, a nondispersive response for the composite device can be obtained over a significant bandwidth (see Fig. 9b). By adjusting the bias field for either or both delay lines, the time delay for the composite device can also be controlled, as shown in Fig. 9c for an actual device operating at 3 GHz.

MSW Filters

Because the bandwidth of magnetostatic-wave delay lines can be controlled by the transducer design, these devices can be used in tunable wideband and narrowband filters at microwave frequencies. In fact, there is very little competition from other technologies for MSW tunable wideband filters.

When multielement grating transducers are used to make narrowband low-loss filters, bandwidths as narrow as 15 MHz have been obtained for the S (2 to 4 GHz), C (4 to 8 GHz), X (8 to 12 GHz), and K (18 to 27 GHz) frequency bands. Fig. 10 shows one such filter using seven-element gratings on a 254-μm-thick sapphire substrate. Fig. 11 shows a multitrace plot of the frequency response of this filter from 3 to 7 GHz. The ripples are as low as 0.1 dB, peak-to-peak, and the near-sidelobe rejection is greater than 25 dB.

MSW Resonators

One-port and two-port tunable magnetostatic-wave resonators can be fabricated using microstrip transducers for coupling and periodic etched-groove gratings as frequency-selective reflectors.[8]

At Hewlett-Packard, a special MSW resonator structure was developed to reduce the insertion loss and increase the Q. This device, called a straightedge resonator (SER),[9]

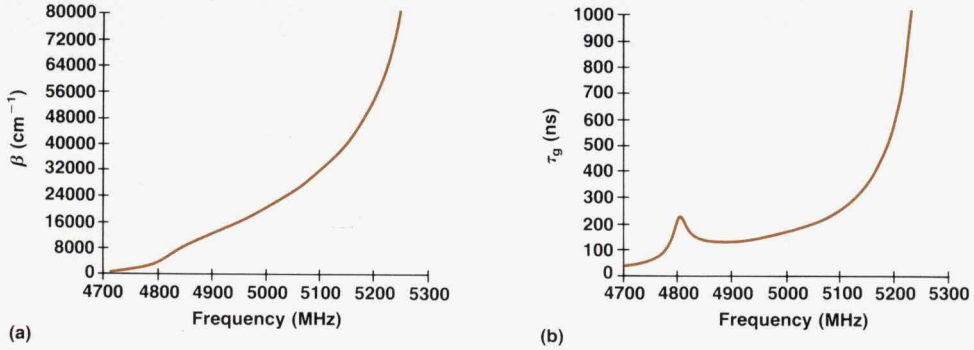


Fig. 5. *Dispersion (a) and group delay (b) versus frequency for a magnetostatic-surface-wave delay line using a YIG film 20 μm thick placed 254 μm away from the ground plane. The applied magnetic bias field is 1000 oersteds and $4\pi M$ = 1760 gauss.*

Spin Waves and Magnetostatic Waves[1]

Fig. 1a is a schematic representation of a ferrimagnetic material of thickness t. A chain of magnetic dipoles is shown precessing in a uniform ferrimagnetic resonance mode under the influence of an external bias field H_o and an RF field, h_{RF}, applied normal to H_o. This uniform precession mode is equivalent to a spin wave of zero order with an infinite wavelength. Increasing the amplitude of h_{RF} will increase the angle of precession with respect to H_o without changing the phase of the precessing dipoles. If the angle of precession of a single dipole is arbitrarily increased (as a result of a localized increase in h_{RF}), the nearest neighboring dipole will try to increase its precession also, by extracting energy from the internal exchange field. This process then continues to other neighboring dipoles, resulting in a volume spin wave of wavelength 2t/3 as shown in Fig. 1b. The process of energy transition from the RF field to the precessing dipole system is followed by an energy transition to the crystal lattice. This latter transition occurs either through the spin-spin relaxation channel or through the spin-lattice relaxation channel. Eventually the energy is converted into heat.

When the interaction between the RF electromagnetic field and the magnetic dipole spin system inside the YIG material is strong, the prevailing coupling mechanism among the spins is the magnetic dipolar field and the exchange effects are negligible. In this case, the spin waves are called magnetostatic. These waves are slow, magnetically dominated, and highly dispersive. They travel with velocities in the 3-to-300-km/s range and exhibit wavelengths from 1 μm to 1 mm.

Reference

1. G. Winkler, *Magnetic Garnets*, Friedr. Vieweg & Sohn, 1981.

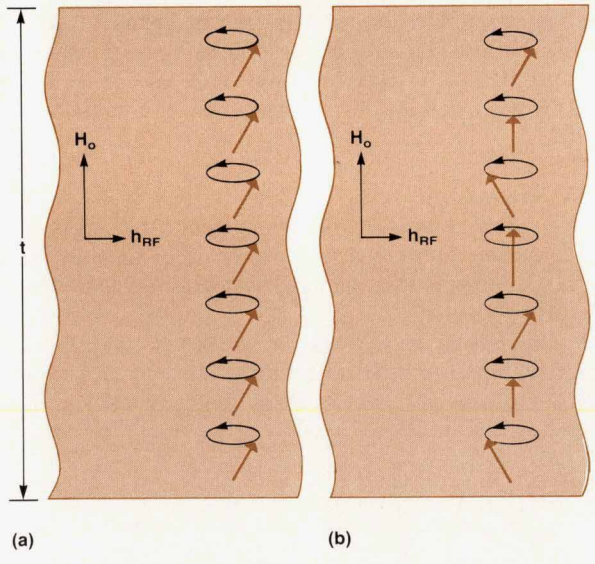


Fig. 1. *Under the influence of uniform magnetic fields (a), the magnetic dipoles in a ferrimagnetic material precess at the same rate and have the same phase. For this case, the wavelength = ∞. If the RF magnetic field is not uniform (b), a volume spin wave is set up where the phase of neighboring dipoles differs slightly and this difference propagates with a wavelength = 2t/3, where t is the thickness of the material.*

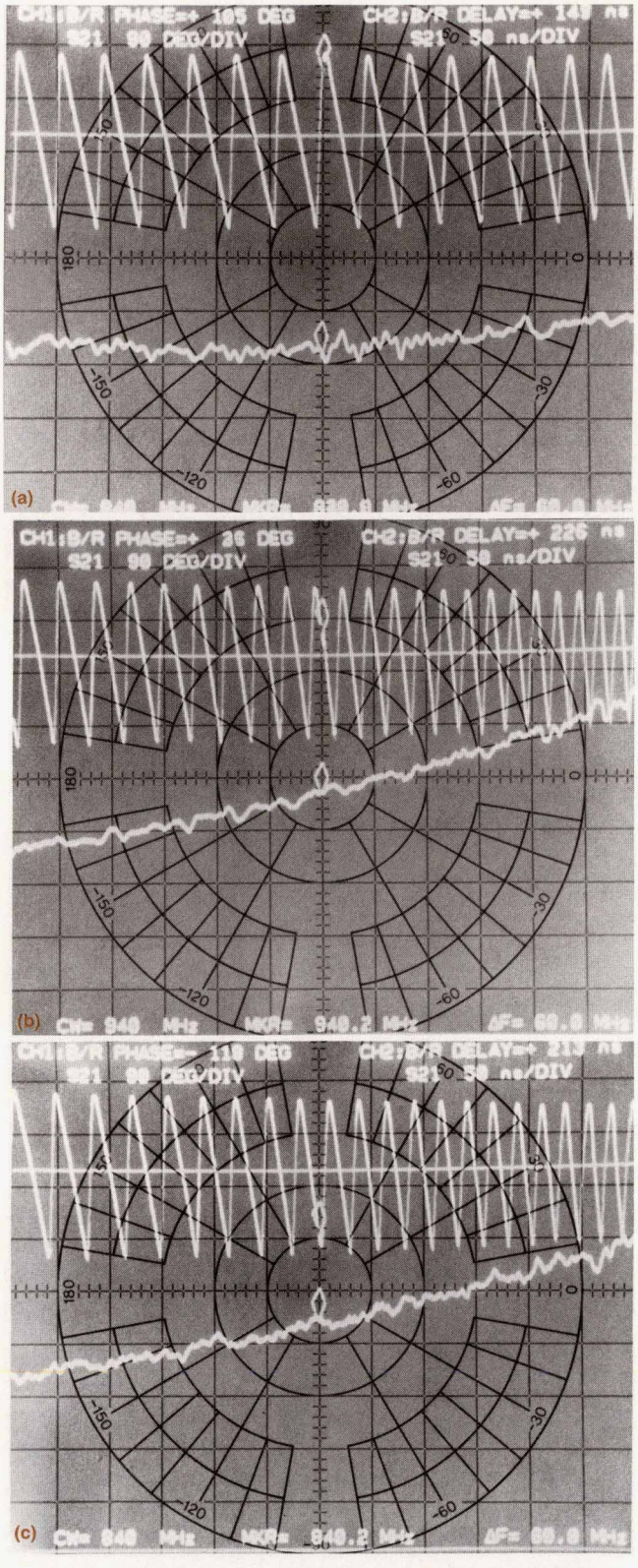


Fig. 6. *Group delay (lower trace, 50 ns/div) and phase (upper trace, 90 degrees/div) characteristics for a tunable magnetostatic-wave delay line versus frequency (10 MHz/div). (a) Center frequency = 840 MHz, H_o = 85 oersteds, relatively constant delay. (b) Center frequency = 940 MHz, H_o = 85 oersteds, dispersive delay. (c) Characteristics of (b) tuned for center frequency of (a) by decreasing H_o to 70 oersteds.*

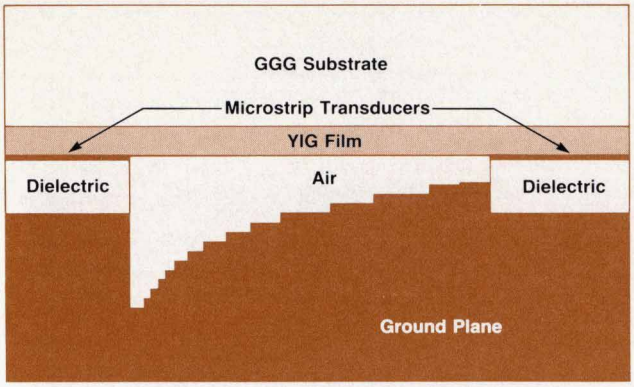


Fig. 7. *Cross section of an MSW delay line using continuously varying YIG-film-to-ground-plane spacing[6] to linearize the dispersion characteristic (see Fig. 8).*

consists of a ferrimagnetic resonant cavity made of a piece of YIG film on GGG cut into a rectangle by a wafer saw. The coupling transducers are gold microstrips patterned on a dielectric substrate, such as sapphire. The substrate is then mounted on an aluminum block, which forms the ground plane. This structure, shown in Fig. 12, is placed between the poles of an electromagnet so that the magnetic field is oriented parallel to the transducers, which permits coupling to magnetostatic surface waves. These waves propagate along the top surface of the YIG film and are reflected onto its other surface at the film's straight edges. A circulating wave pattern results that is resonant if the following condition is met:

$$(\beta_+ + \beta_-)l = 2\pi n$$

where n is an integer, β_+ and β_- are the wave numbers

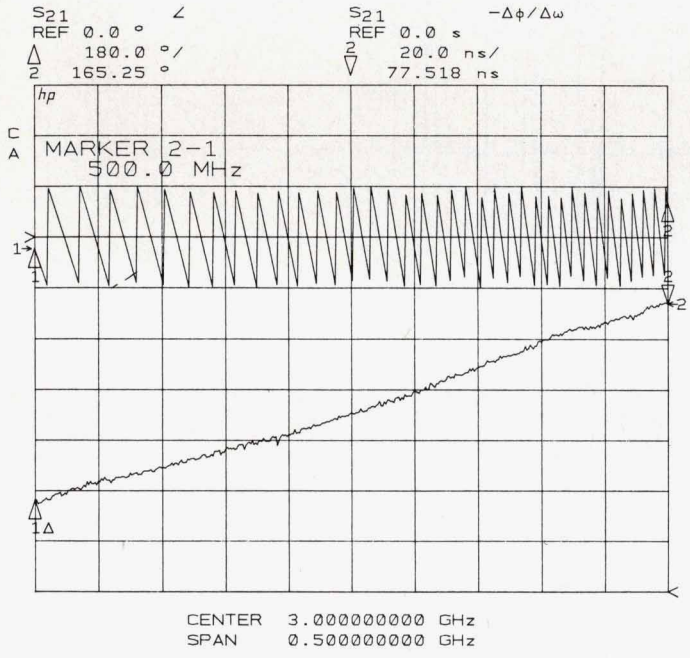


Fig. 8. *Phase (upper trace) and group delay (lower trace) characteristics of MSW delay line constructed with structure shown in Fig. 7.*

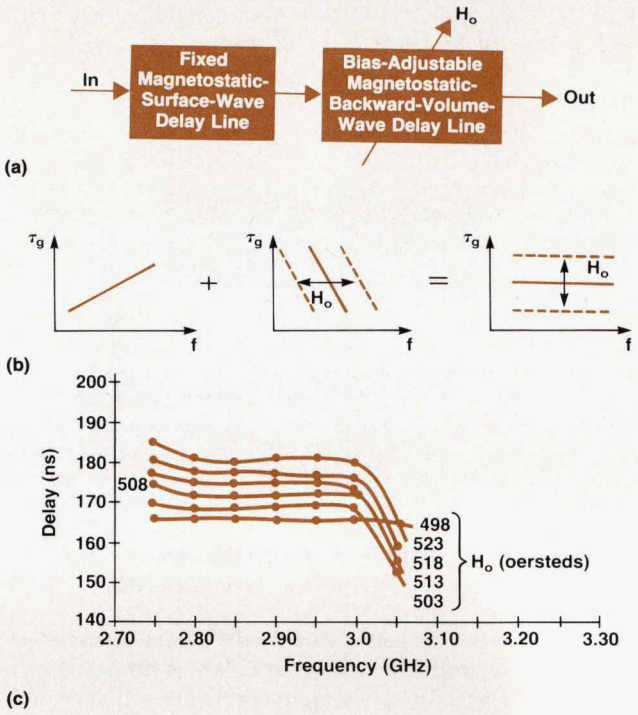


Fig. 9. *(a) By cascading a tunable backward-volume-wave delay line with a fixed surface-wave delay line, a nondispersive, electronically variable time-delay unit can be constructed (Sethares, et al[7]). (b) The combination of the group delay characteristics of the two MSW delay lines in (a) results in a nondispersive response that can be adjusted for different delays by varying the strength of the external bias field H_o. (c) Delay characteristics for a device constructed as outlined in (a).*

for the top and bottom surfaces, respectively, and l is the distance between the straightedge reflectors. Very-high-Q resonators have been demonstrated and Fig. 13 shows a typical transmission response (amplitude and phase) at 8.5 GHz. In addition to their improved performance, these devices are simpler to construct than the etched-groove-reflector resonators.

Signal-to-Noise Enhancers

A signal-to-noise enhancer is a device that exhibits a high attenuation of small noise-level signals and a lower attenuation of signals whose levels are above a certain threshold.[10] This is the opposite of the characteristic of the more familiar power limiter. A typical application for such a device would be in a frequency memory loop where a microwave pulse is to be stored for a specified time.

A signal-to-noise enhancer using a YIG film on GGG is shown in Fig. 14. When a microwave signal is applied to the input port, a magnetostatic wave is launched in the YIG film, which results in an attenuation of the microwave signal. For small input signals, most of the energy is used to launch the MSW, resulting in a large attenuation of these signals at the output port. For larger input signals, only a fraction of their energy is used to launch the MSW wave, resulting in less attenuation at the output port. Fig. 15 shows the measured attenuation versus frequency for different input power levels. Over a frequency range of 2.7 to

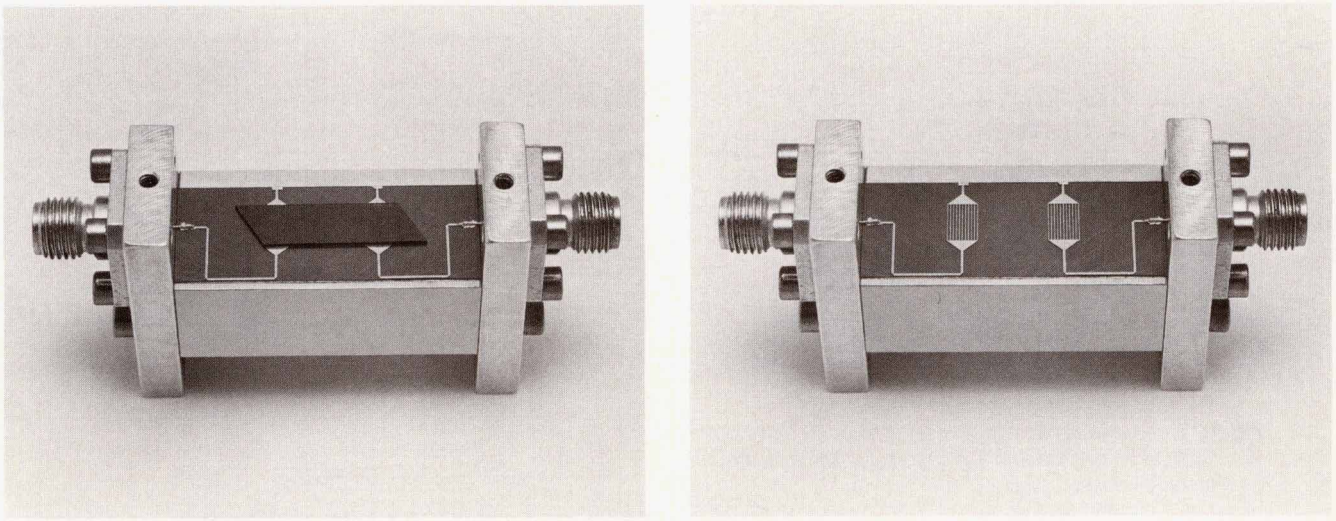


Fig. 10. *Photographs of a narrowband MSW delay-line filter using seven-element transducer gratings on a 254-μm-thick sapphire substrate with YIG film in place (left) and removed (right).*

3.7 GHz, this device can give a 20-dB signal-to-noise enhancement. A similar device using a plate of lithium ferrite instead of the YIG film gives comparable performance over a frequency range of 5.5 to 7 GHz.

MSW Delay-Line Oscillators

Fig. 16 shows a schematic of a microwave oscillator using a magnetostatic-surface-wave delay line for the positive feedback path. The output of the delay line is fed to a solid-state microwave amplifier whose output is connected to a directional coupler. Oscillations occur at frequencies

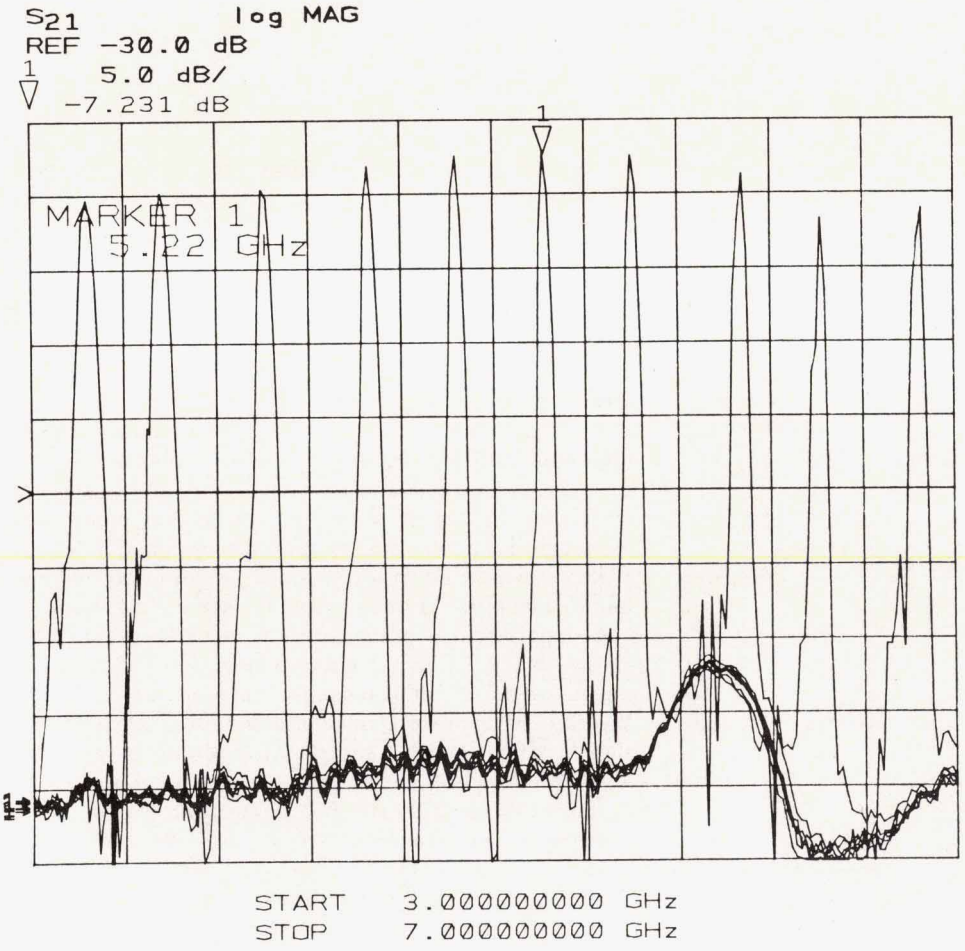


Fig. 11. *Frequency response of the device shown in Fig. 10 at ten values of the applied magnetic bias field.*

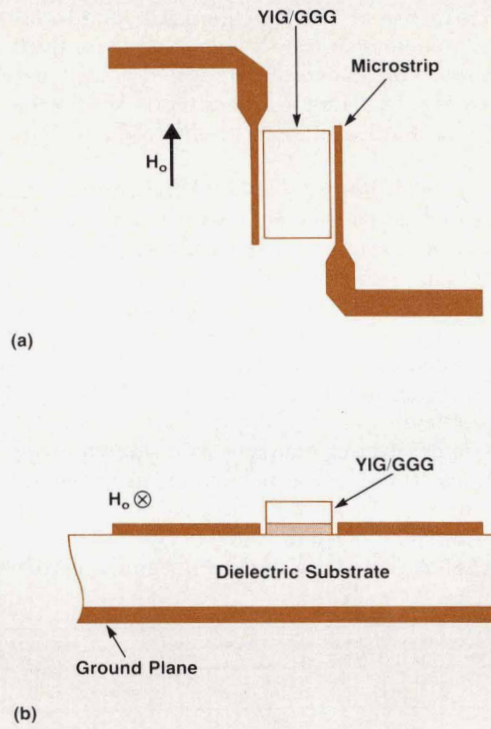


Fig. 12. *Top (a) and side (b) views of basic straightedge MSW resonator structure.*[9]

where the loop gain exceeds unity and the phase shift around the loop satisfies the condition:

$$\phi_{\text{delay line}} + \phi_{\text{amplifier}} + \phi_{\text{coupler}} + \phi_{\text{cables}} = 2\pi n$$

where n is an integer. To achieve single-mode oscillations, the delay line must be a narrowband device. In the example

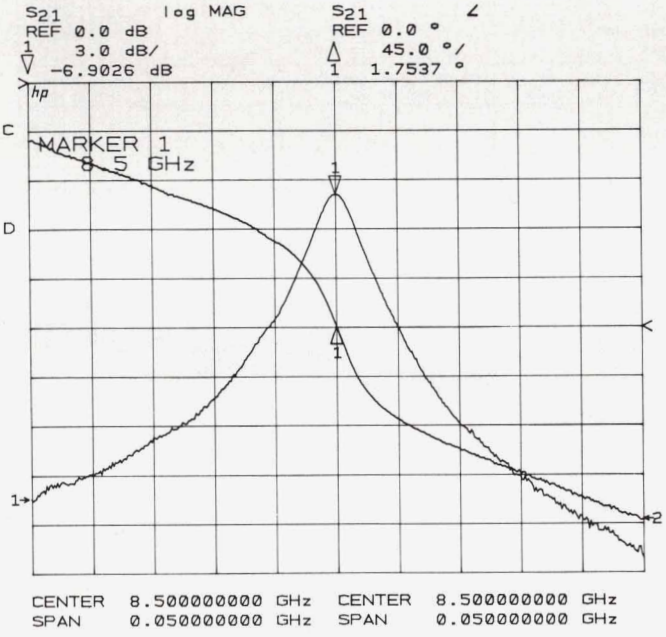


Fig. 13. *Typical amplitude (curve 1) and phase (curve 2) response for an 8.5-GHz straightedge MSW resonator.*

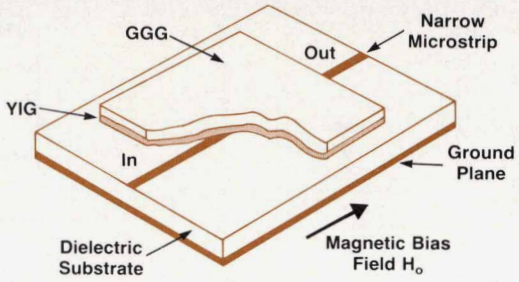


Fig. 14. *Basic structure of an MSW signal-to-noise enhancer (Adam, et al*[10]*).*

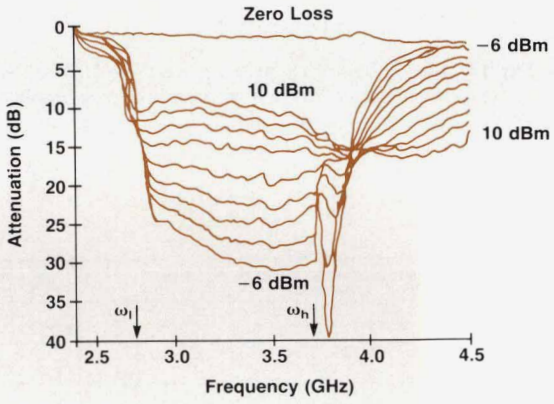


Fig. 15. *Response versus frequency of an MSW signal-to-noise enhancer for different input power levels from 10 dBm to −6 dBm.*

discussed here, this was achieved by using multielement gratings for the input and output transducers and spacing the YIG film 400 μm away from the transducers.[11] The

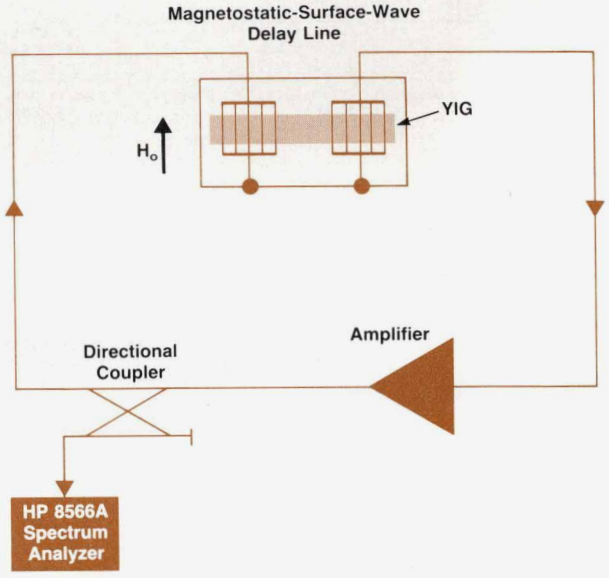


Fig. 16. *Schematic of a microwave oscillator using a magnetostatic-surface-wave delay line for the frequency-control element. Oscillation occurs for frequencies where the total phase shift around the circuit is a multiple of 2π radians.*

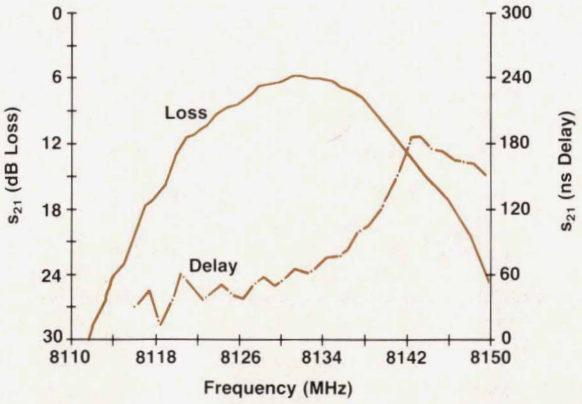


Fig. 17. *Amplitude and delay response of an MSW delay line using an 18-μm-thick YIG film, a 254-μm-thick sapphire substrate, seven-element microstrip transducers, and an external magnetic bias field of 2140 oersteds.*

frequency response of a delay line using an 18-μm-thick YIG film, seven-element transducers, a 254-μm-thick sapphire substrate, and an external bias field of 2140 oersteds is shown in Fig. 17. The 3-dB bandwidth is 15 MHz, the insertion loss is 6 dB, and the near-sidelobe rejection is 25 dB.

When this delay line was used in the configuration of Fig. 16, the oscillation frequency was tunable from 4 GHz to more than 24 GHz. A set of spectrum analyzer measurements is shown in Fig. 18 for oscillations at 4.77, 10.14, 16.43, and 23.78 GHz. Frequency jumping was observed at several points, which was attributed to the excessive delay associated with the cables and the multistage amplifier. Integrating this oscillator circuit should eliminate this problem.

The oscillator's tuning characteristic, shown in Fig. 19, exhibits an essentially linear behavior. The phase noise of the oscillator was measured using an HP 3047A Phase Noise Measurement System and an HP 11729B Carrier Noise Test Set. An excellent single-sideband noise of better

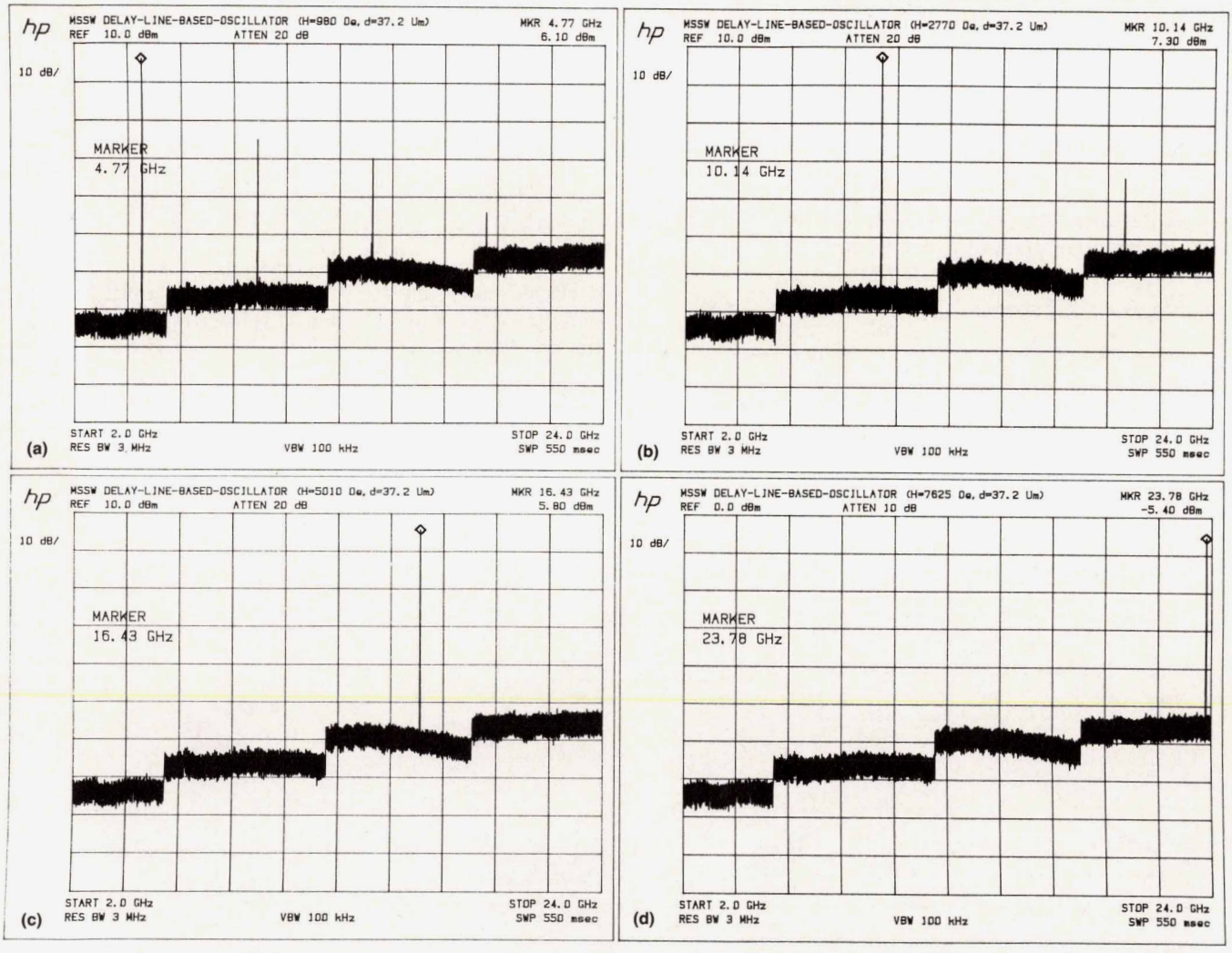


Fig. 18. *Frequency response of circuit shown in Fig. 16 using the MSW delay line whose characteristics are shown in Fig. 17. This oscillator is tunable from 4 GHz to more than 24 GHz, as shown by the spectrum analyzer plots for oscillation at (a) 4.77 GHz, (b) 10.14 GHz, (c) 16.43 GHz, and (d) 23.78 GHz.*

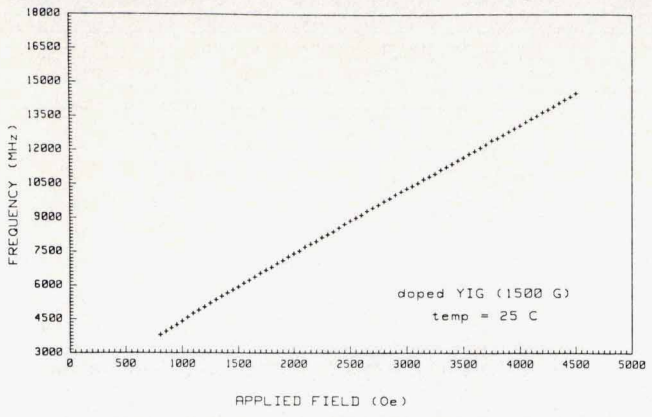


Fig. 19. *Tuning characteristic of oscillator described by Figs. 16, 17, and 18.*

than −110 dBc was measured for a 10-kHz frequency offset when the oscillator was tuned to 5.5 GHz (Fig. 20). This superior phase-noise characteristic and the simple IC-like fabrication and assembly of magnetostatic-surface-wave delay lines make magnetostatic-wave oscillators an attractive alternative to YIG-sphere oscillators.

MSW Frequency Multipliers

Another application for narrowband and wideband MSW delay-line filters is in frequency multiplication. Currently, microwave frequency multiplication is achieved using a step-recovery diode to generate the harmonics of an input signal. The diode is then followed by a YIG-sphere filter to choose the desired harmonic. One problem encountered in this configuration is the connection that must be made between the microstrip output from the step-recovery diode and the coupling loop around the YIG sphere. Another problem is the large shape factor of the YIG-sphere filter, which sometimes results in unacceptable harmonic suppression. If an MSW delay-line filter is used instead of the YIG-sphere filter, both problems can be solved. Fig. 21a shows the output of such a multiplier for an input signal of 1.682 GHz. The chosen output is the second harmonic and the suppression of the other harmonics is 50 dB. Fig. 21b shows the output of the same multiplier for the same input signal, but with the MSW filter tuned for the third harmonic. The harmonic suppression is again 50 dB.

Conclusions

Magnetostatic-wave devices and systems offer special capabilities in the UHF and microwave frequency bands, and have potential in millimeter-wave applications. Tunable MSW delay lines provide advantages over coaxial cables, especially at high microwave frequencies. Dispersive delay lines are attractive for pulse compression applications. Wideband tunable filters can be used in microwave sweepers and amplifiers for harmonic suppression, while narrowband filters can be used as preselection elements. Signal-to-noise enhancers provide the capabilities of improving the signal-to-noise performance of signal generators and the performance of frequency memory loops.

Systems built using MSW devices should find wide acceptance in instrumentation and communications applications because of the compatibility of the technology with existing microwave integrated circuits. As described above, tunable oscillators covering the 500-MHz-to-26-GHz frequency range are feasible. In addition to their excellent phase noise performance, these systems are easier to build than YIG-sphere oscillators, which have stringent polishing and mounting requirements. Another example is in the area of frequency multiplication, where an MSW filter can be used to select the desired harmonic at the output of a step-recovery diode.

Acknowledgments

The authors would like to acknowledge J.D. Adam, J.M. Owens, and J.C. Sethares for allowing the use of graphs they had published. We are also grateful to Johnny Ratcliff,

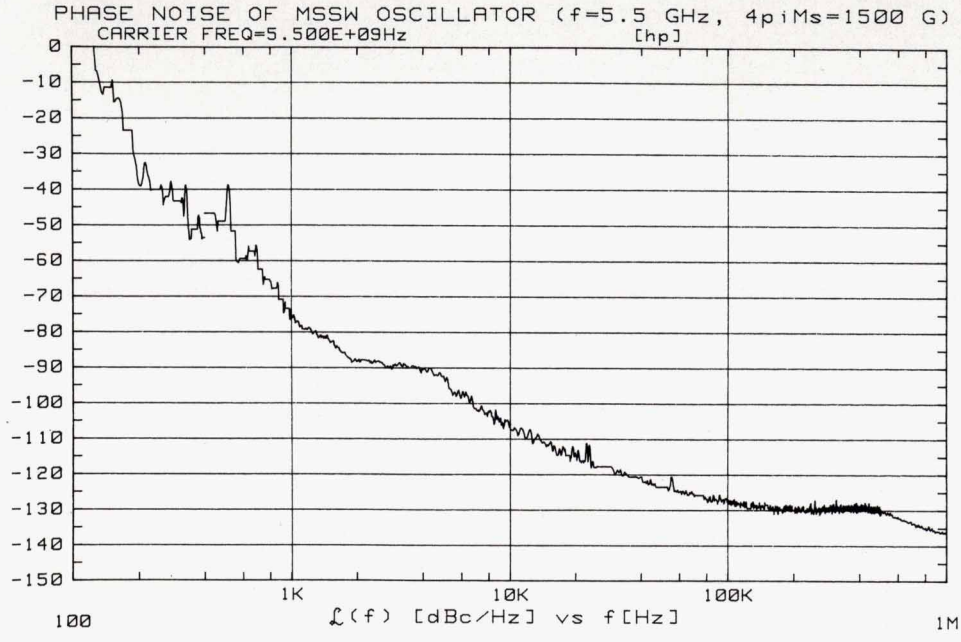


Fig. 20. *Phase noise characteristic of oscillator described by Figs. 16, 17, and 18, tuned to a frequency of 5.5 GHz.*

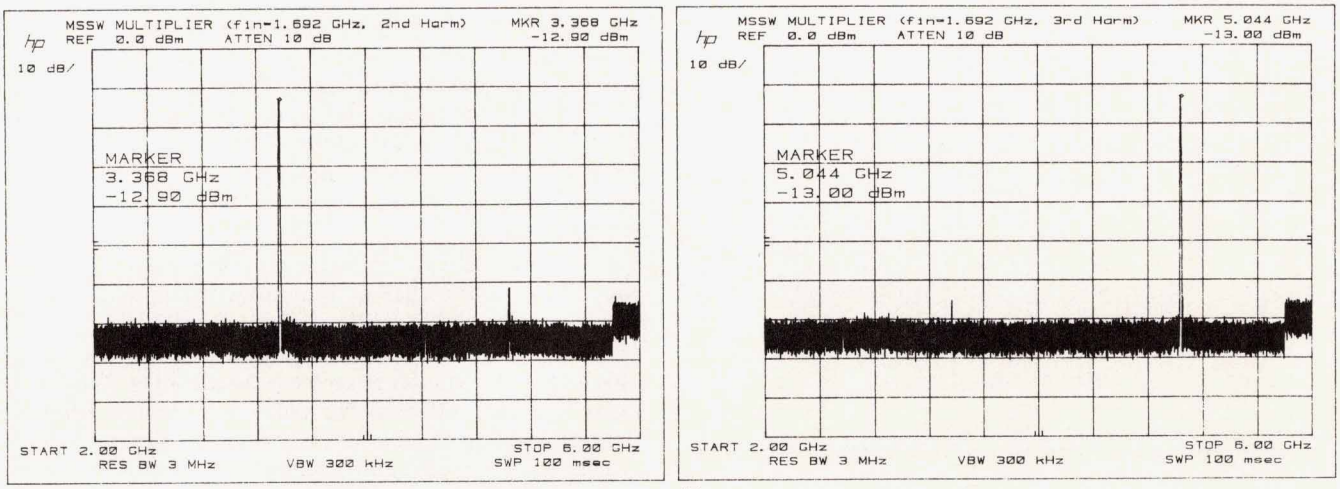


Fig. 21. *Response of MSW frequency multiplier tuned for the second (a) and the third (b) harmonics of a 1.682-GHz input signal generated by a step-recovery diode.*

Hylke Wiersma, and Elena Luiz for their fabrication and assembly of MSW devices. We are also grateful to Eli Reese for testing the oscillators.

References

1. B. Lax and K.J. Button, *Microwave Ferrites and Ferrimagnetics*, McGraw-Hill, 1962, chapter 4.
2. W.S. Ishak, "Microwave Signal Processing Using Magnetostatic Wave Devices," *Proceedings, 1984 IEEE Ultrasonics Symposium.*
3. J.D. Adam and J.H. Collins, "Microwave Magnetostatic Delay Devices Based on Epitaxial Yttrium Iron Garnet," *Proceedings of the IEEE*, Vol. 64, May 1976, pp. 794-800.
4. J.M. Owens, et al, "Magnetostatic Waves, Microwave SAW?," *Proceedings, 1980 IEEE Ultrasonics Symposium*, pp. 506-513.
5. W.S. Ishak, "Magnetostatic Surface Wave Devices for UHF and L Band Applications," *IEEE Transactions on Magnetics*, Vol. MAG-19, September 1983, pp. 1880-1882.
6. K.W. Chang, J.M. Owens, and R.L. Carter, "Dispersive Time Delay Control of Magnetostatic Surface Waves by Variable Ground Plane Spacing," *Electronics Letters*, Vol. 19, July 1983, pp. 546-547.
7. J.C. Sethares, J.M. Owens, and C.V. Smith, Jr., "MSW Nondispersive Electronically Tunable Time Delay Elements," *Electronics Letters*, Vol. 16, October 1980, pp. 825-826.
8. J.P. Castera and P. Hartmann, "Magnetostatic Wave Resonators," *Proceedings, Microwave Magnetics Workshop*, June 1981, pp. 218-228.
9. E. Huijer and W.S. Ishak, "MSSW Resonators with Straight Edge Reflectors," *IEEE Transactions on Magnetics*, Vol. MAG-20, November 1984, pp. 1232-1234.
10. J.D. Adam and S.N. Stitzer, "A Magnetostatic Wave Signal-to-Noise Enhancer," *Applied Physics Letters*, Vol. 36, March 1980, pp. 485-486.
11. W.S. Ishak, "4-20 GHz Magnetostatic Wave Delay Line Oscillator," *Electronics Letters*, Vol. 19, October 1983, pp. 930-931.

Disc Caching in the System Processing Units of the HP 3000 Family of Computers

Disc caching uses the excess main memory and processor capacity of the high-end HP 3000s to eliminate a large portion of the disc access delays encountered in an uncached system.

by John R. Busch and Alan J. Kondoff

THE PRICE/PERFORMANCE RATIO of processors and the cost per bit of semiconductor memory have been falling rapidly in recent years, while moving-head discs have been getting denser but not much faster. The resulting increase in the ratio of access times between secondary and primary storage, coupled with a reduction in the required number of secondary storage servers, has caused low processor utilization in many systems. Proposals to address this problem have focused on:

- Reducing the basic disc access time through the use of higher-speed actuators and rotation rates or through multiple actuators and heads.
- Reducing the effective disc access time by front-ending one or more discs with a semiconductor disc cache providing multiple track buffering or by inserting an intermediate level in the storage hierarchy between the discs and main memory using CCD (charge-coupled device), magnetic bubble, or semiconductor RAM technology.
- Reducing the frequency of disc access through local and global file buffering in the main memory.

The research and development effort reported here examined alternative approaches to exploit current trends in processor and memory technology to realize significant improvements in system price/performance ratio, and applied the results to the HP 3000 family of computers. The solution we selected for the HP 3000 family was to apply the excess main memory and processor capacity of the newer systems to prefetch and cache disc regions and deliver data when requested at main memory speed rather than disc access speeds.[1] The approach eliminates not only a significant portion of the traffic between primary and secondary storage, but also the bulk of process delays waiting for the traffic that remains. Through the integrated management of data base, file, and transient object space, it matches data management requirements and the architecture of the HP 3000 family with the current trends in processor and memory technologies.

In this report, we analyze the alternatives for exploiting processor and memory technology trends with respect to cost, performance, and reliability. We discuss disc caching design alternatives including fetch, replacement, and write handling policies. We present an overview of the tools

Glossary

Read hit rate. The probability that the disc file information to be read will be found in a higher-speed memory and will not have to be obtained from a slower one.

Read miss. Failure to find the desired data in a higher-speed memory.

Nowait. A type of input/output in which a process requests an I/O operation and then proceeds with other processing while the I/O operation completes.

Write wait probability. The probability that a process will have to wait for a disc write operation to complete after it has initiated a nowait write.

File mapping. A type of disc caching in which disc files are addressed as virtual memory locations, with the processor hardware and memory manager performing physical location and file caching functions.

Disc domain. A contiguous section of disc storage containing file data.

File extent. A sequence of file records stored in a contiguous disc region.

Posting. Updating a record or file in its permanent home in nonvolatile disc storage.

MIPS. Million instructions per second.

LRU. Least recently used. A criterion for removing cached disc domains from cache storage to make room for currently needed domains.

Shadow paging. A transaction recovery mechanism in which file changes of uncommitted transactions are saved in temporary disc locations until transaction commit time, when the temporary changes are made permanent.

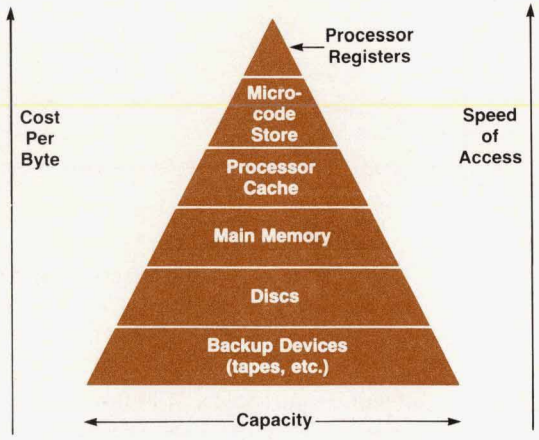


Fig. 1. *Standard computer system storage hierarchy.*

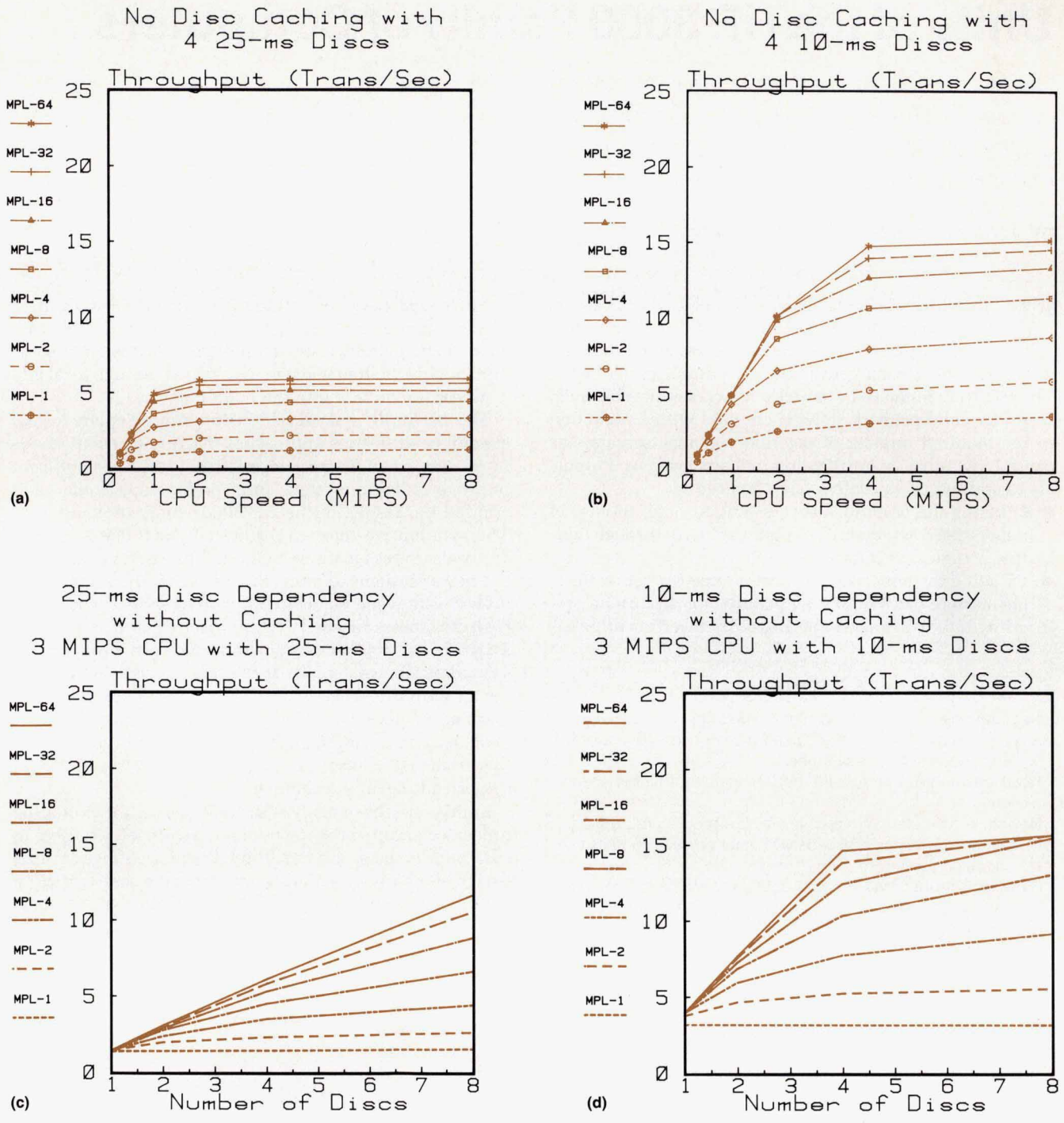


Fig. 2. *Performance of systems using conventional memory hierarchies without disc caching. Transaction throughput is shown as a function of effective multiprogramming level (MPL), the number of processes that can be concurrently executed. The top graphs show throughput versus CPU speed for a fixed number of discs. The bottom graphs show throughput versus number of discs for a fixed CPU speed.*

developed for the performance analysis of the alternatives, and we present measurements of the system and subsystem behavior and performance of our implementation for the HP 3000 family.

Disc caching is standard on HP 3000 Series 42, 48, and 68 systems and on HP 3000 High-Performance Series 39 systems. It is supported by the MPE-V/P and MPE-V/P Delta 1 operating systems, and by all releases of the MPE-V/E

(continued on page 24)

Disc Cache Performance Tools

To evaluate the performance of alternative system architectures and management policies, performance measurement and analysis tools were constructed. Instrumentation was defined, a statistics updating and reporting subsystem was designed and constructed, and display tools were specified and built. Workload generators were collected, and workload characterizers were constructed. A simulation was built to analyze secondary storage caching, and a system model was constructed for analysis of the impact of disc caching across a broad range of configurations and workloads. Fig. 1 shows the relationship of the various performance measurement and analysis tools that were constructed for this research.

The measurement subsystem supports low-overhead, event-driven statistics gathering. Statistics including resource queue length and service time histograms, service class compositions, transition distributions, and resource-specific event frequencies are supported by the processor, main memory, disc, and semaphore managers. A statistics display tool allows the measurer to specify the range of statistics to be measured and the desired duration of measurement.

To analyze the disc subsystem in detail, a trace-driven disc workload characterizer and disc cache simulator were specified and built. These tools provide insight into the workload's disc traffic and the impact of various forms of disc caching. These tools consume disc access trace files as input. A trace record contains the time, accessor class, access transfer count, access function, and access location for each disc access initiated by the system being traced.

The disc workload characterizer produces a profile of disc I/O workload characteristics. This includes a breakdown of the workload into its access type components (data base, directory, sequential access, etc.), interreference times, distributions of transfer sizes, active extent sizes, and extent LRU (least recently used) stack reference depths.

The disc cache simulation allows refined control over cache management policies (rounding, extent adherence, write handling, fetch sizes for each access type, and flush control). Any subset of the accessors can be cached, so that the localities of the access methods can be investigated in isolation. The simulator decomposes the disc references into modes and access functions and gives cache hit information for each access type. In addition to the cache performance information for the access types, the disc cache simulator also gives cache behavior information, including distributions of cache entry counts, cache reference LRU stack depths, and cache interreplacement times.

The hit rates obtained from the simulations are used as the processor-to-disc and processor-to-cache transition probabilities in the system model. Since the hit rates are obtained through simulation, main memory size and contention do not have to be captured explicitly by the system model. This significantly reduces the complexity of the system model.

A custom analytic system model was required so that a disc cache service station could be explicitly included. This allows the user to specify the system configuration and workload characteristics, and produces global performance estimates as well as resource demand and utilization statistics. The system model explicitly captures the effects of alternative secondary storage caching mechanisms, including external caching of discs through an intermediate storage level and internal caching of discs through file mapping or explicit caching in primary memory.

The system model queries for workload and configuration information. These inputs are obtained from knowledge of the installation, from the disc cache simulation, and from the statistics collection and reduction tools. Inputs are specified for processor speed, disc configuration and speed, channel speed, processor and disc service demand, disc cache overhead, and disc cache

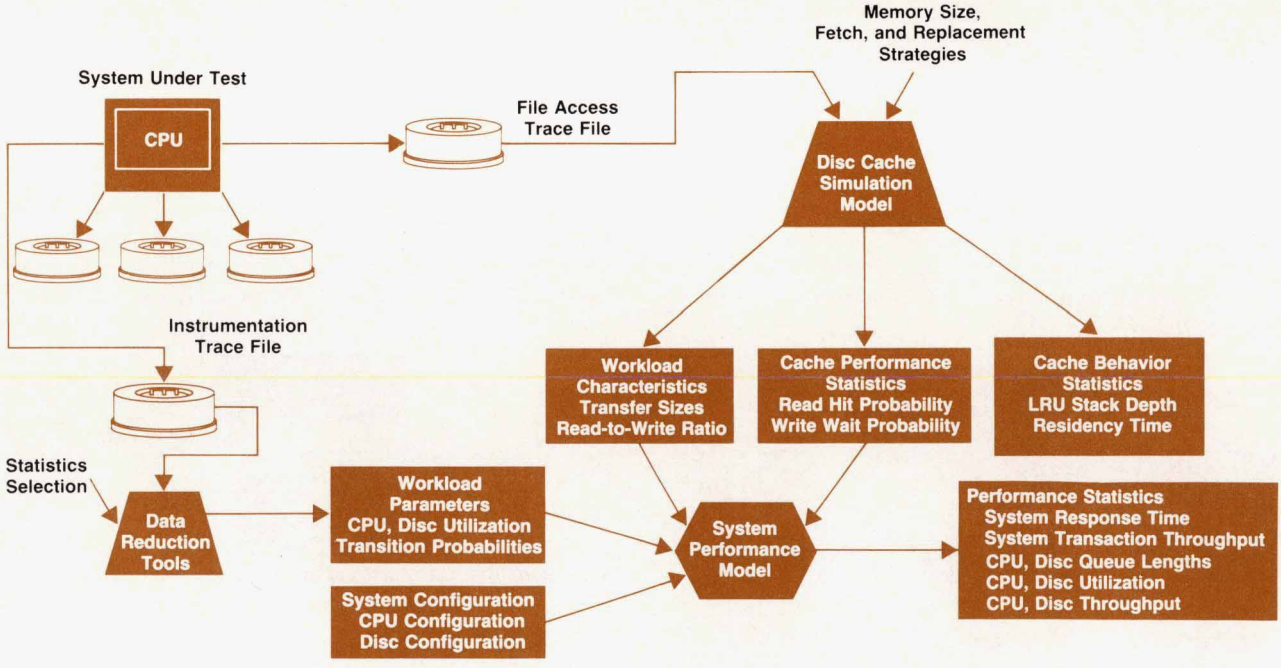


Fig. 1. *Performance measurement and analysis tools developed to evaluate alternative system architectures and management policies.*

read hit rate, fetch size, and post handling method.

Given the input system configuration and workload characterization, the system model program constructs a closed queueing model. The total workload is aggregated, and the network is decomposed into a simple central server model. For internal caching, the disc cache service station is absorbed into the processor station. The resulting central server model is solved through a standard iterative procedure[1] which iterates on the effective multiprogramming level.

Performance and behavioral characteristics are calculated for each server and for the system as a whole. For each iteration on the mean multiprogramming level, the utilization, mean queue length, throughput, and response time of each server are calculated, as are the overall system response time and throughput.

Serialization caused by contention for access control semaphores (e.g., data base locks) can significantly limit the effective multiprogramming level. The impact of serialization is not explicitly captured by the model. Its effects can be approximated by reducing the effective multiprogramming level.

Reference

1. S.S. Lavenberg, *Computer Performance Modeling Handbook*, Academic Press, New York, 1983.

(continued from page 22)

operating system. It includes software, additional memory, and in some cases, additional firmware.

Technology Trends

Fig. 1 (page 21) shows a conventional computer system storage hierarchy. Storage hierachies provide cost-effective system organizations for computer systems. Each successive level of a storage hierarchy uses lower-cost, but slower, memory components. By retaining frequently accessed code and data in the higher-speed memories, the system can operate at speeds close to the access times of the fastest memories while most of the storage capacity is in lower-cost, lower-speed memories. The price and performance of a computer system are often dominated by the organization and management of its storage hierarchy.

Achievable system performance is a direct function of processor speed and utilization. Processor utilization, and so the effective exploitation of high-speed processors, is limited primarily by the waiting time caused by misses at various levels of the storage hierarchy (i.e., the desired data is not found in a higher-speed memory and must be obtained from a slower one). Thus, for optimal system price/performance ratio, the processor speed and the capacity, speed, and management of the levels of the storage hierarchy must be matched.

Traditional solutions for a low hit rate at a certain level of the storage hierarchy include improving the management policies of the levels, increasing the capacity of the level incurring the low hit rate, speeding up the access time of the next lower level of the hierarchy, and introducing a new level into the storage hierarchy. Cost and technology determine which alternative or combination of alternatives is appropriate.

Processor speeds are rapidly increasing, and costs are dropping. The order-of-magnitude advances in processor speeds in recent years have not been matched by proportional advances in disc access speeds. There is an access time gap of five orders of magnitude, and widening, between a main memory reference and a disc reference in most current computer system storage hierarchies. By comparison, the access time gap between a processor cache reference and a processor cache miss resulting in a cache block fetch from main memory is normally less than an order of magnitude.

Disc densities are rapidly increasing, and large-capacity discs offer a fourfold price advantage over small-capacity discs. This trend in disc technology exacerbates the hierarchy imbalance. Since the disc subsystem often represents over 50% of the system cost, the trade-off towards a small number of high-capacity discs is attractive. However, the demand for a disc grows with its capacity. Thus, realizing the cost advantages of a few large-capacity discs over several small-capacity discs reduces the potential for parallel service. This increases the mean disc queue lengths, and thereby the expected values for disc service response time.

Replacing discs with secondary storage devices employing CCD, magnetic bubble, or semiconductor RAM technology has shown limited cost-effectiveness. Bubble and CCD memories have not been able to keep pace with the density improvements and resulting drop in cost per byte of semiconductor RAM, while the cost per megabyte of semiconductor RAM is still two orders of magnitude greater than that of discs. Leading-edge technologies[2,3] offer the potential for high-density discs with mean access times approaching 10 ms. Thus, magnetic-media secondary stor-

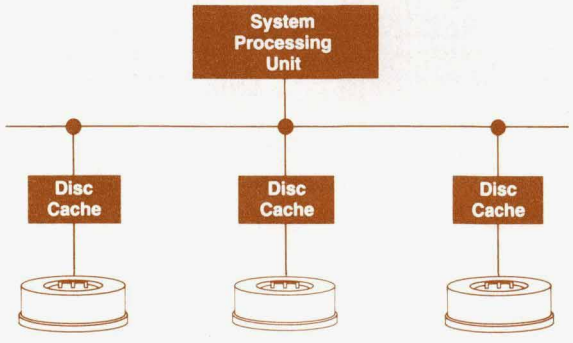


Fig. 3. *External disc caching using a local buffer for each disc.*

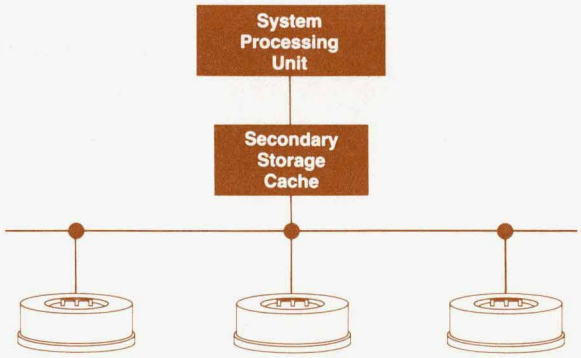


Fig. 4. *External caching at an intermediate storage level.*

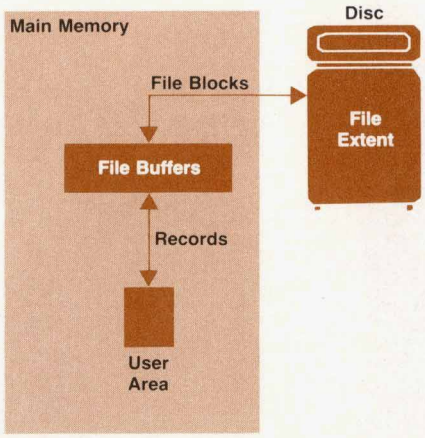


Fig. 5. *Many systems provide limited disc caching in main memory on a subsystem, file, or application basis.*

The MPE-IV Kernel

The kernel of an operating system provides higher-level system software including applications, data bases, and file systems with a virtual interface to the hardware and primitive system resources. The details of system configuration, of hardware interfacing, and of controlling access to system resources are hidden by the kernel.

The kernel multiplexes the system resources among the competing processes through the application of service policies. The overall performance of a general-purpose computer system is to a large extent determined by the effectiveness of these policies.

As the HP 3000 began to expand into a family of computers, a new kernel was required that would support its evolution. A research effort was undertaken, resulting in the MPE-IV kernel. The research approach and the design and performance of the resulting kernel are described in reference 1. Our implementation of disc caching for the HP 3000 family naturally extends the algorithms, measurement tools, and facilities of this kernel.

The kernel structure is based on decomposing the system into resource management modules that can be easily replaced, while enforcing within a module strong separation of policy from mechanism so that new algorithms can be implemented in a leveraged manner. Performance was sought through good algorithms, not instruction-level optimization.

Algorithms for main memory management, processor management, disc access and space management, interprocess communication, and semaphore management were developed that cooperate to optimize global system performance objectives rather than local performance objectives.

To achieve this cooperation, global priority assignments are made to processes based on external policy specifications. These process priority assignments are then reflected in the service requests for the various system resources so that local management decisions contribute to the global objectives. The resource management algorithms of the MPE-IV kernel are discussed in detail in reference 2.

References

1. J.R. Busch, "The MPE IV Kernel: History, Structure and Strategies," *Proceedings of the HP 3000 International User's Group Conference*, Orlando, April 27,1982, reprinted in *Journal of the HP 3000 International User's Group*, Vol. 5, no. 3-4, July-December 1982.

2. J.R. Busch, *Integrated Resource Management Algorithms for Computer Systems*, Ph.D. Dissertation, University of California, Los Angeles, 1984. University Microfilms International, Ann Arbor, Michigan, no. 8420156.

age devices and semiconductor primary memories can be expected to remain dominant in computer storage hierarchies in the years to come.

The results of the direct application of evolving technology are shown in Fig. 2 (page 22). These graphs use the analytic system model described in the box on page 23. The graphs show families of curves for increasing effective multiprogramming level (the number of processes in memory demanding service that can be concurrently executed). The top graphs show transaction throughput as a function of processor speed for a fixed number (4) of 25-ms and 10-ms discs. The bottom graphs show transaction throughput as a function of the number of 25-ms and 10-ms discs at a fixed processor speed of 3 MIPS.

A fixed channel bandwidth of 2M bytes/s is assumed for all the runs. Other assumptions were 10,000 instructions per disc visit, five reads per write access, 1K-byte mean transfer size, and five disc accesses per transaction. These are somewhat characteristic of the referencing patterns of the HP 3000 family, although observed variations are wide. Changes in these parameters shift the curves, but the general characteristics are the same.

We see that with conventional storage hierarchy management with this workload and with four 25-ms discs, effective processor utilization extends only to 1 MIPS, and beyond 1 MIPS, performance is linear with respect to the number of discs. With faster discs (four 10-ms discs), effective processor utilization extends through 4 MIPS, with a sharp dependency on the number of 10-ms discs for the higher multiprogramming levels. High effective multiprogramming levels are required throughout to exploit processor capacity.

This indicates that for effective utilization of higher-speed processors using conventional storage hierarchies and management techniques and high effective multiprogramming levels, ever faster discs in greater numbers are required. As discussed above, the technology trends in secondary storage devices do not support these requirements.

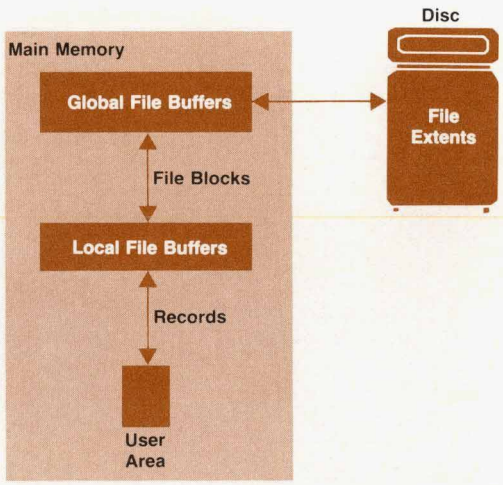


Fig. 6. *Some systems provide centralized buffering schemes, setting aside a portion of main memory for global file buffers.*

Alternative Balancing Approaches

System throughput is directly proportional to effective multiprogramming level, and inversely proportional to processor and disc response times and disc visit frequencies. Consequently, efforts to overcome the limitations in exploiting the trends in processor and memory technologies have focused on:

- Reducing the effective secondary storage access time through external disc caching techniques
- Reducing the number of secondary storage visits per transaction through internal disc caching in the primary memory
- Sustaining large effective multiprogramming levels through improved concurrency management schemes.

In the next two sections, methods of reducing effective disc access time and visit frequencies are examined. Modern methods for achieving high concurrency with data access in transactional systems are addressed in references 4 through 8.

External Caching Techniques

Caching techniques have been applied to the disc subsystem to reduce the effective disc access time to secondary storage. Caching has been implemented in discs, in controllers, and as stand-alone system components interfacing one or more secondary storage devices to the main memory.

The use of a local cache per disc is depicted in Fig. 3. Smith[9] discusses buffering discs using bubble, CCD, or electron beam memories. He concludes that three buffers, each a cylinder in size, would produce a hit ratio on the order of 96%, with LRU (least recently used) working well as a replacement policy. IBM has announced an intelligent cached disc featuring a 384K-byte microprocessor-driven controller that optimizes seeks and caches recently referenced data.[10] Krastins[11] discusses a cache consisting of 1M to 2M bytes of RAM that is integrated with the disc controller. The cache buffers full physical tracks. He reports a hit rate of 85%, a mean access time of 8 to 12 ms, and hit processing time less than 1 ms.

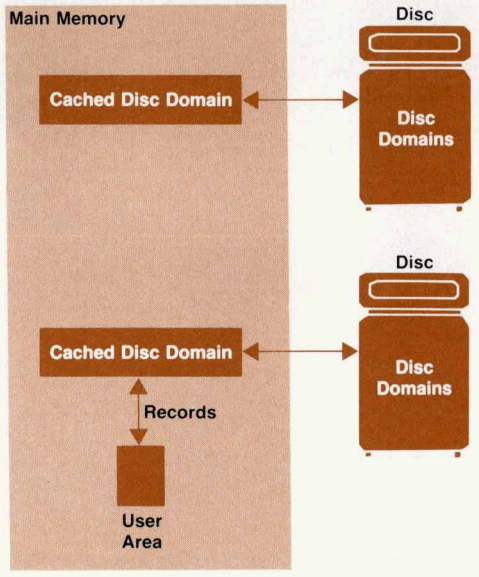


Fig. 7. *Explicit global disc caching in main memory.*

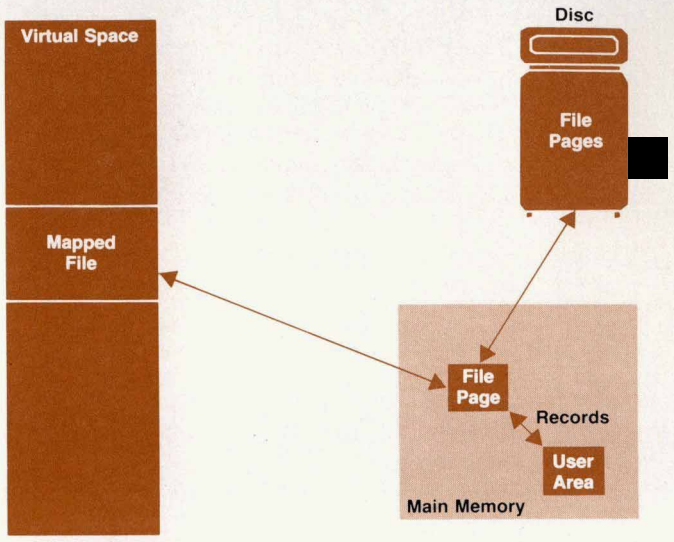


Fig. 8. *Internal caching by means of file mapping.*

The use of a cache front-ending the disc subsystem is depicted in Fig. 4. This can be viewed as inserting a new level into the storage hierarchy.

The use and organization of a CCD memory level that could serve to close the gap between semiconductor and disc access times are discussed in references 12 and 13, and the use of magnetic bubble memory systems for this purpose is discussed in references 14 and 15. But since bubble and CCD memories have not been able to keep pace with the density improvements and the drop in the cost per byte of semiconductor RAM, they have not qualified in gap-filling efforts. Rather, the technology of choice for external disc caches has been semiconductor random access memory.

Hugelshofer and Schultz[16] describe such a semiconductor disc cache marketed by Computer Automation, Inc. It consists of 2M bytes of RAM placed between the processing system and up to four moving-head disc drives. They quote access times of 4 ms on a hit, with up to an 85% hit ratio. It is packaged with four RAM cards and an LSI-2 I/O processor.

The use of external caching, either locally or through an intermediate storage level, reduces the effective secondary storage access time on read hits, and potentially decreases the access time on writes, provided immediate physical update of the disc media is not required on a write access before signaling transfer completion to the processor.

Using Large Primary Memories

With the improvements in memory densities and access times, very large main memories can be cost-effective provided they can be exploited to reduce the traffic between main and secondary storage. Techniques to exploit main memory for this purpose include auxiliary local buffering in applications and subsystems and global disc caching through explicit caching or file mapping.

Systems have conventionally provided limited caching of the discs in main memory through explicit buffering on

a subsystem, file, or application basis as shown in Fig. 5.

Centralized buffering schemes are employed in the Berkeley 4.2 UNIX file system for the Digital Equipment Corporation VAX computer family[17] and in a product for the IBM Personal Computer.[18] As shown in Fig. 6, a fixed portion of main memory is set aside for global file buffers. When a file block read from disc is requested, the global buffers are checked first. If the requested block is present in a global file buffer, the read is satisfied with a move from the buffer to the user's address space. Otherwise, a global buffer is freed, a disc read of the block is initiated into the selected global buffer, and when the read completes, the data is moved into the user's space. File block writes are performed through global buffers as well.

In global disc caching, main memory partitions disappear and cached disc regions containing file data are centrally managed with the pieces of transient objects by the main memory manager. In this approach to disc caching in main memory, pieces of the disc are mapped as data objects, and placed and replaced by the normal memory management algorithms like those used for code, stacks, etc. This approach is shown in Fig. 7.

Global disc caching in main memory can be either explicit beneath the disc access interface or implicit

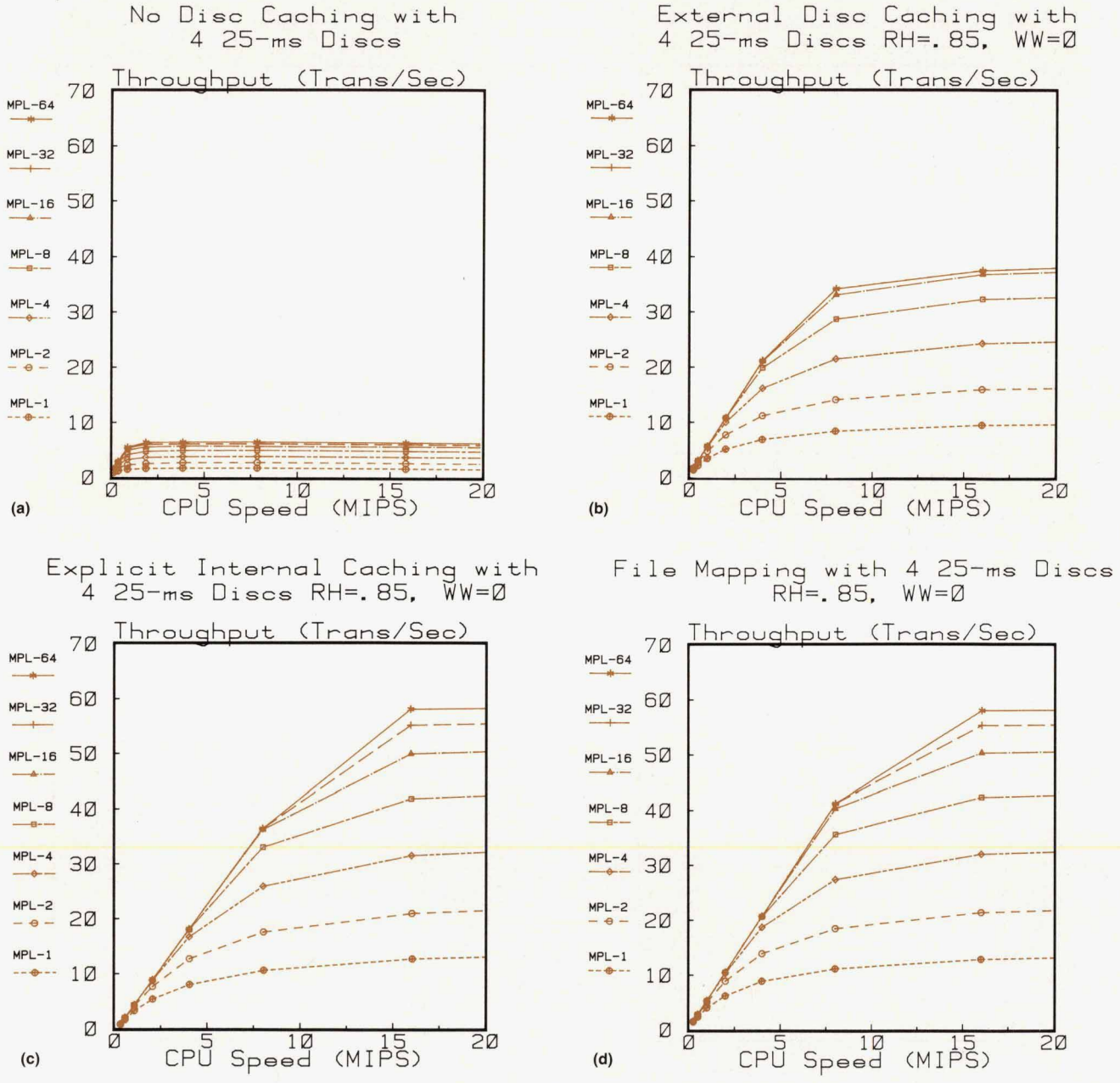


Fig. 9. *Performance of various disc caching alternatives compared to conventional storage management. Assumptions are four 25-ms discs, read hit rates of 85%, and write wait probabilities of zero.*

through the use of file mapping.

With explicit disc caching in main memory, the access methods continue to address the discs, but a software translation mechanism locates the required cached disc domain in main memory and moves the data between the cached region of the disc and the data area of the access method. This approach was implemented in the breadboard system for experimentation purposes, as described later in this report.

With architectures supporting large virtual address spaces, pieces of files or discs or entire files or discs can be mapped directly into the address space. The location of a piece of a file or disc in main memory is then handled by the virtual-to-physical address translation hardware, and the normal memory management mechanisms handle fetching and replacing of pages of files or discs. This approach to secondary storage caching is depicted in Fig. 8. This approach was first employed in Multics,[19] and more recently as described in reference 20.

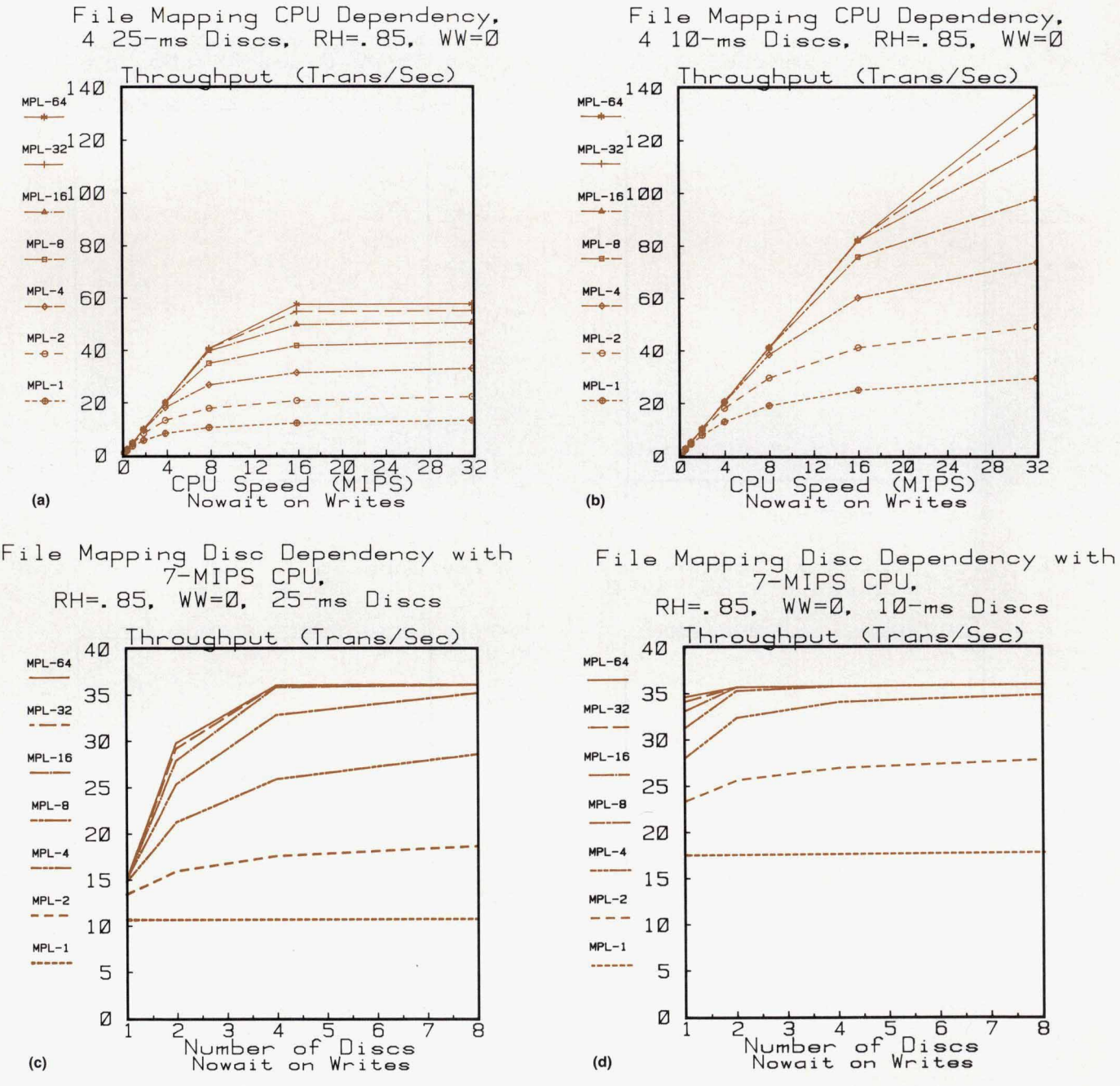


Fig. 10. *Dependency on disc access time and the number of discs is dramatically reduced with disc caching, as shown here for file mapping.*

Requirements of Transaction Management

Modern transactional data base systems guarantee that the data base is left in a consistent state in the event of a transaction abort or a system/disc failure. Solutions to the problem of balancing the storage hierarchy must preserve the consistency and recovery principles of the transactional system.

The standard mechanism used to achieve transaction recovery is the use of a write-ahead log.[21] In write-ahead logging, before the data in the data base is modified, copies

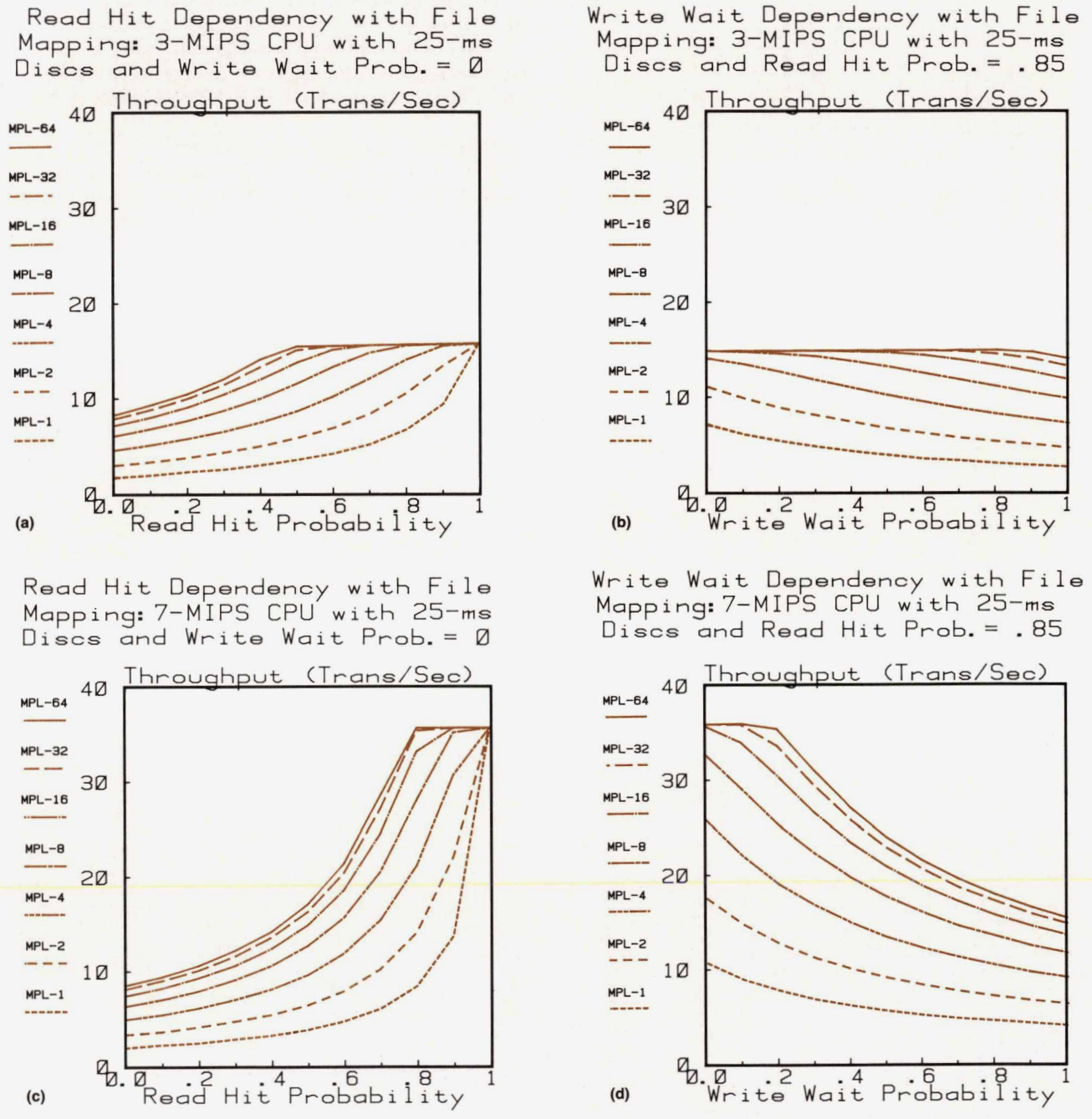
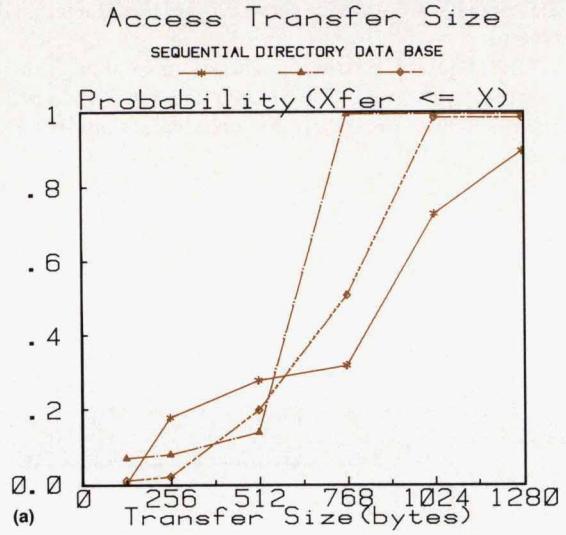
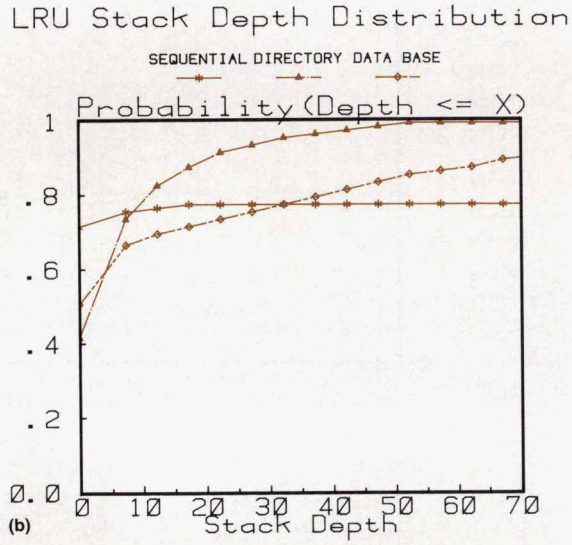


Fig. 11. *Sensitivity of file mapping to read hit rates and write wait probabilities increases with processor speed. A system using the internal disc caching techniques needs to exploit the known methods of achieving high read hit rates and low write wait probabilities.*





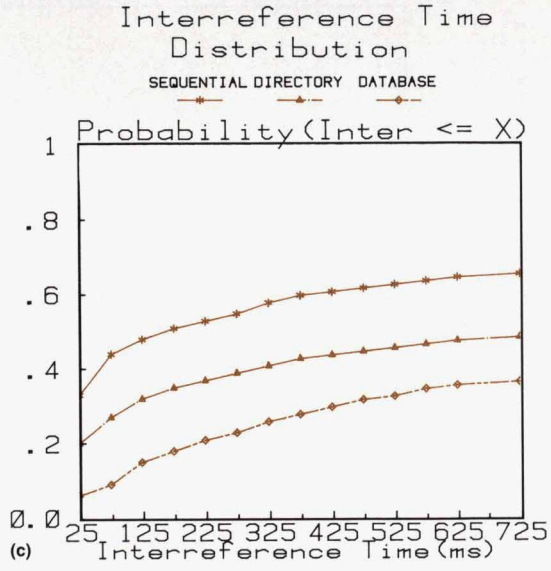


Fig. 12. *Workload characteristics of the HP 3000 family of computers without disc caching.*

of the old and new images of the data are written to a log file, along with identification information. If the transaction aborts, the old images of data modified by the transaction are read from the log and restored into the data base, thereby "undoing" the actions of the transaction. If the system crashes, a utility goes through the log file and undoes any actions of uncommitted transactions. If a disc fails, the data base is restored to its last backed-up state and the actions of committed transactions are redone using the log file.

For transaction recovery through write-ahead logging to work, the before and after images of the data along with identifying information must be posted to the disc before the data base is updated with the new values. Most data base systems issue the log write, then wait for the physical post to complete before issuing the data base write.

Comparative Analysis of Alternatives

In the following sections we evaluate the alternatives discussed above based on performance, cost, and reliability. We identify complementary approaches, and identify workload and configuration sensitivities affecting the suitability of alternatives.

Evaluation with Respect to Performance

Evaluating with respect to performance, we saw in the preceding section that effective processor capacity utilization drops rapidly with increasing processor speed if conventional storage management is employed. Faster discs and more discs provide benefit in such systems, but the benefits are not commensurate with those obtained through secondary storage caching. This is demonstrated in the

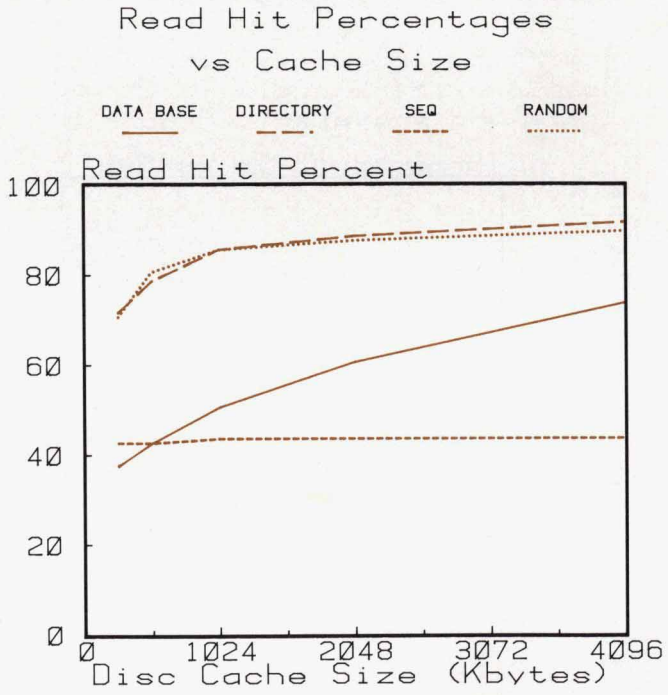


Fig. 13. *Potential benefit of disc caching for the HP 3000 family is shown by this graph of read hit rates that can be achieved by various access methods as a function of cache capacity.*

modeling results shown on page 27. The graphs show transaction throughput as a function of processor speed with conventional storage management (Fig. 9a), external disc caching (Fig. 9b), and caching of discs in the primary memory via explicit global disc caching (Fig. 9c) and file mapping (Fig. 9d). Assumptions are the previously specified workload characteristics, four 25-ms discs, read hit rates of 85% on disc file accesses, write wait probabilities of 0, 1K-byte mean transfers without caching, and 4K-byte transfers with LRU replacement with caching. All of the secondary storage caching alternatives are able to provide effective processor utilization at processor speeds several times those that can be effectively used without caching with equivalent processor/disc configurations. Further, disc caching in the primary memory, either explicit or through file mapping, is able to provide effective processor utilization at speeds twice those that can be effectively used by external caching, assuming the same overheads and hit and wait rates. This is a result of the access time differential between main and external cache storage and the higher degree of parallelism in cache access with primary storage caching.

The overhead of internal caching becomes negligible as processor speeds increase, whereas the overhead in external caches stays fixed. However, with slower processors and a balanced configuration leading to high processor utilization without caching secondary storage, explicit internal caching degrades performance relative to no caching. The added overhead of locating the disc region in memory and moving the data to the target area does not pay off when the processor utilization is high. This effect was observed in the HP 3000 internal caching measurements. The low-end family members degraded in performance when explicit internal disc caching was enabled. External caching outperforms explicit internal caching in high utilization ranges as well, since the data transfer is performed in parallel with processor use, so the move overhead does not consume valuable processor time.

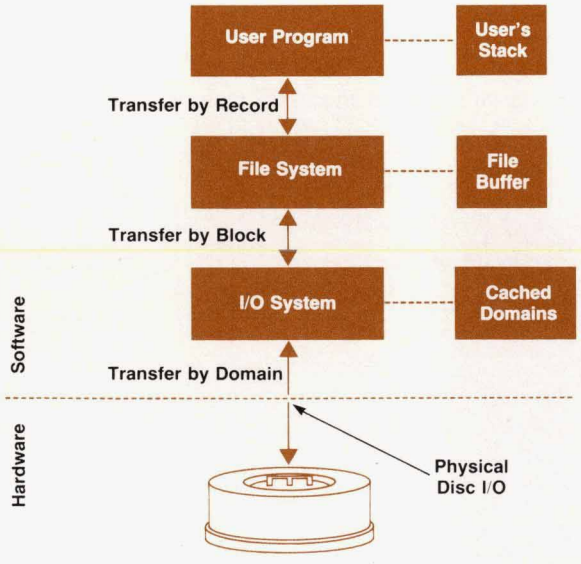


Fig. 14. *Explicit internal disc cache interfaces in the breadboard MPE operating system kernel developed for this study.*

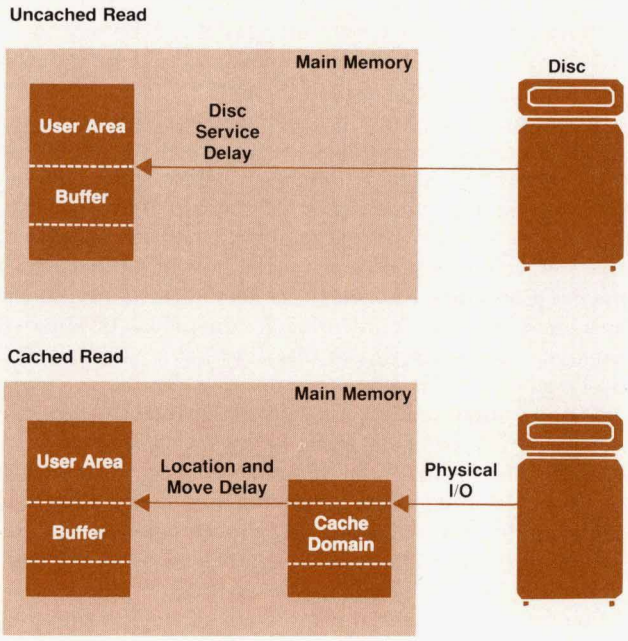


Fig. 15. *Disc read caching in the breadboard MPE kernel.*

The dependency on disc access time and the number of discs is dramatically reduced with the secondary storage caching alternatives. This is shown for file mapping in Fig. 10 for the workload, disc subsystem, and read hit and write wait probabilities used in the previous modeling results. The behaviors of external caching and explicit internal caching are similar. By contrast with the linear dependency on the number of discs seen in Fig. 2 for a 3-MIPS processor without secondary storage caching, full processor utilization can be achieved with only a few discs at 7 MIPS with secondary storage caching. Reducing disc access time from 25 ms to 10 ms increases the effective utilization range from 1 MIPS to 4 MIPS without caching (Fig. 2b). With caching, reducing the mean disc service time extends the effective utilization beyond 30 MIPS with file mapping (Fig. 10b), but processors through 16 MIPS can be effectively used with a few 25-ms discs (Fig. 10a), provided that adequate read hit rates and write wait probabilities can be achieved.

The performance impact of read hit rates and write wait probabilities with systems employing secondary storage caching is significant. Fig. 11 shows the sensitivity of file mapping to the read hit and write wait probabilities. These are shown for two processor speeds, 3 MIPS and 7 MIPS. The sensitivity to read hit and write wait probabilities increases with processor speed, and there is a rapid drop in effective processor utilization for higher-speed processors as the read hit probability decreases and the write wait probability increases. This sensitivity to read hit rate and write wait probability decreases with reductions in disc access time. Thus faster drives enhance caching alternatives, and indeed are required if sufficiently high read hit rates and sufficiently low write wait probabilities cannot be achieved. The characteristic dependencies of the other caching alternatives on read hit rates and write wait probabilities are proportional to those presented here for file mapping.

Special mechanisms can be employed to achieve high read hit rates and low write wait rates. Since the secondary storage caching alternatives are sensitive to these parameters, it is important to exploit them.

Integrating cache fetch, replacement, and write handling with transaction management requirements is achievable with internal caching as described below. With external caching, on the other hand, this would be difficult to achieve, and high overhead would be incurred because of the protocols required between the host and the external device. Consequently, cache memory utilization and the sustained effective multiprogramming level of internal caching can be superior to that achievable with external caching.

With internal caching, the size of a fetched disc domain on a read miss can be tailored to the structure of the data, based on knowledge of the storage layout (e.g., fetch extents instead of fetching tracks containing unrelated data or only pieces of the required extent). The replacement policy can exploit operating system knowledge of access patterns (e.g., the policy can flush a cached disc domain from the cache memory after sequential reference and on file purging). Write posting order and write and read priorities can be adjusted to meet the current needs of transaction management.

Evaluation With Respect To Other Factors

Development of multiple local buffer management schemes is redundant, and the buffering capacity is localized and unresponsive to current memory loading conditions. The amount of memory devoted to the buffering of a specific file or subsystem would likely be either excessive or insufficient at any given moment, depending on the current memory availability and the current workload demand and priority structure. The problems are analogous to those encountered with older memory replacement policies based on fixed partitioning.

Global file buffers require a fixed partition of main memory between swap space and buffer space, and therefore are not responsive to dynamic memory load conditions. They suffer from fixed-partition problems as do local file buffers. Furthermore, they provide memory management functions, including space allocation, replacement, and disc access initiation and interrupt fielding. Therefore, developing efficient algorithms for the global buffer manager is redundant to the development of the manager of memory for transient code and data. Moreover, as is discussed in the next section, global file buffers are not amenable to supporting efficient local caching in decentralized systems.

Explicit internal disc caching and file mapping overcome the fixed-partition problems, so they can be expected to achieve higher read hit rates with dynamic workloads. Furthermore, it is easier to integrate disc access posting order and priority adjustments with these schemes.

Explicit internal disc caching in main memory requires a translation mechanism to locate a required cached disc domain in main memory. In large main memories there can be several thousand cached disc domains present at any moment, and locating a cached domain through a translation of disc address to main memory address can be expensive in processor cycles. Additionally, if the processor has a cache, the location sequence can flush a large portion of the cache while chasing through list structures. However, explicit internal disc caching can be performed with any processor architecture.

File mapping appears to offer the best of the alternatives for caching of discs in main memory. It leverages the main memory management mechanisms, has no limits on main memory space applied to file caching, and has no overhead for location of file domains in main memory. Applicability of this approach is limited, however, to processor architectures with large virtual address spaces ($\geq 2^{48}$ bytes to map many large files) and amenable protection mechanisms to restrict access to mapped-in files. Stonebraker[22] discusses problems with performing transactional management on mapped files with 32 bits of address space. His studies were based on the VAX family. He found that the overhead of mapping in and out pieces of files to accommodate the protection and space limitations exceeded the overhead of explicit data base buffering.

Evaluating with respect to cost, using faster discs and more of them to overcome the performance limitations without caching is clearly not cost-effective. External caching is more expensive than internal caching since additional power, cooling, cabinetry, and electronics are required in addition to the extra memory required for the

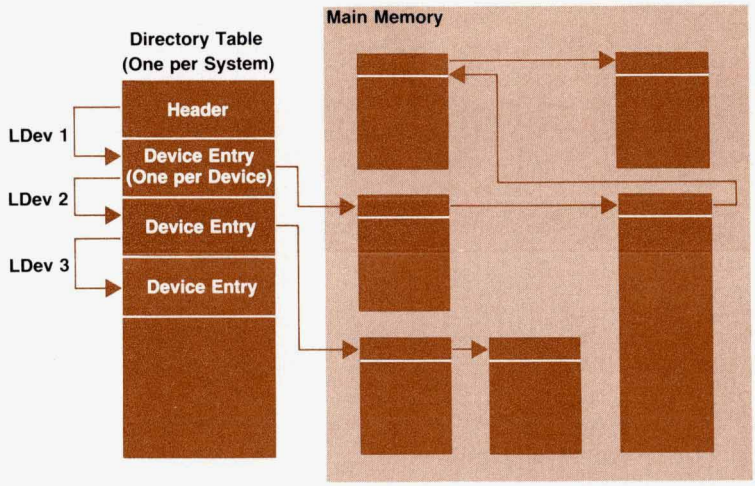


Fig. 16. *Cached disc domain location mechanism used in the breadboard MPE kernel. A separate list is maintained for each disc, identifying the memory regions corresponding to currently cached domains from that disc.*

caching. Also, more memory would be required to achieve the same hit rates in external caching unless the cache management is integrated with the operating system policies.

Evaluating with respect to reliability, internal caching introduces no new hardware components into the system. The reliability is identical to the uncached system, whereas any of the external cache architectures necessarily degrades system reliability by introducing additional hardware components. The software/firmware complexity of explicit internal and external caching is roughly the same, so reliability degradation because of cache management is comparable for these alternatives, while file mapping is simpler and more reliable. With external caching, the posting strategy of the peripheral cache is not integrated with the system posting strategy, so a consistent level of integrity is not guaranteed for transactional data base systems unless post order and wait for media update are observed, thereby impacting the write handling performance of peripheral cache alternatives.

Applicability to the HP 3000 Family

The research and development leading to the MPE-IV kernel[23] examined alternative algorithms for processor, main memory, disc, and semaphore management. This research focused on the interactions between algorithms managing these basic system resources, and determined an algorithm set for these resources that provides good performance through algorithm integration. The MPE-IV kernel is briefly described in the box on page 25. Above the management of these basic system resources, subsystems and applications manage data to provide extended system functionality. The data is kept in structured permanent and temporary objects including files, data bases, stacks, and heaps. Concurrent data access is managed through locking or versioning. Recovery from transaction aborts or system failures is handled through checkpointing, write-ahead logging, and roll-forward/roll-back recovery. To exploit the system price/performance potential offered by evolving technology, basic kernel main memory, disc, and processor management needs to be extended and integrated with subsystem and application data and recovery management. This was well demonstrated in our follow-on research to MPE-IV.

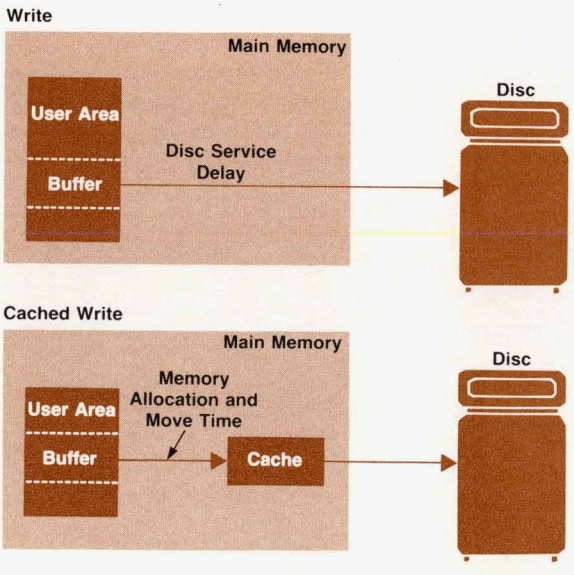


Fig. 17. *Disc write caching in the breadboard MPE kernel.*

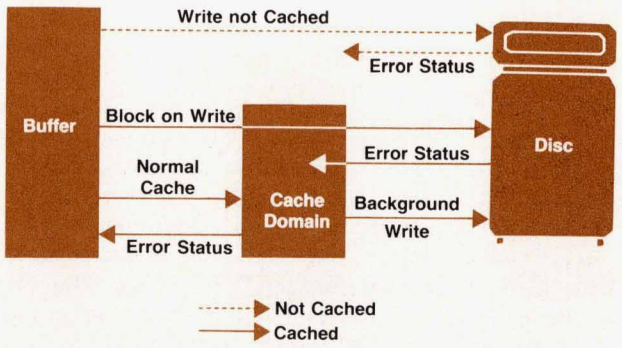


Fig. 18. *Write protocol for internal disc caching in the HP 3000.*

After the management strategies for main memory, disc, processor, and semaphore resources were integrated in MPE-IV to produce balanced resource utilization and significant performance improvement across the HP 3000 family, a new high-end system, the Series 64, was introduced. Its processor is twice as fast and its memory is four times as large as the previous top of the line. The system performance improvements realized with this computer were, however, sublinear with processor speed relative to the previous top-of-the line system. System performance with this high-end computer was found to be very sensitive to disc subsystem throughput and access times, but relatively insensitive to main memory capacity. In spite of the integrated strategies of the breadboard kernel, system performance was not scaling as the family was extended to significantly more powerful systems.

We investigated this scalability problem, and found it to be caused by a lack of scalability in the algorithms used for subsystem and application level buffering and recovery management. Processor utilization was limited by subsystems making processes wait for disc accesses to complete. The disc write accesses were being generated by local buffer replacement and write-ahead log management. The disc read accesses were initiated to resolve local buffer read misses.

We wished to extend the basic algorithm integration for processor, main memory, semaphore, and disc management to exploit current technology trends to realize significant improvements in system price/performance ratio. We examined disc buffering, intermediate storage levels between primary and secondary storage, disc caching in excess primary memory of the system processing unit, and integrating kernel resource management with higher-level data management.

Potential For Secondary Storage Caching

Our analysis, presented in the preceding sections, indicates that explicit internal caching is the best alternative for the HP 3000 family, given its architecture (object sizes must be ≤64K bytes). We proceeded to investigate whether

the workload characteristics were amenable.

Fig. 12 indicates the salient workload characteristics of the secondary storage referencing characteristics of the HP 3000 family. These characteristics were obtained using the measurement tools described in the box on page 23. Fig. 12a shows the transfer sizes for the major access methods (sequential, directory, data base). The distributions indicate that relatively small transfers are performed between primary and secondary storage by all the access methods, with 90% of the transfers less than 1K bytes. Fig. 12b shows the spatial locality of access method references to secondary storage. This stack depth distribution presents the probability of a reference at a given depth into a stack of disc references ordered on a least recently used basis. The distribution indicates that over 70% of all disc references are within the last 30 disc regions referenced. Fig. 12c shows the temporal locality of references of the access methods. It indicates that 50% of the disc references are to regions that have been referenced within the last second.

The resulting potential of secondary storage caching for the HP 3000 family is indicated in Fig. 13. This figure shows the read hit rates that can be achieved by the various access methods as a function of cache capacity. The data for this graph was obtained using the disc cache simulation model on disc access traces obtained from an internal HP production facility. For this installation, mean read hit rates of 85% are achievable with a few megabytes of memory applied to caching disc regions. Results using other trace data and field measurements of our system in use show read hit rates typically in the 65-85% range and read-to-write ratios ranging from 3:1 to 5:1.

The Explicit Internal Disc Cache Implementation

The MPE-IV kernel base was used to investigate principles and integrated approaches to caching discs in the main memory. Algorithm interactions were observed and improvements developed. Differences between disc and main memory caching and areas in which architectural improvements would be of benefit were noted.

The resource management mechanisms and strategies in the MPE-IV kernel provided an efficient, extensible research base. These strategies were extended to support explicit disc caching in primary memory. The resulting mechanisms and strategies integrate kernel and data management. Knowledge of file structure and access type is exploited to enhance prefetch and replacement decisions for disc domains. Data recovery protocols are supported without wait for posting through kernel adherence to post order constraints. Priorities of disc accesses are adjusted to reflect the changing urgency of read and write requests because of data management locks or commits. Process priorities are adjusted to reflect lock states.

The overall structure is shown in Fig. 14. The user program requests records. The file system maintains local file buffers, and on buffer misses or replacements, accesses the I/O system to initiate buffer reads or writes. Beneath the I/O interface, pieces of the disc memory are cached in main memory. Actual disc transfers are initiated by the cache manager.

Read Handling

Fig. 15 shows read processing with and without caching enabled. Without caching, the disc transfers data directly to the buffer, but a disc access delay is incurred. With caching, the disc transfers data to an area of memory reserved for the disc domain, and then the data is moved to the buffer using the processor. On read hits, data access requires locating the data in memory and moving it, rather than incurring a disc delay. This is performed on the current process's stack without a process switch.

Locating Cached Disc Regions

The cached disc domain location mechanism employed in the breadboard kernel is depicted in Fig. 16. A separate list is maintained for each disc that identifies the memory regions corresponding to currently cached domains from the disc. The list is ordered by increasing disc address, and a microcoded link-list-search (LLSH) instruction is used to locate a required region in the list.

This location scheme requires about 500 instructions for setup and cleanup, plus two memory references and two compares in the LLSH instruction for each cached domain in the list. Thus the overhead of translation increases with the memory size. This is not a particularly good feature. Thousands of domains can be cached for each disc in a large memory, so the overhead of translation can become significant. In hindsight, more attention should have been paid to the location mechanism. The overhead would be

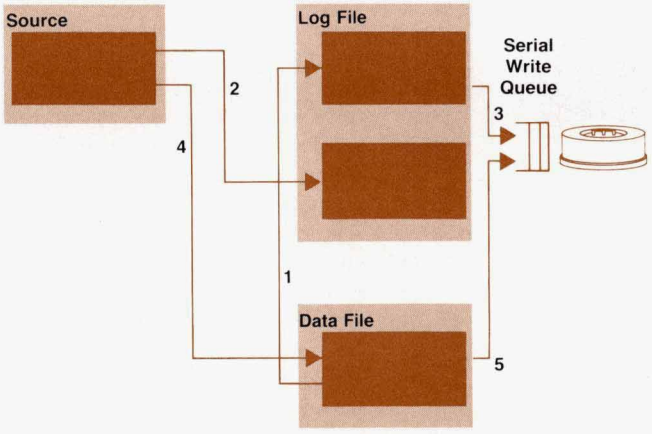


Fig. 19. *Sequence of operations for write-ahead logging with serial posting. This mechanism was developed to minimize the probability of waiting for disc writes to complete.*

reduced significantly by employing a hashing rather than a linked-list algorithm, but this is complicated by the variable size of cached disc regions. Architectures supporting file mapping in virtual space eliminate this overhead altogether.

Write Handling Mechanisms and Policies

Write handling with and without internal disc caching is shown in Fig. 17. Without caching, the transfer occurs directly to the disc from the buffer. The transfer can be initiated with or without wait, with explicit completion synchronization occurring later.

When a write is initiated with internal disc caching, either a disc domain already containing the specified disc addresses or a free area of memory is allocated and the data is moved into that area. A write to the disc is issued, and the process can continue to execute. The write is considered complete once the memory move occurs unless the file has been specified as requiring physical post before completion notification. Note that, thanks to the support of variable-sized segments in the HP 3000 architecture, a write does not require a fetch of the disc region. In paged systems, a write of a partial page requires that the page be fetched before the write can be initiated, thereby incurring an additional disc access on a write miss.

If there is currently a write pending against the specified disc domain, the process's request is queued until the pending write is posted to disc. The process may continue executing until it requests a wait for its write completion and the request is still queued. If the disc domain to be written is not currently cached, an available region of memory is obtained to map the corresponding disc image—i.e., no fetch of the disc domain to be written is required. When the move effecting the write takes place from the process's data area to the cached image of the disc, a post to the disc is initiated. Only the portion of the cached disc image that is modified by the write is posted. After the move to the disc image is performed and the post to disc is initiated, the writing process is allowed to continue running without having to wait for the physical post update to complete. This is all handled on the current process's stack, without even a process switch.

A write-through policy was chosen for our implementation of internal disc caching. The post request to disc is issued at a background disc priority. The priority of a pending post request is raised if the process waits for the post to complete or the region is required by the main memory replacement algorithm. Thus, issuing a physical post when the write is performed rather than waiting for replacement has no negative impact. Only idle disc capacity is used. Furthermore, pending writes of adjacent disc regions can be gathered into a single physical transfer, thereby reducing the total number of required physical accesses.

Write-through also has another benefit. The transaction recovery mechanisms require synchronization on physical commit at some point, so performing these early saves commit delays.

The write protocol is shown in Fig. 18.

We saw in the preceding sections the significant impact of having to wait for writes. To minimize the probability of waiting for disc writes to complete, we developed a special mechanism that allows the specification of posting order constraints at post time or on a file basis so that posting can proceed without wait if only a posting order constraint exists. With this facility, the post of the log can be issued on a nowait basis, and the data base write can be executed immediately. The kernel guarantees that posting order within a serial write queue matches the chronological order of post initiation. The sequence for write-ahead log use of this facility is depicted in Fig. 19.

To ensure disc consistency, only one write access for a serial write queue can be pending at a time. This is depicted in Fig. 20. Since this limits parallel use of available disc drives, multiple serial write queues for unrelated postings are beneficial.

Cache Fetch, Placement, and Replacement

The kernel's main memory placement and replacement mechanisms were extended to handle cached disc domains in the same manner as segments. Thus, cached disc domains can be of variable size, fetched in parallel with other segments or cached disc domains, garbage collected, and

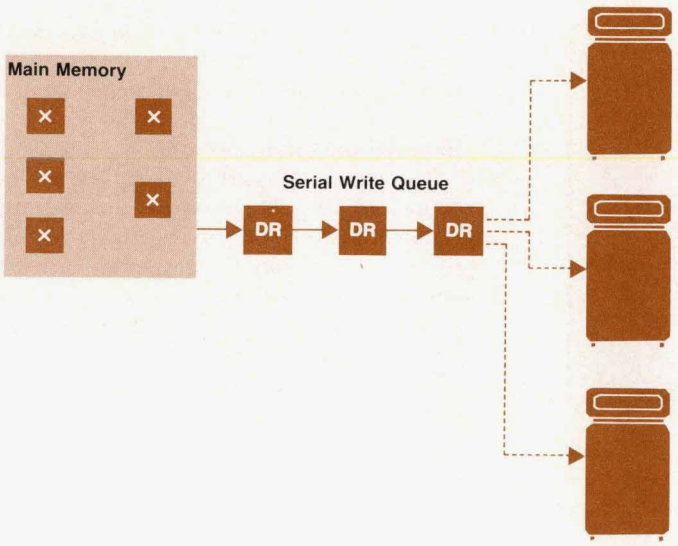


Fig. 20. *To ensure disc consistency, only one write access for a serial write queue can be pending at any time. Multiple serial write queues can be used to allow parallel use of available disc drives for unrelated postings.*

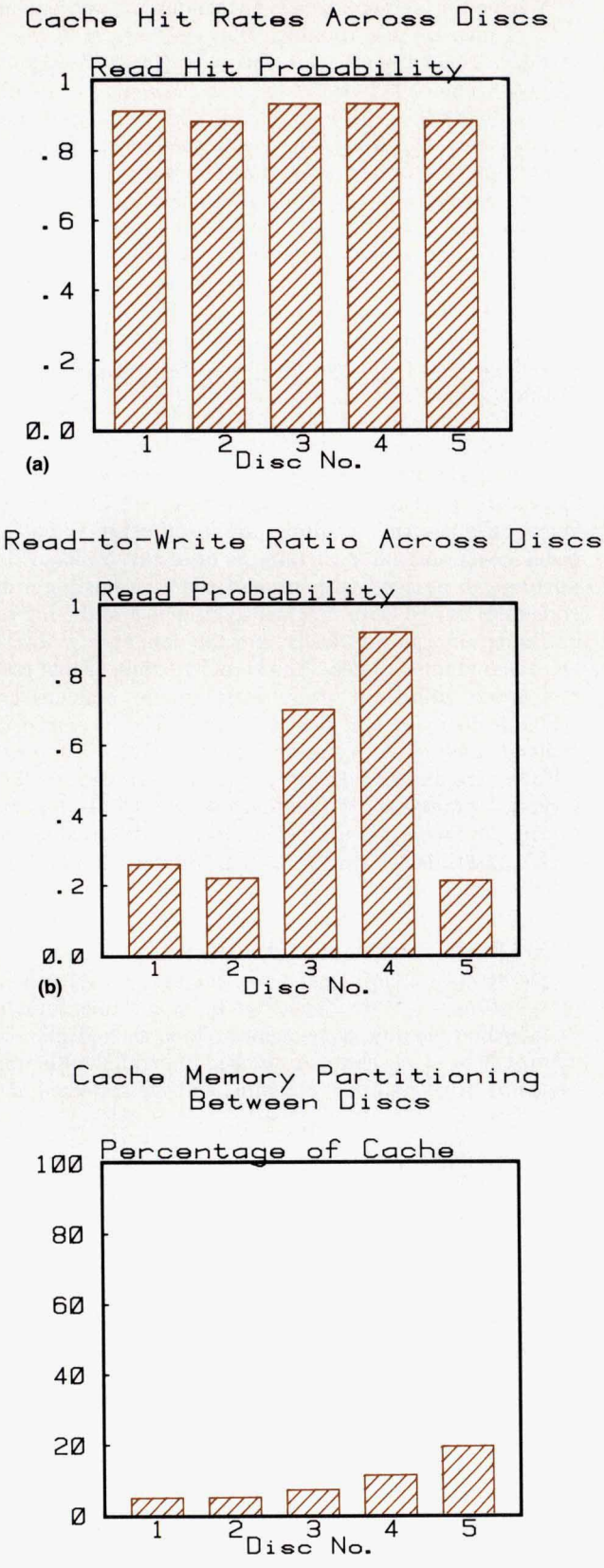


Fig. 21. *Dynamic partitioning of cache operations across the system discs of an HP 3000 system.*

replaced in an integrated manner with stacks, data segments, and code segments. The relative allocation of main memory between stack, data, code, and cached disc domain objects is entirely dynamic, responding to the current workload requirements and current memory availability.

Fetching and replacing differ from processor caching of main memories. Rao[24] discusses the impact of replacement algorithms on cache performance for processor caching of main memory, finding that the replacement algorithm has a secondary effect on cache performance.

With internal disc caching, when a request is made to read data that is not currently cached, the fetch strategy uses knowledge of file blocking, extent structure, access method, and current memory loading to select the optimal size of disc domain to be fetched into memory. The fetch is performed in an unblocked manner so that the requesting process or another process can run in parallel with the cache fetch from disc. With processor caches, the processor must idle on a cache miss since the process switch time exceeds the cache miss processing time.

No special treatment of cached domains is required for the replacement algorithm. It naturally protects the transient objects in smaller memories because of their smaller interreference times, and uses large memories for extensive write and read caching with the relative amount of cached disc space per device skewing naturally to the heavily referenced devices. This dynamic partitioning feature is very significant, since device skewing is common both in terms of referencing frequency and the spatial locality of references.

Fig. 21 shows the cache hit ratios, the read-to-write ratio, and the partitioning of main memory allocated for cached domains across the discs of an HP 3000 Computer System. The normal LRU-type replacement algorithm does fine in responding to the variable demand requirements of the various disc volumes.

Explicit prefetching and flushing for sequential access was found to improve hit rates through simulation studies on standard trace files, whereas special prefetching and flushing for other access modes did not.

This policy was implemented in the kernel. When a process finishes referencing a cached disc domain in sequential mode, the domain is flushed immediately from main memory since it won't be needed again. In this way, memory utilization is improved over that achievable with the kernel's standard LRU-type replacement algorithm.

External Caching Controls

The external controls for caching allow caching to be enabled/disabled against specific discs, display the current status of disc caching, set posting policies on a system and file basis, and control the roundoff fetch sizes for random and sequential access.

Defaults for the tuning parameters were selected based on simulations of disc access traces using the simulation model. Good defaults for random fetch sizes were found to be 4K bytes and for sequential fetch sizes, 24K bytes. Large prefetches were found to pay off significantly for sequential, but not for random type accesses. Rounding the fetching above the requested block was found to be superior for all access methods to fetching below the requested block

or centering on the requested block. The choice of tuning parameters is, as always, an adjustment to the access patterns of the particular subsystems and data bases.

Performance of the Disc Cache Implementation

With these mechanisms and strategies, the extended kernel significantly reduces the traffic between the main memory and secondary disc storage and significantly reduces delays in reading or writing disc information. Read hit rates up to 85% are common for file, data base, and directory buffer fills. These read hits eliminate up to 65% of the disc accesses with a 5:1 read-to-write ratio. Because of the caching of writes, most delays for posting are eliminated. Together, the read hits and cached writes eliminate 90% of process delays caused by disc accessing. This dramatically reduces semaphore holding times, which especially benefits data base systems.

The impact on system performance over the noncached kernel is shown in Fig. 22. Throughput improvements of 50% and response time reductions of 5:1 are standard on

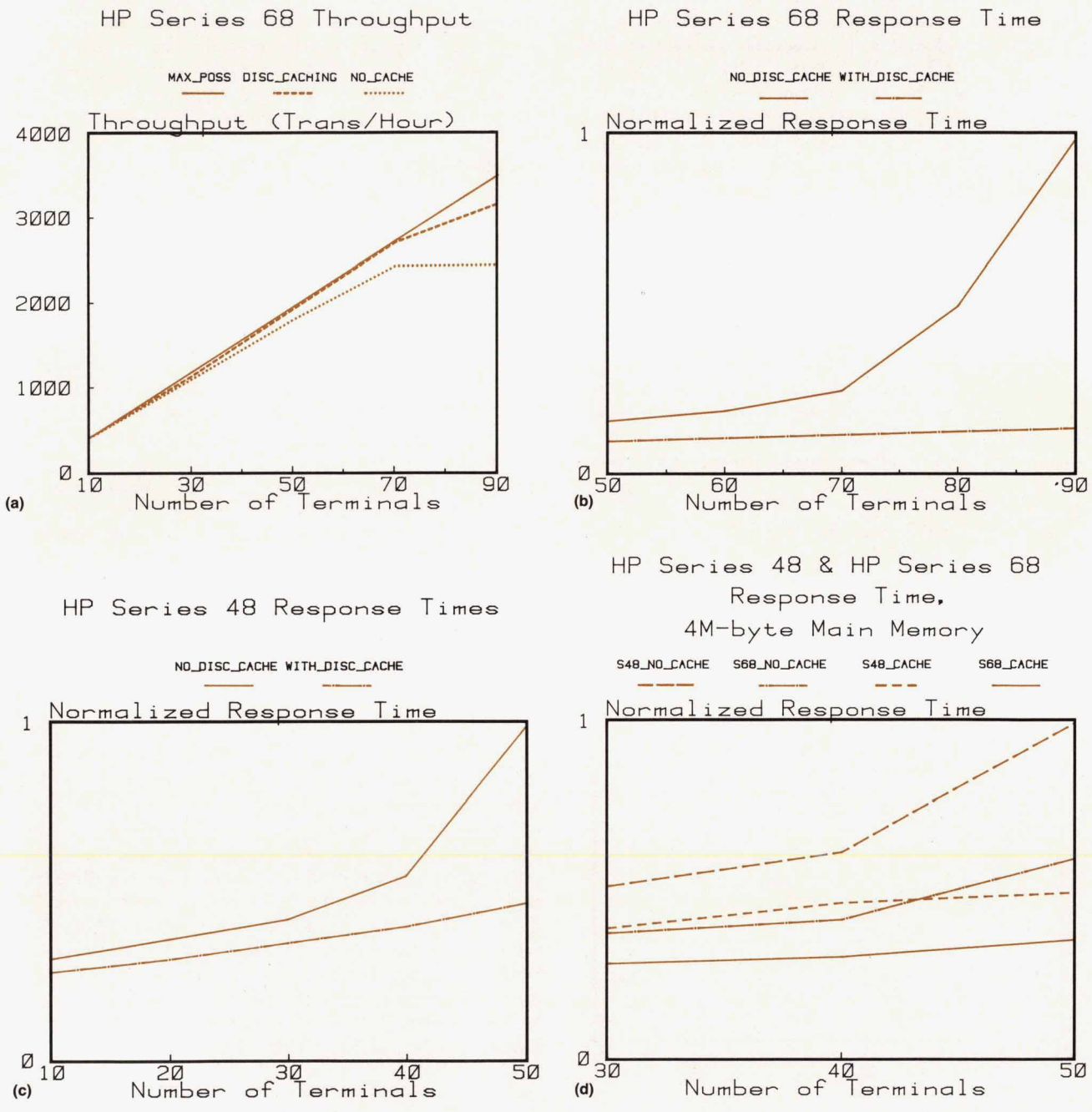


Fig. 22. *Performance of HP 3000 Series 48 and 68 systems with and without disc caching. Throughput improves most on the larger Series 68 system with caching.*

the high end (Series 68) while the midrange system (Series 48) gets about 25%. For the low-end system (Series III or 33, not shown in Fig. 22), performance actually degrades, thus demonstrating the scaling of performance with processor speeds. However, the midrange system with kernel disc caching outperforms the high-end machine without kernel disc caching.

Fig. 23 compares the kernel with and without caching for configurations with four 125M-byte discs or two 400M-byte discs. The cached system with two disc servers outperforms the uncached system with four disc servers. Moreover, the cached system performance is the same with two or four discs, thereby demonstrating that caching can exploit the system cost advantages of large-capacity discs.

A benchmark comparison[25] was made between an IBM 3033 (5 MIPS) and an HP 3000 Series 68 (1 MIPS) with and without applying disc caching in main storage. The benchmark ran Compserv's Materials Request and Planning software developed for both the HP 3000 and the IBM 3033 systems. The benchmark consists of a materials request and planning batch job that performs a large data base implosion to determine the parts, lead times, and orders to manufacture a specified set of products subject to a master production schedule. This type of application tends to be highly I/O intensive since it accesses the data base in an inverted manner, starting with schedules and searching for qualifying parts.

The benchmark ran in 28 hours on the IBM 3033, 49 hours on the HP 3000 Series 68 without kernel disc caching, and in 27.4 hours on the HP 3000 Series 68 with kernel disc caching. That the HP 3000 Series 68 could outperform the IBM 3033 in spite of the fivefold difference in processor

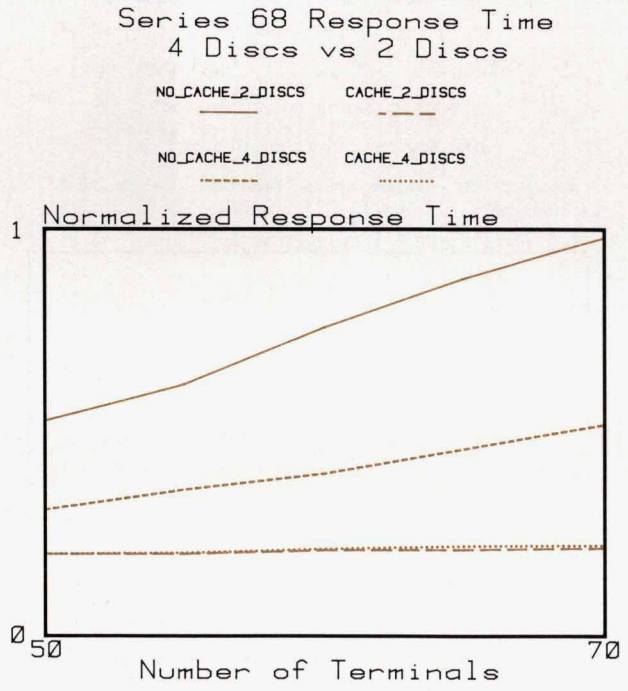


Fig. 23. *Effects of multiple discs and disc caching in the HP 3000 Series 68. Sensitivity to the number of discs is dramatically reduced by caching.*

speed indicates that caching by the kernel of disc domains in excess main storage with the cache management policies introduced in this research has potential application in other computer families employing locally managed, limited caching of data items by data base or file systems. Data base and file system caching is limited to a fixed capacity and applies policies that optimize for the standard access approach. When more resources are available, such as main memory for a stand-alone batch job, and access is nonstandard, as in this inverted-access case, localized caching policies do not respond.

In small memory systems under memory pressure and with slow processors, disc caching degrades performance. The cost of locating the cached disc domain in main memory and moving the domain gets added to the disc access time. With low hit rates and slow processors, this overhead exceeds the benefits of caching the discs. This overhead is not present in architectures supporting file mapping.

Conclusions

This research and development effort examined alternative approaches to exploit current trends in processor and memory technology to realize significant improvements in system price/performance ratio. The results were applied to the HP 3000 family.

Caching secondary storage in primary main memory through explicit internal caching or file mapping, coupled with integrated kernel and data management algorithms, was found to provide the most cost-effective and best performing alternative, given current technology trends. These approaches are supplemented by improvements in secondary storage device characteristics.

Secondary storage cache management can be integrated with data management to obtain significant incremental performance improvements. Specific improvements were identified by integrating write-ahead logging with disc access scheduling through the adherence to posting order constraints, bumping the priority of affected disc post requests at commit time, and using extent boundary and access type knowledge for selecting the fetch size on read misses and flushing on sequential access.

Effective utilization of high-speed processors requires sustaining a high multiprogramming level. Improved concurrency control schemes, such as late binding, granular locking for updates, and versioning for queries, help achieve this. Data base locking can be integrated with kernel resource management by raising the priority of a lock holder when a more urgent process queues on its lock.

Providing caching of discs in main memory differs from caching main memory for processors in several ways. The delay in resolving a read miss is highly state-dependent, requiring disc queueing and head positioning components. Since the read miss resolution time is long compared with the time required to state save the current process and launch another process, other ready processes can be run if the effective multiprogramming level is sufficiently high, allowing productive processor utilization while resolving read misses. Special knowledge of access patterns can readily be exploited to improve cache fetch and replacement strategies. Write misses can be resolved without requiring a fetch from the backup store if variable-sized allocation

is used, thereby using large caches to eliminate process waits on write misses. Special constraints exist on the write handling policies for updating the backup memory because of the requirements of write-ahead logging and shadow paging in transactional systems. Classical techniques used to model processor caches did well in analyzing hit rates, but more complex models were required to capture queueing and nowait effects found in secondary store caching.

The implementation of explicit internal disc caching for the HP 3000 family graphically demonstrates the results the analysis predicts. The effective utilization of higher-speed processors and reduced dependency on the number of discs is clearly demonstrated. The implementation demonstrates further that integrated internal caching of secondary storage can be accomplished with conventional hardware and software architectures. Disc data management integration with internal cache management can be achieved in existing systems by providing special functions to allow subsystems and applications to influence the management of caching, disc scheduling, and main memory management to meet their performance and recovery objectives.

We feel that the major contributions of this research and development effort are the performance analysis techniques, the characterization of secondary storage caching alternatives, and the mechanisms we developed to integrate the requirements of high-level data management with the basic kernel management of processor, disc, primary memory, and semaphore resources. Our leveraging of kernel mechanisms for transient code and data management resulted in reduced development time, highly parallel cache management, and a smooth, dynamic partition of main memory. Our extensive testing (over a year) in varied internal production sites resulted in a high-quality product. However, our implementation of the location mechanism is inefficient for large memories, and further work could be done on data bases and applications to exploit more fully the underlying caching.

The performance advantages of secondary storage caching will be needed in decentralized as well as centralized transaction systems. Caching secondary storage in decentralized systems introduces problems of cache consistency analogous to those encountered in multiprocessors with private caches. Analyzing and building solutions to realize the performance benefits of caching secondary storage in decentralized configurations presents us with a challenging follow-on to the applied research reported here.

Acknowledgments

We'd like to acknowledge our R&D manager Howard Smith and our marketing manager Nancy Anderson for recognizing the potential of our project and supporting its delivery. Our thanks go to Jeff Byrne, our product manager, and to Sam Boles and Becky McBride of the Business Development Group, who were really helpful in providing ongoing encouragement and support.

References

1. J.R. Busch and A.J. Kondoff, "MPE Disc Cache: In Perspective," *Proceedings of the HP 3000 International User's Group Conference*, Edinburgh, October 1983, reprinted in *Journal of the HP 3000 International Users Group*, Vol. 7, no. 1, January-March 1984, pp 21-24.
2. A.E. Bell, "Critical Issues in High-Density Magnetic and Optical Data Storage," *Laser Focus/Electrooptics*, August 1983, pp 61-66 and September 1983, pp 125-136.
3. R. Roseberg, "Magnetic Mass Storage Densities Rise," *Electronics*, October 29, 1984.
4. M.J. Flynn, J.N. Gray, A.K. Jones, K. Legally, H. Opderbeck, G.J. Popek, B. Randell, J.H. Saltzer, and H.R. Wiehle, "Operating Systems : An Advanced Course," *Lecture Notes In Computer Science*, Springer-Verlag, 1978.
5. W.W. Chu and P.P. Chen, *Tutorial : Centralized and Distributed Data Base Systems*, IEEE Computer Society.
6. H.T. Kung and J.T. Robinson, "On Optimistic Methods for Concurrency Control," *ACM Transactions on Database Systems*, Vol. 6, no. 2, June 1981, pp 213-226.
7. R. Agrawal, M.J. Carey, and D.J. DeWitt, *Deadlock Detection Is Cheap*, Draft, Computer Science Department, University of California, Berkeley, 1983.
8. A. Chan and R. Gray, "Implementing Distributed Read Only Transactions," submitted for publication, 1983.
9. A.J. Smith, "A Modified Working Set Paging Algorithm," *IEEE Transactions on Computers*, Vol. C-25, no.9, September 1976, pp 907-913.
10. "Series/1 Gets Top CPU, Extended Networking; System/38 Also Capped," *Computerworld*, Vol. 17, no. 15, April 1983.
11. U. Krastins, "Cache Memories Quicken Access to Disk Data," *Electronic Design*, May 1982, pp 41-44.
12. H.R. Crouch and J.B. Cornett, Jr., "CCDs in Memory Systems Move into Sight," *Computer Design*, September 1976, pp 75-80.
13. G. Panigrahi, "Charge-Coupled Memories for Computer Systems," *Computer*, April 1976, pp 33-42.
14. H. Chang, "Capabilities of the Bubble Technology," *Proceedings of the National Computer Conference*, 1974, pp 847-855.
15. T.C. Chen, "Magnetic Bubble Memory and Logic," *Advances in Computers*, Vol. 17, 1978, pp 224-282.
16. W. Hugelshofer and B. Schultz, "Cache Buffer for Disk Accelerates Minicomputer Performance," *Electronics*, February 1982, pp 155-159.
17. M. K. Mc Kusick, S.J. Leffler, and W.N. Joy, *The Implementation of a Fast File System for UNIX*, Computer Systems Research Group, Department of Electrical Engineering and Computer Science, University of California, Berkeley.
18. "Cache/Q by Techne the Software Accelerator," Techne Software Corporation, advertisement, 1983.
19. E.E. Organik, *The Multics System*, MIT Press, 1972.
20. E. Basart, D. Folger, and B. Shellooe, "Pipelining and New OS Boost Mini to 8 MIPS," *Mini-Micro Systems*, September 1982, pp 258-269.
21. J.N. Gray, "Notes on Data Base Operating Systems," *Operating Systems: An Advanced Course*, edited by R. Bayer, M. Graham, and G. Seegmuller, Springer-Verlag, 1979, pp 393-481.
22. M. Stonebraker, "Virtual Memory Transaction Management," Memorandum No. UCB/ERL M83/74, Electronics Research Laboratory, University of California, Berkeley, December 19, 1983.
23. J.R. Busch, *Integrated Resource Management Algorithms for Computer Systems*, Ph.D. Dissertation, University of California at Los Angeles, 1984. University Microfilms International, Ann Arbor, Michigan, no. 8420156.
24. G.S. Rao, "Performance Analysis of Cache Memories," *Journal of the Association for Computing Machinery*, Vol. 25, no. 3, July 1978, pp 378-395.
25. "Series 64 With Disc Caching Beats IBM 3033 in Batch MRP Run," *Hewlett-Packard Computer News* (internal publication), October 1, 1983.

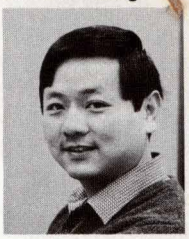
Authors

February 1985

Elaine C. Regelson

Elaine Regelson has been a member of the HP technical staff since 1981. Before working on HP TechWriter, she did investigation and development work on various technical-computer-aided work tools. Before joining HP she was a project manager for advanced speech recognition and generation training systems for air traffic controllers. A 1972 graduate of the University of California at San Diego, she holds a BA degree in biology and has taught computer science, math, and English and Scottish country dancing. Elaine was born in Santa Monica, California and now lives in Fort Collins, Colorado. She's married and has three stepchildren. Besides folk and traditional country dancing, her interests include singing, gardening, and playing the violin. She serves on the advisory board of the Colorado State Science Fair and as a volunteer at a community crisis and information center.

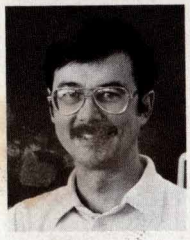
Kok-Wai Chang



Born in Hong Kong, Bill Chang studied physics at National Chung King University, from which he earned a bachelor of science degree in 1977. He continued his studies at Cambridge University and Texas Christian University, and received degrees from both institutions, including his PhD in 1982. Before coming to HP Laboratories in 1984 Bill was a research associate at the University of Texas at Arlington. His work at HP has focused on magnetostatic-wave devices. He is also the author of a number of papers on molecular spectroscopy and MSW devices. He is now a resident of Sunnyvale, California, is married, and has one son.

4 — HP TechWriter

Roy E. Anderson



Before contributing to the development of HP TechWriter, Roy Anderson helped define HP's Command Set 80 and Logical Interchange Format for mass storage peripherals, developed the CS-80 magnetic tape interface and utilities for the HP 9845 Computer, and helped implement CS-80 on HP's Shared Resource Management network. Born in Seattle, Washington, he received his BS degree in industrial engineering in 1971 from Montana State University and his MS degree in computer science in 1975 from the University of Oklahoma. Before joining HP in 1979, he served four years in the U.S. Air Force, where he did communications computer programming and attained the rank of captain. He has also done operating systems design for a large computer firm, taught programming at Wichita State University, and authored an AFIPS conference paper on software engineering. Roy is married, has two children, and lives in Fort Collins, Colorado. He enjoys skiing, woodworking, camping, and "just being in the mountains."

10 — Magnetostatic-Wave Devices

Waguih S. Ishak

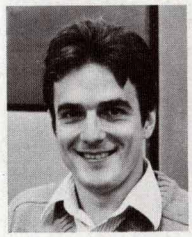
Waguih Ishak was born in Cairo, Egypt and earned degrees from Cairo University (BSEE 1971) and Ain Shams University (BS mathematics 1973). He also studied at McMaster University, from which he received his master's and PhD degrees in electrical engineering (1975 and 1978). After joining HP in 1978, Waguih worked at HP Laboratories on bubble memories and SAW filter design, and then became involved in magnetostatic-wave (MSW) devices. Presently, he is project manager of the Sources and Signal Processing Group. He is the author of more than 20 papers on magnetic bubbles, numerical optimization, SAW devices, and MSW devices. Waguih now lives in Cupertino, California, is married, has two sons, and likes to play soccer.

21 — Disc Caching

John R. Busch

John Busch has been with HP since 1976. Since joining HP, he has been involved in commercial system research and development. He is currently a section manager with HP's Information Technology Group. John holds BA and MS degrees in mathematics and a PhD in computer science from the University of California at Los Angeles. He is married, has one child, and enjoys tennis and skiing.

Alan J. Kondoff



Alan Kondoff has been with HP since 1976. Before entering into systems development in 1981, he was a systems engineer in the field. He is currently a project manager in the Information Technology Group laboratory. Alan holds a BS degree in electrical and computer engineering from the University of Michigan. He is married, has two daughters, and enjoys boating, water and snow skiing, and swimming.

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

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

HEWLETT-PACKARD JOURNAL

February 1985 Volume 36 • Number 2

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 subscribe, change your address, or delete your name from our mailing list, send your request to Hewlett-Packard Journal, 3000 Hanover Street, Palo Alto, CA 94304 U.S.A. Include your old address label, if any. Allow 60 days.

5953-8532