A HIGH-SPEED INTEGRATED CIRCUIT SCRATCHPAD MEMORY

I. Catt, E. C. Garth and D. E. Murray

Motorola, Inc., Semiconductor Products Division
Phoenix, Arizona

INTRODUCTION

Computer systems are presently being designed and fabricated using one- to two-nanosecond current-mode logic gates. High-speed scratchpad memories are required in order to utilize this circuit speed effectively.

This paper describes an integrated circuit memory containing 64 words of 8 bits per word, which is compatible in respect to both speed and signal level with high-speed current-mode gates. The memory has a nondestructive read cycle of 17 nanoseconds and a write cycle of 10 nanoseconds without cycle overlap. This is considerably faster than previously reported integrated circuit memories.1-4

In addition to high speed, a large degree of system flexibility is achieved by using an integrated storage flip-flop of the type described in this paper. Multiple, independently addressed read channels can be included in the memory allowing simultaneous access to more than one storage location. Separate write address decoding allows a write operation to take place at the same time that one or more read operations are occurring. Possible applications of simultaneous read and write operations will be described later, along with certain constraints.

A description of the device, circuit, package, and system design required to implement a memory of the type referred to above is contained in the following sections.

SYSTEM ORGANIZATION

A block diagram of the memory system is shown in Fig. 1. As can be seen, the memory has a dual read capability, i.e., any two storage locations can be addressed and read simultaneously. The address decoding is performed in two levels of logic to limit the fan-out to eight and the fan-in to three.

The memory is designed for a particular application and operates in what is referred to as a read-decision-write cycle. This means that following a read operation, a decision is made whether or not to write predetermined data into a predetermined address.

Figure 2 contains a timing diagram for the memory system. In a read-decision-write cycle, an A-Channel and/or B-Channel read address, a write address, and write data are presented to the memory prior to the receipt of a read initiate A or read initiate B pulse. The read initiate pulse starts the memory cycle by clocking the read address into the read address register. The address is decoded and the data contained in the selected word gated onto the output lines where it is then OR'ed to the output data register. The appropriate delayed read initiate pulse clocks the information into the output data.
register. The information is then sent to the computer which takes 15 nanoseconds to make the decision of whether or not to write. If the computer decides to write, a write command pulse is supplied to the memory. The write command combines with the write address, which is clocked into the write address register at the appropriate time by a delayed read initiate pulse, and allows the write operation to proceed. The write data is also clocked into the write data register by the same delayed read initiate pulse.

As the timing diagram indicates, the written information is not available to be read from the memory until the third cycle. In the particular application for which the memory is designed, it is desirable to have access to the written information on the second cycle. This is accomplished by providing circuitry external to the memory which bypasses the memory whenever a read and write operation occur in the same storage location. The external circuitry contains registers which store the read and write addresses and the new data being written into the memory. When the read and write addresses match during a cycle when a write command is supplied to the memory, the memory output is inhibited and the new data is gated out of the external register and used in place of the memory output. With the external bypass route, the memory performs a full read-decision-write cycle in 17 nanoseconds, the read cycle time.

The block diagram of Fig. 1 indicates a master clear capability. The master clear circuitry can clear the entire memory to ZEROS in 10 nanoseconds. This feature is provided only for the particular application, and it is not required to clear a word to ZEROS prior to writing, since writing is a jam transfer operation.

The output data fan-in block of Fig. 1 is the 64-input OR required in each of the eight bit posi-
tions to connect the output from all 64 words to the single output data register flip-flop. The method of implementing this OR is discussed later.

The write command fan-out and write data fan-out blocks are simply extra stages of logic to limit the required fan-out on any node to eight.

MEMORY CIRCUITS

Memory Bit Circuit

Two memory bit circuits such as those shown in Fig. 3 are interconnected on each integrated circuit die to form one bit of two different words. All inputs and outputs are directly compatible with current-mode logic gates. Each bit consists of a gated flip-flop which drives two current-mode read gates.

One of the gated flip-flops is made up of transistors T₁, T₂, T₃, T₄, T₅, T₆, T₈, T₉. Transistors T₈ and T₉ form a current-mode gate which operates one diode drop below normal logic levels. The base of transistor T₈ is connected to the bias voltage, V_BB, through the diode formed by transistor T₁₀, and the base of transistor T₉ is connected to the word write input, WW₁, through the emitter-follower level shift formed by transistor T₇. In the quiescent state, WW₁ is low and transistor T₉ conducts the current down through transistor T₂ or T₅ depending on the state of the flip-flop. During a write operation, the WW₁ input is pulsed high and the current is switched into transistor T₈. This causes either transistor T₁ or T₇ to conduct depending on whether the bit or bit input is high. A ONE is written when the bit input is high and a ZERO when the bit input is high. At the end of the WW₁ pulse, the current is switched back into transistor T₉ and the flip-flop is latched.

Figure 3. Schematic of memory bit circuit.
in the desired state. The storage flip-flop and the two AND functions required to write selectively into any bit location are thus achieved at the expense of only one unit of circuit power. The delay from the word write input to the slowest output of the flip-flop is approximately 1.7 nanoseconds resulting in a power-speed product of 100 picojoules for the entire function.

It should be mentioned that the bit input can be connected to \( V_{BB} \), thus reducing the number of inputs by one. A ONE is then written when the bit input is high and a ZERO when the bit input is low. The use of single-ended inputs reduces the complexity of the printed circuit board layouts considerably. This is not done in the present system because data taken on a discrete component breadboard indicates that one side of the flip-flop is significantly slower when single-ended inputs are used. This effect is largely due to stray capacitance in the breadboard, however, and the monolithic circuits show very little difference in switching times between single-ended and complementary inputs. Future systems will be constructed using single-ended inputs.

Transistors \( T_{11} \) and \( T_{12} \) are used to implement the clear function. When the clear input is pulsed high, the current is switched from transistor \( T_9 \) into transistor \( T_{11} \) thus forcing the flip-flop into the ZERO state. It should be noted that if the collector of transistor \( T_{11} \) were connected to another set of bit and bit transistors, a dual write function would be achieved. The delay from the clear input to the flip-flop output is the same as for the write delay, approximately 1.7 nanoseconds.

The read gates are current-mode gates which are capable of driving a 50-ohm load with a delay of 1.5 nanoseconds. Corresponding outputs from the two bits are connected together on the die to produce a tied-emitter OR. Eight such groups are then fed into an 8-input OR gate to achieve the desired 64-input OR. The 8-input gate drives the output data register.

Figure 4 shows a photomicrograph of an integrated circuit die containing two complete memory bits. Double layer metallization is used to provide the interconnections on the die. The die is 50 by 55 mils in size.

A single die is placed in a standard 14-lead flatpack and all leads are used. The total power dissipated in the memory bit package is approximately 275 milliwatts. The thermal design of the memory is discussed in the packaging section.

Register Circuit

The register circuit is a simple modification of the memory bit circuit, and a variation of the same integrated circuit mask set is used to produce both. Two register circuits like those shown in Fig. 5 are contained on each die. As can be seen, the modifications consist of deleting the read gates, eliminating the clear input, and altering what is now the clock input. The resistor values are also changed to allow the circuit to drive a 50-ohm load. The change in resistor values tends to increase the power dissipation; however, the elimination of the read gates reduces the overall circuit power substantially. The total power per die is approximately 150 milliwatts. As before, one die is placed in each 14-lead flatpack. A photomicrograph of a register circuit die is shown in Fig. 6. Two-layer metallization is again used to perform the desired interconnections.

The modifications to the clock input were made to allow tapping of transmission lines with a minimum of reflections. The flip-flops in any particular register are usually grouped in a cluster. If all the flip-flops are connected to the end of a long transmission line, a large reflection results from the capacitive loading. Figure 7 shows how the emitter-follower level shift transistor can be used to buffer the flip-flops from the line. Only one emitter follower is connected to the line and its emitter is used to drive all of the flip-flops in the register.

Gate Circuit

All of the decoding and fan-out trees in the memory are implemented using three-input current-mode gates like the one shown in Fig. 8. Two gates are included on each die as the photomicrograph of Fig. 9 indicates. Only three of the four input transistors shown are actually utilized, and a single layer of metal is used to interconnect the circuit. The gates are capable of driving a 50-ohm load with a typical delay of 1.35 nanoseconds. One die is bonded in each 14-lead flatpack and the total power per package is approximately 100 milliwatts.

PACKAGING

Since a signal experiences a delay of one nanosecond when traveling a distance of 6 inches down a transmission line on a printed circuit board made of epoxy glass, the physical layout of the system is
carefully designed to minimize line lengths. In the final design, a distance of about 18 inches is traveled in each of the 17-nanosecond read cycles, resulting in a time loss due to signal propagation of some 3 nanoseconds compared to 14 nanoseconds of delay through circuits. This gives a ratio Line Delay/Circuit Delay $= \frac{3}{14} = 0.21$, and makes the ratio of line delay to total delay $\frac{3}{17}$ or 0.18. This is quite good since a reasonable ratio for line delay to circuit delay is up to $\frac{1}{3}$, or 0.33, making the ratio of line delay to total delay $\frac{1}{4}$ or 0.25. It should not be much higher for the following reasons:

1. Inasmuch as faster circuits are more expensive, it is inefficient to pay for circuit speed which is subsequently lost in interconnections.

2. The time taken for reflections to damp-out in a transmission line increases with line length. If lines are long compared to circuit delays, line characteristic impedances and line terminations have to be more accurate, resulting in extra cost.

3. The problem of mismatch due to branches in a transmission line is greater with longer lines.

**Basic Packaging Philosophy**

In a memory, the worst-case delay is paramount. Nothing is gained by making access time to one memory word less than access time to the worst-case memory word. A rule was formulated which
was of great help in organizing the physical layout of the memory:

The fastest memory layout is that where access time to every word is the same, i.e., every path is a worst case delay path. This approach avoids timing problems by eliminating signal skew.

A 2-dimensional, 64-location storage array is represented in Fig. 10a. Assume that the word-select information enters the array at A and the output data leaves the array at B. If it is further assumed that signal propagation is restricted to the X and Y dimensions (no propagation along diagonals), it is clear that the distance from A to any storage location to B is a constant. For example, the length of path ACB is equal to the length of path ADEFB.

It is required that points A and B be on one connector at one edge of the memory. To meet this requirement, the memory array of Fig. 10a is folded at the centerline, GH, putting points A and B on opposite ends of a rectangle as shown in Fig. 10b.

As shown in Fig. 10c, there are always two paths between two points such as A and E. In address decoding, half of the address lines are sent each way and the final level of decoding performed in a two-input AND gate located at the memory word. This is similar to the well-known coincident current decoding used in magnetic core stores. Write command and write data can travel either route from A to E equally well. Likewise, output data can take either route from E to B.

The foregoing refers only to the layout of the memory storage elements and the second level of decoding. Consideration is now given to the addition of address registers and first-level decoding required at the input to the memory and the output fan-in and registers required at the output of the memory. Figure 10d shows the location of the address registers, first-level decode, output fan-in, and output data registers. A typical set of paths through the memory is as follows:

1. Read initiate enters at the left edge of the memory.
2. Half of the first level decode information (X-decode) is sent up the left edge of the memory and horizontally across the array.
3. Half of the decode information (Y-decode) is sent vertically up through the array.
4. The output data is sent to the output fan-in gates which then drive the output data register.

**System Implementation**

To achieve equal path lengths from any word in the memory array to the output, it is necessary to send the output data from all storage locations to the far end of the card in the Y dimension, and from there back to the near end. By making the Y dimension very short, it is possible to simplify the wiring by not sending the output data to the far end of the array. Signal skew is still very small as long as the difference in path lengths is small compared to the total line delay plus circuit delay. This is achieved by placing the memory circuits on small cards in the third (Z) dimension. Making the cards removable from the mother card results in ease of maintainability.

The system packaging dimensions are somewhat fixed by a requirement that the memory be compatible with an existing larger system. Two of the overall system dimensions are limited to 4.5 and 6.875 inches, and cooling is specified as forced air.
A detailed discussion of cooling follows in another section.

A convenient division of the system into multiple groups of identical logic results in 12 multilayer printed logic cards interconnected by plugging into the 13 layer mother card as shown in Fig. 11. The mother card consists of seven signal and six voltage planes to form interconnections between the logic card connectors spaced on 0.2-inch centers.

The basic printed logic card, measuring 1.6 × 6.5 inches, interconnects three rows of up to 16 ceramic flat packages. This arrangement, shown in Fig. 12, results from the compatibility requirement and the ease with which the memory logic divides into sections. Three signal and three voltage planes are required. The inner signal plane is designed to be 55 ohms characteristic impedance, while the surface conductors are designed to be 75 ohms. All critical path logic is routed, where possible, on the higher impedance surface conductors to reduce the circuit input capacitance loading effect. On the long lines, the distributed loading reduces the effective impedance to approximately 50 ohms. Conductor widths...
are 6 mils on the inner plane and 8 mils on the outer planes. The tolerance on the conductor widths is ±½ mil. The overall card thickness is 50 mils. Each circuit mounting pad has a plated-through hole to serve as a signal crossover when necessary and to add mechanical strength. The leads on the packages are bent flush with the bottom of the package and then reflow soldered on top of the multilayer card. This allows the packages to be mounted and replaced very easily.

Since memory is highly repetitive, division into sections can be made in many ways. The memory elements are divided into eight groups of eight words, each group being contained on one of the small logic cards. The 64 bits on each card are sub-divided into four groups as shown in Fig. 13.
Each subgroup contains two words and the last level of read, write, and clear logic for both words.

Three other card types of like dimensions contain the input/output circuits. In keeping with the previously mentioned layout philosophy, the input logic enters the matrix of memory logic from one corner while all outputs are taken from the opposite corner. Therefore, when viewing the logic cards from the package side, the input circuits are located on the leftmost side while the outputs are on the right side. Logic for these three card types divides naturally into read input/output, write address, and write data.

The read input/output card contains the read address register, the first-level read decode, the output data fan-in, and the output data register. Two identical read input/output cards are required because of the dual read channel requirement. Timing for enter-

Figure 10. Two-dimensional storage array.
ing output data into the output registers is obtained by suitably delaying the read initiate signal by the same time required to access the memory. This is accomplished through use of binary weighted lengths of printed circuit delay line suitably connected in series to form any desired delay to the nearest one-half nanosecond. This can be seen on the artwork for the inner plane of the read input/output card shown in Fig. 14.

The write address card contains the write address register, the first-level write decode, the first two levels of write command fan-out, and the first two levels of clear fan-out. As in the case of the read input/output card, the write address logic card has its inputs on the left side to equalize the length of all write paths. Write address information is entered under control of the read initiate signal. For proper timing, this signal is delayed through use of variable-length printed circuit delay lines.

The write data logic card contains the write data register and the write data fan-out gates. Each data bit enters a register located in the card's center and is then distributed to gates located at points across the card for an even distribution to the memory circuits.

In the preceding discussion, the many practical considerations leading to the exact dimensions of the overall memory package are described. It should be mentioned that a more analytical approach can be taken to determine if these dimensions are close to the values required to produce the minimum total path length through the memory. For minimum path length, an optimum ratio exists between the X and Y dimensions on the mother card (refer to Fig. 10d)

and the Z dimension on the plug-in logic cards. By relating both the total path length and the area required to contain the circuits to the X, Y, and Z dimensions, it can be shown that the optimum ratio of $X:Y:Z$ should be $6:3:1$. The ratio obtained using the actual package dimensions is $6:2.2:1.05$. The fact that the optimum ratio is not used results in only 0.2 inches (40 picoseconds) of extra distance that must be traversed during each read access. This is insignificant compared to the total path length.

The 14-pin ceramic flat package is used as the circuit package because (1) sufficient leads are available to handle the required circuits, (2) the heat transfer characteristics are adequate as discussed in the Thermal Study section, and (3) the reflow solder attachment technique allows ease of assembly and repair.

Separate terminating resistors are required since the circuits are designed for driving a 50-ohm trans-
Figure 14. Read input/output logic card artwork.
mission line system. However, many lines are sufficiently short to be treated as lumped allowing a higher impedance termination. The 50-ohm terminations are resistive divider networks connected across the positive and negative supplies, giving a termination voltage slightly more negative than the negative logic level. Higher impedance terminations are single resistors returned to the negative supply. One-tenth watt carbon resistors are attached to the boards by reflow soldering in the same manner as the packages are attached. Lead lengths are kept to a minimum to reduce stray inductance and capacitance.

Provision is also made for attaching power supply bypass capacitors. In addition, the \( V_{CC} \) and \( V_{EE} \) voltage planes are separated by only two mils of dielectric resulting in a relatively large distributed capacitance throughout the memory.

The copper-clad epoxy material is selected for the thickness required to construct striplines of the desired characteristic impedance. Only normal control of etching is required to achieve the \( \pm \frac{1}{2} \) mil tolerance on line width. Holes in the logic cards before plating are only 16 mils to reduce lumped capacitances and to allow routing of printed runs between holes on the inner plane.

Logic card connectors are required with a sufficiently large number of pins available to distribute voltages and signals across the entire width of the card to reduce transmission line discontinuities. A close card spacing is also required. These requirements are met using a double row strip connector having contacts on 50-mil centers. The connector thickness allows card spacing to be on 0.2-inch centers.

Prior to the design of the memory, a detailed study of stripline characteristics was made. This enabled the layout and fabrication of a system with predictable characteristic impedances and signal cross coupling. Signal cross coupling is limited to less than 5% in all cases. The high package density greatly reduces the number of crosstalk and termination problem areas.

**Thermal Study**

A thermal study was required to show feasibility of the proposed system packaging configuration. Due to the very fast circuit operation, steps were taken to reduce signal propagation times between circuits resulting in an increased power density. In addition, the cooling method was specified as forced air at 70°F and a pressure of 0.2 inches of water. This left thermal resistance, air flow pattern, and cooling surface area as variables.

A total of 50°C junction to ambient drop was set as a design goal to limit the junction temperature to less than 75°C. Therefore, an anticipated maximum 0.3 watt per package limited the junction to ambient thermal resistance to less than 166°C per watt. A standard production type 0.25 by 0.25 inch ceramic flatpack was examined and found to have a junction to case thermal resistance of 30 to 40°C per watt depending upon the quality of the die bond. This is acceptably low if the case ambient thermal resistance is 125°C per watt or lower.

To properly simulate the actual system's environment, multilayer thermal test cards (Fig. 15) were constructed which contained voltage planes in the same manner as the proposed system. Circuits mounted on these cards contained an isolated diode whose forward voltage drop could be measured as a function of temperature. This is shown schematically in Fig. 16. During testing, three fully populated cards were spaced on 0.2-inch centers and subjected to 25°C air under a blower pressure of 0.2 inches of water. Temperature measurements were made on the center card.

Four package attachment techniques were explored. In each case, the package leads were attached by reflow soldering to the test card. In the first test, each package was mounted with the bottom or circuit side of the package closest to the card. The test measurements showed about 190°C per watt thermal resistance with an insignificant variation along the length of the card. The second test involved a simple inversion of the package placing the chip side of the package away from the card and in the air stream. The result was an improved 167°C per watt. Though this mounting technique was acceptable, others were explored.

To effect further improvement, the packages were heat-sunk to the card to gain cooling area improvement. A plated-through hole was placed beneath each package which was connected to a copper voltage plane on the opposite side of the card. Heat transferred readily to all parts of the card thereby using both sides of the card as cooling surfaces. A nonsetting thermal joint compound was applied beneath each package during assembly. The result was a thermal resistance of 131°C per watt.

The second heat-sinking technique involved spe-
Figure 15. Thermal test card.
pecially prepared circuit packages with metallized bottoms which were soldered to the plated-through holes with a low temperature solder. Though requiring a more involved attachment technique, test results showed a significant improvement at 88°C per watt.

The thermal joint compound is used in the final system since it provides a margin of safety and presents no unusual assembly problems.

TEST RESULTS

Partially Populated Model

The printed circuit boards, as described in the Packaging section, are designed and fabricated for a 64-word memory of 8 bits per word. It is felt that four paths through the memory are sufficient to test the memory circuits and interconnections. The memory boards are populated in a manner such that four words of two bits each can be accessed for both read and write. Each circuit node in the four paths is fully loaded with fan-out so that worst-case circuit delays are present. The block diagram of Fig. 17 illustrates the circuits included in the partially populated model. The circuits are positioned such that the worst-case path lengths are encountered. Small variations in path length exist even though the memory is designed to minimize this effect. The difference
between the longest and shortest paths is in the order of 0.5 nanoseconds.

Worst case conditions on the following parameters are important in the operation of the memory:

- Circuit loading
- Stripline delay
- Crosstalk
- Reflections
- Power supply tolerance and noise
- Junction temperature

As discussed above, the worst-case circuit loading and stripline delays are included in the partially populated model.

The use of partially populated, full-size system boards also allows a fairly thorough evaluation of crosstalk and reflection problems. As mentioned earlier, the striplines are designed with the goal of keeping crosstalk below a certain level. The test results indicate that this effort is very successful since even the longest lines have crosstalk of well under the design goal of 5%. The memory is overdesigned in this respect since a 1-nanosecond risetime is assumed in the design, while a 2-nanosecond risetime is actually achieved in the system. All lines of under 2 inches in length are treated as lumped and a 180-ohm output resistor used for the gate load. All lines over 2 inches are treated as distributed and terminated in 50 ohms. As previously described, the actual striplines are designed for a 75-ohm characteristic impedance in areas of distributed load so that the loaded line is close to the desired 50 ohms. The reflections due to connectors, loads, and terminations are also well under 5% of the total signal amplitude.

The positive power supply can be varied by ±10% and the negative supply by ±20% before an error is detected. Also, the power supplies are almost entirely free from noise resulting from current transients. This is due to the use of the thin 2-mil dielectric between voltage planes and miniature tantalum capacitors distributed around the cards. It is felt that the results achieved in the partially populated model are more than satisfactory and that no problems will arise when the remainder of the memory is populated.

Since the full power dissipation is not present in the memory, worst-case temperature effects cannot be determined. The individual circuits have been tested at temperatures in excess of those predicted by the thermal studies previously described, however, with no adverse effects on the switching characteristics. The DC circuit characteristics are also within specification at the elevated temperatures. On the basis of these tests, no problems are anticipated in regard to higher junction temperatures in the full system.

Memory Exerciser

The memory exerciser is designed to test the memory with various patterns of read addresses, write addresses, and write data. Counters containing two stages are used to determine the read and write addresses. The two counters count in opposite directions and can be preset to any value. A third stage on the write address counter switches between two patterns of write data. The two patterns are simply mechanical switches which can be set to ONE or ZERO independently. The pattern bit in the counter changes state only after all four memory locations have been selected. This results in the pattern A information and the pattern B information being written into all four locations alternately. The exerciser compares the actual information obtained during a read cycle to the expected data and stops the machine if an error occurs.

Memory Operation

Figure 18 shows the read initiate pulse supplied by the memory exerciser, the memory bit output at the point of the tied emitter OR, and the output of the output data register. Thirteen nanoseconds are required from the time the read initiate clocks the read address into the register until the stored information is available at the input to the output data

![Memory waveforms](From the collection of the Computer History Museum (www.computerhistory.org))

Figure 18. Memory waveforms.
register. The delayed read initiate arrives at the output data register one nanosecond later and clocks the stored information into the register. The data is available at the terminals of the memory 17 nanoseconds after the start of the read initiate pulse.

It is important to note that the memory bit output is directly compatible with the current mode logic and no amplification is required. This eliminates the usual memory noise problems and attendant recovery time periods allowing the extremely fast memory speeds. This is simply an extension of the same technology used in the remainder of the computer applied to the implementation of a high speed bank of independently addressed registers to be used in various critical points in the computer. A later section gives examples of the use of a memory of this type.

In the read-decision-write cycle, a delayed read initiate pulse clocks the write address and write data into the appropriate registers. The exerciser then supplies a write command pulse 15 nanoseconds after the read is completed. The write operation is completed approximately 8 nanoseconds after the start of the write command pulse.

In a more general application, the write command would be used to clock the write address and write data into the registers. This is a longer delay path and 10 nanoseconds would be required to perform a write operation.

The memory exerciser contains no bypass route in case the read address and write address are identical during a cycle. If this occurs, the old information is obtained during the read operation since the read is completed before the memory bit changes state (refer to Fig. 2). The memory bit changes state during the first portion of the next read cycle just prior to the time the address information filters down to the memory bit read gate. The error detection circuitry is designed to account for this type of operation.

The only limitation placed upon simultaneous read and write operations is that the memory bit cannot be accessed for a read operation during the period of time (approximately 2 nanoseconds) when the memory bit is actually changing state. As long as the system designer knows whether the memory bit is accessed before or after the change of state, he can use the memory output accordingly. This allows a large degree of flexibility on the part of the system designer.

It should be noted that the registers at the inputs and outputs to the memory are included for generality only. In most cases, these registers are contained in the logic portion of the computer and can be eliminated from the memory, assuming the proper timing can be guaranteed. This, of course, allows faster operation of the memory.

In certain instances, cycle overlap could be used to decrease the read and/or write periods. The actual read access time and/or write time would remain the same; however, the rate of operation would be increased by taking advantage of the read address, write address, and input data storage time in the various decode and fan-out trees.

SYSTEM APPLICATIONS

The memory described thus far is designed for a particular application requiring a read-decision-write cycle and 64 words of 8 bits per word. As previously mentioned, other modes of operation for different applications are possible. The memory capacity can also be altered quite easily.

The number of bits of storage can be increased in several ways. A modular approach can be used by connecting $64 \times 8$ memories in parallel or the memory boards can be redesigned to accept the larger number of bits. The modular approach is particularly applicable to the distribution of small memories of various sizes throughout a large computer. It is possible to construct a $64 \times 32$ memory using either of the above approaches with a cycle time of approximately 20 nanoseconds. The extra 3 nanoseconds result from one stage of logic plus line delay.

The memory has all of the usual applications of scratchpad memories, but its two read channels suggest other special applications.

The arithmetic unit of a computer operates on two words at a time. A dual read operation could provide both words to the arithmetic unit in only one memory access time. The simultaneous write feature would allow the memory to be written into at the same time the read operations were taking place. The data being written could come either from the main memory or the arithmetic unit.

In some computers, a separate adder is used for index and dynamic memory relocation operations. A memory with dual read capability could provide...
the index quantity and relocation coefficient to the adder simultaneously. The adder would then add the above two quantities to the address in one step. The dual read operation could also be used to provide both halves of a word for double word length operations.

All of the above applications are used to reduce the effective memory access time by one-half.

CONCLUSIONS

This paper describes an extremely high-speed memory that is very flexible in application. The physical size required to implement a flip-flop memory of this speed can only be obtained using densely packaged integrated circuits. The use of a storage element operating at standard logic levels contributes greatly to the high speed and allowed the development of this memory in the relatively short period of 10 months. The memory speed, flexibility, and development time are all improvements over prior attempts to implement under-100-nanosecond memories using magnetic techniques.

ACKNOWLEDGMENTS

The authors extend their thanks to all of the persons in the Integrated Circuit Research and Development group at Motorola who contributed to the design and construction of the memory described in this paper. The efforts of Howard Hannah, Robert Hawn, Dave Turcotte, and Anthony Ziolekowski are particularly appreciated.

REFERENCES
