A 180-MHz 0.8-μm BiCMOS Modular Memory Family of DRAM and Multiport SRAM

Allan L. Silburt, *Member, IEEE*, Richard S. Phillips, *Member, IEEE*, G. F. Randall Gibson, *Member, IEEE*, Steven W. Wood, *Member, IEEE*, Armin G. Bluschke, *Member, IEEE*, James S. Fujimoto, *Member, IEEE*, Stephen P. Kornachuk, *Member, IEEE*, Benoit Nadeau-Dostie, *Member, IEEE*, Rajesh K. Verma, and Peter M. J. Diedrich, *Member, IEEE*,

***Abstract*—This paper describes a family of modular memories with a built-in self-test interface designed using a synchronous self-timed architecture. This approach is ideally suited to modular memories embedded within synchronous systems due to its simple boundary specification, excellent speed/power performance, and ease of modeling. The basic port design is self-contained and extensible to any number of ports sharing access to a common-core cell array. The same design has been used to implement modular one-, two-, and four-port SRAM's and a one-port DRAM based on a four-transistor (4-T) cell. The latter provides a 45% core cell density improvement over the one-port SRAM. Nominal access and cycle times of 5.5 ns for 64-kb blocks have been shown for a 0.8-μm BiCMOS process with no memory process enhancements. System operation at 100 MHz has been demonstrated on a broad-band time-switch chip containing 96 kb of two-port SRAM.**

I. Introduction

MODULAR memory is an essential component of any digital IC design library. Telecommunications circuits make extensive use of embedded RAM in a wide variety of applications including processor data and microcode stores, time-switch cores, register files, elastic stores, buffers, finite state machine registers, and caches. The ability to create any memory size and configuration provides system designers with the power and flexibility required to optimize designs for custom applications. The availability of a multiport capability adds an additional dimension of flexibility that can often simplify a system design. Systems are further simplified when embedded memories present a synchronous interface in which a minimum number of timing-critical signals are presented at its boundary.

Thus, in developing the modular RAM family for a mixed-signal 0.8-μm BiCMOS ASIC process technology [1], the following objectives were set out:

Manuscript received July 23, 1992; revised October 30, 1992.

A. L. Silburt, G. F. R. Gibson, S. W. Wood, A. G. Bluschke, J. S. Fujimoto, S. P. Kornachuk, and B. Nadeau-Dostie are with Bell-Northern Research, Ottawa, Ont., K1Y 4H7, Canada.

R. S. Phillips is with Bayview Technologies Inc., Constance Bay, Ont., K0A 2M0, Canada.

R. K. Verma was with Bell-Northern Research, Ottawa, Ont., K1Y 4H7, Canada. He is now with ATI Technologies, Scarborough, Ont., M1W 3S2, Canada.

P. M. J. Diedrich is with the Massachusetts Institute of Technology, Cambridge, MA 02139 on leave from Bell-Northern Research, Ottawa, Ont., K1Y 4H7, Canada.

IEEE Log Number 9206360.

TABLE I
RAM Family Modularity Range

Parameter	Min	Step	Max
Words	16	–	8k
I/O Bits per Word	1	1	64
Total Bits	16	–	64k
Rows	8	2	256
Columns	2	2	256
Columns per I/O Bit	2	2	32
Row Address Bits	3	1	8
Column Address Bits	1	1	5

1) Design an interface which would enable system designers to easily deploy embedded RAM's within a predominantly synchronous design environment in such a way that the system function and timing are easily predicted, implemented, and tested.
2) Achieve 100-MHz worst-case system operation (4.5 V, 105°C) for a broad-band time switch containing 96K of two-port SRAM with memory power dissipation less than 500 mW.
3) Create a reusable design with fine grain modularity up to a maximum 64K contiguous block size as specified in Table I.
4) Provide multiport functional capability (up to four ports).
5) Exploit the existing BiCMOS process technology, which has no memory enhancements in its flow, to yield a high-density memory beyond that achievable with a standard six-transistor (6-T) cell.

These objectives were achieved using a synchronous self-timed design style. The result is a four-member family of DRAM and multiport SRAM built around 6-, 8-, and 12-transistor SRAM cells (for one-, two-, four-port functions) and a 4-T one-port DRAM cell. Each member has virtually identical schematics and is capable of operating at 5.5 ns or better nominal cycle times over the entire modularity range listed in Table I.

The achievement of a single, reusable design covering this broad range of size and functionality has not been previously reported. Previous works on modular memory such as [2]–[5] have not employed internal self-timing and thus do not have the simple synchronous interface presented here. Although other fixed-block-size, self-timed RAM's have been reported

0018–9200/93$03.00 © 1993 IEEE

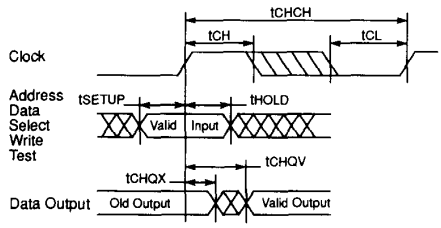


Fig. 1. Read cycle timing specification.

[6], [7], none has proposed a means of modularizing this technique. Since self-timing requires careful control of matched delay paths through a RAM, obtaining high-speed performance while maintaining adequate operating margins over process, environmental range, and size requires careful development of core cell, sense scheme, and modular self-timing technique. The interface specification and generic port architecture used to achieve these requirements will be described in Sections II and III. The implementation of multiport SRAM and DRAM is presented in Sections IV and V. Although fixed-size embedded DRAM has been reported [8], creating embedded modular DRAM in a standard ASIC process has received little attention in the literature. It can offer substantial density benefit and, using a 4-T core cell, can be made compatible with SRAM circuitry, having similar speed and power characteristics. The benefits of this approach versus a 1-T cell for this process are discussed in Section V and the Appendix. The testing of embedded memory is simplified by predefining the chip-level test strategy and building the necessary test functions into the RAM. The unique test modes for the family are described in Section VI. Measured results for stand-alone blocks as well as a set of embedded arrays are presented in Section VII.

II. Interface Specification

A simplified timing diagram for a read cycle on a generic port is shown in Fig. 1. Every input pin (address, data, select, write, and test) is registered and a signal behavior analogous to an edge-triggered flip-flop is achieved. All ports behave identically and have the same parametric values for read and write cycles. Internal self-timing assures insensitivity to the exact positioning of the falling clock edge. When a cycle is complete, the memory reverts to a low-power standby mode in which no dc power is consumed.

This interface specification was chosen because it brings several benefits to the system designer. The high-level model is simple and has few critical parameters. As a result, timing at the boundary can be easily verified using static analysis or a minimal set of functional simulations. Power dissipation varies linearly with clock frequency, is independent of duty cycle, and is less sensitive to process and temperature variations compared to nonself-timed memories. Cycle times, as specified for the block, are easily realized in the system since there are no additional input pin timing relationships to meet other than setup and hold with respect to the clock. An optional modification of the output buffer yields a pipelined version that provides faster effective access times at the expense of one clock cycle latency.

A synchronous memory block select pin is also included to allow block subdecoding at the system level. Since unselected blocks draw no current, area may be traded for power savings when constructing large memory functions. Two test modes are included to provide compatibility with the scan and memory BIST test methodologies described in [9].

III. Generic Port Architecture

A. Self-Timing

In order to achieve the synchronous self-timed port behavior, a control system has to be implemented within the RAM to manage the access operations to the inherently asynchronous core. Such a system requires "gating points" along the data path to synchronize access events with the arrival of global control signals. This poses a significant challenge in the design of a modular RAM family when high speeds are required. Since a "gating point" represents, in effect, the convergence of a race between the data or address and a clock derivative, the arrival times of these signals at the gating points must be adequately margined over process, temperature, voltage, mask dimensional variation and, for modular RAM, size and configuration. Unacceptable speed penalties are incurred if too many gating points and global signals exist in the design. Hence, in order to achieve high-speed operation, the following control optimizations were applied:

a) the number of gating points along the data/address path was minimized, and
b) the number of global control signals was minimized in order to reduce the margining required along different stages of the access path.

In the optimized control system (Fig. 2) the number of gating points on the critical path has been reduced to two: the rising edge of the system clock, which initiates the cycle, and the internally derived self-timed reset signal, which initiates sensing. The number of global control signals is reduced to four: the system clock, the internal reset, the sense signal, and write signal. All other control signals are locally generated. Each memory signal path operates independently, simplifying internal timing and allowing parallel operation so that data and address paths can be independently optimized. The internal self-timing circuit is comprised of several "model" paths that track the propagation of signals through the memory. These paths maintain tightly margined operation over the modularity range, process, and operating conditions.

The memory operation can be divided into three separate phases: setup, core access, and reset. During setup, the address inputs are predecoded. On the rising edge of clock, all input signals are latched, the sense amplifier is preset and enabled, and the predecode information is propagated to the word-line and column decoder blocks. The model timing signals converge on the control block where they are combined to generate the internal reset pulse. The rising edge of the reset pulse triggers sensing and, in parallel, resets the address predecoders reestablishing bit-line and data-bus precharge. Deactivation of the model decode paths creates the falling edge of the reset pulse, indicating that the circuit is ready for

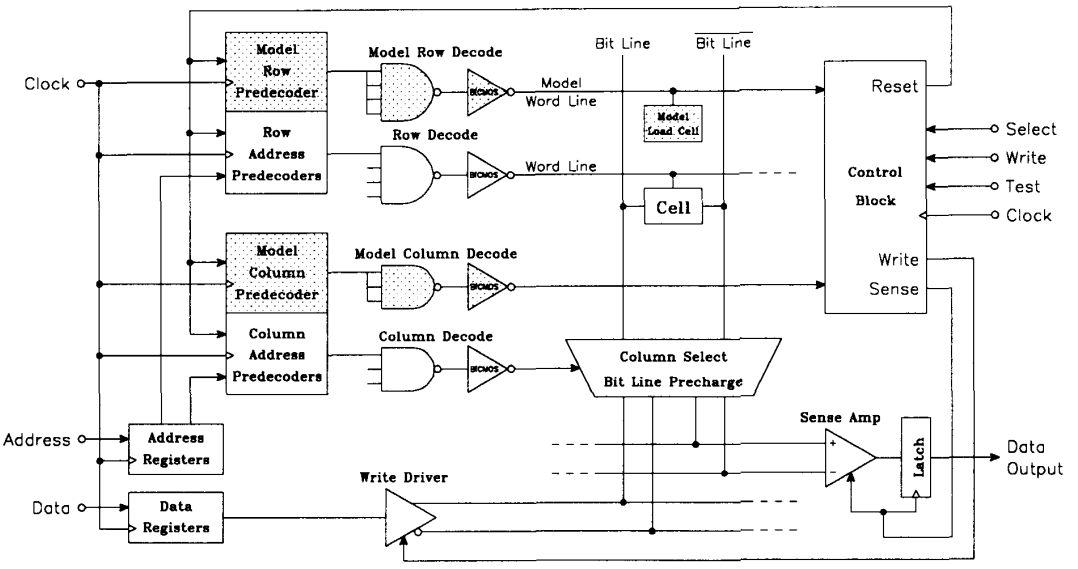


Fig. 2. Generic port architecture block diagram.

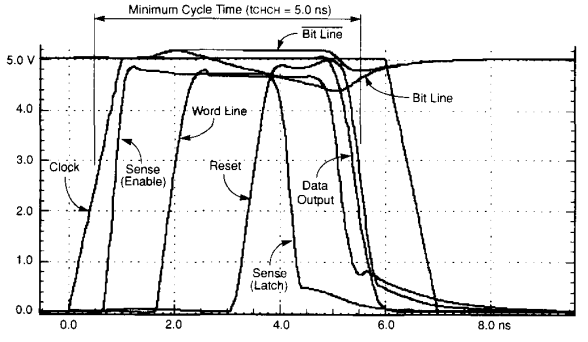


Fig. 3. Simulated read cycle for an 8K × 8 one-port SRAM under nominal conditions.

the next clock edge. In the case of a write, the rising system clock enables the write driver while the reset signal disables the write control signal, terminating the write drive. In this manner, the memory is fully deactivated and returned to its quiescent state with no dc power consumption. A simulated cycle (Fig. 3) showing the internal self-timing signals of an 8K × 8, one-port SRAM demonstrates 200-MHz operation under nominal conditions.

The self-timed reset function provides the added benefits of clock duty-cycle independence, linear power versus frequency characteristics, and reduced sensitivity to variations in operating conditions. This is due to the fact that the length of the active cycle depends directly on the speed of the signal paths in the memory as determined by the process, temperature, and voltage conditions. Hence, an almost constant amount of charge is used during any cycle, yielding a linear power versus frequency characteristic. Fig. 4 shows simulated variation of both access time and average power over voltage, temperature, and process conditions for an 8K × 8 one-port SRAM. A variation of less than 7% in average power is predicted over the complete process and temperature range. The variation is less than 4% over the operating voltage range (once the contribution due to CV^2f is removed).

B. Model Path

The internal self-timing circuitry tracks the address decoding operation using the model address decode, column decode, and word-line paths shown in Fig. 2. One goal of the implementation was to maximize the reusability of active circuit blocks in the model path while accurately tracking delay of critical signal paths. The use of model signal paths comprised of identical circuitry and physical routing allows this goal to be achieved in a modular fashion. Identical predecode buffers drive both model and address predecode signals to the row and column decoder arrays. By accurately loading the model predecode signals with equivalent decoder loads and busing them the full length of the array along with the active predecoded address signals, the transmission-line loading effects are simulated. Word-line delay is simulated by the model row decoder and model word line, which is bused across the top of the memory array and loaded with slightly modified memory cells for maximum modeling accuracy. A model column select signal is similarly generated to simulate the column decode operation.

The ability of the internal model signal paths to track loading and delay over the modularity range provides a signal development period proportional to the size of the memory (see Fig. 5). The sense-amplifier input signal, which is strongly dependent on bit-line load, varies between 130 mV in one extreme (8K × 1 configuration) and 860 mV in the other extreme (32 × 1 configuration). Note that simulations used to derive the above data include the effects of worst-case side-by-side transistor mismatch in the core cell and sense amplifier.

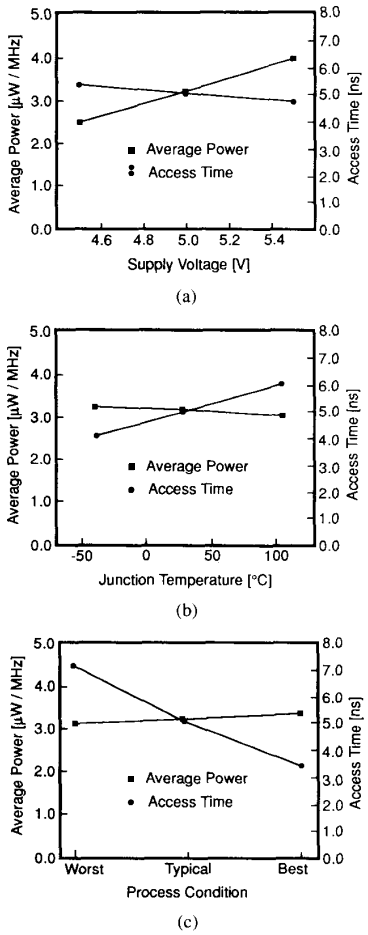


Fig. 4. Simulated average power and access time for an 8K × 8 one-port SRAM over (a) supply voltage, (b) junction temperature, and (c) process variation.

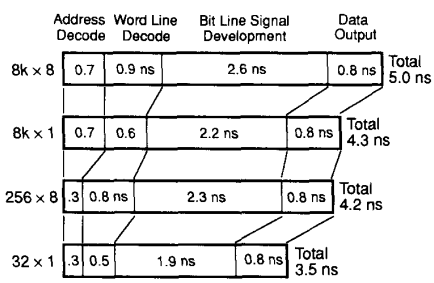


Fig. 5. Read cycle breakdown for four sizes of one-port SRAM under nominal conditions.

C. Address Path

A pulsed predecoded addressing scheme is used to limit decoding power and design complexity while achieving minimum decode delay over the modularity range. The address path makes extensive use of BiCMOS buffers for driving heavily loaded predecoded address lines and word lines. The relative insensitivity of these drivers to load capacitance yields a weak dependence of the access-time and cycle-time specifications on array size. The address inputs are predecoded in pairs and enabled by the system clock. In this manner, a minimum number of predecoded address lines are pulsed in a given cycle: one of four per address bit pair as compared to two of four if predecoding were not used. Static decoding of the predecoded address lines is used to select the active word line and columns, avoiding the requirement for the large decode clock buffers and the additional gating point that would be required by a dynamic decoder.

D. Data Path

A clocked-latch style CMOS sense amplifier (Fig. 6) was chosen for use in the memory family. The clocked latch has the advantage that it combines both the sensing function and the output latch, thus shortening the overall data path. BiCMOS sensing schemes, such as those often used in BiCMOS SRAM's [10], [11], are not suitable for this application, which requires fast activation and shutdown for zero standby current. The large collector-to-substrate capacitances that must be charged and discharged on every cycle make switched bipolar differential stages slower to cycle than a CMOS latch in this application.

The rising edge of the sense control signal triggers the timing chain associated with each sense circuit, initiating a sequence of local control signals to isolate and precharge the sense latch ($C4$ and $C3$), disable the read data-bus precharge ($C1$), and open the PMOS data-bus access devices ($C2$). The sense-amplifier latching clock ($C4$) is initiated on the falling edge of the sense control signal, which is in turn triggered by the self-timed reset sequence. Concurrently, the sense latch is isolated from the data bus via $C2$. The delay through the self-timing network ensures that an adequate signal is present on the read data bus prior to latching the sense amplifier.

The column select signal controls both the local bit-line precharge and the column access function, simplifying the overall system timing by eliminating the need for a global precharge clock and a resulting gating point. If a column is not selected, the read and write access remains off while the bit-line precharge devices are on, acting as bit-line clamps. As a consequence, dc power is consumed in all unselected columns of the memory during a read or write cycle.

Calculated precharge power dissipation for both the local and global precharge cases is compared in Fig. 7. In the global precharge case, the power per column is comprised of CV^2f power in the unselected bit lines as well as the per-column power dissipated in the global clock line and precharge device drivers. An unselected bit-line voltage swing of 800 mV (due to read current) is used in the global precharge case while a 100-mV swing (due to bit line clamping) is used in the local precharge case. The 800-mV bit-line swing in the global precharge case is chosen as a sufficiently conservative lower bound for the modularity range of the design. The results indicate a 20% power savings in larger memories using local precharge, while the advantage of global precharge in smaller memories is insufficient to justify the additional gating point and global signal required.

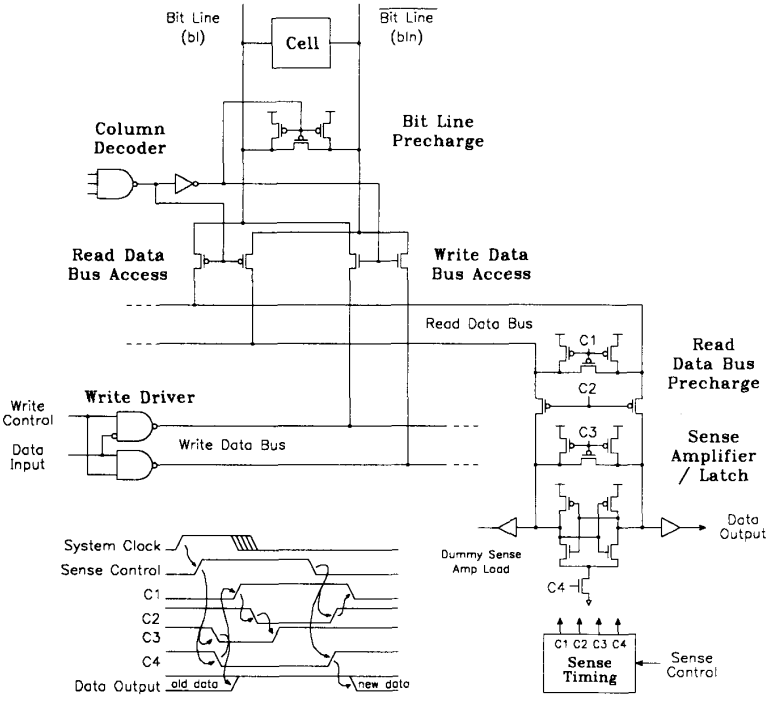


Fig. 6. Generic port data-path schematic.

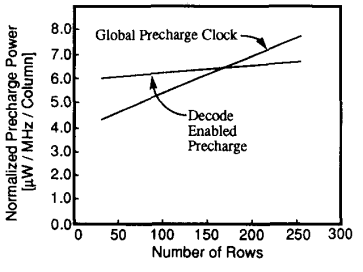


Fig. 7. Comparison of local versus global precharge schemes.

The data path uses separate read and write data buses that allow bit-line access optimization. On reads, the bit-line differential is limited to less than 1 V from the V_{DD} precharge level before the word line is deselected by the self-timing loop. The PMOS read data-bus access devices pass this signal to the sense amplifier whose common-mode rejection ratio permits sensing within this narrow band below V_{DD}. The NMOS write access devices remain off since both write data lines are held high. This isolates the write driver loading from the signal path avoiding degradation of the bit-line slew rate. These access devices also clamp bit lines, limiting the active swing to a maximum of a substrate-bias-enhanced V_{tn}. During a write, the NMOS write access devices allow the full ZERO level to be written to the bit lines. The PMOS read access devices have negligible impact on write performance and power due to their small size.

The buffered data output isolates the sense latch from any external disturbances and provides additional drive. Alternatively, an optional register stage may be added to pipeline the data output. In this manner, the access time (t_{CHQV}) may be reduced to the clock-to-q delay of the flip-flop stage at the expense of a single-clock-cycle latency.

IV. Multiport Implementation

A symbolic layout for each RAM type and its associated core cell schematic is shown in Fig. 8 along with a core cell area and memory performance comparison in Table II. One access transistor pair is added per core cell to provide fully independent read/write capability for each port. The symmetry this imparts to a multiport memory yields simplicity in design, layout, simulation, and built-in self-test circuitry. Each port functions independently with its own control block and model signal path yielding a highly extensible design with broad system applications. Each core cell is margined for stability to permit simultaneous reads from the same address on all ports. Simultaneous writes to the same address location are prohibited. All circuitry outside the core cell is replicated on a per-port basis. Only the layout base cells are unique for each RAM family.

Several multiport SRAM designs reported in the literature have dedicated read or write only ports and use single-ended cell access schemes [2]–[4], [6] in order to gain claimed density advantages. Single-ended read ports such as these induce slower overall access times because of their sensitivity to power supply noise as well as the extra margining required for PMOS/NMOS process skews in the sense amplifier. To prevent inadvertent writes to unselected columns, the write

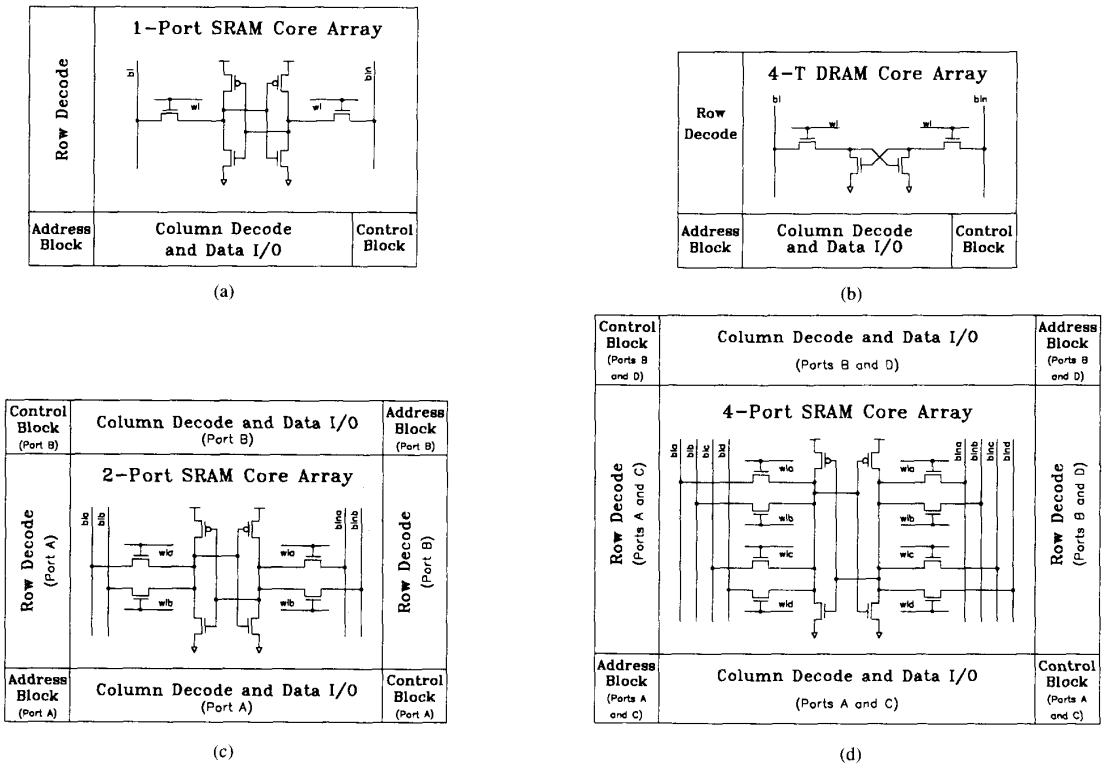


Fig. 8. Symbolic layout diagram with core cell schematics for (a) one-port SRAM, (b) 4-T DRAM, (c) two-port SRAM, and (d) four-port SRAM.

TABLE II
RAM CORE CELL AREA AND PERFORMANCE COMPARISON FOR 8K × 8 BLOCK SIZES AT NOMINAL PROCESS, 27°C, AND 5.0 V

Type	Cell Area [μm^2]	tCHQV [ns]	tCHCH [ns]	Pave [mW / MHz / Port]
4-T DRAM	73.2	4.9	5.2	3.3
1-Port SRAM	131.3	4.8	5.0	3.3
2-Port SRAM	197.8	5.3	5.5	3.5
4-Port SRAM	397.8	5.4	5.5	3.7

ports of these cells require both row and column word-line access thus increasing the area of the memory periphery.

In a 0.8-μm triple-metal process, the density advantages of such single-ended architectures are negligible. For example, the symmetrical, fully differential, two-port cell is 197.8 μm^2, whereas a single-ended, read-only, and write-only two-port cell as in [6] is 183.7 μm^2. This magnitude of density advantage does not warrant the slower read access times and larger write peripheral circuit areas required by the single-ended cell architecture.

The inherent read/write nature of all ports in the multiport architecture chosen simplifies the testing of the memory. Each port can be tested as a stand-alone one-port read/write SRAM independent of its read-only or write-only configuration in the system. Dedicated port-to-port fault testing is simplified with the generic read/write ability on all ports as described in Section VI.

The layout architecture of the generic port periphery places the address block in one corner with the row and column decoders extending from this block along adjacent sides of the core cell array. The data I/O circuitry is integrated into the same layout block as the column decoders. The timing control circuitry is located at the far end of the column decode block so that it may receive the far end of the model word line and model Y-decode output as its main timing interlocks. By placing the control block far from the word-line drivers and column address drivers, any RC delay effects on word-line and column addresses are compensated for in the model timing path.

Ports are mirrored around the cell array, as shown in Fig. 8, to achieve the multiport configurations. No multiport DRAM architectures were considered since the area advantages gained by removing the PMOS load devices become much less significant as one adds extra access devices, word lines, and bit-line pairs for the additional ports.

V. DRAM IMPLEMENTATION

Although little has been published on modular DRAM design, embedding DRAM in place of SRAM can often reduce ASIC die size [8]. In the absence of any DRAM-specific processing (trenches, stacked capacitors, thin oxides, etc.), a simple planar 1-T cell could be used whose data storage node is the channel of the MOSFET capacitor (Fig. 9(a)). Alternatively, a 4-T cell similar to a conventional SRAM cell

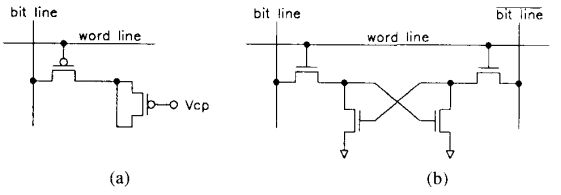


Fig. 9. ASIC DRAM cells: (a) 1-T cell, and (b) 4-T cell.

with no pull-up load devices could be employed (Fig. 9(b)). The two approaches have widely different density, power, speed, and radiation tolerance characteristics. A PMOS array is the best configuration for the 1-T cell in this process because of the inherent noise immunity of the enclosing N-tub while a 4-T cell formed from NMOS transistors was considered because it results in superior current drive onto the bit line.

A comparative analysis of the two options is presented in the Appendix which shows the advantage of using a 4-T DRAM cell over a 1-T cell in an ASIC process. Further, significant advantage is gained when one considers that the 4-T cell fits easily into the modular framework of the SRAM family. Virtually the same peripheral circuits can be used in a 4-T DRAM design as in a 6-T SRAM design.

VI. Test Features

The test features of the memory are designed specifically for a chip test strategy based on built-in self-test (BIST) for embedded RAM's and scan techniques for the logic surrounding them [12]. The BIST technique [9] provides for the sharing of one BIST controller among several memory blocks on a chip. This serves primarily to save chip area. The BIST controller is automatically synthesized and laid out with the rest of the system functions outside the memory block boundary. This provides flexibility in adjusting the test algorithms over the life of the memory design. Dedicated BIST circuit layout is not used since it requires considerable effort to design, lay out, and characterize. There are three test features required to support this strategy: 1) input flip-flops with observable outputs, 2) a data-path shunt mode, and 3) a shadow write mode.

All of the registered inputs are made observable by providing an additional dedicated output for each input pin. This allows these registers to be added to the scan chain to ensure fault coverage of all of the preceding logic and interconnect up to the block boundary. For example, Fig. 10 shows how the address input register can be shared in scan mode (as part of the scan register chain), memory BIST mode (as part of the address counter), and normal functional mode while inserting only one multiplexor delay into the user's critical timing path.

In the data-path shunt mode, the registered input data bypass the memory core and are transferred directly to the data output pins. This is used during scan mode to provide direct control of the memory output pins ensuring fault coverage of the down-stream interconnect and logic.

The shadow write mode is specific to multiport memories. The layout of the multiport core cells to achieve maximum density requires adjacency at minimum pitch of various lines.

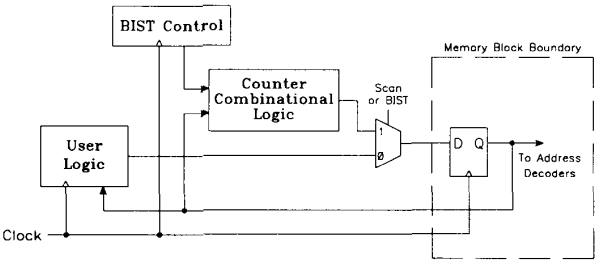


Fig. 10. Memory interface showing BIST, scan, and functional mode usage of memory address input registers. Registers on the other inputs may be similarly exploited.

TABLE III
0.8-μm BiCMOS Process Summary

Vertical NPN:	Emitter size (min)	0.8 × 4.0 μm
	F_t	11 GHz
NMOS and PMOS LDD:	Lmin	0.8 μm
	V_{tp} / V_{tn}	-0.85 / 0.80 V
	t_{ox}	17.5 nm
Interconnect:	Metal 1 line 0.8 μm, pitch 2.0 μm	
	Metal 2 line 0.8 μm, pitch 2.0 μm	
	Metal 3 line 1.0 μm, pitch 2.4 μm	
	Stacked via and contact structures	
	Local Interconnect	

Inductive fault analysis shows that defects causing shorts between adjacent word lines or adjacent bit lines with the same address on opposite ports are a dominant failure mode. This type of fault is difficult to detect under normal operating conditions since the relative timing of the short word-line pulses from each port are critical in sensitizing the fault. This mode provides a means of reliably testing for these defects with minimal change to the one-port BIST algorithms. When a port is in shadow write mode, its self-timing loop and address decoders are disabled, all word lines remain off, and all bit lines are driven statically with the pattern on the data inputs. A normal test sequence is applied to one port while shadow writes are applied to all others with an inverted data pattern. If the memory is fault-free, no cell will be disturbed. Since the data and address to be applied to the two ports are simply derived from a single source, it adds little to the BIST controller size and complexity. A multiport memory is tested as if it was a collection of one-port blocks. While any port is under test, all other ports are in shadow write mode.

VII. Results

An evaluation vehicle for the multiport SRAM and DRAM (Fig. 11) was fabricated in the 0.8-μm BiCMOS process summarized in Table III. The chip contains a number of blocks that span the modularity space of the memory family and has been used to verify the functionality of both the SRAM and DRAM designs.

Performance measurements were made on a stand-alone 3K × 8 nonpipelined two-port memory block under nominal conditions. An oscilloscope trace showing an access time of 4.8 ns is presented in Fig. 12. Average power dissipation

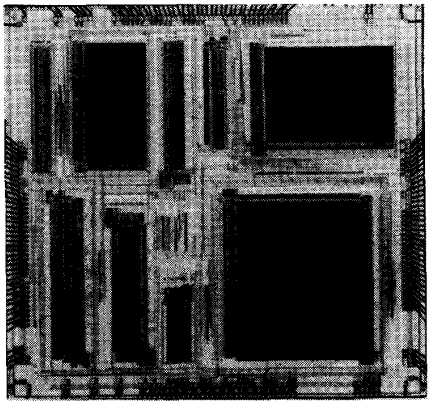


Fig. 11. Die photo of DRAM and multiport SRAM evaluation vehicle.

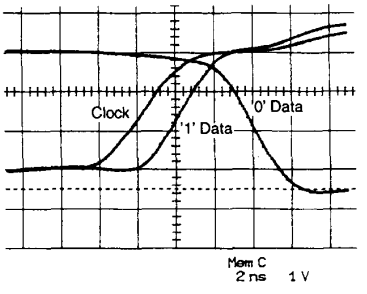


Fig. 12. Access time waveforms for 3K $\times$ 8 two-port SRAM test block under nominal conditions. Measurements are made through TTL-compatible I/O test paths.

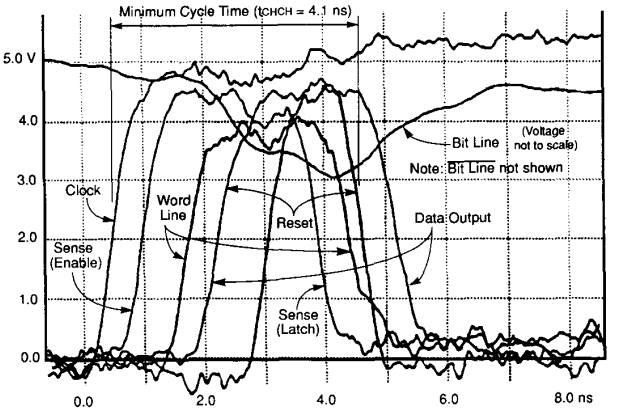


Fig. 13. E-beam voltage contrast waveforms for 3K $\times$ 8 two-port SRAM test block under nominal conditions.

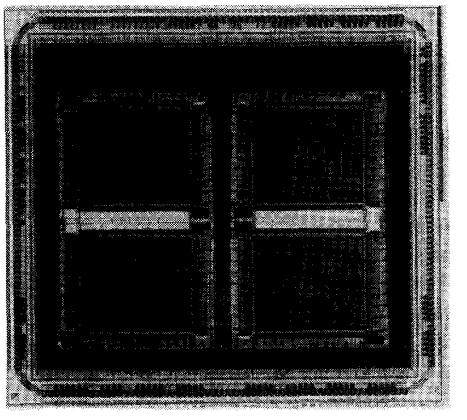


Fig. 14. Die photo of broad-band time-switch chip with four 3K $\times$ 8 two-port SRAM's and embedded BIST circuitry.

was found to be 3.5 mW/MHz under the same conditions. Cycle time measurements were made using an e-beam voltage contrast system to trace internal self-timing signals (Fig. 13). The cycle time, measured between the rising clock edge and the falling edge of the internal reset signal, is determined to be 4.1 ns, confirming greater than 200-MHz performance for this block. Timing measurements from the critical internal signals (Fig. 13) agree to within 5% of simulated values.

System-level performance at 100 MHz has been verified on a memory management chip (Fig. 14) from a broad-band time-switch core containing a 12K $\times$ 8 two-port SRAM subdivided into four 3K $\times$ 8 blocks. A single-memory BIST controller, shared between the four blocks, has also been successfully exercised at the 100-MHz clock rate.

VIII. Conclusions

The design of a synchronous, self-timed, modular memory family has been presented. This architecture is ideally suited to memories embedded within synchronous systems due to its simple interface specification and excellent speed/power performance. Optimization of the control system and prudent use of BiCMOS circuits have resulted in a family capable of 180-MHz operation in an ASIC process. The self-timed architecture results in clock duty-cycle independence, linear power versus frequency characteristics, as well as uniform power dissipation over a wide range of operating conditions. The generic port design is self-contained and extensible to any number of ports sharing access to a common core cell array. Modular DRAM has been implemented in the same architecture based on a 4-T cell which provides a 45% core cell density improvement over the one-port SRAM with similar performance.

The performance of the architecture has been verified in the implementation of a broad-band time-switch running at 100 MHz and a 3K $\times$ 8 two-port test block. Access and cycle times were measured at 4.8 and 4.1 ns, respectively. Functionality of the modular self-timed architecture has been verified on a modular memory test vehicle.

Appendix

DRAM Cell Comparison

An analysis of the 1-T and 4-T DRAM cells suitable for integration into the 0.8-μm BiCMOS process is presented below. Cell density, power, speed, and radiation tolerance characteristics were considered.

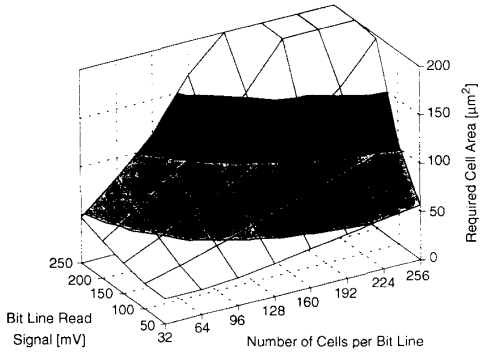


Fig. 15. 1-T DRAM cell size versus bit-line length.

Cell Size

A 1-T cell manufactured in an ASIC process is usually governed by a different set of electrical rules than those found in conventional commercial DRAM's, notably:

1) The drain–source and gate–source voltage on MOSFET transistors is limited to approximately 6 V, eliminating the use of boosted word lines.
2) The absence of thin oxides (< 10 nm) dedicated to the core array implies the use of larger area storage capacitors that use conventional MOS structures.
3) The absence of a heavy depletion implant for the storage capacitor (combined with 1, above) limits the low-level voltage stored in a PMOS 1-T cell to a substrate-bias-enhanced V_{tp} above ground.

A series of planar 1-T cell layouts was developed to yield a range of read signals when connected to a variable length bit line. The required area for the 1-T cell as a function of read signal amplitude and bit line length is shown in Fig. 15. The calculations assume full charge storage in the DRAM cell before charge sharing with the bit line and a midpoint bit-line reference level that is exactly halfway between a stored ONE and a stored ZERO.

A 4-T DRAM cell does not have the same size-versus-signal constraint placed on its design since it does not rely on charge sharing with the bit line to establish a read signal. A 73.2-μm^2 4-T cell was laid out using a local interconnect layer.

We assume the minimum acceptable read signal for a 1-T cell array to be 150 mV to provide sufficient margin for cell data leakage, bit-line common-mode variances, sense-amplifier mismatch, and supply variations. Therefore, to exceed the density of a 4-T cell array, a planar 1-T cell array can have no more than 100 cells per bit line (Fig. 15) for robust operation. Several subblocks of 1-T cells would therefore be required for large memory arrays (as compared to a contiguous 4-T cell array) thus further reducing the 1-T cell's density advantages.

α-Particle Upset Immunity

The minimum size of a 1-T DRAM cell is further constrained when one considers storage sensitivity to α particles. From the work done in [13]–[15], the amount of charge collected by various layouts of planar 1-T DRAM cells was calculated. As the area of the DRAM capacitor increases, its critical stored charge (Q_{crit}) increases, as does its ability to collect stray minority carriers resulting from an α-particle strike ($Q_{\max}$). These two values are plotted as a function of cell area in Fig. 16. The crossing of the two curves defines the minimum allowable cell area for a 1-T DRAM cell.

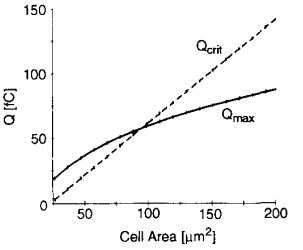


Fig. 16. Alpha-particle sensitivity of 1-T DRAM cell.

The calculations used in generating the data of Fig. 16 assume the following approximations:

1) α-particles radiated from Americium with maximum energies of 5 MeV were considered.
2) The relatively low resistivity of the N-tub in which the PMOS cells reside significantly reduces the drift "funnel" currents caused by the α-particle strike.
3) Incident angles of 20° to the silicon surface double the amount of collected carriers when compared to perpendicular strikes.
4) The buried layer at the bottom of the N-tub does not significantly reduce the number of collected carriers caused by very shallow angle strikes.
5) All 1-T DRAM cell layouts were approximated to circular collectors for the purposes of estimating the amount of collected charge.
6) The stored cell charge, Q_{crit}, is calculated as one half the maximum to allow for leakage over the refresh period.

From Fig. 16, one can see that a 1-T DRAM cell with an area greater than 95 μm^2 is required to ensure the minimum charge stored in the cell is greater than that collected by a shallow α-particle strike. By contrast, the 4-T DRAM cell, while storing a relatively low Q_{crit} (25 fC), has a much lower collection area (6.5 μm^2) than the storage capacitor in a 1-T cell. From the data presented in [13], one can estimate the $Q_{\max}$ of this cell to be 14 fC, well below the Q_{crit} limit. Also, the 4-T cell being differential will tend to collect a common-mode Q_{crit} which is far less damaging to the stored data integrity. A 4-T cell therefore appears more immune to α-particle-induced substrate/tub noise than a comparable planar 1-T cell.

Power

A 1-T cell DRAM design that employs "midpoint" sensing (defined here as being half way between stored ZERO and ONE levels in the cell) sees the bit-line pair precharged to this midpoint and then regeneratively driven to cell V_1 and V_0 levels by a cross-coupled sense amplifier. This occurs once per clock cycle. The total charge injected onto the bit lines

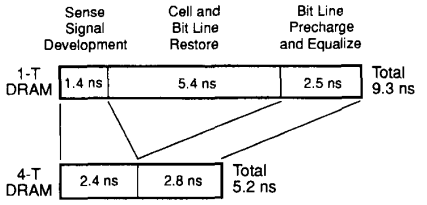


Fig. 17. Read cycle times of 1-T and 4-T DRAM cells.

from the positive supply per cycle is thus

$$Q_{1T} = C_{BL1T}\left(\frac{V_1 - V_0}{2}\right)$$

where C_{BL1T} is the total bit-line capacitance of the 1-T cell array. The corresponding charge per cycle for a 4-T cell array is

$$Q_{4T} = C_{BL4T} \times V_{\text{swing}}$$

where C_{BL4T} is the total bit-line capacitance of the 4-T cell array and is generally twice as large as C_{BL1T} since each bit line connects to twice as many cells. V_{swing} is the voltage swing of the ZERO bit line and, assuming the read conditions of the SRAM cell used in this design, V_{swing} will be approximately one-half of $(V_1 - V_0)/2$ of the 1-T cell array. Thus, the total bit-line read switching power for both cell types is approximately equal.

During write conditions, however, V_{swing} of the 4-T array tends to be 2.5 times that of the corresponding voltage swing of the 1-T array, making the write cycle power 5 times as great for the selected columns. For a 5-b Y decode, this amounts to roughly a 12% increase in write power for a 4-T array.

Speed

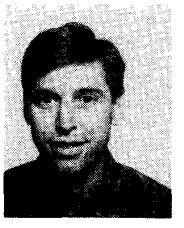
Fig. 17 shows the relative breakdown of simulated read cycle times for 64K 1-T and 4-T cell arrays. These clearly show that the 4-T DRAM cell is significantly faster than the 1-T based DRAM, largely due to the fact that the read is nondestructive in the 4-T case, and refresh occurs in parallel with bit-line signal development.

ACKNOWLEDGMENT

The authors would like to express their thanks to R. Gibbins for his help with the testing and characterization of the memories. Both R. Gibbins and D. Somppi made helpful comments on the manuscript.

REFERENCES

[1] R. Hadaway *et al.*, "A sub-micron BiCMOS technology for telecommunications," in *Proc. European Solid-State Device Res. Conf.* (Montreux), 1991, pp. 513–516.
[2] T. Dao and F. Sevejda, "A dual-port SRAM compiler for 0.8-μm 100K BiCMOS gate arrays," in *Proc. IEEE Custom Integrated Circuits Conf.*, 1991, pp. 22.4.1–3.
[3] H. Hana and S. Hussain, "High speed multi-port static RAM silicon compiler," in *Proc. IEEE Custom Integrated Circuits Conf.*, 1989, pp. 23.6.1–4.
[4] H. Shinohara *et al.*, "A flexible multi-port RAM compiler for datapath," in *Proc. IEEE Custom Integrated Circuits Conf.*, 1990, pp. 16.5.1–4.
[5] J. Tou *et al.*, "A submicrometer CMOS embedded SRAM compiler," *IEEE J. Solid-State Circuits*, vol. 27, no. 3, pp. 417–424, Mar. 1992.
[6] S. Shinagawa *et al.*, "A multi-speed digital cross-connect switching VLSI using new circuit techniques in dual port RAM's," in *Proc. IEEE Custom Integrated Circuits Conf.*, 1991, pp. 3.4.1–4.
[7] C. Ohno, "Self-timed RAM: STRAM," *Fujitsu Sci. Tech. J.*, vol. 24, no. 4, pp. 293–300, Dec. 1988.
[8] D. Lee and R. Katz, "Using cache mechanisms to exploit nonrefreshing DRAM's for on-chip memories," *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 657–661, Apr. 1991.
[9] B. Nadeau-Dostie *et al.*, "Serial interfacing for embedded memory testing," *IEEE Design Test Comput.*, pp. 52–63, Apr. 1990.
[10] T. Akioka *et al.*, "A 6-ns 256-kb BiCMOS TTL SRAM," *IEEE J. Solid-State Circuits*, vol. 26, no. 3, pp. 439–443, Mar. 1991.
[11] M. Takada *et al.*, "A 5-ns 1Mb ECL BiCMOS SRAM," *IEEE J. Solid-State Circuits*, vol. 25, no. 5, pp. 1057–1062, Oct. 1990.
[12] E. B. Eichelberger and T. W. Williams, "A logic design structure for LSI testability," in *Proc. 14th Design Automation Conf.*, June 1977, pp. 462–468.
[13] S. Kirkpatrick, "Modeling diffusion and collection of charge from ionizing radiation in silicon devices," *IEEE Trans. Electron Devices*, vol. ED-26, no. 11, pp. 1742–1753, Nov. 1979.
[14] C. M. Hsieh *et al.*, "A field-funneling effect on the collection of alpha-particle-generated carriers in silicon devices," *IEEE Electron Device Lett.*, vol. EDL-2, no. 4, pp. 103–105, Apr. 1981.
[15] S. Fu *et al.*, "Alpha-particle-induced charge collection measurements and the effectiveness of a novel p-well protection barrier on VLSI memories," *IEEE Trans. Electron Devices*, vol. ED-32, no. 1, pp. 49–54, Jan. 1985.



Allan L. Silburt (S'77–M'82) was born in Toronto, Ont., Canada, in 1957. He received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Ont., Canada, in 1980 and the M.Eng. degree in electronics from Carleton University, Ottawa, Ont., Canada, in 1982.

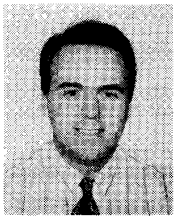
From 1982 to 1986 he was with Mosaid Inc., Carp, Ont., Canada where he worked on tools for the design, modeling and analysis of MOS memories and served as a company director. In 1987 he joined the Advanced Systems and Technology group at Bell-Northern Research Ltd., Ottawa, Ont., Canada, and worked on the development of built-in self test techniques for embedded RAM. From 1988 through 1991 he managed the memory development group in the design and deployment of several families of modular static, dynamic, and nonvolatile memory in CMOS and BiCMOS technologies. In 1992, he formed the Hardware Verification Technology group at BNR to work on the development of advanced methods for modeling and verification of silicon intensive systems.

Richard S. Phillips (M'88) was born in Windsor, Ont., Canada, in 1961. He received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Ont., Canada, in 1985.

From 1985 to 1989 he was with Mosaid Inc., Carp, Ont., Canada, where he worked on the design of 256-kb and 1-Mb DRAM's, application-specific SRAM's and video products. In 1989 he formed his own consulting company, Bayview Technologies Inc., Constance Bay, Ont., Canada, and works with North American silicon suppliers in the areas of memory, video, and analog circuit design.

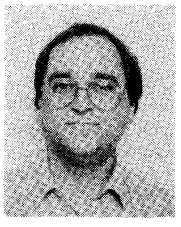
G. F. Randall Gibson (S'85–M'88) was born in Ottawa, Ont., Canada, in 1964. He received the B.Eng. degree in electrical engineering from Carleton University, Ottawa, Ont., Canada, in 1987.

He joined the design analysis group at Mosaid Inc., Carp, Ont., Canada, in 1987 where he was involved in design and failure analysis of SRAM, DRAM, and RAMDAC products. In 1989 he joined the memory development group at Bell-Northern Research, Ottawa, Ont., Canada, where he has been involved in the development of high-speed and low-power embedded memory including DRAM, multiport SRAM, ROM, and PROM.



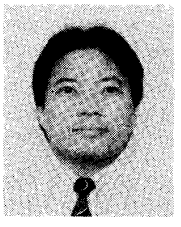
Steven W. Wood (S'85–M'92) was born in Toronto, Ont., Canada, in 1965. He received the B.A.Sc. and M.A.Sc. degrees in electrical engineering from the University of Toronto, Ont., Canada, in 1989 and 1992, respectively. His research at the University of Toronto focused on circuit design for ferroelectric nonvolatile memories.

In 1992, he joined Bell-Northern Research, Ottawa, Ont., Canada, where he is currently working on high-speed embedded BiCMOS SRAM design. His current interests include ferroelectric memory design and high-speed digital processing architectures.



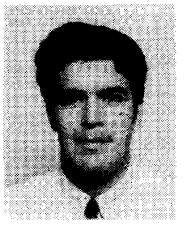
Armin G. Bluschke (M'83–S'86–M'87) was born in Hamilton, Ontario, Canada, in 1959. He received the B.Eng. degree from McMaster University, Hamilton, Ont., Canada, in 1981 and the M.S. degree from the University of California, Berkeley, in 1986.

From 1981 to 1984 and 1986 to 1990 he was with Gennum Corp. (LTI), Burlington, Ont., Canada, designing low-voltage bipolar amplifiers. Since 1990 he has been with Bell-Northern Research, Ottawa, Ont., Canada, in the memory development group. His present work includes the design of high-speed BiCMOS embedded memories as well as low-voltage CMOS SRAM.



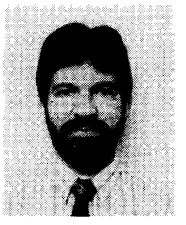
James S. Fujimoto (S'85–M'87) was born in Ottawa, Ont., Canada in 1963. He received the B.Eng. degree in electrical engineering from Carleton University, Ottawa, Ont., Canada, in 1986.

From 1986 to 1988 he was employed with Mosaid Inc., Carp, Ont., Canada, as a Design Analysis Engineer studying competitive commodity SRAM and DRAM designs. In 1988 he moved to Bell-Northern Research Ltd., Ottawa, Ont., Canada, where he has been involved in high-speed and low-power SRAM designs in two generations of the company's internal modular memory offerings. His current interests are in specialized application-specific memory functions for telecommunications and optimizations in CAD for IC design.



Stephen P. Kornachuk (S'88–M'91) was born in Boston, MA, in 1967. He received the B.A.Sc. degree in electrical engineering with a management option from the University of Ottawa, Ont., Canada, in 1990.

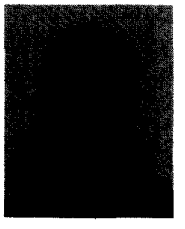
He has been with the memory development group at Bell-Northern Research, Ottawa, Ont., Canada, since joining as a student in 1989. Since that time, he has worked on the design of high-speed modular memories and memory test chips.



Benoit Nadeau-Dostie (S'80–M'85) received the Ph.D. degree in electrical engineering from the Université de Sherbrooke, Québec, Canada, in 1985.

He then joined the Department of Electrical Engineering of the Université Laval (Vision and Digital Systems Lab). Since 1986 he has been working for Bell-Northern Research, Ottawa, Ont., Canada, on the development of new design-for-testability techniques for communication systems. He has published several papers in the fields of design-for-testability (DFT) and built-in self-test (BIST) including four U.S. patents.

Dr. Nadeau-Dostie is a member of the Ordre des Ingénieurs du Québec. He has served on the program committee of several international conferences and workshops.



Rajesh K. Verma was born in Patna, India, in 1966. He received the B.A.Sc. degree in electrical engineering in 1988 from the University of Ottawa, Ont., Canada.

From 1989 to 1992, he was a member of the memory development group at Bell-Northern Research, Ottawa, Ont., Canada, where he worked on the design of CMOS and BiCMOS memories. Presently, he is working in the IC Design Group at ATI Technologies Inc., Scarborough, Ont., Canada.

Peter M. J. Diedrich (S'84–M'85) received the B.Eng. and M.Eng. degrees in electrical engineering from Carleton University, Ottawa, Ontario, Canada, in 1985 and 1989, respectively. Currently he is pursuing the MBA degree at the Massachusetts Institute of Technology in Cambridge, MA.

He was a member of the memory development group at Bell-Northern Research, Ottawa, Ont., Canada, where he worked on the development of high- speed synchronous SRAM's.