

Multipurpose DRAM architecture for optimal power, performance, and product flexibility

by W. F. Ellis
J. E. Barth, Jr.
S. Divakaruni
J. H. Dreibelbis
A. Furman
E. L. Hedberg
H. S. Lee
T. M. Maffitt
C. P. Miller
C. H. Stapper
H. L. Kalter

An 18Mb DRAM has been designed in a 3.3-V, 0.5- μ m CMOS process. The array consists of four independent, self-contained 4.5Mb quadrants. The chip output configuration defaults to 1Mb $\times$ 18 for optimization of wafer screen tests, while 3.3-V or 5.0-V operation is selected by choosing one of two M2 configurations. Selection of 2Mb $\times$ 9 or 1Mb $\times$ 18 operation with the various address options, in extended data-out or fast-page mode, is accomplished by selective wire-bonding during module build. Laser fuses enable yield enhancement by substituting eight 512Kb array I/O slices for nine in each quadrant of the 18Mb array. This substitution is independent in each quadrant and results in 1Mb $\times$ 16 operation with 2Mb $\times$ 8, 4Mb $\times$ 4, and 4Mb $\times$ 4 with any 4Mb independently selectable (4Mb $\times$ 4 w/4 CE). Input and control circuitry are designed such that performance margins are constant across output and

functional configurations. The architecture also provides for "cut-downs" to 16Mb, 4.5Mb, and 4Mb chips with I/O and function as above.

1. Introduction

In the past, DRAM products required few functional or I/O configuration options because mainframe computers were the primary application. Because of the explosive growth in portable and personal computing products, today's DRAM designs must provide functionality that covers the product spectrum from PDAs to mainframes. DRAM manufacturers must supply products that can satisfy various customer requirements for operating voltage, input/output configuration, and function while maintaining high performance and low power. This paper describes a DRAM architecture with design features that provide a single-chip design with the flexibility to meet these market demands.

By the early 1980s, DRAM designers had begun to offer enhanced functions such as page mode and nibble mode

©Copyright 1995 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/95/\$3.00 © 1995 IBM

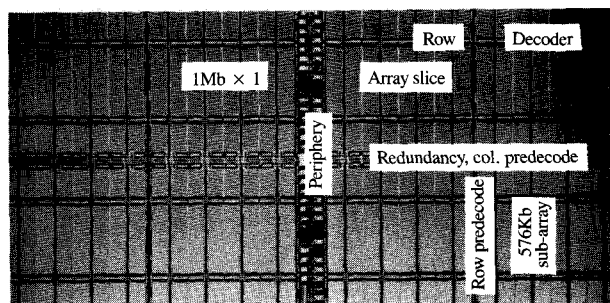


Figure 1

Micrograph of 18Mb DRAM chip.

Table 1 18/16Mb options.

I/O	Addressing	Function
×16/18	12/8	EDO
×16/18	12/8	Fast page
×16/18	10/10	EDO
×16/18	10/10	Fast page
×8/9	12/9	EDO
×8/9	12/9	Fast page
×8/9	11/10	EDO
×8/9	11/10	Fast page
×4.1CE	12/10	EDO
×4.1CE	12/10	Fast page
×4.1CE	11/11	EDO
×4.1CE	11/11	Fast page
×4.4CE	12/10	EDO
×4.4CE	11/11	Fast page
×1*	12/12	Static column
×1*	12/12	Fast page

*@ metal mask select.

[1], as well as demonstrating higher performance [2, 3]. With the move to CMOS technology, DRAMs were presented featuring wider I/O and low standby power [4]. As the DRAM market broadened in the late 1980s, limited functional and I/O configurations selectable by wire bond and/or metal mask were presented [5]. In the past, strong downward pressure on DRAM prices kindled interest in "cutting down" product designs to the bit densities of the previous generation. For example, a 16Mb DRAM chip can be cut down to a 4Mb chip that is of substantially smaller area than the maturing 4Mb technology can provide. Also, a properly architected 18Mb chip can be readily cut down to 16Mb, eliminating the cost and complexity of separate design efforts for each chip. The chip discussed here is manufactured in a 0.5- μ m CMOS process, using silicided polysilicon, two levels of metal, trench storage capacitors [6], and shallow-trench isolation [7, 8]. **Figure 1** is a micrograph of the 18Mb DRAM chip.

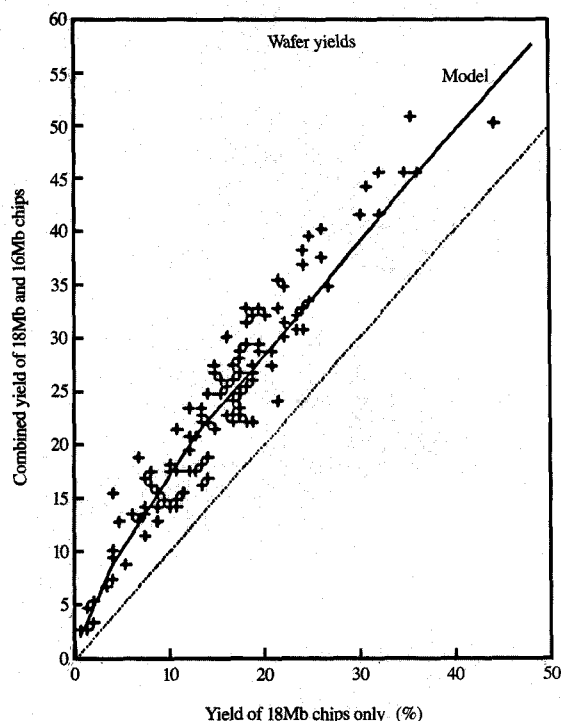


Figure 2

Yield enhancement of 18/16Mb chip during early production.

The chip is divided in half by the center vertical periphery area, which contains the pads, addressing, control, data steering, and I/O circuitry. This results in high-performance propagation of signals in that the longest net is equal to the height of the chip. Mask misalignment is minimized in the chip center regions, providing reduced defect sensitivity and tighter parametric distributions [9]. The peripheral circuits are designed using predefined second-metal signal nets and power buses. Input receivers are located immediately adjacent to the I/O pad [3]. The control and address signals are buffered into the left/right center horizontal stripe, in which word/bit redundancy and column predecode circuits are segmented to independently service the eight array sub-blocks. These nets are equivalent in length to those located in the center vertical region. The widths of nets serving large gate loads were customized to eliminate timing skews.

Each chip quadrant consists of a 4.5Mb array organized as 512 Kb $\times$ 9, with the data lines running horizontally to the center vertical peripheral region. Each data line fans into each of eight 576Kb quadrant sub-array blocks. The center vertical stripe in the quadrant contains two sets

of row predecode circuitry to service the double row of decode stripes which drive polycided word lines. This segmentation architecture optimizes the efficiency of fault-detection schemes such as parity and ECC in the computers using this chip. The chip functions and features are described in Section 2. Circuit design is described in Section 3, with subsections for address path, array design, and data path. Section 4 describes hardware results for the 18Mb DRAM and the 16Mb and 4.5Mb cut-downs.

2. Chip functions and features

The functional and input/output configurations of the 18Mb DRAM are selected via four program pads, PGM0–PGM3, located in the peripheral area of the chip. These pads are clamped either to V_{ss} or V_{cc} through long and narrow devices. The default state of the pads at the wafer level configures the chip as $\times 18$, low-power addressing (12/8) in fast page mode without write-per-bit (WPB), for low-power products. As shown in Table 1, the PGM0–PGM3 pads permit sixteen different functional and I/O configurations to be created by wire-bonding the desired pad(s) to the appropriate state. A second-level-metal (M2) mask is required to activate the $\times 4$ w/4CE and $\times 1$ options along with 3V/5V, static column mode, and WPB selection. The initial 18Mb DRAM design features laser fuses to provide 9/8 substitution in each 4.5Mb quadrant. Chips containing quadrants with a faulty 512Kb I/O slice can thus be reconfigured as a 16Mb chip. This fault-tolerance technique increases yield [10], as shown in Figure 2. The resulting chips are then put into the early 400-mil JEDEC standard packages for 18/16Mb DRAMs. The chip architecture also provides for the efficient cut-down of the 18Mb design to 16Mb. Toward this goal, the hierarchical chip design data are nested such that two $1\text{Mb} \times 1$ array slices, centered between the row decode stripes and the contiguous periphery, can be deleted from the full-chip data file. Because of predefined wiring channels in the center vertical periphery, the resultant 4Mb segments can then be stepped and remerged with the rest of the chip. With the two resulting chip designs, the 9/8 steering circuitry is replaced with self-timed refresh (sleep mode) circuits. The laser fuses are now used to program the STR frequency. The minimal time required to run design ground-rule and physical/logic checking programs then defines the 16Mb DRAM design schedule. The schedule for functional qualification of the 16Mb cut-down is similarly reduced because the I/O and control circuitry will already have been qualified on the 18Mb chip. The 16Mb cut-down DRAM chips are put into the present 300-mil JEDEC standard packages with functional selection via the chip PGM pads.

By eliminating six of the eight array sub-blocks in each quadrant, the 18Mb DRAM can also be cut down to 4.5 Mb, as shown in Figure 3. Because of package requirements, the center vertical periphery must be

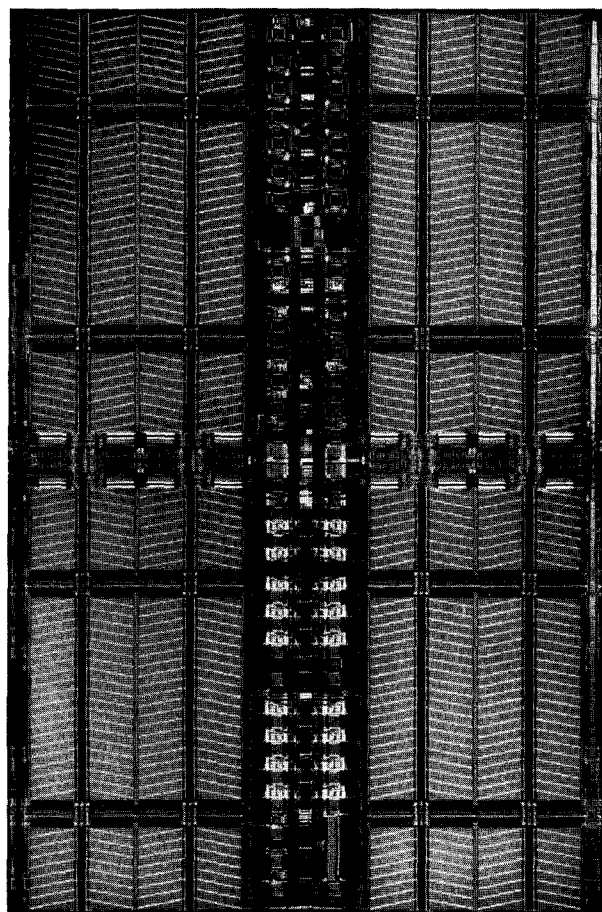


Figure 3

Micrograph of 4.5Mb DRAM chip.

Table 2 4.5Mb DRAM options.

Organization	Standard		Package
	Address	Page depth	
256K $\times$ 18	9/9	512	400 $\times$ 725-mil TSOP-44/40 400 $\times$ 1025-mil SOJ-40
256K $\times$ 16	9/9	512	400 $\times$ 725-mil TSOP-44/40 400 $\times$ 1025-mil SOJ-40
512K $\times$ 9	10/9	512	400 $\times$ 725-mil SOJ/TSOP-28
512K $\times$ 8	10/9	512	400 $\times$ 725-mil SOJ/TSOP-28
1M $\times$ 4	10/10	1024	300 $\times$ 675-mil SOJ/TSOP-26/20
1M $\times$ 4, 4CE	10/10	1024	300 $\times$ 675-mil SOJ/TSOP-26/20

redesigned such that address pads/circuits are located toward the top and the 18 data I/Os are toward the bottom. Because of the design symmetry and predefined wiring in this region and the independence of the array sub-block

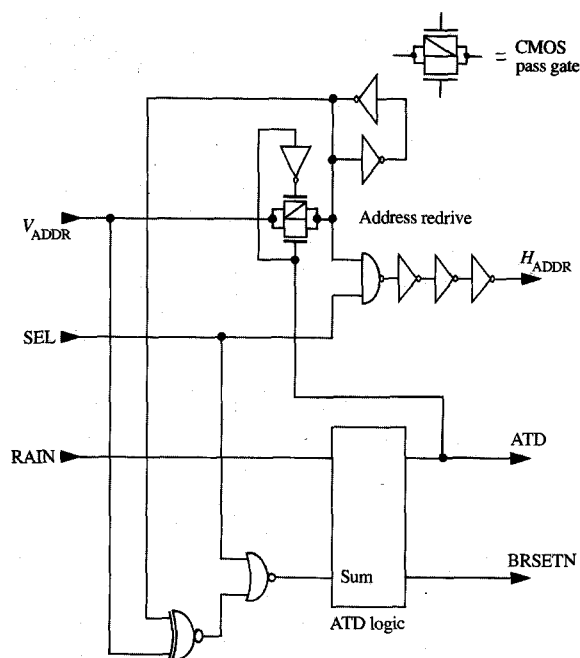


Figure 4

Self-interlocked ATD circuit.

design, a 45.3-mm^2 ($5.5 \times 8.23\text{-mm}$) 4.5Mb DRAM (shrinkable to 36.8 mm^2) was created in four months by a single engineer. Utilization of the PGM pads gives the functional capabilities shown in Table 2.

3. Chip functional design

The flexibility of this architecture to provide multiple functional and design cut-down options requires careful circuit design for optimizing power, performance, and reliability. The power-and-ground distribution is designed to reduce noise that can affect reliable chip operation. By replicating the DRAM storage trench and pass gate, a 50-pF decoupling capacitor can be constructed in $115 \times 35\text{ }\mu\text{m}$. Inclusion of the DRAM cell pass gate in the decoupling capacitor structure provides current limiting in the event of a trench-to-substrate defect. The trench dielectric reliability is more than 300 times that of the planar dielectric, based on field return data. Therefore, designing the decoupling capacitor to be structurally identical to the DRAM array provides high yield and reliability. Such capacitors are placed at each input pad and off-chip driver pad, providing 2.5 nF of local V_{DD} decoupling in the center vertical peripheral region. Precharge of the bit lines in the p-array to V_{DD} [9–11] provides a series-equivalent decoupling capacitance of

7 nF to the global chip-power bus. Timing skews are minimized by custom design of the individual signal nets and the associated driver such that $T_{\text{RISE}} = T_{\text{FALL}}$. Optimized net performance minimizes power and ground noises created by transient, undefined states on logic circuit inputs. Use of control clocks statically timed from the input pad, together with carefully designed timing interlocks, provides high performance while eliminating functional sensitivity to process-parametric variations. For integrated circuit fabricators, the density of smaller defects is generally substantially greater than that of larger defects [12–14]. Redundant word/bit lines integrated into the array are used to fix failures caused by these defects. However, up to 20% of the chip area can consist of peripheral circuits to which these fault-tolerance techniques are not efficiently applicable. Therefore, in the peripheral regions, relaxed design ground rules provide lower defect densities and, therefore, higher yield and reliability [15–17]. These architectural features combined with the following circuit techniques result in a chip with extremely high circuit-limited yields.

• Address path

The address receiver circuits are located immediately adjacent to the associated I/O pad, providing reduced input capacitance loads. This also eliminates coupling-induced receiver input-level sensitivities and timing skews, which result when card-signal nets are bused on-chip to distant receivers. The address pad true state is buffered and then propagated in the chip center vertical peripheral region. This address bus is then buffered into either the left or the right center horizontal region over the redundancy circuits, and further buffered into word predecode circuits in the center vertical quadrant region. The word predecode is enabled by row address interlock (RAIN), which is generated by a physical copy of a word redundancy circuit, located at the left or right chip edge. Therefore, in spite of process-parametric variations and RC net delays, the slowest redundant word select signal will always be valid before word predecode is enabled. The RAIN interlock is also propagated to the chip center vertical peripheral region to enable the address transition detect (ATD) circuit and also to enable column addressing onto the address bus.

Figure 4 shows the ATD circuit used to reliably detect column address transitions. The ATD circuit is integrated into the left/right address redrive circuits located in the center of the chip. After the row address is decoded, the input pass gate to the address redrive latch is closed, isolating the latch state from the center vertical address bus state. These two states form the input to the XOR gate, which detects an address transition when the address net and latch states are unequal. The address transitions are then summed by a static NOR to enable the BRSETN

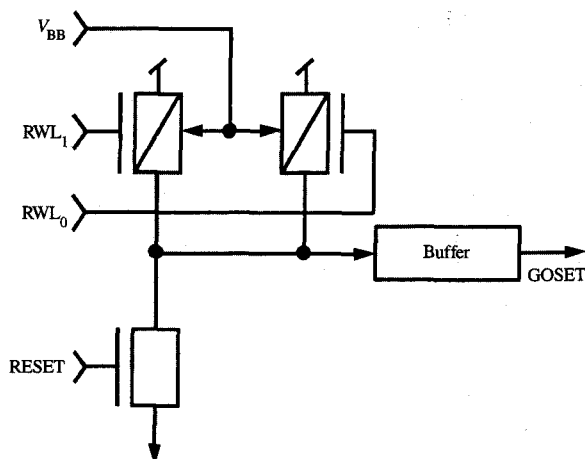


Figure 5

Word line interlock system.

(bit-reset-NOT) signal, which resets the column predecode circuits in response to an address transition and disables the column predecodes. At this time, address redrive latches are reconnected to the vertical address bus to update the center horizontal address bus. Integration of the ATD into the centrally located address redrive circuits and utilization of a single-state address-bus architecture results in enhanced performance with reduced area and power.

• Array design

The array is segmented to reduce the types of failure that can propagate across chip I/O sub-arrays. This is especially important for the efficient operation of parity and ECC in systems using DRAM products with byte-wide ($\times 8$, $\times 9$) or word-wide ($\times 16$, $\times 18$) output configurations. The substrate plate trench cell (SPT) [6] technology provides a 70-fF storage cell for a bit line capacitance of 250 fF. The p-array bit lines are precharged to V_{DD} to provide high performance, high reliability, and low power [9–11]. With bit lines and word lines precharged to the same bias level, defects that cause word line-to-bit line shorts do not contribute to chip standby current. In array designs where the word line and bit line are precharged to different bias levels, e.g., half- V_{DD} bit line precharge, these word line-to-bit line defects can affect the bit line precharge bias level, degrading array signal development and sense-latch sensitivity. Another consequence is unfixable standby-current yield loss in low-power parts [18].

Full- V_{DD} bit line precharge provides higher voltage overdrive of the array-cell transfer device, resulting in

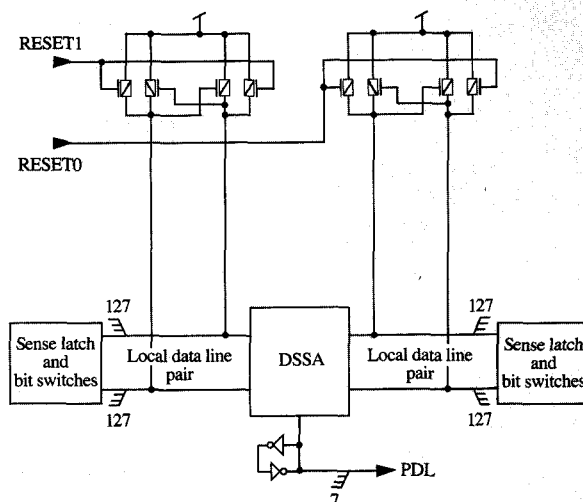


Figure 6

Local data line system.

faster coupling of the stored data onto the bit lines. A word line interlock system, shown in Figure 5, is designed to accurately time the development of the bit line data signal. The reference word lines (RWL_0 , RWL_1) are used as inputs to exact replicas of the array transfer device, thereby allowing bit line signal development to be accurately timed across electrically, thermally, or process-parametrically induced variations in word line RC delay and/or device V_t . Accurate word line interlocking is a key aspect of proper DRAM design and is required for enhanced yield, reliability, and performance.

Because of the optimal sense-latch sensitivity that results from full- V_{DD} bit line precharge, activation of the sense clocks by the word line interlock results in rapid and efficient amplification of the bit line signal. After amplification, while the selected array word line remains at V_{SS} , the selected and unselected reference word lines are equalized together to $1/2 V_{DD}$ for the remaining duration of the array select time. At the initiation of an array restore, the reference word lines are clamped to V_{SS} as the array word line is boosted below V_{SS} . This provides a controlled write-back for the reference and array cells, resulting in consistent stored voltage levels. This eliminates late-write-induced array data pattern sensitivities. Word line boost at the initiation of array restore also reduces the duty factor for boosted-voltage-induced oxide stresses, resulting in improved array reliability. Because of the high overdrive that results from V_{DD} precharge, p-FET devices provide high-performance array restore and equalization of the selected array bit lines. The unselected arrays substantially

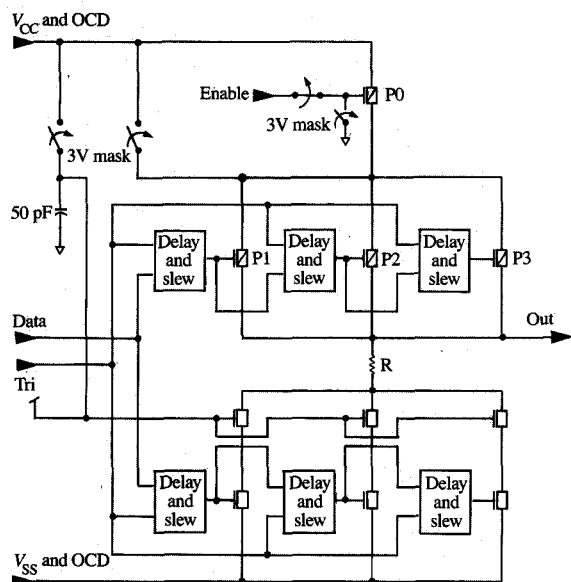


Figure 7

Off-chip driver schematic.

decouple the V_{DD} bus, reducing noise and allowing high-performance operation of the 18Mb DRAM array in the various product address-configuration options.

• Data path

Because the chip is architected as $1\text{Mb} \times 18$, each quadrant contains nine data I/O sub-arrays, each serviced by a single primary data line (PDL). Each PDL fans into eight digital secondary sense amplifiers (DSSA) located in the quadrant array sub-blocks. The DSSA serves two local data line pairs, each of which services 128 sense latches. The DSSA, shown schematically in **Figure 6**, results in low-power, high-performance transfer of the sense-latch data to the PDL. The local data lines feature a p-MOS half-latch and devices to reset the local data lines to V_{DD} upon detection of an address transition. Selection of a bit switch discharges the local data line true or complement through the sense latch, which results in the digital transfer of the sense-latch data to the lightly loaded local data line pair. The local data line state is then buffered onto the 2.5-pF PDL through the tri-state driver. From the selection of the bit switch, transfer of the sense-latch data to the PDL occurs in 3.0 ns for devices made by the nominal process at 2.9 V and 85°C. A small latch maintains the PDL data state after the PDL driver sets it, in response to the next column-address access. The PDL state does not change unless the newly addressed sense latch contains the

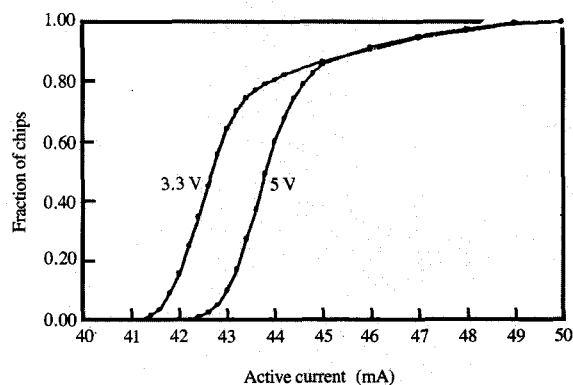


Figure 8

Measured active current distributions.

opposite data state. This eliminates a major source of power consumption in word-wide ($\times 16$ or 18) DRAMs, the selection and reset of primary data lines during column addressing.

The nine quadrant PDLs are wired to the $9/8$ quadrant steering circuits. A laser fuse disables the data I/O circuit associated with the $1\text{Mb} \times 1$ array slices located along the top and bottom chip edges in **Figure 1**. Thus, a single laser fuse reconfigures the 18Mb DRAM for 16Mb operation (which, in spite of the $9/8$ array operating current overhead, still meets industry specifications for power/performance). The flexibility to "steer out" defective I/O sub-arrays is provided by a bank of eight additional laser fuses for each quadrant. This design feature is most feasible in arrays that have been carefully designed and segmented to minimize failure types that affect more than one of the nine quadrant data I/O sub-arrays. By utilizing this fault-tolerance technique [10], 16Mb product yield is obtained from nearly all good 18Mb chips. As shown in **Figure 2**, early program yields were improved by up to 75% by utilization of this design feature.

With the $1\text{Mb} \times 18$ chip laser-reconfigured as $1\text{Mb} \times 16$, additional addressing, function, and output configuration options are provided by wire-bonding of the appropriate program pads (PGM0-3). For example, wiring PGM1 to V_{CC} at module build reconfigures the laser-steered $1\text{Mb} \times 16$ to $2\text{Mb} \times 8$ with addressing reconfigured to 12 row and 9 column bits. The ninth column address bit allows 8-to-4 decoding of the quadrant data I/O arrays. If a module organized as $4\text{Mb} \times 4$, 1CE is desired, PGM2 is wire-bonded to V_{CC} at module build. The addressing reconfigures to 12 row and 10 column bits with column

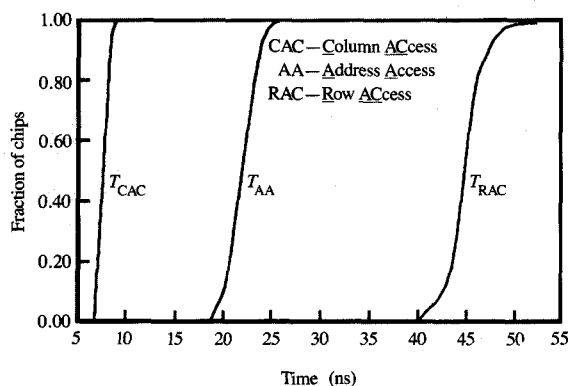


Figure 9

Timing distributions observed on 16Mb DRAM chips.

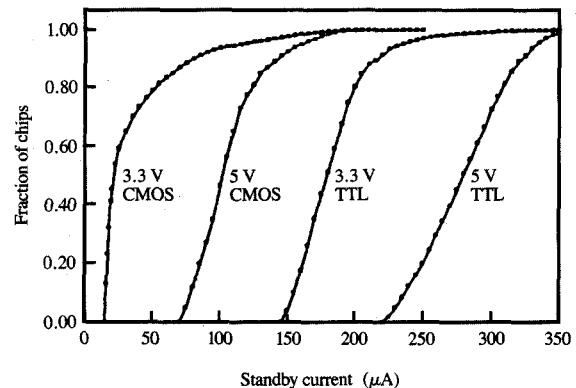


Figure 10

Measured standby current distributions.

addresses 9 and 10, allowing 8-to-2 decoding of the quadrant data I/O. Table 1 illustrates the ability of this design, in conjunction with the flexibility of lead-on-chip packaging technology [19, 20], to provide multiple products from a single 18Mb DRAM chip design.

The flexibility of this design to provide high performance with multiple module-output configuration options required careful design of the total off-chip driver (OCD) circuit/chip/package system, including consideration of 3-V or 5-V operation. The JEDEC standard module pin-outs for 16/18Mb DRAMs organized as $\times 1$, $\times 4$, $\times 8/9$ provide two V_{CC} and two V_{SS} pins. For $\times 16/18$ organization, an extra V_{CC} and V_{SS} pin are added to alleviate internal module noise from switching the extra I/O loads. The eighteen OCDs are located at the pads along the right side of the center vertical peripheral region, shown in Figure 1. The OCDs share a common V_{SS} bus which is separated at the V_{SS} pads from the global ground bus, to reduce IR drops and transient di/dt -induced noise that can affect logic and input receiver performance.

For 5-V operation, the final metal mask provides a V_{CC} bus tied to the V_{CC} pads and shared by the OCDs, which is separate from the global V_{DD} power bus. Voltage regulators supply 3.3 V to the global V_{DD} power bus, which is substantially decoupled by the array and the 50-pF trench decoupling capacitors located at each peripheral pad. The OCD, shown in Figure 7, features three stages with input slew-rate control of the individual stage inputs. A staged output drive and input slew-rate control reduce parasitic V_{SS} and V_{CC} noise induced when the module outputs are switched from nominal voltage levels. However, for reliable switching of outputs that have

precharged to the maximum voltage levels, series damping resistors are required to moderate di/dt -induced V_{SS} and V_{CC} noise. Transistor P0 acts as a series damping resistor between V_{CC} and P1-3, the output pull-up stages. A damping resistor in series with the three output pull-down stages moderates parasitic noises on the separate V_{SS} bus which is shared by the OCDs.

The alternate final metal mask, for 3-V operation, disables the voltage regulators and connects the V_{CC} and global V_{DD} buses. The series damping device P0 is disabled, and the V_{CC} bus is wired directly to the sources of P1-3, along with 50 pF of direct local decoupling and an additional 7 nF of global decoupling provided by the unselected p-arrays. This configuration results in low noise on the power bus, which provides high performance for the 3-V products. A data line interlock (DLINT) is used to ensure efficient access of the array data by holding the OCD in tri-state until the OCD input latch is updated with valid data. The DLINT is constructed by locating an extra primary data line (PDL) in each quadrant, near the chip edge. This PDL fans out of tri-stateable, dummy digital secondary sense amplifiers (DSSA) in each of the eight quadrant array sub-blocks. The local data line pairs closest to the chip edge are NANDed to detect a data transition, which is buffered from the selected array sub-block onto the DLINT net. Left/right DLINT nets are combined at the top/bottom of the center vertical periphery to time the top/bottom bank of nine OCDs. Since the DLINT interlock comes from the array providing the data, it will track with electrically, thermally, and process-parametrically induced variations in data path timing. This results in accurate timing of the OCD for reliable high-performance operation.

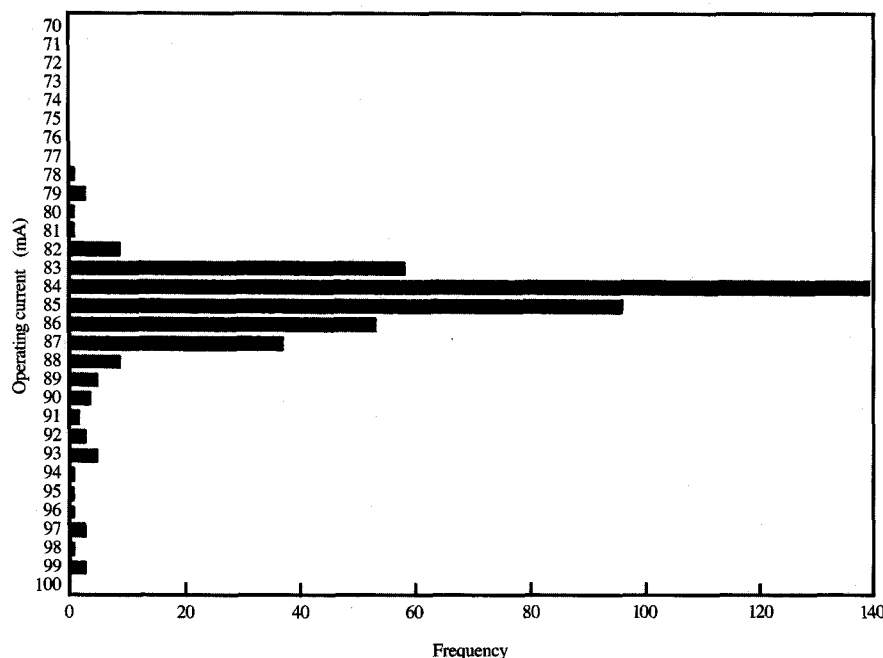


Figure 11

256Kb $\times$ 18 measured current.

4. Hardware results

Hardware results for the 16Mb DRAM, 1Mb $\times$ 16, 12/8 addressed device, under worst-case operating conditions for a 150-ns cycle, are shown in Figures 8 and 9. The 18Mb chips demonstrate similar performance with approximately 9% higher active operating current. Figure 10 shows worst-case results for 3-V and 5-V chips measured with CMOS and TTL input levels. The majority of the 3-V standby current distribution for CMOS input levels, which approaches 10% of the industry specification of 200 μ A, demonstrates that this design is well suited for low-power applications such as portable electronic equipment. Chips that are configured into the other 18/16Mb product options also show similarly superior results.

The 1Mb $\times$ 18 DRAM, as described above, was also cut down to provide a 256Kb $\times$ 18, 9/9 addressed device, which was manufactured for both 3-V and 5-V applications. Figures 11 and 12 illustrate the functionality to worst-case voltage, temperature, and loading for a 150-ns cycle. The active current is half of the industry specification, with a 30% faster first access time. Utilization of the PGM pads at module build provides the product configurations shown in Table 2. This part, if

fabricated in the present 0.5- μ m CMOS technology, would be 36.8 mm², quite possibly the smallest 4.5Mb DRAM in the world.

Conclusions

The architecture and design techniques implemented on this chip provide substantial functionality and flexibility. The lead-on-chip packaging technology [19, 20] utilizes these capabilities to provide multiple product options. The careful design and segmentation of the array permit the use of 9/8 steering in the quadrants to provide 16Mb product yield from nearly all good 18Mb chips. Forethought in the architecture of the design data has resulted in a straightforward methodology for cut-down to a 16Mb chip design for enhanced productivity. This also allows the capability to cut down the 18Mb chip to a highly functional 4.5Mb DRAM chip. The chip periphery is customized for high performance as well as high yield. Reliable functionality and insensitivity to electrically, thermally, and/or process-parametrically induced timing skews are provided by carefully designed timing interlocks. Use of the physical chip structures, such as reference word lines, ensures the precision of the interlock design. Precharge

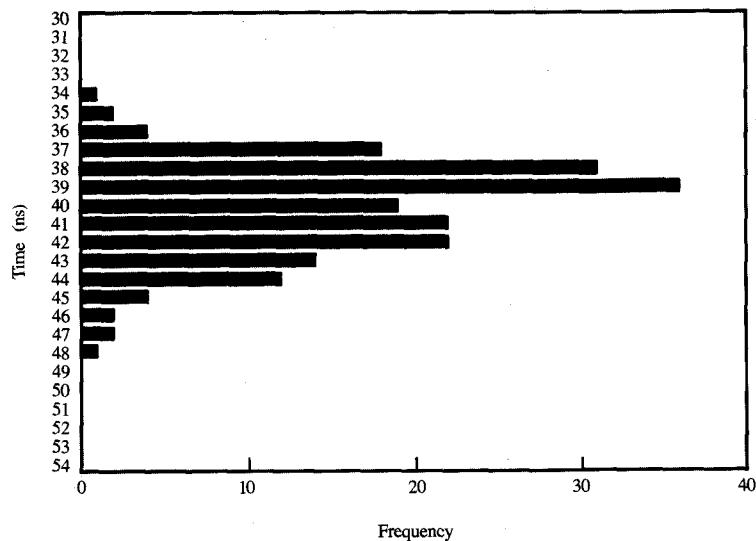


Figure 12

256Kb x 18 first access.

of the p-array bit lines to V_{DD} results in reliable, high-performance, low-power operation. The innate strengths of this design and architecture have been demonstrated by fully functional module operation at 60-ns performance, in a complete 2.5-V manufacturing test [21]. These architecture and design techniques result in a single 18Mb chip design capable of providing DRAM products of bit-wide to word-wide output configuration with multiple functional and power supply options, demonstrating both low power and high performance. The chip design also provides for cut-downs to 16Mb and 4.5Mb DRAM products with functions and features as above. This chip, whose properties are summarized in **Table 3**, demonstrates the ability to cover the broad spectrum of today's DRAM product requirements.

Acknowledgments

The authors would like to express their appreciation to the IBM Burlington facility product characterization, mask manufacture, process development, device development, and manufacturing areas, without whose efforts the hardware verification would not have been possible.

References

1. K. Fujishima, H. Ozaki, H. Miyatake, S. Uoya, M. Nagatomo, K. Saito, K. Shimotori, and H. Oka, "A

Table 3 Multipurpose DRAM product characteristics.

Chip	18 Mb → 16 Mb, 4.5 Mb (cutdowns)
Technology	0.5- μ m CMOS; two-level metal (Al), polycide gate (W)
Memory cell	2.52 $\times$ 1.33 μ m ² (trench)
Chip size	7.34 $\times$ 15.25 mm ² (18 Mb), 6.69 $\times$ 15.25 mm ² (16 Mb)
Supply voltage	5 V, 3.3 V (2.5 V)
Access time	45 ns @ 2.9 V, 85°C
Active current	43 mA @ 3.7 V, 10°C
Standby current	100 μ A @ 5.6 V, 30 μ A @ 3.7 V

- 256K Dynamic RAM with Page-Nibble Mode," *IEEE J. Solid-State Circuits* SC-18, No. 5, 470-477 (October 1983).
2. D. Galloway, B. Hartman, and D. Wooten, "64K Dynamic RAM Speeds Well Beyond the Pack," *Electron. Design*, pp. 221-225 (March 19, 1981).
3. H. Kalter, P. Coppens, W. Ellis, J. Fifield, D. Kokoszka, T. Leasure, C. Miller, Q. Nguyen, R. Papritz, C. Patton, J. Poplawski, S. Tomashot, and W. Van Der Hoeven, "An 80-ns 1-Mb DRAM with Fast-Page Operation," *IEEE J. Solid-State Circuits* SC-20, No. 5, 914-923 (October 1985).
4. H. Kawamoto, T. Shinoda, Y. Yamaguchi, S. Shimizu, K. Ohishi, N. Tanimura, and T. Yasui, "A 288K Pseudostatic RAM," *IEEE J. Solid-State Circuits* SC-19, No. 5, 619-623 (October 1984).
5. K. Shimohigasi, K. Kimura, Y. Sakai, H. Tanaka, K. Yagi, M. Ishihara, K. Miyazawa, S. Shimizu, and

J. Murata, "A 65-ns CMOS DRAM with a Twisted Driveline Sense Amplifier," *1987 IEEE International Solid-State Circuits Conference, Digest of Technical Papers*, pp. 18-19 (1987).

6. D. Kenney, E. Adler, B. Davari, J. De Brosse, W. Frey, T. Furukawa, P. Geiss, D. Harmon, D. Horak, M. Kerbaugh, C. Koburger, J. Lasky, J. Rembetski, W. Schmittek, and E. Sprogis, "16-Mb Merged Isolation and Node Trench SPT Cell (MINT)," *Symposium on VLSI Technology, Digest of Technical Papers*, pp. 25-26 (May 1988).
7. B. Davari, C. Koburger, T. Furukawa, Y. Taur, W. Noble, A. Megdanis, J. Warnock, and J. Mauer, "A Variable-Size Shallow Trench Isolation (STI) Technology with Diffused Sidewall Doping for Submicron CMOS," *International Electron Devices Meeting (IEDM), Digest of Technical Papers*, pp. 92-95 (December 1988).
8. P. Bakeman, A. Bergendahl, M. Hakey, D. Horak, S. Luce, and B. Pierson, "A High Performance 16-Mb DRAM Technology," *Symposium on VLSI Technology, Digest of Technical Papers*, pp. 11-12 (June 1990).
9. W. Ellis, H. Kalter, and C. Stapper, "Design for Reliability, Testability, and Manufactureability of Memory Chips," *Proceedings of the Annual Reliability and Maintainability Symposium*, 1993, pp. 311-319.
10. T. Kawada, Y. Takahashi, N. Tsuda, M. Waki, and N. Hagiwara, "A Pattern Matching Processor Array with Defect Tolerance," *IEEE International Solid-State Circuits Conference, Digest of Technical Papers* **29**, 90-91 (1986).
11. H. Kalter, C. Stapper, J. Barth, J. Di Lorenzo, C. Drake, J. Fifield, G. Kelley, S. Lewis, W. Van Der Hoeven, and J. Yankowsky, "A 50-ns 16-Mb DRAM with a 10-ns Data Rate and On-Chip ECC," *IEEE J. Solid-State Circuits* **25**, 1118-1128 (October 1990).
12. C. Stapper, F. Armstrong, and K. Saji, "Integrated Circuit Yield Statistics," *Proc. IEEE* **71**, 453-470 (April 1983).
13. H. Parks and A. Burke, "The Nature of Defect Size Distributions in Semiconductor Processes," *Proceedings of the International Semiconductor Manufacturing Science Symposium*, May 1989, pp. 131-135.
14. R. Glang, "Defect Size Distribution in VLSI Chips," *IEEE Trans. Semicond. Manuf.* **4**, 265-269 (November 1991).
15. C. H. Stapper, "Modeling of Defects in Integrated Circuit Photolithographic Patterns," *IBM J. Res. Develop.* **28**, No. 4, 461-475 (July 1984).
16. A. Ferris-Prabhu, "Role of Defect Size Distributions in Yield Modelling," *IEEE Trans. Electron Devices* **ED-32**, No. 9, 1727-1736 (September 1985).
17. C. Kooperberg, "Circuit Layout and Yield," *IEEE J. Solid-State Circuits* **23**, No. 4, 887-892 (August 1988).
18. G. Kitsukawa, M. Horiguchi, Y. Kawajiri, T. Kawahara, T. Akiba, Y. Kawase, T. Tachibana, T. Sakai, M. Aoki, S. Shukuri, K. Sagara, R. Nagai, Y. Ohji, N. Hasegawa, N. Yokogama, T. Kisu, H. Yamashita, T. Kure, and T. Nishida, "256-Mb DRAM Circuit Technologies for Field Applications," *IEEE J. Solid-State Circuits* **28**, No. 11, 1105-1113 (November 1993).
19. R. Pashby, D. Phelps, S. Samuelson, and W. C. Ward, "Package Semiconductor Chip," U.S. Patent 4,862,245, August 29, 1989.
20. W. C. Ward, "Volume Production of Unique Plastic Surface Mount Modules for the 80-ns 1-Mb DRAM Chip by Area Wire Bond Technique," *Proceedings of the 38th IEEE Electronic Components Conference*, 1988, pp. 552-557.
21. W. Ellis, A. Adler, and H. Kalter, "A 2.5-V 16-Mb DRAM in 0.5- μ m CMOS Technology," *IEEE Symposium on Low Power Electronics, Digest of Technical Papers*, pp. 88-89 (1994).

Received June 6, 1994; accepted for publication November 18, 1994

Wayne F. Ellis *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (WELLIS at BTVLABVM)*. Dr. Ellis joined the IBM System Products Division in 1974 as a graphics technician after he received the A.A.S. in electrical technology from Hudson Valley Community College, Troy, New York. After receiving the B.S.E.E. in 1977 from Union College, Schenectady, New York, he joined the IBM development laboratory in Essex Junction, Vermont. Dr. Ellis attended the University of Vermont on a Resident Study Fellowship from 1989 to 1991 and 1992 to 1993; he received the M.S.E.E. in 1991 and the Ph.D. in materials science in 1993. In 1991-1992, he returned to work on the architecture and design of the 18/16Mb DRAM product. His work at IBM has dealt with design and development of DRAM technology test chips, DRAM products, and technology structures for fault tolerance. He has worked extensively on low-power product design techniques, as well as advanced circuit and system functional integration issues. He has also been involved in the development of Design for Reliability, Testability, and Manufacturability techniques, and has published several papers on this topic. Dr. Ellis holds nine patents and has more than 35 publications.

John E. Barth, Jr. *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (JBARTH at BTVLABVM)*. Mr. Barth received the B.S.E.E. degree from Northeastern University, Boston, in 1987, and the M.S.E.E. degree from National Technological University (NTU), Fort Collins, Colorado, in 1992. During his B.S. degree work, he was a cooperative student from 1984 to 1985 at the Timeplex Development Laboratory in Rochelle Park, New Jersey, where he wrote data communications software and network monitoring software. He also held a cooperative student assignment in 1986 at the IBM development laboratory in Essex Junction, Vermont, where he was involved in the design and characterization of the 1Mb DRAM product. After receiving his B.S. degree, he joined the IBM development laboratory in Essex Junction as a full-time employee working on the design of a 16Mb DRAM product featuring embedded ECC and SRAM cache. Following this he worked on array design for the 16/18Mb DRAM product. Mr. Barth's most recent responsibilities include design automation tool development and architecture and the design of an advanced logic ASIC featuring 24Mb of imbedded, wide-I/O, high-performance DRAM.

Sri Divakaruni *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (SDIVAKAR at BTVVMOFS)*. Mr. Divakaruni received the B.S. degree in materials and metallurgical engineering from the Indian Institute of Technology, Madras, in 1986 and the M.S. in electrical engineering from Rensselaer Polytechnic Institute in 1988. Upon graduation, he joined the IBM development laboratory in Essex Junction, Vermont, and worked on bipolar CMOS technology development before joining the Advanced Memory Design team. Mr. Divakaruni is currently the manager of the IBM Memory Development team that designs and develops 18/16Mb derivative memory products, including 3D memory and macros for integrated cache applications.

Jeffrey H. Dreibelbis *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (B224833 at BTVLABVM).* Mr. Dreibelbis received the B.S. degree in electrical engineering from Lehigh University, Bethlehem, Pennsylvania, in 1973. That same year, he joined the U.S. Air Force, serving as a communications electronics engineer for the Air Force Communications Service at Griffiss AFB, Rome, New York. In 1977 Mr. Dreibelbis joined the semiconductor development laboratory of the IBM Burlington facility, where he first worked in n-MOS SRAM and DRAM product design. Most recently, he has worked on embedded CMOS SRAM development projects and on the CMOS 16Mb DRAM product development team. He is currently a senior engineer in Advanced Memory Design. Mr. Dreibelbis is a member of Tau Beta Pi and Eta Kappa Nu.

Anatol Furman *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452.* Mr. Furman received his B.S. in electrical engineering from the University of Massachusetts in 1963. He joined the IBM Data Systems Division in Poughkeepsie, New York, also in 1963, to work on the design of discrete (magnetic core) memories. In 1970 he was assigned to the IBM Burlington facility, where he worked on 1Kb through 64Kb DRAM products. From 1979 to 1990, he was involved in the design and development of RISC microprocessors; he later worked on 18Mb and 16Mb DRAM products and, prior to his retirement in 1994, DSP products.

Erik L. Hedberg *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (ERIK at BTVLABVM).* Mr. Hedberg received the B.S. degree in electrical and biomedical engineering from Worcester Polytechnic Institute, Worcester, Massachusetts, in 1978. He also received the M.S. degree in biomedical engineering from the University of Miami, Coral Gables, Florida, and the M.S.E.E. degree from Duke University, Durham, North Carolina, in 1980 and 1982, respectively. In 1983 Mr. Hedberg joined the IBM Burlington facility in Essex Junction, Vermont, as an SRAM designer. In 1987 he began work in defining and designing array built-in self-test (ABIST) for various static RAMs. In 1991, as a staff engineer, he joined the DRAM design team to produce the first IBM vendor-identical 16Mb chip. He is currently the lead design engineer for DRAM and SRAM memory cube development.

Hsing-San Lee *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (HSLEE at BTVLABVM).* Dr. Lee received the B.S. degree in electrical engineering from the National Taiwan University in 1956 and the M.S. and Ph.D. degrees in electrical engineering from Ohio State University, Columbus, in 1960 and 1964, respectively. He was an assistant professor at Ohio State University from 1964 to 1965 and later a member of the technical staff at Bell Telephone Laboratories, Whippany, New Jersey. In 1968, he joined IBM East Fishkill, New York, in the area of bipolar device development. He later joined the IBM General Technology Division in Essex Junction, Vermont, to work on exploratory FET development and advanced high-performance memory. In 1985, Dr. Lee was assigned to the development of advanced high-performance CMOS SRAM and in 1988 to fault-tolerance in high-density DRAMs. He is currently a senior engineer at IBM, responsible for future memory product development. He has published papers in several technical areas, including short-channel threshold voltage, CCD charge control model, merged charge memory, 1Mb SRAMs, and fault-tolerant memory chips. Dr. Lee holds five levels of IBM Invention Achievement Awards and is a senior member of Sigma Xi.

Thomas M. Maffitt *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (B078478 at BTVLABVM).* Mr. Maffitt received the B.S.E.E. degree from the University of Notre Dame in South Bend, Indiana, in 1979. After graduation, he joined IBM in Essex Junction, Vermont. He has held various assignments in memory chip design, including a 25-ns 256Kb SRAM product which featured 3V/5V operation and $\times 1$, $\times 4$, and $\times 8$ configurations selectable via laser fuses. He has also worked on various embedded SRAM macros for a 0.5- μ m logic design system application. His most recent assignments have been data path architecture and design for a 16/18Mb DRAM product. He is currently working on productivity enhancements for the 16/18Mb design.

Christopher P. Miller *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (CPMILLER at BTVLABVM).* Mr. Miller received the B.S.E.E. degree from Rensselaer Polytechnic Institute, Troy, New York, in 1978 and the M.S.E.E. degree from the University of Vermont, Burlington, in 1993. In 1978 he joined the IBM Data Systems Division as a Product Assurance engineer in Poughkeepsie, New York, where he was involved with functional testing of memory chips. In 1979 he joined the DRAM development program in the IBM General Technology Division, developing 64Kb memories. Subsequent projects included an experimental 256Kb chip, an experimental 512Kb chip using a push-plate memory cell, and the IBM 1Mb silicon gate DRAM. Recently he has worked on the IBM 16Mb DRAM program and the 4.5Mb DRAM built in the same process, as well as methods to modify memory chips for three-dimensional packaging concepts. Mr. Miller is currently working on 16Mb derivative products including SDRAM and cached DRAM concepts at the IBM Microelectronics Division facility in Essex Junction, Vermont.

Charles H. Stapper *Retired, formerly IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (stapper@btrlabvm.vnet.ibm.com).* Dr. Stapper received the B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology in 1959 and 1960. After completion of these studies, he joined the IBM development laboratory at Poughkeepsie, New York. From 1965 to 1967, he studied at the University of Minnesota on an IBM resident study fellowship. Upon receiving the Ph.D. degree in 1967, he rejoined IBM at the development laboratory in Essex Junction, Vermont. His work at IBM dealt with magnetic recording, computer memories, and computer memory components. His major contributions are in the development of yield models for integrated circuit manufacturing. He has used these models for productivity optimization of SRAMs and DRAMs with redundancy and error-correcting codes, planning and controlling of memory chip fabrication, and planning the production of gate arrays, logic chips, and multiprocessor chips. Dr. Stapper has been co-guest editor of a special issue on High-Yield VLSI Systems of the *IEEE Transactions on Computers*, and co-chairman of the 1989 International Workshop on Defect and Fault Tolerance in VLSI Systems. He is an editor of the *Journal of Electronic Testing Theory and Application (JETTA)*. Dr. Stapper is the principal author of a paper that won the 1990 Best Paper Award in the *IEEE Journal of Solid-State Circuits*.

Howard L. Kalter *IBM Microelectronics Division, Burlington facility, Essex Junction, Vermont 05452 (HKALTER at BTVVMOFS).* Mr. Kalter received his B.S.E.E., M.S.E.E., and D.Eng. degrees from the University of Florida, Gainesville, in 1966, 1968, and 1970, respectively. Before graduation, he worked for Martin Marietta in Orlando on digital signal troposcatter communication systems. After

graduation, he joined the IBM Corporation in Essex Junction, Vermont. In his career he has had various assignments in memory chip design on both static and dynamic memory, logic, automated logic wiring, memory systems, software rule development for card design, and alterable logic and memory devices. Mr. Kalter has presented several papers on dynamic memory and has been a participant in several evening panel sessions at both the International Solid-State Circuits Conference and the Symposium of VLSI Circuits. Since 1992 he has been a member of the ISSCC Memory Subcommittee. He is a co-recipient of the IEEE Solid-State Circuits Council 1989-1990 Best Paper Award for the paper entitled "A 50ns 16Mb DRAM with a 10ns Data Rate and On-Chip ECC." He is also a co-recipient of the 1991 P. K. McElroy Award. He has 18 patents issued and eight pending in the U.S., and more than 65 publications. Mr. Kalter is an IBM Fellow and is manager of circuit design in Advanced Memory Development in the IBM Semiconductor Research and Development Center. He is a member of Eta Kappa Nu and Tau Beta Pi.