A. Bhattacharyya
D. P. Gaffney
R. A. Kenyon
P. B. Mollier
J. E. Selleck
F. W. Wiedman

# 1/*N* Circuit and Device Technology

*The 1/N memory cell is the bipolar analog of the FET one-device cell. A thin dielectric and doped polysilicon are combined with bipolar technology to achieve a vertically integrated, high-density, fast-performance memory chip. The circuit design, device structure, and processing implementation for a 64K-bit dynamic, 1/N fractional-device, experimental bipolar memory are presented. Test results for several geometrical and structural variations, including 16K-bit storage arrays, are given.*

## Introduction

The 1/*N* fractional device memory cell [1] reverses a trend in the evolution of semiconductor memories. FET main memory technologies achieve relatively higher density at lower cost but have lower performance compared with bipolar memory. Also, FET memory fabrication cost has increased with increased processing complexity. Bipolar main memory, on the other hand, achieves relatively higher performance but has poorer density and higher cost than FET memories. Integrated Injection Logic ($I^2L$) [2, 3] realizes improved bipolar density, but $I^2L$ memories exhibit FET-like performance.

By using normal bipolar devices for all support circuits with the $I^2L$ multiemitter inverted npn transistor for cell density, bipolar performance is maintained as FET-like density is achieved. The 1/*N* memory cell is the bipolar analog of the FET one-device memory cell [4] but with the transistor and the capacitor transposed. With the advent of the $n^+$-doped polysilicon as the extended storage plate (ESP) [5], the thin dielectric storage area dependency on emitter area was removed, thereby reducing operating voltage and improving performance while maintaining the high density of 315 $\mu m^2$/bit.

The major technical strong points for the 1/*N* cell concept emerge from 1) a vertically integrated memory cell which affords extreme cell density; 2) a bipolar switching device with associated high speed; and 3) a cell integration scheme employing metal bit and word lines. The extreme density of this cell design presents a number of technical challenges, some of which follow:

1. The sensing scheme has to fit a narrow bit pitch (12 $\mu$m).
2. The word driver and decoder have to fit on a narrow word pitch ($\approx$15 $\mu$m).
3. The physical layout of the cell constrains the area of the storage capacitor, which in turn limits the signal out of the cell.
4. The bipolar switch has to work in the inverse mode ($I^2L$), since this is how the high density is achieved [6].
5. The junction leakage from the emitter of a bipolar device has to be low enough ($\leq$5 pA) to support a dynamic cell.
6. A 20-nm insulator (nitride), deposited over a doped polysilicon surface, must maintain good insulating properties.

**Table 1** General process sequence.

1. p wafer.
2. $n^+$ subcollector *mask*, diffusion and reox.
3. $p^+$ buried isolation *mask* and diffusion.
4. n epi growth and reox.
5. $n^+$ reach through *mask*, diffusion and reox.
6. p base and top isolation *mask*, diffusion, reox and nitride deposition.
7. All contacts *mask* in nitride only.
8. $n^+$ emitter *mask* and diffusion.
9. $n^+$ poly deposition and *masked*.
10. Oxide dip etched.
11. 20 nm silicon nitride deposition and *masked*.
12. First metal evaporated and *masked*.
13. Nitride and poly plasma etched.
14. Low-temperature pyro deposition and *masked*.
15. Second metal evaporated and *masked*.

The characteristics of the $1/N$ cell are investigated in a 64K-bit experimental chip design, including decoders, drivers, high-resolution sense amplifiers, and input/output circuits, which is organized as 8K words by 8 bits. For ease of testing this chip, the address true/complement generators and timing circuits were not included, but all array bits are addressable. The goals for this chip design were an access time of 50 ns, a cycle time of 100 ns, and a power dissipation of approximately 500 mW, using a single 5-V power supply. Arrays of 16K bits of the storage structure have been fabricated and stress-tested for 1000 hours. The experimental 64K-bit test site is being fabricated. The first part of this paper describes the design of the memory including the array cell, dummy cell, word driver, sense latches, and bit decoder. This is followed by a presentation of the result of computer simulation, using a circuit analysis program, and hardware testing of the array cell, noise, word driver, latches, and leakage mechanisms. Extensions and applications of this technology are also discussed.

## Array design

- *Array cell structure*

The $1/N$ memory cell [1, 7] shown in Fig. 1 combines the inverted bipolar structure of $I^2L$ circuits [2, 3] (Table 1) with the thin dielectric technology of single-device FET memory [4]. The amortization of the contacts of the base and collector regions over $N$ emitters, Fig. 2, yields the high density from which the cell name $1/N$ is derived. The penalty for increased density is the increased base voltage drop due to lateral base resistance and base-collector parasitic current as more emitters are added per base contact [8, 9]. Second-level metal is used to stiffen the base stripes, which are orthogonal to the bit lines; therefore,

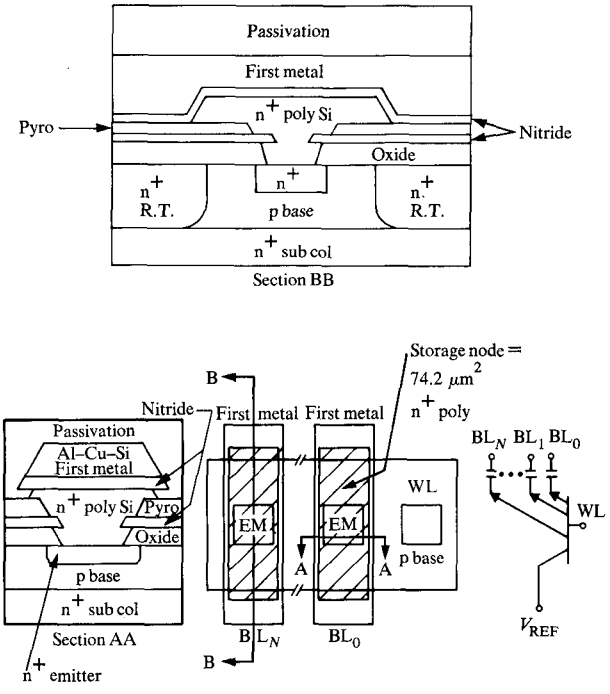


**Figure 1** Self-aligned extended storage plate $1/N$ memory cell.

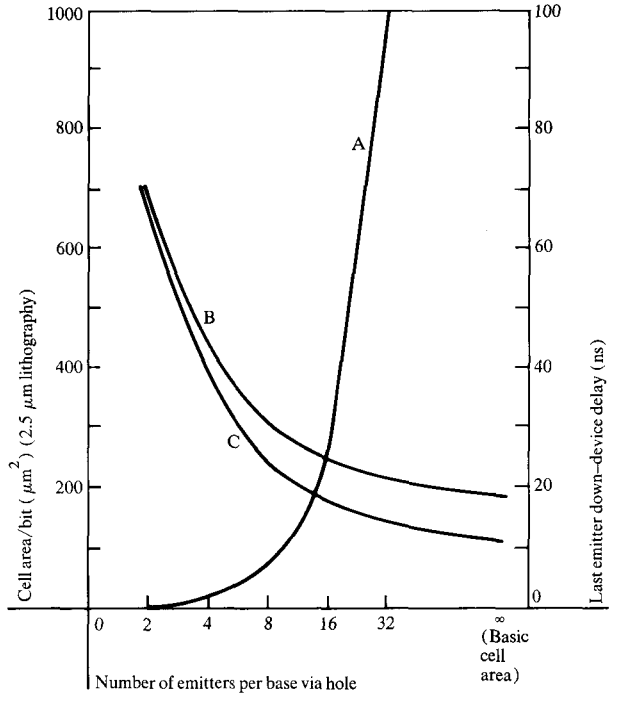


**Figure 2** Memory cell density and performance for various types of cells in 2.5-$\mu$m ground rules. Curve A: device delay; curve B: cell density for a 3.5-$\mu$m × 3.5-$\mu$m emitter; curve C: cell density for a 1-$\mu$m × 1-$\mu$m emitter.

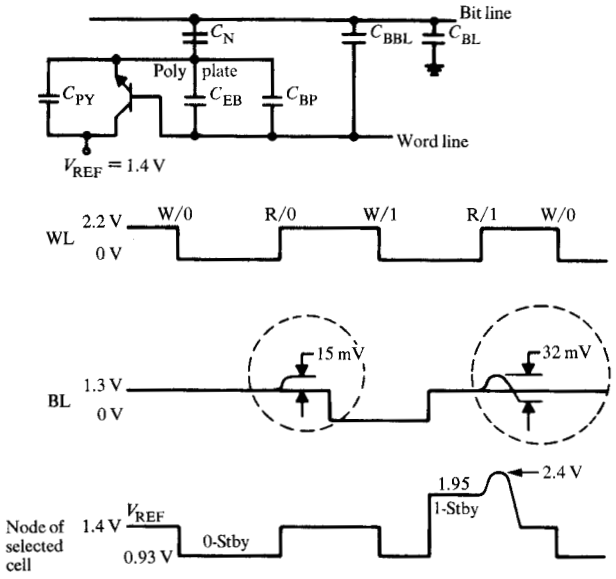


**Figure 3** $1/N$ cell operation. Design has 129 cells per bit line. Emitter dimensions are 3.5 $\mu$m × 3.5 $\mu$m. The silicon nitride film thickness is 20 nm.

the density-performance tradeoff (Fig. 2) can be employed. The common collector (Y-bus or $V_{REF}$) is also stiffened with first-level metal parallel to the bit lines to minimize voltage drops and improve performance with less than 10% impact on density. A deep $n^+$ reach-through diffusion [6] surrounds each base stripe to increase the inverse current gain of the npn transistor to greater than 10. With these performance improvements, $N = 8$ is the density-performance design point (see Fig. 2).

The original cell design had the thin nitride film in direct contact with the silicon emitter; therefore, the storage area was strongly dependent upon the emitter junction area. With this type of structure, the parasitic junction capacitance dominated the capacitor ratio ($C_N/C_{EB}$) < 1. By electrically extending the emitter with $n^+$-doped polysilicon (ESP) [5], the storage node capacitance has been uncoupled from the emitter junction capacitance (Fig. 1); hence the capacitor ratio ($C_N/C_{EB}$) is greater than 1 even for a regular bipolar process with a 20-nm nitride film thickness. To reduce the storage capacitance tolerance, a plasma etch step was added to produce self-alignment of the two capacitor plates [5]. Initially, the $n^+$ polysilicon is defined in stripes covering all emitters in each base pocket (Fig. 1). After the nitride film and metal are deposited and defined, the nitride and polysilicon are plasma-etched out between the metal bit lines. Because this polysilicon etchant will attack the silicon substrate, 100% metal and polysilicon coverage of all emitter and collector contracts is required in the support as well as in the array. Since all $n^+$ diffusions are 100% covered, the base isolation and Schottky barrier diodes can be dip-etched without a mask step before deposition of the thin nitride film.

To protect the exposed edge of the thin nitride under the first-level metal, low-temperature plasma pyrolytic oxide was used as the passivation between the two levels of metal. The storage structure reliability data will be covered in a later section.

### *Array cell operation*

A typical set of 64K array operating voltages is shown in Fig. 3. In general, the array cell transistor operates in the inverse mode only during one transition and the emitter-base thin dielectric capacitor-divider controls all other transitions.

*Write a zero* To write a *zero* or low charge state, the selected npn transistor is initially ON with its base at 2.2 V, collector and emitter at 1.4 V. Since the thin nitride capacitor is in series with the emitter, in the steady state there is no emitter current. As a result the emitter voltage is typically within 5 mV of the collector voltage by the following relationships, which are process- and geometry-dependent:

$$I_e = 0; \text{ then } I_u = \alpha_I I_d, \text{ where}$$

$$I_u = I_{c0}\,[e^{-qV_{cb}/kT} - 1] \text{ and } I_d = I_{e0}\,[e^{-qV_{eb}/kT} - 1].$$

Also under this condition, the base current has the dependency $I_b = I_d(1 - \alpha\alpha_I) = I_c$, which yields the maximum lateral base voltage drop. As the word line starts to drop to zero volts, the transistor turns off, but the word line voltage swing is capacitively divided by the emitter-base junction and the storage nitride capacitors onto the floating emitter. This coupling ratio is called the word line coupling factor,

$$\Theta_{WL} = \frac{C_{EB} + C_{BP}}{(C_{EB} + C_{BP}) + (C_N + C_{PY})} = 0.212,$$

evaluated for typical values shown in parentheses, where $C_{EB}$ is the emitter-base junction capacitance (45 fF); $C_{BP}$ is the base to polysilicon dielectric capacitance (6.3 fF); $C_N$ is the polysilicon to metal thin nitride capacitance (190 fF); and $C_{PY}$ is the polysilicon to collector dielectric capacitance (0.6 fF). This storage node (floating emitter) down-level is the *zero* standby voltage. Its value is typically 0.93 V, as determined by the following relationship:

$$V_N(\text{0-Stby}) = V_{REF} - \Theta_{WL}\Delta V_{WL}.$$

The thin nitride is at its most positive stress, which is

$$\frac{V_{BL} - V(\text{0-Stby})}{20 \text{ nm}} = 0.2 \text{ MV/cm}.$$

*Read a zero* To read a *zero*, the word line is again raised to its up-level, which couples back into the floating emitter the same voltage swing that was coupled out during the write operation. Except for the potential drop due to leakage during standby, the selected emitter will return to the reference voltage on the collector, 1.4 V. If the emitter level is low due to leakage, it will rise to the reference voltage by normal emitter follower transistor action. Since the word line swing is also coupled into the bit line and the bit line is much larger than the storage capacitor, the read *zero* signal on the bit line $\Delta V_{BL}$ is typically 15 mV above the steady state level. This read *zero* coupling factor is given as

$$\Theta_{R0} = \frac{C_{BBL} + \dfrac{C_N(C_{EB} + C_{BP})}{C_N + C_{EB} + C_{BP}}}{C_{BL} + C_{BBL} + \dfrac{C_N(C_{EB} + C_{BP})}{C_N + C_{EB} + C_{BP}}} = 0.068,$$

where $C_{BBL}$ is the base to metal dielectric capacitance (0.3 fF); $C_{BL}$ is the total bit line capacitance including the sense amplifier input capacitance (5.8 pF); and $\Delta V_{BL}$ (Read-0) = $\Theta_{R0}\Delta V_{WL}$.

*Write a one* To write a *one* or a high charge state, after the emitter has been clamped to the reference voltage the selected bit line is dropped to zero. This action charges the storage capacitor to its high-charge state. Since the emitter is clamped by the ON transistor, the emitter does not move during this bit line transition. The thin nitride is now at its most negative stress condition, 1.4 V/20 nm = 0.7 MV/cm. The selected word line returns to its down-level, which turns off the selected transistor and couples down the floating emitter to 0.93 V, the same as in the write *zero* operation. Then the bit line returns to its high level, typically 1.3 V. This voltage swing is coupled into the floating emitter by the following capacitor divider ratio, called the write factor:

$$\Theta_W = \frac{C_N}{C_N + C_{EB} + C_{BP} + C_{PY}} = 0.785.$$

The resulting emitter potential is called the *one* standby voltage, 1.95 V, which is the maximum reverse bias emitter base junction leakage condition:

$$V_N(\text{1-Stby}) = V_N(\text{0-Stby}) + \Theta_W\Delta V_{BL}.$$

*Read a one* To read a *one*, the selected word line is raised to its up-level, which couples the same voltage swings onto the floating emitter and bit line as in the read *zero* case via $\Theta_{WL}$ and $\Theta_{R0}$, respectively. Since the *one* standby voltage on the emitter is higher than the *zero* standby voltage, the floating emitter when selected will couple up to 2.42 V via $\Theta_{WL}$. With the base-collector junction already forward-biased and the emitter above the 1.4-V level of the collector, the inverted transistor will discharge the emitter down to the collector voltage. This transistor action will draw current from the bit line, which results in the *one-zero* window of 32 mV. This is the only place in the total cell operation where inverted transistor action takes place. Since the storage capacitor is in series with the bit line capacitor, the fraction of the emitter voltage drop that results in a bit line drop is given by the read *one* factor,

$$\Theta_{R1} = \frac{C_N}{C_N + C_{BL}} = 0.032.$$

The resulting read *one* level on the bit line is 17 mV below the standby level of 1.3 V:

$$\Delta V_{BL}(\text{Read-1}) = \Delta V_{BL}(\text{Read-0}) - \Theta_{R1}[\Theta_W\Delta V_{BL}].$$

With this nominal set of parameters, a zero-level-detecting sense amplifier (latch) would be sufficient; but with process variations the *zero-one* window does not stay symmetrical about the 1.3-V level. Therefore, there is a need for the dummy cell and differential sensing, which will be covered in a subsequent section.

Finally, two leakage constraints that limit the operating voltages are reverse leakage of the emitter-base junction (1-Stby) and forward emitter-base junction leakage (lowest emitter potential of a half-selected cell). Results of hardware measurements are covered in the section on cell leakage. The half-select condition occurs when a *zero* is stored in a cell whose base is unselected but whose bit line drops to zero level when a *one* is being written into another cell on the same bit line. The emitter of the cell receives a negative write factor $\Theta_W$. The following constraints apply:

$$V_{eb_{max}}(\text{reverse}) = V_N(\text{1-Stby}) = V_{REF} - \Theta_{WL}\Delta V_{WL} + \Theta_W\Delta V_{BL};$$

$$V_{eb_{max}}(\text{forward}) = V_{REF} - \Theta_{WL}\Delta V_{WL} - \Theta_W\Delta V_{BL},$$

where $\Delta V_{WL} = V_{REF} + V_{BC}(\text{ON})$.

The minimum *zero-one* window signal on the bit line as required by the sense amplifier = $[\Theta_{R1}\Theta_W\Delta V_{BL}]_{min}$. With these design constraints the operating voltages have the following constraints:

$$V_{WL} \leq \frac{V_{bc}(\text{selected}) + V_{eb}(\text{forward half-selected}) + \Theta_W V_{BL}}{1 - \Theta_{WL}};$$

$$V_{BL} \geq \frac{\min \textit{ zero-one } \text{window signal}}{\Theta_{R1}\Theta_W}.$$

The design given above, $V_{WL}$ = 2.2 V and $V_{BL}$ = 1.3 V, is based on $V_{bc}$ = 0.7 V, $V_{eb}$ = 0.1 V, and *zero-one* = 20 mV minimum.

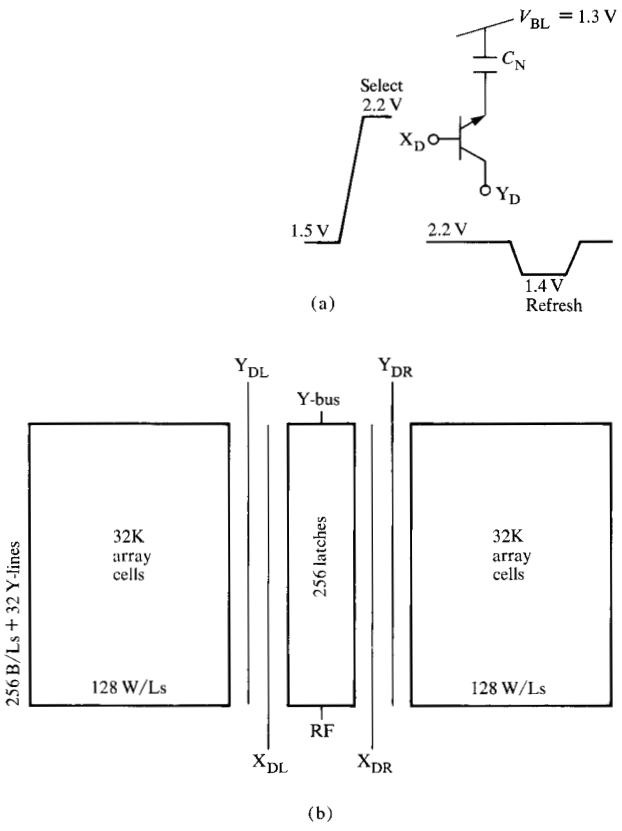


**Figure 4** Dummy cell organization: (a) partial geometry; (b) array organization.

### Dummy cell structure and operation

Since the storage capacitor in the $1/N$ memory is between the bit line and the switching device, the same cell device can be used for the dummy cell without the addition of a clamping device required by one-device FET memories. The only structural difference is the dummy common collector pocket, which is isolated from the array pocket, and is stiffened by a second-level-metal line adjacent to the second-level-metal dummy word line (Fig. 4). For the purpose of simpler support circuits, the dummy word line up-level (select) and collector up-level are both at the regular word line up-level (2.2 V). Since neither npn junction is ever forward-biased, no transistor action will take place during memory read-write-standby operations. The dummy word line down-level is chosen such that its word line swing, with the same timing as the regular word line, couples a voltage midway between a read *zero* and a read *one* of a regular memory cell into the opposite side of the latch, which yields the differential signal across the sense latch. For the midway condition

$$\Delta V_{BL}(\text{dummy-0}) = \frac{\Delta V_{BL}(\text{Read-0}) + \Delta V_{BL}(\text{Read-1})}{2},$$

where

*zero-one* window $= \Theta_{R1}\Theta_W\Delta V_{BL}$, and

$\Delta V_{BL}$ (dummy-0) $= \Theta_{R0}\Delta V_{XD}$.

Then

$$\Delta V_{XD} = \frac{\textit{zero-one}\ \text{window}}{2\Theta_{R0}} = \frac{\Theta_{R1}\Theta_W\Delta V_{BL}}{2\Theta_{R0}}.$$

Note that if the *zero-one* window is greater than $2 \times \Delta V_{BL}$ (Read-0), $\Delta V_{XD}$ is negative (typical nominal $\Delta V_{XD} = -0.24$ V).

Since there is no transistor action and the dummy emitter is floating, the dummy cells will have to be refreshed at the same frequency as that of the regular memory cells but not restored after every read/write cycle. All dummy cells on both sides can be refreshed by pulling $Y_{DL}$ and $Y_{DR}$ down to the regular Y-line voltage ($V_{REF} = 1.4$ V) and pulling $X_{DL}$ and $X_{DR}$ up to their regular up-level. This turns on all dummy cells in the inverse mode. Turning on the load devices of all latches, which will be covered later, writes a *zero* state into all dummy cells simultaneously.

If the regular polysilicon storage plate is large enough, a geometry half-capacitor dummy cell can be employed that uses the regular word line voltage swing. There is a design tradeoff between dummy and word line voltage tracking or parasitic capacitance and geometry tracking. Both types of dummy cells have been included in the design.

### Word design

- *Word drive circuits*

Considerations that influenced the driver design were the tight array word line pitch, Table 2, the need to provide a fast driver with flat current drive capability out to 14 mA, no driver circuit standby power dissipation, minimum selected driver power dissipation, rapid word line recovery, and array noise minimization.

Figure 5 shows the driver, matrix decode, and recovery network schematic. The word drive circuit uses a Schottky clamped pnp selection transistor T3 with its collector supplying the base current of a Schottky clamped npn emitter follower T4. The pnp clamp Schottky diodes (D4s) have their cathodes brought to an outside pad on the test cross section so that the clamp level can be adjusted easily during characterization. The pnp selection device was chosen so that the driver circuit standby

power dissipation is zero since all unselected devices are OFF. The pnp device had a nominal beta ($\beta$) of eight at 100 $\mu$A of collector current. The Schottky clamped npn emitter follower had a nominal $\beta$ of 100 at 14 mA. The emitter stripe length was 40 $\mu$m with first-level-metal base contact stiffening along the 40-$\mu$m dimension to minimize base resistance. To provide this feature while at the same time meeting the 15.5-$\mu$m-array word line pitch, the emitter follower and pnp pre-driver and Schottky diodes D2, D3, and D4 were placed in a tandem arrangement. Two driver networks were placed behind the two drivers closest to the array, and their outputs were routed through the devices closest to the array. The additional capacitance encountered with the longer output driver wiring was not significant when compared with the array load. With this arrangement, the four-driver layout is compatible with the array word lines (4 × 15.5 $\mu$m).

- *Recovery*

A switched recovery network was chosen to minimize driver power dissipation. Two separate networks are used. The first controls the odd-numbered lines, while the second controls the even-numbered group. Upon driver selection, the appropriate recovery network is gated off, enabling the driver to supply full current to the array base pockets. Emitter-follower charge-up device T7 supplies the necessary drive to reverse-bias the Schottky devices D2 and D3 for each of the 64 lines during this interval. During the word line recovery portion of the cycle, the recovery network is switched on, providing good current sinking capability through paralleled npn devices T5 and T6 while forward-biased Schottky diodes D2 and D3 remove the stored charges in the array and in the npn word driver, respectively.

- *Noise minimization*

The tandem driver layout and odd-even recovery network arrangement provide a convenient method for ensuring that the unselected word lines on either side of the selected line (which are the most noise-disturb sensitive) are held at ground. For the test chip, 64 word lines were allowed to float. To provide a still better noise-disturb margin, the recovery networks could have additional levels of decoding which would limit the number of unselected word lines allowed to float.

- *Decoders*

Decoding is accomplished with a matrix transistor approach (see Fig. 5). One matrix device is used for every word line. For a 64K-bit chip with 128 word lines per array side, a 1 of 16 and 1 of 8 decoding arrangement would be used. For the test cross section, to provide ease of testing at the expense of eight additional input pads, only the 1 of 8 decoding is used. This selects 1 of 8 possible

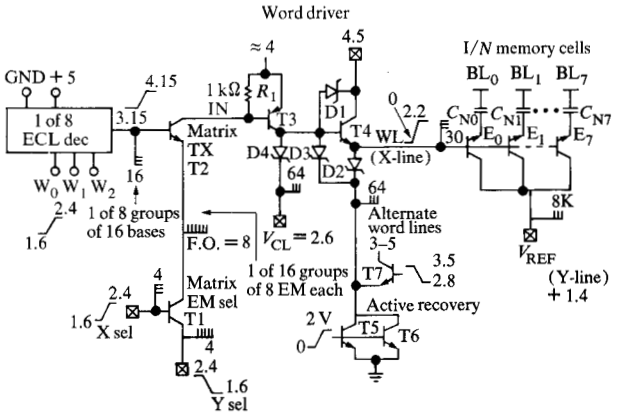


**Figure 5** Word driver and recovery network.

**Table 2** Summary of the 64K RAM's characteristics.

| | |
|---|---|
| Organization | 8K words × 8 bits |
| Cell size ($N$ = 8) | 315 $\mu m^2$ |
| Chip size | 30 $mm^2$ |
| Access time | 50 ns |
| Cycle time | 100 ns |
| Supply voltage | +5.0 V |
| Operating power | 500 mW |
| Refresh | 257 cycles/3 ms |
| Minimum sense signal | ±15 mV |
| Word pitch | 15.5 $\mu$m |
| Bit pitch | 11.0 $\mu$m |
| Lithography | 2.5 $\mu$m |

groups of 16 matrix transistor bases (high level). Selection of 1 of 16 possible groups of 8 matrix transistor emitters is done with an additional matrix of 16 devices shown representatively as T1 and arranged in a 4 × 4 layout. Eight pads are used to provide direct base and emitter inputs.

## Sense and I/O

- *Sense latches*

The circuit design of the sense amplifier (Fig. 6) is highly constrained by the bit pitch given in Table 2. The sense latches are arranged four wide in a staircase pattern so that each sense amplifier is four bit pitches high. The layout allows the bit lines to be on first-level metal and permits the use of the low-impedance, low-capacitance, second-level-metal phase lines that are orthogonal to the bit line axis. The use of metal busses for the emitter lines allows much higher initial turn-on current for the latch.

Latch operation may be understood by referring to Fig. 6 and to the operating waveforms shown in Fig. 7. The latch is operated dynamically in order to maintain low

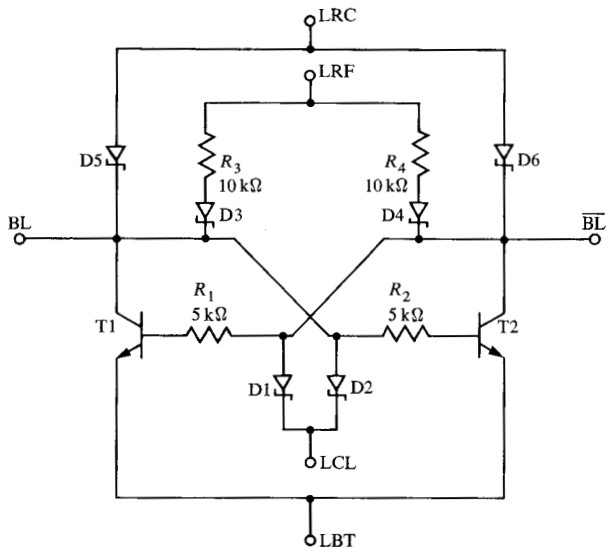


**Figure 6** Latch circuit.

quiescent power dissipation. The following abbreviations are used: LBT (latch bottom), LCL (latch clamp), LRF (latch refresh), and LRC (latch recover).

In the quiescent state, LBT and LRC are high and LRF is low. LRC is set to 3.6 V and the bit lines are clamped at 3.0 V by the bleed current through the Schottky barrier diode pairs, D5-D1 and D6-D2. This bleed current, on the order of 10 $\mu$A, decreases the bit line sensitivity to capacitively coupled noise.

In a read cycle, the LRC line is dropped to 0 V before the cell is selected so that the bit lines are no longer clamped. The cell and dummy cell word lines are selected and LBT is ramped down, turning on both T1 and T2 with initial emitter currents determined approximately by the diode junction equations. A 10-mV differential in bit line potentials is sufficient to give a 1.47:1 ratio of collector currents. This current difference is amplified by the latch feedback and the initially low bit line discharges much faster than the higher bit line. After approximately 12 ns, the base-emitter junction of one npn inverter is back-biased, which completes the dynamic phase of the latch operation. The LRF line is then raised, putting a 10-kΩ load on the bit lines and restoring the high bit line to the desired *zero* refresh level for the $1/N$ cell. After the cell is refreshed, the word line is returned to its low level, isolating the refreshed information on the cell storage node. The LBT line is subsequently raised, shutting off T1 and T2. The LRC line then gives a fast recovery of the bit lines. The diodes D3 and D4 block the shunting path during the period the cell is read, eliminating signal loss.

A write operation is described in the section on the bit decode circuitry which follows. It essentially consists of setting the latch with a large signal and using the latch output to write the cells.

The Schottky-diode clamp-recovery system is used in a unique way. Diodes D1, D2, D3, and D4 contribute no additional cost in area in the layout. The recovery system provides substantial power reduction: from 3.31 mW (worst case) without the diodes to 1.17 mW (worst case) with them, as shown for a 100-ns read-refresh cycle (see Fig. 7).

The overall design dictated that the latch must be able to detect differential signals of $\geq$15 mV. With this signal level, the latch switches to its full output signal level in $\leq$15 ns. The latch furnishes the proper refresh voltage levels (*zero* = 2.75 V and *one* = 1.25 V) to refresh the cell.

- *Bit decode*

The purpose of the bit decode circuitry is to establish a bidirectional connection between a small number of chip pads (*i.e.*, eight data I/Os) and a portion of the total number of sense latches (*e.g.*, eight out of 256) defined by the corresponding address bits. Its function is an 8-bit × 32-way bidirectional multiplexor. The general circuit configuration is shown on the left in Fig. 8. The combination of Schottky DTL with a dual-function npn transistor gives a compact and efficient read/write access method.

In read operation, the transistor is a grounded base amplifier, the emitter of which is driven by a sense latch. This amplifier is on only if its base is selected at up-level by the address bits. The data are then available at the common subcollector, ready for transmission to the off-chip driver.

In write operation, the transistor is also a grounded base amplifier, the common subcollector pocket of which is driven by the write data so that the transistor is operated in the inverse mode. The data are, therefore, available on the emitter diffusion of the transistor whose base is selected at up-level by the address bits; a low level is forced on the left side of the selected latch if the subcollector diffusion is pulled down by the write data input.

Notice that only one side of each of the latches is driven by the multiplexors. Hence, it is necessary to preset the latches before writing by pulling bases and subcollectors of the multiplexors to the up-level, causing 256 emitter followers to force up the left side of the sense latches.

## Simulation and hardware results

• *Array cell*

The initial multiemitter npn hardware used a conventional bipolar process with and without a shallow $n^+$ reach-through diffusion. The parasitic lateral pnp ($n^+$ absent) current gain $\beta$ was 0.8, which was not acceptable; therefore, there was a need for the $n^+$ diffusion. The total npn-device inverse $\beta$ was improved from 0.5 to 2 when the shallow diffusion was added. This confirmed the need for the deep reach-through diffusion [6]. The thin (20-nm) silicon nitride film (Fig. 9) directly deposited on the $n^+$ emitters exhibits the following characteristics:

Capacitance >2.3 fF/$\mu m^2$,
Leakage <10 fA/$\mu m^2$,
Breakdown >10 V.

This hardware demonstrated the lateral base voltage drop due to extrinsic base resistance and parasitic diode current [8, 9].

The next test site hardware concentrated on geometry and process variations for the array cell $I^2L$ [8] circuits. The $\beta_I$ for the 2.5- × 4-$\mu$m (mask) emitter transistor on a 2.7-$\mu$m epitaxial layer increases from 0.3 to 1.5 when a deep $n^+$ reach-through diffusion is added. The $\beta_I$ for the same device on 2.2-$\mu$m epi increases from 1.0 to 8.6. The emitter-base leakage was shown to be less than five pA per storage node, which is acceptable for the memory design. Its geometry and structural dependencies were also studied (Fig. 10). The emitter current fall-off ($I_{E1}/I_{E8}$) due to lateral base voltage drop for $N = 16$ increases from 1.6 to 5.0 as the base pocket current is increased from 0.1 to 1.0 mA.

To measure cell dynamic performance, a cell npn transistor without the storage capacitor was used with a measured probe capacitor, since any probe capacitance had a major effect upon the array cell. Since the array cell is nonsaturated during most of its read *one* transition, the transistor acts as a constant current source discharging a fixed capacitor. After the large probe capacitor is precharged high, the cell transistor discharges it at a constant rate $I = C_p(\Delta V_1/\Delta t_1)$, where $C_p$ = probe capacitor, $\Delta V_1$ = change in capacitor voltage, and $\Delta t_1$ = the time for this voltage change. Since the same transistor emitter current discharges the storage node (thin dielectric and emitter-base junction capacitors), the following array cell read transition delay can be predicted from the hardware measurements:

$$\Delta t_a = \left(\frac{C_N + C_{EB}}{C_p}\right)\left(\frac{\Delta V_a}{\Delta V_1}\right)\Delta t_1,$$

where $C_N$ = thin dielectric capacitor, $C_{EB}$ = emitter-base

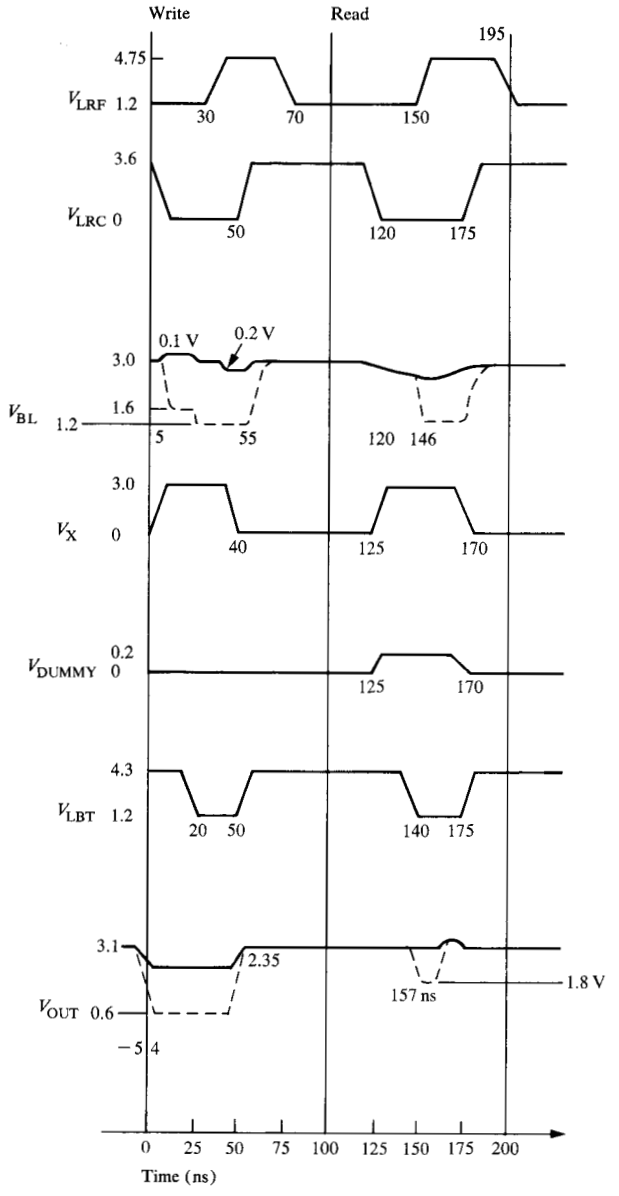


**Figure 7** Latch operating waveforms.

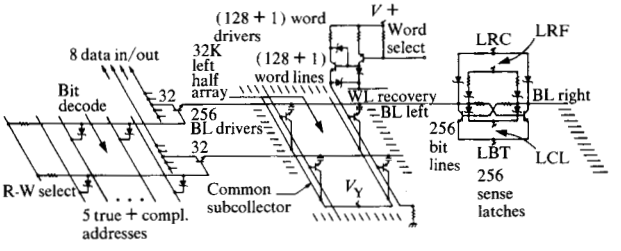


**Figure 8** $1/N$ memory circuit cross section for the 64K design.

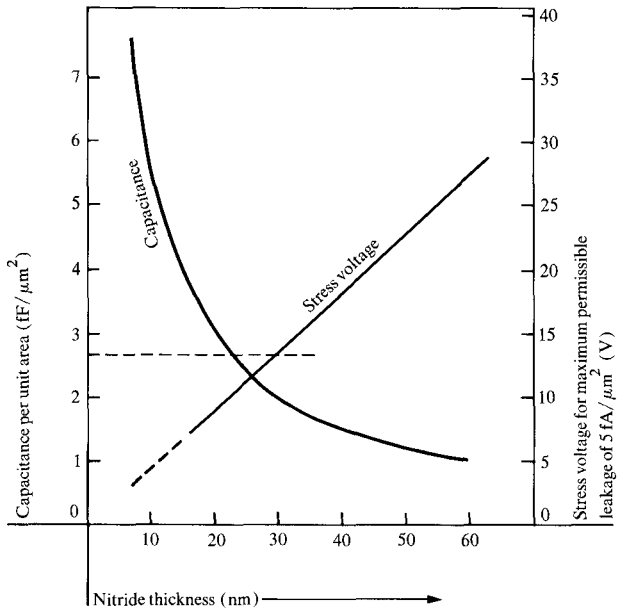


**Figure 9** 64K $1/N$ memory design (ESP).

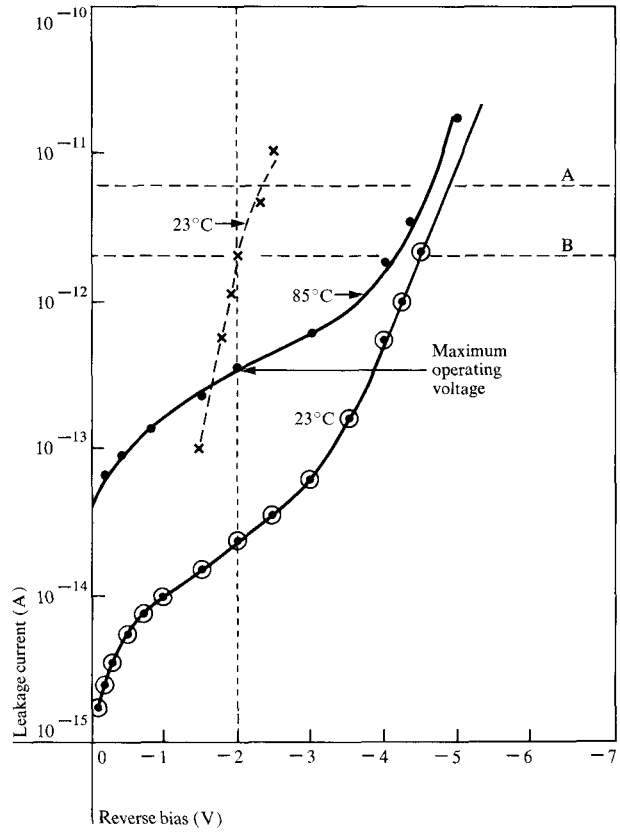


**Figure 10** Emitter-base junction leakage measurement for an emitter size of 3.0 $\mu$m × 4.0 $\mu$m. A—Maximum allowable leakage current for retention time of 3 ms. B—Maximum allowable leakage current for retention time of 10 ms. Curve (×): Conventionally diffused base and emitter at 23° and 85°C. For doubly implanted base and shallow emitter, curve (•) shows results at 85°C and curve (○) shows results at 23°C.

junction capacitor, $\Delta V_a$ = actual array cell storage node voltage swing, and $\Delta t_a$ = the actual cell delay due to transistor action. The other factor in array cell access time is the base pocket propagation delay. These two factors are combined in the results of simulation shown in Fig. 2.

In order to reduce the emitter-base capacitance, a doubly implanted base npn [10] was fabricated with capacitance reduced by a factor of 2.5 over the standard process. The range of emitter current fall-off ($I_{E1}/I_{E8}$) increased from 4.0 to 8.0 as the base pocket current was increased from 0.1 to 10 mA with extrinsic base resistance increased by a factor of 5. This range was not as great as was predicted by Gaffney and Bhattacharyya's model [8] because they assumed a constant $\beta_I$ while the actual $\beta_I$ of the hardware tripled. This structure had much lower junction leakage (Fig. 10).

With the advent of the polysilicon-extended storage plate (ESP) [5], several array storage structures and area capacitors were fabricated. Arrays consisting of 1K, 2K, 4K, 8K, and 16K storage capacitors and area capacitors equal in storage area to the 16K arrays were fabricated on $n^+$ polysilicon storage plates and, for comparison, on $n^+$ emitter diffusions as well. After the initial infant mortality failures were eliminated, approximately 5% additional failures occurred for the 5-volt 150°C 1000-hour stress test. Note from the array operation section of this paper that the maximum dielectric stress voltage is 1.4 V, which is much lower than 5.0 V. With the results from a voltage and temperature stress matrix, the corresponding acceleration factors could be calculated to relate the mean time of failure of the accelerated test to the real machine environment. The 16K-equivalent-area capacitor infant failures were similar to the 1K arrays rather than to the 16K arrays, but all arrays and area capacitors over polysilicon islands or emitter diffusions had the 5% long-term cumulative failures. From the results, the infant failures appear to be geometry-dependent and not bulk- or area-dependent.

Since the geometry of the dummy cell is the same as that of the array cell, the array cell model was used for the latch or array simulations which were performed.

- *Noise compensation*

Since any variations in the base-collector junction voltage will cause variations in array cell performance, the word line (WL) and the common collector (Y-bus or $V_{REF}$) distribution are very sensitive. In each base pocket, the base and collector contacts are placed on opposite ends so that the voltage drops subtract, to reduce the skew. A typical 4-mm × 8-$\mu$m WL driving 0.4 mA into each of 32 base pockets has a 100-mV drop due to second-level metal re-

sistance. The effect of this voltage drop on the array performance was minimized by placing an equivalent drop in the Y-bus. First-level-metal stiffening of the Y-bus across the array was essential to minimize collector drop, which is perpendicular to the WL drop. Then the width of the second-level-metal Y-bus which passes over the latches (Fig. 11) was varied. The Y-bus was driven from the top while the word line was driven from the bottom. With these compensations and the method of word driver selection (to be covered in a subsequent section), the worse-case bit line common mode noise calculated was 40 mV, while the differential noise was 1.5 mV. With a differential sense amplifier (latch), these noise levels are acceptable.

- *Word drive circuit simulation*

The goal for chip access time was 50 ns. To meet it required supplying the turn-on base currents for 256 cell transistors and charging the associated word line parasitic capacitances (typically 95 pF) through a 2.2-V excursion in less than 15 ns. Array simulations showed that the 14-mA emitter-follower driver was needed to meet this objective.

With a goal for chip cycle time of 100 ns, approximately 30 ns are required to remove the array-cell stored charge and turn off all previously selected cells. Array simulations showed that the discharge transistor had to sink ≈26-mA peak currents to meet the recovery goal.

- *Latches*

Latch design evolved in response to process variations which were driven by cell and array requirements. The design feasibility was checked by a combination of simulation and hardware measurements.

Initial designs used lateral pnp transistors as cross-coupling elements and inverted npns. The latch was essentially a cross-coupled pair of $I^2L$ inverters. This circuit could be strobed at the pnp bases to allow a full voltage swing at the collectors. Our initial hardware verification of this design showed that high-performance $I^2L$ inverters were not compatible with satisfactory $1/N$ cell performance (most notably with respect to collector leakage). With a process optimized for the array, switching time of the latch was almost an order of magnitude longer (150–200 ns) than design simulations based on original npn inverter performance assumptions. A second hardware evaluation was made by reconnecting some of the original test wafer to contain latches with noninverted npn devices. Evaluation of this hardware showed improved sensitivity and performance but showed that the pnp parameters were very poorly controlled. Furthermore, a combination hardware-simulation approach demonstrated a

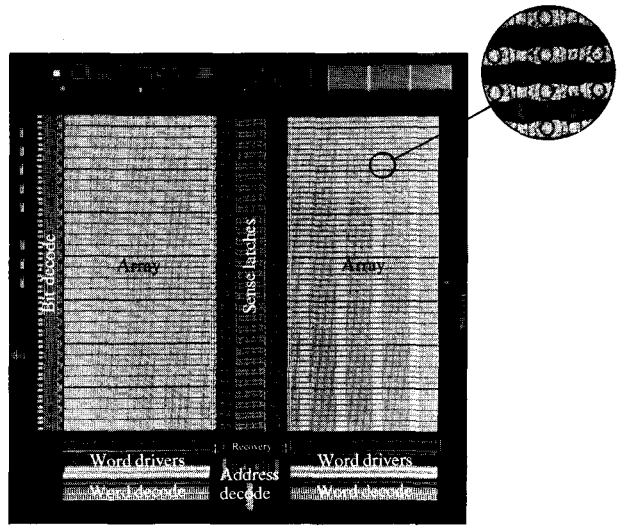


**Figure 11** $1/N$ 64K memory chip with the detail of cell area shown in the enlargment.

"memory" phenomenon in the sense amplifier due to charge storage from the previous cycle. (Since the $1/N$ signal is at such low charge levels, it is easily disturbed.)

The pnp approach was abandoned in favor of a resistor-coupled approach using ion-implanted resistors. Simulation showed that a satisfactory sensitivity-power tradeoff could be achieved during sensing, but that the bit lines did not recover to a balanced value quickly enough with this approach. (The small signal amplitude demands a close recovery, <1 mV.) Various forms of "shorting bars" between bit lines were proposed but the desired precision was lacking. A favorable process change provided the solution: use of the Schottky diodes D5 and D6 (Fig. 6) to recover the bit lines. Thus, fast recovery can be combined with the low power of resistor-coupled sensing. Since the refresh levels do not have to be as precisely set as the bit line recovery levels, diffused resistors can be used. Furthermore, the power used in pulling up the bit line is 1/2 $CV^2$, the charging energy, so no additional power cost is incurred.

The major considerations of these final designs were as follows: In order to minimize access time, it was found that use of a single-slope ramp for LBT, turning both transistors ON initially, gave a faster response than the dual-slope ramp suggested by Lynch and Boll [11]. Since bipolar transistors are essentially charge-controlled, the need to establish working charge levels in the devices dominates the design factors for IGFET latches [11]. In addition, it was found that there is a complex relationship be-

tween the rate of fall of LBT and the values of $R_1$ and $R_2$. For $R_1$ and $R_2$ matched within 2%, there is a decreasingly narrow range of $(dV_{LBT}/dt)$ for optimum transient response as resistance increases. The lower bound is set by the poor transient response of the npns at low currents ($I_{LBT}$ is dominated by $CdV_{LBT}/dt$ factors), and the upper bound by the influence of imbalance in the $IR$ drops. A value of $R_1 = R_2 = 5$ kΩ was chosen as a level low enough to give a design which is not critically dependent on the $dV_{LBT}/dt$ value for good transient performance. The value of $R_3 = R_4$ then follows from a dc analysis to determine the required refresh level which sets $R_3 = 2R_2$ for this design.

• *Cell leakage mechanisms critical to 1/N retention time*
The factors that contribute to the gain or loss of charge on the cell transistor floating emitter node and that subsequently determine retention time and memory availability are 1) storage dielectric leakage, 2) emitter base junction leakage, 3) low current pipes, and 4) the $I_{CEX}$ bias condition. Conditions 1 and 2 produce a charge loss which reduces the stored *one* level while conditions 3 and 4 increase the *zero* level through a floating node charge gain. Before the measurement results are discussed, the $I_{CEX}$ bias condition will be described.

During the write mode, there is a potential half-select *zero* data disturb condition. The disturb condition is analogous to an $I_{CEX}$ bias condition on the unselected cells since the epitaxial layer is at +1.4 V, the cell base at ground, and the emitter node at −0.2 V. This concern can be eliminated by proper choice of bit line write voltage levels and a device requirement that cell transistor conduction at 0.2 V base-emitter forward bias and $V_{CE}$ = 1.6 V should be less than 15 pA. The 15-pA leakage criterion is calculated from the simple relation $CV = \int idt$ and design assumptions of a 235-fF total floating node capacitance, a pessimistic 50 ns or 50% duty cycle during which the bit line is held low, an allowable 10% loss of signal window, or 100 mV, and a 3-ms retention time. This is a pessimistic estimate since the reduced forward base-emitter junction bias tends to cut off the conducting device as the floating node tends to charge up.

*Storage dielectric leakage measurement* Insulator leakage measurements on 3.5-μm × 3.5-μm emitter structures at −2 V reverse bias and 85°C showed currents ≤1 fA/cell. The data confirmed the intuitive feeling that junction leakages are by far the biggest contributor.

*Emitter-base junction leakage measurements* Figure 10 shows leakage results at both room temperature and +85°C for a 3-μm × 4-μm device with a two-step implanted boron base and implanted arsenic emitter. Deep base implant energy was 300 keV with a dose of $1 \times 10^{13}/cm^2$, while the shallow base implant was 40 keV with dose of $1 \times 10^{14}/cm^2$. Extrinsic base resistance was 1 kΩ/□.

Measured leakage at 85°C and −2 V reverse bias is observed to be approximately 1/20th of the needed value of 6 pA, providing greater than an order of magnitude safety at worst-case operating temperatures.

The curve also shows that up to −3 V reverse bias, there is a 20 to 35× increase in current with temperature increasing from 23° to 85°C, indicating depletion-layer carrier generation as the major leakage contribution, while at higher reverse voltages tunneling is the major contributor.

Shown also for comparison are measurement data taken on a conventionally diffused 200 Ω/□ extrinsic base resistance structure with emitter $X_j$ = 1.5 μm. It can be observed that even at room temperature this process would only marginally meet retention time specifications.

*$I_{CEX}$ bias conduction* Measurements over the temperature range 20° to 85°C were made on a sample of 10 devices with 3-μm × 4-μm emitters and conventionally diffused processing (extrinsic base $R_S$ = 200 Ω/□, emitter $X_j$ = 1.5 μm). With the test conditions of 0.2 V base-emitter bias and $V_{CE}$ = 1.6 V, all the sample devices had leakages <10 pA at 85°C.

*Low-current pipes* Bipolar devices have a collector-emitter leakage mechanism known as pipes. For a dynamic memory like 1/N with stringent leakage requirements, the question is whether for a given process the density of pipes is tolerable. If pipe density were to vary inversely with lower-level leakage, severe yield penalties would result. Fortunately, no inverse relationship was found in low-current measurements, either on diffused or implanted structures. The same density of pipes was found at nA and mA current levels as for pA levels; therefore, a nA-level testing screen was found adequate to sort defective from good product.

*Process dependency* Careful base doping profile tailoring is required to balance the retention time requirements with chip performance goals. This can be done either through diffusion or implantation techniques. No fundamental limitations were observed in our measurements to indicate that good retention time and yield could not be achieved for the 1/N bipolar structure.

## Summary

This experimental 64K-bit 1/N memory chip with its circuit schematic (Fig. 8) and layout (Fig. 11) is currently being fabricated using the process as outlined in Table 1. For ease of testing this test chip, the address true/com-

plement generators and timing circuits were not included; if they were added, this design would be a full-function chip. Design objectives are summarized in Table 2. The array and dummy cell operating voltages have been shifted up to accommodate the support circuit design levels.

There are several possibilities for extensions of this design into future bipolar process technologies. As the emitter area decreases, the cell operating voltages and size decrease (Fig. 2). For example, with 2.0-$\mu$m lithography, except for 1.0-$\mu$m emitter openings, the cell area decreases to 140 $\mu m^2$ and chip area drops to 19 $mm^2$. The implanted structure (Fig. 10) reduces the emitter-base junction capacitance and lowers the leakage, which also improves the cell operating conditions. Another example for process extension is that the $n^+$-doped polysilicon can be used as the emitter diffusion source [12].

This memory technology would be most useful in small fast systems where cache memory is undesirable or when the cache becomes too large due to cache/main performance mismatch.


## Acknowledgments

The authors wish particularly to thank M. Sears, M. Desilets, and B. Deliduka for their testing and design support; B. Deliduka for his low-current leakage characterization; D. Chesebro, C. Wexelblatt, J. Wursthorn, C. Kroll, H. Geipel, S. Kingsland, J. Lajza, P. Bouffard, J. Dougherty, J. Underhill, F. Soychak, D. Purcell, C. Osborne, and M. Ravitz for their processing support; W. Cochran for his help in circuit modeling; W. Pricer for his circuit innovations; and H. Berger, R. Jaeger, H. Lee, and F. Buckley for their valuable suggestions. We wish also to thank B. Agusta and W. Pricer for their enthusiastic management support of this project, and the many others who helped make this project successful.


## References


1. W. D. Pricer and J. E. Selleck, "Multiple Element Charge Storage Memory Cell," U.S. Patent 3,979,734, 1976.
2. H. H. Berger and S. K. Wiedmann, "Merged Transistor Logic—A Low Cost," *1972 IEEE ISSCC Digest of Technical Papers* **15,** 90-91 (1972).
3. C. M. Hart and A. Slob, "Integrated Injection Logic—A New Approach to LSI," *1972 IEEE ISSCC Digest of Technical Papers* **15,** 92-93 (1972).
4. R. H. Dennard, "Field Effect Transistor Memory," U.S. Patent 3,387,286, 1968.
5. A. Bhattacharyya and P. Landler, "Enhanced Storage Medium for a Memory Cell," *IBM Tech. Disclosure Bull.* **20,** 5137 (1978).
6. A. Schmitz and A. Slob, "The Effect of Isolation Regions on the Current Gain of Inverse NPN-Transistors Used in $I^2L$," *IEDM Tech. Digest,* 508-510 (1974).
7. J. E. Selleck, R. A. Kenyon, D. P. Gaffney, F. W. Wiedman, A. Bhattacharyya, and P. Mollier, "64K Dynamic, $1/N$ Fractional Device, Bipolar Memory," *1980 IEEE ISSCC Digest of Technical Papers* **23,** 220-221 (1980).
8. D. P. Gaffney and A. Bhattacharyya, "Modeling Device and Layout Effects of Performance Driver $I^2L$, *IEEE J. Solid-State Circuits* **SC-12,** 155-162 (1977).
9. W. H. Matthews and R. P. Mertens, "Modeling the Dynamic Behavior of the $I^2L$ Inverter," *1977 IEEE ISSCC Digest of Technical Papers* **20,** 44-45 (1977).
10. Richard S. Payne, R. J. Scavuzzo, K. H. Olson, J. M. Nacci, and R. A. Moline, "Fully Ion-Implanted Bipolar Transistors," *IEEE Trans. Electron Devices* **ED-21,** 273-278 (1974).
11. W. T. Lynch and H. J. Boll, "Optimization of the Latching Pulse for Dynamic Flip-Flop Sensors," *IEEE J. Solid-State Circuits* **SC-9,** 49-55 (1974).
12. J. Graul, A. Glasl, and H. Murrmann, "Ion Implanted Bipolar High Performance Transistors with Polysil Emitter," *IEDM Tech. Digest,* 450-454 (1975).



*Received April 30, 1979; revised December 6, 1979*



*D. P. Gaffney, R. A. Kenyon, J. E. Selleck, and F. W. Wiedman are located at the IBM General Technology Division laboratory, Essex Junction, Vermont 05452. A. Bhattacharyya is located at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598. P. B. Mollier is with Compagnie IBM France, Corbeil-Essonnes, France.*