

Non-Volatile Memories : State-of-the-Art

Autor(en): **Gnädinger, F. P.**

Objektyp: **Article**

Zeitschrift: **Bulletin des Schweizerischen Elektrotechnischen Vereins, des Verbandes Schweizerischer Elektrizitätsunternehmen = Bulletin de l'Association Suisse des Electriciens, de l'Association des Entreprises électriques suisses**

Band (Jahr): **75 (1984)**

Heft 3

PDF erstellt am: **30.05.2024**

Persistenter Link: <https://doi.org/10.5169/seals-904347>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Non-Volatile Memories – State-of-the-Art

F. P. Gnädinger

This paper deals with high density EEPROMs built in MOS technology, a very small segment of all non-volatile memories. No leading part of the 64 K density has emerged yet. Even the best technology choice is still undecided. It is very important to realize that for large memory arrays the freak distribution and not the average behavior under worst case conditions is the limiting factor.

Der Aufsatz behandelt EEPROM hoher Dichte in MOS-Technologie, einen schmalen Ausschnitt aus allen nicht-flüchtigen Speichern. Bisher hat sich noch keiner der bekannten 64-K-Speicher durchgesetzt. Sogar die Wahl der besten Technologie ist noch nicht entschieden. Wesentlich ist, dass für grosse Speicher-Arrays die Verteilung der Schwachstellen und nicht das mittlere Verhalten unter ungünstigsten Bedingungen die massgebende Begrenzung bilden.

L'exposé traite des EEPROM en technologie MOS à haute densité qui ne représentent, en fait, qu'une petite fraction de toutes les mémoires non volatiles. Jusqu'à présent, aucun composant mémoire 64 K ne s'impose; même le choix de la meilleure technologie est ouvert. Ce qui importe, c'est que les grandes mémoires sont limitées par la distribution des défauts et non par leur comportement moyen dans les conditions les plus sévères.

This paper has been presented at the Fall 1983 Meeting of the IEEE Switzerland Chapter on Solid-State Devices and Circuits, October 1983, at Bern.

Author's address

Dr. Fred P. Gnädinger, Vice President of R&D, Immos Corp., Colorado Springs, P.O. Box, CO 80935-6000 (USA).

1. Introduction

Non-volatile memories have always represented a substantial portion of the total memory market. Various technologies have been employed over the years, including magnetic core technology, magnetic bubbles and, of course, magnetic tapes and discs for mass storage. Of the newer technologies, MOS semiconductor non-volatile memories dominate the market. Other alternatives such as optical discs and similar approaches are still in their infancy.

In this paper, only semiconductor memories fabricated using MOS technology will be covered. Figure 1 shows the actual and projected market potential for MOS type non-volatile memories. The total market is expected to grow from 2 billion dollars in 1981 to 7.9 billion dollars in 1985. Non-volatile memories provide the largest segment of this market. Their market share is expected to grow from 44% in 1981 to 63% in 1985. This segment, however, includes both electrically programmable and erasable non-volatile memories (EEPROMs), UV erasable programmable memories (EPROMs) and mask programmable or fusible link memories (ROMs and PROMs). The most advanced non-vol-

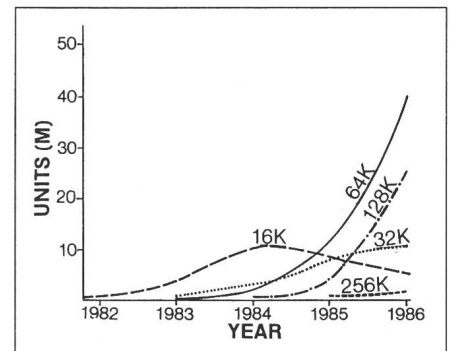


Fig. 2 Total market projection of MOS EEPROM

atile memories are undoubtedly the EEPROMs, and in the remainder of this paper the emphasis will be on state-of-the-art EEPROMs. Figure 2 shows a market projection of the various types (densities) of EEPROMs. The most advanced part in volume production today is the 16 K EEPROM with 16,384 memory storage locations. The next generation parts are projected to be available in 1984 and later. Several major semiconductor companies have 64 K and possibly 128 K, or even 256 K, EEPROMs under development. As is evident from Figure 2, the 32 K EEPROM is not considered to be an industry standard part, but may serve as an intermediate step until the 64 K parts are available.

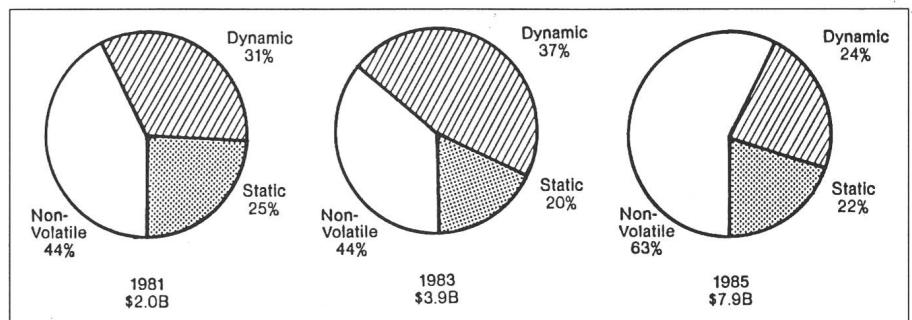


Fig. 1 Actual and projected market potential for MOS type non-volatile memories
Non-volatile includes ROMs, EPROMs and EEPROMs

	SEEQ	XICOR	HITACHI	INMOS
Configuration	8 K × 8	8 K × 8	8 K × 8	8 K × 8
5 V Only (Internal High Voltage)	Yes	Yes ¹⁾	Yes	Yes
Latched Address and Data (Write)	Yes	Yes	Yes	Yes
Latched Address (Read)	No	No	No	Yes
Byte Modify	Yes	Yes	Yes	Yes
Page Write	No	Yes	Yes	Yes
Byte Erase	Yes	Yes	Yes	No
Page Erase	No	Yes	No	Yes
Block Erase	Yes	Yes	Yes	Yes
Internal Timing of Write Cycle	No	Yes	No	No
Automatic Internal Erase	No	Yes	No	No
Read Access Time	200 ns	350 ns		200 ns
Chip Rewrite Time	8–80 s	2.6 s		1.3 s
Active Current	110 mA	110 mA		150 mA
Standby Current	40 mA	50 mA		20 mA

¹⁾ Requires High Voltage (Abnormal Input Levels) on $\overline{WE} + \overline{OE}$ for Chip Erase

2. 64 K EEPROMs

At the time of this writing, four companies have announced 64 K EEPROMs and are apparently close to shipping samples to customers. Table I gives a product comparison of these four products. It is interesting to note that the only commonality of the four parts in the configuration as a byte wide memory (8K×8) and the 5 V only operation. The high voltage required for programming and erasing is generated internally in all four cases. The remainder of the specifications is quite different for the four parts. It will be interesting to find out which part will finally emerge as the industry standard. The technology chosen to build these 64 K EEPROMs (table II) is also quite different, two companies (INMOS and HITACHI) opting for the charge storage in a nitride layer as the basic storage mode (MNOS, SONOS, NITROX, etc.), and two companies (SEEQ and XICOR) employing some form of floating gate technology as the basic storage mode.

The floating gate technology chosen by XICOR is particularly interesting, since it uses a textured poly approach to enhance the program and erase currents by providing local field enhancements. Three levels of polysilicon are used with the second poly layer serving as the floating gate (fig. 3). Poly 1 is the write bit line and Poly 3 the word line. Employing proprietary processing steps, bumps are formed on top of Poly 1 and Poly 2. The field enhancement achieved with these bumps allows use of a fairly thick (800...1000 Å) oxide between Poly 1 and Poly 2, and Poly 2 and Poly 3. This in turn guarantees good data retention and manufacturing yield (low oxide defect density). The triple poly process also makes the technology well scalable since it relaxes the area requirements to achieve an adequate coupling ratio. Standard floating gate technology with only two levels of polysilicon has the drawback that in order to have an adequate coupling ratio for programming and erasing, the capacitance between the word line and the floating gate has to

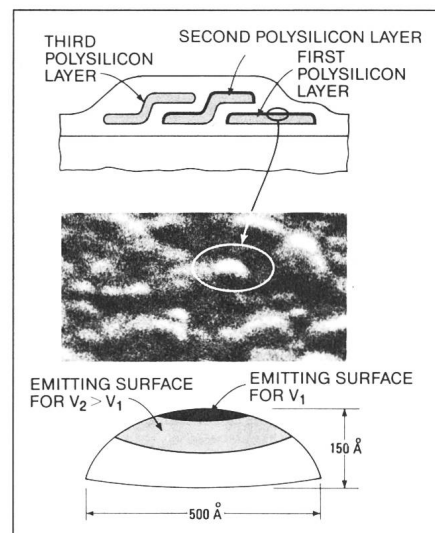


Fig. 3 64 K EEPROM: Floating Gate Technology with Textured Poly
(Courtesy of XICOR Corporation)

be much larger than the capacitance between the floating gate and the channel region, which makes it difficult to scale the cell to minimum lateral dimensions. However, a triple poly process has some inherent disadvantages in manufacturing, such as impaired metal step coverage, which negates some of the advantages mentioned.

A major concern with the textured poly floating gate technology is the write disturb for half selected cells in a memory array. Figure 4 shows the schematic of the XICOR memory cell. The programming voltage V_P between word line and floating gate, which is generated on chip, is typically 35 V for a selected cell that is to be programmed. For half selected cells, that means all the other cells in a row or column that are not to be programmed, the voltage V_P is in the range of 15...20 V. If the memory is specified to have ten years retention and 1 ms write pulses are used—as is the case for the

	SEEQ	XICOR	HITACHI	INMOS
Basic Technology	Si-Gate nMos	Si-Gate nMos	Si-Gate nMos	Si-Gate nMos
Storage Mode	Floating Gate	Floating Gate (Textured Poly)	MNOS (SONOS)	MNOS (NITROX)
Levels of Interconnect	2 Poly, 1 Metal	3 Poly, 1 Metal	2 Poly, 1 Metal	2 Poly, 1 Metal
Isolation	—	—	P-Well, N-Sub	P-Well, N-Sub
Highest Internal Voltage (typical)	18 V	35 V	16 V	20 V
Redundancy	Yes	No	No	Yes
Design Rules	2.5 μm	3.5 μm	2 μm	2.7 μm
Cell Size	168 μm^2	184 μm^2	180 μm^2	135 μm^2
Chip Size	22.4 mm ²		30.0 mm ²	27.6 mm ²

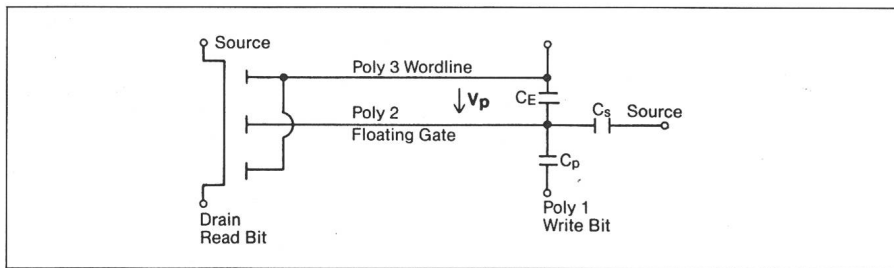


Fig. 4 Cell Schematic of Textured Poly Floating Gate 64 K EEPROM

Selected cells: $V_p = 35$ V, half selected cells: $V_p = 15-20$ V.
For 10 years retention and 1 ms write pulse:

$$\frac{I(V_p = 35 \text{ V})}{I(V_p = 15-20 \text{ V})} = \frac{10 \text{ years}}{1 \text{ ms}} = 3 \times 10^{11}$$

(Courtesy of XICOR Corporation)

XICOR part—the requirements for the dependence of the write current on the write voltage are quite severe. As shown in figure 4, the ratio of the programming current for a selected and a half selected cell must be at least 3×10^{11} . Fortunately, the *Fowler-Nordheim* Tunneling Mechanism which governs the current transport through the oxide has an exponential relationship between current and applied field, as shown in figure 5. The textured poly helps greatly in achieving a very strong current dependence, particularly in the reverse mode compared to standard, planar surfaces.

As mentioned before, the INMOS and HITACHI 64 K EEPROMs employ charge storage in a nitride layer

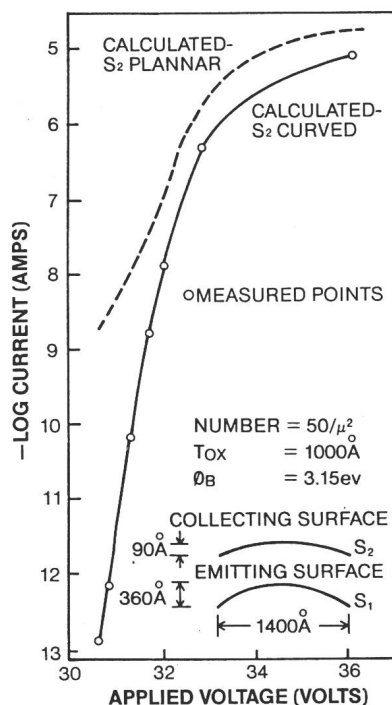


Fig. 5 Reverse Tunneling Characteristics for Textured Poly Floating Gate Technology
(Courtesy of XICOR Corporation)

as the memory mechanism. This technology is at least 15 years old and has been used mostly for small and medium scale integrated circuits in consumer and military applications. The traditional approach uses a metal gate MOS technology (MNOS), whereas the technologies chosen to build high density EEPROMs all employ polysilicon as the gate material in order to achieve the necessary packing density. These modern versions of MNOS are called SONOS, SNOS, NITROX, etc., by the respective vendors.

Figure 6 shows a plane view and figure 7 a cross-section of the NITROX memory cell used in the INMOS 64 K EEPROM (IMS 3630). The cell measures $135 \mu\text{m}^2$ and is built with $2.7 \mu\text{m}$ design rules. It uses two levels of polysilicon with the memory cell built with the second poly level. Using three different gates—an isolate, a storage, and a select gate—it was possible to have a common active area and one contact per cell that is shared with the adjacent cell. This feature helps greatly to achieve a very small cell area.

Apart from the main advantage of being highly scalable, the MNOS or NITROX technology offers good manufacturability because of the in-

Fig. 6 Cell Layout of 64 K EEPROM Cell in NITROX Technology
(Courtesy of INMOS Corporation)

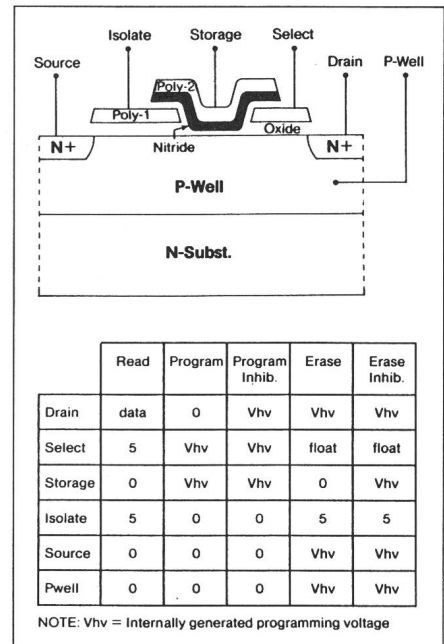
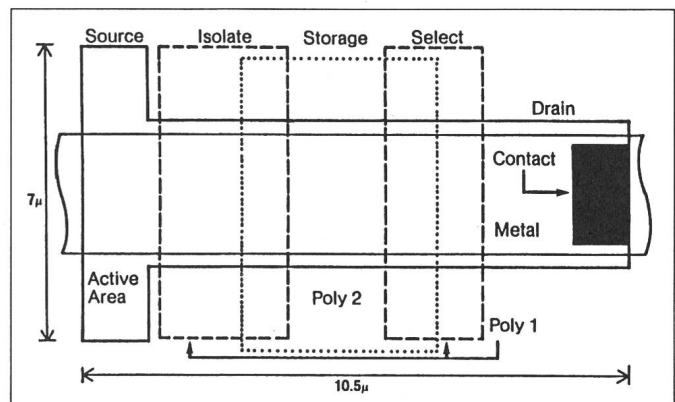


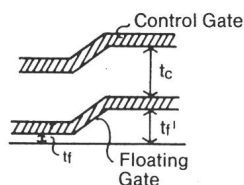
Fig. 7 Cross Section and Cell Operation of 64 K EEPROM Cell in NITROX Technology
(Courtesy of INMOS Corporation)

herently low defect density associated with the nitride-oxide sandwich. MNOS also is proven to have excellent endurance, i.e., write-erase cycles. The main concern with MNOS is retention, the ability to retain information over extended periods of time under worst case conditions, since the same traps that store charge are also responsible for leakage and charge loss. However, a large amount of effort has been expended over the last years to improve and control retention in MNOS. Processing steps such as hydrogen annealing have been developed and high temperature processing has been optimized such that a retention of at least ten years is now guaranteed by most suppliers of MNOS EEPROMs. Retention and endurance will be treated further in this paper.

It is interesting to speculate what might happen if the two competing



CONVENTIONAL:

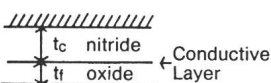


Electric Field for Tunneling:

$$E = V / \left(t_f + \frac{E_f}{E_c} t_c (r - (1-r) \frac{t_f}{t_f}) \right)$$

$$r = A_f / A_c$$

SCALED:



For Scaling:

- (1) Let $r \rightarrow 1$ to minimize area
- (2) Increase control gate dielectric constant; i.e. substitute nitride for oxide

then

$$E = V / \left(t_f + \frac{E_f}{E_c} t_c \right)$$

$$= \text{MNOS}$$

technologies—floating gate and MNOS—are scaled aggressively for future EEPROMs with even higher packing densities. In figure 8 the electric field across the tunneling oxide is given as a function of layer thicknesses and the area ratio r of control gate A_c and floating gate A_f . For aggressive scaling, $r \rightarrow 1$ in order to minimize the area requirement for the cell. To maximize the coupling efficiency between control gate and floating gate, it is advantageous to increase the dielectric constant of the control gate dielectric. Suggestions to use silicon nitride, tantalum oxide, or other dielectric materials with higher dielectric constants instead of silicon dioxide have been made in the literature. If silicon nitride is substituted for silicon dioxide and let $r \rightarrow 1$, equation (1) in figure 8 reduces to equation (2), which is exactly the field expression as a function of dielectric thicknesses for the MNOS structure. This means that in the ultimate scaling limit, floating gate and MNOS technologies may converge.

As an example of a state-of-the-art EEPROM, the functional block diagram of the INMOS 64 K EEPROM is given in figure 9. The memory array uses the NITROX memory cell for charge storage and is organized as 128 rows with each row having 64 bytes. Adjacent to the storage array are column latches used to perform high speed page mode programming. For all cycles, the chip enable clock \overline{CE} latches the addresses, the \overline{PE} and \overline{CS} control functions, and triggers the internal timing generator. The \overline{PE} control signal is asserted only when modification of the non-volatile data storage is desired. When a \overline{PE} cycle is initiated, the \overline{PE} controller executes one of four modification modes, which is

determined by the states of latched addresses A_4 and A_5 . The four modes incorporated on the part are (1) Full-Array Erase; (2) Row Erase; (3) Load-Latches-Row-Erase; (4) Program. The output enable control \overline{OE} is used to control the common I/O data lines, while the write enable \overline{WE} is used to load data into the device in preparation for a non-volatile memory modification. The biases required for all cell functions are also given in figure 7.

The high voltage required for program and erase is generated on chip with a high voltage generator that mul-

tiplies the power supply voltage V_{DD} four times. Column and row pumps decode the high voltage to the individual column and rows. The column and row pumps are designed such that there is no DC loading during deselect.

3. Retention and Endurance

Retention is defined as the time span the memory can store information under worst case conditions. Worst case usually means elevated temperature, maximum bias, and read and/or write disturb. It is important to note that for state-of-the-art EEPROMs with more than 64 000 storage cells per chip, the average or mean retention is not meaningful. A part that shows perfectly adequate retention when all cells are averaged may have a few single bits that show degraded retention characteristics, making the part unacceptable. The only way to ensure that all memory locations show adequate retention characteristics is to develop test screens and accelerated test methods in order to eliminate parts that have single or freak bits with unacceptable retention characteristics. Then, effort has to be expended to decrease the defect density associated

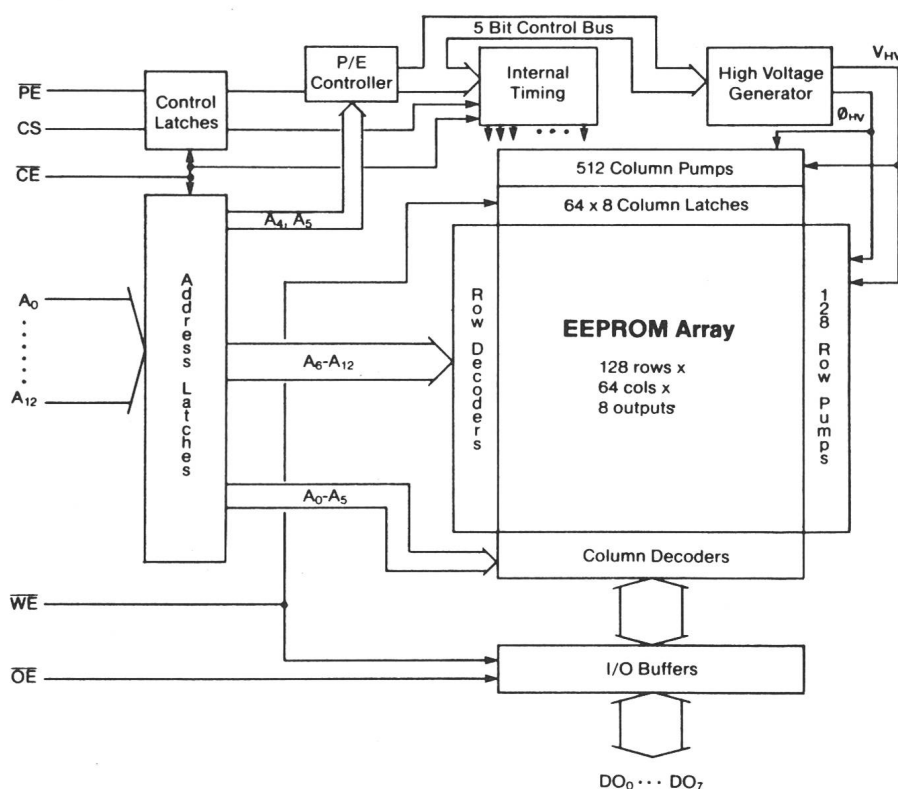


Fig. 9 Functional Block Diagram of IMS 3630
(Courtesy of INMOS Corporation)

with those retention freaks and thus improve the manufacturing yield.

Endurance is usually defined as to how many read/write cycles can be performed before information is lost under worst case conditions. Endurance and retention are not independent, they are interrelated. Usually retention decreases with increasing write/erase cycles. For some technologies, however, it has been reported that retention initially increases with write/erase cycling. The same argument that was made for retention freaks can be made for endurance freaks: it is unimportant what the main population exhibits, the few single bits

that may show degraded performance under write/erase cycling ultimately determine whether a part is good or bad. Again, test screens have to be developed in order to eliminate endurance freaks.

A major advantage of MNOS over floating gate is now apparent: it is much easier to develop retention screens than endurance screens, particularly in a production environment where test times have to be kept short. While MNOS shows predominantly retention freaks, floating gate is mostly plagued by endurance freaks.

An important product feature implemented on the INMOS 64 K EE-

PROM helps greatly in screening product for single bit retention failures. By applying an out of spec test voltage on a particular pin, a built-in margin test circuit is activated that connects the output pad \overline{OE} directly to Poly 2. The actual memory cell window for each cell in the array can now be measured by applying subsequent read cycles. This feature is advantageous both for the supplier who wants to ensure that the product is free of single bit problems, and the customer who can perform meaningful incoming inspection testing to characterize the retention/endurance characteristics of the part.