

In search of a better DRAM: evolving to floating bodies

WIDELY INVESTIGATED FLOATING-BODY MEMORIES APPEAR TO BE COMPELLING REPLACEMENTS FOR CONVENTIONAL DRAMs. A NEW FLOATING-BODY MEMORY USES THE INTRINSIC BIPOLAR TRANSISTOR TO STORE SIGNIFICANTLY GREATER CHARGE.

The memory industry has crammed more and more memory bits onto ever smaller die and is selling those slivers of silicon for a few cents each. Currently, 1-Gbit and even 4-Gbit DRAMs (dynamic random-access memories) are available. Process engineers have been able to achieve these goals, especially with respect to the capacitor element, which has become more difficult to scale, a problem that gets worse as device geometries shrink. These advances employ the basic one-transistor DRAM, which Robert H Dennard, PhD, a fellow at the IBM Thomas J Watson Research Center (www.watson.ibm.com), created in 1966. In 1970, Intel (www.intel.com) released the first DRAM chip, a 1-kbit PMOS device. Since that time, the basic DRAM building block has comprised a single transistor and an increasingly complex capacitor.

Scaling introduces yet another major problem for DRAM manufacturers: leakage current. In both the bit cell and its supporting circuitry, leakage becomes more significant as CMOS (complementary metal-oxide-semiconductor) processing nodes progress from 90 nm through 78, 50, and 45 nm. Manufacturers are now discussing building memory chips at the 32-nm node. At this point, leakage in traditional designs will become a difficult problem and prohibitively expensive to counteract, requiring new architectures, changes to standard operating specifications, and significant process evolutions.

The problems of scaling and leakage, as well as device size, rest fundamentally with the basic transistor-plus-capacitor

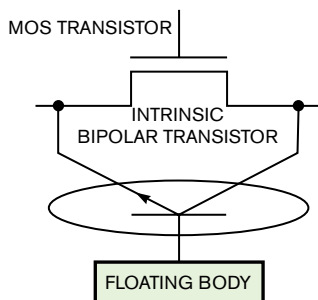


Figure 1 The use of an intrinsic bipolar transistor results in a superior floating-body-memory design.

building block. Although the transistor element is theoretically infinitely scalable, at least for the foreseeable future, the capacitor is not. Manufacturers can fabricate capacitors as either high stacks or deep trenches. However, if the overall bit cell shrinks due to increased density or a smaller process node, the capacitor must become higher or deeper to maintain the minimum charge necessary for reliable operation.

The memory industry is fast approaching the scaling limits for the capacitor element, and it is therefore time for a new approach. There appears to be a groundswell of opinion in the commercial world and the analyst community that floating-body-based

memories may provide the answer. The floating-body effect is the dependence of the body potential of a transistor, using the SOI (silicon-on-insulator) technology, on the history of its biasing and the carrier-recombination processes. The transistor's body forms a capacitor against the insulated substrate. The charge accumulates on this capacitor and may cause adverse effects—for example, opening parasitic transistors in the structure and causing off-state leakages, resulting in higher current consumption and, in DRAM, the loss of information from the memory cells.

Multiple significant companies, notably Samsung (www.samsung.com), Intel, and STMicroelectronics (www.st.com), have published work on the subject, and at least three pa-

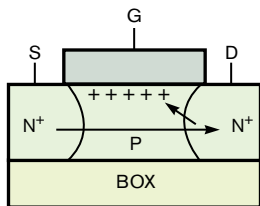


Figure 2 The write cycle for a one data pattern generates current flow through the transistor body.

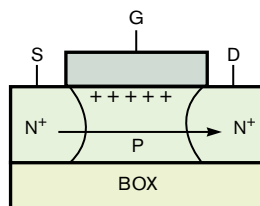


Figure 3 A read-cycle mechanism senses bipolar current flow.

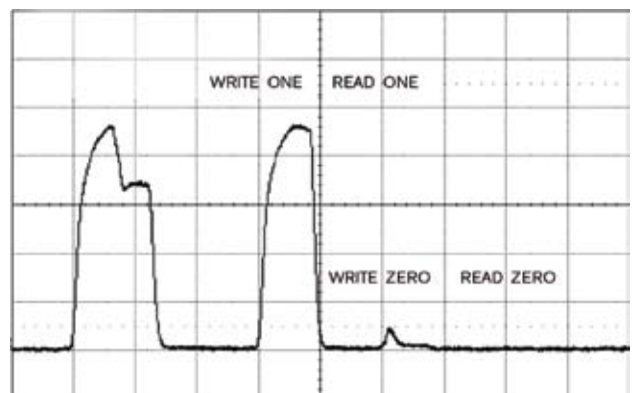


Figure 4 Read and write currents are nearly identical.

pers presented at the International Electron Devices Meeting, which took place in December 2008 in San Francisco, addressed floating bodies. Equally significant, at least one major stand-alone-memory maker, Hynix (www.hynix.com), and one processor company, AMD (www.amd.com), have signed

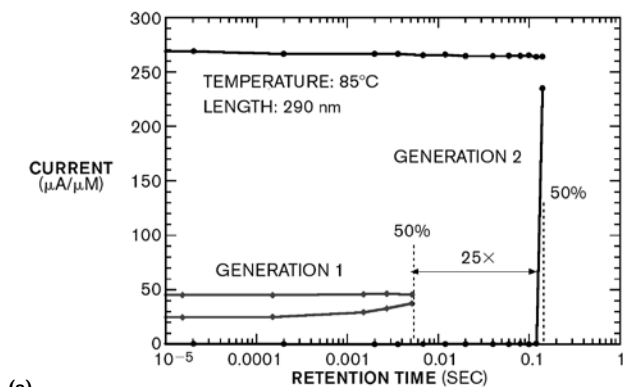
licenses with a floating-body-memory company to develop products.

The floating-body effect naturally occurs in transistors fabricated on SOI substrates, leading to the accumulation of charge in the transistor body. Recently introduced device types such as FinFETs (fin-shaped field-effect transistors) and surround-gate, or pillar, transistors also demonstrate a floating-body effect, even when you implement them on traditional bulk-CMOS substrates. For these reasons, engineers no longer regard the floating-body effect as a parasitic nuisance. Many engineers have tried to manipulate and enhance the body charge so that they can use it to reliably store a logical “state” and function as a memory element. The objective is a memory bit cell that comprises only one transistor, fundamentally the simplest and most scalable of all semiconductor devices.

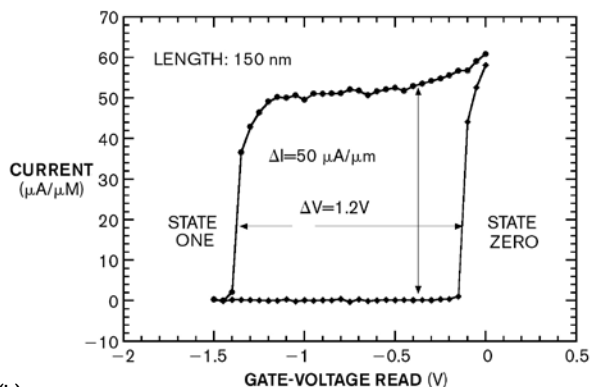
Although a number of companies have investigated floating-body memories, the conventional approach is unlikely to lead to a manufacturable product. A second generation of floating-body memory must be able to store a significantly larger charge in a smaller transistor. This increased amount of charge greatly improves the amount of time the memory can retain its state as well as the signal margin between a one state and a zero state. Other improvements include faster reads and writes and reduced write power consumption.

PRINCIPLES OF OPERATION

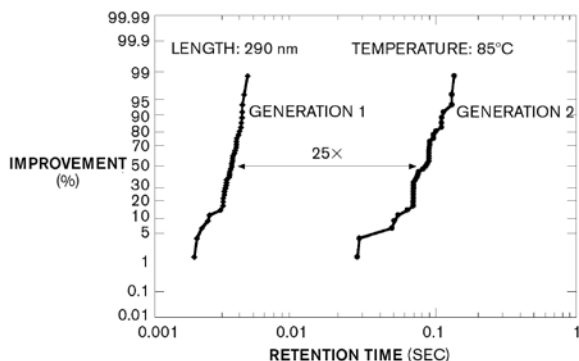
Early attempts at realizing usable floating-body memories employed a MOS transistor to pass current and create charge in the body using impact ionization. Although you can demonstrate memory performance using such a technique, the amount of charge you create with this method is insufficient to create a robust and manufacturable memory device. A superior approach is to use the bipolar transistor intrinsic in



(a)



(b)



(c)

Figure 5 A retention-measurement comparison of first-generation (MOS-transistor element only) and second-generation (added bipolar element) floating-body memories reveals the second generation’s superiority (a). The bipolar-inclusive floating-body approach also delivers a more substantial programming window (b), along with a 25-fold improvement in retention time (c).

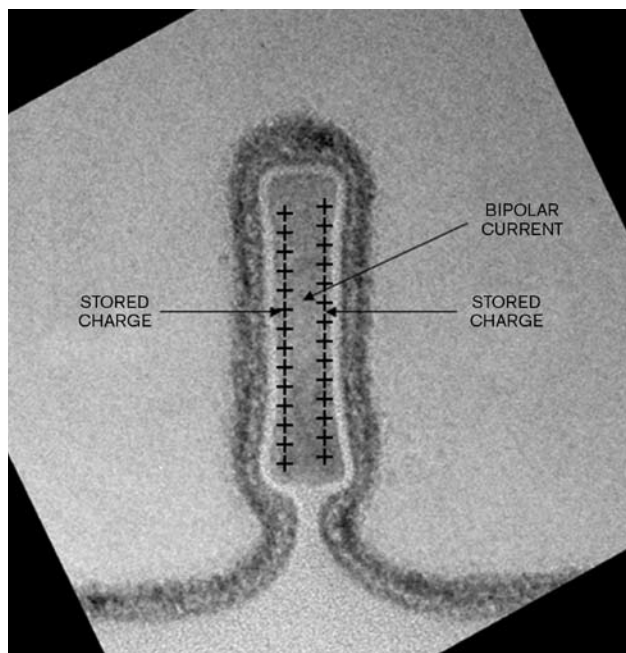


Figure 6 A FinFET design is amenable to floating-body techniques.

the SOI-MOS structure to create charge (Figure 1). This approach allows the creation of a much larger charge and the ability to store more charge because of the increased capacitance of the memory cell.

If you consider an N-channel device, the N+ source, the P-type body, and the N+ drain form the emitter, base, and collector, respectively, of an NPN (negative-positive-negative) bipolar transistor. The body of the MOS transistor is the base of the bipolar transistor and acts as a storage node. Writing a one into a second-generation bipolar floating-body-memory cell triggers the intrinsic bipolar transistor, causing current to flow throughout the transistor body. This approach differs significantly from the behavior of MOS, in which current flows only at the interface. Charge collects at the interface due to the slight bias at the gate. The impact-ionization effect that creates an excess of majority carriers in the floating body is more efficient in this bipolar bit-cell structure, quickly charging the body and therefore resulting in rapid writes (Figure 2).

You can read a floating-body memory using a similar mechanism, which senses the bipolar current through the transistor (Figure 3). Write current is close in value to the read current, and the latch-up characteristic of the intrinsic bipolar quality causes the behavior of the memory cell to appear nearly digital (Figure 4).

CELL MARGIN AND SCALABILITY

Bipolar floating-body memories have a significantly higher operating margin than traditional floating-body devices, al-

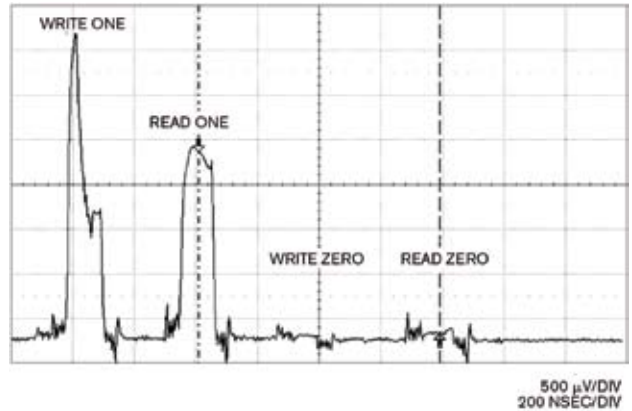


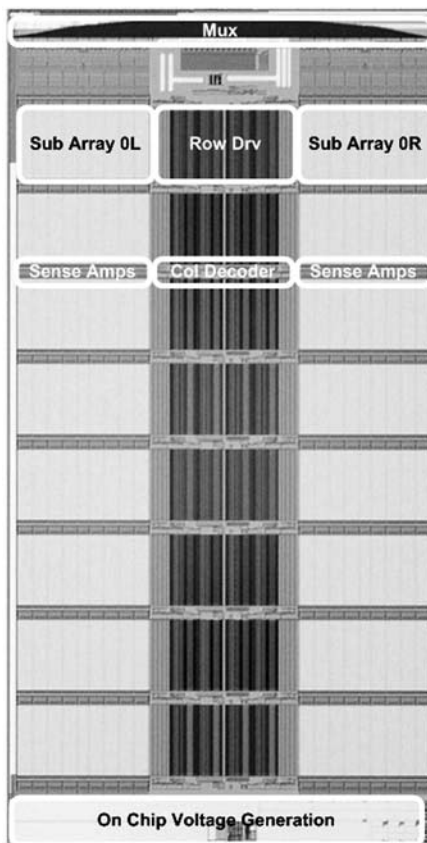
Figure 7 The read and write currents for a FinFET with a length of 55 nm and width of 11 nm reveal its robust capabilities.

lowing them to create faster memory bit cells. The operating margin is better for two reasons: The storage charge is higher, and the difference between the one and the zero states is much larger because the bipolar element is a better amplifier than a MOS device (Figure 5). A high cell margin eases sensing using a simple sensing scheme. The approach also simplifies the sense amplifier's design. When you implement floating-body memories with conventional planar transistors, the memories'

large voltage margin helps to mitigate the negative effects of process fluctuations, which become increasingly significant at smaller process geometries. Perhaps just as important, the new bipolar floating-body-memory designs are compatible with advanced, nonplanar devices, such as FinFET, multigate FET, and gate-all-around FET. Older floating-body designs work on only thin-film planar devices.

FINFETs AND ARRAYS

FinFETs, surround-gate transistors, and other similar 3-D structures should provide the basis for future stand-alone memories as the industry scales to ever smaller process geometries. Working at the 54-nm node, some DRAM companies are using FinFET designs, and most suppliers should follow in the next five years. Manufacturers can fabricate FinFETs and pillar transistors on both SOI and conventional bulk substrates. In a FinFET or trigate-based Z-RAM (zero-capacitor RAM), the



TECHNOLOGY	45-NM PARTIALLY DEPLETED SOI
MEMORY TRANSISTOR	WIDTH: 112 NM, LENGTH: 112 NM, OXIDE THICKNESS: 3 NM
CELL SIZE	0.480×0.192 MICRONS (0.0922 MICRONS ²)
MACRO SIZE	1323×718 MICRONS (FOR 4.44 MBITS)
DENSITY	0.21 mm ² /MBIT
SUBARRAY	LOCAL SENSING WITH 260 ROWS/BIT LINE AND 1120 BITS/ROW
REDUNDANCY	FOUR ROWS AND 16 COLUMNS PER SUBARRAY
ECC	9 ECC CHECK BITS PER WORD
SENSING	VOLTAGE, EIGHT-BIT-LINE PITCH, CORE TRANSISTORS
ROW DRIVERS	FOUR-WORD-LINE PITCH, WORD-LINE DRIVERS WITH BT DUAL-GATE OXIDE DEVICES
INTERFACE	140-BIT DATA WORD, 15-BIT ADDRESS, 7-BIT CONTROL
PERFORMANCE	4-NSEC RANDOM ACCESS, 2-NSEC* READ LATENCY, 2-GHz CYCLE TIME

*WORD LINE HIGH TO SENSE AMPLIFIER'S OUT TIME

Figure 8 You can implement a 4-Mbit macro in a 45-nm SOI process.

charge accumulates under the transistor gate, and the current flows in the middle of the fin structure. The fin stores charge throughout the structure, permitting excellent control of the bipolar current (**Figure 6**). Bipolar floating-body-memory implementations using FinFET structures tend to exhibit an approximately 30-fold increase in margin, leading to a proportional increase in signal and device speed. **Figure 7** shows the cell current during write and read. Even an undoped cell with an 11-nm fin shows clean digital behavior and a good margin. Floating-body memory continues to scale with technology whether you fabricate on planar or 3-D structures.

For the successful implementation of floating-body memories, it is crucial to create memory-bit-cell arrays. Original floating-body implementations, due to the limited possible margin between the one and the zero states, are relatively slow and therefore unsuitable for combining into useful arrays. However, new bipolar floating-body-memory cells, which produce superior signal margins due to the large current gain available from the bipolar device, enable more robust arrays (**Figure 8**).

The macro for these memories supports a 2-nsec read latency and a 4-nsec write latency. It also implements a high-speed page mode to read a 140-bit word during every 2-GHz-processor clock cycle, using a shared-I/O bus and data latches at the interface. This macro supports consecutive memory operations at 4 nsec and a burst of four words during page mode. The 4-Mbit memory macro comprises 16 subarrays of 256×1024 -bit,

Go to www.edn.com/ms4332 and click on Feedback Loop to post a comment on this article.

For more technical articles, go to www.edn.com/features.

or 256-kbit, cells (**Figure 8**). The exact implementation, 260×1120 bits, includes redundancy and ECC (error-correcting code). Subarrays share two banks of sense amplifiers at each end of every subarray and each bank with adjacent subarrays. A combination of column and row operations allows the access of four 140-bit words by simultaneously activating two subarrays on the right and left sides of a shared block of row drivers.

A subset of combined read- and write-restore operations constitutes a refresh operation. A control block outside the 4-Mbit macro initiates refresh. A low-power refresh mode minimizes the voltage swings on inactive bit lines between the read- and write-restore operations of the refresh cycle. The memory macro has a refresh cycle of 1 msec at 105°C, along with three external power supplies of 1.1, 2.6, and 0.5V. A dedicated on-chip voltage-generation block attaches to the memory macro and supplies the required operational voltages. **EDN**

AUTHOR'S BIOGRAPHY



Serguei Okhonin, PhD, is co-founder and chief scientist of Innovative Silicon, where he has worked for six years. He has a master's degree in physics from Novosibirsk State University (Novosibirsk, Russia) and a doctorate from the Swiss Federal Institute of Technology (Zurich, Switzerland). His personal interests include skiing, rock climbing, and mountaineering. You can reach him at sokhonin@z-ram.com.