Experimental Memory Chips Reach 1 Megabit

As they become larger, memories become an increasingly important part of the integrated circuit business, technologically and economically

The well-chronicled advance of the microelectronics industry is evidenced most visibly in the appearance every 3 years or so of a new generation of dynamic random access memory (DRAM) integrated circuits. With a storage capacity measured in units of 1024 or 1K binary bits, the first DRAM to be competitive with magnetic core memory in computers was a 1K chip in the early 1970's. Each succeeding generation has had quadruple the capacity of its predecessor.

Today we are in the midst of the 64K generation. However, 256K DRAM's are beginning to be shipped to the first customers in small quantities. And 1024K or megabit DRAM's are already on the horizon. At the International Solid State Circuits Conference, held in February in San Francisco, three Japanese firms discussed the first experimental 1-megabit (1-Mb) DRAM chips. And last month, the first American entry arrived with the announcement by IBM that it had made a working 1-Mb DRAM.

Economics provides a perspective on the importance of DRAM's. Consider the Sturm und Drang in the U.S. microelectronics industry during the heyday of the 16K chip in the late 1970's, when the Japanese captured 40 percent of the market and thereby became serious competitors. In 1979 worldwide DRAM sales amounted to $650 million, according to figures cited by Daniel Klesken of Montgomery Securities, a San Francisco financial firm, at a forum on semiconductor memories that was held just prior to the conference.

The stakes are rising rapidly. In 1984, the Japanese are expected to maintain a 70 percent share of a $3.4 billion market dominated by the 64K DRAM. And in 1989, during the 256K era, Klesken projects DRAM sales to rise to $9.4 billion. Moreover, while DRAM's are only one of a growing family of memory chips, they should account for almost half of all semiconductor memory sales this year and almost 20 percent of all integrated circuits. And these percentages are expected to grow in the future.

Fueling the explosive growth in the appetite for memory chips has been an exponentially decreasing cost per stored bit due to the growing storage capacity at a relatively constant chip price. Speaking at the forum, Robert Frankenberg of the Hewlett-Packard Company extrapolated current trends in memory growth. He estimated that a main memory for multiuser computer systems would increase to 100 megabytes by 1990, up from about 8 megabytes in 1982.

Main memories for personal (one-user) computers would soar during the same period to 5 megabytes. Since a byte is 8 bits, this adds up to 40 1-Mb DRAM's in a personal computer. For comparison, today's IBM PC has space for 32 64K DRAM's or 256 kilobytes. Among the capabilities a larger main memory would give to a personal computer, Frankenberg listed ease of use, access to large data bases, fast performance, voice and image transmission, and sophisticated graphics.

The stakes are rising rapidly. In 1984, the Japanese are expected to maintain a 70 percent share of a $3.4 billion market.

What does a DRAM do? Random access memories of any type, DRAM or other, allow direct access to any item of stored information without having to sift through unwanted items on the way, as in serial access devices such as cassette tapes. Most semiconductor memories are random access memories. The basic structure comprises two orthogonal arrays of parallel electrically conducting wires imprinted on a silicon chip. A memory cell connects the wires at each intersection of the arrays.

Within this basic structure, there are numerous subspecies of random access memory, according to the nature of the memory cell. The two main distinctions are between volatile and nonvolatile memories and between read/write and read only or mostly memories. In practice, the two distinctions tend to make the same cut through the memory varieties. Read/write memories are almost always volatile; that is, all information is lost when the computer is turned off or there is a power failure. Read only or mostly memories must be nonvolatile.

A DRAM is a volatile, read/write memory. The binary information in a DRAM cell consists of the presence of a stored electrical charge on a capacitor. Two different charge levels correspond to the binary 0 and 1. Since the charge tends to leak away, often as fast as a few milliseconds, it must be regenerated or refreshed at regular intervals. The refreshing operation is the dynamic part of DRAM. DRAM's comprise the largest segment of the semiconductor memory market because the small memory cell allows chips of large memory capacity and because their general applicability permits identical chips to be manufactured in volume and hence cheaply.

Imagine the wires of the orthogonal arrays to run horizontally and vertically. A horizontal wire is called a word or select line, while a vertical wire is a bit or data line. Decoder circuits translate a request to read from or write into a specific memory cell into electrical signals for the appropriate word and bit lines. The DRAM memory cell at the intersection of the word and bit lines comprises one metal-oxide-semiconductor (MOS) access transistor and a storage capacitor (see figure). Note that one plate of the capacitor is attached to a voltage source or to ground by way of a wire in a third array.

During the writing operation, the access transistor, which normally is off and does not let a signal pass, is turned on by a voltage pulse on its word line. Another voltage pulse from the memory's input/output circuits that corresponds to the data to be stored reaches the capacitor through the bit line and the on transistor. The voltage results in a stored charge on the capacitor.

For reading, the access transistor is again turned on by a voltage pulse on its word line. This connects the capacitor by way of the bit line to a sense amplifier, which detects the voltage corresponding to the stored charge and passes on a signal to the input/output circuits. The sense amplifier also refreshes the memory cell by restoring the bit line to the original voltage written into the cell.

Despite the geometric regularity of the memory array, DRAM's are quite complex devices. The complexity derives from the need to properly time the voltages on all wires so that the writing, reading, and refreshing operations take

place without errors. Access times are of the order of 100 nanoseconds.

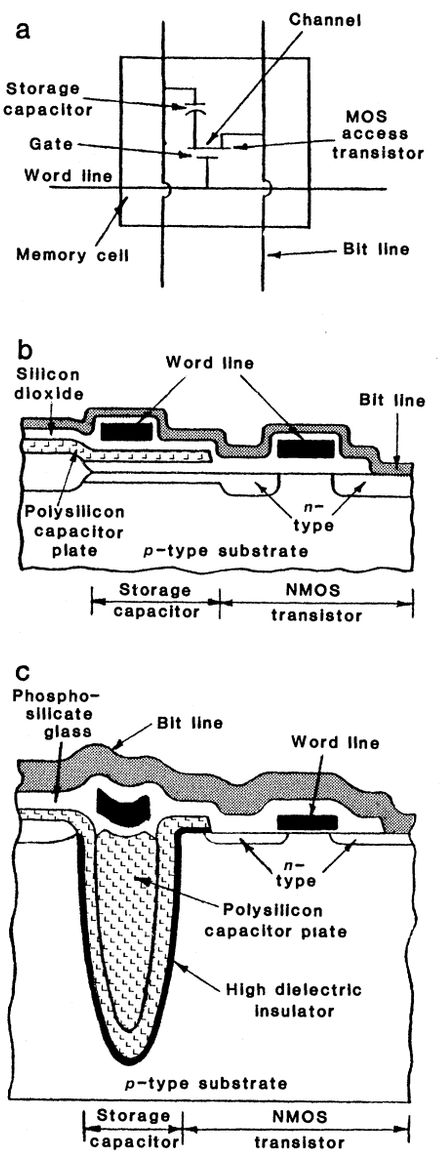
Although neatly represented on a circuit diagram, the memory cell presents a considerably nastier problem when implemented in physical form. Miniaturization heightens the difficulties. Consider the storage capacitor. For densely packed arrays, one wants small capacitors. In the progression from 16K to 256K DRAM's, the area of the memory cell (transistor plus capacitor) has shrunk from about 450 square micrometers to less than 100 square micrometers now and perhaps 50 square micrometers as technology improves.

The size cannot shrink indefinitely, however. Stored charge is proportional to the capacitance, which is proportional to the area of the capacitor plates. A small stored charge is more difficult for the sense amplifiers to detect, suggesting a lower limit to capacitor size. Part of the difficulty is that the bit line has a capacitance of its own, 10 to 20 times larger than that of the storage capacitor. During the reading operation, part of the stored charge drains out of the storage capacitor and charges the bit line capacitance, so that the signal reaching the amplifier is reduced to the nominal value times the ratio of the storage to bit line capacitances. This is also one reason why refreshing is needed.

An additional problem comes from bombarding alpha particles. Effects due to alpha particles, which come from radioactive contaminants in the materials making up the packages that integrated circuits rest in, as well as from cosmic ray–induced nuclear reactions in the chip itself, first became annoying in 16K DRAM's and serious in 64K devices. An alpha particle passing through a memory cell induces a charge. Whether the induced charge changes the content of the cell depends on how large it is compared to the stored charge. A 50-femtofarad capacitor charged to 5 volts stores about 1.5 million electrons, just enough to withstand a typical alpha hit. As it happens, 50 femtofarads is about what the capacitance in 256K DRAM memory cells is.

At the solid-state circuits conference, Kiyoo Itoh of the Hitachi Central Research Laboratory in Tokyo discussed how he and six colleagues addressed these and other problems in a 1-Mb DRAM. According to Itoh, the signal to be detected by a sense amplifier has been dropping—for the reasons discussed above—from 300 millivolts in 16K DRAM's to a projected 11 millivolts in a 1-Mb chip. The Hitachi engineers managed to raise the signal back up to 250 millivolts by increasing the storage capacitance by a factor of 7.5 and reducing the bit line capacitance by a factor of 3.

Ordinarily, the capacitor in a DRAM memory cell is built around a thin layer (a few hundred angstroms) of silicon dioxide as the dielectric. One end of the access transistor, which may extend some distance under the silicon dioxide, and a layer of electrically conducting polycrystalline silicon, which extends over the silicon dioxide, act as the plates of the capacitor (see figure). The capacitance could be increased by decreasing the oxide thickness. Engineers at IBM's General Technology Division, Essex Junction, Vermont, built a capacitor only 150 angstroms thick in their 1-Mb DRAM. However, too thin a dielectric heightens the chance of electrical breakdown of the oxide when the gate voltage is applied.



DRAM cell

(a) Schematic diagram of memory cell. (b) Conventional physical implementation of the cell. (c) Hitachi's trench capacitor with a larger plate area and hence larger storage capacity. The MOS transistor is an n*-channel device. With a positive voltage on the gate (word line), normally sparse electrons in the* p*-type silicon are attracted to the area under the gate. This forms an* n*-type channel between the* n*-type regions so that current can flow between them, thereby turning the transistor "on." [Drawing by Eleanor Warner]*

To boost the capacitance, the Hitachi engineers fabricated a 4.5-micrometer-deep trench that they lined with a "high dielectric" insulator consisting of silicon nitride sandwiched between silicon dioxide layers. The dielectric reaches over the edge of the trench to slightly overlap one end of the transistor, which is the location of one capacitor plate. The trench was filled in with electrically conducting polycrystalline silicon to form the other plate of the capacitor. By effectively increasing the area of the capacitor in this way, they achieved a capacitance of 60 femtofarads in a memory cell only 21 square micrometers in area, as compared to IBM's 56-square-micrometer cell.

Trench capacitors were also featured in another 1-Mb DRAM described at the conference by Junzo Yamada and four co-workers from Nippon Telegraph and Telephone's Atsugi Electrical Communication Laboratory near Tokyo. The NTT engineers built a memory cell of 20 square micrometers that had a capacitance of 30 femtofarads by means of a trench capacitor of a different design than Hitachi's. To ensure a usefully large signal for the sense amplifiers, they also lowered the bit line capacitance.

To achieve this, the NTT engineers devised a circuit trick that effectively shortened the bit line. All other things equal, the capacitance of a wire is proportional to its length. In this way, the engineers achieved a bit line capacitance of only 270 femtofarads. The Hitachi group did only slightly worse, 290 femtofarads, by a different strategem that also acted to shorten the bit line.

Another aspect on which the Hitachi and NTT engineers took similar views is the need for reduced voltages. Since the late days of the 16K DRAM era, the standard operating voltage has been 5 volts. But as device sizes shrink, problems can arise. As the channel length between the ends of the transistor has dropped from 6 or 7 micrometers in 16K chips to about 1 micrometer in 1-Mb DRAM's, "short channel effects" come increasingly into play. For example, "hot electrons" with nonthermal energy distributions that are produced by high

electric fields can change the turn-on/turn-off characteristics of a transistor. Such fields are increasingly produced near the ends of the channel as the transistor shrinks.

Solutions up to now have involved modifying the transistor structure in such a way that the field is reduced. But eventually, it will be necessary to reduce the operating voltage. Both Hitachi and NTT have incorporated circuitry on their DRAM's that lowers the 5 volts from the external power supply to 3.7 volts and 3 volts, respectively. As discussed by several participants in one of the traditional "informal" discussion sessions at the solid-state circuits conference, no manufacturer wants to be the first to change an industry standard. For the moment, on-chip voltage conversion preserves the standard power supply that types of chips other than 1-Mb DRAM's will continue to use for a while.

In one way, Hitachi and NTT did take quite different approaches to their 1-Mb DRAM's. NTT incorporated error-correcting circuits on its chip to test for and fix errors due to alpha particles. The presence of error-correcting circuits somewhat relaxes the need for the highest possible cell capacitance as a means of resisting the effects of alpha particles.

The third 1-Mb DRAM was presented by Masumi Nakao and five associates from the NEC Corporation in Kawasaki. With an access time of 120 nanoseconds, the NEC DRAM comes between Hitachi (90 nanoseconds) and NTT (140 nanoseconds). But because of its large memory cell (44 square micrometers with a capacitance of 50 femtofarads), the overall chip size is larger by quite a lot—76 square millimeters as compared to 46 square millimeters (Hitachi) and 52.5 square millimeters (NTT). Chip size is a major concern. Large chips carry higher fabrication costs and higher probabilities of defects. And they will not fit in the standard memory packages.

IBM's chip is the largest of all at 81 square millimeters. But the company makes chips only for its internal use and has developed sophisticated nonstandard chip carriers for its unorthodox logic and memory integrated circuits.

NEC's chip illustrates one approach to the increasingly serious problem of what materials to use for word and bit lines, as well as other interconnections in DRAM's and other types of integrated circuits as miniaturization proceeds apace.

In 16K DRAM's, engineers used aluminum for the word lines. However, for convenience in the fabrication process, they used electrically conducting silicon for the bit lines. Since then, the desire to minimize bit line capacitance and to keep the resistance down as dimensions decreased has caused most engineers to switch to aluminum bit lines in the upcoming generation of 256K DRAM's. IBM is an exception. It has stuck with its traditional manufacturing process even in its 1-Mb chip.

Use of aluminum for both bit and word lines has some drawbacks. Aluminum does not withstand high temperatures once it is laid down, so use of this metal for both lines requires a production process that has no high-temperature steps following the deposition of aluminum. A different response has been to use for the word line a refractory metal that withstands high temperatures, a refractory metal silicide, or even a multilayer structure comprising both polycrystalline silicon and silicide. In its 1-Mb DRAM, for example, NTT used aluminum for the bit lines and molybdenum for the word lines. NEC, however, chose to go all the way with aluminum, using it for both word and bit lines.

It takes annual R&D expenditures of $50 million to remain in the DRAM game.

Just as interesting is the organization of the NEC and IBM DRAM's. The normal organization, that of the Hitachi and NTT chips, is called ×1, which means that only one bit in a sequence of bits is accessed and read out at a time. Computers are organized to think in words, with words ranging from 8 to 64 bits, so it would be faster to have access to a word at a time. Large computers with huge numbers of DRAM's accomplish this by assigning a different DRAM to each bit of a word. A 32-bit mainframe for example, would access 32 DRAM chips for each word. Personal computers, which typically have 8- or 16-bit words, cannot afford this luxury since they do not usually have space to expand their memory in blocks of 8 or 16 DRAM chips.

The result is a trend (already established in other types of memory chips) toward offering DRAM's in several organizations. The new ones allow access to several bits on one DRAM simultaneously. A ×4 allows access to 4 bits, a ×8 to 8 bits, and so on. Manufacturers are offering new versions of their established 64K DRAM's that do this, and ×4 and ×8 versions of 256K DRAM's already exist. NEC's 1-Mb DRAM is organized as a 128K ×8 memory. IBM's chip is manufacturable in ×1, ×2, and ×4 organizations. To avoid separate production lines, choice of organization comes only in the last chip processing step.

One feature that none of the three companies availed themselves of is redundancy. The problem of yield (percentage of defect-free chips) is a crucial one in integrated circuits but especially so for DRAM's, where the ability to make chips inexpensively is paramount in a highly competitive business. Redundancy is the practice of placing extra rows or columns of memory cells on the chip. If a defective cell is found in one row or column, it can be disconnected and replaced by its unused counterpart. Rewiring of the finished chip to accomplish this is by means of "fusible links" that can be blown out by passage of a large electric current or irradiation by a laser. Almost all 256K DRAM's make use of redundancy, and 1-Mb devices surely will as well.

As it happens, of the four experimental DRAM's, only IBM's conservative chip was fully functional; that is, every bit could be accessed. Nonetheless, 3 years ago, 256K DRAM's were in a similar stage of development but are now nearing commercial availability. Evidently, 1-Mb DRAM's are not far off.

When they do arrive, who will make them? A few years back, in the 16K DRAM era, forecasters noted that the integrated circuit business was becoming highly capital intensive, partly because of the advanced and expensive equipment that would be needed to make future circuits with features, such as the width of metal bit lines, of the order of 1 micrometer. As a result, they said, only a few successful firms would survive. The capital intensive part was true, as Gordon Moore, chairman of the Intel Corporation, the Santa Clara, California, company that started the DRAM business, noted at the memory forum. It now takes several hundred million dollars of capital and annual R&D expenditures of $50 million to remain in the DRAM game.

However, the memory market has become a huge and inviting target. Where there were 17 suppliers of 16K DRAM's, there are 18 for 64K chips, and Montgomery Securities' Klesken expects 22 for 256K devices, including possibly four new Korean government-sponsored start-ups. The prospects are thus good for continued successful American participation, even if the United States never again dominates the way it did until the late 1970's.—**ARTHUR L. ROBINSON**