CMOS Future for Microelectronic Circuits

Low power consumption of complementary metal-oxide-semiconductor integrated circuits drives next stage of ultraminiaturization

Looking for a winner? Consider complementary metal-oxide-semiconductor (CMOS) integrated circuits.

MOS transistors are the feedstock for the manufacture of the most densely packed integrated circuits, computer memory chips. According to figures compiled by Lane Mason of Dataquest, Inc., a Cupertino, California, market research firm, MOS memory sales amounted to $4.02 billion last year, with CMOS making up just 14 percent of the total. Projections for 1988, when marketing of the next generation of even more highly miniaturized chips will be in full swing, have MOS memory sales almost quadrupling to $15.5 billion. However, the CMOS share will be a whopping 70 percent!

A growing number of integrated circuit manufacturers are hopping on the CMOS bandwagon. Some are pledging to develop only CMOS parts from here on out. For example, just prior to the annual International Solid-State Circuits Conference, held most recently in San Francisco in February, the Intel Corporation, Santa Clara, California, discussed with the press its intention to concentrate entirely on CMOS, once current projects using the predecessor *n*-channel MOS (NMOS) technology are completed. Intel backed up its claim by announcing last month that it was taking orders for a 64K dynamic random access memory (DRAM) implemented in CMOS. (Memory storage is measured in units of 1024 = 1K binary bits.)

DRAM's are especially significant because they are the largest selling type of integrated circuit. Intel's 64K CMOS chip is the first CMOS DRAM commercially available. At the solid-state conference, three Japanese firms presented preliminary data on experimental 1024K (1 megabit) DRAM's, and last month IBM disclosed that it too had a 1-megabit chip (*Science*, 11 May, p. 590). IBM's and two of the Japanese devices were implemented in NMOS, while one was in CMOS, but commercial versions of 1 megabit DRAM's are likely to be all in CMOS, according to most observers.

Memories are not the only home for CMOS. Digital logic circuits, microprocessors, and analog (linear) circuits are also expected to benefit from this venerable but only recently highly appreciated technology. However, new ideas tend to turn up first in memories, and the cost of CMOS versions of at least one kind of memory chip, the static random access memory (SRAM), is nearing that of its NMOS antecedents.

CMOS is not a new idea, having been part of the MOS integrated circuit scene almost from the beginning. In the mid-1970's, however, RCA's solid-state division in Somerville, New Jersey, was nearly alone in proclaiming its virtues. At the time, low power consumption was the principal benefit conferred by the use of CMOS. This made it ideal for digital watches and the like, which had to be battery powered. But NMOS had better performance, was simpler to manufacture, and lent itself more readily to the miniaturization (more computer power on a chip without an increase in the price of the chip) that fueled the microelectronics revolution.

Over the course of time, power consumption has taken on an increasingly critical role in the march of miniaturization. In short, nearly all the advantages of NMOS have evaporated as engineers have struggled to deal with this problem. CMOS is on the verge of being the higher performance technology, while it maintains its traditional low power consumption.

Fear of the electricity bill is not the main power consumption issue in today's microelectronics. Power consumed appears as waste heat. The metal pins that connect integrated circuits to a circuit board can dissipate up to a few watts. A chip generating much more than that will overheat with concomitant effects on performance and reliability. Moreover, the cost and size of the computer or electrical system built around integrated circuits depends sensitively on the heat generated during circuit operation. Writing in the 5 April issue of *Electronics*, Walter Niewierski of the Harris Semiconductor Corporation, Melbourne, Florida, illustrated the effects of heat generation on a microcomputer system.

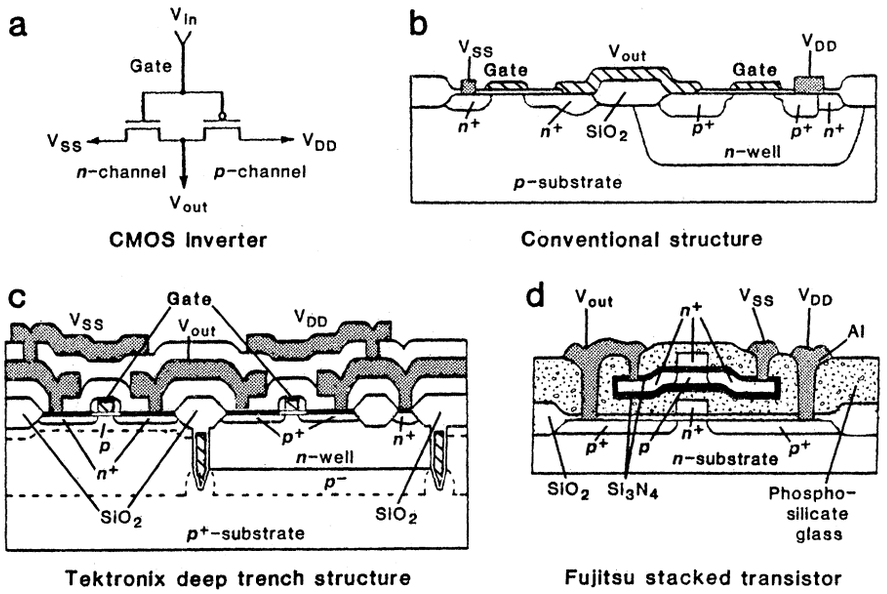
Niewierski considered a 16-bit microprocessor implemented in NMOS that generated about 1.7 watts. If memory chips and other peripheral circuits (also mostly NMOS) are counted in, the total power consumption could be as high as 50 watts. Chip temperatures would rise about 50°C, requiring heat sinks, fans, and vents.

If the same microcomputer were constructed around CMOS chips, power consumption would plummet to less than 10 percent of the previous value. Chip temperatures would increase no more than 5°C, obviating the need for heat sinks, fans, and vents. With CMOS, Niewierski concluded, a bulky, power hungry microcomputer could turn into a lightweight, battery-powered, portable unit that could even be sealed against dust and other environmental insults. As it happens, Harris manufactures a 16-bit CMOS microprocessor, the only one now available.

A growing number of integrated circuit manufacturers are hopping on the CMOS bandwagon. Some are pledging to develop only CMOS parts from here on out.

An inverter circuit shows why CMOS is such a power saver. An inverter reverses the high and low voltages corresponding to binary 1's and 0's in digital electronics; that is, a binary 0 input becomes a binary 1 output and a 1 becomes a 0. A CMOS inverter comprises one *n*-channel and one *p*-channel transistor connected in series drain to drain (see figure). The input signal is applied simultaneously to the gates of both transistors. The output signal is taken from the node between the two transistor drains.

A high positive input voltage turns on the *n*-channel but not the *p*-channel transistor, while a low positive (relative to the voltage on its source) voltage turns on the *p*-channel but not the *n*-channel transistor. Except for a brief period during the switching between on and off



CMOS inverter

The inverter is the basic CMOS circuit and comprises a pair of n- *and* p-*channel transistors. Electrons carry the current in* n-*type silicon, whereas positively charged "holes" (missing bonding electrons) do the job in* p-*type material. When an* n-*channel MOS transistor is "on," electrons flow from one end, the source, to the other, the drain. Both the source and drain are islands of* n-*type silicon on a* p-*type substrate, the chip. Because of the* p-*type channel between the source and drain, current normally does not flow, and the transistor is "off." Application of a positive voltage to the gate electrode, which overlays a thin layer of silicon dioxide just above the channel, repels holes and concentrates what few electrons there may be, thereby forming an* n-*type channel just under the gate and turning the transistor on. A* p-*channel MOS transistor is the mirror image of an* n-*channel device, with a* p-*type source and drain on an* n-*type substrate, and a negative gate voltage turns it on. [Drawing by Eleanor Warner]*

states, one or the other transistor is always off and current flow through the inverter is minimal. This is in contrast to an NMOS inverter, in which current does flow as long as the input is active.

The lower power of CMOS confers an extra blessing on circuit designers beyond portability and reliability. Some years ago, in an effort simultaneously to increase packing density, increase speed, and save power in NMOS circuits, designers settled on dynamic rather than static circuits. To illustrate the difference, consider a register (temporary storage site for data or instructions) in a microcomputer CPU.

In a static circuit the contents of the register remain fixed until new information arrives to be stored (unless the power goes out or the computer is turned off, then all information is lost). In a dynamic circuit, the contents of the register leak away and must be periodically refreshed.

The advantage of dynamic circuits is that they do not draw current between refreshings; the disadvantage is that refreshing requires additional circuitry including clocks to synchronize the updating with the operation of the register and thereby makes the designer's job harder. As illustrated by the inverter example, CMOS circuits naturally do not draw power and can easily be implemented as static circuits without the need for clocking circuitry.

Circuit complexity has serious effects in both design and physical implementation of a chip. With two types of transistors and consequently more total transistors required to accomplish a given circuit function, CMOS has historically been a more complex technology. But, as the number of transistors on a chip exceeds about 75,000, the burden on NMOS becomes so great that engineers are finding CMOS to be more attractive and less costly, David House, an Intel vice president, told those attending a forum on semiconductor memories that was held just before the solid-state conference.

At the same time, performance of CMOS integrated circuits has been closing the gap with that of NMOS. SRAM's are a case in point. As befits its name, a SRAM is a static circuit. Data is stored in a circuit called a flip-flop, which remains in whichever of two states (high and low output voltage) it is in until a new input signal arrives. CMOS has made its greatest inroads in SRAM's. As it happens, SRAM's are the high-speed memory technology, in contrast to DRAM's, which generally have larger storage capacities, cost less, but are slower.

At the solid-state conference, for example, five of the seven SRAM's presented were CMOS devices. The largest storage capacity was 256K bits in a chip from the Toshiba Corporation, Kawasaki, Japan. The Toshiba SRAM was the slowest of the lot with an access time of 46 nanoseconds. The fastest chip was a 64K SRAM from Hitachi, Ltd., Tokyo, with an access time of 20 nanoseconds, which equaled that of an NMOS 64K SRAM from the IBM Yorktown Heights Laboratory.

Significant as these advantages may be, heat generation will eventually limit how far even CMOS miniaturization can be taken because heat is also generated during the switching operation of all transistors. Under today's rule for miniaturization wherein the voltage applied to each transistor remains constant while device dimensions shrink linearly and the number of transistors on the chip increases quadratically, the total heat generated by switching varies inversely with the linear dimension. Sooner or later, the heat will exceed the limit.

At this point, engineers must resort to reducing the rate of increase in the number of transistors, slowing down the switching speed, lowering the operating voltage, or finding more efficient heat dissipating techniques. Chips in supercomputers, for example, which operate at the highest speeds and consequently dissipate the most heat, are seldom as densely packed with transistors as those in other machines and are sometimes refrigerated with liquid coolants.

Use of CMOS rather than NMOS staves off having to make such choices or at least makes them less critical in the next generation or two of integrated circuits. But, if miniaturization is to proceed apace, some problems unique to CMOS remain to be solved. One of these is how to pack the *n*-channel and *p*-channel MOS transistors more closely together without degrading their electrical characteristics.

At present, CMOS comes in two forms. If the substrate is *p*-type, the *n*-channel transistor can be fabricated directly, but a large lake (or well) of *n*-type silicon must be made on the substrate for the *p*-channel transistor. This is called *n*-well CMOS. Or, the substrate may be *n*-type with a *p*-type well. In either case, a wide barrier of insulating silicon dioxide separates the *n*- and *p*-channel transistors to keep them electrically isolated and thereby determines the minimum distance between them.

Even with the oxide isolation, the transistors remain in electrical communication under certain circumstances,

leading to a pathological condition called latch-up. Latch-up occurs because the CMOS structure results in the unintentional formation of a silicon controlled rectifier (SCR). SCR's are highly useful devices for switching large currents on and off in power supply controllers and the like. They comprise a series of four alternating *n*-type and *p*-type silicon segments. In *n*-well CMOS devices, for example, the parasitic SCR is formed by the source of the *n*-channel transistor (*n*-type), the *p*-type substrate, the *n*-well, and the source of the *p*-channel transistor (*p*-type).

Latch-up refers to the condition when a transient current accidentally turns the SCR on. This results in a large current that lasts until the condition is discovered and the SCR is turned off. Unfortunately, there is no automatic test for latch-up when the chip is in operation. In the worst case, the current can overheat the device and destroy it. The susceptibility to latch-up grows as transistor size shrinks.

Ways to deal with latch-up are hotly debated. At the International Electron Devices Meeting held in Washington, D.C., last December, Tadanori Yamaguchi and four co-workers from Tektronix, Inc. of Beaverton, Oregon, presented an elegant implementation of a previously proposed method that simultaneously allows tighter packing of *n*- and *p*-channel transistors and resists latch-up. In place of the chunk of silicon dioxide insulator, they etched a deep trench in the silicon substrate, lined it with silicon dioxide, and filled it with polycrystalline silicon. Ordinarily, the spacing between *n*- and *p*-channel transistors is from 5 to 9 micrometers. The trench is only 2 micrometers wide.

The Tektronix engineers also used a so-called twin-well structure that will probably be part of any solution to latch-up. The substrate is heavily *p*-type (has a high electrical conductivity). On top of the substrate, they epitaxially deposited a lightly *p*-type layer (with a low conductivity). This epitaxial layer serves as a *p*-well for the *n*-channel transistor. An *n*-well is also created in the epitaxial layer for the *p*-channel transistor. The deep trench extends down though the epitaxial layer (6 micrometers) to provide the electrical isolation, even though the surface separation between transistors is small (see figure).

Rather than burrowing down into the silicon, it is also possible to isolate closely spaced transistors by building upward, as demonstrated at the electron devices meeting by engineers from the NEC Corporation, Kawasaki, Japan. The team calls its technique selective epitaxial growth. The idea is to grow a 2-micrometer-thick layer of silicon dioxide on top of a (in their case) heavily *p*-type silicon substrate. The silicon dioxide has openings etched in it where the transistors are to go. Lightly *p*-type and *n*-type silicon is epitaxially deposited in the openings to act as *p*- or *n*-wells, according to whether an *n*- or *p*-channel transistor is to be formed. The silicon dioxide remains as the isolation material.

The ultimate in eliminating the spacing between transistors is to build one on top of the other. Like trench isolation, stacking transistors on top of one another is not a new idea, but it is getting more practical to do as fabrication technology improves. Accurately etching deep trenches with small dimensions is made possible by an increasingly popular dry etching technique called reactive ion etching rather than the old style wet chemical method. Ions from a plasma react with silicon to form volatile molecules that then depart, leaving a cavity.

Stacking transistors is accomplished by means of a laser technique known as laser recrystallization. Two groups reported on new developments in laser recrystallization at the electron devices meeting. The problem that laser recrystallization solves is how to create a layer of crystalline silicon for the upper transistor. Insulating materials that cover the first transistor, such as silicon dioxide, are amorphous, so silicon cannot be epitaxially deposited. However, it is possible to put down a layer of polycrystalline silicon. Irradiation with a focused, scanning laser beam melts the silicon, which then resolidifies as a single crystal if the laser power and scanning speed are correctly adjusted.

One approach taken by engineers at the Matsushita Electrical Industries Company, Ltd., in Osaka, Japan, started with a polycrystalline silicon "heat sink" to cover the bottom layer, which is made in the usual way on a silicon substrate. The heat sink is covered by silicon dioxide for electrical insulation, and a polycrystalline silicon island is deposited on the silicon dioxide. The island is laser recrystallized. The heat sink simultaneously protects the already completed lower level transistor from heating during laser recrystallization of the polycrystalline silicon in the upper level and provides a smooth substrate for the second layer.

The Matsushita group made simple CMOS test circuits (shift registers) in two ways. In one, conventional CMOS circuits resided independently in the top and bottom layers. In the other, which is the stacked configuration, the *p*-channel transistors of the CMOS pairs all lay in the bottom *n*-type substrate layer, while the *n*-channel transistors were all in the upper laser-recrystallized layer.

Researchers at Fujitsu, Ltd., in Kawasaki, Japan, took a different tack in laser recrystallization. Rather than irradiating the polycrystalline silicon directly, they covered the material with a "cap" that indirectly transferred the heat to the polycrystalline silicon. They claimed this method better promoted the formation of a single crystal. The engineers achieved a relatively smooth surface for the upper layer by means of phosphosilicate glass, which can be made to "flow" and thereby reduce rough edges. Silicon nitride insulator completely protects the upper transistor. In this way they made simple test circuits (ring oscillators) with the *n*-channel transistor of the CMOS pair lying over the *p*-channel transistor in an *n*-type substrate (see figure).

Despite their place at the center of today's high-tech world, integrated circuit manufacturers tend to the conservative side when considering new technologies.

When might any of these or other novel ideas appear in commercially available CMOS circuits? Despite their place at the center of today's high-tech world, integrated circuit manufacturers tend to the conservative side when considering new technologies. Rick Davies, a CMOS manager at Texas Instruments in Dallas, points out that manufacturers want to be sure that they can mass-produce chips reliably and inexpensively before they will change an already working fabrication process, despite the promise of laboratory devices. But it seems that, if CMOS is to lead microelectronics across the frontier toward chips with several million transistors squeezed onto them, some of these ideas will have to be implemented sooner or later.—**Arthur L. Robinson**