from *SIAM News*, Volume 31, Number 4

Designing and Producing the Microcircuits Of the Future: Part I

By James Case

Counting "captive" purchases by one division of a firm from another, worldwide semiconductor sales for 1995 exceeded $150 billion, with American firms claiming some 43% of the total and employing about 240,000 individuals. U.S. sales alone exceed the gross domestic products of many third-world nations. In order that the industry may continue to grow and prosper, along with the hardware, software, and other "downstream" industries, vast amounts of R&D remain to be performed.

Lest the U.S. cease to be a leader in the field, all sectors of the American semiconductor community have begun to cooperate in the production and updating of a document known as the National Technology Roadmap for Semiconductors, published and distributed by the Semiconductor Industry Association (SIA), headquartered in San Jose, California. Its purpose is to analyze and coordinate the R&D effort in such a way as to avoid undue duplication of effort, and to alert all concerned to potential barriers to progress.

The roadmap first appeared in 1992 as a pair of documents, which were combined and updated in 1994. Revised versions were prepared in 1997 and (presumably) will continue to be undertaken at three-year intervals. The 1994 edition focused primarily on the technology required for the design and production of complementary metal oxide semiconductor (CMOS) integrated circuits. "These products," the introduction notes, "constitute over 75% of the world semiconductor market and therefore determine mainstream technology. This mainstream provides the primary advancements for other semiconductor products, such as compound semiconductor, microwave, and linear devices. Therefore, the roadmap serves as a technology guide for all semiconductor products."

(A FET, by the way, is a "field effect transistor," while a MOSFET is a FET composed of metal oxide semiconductors. MOSFETs come in two flavors, known as n-channel MOSFETs and p-channel MOSFETs. A CMOS is a logic gate that incorporates various numbers of both in such a way that the combination dissipates no power in any "logic state." CMOS integrated circuits are expected to remain at the "cutting edge" of technology at least until the year 2010.)

The Remarkable Longevity of Moore's Law

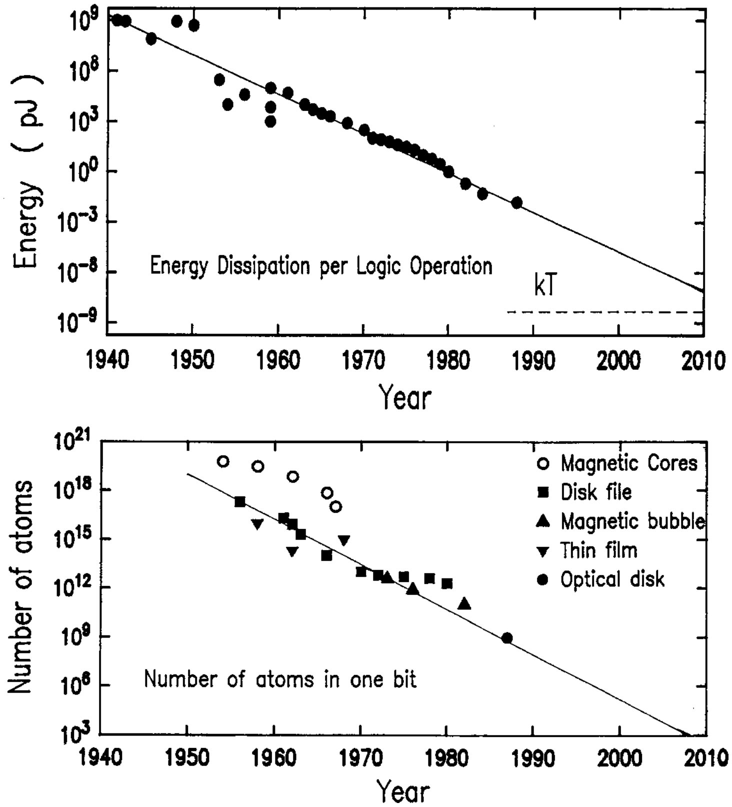
Historically, the first chips of each generation to reach market have been dynamic random access memory (DRAM) chips. Accordingly, profits from their sale have routinely been used to finance the development of other products of the same generation. In the foreseeable future, however, logic circuitry is expected to drive technological progress. The 1994 roadmap assumed (correctly, as it turned out) that the first DRAMs with minimum operational features measuring 0.35 mm would be shipped in 1995, and that they would be followed by chip features as small as 0.25 mm in 1998, 0.18 mm in 2001, 0.13 mm in 2004, 0.10 mm in 2007, and 0.07 mm in 2010!

These projections are in rough accord with "Moore's law," according to which the number of transistors that can be packed onto a single chip will double every 18 months. This rule of thumb has held more or less true since about 1965, and may even be pessimistic as it applies to the immediate future. A companion law suggests that the cost of a single bit of memory is halved every 18 months. If all goes as expected, observes Microsoft CEO Bill Gates, computing speeds will increase 10,000-fold by the year 2010—computations that now require an entire day would be completed in less than ten seconds. Already, he says, various laboratories have begun to experiment with "ballistic" transistors that have switching times on the order of a femtosecond (10^{-15} second), which is about 10 million times faster than the transistors currently in use.

Because computer engineers tend to use the terms "model" and "simulate" more or less interchangeably, the roadmap lists "simulation needs" for everything from microcircuits to manufacturing processes. The emphasis in this article (the first of two parts) is on the former. Somewhat speculative in nature, it deals with problems unlikely to become important for some time yet, while the second part will focus on more concrete and immediate concerns.

Although progress in microchip technology has conformed (more or less) with Moore's law since about 1965, that progress has been anything but smooth. On the contrary, the dreaded "show stopper"—the technical difficulty that would bring the entire circus to a standstill—has been predicted many times. But each time, the required breakthrough has been forthcoming.

Most recently, IBM and Motorola have announced techniques for etching copper instead of aluminum "wires" onto silicon wafers, thereby permitting conductor thickness to be reduced from the current 25 or 35 microns to 20 microns or less. Had this not been done, progress in computer design might have slowed perceptibly, since aluminum lacks the



Figures 1 (top) and 2 (bottom). *Miniaturization and integration since 1940.*

conductivity to be effective in such narrow strands. Earlier attempts with copper had failed because of the tendency of that and other good conductors to diffuse away into their silicon surroundings. The new IBM design succeeds where others have failed by insulating copper and silicon elements from one another.

At almost the same time, Intel announced the development of transistors that, by recognizing two intermediate states—partly off (1/3) and partly on (2/3)—between the 0/1 binary states, can store two bits of information instead of one at each memory address. As a result, existing production facilities will soon be using current technology to turn out memory chips with 32 or even 64 million transistors apiece, instead of the present 16 to 32. Moreover, some Intel sources believe that even more discriminating transistors will soon be possible! In short, after remaining remarkably accurate for more than 30 years, Moore's law may prove *unduly pessimistic* over the next decade or two!

The uncertain nature of research funding is not the only hazard to continued progress. Production costs are rising toward the point at which technically feasible improvements may become uneconomic. For instance, the costs of building semiconductor factories seem to be doubling about every three years, or at half the rate of Moore's law. Intel is spending $1.1 billion on its new plant in Hillsborough, Oregon, and $1.3 billion on another in Chandler, Arizona. Samsung and Siemens are said to be building plants that will cost $1.5 billion to finish, and Motorola is planning a facility likely to cost $2.4 billion. This means that factories built in 2010 will cost about 20 billion 1997 dollars, or 30 billion future dollars, assuming a 3% average rate of inflation.

It is less than obvious that anticipated industry profits will justify such investments. The profits in question seem to run in cycles of about six years, in which an initial period of high (but declining) investment and low (but increasing) profit precedes a period of high but declining profit and low but increasing investment. Intel, for instance, experienced three [1] such cycles between 1971 and 1994. Even at Intel, however, each successive cycle seems to have lagged behind its predecessor in profitability, indicating that reinvestable profits are in decline, while investment requirements are growing. It is by no means an encouraging trend.

Beyond Transistor Counts

Measures of circuit miniaturization and integration other than transistor count are also improving at exponential rates. Because denser chips are also faster chips, the number of floating-point operations achievable per second is roughly quadrupling every 18 months. Figures 1 and 2 reveal that the amount of energy dissipated per logic operation, and the number of atoms required to store a single bit of information, are also improving exponentially.

It is said that memory chips now under development will have capacities comparable to those of current magnetic disks. If and when such chips become operational, it will be possible to build computers without any moving parts at all, and to mount an entire computer on a single chip. Such computers will, of course, operate significantly faster than any of the current ones. Figure 2, which records the number of atoms required to store a single bit of information, is particularly enlightening. To continue the current trend even to the year 2010, it will be necessary to store single bits of information with only a few hundred atoms each.

According to Bill Gates, the ultimate memory device would be a "single electron transistor," within which a bit of information is indicated by the presence or absence of a single electron. A somewhat more modest objective is suggested by Eric Drexler, the prophet of nanotechnology, who has described an entire atomic-scale computer in which bits of information are represented by tiny rods, consisting of only a few atoms each, which may (like a turtle's head) be in either the "protruding" or the "withdrawn" state.

Whereas the existence of atomic- and even subatomic-scale computer components does not contradict the laws of physics, as currently understood, nobody has even a clue as to how to build them. Currently available chips, for all their miniaturization, are still "macroscopic" devices, and the circuits etched on them still behave more or less like the circuits of ordinary experience, composed of wires, batteries, switches, and other off-the-shelf elements. Such circuits can be, and routinely are, analyzed by techniques not unlike the familiar "circuit analysis" of high school and college physics. Those methods are moderately well understood, and will be discussed in the second part of this article.

On the Mesoscopic Scale

Between the "macro" domain of everyday experience and the atomic or even subatomic domain of nanotechnology, however, lies what is coming to be known as the "mesoscopic" domain of physical phenomena. Although science is only beginning to understand what goes on in this recently identified domain, and the relevant mathematics has yet to be identified, semiconductor technology is already poised to enter.

The mesoscopic domain differs from the familiar macroscopic one in that the dual (particle/wave) nature of electrons cannot be ignored. Whereas electric currents in wires behave more or less in accord with Ohm's law in existing integrated circuits, they no longer do so in significantly smaller devices. There, the process seems to be better described in terms of the Feynman path formulation of quantum mechanics, in which the "amplitude" for the propagation of a particle between two points is the sum of the classical amplitudes over all possible paths connecting the two points, and the (real) probability of such propagation is the modulus of that (complex) amplitude.

For a free particle, the dominant paths seem to lie within a (straight) cylindrical tube connecting the endpoints, whose cross-sectional radius is on the order of the particle's de Broglie wavelength, l. Such particles move more or less freely through regular atomic lattices but are randomly scattered when they encounter the occasional impurities that turn silicon into a semiconductor. Electrons thus appear to execute random walks from one scattering site to another.

One of the surprises of the new domain has to do with the quantum mechanical nature of electromagnetic phenomena. Maxwell's equations, held to govern all such phenomena, are traditionally expressed as a system of 12 first-order PDEs involving the six components of the electric field (E) and magnetic field (B) in a given region of space. While six of the equations assert the existence of a scalar potential f and a vector potential V, such that $E = \nabla\backslash\phi$ and $B = \nabla \times V$, the remaining six express the natural laws according to which currents and charges interact with E and B.

The equations can be rewritten in terms of the four-vector $A = \phi \oplus V$, which is known as the electromagnetic potential. The original significance of A was merely that, when so rewritten, Maxwell's equations reduce to a system of only four second-order PDEs involving only four unknown functions of position and time. All that changed in 1959, however, when physicists Yakir Aharonov and David Bohm deduced (directly from the Schrödinger equation) that a magnet shielded in such a way as to exert no force on another nearby magnet would still alter the quantum mechanical phase (the θ in $\sin(\omega t + \theta)$) of an itinerant electron wave! The quantity A, introduced merely for mathematical convenience, is in fact a more fundamental physical quantity than either the E or the B field, and penetrates shields through which the others cannot pass. Their prediction was quickly confirmed as it applies to electrons traveling in a vacuum and, in the last decade, to electrons traveling through very thin conducting wires at low temperature.

Magnetoresistance

An important consequence of the Aharonov–Bohm effect is that thin cold conducting wires possess no property analogous to electrical resistance. The closest of kin seems to be a quantity known as "magnetoresistance," which varies dramatically with applied electromagnetic potential. An experiment conducted at IBM in 1985 demonstrated as much.

In the IBM experiment, a gold ring with inside and outside diameters of 0.78 micron and 0.86 micron, respectively, was etched on a silicon wafer. Input and output leads positioned on opposite sides of the ring allowed current to pass from one side to the other in either of two ways. The current strength and voltage drop across the ring were both measured.

Ordinarily, the ratio R between them would be a constant more or less independent of current, voltage, and ambient conditions, and would be called resistance. Here, however, R was observed to vary substantially with the strength of the vector potential generated by a magnetically shielded solenoid passing through the center of the ring, perpendicular to the plane of the chip. Outside the shield of the solenoid, the vector potential would be expected to curl around the ring, and to increase the phases of electron waves traversing one side of the ring, while decreasing the phase of those traversing the other side. If the induced phase difference is an even multiple of π, the waves fully reinforce one another and double the observed current. But if the phase difference is an odd multiple of π, the waves cancel one another and contribute nothing to the observed current. Intermediate phase differences should produce intermediate effects. The observed current, then, should oscillate about its average value as the strength of the vector potential is increased. These expectations were confirmed by observed variations.

In the original experiment, the magnitude of the oscillations was a mere tenth of 1% of the average (over several oscillations) resistance, but in subsequent experiments, with different materials, the fraction has been as large as 50%. The magnitude and frequency of the oscillations compare favorably with theoretical predictions expressed in terms of Planck's constant, $\hbar$, and the charge e of an electron. Indeed, oscillations in the resistance of various rings are independent of their average resistances and approximate $\hbar/e^2$. The shape of the smoothed curve, about which the observed resistances appear to oscillate, is not universal—it depends on the number and location of scattering impurities and changes from one ring to another.

The physics of mesoscopic systems, which are large on the atomic scale but small enough that the quantum mechanical coherence of traveling electron waves cannot be ignored, remains in its infancy. The fluctuation of electrical resistance in response to changing electromagnetic potential, chemical potential, and other seemingly relevant factors is but one of many new phenomena to be investigated. Although few others have been studied in comparable detail, the quest for smaller and faster computer chips appears to lead directly into this new and largely uncharted territory. The construction of working prototypes is unlikely to contribute significantly to the development of the devices in question—the design process seems certain to rely almost exclusively on mathematical modeling and simulation.

James Case is an independent consultant who lives in Baltimore, Maryland.