The Application of Transistor Technology to Computers

ROBERT A. HENLE, FELLOW, IEEE, IRVING T. HO, WILLIAM S. JOHNSON, MEMBER, IEEE, W. DAVID PRICER, SENIOR MEMBER, IEEE, AND JAMES L. WALSH, SENIOR MEMBER, IEEE

Abstract—**The density of integrated circuits has improved in the last 15 years by 4 to 5 orders of magnitude. This paper discusses four primary computer technology areas that have helped bring about this advance and which have significantly changed the nature of computer components: FET logic, FET memory, bipolar logic, and bipolar memory. Projections are made for future trends in these technologies.**

Index Terms—**Bipolar logic, bipolar memory, FET logic, FET memory, integrated circuits, semiconductor circuits.**

THIS paper presents an overview of what has happened to the application of monolithic circuits and devices in the decade that monolithic circuits have been available from manufacturers for computer use. The first monolithic products were preceded by a development period of four to six years, judging by the 1961/1962 Records of the International Solid-State Circuits Conference. The impetus for this development was the tremendous potential of silicon planar technology, namely, the fabrication of many circuits simultaneously on a single silicon substrate. These developments have had a dramatic impact on computer logic and memory technology. The vacuum tube logic in the IBM 704 (1955) had a circuit speed of 200 ns and a memory of 18K bytes, with a cycle of 12 μs. The technology utilized in the 370/168 (1972) has a circuit speed of 4 ns with a memory of 8390K bytes of FET storage, with a cycle time of 160 ns. The throughput per dollar had improved by about two orders of magnitude in this period. The complexity of integrated circuits has approximately doubled every year since their introduction [1].

In this paper we describe progress in four primary technology areas—FET logic, FET memory, bipolar logic, and bipolar memory—and make some projections for what the future might hold in each area.

FET Logic

Field-effect transistors (FET's) operate on a simple and easily understood principle: application of a voltage to a capacitively coupled electrode creates an electric field that alters the number of charge carriers in a semiconductor and thereby modulates its conductivity. The capacitive electrode has been realized in practice as a p-n junction (JFET: junction FET), a Schottky barrier (MESFET: metal-silicon FET), or a plate separated from the semiconductor by an insulating layer (IGFET: insulated gate FET). This last is usually realized as a metal plate insulated from the semiconductor by silicon dioxide (MOSFET: metal-oxide-semiconductor FET). Recent advances have replaced the metal in many cases by polycrystalline silicon; the term MOSFET is generally applied loosely to these silicon-gate FET's as well.

The basic MOSFET physical structure is shown in Fig. 1. The source and drain are n-doped regions in a p-type substrate. The channel is normally p-type and prevents conduction from source to drain. The application of a positive bias to the gate, however, attracts electrons into the channel, making the channel appear temporarily n-doped and allowing conduction from source to drain. Such a device is an n-channel enhancement-mode MOSFET. Devices are also made on n-type substrates with p-type source and drain. These devices require a negative gate bias to attract holes into the channel to initiate conduction and are known as p-channel enhancement-mode MOSFET's. Because electrons have higher mobility than holes, n-channel devices have superior performance to p-channel devices and are preferred in (non-CMOS) circuits.

In modern devices, the doping in the channel region is often controlled by ion-implantation. This advance allows very accurate control of threshold voltage and, in particular, allows the conversion of the channel region to the same type doping as the source and drain. Such devices are in a conducting state when the gate is at zero bias. They require nonzero gate bias to deplete the channel and turn them off (negative bias for n-channel and positive bias for p-channel). These are depletion-mode devices.

Only majority carriers are important in the operation of FET's, unlike bipolars. Besides making the devices conceptually simpler, this difference also simplifies device modeling and operation, which, in turn, simplifies design. Also because of this dependence on majority carriers, the intrinsic speed of the FET is extremely high, being limited by the channel transit-time and unaffected by delays due to generation and recombination of minority carriers.

Fig. 2 is a plot of logic delay per stage as a function of channel length. At the bottom of the figure the ideal delay of a single MOSFET is shown under the simple approximation that the electric field is uniform along the channel. Basically the transit delay varies as the square of the channel length and inversely as the applied voltage, unless the electric field exceeds 10^4 V/cm, in which case the carrier

Manuscript received May 19, 1976; revised August 3, 1976.

R. A. Henle, I. T. Ho, and J. L. Walsh are with the IBM System Products Division, East Fishkill, Hopewell Junction, NY 12533.

W. S. Johnson was with the IBM System Products Division, East Fishkill, Hopewell Junction, NY 12533. He is now with Motorola, Inc., Phoenix, AZ.

W. D. Pricer is with the IBM Systems Products Division, Burlington, VT.

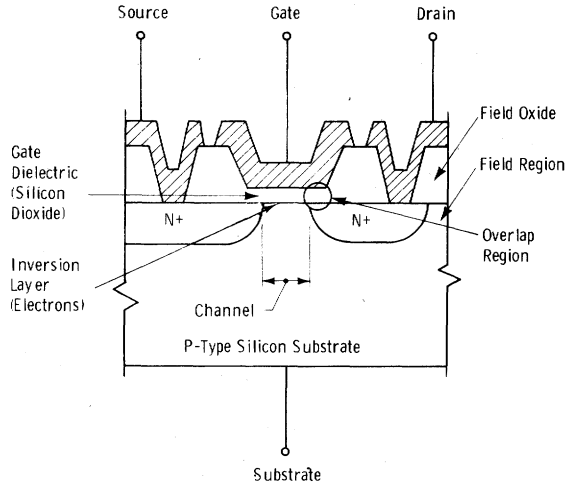


Fig. 1. Physical cross section of a metal-gate n-channel MOSFET.

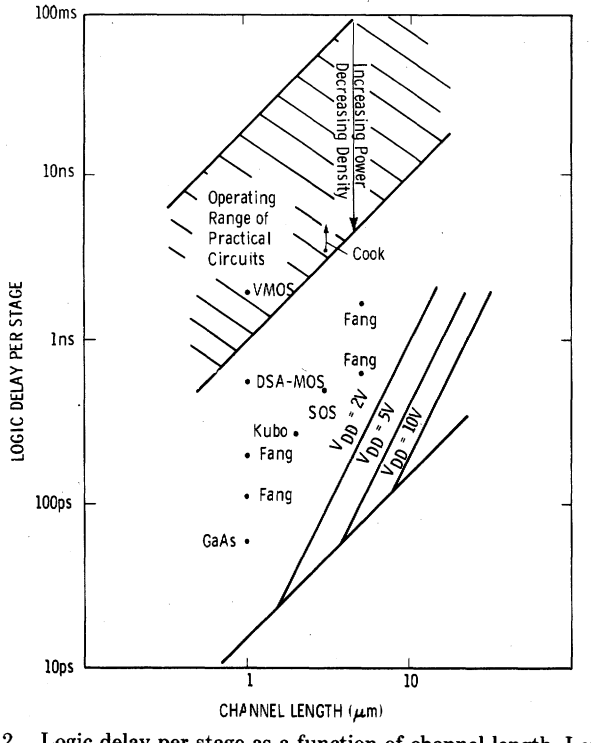


Fig. 2. Logic delay per stage as a function of channel length. Lower curves are theoretical limits due to transit-time delay in a single device. Shaded area represents performance of fully loaded practical circuits. Operating points within this area are determined by power and density tradeoffs. All points except "Cook" are experimental measurements on lightly loaded ring oscillators; "Cook" is from a theoretical study of fully loaded circuits [25]. The other references are GaAs [49], Fang [39], DSA-MOS [43], VMOS [44], Kubo [52], and SOS [45].

velocity saturates [2] at 6.5×10^6 cm/s. This figure indicates essentially that at today's channel lengths, around 5 μm, the intrinsic device delay is 100 ps; at channel lengths around 1 μm or below, a reasonable projected value for five to ten years hence, the intrinsic delay is about 10 ps. These delays compare well with bipolar delays. However, unlike the case with bipolars, the actual circuit delays achieved in fully loaded logic nets are about 100 times slower than the intrinsic delay. Thus, roughly 10 ns logic delays are being achieved with today's MOSFET logic technology and, barring significant circuit improvements, we can expect 1–2 ns logic delays to be achieved with the 1 μm channels of the future. The approximate operating region for practical circuits is shown by the shaded area in the upper part of Fig. 2. The data points in Fig. 2 are experimental measurements on ring oscillators and will be discussed later.

While the basic concept of MOSFET can be traced to the 1930's [3], the device proved impractical to build for about 25 years because natural semiconductor surfaces contain large densities of electronic states that trap the carriers attracted by the gate and hold them immobile [4].

In 1959, Atalla *et al.* [5] reported that these traps could be controlled by stabilizing silicon surfaces with thermally grown silicon dioxide, and in 1960, Kahng and Atalla [6] reported on the use of a thermally oxidized silicon structure to build a field-effect transistor. Investigations into this new device were pursued vigorously over the next several years [7]–[10].

The basic FET logic gate seems to trace from the work of Wallmark and Marcus [11] who reported in 1959 on a basic gate composed of a JFET driver and a floating-gate JFET pull-up device. Five years after these breakthroughs about 25 different MOSFET logic parts were commercially available [12].

MOSFET Logic Circuits

The three general classes of MOSFET logic are static, complementary, and dynamic (multiphase) logic. They are illustrated schematically in Fig. 3, which shows a basic two-input NOR gate in each logic type. Static logic [Fig. 3(a)], uses parallel enhancement-mode input devices (drivers) and a load device, which can take several forms. The basic operation of this circuit is as follows. If either input, A or B, has voltage applied, then the output node is connected to ground; if neither A nor B has voltage applied, then the load device charges the output node to a high voltage. The simplest form for the load device is an enhancement-mode device with its gate tied to V_{DD} (saturated load) [12]. The disadvantage of this configuration is that the device turns itself off as the output node rises in voltage. In fact, the node can rise no higher than $V_{DD}-V_T$, where V_T is the threshold voltage of the device. An improved load device can be realized by connecting the gate to a positive power supply of higher voltage than V_{DD} (linear load) [13], [14]. This keeps the load device conducting at all times and allows the output node to charge all the way to V_{DD}. However, this form requires a second power supply and additional interconnections.

Recent processing advances, such as ion implantation, have made it possible to adjust different devices on the chip to different threshold voltages. The problem of the load device shutting itself off can be overcome by this approach: by adjusting the threshold so that it conducts even at zero gate bias; in other words, by making it depletion-mode. With this device it is best to tie the gate to the source. In this way the device acts nearly as a constant

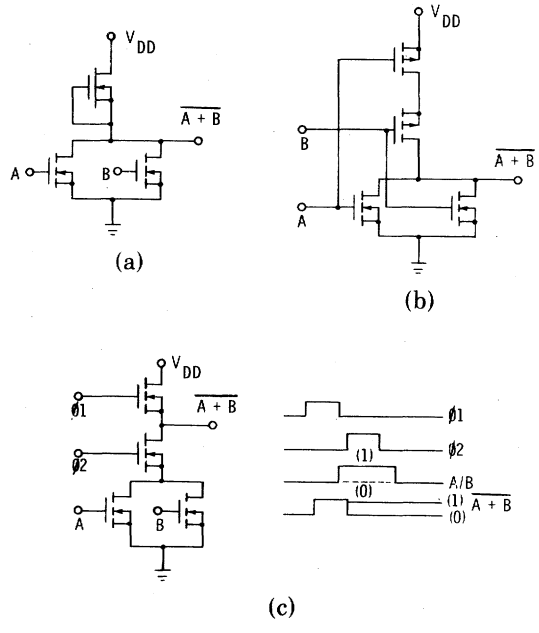


Fig. 3. Schematics of three basic classes of MOSFET circuits. (a) Static, (b) complementary, and (c) dynamic (four-phase) logic. (c) includes a timing diagram. A device with a dashed bar is enhancement-mode; with a solid bar, depletion-mode. An arrow toward the bar represents n-channel; away from the bar, p-channel.

current source and allows improved performance at the same dc power dissipation as enhancement-mode loads [15], [16]. The circuit in Fig. 3(a) shows the depletion-load form.

The best load device would be one that shuts itself off when the output node is down, because it is in this state that the NOR gate in Fig. 3(a) dissipates dc power. This can be accomplished by switching the load as well as the driver with the incoming signal, as shown in Fig. 3(b). Since the load must be switched out of phase with the driver, however, it must be switched on by the opposite polarity voltage step. If the drivers are n-channel, then the loads must be p-channel; in other words, complementary types of devices are needed, leading to the name complementary MOS, or CMOS.

The complementary NOR gate [17]–[19] in Fig. 3(b) operates as follows: if A is positive, then the lower n-channel device is on, and the upper p-channel device is off. (Note that p-channel devices effectively have their sources at their upper terminal and that n-channel devices have their sources at their lower terminal.) Similarly for input *B*. It is readily apparent that the output is only tied to V_{DD} and is isolated from ground when both *A* and *B* are low. Note also that no combination of inputs allows any dc current to flow. Thus CMOS logic circuits dissipate no dc power. This is their principal advantage, the price of which is increased fabrication complexity.

Another method of avoiding dc power is to switch the load with a clock instead of the incoming signal. Dynamic (multiphase) logic accomplishes this in one of several ways. Luscher and Hofbauer [20] described a method that uses capacitor pull-up devices. Most methods utilize MOSFET load devices with clocks switching their gates. While four-phase circuits [21], [22] are the most common, circuits with two to six phases [23], [24] have been described.

Quantitative comparison of the different MOSFET logic circuit approaches is complicated by the large number of differing design criteria for each. Such comparisons must be made for a given application and loading condition.

Power comparisons are complicated by the fact that power dissipation in CMOS and dynamic circuits varies linearly with operating frequency, whereas static logic dissipation depends only weakly on operating frequency. Generally at low frequencies, in applications where delay is not a primary concern, CMOS and dynamic MOSFET circuits dissipate far lower power than static logic for the same performance. However, in achieving fast delays and high operating frequencies ac power increases, and the relative disadvantage of the static circuit in dissipating dc power is diminished. Furthermore, since static logic uses fewer devices than CMOS and does not have clock lines to drive, as dynamic circuits do, its ac power dissipation is generally less for a given performance than that of other types. A useful study comparing these logic circuits was performed by Cook *et al.* [25].

MOSFET Process Development

Improvements in MOSFET devices and processing techniques generally aid all of the circuit families discussed. Two processing advancements have been fundamental to the evolution of MOSFET's: processing cleanliness and photolithographic resolution. These have occurred steadily over the years rather than resulting from a single innovation.

Processing cleanliness is difficult to quantify but is of unquestionable importance. If one realizes that a density of foreign atoms corresponding to 1/10 000 of a monolayer at the $Si\text{-}SiO_2$ interface shifts the threshold voltage more than 0.6 V (assuming a 1000 Å gate oxide), it is easy to understand that meticulous cleanliness is indispensible in MOSFET processing.

Reduction in lateral dimensions is the simplest way to improve density, power dissipation, and performance, all at the same time. To improve resolution by a factor of two allows both length and width of a MOSFET device to be reduced by the same factor if parasitic effects, which have been relatively unimportant in the past, are ignored. Such an improved device conducts the same current at the same voltages as the original device, but represents only one quarter the capacitance and occupies only one quarter the area. If we assume that the increased density allows wire lengths and widths, as well as diffusion sizes, to be scaled similarly, then the total load capacitance is quartered. Thus performance as well as density improves by a factor of four. In this example, power dissipation is held constant, but power and performance can always be traded off by adjusting device widths. Lithography improvements have been very important to MOSFET device evolution. Unfortunately, however, parasitic effects, which are magnified at small dimensions, preclude this simple form of scaling

from continuing indefinitely. This problem will be discussed in the next section.

Many MOSFET improvements have come about, not by continual process improvement, but by innovation or development that can be fixed fairly well in time. Several that stand out are thermal-oxide stabilization, phosphosilicate-glass stabilization, silicon-on-sapphire, self-registered gates, silicon gate, and ion-implantation doping.

Control of interface states by thermal oxidation of the silicon surface [5], as discussed earlier, was the breakthrough that made MOSFET device development practical. Another stabilization problem affecting MOSFET's is the movement, under an electric field, of charged ions within the oxide dielectric. These ions result from contamination, mostly by sodium, during processing. Their movement causes the threshold voltage to shift with time and results in undesirably high thresholds on p-channel devices and undesirably low thresholds on n-channel. In fact, because of the work function potential of aluminum relative to silicon, n-channel enhancement devices normally have threshold voltages near 0 V. A small amount of mobile ion drift can cause these devices to become depletion-mode. Thus reliable n-channel enhancement-mode devices were originally very difficult to make and were avoided by manufacturers for some time after their discovery despite their superior performance capabilities.

Although cleanliness in processing is clearly a solution to the mobile ion problem, the extreme sensitivity of MOSFET devices to even minute amounts of contamination makes this answer difficult to achieve. A very practical solution to the problem turned out to be to cover the gate oxide with a thin layer (about 100 Å) of deposited phosphosilicate glass [26], [27]. This material is a getter for alkali ions; that is, it attracts the ions and holds them immobile. As long as the ions are held at the outside surface of the gate dielectric, near the metal, they have no effect on the electrical properties of the device. This process improvement helped p-channel devices but, most important, it made n-channel devices practical.

A development that promised a substantial improvement in performance was the realization of MOSFET devices in single-crystal silicon films grown on insulating sapphire crystals (SOS) [28]. This technique allows almost perfect isolation of devices and nearly eliminates junction and wiring capacitance. SOS combines well with CMOS and has been successfully produced commercially. Nevertheless, high substrate cost, lower silicon quality, and process difficulties, as well as continual improvement in silicon substrate MOSFET devices, have kept this approach from becoming dominant.

Self-registered gates must rank as one of the most important MOSFET innovations. In the original approach to MOSFET device fabrication the source/drain was diffused, an oxide was grown, and then the gate area was defined by a separate lithography step (see Fig. 1). Since the designer had to insure that the thin oxide gate region overlapped the source and drain, with substantial misalignment of the gate mask during lithography, he was forced to provide a large overlap region. This region results in a large capacitance between the gate electrode and the source and drain, which is parasitic to logic circuits and deteriorates their performance.

The solution to this problem was reported by Bower and Dill [29], [30], who discussed the then new technique of ion-implantation doping. Their structure was similar to that shown in Fig. 4. Basically the gate electrode (aluminum in their case) was applied *before* the source/drain regions were introduced. Then, using the gate as a mask, the source and drain were ion-implanted. Thus registration was automatic and the overlap region was reduced to a minimum.

The drawback of aluminum in this application is that subsequent heat treatments are limited to about 500°C, and the damage inevitably caused by implantation cannot be completely annealed. However, the development of polycrystalline silicon as a gate material solved this problem nicely. Since polysilicon can withstand diffusion temperatures, self-registration could be combined with diffused source/drains [31]–[33]. (In Fig. 4 the thin oxide outside the gate must be etched before diffusion and then regrown afterwards.) Yet another successful method of self-registration utilizes silicon nitride as a diffusion mask to define the source and drain, and then as an oxidation mask to define the gate region, in subsequent process steps [34].

The silicon gate method described above has become an industry standard because it offers many benefits in addition to self-registration: 1) if p-type polysilicon is used, its work function is such as to lower p-channel thresholds and increase n-channel thresholds; 2) because the polysilicon can be deposited immediately after gate oxide growth, contamination of the gate oxide is minimized; 3) by oxidizing the polysilicon, aluminum metallization can pass over it without making connection, thereby providing an extra level of interconnection.

Probably the greatest advance of all for MOSFET devices has been ion implantation. In this technique ionized dopant atoms, accelerated in a vacuum, impinge on the semiconductor and bury themselves beneath the surface. The fundamental advantage of this technique is that it gives control of both dopant concentrations and depths to electrical rather than chemical sources.

The use of implantation for self-registration was discussed above; of course, the technique can be combined with silicon gate as well as aluminum gate. Other applications of implantation to MOSFET have proven more important, however. These are all related to the following fundamental properties of the technique: accurate control of extremely small quantities of dopant (to a resolution of about one millionth of an atomic monolayer), attainment of extremely shallow doping depths (less than 1000 Å), ability to place dopant into the silicon through intervening oxide layers, and high uniformity and reproducibility (approximately 1 percent).

Besides self-registration ion implantation has been used

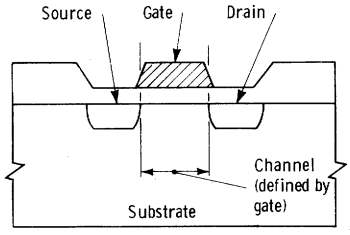


Fig. 4. Physical cross section of a self-registered gate structure. Gate could be metal or polysilicon. The source/drain could be formed by ion implantation or diffusion.

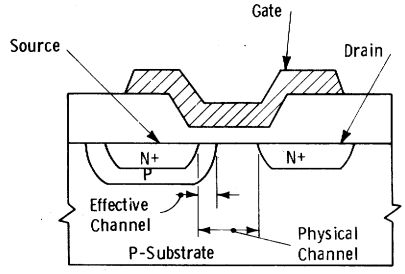


Fig. 5. Physical cross section of a DMOS structure.

with great success to adjust threshold voltage [35], [36], prevent punch-through in short channel devices [37], produce depletion-mode devices [15], [16], [36], and improve field threshold characteristics [38].

The use of these techniques improves performance significantly because virtually every component of parasitic capacitance can be reduced. Self-registration reduces overlap capacitance. n-channel devices become as easy to fabricate as p-channel because threshold voltages can be set at will. By starting with higher resistivity substrate material, junction capacitance is reduced. (This approach must be combined with both threshold adjustment and field doping to control device characteristics and parasitics properly.) Finally, depletion-mode load devices are easily fabricated.

Fang and Rupprecht [39] combined all of these techniques with 1 μm lithography to demonstrate the fastest delays yet obtained for silicon NOR gates, 115 ps. As can be seen in Fig. 2, this result, obtained from an 11-stage ring oscillator, would unfortunately deteriorate by about a factor of ten in a practical loaded circuit. Nevertheless, even nanosecond speeds were unobtainable with silicon substrates before ion implantation.

Channel lengths of 1 μm are beyond the reach of today's lithography, except in the research laboratory. A device that obtains equivalent channel lengths from the point of view of electrical characteristics, without the necessity of such high-resolution lithography, is DMOS (double-diffused MOS) [40]–[42]. This structure, shown in Fig. 5, has a lightly doped p-type substrate, and a more heavily p-doped region diffused through the same hole as the source. Although the effective channel length depends upon bias conditions, it is equal to the width of the p-region over a useful range. Ohta *et al.* [43] fabricated ring oscillators by combining a silicon-gate self-aligned form of DMOS (DSA-MOS) [41] device with normal depletion MOSFET load devices and achieved 0.56 ns delays. This result, plotted in Fig. 2, does not match that of Fang and Rupprecht because the gate capacitance is determined by the actual channel length (~3 μm), not the electrical channel length (~1 μm).

Another form of DMOS has its channel formed on the sloping wall of a v-groove in an epitaxial layer (VMOS) [44]. This form has the advantage of extremely large channel width in a small device area. A point is plotted in Fig. 2 for a VMOS ring oscillator, but a comparison with the other 1 μm points is unfair since this device was fabricated with 10 μm lithography. A reduction to 3 μm lithography would reduce gate capacitance by nearly a factor of 10 and put this point well within the subnanosecond region.

High-speed ring oscillators have also been realized in SOS [45]. The data of Pomper and Tihanyi in Fig. 2 were taken from ion-implanted, self-registered devices with 3 μm channel lengths.

Gallium arsenide exhibits electron mobilities around five times those in silicon, which could ultimately result in devices with five times the performance. Schottky-barrier depletion-mode FET's (MESFET's) have been successfully fabricated in GaAs [46]–[49], and logic delays as fast as 60 ps have been measured for 1 μm channel lengths [48]. Logic circuits with only depletion-mode transistors require a level-shifting stage, which consumes power and creates delay. It is unfortunate that wiring capacitance is no better in GaAs than in silicon. Thus, achievable delays in practical integrated circuits may never be much better than in silicon.

Future Prospects

In the future, MOSFET's will probably become totally ion-implanted and will certainly benefit from improved lithography expected to come about from projection printing and use of electron-beams [50]. However, scaling devices to line widths smaller than about 3 μm is complex and requires changes in vertical dimensions and doping concentrations commensurate with the lateral dimensions and doping concentrations commensurate with the lateral reduction [51]. Improvements in circuits, such as the voltage-compensating circuits [52], [53] recently discussed and others that use implanted devices more imaginatively, are expected to significantly improve power-delay products even for today's devices.

Logic delays in the low nanosecond range should be easily achieved over the next few years. These will be combined with circuit densities extending from today's 1000–2000 gates/chip to 10 000. The higher chip power levels, however, will require better packaging methods. Logic delays in the subnanosecond region, for fully loaded logic circuits, will be difficult to attain.

FET Memories

While the development of bipolar memories initially may have led that of FET memories, it soon became evi-

dent that FET memory arrays possessed a clear advantage in density and processing simplicity over bipolar equivalents. During the late 1960's, a general debate arose concerning the technology interface between FET memory arrays and their supporting accessing circuitry. The use of bipolar support circuits could greatly enhance memory performance, frequently at little increase in cost. This influenced designers to make as many support circuits as practical in the faster bipolar technology.

Substantial effort was expended to develop denser interconnecting techniques rather than resort to a compound process capable of both bipolar and FET circuit technology. By the early 1970's, it had become clear that advances in array density would outpace those in interconnection technology. In the resulting compromise, support circuitry packaged on-chip became FET technology and almost all off-array-chip circuitry became bipolar technology. Common array/support circuit memory chips proved a mixed blessing. On one hand, system performance was necessarily slower and on-chip support circuitry frequently had to be replicated on a chip-to-chip basis. Typically, a fully decoded integrated memory chip is only 50 per cent array. The rest is kerf, interconnection pads, and support circuitry. On the other hand, incorporating (and replicating) some of the decode on chip did decrease the interconnection cost. Perhaps more important, it decreased the numerical size of the interconnection interface, and with it the attendant reliability exposure. This advantage became more significant when it was realized that such highly functional memory chips were particularly amenable to error detecting and correcting techniques [54]. These techniques coupled with the highly functional if somewhat redundant design of the memory chips allowed semiconductor memory to approach the reliability achievements of magnetic core memories. This was a significant and perhaps necessary step in their widespread acceptance.

From earliest inception, it was clear that semiconductor memories would be able to out-perform magnetic core memories. The point in their development when they would first become cost competitive was less distinct, in part because each technology's costs varied differently with size.

In general, magnetic core memories could be characterized by a large initial cost and small incremental cost [55] (see Fig. 6), whereas semiconductor memories displayed a very small initial cost but a comparatively large incremental cost. In general, the point of cost competiveness for many applications appears to have been reached at a level of integration somewhere between 1024 and 4096 bits/chip.

Practical and commercially viable random-access cells were achieved through better device control, and sometimes with the inclusion of complementary devices [56]. The particularly cost-sensitive nature of memory applications has favored static memories of a single conductivity type (n-channel or p-channel devices). Improved costs came from the adaptation of certain circuit techniques originally developed in shift register applications. Fig. 7 illustrates a family of circuits using capacitance storage and ratioless devices. Transistor action is used to steer charge to the storage node and to amplify and nondestructively sense that charge. The introduction of dynamic random-access memory cells, while economically attractive, did incur at least one operational penalty. Some system time had to be set aside on a regular basis to replenish (refresh) any lost charge. In applications where this overhead time was limited, a more complex circuit—still using charge storage but with internal refresh—could be used (see Fig. 8). This circuit has the additional attraction of responding faster than three-device cells in arrays of comparable size. Other circuit techniques such as charge pumps were introduced to counter the refresh penalty [57].

These memory circuits, simple as they are, were antedated by one simpler still. In the mid-1960's, Dennard [58] proposed a bank of thin dielectric capacitors accessed by a matrix of FET bilateral switches. The lack of MOSFET parameter control and suitable sensing techniques hindered the early development of arrays of this type. By 1973, however, the gradual stabilization of process control had allowed the fabrication of practical arrays up to 4000- and 8000-bit chips [59], [60] and had encouraged some manufacturers to produce them in the faster n-channel technology. At present, arrays of these so-called one-device cells appear to have become the mainstream of product development [61].

Concurrently, emphasis has shifted somewhat from cell development to support circuits and particularly sense amplifier development. Cell structure appears to have been depleted of transistors down to an irreducible minimum. The leverage for performance improvement, and density improvement, to some degree, now rests with the support and sense amplifier circuits.

Semiconductor memories, unlike their magnetic core predecessors, normally do not retain their stored information once power is removed. This deficiency has not hindered their broad acceptance in most applications. Nonvolatility [62] is available at increased process complexity and some technology risk through the use of charge trapping in various gate dielectrics. To date, such approaches have met with limited acceptance.

Future Prospects

Many excellent articles on the future of semiconductor technology in general and monolithic memory in particular have been published [63]–[66]. The increased productivity of FET memory during the past decade is phenomenal by any measure. Memory chip densities have typically doubled every year. Clearly, such rapid advances cannot go on forever. The work of Hoeneisen and Mead [65] suggests that we are perhaps an order of magnitude in linear dimensions away from some ultimate transistor. The historic growth rate (see Fig. 9) is starting to show some perceptible but not yet decisive deceleration. Based on these two observations alone, we might hasten to conclude that we

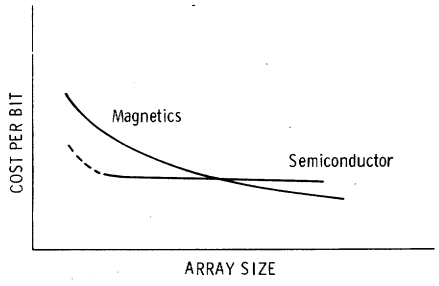


Fig. 6. Cost comparison of core versus FET memory.

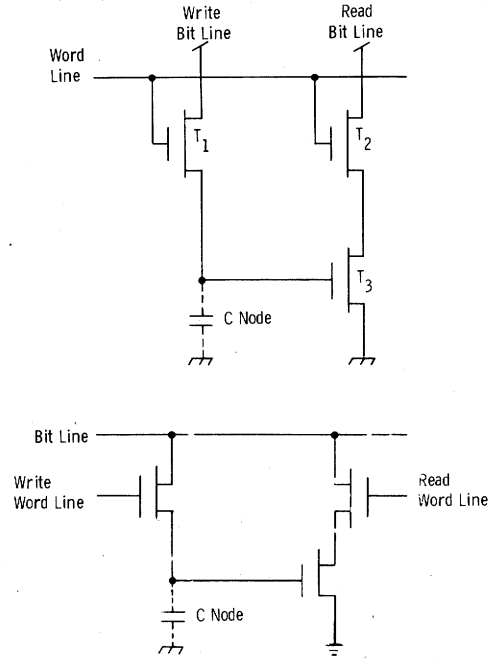


Fig. 7. Capacitance storage memory circuits.

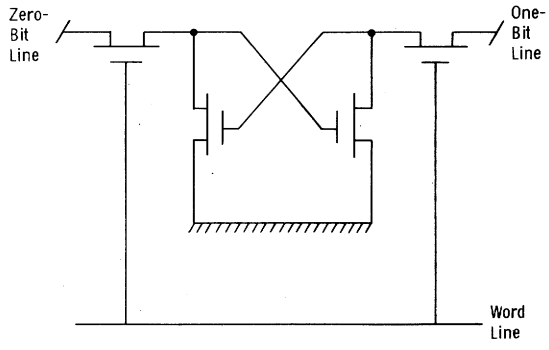


Fig. 8. Charge storage with internal refresh.

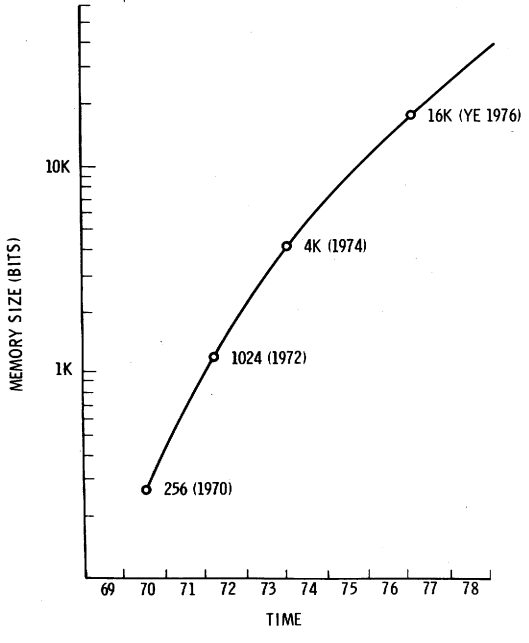


Fig. 9. Memory growth rate.

could extrapolate the future on a simple trajectory to an asymptotic transistor limit.

A review of past progress that has brought us this far, however, reveals a strong reliance on circuit simplification and improvisation and a somewhat lesser dependence on device miniaturization. This trend can be projected to continue into the 1980's with a continued decline in cost per circuit.

Bipolar Logic

The first monolithic bipolar logic products averaged less than six circuits per chip. Chips were about 75 mils on edge and contained 12–16 input/output connections. At this modest level of integration the logic functions performed on chip were limited to simple primitive AND, OR, Invert functions. The most commonly used circuits were DTL, RTL, ECL, T^2L. Except for the T^2L circuit, they were carry-overs from discrete component technology. ECL circuits were used where high performance was required. DTL and T^2L competed for cost performance applications, and RTL was used for lower speed, cost-oriented applications. Typical early RTL, DTL, and T^2L circuits are shown in Figs. 10–12. In the DTL circuit shown in Fig. 11, an AND-OR logic statement is performed with diodes and the transistor is used to invert and amplify. This was probably the most commonly used discrete component transistor logic circuit, and it is not surprising that it was carried over to monolithic form. Early handbooks [67] quoted performance around 30 ns per gate at 10 mW of power dissipation. The output device was usually driven into saturation. The DTL circuit is characterized by good logic capability; that is, a single circuit provides an AND-OR-Invert function. However, it requires more silicon area and has a poorer power performance ratio than T^2L.

The T^2L circuit of Fig. 12 is well known [68] because of its adaptability to integration (less silicon area), its low speed-power product, and its good parameter tolerance. The output drive is provided by a push-pull circuit comprising T1, T2, D, and R. Depending on the value of $+V_C$, T1 and D may be off when T2 is full on, or T1 and D may be used to limit the down-level of T2 for larger values of $+V_C$, in which case the resistor provides current limiting. R also provides some degree of short circuit protection. It is important to balance the input time constants at T1 and T2 so that T1 does not turn on before T2 turns off and

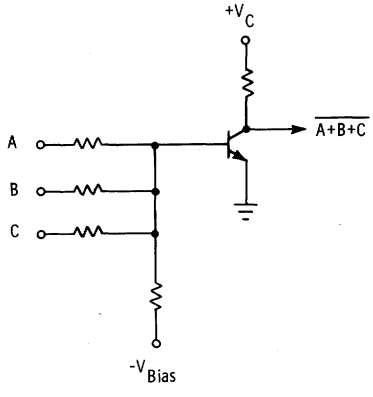


Fig. 10. Early RTL circuit.

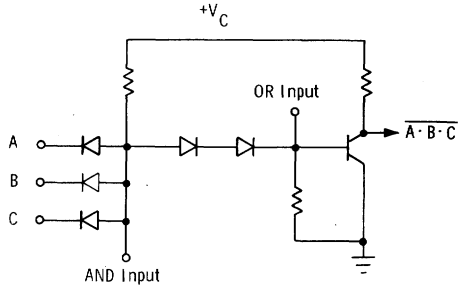


Fig. 11. Early AND-OR DTL circuit.

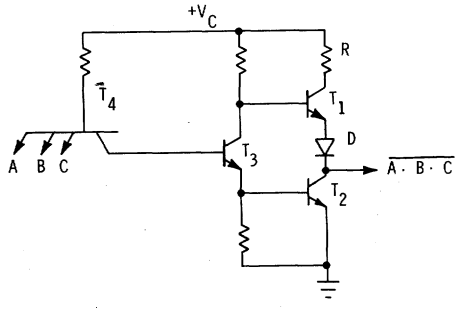


Fig. 12. Early T²L circuit.

cause a high current path from V_C to ground. Experience in design and chip layout has shown this to be a solvable problem. Push-pull drivers were desirable at low levels of integration to drive large capacitances and long signal lines. They contributed most of the total power dissipated by the chip. Today, drivers of this sort are still necessary and they still contribute significantly to the total power dissipation, but they are few in number compared with the logic circuits that do not drive off chip. Early T²L circuits had performances in the 10 and 20 ns range at about 15 mW/gate.

The RTL circuit, because of its low performance and limited margins, did not enjoy the success of DTL or T²L circuits. By and large it has been replaced by FET logic circuits.

A representative ECL circuit is shown in Fig. 13. Additional inputs are provided by paralleling transistors, as shown by T1 and T2. A relatively fixed current flows through R3 to the $-V$ power supply. This current is switched from T1 or T2 to T3 by application of an input

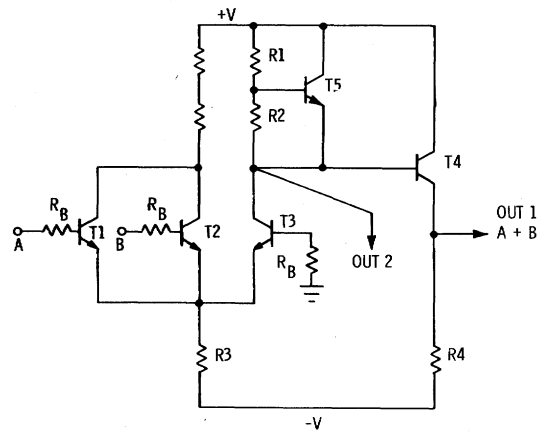


Fig. 13. ECL circuit configuration.

signal; that is, if the inputs at T1 and T2 are negative, T3 conducts. Conversely, if the inputs to either T1 or T2 or both are positive, they conduct and T3 is off. The function of T5 is to clamp the down-level signal at the collector of T3 and to provide an output at this point that can be wired to other similar circuit outputs to form a wired AND function. The function of T4 is to provide drive capability, but it may also be wired to other similar outputs to form a wired OR function. The entire load network of R1, R2, R4, T4, and T5 could also be replicated at the common collector of T1 and T2 to provide two more outputs, which would then be the logic complements of outputs 1 and 2. The required minimum input signal is small, being the difference between the "on" and "off" values of the emitter base diode voltage. Representative values would be ±300 mV. The delays from input to both normal and complementary outputs are about equal for equal loading.

The advantages of the ECL circuit are its high performance, its versatile logic capability, its good noise rejection capability, and its tendency to reduce switching transients on power supply lines because circuit current is never switched off. The disadvantages are high power dissipation as compared with circuits in which the current is switched off completely, high component count, which translates to large silicon area used per circuit, and possible instability. The input impedance of the circuit is a product of beta (base-to-collector current gain) and the emitter load, both of which vary in phase and magnitude with frequency. This product exhibits a negative resistance, often over a significant frequency range. It is common practice to place a resistor, R_B in Fig. 13, close to each base so that lines connected to the circuit do not see a negative resistance. This is done at a small loss in performance. It is usually required at low levels of integration where there are many long lines in a system, generally never perfectly terminated, and therefore capable of providing an inductance sufficient to cause the circuit to oscillate. Experience with various computer packages has shown this to be a completely solvable problem. The power dissipation may be significantly reduced by using a transistor current source in place of R3 in Fig. 13. This also makes the current switched in-

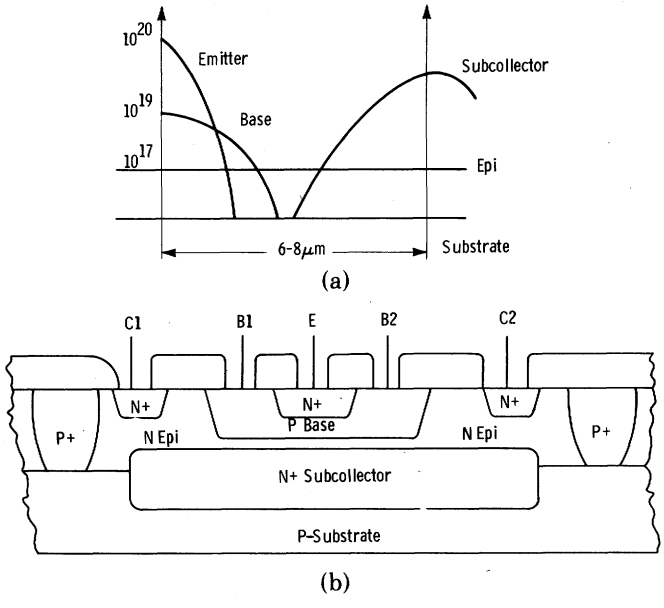


Fig. 14. (a) Typical early monolithic transistor profile. (b) Typical early monolithic transistor cross section.

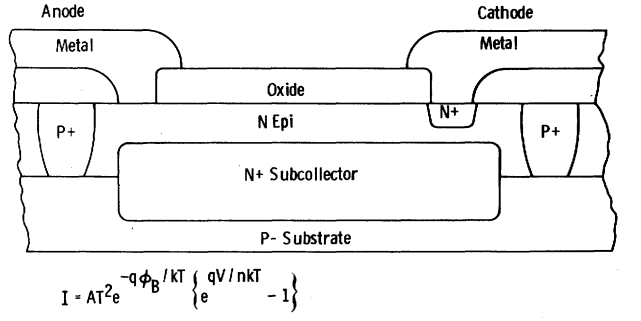
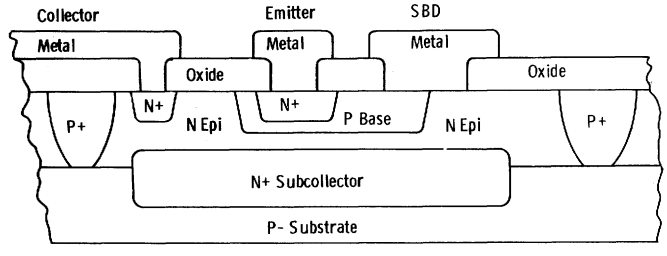


Fig. 15. A Schottky barrier diode.



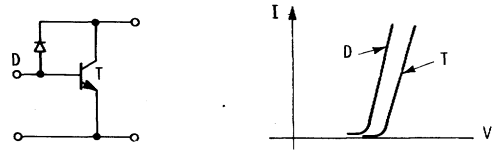


Fig. 16. SBD antisaturation clamp on transistor.

dependent of the input signal. The ECL circuit is the fastest circuit (smallest delays) in use today.

The device profile and cross section shown in Fig. 14(a) and (b) are representative of an average transistor used in the first monolithic products. Briefly, the n^+ subcollector was arsenic or antimony, the epitaxial layer was 6–8 μm thick, the base was boron, and the emitter was phosphorus. p^+junction isolation was used and F_TS were approximately 500 MHz. Resistors were made with the base diffusion, which ranged from 250 to 400 Ω/□. The circuit performance ranged from about 5 ns for the highest powered ECL circuit to about 25 ns for the lowest powered T^2L circuit.

During the late sixties the level of integration increased in monolithic products. Perhaps the most significant outward sign of monolithic circuit progress was the increasing availability of a wide range of custom-designed chips containing a wide variety of logic functions. By 1970, these chips were available by circuit family with compatible power supplies and rules for interconnection across a performance range of 1–2 to 35–50 ns.

By 1970, the Schottky barrier diode (SBD) was being used in monolithic circuits. A typical cross section of an SBD is shown in Fig. 15. This majority-carrier device has current densities of ~2000 A/cm^2 and switches very fast. The diode drop is a function of the metal used to contact the silicon epi surface, as shown by the ϕ_B term in the equation in Fig. 15. The SBD has had a large impact on both T^2L and DTL circuits [69], [70]. First the diodes were used to clamp grounded emitter amplifiers out of deep saturation. In Fig. 16, if diode D has a drop slightly below the base collector diode of T, it will conduct before the base collector diode goes into deep saturation and essentially eliminate the minority carrier storage delay effect from the transistor. Furthermore, the diode occupies a very small amount of silicon area. Fig. 16 shows a transistor cross section in which a split base collector contact is used to form the diode.

Fig. 17 shows a Schottky clamped T^2L circuit from a 1971 catalog [71]. The quoted performance was typically 3 ns at 20 mW of power dissipation. Similar improvements were also realized with DTL circuits.

During the 1960's and 1970's, the trend to higher levels of integration continued. The logic subassembly of the late 1960's grew in complexity until today's design activity centers on subassemblies such as complete micprocessors [72], which have about 500–1000 times as many components as their predecessors of the early sixties.

Clearly, the bipolar technology of today has greatly improved, allowing new and superior products. These technology improvements could be listed as follows: significant improvement in photolithography capability, which has made it possible to fabricate much smaller devices; significant improvement in diffusion and epi thickness control, which has resulted in narrower base widths, smaller sidewall capacitances, and faster devices; further reduction of sidewall capacitance by use of dielectric isolation, which has further improved device performance; the development of multilayer metallization techniques without which complex chips could not be wired; the use of ion implantation to form junctions and to adjust impurity profiles in structures to great accuracy; and the ability to grow large-diameter single-crystal silicon wafers.

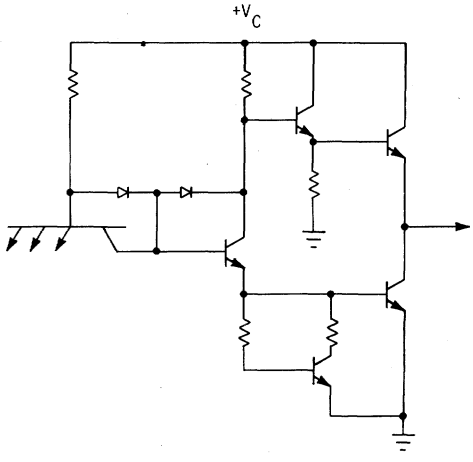


Fig. 17. Schottky clamped T^2L circuit.

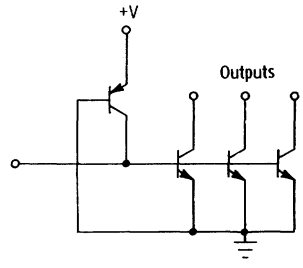


Fig. 18. Basic I^2L circuit.

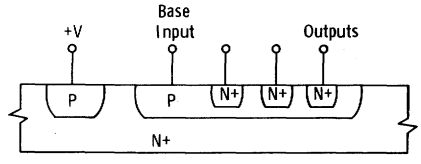


Fig. 19. Basic I^2L structure.

ECL and T^2L circuits are still the most used today. A new circuit [73], [74], I^2L, has created considerable interest. A Schottky diode version of this circuit [75] offers improvements in performance. ECL circuits [76] have been improved by the addition of reference bias generators on chip, which provide tracking with temperature and therefore better circuit capability. Otherwise the circuit itself has not changed. However, the performance realized with today's devices is in the 0.5–1.0 ns range for on-chip delays, as compared with 5.0 ns in 1962–1964. T^2L circuit performance is now down to 1–2 ns on chip. The basic circuits are still the same, except at higher levels of integration the output driver in the T^2L circuit of Fig. 17 is not used for circuits that do not drive off-chip.

A basic I^2L circuit is shown in Fig. 18 and a structure in Fig. 19. A logic circuit is formed by paralleling outputs of several of the basic circuits. Note from the structure of Fig. 19 that the output device is operated in the inverted mode and outputs are taken from the normal emitters. In this way the base of the p-n-p is tied to the substrate and ground. The advantage of the basic I^2L circuit is that it probably occupies less silicon area than any other circuit, is simply fabricated (4 masks), and requires only one power supply. In its simplest form it offers a challenge to FET circuits in fabrication complexity, performance, and superior speed-power products.

To summarize the discussion of today's bipolar circuits, we observe that the faster circuits offer less than 1 ns on-chip delays, and the slowest circuits overlap the fastest FET logic circuits. When a circuit with today's devices has an on-chip delay of much greater than 5.0 ns for moderate loads it generally is a deliberate action on the part of the designer to gain lower power dissipation for greater circuit density or perhaps greater noise immunity.

Future Prospects

Addressing the question of what will happen during the next decade, it should be possible to push performance down into the low-to-mid-picosecond range with a high-performance capability between 200 and 300 ps. The cost performance logic circuits should replace what is high performance today, which will probably center around 700 ps. There probably will be new circuits, but both ECL and the DTL circuits with low drop SBD's should be able to achieve the high-performance limit stated above. Table I compares an average 1976 device with what the device of 1985 might be.

Performance of the device itself will be limited by base resistance and substrate plus sidewall capacitance. The practical problem of contacting the base and emitter with metal lines will determine how close together these contacts can be and will fix a lower limit on base resistance.

Epitaxial thickness will probably be defect-limited and the control on epi thickness and the occurrence of pipes [77] will fix a lower limit on base width and, to some extent, sidewall capacitances. The small emitter size will result in a smaller device and a direct reduction of capacitance to substrate and base-to-collector capacitance. At the same current density then, the capacitances over all will be significantly smaller and the performance will improve.

Given this new device, then what? For high-performance applications the most severe limits would seem to be in packaging chips: specifically, the number of input/output terminals, how to fan out from the chip, and how to cool the chip. However, the product improvements possible with circuits using the device projected in Table I will create the impetus to push forward. The level of integration will be fixed by manufacturing yields.

In cost performance computer applications, the packaging limit should not be so severe, since the higher the level of integration, the fewer chips per system. If the integration level on chip is a significant percentage of the total circuits in the system, off-chip drivers will be minimized and good use can be made of the device at lower power levels.

Bipolar Memory

Bipolar memories evolved basically from bipolar transistor flip-flops. When flip-flops are arranged in array, and encoding, decoding, control, and output buffering circuits

TABLE I
Comparison of Average 1976 Transistor with What the 1986 Device May Be

	1976	1986
Emitter size (μm)	3 x 8	1 x 2
Base width (μm)	2200	1000
Epi thickness (μm)	2.0	1.0
F_T (GHz)	2.5	> 5

are included in a monolithic chip, an integrated semiconductor random-access memory (RAM) results. Because of the compatibility between the peripheral circuits and the flip-flop array (unlike the FET memories discussed earlier), the peripheral circuitry is much simpler than that required for core or other magnetic memories. Because of the high speed of bipolar transistors, the performance of bipolar memory greatly exceeds that of core memory. Memory access time has been brought to under 100 ns from that of core memory at around 1 μs.

Nondestructive read is obtained automatically because of the characteristic of the flip-flops. Initially, the cost of bipolar memory was the biggest problem; R. Rice publicized [78] a $5-a-bit scratchpad bipolar memory at a panel discussion during the 1965 International Solid-State Circuits Conference (ISSCC). B. Agusta [79] reported a 16-bit random-access bipolar memory chip at the October 1965 Professional Group on Electron Devices (PGED) Conference, Washington. As integrated circuit technology developed further, silicon chip size grew larger and each circuit occupied less area. Cost per bit of bipolar memory decreased as density increased. Total power consumption and removal of the heat generated by it has become a more limiting factor than the photolithographic limit imposed by the integrated circuit technology. Bipolar memories, because of their high speed, are generally used in scratchpad and buffer memories, caches, control stores. For the main memory application, the denser and less power-consuming, yet slower, FET memories dominate. Roughly, the n-channel FET memory chips operated in a dynamic mode are four times denser (4K versus 1K bits) and eight times slower (200 versus 25 ns). In order to increase density and to reduce power consumption, bipolar memory operated in a dynamic mode has been developed [80]–[83]. Dynamic mode operation loses the nondestructive read feature and requires periodic regeneration or refreshing. The high density decreases per-bit memory cost with some decrease in performance.

Because of its high speed, and its compatability with computer logic circuits, bipolar memories find usage in content-addressable-memory (CAM) or associative memory [84], [85]. CAM performs the functions of both memory and logic.

In a computer-control store, where high performance is most desirable and the information stored rarely changes, bipolar read-only-memory (ROM) fits in nicely [86], [87].

Bipolar ROM may also take the form of being programmable (PROM). Each single transistor that contains a bit of information in the memory array may have its emitter contact connected or not connected.

Static Mode Random-Access Memory

A full description of a bipolar RAM [88] was made by Perkins and Schmidt in 1965. The memory cell contained three bipolar transistors at a comparatively slow 150 ns access time. Publication of a more complex six-transistor memory cell, at a reading speed of 60 ns with 8 bits on a chip, soon followed [89]. A high-performance bipolar memory cell at 17 ns read-write cycle time utilizing current mode operation was described at the 1967 Fall Joint Computer Conference [90]. The higher speed is obtained at a higher power consumption rate. A much simpler memory cell with only two direct-coupled bipolar transistors was shown [91] to pack 16 bits on a chip and to operate at 100 ns read-write cycle time. The bipolar RAM used in IBM Systems/360 and 370 had a packing density of 64 bits per chip [92]–[94], with a chip size of 112 mils square. A cell is shown in Fig. 20. A 150K bit basic storage unit exhibits a worst-case performance of 40 ns access time and less than 60 ns cycle time.

At the 1970 ISSCC meeting, Green and Phan presented a bipolar RAM in the emitter-coupled mode [95], as shown in Fig. 21. The RAM's were developed for the Illiac IV computer, with 100 ns access and 200 ns cycle time for a memory system of 2048 words by 64 bits. The typical power consumption of the 256-bit chip is 500 mW. At the same conference, Lynes and Hodges showed a Schottky-diode-coupled bipolar transistor memory [96].

A switched collector impedance memory cell that also features an emitter-coupled configuration was presented by Tangiguchi *et al.* [97]. A switching transistor is added in parallel with the two load resistors of a conventional emitter-coupled cell. The new cell has a greatly improved ratio of read-out current to standby current of the memory cell.

Tsang [98] described in 1974 a 1024-bit bipolar RAM utilizing pinched-base resistors that resulted in lower power and higher density. In the same period, another 1K bipolar RAM developed by Magumi *et al.* [99] had a chip access time of 25 ns and operated in an emitter-coupled mode. Small geometry with 2.5 × 2.5 μm minimum emitter size was employed.

An emitter-coupled RAM utilizing parallel diodes with ion-implanted load resistors was reported recently by Rathbone *et al.* [100]. It features a 1024-bit with 15 ns access time. Oxide isolation technology has been adopted for performance improvement. The sizable resistance (40 K Ω) of the ion-implanted resistors reduces the power consumption during standby, and the paths through parallel diodes allow large read or write current when the memory is being actively used. The cell is shown in Fig. 22.

A novel bipolar memory cell of extremely low power dissipation and small size was presented by Wiedmann and Berger [101] at the 1971 ISSCC meeting. The memory uses

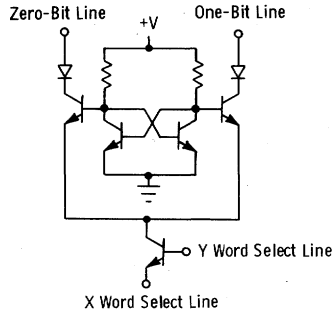


Fig. 20. Memory cell with two-dimensional word selection.

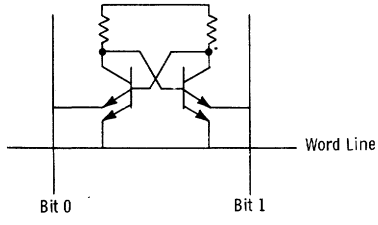


Fig. 21. Memory cell of emitter-coupled cell.

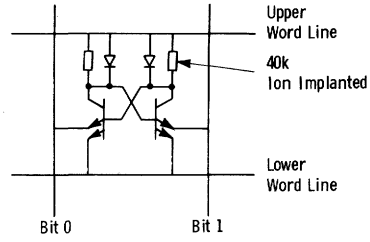


Fig. 22. Parallel diode ECL memory cell.

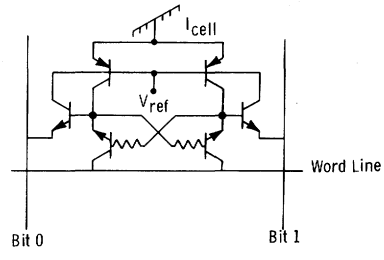


Fig. 23. MTL memory cell.

inversely operated n-p-n transistors as flip-flop transistors and lateral p-n-p transistors as load devices. The entire array needs only one isolation region. This memory cell is shown in Fig. 23. The circuit configuration was later labeled merged transistor logic (MTL) or integrated injection logic (I^2L) [73], [74], [101], [102].

Dynamic Mode Random-Access Memory

In a static-mode RAM cell, steady-state current is maintained for storing information in the flip-flop. In dynamic-mode operation, information is stored in capacitances and no steady-state current is required. Total power consumption is drastically reduced. Cell size will also decrease because many elements can be eliminated in a dynamic cell. The information stored in the capacitances is an electric charge, which gradually discharges. Regeneration or refreshing operations must be carried out periodically.

In one version of dynamic RAM [80], the flip-flop remains but the load resistors have been eliminated. Information in the form of electric charges is stored mainly in the capacitances of base-emitter p-n junctions.

Another version of bipolar dynamic RAM eliminates the flip-flop entirely. The storage cell contains a single transistor with a floating base [81]. Extremely high density is obtainable from this memory. In writing a "0," charges are trapped in the floating base by pulsing the collector. In writing a "1," the trapped charges should be reduced to a minimum. Junction breakdown is utilized for the charge reduction operation.

The third version of a bipolar dynamic memory cell [82], [83] bears a similarity to MTL or I^2L technology. The cell contains a lateral p-n-p and a vertical n-p-n transistor. Each cell is actually smaller in size than a single transistor. The circuit configuration is shown in Fig. 24. Information is stored in the base-collector junction. The p-n-p transistor controls the charge flow, and the n-p-n transistor provides internal amplification.

Associative or Content-Addressable Memory

The associative memory or CAM is organized for parallel search. An associative cell is usually capable of holding three states: 0, 1, and don't care. One obvious approach [84] is to utilize two conventional memory cells to perform the function of one associative cell with 75 percent efficiency (using three out of the possible four states). Another approach is to specially design an associative cell with the capability of responding to these three input states. A rather simple three-transistor [85] associative cell is shown in Fig. 25. Associative interrogation is performed by pulsing one of the two-bit lines in every pertinent bit position. A mismatch signal will appear in the word sense line when the stored information in the cell does not match the desired content. When neither bit line is pulsed, this particular bit will not give out a mismatch signal. It is the don't-care response.

Read-Only Memory and Programmable Read-Only Memory

Bipolar ROM arrays may be built with extremely high density. Each bit needs only one transistor. In addition, a whole row or whole column may have a common base. One-dimensional isolation in the array is sufficient to achieve a density improvement. Roughly, a ROM array may be about four times denser than a RAM array for the same chip size.

The memory array of a 1024-bit bipolar ROM [86] is shown in Fig. 26. A common collector is shared by all the transistors in the array. The make or break of the emitter contact represents a 0 or a 1. The programming may be

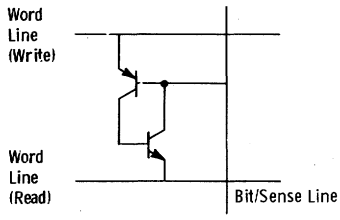


Fig. 24. Dynamic memory cell.

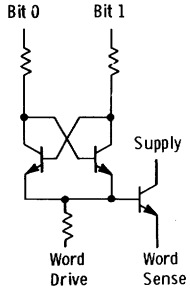


Fig. 25. Associative memory cell.

done at the metallization level or in the field. In the latter case, fusable contacts are employed. Current of much larger magnitude than that of regular operation may be directed toward the contact point to burn out the contact.

Future of Bipolar Memories

In projecting the future of bipolar memories it seems safe to assume that they will continue to play a major role for high-speed cache requirements. It is difficult to assess whether they will displace a significant portion of the main memory market currently dominated by FET technology. One very large computer developed by Cray Research, Inc. utilizes a 50 ns bipolar main memory [103]. To displace the FET RAM, bipolar memories will have to compete on a cost and performance basis. Bipolar devices are susceptible to a defect not present in FET's [77]. Minimization of this problem, known as pipes, is necessary for bipolars to achieve the required yields. Cell sizes will need to compete with those of dynamic FET RAM's and this will depend to a great extent on the ingenuity of the designers. Bipolar memories today have reached no limiting boundaries, and we can expect to see continued improvement in performance and density.

Conclusion

In the last 15 years, we have seen silicon technology go from a single device on a chip to complexities of the order of 10^4–10^5 devices per chip, a change of four to five orders of magnitude in complexity since the fabrication of the first integrated circuits in the early 1960's. It is projected [50] that a potential eight orders of magnitude development is possible, with no fundamental obstacles to 10^7–10^8-device integrated circuits. For the future, integrated circuits will remain a very dynamic changing technology, with an impact only partially perceived at this time.

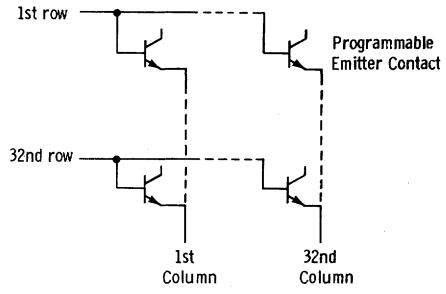


Fig. 26. Bipolar ROM array.

References

[1] G. R. Moore, "Progress in digital integrated electronics," in *Proc. IEEE Int. Electron Devices Meeting,* Dec. 1975.
[2] F. F. Fang and A. B. Fowler, "Hot electron effects and saturation velocities in silicon layers," *J. Appl. Phys.,* vol. 41, p. 1825, 1970.
[3] J. E. Lilienfeld, U. S. Patent 1, 745, 175, 1930 and U. S. Patent 1, 900, 018, 1935.
[4] W. Shockley and G. L. Pearson, "Modulation of conductance of thin films of semiconductors by surface charges," *Phys. Rev.,* vol. 74, p. 232, 1948.
[5] M. M. Atalla, E. Tannenbaum, and E. J. Schreibner, "Stabilization of silicon surfaces by thermally grown oxides," *Bell Syst. Tech. J.,* vol. 38, p. 749, 1959.
[6] D. Khang and M. M. Atalla, "Silicon-silicon dioxide field induced surface devices," *IRE Solid-State Device Res. Conf.,* Carnegie Inst. Technol., Pittsburgh, PA, 1960.
[7] H. K. J. Ihantola, "Design theory of a surface-field-effect transistor," Stanford Electron. Lab., Stanford Univ., Stanford CA, SEL Tech. Rep. 1661-1, 1961.
[8] S. R. Hofstein and F. P. Heiman, "The silicon insulated-gate field-effect transistor," *Proc. IEEE,* vol. 51, p. 1190, 1963.
[9] C. T. Sah, "Characteristics of the metal-oxide-semiconductor transistor," *IEEE Trans. Electron Devices,* vol. ED-11, p. 324, 1964.
[10] H. K. J. Ihantola and J. L. Moll, "Design theory of a surface field-effect transistor," *Solid-State Electron.,* vol. 7, p. 423, 1964.
[11] J. T. Wallmark and S. M. Marcus, "Integrated devices using direct-coupled unipolar transistor logic," *IRE Trans. Electron. Devices,* vol. ED-8, p. 98, 1959.
[12] D. E. Farina and D. Trotter, "MOS integrated circuits save space and money," *Electronics,* vol. 38, p. 84, Oct. 1965.
[13] M. Axelrod, "Integrated IGFET logic circuit with linear resistive load," U. S. Patent 3 406 298, Oct. 15, 1968.
[14] H. C. Lin and C. J. Varker, "Normally-on load device for IGFET switching circuits," *NEREM Rec.,* vol. 11, p. 124, 1969.
[15] J. MacDougall, K. Manchester, and R. B. Palmer, "Ion implantation offers a bagful of benefits for MOS," *Electronics,* vol. 43, p. 86, June 22, 1970.
[16] T. Masuhara, M. Nagata, and N. Hashimoto, "A high-performance n-channel MOS LSI using depletion-type load elements," *IEEE J. Solid-State Circuits,* vol. SC-7, p. 224, 1972; also *Int. Solid-State Circuits Conf., Digest Tech. Papers,* 1971, p. 12.
[17] F. M. Wanlass and C. T. Sah, "Nanowatt logic using field-effect metal-oxide semiconductor triodes," *Proc. Int. Solid-State Circuits Conf.,* 1963, p. 32.
[18] J. R. Burns, "Switching response of complementary-symmetry MOS transistor logic circuits," *RCA Rev.,* vol. 25, p. 627, 1964.
[19] R. W. Ahrons and M. M. Mitchell, "MOS micropower complementary transistor logic," *Int. Solid-State Circuits Conf., Digest Tech. Papers,* 1965, p. 80.
[20] J. Luscher and H. Hofbauer, "Monolithic micropower frequency divider using P-MOS transistors," *Int. Solid-State Circuits Conf., Digest Tech. Papers,* 1966, p. 116.
[21] J. Karp and E. deAtley, "Using four-phase IC logic," *Electron. Des.,* vol. 15, p. 62, 1967.
[22] L. Cohen, R. Rubenstein, and F. Wanlass, "MTOS four-phase clock systems," *NEREM Rec.,* vol. 9, p. 170, 1967.
[23] B. G. Watkins, "Low-power multiphase circuit technique," *IEEE J. Solid-State Circuits,* vol. SC-2, p. 213, 1967.

[24] W. N. Carr and J. P. Mize, *MOS/LSI Design and Applications.* New York: McGraw-Hill, 1972, p. 152.

[25] P. W. Cook, D. L. Critchlow, and L. M. Terman, "Comparison of MOSFET logic circuits," *IEEE J. Solid-State Circuits,* vol. SC-8, p. 348, 1973.

[26] D. R. Kerr, J. S. Logan, P. J. Burkhardt, and W. A. Pliskin, "Stabilization of SiO_2 passivation layers with P_2O_5," *IBM J. Res. Develop.,* vol. 8, p. 376, 1964.

[27] E. H. Snow and B. E. Deal, "Polarization phenomena and other properties of phosphosilicate glass films on silicon," *J. Electrochem. Soc.,* vol. 113, p. 263, 1966.

[28] C. W. Mueller and P. H. Robinson, "Grown-film silicon transistors on sapphire," *Proc. IEEE,* vol. 52, p. 1487, 1964.

[29] R. W. Bower and H. G. Dill, "Insulated gate field-effect transistors fabricated using the gate as a source drain mask," presented at the IEEE Int. Electron Devices Meeting, Oct. 1966.

[30] R. W. Bower, H. G. Dill, K. G. Aubuchon, and S. A. Thompson, "MOS field-effect transistors formed by gate masked ion implantation," *IEEE Trans. Electron Devices,* vol. ED-15, p. 757, 1968.

[31] J. C. Sarace, R. E. Kerwin, D. L. Klein, and R. Edwards, "Metal-nitride-oxide-silicon field-effect transistors, with self-aligned gates," *Solid-State Electron.,* vol. 11, p. 653, 1968.

[32] F. Faggin, T. Klein, and L. Vadasz, "Insulated gate field effect transistor integrated circuits with silicon gates," in *Proc. IEEE Int. Electron Devices Meeting,* Oct. 1968, p. 22.

[33] L. L. Vadasz, A. S. Grove, T. A. Rowe, and G. E. Moore, "Silicon-gate technology," *IEEE Spectrum,* p. 28, Oct. 1969.

[34] J. A. Appels, H. Kalter, and E. Kodi, "Some problems of MOS technology," *Philips Tech. Rev.,* vol. 31, p. 225, 1970.

[35] K. G. Aubuchon, "The use of ion implantation to set the threshold voltage of MOS transistors," in *Proc. Int. Conf. Properties and Use of MIS Structures,* 1969, p. 575.

[36] M. R. MacPherson, "The adjustment of MOS transistor threshold voltage by ion implantation," *Appl. Phys. Lett.,* vol. 18, p. 502, 1971.

[37] J. M. Shannon, "Ion-implanted high frequency MOS transistors," *Philips Tech. Rev.,* vol. 31, p. 267, 1970.

[38] J. D. Sansbury, "MOS field threshold increase by phosphorus-implanted field," *IEEE Trans. Electron Devices,* vol. ED-20, p. 473, 1973.

[39] F. F. Fang and H. S. Rupprecht, "High performance MOS integrated circuit using the ion implantation technique," *IEEE J. Solid-State Circuits,* vol. SC-10, p. 204, 1975.

[40] T. P. Cauge and J. Kocsis, "A double-diffused MOS transistor with microwave gain and subnanosecond switching speeds," presented at the IEEE Int. Electron Devices Meeting, Oct. 1970.

[41] Y. Tarui, Y. Hayashi, and T. Sekigawa, "DSA enhancement-depletion MOS IC," presented at the IEEE Int. Electron Devices Meeting, Oct. 1970.

[42] H. J. Sigg, G. D. Vendelin, T. P. Cauge, and J. Kocsis, "D-MOS transistor for microwave applications," *IEEE Trans. Electron Devices,* vol. ED-19, p. 45, 1972.

[43] K. Ohta *et al.,* "A high-speed logic LSI using diffusion self-aligned enhancement depletion MOS IC," *IEEE J. Solid-State Circuits,* vol. SC-10, p. 314, 1975.

[44] T. J. Rodgers and J. D. Meindl, "VMOS: High-speed TTL compatible MOS logic," *IEEE J. Solid-State Circuits,* vol. SC-9, p. 239, 1974.

[45] M. Pomper and J. Tihanyi, "Ion-implanted ESFI MOS devices with short switching times," *IEEE J. Solid-State Circuits,* vol. SC-9, p. 250, 1974.

[46] C. A. Mead, "Schottky barrier gate field-effect transistor," *Proc. IEEE,* vol. 54, p. 307, 1966.

[47] W. W. Hooper and W. I. Lehrer, "An epitaxial GaAs field-effect transistor," *Proc. IEEE,* vol. 55, p. 1237, 1967.

[48] K. E. Drangeid, R. Somnerhalder, and W. Walter, "High-speed gallium-arsenide Schottky-barrier field-effect transistor," *Electron. Lett.,* vol. 6, p. 228, 1970.

[49] R. L. VanTuyl and C. A. Liechti, "High-speed integrated logic with GaAs MESFETs," *IEEE J. Solid-State Circuits,* vol. SC-9, p. 269, 1974.

[50] I. E. Sutherland, C. A. Mead, and T. E. Everhart, "Basic limitations in microcircuit fabrication technology," Rand Corp., Santa Monica, CA, WN-9404-ARPA, Feb. 1976.

[51] R. H. Dennard *et al.,* "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE J. Solid-State Circuits,*

[52] M. Kubo, R. Hori, O. Minato, and K. Sato, "A threshold voltage controlling circuit for short-channel MOS integrated circuits," *IEEE Int. Solid-State Circuits Conf. Digest,* p. 54, 1976.

[53] E. M. Blaser, W. M. Chu, and G. Sonoda, "Substrate load gate voltage compensation," *IEEE Int. Solid-State Circuits Conf. Digest,* p. 56, 1976.

[54] J. H. Pomerene and R. W. Melvill, "Error correctable and repairable data processing storage system," U. S. Patent 3 436 734, Apr. 1, 1969.

[55] L. Terman, "MOSFET memory circuits," *Proc. IEEE,* vol. 59, pp. 1044–1058, July 1971.

[56] J. Wood and R. G. Ball, "The use of insulated-gate field-effect transistors in digital storage systems," *ISSCC Digest Tech. Papers,* pp. 82–83, 1965.

[57] H. M. Burke and G. J. Michon, "Charge pump random-access memory," *ISSCC Digest Tech. Papers,* pp. 16, 17, and 208, 1972.

[58] R. H. Dennard, "Field-effect transistor memory," U. S. Patent 3, 387, 286, June 1968.

[59] L. Boonstra, C. W. Lambrechtse, and R. H. W. Salters, "A 4096 one transistor per bit random-access memory with internal timing and low dissipation," *IEEE J. Solid-State Circuits,* vol. SC-8, no. 5, pp. 305–310, 1973.

[60] W. R. Hoffman and H. L. Kalter, "An 8K random-access memory chip using the one-device FET cell," *IEEE J. Solid-State Circuits,* vol. SC-8, no. 5, pp. 298–305, 1973.

[61] C. N. Ahlquist *et al.,* "A 16K dynamic RAM," *ISSCC Digest Tech. Papers,* pp. 128–129, 1976.

[62] J. J. Chang, "Nonvolatile semiconductor memory devices," *Proc. IEEE,* vol. 64, pp. 1039–1059, July 1976.

[63] E. G. Fubini and M. G. Smith, "Limitations in solid state technology," *IEEE Spectrum,* pp. 55–59, May 1967.

[64] R. Keyes, "Physical limits of digital electronics," *Proc. IEEE,* vol. 63, pp. 740–767, May 1975.

[65] B. Hoeneisen and C. A. Mead, "Limitations in microelectronics," *Solid-State Electron.,* vol. 15, pp. 819–829 and 891–897, 1972.

[66] D. A. Hodges, "A review and projection of semiconductor components for digital storage," *Proc. IEEE,* vol. 63, pp. 1136–1147, Aug. 1975.

[67] R. G. Gibson *et al., Integrated Circuits Handbook.* Boston, MA: Boston Tech. Publ., 1966.

[68] R. G. Hibberd, *Integrated Circuits* (Texas Instruments Electronic Series). New York: McGraw-Hill, 1969, p. 81.

[69] Y. Tarui, Y. Hayashi, H. Teshiura, and T. Sekigawa, "Transistor Schottky barrier diode integrated logic circuits," *IEEE J. Solid-State Circuits,* vol. SC-4, pp. 3–12, Feb. 1969.

[70] R. A. Heald and D. A. Hodges, "Design of Schottky barrier diode clamped transistor layout," *IEEE J. Solid-State Circuits,* vol. SC-8, pp. 269–275, Aug. 1973.

[71] Texas Instruments Series 545/745, Bull. DL-S 7111414, May 1971.

[72] "TRW system sets gate-density record for bipolar logic," *Electronics,* p. 29, Mar. 20, 1975.

[73] H. H. Berger and S. K. Wiedmann, "Merged transistor logic, a low-cost bipolar logic concept," *IEEE J. Solid-State Circuits,* vol. SC-7, pp. 340–346, Oct. 1972.

[74] K. Hart and H. Slol, "Integrated injection logic, a new approach to LSI," *IEEE J. Solid-State Circuits,* vol. SC-7, pp. 346–351, Oct. 1972.

[75] "Bell Labs develops Schottky I^2L logic to replace TTL arrays," *Electronics,* Sept. 4, 1975.

[76] V. A. Dhaka, J. E. Muscliushe, and W. Kowens, "Subnanosecond emitter-coupled logic gate circuit using isoplanar II," *IEEE J. Solid-State Circuits,* vol. SC-8, pp. 368–372, Oct. 1973.

[77] F. Barson, M. S. Hess, and M. M. Ray, "Diffusion pipes in silicon NPN structures," *J. Electrochem. Soc.,* vol. 116, Feb. 1969.

[78] "Scratchpads," *Electronics,* p. 46, Mar. 8, 1965.

[79] B. Agusta, "A 16-bit monolithic memory array chip," Professional Group on Electron Devices, Washington, DC, Oct. 1965.

[80] H. H. Henn, "Bipolar dynamic memory cell," *IEEE J. Solid-State Circuits,* vol. SC-6, pp. 297–300, Oct. 1971.

[81] J. Mar, "A two-terminal transistor memory cell using breakdown," *IEEE J. Solid-State Circuits,* vol. SC-6, pp. 280–283, Oct. 1971.

[82] W. B. Sander, J. M. Early, and T. A. Longo, "A 4096 × 1 (I^3L) bipolar dynamic RAM," *ISSCC Digest Tech. Papers,* pp. 182–183, Feb. 1976.

[83] W. F. Beausoleil, I. T. Ho, T. S. Jen, and W. D. Pricer, "Two-device

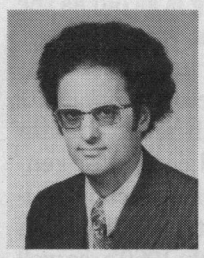
monolithic bipolar memory array," U. S. Patent 3 697 962, Oct. 10, 1972.
[84] P. L. Gardner, "Functional memory and its microprogramming implications," *IEEE Trans. Comput.*, vol. C-20, pp. 764–775, July 1971.
[85] A. W. Bidwell and W. D. Pricer, "A high-speed associative memory," *ISSCC Digest Tech. Papers*, vol. 10, pp. 78–79, Feb. 1967.
[86] A. Bergh, J. C. Barrett, and J. E. Price, "Design considerations for a high-speed bipolar read-only memory," *ISSCC Digest Tech. Papers*, pp. 66–67, Feb. 1970.
[87] G. Luecke, J. P. Mize, and W. N. Carr, *Semiconductor Memory Design and Applications*. New York: McGraw-Hill, 1973, pp. 164–166.
[88] H. A. Perkins and J. D. Schmidt, "An integrated semiconductor memory system," in *1965 Fall Joint Computer Conf., AFIPS Conf. Proc.*, vol. 27. Montvale, NJ: AFIPS Press, Nov. 1965, pp. 1053–1064.
[89] G. A. Potter, J. Mendelson, and S. Sirkin, "Integrated scratch pads sire a new generation of computers," *Electronics*, vol. 39, pp. 118–126, Apr. 4, 1966.
[90] I. Catt, E. C. Garth, and D. E. Murray, "A high-speed scratchpad memory," in *1966 Fall Joint Computer Conf., AFIPS Conf. Proc.*, vol. 29. Montvale, NJ: AFIPS Press, Nov. 1966, pp. 315–331.
[91] J. E. Iwersen, J. H. Wuorinen, B. T. Murphy, and D. J. D'Stefan, "New implementation of bipolar semiconductor memory," *ISSCC Digest Tech. Papers*, pp. 74–75, Feb. 1967.
[92] J. K. Ayling, R. D. Moore, and G. K. Tu, "A high-performance monolithic store," *ISSCC Digest Tech. Papers*, pp. 36–37, Feb. 1969.
[93] B. Agusta, "A 64-bit planar double-diffused monolithic memory chip," *ISSCC Digest Tech. Papers*, pp. 38–39, Feb. 1969.
[94] A. S. Farber and E. S. Schlig, "A novel high-performance bipolar monolithic memory cell," *IEEE J. Solid-State Circuits*, vol. SC-7, pp. 297–298, Aug. 1972.
[95] F. S. Green, Jr. and N. U. Phan, "Megabit bipolar LSI memory system: IV," *ISSCC Digest Tech. Papers*, pp. 40–41, Feb. 1970.
[96] D. J. Lynes and D. A. Hodges, "Memory using diode-coupled bipolar transistor cells," *IEEE J. Solid-State Circuits*, vol. SC-5, pp. 186–191, Oct. 1970.
[97] K. Tangiguchi, A. Hotta, and I. Imaizumi, "Switched collector impedance memory," *IEEE J. Solid-State Circuits*, vol. SC-6, pp. 289–296, Oct. 1971.
[98] F. Tsang, "A 1024-bit bipolar RAM," *ISSCC Digest Tech. Papers*, pp. 200–201, Feb. 1974.
[99] H. Magumi, J. Nokubo, and K. Okada, "A 25 ns read-access bipolar 1K/bit TTL RAM," *ISSCC Digest Tech. Papers*, pp. 208–209, Feb. 1974.
[100] R. Rathbone, H. Ernst, H. Glock, and U. Schwabe, "A 1024-bit ECL RAM with 15 ns access time," *ISSCC Digest Tech. Papers*, pp. 188–189, Feb. 1976.
[101] S. K. Wiedmann and H. H. Berger, "Small-size low-power bipolar memory cell," *IEEE J. Solid-State Circuits*, vol. SC-6, pp. 283–288, Oct. 1971.
[102] S. K. Wiedmann, "Injection-coupled memory, a high-density static bipolar memory," *IEEE J. Solid-State Circuits*, vol. SC-8, pp. 332–337, Oct. 1973.
[103] "Cray-1: the smaller supercomputer," *Computer*, p. 53, Mar. 1976.

Robert A. Henle (A'52–M'57–F'66) received the M.S.E.E. degree from the University of Minnesota, Minneapolis, in 1951. He joined IBM and worked on the development of transistor circuits and digital and analog applications. He was named an IBM Fellow in 1964, and since then he has been concerned mainly with the application of LSI technology to computers. He is presently manager of advanced technology in the IBM System Products Division, East Fishkill, Hopewell Junction, NY.

Irving T. Ho received the B.S.E.E. degree from National Amoy University, Fukien, China, and the M.S. and Ph.D. degrees in 1957 and 1961, respectively, from Stanford University, Stanford, CA.

His experience includes being a senior research engineer and group leader at the Boeing Company, Seattle, WA, in microwave transmission and detection; research associate at Stanford, in parametric device and plasma research; and consultant to Fairchild Semiconductors. Since joining IBM in 1963, he has worked in the exploratory development of semiconductor memories, computer circuitries and subsystems, array logic, and integrated circuit interconnections. He has received ten IBM invention achievement awards. He has been a senior engineer since 1969. He served as a visiting lecturer at Stanford (summer 1974) and, while on leave from IBM System Products Division, East Fishkill, Hopewell Junction, NY, was the Far East Chair Professor at National Taiwan University, Taipei, in the academic year 1974–1975.

Dr. Ho is a member of Sigma Xi.



William S. Johnson (S'63–M'68) received the B.S. degree in electrical engineering with high honors in 1964 from Swarthmore College, Swarthmore, PA, and the M.S. and Ph.D. degrees in electrical engineering in 1965 and 1969, respectively, from Stanford University, Stanford, CA.

His research at Stanford was among the earliest work in ion implantation and resulted in a thesis on ion damage phenomena and a book, *Projected Range Statistics*. After one year as a research associate at Stanford, he joined the IBM System Products Division, East Fishkill, Hopewell Juntion, NY, in 1970. His work at IBM, in development and management, included application of ion implantation to many types of advanced semiconductor devices and integrated circuits, particularly MOSFET's. He joined Motorola, Inc., Phoenix, AZ, in July 1976.

W. David Pricer (S'59–M'60–SM'71) received the B.A. degree from Middlebury College, Middlebury, VT, and the B.S. and M.S. degrees from the Massachusetts Institute of Technology, Cambridge, both in 1959.

Upon graduation he joined IBM and has worked in tunnel diodes, thin-film and ferrite memories, read-only control stores, and computer switching. As a senior engineer at the IBM System Products Division, Burlington, VT, he is manager of an advanced technology group working in semiconductor memories.

Mr. Pricer is a member of Sigma Xi.

James L. Walsh (A'54–M'60–SM'70) received the B.S.E.E. degree from the University of Rhode Island, Kingston, 1949, and the M.A. degree in physics from Hofstra University, Hempstead, NY, in 1952.

He joined IBM System Products Division, East Fishkill, Hopewell Junction, NY, in 1952 and is now a senior engineer and manager of exploratory devices and circuits.