

◆ Key Steps to the Integrated Circuit

*C. Mark Melliar-Smith, Douglas E. Haggan,
and William W. Troutman*

This paper traces the key steps that led to the invention of the integrated circuit (IC). The first part of this paper reviews the steady improvements in the performance and fabrication of single transistors in the decade after the Bell Labs breakthrough work in 1947. It sketches the various developments needed to produce a practical IC. Some of the discoveries and developments discussed in the previous paper (“The Foundation of the Silicon Age” by Ian M. Ross) are briefly reviewed here to show how they fit on the critical path to the invention of the IC. In addition, the more advanced processes such as diffusion, oxide masking, photolithography, and epitaxy, which culminated in the planar process, are summarized. The early growth of the IC business is touched upon, along with a brief statement on the future limits of silicon IC technology. The second part of this paper sketches the various problems associated with the quality and reliability of this technology. The highlights of the semiconductor reliability story are reviewed from the early days of germanium and silicon transistors to the current metal-oxide semiconductor IC products. Also described are some of the process, packaging, and alpha particle problems that were encountered and solved before arriving at today’s semiconductor products.

Introduction

Improvements in the performance and fabrication of single transistors occurred steadily in the decade after the Bell Labs breakthrough work in 1947. Looking back over the various improvements, it is possible to sketch the development path to the integrated circuit (IC). The following set of discoveries illustrates the pattern of problem-solving progress that ultimately led to the first practical, cost-effective ICs.

The nurturing environment needed to produce these discoveries first existed at Bell Labs. However, with the growth of the semiconductor industry in California, a similar environment grew quickly there as well. Furthermore, the concept of starting a company to reap the rewards of a new or improved technology emerged after World War II. That concept was honed to a razor’s edge through the 1950s and 1960s and continues even to the present. Entrepreneurial opportunities for creating wealth through company start-ups fueled many technolog-

ical innovations of the late fifties and early sixties.

Some of the discoveries discussed in the previous paper (see “The Foundation of the Silicon Age” by Ian M. Ross) are briefly mentioned here to show how they fit on the critical path to the invention of the IC. For example, Pfann’s zone refining process,^{1,2} as modified and adapted for silicon by Theuerer,^{3,4} provided the pure, low-defect, raw material needed for multi-transistor devices. In the same way, the additional processing steps needed to produce an IC were drawn mostly from the technology base developed to produce low-cost, high-performance, reliable transistors.

Putting the Key Technologies in Place

The key process steps that led to the development of the IC are summarized below.

Diffusion of Donor and Acceptor Atoms

Early in the evolution of the technology that enabled the transistor to be realized, it became clear

that the base layer had to be thin to achieve acceptable speed. A base thickness of about 10 microns was needed to yield a frequency response approaching 10 MHz. For a higher frequency, the base needed to be proportionally thinner.

In 1952, C. S. Fuller published studies⁵ showing that dopants could be introduced in a controlled manner to very shallow depths by diffusion of donor and acceptor atoms. The method included surrounding the semiconductor with a vapor containing the desired dopants at an elevated temperature, which enhanced the diffusion of the dopants into the semiconductor surface. The temperature could be chosen to create diffusion rates that would provide precise control of the depth of penetration. In 1954, using this diffusion process, C. A. Lee made the first diffused-base germanium mesa transistor.⁶ This device had a cutoff frequency of 500 MHz, a factor of ten faster than the best alloy transistors of the time.

Oxide Masking, Photolithography, and Epitaxy

In 1955, C. J. Frosh and L. Derick made an important contribution by noting that silicon dioxide could act as a diffusion mask.^{7,8} A few-thousand-angstrom layer of silicon dioxide on the surface of a silicon wafer could mask the diffusion of certain donor and acceptor atoms into the silicon. Frosh and Derick further demonstrated that diffusion would occur unimpeded through windows etched in the oxide layer. Somewhat later, J. Andrus and W. L. Bond showed that a specific photoresist deposited on the oxide surface would prevent etching of the oxide.^{9,10} Hence, optical exposure of the resist through photo masks created precise window patterns in the oxide, which in turn provided control over areas in which diffusion would occur. In 1955, M. Tanenbaum and D. E. Thomas developed the all-diffused silicon transistor¹¹ based on diffusing the impurity atoms into the silicon wafer and protecting the active elements with an oxide barrier. The announcement of these results created great excitement throughout the infant semiconductor industry.

The speed of a diffused-base mesa transistor was limited by the series resistance of the collector region. Yet high resistance was known to be essential to provide a suitable breakdown voltage for the transistor.

Panel 1. Acronyms, Abbreviations, and Terms

DRAM—dynamic random access memory
DTL—diode transistor logic
ECL—emitter coupled logic
FET—field-effect transistor
IC—integrated circuit
IEEE—Institute of Electrical and Electronics Engineers
KS—Kearney specification
LSI—large scale integration
MOS—metal-oxide semiconductor
MOSFET—metal-oxide semiconductor field-effect transistor
NMOS—MOS with n-type transistors
p-n—positive-negative (junction)
RTL—resistor transistor logic
SRAM—static random access memory
TTL—transistor transistor logic
VLSI—very large scale integration

Furthermore, because the collector contact for the mesa transistor was made through the bottom of the die, the full thickness of the high-resistance wafer was in series with the relatively narrow active collector region.

The ability to make the wafer thinner to reduce the resistance was limited by what could be achieved in manufacturing. The thinnest possible wafer that was not too fragile for manufacturing was on the order of 200 microns thick; for an ideal collector resistance, the thickness would need to be one tenth of that. A thin, lightly doped, high-resistance collector in series with a highly doped, low-resistance substrate was needed. A group led by Ian Ross at Bell Labs initiated the development of *epitaxial* technology. The work of this group built on earlier work by Teal and Christensen, who had worked on germanium epitaxy in the 1940s.⁸

Epitaxy comes from the Greek words *epi*, meaning “upon,” and *taxis*, meaning “ordered.” The object of the epitaxial process was to grow a thin semiconductor film with a light impurity concentration on a wafer of much higher doping concentration. This approach allowed the series resistance of the substrate to be reduced while maintaining the desired high-resistivity characteristics of the collector. At the Solid State Device Conference in 1960, Theuerer, Kleinack, Loar,

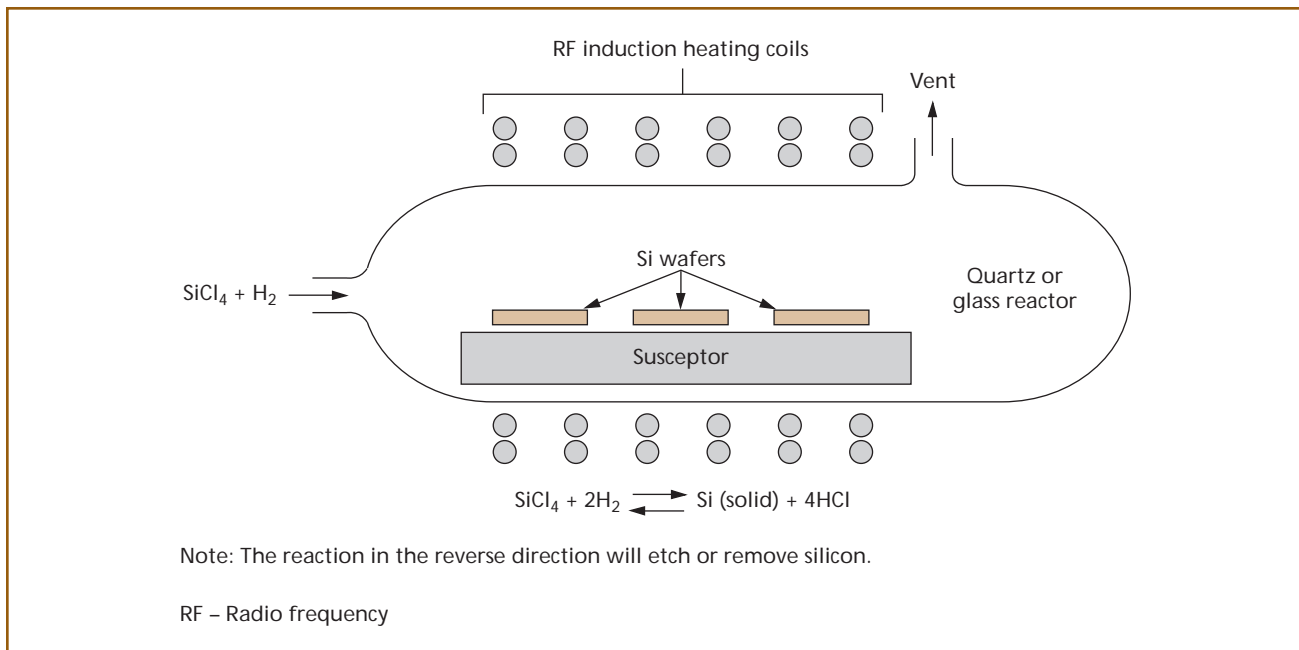


Figure 1.
The epitaxial process.

and Christensen described the first practical transistor having an epitaxial layer.⁸

The process was performed in a quartz reactor able to be heated by radio-frequency induction. For silicon transistors, a high-temperature gas, such as silicon tetrachloride, was flowed past heated wafers, thus causing the atoms to have sufficient mobility to orient themselves to the crystal structure of the silicon wafer. Hydrides of the impurity atoms were added to the silicon tetrachloride transport gas to supply the desired impurity atoms (see **Figure 1**). Today, this process is called *chemical vapor deposition*, and is used extensively in device fabrication to achieve abrupt changes in doping concentration.

Shockley Goes West

William Shockley and Stanley Morgan led the solid-state research group formed at Bell Labs in 1945 by Mervin Kelly. In 1955, Shockley left Bell Labs for Palo Alto, California, to form Shockley Semiconductors. He quickly assembled a strong group of young scientists and engineers and set about developing silicon-based semiconductor components.

In 1956, Shockley shared the Nobel Prize in physics with Walter Brattain and John Bardeen.

Armed with this recognition, Shockley led the engineers at the new company in the direction of a pet project—namely, a four-layer silicon diode. However, the team soon became restless and, in 1957, eight key members, led by Robert Noyce, left Shockley Semiconductors to form Fairchild Semiconductors. Using the diffusion techniques developed at Bell Labs, Fairchild concentrated on high-performance mesa transistors. This small company was fortunate in that it landed an order from IBM to produce diffused mesa transistors for certain computer products. It is interesting to note that after being in business for only two years, Fairchild was a profitable company, with revenues over half a million dollars.

The Planar Process

By the second half of the 1950s, three major methods of transistor fabrication existed: the *grown junction*, the *alloy junction*, and *diffusion* (see **Figure 2**). By that time, the grown-junction process was nearly obsolete and was being replaced by the alloy junction and the mesa diffused junction processes. Bell Labs had invented the double-diffused mesa transistor based on the diffusion process. This device was made by growing a lightly doped epitaxial collector layer on

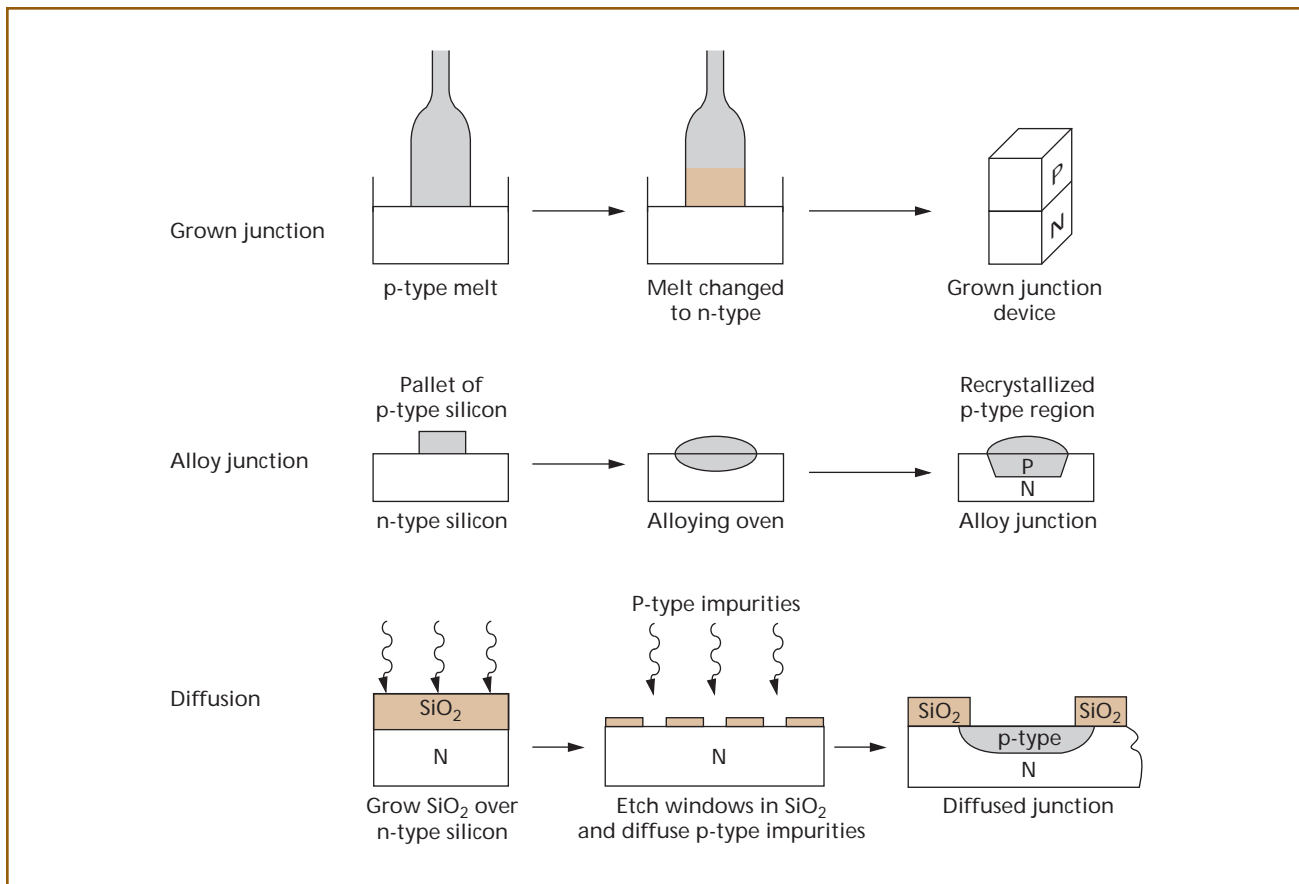


Figure 2.
Discrete transistor fabrication.

top of a highly doped wafer. Successive oxide masking and diffusion steps were used to diffuse the base and emitter regions. A layer of wax or other inert material was put over the active base and emitter areas, and the surrounding silicon was etched away using a strong acid. The result was a raised *mesa*, or broad, flat-topped elevation, containing the base and emitter junctions over a lightly doped collector. Wires were then attached to the base and emitter areas and the collector was contacted through the bottom of the die (see **Figure 3**). This technique produced very-high-performance transistors.

Both Texas Instruments and Fairchild used mesa techniques to fabricate their high-performance, commercially successful transistors in the late 1950s. However, the mesa process had a serious shortcoming in that the junctions were all left exposed. Dust, various process residues, and human oils (fingerprints) all were able to easily contaminate the exposed junctions.

The exposed junctions were not only a reliability problem, but also a source of production rejects or manufacturing yield loss. It was in this environment that Jean Hoerni of Fairchild proposed a unique solution. Instead of etching away the silicon and producing the mesa areas with their exposed junctions, silicon dioxide was used to cover the junctions.

Hoerni's solution, known as the *planar process*, which is still the state of the art in IC production today, exploited the repetition of three basic steps. First, a layer of silicon dioxide was grown over the whole silicon wafer. Second, the oxide-coated wafer was patterned using photoresist techniques and open windows were etched in the oxide. Third, impurities were diffused into the underlying silicon through the window areas opened in the oxide (see **Figure 4**). Also, as shown in step 4 of Figure 4, impurities were diffused on top of each other to create layers of differing semiconductor type. The final step, step 5, shows

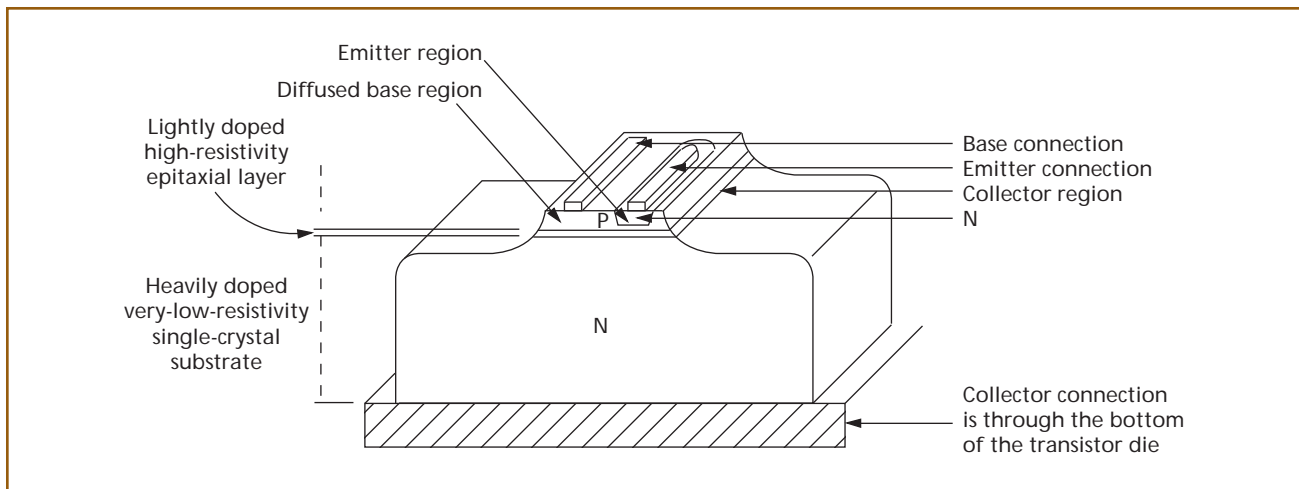


Figure 3.
Epitaxial diffused mesa transistor.

the diffused base and emitter areas of the bipolar transistor formed over the buried collector. Electrical contacts were applied to the base and emitter areas from the top. Gold or aluminum metalization evaporated on the surface of the transistor over silicon dioxide was patterned using a mask and photoresist, followed by etching to make the fine base and emitter connection leads. The collector was simply contacted from the bottom of the silicon, usually by mounting the transistor die on a metallic header.

The planar process, though similar to the mesa process, fundamentally differed in that the device was flat. This attribute meant that the electrical connections could be made by depositing an evaporated metal film and patterning that metal using a photolithographic etching process similar to that used to pattern the oxide. Furthermore, the engineers at Fairchild found that leaving the final oxide mask on the device protected the sensitive junctions from many of the contaminants that plagued the quality and reliability of the mesa transistor.

The planar process was indeed a major breakthrough even though it incorporated no new process steps. All the process steps and procedures had been previously developed at Bell Labs. Initially, the performance of planar transistors fell short of the performance of mesa transistors. However, in time, the high yield, low cost, and reliability of planar transistors made up for the slight loss in performance. Over time, as lithography improved, the planar transistor was able to surpass the mesa transistor with respect to performance, also.

Putting The Pieces Together

Early in 1958, Jack Kilby left Centralab's semiconductor operation to join Texas Instruments in Dallas. At that time, government-funded, electronic research programs were focused heavily on miniaturization. This emphasis was driven by the Russians' stunning success on October 5, 1957, with Sputnik I. The first Russian satellites orbiting Earth and the large booster rockets they possessed galvanized the United States' military efforts for miniature electronic circuits.

Kilby's view of all miniaturization projects he had seen was that they were "kluges" that never addressed the real problems. In July of 1958, he wrote in his lab notebook: "Extreme miniaturization of many electric circuits could be achieved by making resistors, capacitors, transistors, and diodes on a single slice of silicon."¹² He then proceeded to show in the next five pages of his notebook how this could be done. By the end of August, a simplified version of his circuit had been built using wire bonds for interconnection (see **Figure 5**). It could be said that this was an idea whose time had come because, in January of 1959, ideas about another version of an IC were proposed by Robert Noyce at Fairchild.

The idea came to Noyce as he was considering the ramifications of a new transistor processing technology under development at Fairchild. To improve on the Bell Labs mesa transistor, Hoerni was developing the planar process. By being able to bury the active ele-

ments in silicon dioxide using the planar process, Noyce was able to focus on the key issue of *interconnection*. The Kilby circuit required the application of wires to make circuit connections that, unfortunately, required more space than the circuits themselves. Noyce's idea was to use aluminum or gold metal that could be photo defined and etched as a step in the batch processing of the semiconductor device (see **Figure 6**). The transistor had become the integrated circuit.

Litigation between Fairchild and Texas Instruments over IC patents lasted into the mid-1960s. Ultimately, a compromise was reached. The settlement recognized both Noyce and Kilby as inventors sharing in the rights, with each having title to different aspects of the invention. In 1961, just fourteen years after the invention of the transistor, the semiconductor industry surpassed \$1B in revenue.

Toward Large Scale Integration

Initially, the IC was viewed as much too costly to be practical for applications other than those placing the highest premium on miniaturization. The high cost was based on very low transistor manufacturing yield. It was not uncommon to have single transistor yields as low as 20%. Various developers asked the following question: What would be the yield if an IC having one hundred or even one thousand transistors were fabricated as a single die? Reliability was also a concern. It was argued that the reliability of an IC would approximate the reliability of a discrete transistor proportionally degraded by the power of the number of transistors. Both arguments assumed that the yields and failure rates of ICs would be governed solely by random events, which turned out not to be the case. The reliability issue plagued the industry for years and imposed costly hermetic restrictions on device packaging. Improvements in these areas are highlighted later in this paper.

At Bell Labs, M. M. Atalla had a group that studied the surface properties of silicon in the presence of a silicon dioxide layer. They suspected that growing an oxide layer under very clean and controlled circumstances might lead to a reduced density of surface states and also serve to protect the surface against sub-

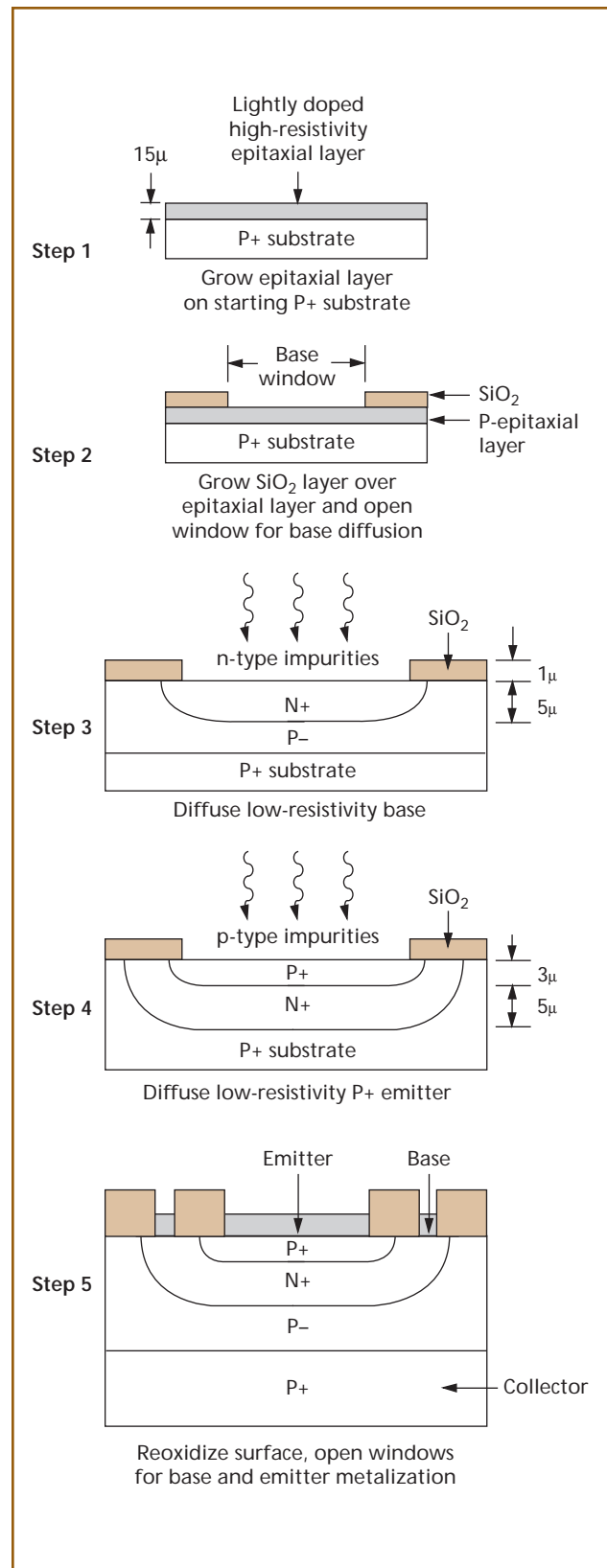


Figure 4. The planar process.

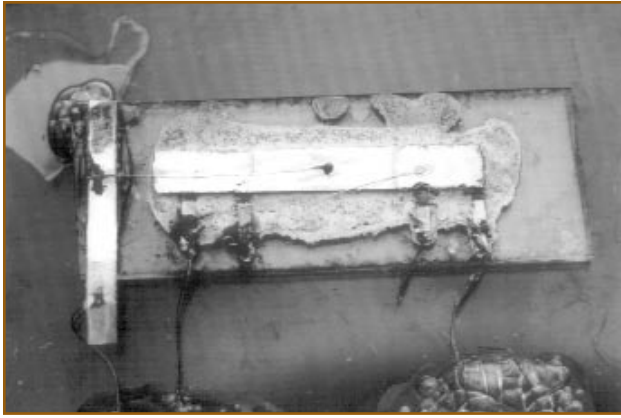


Figure 5.
Jack Kilby's first integrated circuit.

sequent contamination. Data confirming this hypothesis was obtained as early as 1959.¹³ However, the process was difficult to control and not very reproducible. In 1960, Hoerni, at Fairchild, showed that simply leaving the oxide diffusion mask layer in place produced a startling improvement in reliability without any further treatment.^{14,15,16}

Other investigators made incremental reliability improvements over the next few years, with the most fundamental advance coming from J. V. Dalton at Bell Labs in 1966.¹⁷ He showed that an overcoating of silicon nitride would provide an effective seal against alkali ions, mainly sodium. Silicon nitride provided a tough surface passivation and paved the way for the era of low-cost plastic packages with acceptable reliability.

The work by Atalla in 1959 on surface states led to the fabrication of a new field-effect device.^{18,19,20} Atalla and Kahng applied voltage to a metal contact overlaying an oxide layer grown on the silicon surface. The density of surface states was sufficiently low that they were able to create a substantial inversion layer in the underlying silicon. This was the first operation of a metal-oxide semiconductor (MOS) transistor. However, the era of the field-effect transistor (FET) would not come until much later. In the early 1960s, MOSFETs could not compete with bipolar transistors in performance. Initially, MOSFET transistors were used only in limited applications that could take advantage of their unique electrical characteristics.

It took several years to show that the feared problems of yield and reliability caused by random defects did not cause ICs to be priced too high to be suitable for

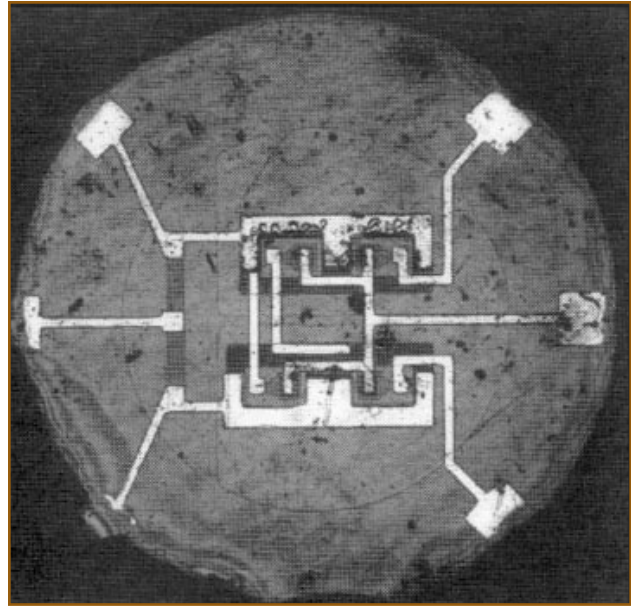


Figure 6.
Early Fairchild resistor transistor logic integrated circuit.

high-volume electronic systems. As it turned out, neither yield nor reliability was dominated by random events. In the case of yield, defects tended to congregate, and large areas of a wafer contained effectively 100% perfect devices. If the area of the IC was small compared to the area of good material on the wafer, the yield would have little dependence on the size of the IC.

Growing The IC Business

By the mid-1960s, all the transistor principles and processes were in place to produce ICs. From this time on, it took much ingenuity, effort, and investment to apply and refine these principles and processes, but no major breakthroughs were needed. Jack Kilby, speaking about the buildup of IC production, said in 1976: "It should be noted that one of the strengths of the integrated circuit concept has always been that it could draw on the mainstream efforts of the semiconductor industry."²¹ IC fabrication adapted all the techniques of crystal growing, epitaxy, and lithography from transistor fabrication.

However, one more critical and interesting choice remained—transistor type. Many developers favored the planar transistor because of its high current gain and potential for high performance. This issue was not immediately settled. The early winner appeared to be the bipolar transistor. Many bipolar logic families were

rapidly developed: resistor transistor logic (RTL), then diode transistor logic (DTL), followed by emitter coupled logic (ECL), and finally transistor transistor logic (TTL). TTL went through several modifications driven by the need to improve speed and lower power dissipation. However, the advent of semiconductor memory devices in the early 1970s intensified efforts to find a more appropriate transistor.

Memory devices tended to be much more complex than the early logic gates and registers used by the computer industry. Storage devices needed a transistor that could be reliably produced with high yield from a simple process.

The showdown came with the 1K static random access memory (SRAM). An early, high-speed 1K bipolar SRAM from Fairchild called the 93415 found competition from an Intel pin-compatible device called the 2115, which had been built using an NMOS process. (NMOS is a MOS using n-type transistors.) Intel's 2115 demonstrated that a MOS device that had scaled-down linear dimensions could compete in performance with a bipolar device.

The simpler and smaller MOS transistor was easier to fabricate using leading-edge design rules. Therefore, performance deficiencies were minimized and, more importantly, a large number of MOS transistors could be fabricated on a chip of a given size.

The dynamic random access memory (DRAM), starting in the early 1970s with the first high-volume device, the Intel 1103, introduced the process of scaling successive generations of the MOS transistor's physical dimensions. DRAMs acted as the cutting edge for finer lithography and improvements in processing technologies, particularly etching. In addition, the physical size reduction needed for DRAMs became the process driver for other circuits, notably microprocessors, starting with the Intel 4004 (see **Figure 7**).

One observation from scaling MOS transistors and increasing levels of integration was captured by Gordon Moore of Intel and later became known as Moore's Law. In a 1965 edition of *Electronics*, Moore published a semi-log plot of the number of transistors on a silicon chip as a function of the date of first chip availability. The result was a straight line almost doubling nearly every year. In later years, the rate relaxed

to doubling in 18 to 24 months. This pace continues today and is expected to continue well into the twenty-first century.

The technology undergirding Moore's Law has had a major influence on the electronics industry. This unrelenting, continuous improvement in functional density with lower cost per function provided management with a simple yardstick against which to allocate a variety of resources. It provided the rhythm that set the timing of the development processes, the introduction of new products, and the erection of fabrication facilities.

Projecting forward from today using Moore's Law, the Semiconductor Industry Association suggests we will reach some physical limits around the year 2007, when the transistor will be 60 years old. At that time, the minimum physical transistor dimensions will be somewhat below 0.1 micron and chips might be expected to have on the order of 20 billion transistors. The chip size could be as large as 10 square centimeters in area and will be fabricated on a wafer that will be 12 inches or more in diameter.

Ensuring Quality and Reliability

Mervin Kelly had reliability concerns on his mind when he formed the solid-state research group at Bell Labs in 1945. He strongly felt that the poor reliability of the vacuum tube would be a limiting factor in the growth of telecommunications systems. Kelly's vision would take a number of years and much hard work after the transistor invention to be realized.

Quality and reliability were vexing problems with transistors in their early days. Point contact transistors were inherently unstable electrically and prone to failure. The first junction transistors had much more stable electrical characteristics than point contact transistors but their reliability was no better. The twin problems of *quality*, the ability to consistently meet a specification, and *reliability*, the ability to meet that specification over the useful life of the system, demanded the allocation of significant resources from the new industry. Surprisingly, these resources, in terms of engineering time and specialized equipment, often came from the users of the new technology rather than the makers of the technology. It was

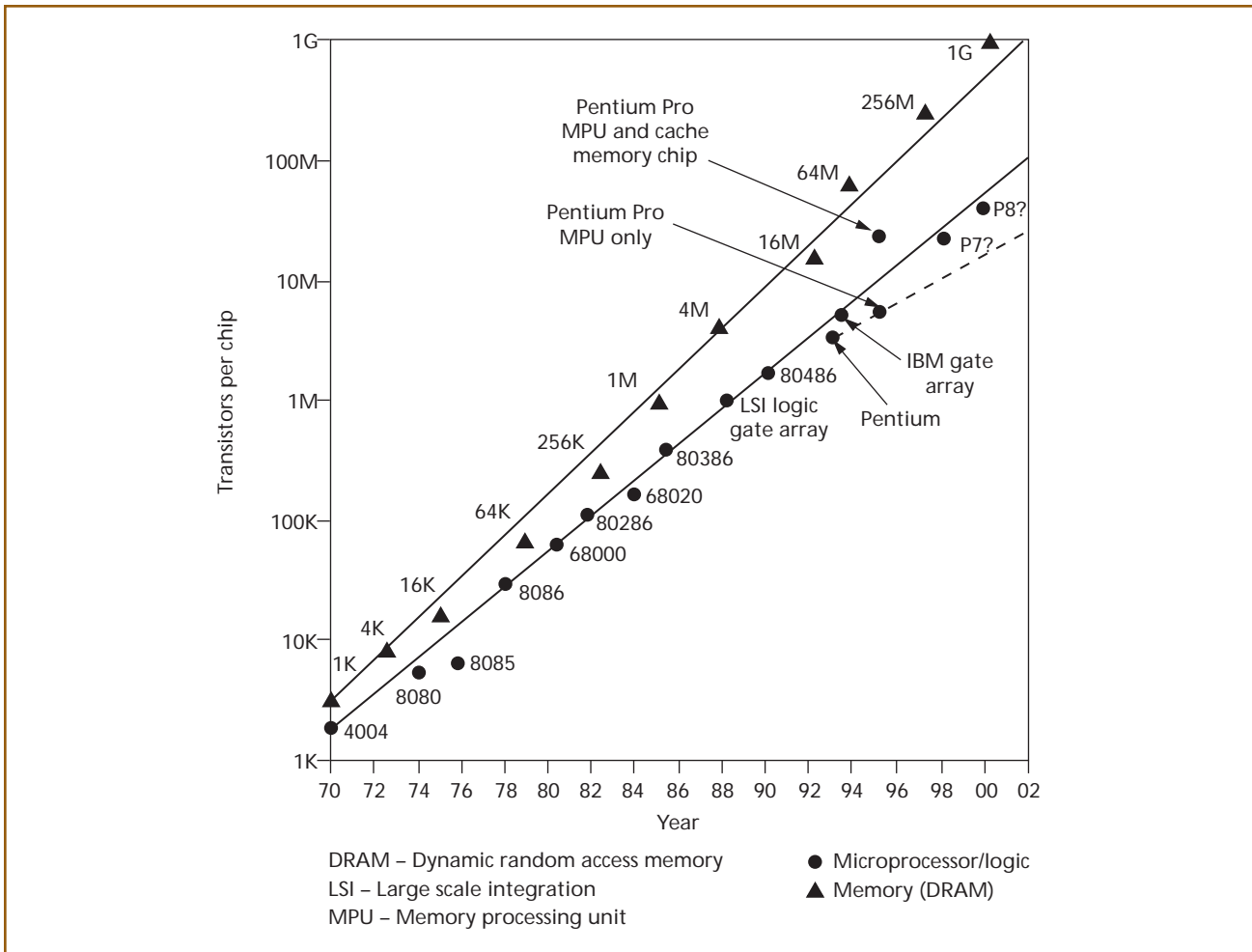


Figure 7. DRAM and microprocessor roadmap.

almost a “buyer beware” relationship, with the manufacturer of semiconductors heralding its new device or fabrication process, and the user being left to find the technology’s weakness. It was not uncommon for a system builder to re-test or grade transistors received from a semiconductor supplier to ensure their quality. Furthermore, as many of the generic semiconductor reliability problems were found by the users as were found by the manufacturers. For this reason, many of the reliability qualification tests for devices were the result of users of the devices developing techniques to ensure the systems built with those devices would, in turn, meet their own customers’ expectations.

In all the early transistors, the surface of the semiconductor was completely exposed. Airborne dust, salts from human handling, and residual processing

materials were contaminants that reduced their useful life. Many of these contaminants became chemically active in the presence of moisture, which necessitated the use of costly hermetic packages.

The following are examples of some of the quality and reliability problems encountered and solved over the years by both the manufacturers and the users. This list is by no means complete, but rather an attempt to select examples of the more serious problems that, when they were solved, allowed the industry to move to higher levels of both quality and reliability.

The Early Years: Germanium

Some of the earliest heating experiments with germanium crystals at Bell Labs produced unexplained, gross changes in resistivity that, because they were not

understood, were attributed to “thermium.” A related unexplained problem that affected minority carrier lifetime was attributed to “deathnium.” Later, these twin problems were traced to minute amounts of copper from the surface of the germanium.²² Because thermium and deathnium were related to heating, the degradation they caused prevented the production of high-frequency diffused germanium devices for several years.

This early work made clear the necessity for ultra-pure materials fabricated in ultra-clean environments for high-temperature processes to be successful. The need for cleanliness and purity on a scale that had never before been seen in routine manufacturing would become a dominant characteristic of the semiconductor industry.

Operational instability in early alloy junction transistors was eventually traced to small amounts of ionic impurities coupled with moisture on the transistor surface.²³ The solution to this problem was to package the devices in a hermetic enclosure. The packaging process included flushing the transistor repeatedly with a dry gas and then evacuating the interior before hermetic sealing. These procedures were identical to those needed for vacuum tubes and were a grave disappointment to transistor manufacturers because they added significant cost. There never would be a low-cost, non-hermetic, reliable package for germanium transistors as there eventually was for silicon devices.

One of the early challenges for the industry was to determine how to measure properly a transistor’s reliability without testing it to the point that it wore out. The device chosen as a test vehicle for learning more about reliability measurement at Bell Labs was the carefully produced germanium 2N559 diffused-base transistor used in the Nike Zeus program. It was made at the Laureldale shop of Western Electric, which later became the Western Electric Reading Works. B. T. Howard and G. A. Dodson were faced with a product that had a failure rate less than 0.01% per 1000 hours. Their task was to determine how much less than 0.01% and to establish reliable proof.

They devised a method of step-stressing a limited sample of devices for fixed time intervals. The temperature was stepped up until all the devices failed. From

the failures, Howard and Dodson were able to derive activation energy values for the various failure mechanisms found. Using the activation energies, they were then able to predict the life of the devices at the actual temperature of use.²⁴ After refinement, this became the standard method for determining reliability for the industry. It should be noted that D. S. Peck, using this technique, found the 2N559 to have an actual failure rate of 0.001% per 1000 hours or 10 failures per 10⁹ hours. This was quite respectable, even for our times, considering the year was 1957.

The Early Silicon Age

A particularly troublesome failure mechanism associated with early silicon bipolar transistors was called *channels*. These were surface inversion layers caused by mobile ions on the surface of the transistor that collected over the base region and had the effect of producing collector-to-emitter shorts. M. M. Atalla led a group at Bell Labs that was studying surface properties and related oxide growth. These studies showed that thermally grown oxides over clean silicon surfaces had the effect of providing surface passivation for the transistor.¹³ Thus, a p-n junction that emerged at the surface of a semiconductor could be protected from channels by a covering oxide. This was an important discovery that was later incorporated in the planar process. It was this process, developed at Fairchild, that made ICs practical. Years later, after much refinement by many contributors, the use of oxide and oxide nitride surface passivation processes made reliable plastic packages practical for most silicon technologies, covering a range of device applications.

With the ramp-up in the use of NMOS ICs beginning in the 1970s, the standard for process cleanliness was raised to new heights. This included not only the MOS devices’ higher sensitivity to dust and dirt caused by the finer lithography and larger die, but also their much higher sensitivity to chemical contaminants. For early MOS transistors, one of the “killer” problems was device threshold shift, which could occur under normal conditions. The problem was found to be activated by heat and quickly showed up under elevated life tests. Therefore, temperature-bias stress tests were developed to qualify process changes and used as periodic reliability monitors to ensure that processing

equipment had not become dirty and thus contaminated MOS wafer lots in process.

The cause of device threshold shifts was identified as mobile alkali ions such as sodium, potassium, and lithium. Sodium was a particular problem because it was so common and only minute amounts could produce unwanted threshold shifts. Capacitance-voltage tests were developed as process monitors to test for the problem at the wafer level. Ultimately, gettering techniques such as adding a few percent of phosphorus to oxides brought the problem down to very low levels.²⁵

Packaging Problems

Over the years, many reliability problems were directly or indirectly traced to device packaging. An early problem was called “purple plague,”²⁶ which caused gold wire bond failures in devices during operational life. This was a widespread industry problem resulting in elaborate screening techniques being developed to weed out weak devices and guarantee that a population of transistors (and later, ICs) did not have the problem.

The name purple plague came from the mottled purple color of the gold wire bond material at the failure site. Subsequent studies showed the failure mechanism to be the result of the Kirkendall effect, the diffusion of vacancies through a metallurgical system. These vacancies sometimes aggregated, causing local heating and fusing or voids in the gold metal wire. Elaborate procedures were put in place to minimize the probability of the occurrence of purple plague, along with various quality-control screens on products to verify reliability of wire bond connections. Ultimately, the industry used aluminum wire to bond to aluminum metalization on the device, which eliminated the gold-aluminum intermetallic materials and solved the problem.

Prior to the widespread use of plastic packages, numerous problems existed with the hermetic cavity packages required for reliable telephone, computer, and military applications. One problem was related to the entrapped moisture inside the package that, under the right conditions, condensed on the surface of the semiconductor and caused electrolytic corrosion of the metalization. A parallel failure mechanism was caused by the low quality of the package seal, which allowed

moisture to enter the cavity and produce corrosion. In short, leaks in the sealing of the package or that developed during the use of the device allowed contaminants and moisture to enter the cavity and caused the device to fail.

Procedures were developed to monitor water vapor content inside the package and to measure package leaks, which, together with improved manufacturing processes, reduced this failure mechanism to an acceptable level. In addition to unwanted moisture inside the package, unwanted particles were inside the package cavity that, if they were conductive and large enough, caused electrical shorts. To test for loose particles inside a small transistor or IC package, vibration and acoustic detection methods were developed. It was in quality and reliability areas such as these that military specifications began to play a useful role.

Military specifications evolved into documents such as Military Standard 883 that defined the semiconductor device reliability tests in standard terms.^{27,28} A user or maker of devices could set up a test apparatus and correlate the results. Even for systems not intended for the military, the applicable device quality and reliability characteristics could be selected from a military standard. These characteristics could then be applied to the system semiconductor component specifications, to the extent appropriate, by reliability engineers. In this way, semiconductor components with quality and reliability features suitable for their application could be specified for both telephone central office equipment and mainframe computers.

Bell Labs went a step further with specifications when they developed the “KS” specifications. The origin of the term *KS* is not clear but some have proposed that it meant “Kearny specification,” after the Western Electric plant in Kearny, New Jersey. Whatever “KS” stood for, it was a unique form of specification in that, rather than defining a result, it defined a manufacturing process and sequence. Since most components used by Western Electric in those days were made by Western Electric, only a few preferred suppliers produced KS parts. At first, because of the volume of parts used, suppliers were willing to conform to the KS requirements. However, in time, the KS semiconductor specifications became unattractive to many suppli-

ers simply because they defined the manufacturing process rather than the desired attributes of the component. In a fast-moving industry such as the semiconductor industry, manufacturers did not want to be held to obsolete processes that required extensive approval from a customer prior to change. KS specifications worked best in the early days of semiconductor manufacturing, when most of the best process steps were developed by Bell Labs.

DRAMs Become the Quality and Reliability Drivers

In the early 1970s, the IC industry had advanced to the point where semiconductor memories became a cost-effective alternative to ferrite cores for computer memory systems. The IC memory business grew rapidly and memory devices quickly became the largest segment of the large scale integration (LSI) market. Memory devices, because of their need to use leading-edge fabrication processes and packaging, became the quality driver for the IC industry. Memory devices had all the process and package quality and reliability problems of any LSI component, along with some that were unique to memory, such as data and address pattern sensitivities. These were prevalent in the first-generation 1K DRAM devices and resulted in elaborate test patterns and equipment being designed to uncover design and process weaknesses. The young IC memory and computer industry cut its teeth on the test equipment and qualification procedures needed to produce and consume high-volume DRAM devices.

Alpha Particles. The second-generation 4K DRAM brought a new and totally unexpected problem called a *soft error*. The soft error was often called a failure mechanism but that was not strictly true, because a failure generally implies a permanent change that limits or stops the useful life of the device. This phenomenon produced an error in a particular memory location that might never again occur in that location during the life of the device.

It was originally noticed that large memory systems containing 4K DRAMs were producing random errors at fairly regular intervals. A great deal of checking was performed for system margins, noise conditions, and device application procedures without any definitive results. Finally, Tim May and Murray Woods of Intel hit on the idea that ionizing radiation or alpha particles

from the packaging materials were striking the surface of the memory chip and producing electron/hole pairs that were altering the stored charge in the DRAM cell. The source of alpha particles in the packaging materials was minute amounts of radioactive thorium and uranium. The soft error mechanism had been found.²⁹ Over the next few years, DRAM cell design, packaging materials, and die coating materials were all actively pursued as solutions to the alpha particle problem.

High-Reliability, Low-Cost Packaging. It was in this environment of alpha particles and packaging materials that DRAMs became the driver for one of the major packaging changes in the IC industry. For reasons of cost as well as soft errors, there was a need to develop high-quality, low-cost, plastic packages. Plastic was not new to the IC business, but up to this time it had only been used in low-cost, small- and medium-scale integration TTL bipolar logic circuits. The nature of those TTL applications was far more benign than that of MOS memory devices.

The use of plastic for MOS DRAM memory did not come without significant development work on both plastic package failure mechanisms and packaging quality assurance procedures. Methods were needed that would accelerate device failures with known acceleration factors. Out of this need for rapid evaluation techniques, the industry developed pressure-cooker (100°C, 100% relative humidity, and elevated pressure) and 85/85 (85% relative humidity and 85°C) life tests.

Procedures to determine the acceleration factors for both pressure-cooker and 85/85 life tests were heatedly debated in the industry. During this time, the IEEE Reliability Physics Symposium was the scene of numerous debates over the relative merits of tests, test results, and their relevance to actual use conditions. However, this debate and the work that evolved set the stage for the surface passivation, pure plastic materials, packaging processes, and reliability tests that we use for all high-quality, high-reliability, low-cost IC packages in the industry today.

Conclusion: The Original Objective

Quality and reliability for semiconductors have come a long way over the past fifty years. However, it

is good to remember that Mervin Kelly, of Bell Labs, who originally formed the solid-state research group in 1945 that led to the invention of the transistor, saw the poor reliability of the vacuum tube as a major reason for searching for a solid-state replacement. His vision for a reliable electronic amplifier and switch has been fulfilled beyond anyone's imagination by the VLSI technology of our day.

References

1. W. G. Pfann, "Principles of Zone Melting," *Transactions of the American Institute of Mechanical Engineers*, Vol. 194, July 1952, pp. 747-753.
2. W. G. Pfann, "Process for Controlling Solute Segregation by Zone-Melting," U.S. Patent 2,739,088, filed Nov. 16, 1951, issued Mar. 20, 1956.
3. H. C. Theuerer, "Removal of Boron from Silicon by Hydrogen Water Vapor Treatment," *Transactions of the American Institute of Mechanical Engineers*, Vol. 206, 1956, pp. 1316-1319.
4. H. C. Theuerer, "Method of Processing Semiconductive Materials," U.S. Patent 3,060,123, filed Dec. 17, 1952, issued Oct. 23, 1962.
5. C. S. Fuller, "Diffusion of Donor and Acceptor Elements into Germanium," *Physical Review*, Vol. 86, Second Series, Apr. 1, 1952, pp. 136-137.
6. C. A. Lee, "A High Frequency Diffused Base Germanium Transistor," *Bell System Technical Journal*, Vol. 35, No. 1, Jan. 1956, pp. 23-34.
7. I. Derick and C. J. Frosh, "Oxidation of Semiconductive Surfaces for Controlled Diffusion," U.S. Patent 2,802,760, filed Dec. 2, 1955, issued Aug. 13, 1957.
8. H. C. Theuerer, J. J. Kleimack, H. H. Loar, and H. Christensen, "Epitaxial Diffused Transistors," *Proceedings of the IRE*, Vol. 48, Sept. 1960, pp. 1642-1643.
9. J. Andrus and W. L. Bond, "Photograving in Transistor Fabrication," *Transistor Technology*, Vol. III, edited by F. J. Biondi, D. Van Nostrand, Princeton, New Jersey, 1958, pp. 151-162.
10. J. Andrus, "Fabrication of Semiconductor Devices," U.S. Patent 3,122,817, filed Aug. 15, 1957, issued Mar. 3, 1964.
11. M. Tanenbaum and D. E. Thomas, "Diffused Emitter and Base Silicon Transistors," *Bell System Technical Journal*, Vol. 35, No. 1, Jan. 1956, pp. 1-22.
12. M. Riordan and L. Hoddeson, "The Monolithic Idea," *Crystal Fire*, W. W. Norton and Company, New York, 1997, pp. 254-275.
13. M. M. Atalla, E. Tannenbaum, and E. J. Scheibner, "Stabilization of Silicon Surfaces by Thermally Grown Oxides," *Bell System Technical Journal*, Vol. 38, No. 3, May 1959, pp. 749-783.
14. J. A. Hoerni, "Planar Silicon Diodes and Transistors," *IRE Transactions on Electron Devices*, Vol. 8, No. 2, Mar. 1961, p. 178.
15. J. A. Hoerni, "Planar Silicon Diodes and Transistors," Professional Group on Electron Devices, Washington, D. C., Oct. 1960.
16. J. A. Hoerni, "Method of Manufacturing Semiconductor Devices," U.S. Patent 3,025,589, filed May 1, 1959, issued March 20, 1962.
17. J. V. Dalton and J. Drobek, "Structure and Sodium Migration in Silicon Nitride Films," *Journal of the Electrochemical Society*, Vol. 115, Aug. 1968, pp. 865-868.
18. M. M. Atalla, "Semiconductor Devices Having Dielectric Coatings," U.S. Patent 3,206,670, filed Mar. 8, 1960, issued Sept. 14, 1965.
19. D. Kahng, "Electric Field Controlled Semiconductor Device," U.S. Patent 3,102,230, filed May 31, 1960, issued Aug. 27, 1963.
20. D. Kahng and M. M. Atalla, "Silicon-Silicon Dioxide Field Induced Surface Devices," Solid State Research Conference, Pittsburgh, Pennsylvania, June 1960.
21. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Transactions on Electron Devices*, July 1976, pp. 648-654.
22. R. A. Logan and M. Swartz, "Restoration of Resistivity and Lifetime in Heat Treated Germanium," *Journal of Applied Physics*, Vol. 26, No. 11, Nov. 1955, pp. 1287-1289.
23. *A History of Engineering and Science in the Bell System, Electronics Technology (1925-1975)*, edited by F. M. Smith, AT&T Bell Laboratories, 1985, p. 39.
24. G. A. Dodson and B. T. Howard, "High Stress Aging to Failure of Semiconductor Devices," *Proceedings of the 7th Symposium On Reliability and Quality Control in Electronics*, Jan. 1961, pp. 262-272.
25. E. Snow, A. S. Grove, B. E. Deal, and C. T. Sah, "Ion Transport Phenomenon in Insulated Films," *Journal of Applied Physics*, Vol. 36, No. 5, May 1965, pp. 1664-1673.
26. *A History of Engineering and Science in the Bell System, Electronics Technology (1925-1975)*, p. 67.
27. MIL-STD-19500, "General Specification for Semiconductor Devices."
28. MIL-STD-883, "Test Methods and Procedures for Microelectronics."
29. T. C. May and M. H. Woods, "A New Physical Mechanism for Soft Errors in Dynamic Memories," *Reliability Physics, 16th Annual Proceedings*, Apr. 1978, pp. 33-40.

(Manuscript approved October 1997)

C. MARK MELLIAR-SMITH, who was chief technical officer for Lucent's Microelectronics Group when this paper was developed, is currently President and Chief Operating Officer of SEMATECH. Prior to joining SEMATECH, he also served as executive director of the Integrated Circuits Division of Bell Labs. During his career in telecommunications, he has held a wide variety of assignments, including work in fundamental research, electronic and photonic device development, and manufacturing and business unit management. Dr. Melliar-Smith holds B.S. and Ph.D. degrees in chemistry from the University of Southampton in England and an M.B.A. from Rockhurst College in Kansas City, Missouri.



DOUGLAS E. HAGGAN is a technical manager in the Electronics Technology and Planning Center, Office of the Chief Scientist, in Lucent's Microelectronics Group in Murray Hill, New Jersey. He is responsible for device technology. Mr. Haggan's work experience outside Lucent includes various engineering and management positions. He worked for Burroughs Corporation before joining Western Electric and subsequently Bell Labs. At Bell Labs, he worked as a supervisor in the Integrated Circuit Customer Service Department, in the Microelectronics Account Management organization, and in the Integrated Circuit Reliability Department. Mr. Haggan has a B.S.E.E. from Bucknell University in Lewisburg, Pennsylvania, and he attended graduate school at both the University of Delaware and Rutgers University. He has been active in the IEEE International Test Conference and has received three patents in the area of electronic circuits.



WILLIAM W. TROUTMAN is head of the Technology Analysis Department, Office of the Chief Scientist, in Lucent's Microelectronics Group in Murray Hill, New Jersey. He is responsible for technology and business issues. Upon joining the company, he worked on electromagnetic pulse effects on ground-based missiles and radar at Bell Labs in Whippany. He later moved to Murray Hill and headed a group designing various analog and digital circuits. He also played a key design role in the BellMac-32 microprocessor chip set. Dr. Troutman earned a B.S. in engineering from Pennsylvania State University in University Park, an M.S. in engineering from the Massachusetts Institute of Technology, and a Ph.D., also in engineering, from Cambridge University in England. ♦

