The Role of Fairchild in Silicon Technology in the Early Days of "Silicon Valley"

GORDON E. MOORE, LIFE FELLOW, IEEE

Invited Paper

Fairchild Semiconductor was founded in 1957 by a group originating from Shockley Semiconductor Laboratory, the first organization attempting to exploit silicon transistor technology in the region at the base of the San Francisco peninsula now often referred to as "Silicon Valley." Fairchild produced the first commercial silicon mesa transistors and invented the "planar" process that formed the basis of practical integrated circuits. Several of the key directions in silicon device technology originated at Fairchild Semiconductor Corporation and its successor organization, the Semiconductor Division of Fairchild Camera and Instrument Corporation. This paper describes the author's recollections of some of the related events.

Keywords*—*Fairchild, Hoerni, integrated circuit, Noyce, planar, Shockley, transistor.

I. INTRODUCTION

For several years in the late 1950's and 1960's, Fairchild Semiconductor was an important contributor to the development of silicon device technology and related products. It introduced the first silicon mesa transistor to be made commercially, the first planar transistor, and the first commercial integrated circuit, as well as performed much of the research that has lead to stable interfaces necessary for today's metal–oxide–semiconductor (MOS) transistors. I had the good fortune to be part of this important chapter in semiconductor history and would like to take this opportunity to record some of my recollections.

A group of eight of us founded Fairchild Semiconductor Corporation in September 1957 with the financial support of the Fairchild Camera and Instrument Corporation. The initial goal of the new venture was to develop, manufacture, and sell double-diffused silicon transistors. The idea for the mesa transistor was not new. Tannenbaum and Thomas of Bell Laboratories [1] had demonstrated that double-diffused n-p-n transistors with useful electrical characteristics could be produced by diffusing both base and emitter layers over the entire surface of a silicon wafer. They made contact to the intermediate p-type base layer by alloying stripes of aluminum that had been evaporated onto the silicon surface through the emitter layer, taking advantage of the fact that the regrown silicon from the alloying step was doped with aluminum to make an ohmic contact to the base, but a rectifying junction with the n-type silicon constituted the emitter layer. Individual transistor areas were separated by placing wax dots over a portion of the aluminum base contact and a portion of the exposed emitter region to act as a mask and then etching through the emitter and base diffused layers into the original n-type silicon. This resulted in an array of flat-topped transistors called mesas and, hence, the mesa transistor. Since this was a batch process wherein the entire top surface of a silicon wafer could be processed at the same time to make several hopefully identical structures, it offered the promise of an efficient production process. Our goal was to take this basic idea of a double-diffused mesa transistor and turn it into a product that could be manufactured reproducibly.

The founding group of eight consisted of a metallurgist, S. Roberts; three physicists, J. Hoerni, J. Last, and R. Noyce; an electrical engineer, V. Grinich; an industrial engineer, E. Kleiner; a mechanical engineer, J. Blank; and me, a physical chemist.

We had been recruited, along with others, by W. Shockley, the inventor of the junction transistor, for his 1955 start-up, the Shockley Semiconductor Laboratory, in Palo Alto, CA. Most of the group arrived at Shockley Labs in early 1956. Of the Fairchild founders, only Noyce had previous semiconductor experience, and that had been with germanium devices. In 1955, because of the consent decree the parent organization, AT&T, entered into with the U.S. government, Western Electric was required to license their semiconductor patents, including those originating at Bell Laboratories, on favorable terms. Bell Labs opened their research results to all the Western Electric licensees that were interested and held a conference to review the technology. Shockley Labs had the documents from this "Diffusion Conference," and starting with this and the information in the literature, we set out to try to develop a

Manuscript received September 27, 1997.
The author is with Intel Corporation, Santa Clara, CA 95052-8119 USA.
Publisher Item Identifier S 0018-9219(98)00752-X.

0018–9219/98$10.00 © 1998 IEEE

Fig. 1. A current photograph of the building where the Shockley Semiconductor Laboratory started operation in 1955.

double-diffused silicon transistor. There was a lot to do to develop the technology to allow such a device to be made. Converting the bare building shell that housed Shockley Labs into a laboratory (Fig. 1) where semiconductor technology could be developed and refined was a challenge, but it was an excellent opportunity for us novices to learn by experiencing many of the pitfalls the technology held in store.

Because of some internal problems that could not be resolved satisfactorily, the eight Fairchild founders resigned from Shockley Labs after about a year and a half. While our initial direction at Shockley Labs had us working toward a diffused silicon transistor, the focus drifted toward another Shockley invention from his Bell Lab days, the four-layer diode, rather than a transistor, although Shockley originally had a bipolar transistor as his objective. Our relatively brief time at Shockley Labs allowed those of us new to semiconductors an opportunity to become familiar with some of the technology involved and to gain confidence that a double-diffused silicon transistor could compete successfully with the transistors being made by other approaches.

None of our group of eight scientists and engineers had significant management experience prior to Fairchild. Accordingly, a first goal after founding Fairchild Semiconductor Corporation was to bring in someone who could fill this void. We advertised for a general manager with appropriate background and experience, eventually hiring E. Baldwin from Hughes Aircraft's semiconductor operation, where he had been engineering manager. (At that time, Hughes Aircraft was a leading manufacturer of silicon diodes.)

We leased a newly constructed 14 000 square foot tilt-up shell in Palo Alto (Fig. 2), about a mile from Shockley Labs, and began the task of equipping it for semiconductor research, development, and, hopefully, production. While still a long way from a modern cleanroom, we took care to keep some areas reasonably clean and dust free. For example, nitrogen, oxygen, and forming gas were piped throughout the building from central sources of manifolded cylinders in cleaned copper pipe soldered without flux to minimize contamination.

Several processes had to be developed to allow the production of a transistor. We needed good quality single-crystal silicon wafers. There being no commercial source of silicon crystals, we had to build a crystal grower and produce our own from hyperpure polycrystalline silicon, which could be purchased. Roberts set up a crystal pulling system and shortly began producing single-crystal ingots ranging in diameter from about 3/8 in to 3/4 in, often all in one ingot.

We had decided to try to use photolithography to make the patterns on the silicon wafers. This was an extrapolation to much finer dimensions of the resist and etching technology that had been developed to make printed circuit boards. The technique had been used to etch holes in silicon oxide on wafers at Bell Laboratories and to make diode arrays at the Diamond Ordinance Fuse Laboratories by J. Nall, who later joined us at Fairchild. These applications required only a single layer to be produced by the technique. We hoped that we could develop the technology to use photolithography multiple times to make the various layers required for transistors.

The precise introduction of impurities into silicon by diffusion requires good temperature control at high temperatures. We knew from our Shockley Labs experience that there were no commercially available furnaces with the desired uniform hot zone and high operating temperatures we would need for diffusion, so we would have to design and build our own. Beyond that, metal films had to be applied by vacuum evaporation or other means, and a technology for packaging the finished transistors to protect them from the ambient would be needed. We had our

(a)

(b)

Fig. 2. (a) Current photograph of the original Fairchild Semiconductor building. (b) This building has been recognized as California's one thousandth historical landmark as the site where Noyce conceived the planar integrated circuit.

general outline of where we wanted to go, but there was a lot of work ahead to understand the materials and devices.

We divided the work to fit the backgrounds of the group. Roberts took responsibility for growing and slicing silicon crystals and for setting up a metallurgical analysis laboratory. Noyce and Last took on the lithography technology development, including mask making, wafer coating, exposure, development, and etching. Grinich set up electrical test equipment, consulted with the rest of the group on our electronic questions, and taught us how to measure various transistor parameters. Kleiner and Blank took charge of the facilities and set up a machine shop to make the equipment and fixtures we could not purchase. I took on the diffusion, metallization, and assembly technology

development. Hoerni, our theoretician, sat at his desk and thought.

The contrast from our first efforts to current industry practice is extreme. For example, for Noyce and Last to produce the first masks through which to expose photoresist, they sorted through the inventory of 16-mm movie camera lenses in a San Francisco camera store to select the three that matched most closely in focal length. These were then mounted in a rigid frame and used to make a set of three masks by stepping the image over the surface of three photographic plates. Because of the rigid frame, any array misalignment in one mask was reproduced in the other two, so that patterns printed with the resulting mask set could align, even if the individual masks were not perfect. For wafer exposure, the individual masks were mounted in metal frames with three indexing points to contact the wafer edges, assuring alignment. Two of these points touched a flat segment in the otherwise nearly round silicon wafers, while the third contacted the rounding edge 90° around the circle from the flat. To print the pattern on the wafer surface, the photoresist-coated wafer was placed in contact with the gelatin side of the photographic plate that constituted the mask and the combination was exposed to a strong light source. The first devices we made used minimum feature sizes of 0.005 in ($\sim$125 μm) with alignment tolerances half as large.

Early efforts to make junctions yielded poor electrical characteristics. The breakdown of the junctions in the reverse-bias direction was "soft," with the current rising rapidly with voltage, to levels not conducive to good transistor performance and not at all following the theoretically expected I–V characteristics. We had seen this problem at Shockley Labs. Something was badly wrong with our diffused junctions. It was not just that the carrier lifetime was short in our silicon. That would have made for higher leakage currents, but they would have had the correct functional dependence on voltage. Besides, photocurrent decay measurements indicated reasonable carrier lifetimes. We had a problem that there was conduction that should not be there at all. While it varied in magnitude from run to run, this excess current always dominated. Diffusion was my responsibility, and I was having a tough time making any progress on eliminating this leakage current.

A paper by Bemski and Struthers from Bell Labs [2] described how nickel plating on a silicon wafer and heating could be used to increase minority carrier lifetime by removing residual traces of gold, a known lifetime killer, from the bulk of the wafer. While lifetime was not our problem, Noyce suggested to me that I try nickel plating the back of a wafer to see if that would help. While I was not very enthused about introducing such a potential contaminant into my presumably clean diffusion furnaces, I was out of ideas, and so gave Noyce's a try. Much to my surprise, the resulting junctions were the best I had ever seen. They had low leakage and sharp breakdowns and obeyed the current–voltage relationship expected for space-charge generated current. The nickel certainly solved the soft junction problem.

In thinking back to my earlier attempts to make diffused junctions at the Shockley Laboratory, I remembered that my best "hard" junctions occurred on wafers where the gallium diffusion source had run out of control depositing droplets of molten gallium over the surface of the wafer, each droplet dissolving a little pit in the surface. These wafers looked terrible with all the pits, but between pits, there was enough flat area to make clean mesa diodes that behaved as they should. There was something about having a puddle of metal on the wafer during diffusion that cleaned up the junctions. Evidentially, the liquid metal was acting as a sink for some of the rapidly diffusing metallic impurities that were responsible for my soft junctions. Nickel "gettering" became an important part of our early device manufacture. Soft junctions were a thing of the past.

The mesa transistor required emitter and base contacts on the top surface, with the collector contact to the back of the transistor die. The Bell Labs [1] device that had been described used aluminum (an acceptor impurity that makes p-type silicon) to make ohmic contact to the p-type silicon constituting the base of the n-p-n transistor and a gold–antimony alloy to make contact to the emitter. Antimony dissolved in silicon acts as a donor making n-type material. Using two different contact metals complicates manufacture, both from the processing and from the subsequent attachment of electrodes. Accordingly, I was trying to come up with a single-metal system to make contacts to both the p-type base of the n-p-n transistor and the heavily doped n-type emitter. To achieve this delicate balance, I was working with various alloys to try to find one that would behave differently electrically in the emitter and base regions, making ohmic contacts to both. Here again, I was running out of ideas. My occasional partial successes were difficult to duplicate. It was beginning to look like we might have to go to a two-contact-metal solution, when Noyce suggested that I try aluminum for both. Now, we all knew that aluminum would make a rectifying contact to n-type material. Noyce, with his background in semiconductors, knew it better than any of us. Nevertheless, I tried evaporating a film of aluminum on both the base and emitter areas of the transistors and alloying the film to the silicon by passing it through a furnace at a temperature above the silicon-aluminum eutectic temperature but below the melting point of aluminum. Much to my amazement, the contacts to both the base and emitter were nearly perfect. Aluminum did make a marvelous low-resistance ohmic contact to the emitter as well as to the base. While it was several years before we understood the physics involved, that was not important at the time. What was important was that we had a reproducible process to make contacts to our transistors. Although many other metal contact and interconnection schemes have been tried subsequently, aluminum remains the principal conductor used in integrated circuits today.

We were developing both n-p-n and p-n-p mesa transistors in the beginning, not knowing which would be better. Both were working, but both had problems. The aluminum solution for contacts worked in a straightforward manner

only for the n-p-n, since it did not make ohmic contact to the relatively lowly doped base region of the p-n-p. Principally for this reason, our first product was the n-p-n, although it was followed shortly by a similar p-n-p using a more complicated process for contacts.

Our first transistor was a fairly large device capable of switching enough current to drive magnetic core memories. It was targeted at an application in the Government Systems Division of IBM in Owego, NY. Their need was for a silicon transistor to switch 150 mA to drive magnetic core memories. Fortunately, this high current requirement put the size of the device in a range where it was practical for us as a first target for our photolithography. We successfully shipped the first 100 of these transistors in 1958. This device was designated the 2N696 and, along with a higher gain selection called the 2N697, was a significant success, finding a broad range of applications in analog as well as digital circuits. This high-current device was followed by smaller mesa transistors giving higher switching speeds at lower currents in logic applications

It seems that there is always the need for more speed in electronics. The switching speed of these early mesa transistors was limited by the time necessary for the charge injected into the collector region to decay. This time was a function of both geometry and carrier lifetime in the collector. Hoerni had the idea that if we could kill the minority carrier lifetime, the transistors would turn off much faster, and he calculated that we did not need as long a lifetime as we measured in our devices. This was certainly new thinking. The conventional wisdom was that greater lifetime was always better. The double-diffused transistor with its lightly doped collector region, however, changed the rules.

Knowing that gold diffused rapidly in silicon and was an effective carrier recombination center, Hoerni evaporated gold on the back of a wafer in place of the nickel and shoved it into one of the diffusion furnaces. To my surprise, gold had the same favorable effect on the junction characteristics as did nickel. Again, a puddle of molten metal did the trick: the junction breakdowns were sharp. On the other hand, the carrier lifetime was much shorter than in the nongold-doped transistors. They shut off much more rapidly, allowing faster switching times, and there was still enough lifetime that the transistors had adequate current gain. Gold-doped transistors became the standard high-speed switches.

In mesa transistors, the emitter-base junction is exposed on the top surface between the metal contacts, while the base-collector junction intersects the sides of the mesa (Fig. 3). The regions of high electric fields where the junction comes to the surface are sensitive to contamination. Contamination of the emitter-base junction can decrease the gain of the transistor dramatically. In the case of the collector junction, the breakdown voltage and leakage characteristics can change. We noted a problem that some of the transistors packaged in hermetically sealed cans in dry nitrogen showed very unstable collector junction characteristics. Breakdown voltages sometimes decreased by several tens of volts and became unstable when observed

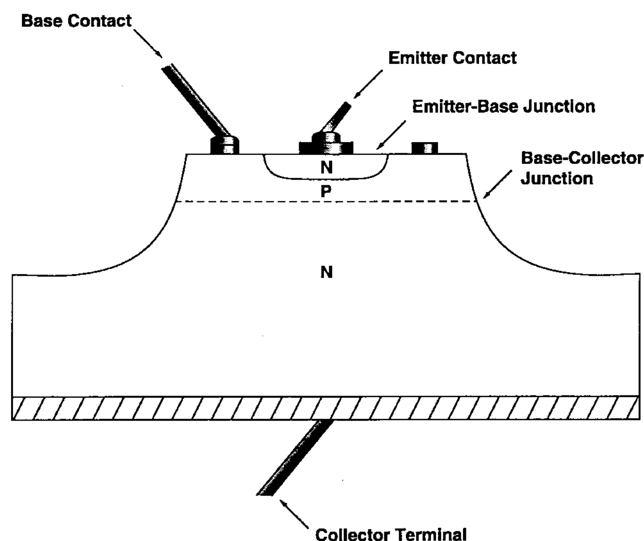

Fig. 3. Schematic cross section of an early mesa transistor made by Fairchild Semiconductor Corporation.

on an oscilloscope, potentially a major reliability problem. We formed a task force to try to understand and correct this problem. One of our technicians, B. Robson, carefully cut the can off one of the bad devices and examined it under a microscope. He noticed a spot of light emitted from the side of the mesa when the transistor was biased into breakdown. He shut off the power and saw a tiny particle on the side to the mesa at the point of the light emission. Carefully removing the particle and reapplying power, he found that the original high breakdown voltage was restored. The particle, evidentially attracted by the high electrical field where the junction came to the surface, was causing the premature breakdown of the junction. Now we knew the cause of the low breakdown. All we had to do was eliminate all the sources of particles.

First, we learned how to make particles cause the low breakdown. By taking almost any packaged mesa transistor and tapping it sufficiently vigorously with a pencil, we could produce unstable voltage characteristics. Particles shaken loose in the transistor can land on the junction area, causing premature breakdown. We undertook to eliminate all the potential sources of particles we could find. These included little bits of plating, solder, or whatever else might come out of the package. The pencil-tapping test was automated and enhanced. By careful cleaning and vigorous agitation of all the parts of the transistor and package before welding it closed, we were able to improve the incidence of this problem significantly, but we could never seem to get completely rid of it. Perhaps the electric spot welding used to seal the package produced enough particles that there would always be a residual problem. Since many of these devices were destined for applications where very high reliability was required, we were concerned.

II. The Planar Transistor

Actually, we had had the ultimate solution to this problem since a few months after Fairchild was founded. While

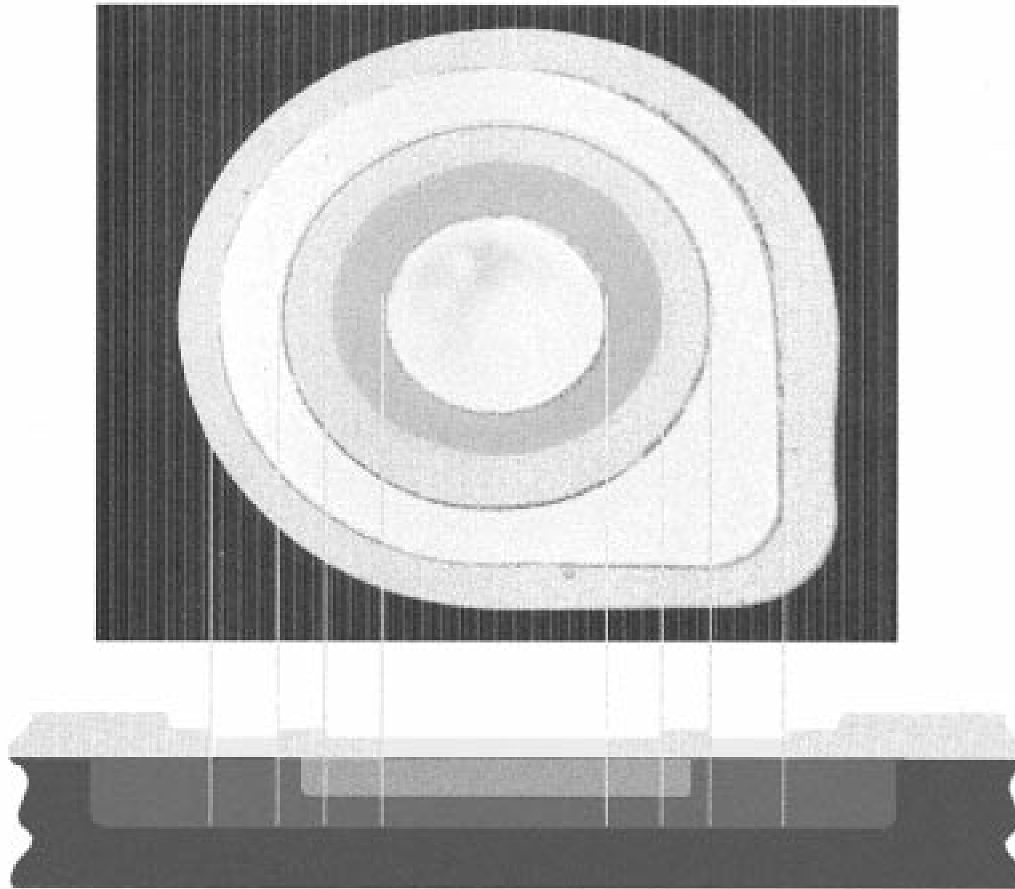

Fig. 4. Photomicrograph of the first planar transistor. The diameter of the circle that forms most of the outside ring is 0.030 in. The light areas are aluminum emitter and base electrodes. (From "A Solid State of Progress," Fairchild Camera and Instrument Corporation, 1979.)

most of us were setting up the facilities and developing the early processes, Hoerni was drawing in his notebook. He sketched a double-diffused transistor made without a mesa. Instead, he proposed a structure where the collector area was defined by an oxide-masked diffusion from the top surface, similar to the way we made the emitter diffusion on our mesa transistors. He proposed, however, that instead of dissolving the silicon dioxide layer from the entire surface, it be left in place, except in those areas where contacts were to be made. Since the diffused junctions actually went under the edge of the oxide layer a distance about equal to their depth into the wafer, the junction's intersection with the surface would actually be covered by the oxide [3]. The conventional wisdom was that the oxide after diffusion was contaminated and should be removed. Also, the technology to leave the oxide in place required the flexibility of the photolithographic technology to accomplish. Even at Fairchild, we did not attempt to make Hoerni's planar structure for some time after he proposed it. The original photolithography we developed only made three masks in a set. To make Hoerni's planar transistor took a set of four: the first to define a hole in the oxide layer to delineate the base diffusion, the second to define a smaller oxide hole for the emitter diffusion in the regrown oxide, the third to open areas where the contacts were to be made, and the fourth to etch the contact pattern in the aluminum film. Fig. 4 shows a photomicrograph of the first commercial planar transistor.

First experiments toward realigning Hoerni's planar device left the oxide only on the emitter junction. This could be tried on mesa transistors by taking advantage of the oxide over the base's being thicker than over the emitter region. A careful etch left some oxide over the base, and the region where the emitter-base junction came to the surface. The current gain of transistors with this thin oxide over the junction was much more stable and at the high end of the range for the particular batch of devices. Hoerni's oxide-over-junction idea really seemed to work, at least on the emitter junction. When we could try it over both junctions, the results were fantastic. The breakdown voltages and gain were stable. The devices were far less sensitive to the ambient than any previous transistors. Particles in the package too small to short between contacts had no effect. We had an approach that would allow us to make the highly reliable transistors that were being specified for the Minuteman intercontinental missiles.

There were still problems, however. Yields of planar transistors were very low initially. The masking steps were far more sensitive to pin holes in the oxide films than had been the case for the mesa transistors. Considerable cleanup of the manufacturing process to eliminate particles from the

resist material used in photolithography and dust from the various processes and ambient was necessary to get yields comparable to the older mesa devices. With the improved electrical characteristics and stability, however, it was well worth the effort.

III. The Integrated Circuit

In the late 1950's, there was considerable interest in shrinking the size of electronic systems, especially for military applications. Several approaches toward miniature packages were being pursued, and there was a large effort sponsored by the Department of Defense to make functional electronic devices to replace assemblages of components. When the patent application for the planar transistor and the process for making it was being prepared for filing, having been made aware of our view of the importance of this invention, the patent attorney asked if we had thought through the implications of the invention to be sure that we were covering it sufficiently. To consider this question, Noyce assembled a meeting of the key technical people in the Research and Development Department at Fairchild to discuss how Hoerni's invention might be extended. (I was not at that meeting, since my responsibility at the time was to oversee the engineering effort to get our transistors into production.) During this session, Noyce described how the planar idea could be extended to make complete circuits rather that just individual components.

To accomplish this in a general way required two major extensions—electrical isolation and interconnection. In a silicon wafer of planar transistors, the collector region is common to all of them. For n-p-n transistors, this was the n-type material of the original wafer. For most circuits, the collectors have to be separated electrically. Noyce proposed that extra junction pairs be inserted in the wafer between transistors or other components so that one or the other junction would be biased in the nonconducting direction irrespective of the relative voltages, thus eliminating current flow. He also realized that it should be possible to run the aluminum metallization from the contacts of the transistor over the regions where the junctions reach the silicon surface using Hoerni's planar oxide as an insulator to separate the conductor from the junction. It was straightforward to see how to make diffused resistors and small capacitors with the same processing, thus allowing the construction of complete circuits in a single monolithic silicon chip. At this meeting, Noyce described all the additional features necessary to extend the planar transistor technology to make a practical structure for an integrated circuit. He was eventually awarded a patent on the key idea of interconnection [4], although his insight was much broader.

Shortly after Noyce's invention, Baldwin left to found Rheem Semiconductor and Noyce became general manager of Fairchild Semiconductor Corporation. I replaced him as head of the R&D Department, inheriting the job of realizing his integrated circuit invention.

It was easier to draw the structures on a blackboard than it was to construct them in a silicon wafer. The junction isolation was especially tricky. We could see two ways to construct the isolated transistors Noyce envisioned. Probably the easiest way was to perform a triple diffusion from the front surface of the wafer. In addition to the emitter and base diffusions done for the planar transistors, one could also do a deeper diffusion at lower surface concentration to make isolated collector regions. Calculations showed, however, that this would be difficult to control, and the resulting transistors would have a very large parasitic series resistance in their collectors, limiting performance.

We chose to implement a second approach where the transistors would be made with the collector region composed of the original silicon wafer isolated by a grid formed by diffusing completely through the wafer from both top and bottom. On the top of the wafer, a pattern of silicon dioxide masked regions to contain transistors and other components. Boron, a relatively fast-diffusing dopant, was diffused from both the front and the entire back surfaces of the wafer until the diffusion fronts met in the center of the wafer. This left wells of the original n-type material on the top wafer surface where the oxide mask was located, each well surrounded by isolating p-type material. Diffusion through the wafer favored thin wafers. We found that we could work with wafers as thin as about 80 μm. At this thickness, a 24-hour diffusion was sufficient to complete the isolation structure. Transistors and other components were then made by the regular planar processes in the isolated n-type wells. Fig. 5 shows a photomicrograph of an early circuit made by this process, and Fig. 6 shows a schematic cross section of the structure at various stages in the manufacturing process [5]. Beyond this stage, planar transistors and other components were constructed in the isolated n-type regions.

This first family of integrated logic circuits, called Micrologic [6], consisting of five simple logic functions, was introduced in 1961. The circuits were packaged in modified TO-5 transistor packages with up to eight leads. The resistor-coupled circuitry employed was chosen for its simplicity to implement. Early units were expensive and did not offer performance advantages over conventional circuitry. The real advantage that the early integrated circuits offered was packing density. The first customers were mostly military contractors. The computer that went to the moon with the Apollo astronauts was built using the three-input NOR gate from the Micrologic family of circuits.

About the time that these first integrated circuits were being developed, the idea of epitaxial growth and the epitaxial transistor arose. By this process, one grows additional silicon on the surface of the wafer continuing the underlying crystal lattice. Since the added layer can be doped during growth, epitaxy, contrary to diffusion, offers the ability to deposit layers that are more lightly doped than the substrate wafer. By changing the impurity concentration while growing an epitaxial layer, it is possible to control the doping profile in the grown film.

The use of epitaxial growth to make a transistor was first described by a research team from Bell Labs at the 1960 Solid State Device Research Conference [7]. By growing a thin epitaxial film of relatively lightly doped silicon on a

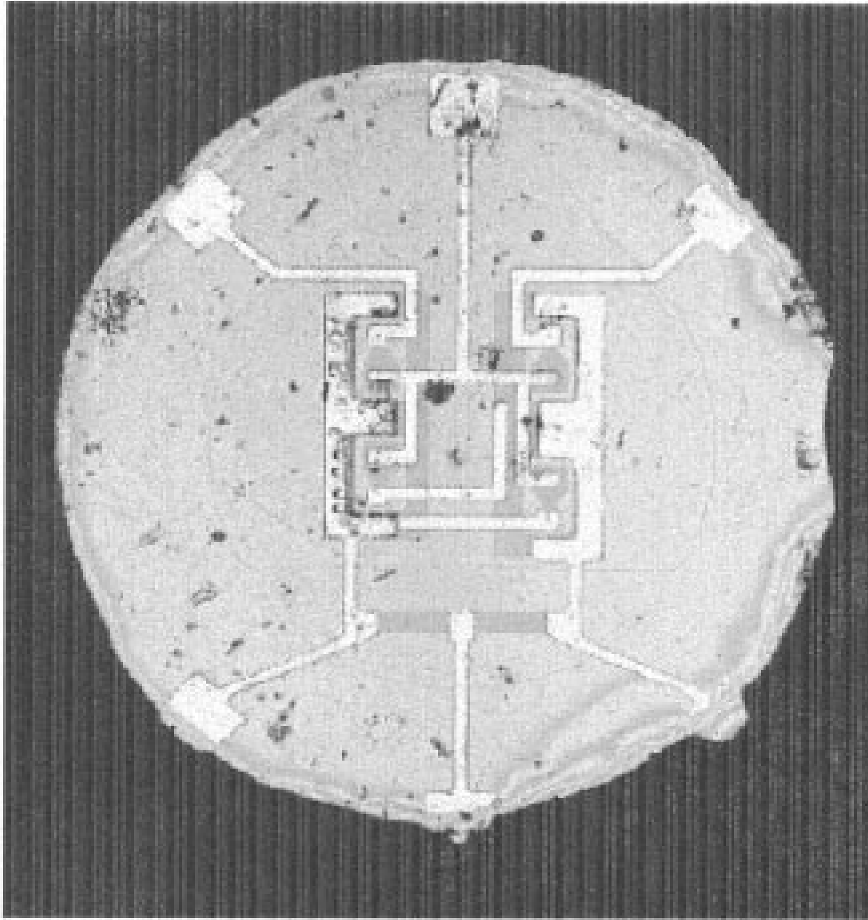

Fig. 5. Photomicrograph of one of the first planar integrated circuits made at Fairchild. This is a flip-flop circuit. Some of the aluminum interconnection metal has been damaged during the etching operation to form a circular chip of silicon to plane into a transistor can modified to have more leads. (From "A Solid State of Progress," Fairchild Camera and Instrument Corporation, 1979.)

more heavily doped substrate wafer, they showed that the parasitic collector resistance in a double-diffused transistor could be greatly reduced, thereby increasing the current handling of a transistor switch. At Fairchild, we also had the idea of the epitaxial transistor, but we were not set up to grow epitaxial silicon internally. We tried to procure silicon with a layer of lightly doped n-type on a heavily doped n-type substrate, into which we wanted to diffuse base and emitter regions, but our silicon crystal vendor (by this time there was a supply of single-crystal silicon) only wanted to supply material with epitaxial grown junctions. Had we been able to get the substrates, we would have beaten the Bell Labs team to the epitaxial transistor. Several years after the announcement of the epitaxial transistor, I learned that their work had been done just prior to the announcement. Our conception of the idea at Fairchild was several months earlier.

The real impact of epitaxial growth, however, was on integrated circuits. It was no longer necessary to diffuse isolation clear through the silicon wafer. We could grow a relatively thin n-type layer of the correct doping level for the collector region on a p-type substrate and make isolated regions by diffusing boron through the epilayer. Not only was it possible to grow an n-type film on a p-type substrate but it proved practical to diffuse patterns of impurities into the substrate prior to growing the epilayer. When combined with masked diffusion, epitaxial growth enabled nearly complete flexibility of doping profiles in three dimensions. This allowed integrated circuits to proliferate, and Fairchild, as well as other semiconductor companies, introduced a plethora of digital and analog functions that offered increasing advantages in performance and cost in addition to packing density and weight.

IV. The MOS Transistor

At the same meeting that the epitaxial transistor was first announced, Kahng and Atalla from Bell Labs described a MOS transistor [8]. This was a working version of the insulated-gate field-effect transistor that had been the goal of the research that led to the original invention of the transistor in 1947. This device consisted of two diffused regions, the source and drain, with a gate electrode covering the region between them but insulated from the substrate by a layer of silicon oxide. Voltage applied to the gate could control the conduction between the source and drain regions acting as an amplifier or a switch. The thermally grown

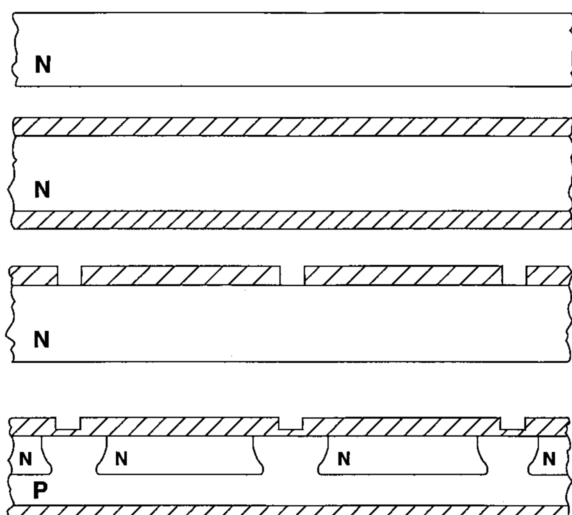

Fig. 6. Schematic cross section of steps in preparation of the isolation structure made by diffusion completely through the silicon wafer that was used in the earliest commercial integrated circuits. Devices such as planar transistors were formed in the isolated wells of n-type silicon.

oxide layer had decreased the density of surface states to trap electrons at the surface of the silicon sufficiently that the long-pursued field-effect device initially patented by Lilienfeld in 1926 [9] could finally be made to function.

At Fairchild, at the time that the Kahng and Attala paper was presented, we had been working with various electrodes on top of the oxide layers over junctions in an attempt to understand some of the characteristics of the interface. While the planar transistor structure solved some of the biggest problems with double-diffused transistors, there were several that persisted. Particularly over the lightly doped collector region of high-voltage p-n-p planar transistors, an inversion layer sometimes developed, effectively extending the base region to the edge of the die. Inversion layers were not a new phenomenon. For example, early grown-junction transistors had problems with such layers developing on the surface of the base's shorting the emitter to the collector. In fact, the reason that Fairchild's first transistor had the base contact completely surrounding the emitter was to eliminate the possibility of such inversion layers.

By putting electrodes over the collector junction and applying appropriate voltages, inversion could be controlled and the junction breakdown voltage could be modulated. Similarly, by applying voltage to an electrode over the emitter-base junction, the gain of the transistor could be modulated. C. T. Sah, who was directing our experiments with gate electrodes over junctions, even patented a tetrode transistor that had as a fourth electrode an insulated control gate on top of the oxide covering the emitter-base junction [10].

With Kahng and Atalla's paper, we started looking at the MOS transistor. It was easy to make a device that worked, but we could not make one whose characteristics were stable. Clearly, something was happening at the silicon-oxide interface that we did not understand that caused the device parameters to drift under bias. The effect was accelerated by temperature. These devices were made using the usual planar processing techniques, including photolithography, oxide-masked diffusion, and aluminum metallization deposited by evaporating the metal from a heated tungsten filament in a vacuum chamber.

In an effort to vary the threshold voltage by changing the work function of the gate electrode, we tried different metals. Coincidentally, we found that devices made with molybdenum gates were far more stable than the ones with aluminum gates. Clearly, something was different. The easiest way to evaporate the more refractory metals, such as molybdenum, was to heat the source with an electron beam. The devices made with electron-beam evaporation were pretty stable. Even when we evaporated aluminum with e-beam heating, the devices were much more stable than with filament evaporation. We began to suspect that some impurity in the filament-evaporated aluminum was responsible for the instability. Further experiments with purposely induced contamination led us to alkali ions, especially sodium, as the culprits. A heated tungsten filament for aluminum evaporation gave off enough impurities to cause the problem, while electron-beam heating did not contaminate the hyperpure aluminum used as an evaporation source. Considerable additional, careful work was required to identify the sources of impurities and other contributors to parameter drift to allow the production of reproducible, stable MOS transistors.

Controlling what was happening at the silicon–silicon–oxide interface was becoming increasingly important, and a variety of observations convinced us that there was much we did not understand. This led to a major research effort to study in detail what was happening. We expanded the group working in this area with the charter of learning how to control and stabilize the electrical characteristics of the silicon–silicon–oxide interface. Sah originally was in charge of the research efforts. A. S. Grove, E. Snow, and B. Deal were hired to expand the group. Sah moved to the University of Illinois in 1963, and leadership of the group was taken over by Grove. For a time, Sah visited about monthly to consult with the Fairchild researchers. Over the next few years, this team developed an increasingly detailed understanding of the phenomena taking place in the metal–oxide–silicon system that forms the underpinning for modern MOS devices [11].

At the same time that we were working on the instabilities in MOS devices, we were becoming more excited about their potential applications. F. Wanlass, a very creative individual, joined Fairchild in 1962 and began investigating how they could be used. He proposed a wide variety of applications and circuits exploiting the unique electrical characteristics of MOS transistors, probably the most important of which was the idea of complementary MOS circuitry. By using both n-channel and p-channel normally off MOS transistors, one could make circuits that only used significant power when switching. This was reported at the 1963 Solid State Circuit Conference [12] and has become the dominant circuit form used in microprocessors

and memories today, in addition to having enabled very low power electronics, such as digital watches.

In spite of its contributions to the understanding of MOS structures and the first commercial introduction of a MOS transistor [13], Fairchild never was a major participant in the market for MOS devices. In my opinion, the main cause of this was Fairchild's great success in bipolar circuits. The market for integrated circuits was growing and expanding rapidly in the mid- and late 1960's. There was strong competition, and Fairchild was expanding rapidly to keep its industry-leading position. MOS technology, while superficially very similar to that for bipolar devices, required special attention in areas where bipolar circuits were robust. Accordingly, it was not a simple substitution of MOS for bipolar. In addition, in the classic Silicon Valley manner, many of the key players in the development of MOS at Fairchild, including Wanlass, left to form other companies to exploit the new technology.

We were very lucky at Fairchild to get off on the right technological track. Silicon was the right material, and oxide-masked diffusion allowed batch fabrication of transistors. The planar structure not only solved many of the vexing problems of the mesa transistor but was the path to the practical integrated circuit and, with expanded knowledge, the MOS transistor structure. For a decade or so during the rapid evolution of silicon device technology, Fairchild Semiconductor Corporation and its successor, the Semiconductor Division of Fairchild Camera and Instrument Corporation, played a key role.

References

[1] M. Tannenbaum and D. E. Thomas, "Diffused emitter and base silicon transistors," *Bell Syst. Tech J.*, vol. 35, no. 1, pp. 1–15, Jan. 1956.

[2] G. Bemski and J. D. Struthers, "Gold in silicon," presented at the Electrochemical Society Meeting, Buffalo, NY, Oct. 1957.

[3] J. Hoerni, U.S. Patent 3 025 589.

[4] R. N. Noyce, U.S. Patent 2 981 877, Apr. 25, 1961.

[5] E. Keonjian, Microelectronics. New York: McGraw-Hill, 1963, p. 306.

[6] "Micrologic," Fairchild Semiconductor Division, Fairchild Camera and Instruments, Mountain View, CA, Aug. 1961.

[7] H. C. Theuerer, J. J. Kleimack, H. H. Loar, and H. Christensen, "Epitaxial diffused transistors," presented at the Solid State Research Conf., Pittsburgh, PA, June 1960. See also *Proc. IRE*, vol. 48, pp. 1642–1643, Sept. 1960.

[8] D. Kahng and M. M. Atalla, "Silicon-silicon dioxide field induced surface devices," presented at the Solid State Research Conf., Pittsburgh, PA, June 1960.

[9] J. E. Lilienfeld, "Method and apparatus for controlling electric currents," U.S. Patent 1 745 175, Jan. 18, 1930.

[10] C. T. Sah, "A new semiconductor tetrode, the surface-potential controlled transistor," *Proc. IRE.*, vol. 49, pp. 1623–1634, Nov. 1961.

[11] A. S. Grove, *Physics and Technology of Semiconductor Devices*. Wiley, 1967.

[12] F. M. Wanlass and C. T. Sah, "Nanowatt logic using field-effect metal-oxide semiconductor triodes," in *Tech. Dig. IEEE 1963 Int. Solid State Circuit Conf.*, Philadelphia, PA, Feb. 20, 1963, pp. 32–33.

[13] C. T. Sah, "Evolution of the MOS transistor-from conception to VLSI," *Proc. IEEE*, vol. 76, pp. 1280–1326, Oct. 1988.

Gordon E. Moore (Life Fellow, IEEE) was born in San Francisco, CA, on January 3, 1929. He received the B.S. degree in chemistry from the University of California, Berkeley, in 1950 and the Ph.D. degree in chemistry and physics from the California Institute of Technology, Pasadena, in 1954.

He joined the Technical Staff of the Applied Physics Laboratory at Johns Hopkins University in 1953, where he did basic research in chemical physics. He joined Shockley Semiconductor Laboratory shortly after its founding in 1956 in Palo Alto, CA, working on semiconductor process technology with W. Shockley, coinventor of the transistor. He cofounded Fairchild Semiconductor Corporation in Mountain View, CA, in 1957, serving as Manager of the Engineering Department until 1959, when he became the Director of Research and Development. Fairchild produced the first commercial integrated circuit during this period. In July 1968, he cofounded Intel Corporation to develop and produce large-scale integrated (LSI) products, beginning with semiconductor memories. Intel has gone on to produce a number of products based upon LSI technology, including the world's first microprocessor. He was Executive Vice President of Intel until 1975, when he became President and Chief Executive Officer. In April 1979, he became Chairman of the Board and Chief Executive Officer, holding the latter position until April 1987. He remained Chairman until 1997, at which time he became Chairman Emeritus. He is a Director of Transamerica Corporation, Gilead Sciences, and Varian Associates, Inc. He is Chairman of the Executive Committee of Conservation International and Chairman of the Board of Trustees of the California Institute of Technology.

Dr. Moore is a member of the National Academy of Engineering. Among the awards he has received during his career are the Distinguished Alumnus Award from the California Institute of Technology, the W. W. McDowell Award from the IEEE Computer Society, the IEEE Frederick Phillips Award, the AFIPS Harry Goode Award for Leadership in Science, the IEEE Computer Pioneer Medal, the ASM Medal for the Advancement of Research from the American Society for Metals, and the Founders Award from the National Academy of Engineering. In November 1990, he received the National Medal of Technology from President G. Bush. In February 1993, he received the IEEE John Fritz Medal. In 1997, he received the IEEE Founders Award.