

The Origins of Diffused-Silicon Technology at Bell Labs, 1954-55*

by Nick Holonyak, Jr.

Silicon-based transistor and integrated circuit technology has grown so large and become so important that it is difficult to recognize how it all began. In spite of the obvious differences between germanium and silicon, it was not evident during the early 1950s in what form and substance the transistor would prevail. What was the need for silicon—at the time such an intractable, peculiar new technology?

In 1954-55, however, the requirement for low-leakage switching devices at Bell Telephone Laboratories led directly to the exploration of impurity-diffusion and aluminum-metallization technology to make silicon transistors and *p-n-p-n* switches. This technology, a more or less ideal thin-layer technology that can be referenced from a single surface, led to the 1955 discovery of the protective silicon-dioxide layer, oxide masking and patterning, and ultimately to the silicon integrated circuit.

In this article, I recount some of the exploratory diffused-impurity silicon device development at Bell Labs that helped to establish the fundamental basis for microelectronics. It draws upon my direct experiences in the group in the Device Development Department supervised by John L. Moll, which was concerned with switching devices. Silicon diffused-impurity devices made there in 1954-55 are described, including work and data not broadly known. Much of this technology was soon carried across the country to what became known as Silicon Valley.

Silicon Technology and Switching Devices

In the early 1950s, laboratories were hiring new PhD scientists and engineers to work on the transistor and, finding many of the new recruits lacking in semiconductor background, began short courses and training programs in semiconductor studies. In the fall of 1954, James M. Goldey and I joined Moll's switching-device group and, because of prior semiconductor device experience, were ready to start building and learning about silicon switching devices. Moll, of course, had already written (with J. J. Ebers) the Ebers-Moll equations describing the switching operation of transistors¹ and had given much thought to why silicon devices must be developed to replace

germanium, particularly for switching functions. The energy band gap of germanium was too small (0.65 eV) compared to silicon (1.11 eV), and thus yielded an off-state switch—a reverse-biased junction—that was too leaky. We knew, as Moll had argued, that we had to work with silicon to build switching transistors, especially the *p-n-p-n* switch. This four-layer diode was of substantial interest then because it appeared capable of competing with a gas-tube cross-point switch that had attracted interest within Bell Labs.^{2,3}

At first I proposed to build a germanium *p-n-p-n* switch, but Moll objected on the grounds of the leakiness of this element due to its small energy gap. An alternative was to fabricate quickly a silicon *p-n-p-n* switch by introducing a point contact on the collector body, near the collector junction, of a grown-junction *n-p-n* transistor. This approach would work in principle, Moll agreed, but he noted the inherent problem of poor reproducibility because of the point contact. He knew we had to employ silicon and, if necessary, generate all the requisite diffusion and metallization technology to realize sophisticated *n-p-n*, *p-n-p* and *p-n-p-n* device structures. Small silicon crystals were then available, typically almost two inches in diameter, but a process technology yielding the switching devices we wanted did not yet exist.

At the time we began to construct diffused-impurity silicon switching devices, diffused-silicon solar cells had been demonstrated,⁴ and Prince and colleagues were then developing diffused-silicon rectifiers.⁵ In addition, Lee and co-workers were attempting to make a germanium transistor with a diffused base and an evaporated (alloyed) emitter.⁶ Thus there was reason to think that diffused-silicon devices could indeed be made, particularly if impurity diffusion could be developed further and sophisticated contacting schemes be devised. In the latter case, metal evaporation appeared most desirable because either a contact or a shallow *p-n* junction (using thin-layer precision alloying) could be realized. Most important, metal evaporation could be done in a uniform patterned format.

The kind of device technology outlined here, which concerned us in 1954, has the important advantage, particularly for low-power devices, that all of the vital junction construction could be referenced and realized (precisely!) from one side of the crystal wafer.⁷ This indicated that precise dimensions could be achieved, which is not possible in the case of an alloy transistor with junctions constructed, more or less imprecisely, on both sides of a relatively thick crystalline wafer.

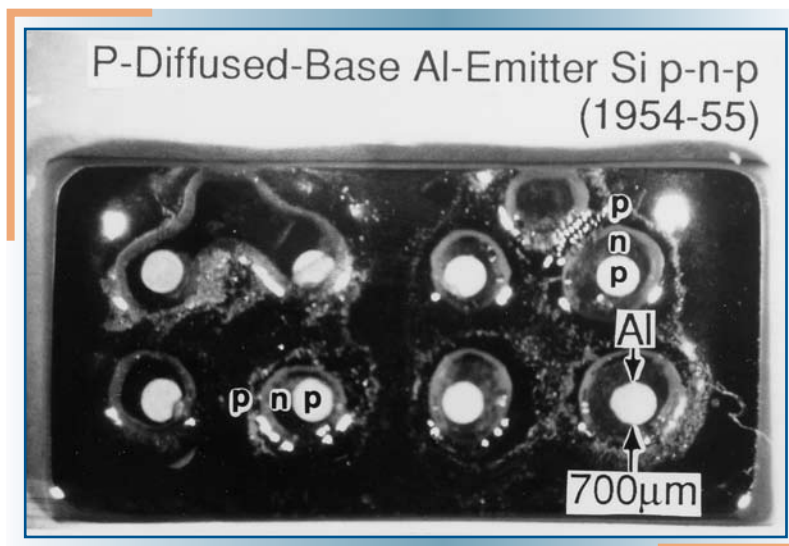


Fig. 1. Aluminum evaporated and alloyed on a *p*-type silicon wafer that had first been diffused and made *n*-type with phosphorus. The etched mesas are low-gain *p-n-p* transistors fabricated in late 1954.

*Adapted from N. Holonyak, Jr., "Diffused Silicon Transistors and Switches (1954-55): The Beginning of Integrated Circuit Technology," in David G. Seiler *et al.*, Editors, *Characterization and Metrology for ULSI Technology*, 2003 International Conference on Characterization and Metrology for ULSI, AIP Conference Proceedings Vol. 683, (College Park, MD: American Institute of Physics, 2003), pp. 40-60. Also published in C. L. Claeys *et al.*, Editors, *ULSI Process Integration III*, ECS Conference Proceedings No. PV 2003-06 (2003), pp. 68-106. Please consult this paper for a more complete discussion of the early work in diffused-base silicon transistors and switches.

Discovery of the Silicon Dioxide Layer

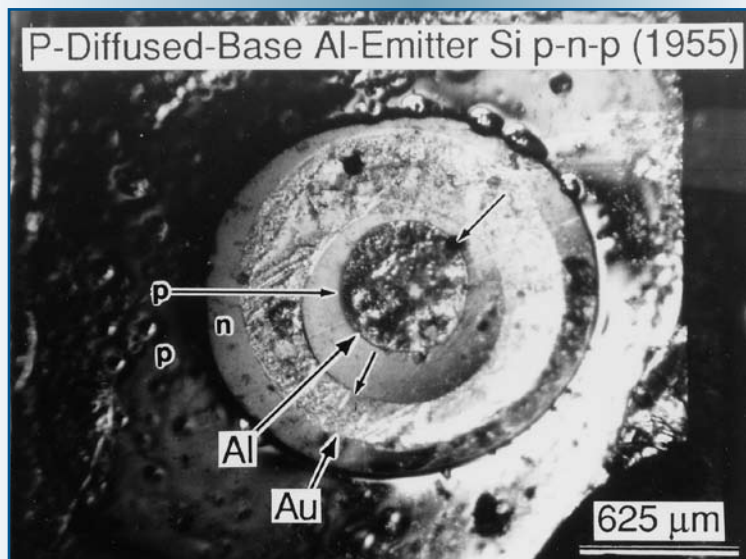


Fig. 2. Diffused-base $p-n-p$ silicon mesa transistor made in early 1955 using evaporated gold-plus-antimony ring base contact and aluminum-evaporated and alloyed emitter.

Using the evaporator in Moll's group, Goldey began trying to determine how to evaporate gold-antimony on silicon in an attempt to realize n -type contacts or shallow (n on p) $n-p$ junctions. In the research department, Morris Tanenbaum was attempting to evaporate and alloy aluminum on "hot" (over 660°C) silicon, but he had failed to realize continuous metallization across the silicon surface. After Moll obtained Tanenbaum's permission for us to use his evaporator, we quickly solved the problem of evaporating aluminum on silicon, either hot or cold. We also made precision, shallow alloyed p -type contacts or shallow (p on n) $p-n$ junctions. By late 1954 Goldey and I had solved the problem of making metal contacts on silicon and forming uniform, shallow $p-n$ (or $n-p$) junctions or shallow ohmic contacts. I wrote a Bell Labs memorandum with Tanenbaum on this aluminum metallization and shallow junction formation on silicon;⁸ Goldey included this material and further results in a later report.⁹

The prototype devices we made in late 1954 and early 1955 included a diffused-base, alloyed-emitter transistor. In March 1955, these achievements—along with those of our colleagues Calvin Fuller, Carl Frosch and Tanenbaum—helped convince transistor-development head Jack Morton (after strenuous argument) to redirect all work towards diffused-base silicon transistor technology. Figure 1 is a photograph of our rudimentary silicon $p-n-p$ diffused-base alloyed-emitter transistor; the emitter is aluminum evaporated and shallow-alloyed into the silicon. In creating this device, phosphorus was first diffused into a p -type silicon wafer, forming an n -type base layer. The wafer was masked and aluminum was evaporated and alloyed onto the eight

spots on the wafer to form (after etching) separate $p-n-p$ mesa transistors.

Made before Christmas of 1954, our first diffused-base silicon transistors had ratios of collector current to emitter current of only 0.1–0.2. Small improvements that came soon thereafter were needed to make these transistors fully functional (See Fig. 2). Double-diffused silicon $n-p-n$ transistors, fabricated by Tanenbaum and D. E. Thomas using wafers that Fuller had prepared by simultaneously diffusing aluminum and antimony into them, achieved higher ratios earlier,¹⁰ but they required that the p -type base contact be alloyed through the n -type emitter, thus resulting in some compromise. If the evaporated aluminum contact on the double-diffused $n-p-n$ device was not alloyed deep enough and thus formed a shallow $p-n$ junction on the top n -type layer, an accidental $p-n-p-n$ switch resulted.

Many of the diffused-silicon wafers for our $p-n-p$ transistors (and $p-n-p-n$ switches) were prepared by Frosch using a phosphorus "pre-deposition" diffusion followed by a higher-temperature "drive-in" diffusion. When we asked him to reduce the pre-deposition time and temperature and to increase the drive-in time and temperature, we either lost the n -type layer completely, or the wafer was pitted, eroded or destroyed during the dry-gas process. This was a serious problem in late 1954 and early 1955, but Frosch solved it by a major discovery he made a few months later.

In the meantime, we were able to devise a practical solution to our problem of constructing a diffused-base alloyed-emitter silicon $p-n-p$ device using a trick to cope with the high phosphorus concentration that Frosch was introducing. But our solution to the problems of diffusion in silicon as well as Tanenbaum's $n-p-n$ transistors¹⁰ fabricated from Fuller's double-diffused wafers did not matter much in the long run, except to demonstrate the feasibility of diffused-silicon devices. A far more important development was about to emerge that would fundamentally change semiconductor technology, making all existing diffusion procedures obsolete.

Frosch, the man chiefly responsible for this advance, was a consummate process chemist familiar with many types of procedures and had been working for several years on impurity diffusion in silicon with his technician Lincoln Derick (See Fig. 3). In spite of his considerable experience with dry-gas diffusion procedures, however, Frosch regularly reduced many of our silicon wafers to cinders, particularly at high temperatures near or above 1100°C .



Fig. 3. Calvin Fuller, Carl Frosch, and Link Derick with an early diffusion furnace. (Courtesy AT&T Archives and History Center)

Holonyak, Jr.
(continued from previous page)

Because we had mastered making a diffused-base alloyed-emitter silicon *p-n-p* transistor despite the problems with diffusion, one of the *p-n-p-n* configurations that we could explore was a simple modification of this transistor. We could fabricate the transistor on either side of a *p*-type substrate wafer after it was first prepared

encountered the same long face as he commented, "Well, we did it again." By this I thought he meant that my wafers had been destroyed once more. But then he cracked a broad smile and showed me the silicon wafers in his hand. They were surprisingly smooth and green in color—due to interference in a thin surface layer! He and Derick had inadvertently switched from dry-gas impurity diffusion to wet-ambient (water vapor plus carrier

blocked by the natural protective oxide and what others (gallium, for example) could penetrate it. Once this layer had been recognized, it was easy to devise various schemes to promote or block impurity diffusion. The process was so flexible that planar *n*-type regions of almost any desired shape and pattern could be prepared on *p*-type silicon wafers—or the opposite, *p*-type diffused regions on *n*-type silicon. Suddenly, all other diffusion procedures had been rendered obsolete.

Frosch and Derick had done nothing less than lay the groundwork for a revolution in semiconductor technology. In fact, it is the oxide layer on silicon that is the very basis—the vital foundation—of today's integrated circuit or microchip, and of essentially all the silicon devices so critical to the electronics industry. Based on a long series of experiments and studies, they eventually published an epochal paper entitled "Surface Protection and Selective Masking During Diffusion in Silicon" in the *Journal of The Electrochemical Society*.¹¹ But in June 1955 they had prepared an extensive Bell Labs memorandum to report their results to their colleagues.¹² Much of this text is not the kind that ever gets published in an official paper, but some of it is well worth quoting. For example, on p. 15, they wrote:

Thin silicon slices also were diffused with Sb for N. Holonyak for preliminary device development investigations. These were diffused for 2, 5, and 16 hours respectively at 1300°C in N₂ saturated with water vapor at room temperature. After diffusion, these slices were green in color with an excellent surface appearance. These layers were reported to have resistivities of from 10 to 20 ohms per square. The diffusion layers were reported to be uniform in thickness being 0.26, 0.39 and 0.76 mils respectively for the 3 heating times . . . Holonyak was able to produce very promising cross-point switches from some of these Sb diffusions.

From my earlier discussion, it should be clear how we converted Frosch's diffused-silicon *n-p-n* wafers into functioning *p-n-p-n* switches.^{2,3}

We knew of Frosch's work at first hand because of our exploratory silicon device-development work and thus realized immediately what he and Derick had accomplished. All of us familiar with this work realized it was important, but in truth none of us—least of all Carl, in his exceeding modesty—could envision its enormous future impact. Frosch and Derick even wrote the following statement on p. 20 of their Bell Labs memorandum:¹²

In addition to the possibilities of process simplification, the protective quartz envelope added during the heating may be useful for protecting an electrical device

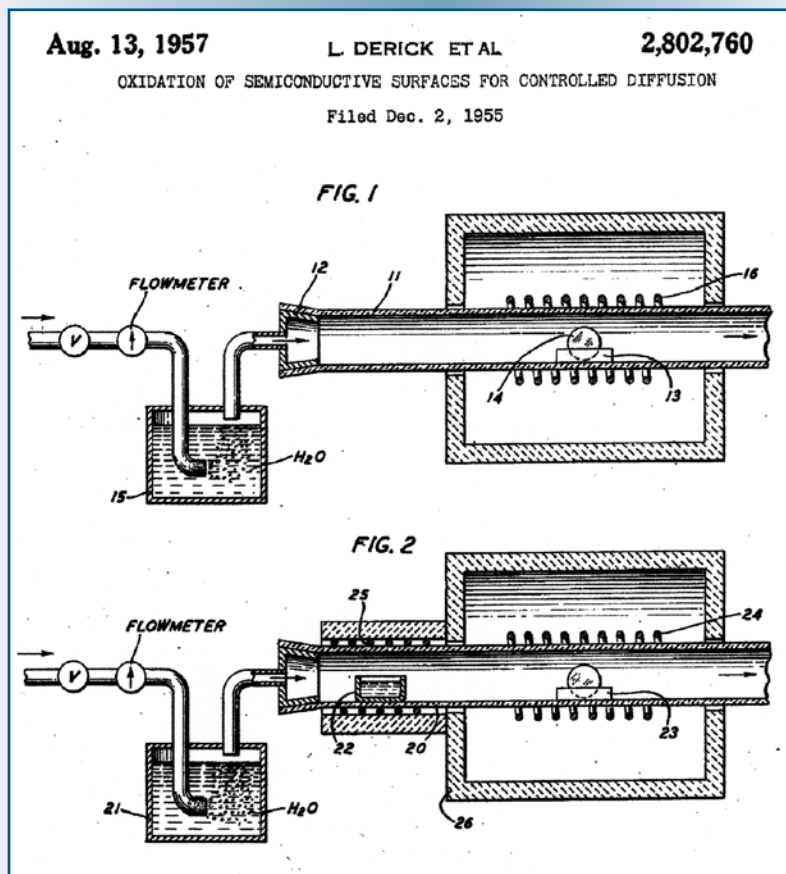


FIG. 4. Cross-sectional drawing of a diffusion furnace from Derick and Frosch's patent, "Oxidation of Semiconductive Surfaces for Controlled Diffusion," U. S. Patent No. 2,802,760, awarded 13 August 1957.

with symmetrical *n*-type diffused regions on both sides. After aluminum evaporation and alloying of a *p*-type emitter on the *n*-type diffused base, a *p-n-p-n* switch resulted.

During the process of making these *n*-type diffusions, particularly the drive-in phase of the donor diffusion at higher temperature in a hot, dry gas (typically hydrogen), Frosch would seriously damage our wafers. They would be eroded and pitted, or even totally destroyed. Every time this happened, the loss was immediately apparent from the expression on his face when he told us the bad news. We would make some adjustments, get more silicon wafers ready, and try again.

One day early in the spring of 1955, I came into Frosch's laboratory and

gas) diffusion. It happened because the exhaust hydrogen accidentally ignited and flashed back into the diffusion chamber, thus making water vapor that covered, reacted with, and protected the silicon with an oxide layer. It was a golden accident.

This wet-ambient process, which was then quickly evaluated and adopted, created a protective silicon-dioxide (SiO₂) layer that could be selectively removed for diffusion into the bare regions, which could then be resealed with oxide for high-temperature annealing or further diffusions (See Fig. 4). Many processing sequences were devised using the protective oxide, which, of course, prevented surface pitting and erosion. Frosch and Derick ascertained which impurities were

from atmospheric conditions. For example, the device might prove more stable if left enclosed in such a quartz envelope. However, it may not be possible to make all of the necessary electrical contacts through the quartz. In these cases some protection may be retained by the removal of a small area of the envelope for the application of the contacts.

As far back as mid-1955, then, Frosch had anticipated modern planar, oxide-protected device processing. He appreciated almost immediately the oxide layer's importance. It is debatable whether any other contribution has had as much to do with the microchip's existence and today's microelectronics industry as the silicon-dioxide layer. One can easily see this by asking the probing question: "Remove the oxide layer, imagine that it doesn't even exist, and then what would be left of silicon technology?" Silicon itself is, of course, the critical ingredient, followed by its unique natural oxide, without which little of today's thriving semiconductor industry would ever have begun to exist.

Further Developments in Diffusion Technology

Although it may appear that everything needed to build diffused-silicon devices had come easily and quickly by the spring of 1955, there is a lot more to the story. For example, silicon itself had some then-unknown defect center that was limiting minority-carrier lifetimes, and George Bemski in Moll's group was trying to understand the problem. He eventually showed that gold was the impurity most responsible.¹³

By the time of Frosch's discovery of oxide protection and masking of silicon, our diffused-device work had reached the stage where it had become necessary to re-evaluate the overall device-development program and decide whether to shift major efforts away from germanium alloyed and grown-junction transistors towards diffused-silicon devices. This was not a trivial decision. What if these devices ultimately proved unreliable? Jack Morton returned home from a European trip and made the crucial decision that March. By then, there just was too much that diffused-silicon device technology offered for it to be denied. A new transistor technology was clearly emerging that would never be replaced, but would only develop further.

We had learned enough about diffused-silicon transistors by the summer of 1955 that Moll asked me to prepare a preliminary design "theory" to permit our colleagues to know what could be expected. With R. M. LeLacheur, I wrote a Bell Labs

memorandum¹⁴ discussing some of the design variables and parameters involved in these transistors. Based on expectations that we could easily realize base thickness of 2.5 micrometers or less, we projected cutoff frequencies above 100 MHz. Diffused-silicon transistors indeed proved capable of such high-speed operation, particularly as the technology was continually refined and transistor dimensions reduced.

The value of oxide masking was appreciated immediately at Bell Labs; it led eventually to patents and licenses issued later to 160 companies. After the licensees were shown the new diffused-silicon device technology at the third Transistor Technology Symposium, held for their benefit in January 1956, a few processing-equipment makers—particularly those selling evaporators and diffusion furnaces—began to notice a spurt in sales. The diffused-silicon device technology was evidently spreading. Perhaps nobody was treated as generously as William Shockley, who was given the new silicon technology to use at his West Coast start-up company, Shockley Semiconductor Laboratory.^{15,16} This ultimately unsuccessful firm has been incorrectly viewed by many observers as the beginning of diffused-silicon devices and the origin of Silicon Valley—not the actual Bell Labs source of the technology.

Besides the features described here, there were other experiments and accidental discoveries that became part of the technology but were not reported. For example, in one experiment I evaporated aluminum on top of the oxide surface of one of the silicon wafers Frosch described as covered with green oxide,¹² expecting to form a junction right through the oxide. Instead I got aluminum metallization on the silicon-dioxide insulating surface; this was immediately obvious because molten aluminum did not contact the silicon, in fact, but was merely deposited from the vapor as "cold" aluminum on the oxide surface. I did not report this result, however, and was later chagrined to find others claim aluminum metallization on the silicon-dioxide layer as their own original research result.

Another area of silicon technology in which Bell Labs made major contributions was photolithography. A few doors down from where we worked in 1954-55, J. Andrus and W. L. Bond were developing photoengraving techniques to define geometrical patterns on metal and semiconductor surfaces.¹⁷ Bond thought that photolithography would have an important role in device processing, and he proved correct. Because of the exploratory nature of our initial diffused-silicon device experiments, however, it would have been a nuisance to have used photolithography in our early work. We had already introduced many unconventional procedures

in making these devices, and photolithography had not yet become necessary. We just wanted to demonstrate feasibility and were not then ready to employ all the processing techniques that are now in hindsight so valuable.

In summary, Frosch and Derick's discovery of the protective oxide layer and its masking was a fundamental, revolutionary development in semiconductor technology. Without it, all the forms of planar, metal-oxide-semiconductor (MOS) devices that came later, and are now so prevalent, would not exist. In the case of the complementary MOS (or CMOS) structure, the breakdown inherent in this form of device is explained by the *p-n-p-n* switching mechanisms we elucidated in 1954-55. And it was probably the silicon *p-n-p-n* switch that more than anything else brought diffused silicon technology to the West Coast, via Shockley and his company.

I am pleased to have worked in John Bardeen's laboratory and then with John Moll, who recognized a revolutionary new silicon technology prior to and at the moment of its creation. I am particularly pleased to recount here the work of Carl Frosch and his great contribution, and to be able to recall Moll's seminal role in guiding and introducing this historic technology. I thank M. Riordan and H. Huff for their interest and editorial advice. ■

About the Author

NICK HOLONYAK, JR. is John Bardeen Professor of Electrical and Computer Engineering and Physics at the University of Illinois (Urbana). Bardeen's first graduate student, Holonyak has worked on semiconductor materials and devices for 55 years. He is recognized for his work on *p-n-p-n* switches, tunnel diodes, the III-V (visible) alloy laser and LED, the quantum-well laser, impurity-induced heterolayer intermixing, the Al-based III-V oxide, and the transistor laser. He has received many awards, including the IEEE Medal of Honor, the Japan Prize, the Global Energy Prize, and the U.S. Medals of Science and of Technology. He was the 1983 recipient of the ECS Gordon E. Moore Medal for Outstanding Achievement in Solid-State Science and Technology (formerly the Solid State Science & Technology Award). He may be reached at blpayne@uiuc.edu.

(continued on next page)

