The invention of the transistor

Michael Riordan

Department of Physics, University of California, Santa Cruz, California 95064

Lillian Hoddeson

Department of History, University of Illinois, Urbana, Illinois 61801

Conyers Herring

Department of Applied Physics, Stanford University, Stanford, California 94305

[S0034-6861(99)00302-5]

Arguably the most important invention of the past century, the transistor is often cited as the exemplar of how scientific research can lead to useful commercial products. Emerging in 1947 from a Bell Telephone Laboratories program of basic research on the physics of solids, it began to replace vacuum tubes in the 1950s and eventually spawned the integrated circuit and microprocessor—the heart of a semiconductor industry now generating annual sales of more than $150 billion. These solid-state electronic devices are what have put computers in our laps and on desktops and permitted them to communicate with each other over telephone networks around the globe. The transistor has aptly been called the ''nerve cell'' of the Information Age.

Actually the history of this invention is far more involved and interesting than given by this ''linear'' account, which overlooks the intricate interplay of scientific, technological, social, and personal interests and developments. These and many other factors contributed to the invention of not one but two distinctly different transistors—the point-contact transistor by John Bardeen and Walter Brattain in December 1947, and the junction transistor by William Shockley a month later.[1] The point-contact transistor saw only limited production and never achieved commercial success. Instead, it was the junction transistor that made the modern semiconductor industry possible, contributing crucially to the rise of companies such as Texas Instruments, SONY, and Fairchild Semiconductor.

Given the tremendous impact of the transistor, it is surprising how little scholarship has been devoted to its history.[2] We have tried to fill this gap in recent publications (Herring, 1992; Riordan and Hoddeson, 1997a, 1997b). Here we present a review of its invention, emphasizing the crucial role played by the postwar understanding of solid-state physics. We conclude with an analysis of the impact of this breakthrough upon the discipline itself.

I. PRELIMINARY INVESTIGATIONS

The quantum theory of solids was fairly well established by the mid-1930s, when semiconductors began to be of interest to industrial scientists seeking solid-state alternatives to vacuum-tube amplifiers and electromechanical relays. Based on the work of Felix Bloch, Rudolf Peierls, and Alan Wilson, there was an established understanding of the band structure of electron energies in ideal crystals (Hoddeson, Baym, and Eckert, 1987; Hoddeson *et al.*, 1992). This theory was then applied to calculations of the energy bands in real substances by groups of graduate students working with Eugene Wigner at Princeton and John Slater at MIT. Bardeen and Frederick Seitz, for example, wrote dissertations under Wigner, calculating the work function and band structure of sodium; studying with Slater, Shockley determined the band structure of sodium chloride (Bardeen, 1936; Shockley, 1936; Herring, 1992). By the mid-1930s the behavior of semiconductors was widely recognized to be due to impurities in crystals, although this was more a qualitative than quantitative understanding. The twin distinctions of ''excess'' and ''defect'' semiconductors could be found in the literature; their different behavior was thought to be the result of electrons added to the conduction band or removed from the valence band by impurity atoms lodged in the crystal lattice (Wilson, 1931; Mott and Jones, 1936).

There were a few solid-state electronic devices in use by the mid-1930s, most notably the copper-oxide rectifier, on which Brattain worked extensively at Bell Labs during that period (Brattain, 1951). Made by growing an oxide layer on copper, these rectifiers were used in AC-to-DC converters, in photometers and as ''varistors'' in telephone circuitry made for the Bell System. But the true nature of this rectification, thought to occur at the interface between the copper and copper-oxide layers, was poorly understood until the work of Nevill Mott (1939) and Walther Schottky (1939) showed the phenomenon to be due to the establishment of an asymmetric potential barrier at this interface. In late 1939 and early 1940, Shockley and Brattain tried to fabricate a

[1]This paper is based in large part on Riordan and Hoddeson (1997a). The best scholarly historical account of the point-contact transistor is that of Hoddeson (1981); on the invention of the junction transistor, see Shockley (1976).

[2]In addition to the above references, see Bardeen (1957), Brattain (1968), Shockley (1973,1976), Weiner (1973), Holonyak (1992), Riordan and Hoddeson (1997b), Ross (1998), and Seitz and Einspruch (1998b). Scholarly books that cover the topic well include those of Braun and MacDonald (1978) and Seitz and Einspruch (1998a).

solid-state amplifier by using a third electrode to modulate this barrier layer, but their primitive attempts failed completely.

One of the principal problems with this research during the 1930s was that the substances generally considered to be semiconductors were messy compounds such as copper oxide, lead sulfide, and cadmium sulfide. In addition to any impurities present, there could be slight differences from the exact stoichiometric ratios of the elements involved; these were extremely difficult, if not impossible, to determine and control at the required levels. Semiconductor research therefore remained more art than science until World War II intervened.

During the War, silicon and germanium rose to prominence as the preferred semiconductors largely through the need for crystal rectifiers that could operate at the gigahertz frequencies required for radar receivers. Driven by this requirement, the technology of these two semiconductor materials advanced along a broad front (Torrey and Whitmer, 1948). Where before the War it was difficult to obtain silicon with impurity levels less than one percent, afterwards the DuPont Company was turning out 99.999 percent pure silicon (Seitz, 1994, 1995; Seitz and Einspruch, 1998a). The technology of doping silicon and germanium with elements from the third and fifth columns of the periodic table (such as boron and phosphorus) to produce *p*-type and *n*-type semiconductor materials had become well understood. In addition, the *p*-*n* junction had been discovered in 1940 at Bell Labs by Russell Ohl—although its behavior was not well understood, nor was it employed in devices by War's end (Scaff and Ohl, 1947; Scaff, 1970; Riordan and Hoddeson, 1997a, 1997c).

There was also extensive research on semiconductors in the Soviet Union during the same period, but this work does not seem to have had much impact in the rest of Europe and the United States (Herring, Riordan, and Hoddeson, n.d.). Of course, contributions of well-known theorists, such as Igor Tamm on surface-bound electron levels and Yakov Frenkel on his theory of excitons, attracted wide interest (Tamm, 1932; Frenkel, 1933, 1936); published in German and English, they were quickly incorporated into the corpus of accepted knowledge.

But the work of Boris Davydov on rectifying characteristics of semiconductors seems to have eluded notice until after the War, even though it was available in English-language publications (Davydov, 1938). Working at the Ioffe Physico-Technical Institute in Leningrad, he came up with a model of rectification in copper oxide in 1938 that foreshadowed Shockley's work on *p*-*n* junctions more than a decade later. His idea involved the existence of a *p*-*n* junction in the oxide, with adjacent layers of excess and deficit semiconductor forming spontaneously due to an excess or deficit of copper relative to oxygen in the crystal lattice. Nonequilibrium concentrations of electrons and holes—positively charged quantum-mechanical vacancies in the valence band—could survive briefly in each other's presence before recombining. Using this model, Davydov successfully derived the current-voltage characteristics of copper-oxide rectifiers; his formula was essentially the same as the one that Shockley would derive a decade later for *p*-*n* junctions (Shockley, 1949). But his cumbersome mathematics and assumptions may have obscured the importance of his physical ideas to later workers. Bardeen, for example, was aware of Davydov's publications by 1947 but does not seem to have recognized their significance until a few years later.

II. THE INVENTION OF THE POINT-CONTACT TRANSISTOR

Both the point-contact transistor and the junction transistor emerged from a program of basic research on solid-state physics that Mervin Kelly, then Bell Labs Executive Vice President, initiated in 1945. He recognized that the great wartime advances in semiconductor technology set the stage for electronic advances that could dramatically improve telephone service. In particular, he was seeking solid-state devices to replace the vacuum tubes and electromechanical relays that served as amplifiers and switches in the Bell Telephone System (Hoddeson, 1981; Riordan and Hoddeson, 1997a). He had learned valuable lessons from the wartime efforts at Los Alamos and the MIT Radiation Laboratory, where multidisciplinary teams of scientists and engineers had developed atomic bombs and radar systems in what seemed a technological blink of an eye (Hoddeson *et al.*, 1993).

Kelly perceived that the new quantum-mechanical understanding of solids could be brought to bear on semiconductor technology to solve certain problems confronting his company. ''Employing the new theoretical methods of solid state quantum physics and the corresponding advances in experimental techniques, a unified approach to all of our solid state problems offers great promise,'' he wrote that January. ''Hence, all of the research activity in the area of solids is now being consolidated . . . .'' (Riordan and Hoddeson, 1997a, pp. 116–117.) At the helm of this Solid State Physics Group he put Shockley and chemist Stanley Morgan. Soon Brattain and Bardeen joined a semiconductor subgroup within it headed by Shockley.

While planning the new solid-state group in April 1945, Shockley proposed a device now called the ''field-effect'' transistor (Shockley, 1976; Hoddeson, 1981). Here an externally applied transverse electric field is arranged so that it can increase or decrease the number of charge carriers in a thin film of silicon or germanium, thus altering its conductivity and regulating the current flowing through it. By applying suitable voltages to two circuit loops passing through this semiconductor material, Shockley predicted that an input signal applied to one loop could yield an amplified signal in the other. But several attempts to fabricate such a field-effect device in silicon failed. So did Shockley's theoretical attempt to explain why, on the basis of Mott and Schottky's rectification theory, his conceptual field-effect device did not work as predicted (Hoddeson, 1981, pp. 62–63).

In October 1945 Shockley asked Bardeen, who had just joined the group, to check the calculations that he had made in an attempt to account for the failure of his field-effect idea. By March 1946 Bardeen had an answer. He explained the lack of significant modulation of the conductivity using a creative heuristic model, based on the idea of "surface states" (Bardeen, 1947). In this model, electrons drawn to the semiconductor surface by the applied field become trapped in these localized states and are thus unable to act as charge carriers.[3] As Shockley (1976, p. 605) later recalled, the surface states "blocked the external field at the surface and . . . shielded the interior of the semiconductor from the influence of the positively charged control plate."

But were these postulated states real? If so, how did they generally behave? These questions became intensely interesting to the Bell Labs semiconductor group, which in the following months responded to Bardeen's surface-state idea with an intensive research program to explore this phenomenon. Bardeen worked closely on the problem with the group's experimental physicists, Brattain and Gerald Pearson.[4]

On 17 November 1947, Brattain made an important discovery. Drawing on a suggestion by Robert Gibney, a physical chemist in the group, he found that he could neutralize the field-blocking effect of the surface states by immersing a silicon semiconductor in an electrolyte (Brattain, 1947b, pp. 142–151; 1968). "This new finding was electrifying," observed Shockley (1976, p. 608); "At long last, Brattain and Gibney had overcome the blocking effect of the surface states." Their discovery set in motion events that would culminate one month later in the first transistor.

Four days after this discovery, Bardeen and Brattain tried to use the results to build a field-effect amplifier. Their approach was based on Bardeen's suggestion to use a point-contact electrode pressed against a specially prepared silicon surface. Rather than the thin films employed in the 1945 experiments by Shockley and his collaborators, Bardeen proposed the use of an *n*-type "inversion layer" a few microns thick that had been chemically produced on the originally uniform surface of *p*-type silicon. Because charge carriers—in this case, electrons—would have higher mobility in such an inversion layer than they had in vapor-deposited films, Bardeen believed that this approach would work better in a field-effect amplifier (Bardeen, 1957). In particular, this layer would act as a shallow channel in which the population of charge carriers could be easily modulated by an applied external field. The device tested on 21 November used a drop of electrolyte on the surface as one contact and the metal point as the other; Bardeen and Brattain obtained a small but significant power amplification, but the device's frequency response was poor (Bardeen, 1946, pp. 61–70).

The next crucial step occurred on 8 December. At Bardeen's suggestion, Brattain replaced the silicon with an available slab of *n*-type, "high-back-voltage" germanium, a material developed during the wartime radar program by a research group at Purdue directed by Karl Lark-Horovitz (Henriksen, 1987). They obtained a power gain of 330—but with a negative potential applied to the droplet instead of positive, as they had expected. Although the slab had not been specially prepared, Bardeen proposed that an inversion layer was being induced electrically, by the strong fields under the droplet. "Bardeen suggests that the surface field is so strong that one is actually getting *P* type conduction near the surface," wrote Brattain (1947b, pp. 175–176) that day, "and the negative potential on the grid is increasing the *P* type or hole conduction."[5] This was a crucial perception on Bardeen's part, that holes were acting as charge carriers within a slab of *n*-type germanium.

Later that week Brattain evaporated a gold plate onto a specially prepared germanium slab that already had an inversion layer. In an attempt to improve the frequency response by eliminating the sluggish droplet, he employed instead a thin germanium-oxide layer grown on the semiconductor surface. He thought the gold would be insulated from the germanium by this layer, but unknown to him the layer had somehow been washed away, and the plate was now directly in contact with germanium. This serendipitous turn of events proved to be a critical step toward the point-contact transistor (Hoddeson, 1981).

The following Monday, 15 December, Bardeen and Brattain were surprised to discover that they could still modulate the output voltage and current at a point contact positioned close to the gold plate, but only when the plate was biased *positively*—the opposite of what they had expected![6] "An increase in positive bias *increased* rather than decreased the reverse current to the point contact," wrote Bardeen (1957) ten years later. This finding suggested "that holes were flowing into the germanium surface from the gold spot and that the holes introduced in this way flowed into the point contact to enhance the reverse current. This was the first indication of the transistor effect."

---

[3]Previous work on surface states had been done by Tamm (1932) and Shockley (1939). Bardeen, however, was the one who applied these ideas to understanding the surface behavior of semiconductors (Bardeen, 1946, pp. 38–57; 1947).

[4]The research program is described in the laboratory notebooks of Bardeen (1946), Brattain (1947b, 1947c), Pearson (1947), and Shockley (1945); it is summarized by Hoddeson (1981). The sequence of steps to the point-contact transistor detailed here largely follows the account in Hoddeson (1981) and Riordan and Hoddeson (1997a).

[5]This was the first recorded instance we can find in which Bardeen and Brattain recognized the possibility that holes were acting as charge carriers. Note that Bardeen still proposed that the flow occurred within a shallow inversion layer at the semiconductor surface.

[6]Hoddeson (1981, p. 72) states that this event occurred on Thursday, 11 December. A closer examination of Brattain (1947b, pp. 183–92) indicates that there was a period of confusion followed by the actual breakthrough on 15 December. See Riordan and Hoddeson (1997a), Chapter 7, for a more complete discussion of this sequence of events.

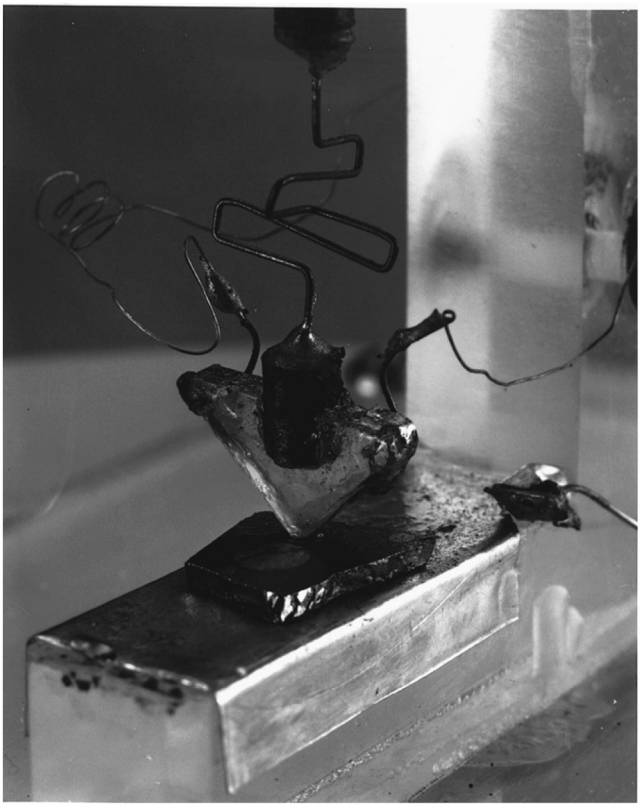

FIG. 1. Photograph of the point-contact transistor invented by Bardeen and Brattain in December 1947. A strip of gold foil slit along one edge is pressed down into the surface of a germanium slab by a polystyrene wedge, forming two closely spaced contacts to this surface. (Reprinted by permission of AT&T Archives.)

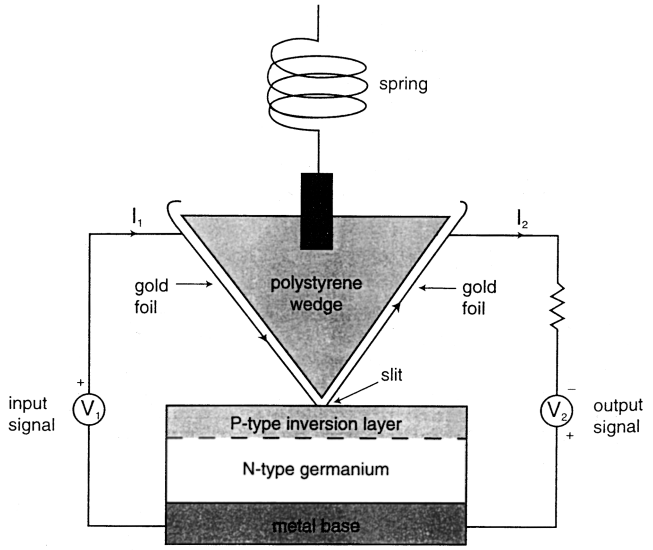

FIG. 2. Schematic diagram of the first transistor (Fig. 1). The signal current I_1 flows through the input circuit, generating holes in a *p*-type inversion layer that modulate the flow of current I_2 in an output circuit. (Reprinted from M. Riordan and L. Hoddeson, *Crystal Fire*.)

Although Brattain and Bardeen failed to observe power amplification with this configuration, Bardeen suggested that it would occur if two narrow contacts could be spaced only a few thousandths of an inch apart. Brattain (1947b, pp. 192–93) achieved the exacting specifications by wrapping a piece of gold foil around one edge of a triangular polystyrene wedge and slitting the foil carefully along that edge. He then pressed the wedge—and the two closely spaced gold contacts—down into the surface of the germanium using a makeshift spring (see Figs. 1 and 2). In their first tests, made on 16 December, the device worked as expected. It achieved both voltage and power gains at frequencies up to 1000 Hz. The transistor had finally been born. A week after that, on 23 December 1947, the device was officially demonstrated to Bell Labs executives in a circuit that allowed them to hear amplified speech in a pair of headphones (Brattain, 1947c, pp. 6–8; Hoddeson, 1981).

III. THE FLOW OF CHARGE CARRIERS

An important issue that has engendered much recent debate is how Bardeen and Brattain conceptualized the flow of charge carriers while they were developing the first transistor. Memory is imperfect, and later accounts are often subject to what is called "retrospective realism,"[7] a process whereby conjectures become imbued with an aura of certainty, or embellished with details that became known only at a later time. Fortunately, we have available several telling entries that Bardeen, Brattain and Shockley made in their laboratory notebooks during those pivotal weeks before and after Christmas 1947.[8]

On 19 December, three days after the first successful test of their device, Brattain (1947c, p. 3) wrote: "It would appear then that the modulation obtained when the grid point is bias+is due to the grid furnishing holes to the plate point." By grid point and plate point, he was referring to what we now call the emitter and collector: he was obviously using a familiar vacuum-tube analogy. Although we cannot determine from this passage exactly how he conceived the details of their flow, we can be sure he understood that holes were responsible for modulation.

Bardeen gave a more detailed explanation in a notebook entry on 24 December, the day after the team made its official demonstration. After describing their setup, which used a slab of *n*-type germanium specially prepared to produce a shallow inversion layer of *p*-type conductivity near its surface (see Fig. 3), he portrayed the phenomenon as follows (Bardeen, 1946, p. 72):

> When *A* is positive, holes are emitted into the semi-conductor. These spread out into the thin *P*-type layer. Those which come in

[7]This phrase and concept is due to Pickering (1984).

[8]Some of the entries in Brattain's notebooks during those critical weeks in December 1947 are written in Bardeen's handwriting. The two obviously were working side by side in the laboratory.

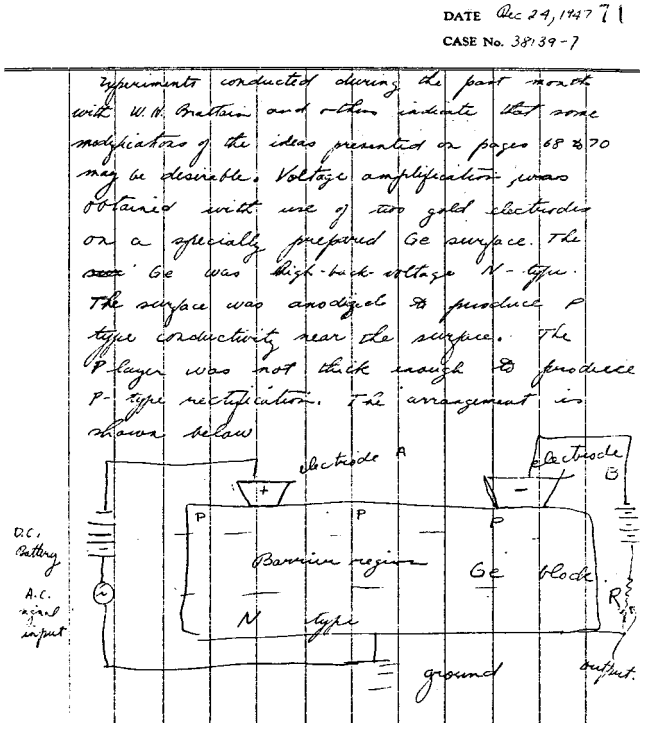

FIG. 3. Entry in Bardeen's lab notebook dated 24 December 1947, giving his conception of how the point-contact transistor functions. (Reprinted by permission of AT&T Archives.)

> the vicinity of *B* are attracted and enter the electrode. Thus *A* acts as a cathode and *B* as a plate in the analogous vacuum tube circuit.

Again it is clear that Bardeen also attributed the transistor action to the holes, but he went a step farther and stated that the flow of these holes occurs within the inversion layer.

This emerging theory of the transistor based on the flow of holes at or near the surface of the germanium developed further during the following months, the period in which Bell Labs kept the discovery of the transistor "laboratory secret," while patent applications were being drawn up. A drawing found in Bardeen and Brattain's patent application of 17 June 1948 (revised from a version submitted on 25 February) suggests that although the flow of charge carriers was thought to occur largely within the *p*-type inversion layer, they were by this time allowing that some holes might diffuse through the body of the *n*-type germanium. They state (Bardeen and Brattain, 1948a):

> . . . potential probe measurements on the surface of the block, made with the collector disconnected, indicate that *the major part* of the emitter current travels on or close to the surface of the block, substantially laterally in all directions away from the emitter . . . .

In a famous letter submitted to the *Physical Review* on 25 June 1948, they wrote (Bardeen and Brattain, 1948b) that as a result of the existence of the shallow *p*-type inversion layer next to the germanium surface, "the current in the forward direction with respect to the block is composed in large part of holes, i.e., of carriers of sign opposite to those normally in excess in the body of the block." In a subtle shift from their earlier conception, they envisioned that holes flow predominantly in the *p*-type inversion layer, but with a portion that can also flow through the *n*-type layer beneath it.

It is not clear from these entries just how and why this shift occurred. But both the revised patent application and the *Physical Review* letter are dated well after Shockley's conception of the junction transistor in late January and a crucial mid-February experiment (discussed below) by John Shive.

IV. THE CONCEPTION OF THE JUNCTION TRANSISTOR

During the weeks that followed the invention of the point-contact transistor, Shockley was torn by conflicting emotions. Although he recognized that Bardeen and Brattain's invention had been a "magnificent Christmas present" to Bell Labs, he was chagrined that he had not had a direct role to play in this obviously crucial breakthrough. "My elation with the group's success was tempered by not being one of the inventors," he recalled a quarter century later (Shockley, 1976). "I experienced frustration that my personal efforts, started more than eight years before, had not resulted in a significant inventive contribution of my own."

Since the failure of his field-effect idea more than two years earlier, Shockley had paid only passing attention to semiconductor research. During the months before the invention, he had mainly been working on the theory of dislocations in solids. He had, however, thought about the physics of *p*-*n* junctions and their use in such practical devices as lightning arrestors and high-speed thermistors (Shockley, 1945, pp. 71, 76–78, 80, 88–89).

Brattain and Gibney's discovery in November 1947 stimulated Shockley's thinking. A few days after that he suggested fabricating an amplifier using a drop of electrolyte deposited across a *p*-*n* junction in silicon or germanium; this approach worked when Brattain (1947b, pp. 169–70) and Pearson (1947, p. 75) tried it. On 8 December 1947, more than a week before the point-contact transistor was invented. Shockley (1945, p. 91) outlined an idea in his laboratory notebook for an *n*-*p*-*n* sandwich that had current flowing laterally in the interior *p*-layer and with the *n*-layers around it acting as control electrodes.

The 16 December invention of the point-contact transistor and Bardeen's interpretation of its action in terms of the flow of holes galvanized Shockley into action. Bardeen's above-quoted analogy with the operation of a vacuum tube—in which the current carriers were holes instead of electrons—was in fact due to Shockley,[9] who applied it in his first attempt at a junction transistor, written in a room in Chicago's Hotel Bismarck on New

[9]Bardeen (1946) credits Shockley with this suggestion in his notebook on p. 72.

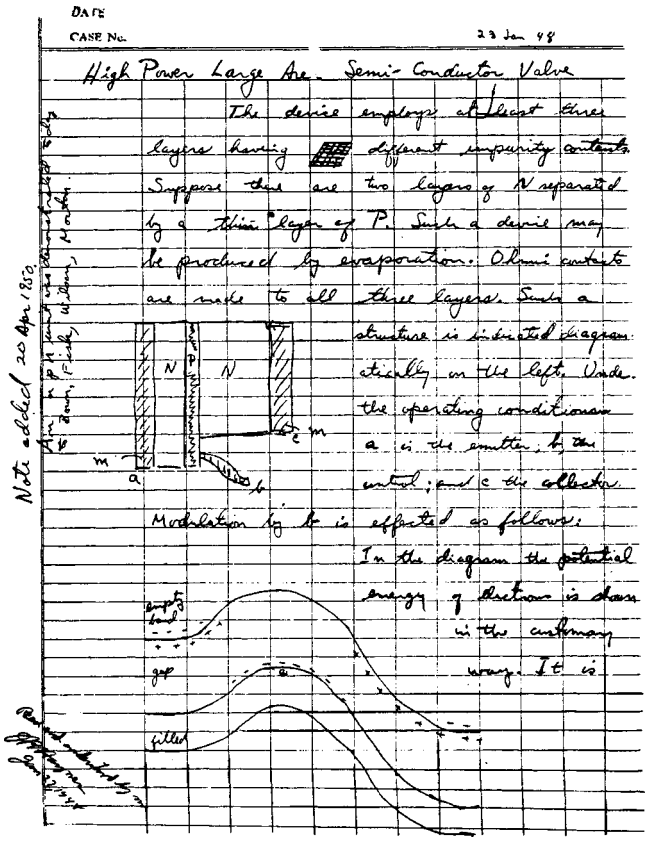

FIG. 4. Entry in Shockley's lab notebook dated 23 January 1948 recording his conception of the junction transistor. He wrote this page at home on a piece of paper, which he later pasted into his notebook. (Reprinted by permission of AT&T Archives.)

Year's Eve of 1947. In this first stab at a junction transistor, one can see a clear analogy with a vacuum tube; its "control" electrode acts as a grid to control the flow of holes from a "source" to a "plate" (Shockley, 1945, pp. 110–13). On this disclosure of a *p-n-p* device, however, Shockley (1976) admitted that he had "failed to recognize the possibility of minority carrier injection into a base layer . . . . What is conspicuously lacking [in these pages] is any suggestion of the possibility that holes might be injected into the *n*-type material of the strip itself, thereby becoming minority carriers in the presence of electrons."

A little more than three weeks later, this time working at his home on the morning of 23 January, 1948, Shockley conceived another design in which *n*-type and *p*-type layers were reversed and electrons rather than holes were the current carriers (see Fig. 4). Applying a positive voltage to the interior *p*-layer should lower its potential for electrons; this he realized would "increase the flow of electrons over the barrier exponentially" (Shockley, 1945, p. 129). As Shockley (1976) observed nearly thirty years later, this *n-p-n* sandwich device finally contained the crucial concept of "exponentially increasing minority carrier injection across the emitter junction." Minority carriers, in this case the electrons, had to flow in the presence of the dominant majority carriers—the holes of the *p*-type layer.

V. A CRUCIAL EXPERIMENT

Almost another month passed before Shockley revealed his breakthrough idea to anyone in his group other than physicist J. Richard Haynes, who witnessed the entry in his logbook. Why did Shockley keep the information to himself? Did he recognize that he had made a major conceptual advance but decide to keep it quiet to give himself more time to follow up its theoretical and practical ramifications? Was he afraid that Bardeen and Brattain were so close to making a similar discovery themselves that knowledge of his idea would push them to publish before him?[10] Or was he simply so unsure of the idea that he avoided discussing it with them until he could think about it further? We do not know.

In order to function, Shockley's *n-p-n* device required additional physics beyond that involved in the point-contact transistor. It was crucial to understand that minority carriers are able to diffuse *through* the base layer in the presence of majority carriers. Bardeen may have in fact had such an understanding, but it is not obvious from his logbook entries at the time. And at the time, he and Brattain were preoccupied with preparing patent documents dealing with their point-contact device. They still apparently believed that nearly all the hole flow occurred in a micron-deep *p*-type layer at the semiconductor surface.

Evidence for the required diffusion of the minority carriers into the bulk material was not long in coming. In a closed meeting at Bell Labs on 18 February 1948, physicist John Shive revealed that he had just tested a successful point-contact transistor using a very thin wedge of *n*-type germanium, but with the emitter and the collector placed on the *opposite* faces of the wedge (Shive, 1948, pp. 30–35). At the position where the two contacts touched it, the wedge was only 0.01 cm thick, while the distance between these points along the germanium surface was much larger. Shockley immediately recognized what this revelation meant. In this geometry, the holes had to flow by diffusion in the presence of the majority carriers, the electrons in the *n*-type germanium, through the bulk of the semiconductor; they were not confined to an inversion layer on the surface, as Bardeen and Brattain had been suggesting occurred in their con-

[10]Nick Holonyak gave a reasonable argument that once the point-contact transistor and the notion of minority carrier flow were on hand, the *p-n* junction transistor was "bound to follow." He recalled a dinner in the mid-1980s in Urbana at which Bardeen stated that he and Brattain had planned to move on to that device as soon as they completed their time-consuming work of preparing patents for the original transistor, only to find that Shockley had already tied up this area of work with the Bell Labs patent attorney (who, in John's words, "was in Shockley's pocket"). Holonyak interview by L. Hoddeson, 10 January 1992.

traption. ''As soon as I had heard Shive's report,'' Shockley (1976) recalled, ''I presented the ideas of my junction transistor disclosure and used them to interpret Shive's observation.''

This experiment may have injected a heady dose of urgency into the Bell Labs solid-state physics group. On 26 February, the company applied for four patents on semiconductor amplifiers, including Bardeen and Brattain's original application on the point-contact transistor. Their two landmark papers, ''The Transistor, a Semi-Conductor Triode'' (Bardeen and Brattain, 1948b) and ''Nature of the Forward Current in Germanium Point Contacts'' (Brattain and Bardeen, 1948), were sent to the *Physical Review* four months later, on 25 June. One day later, Bell applied for Shockley's patent on the junction transistor, and on 30 June it announced the invention of the transistor in a press conference.

In July 1948 Shockley proved that hole ''injection'' (as he dubbed the flow of minority carriers in transistor action) was indeed occurring in *n*-type germanium. Working with Haynes, he showed that the charge carriers traveling from the emitter to the collector were in fact ''positive particles with a mobility of about 1.2 $\times 10^3$ cm^2/volt-sec'' (Haynes and Shockley, 1949). Their paper was published in early 1949 together with Shive's article (Shive, 1949) on the two-sided transistor. Much of this research was discussed in detail in *Electrons and Holes in Semiconductors, with Applications to Transistor Electronics* (Shockley, 1950), which became the bible of the new discipline.

Shockley had another blind spot to overcome in his thinking about minority carriers before it finally became possible to fabricate working junction transistors. One of the problems behind the failure of his field-effect transistor had been how slowly charge carriers diffused through the polycrystalline silicon and germanium films used in the early experiments. Gordon Teal, a Bell Labs physical chemist, recognized the merits of using single crystals of germanium and silicon (Teal, 1976; Goldstein, 1993). He realized that in polycrystalline films minority carriers cannot survive long enough to make it from emitter to collector in sufficient numbers, but that they would have lifetimes 20 to 100 times longer in single crystals. Teal tried to convince Shockley of this critical advantage, but Shockley ignored his suggestion.

Fortunately Jack Morton, an engineer who headed the Bell Labs efforts to develop the point-contact transistor into a commercially viable product, took Teal seriously and in late 1949 gave him a small amount of support to pursue this avenue. Working with physical chemist Morgan Sparks, Teal modified the crystal-growing machine that he and a colleague had developed for pulling single crystals out of molten germanium (Goldstein, 1993). This alteration allowed them to dope the germanium in a controlled manner and thereby fabricate the first practical *n*-*p*-*n* junction transistor in April 1950. On the date of its demonstration, 20 April 1950, Shockley (1945, p. 128) penned a note in the margin of his 23 January 1948 entry (see Fig. 4): ''An *n*-*p*-*n* unit was demonstrated today to Bown, Fisk, Wilson, Morton.'' [11]

VI. CONCLUSIONS

On 10 December 1956 Shockley, Bardeen, and Brattain (in that order) were awarded the Nobel prize in physics for their ''investigations on semi-conductors and the discovery of the transistor effect.'' [12] Taken together, their physical insights into the flow of electrons and holes in the intimate presence of one another were what made the invention of the transistor possible. Following Brattain's initial experiment indicating that the surface states could be overcome, Bardeen recognized in early December 1947 that holes could flow *as minority carriers* in a surface layer on a slab of *n*-type germanium; they employed this understanding to invent the point-contact transistor. But the possibility of minority-carrier *injection* into the bulk of the semiconductor, which made the junction transistor feasible, apparently occurred first to Shockley. In 1980 Bardeen reflected on the two interpretations of transistor action:

> The difference between ourselves [Bardeen and Brattain] and Shockley came in the picture of how the holes flow from the emitter to the collector. They could flow predominantly through the inversion layer at the surface, which does contain holes. And the collector would be draining out the holes from the inversion layer. They could also flow through the bulk of the semiconductor, with their charge compensated by the increased number of electrons in the bulk . . . .[13]

It was this detailed understanding of semiconductor physics, which emerged in the course of a basic research program at Bell Labs, that overcame the barriers that had foiled all previous attempts to invent a solid-state amplifier.

It is important to recognize, however, that this physical insight was applied to a new technological base that had emerged from World War II. The very meaning of the word ''semiconductor'' changed markedly during that global confrontation. Where before the War, scientists commonly used the word to refer to compounds

[11]This work was published (Shockley, Sparks, and Teal, 1951) in *Physical Review* over a year later, after a microwatt junction transistor operating at 10 kHz had been announced to the press. For the full story of the invention and development of the junction transistor, see Riordan and Hoddeson (1997a), Chapter 8.

[12]Quoted from Felix Belair, Jr., ''Nobel Physics Prize Goes to 3 Americans; 2 Chemists Honored,'' *The New York Times*, 2 November 1956, p. 1.

[13]Interview of Bardeen by L. Hoddeson, 13 February 1980 (AIP Niels Bohr Library archives, College Park, MD), p. 2.

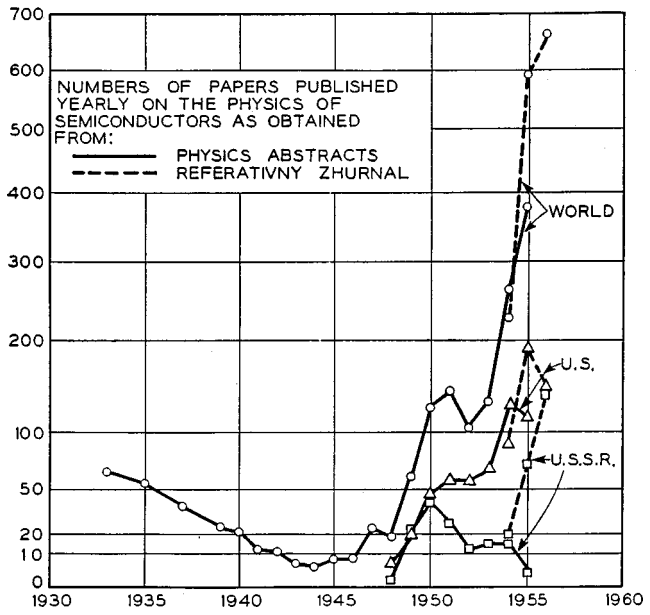

FIG. 5. Variation with time in the annual numbers of papers on semiconductor physics listed in *Physics Abstracts* and (for 1954–56) in the Russian publication *Refarativny Zhurnal*. Symbols represent the number of papers from the United States, the Soviet Union, and the entire world, as indicated. The correspondence to actual publication rates is only rough, as the abstract journals fluctuate in the breadth of their coverage and in the time lag from publication of the papers to appearance of the abstracts.

such as copper oxide, lead sulfide (or galena), and cadmium sulfide, afterwards it meant silicon and germanium doped with small amounts of highly controllable impurities. These crucial technological advances were mainly due to the work of physical chemists and electrochemists working in relative obscurity (Scaff and Ohl, 1947; Scaff, 1970; Seitz, 1995; Seitz and Einspruch, 1998a). Thus the ''linear model'' of technological development—wherein scientific research precedes technological development, from which useful products emerge—does not encompass very well what happened in the case of the transistor.

This new technology and the invention of the transistor have influenced the progress of science in many ways through the revolutionary impact of computers and electronic information processing. A more immediate impact was the stimulus on the field of solid-state physics that came in the few years after the breakthrough; a rough measure of this stimulus is given by the publication statistics plotted in Fig. 5 (Herring, 1957). The publication rate for research papers in all fields of physics showed a sizable decline during the combat years followed by a postwar recovery to a level above the prewar rate—as one might expect due to lessened monetary and manpower resources during the War, followed by eventual return to a slowly expanding peacetime rate. In contrast, the publication rate in semiconductor physics suffered a gradual decline even in the pre-war years and almost disappeared during the War, but it recovered to a nearly level value in the period 1950–53 and then rose again spectacularly. We can reasonably attribute this great burst of activity to an increase in the number of people working in the field, and in the industrial and governmental support for such research.[14]

The research that led to the transistor had a psychological and intellectual impact that not only accelerated the growth of semiconductor research but also stimulated work in other areas of solid-state physics. Like semiconductors, most of these areas had seemed ''dirty'' to many physicists because relevant measurements were sensitive to factors such as the purity of materials, cleanliness of surfaces, and perfection of crystals, which made the phenomena too complicated to be understood in terms of simple theories. The research involved in the invention and development of the transistor showed that materials and experimental conditions could indeed be controlled, after all, and that many phenomena, such as the behavior of p-n junctions could be interpreted quantitatively using soundly based theories. Awareness of these advances was probably a major factor in the enthusiasm and resulting wave of publication that swept through the solid-state community in the early 1950s.[15]

The transistor discovery has clearly had enormous impact, both intellectually and in a commercial sense, upon our lives and work. A major vein in the corpus of condensed-matter physics quite literally owes its existence to this breakthrough. It also led to the microminiaturization of electronics, which has permitted us to have powerful computers on our desktops that communicate easily with each other via the Internet. The resulting globalization of science, technology, and culture is now transforming the ways we think and interact.

ACKNOWLEDGMENTS

We thank William Brinkman, Nick Holonyak, Howard Huff, and Frederick Seitz for helpful discussions. This work was supported in part by grants from the Alfred P. Sloan Foundation and the Richard Lounsbery Foundation.

[14]Note that a similar rise appears in the world totals of papers on semiconductor physics, as listed by the Soviet abstract publication *Referativny Zhurnal*. But when one compares the curves for papers published in the United States with those in the Soviet Union, the rise in the latter seems to begin about a year later, a modest delay in view of the poor communication between scientists of the two countries during the Stalin years.

[15]There were, of course, other favorable influences, such as the new availability of microwave tools. And the Cold War confrontation of the United States and Soviet Union probably also played a part. Another indication of the change in perspective can be seen in some of the statistics on ten-minute papers presented at meetings of the American Physical Society on non-semiconductor solid-state work done at governmental or industrial laboratories (other than Bell Labs): in 1949–50, about a sixth of such papers were theoretical; in 1956, about a third.

REFERENCES

Bardeen, J., 1936, "Theory of the work function II: the surface double layer," Phys. Rev. **49**, 653-663.

Bardeen, J., 1946, Bell Labs Notebook No. 20780 (AT&T Archives, Warren, NJ).

Bardeen, J., 1947, "Surface states and rectification at a metal-semi-conductor contact," Phys. Rev. **71**, 717-727.

Bardeen, J., 1957, "Semiconductor research leading to the point-contact transistor," in *Les Prix Nobel en 1956*, edited by K. M. Siegbahn *et al.* (P. A. Nordstet & Sons, Stockholm), pp. 77–99; edited and reprinted in Science **126**, 105-112.

Bardeen, J., and W. H. Brattain, 1948a, "Three-electrode circuit element utilizing semiconductive materials," US Patent No. 2,524,035 (Washington, DC).

Bardeen, J., and W. H. Brattain, 1948b, "The transistor, a semi-conductor triode," Phys. Rev. **74**, 230-231.

Bardeen, J., and W. H. Brattain, 1949, "Physical principles involved in transistor action," Phys. Rev. **75**, 1208-1225.

Brattain, W. H., 1947a, "Evidence for surface states on semiconductors from change in contact potential on illumination," Phys. Rev. **72**, 345.

Brattain, W. H., 1947b, Bell Labs Notebook No. 18194 (AT&T Archives, Warren, NJ).

Brattain, W. H., 1947c, Bell Labs Notebook No. 21780 (AT&T Archives, Warren, NJ).

Brattain, W. H., 1951, "The copper oxide rectifier," Rev. Mod. Phys. **23**, 203-212.

Brattain, W. H., 1968, "Genesis of the transistor," Phys. Teach. **6**, 109-114.

Brattain, W. H., and J. Bardeen, 1948, "Nature of the forward current in germanium point contacts," Phys. Rev. **74**, 231-232.

Braun, E., and S. MacDonald, 1978, *Revolution in Miniature: The History and Impact of Semiconductor Electronics* (Cambridge University Press, Cambridge).

Davydov, B., 1938, "On the rectification of current at the boundary between two semi-conductors," C. R. (Dokl.) Acad. Sci. URSS **20**, 279-282; "On the theory of solid rectifiers," 1938, **20**, 283-285.

Frenkel, J., 1933, "Conduction in poor electronic conductors," Nature (London) **132**, 312-313.

Frenkel, J., 1936, "On the absorption of light and the trapping of electrons and positive holes in crystalline dielectrics," Phys. Z. Sowjetunion **9**, 158-186.

Goldstein, A., 1993, "Finding the right material: Gordon Teal as inventor and manager," in *Sparks of Genius: Portraits of Electrical Engineering*, edited by F. Nebeker (IEEE Press, New York), pp. 93–126.

Guerlac, H., 1987, *Radar in World War II* (AIP Press, New York).

Haynes, J. R., and W. Shockley, 1949, "Investigation of hole injection in transistor action," Phys. Rev. **75**, 691.

Henriksen, P. W., 1987, "Solid state physics research at Purdue," Osiris **2:3**, 237-260.

Herring, C., 1957, "The significance of the transistor discovery for physics," paper presented at a Bell Labs symposium on the Nobel prize (unpublished).

Herring, C., 1992, "Recollections from the early years of solid-state physics," Phys. Today **45**(4), 26-33.

Herring, C., M. Riordan, and L. Hoddeson, "Boris Davydov's theoretical work on minority carriers," n.d. (unpublished).

Hoddeson, L., 1981, "The discovery of the point-contact transistor," Hist. Stud. Phys. Sci. **12**, 41-76.

Hoddeson, L., G. Baym, and M. Eckert, 1987, "The development of the quantum mechanical theory of metals," Rev. Mod. Phys. **59**, 287-327.

Hoddeson, L., *et al.*, 1992, *Out of the Crystal Maze: Chapters in the History of Solid State Physics* (Oxford University Press, New York).

Hoddeson, L., *et al.*, 1993, *Critical Assembly: A Technical History of Los Alamos during the Oppenheimer Years, 1943–45* (Cambridge University Press, New York).

Holonyak, N., 1992, "John Bardeen and the point-contact transistor," Phys. Today **45**(4), 36-43.

Mott, N. F., 1939, "The theory of crystal rectifiers," Proc. R. Soc. London, Ser. A **171**, 27-38.

Mott, N. F., and H. Jones, 1936, *Theory of the Properties of Metals and Alloys* (Oxford University Press, Oxford).

Pearson, G., 1947, Bell Labs Notebook No. 20912 (AT&T Archives, Warren, NJ).

Pickering, A., 1984, *Constructing Quarks: A Sociological History of Particle Physics* (Edinburgh University Press, Edinburgh).

Riordan, M., and L. Hoddeson, 1997a, *Crystal Fire: The Birth of the Information Age* (W. W. Norton, New York).

Riordan, M., and L. Hoddeson, 1997b, "Minority carriers and the first two transistors," in *Facets: New Perspectives on the History of Semiconductors*, edited by A. Goldstein and W. Aspray (IEEE Center for the History of Electrical Engineering, New Brunswick, NJ), pp. 1–33.

Riordan, M., and L. Hoddeson, 1997c, "The origins of the pn junction," IEEE Spectr. **34**(6), 46-51.

Ross, I., 1998, "The invention of the transistor," Proc. IEEE **86**, 7-28.

Scaff, J., 1970, "The role of metallurgy in the technology of electronic materials," Metall. Trans. A **1**, 561-573.

Scaff, J., and R. S. Ohl, 1947, "The development of silicon crystal rectifiers for microwave radar receivers," Bell Syst. Tech. J. **26**, 1-30.

Schottky, W., 1939, "Zur halbleitertheorie der sperrschict- und spitzengleichrichter," Z. Phys. **113**, 367-414.

Seitz, F., 1994, *On the Frontier: My Life in Science* (AIP Press, New York).

Seitz, F., 1995, "Research on silicon and germanium in World War II," Phys. Today **48**(1), 22-27.

Seitz, F., and N. Einspruch, 1998a, *Electronic Genie: The Tangled History of Silicon* (University of Illinois Press, Urbana).

Seitz, F., and N. Einspruch, 1998b, "The tangled history of silicon and electronics," in *Semiconductor Silicon/1998*, edited by H. R. Huff, U. Gösele, and H. Tsuya (Electrochemical Society, Pennington, NJ), pp. 69–98.

Shive, J. N., 1948, Bell Labs Notebook No. 21869 (AT&T Archives, Warren, NJ).

Shive, J. N., 1949, "The double-surface transistor," Phys. Rev. **75**, 689-690.

Shockley, W., 1936, "Electronic energy bands in sodium chloride," Phys. Rev. **50**, 754-759.

Shockley, W., 1939, "On the surface states associated with a periodic potential," Phys. Rev. **56**, 317-323.

Shockley, W., 1945, Bell Labs Notebook No. 20455 (AT&T Archives, Warren, NJ).

Shockley, W., 1949, "The theory of *p-n* junctions in semiconductors and *p-n* junction transistors," Bell Syst. Tech. J. **28**, 435-489.

Shockley, W., 1950, *Electrons and Holes in Semiconductors, with Applications to Transistor Electronics* (Van Nostrand, New York).

Shockley, W., 1973, in *Proceedings of the Second European Solid State Device Research Conference* (Institute of Physics, London), pp. 55–75.

Shockley, W., 1976, "The path to the conception of the junction transistor," IEEE Trans. Electron Devices **ED-23**, 597-620.

Shockley, W., M. Sparks, and G. Teal, 1951, "P-N junction transistors," Phys. Rev. **83**, 151-162.

Tamm, I., 1932, "Uber eine mögliche art der elektronenbindung an kristalloberflächen," Phys. Z. Sowjetunion **1**, 733-746.

Teal, G., 1976, "Single crystals of germanium and silicon—basic to the transistor and integrated circuit," IEEE Trans. Electron Devices **ED-23**, 621-639.

Torrey, H. C., and C. A. Whitmer, 1948, *Crystal Rectifiers* (McGraw-Hill, New York; republished in 1964 by Boston Technical Publishers, Boston).

Weiner, C., 1973, "How the transistor emerged," IEEE Spectr. **10**(1), 24-33.

Wilson, A. H., 1931, "The theory of electronic semi-conductors," Proc. R. Soc. London, Ser. A **133**, 458-491; 1931, "The theory of electronic semi-conductors—II," **134**, 277-287.