

THE EARLY DAYS OF THE TRANSISTOR: 1946 TO 1955

— Part 1 —

The OTB is pleased to bring you a transcript of a presentation on the development of the transistor by Dr. John Bardeen, a co-inventor of the device that revolutionized the technology of electronics and changed our world. Dr. Bardeen honored the AWA by delivering this talk at the 1986 Annual Conference, where it was tape-recorded by AWA Member John W. Weyer III. John provided us with a partial transcript of his recording, and in our May 1998 issue (honoring the golden anniversary of the transistor's invention), we asked for help in piecing together a more complete version.

Several members contacted the OTB with offers of assistance, and a very exciting find turned up! Dr. Bardeen had brought to the conference some manuscript copies of his presentation, including copies of all the illustrations, and made them available as handouts. One of these, autographed and specially presented by Dr. Bardeen,

was provided to the OTB by past AWA President Bruce Roloson. The manuscript is based on an identical talk given to the IEEE student branch at the University of Illinois in 1979. It is somewhat formalized for the printed page, but otherwise a faithful copy of the speech.

Among the people who offered assistance with the Bardeen material was Dr. Frederick Seitz of The Rockefeller University (New York, NY). As it happened, Dr. Seitz had used another taped copy of the talk as part of the source material for the book *Electronic Genii—the Tangled History of Silicon* (University of Illinois Press, 1998), which he co-authored (see the review in our previous issue). Learning of the existence of the manuscript copy, he volunteered to secure the necessary permissions for the OTB to print it. This he did, and part one of the talk follows. The conclusion will appear in our February issue.

—MFE

Introduction

The point-contact transistor, the first bipolar transistor, was discovered in a program of basic research on solid state physics initiated at the Bell Telephone Laboratories just after World War II. The program was due in large part to M. J. Kelly, who was director of research and later president of the Bell Laboratories. He felt that one could improve the properties of materials from an understanding of their electronic and atomic structure brought about by application of quantum mechanics. The group formed to work on the program included physicists and chemists as well as theorists with an understanding of the quantum theory of solids. The basis for a theoretical understanding of solids had been developed in the late twenties and thirties, but generally there was a wide gap

between theory and experiment.

Semiconductors was one of several areas the group was involved in, others being magnetism, piezoelectricity and dielectrics. There was a distant hope of using semiconductors to make an amplifying device to replace a vacuum-tube triode. A triode is essentially an electrical valve in which a voltage applied to one electrode can be used to control a current flowing between the other two in such a way as to produce amplification of an electrical signal. The finite lifetime and unreliability of vacuum tubes was limiting the complexity of circuits that could be used in the telephone system.

The head of the semiconductor group was William Shockley. The principal experimentalists were Walter Brattain and Gerald Pearson. A

physical chemist, Robert Gibney, was added a little later. I joined Bell Laboratories to work with the group just after the end of the war in late 1945. I had worked on the theory of metals before the war but had done little work on semiconductors. I became interested in semiconductor problems through associations with Brattain, Pearson and Shockley. None of us had worked on semiconductors during the war, so we formed a study group to learn about the subject and met at least once a week.

A sound theoretical foundation for understanding the properties of semiconductors was available from work done during the 1930s based on the Bloch theory of energy bands for electrons moving the periodic field of a crystal lattice. In developing his theory, given in 1928, Bloch was concerned primarily with metallic conduction. In 1931, Alan Wilson adopted the band theory to semiconductors. In the ideal crystal at absolute zero, the valence electrons that bond the crystal together completely fill a band of energy levels for the electrons called the valence band. There is an energy gap to the next higher band called the conduction band (Fig. 2). Conduction can take place in two different ways, by electrons in the conduction band and by missing electrons or holes in the valence band. The holes behave in all respects like particles of positive charge.

At high temperatures, electrons may be thermally excited from the valence to the conduction band, so that there are equal concentrations of conduction electrons and holes. Such conduction is called intrinsic. As illustrated in Fig. 2, the electrons and holes form a neutral plasma of mobile charges. Electrons may also be excited from the valence to the conduction band by absorption of light of appropriate frequency, giving rise to photoconductivity. Brattain and I discovered another way of increasing the concentrations of electrons and holes, by current flow from an appropriate contact. This is

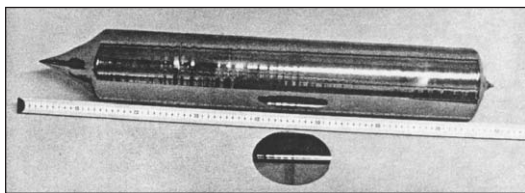


Fig. 1. State-of-the-art single crystal of silicon, diameter 12.5 cm. A wafer of silicon on which circuits are formed is shown below the meter stick. (Courtesy Monsanto Co.)

the principle of bipolar transistor.

Wilson also recognized that carriers can be introduced by impurities that go into the lattice in the form of positive or negative ions, as illustrated in Fig. 3. The charge of positive ions (called donors) is compensated by electrons in the conduction band. Since the carriers are negatively charged, semiconductors with donor impurities are called N-type. If the impurities are negatively charged, the charge is compensated by positively charged holes in the valence band and the semiconductor is called P-type. Impurities in parts per million or less can have a significant effect on the conductivity. Conduction introduced by impurity ions is called extrinsic.

It had long been known that a contact between a metal and a semiconductor can be rectifying, but an adequate explanation was not given until the late thirties by N. F. Mott and more completely by W. Schottky. Some of Schottky's papers with his co-worker, Spenke, did not become available in the West until after the war. Schottky's picture of the energy bands in the vicinity

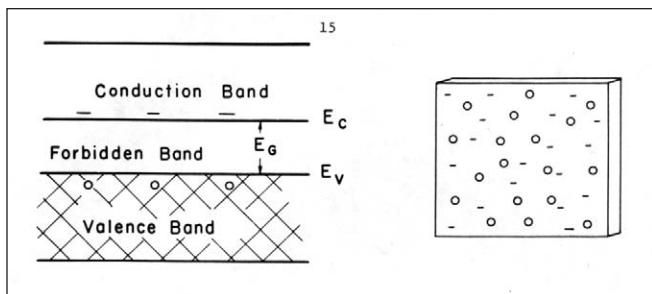


Fig. 2. Energy band diagram of an intrinsic semiconductor with equal concentrations of conduction electrons and missing electrons or holes in the valence band. As shown schematically on the right, the negative conduction electrons and positive holes form a neutral plasma.

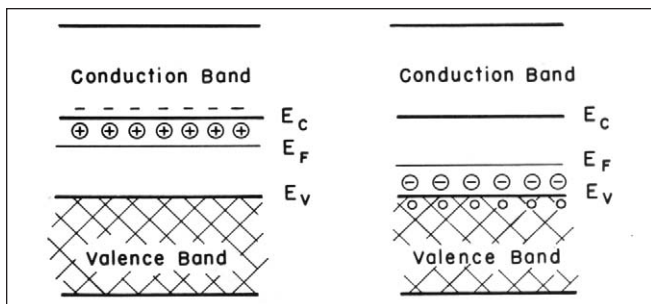


Fig. 3. Energy band diagram for N-type and P-type extrinsic semiconductors. The charge of the free carriers (electrons or holes) is compensated by an equal concentration of ionic impurities.

of a metal in contact with an N-type semiconductor is shown in Fig. 4, with the space direction perpendicular to the contact. An appreciable energy, X , of the order of 0.5 eV, is required to take an electron from the Fermi level, the energy of the uppermost occupied levels of the metal, and place it in the conduction band of the semiconductor. There is a region of positive space-charge from uncompensated donor ions in the semiconductor adjacent to the metal, with a resulting change in electrostatic potential. The direction of easy flow is that with the semiconductor negative with respect to the metal; a negative voltage raises the levels of the semiconductor relative to the metal and allows flow of electrons over the barrier into the metal. With a positive voltage, electrons must surmount the barrier, X , to enter the semiconductor; this is the direction of high resistance. Signs of carriers are reversed for a rectifying contact with a P-type semiconductor.

If, as illustrated, the barrier is such as to bring the top of the valence band close to the Fermi level, there will in equilibrium be an appreciable population of holes in the semiconductor adjacent to the metal. There is a change from N-type conduction in the bulk to P-type in the "inversion layer." In this case, in the direction of easy flow, the current may consist predominantly of electrons going from the valence band rather than the conduction band, leaving holes in the valence band. These flow into the bulk of the

semiconductor increasing the conductivity. It is the discovery of this effect that led to the first point-contact and junction transistors, but as we shall see the discovery came about in a roundabout way. Mott and Schottky considered the flow of only one type of carrier.

In our review of background material, we became acquainted with the large amount of work that had been done during the

war on silicon and germanium as "cat's whisker" detectors for radar. An extensive study of these elemental semiconductors had been carried out at university, government and industrial laboratories. The study of silicon was initiated by Russell Ohl at Bell Laboratories just before the war and was carried further at the Radiation Laboratory at MIT, the University of Pennsylvania, Bell Labs, and elsewhere. The study of germanium was initiated by Karl Lark-Holovitz at Purdue. By 1946 it was possible to produce relatively pure polycrystalline ingots of silicon and germanium and to control their electrical properties by introduction of appropriate donor and acceptor impurities.

Because of this background, we decided to concentrate our research efforts on germanium and silicon. An important advantage of these ma-

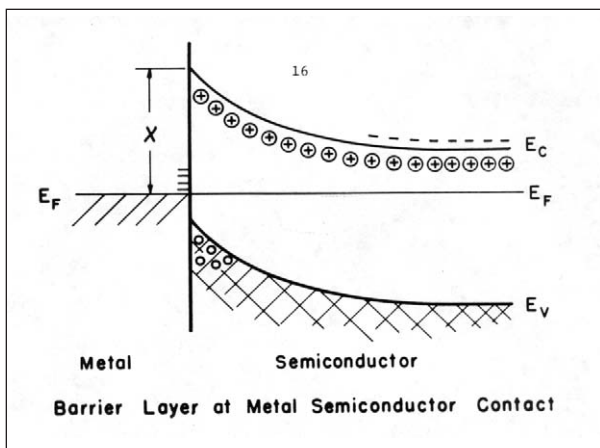


Fig. 4. Energy band diagram of a Schottky barrier at a metal-semiconductor rectifying contact, showing an inversion layer of P-type conduction at the interface.

terials is that they can be made either N- or P-type, depending on the nature of the impurities introduced. This is not true of many semiconductors. Another is that, being elements, they are easier to purify.

Not long after I arrived at Bell, Shockley asked me to check some calculations he had made earlier in the year (1945) on the possibility of making what is now known as a thin film "field effect" transistor. A semiconductor in the form of a thin film forms one plate of a parallel plate condenser. Shockley suggested that one could control the conductance of the film by a transverse electric field from a voltage applied across the condenser. With reasonable dimensions, the induced charge on the film should be comparable to the total charge of the carriers normally present in the semiconductor. Thus if the induced charge consists of mobile carriers, one should be able to make substantial changes in the conductance. With suitable geometry and materials, the effect should be large enough to produce amplification of an a.c. signal.

Various attempts to observe the effect all failed. This suggested to me that the induced charges were trapped in surface states for electrons at the surface of the film and thus were not available for conduction. I found that only a small concentration of surface states, less than one per hundred surface atoms, would be sufficient to effectively shield the interior of the film from the applied field. If there were surface states, other consequences should be observable; a Schottky space-charge layer should exist at the free surface, with the charge in the layer compensated by a charge in the surface states. It is possible to have an inversion layer of opposite conductivity type on a free surface. If the barrier is determined by surface states, rectification characteristics should be independent of the metal used for the contact.

Birth of the Point-Contact Transistor

It was decided to initiate a basic research program on both bulk and surface properties, with Pearson primarily responsible for bulk and Brattain for surface properties. Experiments carried out by Brattain helped to verify the surface states hypothesis. One may get information about the surface barrier from the contact potential and the way it changes with illumination of the surface. Brattain and Gibney found that the surface barrier could be altered if the voltage is applied through an electrolyte adjacent to the surface. Evidently, ions piling up at the surface gave a field sufficiently large to penetrate the surface



**JOHN BARDEEN
1908-1991**

Dr. John Bardeen received his B.S. (1928) and M.S. (1929) in Electrical Engineering from the University of Wisconsin. The next

three years were spent doing geophysical research at Gulf Research Laboratories. Returning to graduate studies in 1933, he was awarded a Ph.D. in Mathematical Physics by Princeton University in 1936. After three years as a Junior Fellow of the Society of Fellows at Harvard University, he joined the faculty of University of Minnesota, serving as Assistant Professor from 1938 to 1941. Bardeen left academia to work with Naval Ordnance Laboratories during the war years (1941-1945).

Dr. Bardeen joined Bell Laboratories in 1945, and it was here that his epochal solid state research was done. He left Bell in 1951 to become a professor both of Electrical Engineering and Physics at The University of Illinois. In these dual capacities, he established distinguished programs for both solid state research and the theoretical aspects of macroscopic quantum systems. He became Professor Emeritus in 1975.

This remarkable man was the recipient (with co-workers W.H. Brattain and Shockley) of a Nobel Prize for the invention of the transistor (1956). In 1972 he received an unprecedented *second* Nobel Prize (with L.N. Cooper and J.R. Schrieffer) for his work on the theory of superconductivity. Bardeen's other awards and honors are too numerous to mention here, but he continued to publish original scientific papers until his death in 1991.

states. This meant, in principle, that with suitable geometry one should be able to change the number of free carriers by an applied field and thus make a field-effect amplifier.

Experiments carried out with Pearson showed that the mobilities of the carriers in the thin evaporated films used to try to observe the field effect

(continued on page 21)

Radialamp, November 88
and May 92

RCA, June 84, November 87
and Nov. 89

Radio Cabinet, August 96

Receivership and
bankruptcy, February 87

Reichmann, May 91

Remo, May 94

Resonohone, March 84

Rice and Hochster, May 89

Riley-Klotz, August 94

Rola, September 80

Saal, August 86, Oct. 90

Seattle made, March 85

Sea shell horns, June, 82

Specialty Co. March 85

Standard Metal, August 90
and October 91

Star Radio, March 85

Stentorphone, August 93

Stromberg-Carlson,
September 83 and
May 98

Subcontracted parts,
August 90

Timbre-tone, February 91

Timmons, September 82

Tonks Bros., June 82

Tower, May 89 and
November 93

Trade names of horns,
June 83

Trimm, September 84,
November 96

Trinity Radio, August 98

United Motors Service,
March 85

Universal High Power,
March 85

Victor Talking Machine,
December 82 and
September 83

Victory Radio-Electro,
March 83

Western Electric, June 81
and December 82

Westinghouse, March 86

Wilson Utensil, March 84

Winkler-Reichmann, May 91

World Loud Speaker,
February 90

Yahr-Lange, June 86



THE EARLY DAYS OF THE TRANSISTOR, *continued from page 19*

were very small compared with those in good bulk materials, so that the change in conductance of the film from a transverse electric field would be small even if surface states were not present. To get the effect of a thin film in bulk material, I suggested that one use an inversion layer on the surface. It was known that with proper surface treatment one could get a thin inversion layer of N-type conductivity a few hundred Angstroms thick on the surface of a block of P-type silicon. In order to contact the inversion layer I suggested using a cat's whisker contact.

In discussions with Brattain, we decided to try the arrangement illustrated in Fig. 5. A small drop of electrolyte surrounded, but was insulated from, a metal point contact which made a rectifying contact with the P-type block. When biased in the reverse direction, metal positive, the current consists of holes flowing from the interface to the interior or of electrons flowing in the inversion layer to the contact. By applying a field through a voltage applied to a wire in con-

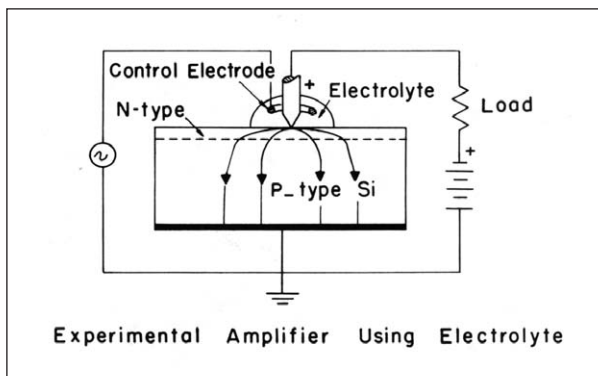


Fig. 5. Experimental field effect amplifier with a drop of electrolyte surrounding but insulated from a metal point contact.

tact with the electrolyte, we hoped to be able to change the concentration of electrons in the inversion layer and thus the reverse current. A negative voltage should decrease the concentration of electrons and thus the current and a positive voltage should increase it. The experiment was successful the first time it was tried. This was the first solid state amplifier and it demonstrated that the field effect principle is a valid one.

To be continued in the May, 1999 issue.



THE EARLY DAYS OF THE TRANSISTOR: 1946 TO 1955

— Part 2 (Conclusion) —

In the last issue of the OTB, we were pleased to bring you Part 1 of a presentation on the development of the transistor by Dr. John Bardeen (1908-1991), co-inventor of the device that revolutionized the technology of electronics. Dr. Bardeen honored the AWA by delivering this talk at our 1986 annual conference. We are grateful the Bardeen Family for granting per-

mission to reprint the talk here and to Dr. Frederick Seitz of The Rockefeller University (New York, NY) for contacting the family to arrange permission. Part 1 concluded with the experiment that resulted in the first solid state amplifier, demonstrating the validity of the field effect principle.

— MFE

Birth of the Point-Contact Transistor (continued)

Brattain and I next decided to try the experiment with N-type germanium, which was known to make a good rectifying contact to a metal point contact. Although we had no prior knowledge that a P-type inversion layer might be present on the surface, the experiment was successful. In this case a negative voltage applied to the electrolyte increased the reverse current and a positive voltage caused a decrease, as expected from the field effect.

Because of the slow response time of the electrolyte, amplification occurred only at very low frequencies. In order to eliminate the electrolyte, we attempted to grow a thin oxide film on germanium and evaporate a gold electrode on the surface. The idea was to test the field-effect principle by placing a point contact in close proximity. The principle is that of present day MOS (metal-oxide-semiconductor) transistors.

It was found that the gold spot was not insulated from the surface but made ohmic contact with the germanium. We decided to see what would happen anyway. Brattain placed the point contact biased in the reverse direction in close proximity to the spot and applied a small voltage between the gold electrode and a large-area low-

resistance contact on the base of the block of germanium. It was found that the current from the gold electrode had a large effect on the reverse

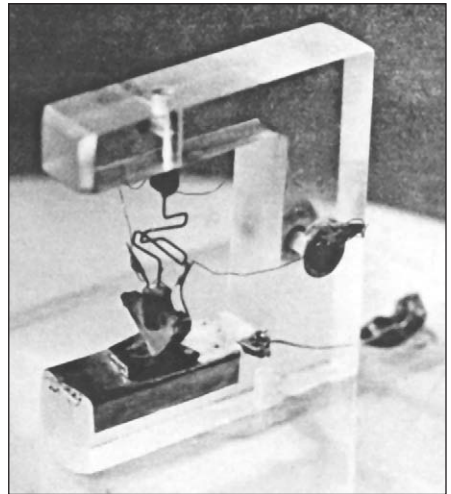


Fig. 6. Laboratory demonstration of the first bipolar transistor.

current to the point contact, but that the effect was in a direction opposite to that expected for the field effect; a positive voltage on the gold electrode increased rather than decreased the current flowing to the point contact biased negatively. We concluded that holes must be flowing from the gold contact. There was some voltage but no power amplification. This was the first indication of the transistor effect, a change in conductance from current flow from an appropriate contact.

The geometry used is not one expected to be very efficient for transfer of holes. Brattain and I next decided to try two parallel line contacts placed very close together. A gold foil was wrapped around a plastic wedge and the foil was slit where it passed around the knife edge. As shown in Fig. 6, a spring was used to press the wedge on the germanium surface, and connections were made to the foil on each side of the wedge. On contact, the hole emitter was biased in the forward direction (positive) and the other, the collector, in the reverse (negative). The experiment worked the first time it was tried; there was both voltage and current amplification of an input signal, with an overall power gain of about 20 DB. The point-contact transistor was born. Note in the illustration that we were still using the same block of germanium with the evaporated gold spot, which in this case was not playing any role.

It was soon realized that parallel contacts were not necessary. If the reverse resistance was not too high, the field of the collector current should draw in holes from the surrounding region of the germanium crystal. We then changed the design to two point contacts placed close together, as illustrated schematically in Fig. 7. A signal applied between the emitter and the base electrode appeared in amplified form across a high-resistance load between the collector and base.

The point-contact transistor was demonstrated to top executives of the Bell Labs on December 23, 1947, and this day is taken as the date of the invention, although the first experiments were

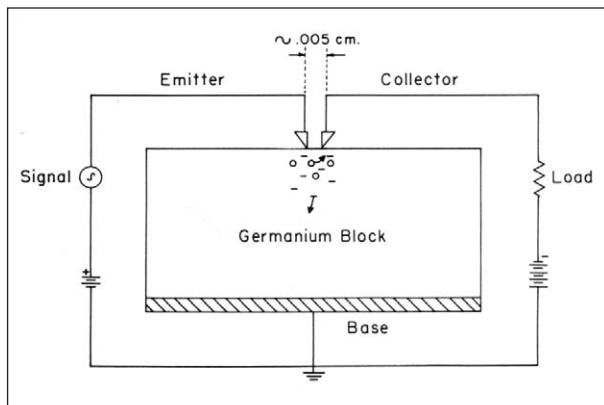


Fig. 7. Schematic diagram of a point-contact transistor. Holes flow from the emitter biased in the forward direction (+) to the collector biased in the reverse direction (-) relative to the base electrode on the N-type germanium block.

carried out a week earlier. The discovery created great excitement. A group was set up under Jack Morton to develop the point-contact transistor into a practical device. In the first model, called type A, the device was packaged into a small can about as big as the eraser on the end of a pencil. Most of the structure was to hold the point contacts in place. The active region in the germanium adjacent to the contacts had a dimension of the order of tens of μm , not much larger than present-day transistors.

The first public announcements and publications were made in June, 1948, after patent applications had been filed worldwide. At a press conference in New York City, there were demonstrations of the transistor as an amplifier and oscillator as well as in a radio receiver without vacuum tubes. Devices at that time operated to about 15 Mc. The lay press paid little attention, but there was a great stir in the technical press.

The Junction Transistor

At first it was uncertain how the holes introduced at the emitter contact flowed to the collector. Did they flow in the inversion layer or through the bulk with their space charge compensated by a corresponding increase in electron concentration? In an attempt to understand the mechanism, Shockley devised another configuration, the junction transistor, in which the entire

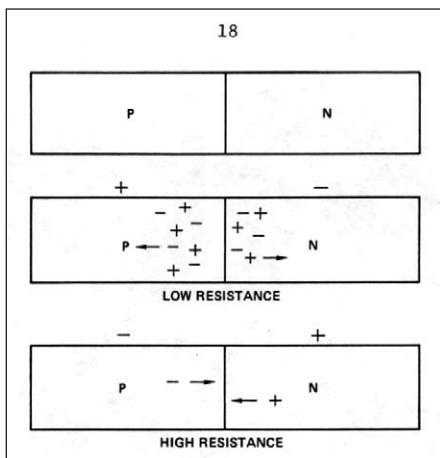


Fig. 8. Current injection and rectification at a P-N junction (see text).

process occurs in the bulk of the semiconductor. He suggested replacing their point contacts with PN junctions.

As illustrated in Fig. 8, a PN junction is a boundary between two regions of a semiconductor which differ only in the nature of the impurities present. In the P-region, the impurities are acceptors and conduction is by holes. In the N-region, impurities are donors and conduction is by electrons in the conduction band. The impurities may be present in concentrations as small as parts per million or less.

A PN Junction is rectifying when the direction of easy flow is that with the P-side positive with respect to the N-side. Holes flow from the P-side to the N-side and conduction electrons from the N-side to the P-side, increasing the conductivity. In the reverse direction, electrons would have to flow from the P-side to the N-side or holes from the N-side to the P-side. But there are very few electrons on the P-side and very few holes on the N-side, so the current is small.

An NPN junction transistor in simple form is shown in Fig. 9. A narrow region of P-type conductivity separates two N-type regions, with low-resistance contacts made to each region. One junction, biased in the forward direction, acts as the emitter and the second, biased in the reverse direction, as the collector. Electrons injected into the P-type base layer diffuse to the collector and contribute to the collector current rather than flowing to the base contact. Thus a voltage applied between the base and emitter serves to control the current flow between emitter and collector.

Shockley got the idea for the junction transistor within a month or so after the invention of the point-contact transistor, but kept the idea to himself while he was working out the details. In the meantime, John Shive tried making a transistor with point contacts on opposite faces of a thin wedge of germanium. He found that it operated in a similar manner to one with both contacts on the same side. He described his results to the group in a seminar talk on February 22, about a month after Shockley had described the junction transistor concept in his notebook. It was obvious from Shive's talk that the holes were flowing through the bulk. Shockley then disclosed his ideas concerning the Junction transistor.

At first Shockley thought his design might be useful for a high-power transistor. It was only later that it was realized that the junction transistor had excellent characteristics in its own right. He later worked out the theory in detail and published the results in 1949. When the first junction transistors were made, in 1950, it was found that they performed very close to predictions. Fabrication required great advances in semiconductor technology, including in particular the growth of high-quality single crystals.

Early Developments

For two or three years the major development effort in Jack Morton's group was on point-contact transistors. Shockley's group was concerned mainly in developing a better scientific understanding of the flow of electrons and holes in semiconductors under the nonequilibrium conditions that prevail under carrier injection. In retrospect, there was a large gap between the scientific and development efforts in that no one was responsible for the development of semiconductor technology independent of any particular device.

The point-contact design was arrived at because the experiments were very easy to do within the limits of known technology. Since there was a built-in positive feedback in the device, the characteristics were far from the optimum described in a triode. It was difficult to design to a set of prescribed characteristics; there was insufficient flexibility in design. Initially, reproducibility and reliability were poor. Point-contact transistors had some very limited applications in the telephone system, but had no widespread use. The year 1952 marked the turning point to junction transistors.

Transistor development might have taken quite a different course if the need for the development of semiconductor technology had been recognized earlier. Gordon Teal, who wanted to

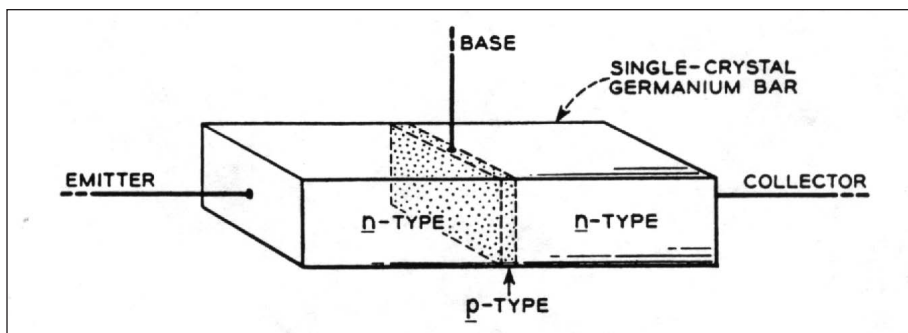


Fig. 9. Schematic diagram of an NPN junction transistor. Voltage applied to the P-type base layer controls the current of electrons between emitter and collector.

grow crystals, could find no support in the beginning and finally had to bootleg this project. Shockley thought that he could cut what small crystals he would need for polycrystalline ingots and gave Teal no more than moral support. The metallurgists thought they could produce as good a material as desired with the ingots. After the first single crystals were grown, Morton supported the project because he felt that use of single crystal material would improve the reproducibility of point-contact transistors. After they became available and their excellent qualities recognized, single crystals were used in all scientific work. The first junction transistors were made by doping the crystal appropriately as it was being grown by pulling from a melt.

The technology of growing single crystals of germanium, silicon and other semiconductors has developed continuously over the years. With increasing crystal perfection and purity, the mobility of the electrons and holes increased linearly with time for some years, leveling off to values characteristic of the ideal crystal around 1954. The size of the crystals has also shown a steady increase from dimensions of a few centimeters in the early days to the present state-of-the-art crystal shown in Fig. 1 (*in Part 1, Feb. 1999 issue—Ed.*).

Recognizing that transistors would have broad applications far beyond the telephone system, Bell adopted a policy that made it easy to obtain licenses which gave access not only to the patents but also to the underlying technology. A large number of companies in this country and abroad did obtain licenses and many initiated or expanded their research programs on semiconductors. This infusion of ideas and approaches from many different sources did a great deal to stimulate the rapid growth of the technology. While many of the key advances came from

Bell, many came from other organizations. For example, the first alloy junction and diffused junction transistors were made at General Electric by R. N. Hall and co-workers. The first alloy junctions to single crystal silicon were made by Pearson at Bell, but the first commercial silicon junction transistors were made by Teal and co-workers at Texas Instruments.

Bell Labs were also instrumental in stimulating research and teaching on semiconductors at universities. I took part in a summer school held at Bell Labs in 1951 for university faculty which included both lectures and laboratory work. The rapid expansion of university programs not only advanced understanding of the basic science and technology, but also trained students who later took positions in industry and government laboratories.

The broad base required for development of a large new technology was established. Government played an important role not only in providing research funds to universities and industry, but also in providing a market at a critical time for transistors when costs were too high for the bulk of the consumer markets. The semiconductor revolution is continuing and is expected to continue well into the eighties. While the fundamentals for both bipolar and field-effect transistors were established during a short period in 1947 and 1948, the great advances that have taken place resulting from the work of many people in different places have gone far beyond the wildest dreams any of us could have had at that time.

In looking for alternative sources of energy to meet our future needs, we should not look for easy answers or short term solutions, but build a broad base in university, government and industry to develop the required science and technology. Many new innovations will be required to bring them to fruition. ☐