The Discovery of the Point-Contact Transistor
Author(s): Lillian Hoddeson
Source: *Historical Studies in the Physical Sciences*, Vol. 12, No. 1 (1981), pp. 41-76
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/27757489
Accessed: 17/06/2014 03:59

LILLIAN HODDESON

The discovery of the point-contact transistor

TWO FEATURES STOOD OUT in the reorganization of The Physical Research Department of the Bell Telephone Laboratories, which Mervin Kelly, the recently appointed Executive Vice President, announced in July 1945 to a surprised staff: an orientation towards basic studies in the new solid-state subfield of modern physics, and an emphasis on multidisciplinary team research.[1] The discovery of the point-contact transistor thirty months later by John Bardeen and Walter Brattain was the first significant achievement of the new approach.

In what follows I attempt to unravel and present the crucial scientific and institutional steps that led to this discovery. The principal sources are laboratory notebooks and other archival material in the private collection of the Bell Telephone Laboratories; interviews with the major participants and their contemporaries at Bell Laboratories in the 1930s and 1940s; and published literature. This literature can be

* Department of Physics, 1110 West Green Street, University of Illinois at Urbana-Champaign, Urbana, IL 61801; and Fermi National Accelerator Laboratory, Box 500, Batavia, IL 60510.

The following abbreviations are used: AIP, American Institute of Physics, Niels Bohr Library; BLC, Bell Telephone Laboratories, Private Collection; IEEE, Institute of Electrical and Electronics Engineers; NB, Bell Laboratories Notebook (in BLC); *PR, Physical review; PRSL,* Royal Society of London, *Proceedings; ZP, Zeitschrift für Physik.* All interviews together with associated files are at AIP and unless otherwise stated were conducted by L. Hoddeson. I am much obliged to the many persons who gave their time to be interviewed, and to Bell Laboratories and AIP for help in processing the transcripts.

I wish also to thank Gordon Baym, Walter Brattain, Joseph Burton, Mort Fagen, Paul Forman, John Heilbron, Russell Ohl, Philip Platzman, Kiiti Siratori, Charles Slichter, Carl Thurmond, Joan Warnow, Spencer Weart, and Charles Weiner for commenting on drafts; Bell Laboratories for access to the BLC; Joseph Burton, Bernard English, Mort Fagen, Norma McCormick, Michael Majorossy, George Schindler, Ruth Stumm, and Mary Lynn Toto for assistance in using it; and Viola Graeper for help in using the archives of the American Telephone and Telegraph Company.

1. Organization chart, Physical Research Department, 1100, 2 Jul 1945 (BLC); interview with Dean Wooldridge.

HSPS, 12:1 (1981) 0073-2672/81/020041+36$00.50

divided into three categories: the handful of classic scientific papers,[2] some dozen retrospective accounts by the participants,[3] and hundreds of articles or sections of books by nonparticipants.[4] Only a few historians of science have contributed to this literature.[5]

1. SCIENTIFIC BACKGROUND AND INSTITUTIONAL SETTING

In 1945 solid-state physics had only recently assumed an intellectual and institutional identity as a research field. A rich body of data had been growing for more than a century out of studies in initially disconnected areas, including electrical and thermal conduction, optical and mechanical properties of solids, crystallography, magnetism and metallurgy. In addition many centuries of work by artists and craftsmen had produced a vast fund of empirical knowledge about the behavior of materials.[6] But before 1927, none of the conceptual schemes proposed to organize this data could claim wide success; they would fit one set of observations but not others, and different theories often contradicted each other.[7] A unified framework was lacking.

A rapid succession of fundamental papers on the quantum electron theory of metals, the most fundamental written between 1926 and 1932 by Wolfgang Pauli, Werner Heisenberg, Arnold Sommerfeld, Felix Bloch, and Rudolf Peierls, resulted in a skeletal quantum theory of solids. The theory solved most of the outstanding fundamental puzzles of the last quarter century concerning ideal metals. By about 1933 this work had provided the necessary framework, and the pace of develop-

2. The classic articles are J. Bardeen and W. Brattain, "The transistor, a semiconductor triode," *PR, 74* (1948), 230-231; "Nature of the forward current in germanium point contacts," ibid., 231-232; "Physical principles involved in transistor action," *PR, 75* (1949), 1208-1225, reprinted in E. J. M. Kendall, *Transistors* (Oxford, 1969).

3. E.g., Brattain, "Genesis of the transistor," *The physics teacher, 6* (1968), 108-114, and "One researcher's personal account," *Adventures in experimental physics, 5* (1976), 3-13; W. Shockley, "The invention of the transistor: An example of creative-failure methodology," European Solid State Device Research Conference, 2nd, *Proceedings* (Lancaster, 1972), 55-75, and "The path to the conception of the junction transistor," IEEE, *Transactions, ED-23* (1976), 597-620; Bardeen, "Semiconductor research leading to the point contact transistor," *The Nobel lectures, 1942-1962* (Amsterdam, 1964), 313-386.

4. The official Bell Laboratories account for the historical record is W. Gorton, "The genesis of the transistor," Bell Labs Memorandum for Record, 27 Dec 1949, 1100, WSC-XB (BLC). Other helpful accounts include M. Kikuchi and Y. Uemura, *Theory and application of semiconductors: Progress of the last half century* (Tokyo, 1960) (in Japanese); and R. Nelson, "The link between science and invention: The case of the transistor," in Universities National Bureau Commission for Economic Research, *The rate and direction of inventive activity: Economic and social factors* (Princeton, 1962), 549-583.

5. E.g., C. Weiner, "How the transistor emerged," IEEE, *Spectrum, 10* (1973), 24-33; Ernest Braun and Stuart MacDonald, *Revolution in miniature: The history and impact of semiconductor electronics* (Cambridge, 1978).

6. Cyril S. Smith, "The prehistory of solid-state physics," *Physics today, 18:12* (1965), 18-30.

7. Cf. W. Hume-Rothery, *The metallic state* (Oxford, 1931).

ment of fundamental theory slowed down.[8] Many of the earlier contributors switched to other fields. Political factors, notably Fascist policies, amplified the transition stimulated by the challenging fundamental problems of new fields like nuclear physics and quantum electrodynamics.

Solid-state physics now changed course. New theorists, unlike those who had made the quantum theory of metals, emphasized approximate calculations of the properties of real, imperfect solids. In the United States this emphasis influenced graduate students at the two principal centers for solid-state theory in the early 1930s, Princeton and the Massachusetts Institute of Technology. At Princeton, Eugene Wigner trained Conyers Herring, Bardeen, and Frederick Seitz, three leading members of the first generation of American physicists who would refer to themselves as solid-state theorists.[9] At MIT, John Slater educated the first large group of Americans to work on the quantum theory of solids, including William Shockley, Harry Krutter, and Marvin Chodorow.[10] Another group with a practical emphasis grew up in Britain around Harry Jones and Nevill Mott at Bristol. Academic efforts interacted with industry to start a field that has become the most heavily populated in physics.

In 1943 a group of solid-state physicists, anticipating "rapid growth after the war," urged that the American Physical Society (APS) create a "nucleus around which future development might concentrate."[11] The first evident result was a "Symposium on the Solid State" at the meeting of the APS in January 1945. Two years later the APS established a division devoted to the physics of solids.[12] The war had helped to restore rapid progress to solid-state physics by demonstrating the effectiveness of collaborative work by theorists and experimentalists on complex problems pertaining to radar and the atomic bomb, by supporting development of new methods of producing pure and perfect sam-

8. L. Hoddeson and G. Baym, "The development of the quantum mechanical electron theory of metals: 1900-28," *PRSL, A371* (1980), 8-23; A. H. Wilson, "Solid state physics 1925-33: Opportunities missed and opportunities seized," ibid., 39-48. The most influential early technical review was A. Sommerfeld and H. Bethe, "Elektronentheorie der Metalle," *Handbuch der Physik, 24:2* (Berlin, 1933), 333-622; cf. R. Peierls, "Elektronentheorie der Metalle," *Ergebnisse der exakten Naturwissenschaften, 11* (1932), 274-322.

9. Herring, "Recollections," *PRSL, A371* (1980), 67-76; Bardeen, "Reminiscences of early days in solid state physics," ibid., 77-83; Seitz, "Biographical notes," ibid., 84-99; interviews with Bardeen, Herring, and Wigner.

10. J. C. Slater, *Solid state and molecular theory. A scientific biography* (New York, 1975).

11. The proposal was signed by S. Dushman, R. A. Read, F. Seitz, W. Shockley, S. Siegel and R. Smoluchowski, Archives of the Division of Solid State Physics of the American Physical Society (AIP).

12. Weiner, "Transistor" (ref. 5). The first comprehensive textbook account of the subject is Frederick Seitz, *The modern theory of solids* (New York, 1940).

ples of materials on a large scale, and by encouraging new computational and experimental techniques.

After the war solid-state physics progressed most rapidly at institutions where theorists and experimentalists worked closely with each other. At Bell Laboratories, Kelly anticipated the usefulness of strong interaction between theorists and experimentalists in his reorganization of Bell's Physics Research Department. We will examine the work of one subdivision of this department, the semiconductor subgroup.

Semiconductors interested industry because of their abilities to rectify electrical current and to alter their conductivities on addition of impurities or the application of heat, light, or electrical voltages. This last property made semiconductors useful as electronic switches. Both properties had been studied for over a century.[13]

Rectification at the contact between a semiconductor and a metal, which arises from variation of resistance with the direction of current flow, was discovered in 1874 by Ferdinand Braun. This finding, and Braun's discovery in 1877 that the effect depended on the size of the contacts, led to the "cat's whisker" crystal detector, in which a small metal point (the whisker) makes a rectifying contact with a slab of semiconductor.[14] Cat's-whisker detectors were used extensively in radio before vacuum tubes replaced them in the 1930s. They returned to widespread use shortly before World War II and figured prominently in the invention of the point-contact transistor.

Another semiconductor discovery relevant to the transistor was the copper-oxide rectifier, invented in 1926 by L. Grondahl and P. Geiger, who worked at the Union Switch and Signal Company, a subsidiary of Westinghouse.[15] This device consisted of an oxidized plate of copper coated with a graphite or silver electrode. When under an applied voltage, it permits a considerably greater flow of current from the graphite or silver to the copper (the "forward current") than in the reverse direction (the "backwards current"). At first it was not known whether the

13. A beginning of a history of semiconductors has been made by Braun and MacDonald, "Revolution" (ref. 5); Colin Hempstead, *Semiconductors 1833-1919, an historical study of selenium and some related materials* (Ph.D. thesis, University of Durham, 1977); and Charles Süsskind in pieces on radio and electronics published over the last two decades chiefly in IEEE, *Proceedings.* Technical summaries include: Pearson and Brattain, "History of semiconductor research," IRE, *Proceedings, 43* (1955), 1794-1806; Brattain, "Development of concepts in semiconductor research," *American journal of physics, 24* (1956), 421-425; G. Renard, "La découverte et le perfectionnement des transistors," *Revue d'histoire des sciences, 16* (1963), 323-358; Karl Lark-Horovitz, "The new electronics," in F. S. Brackett, ed., *The present state of physics* (Washington, D.C., 1954), 57-127.

14. Süsskind, "Ferdinand Braun: Forgotten forefather," *Advances in electronics and electron physics, 50* (New York, 1980), 240-260.

15. L. O. Grondahl and P. H. Geiger, "A new electronic rectifier," IEEE, *Journal, 46* (1927), 215-222; Grondahl, "Theories of a new solid junction rectifier," *Science, 36* (1926), 306-308.

rectification came from the junction of the copper and copper oxide, from the junction of copper to graphite or silver, or from the body of one of the materials. By 1929 experiments had shown that the relevant processes occur only in a very thin "barrier region" of between 10^{-8} and 10^{-6} cm near the interface of the copper and copper oxide. One knew that the barrier is greater in the direction from copper to copper oxide, but not how near the barrier came to the surface nor which of the surfaces was responsible.[16]

No adequate theoretical explanation of the rectification was proposed during the 1930s. The first quantum theoretical tools for dealing with semiconductors, developed by Alan Wilson in 1931, required considerable sharpening before they could explain the phenomena of rectification and amplification.[17] As late as 1938 the direction of the rectification of semiconductor-metal junctions predicted by the best versions of Wilson's theory was opposite to that observed.

Kelly had recognized the promise of semiconductor research during his tenure as Director of the Vacuum Tube Department (1928-1934). Vacuum tubes and relays underlay the telephone system's tremendous expansion in the 1920s and 1930s: through its ability to rectify and amplify, the vacuum tube enabled the extension of telephony to great distances; through its switching ability, the relay made feasible a complex interconnecting network.[18] However, both devices had inherent limitations. Relays were slow; electron tubes, though fast, were expensive, wasteful of power, unreliable, and bulky. Kelly was interested in exploring semiconductor devices as possible alternatives and talked with Oliver Buckley, then the Director of Research of the Bell Laboratories, about setting up a program in the fundamental physics of solids.[19] But the early 1930s was not the time to create such a program.

As a practical matter, however, Bell scientists had been studying solid-state devices, the various semiconductor "istors"—resistors, varistors, and thermistors—that form the ancestry of the transistor.[20] Brat-

16. W. Schottky and W. Deutschmann, "Leitungs- und Photoeffekte an Sperrschichten," *Physikalische Zeitschrift, 30* (1929), 839-846.

17. Wilson, "The theory of electronic semiconductors," *PRSL, A133* (1931), 458-491, and *A134,* 277-287; "A note on the theory of rectification," *PRSL, A136* (1932), 487-498. Wilson noted the theory's failings in his *Semiconductors and metals* (Cambridge, 1939), 57-62.

18. Kelly, "Contributions of research to telephony: A new look at the past and a glance into the future," Franklin Institute, *Journal, 261* (1956), 189-200, on 191.

19. Interviews with Alan Holden, Richard Bozorth, and Stanley Morgan, who were there at the time.

20. Varistors, used in making rectifiers, are non-ohmic devices whose variable resistance depends on voltage, current, or polarity. Thermistors, used in thermometry, adjustment of amplification gain, microwave power measurement, and radar detection, are resistors sensitive to temperature change.

tain and Joseph Becker began to study copper oxide varistors as modulators in carrier frequency equipment; Richard Grisdale's research on varistors related to the reduction of clicks in telephone transmission; and Gerald Pearson's and J. B. Johnson's study of thermistors responded to Buckley's request for a device that would regulate repeaters on transcontinental cables more sensitively than the thermocouples then in use. These early semiconductor studies at Bell cross-fertilized one another.[21]

The study of copper oxide varistors in the late 1920s and through the 1930s stimulated several attempts to build semiconductor devices. Charles Demarest, an officer in AT&T's Radio Division, learned about Grondahl's and Geiger's copper-oxide rectifier at a meeting of the American Physical Society and, suspecting that it might be useful to the company, asked Russell Ohl, a member of his department, to examine its physics. Ohl found that that copper oxide rectifiers are not useful at high frequencies.[22] Then Maurice Long of Bell's Research Division, and also its Director of Educational Relations, heard Grondahl discuss the new rectifier at the Bureau of Standards. Experimenting with copper oxide samples that Grondahl gave him, Long realized that some characteristics of this device "were so similar to those of vacuum tubes that we ought to find out what made them tick." By 1929 he had convinced the Company to hire Brattain to work with Becker on the copper oxide rectifier.[23] Among the first questions that Brattain and Becker explored were: Is the rectification of copper oxide rectifiers a surface or body effect? Which contacts between semiconductors and metals rectify and which do not? Where does the seat of rectification lie? Why does the Hall effect depend upon direction of rectification? And why do certain samples of Chilean copper not work in rectifiers?[24] Their research answered some of these questions and underlined the need for a better theory than was then available.

Quantum theory of solids at Bell Labs

Bell physicists and physical chemists, including Brattain, Pearson, Addison White, Foster Nix, and Alan Holden, recognized that the new

21. Case reports and internal technical memoranda (BLC); J. Scaff, "The role of metallurgy in the technology of electronic materials," *Metallurgical transactions, 1* (1970), 561-573; interviews with Scaff, Pearson, and Grisdale; Brattain interview by Charles Weiner.

22. Progress reports of Radio Department, R. Ohl's private collection, Vista, CA; Ohl interview.

23. Long to Brattain, 25 Apr 1972, Brattain's private collection, Walla Walla, WA, also in transcript of interview with Brattain by Weiner.

24. Case reports (BLC).

quantum theory was essential to understanding solids and undertook study of quantum physics despite the lack of a formal program in solid-state theory at Bell.[25] For example, Brattain, whose training included introductory quantum mechanics taught by John Van Vleck at the University of Minnesota, attended lectures on the electron theory of metals by Sommerfeld at the 1931 Michigan Summer Symposium in Theoretical Physics. On his return to Bell, Brattain reported what he had learned to a group of about forty, which included executives Kelly and Buckley.[26] Pearson and White took formal courses in quantum physics at Columbia University, then only a subway ride from Bell Laboratories. More often, however, the acquisition of quantum theory came from independent study.[27] Nix and Holden, and Grisdale and White, laboriously went through Slater and Frank's *Introduction to theoretical physics* (1933). The depression actually helped this self-improvement by creating the threat of layoff and competition to avoid it. The depression also provided extra study time on "lay-off days."[28] But Bell researchers in the early 1930s could not through self-study achieve the level of the graduates of Princeton's or MIT's solid-state programs. The natural step would have been to hire recently trained Ph.D.s in solid-state physics. A hiring freeze between 1930 and 1936 made this resolution temporarily impossible.

When the freeze thawed, Kelly, now advanced to Director of Research, recruited several outstanding new graduates. In 1936 Dean Wooldridge came from the California Institute of Technology (Cal Tech) and Shockley from MIT; in 1939, James Fisk from MIT and Charles Townes from Cal Tech. In an organizational move in 1938, presaging his postwar program, Kelly placed Wooldridge and Shockley together with Nix in a research group under unprecedentedly loose supervision. Nominally they were authorized to explore, under the direction of Harvey Fletcher, basic questions of "electronic conduction in solids." But in effect they had investigative freedom.[29] Kelly's recruitment of first-class quantum theorists and his decision to leave them unfettered in the study of solid state were major institutional changes.

25. Interviews with Gerald Pearson, Addison White, Alan Holden and Foster Nix, and (by Weiner) with Brattain; Hoddeson, "The entry of the quantum theory of solids into the Bell Telephone Laboratories," *Minerva* (in press).

26. Memorandum, 15 May 1931, J. Becker to N. Williams, Brattain collection; Brattain to Hoddeson, 10 Jan 1979; S. Goudsmit, "The Michigan symposium in theoretical physics," *Michigan alumnus quarterly review* (20 May 1961), 181-182.

27. Interviews with Holden, Nix, Pearson, White, and Grisdale.

28. In 1932 the work week was reduced from 5.5 to 4 days, in 1934 it rose to 4.5, and in 1936 to 5.

29. Organization chart, 31 Dec 1938 (BLC); interviews with Nix, Wooldridge, and Shockley.

They made possible the establishment of the research program at Bell Laboratories that produced the transistor.

While the group had independence in its choice of research problems, the Company's interests were clearly indicated in its official authorization. The first report on the "Case" (as Bell Laboratories calls an authorized project), dated 2 January 1935, states the reason for the research: "It is believed that such work will be valuable as a background for work on more specific problems, and will enable us to keep up and take advantage of the most recent advances in this field made outside of the Company." The reports in 1935 and 1936 support the authorization of the Case: "The knowledge of materials and processes acquired will ultimately be of value in the solution of engineering problems. There is also some speculative value in the chance discovery or invention of new materials or processes." As it happened[!], the groups' efforts between 1936 and 1940 centered on precisely the problem that Kelly had in mind: developing a microscopic theory for solid-state phenomena relevant to communication devices.

Shockley worked on theories for grid current in vacuum tubes, electron multipliers, and the interaction of electron beams and electromagnetic waves; Wooldridge worked on a theory of secondary electron emission; and Nix and Shockley collaborated on a theory of order and disorder in alloys, particularly in the gold-copper system.[30] Gradually the group went beyond the theory of communication devices. Already in 1937 the work described in the case reports included "other phases of metal physics, especially those having to do with fabrication strains and atomic distribution in alloys." And in 1938 the group's official subject changed to "Physics of the Solid State," and its purpose became the "fundamental study of the physical properties of metals as affected by composition, mechanical working and heat treatment." The authorization explicitly mentioned basic science and its anticipated fruit: "This case covers fundamental research work on the solid state and it is expected that eventually it will aid in the discovery of new materials or methods of processing old materials which will be useful in the telephone business."

In these years, 1936-40, Nix, Shockley, and Wooldridge were concerned to teach themselves and others on the staff (including Holden, Fisk, Brattain, Pearson, White, Richard Bozorth, and Howell Williams) the latest advances in solid-state theory. They set up a study group, modeled after a similiar one at MIT, and met for several hours each week, partly on company time, to review systematically recent texts in

30. "Electronic conduction in solids" and "Physics of the solid state," Case Reports, 1936-40 (BLC).

the quantum theory of solids, including those by Mott and Jones, Richard Tolman, and Linus Pauling.[31]

A third major line of research at Bell that proved important for the invention of the transistor aimed to understand and improve the operation of point contacts. In the late 1930s George Southworth and other members of Bell's radio group at Holmdel, New Jersey, were seeking new methods for the detection of very short (40 cm) radio waves.[32] Vacuum tubes, which had replaced cat's-whisker rectifiers, were insensitive at high frequencies. Southworth decided to reexamine point-contact detectors and constructed one first from parts found while rummaging in the famous secondhand radio market on Cortlandt Street in lower Manhattan. He found them to be much superior to vacuum tubes in the high frequency range.[33] This discovery prompted a large and continuing research effort on point contacts. The attempt to find materials that work best in point contacts led in turn to a program of studies of silicon.

Galena, the material traditionally used in cat's-whisker detectors, was not useful commercially; because of the nonuniformity of its crystals, one had to hunt around on them for spots at which rectification would occur. Southworth interested Ohl in studying other materials. An electrochemist turned radio engineer during World War I, Ohl tested more than 100 materials and found silicon to have the most sensitive point-contact detection properties; sensitive, but again erratic. Using the material then available commercially, he found that the contact between the silicon and the whisker sometimes rectified in one direction, sometimes in the other, and sometimes not at all. With Grisdale's help, Ohl attempted to make the silicon more uniform by purifying it through melting. They found that they needed special fur-

31. Nevill Mott and Harry Jones, *The theory of the properties of metals and alloys* (Oxford, 1936); Richard C. Tolman, *The principles of statistical mechanics* (Oxford, 1938); Linus Pauling, *The nature of the chemical bond* (Ithaca, N.Y., 1939). See interviews with Bozorth, James Fisk, Holden, Nix, Pearson, Shockley, Wooldridge, and White; and with Brattain by Weiner, and by A. Holden and W. J. King.

32. Southworth, *Forty years of radio research* (New York, 1962), 158-160; interviews with Ohl and Scaff.

33. In ordinary large area rectifiers, the ac barrier resistance R_{ac} is inversely proportional, and the capacitance C proportional, to the area; the critical frequency at the upper end of the range of a rectifier, $1/R_{ac}C$, is therefore independent of the area. In a point-contact rectifier, capacitance is again proportional to area, but the resistance in the forward direction (from point to semiconductor) is reduced owing to the geometrical spreading of the current flow lines (the "spreading resistance"), in inverse proportion to the *radius* of contact; the critical frequency thus goes inversely as the radius and can be made very high. J. H. Jeans, *Mathematical theory of electricity and magnetism* (Cambridge, 1925), 356; H. Torrey and C. Whitmer, *Crystal rectifiers* (New York, 1948 [MIT Radiation Laboratory Series, *15*]).

naces, and Ohl in August 1939 engaged Jack Scaff and Henry Theuerer, two of Bell's metallurgists to purify the silicon.[34]

In carrying out this purification by melting in high vacuum, Scaff and Theuerer came upon the strange effect that the direction of rectification varied in an uncontrolled fashion from one silicon ingot to the next. They were producing what came to be recognized as either "n-" or "p-type" silicon. (In n-type semiconductors the majority electrical current carriers are electrons; while in p-type semiconductors the majority carriers are "holes" or electron-deficient states that behave like positive charges.) The behavior remained a mystery until wartime research confirmed Scaff's conjecture that it arose from impurities in the silicon composed of elements in the third column (p-type) and fifth column (n-type) of the periodic table.[35]

In September 1939, while cooling hot silicon ingots very slowly in an attempt to avoid cracking, Scaff and Theuerer accidentally produced a sample in which two parts rectified in opposite directions. As Ohl soon discovered, using an oscilloscope and neon lamp whose current passed through a chopper, this ingot contained a "pn junction," an interface within a crystal between a p-type and n-type region. He demonstrated the strikingly large photovoltaic effect of this junction in Kelly's office by shining a flashlight on it. Brattain recalled that on witnessing this demonstration early in 1940 "we were flabbergasted." The effect was more than ten times greater than they had been getting with normal photoelectric cells.[36]

This demonstration was the climax of Bell's prewar studies pertinent to the transistor. Under Kelly's watchful eye, the three lines of semiconductor research were coming together through the internal communications within the Laboratory. At this point the war claimed the entire research effort of the Laboratory and delayed the unification of the solid-state program for several years. However, the wartime diversion of research did provide a more favorable environment, both at Bell and in the country at large for the prosecution of research.

The most evident alteration was in the physical facilities for

34. Scaff, "Metallurgy" (ref. 21); interviews with Ohl and Scaff; Ohl to Hoddeson, 2 Jan 1979.

35. Brattain and Becker knew that small variations in the proportions of copper and oxygen in cuprous oxide could affect its conductivity, but the application to silicon, a single element, was not obvious; Brattain to Hoddeson, 10 Jan 1979, and Scaff interviews. Subsequent experiments by Ohl and others showed the rectification direction of semiconductors was determined by whether the majority carriers were holes (p-type) or electrons (n-type).

36. Southworth (ref. 32); interviews with Ohl and Scaff, and (by Holden and King) with Brattain; Brattain to Hoddeson, 10 Jan 1979.

research. The Bell Telephone Laboratories had opened in January 1925 at 463 West Street in downtown New York, where the Western Electric Engineering Department had been located. It expanded into neighboring buildings and suffered increasingly from city dirt and noise. In the late 1920s, Bell's first Director of Research, Harold Arnold, had considered relocating the laboratory to a college-like setting in the country, but due to the depression postponed construction. The plan was revived in the late 1930s, and a small laboratory was erected as an experiment on a site in Murray Hill, New Jersey, offering the possibility for expansion.[37] This new laboratory's design contrasted sharply with housing in New York City; among other advantages, it eased communication between specialists in different departments.

The war made the new laboratory a much larger experiment than anticipated. Beginning in November 1941 many of the laboratory's rapidly multiplying research workers went to the Murray Hill facility. In 1945, approximately four years after its initial occupation, Kelly judged the Murray Hill experiment successful: "There is now only a very small number who would choose to return to the New York Laboratory....One hears frequent comments that work is done under less strain and that health is improved." And he believed that location in the country "will place us in a more favorable position for recruitment of research staff."[38]

Of the approximately 1500 military projects at Bell Laboratories during World War II, radar, their principal effort, had the greatest impact on the postwar semiconductor program.[39] It stimulated extensive study of semiconductor phenomena, and of the technology for producing crystals and devices.[40] Ohl, for example, studied silicon diodes to modulate radar signals, while Brattain and Becker investigated silicon point contacts in radar detectors.[41] But other military projects also prompted work with and on semiconductors. Pearson and his co-

37. O. Buckley, unpublished manuscript autobiography, National Academy of Sciences, Washington, D.C., 26-27.

38. Memorandum, Kelly to Buckley, 15 Jan 1945, esp. 18, 25 (AT&T archives).

39. The Bell System carried out about half of all research and development on radar in the U.S. during the war. Cf. M. D. Fagen, ed., *A history of engineering and science in the Bell System: National service in war and peace (1925-1975)* (Murray Hill, N.J., 1978); M. J. Kelly, "Radar and Bell Laboratories," *Bell Telephone magazine, 24:4* (1945-46), 3-37; Buckley, "Bell Laboratories in the war," *Bell Telephone magazine, 23:4* (1944-45), 227-240; and Memorandum, "War work expense for the period 1/1/41-7/30/44," issued 9 Oct 1944 by Bell's Commercial Relations Department, in Buckley papers (BLC).

40. An ample account of wartime semiconductor research in Britain and the United States appears in Torrey and Whitmer (ref. 33).

41. Shockley, "Path" (ref. 3), 604; interview with Ohl and (by Holden and King) with Brattain.

workers, for example, developed thermistors in thin films for night-vision devices.[42]

Westinghouse, General Electric, Sylvania, Du Pont, the University of Pennsylvania, and Purdue University among other laboratories in the United States, and General Electric, British Thompson-Houston, the Telecommunications Research Establishment, and Oxford University in Britain also rigorously pursued semiconductor research connected with radar. Intercommunication created widespread awareness of the importance of semiconductors (including a possible semiconductor amplifier), and of the fierce competition in the field likely to occur after the war.

Well before the war ended, Kelly began to plan to meet the competition. In 1943 he pointed out that until recently research at Bell had not interacted much with other technologies and industries, notably radio. Around 1937, however, Bell's research and development, stimulated by the military research on radar, began to move into radio and then into very high microwave frequency ranges and broader transmission bands. Thus, Bell entered a new phase in which "technical cross-fertilization and economic comparison" were important factors to be taken into account. According to Kelly, the key to success was a large-scale effort in basic solid-state research. Not a little was necessary to insure Bell's "preeminence in research and development."[43]

> We need only to complete our building program as soon as possible; to continue our present policy of generous support of the fundamental scientific researches underlying telephone technology; to maintain the present splendid contacts with pure science groups on the one hand, and our development-design and design groups on the other; and to maintain the present high standard of personnel in this organization.

Kelly's convictions can also be seen in his justifications for the case in "Physics of the Solid State," which continued, though primarily on paper, through the war. The authorizations for 1940-42 read:

> It is not anticipated that this program of solid state employing the most recent concepts of solid-state structure will bring immediate results applicable to our business. However, its method of approach is so basic and may well be of such far reaching importance that we should have such studies in progress as a background for our various materials developments. It is reasonable to expect that ultimately such a fundamental attack will point the way to the production of new materials with properties of importance in telephone uses.

42. Pearson interview.
43. Kelly, "A first record of thoughts concerning an important postwar problem of the Bell Telephone Laboratories and Western Electric Company," 1 May 1943, 1-2; memorandum, Kelly to Buckley, 29 Jul 1949, "Five year program for control of laboratories operations," 5-6 (BLC).

The authorizations of 1943 and 1944 for this group explained that although the work on this case had been practically suspended, it aimed at "fundamental study of the physical properties of metals and other solids, such as electrical, thermal, mechanical, diffusional and structural, particularly crystallographical."

With the strengthening of national commitment to research by wartime successes,[44] those responsible to Bell for the running of its laboratories could afford further to liberalize their "enlightened" attitude toward research and offer greater freedom to the Laboratories' most talented scientists. Staff members were warmly encouraged to participate actively in the larger scientific community; the leadership of the Laboratories understood well that such participation was the most effective way to achieve necessary awareness of the scientific frontier. After the war the number of seminars, journal clubs, and study groups increased, and efforts were also made to provide further education for junior staff.[45] These attitudes and circumstances, together with easy access to a larger staff of scientific and technical specialists in many fields, made the research environment for some Bell scientists superior to that offered at the best universities.[46]

Kelly's reorganization of the Physical Research Department towards the war's end made solid state the leading subject of exploration in the new quasi-university environment. Three of the nine departments into which Kelly divided all of Physical Research—Solid State Physics under Stanley Morgan and Shockley, Physical Electronics under Wooldridge, and Electron Dynamics under Fisk—were authorized to carry out basic solid-state studies. Of course, a mission lurked in the background; as Kelly put it in the 1945 authorization of the new Solid State Department, "the research carried out under this case has as its purpose the obtaining of new knowledge that can be used in the development of completely new and improved components and ap-

44. Cf. V. Bush, *Science, the endless frontier* (Washington, D.C., 1945), and later commentary on it in *Science, 182* (1973), 116, *183* (1974), 798-799, and *191* (1976), 41-46.

45. One seminar in 1945-46 studied J. H. de Boer, *Electron emission and absorption phenomena* (Cambridge, 1935); another, the continuation of the prewar solid state physics study group, studied N. F. Mott and R. W. Gurney, *Electronic processes in ionic crystals* (Oxford, 1940); lecture courses in quantum and statistical mechanics were offered; and a journal club reviewed current articles of interest. Interviews with Herring, Holden, Bardeen, and Robert Gibney. Employees who had been regularly employed for at least one year and who held a BA or BS could enroll in a "Part-time postgraduate study plan" at Columbia; "Resumption of part-time postgraduate study plan 1945-46," 30 Aug 1945, Buckley papers (BLC). The "Communications development training program," or "Kelly College," was established in 1948 as a three-year course for recent graduates in engineering; Kelly, "Five year program" (ref. 43). High-level consultants, i.e., P. Debye and J. C. Slater, also provided instruction; cf. Kelly to Slater, 11 Dec 1944 (BLC).

46. Pearson interview.

paratus elements of communications systems." He noted further the "great possibilities of producing new and useful properties, by finding physical and chemical methods of controlling the arrangement and behavior of the atoms and electrons which compose solids." But he turned the case over to men oriented toward fundamental research. Morgan, a Princeton-trained Ph.D. chemist whom Bell hired in the 1920s, had stressed the importance to Bell's research staff of graduate level scientific study; Shockley, Wooldridge, and Fisk, whom Kelly himself had recruited before the War, had earned Ph.D.s at outstanding American universities and were well-versed in the application of quantum mechanics.

The organizational chart for July 1945 further divided the Solid State Department into four subgroups whose specialties reflected Bell's continuing interests in solid state: magnetism under Bozorth; contacts and carbon microphones under F. S. Goucher; dielectrics as well as physical chemistry of solids under Morgan; piezo-electricity, crystals for oscillators, and propagation of sound in solids under Warren Mason. Each subgroup was designed as a multidisciplinary team—a balanced mixture of individual specialists—so that the group as a whole would have a spectrum of expertise. Such an organization, as Kelly observed, had proved itself at the big wartime research laboratories like MIT and Los Alamos. To adapt it at Bell would be both prudent and progressive.

A fifth subgroup of the Solid State Department did not appear on the chart of July 1945. Devoted to semiconductors and directed by Shockley, it first surfaced in January 1946, on the second chart of the new organization, following certain crucial additions to the staff that made Bell Laboratories a leading international center for theoretical as well as experimental solid-state physics.

In the Fall of 1945 Kelly authorized Wooldridge to invite Herring, a top solid-state theorist and former student of Wigner's, to join the new Physical Electronics group. Wooldridge had known Herring when both were graduate students at Cal Tech, and he was impressed with Herring's theoretical ability. Herring came to Bell for a trial of a few months and by spring 1946 had accepted a permanent position.[47]

Fisk and Shockley convinced Kelly also to offer a position to Bardeen, whom they had known from their Cambridge days in the 1930s. Bardeen was already one of the outstanding solid-state theorists in the country; he joined the semiconductor subgroup of the solid-state group late in 1945.[48] The Murray Hill Laboratories were congested in 1945.

47. Interviews with Wooldridge and Herring.

48. Bardeen took his Ph.D. under Wigner at Princeton in 1936, having worked in the Engineering Division of the Western Electric Company. He met Fisk and Shockley while a postdoctoral Fellow at Harvard in 1935-38 under Van Vleck.

Bardeen had to share an office with Pearson and Brattain, who were again studying rectifiers, varistors, and thermistors. Bardeen was soon trying to explain the data that Pearson and Brattain were gathering on semiconductors.[49]

2. STEPS TO THE INVENTION OF THE POINT-CONTACT TRANSISTOR

Shockley's semiconductor subgroup represented a range of pertinent specialties. Shockley himself and Bardeen were theoretical solid state physicists; Brattain and Pearson, experimental physicists who had worked for more than a decade on semiconductors; Robert Gibney, an experienced physical chemist; and Hilbert Moore, an electronics expert. The team also had two technical assistants, Thomas Griffith and Philip Foy.

The first task was to review the wartime progress. Advances made at Bell during the war that would be crucial to the invention of the transistor include the identification by Ohl and others of the impurities that cause n- and p-type behavior and the development by the metallurgists Scaff and Theurer of techniques for "doping" silicon and germanium. (Doping is the addition of impurities in order to improve rectification through the creation of extra energy levels.) Seitz and coworkers at the University of Pennsylvania and Du Pont, studying silicon, and Karl Lark-Horovitz' group at Purdue, working on germanium, had significantly deepened understanding of the properties and methods of producing both high purity and doped silicon and germanium. The Purdue group had also developed the "high-back-voltage germanium" rectifier, which overcame the low-voltage limitations of the earlier crystal rectifiers and had improved rectification characteristics. Unlike the ordinary crystal rectifiers of that time, which at high voltages not only suffered physical damage but conducted in the reverse direction (their back resistance decreasing with increased negative voltage), the high-back-voltage rectifier of super pure germanium had exceptionally low conductivity in the back direction for applied voltages up to about 100 volts.[50]

Substantial progress had also been made by Mott, Walter Schottky, and B. Davydov in 1938,[51] and by Schottky, Hans Bethe, Robert Sachs,

49. Interviews with Bardeen, Pearson, and Brattain.

50. Torrey and Whitmer (ref. 33), 22, 361-381.

51. The classic papers include N. F. Mott, "The theory of crystal rectifiers," *PRSL, A171* (1939), 27-38; W. Schottky, "Halbleitertheorie der Sperrschicht," *Naturwissenschaften, 26* (1938), 843; "Zur Halbleitertheorie der Sperrschicht- und Spitzengleichrichter," *ZP, 113* (1939), 376-414; "Vereinfachte und erweiterte Theorie der Randschictgleichrichter," *ZP, 118* (1942), 539-592. For further literature see Mott and R. W. Gurney, *Electronic processes in ionic crystals* (Oxford, 1940).

Karl Herzfeld, and others during the war,[52] in developing a theory to explain rectification at a junction between semiconductor and a metal, or between two different semiconductors. Both the early theory of Wilson, Lothar Nordheim, Jakov Frenkel, and A. F. Joffe,[53] and the later Mott-Schottky theory attributed the rectification to the experimentally observed potential barrier at the interface, which was assumed to be caused by a region deficient in electrons. This "space charge rectification layer," was physically bounded by a sheet of positive and one of negative charge.

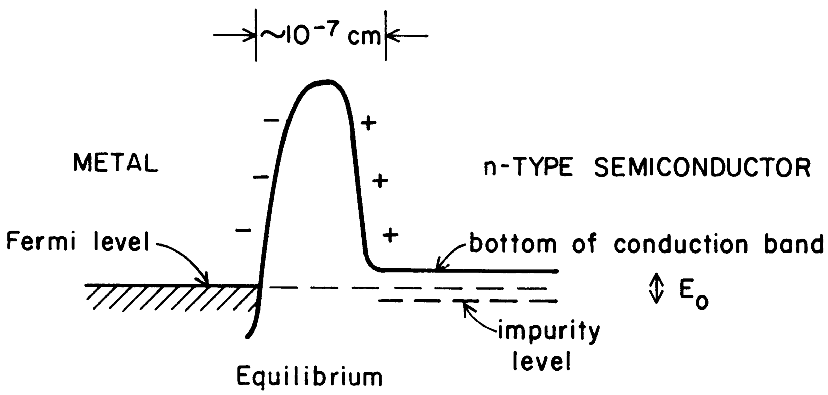

Fig. 1 The Wilson Model.

These sheets set up when the metal and the semiconductor came into contact. Consider a metal to n-type semiconductor interface (figs. 1 and 2) with electrons as the majority carriers. In the metal the electrons occupy energy levels up to the "Fermi level"; in the semiconductor, the "conduction bands," to which the electrons must be excited in order to participate in current flow, are normally empty. A number of electrons also occupy "impurity levels" located at an energy E_0 below the conduction band. Upon contact, electrons flow from the the semiconductor to the metal, thus building up a sheet of positive charge on the semiconductor side and one of negative charge on the metal side, until the density of electrons in the metal in any range more than $E_0/2$ above the Fermi level is the same as in the semiconductor. In equilibrium the Fermi level of the metal is approximately halfway between the impurity and conduction levels of the semiconductor. In both the Wilson and Mott-Schottky models, the space charge layer functions as a rectifier in much the same way as does the region in a vacuum tube diode between the filament and plate. The essential difference between

52. Reviewed in Torrey and Whitmer (ref. 33).

53. Bibliography in A. H. Wilson, *Theory of metals* (Oxford, 1936), and Mott and Gurney (ref. 51).

the two models is the shape of the barrier and the mechanism by which charge carriers traverse it.

Wilson (fig. 1) pictured a quantum-mechanical mechanism in which the charge carriers pass through a symmetrical potential hump about 10^{-7} to 10^{-6} cm wide, which is narrow enough to allow penetration by quantum mechanical tunneling. Experiments confirmed the estimate of the width.[54] A negative voltage applied on the metal side raises the (electron) energy levels in the metal in relation to those in the semiconductor, increasing the number of electrons in the metal that are able to reach unoccupied states in the semiconductor; the net flow of electrons from metal to semiconductor therefore goes up. A negative voltage applied on the semiconductor side raises the energy levels in the semiconductor and increases the flow from the semiconductor to metal. However, in the latter case, although the shift in levels is the same, the current increase is not so large because the density of conduction electrons in the semiconductor is much smaller than in the metal. The theory therefore predicts the direction of easy flow to be from the metal to the semiconductor.[55] The analogy to the rectifying property of a vacuum tube diode lies in the fact that in both cases many more electrons are available for charge flow in one direction than in the other. But the direction of easy flow predicted by Wilson was just opposite to that observed! Furthermore, later experiments made the barrier widths about 10^{-4} or 10^{-3} cm, undermining tunneling as the transfer mechanism.[56]

With these objections in mind, Mott and Schottky reinterpreted the rectification mechanism with an essentially classical model in which the majority carriers do not tunnel through a narrow barrier but are excited thermally over a relatively wide one, about 10^{-4} cm thick.[57] They pictured the hump as asymmetrical, with its maximum near the surface of the metal and its minimum well into the semiconductor. In equilibrium, with no applied voltage (fig. 2a), the conduction electrons on both sides see approximately the same barrier height ϕ_0, and there is no net flow of charge. A *negative* voltage V applied on the metal side (fig. 2b) raises the (negative) electron energy levels in the metal relative to those in the semiconductor and increases the barrier space charge so that the conduction electrons on the semiconductor side see an increased barrier, $\phi_0 + V$, while the corresponding electrons on the metal side still see a barrier ϕ_0. The result is a small decrease of elec-

54. Schottky and Deutschmann (ref. 16).

55. Wilson, *Semiconductors and metals* (ref. 17).

56. Grondahl, "The copper-cuprous oxide rectifier and photoelectric cell," *Reviews of modern physics, 5* (1933), 141-168; F. Waibel, "Über den Aufbau der Sperrschicht beim Kupferoxydulgleichrichter," Siemens-Werken, *Wissenschaftliche Veröffentlichungen, 15:3* (1936), 75-86.

57. Mott, "Theory" (ref. 51).

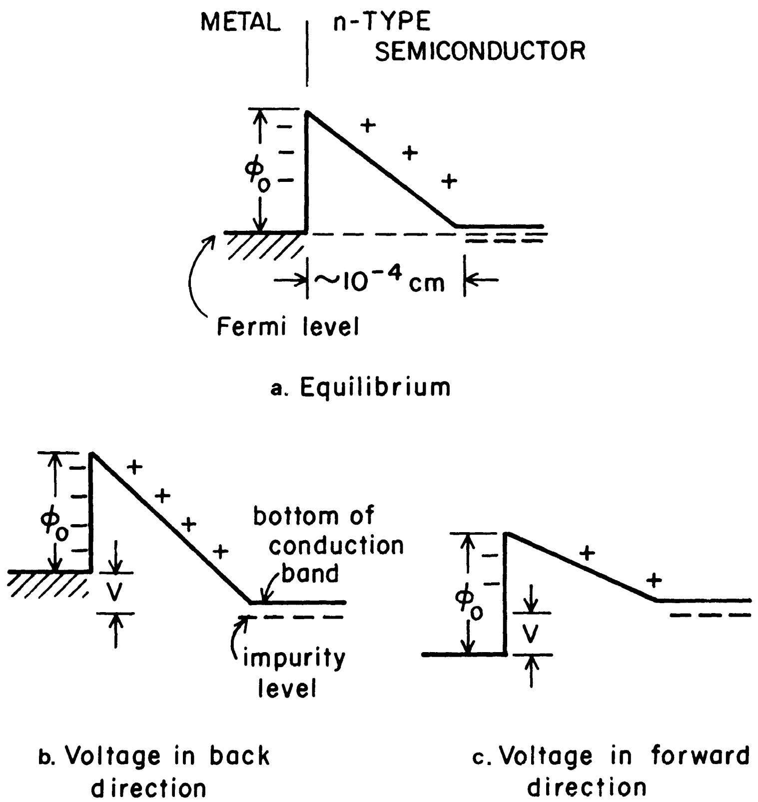

Fig. 2 The Mott-Schottky Model.

tron flow from semiconductor to metal and no change in flow from metal to semiconductor. A positive voltage V applied to metal (fig. 2c) lowers the (negative) electron energy levels in the metal relative to those in the semiconductor and decreases the barrier space charge so that conduction electrons on the semiconductor side see a reduced barrier $\phi_0 - V$; the electrons on the metal side still see ϕ_0. There results a substantial increase of electron flow in the "forward" or "easy flow" direction from semiconductor to metal, while the reverse flow remains the same. The analogy of the space charge layer to the vacuum tube no longer rests on the difference between the numbers of available electrons in the two materials. In one direction the application of a voltage increases the barrier, and in the opposite direction decreases it, rather as if in a vacuum tube, the filament and the plate would approach or recede from one another according to the sign of the applied voltage. In Wilson's model of a rectifier, on the other hand, the difference between the current in the back and forward cases arises because the current in the forward direction is $\exp(eV/kT)$ larger than the current in the back direction, as experiments showed. The Mott-Schottky theory solved the major problems of the earlier rectification theory: it assumed and also offered an explanation for the large observed barrier widths of 10^{-3}-10^{-4} cm; it provided a suitable mechanism for traversing such barriers; and it accounted for the observed direction of easy flow responsible for rectification.

A shortcoming of the theories of Mott and Schottky, and also Wilson, was the focus on the role of the majority carriers in the semiconductor (in our example, the electrons); both models ignored the minority carriers. Davydov took the latter into account in 1938, but his mathematical treatment masked this point, which received little attention at the time.[58] Excluding minority carriers is indeed valid for copper oxide, where their flow is insignificant, but does not do for germanium and silicon. The Bell Laboratories semiconductor group would later learn, essentially by accident, that minority carrier flow is the key to the operation of the germanium point-contact transistor.

The Mott-Schottky theory had other troubles too. It did not yield the experimentally measured rectification nor its variation with change in contact potential. Furthermore, according to the theory, the size of the barrier for any given semiconductor in a rectifying contact depends on the work function of the metal or semiconductor it touches. But experiments showed that properties of germanium and silicon rectifiers often do not depend on their partners.[59] Nevertheless, even with its failures, the Mott-Schottky theory was much better than anything available before, and it helped the Bell semiconductor group to understand why certain attempts to build semiconductor amplifiers during or after the war would not work.[60]

Most of the early amplifier experiments at Bell were based on the simplest analogy between the semiconductor and vacuum tube diode and consisted of trying to extend the analogy by inserting into a semiconductor diode a third element, which, it was hoped, would play the role of the grid in the vacuum tube triode. Brattain and Becker had tried to do so with copper oxide in the 1930s.[61] The hitch, as they perceived it then, was lack of technique for placing a grid in the appropriate place in the barrier region of 10^{-4} cm. By computing the size of the

58. Bardeen, "Trends in semiconductor research," *Advances in semi-conductor science* (London, 1949), 2-6.

59. W. E. Meyerhoff, "Contact potential differences in silicon crystal rectifiers," *PR, 71* (1947), 727-735.

60. During the war, before they knew the relevant papers by Schottky, a few members of the solid state study group at Bell reinvented some of his ideas about barriers; White interview.

61. Brattain interview by Holden and King.

grid needed to fit into this space, Brattain had concluded that there could be no feasible geometry.[62]

The Mott-Schottky theory also stimulated the invention of amplifier designs. Shockley recalls being inspired to think along such lines, at the time he was hired, by a "lecture or pep talk" from Kelly on Bell's need to replace metal contacts (as in relays) by electronic switches.[63] Around 1939 Brattain and Becker interested Shockley in the copper oxide rectifier, and shortly afterwards Shockley conceived a way to employ the space charge rectification layer, as discussed by Mott and Schottky, as a valve to control conduction and achieve amplification. Shockley described the design, which would now be called a "Schottky-gate-field-effect" transistor, in his laboratory notebook on 29 December 1939; the basic idea was to create a region of high resistivity in the way of the current in the circuit of interest and electrically bias this region so that it would regulate current much as the biased grid does in a triode. Shockley wrote:[64]

> It has today occurred to me that an amplifier using semiconductors rather than vacuum is in principle possible. Suppose, for example, that a very fine mesh copper screen is oxidized thus giving a metal grid embedded in oxide and let ohmic contacts be made to the outer surfaces [fig. 3]. Then if the carriers of charge are for convenience regarded as positive, if the grid is made plus, a space charge sheath with carrier deficit forms around it. This gives a region of low conductivity and accounts for high resistance in the reverse direction for the rectifying junction. Suppose the region is so large that it envelops the entire system of grid wires. Next suppose that an additional negative potential is applied to the right hand basic contact. This draws some, but not many more carriers from the grid and few if any from the left electrode because the grid wires are surrounded by the negative sheath which fills the space between them. If on the other hand the grid is at zero; then make right positive draws current, similar as if grid were not there, from the left. We can say that the grid effectively can be used to raise the resistance in its vicinity and thereby hinder the flow of current from left to right. Since the grid is being used in the reverse direction, its resistance is high and it will not consume much power, whereas relatively large currents flowing from left to right can be controlled.

62. In 1938 R. Hilsch and R. W. Pohl managed to insert a grid in alkali halide crystals in which the space charge layer could be made as thick as 1 cm. But since the upper frequency cutoffs of their crystal analogues to the vacuum triode were about 1 cycle/sec, the devices had no immediate application. Hilsch and Pohl, "Steuerung von Elektronenstromen mit einem dreielektroden Kristall und ein Modell einer Sperrschicht," *ZP, 111* (1938), 399-408.

63. Transcript of filmed interview with Shockley, 1972 (BLC).

64. Shockley, "Invention" (ref. 3), 57; Shockley, NB, 29 Dec 1939, reprinted in Shockley, "Path" (ref. 3), 603.

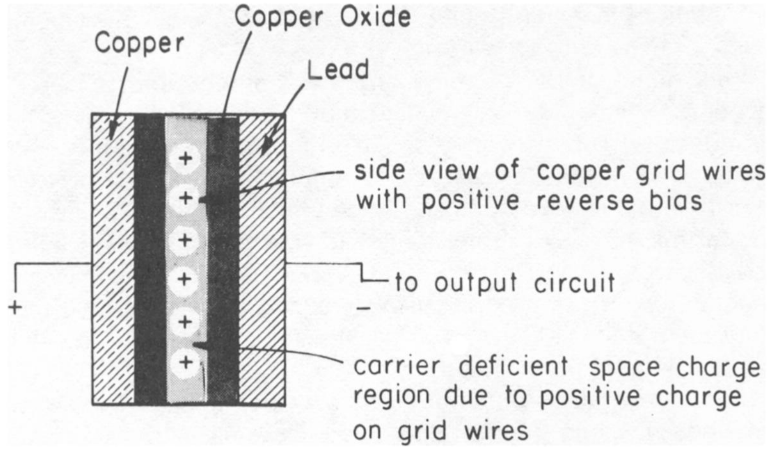

Fig. 3 Shockley's copper oxide amplifier design of 29 December 1939.

Wooldridge recalls watching Shockley experiment with such a device, provisionally constructed using "a piece of copper screen wire that had apparently been out in the elements for years and years because it was heavily oxidized. And... two copper wires... had been brought up to be both trying to touch... the oxide coating on one of those small screen wires... he was already then trying to put some batteries across there and see if he could influence the current flowing by the potential across the third electrode."[65] The experiments did not reveal the predicted effect. But working with copper wires, metal-semiconductor junctions, and batteries connected across pairs of these elements in an attempt to produce amplification, Shockley had before him in late 1939 the basic elements of a transistor.

Other materials besides copper oxide were employed in early solid state amplifier experiments at Bell. In 1938 or 1939, Shockley had tried, in collaboration with Holden, to build an electro-mechanical amplifier using carbon contacts brought together through pressure, controlled by a signal applied to a piezoelectric quartz crystal. They expected, but did not get, a usable output through a change in the resistance of the carbon as a signal was applied.[66] Ohl experimented with silicon. During the war he tried to insert a third element in a diode composed of two silicon surfaces, but the device could not develop a sufficiently low impedance because of its high capacitance.[67] In another wartime experiment, Ohl built a device for amplifying radio signals that employed point-contact detectors having negative resistance (current dropping with voltage increasing) created by thermistor heating. This device

65. Wooldridge interview.
66. Shockley, "Invention" (ref. 3), 57.
67. Ohl interview; Shockley, "Path" (ref. 3), 604.

worked, but its marginal stability made it unsatisfactory for commercial use.

While none of the prewar attempts to build commercially usable semiconductor amplifiers succeeded, they identified relevant problems. In its efforts to solve them, the Bell semiconductor group decided to focus on silicon and germanium, not only because they were much simpler than the copper oxide and selenium studied earlier, but also because wartime advances in purifying and controlling impurities in these materials made it possible to obtain consistently behaving samples. Germanium had been used extensively during the war in high frequency detectors; silicon, because of its larger bandgap, had found fewer uses but was recognized as "a thing of the future."[68]

Characteristically, all members of the group took part in such a decision; indeed, at no point in the studies leading to the transistor did any of them work alone for an extended time. There was almost daily informal discussion on all aspects of the work, as well as frequent interaction with outside specialists, including Scaff and Theuerer who supplied the silicon and germanium.[69]

3. THE TRANSISTOR

The road to the discovery of the point-contact transistor was opened in April 1945 when Shockley sketched the design of a "field-effect" amplifier in his laboratory notebook.[70] (He may have been inspired, as he suggests, by a demonstration of Ohl's amplifier based on thermistor heating of point contacts.)[71] Shockley argued that if the contact potential field of a rectifier could produce a space charge layer at the interface between a metal and a semiconductor, with most of the potential drop on the semiconductor side spread out over a distance of the order of 10^{-4} cm (fig. 2), then an externally applied electric field should be able to create such a barrier region. Furthermore, if the semiconductors were sufficiently thin, changes in the applied field should significantly alter the number of charge carriers and modulate the current passing through the semiconductor, thus providing the basis for an amplifier.

Shockley's notebook entries on the 13th and 16th of April 1945 are schematized in figure 4. The input (or control) voltage V_1, applied to

68. Pearson interview.
69. Brattain interview by Holden and King; Pearson and Bardeen interviews.
70. Shockley, NB 20455, 13-17 Apr 1945, 11-24.
71. Shockley, "Invention" (ref. 3), 58, and "Path" (ref. 3), 604; Ohl interview. Shockley had recently returned to Bell from the Pentagon where he had been working on non-scientific problems.

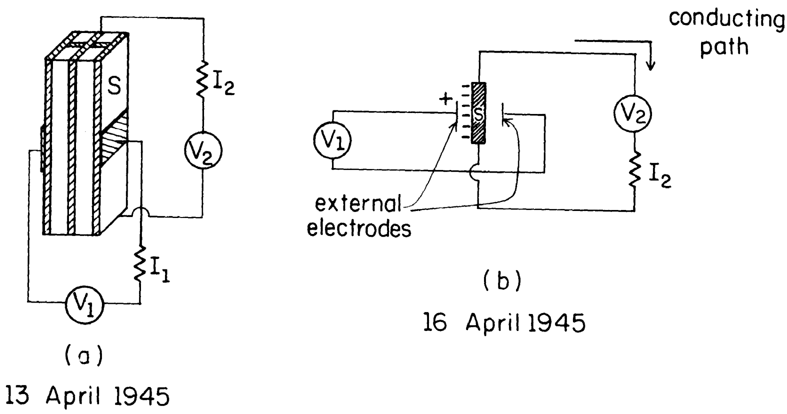

Fig. 4 Shockley's field-effect amplifier, April 1945.

one or more plates placed perpendicular to one or more thin semiconductor layers (S) (shown crosshatched) to form a parallel plate capacitor, attracts or repels charge carriers in the semiconductor. By increasing or decreasing the number of carriers in the semiconductor, the applied field changes the conductance of the semiconductor and regulates the current flow through it. This was the "field effect": with suitably applied voltages and resistance in the two circuits, an input signal could appear amplified in the output circuit (I_2, V_2). The relationship between the number of carriers and conductivity of the semiconductor layer was assumed to be given by the usual formula: $\sigma = ne\mu$, where σ is the conductivity, n the number of carriers per cubic cm, e the electron charge, and μ the mobility of the carriers in bulk samples. Once again, however, experiments to test the design, carried out by J. Hayes, H. McSkimin, W. Yager, Brattain and Ohl, did not reveal the predicted effect. Some important, as yet unknown, phenomenon was being ignored.

Shockley continued to ponder the field-effect amplifier, and on 23 June 1945 calculated that in order for the effect not to be observed, the changes in conductivity had to be less than 1/1500 of that predicted by the Mott-Schottky theory.[72] Sometime during the fall of 1945 or winter of 1945-46, he took his calculation to Bardeen, who verified its accuracy and highlighted the disparity between the existing semiconductor rectification theory and the experiments.

By the afternoon of 18 March 1946, Bardeen had developed a new theory that explained the failure of the initial field-effect experiments. This theory added to the existing model of semiconductor rectification the general postulate that electrons at the surface of a semiconductor

72. Shockley, NB 20455, 23 Jun 1945, 36.

can be trapped in "surface states." These trapped electrons do not participate in the conduction process in the semiconductor and so reduce the carriers available for the field effect. Furthermore, as Bardeen showed, if enough surface states are filled, the sheet of surface electrons forms a barrier, or space charge layer, at the free surface of the semiconductor similar to that at the interface of a metal and semiconductor. This layer electrically shields the semiconductor from the influence of the applied electric field on the control plate, thus canceling most of the effect of the external field and greatly weakening the field effect.[73]

Bardeen's surface state theory also explained other known failures of the Mott-Schottky theory, including the absence of an electrostatic potential between two different contacting semiconductors in equilibrium, and the fact that metal-semiconductor contact potentials are independent of the metal used.[74] In both cases, the explanation lies in the fact that the surface charge adjusts itself so as to cancel the difference of potential between the semiconductor and its surface, and consequently, also between different semiconductors. The failure of Shockley's initial field-effect design had led to a new physical theory, and this theory, by providing an explanation for the failure, renewed the group's confidence that they might achieve the device if they could learn enough about the nature of the surface states.[75]

Pearson, Brattain, and Bardeen undertook an extensive program of research focused on their important and mysterious new clue of surface states. Pearson studied their relation to bulk properties, such as the mobility of holes and electrons in the semiconductor; Brattain explored their effect on surface properties such as contact potential; and Bardeen collaborated closely with both.[76] These studies between March 1946 and November 1947 included the first clear experimental demonstration of the field effect by Pearson in March 1946. Pearson's experiment was based on Bardeen's deduction that, since the trapping of electrons in surface states should increase in a predictable way as the temperature

73. Bardeen, NB 20780, 18 Mar-23 Apr 1946, 38-57; "Surface states and rectification at a metal to semiconductor contact," *PR, 71* (1947), 717-727; I. Tamm, "Über eine mögliche Art der Elektronenbindung an Kristalloberflächen," *Physikalische Zeitschrift der Sowjetunion, 1* (1932), 733-746; W. Shockley, "On the surface states associated with a periodic potential," *PR, 56* (1939), 317-323. As a graduate student Shockley had suggested the possibility of surface states or "dangling bonds" at the free surface of a solid in a vacuum. Bardeen's model treated the oxidized surface of a semiconductor in air, in which Shockley's dangling bonds are already filled.

74. Bardeen (ref. 73); Meyerhof (ref. 59); S. Benzer, "Ge-Ge contacts," *PR, 71* (1947), 141.

75. Bardeen interview.

76. Traced in Pearson, NB 20912; Bardeen, NB 20780; Brattain, NB 18194; Dreher, NB 21373. Cf. Brattain, "Genesis" (ref. 3).

lowered, then at very low temperatures the electrons in surface states would effectively be frozen. They would therefore not come into equilibrium to cancel the time-varying field effect.[77] The effect should accordingly occur at very low temperatures. Pearson confirmed the prediction. The effect was, however, about fifty times smaller than Bardeen's original theory predicted; evidently the carriers had a much lower mobility in thin deposited films than they had in the bulk. The group now understood that two effects caused the large discrepancy between the experiments and calculations: the existence of the surface states *and* the much lower carrier mobility in thin films. A field effect amplifier would have to get around both obstacles.

In collaboration with Bardeen and Moore, Pearson also investigated many other properties of silicon, both pure and doped to different levels with boron or phosphorous.[78] Using accurate x-ray measurements, Pearson showed that both boron and phosphorous replace silicon atoms in the lattice, rather than entering the lattice interstitially; the boron atoms added "acceptor" levels (giving p-type conductivity) and the phosphorous "donors" (giving n-type conductivity). Pearson also contributed to the theory of temperature and impurity scattering in silicon, determining by electric resistivity and Hall effect measurements the temperature variation of the concentration of electrons and holes, the mobility of electrons and holes, and the ionization energy of both acceptors and donors. In still another program of experiments, Pearson and Moore tried, with inconclusive results, to observe directly the space-charge layer produced by the states that Bardeen had predicted at the free surface of semiconductors.[79]

Brattain's long series of experiments on the nature of the contact potential in silicon and germanium formed another essential part of the surface state program. By measuring the contact potential, he gained information about both the nature of the barrier and the density of the surface states; a higher density of surface states requires a greater field to overcome them. In one experiment, performed in April 1947, Brattain showed by means of a photovoltaic effect the existence of the surface space-charge layer that Bardeen had predicted. Light shown on n-

77. Bardeen, NB 20780, 23 Apr 1946, 47-57; Pearson, NB 20912, 22 Apr 1946, 1-11; and interviews with Bardeen and Pearson. A significant effect was not reported until 1953. Cf. Pearson, "A high impedance field-effect silicon transistor," *PR, 90* (1953), 336.

78. Pearson and Bardeen, "Electrical properties of pure silicon and silicon alloys containing boron and phosphorus," *PR, 75* (1949), 865-883; Pearson and Shockley, "Measurements of Hall effect and resistivity of germanium and silicon from 10° to 600° K," *PR, 71* (1947), 142, and "Modulation of conductance of thin films of semi-conductors by surface changes," *PR, 74* (1948), 232-233; Seitz, "The electrical conductivity of silicon and germanium," *NDRC Report*, 14-110, 3 Nov 1942; Lark-Horovitz and V. Johnson, "Theory of resistivity in germanium alloys," *PR, 69* (1946), 258-259.

79. Bardeen interview.

type silicon changed the contact potential. Apparently the holes from the electron-hole pairs produced by the light in the silicon were attracted to the surface by the negative surface charge layer arising from surface states, thus altering the contact potential. The experiment was varied many ways, ultimately in a form that resulted in the transistor.[80]

In May 1947 Brattain carried out experiments on the role of impurities on contact potential. About a year earlier Shockley had pointed out that, if the surface states are not too dense, then the contact potential between n-type and p-type samples of silicon or germanium should change if the doping were intensified. Brattain found that he could thus alter the contact potential of silicon, and his measurements led to an estimate of the surface density.[81]

Brattain also tested whether a surface barrier was already present in germanium before contact with a metal. It appeared to be so since the rectification did not depend upon the metal employed. In another experiment confirming the surface states, Brattain placed two rectifiers back to back. Under an applied voltage, charge held by surface states should build up at the interface; Brattain observed and measured this charge.[82]

Return to field effect experiments

The direct line of crucial experiments began with an experiment carried out by Brattain between 13 and 17 November 1947. While studying the change of the contact potential on illumination with light over a range of temperatures, he noticed that incidental condensation of water on the surface caused considerable hysteresis as the apparatus was taken from high to low temperatures and back again. In an attempt to avoid the hysteresis, which he regarded as a nuisance, Brattain decided to immerse the entire system in various electrolytes and dielectric liquids, including acetone, alcohol, toluene, and distilled water. To his surprise, the photovoltaic effect increased; the mobile ions in the liquids, being very close to the surface, created an electric field strong enough to overcome the shielding of the surface states.[83]

Brattain showed the results to Gibney, who on 17 November 1947 made the important suggestion, recorded in Brattain's notebook, "that I

80. Brattain, NB 18194, 2-9 Apr 1947, 78-99; Bardeen, NB 20780, 29 Sep 1947, 58-60; Brattain, "Evidence for surface states from change in contact potential on illumination," *PR, 72* (1947), 345.

81. Dreher, NB 21373, 19 May-11 Jun 1947, 85-109; Brattain and Shockley, "Density of surface states deduced from contact potential measurements," *PR, 72* (1947), 345.

82. Brattain, NB 18194; Bardeen interview.

83. Brattain, NB 18194, 13-17 Nov 1947, 138-142; interview with Brattain by Holden and King; Brattain, "Genesis" (ref. 3); Gorton, "Genesis" (ref. 4).

vary the D.C. bias on circuit while observing the light effect." (In Brattain's experiment this electrical bias between the semiconductor surface and the reference electrode had been used to counteract the contact potential.) On trial, the two noticed that varying the applied voltage caused a corresponding variation of the photovoltaic effect in cases where the liquids used in the experiment were electrolytes. They thought that the effect might be used in the design of an amplifier. In particular, as Brattain wrote, "it was found that with distilled H_2O, alcohol and acetone, ... a plus potential increased the effect and the opposite potential reduced the effect almost to zero." The role of the electrolyte was to provide ions at the surface of the semiconductor.[84]

The physical phenomenon associated with Gibney's suggestion can be understood from figure 5. Brattain was measuring the contact potential by applying a negative potential bias, V_1, to the reference electrode. The negative electrode attracted positive ions in the electrolyte and repelled negative ions, which, collecting near the surface of the sample of p-type silicon, caused positive charge in the silicon to come to the surface and fall into the positive surface states. The maneuver reduced the number of mobile carriers in the silicon. A very strong electric field existed at the surface owing to the double charge layer. The effect was measured by a photovoltaic experiment: light produced electron-hole pairs in the silicon; the positive surface charge repelled the holes and attracted the electrons; the high electric field at the surface fell, causing a measurable "photovoltaic impulse." Increasing the external field increased the impulse.

Gibney's suggestion brought the group back to the influence of applied fields on the electrical properties of semiconductors, back to an analogy to Shockley's field-effect experiments. Gibney and Brattain realized that if they could project the change in the number of mobile charge carriers in the silicon that they were causing by altering V_1 onto a suitably arranged second circuit, rather than merely observing it as a photovoltaic effect, they might be able to build a field-effect amplifier. Brattain wrote in his notebook, on 20 November, the following disclosure:[85]

> The experiments above indicate that liquid dielectric of low but finite conductivity and high dielectric constant can be used to create high electrical fields at the surface of a semiconductor and thus change the field in the space charge layer in the semiconductor without drawing very large current through the space layer. Such means therefore can be used to

84. Brattain, NB 18194, 17-20 Nov 1947, 142-151; "Genesis" (ref. 3); *PR, 72* (1948), 345; interview with Brattain by Holden and King; Shockley, "Path" (ref. 3).

85. Brattain, NB 18194, 20 Nov 1947, 151-153, resulting in patent #2,524,034 (filed 26 Feb 1948). Cf. Shockley, "Path" (ref. 3), 608, and Gorton, "Genesis" (ref. 4), 4.

> change the resistance of thin layers of the semiconductor or in other words modulate the resistance of such a semiconductor. It is evident that with a semiconductor film of the proper resistance and thickness this field could be used to change the resistance of the film by large factors without drawing appreciable currents or expending appreciable power in the dielectric semiconductor circuit. It is therefore evident that various arrangements of the dielectric semiconductor configuration could be used to amplify given signals. It is of course evident that the liquid dielectric could be replaced by a solid dielectric if one can be found having the proper ionic mobility to form such a dipole layer at the surface of semiconductor.

After almost two years of probing into the nature of the surface states and mobilities, the group appeared to be nearing its goal.

During this period Shockley continued to offer suggestions and explanations about the field effect. However, as a consequence of a visit to European solid-state laboratories in the summer of 1947, particularly Mott's in Bristol, he had become deeply involved in the theory of dislocations and the flow of electrons through alkali and silver halides. He did not participate directly in the field-effect experiments.[86]

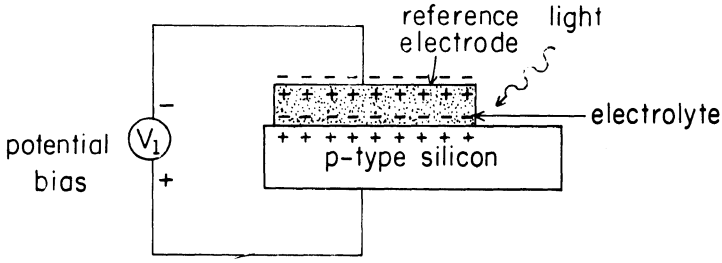

Fig. 5 Gibney's suggestion.

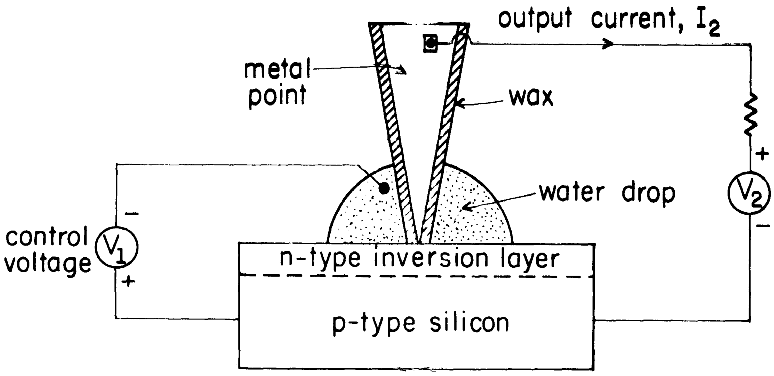

Fig. 6 Bardeen's proposed field-effect amplifier, 21 November 1947.

86. Bardeen interview.

It was Bardeen who suggested how to add a second circuit that would convert the arrangement shown on figure 5 into a field-effect amplifier. As Brattain recalls, one or two days after the group discussed the new results, "Bardeen came to my office with a suggested arrangement."[87] Bardeen's proposal, shown in figure 6, employed a metal point contact and an "inversion layer," a thin region (discussed earlier by Schottky) in which the majority carriers are opposite to that in the bulk, which in certain circumstances forms near the surface of a semiconductor. The control voltage circuit here was the counterpart of the earlier circuit (fig. 5) that applied a bias V_1 to the reference electrode, and the electrolyte was a drop of distilled water. The current in the output circuit (I_2, V_2) passed through the metal point, which was pushed down on an n-type inversion layer on a p-type silicon sample; the point was insulated by a thin layer of wax and thus separated from the electrolyte. Here again, as in Shockley's earlier field-effect design, the control potential V_1, acting between the water and the silicon, changed the number of carriers in the silicon by attracting negative charge from the silicon to the surface states caused by the positive water ions at the surface of the silicon.

As Bardeen recalls, he suggested using a point-contact device at this time simply for "convenience"; he did not intend to go to high frequencies, or to create the ideal arrangement. Considerable art and understanding had been developed in working with point contacts since the return to this device almost a decade earlier by Southworth; a point-contact amplifier design could now be set up and tested in a single day. By using an inversion layer in place of the thin film of semiconductor of Shockley's earlier design, Bardeen hoped to take advantage of the large changes in conductivity that are possible in films, circumventing two problems inherent in film experiments: the practical difficulty of depositing a thin layer of semiconductor and the problem of the low mobility of carriers in such films. Mobility is considerably higher in an inversion layer than in a film. And Brattain knew that an inversion layer could be formed on the surface of silicon.[88]

The experiment consisted of varying the applied or input voltage V_1 and measuring the resulting change (due to the change in the number of carriers in the silicon) in the output voltage V_2. On Friday, 21 November, using this apparatus, they achieved amplification of both current and power, but not of voltage. Bardeen ended his notebook entry of 22 November on an optimistic note:[89]

87. Brattain, "Genesis" (ref. 3).
88. Bardeen interview.
89. Bardeen, NB 20780, 22 Nov 1947, 61-67, on 67; ibid., 23 Nov 1947 (a Sunday!), 68-70, contains the disclosure of the field effect transistor design that led to patent #2,524,033 (filed 26 Feb 1948).

> These tests show definitely that it is possible to introduce an electrode or grid to control the flow of current in a semiconductor. Conditions were far from ideal in this first preliminary test. The drop covered a much larger area than necessary, making the control currents much larger than would be required in a proper design. A factor of 100 or more could readily be obtained. It is not known whether or not the thickness of the N-type layer used in these tests was optimum. Much more could be done, of course, in improving the insulation between the metal point and the electrolyte. The electrode could be put much closer to the point to improve the voltage amplification factor (which is now less than unity).

The device was not yet usable for other reasons too: the water drop tended to evaporate during experimental tests and the apparatus responded only to low frequencies. During the next two weeks they tried many modifications.[90] Gold and tungsten were used as electrodes and coated with Duco lacquer as well as with wax; for the electrolyte, besides water, tetramethyl ammonium hydroxide was tried, and, at Moore's suggestion, various gels. On 26 November Brattain reported that gelatin in water "showed some promise but dried out"; they achieved better results with a mixture—they called it "gu"—of glycol borate and glycol bori-borate. Several geometries were considered, the most promising one having, in place of the water drop, a ring of gold on the "gu" with the metal point, as before, passing through the center of the electrolyte. Other configurations included a drop of the electrolyte on a junction of a p- and an n-type sample, two metal points placed close together on the n-type inversion layer on a p-type sample, and a sandwich of germanium on SiO_2 on top of the n-type silicon inversion layer. Brattain noted on 8 December that, during a luncheon discussion, Bardeen had suggested replacing the silicon by the "high-back-voltage" germanium studied by the Purdue group. Since this high purity germanium made better rectifying contacts than ordinary silicon or germanium, they guessed that its use would improve the voltage amplification ratio. They soon found that an inversion layer would form on the special germanium, which they then used in place of silicon.[91]

On 8 December 1947, with a gold ring electrode on a circular glob of the glycol borate mixture on the special germanium, and the metal point contact passing through the center of the ring, Bardeen and Brattain found a voltage amplification of two and a power amplification of 330.[92] The frequency response remained poor owing to the glycol

90. Brattain, NB 18194, 24 Nov-8 Dec 1947, 156-177.
91. Bardeen interview; Bardeen, "Semiconductor research" (ref. 3).
92. Brattain, NB 18194, 8 Dec 1947, 176.

borate; the device failed to amplify at frequencies much above 10 cycles.[93]

Then, accidentally, working with the glycol borate and high-back-voltage germanium, Bardeen and Brattain observed an unexpected effect of great importance: they found that with a steady bias on the electrolyte, the glycol borate would etch the surface of the germanium, forming, as Gibney pointed out, an oxide film.[94] Brattain recalls: "We could see through the glycol borate that we were anodizing, growing visible interference films. . . . I can remember the green color under the glycol borate."[95] They speculated that this accidentally grown oxide film between the electrolyte and semiconductor might, if insulating, act as a condenser and, since it was very thin, enable them to obtain a very large field with a moderate voltage and so increase the degree of modulation of the current flowing through the inversion layer. In effect the film would replace the glycol borate and might give a higher frequency response.

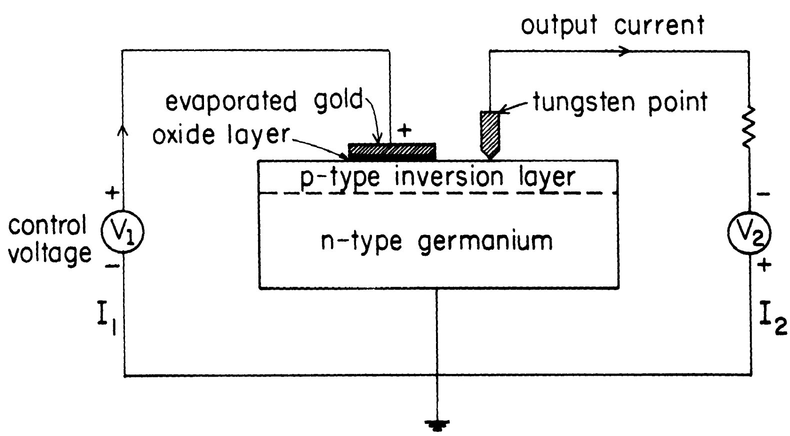

Fig. 7 Bardeen's and Brattain's field-effect model of 11 December 1947, in which accidental contact occurred.

In the next experiment, performed on 11-12 December 1947, the currents were reversed by the replacement of a p-type sample by an n-type sample (fig. 7). An oxide replaced the drop of water. And to serve as the voltage plate in the control circuit, Gibney carefully evaporated a circular spot of gold upon the somewhat thicker oxide film, which had been grown on the germanium using a steady applied voltage. He left a hole in the center of the gold to allow the metal point to contact the germanium as in the previous configuration; but an electri-

93. Brattain interview by Holden and King.
94. Brattain, NB 18194, 10 Dec 1947, 179.
95. Brattain interview by Holden and King.

cal discharge between the point and the gold spoiled the hole, and Brattain resorted to placing the point at the edge of the gold, thus physically separating the two neighboring circuits, as the figure indicates.[96]

The point-contact amplifier

Once again the elements for observing the field effect were present. However, the experiment did not work as planned, for the oxide layer did not insulate. To their great surprise, on 11 December 1947, Brattain and Bardeen observed modulation of the output current and voltage in this configuration. But it occurred only when the gold was biased positively, i.e., oppositely to what was expected![97]

According to the earlier arguments (cf. fig. 5) based on the supposition of a field effect, if the positive bias V_1 on the gold electrode increased, then holes would be driven away from the surface and out of the p-type inversion layer, decreasing I_2. However, the current I_2 in the output circuit, biased in the reverse direction, increased rather than decreased as the positive bias on the gold increased. This indicated that holes—the minority carriers in the bulk material—were flowing *into* rather than out of the inversion layer through some accidental contact with the positively biased gold. Brattain wrote in his notebook on 19 December: "It would appear then that the modulation obtained when the grid point [the gold] is bias + is due to the grid furnishing holes to the plate point [the tungsten]."[98] On realizing that the grid emitted holes, they renamed the control circuit the "emitter" and the output circuit through the tungsten point the "collector"; the corresponding currents were designated I_e and I_c respectively.[99]

The attempt to use the oxide layer as a field-effect capacitor had

96. Brattain, "One researcher's account" (ref. 3), 9; Brattain, NB 18194, 12 Dec 1947, 185.

97. Brattain, NB 18194, 12 Dec 1947, 183-190, esp. 187; Brattain, "Genesis" (ref. 3); Bardeen, "Semiconductor research" (ref. 3), 96; Brattain interview by Holden and King; Bardeen interview. The "accident" is passed over in the official account by Gorton, "Genesis" (ref. 4), 4-5.

98. Brattain, NB 21780, 19 Dec 1947, 4. Brattain and Bardeen's notebooks show that they attributed the effect to the injection of holes. However, Shockley claims ("Path" [ref. 3]) that the "first suggestion that minority carrier injection might be an important mechanism was made five weeks later in my disclosure [23 Jan 1948]." This disparity may be understood by noting that for Shockley the term "minority carrier injection" refers to the detailed mechanism of the phenomenon, which was indeed only understood later after the detailed investigations of Shockley, John Shive, J. Haynes, and E. Ryder. As Bardeen explains (interview) at the time of the first experiments on the point-contact transistor, it was not clear that holes could diffuse three dimensionally through the bulk of the semiconductor.

99. What is now called the "base" contact was originally called the "control" contact because it plays a role analogous to the grid in a vacuum tube. Since control and collector both start with the same letter, the former was changed to base (Bardeen interview).

therefore failed. But growing the oxide film and treating it with glycol borate had helped to produce the large inversion layer at the surface needed for the effect. The device gave some voltage amplification but no power amplification. It showed Brattain and Bardeen that one could build an amplifier on a principle entirely different from the field effect, that of *point-contact*.[100]

The Bell Laboratories group was not the first to observe hole injection. In 1947 a group at Purdue including Ralph Bray, who had studied spreading resistance in the high-back-voltage germanium rectifier during the war, noted that the resistance increased with voltage when the rectifier was biased in the forward direction. This effect, arising from the injection of holes into the germanium, was the key to the point-contact transistor, but the Purdue researchers did not initially understand its physical basis and could not, therefore, use the phenomenon in the design of an amplifier.[101] The Bell group became aware of Bray's experiments before they had completed their crucial experiments. As Bardeen recalls, they realized that the conductivity increased near the point by the injection of holes into germanium and became concerned that Purdue might beat them to what turned out to be their discovery of the first transistor.[102]

The team had yet to produce a usable amplifier—a device in which the change in the collection current not only followed but exceeded the change in the emitter current. They had somehow to make the hole current flowing from emitter to collector to alter the barrier layer at the collector so as to make it easier for electrons to flow in the collector circuit. The climax of Brattain and Bardeen's experiments employed Bardeen's suggestion that they could perhaps obtain greater amplification if two contacts on the germanium were placed only a few mils apart. They believed that they could get a stronger effect if they used line rather than point contacts, but it later became evident that point contacts would do if enough current flowed in the collector. Brattain solved the technical difficulty of getting the line contacts close enough by wrapping a piece of gold tape around the tip of a triangle of

100. Bardeen believes that his discovery with Brattain of the point-contact transistor may have slowed the advancement of the transistor field because it diverted the semiconductor program from junction and field-effect transistors, which subsequently proved to be far more useful commercially. However, had the oxide layer been insulating as expected, the first successful transistor would have been a field-effect rather than a point-contact device, working as do the present MOS (metal-oxide-semiconductor) silicon transistors. This experiment can now be done using silicon, but not germanium (Bardeen interview).

101. "NDRC Report 14-585. Purdue University, covering March 1942 to November 1945," and "Final report of work performed on contract between Feb. 1, 1946-Jan. 31, 1949, Purdue University" (Purdue University Collection); interview with H. Y. Fan.

102. Bardeen interview.

polystyrene, cutting a slit on the gold at the apex with a razor, and filling the cut with wax. The separation was about 4×10^{-3} cm.[103]

The classic experiment, shown in figure 8 and differing from the arrangement in figure 7 only in the geometry and constitution of the contacts, worked the first time Bardeen and Brattain tried it, on 16 December 1947. In one of the first experiments at 1000 cps, the power gain was 1.3 and the voltage gain fifteen.[104] Thus was the point-contact transistor born.

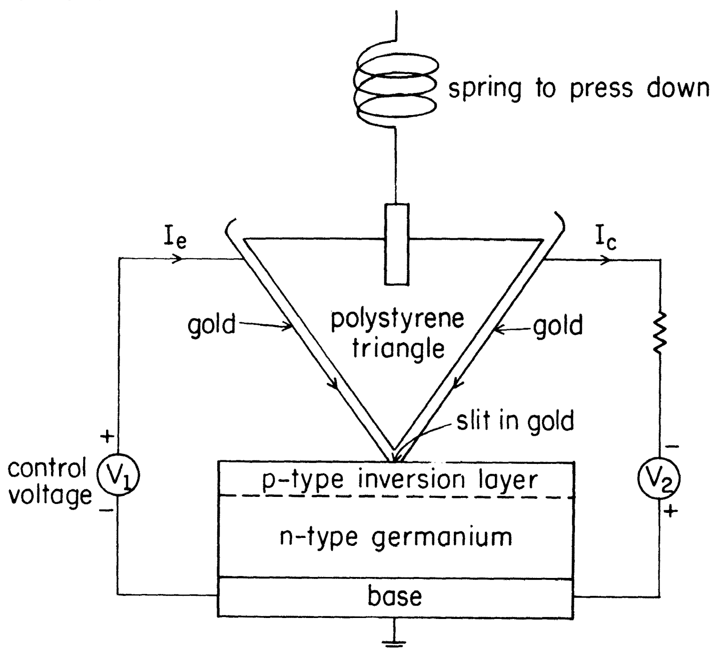

Fig. 8 The first transistor.

After a week of checking their results, Bardeen and Brattain showed them to executives at Bell Laboratories. They connected the input to an audio signal and the output to an oscilloscope. Brattain reported in his notebook on 24 December 1947:[105]

> This circuit was actually spoken over and by switching the device in and out a distinct gain in speech level could be heard and seen on the scope presentation with no noticeable change in quality. By measurements at fixed frequency in it was determined that this power gain was the order of a factor of 18 or greater. Various people witnessed this test and listened

103. Brattain, NB 18194, 16 Dec 1947, 192-193; Brattain interview by Holden and King.

104. Ibid.; Brattain, NB 18194, 16 Dec 1947, 193-194.

105. Brattain, NB 21780, 24 Dec 1947, 6-9; Brattain, "One researcher's account" (ref. 3).

(were present) of whom some were the following; R. B. Gibney, H. R. Moore, J. Bardeen, G. L. Pearson, W. Shockley, H. Fletcher, R. Bown. Mr. H. R. Moore assisted in setting up the circuit and this demonstration occurred on the afternoon of 23 December 1947....

This morning H. R. Moore changed the circuit on scope 7.... With this circuit the device could be made to amplify audio signal and by turning [sic] of the input audio signal...and putting phones across output switch the device could be heard to oscillate. The oscillations were also observed on the scope....

On 24 December Bardeen gave a more detailed explanation:[106]

> Voltage gains up to about 100 and power gains up to about 40 have been observed. The explanation is believed to be as follows. When A [the gold electrode] is positive, holes are emitted into the semi-conductor. These spread out into the thin P-type layer. Those which come in the vicinity of B [the tungsten point] are attracted and enter the electrode. Thus A acts as a cathode and B as a plate in the analogous vacuum tube circuit. The ground corresponds to the grid, so the action is similar to that of a grounded grid tube. The signal is introduced between A (the cathode) and ground (grounded grid). The output is between B (the plate) and ground. The signs of the potentials are reversed from the (sic) those in a vacuum tube because conduction is by holes (positive charge) rather than by electrons (negative charge). The analogy was suggested by W. Shockley.

The transistor was named by John Pierce in Brattain's office. Brattain recalls:[107]

> [We] wanted a name that would fit with "thermistor" and "varistor". Pierce first mentioned the important factor in vacuum tubes that was called "transconductance." Pierce knew that the point contact amplifier was the electrical dual of the vacuum tube. After some thought he then mentioned the dual of transconductance which is "transistance" and then he said "transistor".

Brattain's response was: "That is it!"

The discovery was kept "laboratory confidential" for seven months, for Bell was concerned that the breakthrough might be occurring simultaneously elsewhere, perhaps at Purdue.[108] Meanwhile, Bardeen worked with an attorney to draw up patents. The classic paper announcing the discovery was sent to the *Physical review* with a request to hold back publication until Bell had received word that the discovery would not be classified.

106. Bardeen, NB, 20780, 24 Dec 1947, 71-74, on 72.

107. Brattain to Hoddeson, 10 Jan 1979.

108. Harold Zahl recalls Shockley's asking him in late June 1948 whether the Purdue researchers had already made the discovery. They had not. The military had, however, heard a (false) rumor that the Naval Research Laboratory had done similar work. Zahl, "Birth of the transistor," *Microwave journal,* 9 (1966), 94-96; Bardeen interview.

The military had not supported any of the research leading to the transistor,[109] but Bell felt it necessary to get release because of the invention's potential defense applications. The military saw the point-contact transistor on 23 June 1948, the day after its first demonstration to the entire technical staff at Bell. Harold Zahl, Chief of the Research Branch of the Signal Corps Engineering Division in Fort Monmouth, New Jersey, recalls that during the discussions "The analogy to the discovery of fission was given, and some suggested that we should make another big Manhattan Project and go 'underground.' Personally, I was very opposed."[110] In view of the transistor's broad applications, it was decided not to classify it.

On 30 June, after this decision, news of the discovery was released to the press. On 1 July *The New York Herald Tribune* reported, "The device is still in the laboratory stage but engineers believe it will cause a revolution in the electronics industry!" *The New York Times*, mentioning the discovery in a small item in their "News of Radio" section, reported: "A device called a transistor, which has several applications in radio where a vacuum tube ordinarily is employed, was demonstrated for the first time yesterday at Bell Telephone Laboratories...where it was invented."[111]

From this point on, Bell Laboratories engaged in a major effort to exploit and develop the transistor. A team under Jack Morton began to explore applications. Shockley went on to work out the theory of the pn junction and to predict the junction transistor, which Morgan Sparks built after the discovery in 1950 by G. Teal and J. Little of a practical method of pulling single crystals from a melt of germanium.[112] The original transistor ultimately led to new families of solid-state amplification devices (which would soon eclipse the point-contact transistor's technological significance) and to the creation of whole new industries. It contributed significantly to the mushrooming of basic research in solid-state physics during the 1950s and 1960s, both at Bell Laboratories and in the larger scientific community.

109. Military support for research and development of the transistor at Bell Laboratories began in June 1949 (Fagen [ref. 39], 621-625).

110. Zahl (ref. 108); interview with Bardeen and (by Weiner) with Brattain.

111. In his interview with Weiner and in his letter to Hoddeson, 10 Jan 1979, Brattain recalls the response of the Purdue group. Both he and Bardeen knew about the spreading resistance experiments before 23 Dec 1947. During the six-month secrecy period they heard Benzer read a paper to the American Physical Society on "Photoelectric effects and their relation to rectification at metal-semiconductor contacts" (*PR, 73* [1948], 1256). In later discussions at the meeting, as Brattain recalls, either Bray or Benzer suggested adding a second point on the surface of the semiconductor close to the first and measuring the distribution of potential. Brattain, who had already done this, answered, "Yes, I think maybe that would be a very good experiment."

112. Shockley, "The theory of pn junctions in semiconductors and pn junction transistors," *Bell System technical journal, 28* (1949), 435-489; Shockley, "Path" (ref. 3); G. K. Teal and J. B. Little, "Growth of germanium single crystals," *PR, 78* (1950), 647.