A SEMICENTURY OF SEMICONDUCTORS

Researchers and technologists have grappled with semiconductors for more than 50 years, and their labors have had a revolutionary impact on science and society.

Alan B. Fowler

"One shouldn't work on semiconductors, that is a filthy mess; who knows whether they really exist."
—*Wolfgang Pauli, 1931*[1]

For more than 50 years the symbiotic relationship among semiconductor physics, technology and device engineering has exemplified cooperative activity that spans the continuum of the scientific enterprise, from the purest physics to the marketplace. In physics this activity has led to the observation of unexpectedly rich and intricate phenomena; in technology it has resulted in expanded techniques and invention with broad applications and key contributions to the electronics, communications and computer revolutions. Semiconductors have changed our world profoundly and probably beneficially, touching the lives of almost everyone in it—and not just by fueling advances in data handling and communications but also by making possible such consumer staples as the transistor radio and the compact disk player. Some might choose to designate the last half-century as the nuclear age or the jet age; others might think of it as the age of semiconductor electronics.

Among the first people to apply physics to semiconductors were Alan H. Wilson and Walter Schottky. According to an excellent historical review by Lillian Hoddeson, Gordon Baym and Michael Eckert,[2] by 1922 Eduard Grüneisen had distinguished semiconductors as solids with a maximum in conductivity as a function of temperature, although the term itself had appeared by 1911 and may have been used even earlier. Many of the great quantum mechanicians of the late 1920s and 1930s contributed to the theory of metals and insulators. In the 1930s Wilson, in a series of papers and books noted for their clarity and insight, provided a solid basis for the theoretical understanding of semiconductors. He clarified the concept of holes and electrons, explained the function of the energy gap in insulators and developed the understanding of the role impurities play in converting an insulator to a semiconductor. In the more technical arena, Schottky contributed, along with many other ideas, the best model for rectifying metal–semiconductor contacts.

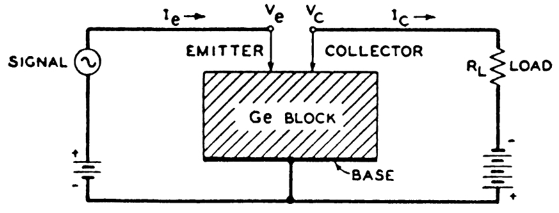
Little semiconductor physics was done in the US before 1940. The indexes of the *Physical Review* for 1935 and 1936 show four entries under "Photovoltaic . . ." in the first year and no references at all to semiconductors in the next. The first year in which "Semiconductors" appears as a topic in the index seems to be 1946, under "Electrical Conductance and Resistance." By contrast, a recent issue of the semimonthly *Physical Review B15* contains over 50 papers on semiconductors, taking up more than 500 pages of text. At present *B15* is "the" international journal of semiconductor research, with more than half the articles on that topic coming from researchers in countries other than the US.

During the war, between 1939 and 1945, there was extensive development of semiconductor technology for use in microwave detectors, which were technologically advanced versions of the "cat's whisker" of early radio. However, real advances in understanding had to await the return of peace. In the US, postwar work was heavily concentrated at Purdue University under Karl Lark-Horovitz and at the Bell Telephone Laboratories under William Shockley, although other laboratories certainly played a role.[3] Shockley had set out to find a solid-state replacement for the vacuum tube. His first approach was to develop a field-effect transistor using silicon films (to make what would now be called a thin film transistor, or TFT). The idea of the field-effect transistor was an old one, going back at least to Charles F. Mott (the father of Nevill Mott), who at the turn of the century attempted and then abandoned a thesis under J. J. Thomson in which he proposed to cause a change of resistance in thin metal films using capacitive induction of charge. He failed to produce a measurable result. So did others who reinvented and tried this experiment for some years to come. Silicon films showed a weak change in conductance, which was less than expected or useful. It was observed in 1947 that the barrier height of a Schottky contact was the same for all metals studied, even though the metals had very different work functions. John Bardeen suggested that this was because surface states, associated with the termination of the semiconductor lattice, pinned the metal Fermi level at a fixed point in the energy gap and thus fixed the height of the barrier. And it was these same states that soaked up the induced charge in the silicon field-effect transistors, immobilizing that charge and making the film nonconducting.

Invention of the bipolar transistor

In 1947, while Bardeen and Walter Brattain[4] were studying these surface states using two metal point-contact probes, they observed bipolar transistor action: the amplification of power carried in by current through a forward-biased, lower-resistance contact to a semiconductor and of power carried out through a reverse-biased, higher-resistance contact. Almost immediately upon making this observation they explained their result in terms of minority-carrier injection, and new eras in both electronics and semiconductor physics were born. This so-called point-contact transistor was difficult to reproduce or understand completely

Alan Fowler recently retired from the IBM Thomas J. Watson Research Center, where he worked for 35 years, most recently as an IBM Fellow.

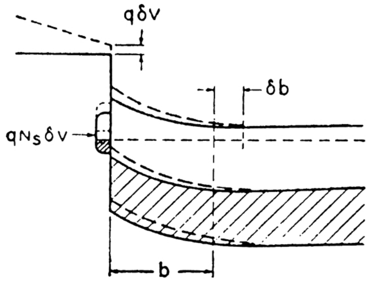


Semiconductor triode schematic from the first paper on bipolar transistors,[4] published in *Physical Review* in 1948. That paper sparked the explosion of work on semiconductors in the middle of this century.

because of the nature of the collector contact, and it never did become industrially important. But its invention was the spark that ignited the explosion of energy devoted to making better crystals, developing better technology and intensifying the scientific study of semiconductors. Thus although the original impetus for the work was technological, it was the scientific studies of surface states that led to the invention of the bipolar transistor and, ultimately, to the multi-billion-dollar electronics and computer industries.

Shortly after the discovery of the point-contact transistor, Shockley invented the junction bipolar transistor,[5] which proved to be not only simpler to describe analytically but also easier to make reproducibly than the point-contact transistor. It was for many years the dominant semiconductor device and is only now being displaced by field-effect devices as the key electronic element of high-performance computer logic circuits.

In the 1950s there was an explosion of both scientific and technological interest in semiconductors. By the late 1950s industrial, government and university laboratories had grown the elemental (Si and Ge) and many compound semiconductors into high-quality single crystals,[6] and most of their bulk properties were well understood. Theories to describe band structure began to appear in the literature, starting with the work of Frank Herman,[7] who used the orthogonalized plane-wave method pioneered by Conyers Herring a decade earlier. Experiments, especially magnetotransport[8] and cyclotron resonance studies,[9] guided the fuller development of the band structure calculations. Researchers (especially Herring) developed satisfactory theories of transport. The first work on localization and on disorder—done by many university, government and industrial laboratories, from many countries—appeared.



Energy level diagram, from the same issue of *Physical Review* as the figure above, illustrates the basis of the field-effect transistor: An external field induces charge in surface states, inhibiting gain. (From W. Shockley, G. L. Pearson, *Phys. Rev.* **74**, 232, 1948.)

In addition there had been during this decade a proliferation of techniques for making, "passivating" (protecting) and packaging transistors for commercial and military use. Much of this innovation came from Bell Labs, but dozens of laboratories—primarily industrial—contributed to the technology. Increasingly toward the end of the 1950s technological attention turned to silicon. This was not because silicon has transport properties superior to those of germanium or the III–V compounds. It doesn't. Nor was it entirely because silicon has a larger energy gap than germanium, although this was certainly a factor in the important military support of silicon technology development.

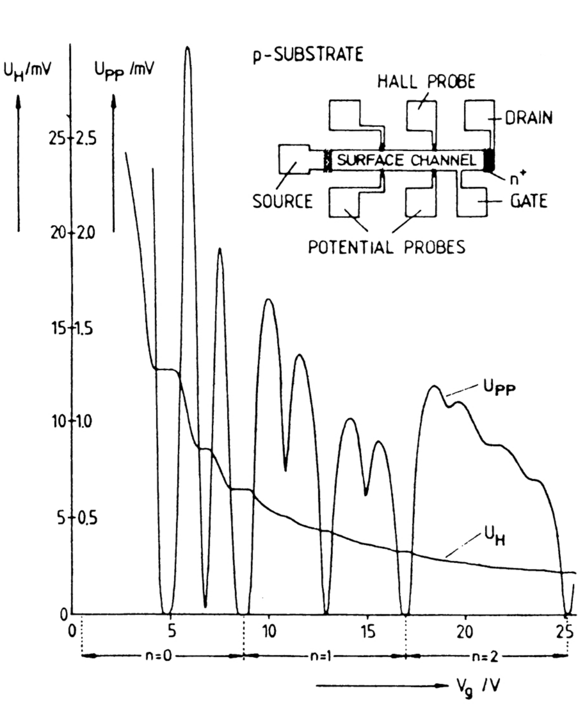
Rather, the reasons for the success and eventual dominance of silicon technology were technological and chemical. The new techniques for transistor and integrated circuit fabrication were peculiarly suited to silicon. Those "planar" techniques were based on the diffusion of dopants into the semiconductor through patterned openings etched into a masking material. The patterns were increasingly produced by photolithography on a silicon dioxide surface layer grown on silicon. Silicon dioxide could be grown in tightly adherent layers on silicon. Furthermore, silicon dioxide is a good (but not perfect) barrier to the diffusion of impurities at fabrication temperatures as well as at operating temperatures, so that it passivates the underlying structures. The interface state densities at the Si–SiO_2 interface are low, so that loss of induced electrons to traps is negligible compared with that in earlier field-effect devices. And SiO_2 is a superb electrical insulator, with high breakdown strength and a high barrier for the injection of electrons from either silicon or metals. Thus by the early 1960s much of the technology that has led to the highly complex and highly integrated electronic circuits of today was available, at least in crude form.

Gallium arsenide technology

For the most part, the technologists—that is, the metallurgists, chemists, engineers and physicists who gravitated toward these fields—developed these silicon-based device fabrication techniques. But what were other semiconductor physicists doing from the 1950s well into the 1980s? Much of their effort went into the study of alternative, "high performance" semiconductors, and most of this work focused on gallium arsenide. GaAs has found use primarily in light-emitting diodes, injection lasers and MESFETs (metal–semiconductor field-effect transistors), used mostly in military and "niche" applications. Even in advanced technologies GaAs has never seriously challenged silicon for general purpose applications. Many physicists invented new devices that were destined to fulfill lesser roles than the silicon bipolar transistor. Some of these devices—such as the tunnel diode invented by Leo Esaki[10] and the Gunn diode—elucidated novel physical phenomena. There were two exceptional devices besides the bipolar transistor that saw large-scale use: the semiconductor injection laser[11] and the MOSFET, or metal oxide–silicon field-effect transistor.[12]

The injection laser[11] grew out of a series of experiments at several labs—notably RCA, GTE and Lincoln Laboratories—all of which contributed to the realization that a direct-bandgap semiconductor was necessary for efficient spontaneous emission and that GaAs diodes exhibited high quantum efficiency in producing light. In 1962 researchers at GE and IBM, and soon after at Lincoln Labs, observed and reported stimulated emission. Although niche uses were quickly found for the laser and for the related light-emitting diodes, it took about 20 years to develop major

Recordings of the Hall voltage U_H, first reported in *Physical Review Letters* in 1980. The resistance is quantized at values corresponding to $h/e^2\nu$, where ν is the number of filled Landau levels. Inset shows a top view of a typical MOSFET device used to obtain the Hall voltage. This measurement now sets the resistance standard. (From ref. 14.)



industrial or consumer uses for the device, in part because the development of these applications had to await much technological innovation and many subsidiary inventions (such as the heterojunction laser). For example, it took many years and extensive materials research to produce semiconductor materials with satisfactory levels of reliability. The creation of communications applications required the development of light fibers and other technologies. Ternary compounds such as InGaAs replaced or supplemented GaAs. The invention of the compact disk player required not only new technology but also the invention of a market. Those who believe that the canonical 17 years between the appearance of a revolutionary invention and its major impact is an unnecessarily long interval would do well to study the history of such devices in detail. Solid-state physicists played a major role in these developments—especially at the outset. At times the roles played by physicists, engineers and materials scientists were indistinguishable, partly because of increased interdisciplinary work in the universities and partly because physicists moved in increasing numbers into electrical engineering and materials science departments from the late 1950s onward.

Although physicists were among the first to study field-effect devices, the first practical implementation, the MOSFET, grew directly out of technological studies done at Bell Labs in the late 1950s.[12] The device itself was invented in 1960, became a major element in computer memories by 1972 and is now the dominant active (or amplifying) semiconductor device. The basic understanding of this device derived from the earlier field-effect devices mentioned above.

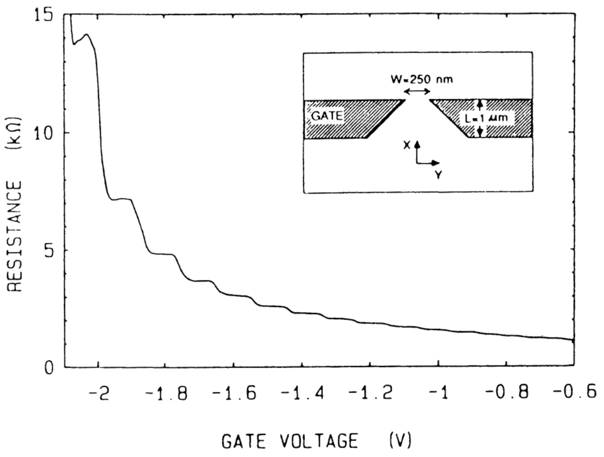
The development of the MOSFET is an example of the enhancement of physics by technology: From 1965—when physicists at IBM first observed two-dimensional electron gases in MOSFETs—until the present, innovative semiconductor physics increasingly depended upon semiconductor device technology. Starting in that year and continuing into the early 1980s the MOSFET remained the primary tool for studying two-dimensional systems.[13] As important as the two-dimensionality was as a phenomenon, equally important was the ability to vary carrier concentrations by more than two orders of magnitude. This capability allowed the study of such properties as mobility, electron–electron interactions and localization over a large range of conditions in a single sample. Work on the transport properties of MOSFETs reached a climax with the discovery of the quantum Hall effect[14] by Klaus von Klitzing, Gerhard Dorda and Michael A. Pepper in 1980. This discovery has found use as a standard for resistance and has become one of the central topics for study in condensed matter physics. The investigation of MOSFETs led to many other important results. The first observations of the ballistic motion of electrons were made in thin oxide films in MOSFETs.[15] Extensive studies of both weak and strong localization, the earliest investigations of one-dimensional systems and the first observations of mesoscopic systems were also made on MOSFET samples.

Heterostructures

An important characteristic of MOSFETs was that the mobile charge is supplied by induction by the gate, so that there is no compensating ionic charge to scatter the electrons, as would be the case in heavily doped semiconductors. This principle also applies to modulation-doped structures[16] made in GaAs–AlGaAs heterostructures, which have far more nearly perfect interfaces than do MOSFETs. Three characteristics of GaAs—greater perfection, lesser interface roughness and charge, and lighter electron mass—result in electron mobilities more than two orders of magnitude larger than those in silicon. GaAs technology had grown out of early efforts by Esaki and Raphael Tsu[17] to make superlattices and the development of molecular-beam epitaxy. Heterojunction resonant tunneling structures[18] grew directly out of this work. GaAs technology led as well to the discovery of the fractional quantum Hall effect[19] and has been the main vehicle in the search for convincing evidence of electron crystallization in a quantum mechanical system. Quantum well lasers—another outgrowth of GaAs technology—are of considerable technical interest.

GaAs–AlGaAs heterojunctions have been the main tool for studying one- and zero-dimensional systems and ballistic effects. It is the heterojunction structure's extremely long mean free paths (greater than 10 microns) that have made these experiments possible. This property has also allowed the observation of many quantum mechanical effects in laboratory-made samples rather than in atoms or molecules. While no practical applications of nanostructures and ballistic effects are yet apparent, there have recently been some exciting developments, such as the observation of quantized conductances in point contacts.[20] Much has been made of the possible use of nanostructures in electronics; however, as yet, no technically viable semiconductor nanostructure has been made or proposed. This is partly because cryogenic temperatures are required and partly because the quantum of resistance (h/e^2) is too large to be useful in high-speed circuits. The field is in its infancy, however.

Starting in the mid-1950s researchers in increasing numbers began to study semiconductor surfaces under extremely clean, ultrahigh-vacuum conditions. Some 40 years of intensive and at times inspired research on these surfaces have resulted in the observation and explanation of a wealth of phenomena and the development of myriad new techniques. These studies were often justified on the basis that



Point-contact resistance of a surface structure. The resistance is quantized at values of $h/e^2\nu$, where ν is the number of quantized channels allowed through the point contact. Inset shows the arrangement of the point contact. (From ref. 20.)

they would contribute to the understanding of interfaces in small devices. (Condensed matter physicists have long been obsessive about justifying even their most basic and elegant physics by its eventual utility.) Analytic and vacuum techniques made possible the development of molecular-beam epitaxy, which may yet play a major technological role. The use of the scanning tunneling microscope has been widely heralded as the technology of the future for building nanostructures; it is expected to surpass electron-beam lithography. For the technique to be useful, however, its developers must find a way to multiplex the tips (where "multiplex" implies the use of multiple processing paths), a step only now being taken for electron beams.

Schottky (metal–semiconductor) barriers have remained for more than 50 years a favorite topic of study and the subject of sometimes acrimonious controversy for semiconductor experimentalists and theorists. The problem has always been that barrier heights are very sensitive to artifacts and defects, making the achievement of an "ideal" interface elusive. Another area of continued interest to many physicists has been amorphous semiconductor films. Their behavior has been easy to study experimentally but difficult to model in detail theoretically. Nonetheless, intensive work starting in the 1960s has led not only to a large body of descriptive science in this field but also to applications in solar cells and the thin film transistors used in liquid crystal displays.

Underlying all of this work have been continued studies, both experimental and theoretical, of defects in semiconductors. From the time of Wilson, solid-state scientists have recognized the crucial role of defects. As time has progressed and computational techniques have improved, it has become more nearly possible to match the theory of defects to experiment.

Challenges ahead

Given the limitations on space in this article, I have been unable to mention most of the various contributions in the vast field of semiconductor science, much less many major papers or contributors. Thousands of physicists have been involved in these efforts, and their work has resulted in a vast edifice of scientific knowledge and technique that has led and will lead to many technological revolutions.

Industrial laboratories have traditionally led much of the research in this field. The number of corporations willing to support such research has steeply declined, and even those research laboratories still in operation have seen their activities much curtailed. The reasons for this cutback are many and complex and generally not understood by the university community, which must be affected in the long run. Miniaturization of devices has brought with it escalating costs for new development. Semiconductor device development has not been slowed as yet by physical limits but rather by economic ones. A slowdown in the pace of technological development will inevitably erode the military advantages of the US, its allies and clients, since these advantages are based to a large extent on technical superiority. Thus the decision to reduce the pace of development, though largely an economic one, is also political and possibly moral.

The health of semiconductor science, for all its many results in "pure" physics, continues to depend on its potential for application. There are still many interesting problems in semiconductor science and many technological challenges. We may expect that as in the past, the new physics being developed will be the precursor to unexpected applications.

References

1. Cited in L. Hoddeson, G. Baym, M. Eckert, in *Out of the Crystal Maze*, L. Hoddeson, E. Braun, J. Teichmann, S. Weart, eds., Oxford U. P., New York (1992), p. 121.
2. L. Hoddeson, G. Baym, M. Eckert, Rev. Mod. Phys. **59**, 287 (1987).
3. For a more nearly complete review of the history of this period, see E. Braun, in *Out of the Crystal Maze*, L. Hoddeson, E. Braun, J. Teichmann, S. Weart, eds., Oxford U. P., New York (1992), p. 443.
4. J. Bardeen, W. H. Brattain, Phys. Rev. **74**, 230 (1948). For one view of the history of the transistor see N. Holonyak Jr, PHYSICS TODAY, April 1992, p. 36.
5. W. Shockley, Bell Syst. Tech. J. **28**, 435 (1949). For another view of the history of the transistor, see W. Shockley, IEEE Trans. Electron Devices **7**, 597 (1976).
6. G. K. Teal, J. B. Little, Phys. Rev. **78**, 63 (1967).
7. F. Herman, Phys. Rev. **88**, 1210 (1952). F. Herman, J. Callaway, Phys. Rev. **89**, 518 (1952).
8. G. N. Pearson, G. H. Suhl, Phys. Rev. **83**, 786 (1951). B. Abeles, S. Meiboom, Phys. Rev. **95**, 31 (1954).
9. G. Dresselhaus, A. F. Kip, C. Kittel, Phys. Rev. **92**, 827 (1953). B. Lax, J. Ziegler, R. N. Dexter, E. S. Rosenblum, Phys. Rev. **93**, 368 (1954).
10. L. Esaki, Phys. Rev. **109**, 603 (1958).
11. R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Salty, R. O. Carlson, Phys. Rev. Lett. **9**, 366 (1962). M. I. Nathan, W. P. Dumke, G. Burns, F. H. Dill Jr, G. Lasher, Appl. Phys. Lett. **1**, 62 (1962). T. M. Quist, R. H. Rediker, R. J. Keyes, W. E. Krag, B. Lax, A. L. McWhorter, H. J. Zeiger, Appl. Phys. Lett. **1**, 91 (1962). For a brief history of the semiconductor injection laser, see R. N. Hall, IEEE Trans. Electron Devices **7**, 700 (1976).
12. For a brief review, see D. Kahng, IEEE Trans. Electron Devices **7**, 655 (1976), and refs. therein.
13. A. B. Fowler, F. F. Fang, W. E. Howard, P. J. Stiles, Phys. Rev. Lett. **16**, 1901 (1966).
14. K. von Klitzing, G. Dorda, M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).
15. G. Lewicki, J. Maserjian, J. Appl. Phys. **46**, 3032 (1975).
16. R. Dingle, H. L. Störmer, A. C. Gossard, W. Wiegmann, Appl. Phys. Lett. **33**, 655 (1978).
17. L. Esaki, R. Tsu, IBM J. Res. Dev. **14**, 61 (1970).
18. L. L. Chang, L. Esaki, R. Tsu, Appl. Phys. Lett. **24**, 593 (1974).
19. D. C. Tsui, H. L. Störmer, A. C. Gossard, Phys. Rev. Lett. **48,** 1559 (1982).
20. B. Van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. T. Kouwenhoven, D. van der Marel, C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988). D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, G. A. C. Jones, J. Phys. C **21**, L209 (1988).