# Highlights in Semiconductor Device Development

## L. Esaki

IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598

Following a brief description of early semiconductor history, the invention of the transitor and subsequent important events are presented in perspective, with emphasis on the role of semiconductor physics in device development.

Key words: diode; field-effect transitor; IMPATT diode; integrated circuit; LASER; LED; photocell; solar cell; solid state; transitor; tunnel diode; III-V compounds.

## I. Introduction

The agricultural civilization in the cultural history of man was said to be the result of two genetic accidents which gave birth to a new species of bread wheat some 10,000 years ago, involving wild wheat and goat grass. Large-scale agricultural activity in man's society followed. Great inventions or discoveries could be considered to be such genetic accidents—mutations. New knowledge, arising from these inventions, often leads to a large-scale engineering effort which eventually has far-reaching consequences in our society. The invention of the transistor by three solid state physicists, Shockley, Bardeen, and Brattain, is one such example. The development of the transistor began in 1947 through interdisciplinary cooperation with chemists, metallurgists, and electronic engineers, at Bell Laboratories. A large-scale development effort for a variety of semiconductor devices followed in a number of institutes throughout the world. Semiconductor know-how, thus established, has revolutionized the whole world of electronics—communications, control, data processing, and consumer electronics.

One of the major achievements of modern physics has been the success of solid-state physics in creating new technologies. Solid-state physics, which involves experimental investigation as well as theoretical understanding of the physical properties of solids, constitutes, by a substantial margin, the largest branch of physics; probably a quarter of the total number of physicists in the world belong to this branch. Semiconductor physics, one of the most important sub-fields of solid-state physics, covers electrical, optical, and thermal properties and interactions with all forms of radiation in semiconductors. Many of these have been of interest since the 19th century, partly because of their practical applications and partly because of the richness of intriguing phenomena that semiconductor materials present.

Point-contact rectifiers made of a variety of natural crystals found practical applications as detectors of high-frequency signals in radio telegraphy in the early part of this century. The natural crystals employed were lead sulphide (galena), ferrous sulphide, silicon carbide, etc. Plate rectifiers made of cuprous oxide or selenium were developed for handling large power [1].[1] The selenium photocell was also found useful in the measurement of light intensity because of its photo-sensitivity.

In the late 1920's and during the 1930's, the new technique of quantum mechanics was applied to develop electronic energy band structure [2] and a modern picture of the elementary excitations of semiconductors. Of course, this modern study has its roots in the discovery of x-ray diffraction by von Laue in 1912, which provided quantitative information on the arrangements of atoms in semiconductor crystals. Within this framework, attempts were made to obtain a better understanding of semiconductor materials and quantitative or semiquantitative interpretation of their transport and optical properties, such as rectification, photoconductivity, electrical breakdown, etc.

During this course of investigation on semiconductors, it was recognized in the 1930's that the phenomena of semiconductors should be analyzed in terms of two separate parts: surface phenomena and bulk effects. Rectification and photo-voltage appeared to be surface or interface phenomena, while ohmic electrical resistance with a negative temperature coefficient and ohmic photocurrent appeared to belong to bulk effects in homogeneous semiconductor materials. The depletion of carriers near the surface primarily arises from the existence of surface states which trap electrons and, also from relatively long screening lengths in semiconductors because of much lower carrier concentra-

---

[1] Figures in brackets indicate literature references at the end of this paper.

tions than in metals. Thus, it is possible to create potential barriers for carriers on the semiconductor surface or at the interface between a semiconductor and a metal contact, or between two semiconductors. The early recognition of the importance of surface physics was one of the significant aspects in semiconductor physics.

## 2. Transistors

Since the rectification in semiconductor diodes is analogous to that obtained in a vacuum diode tube, a number of attempts had been made to build a solid-state triode by inserting a "grid" into semiconductors or ironic crystals—a solid-state analog of the triode tube amplifier [3,4,5,]. Because of the relatively low density of carriers in semiconductors, Shockley thought that the control of the density of carriers near the semiconductor surface should be possible by means of an externally applied electric field between the surface and a metal electrode insulated from the surface—the field effect device. The observed effect, however, was much less than predicted [6]. In 1947, in the course of trying to make a good "field effect" device with two gold contacts less than fifty microns apart on the germanium surface, Bardeen and Brattain made the first point-contact transistor where they discovered a phenomenon—minority carrier injection into a semiconductor [7]. The importance of this phenomenon was soon recognized and led to the invention of the junction transistor by Shockley. The realization of this junction device, which did not occur until 1950 [8], was far more significant than its precursor.

The early version of the junction transistor was presented by Shockley as a post-deadline paper at the Reading Conference on "Semi-Conducting Materials," held July 10 to 15, 1950. This conference is now called "The First International Conference on the Physics of Semiconductors." Shockley's paper, however, was omitted from its Proceedings, apparently because a fabrication method used for the junction transistor was then proprietary at the Bell Laboratories [9].

The Foreword of the Proceedings of the Reading Conference [10] states: "During recent years physicists in many countries have made rapid and important advances in the field of solid state physics. Semiconducting materials, in particular, have become a subject of great interest by reason of their numerous practical application. . . ." Indeed, the development of transistors, as well as the progress in semiconductor physics of Ge and Si, would not have been accomplished without the key contribution of materials preparation techniques. Soon after Teal and Little prepared large Ge single crystals, Sparks successfully made a grown junction transistor at Bell Laboratories [11]. The subsequent development was Pfann's zone refining and

then Theuerer's floating zone method for silicon processing. These developments made it possible to make Ge and Si of controlled purities and crystal perfection.

The early Ge junction transistors had poor frequency response and relatively low reliability. In fabricating these transistors, the grown-junction technique, or the alloying technique, was used to form p-n junctions; in other words, these techniques were used to control the spacial distribution of donors and acceptors in semiconductors. Then a procedure for forming p-n junctions by thermal diffusion of impurities was explored in order to obtain better reproducibility and tighter dimensional tolerances. This technique, indeed, enabled bringing forth the double diffused transistor with desirable impurity distribution, the prototype of the contemporary transistor [12]. Attention was also turned toward Si because of its expected high reliability and improved temperature capability.

In the 1940's, a team at the Bell Laboratories selected elemental semiconductors, Ge and Si, for their solid-state amplifier project, primarily because of the possible simplicity in understanding and material preparation, in comparison with oxide or compound semiconductors. This not only was a foresighted selection but also had important implications: Ge and Si single crystals exhibited long diffusion lengths of hundreds of microns at room temperature, which were prerequisites to the desirable operation of the transistor, because of both reasonably high mobilities of electrons and holes, and long trapless lifetimes of minority carriers. The latter fact may arise from the indirect energy-gap in these elemental semiconductors in contrast with the direct energy-gap in some III-V compound semiconductors which exhibit high rates of radiative combination of electrons and holes. The exploration of the III-V compound semiconductors was initiated through Welker's ingenuity and imagination, in the early 1950's, to produce semiconductor materials even more desirable for transistors than Ge or Si [13]. Although this initial expectation was not quite met, III-V compound semiconductors later found their most important applications in LED (light emitting diodes), injection lasers, Gunn microwave devices, etc.; these devices could not have been achieved through elemental semiconductors.

## 3. Important devices

Now, in order to reach a perspective in semiconductor device development, it may be worthwhile to comment on some selected semiconductor devices in chronological order.

*1) Solar Cells.* In 1940, Ohl observed a photovoltage as high as 0.5V by flashlight illumination in "naturally" grown Si p-n junctions [14]. The modern Si solar cell,

however, was created by bringing together seemingly unrelated activities, namely, large area p-n junctions by Fuller's diffusion method, Pearson's effort for power rectifiers, and Chapin's search for power sources for communication systems in remote locations. According to Smits' article [15], Pearson's diode showed "a conversion efficiency from solar energy to electrical energy of 4 percent. Low as this efficiency may seem today, in 1953, it was very exciting, improving on selenium by a factor of five." Development and production of solar cells were stimulated by the needs of the space program.

In 1972, heterojunction solar cells consisting of p $Ga_{1-x}Al_xAs$-p GaAs-n GaAs, exhibiting power conversion efficiency of 16-20 percent, were reported by Woodall and Hovel [16]. The improved efficiencies were attributed to the presence of the heavily-doped $Ga_{1-x}Al_xAs$ layer, which reduced both series resistance and surface recombination losses. The recent advent of the energy crisis has generated a renewed interest in research and development of solar cells which might be economically viable for terrestrial applications.

*2) Tunnel Diodes.* Interest in the tunneling effect goes back to the early years of quantum mechanics. Phenomena such as rectification, contact resistance, etc. in solids, were explained by tunneling in the early 1930's. However, since theories and experiments often gave conflicting results, not much progress was made at that time. Around 1950, semiconductor p-n junctions generated a renewed interest in the tunneling process. Experiments to observe this process in the reverse breakdown of the junctions, however, were again inconclusive.

In 1957, Esaki demonstrated convincing experimental evidence for tunneling in his heavily-doped (narrow) p-n junction—the tunnel diode [17, 18]. This diode found use in microwave applications because of its differential negative resistance being responsive to high frequencies. The discovery of the tunnel diode not only generated an interest in heavily-doped semiconductors but also helped to open a new research field on tunneling in semiconductors as well as in superconductors.

*3) Integrated Circuits.* In 1958, Kilby initiated the fabrication of a circuit which included a number of transistors, diodes, resistors, and capacitors, all residing on one semiconductor chip [19]. This structure is called the (monolithic) integrated circuit. Around the same time, Noyce and Moore introduced improved fabrication techniques called the "planar" process which enabled the birth of the first modern transistor—a landmark in semiconductor history. It was soon realized that this transistor with dished junctions (extending to the surface) and oxide passivation (protecting the junctions), was most suited for assembling integrated circuits, because metal stripes evaporated over the surface oxide layer could be readily used for interconnection [20].

Integrated circuits of digital as well as linear types have had one of the largest impacts on electronics; they are now the main building block in computers, instrumentation, control systems, and consumer products. According to a recent analysis by Moore [21], their complexity has almost doubled each year, now approaching one hundred thousand components on a single Si chip of, say, a quarter centimeter square, and yet the cost per function has decreased several thousandfold since their introduction at the beginning of the 1960's. Meanwhile system performance and reliability have been tremendously improved.

*4) MOS FET (Metal Oxide Semiconductor Field Effect Transistor Devices.* As mentioned earlier, the transistor was invented while searching for a field-effect device. The field-effect concept originated as early as the 1920's, but no successful device was made in spite of a number of attempts because of the lack of adequate technology.

Thermally-grown $SiO_2$ on Si single crystal surfaces, which was originally developed for oxide passivation of junctions in the later 1950's, was found to be a most suitable insulator for a field effect device by Kahng and Atalla [22]. This insulator, indeed, had relatively low loss and high dielectric strength, enabling the application of high gate field. More importantly, the density of surface states at the Si-$SiO_2$ interface was kept so low that the band bending in Si near the interface was readily controllable with externally applied gate fields. Thus, a simple, yet most practical, Si MOS transistor was created whereby the surface inversion layer conductance ("channel") was modulated by gate voltages. This transistor is called a unipolar device because of no minority carrier involvement; it requires fewer processes in fabrication than the bipolar transistor because of its structual two-dimensionality, and is especially adaptable for large-scale integrated circuits.

Presently, integrated circuits, consisting of MOS FET or MOS based components such as dynamic memory cells [23], charge-coupled devices [24], MNOS (Metal Nitride Oxide Semiconductor) memory cells, etc., are even more extensively used than bipolar transistors, in computer memories, microprocessors, calculators, digital watches, etc., while being challenged by advances in bipolar-based devices such as $I^2L$ (Integrated Injection Logic). As the size of individual FETs has continued to decrease for large integrated circuits with the application of advanced processing techniques, the "channel" distance is shortened to one micron or even less and the oxide thickness is thinned to a few hundred angstroms. If one pushed this to the extreme, new physical problems arise from excessively high fields across thin oxide films as well as in the "channel" direction. There has been some discussion on physical limits in digital electronics [25, 26].

While taking measurements of Si surface transport properties at low temperatures, Fang and Howard discovered

that electrons in the "channel" were two-dimensional [27], which provided a unique opportunity for studying quantum effects [28].

*5) Injection lasers.* Since the early part of this century, the phenomenon of light emission from SiC diodes was recognized and studied, although a practical light emitting diode had not materialized until the development of efficient p-n junctions made of III-V compound semiconductors [29]. Apparently, reports of high-efficiency radiation in GaAs stimulated a few groups to engage in a serious experimental effort to find lasing action in semiconductors: These possibilities were previously discussed [30, 31]. In 1962, the announcement of the successful achievement of lasing action in GaAs came on the same date, independently, from two groups: Hall et al. at General Electric; and Nathan et al. at IBM; and a month later from Quist et al. at Lincoln Laboratory [32]. All of them observed a pulse coherent radiation of 8400Å from liquid nitrogen-cooled, forward-bias GaAs p-n junctions. This occurrence is not suprising in the present competitive environment of the technical community where new scientific information is rapidly disseminated and digested, and new ideas are quickly implemented. There was a two-year interval between the first reports of the Ruby and He-Ne lasers and the announcement of the injection laser.

The performance of the device was improved with incorporation of heterojunctions by Alferov et al [33, 34]. With double-heterostructure the threshold current density for lasing was substantially reduced by confinement of both carriers and photons between two heterojunctions [35]. Finally, in 1970, Hayashi et al. [36] succeeded in operating the device continuously at room temperature. Because of the compactness and the high efficiency of this laser, the achievement paved the way towards many practical applications such as optical (light-wave) communication, signal processing, display and printing. There is a development effort in integrated optics to mount miniaturized optical components, including injection lasers and waveguides, on a common substrate using heterojunction structures of III-V compound semiconductors, analogous to the integrated circuit, for improved signal processing.

*6) Gunn and IMPATT (Impact Ionization Avalanche Transit Time) Microwave Devices*—In 1962, Gunn discovered that, when the applied field across a short bar of reasonably pure n-type GaAs exceeded a threshold voltage of several thousand volts per cm, coherent microwave oscillations could be extracted by synchronizing the random current fluctuations with a resonator [37]. Furthermore, by his ingenious probe technique, he was able to show that the oscillations were related to the periodic formation and propagation of a narrow region of very high field—"domain." It took two years to confirm that Gunn's experimental discovery of oscillations was indeed due to the Ridley-Watkins-Hilsum transferred electron effect, proposed in 1961 and 1962 [38, 39].

As is true of any important discovery, Gunn's work triggered a wide spectrum of experimental and theoretical activity from device physics to microwave engineering. Apparently this achievement rejuvenated the work of microwave semiconductor devices in general, and, in 1964, IMPATT diodes finally started to oscillate—which was rather overdue since Read's proposal in 1958 [40]. The operation of the device was explained on the basis of dynamics of electrons involving the transit time and avalanche. IMPATT and Gunn devices are now widely used in many microwave gears: the former has high power capabilities ($\sim$50mW at 110 GHz), whereas the latter meets low noise requirements.

## 4. Summary

Figure 1 schematically illustrates the development path of a variety of semiconductor devices. It should be noted that the development path of each device appears to have had its own sequence of conception (theory) and observation (experiment): Typically, the theoretical prediction was later confirmed by the experiment, but, in may instances, the experimental discovery came first, followed by the theory and yet, in other instances, the initial idea which led to the discovery was irrelvant to its consequence. Obviously, this article cannot possibly cover all landmarks and indispensable innovations, not to mention a great number of wonderful, but nonworkable ideas. We will summarize some important items in semiconductor devices and processing techniques which have made remarkable progress since 1950: Si devices of new structures, such as controlled rectifiers, solar cells, photodetectors, $I^2L$, etc.; development of novel semiconductor devices, such as injection lasers, Gunn microwave oscillators, Schottky junction FETs, infrared detectors, etc., combined with the investigation of new materials, such as III-V compounds; and the introduction of new processing techniques for device fabrication, such as alloy contacts, etching, thermal diffusion of impurities, vapor and liquid-phase epitaxy, oxide formation, sputtering, photolithography and, more recently, ion etching, molecular beam epitaxy, etc.

Not all of this progress arises from engineering ingenuity and advanced material technology; pioneering research in semiconductor physics has also contributed to each significant development, exploring intriguing phenomena in semiconductors, for example, electron-hole multiplication (avalanche), tunneling, hot electrons, lasing by high carrier injection, two-dimensional electrons on the surface or in a semiconductor superlattice, etc. For a qualitative understanding, semiconductor materials, crystalline or amor-
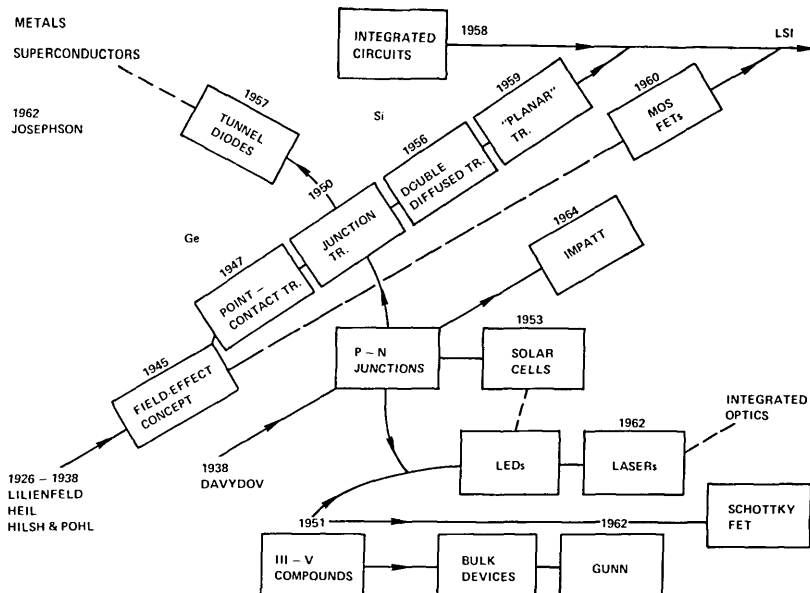
FIGURE 1. Schematic illustration of the development path of a variety of semiconductor devices.

phous, as well as surfaces, have been extensively investigated—often under extreme conditions with advanced instruments; measurements at high pressure or in ultrahigh vacuum, or under the synchrotron radiation fall into this category.

Semiconductor physics has a strong interaction with chemistry, metallurgy, and electrical engineering, and with the broader field of materials science. The separation between basic discoveries and applications in this field of physics is far less distinct than that in some of the other fields of physics. Semiconductor physics has a particularly effective interface with engineering.

"Science is the understanding of nature, whereas engineering is the control of nature." Following this notion, industrial laboratories appear to have played a dominant role as a junction between science and engineering in many technological developments, wherein there may possibly be a kind of gap between them. (Hopefully, this junction will always be forward-biased so that electrons and holes, carrying information, can flow easily from science to engineering and vice versa.) In the field of semiconductors, one may think that the coupling between science and engineering is strong, or that the gap between them is indeed narrow. After all, the semiconductor is a narrow-gap insulator!!

In the preparation of this manuscript, I am indebted to many authors who kindly sent me advance copies of papers which have appeared in the Special Issue of the IEEE Transactions on Electron Devices, July 1976.

## 5. References

Note: IEEE Trans. Electron Devices, Vol. ED–23 abbreviated herein by **ED-23.**

[1] Henisch, H. K. *Rectifying Semiconductor Contacts,* Clarendon Press, Oxford 1957.
[2] Wilson, A. H. Proc. Roy. Soc. A, **133,** 458 (1931).
[3] Lilienfeld, J.E. U.S. Patents 1,745,175 (1926); 1,877,140 (1928); 1,900,018 (1928).
[4] Heil, O. British Patent 439457 (1935).
[5] Hilsh, R. and Pohl, R.W., Zeits. f. Phys. **111,** 399 (1938).
[6] Shockley, W. and Pearson, G. L., Phys. Rev. **74,** 233 (1948).
[7] Bardeen, J. and Brattain, W. H., Phys. Rev. **74,** 230 (1948); Phys. Rev. **74,** 231 (1948); Phys. Rev. **75,** 1208 (1949).
[8] Shockley, **ED-23,** 597 (1976).
[9] Shockley, W., NBS Special Pub. 388, Proc. of Conf. on the Public Need and the Role of the Invention, Montery, CA, p. 47, June 1973.
[10] Proc. of Conf. Semi-Conducting Materials, Reading, 1950, published by Butterworths Scientific Pub. Ltd., London, 1951.
[11] Teal, G.K., **ED-23,** 621 (1976).
[12] Tanenbaum, M. and Thomas, D. E., B.S.T.J. **35,** 1 (1956).
[13] Welker, H.J., **ED-23,** 664 (1976).
[14] Ohl, R. S. U.S. Patent 2,402,662 (1941).
[15] Smits, F. M., **ED-23,** 640 (1976).
[16] Woodall, J.M. and Kovel, H.J., Appl. Phys. Lett. **21,** 379 (1972).
[17] Esaki, L. Nobel Lecture, Dec. 11, 1973, Pub. by Les Priz Nobel, p. 66 (1974).
[18] Esaki, L., **ED-23,** 644 (1976).
[19] Kilby, J. S., **ED-23,** 648 (1976); U.S. Patent 3,138,743 (1959).
[20] Noyce, R. N., U.S. Patent 2,981,877 (1959).
[21] Moore, G. E., Intn'l Electron Devices Meeting, Tech. Digest, Washington, DC, 1975, p. 11.
[22] Kahng, D., **ED-23,** 655 (1976).
[23] Dennard, R. H., U.S. Patent 3,387,286 (1968).

[24] Boyle W. S. and Smith, G. E., **ED-23**, 661 (1976).

[25] Keyes, R. W., Proc. IEEE **63**, 740 (1975).

[26] Wallmark, J.T., *Microelectronics*, E Keonjian, Ed., McGraw Hill, NY, p. 10 (1963).

[27] Fang F. F. and Howard W. E., Phys. Rev. Lett. **16**, 797 (1966).

[28] Fowler, A. B., Fang, F. F., Howard, W. E., and Stiles, P. J., Phys. Rev. Lett. **16**, 901 (1966).

[29] Loebner, E. E.,**ED-23**, 675 (1976).

[30] Bernard, M.G.A. and Duraffourg, G., Physica Status Solidi **1**, 699 (1961).

[31] Dumke. W.P., Phys. Rev. **127**, 1559 (1952).

[32] Hall, R. N., **ED-23**, 700 (1976).

[33] Alferov, Zh. I., Andreev, V. M., Korol'kov, V. I., Portnoi, E. L., and Tret'yakov, D. N., Fiz. Tekh. Poluprov. **2**, 1545 (1968).

[34] Alferov, Zh.I., Andreev, V. M., Portnoi. E. L., and Trukan, M. K., Fiz. Tekh. Poluprov. **3**, 1328 (1969).

[35] Kressel, H. and Hawrylo, F. Z., Appl. Phys. Lett. **17**, 169 (1970).

[36] Hayashi, I., Panish, M. B., Foy, P. W., and Sumski, S., Appl. Phys. Lett. **17**, 109 (1970).

[37] Gunn, J. B., **ED-23**, 705 (1976).

[38] Ridley, B. K. and Watkins, T. B., Proc Phys. Soc. (London) **74**, 293 (1961).

[39] Hilsum, C., Proc. IRE **50**, 185 (1962).

[40] DeLoach, Jr., B. C., **ED-23**, 657 (1976).