IEEE HOME | SEARCH IEEE | SHOP | WEB ACCOUNT | CONTACT IEEE
Membership Publications Services Standards Conferences Careers/Jobs
IEEE SPECTRUM ONLINE
Member Table of Contents Guest Table of Contents Job Site Editorial Staff Advertising Direct Mail Lists
Tue May 10 2005 21:17:57 GMT-0700 (Pacific Daylight Time) Home >> Feature Story
Search
SITE INDEX
FEATURE ARTICLE
WEEKLY FEATURE

Robert Noyce and the Tunnel Diode

A 50-year-old notebook reveals the seed of a great invention

By Leslie Berlin & H. Craig Casey Jr.

I have in my notebooks from [1956] a complete description of the tunnel diode," the speaker told the audience at a symposium on innovation at the MIT Club of New York, in New York City, in December 1976. It was quite a revelation, because the speaker wasn't Leo Esaki, who had won the 1973 Nobel Prize in physics for inventing the tunnel diode in the late 1950s. It was Robert N. Noyce, cofounder of Intel Corp., Santa Clara, Calif.; inventor of the first practical integrated circuit; and a man who, as far as anyone knew before that speech, had no connection to the most storied electronic device never to be manufactured in large numbers.

Engineers coveted the tunnel diode for its extremely fast switching times—tens of picoseconds—at a time when transistors loped along at milliseconds. But it never found commercial success, though it was occasionally used as a very fast switch. As a two-terminal device, the diode could not readily be designed for amplification, unlike a three-terminal transistor, whose circuit applications were then growing astronomically. Nevertheless, the tunnel diode was a seminal invention. It provided the first physical evidence that the phenomenon of tunneling, a key postulate of quantum mechanics, was more than an intriguing theory.

Quantum mechanics, the foundation of modern physics, is an elaborate conceptual framework that predicts the behavior of matter and radiation at the atomic level. One of its most fundamental notions is that the exchange of energy at the subatomic level is constrained to certain levels, or quantities—in a word, quantized.

Many of the core concepts and phenomena of quantum mechanics are almost completely counterintuitive. For example, consider a piece of semiconductor joined to an insulator. From the point of view of classical physics theory, the electrons in the semiconductor are like rubber balls, and the insulator is like a low garden wall. An electron would have no chance of getting over the barrier unless its energy were higher than the barrier's. But according to quantum mechanics, the phenomenon of tunneling ensures that for certain conditions an electron with less energy than the barrier's will not bounce off the wall but will instead tunnel right through it.

Ever since the late 1920s, physicists had debated about whether tunneling really occurred in solids. The tunnel diode offered the first compelling experimental evidence that it did.

When Esaki, then a 49-year-old semiconductor research scientist at IBM Corp., won his Nobel Prize in 1973, neither he nor the Nobel committee had any idea about Noyce's work. Esaki had made a tunnel diode and measured its current versus voltage behavior 16 years earlier, when he was working at the company now called Sony Corp. in his native Japan. The Nobel committee, in fact, dated Esaki's discovery from 1957, roughly contemporaneous with Noyce's recollected work in the same field. Stig Lundqvist of the Swedish Royal Academy of Sciences used the "electrons as balls against the wall" analogy in his speech presenting the 1973 Nobel Prize in physics to Esaki; Ivar Giaever and Brian David Josephson shared the award for discovering different aspects of the tunneling phenomenon in solids.

ALMOST EVERY IMPORTANT DISCOVERY since the start of the industrial age has a contested history. Heinrich Gobel, from a town near Hanover in Germany, filed suit in 1893 claiming that he, not Thomas Edison, had invented the light bulb years earlier in New York City. Something similar has occurred for the airplane, telephone, rotor encryption machine, television, integrated circuit, and microprocessor, to name but a few. Such counterclaims often have merit—invention and research are often group activities, and discoveries regularly appear in different places at almost precisely the same time. And sometimes such claims come from experimenting hacks eager for a measure of recognition for themselves.

Noyce was no hack, obviously—his integrated circuit nestles at the heart of essentially every piece of modern electronics. In fact, the invention of the IC was recognized as a Nobel-level achievement in 2000, when the prize for physics was awarded to Jack S. Kilby, credited by U.S. courts as the coinventor of the IC. Unfortunately for Noyce, he missed his chance to join the pantheon of laureates when he died in 1990; the prizes are not awarded posthumously.

Nor was Noyce pursuing glory when he mentioned his work in his talk at that symposium in 1976. In fact, immediately after claiming to have the invention in his notebooks, Noyce said, "The work had been done elsewhere [by Leo Esaki] and was published shortly thereafter." He had mentioned it in the first place only because he thought the way his boss had handled Noyce's tunnel diode efforts in 1956 "may be instructive in how not to motivate people."

NOYCE'S BOSS AT THAT TIME was William B. Shockley, the brilliant, mercurial, ambitious, autocratic, and eccentric physicist. He was the sort of man who thought nothing of publicly subjecting his employees to lie-detector tests. As the young Noyce had his insight about the tunnel diode, Shockley himself was only weeks away from his own Nobel Prize in physics, awarded for his 1947 invention, along with two colleagues, John Bardeen and Walter Brattain, of the transistor.

Shockley had started Shockley Semiconductor Laboratory in 1955 with the self-proclaimed goal of making a million dollars and seeing his name in *The Wall Street Journal*. Noyce headed the transistor group at Shockley Lab. With a Ph.D. in physical electronics and two years in a transistor research lab at Philco Corp., he was the most experienced semiconductor researcher among Shockley's several dozen employees.

On 14 August 1956, Noyce noted an idea for a "negative resistance diode" in his lab notebook. With most diodes, current increases with increased voltage—the more voltage applied to the device, the more current passes through it.

In a diode, the current under a forward voltage, or bias, is relatively large, while little current results when the bias is reversed. For a semiconductor diode, such behavior is obtained by adding impurity atoms. Esaki's semiconductor was germanium. He used two types of impurities. So-called donor atoms have more electrons in their outer orbits than do the outer orbits of germanium atoms. The excess electrons become free electrons, available for conduction. A semiconductor with an excess of electrons is called *n*-type.

Similarly, if the germanium is doped with impurity atoms that hold fewer electrons in their outer orbits than germanium, the impurity atoms will take away, or accept, electrons from the semiconductor atoms, leaving behind deficiencies of electrons, known as holes. A semiconductor with an excess of holes, each one considered to have a positive charge, is called a *p*-type semiconductor.

Germanium can be doped to create *p*- and *n*-type sections that butt against each other and form what is called a p-n junction. In a *p-n* junction, a potential difference normally builds up across a narrow region near where the *p*-type and *n*-type semiconductors come into contact. This built-in potential sets up a barrier against the passage of holes into the *n*-type material and the passage of electrons into the *p*-type. Applying an external bias across the diode changes the barrier's height. A "forward bias"—obtained by connecting a battery's positive terminal to the *p* side and its negative terminal to the *n* side—lowers the barrier, allowing electrons to flow easily from the *n* side to the *p* side. Reverse the polarity, and the height of the barrier rises, prohibiting the flow of electrons.

Noyce, however, made a startling prediction. First he proposed the existence of a semiconductor whose regions of opposite polarity were each doped with roughly a thousand times more impurities than was usual at the time. When a forward bias increasing from zero was applied to such a heavily doped diode (which Noyce called "degenerate"), he predicted that current would initially increase at a greater rate than for a normal diode. This phenomenon would occur because the high impurity density would, in effect, make it possible for the balls (electrons) to tunnel through the wall (the junction's potential barrier). At some point, increasing the voltage further would decrease the tunneling current, but at still higher voltages, the current would increase because of the nontunneling diode current.

Noyce discussed his ideas with his friend Gordon E. Moore, a chemist who had joined Shockley Lab a day before Noyce. He then brought his notebook to Shockley, fully expecting him to be impressed. Instead, "the boss showed no interest in the idea," Noyce said. The lab was not equipped to do anything profitable with Noyce's thoughts, and besides, Shockley was a fiercely competitive man who resented his employees' pursuing ideas that he had not personally placed on their research agendas. Disappointed, Noyce closed his lab book and "went on to other projects" more in line with Shockley's wishes.

UNTIL NOW, NO ONE OTHER THAN MOORE and Shockley had seen Robert Noyce's 1956 description of a tunnel diode. But Noyce copied his work and saved it. How he managed to copy these pages is unclear—photocopy technology was in its infancy in the late 1950s, and Noyce never made note of going back to his Shockley notebooks later in life—but that the pages are legitimate is indisputable. Leslie Berlin, one of this article's authors, found them in January 2001 tucked in one of Noyce's Fairchild notebooks stored in Santa Clara, Calif., at a company that prefers not to be identified.

Berlin compared these copied pages to the only surviving notebook from Shockley Lab: the book belonging to William Shockley housed in the Special Collections of Stanford University, in California. The pages on which Noyce's ideas are written are clearly from the same type of lab book that Shockley issued to his staff, and the handwriting is undoubtedly Noyce's. This, along with the date of Noyce's work (which correlates with his 1976 comments about it), and Moore's recollections of the event, further validate their authenticity.

A quick comparison of Noyce's notebook pages with Esaki's seminal paper, "New Phenomenon in Narrow Germanium *p-n* Junctions," published in *Physical Review* in January 1958 (and received by that journal in October 1957), shows striking parallels. Both men used an energy-band diagram that represents the electron and hole energies on the *y* (vertical) axis versus their position in the p-*n* junction on the *x* (horizontal) axis [see "The Noyce Diode," two pages from Noyce's notebook].

Noyce's "energy-level" diagram, which is now called an energy-band diagram [on the left-hand page], shows where the electrons and holes are located. It also illustrates the conditions necessary for tunneling current. The upper solid line in the diagram represents the bottom of the semiconductor's conduction band; in this band electrons can move freely as a result of the donor atoms. The lower solid line represents the top of the valence band, where acceptor impurities allow holes to move freely. The separation between the conduction and valence bands is the energy gap, or E_g, and is the range in energy where no electrons or holes are permitted. For this reason, E_g is sometimes called the forbidden gap.

In Noyce's diagram, the Fermi energy, or E_f, represents the energy boundary for most of the holes in the *p*-type semiconductor and most of the free electrons in the *n*-type. For a highly doped, or degenerate, semiconductor, E_f falls below the edge of the valence band and rises above the edge of the conduction band. Electrons "sink" so they fill the lowest energy levels in the conduction band, while holes "float" and fill the highest levels of the valence band. Therefore, it is the holes between the top of the valence band and E_f and the free electrons between E_f and the bottom of the conduction band that are significant for tunneling.

IN GOOD COMPANY: The eight engineers and scientists, including Robert N. Noyce [right] and Gordon E. Moore [standing second from left], who cofounded Fairchild Semiconductor in 1957, are pictured here on the firm's production floor in its early years.

The region between the *p* and the *n* sides where the valence and conduction band edges bend is called the depletion region; this is where the potential barrier exists. This region narrows for large donor and acceptor concentrations and would be less than 10 nanometers for a tunnel diode.

Note that without an applied bias, the holes on the p side are at a higher energy than the electrons on the *n* side. For tunneling to occur, there must be holes at the same energy as the free electrons. But a forward bias (a positive voltage connected to the *p* side), raises E_f and the conduction-band electrons on the n side with respect to E_f on the *p* side by the amount of the bias voltage. Now there are free electrons at the same energy as the holes, and the electrons can tunnel through the potential barrier to holes on the *p* side, resulting in a current. As the forward bias is increased, more free electrons and holes are at the same energy and the tunneling current increases.

Both Noyce and Esaki recognized that as the bias increased further, E_f on the *n* side would be raised further with respect to E_f on the *p* side and the concentration of free electrons at the same energy as holes would diminish and result in a reduced tunneling current, as shown in Noyce's current (*I*) vs. voltage (*V*) plot. At a larger bias, the normal diode current would flow at the voltage E_g in Noyce's plot.

This plot [on the right-hand page] is very similar to the measured *I-V* plot that Esaki shows. This phenomenon of decreasing current with increasing voltage is negative resistance, a characteristic that has been exploited to build oscillators.

ut there was one important difference between Noyce's and Esaki's work. Noyce only

For more on William Shockley, see *Crystal Fire: The Birth of the Information Age,* by Michael Riordan and Lillian Hoddeson (W.W. Norton, 1997).

PHOTOS: INTEL CORP.

URL: http://www.spectrum.ieee.org (Modified: 30 April 2005)