difficult to connect.

Hoerni's approach yielded transistors that were buried beneath a layer of silicon dioxide. From the work of John Moll and Carl Frosch at Bell Labs, the merits of SiO_2 were already beginning to be appreciated as a protective coating, as a diffusion mask for controlling the boundaries of the doping process, and as a good natural insulator that has few inherent charge traps and interface states. Largely because of the desirable properties of its oxide, Si was starting to replace Ge as the semiconductor of choice in electronics (see PHYSICS TODAY, October 1997, page 50).

As Noyce realized, the flat transistors resulting from Hoerni's planar process could be interconnected by depositing metal over etched holes in the oxide. This fabrication process is still used today, "although it's been improved almost beyond recognition," comments Kilby. Noyce, who died in 1990, and Kilby are generally regarded as coinventors of the IC.

Gordon Moore, who left Fairchild with Noyce in 1968 to found Intel, noted in 1965 that the complexity of ICs was doubling roughly every year. With a slight revision in time constant, to about 18 months, this exponential growth of semiconductor devices—known as Moore's law—is still going strong, driven by increased control over processing, the quality of materials, and especially the refinements in lithography that allow reduced feature size and higher packing densities. "The modern IC is a result of the efforts of tens or hundreds of thousands of the best engineers," notes Kilby, "with almost a continuous series of advances."

Mixing and matching

Doping a semiconductor like Si or GaAs provides control over the sign and density of the charge carriers. By combining different semiconductors in heterostructures, one gains control over much more, including the band gap energy, refractive index, carrier mass and mobility, and other fundamental parameters.

Not long after the development of the transistor, ideas for controlling the holes and electrons in them started to emerge—in particular, what's come to be known as bandgap engineering. In his 1948 transistor patent application, Shockley mentioned the idea of using a wider-gap semiconductor for the emitter than for the rest of the transistor, a configuration analyzed by Aleksandr Gubanov at Ioffe in 1951.

Lights out at LEP

As we were going to press, the CERN directorate announced that the 2 November shutdown of LEP was indeed the e^+e^- collider's final termination. The 27-km LEP tunnel will now be made ready for the Large Hadron Collider, which should be providing 14-TeV proton–proton collisions by 2005. LEP had already enjoyed a one-month reprieve in response to recent tantalizing hints of a possible Higgs boson with a mass of about 115 GeV/c^2. But, after heated discussion, it was finally decided that the extra October run did not uncover sufficient confirmatory evidence to justify a costly further extension of LEP's 11-year life.

BERTRAM SCHWARZSCHILD

In 1957, Kroemer, then at RCA in Princeton, New Jersey, returned to an idea he had had a few years earlier: varying the energy gap in semiconductors.[2] He realized that a spatially varying energy gap, and the resulting variations in the semiconductor's conduction and valence bands, creates quasi-electric fields that act differently on electrons in the conduction band and holes in the valence band. The resulting forces on electrons and holes could even act in the same direction.

Kroemer realized that this effect should make it possible to create devices that are fundamentally impossible to achieve in homostructures. "While the examples I could come up with—transistors with a graded bandgap in the base, or with a wide-gap emitter—represented improvements," he told us, "they did not fully exploit the new principle." This would change abruptly six years later with the idea of the double-heterostructure (DHS) laser.

Starting in the early 1960s, much attention in the semiconductor community was aimed at developing a laser. The indirect bandgap of Si and Ge made those elements unlikely candidates, so the focus was on III–V semiconductors such as GaAs. The goal was attained in 1962 by several groups: Robert Hall and coworkers at General Electric Co in Schenectady, New York, Marshall Nathan and colleagues at IBM Corp, and Robert Rediker and coworkers at MIT's Lincoln Laboratory demonstrated lasing in GaAs, and Nick Holonyak Jr and Sam Bevacqua at GE in Syracuse created a visible laser in the alloy GaAsP.

These early devices used forward-biased p–n homojunctions made of a single semiconductor with different dopants. Electrons from the n side and holes from the p side recombined in the junction region, giving off light. The lasers were not very efficient because of high optical and electrical losses. Thus the threshold currents were high and low temperatures were necessary.

On learning of the large losses, Alferov in the Soviet Union and Kroemer in the US—neither of whom had any prior background in lasers—independently and simultaneously proposed the solution: use a DHS. DHS lasers, with a low-bandgap material sandwiched between two wide-bandgap semiconductors, have two significant inherent advantages over homostructures. Because of the lower conduction band energy and higher valence band energy in the middle layer, the electrons and holes are trapped there. But lower bandgap materials also have higher indices of refraction. Thus, in addition to carrier confinement, the DHS also provides optical confinement of the emitted light to the lower-bandgap region. Alferov proposed DHS lasers in a 1963 Soviet patent application with Rudolf Kazarinov.[3] Kroemer published a similar suggestion in the US, and filed for a US patent just one week after Alferov.[4]

Alferov organized an effort at Ioffe to explore heterostructure applications, and many US labs continued work on semiconductor devices. However, Kroemer, working in industry at Varian Associates in Silicon Valley at that time, was refused support for working on DHS. "I was told that the new device could not possibly have any practical applications." He turned his attention to the Gunn effect, but in the early 1970s returned to heterostructures, focusing on heterojunction bipolar transistors (HBTs).

Although the ideas for DHS were there, the technology wasn't. The primary question was which materials to use, and the dominant consideration was matching the lattice constants of the semiconductors. GaAs–GaAlAs eventually emerged as the pairing of choice, once researchers correctly recognized the close values of their lattice constants and realized that, whereas AlAs will oxidize in air, the alloy $Ga_xAl_{1-x}As$ is chemically stable for low to moderate Al concentrations.

Heterostructure research picked up steam in 1967, when Alferov in the Soviet Union and Hans Rupprecht and Jerry Woodall at IBM independently reported the first GaAs–GaAlAs heterojunctions, grown by liquid-phase epitaxy. Whereas the US groups were working on single heterostructures, Alferov's group was committed to DHS.