Special 30th Anniversary

How the Double-Heterostructure Laser Idea got started

Herbert Kroemer

In August 1952, I was hired as “Resident Theorist” by the small semiconductor research group at the Central Telecommunications Technology Office [Fernmeldetechnisches Zentralamt (FTZ)] of the German Postal Service in Darmstadt, Germany. I had just received a doxtoral degree in Theoretical Solid State Physics from the University of Göttingen, with a dissertation on what we would today call hot-electron transport effects, of the kind that were thought to play a role in the collector of the then-new transistor. In the process, I had acquired what by 1952 standards was a good background in semiconductor physics, including the emerging device physics. It was an unusual background for a 1952 Theoretical Physicist, but it was perfect for what was to come.

The FTZ group was at that time working on the first bipolar junction transistors. These early devices were far too slow for practical applications in telecommunications, and I set myself the task of understanding the frequency limitations theoretically—and what to do about them. One approach was to speed up the flow of the minority carriers from the emitter to the collector by incorporating an electric field into the base region, by using a non-uniform doping in the base, which decreased exponentially from the emitter end to the collector end. While working out the details, I realized that

“... a drift field may also be generated through a variation of the energy gap itself, by making the base region from a non-stoichiometric mixed crystal of different semiconductors with different energy gaps (for example, Ge-Si), with a composition that varies continuously through the base.” ([1]; translated)

This was not yet a full general design principle, but a seed had been planted. Because of the absence of any credible technology, I did not follow up on this idea until 1957, after I had joined RCA Laboratories in Princeton, NJ. At that time, the 1954 seed had germinated: I had realized the generality of the design principle hinted at in the above sentence, and published a (widely ignored) paper in the RCA Review (never publish in obscure journals!), which included the figure shown below, and the accompanying paragraph of text [2]:

‘In a homogeneous semiconductor the band slope under an electric field is the same for all bands and, as a result, the forces upon electrons and holes are equal in magnitude and opposite in direction. This is not the case with a varying band gap. If the concept of a varying band gap is a legitimate one, the forces would no longer be equal and opposite. It should, for example, be possible to have force acting only upon one kind of the carriers, or to have a force which acts in the same direction for both (Fig. 1). Electrical forces in uniform crystals can never do this. This is why we call these forces “quasi-electric” forces. They present a new degree of freedom for the device designer which enables him to obtain effects with the quasi-electric forces that are basically impossible to obtain with ordinary circuit means involving only “real” electric fields.’ [Emphasis added]

ECE DEPARTMENT AND MATERIALS DEPARTMENT UNIVERSITY OF CALIFORNIA, SANTA BARBARA, CA 93106

As an example, I quoted the improved bipolar transistor. However, that still did not draw on the full power of the idea expressed in the general design principle that the quasi-electric fields ‘enable the device designer to obtain effects that are basically impossible to obtain using only “real” electric fields.’ It certainly represents major improvements, but does it represent something “basically impossible” otherwise?

Something that was indeed truly impossible to achieve otherwise emerged abruptly in March 1963. I was working at Varian Associates in Palo Alto at the time, where a colleague of mine—Dr. Sol Miller—had taken a strong interest in the first semiconductor junction lasers that had emerged in 1962, a topic then outside my own range of interests. In a colloquium on the topic he gave a beautiful review of what had been achieved, not failing to point out that successful laser action required either low temperatures or short low-duty-cycle pulses, usually both. Asked what the chances were to achieve con-

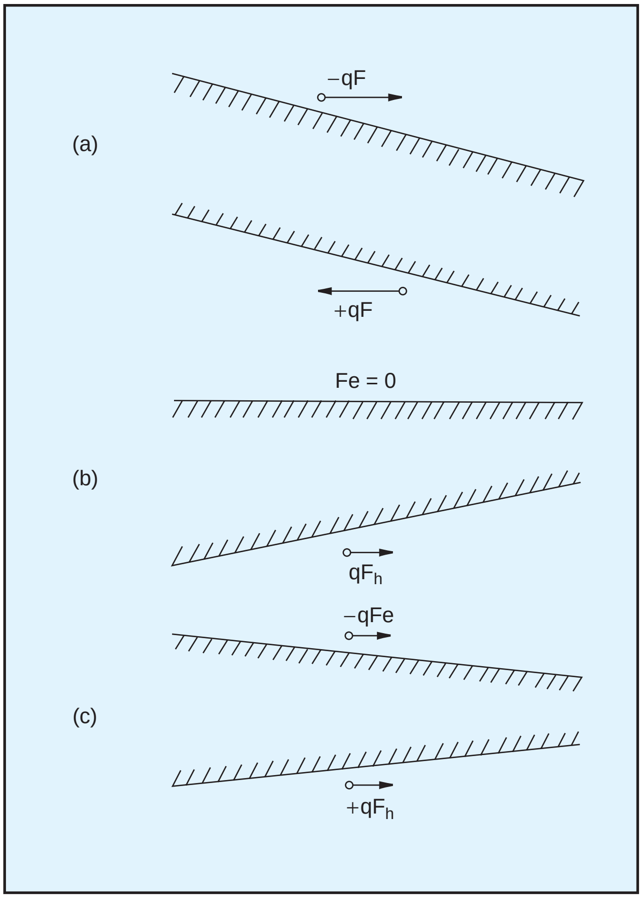

Figure. 1: (a) Effect of a true electric field; (b) and (c): Effects of quasi-electric fields.

tinuous operation at room temperature, Miller replied that certain experts had concluded that this was fundamentally impossible. The argument ran roughly as follows. In order to obtain laser action, a population inversion has to be achieved, which means that, in the active region, the occupation probability of the lowest states in the conduction band has to be higher than that of the highest states in the valence band. A necessary condition for such a population inversion is a forward bias larger than the energy gap. Even then, a population inversion is hard to achieve in an ordinary p-n junction. The carrier concentrations in the active region will always be lower than in the doped contact regions. Inversion, therefore, required degenerate doping on both sides. Even then, both the electrons and holes would diffuse out of the active region immediately into the adjacent oppositely doped region, preventing a population inversion from building up.

I immediately protested against this argument with words somewhat like “but that is a pile of .... , all one has to do is give the injector regions a wider energy gap.” To me, this would cause an electron-repelling quasi-electric field to be present on the p side, and a similar hole-repelling barrier on the n side. Carrier confinement would thus be achieved. In fact, electron and hole concentrations could be much larger than the doping levels in the contact regions (for details, see my Nobel Lecture [4]), and it would become readily possible to create the population inversion necessary for laser action. This double-heterostructure (DH) laser finally represented a device truly impossible with only the real electric fields available in homostructures. Note that the idea for it arose essentially at the instant I had been made aware that there was a problem.

The rest is history.

I wrote up a paper describing the DH idea, along with a patent application. The paper was submitted to Applied Physics Letters, where it was rejected. I was urged not to fight the rejection, but to submit the paper to the Proceedings of the IEEE instead, where it was published [3], but largely ignored. The patent was issued in 1967 [5]. It is probably a better paper than the Proc. IEEE letter. It expired in 1985. Both the IEEE letter and the patent draw on an extensive unpublished corporate report that might be of interest to readers wishing to go deeper into the history of the subject [6].

When I proposed to develop the technology for the DH laser, I was refused the resources to do so, on the grounds that “this device could not possibly have any practical applications,” or words to that effect. It was a classical case of judging a fundamentally new technology, not by what new applications it might create, but merely by what it might do for already-existing applications. This is extraordinarily shortsighted, but the problem is pervasive, as old as technology itself. The DH laser was simply another example in a long chain of similar examples. Nor will it be the last. Any detailed look at history provides staggering evidence for what I have called the Lemma of New Technology:

The principal applications of any sufficiently new and innovative technology have always been — and will continue to be — applications created by that technology.

As a rule, such applications have indeed arisen—the DH laser is just a good recent example—although usually not immediately.

References:

[1] H. Krömer, Archiv d. Elekt. Übertragung 8, 499 (1954).

[2] H. Kroemer, RCA Review 18, 332 (1957).

[3] —, Proc. IEEE 51, 1782 (1963).

[4] —, Revs. Mod. Phys. 73, 783 (2001)

[4] —, US patent 3,309,553 (filed Aug. 16, 1963, issued 1967).

[5] —, Varian Central Research Report CRR-36 (1963); unpublished (available from the author as PDF copy).

Biography: Herbert Kroemer

Herbert Kroemer is Professor of Electrical and Computer Engineering and of Materials at UCSB. He was born and educated in Germany. In 1952 he received a Doctorate in Physics from the University of Göttingen, Germany. Since then, he has worked on the physics and technology of semiconductors and semiconductor devices, especially heterostructures. He originated several key device concepts, including the heterostructure bipolar transistor and the double-heterostructure laser. He is a Member of the National Academy of Engineering and the National Academy of Sciences. He holds honorary doctorates from the Technical University of Aachen, Germany, from the University of Lund, Sweden, from the University of Colorado, and from the University of Duisburg-Essen, Germany. He has received numerous awards, most recently, in 2000, the Nobel Prize in Physics, “for developing semiconductor heterostructures used in high-speed and opto-electronics,” and in 2002 the IEEE Medal of Honor.