Extended Abstracts of the 1994 International Conference on Solid State Devices and Materials, Yokohama, 1994, pp. 397-399

Invited

Devices for the Future: A Peek into the Next Century

Herbert Kroemer

Department of Electrical and Computer Engineering,
University of California
Santa Barbara, CA 93106 USA

The forces that drive device development include both applications-driven and discovery-driven forces. Applications-driven forces have fairly predictable consequences. Discovery-driven forces are much harder to predict, but are likely to play an increasingly important role as the prediction time span becomes longer. Much of the importance of new discoveries will be in new applications *created* by the discoveries; and in the economic system leverage provided by high-performance devices even at small production volumes.

Much of the history of long-term technology forecasts has been a history of failures — the longer the forecast period, the larger the failure. As a rule, the actual *long-term* developments have exceeded any *rationally-based* forecasts, defined as forecasts based on actual knowledge in the field of the technology, rather than wild irrational science-fiction-type guesses. Experts tend to be too conservative!

This is of course not a reason to leave this hazardous business to the science fiction writers, many — not all! — of whom seem to be under few rational constraints, such as respect for established physical laws not likely to be found invalid. Instead, we must take a rational yet creative look at the *forces* that actually drive technological progress.

There are really two quite different forces involved, which I would like to call *applications-driven* and *discovery-driven* forces (often both act in combination).

Applications-driven forces are easy to understand. As an example, take the trend towards computer software capable of handling tasks of higher and higher levels of complexity. There is no foreseeable end to this trend. The rate of progress in this applications-driven direction is heavily dependent on the concurrent development of computer chips that are themselves of increasing levels of complexity. But increasing complexity inevitably means increasing density, which means smaller individual devices, with lower power consumption and more speed, etc. Hence, we have here a very strong applications-driven force acting on the physics and technology of the semiconductor devices that make up the computer chips — indeed, this is one of the strongest and most-widely recognized such force.

I have sometimes heard it said that in the future everything will be done with software, and that hardware will become less and less important, even irrelevant. While I agree with the first part of this statement, I consider the *conclusions* drawn in the second part to be nonsense: The demands on the hardware will continue to increase with no foreseeable limit exactly as long as the demands on the software will continue to increase!

Similarly, in the field of high-speed analog devices, I see no end to the drive for higher and higher frequencies, with similar implications for device technology.

As these development processes continue, we will eventually reach technological or even physical limits to the further improvement of the devices

Some of the current interest in quantum effect devices is stimulated by the realization that the internal dimensions even of conventional types of devices such as FET's are beginning to get smaller than important physical parameters such as electron mean free paths or even the quantum mechanical wavelength of electrons. But quantum effect devices would almost certainly have operational characteristics very different from conventional transistors, calling for new architectures, and I have so far seen very little in the way of *specifics* of how exactly quantum effect devices will contribute to meeting those applications-driven needs for computers that I stated earlier. Being myself an active participant in research on that topic, I can hardly be accused of being a pessimist on the promise of research on quantum devices, but it may very well turn out that the real payoff of that research will be in areas other than those that seem to motivate the research — as has happened many times before. We will simply have to wait for an answer.

These ultimate limits of devices will almost certainly be decided by discoveries not yet made, and be basically unpredictable. Hence, the longer-term development, over periods longer than, say, 10 years should be expected to be increasingly discovery-driven, and I would expect that progress beyond 30 years from now will be due primarily to unpredictable discoveries not yet made.

Anyone who has any doubt about the role of unforeseen discoveries is invited to look at history, the farther back the better. Towards the end of his life, Max Planck once described that, in the 1870's, he was told by one of his professors that there was not much

future left in physics: With the discovery of the principle of conservation of energy, physics was essentially complete, and that all that was left was working out the details. Even a quarter century later, in 1900, the year Planck had taken the first step towards quantum mechanics, almost nothing of the physics and technology dominating the last decade of the 20th century could have been predicted.

But we need not go that far back that far. The creative professional career of a scientist or engineer tends to last about 40 years; and for someone near the end of his career like myself, that is a good "natural time span" for looking back to see what the rate of progress has been during the last 40 years — since 1954 — because that is likely to be again he rate of progress for the next 40 years. In 1954, semiconductors meant germanium, not silicon ("you cannot get rid of that oxide; besides, the mobilities are too low"). The semiconductor laser had not been conceived yet. Much of semiconductor physics did not exist yet, like hot-electron physics (Gunn effect) Going beyond semiconductors, superconductivity was still was not understood, even though it was over 40 years old. Other examples could be given. All the examples given changed their status during the next 10 years, and much of the core of today's technology could at least be rationally anticipated by 1964 if one made some reasonable allowances for the likelihood that progress would continue and that certain remaining difficulties were likely to be solvable.

In the 30 years since 1954 many new discoveries have been made that have, at least so far, not found their way into commercial device technology. But if history is any guide, many of those will also find their way into actual applications. What is hard to predict are individual cases.

There are of course those who deny the applicability of arguments that draw on history as far back as the young Max Planck, or even only 40 years. After all, we now *do* have quantum mechanics, and how many more unmade discoveries can there be? This argument is fundamentally irrefutable, and at one point or another in the future it may indeed become true. But from what progress we have witnessed just in the last 20 years (half a professional career), I see no reason to believe that this saturation point is near. I would rather place my bets on the assumption that the young people coming out of our universities will prove just as innovative as we were, rather than make the arrogant assumption that innovation will dry up as the present generation gradually passes from the scene.

Let us return to my opening paragraph and to the problem of experts tending to be too conservative. I believe that there are several separate reasons for this failure-through-conservatism of most rationally-based technology forecasts.

(1) The longer the forecast period, the larger the fraction of the new technology that is based on new discoveries that could not be rationally predicted. This was always true in the past, and I see no reason to believe that this will soon change.

(2) A much less obvious but perhaps more troublesome reason is that most responsible engineers and scientists tend to view the future utilization of new discoveries in the light of *already existing applications needs*, where the new discovery has little chance to be used in the face of competition with already-existing and entrenched technology. A simple improvement in performance just will not do!. But the history of technology teaches us that new discoveries tend to create *new* applications and that the main applications of new science and technology have usually been such applications *created* by the new science.

Perhaps the most important example of this central historical lesson is the transistor itself. Initially viewed simply as a replacement for electron tubes, and for such pedestrian applications as portable radios, it ultimately *created* the modern computer and the new industrial revolution that followed it.

Another example of a device creating its own application was the double-heterostructure laser. I recall painfully that I was told in 1963 that there was no point in developing a technology for this new concept, because this device would *never* be useful, because of its the low anticipated power and a relatively poor spectral purity. If those skeptics had been right, we would today not have optical fibers, nor compact discs. In fact, the optoelectronics that developed in the wake of the DH laser is likely to be one of the "driving engines" for device development well into the next century.

As a third example — of a different kind — let me also mention the HEMT. It did not live up to the initial expectations many of us had for it as a device for high-speed RAM's. If everything else had been the same, the higher mobilities in GaAs would have given it a considerable speed advantages over Si RAM's. But everything else just wasn't the same, and GaAs HEMT's never could compete with Si RAM's — ultimately not even on speed. What happened instead was that they turned out to be superb low-noise devices for the direct reception of TV signals from satellites, practically creating the industry of those small (if ugly) dishes seen outside many windows worldwide. It would in principle have been possible to do that with Si FET's, but the better noise performance of HEMT's permitted the use of much smaller dishes, and this created a large economic *leverage* that more than made up for the higher cost of the FET itself — not to mention the much better customer acceptance of the smaller dishes. Please keep that concept of *leverage* in mind; I will return to it shortly.

I believe that this pattern of new science creating new devices that create their own applications will continue in the next century. I do not think we can realistically predict which new devices may emerge, but I believe we can create a psychological environment for progress by not always asking immediately what any new science might be good for (and cutting off the funds if no answer full of fanciful promises is forthcoming — a worldwide problem). Instead, we must make it an acceptable answer if the researcher tells us that it is one of the tasks of the research itself to search for new applications. This answer is of course acceptable only if it is a sincere one, rather than being lip service. The attitude necessary for this seems to be more highly developed in Japan than elsewhere. Amongst some physical

scientists in some other societies, there is still too much of a value system where pure research is somehow more respectable than applied research. A change in social values might go a long way helping those societies retaining their global competitiveness.

(3) One of the weaknesses I see in too many attempts to look at the future of semiconductor devices is that new concepts are often judged by whether they can be mass-produced at the huge volumes and low cost that are characteristic of Si integrated circuit technology. This is of course appropriate for concepts that are indeed intended to find their application in the same market as Si integrated circuits, where it is indeed extraordinarily difficult to compete with the existing technology.

But as I pointed out above, the applications of new concepts are more likely to be applications that get *generated* by the new concepts than pre-existing applications, and here the economics is an altogether different one. What matters for the economic viability of the new technology is simply whether the new application can support the R&D cost and the increase in manufacturing cost of that new technology. If a new technology has enough of that crucial economic *leverage* I referred to earlier in the context of HEMT's, it may be economically viable even at a very low manufacturing volume with a high attendant cost per device. For example, if a new but expensive-to-make $500 device would make possible a new $20,000 instrument that could simply not be built without that device, and if there were enough demand for the enhanced capability of that instrument to sell enough of them to permit a recovery of the $500 cost of making each device, then the technology for making the device would be self-supporting, and would have a chance of surviving — never mind that the increase in cost over, say, silicon technology would be huge: The latter could not do the job. Recent history abounds with examples of such high-leverage devices, and one of my predictions is that we will see much more of this, especially in the instrumentation and sensor field, and that high-leverage applications in these fields will be amongst the driving engines of device technology for the next century.

The number of such devices and even their total money value may be miniscule to the number and money volume of Si IC's, but this does not in any way diminish the attractiveness of the device to those working on it: Working on high-leverage special-purpose devices may, in fact, be an attractive career path for a young scientist or engineer. Moreover, it is an excellent way for universities to prepare future scientists and engineers for the technologies of the future — they can learn the technologies of the present on their first industrial job much better than at the university.

Let me say a few more words about that "It can never compete with silicon" syndrome. It is undoubtedly essential that you are able to compete with silicon *if* you want to do something that can be done reasonably well with silicon. But Si is not everything in electronic metallurgy any more than steel is everything in structural metallurgy. Just as steel, the material from which we build automobiles, railroads and ships, was and is likely to remain the basic material for *structural* metallurgy, both in tonnage and money volume, so silicon IC technology is likely to remain the dominant device technology, both in number of chips and money volume. But just as we need other structural metals, such as aluminum, magnesium, titanium, etc., as structural materials for aircraft, spacecraft, etc., we will need to go beyond Si IC technology for numerous applications that are not digital IC type applications.

The reader may have noticed that (throughout this Extended Abstract at least) I have carefully avoided being too specific, but have rather stuck to general principles. There is, however, one area where I *do* wish to stick my neck out a little more specifically: Cryogenic devices, especially devices operating at temperatures not below 77K, offering much higher performance (by several criteria) than room temperature devices. I am quite convinced that such devices have a great future, which will be paced largely by the increasing availability of small self-contained closed-cycle refrigerators (mainly Stirling cycle machines). The development of the latter is rapidly approaching the point that we may begin to view them as just another module inside a piece of electronic equipment, analogous to a power supply. The two principal bottlenecks to their widespread use are cost — and the lack of semiconductor devices actually optimized for low-temperature operation. The first of these problems is likely to follow the classical pattern of dramatic cost reduction in the wake of building up mass production. The solution of the second problem is up to us. The devices that very likely will emerge will not only be superconducting devices using high-T_c superconductors, but also conventional devices such as FET's made from unconventional narrow-gap materials such as InAs. Given these developments, I expect to see cryogenic modules in high-performance desk top work stations, and in similar demanding high-volume applications.

We will probable even see mainframe computers based superconducting devices operating at liquid-He temperatures, mainly in two kinds of environments: (a) The largest and highest-performance computers, where the cost of the cryogenics is only a small fraction of the overall machine cost. (b) Various high-tech environments — increasingly common — where He temperatures are available anyway, and where the utilization of that environment comes at small additional cost, but offers large performance advantages. Both types of example are not likely to emerge as high-volume applications anytime soon, but they are again examples where a large leverage justifies a high expenditure at low-volume.

The oral presentation will discuss a few additional specific device categories, such as opto-electronic devices, microwave devices, and selected sensors.