phys. stat. sol. (a) **202**, No. 6, 957–964 (2005) / **DOI** 10.1002/pssa.200460701

Expert Opinion

Nano-whatever: Do we *really* know where we are heading?

Herbert Kroemer [*]

ECE Department and Materials Department, University of California, Santa Barbara, CA 93106, USA

Received 15 July 2004, revised 21 July 2004, accepted 12 January 2005
Published online 16 March 2005

PACS 61.46.+w, 73.22.–f, 78.67.–n, 81.07.–b, 85.35.–p

The principal applications of any sufficiently new and innovative technology have always been applications that were *created* by that technology, rather than being pre-existing applications, where the new technology simply provided improvements. This historical trend is likely to remain true, and it is likely to be true specifically for the emerging nanotechnology.

Applications yet to be created have been notoriously hard to predict. Historically, a rich set of new (and unforeseen) applications has almost always arisen; this is likely again to be the case in nanotechnology. Hence, nanotechnology should be expected to have a rich future with new applications, but the details of that future are resistant to prediction. The most we can do is to identify some general research *directions* that appear promising – and I will attempt that here – but for the ultimate outcome we will have to wait until it happens: The progress from new technology to new applications is *opportunistic* rather than *deterministic*.

1 Introduction: "Nano-talk = Giga-hype?"

The title of my talk seems to indicate a certain skepticism, and one might wonder whether in old age I have turned into a pessimist who no longer believes in a bright future. Nothing could be farther from the truth. I share the optimism of many of those who are actually actively involved in developing the field, be it in physics, technology, or applications (like many at this conference). My skepticism pertains to the unbelievable hype that has arisen, during the last decade, about the "nano-whatever" field, a hype that exceeds anything I have encountered during my fifty years in solid-state physics and technology. The prefix *nano* suddenly gets attached to everything (this conference is no exception), and we are deluged with predictions about fantastic future applications, often promised for the immediate future.

That wouldn't be so bad if these predictions came predominantly from individuals who are actually knowledgeable about the topic, individuals in the technical community who are actually actively involved in developing the field, be it in physics, technology, or applications (like many at this conference). But if those were the principal speakers, the hype probably wouldn't be so bad. Much of the hype is being generated by outsiders, not actually involved in these three areas, and basically clueless about how the process from science and technology actually works, but simply wishing to profit from the development. I fear a severe backlash when it becomes clear that many of those predictions are unrealistic.

[*] e-mail: kroemer@ece.ucsb.edu, Phone: +01 805 893 3078, Fax: +01 805 893 7990

physica status solidi a

2 From physics and technology to new applications

2.1 Kroemer's *Lemma of New Technology*

Any detailed look at the history of how applications have arisen from new science and technology provides staggering evidence for what I have called, on other occasions [1, 2] my *Lemma of New Technology*:

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

In other words: The principal applications have *not* been of the kind where one improves on something that could already be done on an acceptable level; they have been applications that could previously not been met at all – or were not even recognized.

As a rule, such applications have indeed arisen, although usually not immediately. The problem with truly new applications is that they are extraordinarily hard to predict, and were historically rarely predicted. Ultimately, progress in applications is not *deterministic*, but *opportunistic*, exploiting for new applications whatever new science and technology happen to be coming along.

I will give below three examples from the pre-nano past for my thesis, but let me draw one conclusion right away: We *must* take a long-term look when judging the applications potential of any new technology. The latter must *not* be judged simply by how it might fit into already existing applications, where the new discovery may have little chance to be used in the face of competition with already entrenched technology. Dismissing it on the grounds that it has no known applications will *retard* progress towards the ultimately important applications rather than advancing it!

2.2 Three examples

2.2.1 The transistor

When the transistor was invented – or should I say *discovered*[*]? – it was widely viewed as a replacement for electron tubes or for electromechanical switches. These developments did indeed happen. But the dominant significance of the transistor is that it *created* the modern computer. This development could hardly have been predicted, nor was it predicted at the time: The original transistor was the bipolar transistor, but the computer revolution is based on CMOS transistors whose low dissipation made large-scale ICs possible.

Ironically, the discovery of bipolar transistor action (via minority-carrier injection) grew out of a *failed* attempt to build a field-effect transistor (FET) in germanium. The interface-state density at the interface between the Ge and the gate dielectric was far too high to permit a modulation of the carrier density in the FET channel. There was no reason to expect the situation at a Si/SiO_2 interface to turn out any better, but this is what actually happened, leading to the development of MOS transistors and ultimately (after long doubts) to CMOS.

2.2.2 Semiconductor laser

When, in 1963, I proposed the idea of the double-heterostructure laser (DHL), I was refused resources to develop the necessary technology, on the grounds that this device could not possibly compete with existing lasers and hence would never find any applications. This was of course a failure to recognize the central role of the creation of new applications, expressed in my *Lemma*. Instead, the DHL created its own applications, from the CD player to fiber communications, plus many others. And without fiber communications, there could never have been an internet.

[*] The citation for the 1956 Physics Nobel Prize reads "for the discovery of the transistor effect".

Fiber communications actually required *two* new technologies, the DH laser *and* a technology for low-loss glass fibers. Such synergisms involving two or more new technologies are a frequent feature of new applications. But they increase the overall unpredictability of the path from new technology to new applications.

Once the DHL technology was established, *non-lasing* light-emitting diodes started to invade the entire illumination field, and are in the process of taking over *all* applications that require *colored* light in any case (traffic lights are a good example), as well as the low-wattage white-light field. Such "backdoor invasions" into pre-existing applications are also a common feature of a new technology – *after* it has created its own new applications.

2.2.4 Nuclear magnetic resonance

When Bloch and Purcell received the 1952 Nobel Prize in Physics for the development of NMR, this new phenomenon was a purely scientific discovery, not anticipated to have important applications. Yet in the end it *created* one of the central analytical tools for organic chemistry. But this was only the beginning. Out of this chemical tool grew *medial* magnetic resonance images, one of the most powerful diagnostic tools of modern medicine, honored by the 2003 Nobel Prize in Medicine. This development again required synergism with another technology, the development of superconductor wires capable of retaining their superconductivity in high magnetic fields, based on ideas of Abrikosov, who was honored by the 2003 Nobel Prize in Physics.

2.3 Lessons

From the *Lemma of New Technology* flow a number of lessons [2]. Although we cannot realistically predict which new devices and applications may emerge, we can create an environment encouraging progress, by *not* always asking immediately what any new science might be good for. In particular, we must educate our funding agencies about this historical fact. This may not be easy, but it is necessary. Cutting off the funds if no answer full of fanciful promises is forthcoming is not the road to progress, a fact not altered by disguising this shortsightedness with the fancy name of "strategic research." We must make it an acceptable answer to the quest for applications to defer that answer, and that at the very least a search for applications should be considered a part of the research itself, rather than a result to be promised in advance. Nobody has expressed this last point better than David Mermin in his recent put-down [3]

> "I am awaiting the day when people remember the fact that discovery does not work
> by deciding what you want and then discovering it."

What is *not* acceptable – and what we must refrain from doing – is an attempt to justify the research by promising credibility-stretching mythical improvements in *existing* applications. Most such claims are not likely to be realistic and are easily refuted; they only trigger criticism of just how unrealistic the promises are, thereby discrediting the whole work.

3 Roots of nano-technology

Given the hype surrounding nano-technology (NT), few people realize that some of us have been practicing nano-technology for over 30 years – we just didn't call it nano-technology. By 1974 we saw the first papers on actual experimental devices demonstrating quantum effects, grown by the new molecular-beam epitaxy (MBE) technology. Involving quantization effects, this work was NT *per definitionem*. In the 30 years since then, these techniques have progressed much further; perhaps the most spectacular results being in quantum cascade lasers (see the paper by Capasso at this conference).

A second root of NT is colloid chemistry, which naturally deals with particles on a nanoscale, and which has been remarkably successful in synthesizing “loose” semiconductor nanoparticles with interesting optical properties. I will say more about them later.

But *physical* technologies, like epitaxial growth techniques or colloid chemistry, were joined biological techniques. The synthesis of long DNA strands with pre-set sequence of base pairs is pure NT in biology and medicine. The combination of physical and biomedical techniques will undoubtedly be one of the most important areas of NT overall, and it deserves the highest possible level of attention at a conference like ours, which grew out of the physical root of the technology. Not being personally knowledgeable about biology, I will, for the remainder of my presentation, stick to *physical* NT, but I would strongly recommend that the leaders of this conference routinely include in future meeting one or two prestigious invited speakers from the other side.

4 Back to the future: Beyond a single degree of quantization

One of the characteristic features of more recent development in physical NT is an increasing emphasis on nanoscale effects in more than one spatial dimension, as opposed to the earlier structures where we studied quantization in just one dimension, with the remaining two spatial dimensions remaining macroscopic. Maybe this is the threshold beyond which it becomes legitimate to speak of NT.

4.1 Quantum wires

To me, the beginning of the search for reduced dimensionality was Sakaki’s 1980 “dream” of achieving huge electron mobilities in ideal quantum wires [4], drawing on the idea that in a 1-D transport structure an electron could scatter only forward or backward, with the forward process being inefficient in momentum dissipation, and the latter being improbable.

In the end, this promise remained unfulfilled, due to interface roughness scattering, which had limited the achievable mobilities already in a 2-D electron gas, but which was much more severe in quantum wires whose surface-to-volume ratio is much larger. In effect, the second spatial dimension of quantization, *within* the plane of the wafer, was purely “man-made” by lithography, without any help from Nature.

It was to be the first encounter with what I consider to be the central problem of almost all nanostructures, the problem of statistical fluctuations – geometrical and others – on the atomic level just below the nominal nanometer dimensions of the structure. What we need badly is way to have Nature help us rather than fighting her. One potential set of candidates with nature-made wire dimensions are the various nanotubes currently being explored. It is too soon to offer a verdict on this possibility, but I was encouraged by a paper at the 2003 APS March meeting where single-wall carbon nanotubes were reported, with mobilities higher than in most bulk semiconductors.

4.2 Quantum dots

Much of the current research in nanostructures has turned towards quantum dots (QDs), with three dimensions of spatial quantization, especially for optical applications. The initial research utilized the “self-assembled” QDs that form spontaneously during the initial MBE deposition of a material on a lattice-mismatched substrate. This had the advantage that no nano-scale lithography was needed. But the price for this is that the fluctuation problem becomes even more severe than in the case of quantum wires, with fluctuations both in position and size of the dots. While individual single QDs have very narrow optical lines, the exact energetic location of these lines is subject to fluctuations that are very large compared to this natural line widths. In multi-dot ensembles, these fluctuations cause huge inhomogeneous line width broadening within a dot assembly, making the often-cited delta-function-like density-of-states distribution of a set of identical dots an illusion.

There are of course applications where such fluctuations are acceptable, and I am in fact an admirer of just how much has been achieved along this direction (see the paper by Henini at this conference). But I believe that the full potential of quantum dots for future applications goes far beyond what is achievable with this approach, and that we must ultimately overcome not only the *size* fluctuation, but also the *position* fluctuations. In fact, tight size control is probably unachievable without tight position control, and position control also becomes necessary for applications where the dots need to be contacted electrically. Hence QDs do not relieve us of the continuing need for lithography down to tens-of-nanometers size scales; lithography will remain an integral part of QD technology.

Even with pre-patterning of the substrate, tight size (and shape) control is not assured. If, for size control we simply rely on intercepting all the atoms impinging on a precisely pre-defined collection area, we still have to contend with purely statistical Poisson fluctuations: If N is the average number of atoms (or bonding pairs) per dot, the actual number will fluctuate with a standard deviation of $\sqrt{N}$. For, say, $N = 1000$ (not an unrealistic extrapolation of what we may wish to achieve in the future), this would imply a 3% size variation, probably too large for at least the more demanding applications. Hence, overcoming the Poisson fluctuations must be one of our targets. One simple approach would be to forgo self-assembled dots altogether, and to employ post-growth "cookie-cutter" lithography on extended epilayers of precise thickness to define both size and position of the dots. There are almost certainly other approaches.

5 More challenges

5.1 Lithography alternatives for the nanoscale

Given the continuing need for lithography on a nanoscale, we should pursue alternatives to the extreme-UV photolithography approach currently pursued (for good reasons) by the semiconductor industry. For pure research, *serial* techniques, like e-beam or ion-beam writing, or using a STM-like scanning probe, are acceptable. But for practical applications we need techniques that have the massively parallelity of photolithography is. In fact, it is this parallelity that lies at the root of Moore's Law. I believe that nano-printing is a promising alternative; it is already being pursued by several research groups, and I believe we should pay more attention to it.

5.2 "Loose" nanoparticles

As I mentioned earlier, one of the roots of NT has been colloid chemistry, which by its very nature deals with "loose" nanoparticles, rather than QDs on a substrate. Loose nanoparticles suffer from the same fluctuation problem as QDs on a crystal surface. But they have the advantage that they can, in principle, be sorted after their synthesis. I would consider the development of suitable sorting techniques a matter of high priority, ideally not just sorting by size (via centrifuging), but via selected optical excitation.

Much of the current interest in loose nanoparticles is in their optical properties, at least where a precisely-defined spatial localization is not required. But I also consider them to be potential building blocks for assembly into 3-D nanostructures, similar to the way tennis balls placed into a proper-sized (and shaped) box tend to arrange themselves into a regular lattice. In fact, the achievement of nanoscale 3-D structures with controlled properties, rather than just 2-D structures, is one of the greatest (and so far unsolved) challenges to nanotechnology. Assembling nanoparticles would be one of the alternatives to in-situ synthesis.

Another application worth exploring would be their use as chemical catalysts, including specifically photocatalysts operating under illumination, for example, using sunlight to split water molecules into hydrogen and oxygen. The average photon energy in sunlight is significantly above the 1.23 eV *net* energy needed, but water does not absorb at that energy, due to a high reaction barrier that must be overcome, which requires a suitable catalyst. If NT could provide such a catalyst (and perhaps also provide suitable membranes to separate the H_2 from the O_2), the impact on the world-wide energy problem could be huge.

6 "Other" quantization effects

Throughout the above, the word *quantization* has, at least implicitly, referred to the quantization of energy levels in nanoscale structures. But there are other physical properties that are naturally quantized.

6.1 Charge quantization and Coulomb blockade

The oldest of these is the quantization of electric charge. Classical electronics treats electric current flow as a continuous fluid, and to the limited extent that the discreteness of the electron charge enters classical circuit theory at all, it is as a nuisance effect, in the form of shot noise.

One of the great opportunities of emerging NT is the potential of single-electron electronics. NT is rapidly making it possible to build capacitors sufficiently small that the voltage per electron needed to charge the capacitor becomes larger than the thermal voltage kT/q. For example, for a cube-shaped capacitor with linear dimensions of 10 nm, filled with a dielectric with a dielectric constant of 10, that voltage is 180 mV, already large compared to a room-temperature kT/q of about 26 mV.

Following my *Lemma*, I am not going to speculate what exactly what applications such effects might have, be they in metrology, instrumentation, digital logic, or what-have-you, but I am convinced that research in this field should and will be an active area of future NT.

6.2 Magnetic flux quantization

A second example of "other" quantization effects is the quantization of the magnetic flux contained in a superconducting loop. It has been pointed out that computers based on Josephson-junction circuits manipulating discrete flux quanta could in principle be much faster and have much lower dissipation than a conventional computer based on voltage-state logic in semiconductor circuits [5]. In the absence of room-temperature superconductors, the required cryogenics would of course rule out such a scheme from "everyday" computers, but the cryogenics cost is likely to be acceptable in advanced larger computers where the highest possible performance outweighs all other considerations.

At present, the principal limitation to the implementation of this scheme is the technological one of statistical fluctuations amongst Josephson junctions due to fluctuations in their conventional oxide tunnel barriers. If NT could provide atomic-level control over the barriers, this just might lead to the ultimate high-speed mainframe computer.

6.3 Spintronics?

Electrons have not only charge, they also have spin, and *spintronics* is the field dedicated to attempts to exploit the spin. Its greatest applications triumph has been in the phenomenon of giant magnetoresistance, the basis of all modern magnetic-disk read heads. Spintronics is not necessarily a part of NT, but it overlaps with NT, and is often quoted in a NT context. Hence, it cannot be ignored in a discussion of NT and its potential applications.

As a very active area of modern solid-state physics and technology, spintronics, too, has received a high level of hype, second only to "nano". And as in the case of "nano", I am skeptical about some claims about future device applications, specifically about applications where the spin is utilized in an "imitation mode", namely, using the electron *spin* rather than the electron *charge* for various electronic functions that can be performed perfectly well – or better – by charge alone.

A notorious example is a "spin transistor" imitating an ordinary bipolar transistor, but using spin orientation rather than electron charge as the signal carrier from an "emitter" to a "collector". As a *re-*

search tool for studying spin physics it is beyond reproach, and I am in fact highly in favor of unrestricted pure research of any kind in spintronics. But viewed as a *practical* device, the spin transistor has no chance of competing with an ordinary bipolar transistor in any applications that can be met by the latter. Similar criticisms apply to several other applications that have been proposed for spintronics devices.

My advice to individuals wishing to go beyond pure research in spintronics, and looking for real-world applications of spintronic devices, is to stick to applications that cannot be done without spin, where spin is necessary! Giant magnetoresistance is one of those. I am sure there are others, for example non-reciprocal wave propagation effects like Faraday rotation. And of course there is the application of spintronics to hypothetical future quantum computers.

7 Meta-materials

One area of NT that I believe to be worth pursuing is the field of *meta-materials*, defined as artificially structured quasi-bulk materials with new properties that are basically unattainable in an unstructured homogeneous bulk material. What we normally consider a homogeneous material is, on an atomic level, a highly inhomogeneous structured assembly of individual atoms. Nanostructured meta-materials simply work with a second structural level where the "atoms" on that level are themselves artificial "quasi-atoms" put together from true natural atoms.

A promising example of the latter would be the extension of so-called *photonic crystals* to optical wavelengths. This name refers to structures in which – in their simplest form – the dielectric permittivity varies periodically with position, with a periodicity that is on the order of the wavelength of the radiation inside the medium. (For a review of photonic crystals see [6].)

Wave propagation in periodic media invariably exhibits *band structures*, with allowed and forbidden bands of propagation, regardless of the nature of the waves. Most work on photonic crystals has concentrated on achieving frequency bands inside which the propagation is *forbidden*. But unusual effects also take place inside the *allowed* propagation bands. Just as the allowed bands in an *electronic* band structure contain regions where the *effective mass* of the electrons becomes negative (at least in certain directions), a *photonic* band structure contains regions where a (suitably defined) *effective refractive index* can become negative [7].

Optical gaps and negative refraction have been demonstrated at microwave frequencies, where the structure periods are macroscopic. Their extension to optical frequencies provides a strong impetus for developing nanotechnologies in this direction.

8 The research-vs-applications issue re-visited

Even if the process from science and technology to applications is opportunistic rather than deterministic, we can speed up this process by better cross-discipline communication between scientists, technologists, and application engineers. It is not uncommon that new possibilities arise within science or technology, with its originators having few (or unrealistic) ideas about applications for it. But an applications engineer might be aware of a potential new application that are just waiting for just such a scientific/technological breakthrough to make the application possible.

What I am pleading for is a new form of "triangular workshop" in which these three principal groups involved talk to each other, rather than each group talking only amongst its own members. For example, a scientist might see some scientific possibilities, and might ask the technologists how the necessary structures might be built. Conversely, a technologist might see some possibilities to build something, and might ask the scientists whether such a structure might have interesting properties. Most importantly, the two groups might ask the applications engineers what they could do with such structures if they had them. The process would be turned around, when an applications engineer perceives of a potential new

physica status solidi a

revolutionary application, and would ask the scientists about the scientific possibilities and constraints, and the technologists how the required structure could be built.

For example, given my interest in photonic crystals, I would like to throw the technologists the challenge to develop photonic crystals with interesting refraction properties in the optical regime (infrared and visisble), which would require large-amplitude periodic variations of the local refractive index with a nanoscale periodicity. And I would also throw the engineers the challenge of what applications they might have if we gave them a slab of material having a negative refractive index, or other bizarre optical properties.

9 Conclusion

The title of my presentation asked the question: "Do we *really* know where we are heading?" To which my answer is: "Not exactly." We can identify a significant number of different overall *directions* that appear promising, but I do not believe we can predict which of these directions will be the most promising ones and which – if any – may be blind alleys. And my (or anybody else's) list of directions is almost certainly incomplete; be it because of limited insight, or because new unforeseeable direction may emerge in the course of time. Within any given direction, I do not believe we can foresee just how far this direction will progress.

But does this uncertainty really matter? Hardly! There is enough "out there."

References

[1] H. Kroemer, Rev. Mod. Phys. **73**, 783 (2001).
[2] H. Kroemer, in: Future Trends in Microelectronics: Reflections on the Road to Nanotechnology, Proc. NATO Adv. Res. Workshop, Ile de Bendor (France), 1995, edited by S. Luryi, J. Xu, and A. Zaslavsky, NATO ASI Series E, Vol. 323 (Kluwer Academic Publishers, Dordrecht, 1996) p. 1.
[3] N. D. Mermin, Physics Today **52**, 11 (1999).
[4] H. Sakaki, Jpn. J. Appl. Phys. **19**, L735 (1980).
[5] K. K. Likharev and V. K. Semenov, IEEE Trans. Appl. Supercond. **1**, 3 (1991).
[6] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, Photonic Crystals (Princeton University Press, 1995).
[7] C. Luo, S. G. Johnson, and J. D. Joannopoulos, Appl. Phys. Lett. **81**, 2352 (2002).