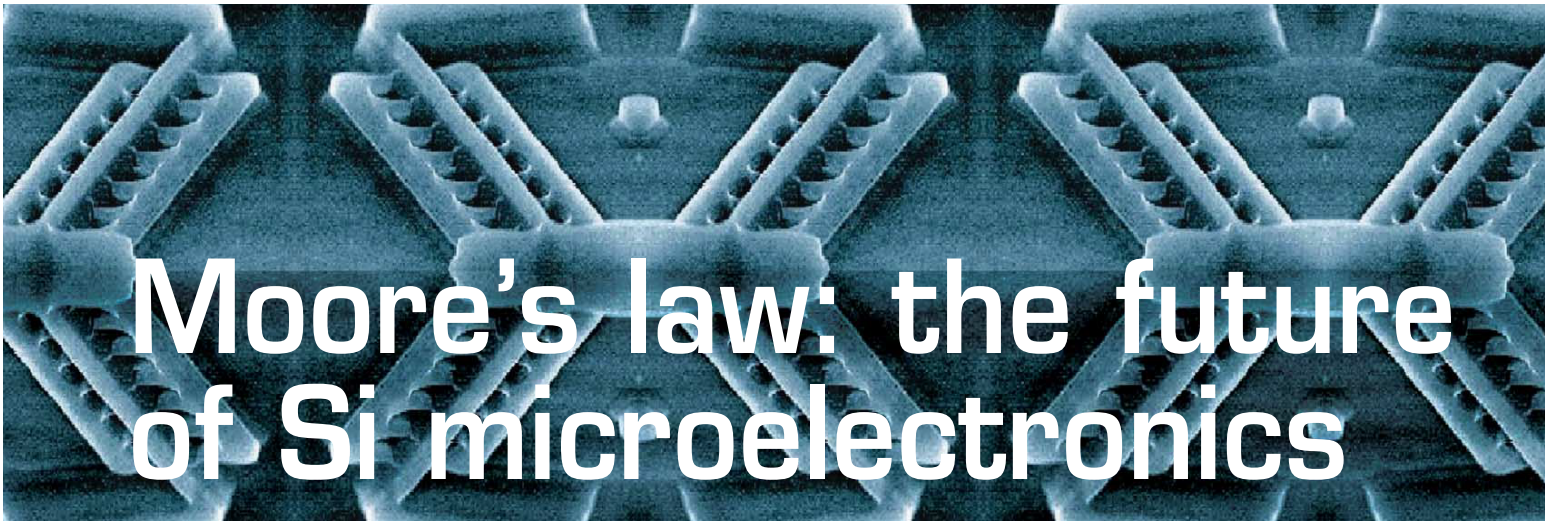

Moore's law: the future of Si microelectronics

Soon after Bardeen, Brattain, and Shockley invented a solid-state device in 1947[1] to replace electron vacuum tubes, the microelectronics industry and a revolution started. Since its birth, the industry has experienced four decades of unprecedented explosive growth driven by two factors: Noyce and Kilby inventing the planar integrated circuit[2,3] and the advantageous characteristics that result from scaling (shrinking) solid-state devices.

Scott E. Thompson and Srivatsan Parthasarathy
SWAMP Center, Department of Electrical and Computer Engineering, University of Florida, Gainsville, FL 32611-6130 USA
E-mail: thompson@ece.ufl.edu, psrivats@ufl.edu

Scaling solid-state devices has the peculiar property of improving cost, performance, and power, which has historically given any company with the latest technology a large competitive advantage in the market. As a result, the microelectronics industry has driven transistor feature size scaling from 10 μm to ~30 nm[4-6] during the past 40 years. During most of this time, scaling simply consisted of reducing the feature size. However, during certain periods, there were major changes as with the industry move from Si bipolar to *p*-channel metal-oxide-semiconductor (MOS), then to *n*-channel MOS, and finally to complementary MOS (CMOS) planar transistors in the 1980s, which has remained the dominate technology for the past two decades. The big challenge going forward is that the end of planar CMOS transistor scaling is near as the transistor size approaches tens of nanometers. How the industry evolves after this limit is reached is unclear.

To address these challenges, present day research is focused on identifying new materials and devices that can augment and/or potentially replace the aging ~50-year-old Si transistor[7]. Two approaches under investigation are: (1) nonclassical CMOS, which consists of new channel materials and/or multigate fully depleted device structures; and (2) alternatives to CMOS, such as spintronics, single electron devices, and molecular computing[8,9]. While some of these non-Si research areas are important and will be successful in new applications and markets[10], it seems unlikely any of the non-Si options can replace the Si transistor for the $300 billion microelectronics industry in the foreseeable future (perhaps as long as 30 years).

This review aims to explain the future of Si microelectronics, key issues at the end of the Si roadmap, and the time frame for possible non-Si technology replacements. We first discuss the state of Moore's law and conventional planar Si transistor scaling limits. Next, we cover the issues at the end of the Si roadmap based on current technology trends. We end, perhaps foolhardily, with an assessment of nonclassical CMOS and alternatives to CMOS. The key takeaway messages are that simple scaling has ended, there is enormous life left in planar Si CMOS technology, and nothing is on the horizon to replace it for mainstream logic applications.

ISSN:1369 7021 © Elsevier Ltd 2006 Open access under CC BY-NC-ND license.

TSMC EXHIBIT 1059

Status of Moore's law

Moore's law is the *empirical* observation that component density and performance of integrated circuits doubles every year[11], which was then revised to doubling every two years[12]. Guided by the scaling rules set by Dennard[13] in 1974, smart optimization, timely introduction of new processing techniques, device structures, and materials (in many areas of the device except the channel), Moore's law has continued unabated for 40 years. Driven by tremendous advances in lithography, the 65 nm logic technology node featuring ~30 nm transistors is currently in high volume production[14,15]. Furthermore, 45 nm and 32 nm technologies with process targets defined to maintain Moore's law are currently under development at several companies. With such small feature sizes in high volume production and under development, Si CMOS technologies are now leading the field of nanotechnology and will continue to do so. Nanotechnology is defined, according to the National Science and Engineering Technology Council (NSET)[16], as:

> *Research and technology development at the atomic, molecular, or macromolecular levels, in the length scale of approximately 1-100 nm range, to provide a fundamental understanding of phenomena and materials at the nanoscale and to create and use structures, devices, and systems that have novel properties and functions because of their small and/or intermediate size. The novel and differentiating properties and functions are developed at a critical length scale of matter typically under 100 nm.*

Si MOSFETs entered the nanometer era around 2000, as seen in Fig. 1 that shows technology node and transistor feature size versus year for the semiconductor industry. For the 0.13 μm technology node, the industry incorporated ~70 nm gate length transistors on average. What is interesting to note is that traditional top-down microelectronics have not only become nanoelectronics but the device dimensions are now comparable to those being explored in the new field of bottom-up nanotechnology and molecular electronics!

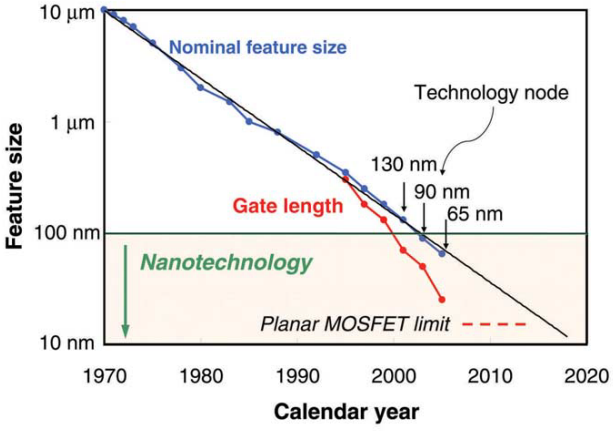

Fig. 1 Logic technology node and transistor gate length versus calendar year. Note mainstream Si technology is nanotechnology. (Published at the International Electron Devices Meeting.)

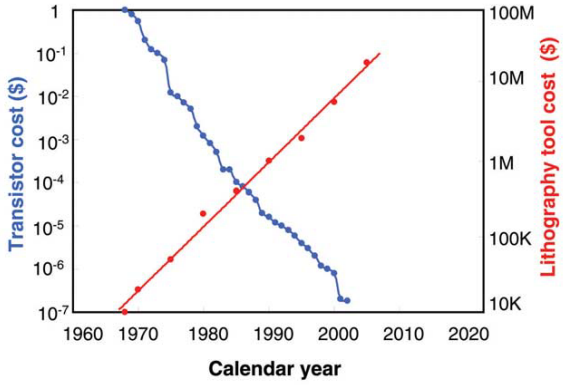

Fig. 2 Transistor cost and lithographic tool cost versus years. Note transistor cost has decreased seven orders of magnitude even while tool cost has increased.

The key driver behind these trends is economics, as pointed out by Moore in 1965[11]. According to Moore, integrated circuits and scaling are "the cheap way to do electronics". Even with large increases in lithography tool cost to fabricate nanoscale CMOS transistors (for example, the cost of lithography steppers increased from $10 000 to $35 million, as shown in Fig. 2), which has lead to modern factories costing $2-3 billion, the cost per transistor has decreased by seven orders of magnitude during the last 40 years (Fig. 2) and is likely to continue to decrease for another decade. However, CMOS transistor scaling must inevitably slow down and finally halt, at least in the traditional sense, as the lithography scale approaches atomic dimensions.

CMOS limits

When starting a discussion of CMOS limits, it is first important to point out that when the limits are hit, this will not be the end of integrated circuits or Moore's law. All it means is that the rate of improvement will change once again (as happened in 1975[12]). Improvements will instead come from areas other than scaling and Si CMOS technology will continue for many decades before a credible alternative arises. For example, even if device density slows, cost per transistor will continue to be reduced through improved tool productivity, cycle time reduction, defect elimination, and possibly, though unlikely, another wafer size conversion by the industry[17], thus further extending Moore's law.

Next it is important to classify the limits as practical or theoretical and into the categories of lithography, transistor, and wiring. In this work, we will focus on transistor limits since this appears to be the most serious issue. At present, lithography will not be the limiter. As a testament as to how far engineers can push mainstream technology, conventional optical lithography enhanced with high numerical aperture, retical enhancement techniques, and double exposure can pattern the 22 nm node[18], putting into question the role of nonoptical lithography techniques such as extreme ultraviolet (EUV) lithography.

Wiring limits, though serious, can be addressed by architecture and adding more metal layers.

Planar CMOS transistor limits

Transistor scaling limits arise from practical limits related to leakage current at small gate lengths. The problem at small gate lengths is that the drain voltage reduces the barrier height at the source, thereby causing a low source-to-channel barrier height even with the gate voltage off, which leads to undesirable, large off-state leakage. This phenomenon is referred to as drain-induced barrier lowering and/or degraded short channel effect (SCE). For evidence that CMOS planar transistors are approaching their minimum practical size, one only need look at the off-state leakage trends for the industry. CMOS was initially promised as a technology that dissipated negligible power in the standby state. In present day high-performance logic technologies designed for microprocessors, the leakage power of CMOS transistors is approximately 20-30 W (out of a total power budget of 100 W). This magnitude of leakage is already at the practical limit since it increases packaging cost (because of cooling) and, even more importantly, energy cost (both in terms of utility bills and the infrastructure to get energy into corporate server computer rooms). To prevent further increases in leakage, the rate of gate length scaling has already slowed in the recent 90 nm and 65 nm technology nodes[14,19,20]. There is no hard limit on the minimum size of a planar CMOS device, but practical considerations on leakage limit the physical gate length to ~20 nm[7].

With the industry already close to the limits of planar transistors, progress can still be achieved using methods other than scaling. One possible option is yet again to move to alternate device structures while still using a Si channel such as multigate fully depleted devices (sometimes called nonclassical CMOS). Without question, these nonclassical devices improve SCE (reduce leakage) because of improved electrostatics from the multigates and ultrathin bodies. Fig. 3 shows a summary of fully depleted devices investigated by the industry during the past two decades. All of these devices are based on a thin layer of Si-on-insulator (SOI). However, this path has not been chosen by the industry at present and keeps getting pushed out to future technology nodes[21,22] because of three difficult and unsolved issues: (1) a thin body leads to higher external resistance, which will be shown to be a significant issue in state-of-the-art nanoscale MOSFETS where the external resistance is becoming comparable in magnitude to the channel resistance[23,24]; (2) significant uniformity, process complexity, and cost issues associated with fabricating multigate devices; and (3) difficulty in engineering the band structure in fully depleted devices using strain (to create a low conductivity mass), which has already become mainstream in planar CMOS to increase carrier mobility (transistor speed)[25].

Instead of moving to a new device structure, the path chosen by the industry is to improve transistor performance without any further shrinking of the transistor gate length by introducing lattice strain into the Si channel. This approach significantly alters the band structure and addresses Si transport deficiencies compared with other high-mobility,

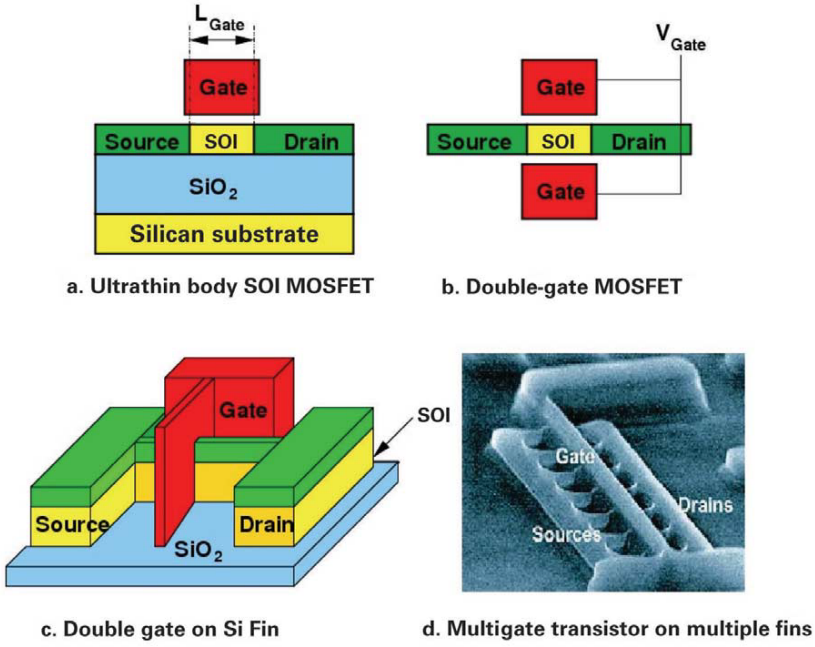

Fig. 3 Emerging alternative fully depleted CMOS structures for continue scaling: (a) one-, (b, c) two-, and (d) three-gate fully depleted devices. (Part (d) courtesy of R. Chau, Intel.)

III-V semiconductors. For example, by altering the position of Si atoms in the face-centered cubic unit cell, the hole conductivity effective mass can be reduced by a factor of ~¼, which improves mobility and results in a 100-200% increase in transistor current[19] and dramatic performance gains. As a result, strained Si is being implemented in nearly all 90 nm, 65 nm, and 45 nm technology nodes. The results have been so successful that it will be difficult for alternative high-mobility channel technologies to compete with strained Si. The adoption of strained Si while keeping the gate length constant still supports the historical transistor density increase and cost reduction (at least initially) since the gate dimension is only a fraction of the transistor pitch (i.e. transistor density is increased by reducing the space between transistors and not the gate length). Maintaining a constant gate length (while scaling the space and pitch) is viable for a few technology nodes, but will eventually lead to scaling being limited by parasitic resistance and capacitance because of the space between transistors becoming too small.

Real limiters to scaling: parasitic resistance and capacitance

With the end of planar Si transistor scaling in sight, it is now possible to give an insight into the real limits on the Si roadmap, which are the parasitic resistance and capacitance generally assumed negligible by scaling theories. This was a good assumption for the past 40 years but will no longer be the case during the next decade. Fig. 4 shows the various parasitic resistances and capacitances present in a planar MOSFET. The sudden rise in parasitics at the end of the roadmap can be qualitatively understood as resulting from the space between neighboring devices decreasing to tens of nanometers since the source/drain and contact size need to be aggressively scaled to support the increased density in the absence of gate length scaling. Fig. 4 depicts the typical design rules of planar Si transistors for the 32 nm technology node. It can be seen that the source/drain contact and the gate are only tens of nanometers apart, which is undesirable in terms of parasitic resistance. Such small contact size leads to higher contact resistance and contact-to-gate capacitance. Historically, the parasitics did not matter since they were much smaller than the channel resistance and capacitance. However, the intrinsic channel capacitance and resistance has decreased dramatically during the past four decades. To the first order, channel resistance and capacitance are both proportional to the gate dimension, which has decreased ~1000 times since the start of Moore's law. Because of such dramatic reductions in channel resistance, the parasitic resistance and capacitance are now becoming comparable and are on course to becoming even larger than the intrinsic device resistance and capacitance (see Figs. 5 and 6). Figs. 5 and 6 were obtained using industry design rules for the 90 nm and 65 nm technology nodes, 0.7 times scaling for future nodes, and equations found in

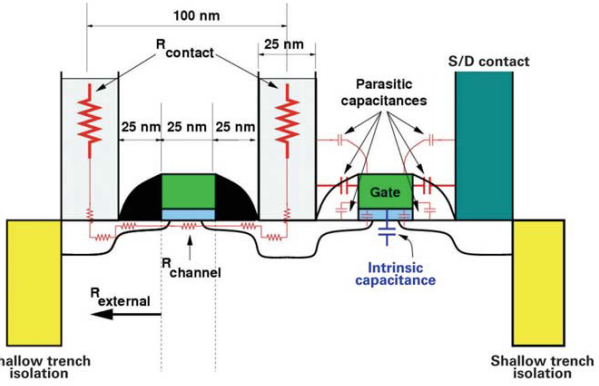

Fig. 4 Planar CMOS schematic showing the various parasitic resistances and capacitances (drawn to scale for the 32 nm process technology node, transistor pitch = 100 nm).

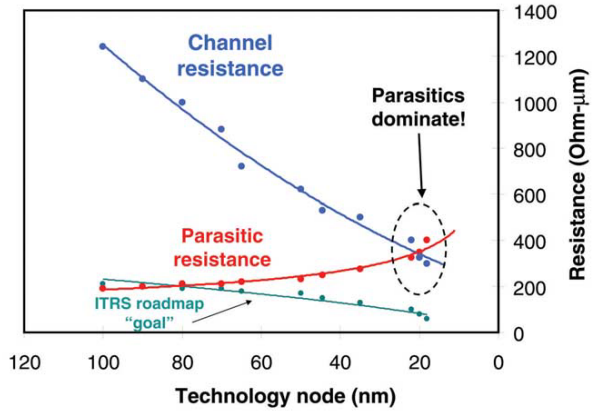

Fig. 5 Estimated total planar CMOS parasitic and channel resistance versus technology node. Note parasitic resistance is on course to become larger than the intrinsic channel resistance.

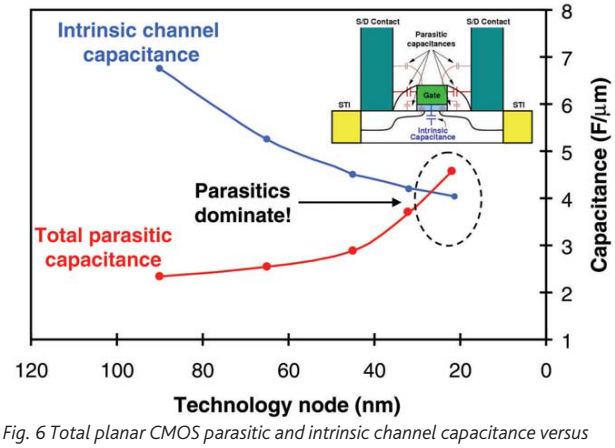

Fig. 6 Total planar CMOS parasitic and intrinsic channel capacitance versus technology node. Note parasitic capacitance is on course to become larger than intrinsic channel capacitance.

references[26-32]. Slot or trench contacts are assumed for the 45 nm technology node and beyond to help minimize contact resistance.

The parasitic scaling trends highlight an interesting area where future transistor research needs to be focused and important metrics for future devices. Much of present day transistor research is focused on new channel materials and devices for very high mobility. The scaling trends suggest the real limiter is parasitics and that this is where the focus needs to be. Once the channel resistance becomes smaller than the external resistance, reducing the channel resistance further, even to zero, has little performance benefit (the same is true for capacitance). Furthermore, and perhaps more importantly, it highlights that new device structures need to be judged on parasitic resistance and capacitance more than on channel transport properties. To date, parasitics are much worse for most new device options and rarely is this taken into account. Take, for example, the recent logic circuits demonstrated in carbon nanotube transistors. Although the channel mobility is several orders of magnitude higher than Si, the fabricated inverters[33-35] are ~10^6-10^{10} times slower than state-of-the-art CMOS because of parasitics! Also, in fully depleted multigate devices fabricated on Si fins, the parasitic resistance is worse than in the planar CMOS transistor[23,24]. For both fully depleted and carbon nanotube transistors, the high parasitic resistance is likely to be a fundamental problem because of the small source/drain volume (the source/drain contacts connect to a Si bulk 'box' that is just 5-10 nm thick in the case of fully depleted devices; in carbon nanotube devices this is even less, with the thickness on the order of one atomic layer).

Prospects to replace Si electronics

With Si CMOS scaling limits in sight, the obvious questions are "What nanotechnology is on the horizon to replace planar Si CMOS transistors and in what time frame could this happen?" Though these are difficult and perhaps foolhardy questions to try and answer, it is important to attempt to do so since this affects a $300 billion worldwide industry and the careers of most engineers.

Addressing the time frame question first, how quickly the industry can adopt a radically different device type depends on how many research and development levels it impacts. Levels are defined by development efforts and organizations that are currently present in the microelectronics industry such as materials, device design, circuit design, computer architecture, and software. As a rule of thumb, past changes that affect one level generally take approximately five or more years with the exception of software, which takes even longer. Examples of recent changes that mostly affected one level and took five years or more are bipolar to planar CMOS transistors, strained Si, high-*k* gate dielectrics, low-*k* back end dielectrics, and computer architecture changes such as hyper threading and the move from single thread performance to dual core microprocessors. Many of the non-Si device types affect multiple research and development levels with most

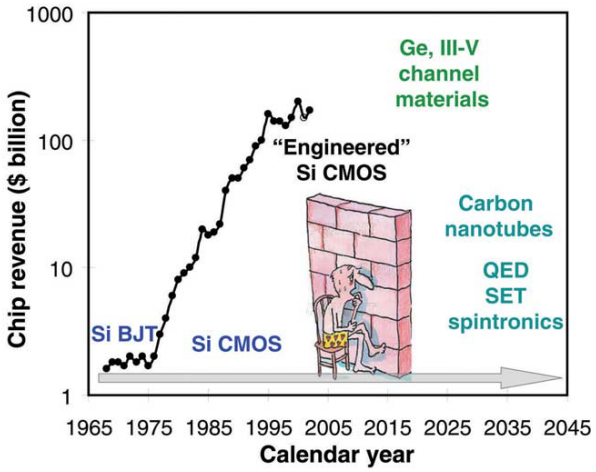

Fig. 7 Si technology industry time line showing semiconductor revenue and the evolution of new devices and possible time frame for new device types (BJT – bipolar junction transistor, QED – quantum electronic device, and SET – single electron transistor).

radical new devices affecting all levels. Changes that affect multiple levels can easily take five years per level and are very difficult. For example, a new device based on single-electron transport or spintronics will likely require a new product architecture and software to take advantage of their unique device attributes (high density and low power) and limitations (small transistor drive current and inability to drive the six to ten layers of Cu interconnects used to wire the 100s of millions to billion of transistors in modern chips). Thus, a radically new device type could easily take over 15-20 years to coordinate the changes among all the levels once the industry has decided to pursue this approach. Since the industry is still more than a decade away from making a decision on which new device type to pursue, it puts the time frame for a radical new nanotechnology device for mainstream logic applications more than 30 years away. When the industry talks about radically new device types that will "revolutionize computing", this type of time perspective is often missed. Fig. 7 attempts to put in perspective the time frame required to implement some of the new radical device types into production.

The second part of the question is what non-Si technology can potentially replace the Si planar device. This is perhaps an even more difficult question to answer but equally important since resources for a society (both people and capital) need to be focused on the best areas for return. Even with these caveats, some conclusions can be drawn. First, if the industry is going to spend several decades at a development cost likely to be greater than $100 billion to get a technology to a level competitive with the already > $1 trillion invested in Si technology during the past 40 years, the new technology is going to have to offer at least an order of magnitude improvement over the planar Si CMOS transistors. For noncharge-based devices relying on radically different computing models, at this time it is difficult to predict the potential

performance advantages. However, for non-Si charge-based devices (Ge, III-V, or carbon nanotube channels), which are viewed as the first possible replacements to the planar Si transistor, it is now becoming possible to estimate the potential improvement over Si. To assess the potential performance, the first order metric often used is the magnitude of the channel mobility in the new device types compared with the surface channel Si electron and hole mobility. Though commonly used this benchmark is, unfortunately, too simplistic and in many cases leads to wrong conclusions. First, high mobility can result from either low conductivity mass or scattering. Only reduced conductivity mass (as opposed to reduced scattering) is important for ballistic-limited transport, which dominates at the end of the Si roadmap. Second, and equally important, is the density-of-states for holes and electrons in the channel material since this affects the amount of inversion charge and the device drive current (both amount and velocity of charge determines the current). Most of the very high mobility materials (for example carbon nanotubes) have very low density-of-states (several orders of magnitude lower than Si) making it difficult to achieve high drive currents. In fact, even in the strained Si technology currently in production, strain is used to create both a low conductivity mass in the channel direction and a high density-of-states by creating a very large conductivity mass in the plane of the transistor perpendicular to the channel direction[25]. Including these considerations, is not clear if any of the new channel materials offer significant improvement over Si. Furthermore, based on the previous discussion of parasitic resistance, few new channel materials can compete since a great deal of development has gone into reducing external resistance for Si CMOS devices by using millisecond anneals to create above solid solubility doping and incorporating nickel silicide (NiSi) and SiGe in the source and drain.

Conclusion

Based on current technology trends, the scaling limits to planar CMOS are clear. Though simple scaling of planar CMOS has ended, the technology will continue approximately on Moore's historical performance trend for another few technology nodes making planar CMOS the leader in commercial nanotechnology. The real limits to planar Si CMOS (and most new device types) are parasitic resistance and capacitance. Thus, alternate device types should be benchmarked on parasitics and not just channel transport properties. Furthermore, radically new device types will require changes along many research and development levels from materials to software. The time frame to implement a radically new device is estimated to be ~30 years. Thus Si CMOS will be the dominant form of nanotechnology for the foreseeable future.

Finally, we must remember Moore's law is not a physical law but a law about economics. Consumer products and emerging markets have become the dominate end markets for semiconductors and will continue to be so for the next decade (versus the military in the 1960s-70s). A key attribute in these markets is price. Si technology is on course in the next decade to offer a billion transistor chips for ~$1, which will be a very difficult technology to displace. mt

REFERENCES

1. Bardeen, J., and Brattain, W. H., *Phys. Rev.* (1949) **75**, 1208
2. Kilby, J. S., and Keonjian, E., *Tech. Dig. IEDM* (1959) **5**, 76
3. Kilby, J. S., *IEEE Trans. Electron Devices* (1976) **23**, 648
4. Ranade, P., *et al., Tech. Dig. IEDM* (2005), 211
5. Oishi, A.., *et al., Tech. Dig. IEDM* (2005), 229
6. Ohta, H., *et al., Tech. Dig. IEDM* (2005), 237
7. Thompson, S.E., *et al., IEEE Trans. Semicond. Manufacturing* (2005) **18**, 26
8. Pearton, S. J., *et al., IEE Proc. Circuits Devices Systems* (2005) **152**, 312
9. Frunzio, L., *et al., IEEE Trans. Appl. Supercond.* (2005) **15**, 860
10. Mann, C. C., *Technology Review* (Nov 2004), 60
11. Moore, G. E., *Electronics* (1965) **38**, 114
12. Moore, G. E., *Tech. Dig. IEDM* (1975) **21**, 11
13. Dennard, R. H., *et al., IEEE J. Solid-State Circuits* (1974) **9**, 256
14. Bai, P., *et al., Tech. Dig. IEDM* (2004), 657
15. Chatterjee, A., *et al., Tech. Dig. IEDM* (2004), 665
16. NSET (Feb 2000), www.nsf.gov/crssprgm/nano/reports/omb_nifty50.jsp
17. LaPedus., M., Industry not ready for 450-mm, says Applied. *EE Times*, Jan 2006, www.eetimes.com/news/semi/showArticle.jhtml?articleID=175803153
18. Lin., B. J., *Tech. Dig. IEDM* (2005), 48
19. Thompson, S. E., *et al., Tech. Dig. IEDM* (2002), 61
20. Thompson, S. E., *et al., IEEE Trans. Electron Devices* (2004) **51**, 1790
21. Lammers, D., FinFETs may never come in, IBM fellow claims. *EE Times*, Dec 2005, www.eetimes.com/showArticle.jhtml?articleID=174903289
22. Lammers, D., Little time, big decisions for 45-nm node. *EE Times*, June 2005, www.eetimes.com/news/semi/showArticle.jhtml?articleID=164900707
23. Dixit, A., *et al., IEEE Trans. Electron Devices* (2005) **52**, 1132
24. Wu, W., and Chan, M., *IEEE Electron Device Lett.* (2006) **27**, 68
25. Thompson, S. E., *et al., IEEE Trans. Electron Devices* (2006) **53**, 1010
26. Kim, S-D., *et al., IEEE Trans. Electron Devices* (2002) **49**, 457
27. Kim, S-D., *et al., IEEE Trans. Electron Devices* (2002) **49**, 467
28. Ng, K. K., and Lynch, W. T., *IEEE Trans. Electron Devices* (1987) **ED-34**, 503
29. Mohapatra, N. R., *et al., IEEE Trans. Electron Devices* (2002) **49**, 826
30. Mohapatra, N. R., *et al., IEEE Trans. Electron Devices* (2003) **50**, 959
31. Sakurai, T., and Tamaru, K., *IEEE Trans. Electron Devices* (1983) **ED-30**, 183
32. Suzuki, K., *IEEE Trans. Electron Devices* (1999) **46**, 1895
33. Bachtold., A., *et al., Science* (2001) **294**, 1317
34. Appenzeller., J., *et al., Phys. Rev.* (2002) **89**, 126801
35. Chen, Z., *et al., Science* (2006) **311**, 1735