Recent Advances in High Performance CMOS Transistors: From Planar to Non-Planar

by Suman Datta

More than 40 years later, Gordon's Moore accurate observation[1]—that the number of transistors in an integrated circuit doubles roughly eighteen months—continues to be the guiding principle of the semiconductor industry. We have almost taken for granted the apparent corollary: as transistor count increases, each transistor becomes smaller, faster, and cheaper. Today, the transistor gate length (L_G) in production is approximately 28 nanometers. Further geometric scaling of conventional silicon MOSFET devices faces many fundamental challenges, such as: excessive gate leakage current, exponentially increasing source to drain sub-threshold leakage current, gate stack reliability and channel mobility degradation from increasing electric field rising dynamic power dissipation (CV^2f) from non-scaled supply voltages, band-to-band tunneling leakage at high body doping levels, device to device variation from random dopant fluctuation effects, and high source-drain access resistance from scaled contact areas limiting on-current. In order to continue and maintain the pace of energy efficient transistor scaling, it is imperative to scale the supply voltage of operation concurrently. However, the supply voltage scaling has slowed in recent years due to two fundamental reasons: (1) lower operating electric field results in both lower carrier density and lower carrier velocity and, hence, less transistor drive current; and (2) lower threshold voltage results in exponentially rising leakage power. Materials scientists and device/process integration engineers jointly have responded to this challenge by aggressively introducing new materials and modifying transistor structures as shown in Fig. 1. In this article we review some of the recent innovations in CMOS transistors that have allowed us to maintain the historical pace of delivering higher transistor performance with increasing energy efficiency.

High-k/Metal Gate Transistors

Until earlier this decade, the scaling and performance trends of the logic transistor were mainly the result of SiO_2 gate oxide scaling as well as source–drain junction and channel doping engineering. Over a period of 15 years, the physical gate-oxide thickness was scaled from about 20 nm to only 1.2 nm in the 65 nm technology node. The nitrogen profile in the 1.2 nm thick nitride silicon oxide was carefully optimized to ensure that the dopant penetration from the doped polysilicon gate electrode was prevented by the nitrogen barrier and yet the scattering of carriers at the interface between the gate oxide and the channel was mitigated. Further, various innovative gate oxide annealing and surface cleaning technologies were also introduced to improve the reliability and reduce defect

(continued on next page)

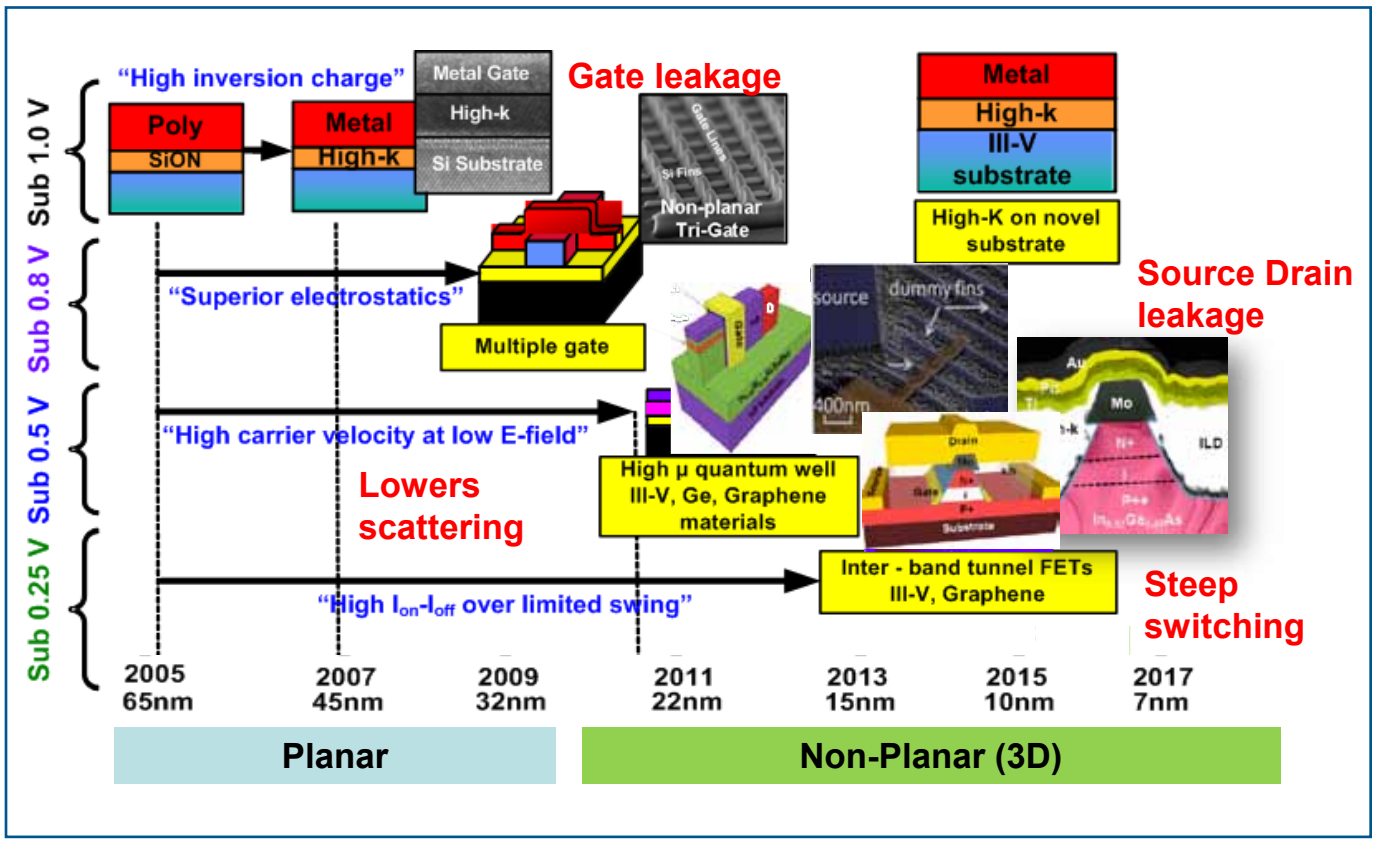

Fig. 1. *Trend in state-of-the-art high performance (HP) CMOS transistor innovation. Transformative changes in materials (high-k dielectric, Ge, III-V channel) and the transistor architecture (3D, Tunnel FET) being implemented and explored to maintain historical rate of performance, density and power scaling.*

density in the ultra-thin nitride silicon dioxide. As the thickness of the SiO_2 gate oxide reached 1.2 nm, which is less than five atomic layers thick, the industry *ran out of atoms* to scale the gate oxide further, since any more reduction in physical oxide thickness will make the gate leakage—due to quantum mechanical tunneling of electrons and holes through the gate oxide—unacceptable for circuit operation and overall power consumption. To solve this problem, an alternative high-k gate dielectric is needed. The higher permittivity (k) of the high-k dielectric allows the device engineer to achieve the same or even lower electrical oxide thickness with a physically thicker dielectric than silicon dioxide and reduce the gate leakage. Early materials research on high-k dielectrics identified hafnium dioxide (HfO_2) due to its excellent thermodynamic stability on silicon as well as sufficient conduction and valence band offsets with silicon.[2] Characterization of transistors with high-k dielectrics and polysilicon electrodes indicated that, while gate leakage can be mitigated by replacing SiO_2 with HfO_2, the soft optical phonons associated with the polarizable hafnium oxygen bonds induced significant reduction of the electron mobility and reduced transistor performance.[3] Researchers had to substitute the polysilicon gate electrodes with metal gates and were able to dynamically screen the longitudinal soft optical phonon modes arising from the high-k dielectric and recover a portion of the channel mobility.[4] Further enhancement of the channel mobility in high-k/metal gate transistors came from channel strain engineering where process induced uniaxial tensile strain for NMOS and compressive strain for PMOS was employed to demonstrate high performance high-k/metal gate CMOS transistors.[5] In addition, metal-gate electrodes with the correct work functions were required to achieve the correct n- and p-channel MOS transistor threshold voltages. Since the work functions of the metal gate electrodes are sensitive to the thermal budget, the process integration engineers had to modify the transistor process flow and introduce a replacement metal gate (RMG) or gate-last scheme to ensure that the metal gate electrodes maintain their work functions throughout the transistor fabrication process. Figure 2 summarizes key research milestones leading up to the introduction of high-k/metal gate transistors into mainstream technology.

Multiple Gate Transistors

While strained silicon and high-k/metal-gate technologies will continue to play significant roles in advancing present CMOS technology, the need for further scaling of the transistors will require the transistor structure itself to evolve. For example, a transition for the present planar structure to non-planar, three-dimensional (3D) structures such as the tri-gate transistor,[6] as shown in Fig. 3, is urgently needed to improve the short-channel performance and enhance scalability. This transition is inevitable since the high-k/metal gate stack with even low equivalent oxide thickness (EOT) is insufficient to control the source to drain leakage once the gate length is aggressively scaled and the source-drain extension regions approach each other. The tri-gate transistors, by design, are fully depleted so that the entire available silicon underneath the gate electrode is depleted of carriers before the threshold condition is reached. Such tri-gate transistors have shown significantly improved electrostatics in terms of sub-threshold slope (SS) and drain induced barrier lowering (DIBL) and hence better scalability than planar transistors. The key knob controlling the short channel effects in 3D transistors are given by the

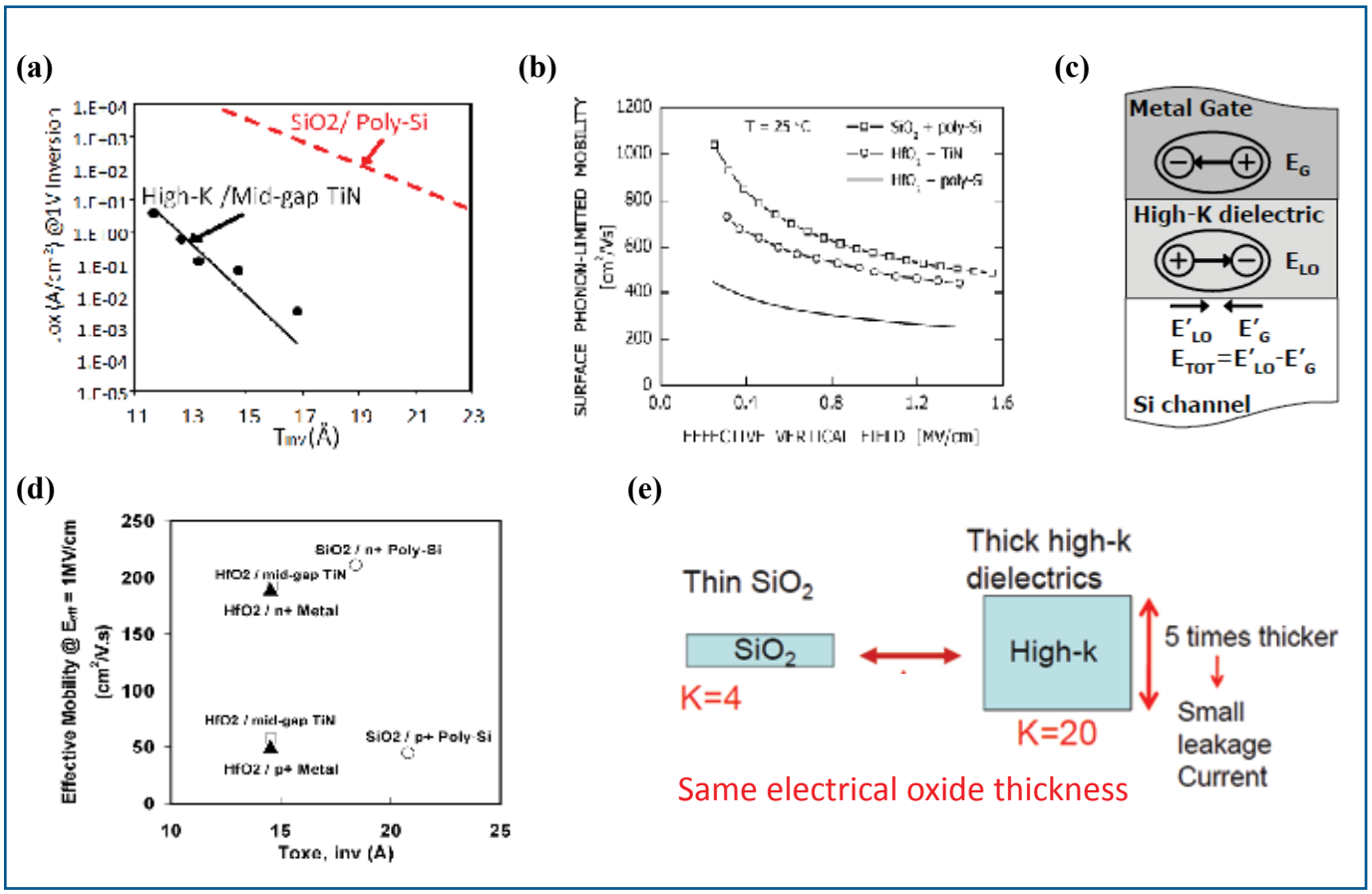

Fig. 2. *(a) Significant reduction in gate leakage achieved by replacing SiO_2 with high-k dielectric. (b) Remote soft optical phonon scattering arising from the polarizable bonds in high-k dielectric/poly-Si gate stack reduces the phonon limited mobility. (c) Metal gate can dynamically screen electrons in the high-k dielectric and improve mobility. (d) Work function engineered gate electrodes with correct NMOS and PMOS threshold voltages integrated using replacement metal gate process flow also exhibit acceptable electron and hole mobilities.[4] (e) Same electrical oxide thickness is achieved with 5 times thicker physically thicker HfO_2 dielectric compared to SiO_2 dielectric.*

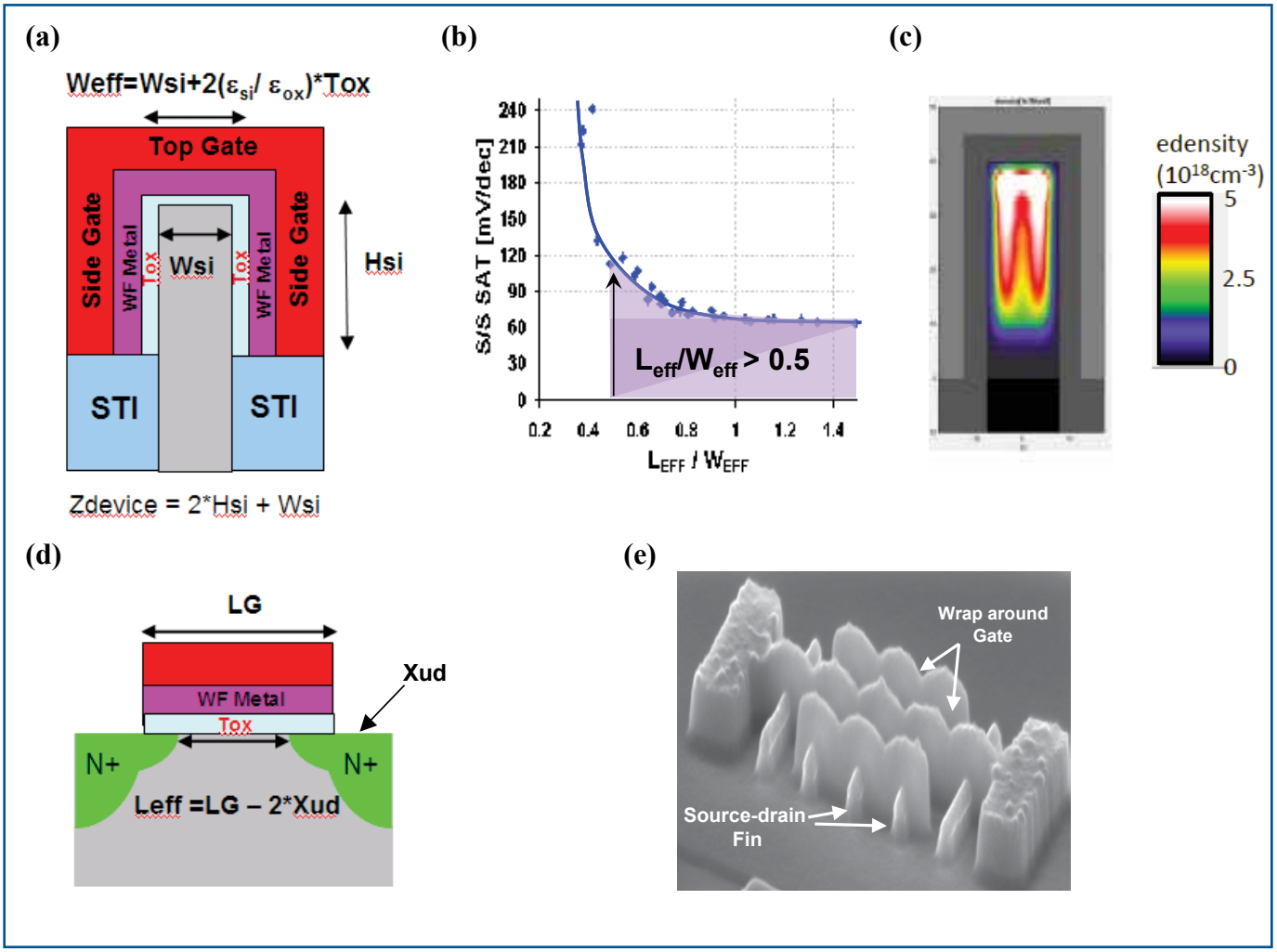

Fig. 3. *(a) Transition from planar to non-planar 3D transistors is motivated by need to control short channel effects. Scaling of the fin width ensures fully depleted operation of the transistor and mitigates the short channel effects. (b) The effective channel length to fin width ratio needs to be greater than 0.5 to maintain short channel effects with very little doping concentration in the channel. (c) Lower doping results in volume inversion and improves mobility at low bias region. (d) Transistor schematic illustrating the effective channel length. (e) Tilted view scanning electron microscopy (SEM) picture of 3D Tri-gate transistor featuring three parallel fins and one active wrap around gate.*

effective channel length (L_{eff}) and the effective fin width (W_{eff}) of the device. Device designers design electrostatically well controlled 3D transistors by keeping the effective channel length to fin width ratio greater than 0.5 as illustrated in Fig. 3. This also allows the device designers to reduce the channel doping which, in turn, reduces the impurity scattering the channel, enhances volume inversion effect and results in higher performance, particularly at lower bias region. Device and process integration engineers were required to re-engineer the process induced strain in 3D transistors due to the densely packed fins with ultra narrow width. Researchers found that a process induced vertical compressive strain improves the electron mobility on the sidewall surface with (110) crystal orientation for 3D NMOS, while the conventional embedded SiGe source drain stressor can significantly improve the hole mobility on the 110 sidewall for 3D PMOS. The combined benefits of the tri-gate CMOS transistor architecture with strained-silicon channels, high-k gate dielectric, metal-gate electrode, and epitaxial grown raised source–drains have been demonstrated, and the resulting CMOS transistors show excellent short-channel characteristics with high drive-current performance.[7] These results demonstrate that the benefits of various innovations can be combined to extend and continue the CMOS scaling and performance trends.

Ballistic Channel Transistors

While high-k/metal gate non-planar 3D transistors will continue to be the work-horse of advanced CMOS in the leading edge and future technology nodes, much interest has been generated and good progress has been made in the research of non-silicon electronic materials to replace the silicon channel for future logic applications, and their potential integration onto the silicon platform. Among the most studied materials in the form of planar quantum-well transistors are Ge, III-V compound semiconductors such as InSb[8] and InGaAs.[9] Figure 4 shows the experimental room temperature electron and hole mobilities measured in these planar quantum-well transistors. MOSFETs with high mobility materials are of interest for reduction of supply voltage of operation such that high drive current, I_{ON}, can be achieved with low overdrive voltage, V_{ON}-V_T. High mobility materials suffer from low density of states due to lower effective mass of the carriers and, hence, need detailed characterization. Preliminary device results show that at equivalent supply voltage of operation at or below 0.5 V, the higher injection velocities in planar high mobility III-V FETs compensate for the lower carrier density and provide higher performance than their Silicon MOSFET counterparts. The advantage of planar III-V FETs over Si MOSFETs was demonstrated early on in the embodiment of HEMTs.[12] However, for logic applications, not only high quality high-k dielectrics need to be integrated with III-V channels to mitigate gate leakage, but also non-planar device configurations need to be employed to ensure scalability to sub-20 nm gate length and beyond. Recently, there have reports on experimental demonstration of

(continued on next page)

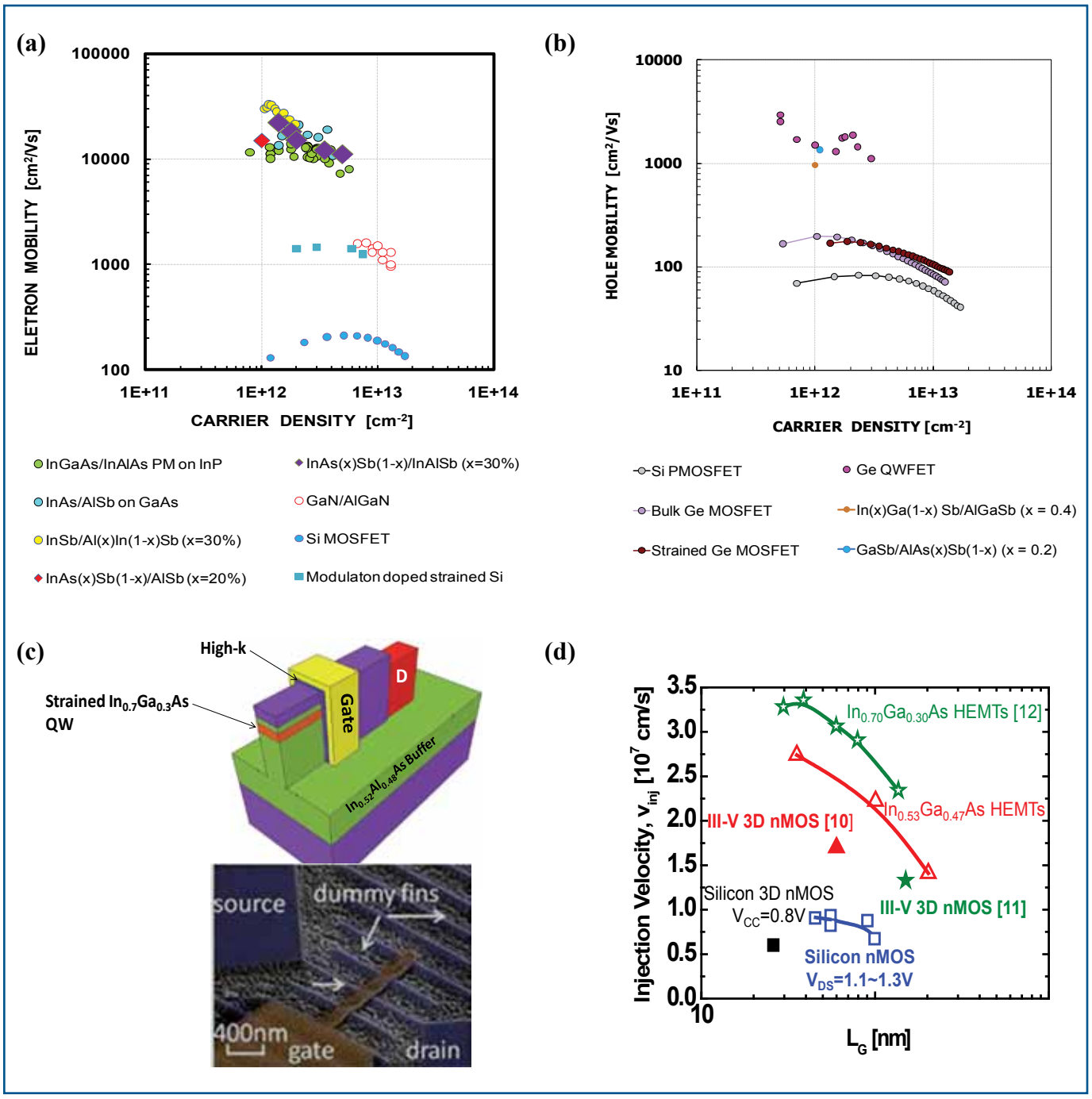

Fig. 4. *Room temperature (a) electron and (b) hole mobilities in Ge and compound semiconductors benchmarked against silicon CMOS. (c) Schematic and scanning electron micrograph of fabricated non-planar 3D transistors incorporating InGaAs (indium gallium arsenide) quantum well channel. (d) Experimentally measured injection velocities of non-planar III-V 3D transistors benchmarked against their Si counterpart. The 3D III-V transistors show higher injection velocities at equivalent gate length while operating at lower supply voltage than Si NMOS.*[10,11]

non-planar III-V MOSFETs, albeit at longer gate lengths than today's advanced 3D Si MOSFETs, which show much enhanced electron injection velocities.[10,11] Various challenges remain to make high mobility, ballistic channel MOSFETs a reality. These challenges include integration of III-V layers selectively on large silicon wafers using a high throughput, manufacturable growth technique such as MOVPE, demonstration of highly reliable high-k dielectric compatible with III-V, demonstration of a high performance pFET compatible with III-V nFET and a viable proves integration scheme, demonstration of the ballistic channel n and pFETs at a highly restricted footprint to justify their insertion at the 7 nm node or beyond.

Tunnel Transistors

As we approach the 7 nm technology node and beyond, the ever increasing transistor count on single chip will require aggressive supply voltage scaling to reduce device level energy consumption in order to stay below the chip level power budget.

(continued on page 46)

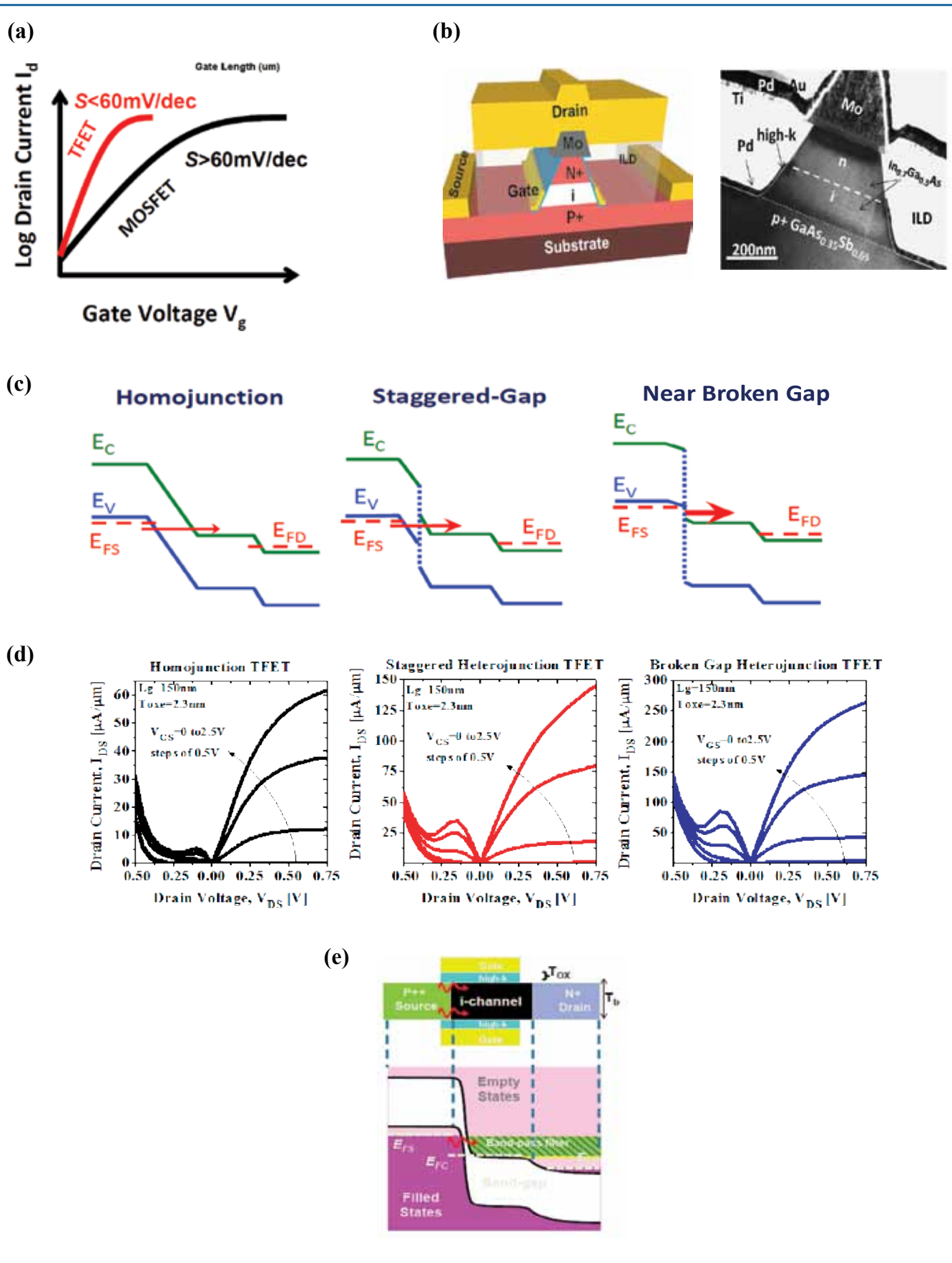

Fig. 5. *(a) Energy band diagram of Tunnel FET (TFET) showing steep switching transfer characteristics. (b) Schematic and cross-section of vertical 3D nanopillar TFET. (c) Schematic band diagram of homojunction, staggered gap and near broken gap heterojunction TFETs. (d) Experimental output characteristics of homojunction, staggered-gap and nearly broken gap heterojunction TFETs showing significant increase in drive current.[15] (e) Energy band diagram of Tunnel FET.*

(continued from page 44)

This warrants the researchers to explore possibilities of steep slope transistors which can operate at very low supply voltages and yet with high on-current to off-current ratios. Recently, non-planar 3D Tunnel FETs have been proposed to implement such transistors.[13] Unlike conventional MOSFETs, the Tunnel FET (TFET) architecture employs a gate modulated Zener tunnel junction at the source which controls the transistor ON and OFF states. This scheme fundamentally eliminates the high-energy tail present in the Fermi-Dirac distribution of the valence band electrons in the p+ source region and allows sub-kT/q steep slope device operation near the OFF state. This allows Tunnel FETs to achieve a much higher ION−IOFF ratio over a small gate voltage swing. A major challenge in the demonstration of high performance Tunnel FET is the limited rate of tunneling across the Zener junction which results in low drive current. Figure 5 illustrates the vertical Nanopillar Tunnel FET architecture used in the fabrication of the TFETs. Nanopillar Tunnel FET provides several features not accessible in conventional lateral device geometry. It allows incorporation of an asymmetric source drain configuration within the transistor structure where the source region composition can be markedly different from those of the channel and drain regions. This is vital for high performance Tunnel FET which requires a heterojunction source, abruptly doped source and channel tunnel junction, precise alignment of the gate edge with the source-channel tunnel junction, ultra-thin body and double gate or surround gate configuration, but requires suppression of gate induced channel to drain tunneling. Recently, we have experimentally demonstrated III-V Nanopillar Tunnel FETs using both homojunction and heterojunction tunnel source regions.[14,15] The results show, for the first time, that the on-current bottleneck in Tunnel FETs can be overcome by careful bandgap engineering, and highlight that high performance and low power Tunnel FETs could be practical in the near future.

Conclusion

The last decade has witnessed tremendous innovation in transistor architecture with the introduction of strained silicon channel, high-k/metal gate stack and non–planar 3D transistor architecture, marking the end of the era of the traditional planar transistor scaling. A power limited era has begun where new materials and new switching mechanisms need to be embraced with the framework of 3D transistors to continue the relentless forward march of technology in shrinking transistors and integrating more functionality on silicon to produce ever higher- performance and more energy efficient computational and memory devices. The natural scaling process will eventually lead us to the realm of reduced dimension materials and quantum engineered devices where both difficult technical challenges and golden opportunities co-exist. To overcome the challenges, research on new nanodevice structures, different device usage models, novel electronic materials and their integration on silicon, are required. Going forward, research on nanoelectronics addressing such fundamental issues to enable high-performance and energy-efficient ULSI applications are going to be more exciting and rewarding than ever. ■

Acknowledgment

The author would like to acknowledge contributions from his colleagues at Intel Corporation for the research on high-*k*/metal gate and multiple gate CMOS transistors. The author would also like to acknowledge contributions from his current and former graduate students at Penn State University for the research on ballistic channel and tunnel transistors.

About the Author

Suman Datta is Professor of Electrical Engineering at the Pennsylvania State University, University Park. He is exploring new materials, novel nanofabrication techniques, new classical and non-classical device structures for CMOS "enhancement" and CMOS "replacement" for future energy efficient, high performance and information processing systems. Datta is a Fellow of IEEE and a Distinguished Lecturer of the IEEE Electron Devices Society. He received the 2012 IBM Faculty Award, the 2012 Penn State Engineering Alumni Society Outstanding Research Award, the 2003 Intel Achievement award. He is the author of over 140 refereed publications and holds 145 US patents related to advanced CMOS technology. He can be reached at sdatta@engr.psu.edu.

References

1. G. Moore, *Electronics,* **38**, 144 (1965).
2. D. Schlom, S. Guha, and S. Datta, *MRS Bulletin,* **33**, 1017 (2008).
3. S. Datta, G. Dewey, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, M. Metz, N. Zelick, and R. Chau, *International Electron Devices Meeting (IEDM) Technical Digest*, 28.1.1 (December 2003).
4. R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz, *IEEE Electron Device Letters,* **25**, 6, 408 (2004).
5. K. A. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C.-H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. Mclntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, K. Zawadzki, *International Electron Devices Meeting (IEDM) Technical Digest*, 247 (December 2007).
6. B. S. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, *IEEE Electron Device Letters,* **24**, 4, 263 (2003).
7. J. Kavalieros, B. S. Doyle, S. Datta, G. Dewey, and R. Chau, *VLSI Technology Symposium Digest of Technical Papers*, 62 (June 2006).
8. S. Datta, T. Ashley, J. Brask, L. Buckle, M. Doczy, M. Emeny, D. Hayes, K. Hilton, R. Jefferies, T. Martin, T. Phillips, D. Wallis, P. Wilding, and R. Chau, *International Electron Devices Meeting (IEDM) Technical Digest*, 763, (December 2005).
9. R. Chau, S. Datta, M. Doczy, et al., *IEEE Transactions on Nanotechnology,* **4**, 2, 153 (2005).
10. L. Liu, V. Saripalli, V. Narayanan, and S. Datta, *IEEE International Electron Devices Meeting (IEDM)*, Washington DC (December 2011).
11. M. Radosavljevic, B. Chu-Kung, S. Corcoran, G. Dewey, M. K. Hudait, J. M. Fastenau, J. Kavalieros, W. K. Liu, D. Lubyshev, M. Metz, K. Millard, N. Mukherjee, W. Rachmady, U. Shah, and R. Chau, *IEEE International Electron Devices Meeting (IEDM)*, Washington DC (December 2011).
12. J. del Alamo, *Nature,* **479***, 7373*, 317 (2011).
13. S. Mookerjea, D. Mohata, R. Krishnan, J. Singh, A. Vallett, A. Ali, T. Mayer, V. Narayanan, D. Schlom, A. Liu, and S. Datta, *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, 949 (December 2009).
14. D. K. Mohata, R. Bijesh , S. Mujumdar, C. Eaton, R. Engel-Herbert, T. Mayer, V. Narayanan, J. Fastenau, D. Loubychev, A. Liu, and S. Datta, *IEEE International Electron Devices Meeting*, Washington DC (December 2011).
15. D. K. Mohata, R. Bijesh, Y. Zhu, M. K. Hudait, R. Southwick, Z. Chbili, D. Gundlach, J. Suehle, J. M. Fastenau, D. Loubychev, A. K. Liu, T. S. Mayer, V. Narayanan, and S. Datta, *IEEE Symposium on VLSI Technology, Honolulu, Hawaii (June 2012).*