

Why BiCMOS and SOI BiCMOS?

by T. H. Ning

Silicon technology development is at a crossroads, following an exponential rate of progress for more than thirty years. While CMOS (complementary metal-oxide-semiconductor) will remain the backbone of digital logic, silicon technology will evolve in directions driven by system needs that are not met by CMOS alone. It is argued that BiCMOS (bipolar complementary metal-oxide-semiconductor), particularly SOI (silicon-on-insulator) BiCMOS, will emerge as an important technology platform for mixed-signal systems. The bipolar transistors will be used primarily for the rf (radio-frequency) and analog functions, while the CMOS will be used for most digital functions. The SOI substrate provides dc and ac isolation, as well as opportunities for further device and circuit innovation.

Introduction

Silicon technology has been the backbone of microelectronics and electronics systems for more than thirty years now. It is apparent that this will remain the case for the foreseeable future. In terms of productivity, the density of devices on a silicon chip has been following Moore's law, doubling roughly every two years since about 1980. **Figure 1** shows the prediction made by Gordon Moore in 1975 [1] of the exponential increase in the density of transistors on chips vs. time. This improvement in density has been enabled primarily by a combination of progress in lithography and innovation in device technology.

Over the years, there have been several major directions in the evolution of silicon digital logic technology. The most prominent ones are listed in **Figure 2**. It is clear that CMOS has been and remains the dominant technology because of its low standby-power dissipation and scaling properties. However, every technology has its limits. In 1975, Moore suggested that a significant slowdown in device and circuit innovation will cause Moore's law to change from density doubling every year to density doubling every two years, as illustrated in **Figure 1**. Thus, it should come as no surprise that, after more than two decades of very rapid progress in scaling, CMOS is finally approaching its scaling limits. In this paper, we consider the characteristics of CMOS near its scaling limits and examine whether BiCMOS will reemerge as an important silicon technology platform.

CMOS near its limits

CMOS for digital applications

CMOS scaling has been driven primarily by digital applications, in which the objectives are relatively simple—smaller, faster, and lower-power devices, preferably at lower cost per circuit as well. It has been shown by simulation that bulk CMOS transistors with channel lengths of about 25 nm can be designed to function well [2]. Recently, CMOS devices with channel lengths close to such small dimensions have been demonstrated.¹ They show that as CMOS is scaled toward its limiting dimensions, it remains excellent for high-performance digital applications. However, they also show

¹ Examples of reported sub-50-nm CMOS devices can be found in the 2000 *IEDM Technical Digest*, pp. 43–68.

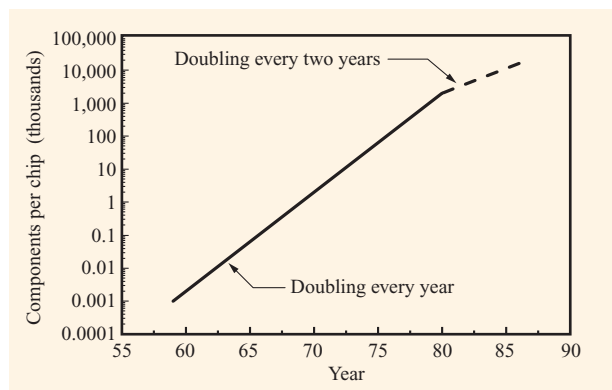


Figure 1

Moore's 1975 projection of the trend in silicon chip complexity. Adapted with permission from [1]; © 1975 IEEE.

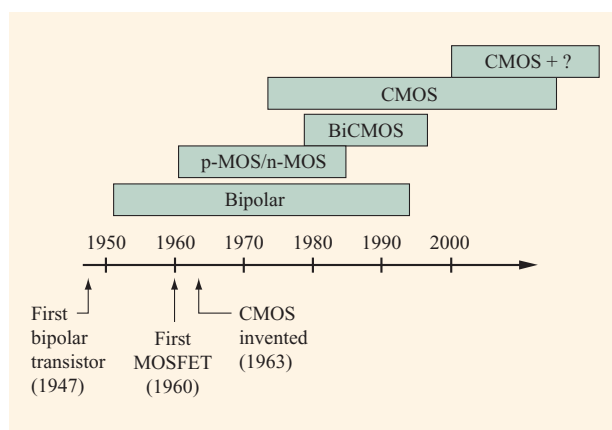


Figure 2

Evolution of silicon logic device platform.

that this performance comes at the expense of much higher device off-current, about $100 \text{ nA}/\mu\text{m}$ at 25°C , significantly compromising the low-standby-power attribute of CMOS. Also, in order to achieve large drive currents, the devices typically use ultrathin gate oxide (1.5 nm or thinner), resulting in gate tunneling current density as large as $10\text{--}100 \text{ A}/\text{cm}^2$. The implication of such a large gate tunneling current for circuit performance and reliability is not clear. What is clear is that we have seen CMOS near its limits.

There have been focused efforts in the industry to develop CMOS device structures with performance potential better than bulk CMOS [3, 4]. Thus far, CMOS on silicon-on-insulator (SOI) has been most successful [5].

More in-depth discussion of the potentials of SOI and other novel ideas for improving CMOS performance can be found elsewhere in this special issue.

CMOS for rf and analog applications

A 25-nm CMOS device is expected to run about three times as fast as a 100-nm (channel length) device, and it will have an f_T (the cutoff frequency at which the current gain is unity) of more than 250 GHz [2]. It should be excellent for high-speed digital circuits. The very high f_T suggests that it should be excellent for high-frequency rf circuits as well. However, the series resistances associated with the source/drain and gate, together with the characteristic short-channel effects associated with very-short-channel CMOS devices, which are discussed further below, limit the attainable $f_{\max}$ [3, 6–8]. This can be seen from the expression for $f_{\max}$ (the maximum oscillation frequency at which the power gain is unity) [9],

$$f_{\max} = \frac{1}{2} \frac{f_T}{\sqrt{\frac{(r_g + r_s)}{r_o} + 2\pi f_T r_g C_{gd}}}, \quad (1)$$

where r_g is the series resistance of the gate electrode, r_s is the series resistance of the source electrode, r_o is the output resistance of the transistor, and C_{gd} is the gate–drain capacitance. Currently there is great interest in developing CMOS with reduced parasitic resistance, including using a metal as the gate electrode. With a metal gate and careful layout to minimize parasitic resistance, the attainable $f_{\max}$ could approach 100 GHz [8].

For low-noise analog circuits, the gate leakage current level should be less than $10^{-8} \text{ A}/\text{cm}^2$ [10]. Thus, CMOS devices near the scaling limit are not suitable for low-noise analog circuits. Also, analog circuit designers prefer to design with power-supply voltages much larger than that of high-speed digital CMOS. This should not be a practical problem because most, if not all, advanced CMOS ASIC (application-specific integrated circuits) processes are designed to be compatible with the input/output voltages of the previous one or two technology generations. This is accomplished by having devices of thick and thin gate oxides on the same chip. The thick-oxide devices can be used for analog and for any special circuits requiring higher voltages and lower gate tunneling currents. Of course, the thick-oxide devices are slower than the thin-oxide ones.

A large output resistance is desirable for many analog circuit applications [7, 8, 11]. However, high-speed CMOS devices usually have a small output resistance, as illustrated in **Figure 3**. The small output resistance is a direct result of the fact that short-channel devices have a large drain-induced barrier-lowering effect which causes

the drain current to increase with drain voltage [12]. As can be seen from Equation (1), a small output resistance also means reduced $f_{\max}$. When large output resistance is needed, designers often use device channel lengths significantly larger than minimum for the technology. Thus, there is a tradeoff between speed and output resistance in CMOS devices, making scaled high-speed CMOS devices undesirable for many analog circuits.

Bipolar needed

It is true that after years of intensive research and development effort on analog and rf CMOS, more and more rf and analog functions are being implemented in CMOS. CMOS is simply the low-cost solution. However, because of the limitations of CMOS devices, designers often use bipolar transistors for many analog and rf circuits in which high performance and/or low power dissipation are required [13, 14]. In most cases, the bipolar transistors are on a separate chip, and the technology could be either GaAs heterojunction bipolar transistor (HBT) or SiGe-base bipolar transistor.

In a regular bipolar transistor, the output resistance is proportional to the amount of dopant in the base. On the other hand, the collector current, and hence the current gain, is inversely proportional to the amount of dopant in the base [12]. Therefore, there is a tradeoff between the output resistance and the current gain.

Current gain can also be increased by suppressing the base current, which depends on the emitter parameters [12]. In the case of a GaAs HBT, the relatively large emitter energy bandgap suppresses the base current so much that the base can be doped quite heavily ($>10^{19} \text{ cm}^{-3}$) to yield low base resistance and high output resistance while still maintaining adequate current gains. In the case of a SiGe-base bipolar transistor, the emitter is the same as a regular Si-base bipolar transistor, namely a polysilicon emitter. As a result, a SiGe-base bipolar transistor normally has the same base current characteristics as a regular Si-base bipolar transistor. However, the Ge profile in the base can be engineered to simultaneously improve the current gain (typically by about 5 $\times$) and the output resistance (typically by greater than 20 $\times$) [12]. **Figure 4** shows the typical current-voltage characteristics of a small-dimension SiGe-base bipolar transistor [15]. It shows good current gain and large output resistance.

Bipolar transistors are fast. As shown in **Figure 5**, SiGe-base bipolar ECL (emitter-coupled logic) circuits can easily have switching delays of less than 10 ps [16]. Compared to the SiGe-base bipolar transistor, the GaAs HBT is inherently faster because of its smaller emitter-base junction capacitance and smaller base resistance. Also,

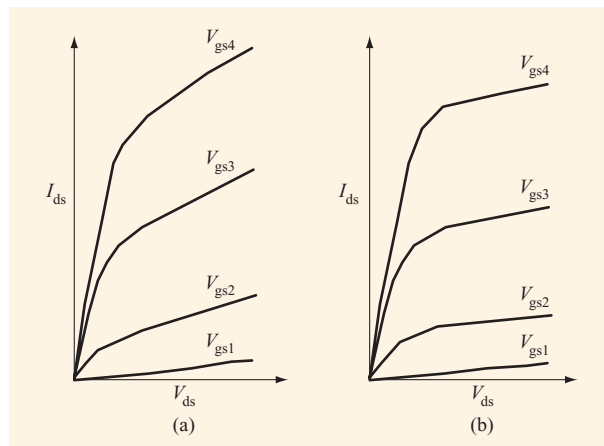


Figure 3

(a) Typical I - V characteristics of a high-speed digital MOSFET. The drain current increases significantly with drain voltage, resulting in a small output resistance. (b) Desired MOSFET I - V characteristics for many analog circuit applications. The output resistance is relatively large.

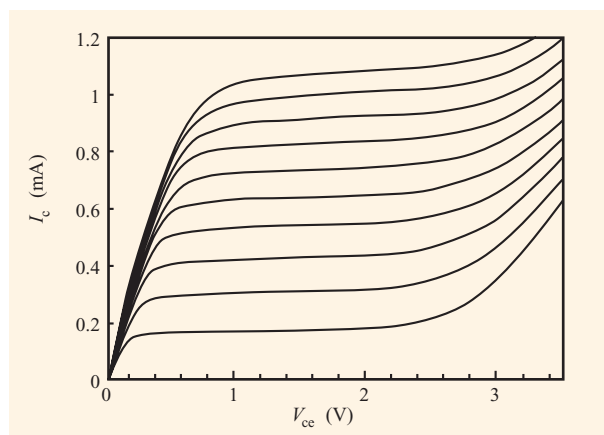


Figure 4

Typical I - V characteristics of a SiGe-base bipolar transistor. Emitter area = $0.25 \times 0.75 \mu\text{m}^2$, and base current = $2 \mu\text{A}/\text{step}$. The collector current is relatively independent of collector-emitter voltage, implying a large output resistance. Reprinted with permission from [15]; © 1999 IEEE.

GaAs HBTs can be scaled to higher collector current densities than silicon bipolar transistors because of the larger energy bandgap of GaAs [17]. However, for many applications, both GaAs HBT and SiGe-base bipolar can meet the requirements. SiGe-base bipolar is preferred either because it is less expensive or because it is integrated with CMOS.

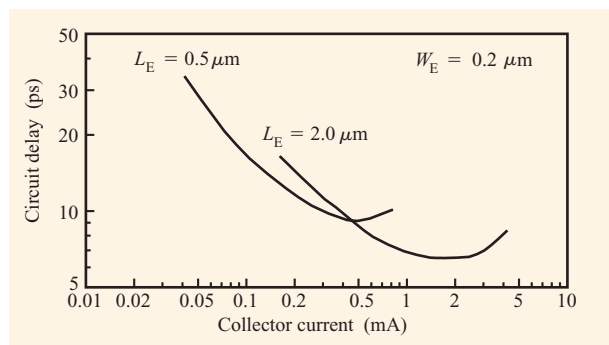


Figure 5

Typical switching delay of self-aligned SiGe-base bipolar ECL circuits. W_E is the emitter stripe width and L_E is the emitter length. Adapted with permission from [16]; © 1999 IEEE.

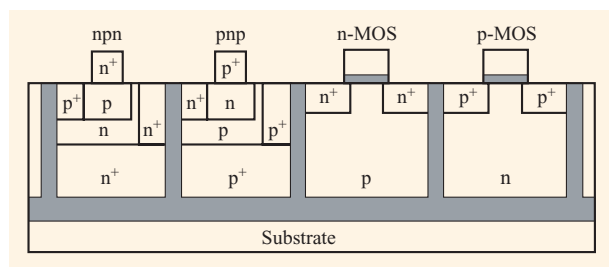


Figure 6

Schematic illustration of an SOI complementary bipolar-plus-CMOS technology.

Why BiCMOS?

As indicated in Figure 2, BiCMOS for digital logic applications was actively developed between the late 1970s and the early 1990s. It was a time when CMOS was a 5-V technology. The logic swing of a CMOS circuit is the same as its power-supply voltage, while the logic swing of a bipolar circuit is typically only 400 mV. At 5 V, the CMOS logic swing is more than ten times that of a bipolar circuit, resulting in CMOS circuits that are much inferior to bipolar circuits for digital logic applications. The idea behind the development of BiCMOS for digital logic applications was to combine the characteristics of high-speed bipolar and low-power CMOS on the same chip. Of course, designers combined bipolar and CMOS devices in novel circuit configurations as well.

However, as CMOS designers began scaling down power-supply voltage in earnest in the early 1990s [18, 19], the performance gap between bipolar and CMOS for digital logic began to narrow. As a result, the need to add bipolar transistors on a CMOS chip in order to improve

the chip or system performance diminished. Since the early 1990s there have been very few reports on developing BiCMOS for digital applications.

The rapid growth of the wireless and communication markets in recent years has rekindled the interest in BiCMOS technology. This time, it is generally recognized that CMOS remains the backbone for building the digital portion of the systems, and bipolar devices are used primarily for the rf and analog functions which, as discussed earlier, are often beyond the capability of CMOS devices.

The often-cited major concern with BiCMOS, as compared to combining CMOS and bipolar chips on a package, is cost. However, the cost factor must be evaluated at the system or final product level. Integrating as many components on a chip as possible has been the goal of integrated-circuit development from the beginning. The reasons are obvious—integration leads to smaller and faster systems that dissipate less power and are more reliable. Thus, there is a natural tendency to integrate the CMOS and bipolar transistors which are required to build the systems on the same chip. As long as using BiCMOS, instead of packaging bipolar devices and CMOS devices, can result in more competitive products, BiCMOS will be an important silicon technology platform. This integration trend gives SiGe-base bipolar a tremendous advantage over GaAs HBT, since SiGe-base bipolar and CMOS can be readily fabricated on the same chip. SiGe-base BiCMOS technology is currently a topic of considerable interest for research and development.

Why SOI BiCMOS?

As discussed above, the drivers for BiCMOS technology are mixed-signal systems. For these systems, isolation of the noise generated in the digital part of the system from the rf and analog parts of the system is critical. A major noise coupling is through the wafer substrate. Great care in design and layout is exercised to minimize substrate-coupling noise, often using high-resistivity wafer substrates [20]. Using SOI, particularly SOI with a relatively thick buried-oxide layer, should minimize substrate-coupling noise. If necessary, a combination of SOI and high-resistivity substrate can be used [21].

With SOI as the starting wafer, the transistors on the chip can be isolated from one another with simple lateral oxide regions or trenches. This makes the integration of npn and pnp bipolar transistors (complementary bipolar) and CMOS devices on the same chip relatively straightforward, as illustrated conceptually in **Figure 6**.

In conventional bipolar technology, usually only a high-performance vertical npn bipolar transistor is available. The pnp transistor is usually made up of two back-to-back lateral p-n diodes. This lateral pnp transistor is very slow and has large parasitic resistance and capacitance.

Nonetheless, lateral pnp transistors are often used in bipolar analog circuits to improve circuit characteristics [11]. Figure 6 suggests that in SOI complementary BiCMOS, the pnp transistor can be a vertical transistor. A vertical pnp transistor could have good performance, perhaps within a factor of 2 of that of a vertical npn. The availability of a high-performance pnp should enable significantly improved bipolar analog circuits.

The lateral pnp transistor is usually not used in bipolar digital circuits at all because its performance is inferior to that of the vertical npn transistor. However, integrating high-performance pnp and npn transistors on the same chip has long been the goal in bipolar technology development [22, 23]. If SOI is not used, the collector of the vertical pnp transistor is connected to the p-type substrate of the starting wafer, and the pnp can be used only in the grounded common-collector mode. Even with this limitation, digital bipolar circuits using a combination of high-performance npn and pnp can have significantly smaller power–delay product than circuits using high-performance npn alone [23, 24]. It is reasonable to expect that truly isolated but integrated high-performance npn and pnp transistors, enabled by SOI, will provide digital circuit designers with innovation opportunities.

Perhaps the most convincing reason for SOI BiCMOS is the fact that SOI as a technology is now ready for broad applications. Driven primarily by high-end CMOS microprocessor applications, SOI technology has been developed [25]. It is natural to expand SOI to other applications that can benefit from this new technology. Wireless and communication systems, or any high-end mixed-signal systems, are likely drivers of SOI BiCMOS development.

Thus far, a few companies have reported research and development efforts on SOI BiCMOS [26–28] and on SOI complementary BiCMOS [29]. SiGe-base vertical npn bipolar transistors, CMOS devices, and passive elements have been successfully integrated on the same SOI chip [26]. Applications to communications [27] and to channel electronics for disk drives [28] have been reported.

Summary

For more than three decades now, silicon technology development has been progressing at an exponential rate, driven by the demand for systems that are ever faster and dissipate less power. For the digital part of these systems, these demands in turn drive CMOS development. Scaled CMOS has been, and is likely to remain, the dominant technology for digital applications.

Recently, there has been renewed interest in the development of BiCMOS, particularly SiGe-base BiCMOS, for mixed-signal system applications. The idea is that the CMOS devices will be used primarily for the digital functions, and the SiGe-base bipolar devices will be

used for the rf and analog functions. For these mixed-signal chips, noise isolation between the digital and the rf/analog parts is critical. SOI or SOI combined with a high-resistivity substrate [21] is advantageous for minimizing noise coupling through the wafer substrate.

Furthermore, SOI allows easy isolation of all of the devices on a chip. Once device isolation is not an issue, one can contemplate integrating various devices on the same chip. In particular, the integration of high-performance npn and pnp bipolar transistors and CMOS devices on the same chip will enable mixed-signal systems that provide higher performance and/or lower power dissipation. With the recent rapid growth of the wireless and communication system markets, research and development of SOI BiCMOS technology is likely to grow.

References

1. Gordon Moore, "Progress in Digital Integrated Electronics," *IEDM Tech. Digest*, pp. 11–13 (1975).
2. Y. Taur, C. H. Wann, and D. J. Frank, "25 nm CMOS Design Considerations," *IEDM Tech. Digest*, pp. 789–792 (1998).
3. H.-S. P. Wong, D. J. Frank, P. M. Solomon, C. H. J. Wann, and J. J. Welser, "Nanoscale CMOS," *Proc. IEEE* **87**, 537–570 (1999).
4. R. D. Isaac, "The Future of CMOS Technology," *IBM J. Res. & Dev.* **44**, 369–378 (2000).
5. G. G. Shahidi, A. Ajmera, F. Assaderaghi, R. J. Bolam, E. Leobandung, W. Rausch, D. Sankus, D. Schepis, L. F. Wagner, K. Wu, and B. Davari, "Partially-Depleted SOI Technology for Digital Logic," *Proceedings of the IEEE International Solid-State Circuits Conference*, 1999, pp. 426–427.
6. C. Wann, L. Su, K. Jenkins, R. Chang, D. Frank, and Y. Taur, "RF Perspective of Sub-Tenth-Micron CMOS," *Proceedings of the IEEE International Solid-State Circuits Conference*, 1998, pp. 254–255.
7. J. N. Burghartz, M. Hargrove, C. Webster, R. Groves, M. Keene, K. Jenkins, D. Edelstein, R. Logan, and E. Nowak, "RF Potential of a 0.18- μm CMOS Logic Technology," *IEDM Tech. Digest*, pp. 853–856 (1999).
8. T. Ohguro, H. Naruse, H. Sugaya, S. Nakamura, E. Morifuji, H. Kimijimi, T. Yoshitomi, T. Morimoto, H. S. Momose, Y. Katsumata, and H. Iwai, "High-Performance RF Characteristics of Raised Gate/Source/Drain CMOS with Co Salicide," *Symposium on VLSI Technology, Digest of Technical Papers*, 1998, pp. 136–137.
9. M. A. Hollis and R. A. Murphy, "Homogeneous Field-Effect Transistors," *High-Speed Semiconductor Devices*, S. M. Sze, Ed., John Wiley & Sons, Inc., New York, 1990, pp. 211–281.
10. *International Technology Roadmap for Semiconductors*, 1999 Edition, Semiconductor Industry Association, San Jose, CA; see http://member.itrs.net/Files/1999_SIA_Roadmap/Home.htm.
11. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 3rd ed., John Wiley & Sons, Inc., New York, 1993.
12. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, New York, 1998.
13. L. E. Larson, "Device and Technology Requirements for Next Generation Communication Systems," *ISSCC Digest of Technical Papers*, 2001, pp. 737–740.
14. X. Li, T. Brogan, M. Esposito, B. Myers, and K. K. O, "A Comparison of CMOS and SiGe LNA's and Mixers for Wireless LAN Applications," *IEEE Custom Integrated*

- Circuits Conference, Digest of Technical Papers, 2001, pp. 531–454.
15. G. Freeman, D. Ahlgren, D. R. Greenberg, R. Grove, F. Huang, G. Hugo, B. Jagannathan, S. J. Jeng, J. Johnson, K. Schonenberg, K. Stein, R. Volant, and S. Subbanna, "A 0.18- μm 90-GHz f_T SiGe HBT BiCMOS, ASIC-Compatible, Copper Interconnect Technology for RF and Microwave Applications," *IEDM Tech. Digest*, pp. 569–572 (1999).
 16. K. Washio, M. Kondo, E. Ohue, K. Oda, R. Hayami, M. Tanabe, H. Shimamoto, and T. Harada, "A 0.2- μm Self-Aligned SiGe HBT Featuring 107-GHz f_{max} and 6.7 ps ECL," *IEDM Tech. Digest*, pp. 557–560 (1999).
 17. T. H. Ning, "History and Future Perspective of the Modern Silicon Bipolar Transistor," *IEEE Trans. Electron Devices* **48**, 2485–2491 (2001).
 18. B. Davari, W. H. Chang, M. R. Wordeman, C. S. Oh, Y. Taur, K. E. Petrillo, D. Moy, J. J. Bucchignano, H. Y. Ng, M. G. Rosenfield, F. J. Hohn, and M. D. Rodriguez, in *IEDM Tech. Digest*, pp. 56–59 (1988).
 19. W. H. Chang, B. Davari, M. R. Wordeman, Y. Taur, C. C.-H. Hsu, and M. D. Rodriguez, "A High-Performance 0.25- μm CMOS Technology: I—Device and Characterization," *IEEE Trans. Electron Devices* **39**, 959–966 (1992).
 20. T. Ohguro, T. Ishikawa, T. Kimura, S. Samata, A. Kawasaki, T. Nagano, T. Yoshitomi, and T. Toyoshima, "High-Performance Digital–Analog Mixed Devices on a Si Substrate with Resistivity Beyond 1 k Ω -cm," *IEDM Tech. Digest*, pp. 757–760 (2000).
 21. A. Viviani, J. P. Raskin, D. Flandre, J. P. Colinge, and D. Vanhoenacker, "Extended Study of Crosstalk in SOI/SIMOX Substrates," *IEDM Tech. Digest*, pp. 713–716 (1995).
 22. S. K. Wiedmann, T. C. Chen, C. T. Chuang, K. Heuber, D. F. Wendel, J. Warnock, G. P. Li, K. Chin, and T. H. Ning, "Sub-300-ps CBL Circuits," *IEEE Electron Device Lett.* **10**, 484–486 (1989).
 23. T. Onai, E. Ohue, Y. Idei, M. Tanabe, H. Shimamoto, K. Washio, and T. Nakamura, "Self-Aligned Complementary Bipolar Technology for Low-Power Dissipation and Ultra-High-Speed LSI's," *IEEE Trans. Electron Devices* **42**, 413–418 (1995).
 24. C. T. Chuang, "Advanced Bipolar Circuits—A Perspective," *Proceedings of the International Symposium on VLSI Technology, Systems, and Applications*, 1991, pp. 394–397.
 25. G. G. Shahidi, "SOI Technology for the GHz Era," *IBM J. Res. & Dev.* **46**, 121–131 (2002, this issue).
 26. K. Washio, E. Ohue, H. Shimamoto, K. Oda, R. Hayami, Y. Kiyota, M. Tanabe, M. Kondo, T. Hashimoto, and T. Harada, "A 0.2- μm 180-GHz- f_{max} 6.7-ps-ECL SOI/HRS Self-Aligned SEG SiGe HBT/CMOS Technology for Microwave and High-Speed Digital Applications," *IEDM Tech. Digest*, pp. 741–744 (2000).
 27. S. Ueno, K. Watanabe, T. Kato, T. Shinohara, K. Mikami, T. Hashimoto, A. Takai, K. Washio, R. Takeyari, and T. Harada, "A Single-Chip 10-Gb/s Transceiver LSI Using SiGe SOI/BiCMOS," *ISSCC Digest of Technical Papers*, 2001, pp. 82–83.
 28. N. Fujii, M. Kuraishi, T. Mochizuki, S. Irikura, and T. Hirose, "A SOI-BiCMOS 800 Mbps Write Driver for Hard Disk Drives," *IEEE Custom Integrated Circuits Conference, Digest of Technical Papers*, 2001, pp. 451–454.
 29. J. A. Babcock, B. Loftin, P. Madhani, X. Chen, A. Pinto, and D. K. Schroder, "Comparative Low Frequency Noise Analysis of Bipolar and MOS Transistors Using an Advanced Complementary BiCMOS Technology," *IEEE Custom Integrated Circuits Conference, Digest of Technical Papers*, 2001, pp. 385–388.

Tak H. Ning IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (ningth@us.ibm.com). Dr. Ning received his Ph.D. degree in physics from the University of Illinois at Urbana–Champaign in 1971. He joined the IBM Thomas J. Watson Research Center at Yorktown Heights, New York, in 1973. Since 1991, he has been an IBM Fellow. He has made contributions to various areas of silicon devices and technology, including bipolar, CMOS, DRAM, SOI, EEPROM, and hot-electron effects. Dr. Ning has received numerous awards, including the 1991 IEEE Jack A. Morton Award and the 1989 IEEE Electron Devices Society J. J. Ebers Award. He is a member of the National Academy of Engineering and a Fellow of the IEEE and of the American Physical Society.

Received June 29, 2001; accepted for publication November 29, 2001