SOI technology for the GHz era

by G. G. Shahidi

Silicon-on-insulator (SOI) CMOS offers a 20–35% performance gain over bulk CMOS. High-performance microprocessors using SOI CMOS have been commercially available since 1998. As the technology moves to the 0.13-μm generation, SOI is being used by more companies, and its application is spreading to lower-end microprocessors and SRAMs. In this paper, after giving a short history of SOI in IBM, we describe the reasons for performance improvement with SOI, and its scalability to the 0.1-μm generation and beyond. Some of the recent applications of SOI in high-end microprocessors and its upcoming uses in low-power, radio-frequency (rf) CMOS, embedded DRAM (EDRAM), and the integration of vertical SiGe bipolar devices on SOI are described. As we move to the 0.1-μm generation and beyond, SOI is expected to be the technology of choice for system-on-a-chip applications which require high-performance CMOS, low-power, embedded memory, and bipolar devices.

Introduction

Deep submicron room-temperature bulk CMOS has been the main technology used for ULSI systems, and CMOS scaling has been the primary tool for improving system performance. Over the last three decades, SOI CMOS has been identified as one possible method for increasing the performance of CMOS over that offered by simple scaling [1]. Prior to the 1990s, SOI had not been suitable as a substrate for mainstream applications. The barriers to its widespread use were many, the main ones being SOI material quality, device design, and the steady progress in bulk CMOS performance through scaling.

In early 1989, the IBM Research Division initiated activity in SOI CMOS with a program focusing on device design and materials research. The results of this early work were the adoption of the partially depleted device design point, the demonstration of a 0.1-μm effective-channel-length (L_{EFF}) CMOS [2], its application to a 512Kb SRAM, significant progress in methodology to characterize the floating-body effects, and notable progress in SIMOX (Separation by IMplantation of OXygen) materials technology which was developed on 5-in. wafers. In the early 1990s, SOI development was transferred to the Advanced Silicon Technology Center (ASTC) of the IBM Microelectronics Division, with the goal of serious assessment of SOI CMOS and increased focus on 8-in. wafer development. In 1994, using a quasi-0.35-μm CMOS SOI technology (i.e., the same oxide thickness and polysilicon gate width as bulk CMOS, but with a lower operating voltage range), a "fully" functional PowerPC 601* was demonstrated over a limited shmoo window (graph identifying functional points in performance–voltage space). That demonstration prompted IBM to initiate a focused program to qualify a 0.25-μm CMOS SOI technology [3]. The qualification focused on the same development areas as a 0.25-μm bulk technology, including yield, reliability, SOI material, modeling and design manual, transfer to manufacturing, SOI-unique device structures (diodes, electrostatic discharge

0018-8646/02/$5.00 © 2002 IBM

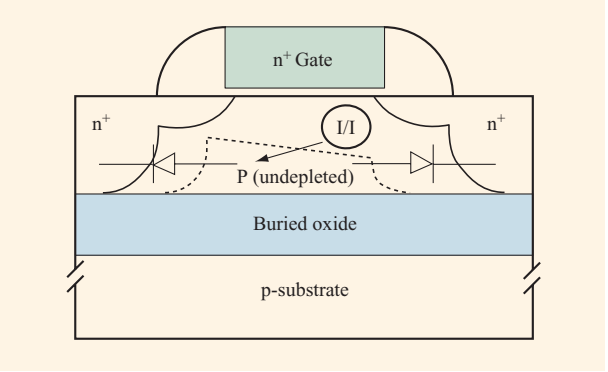

Figure 1

Cross section of n-FET on SOI, showing the body charging mechanisms. I/I = ion-implanted region.

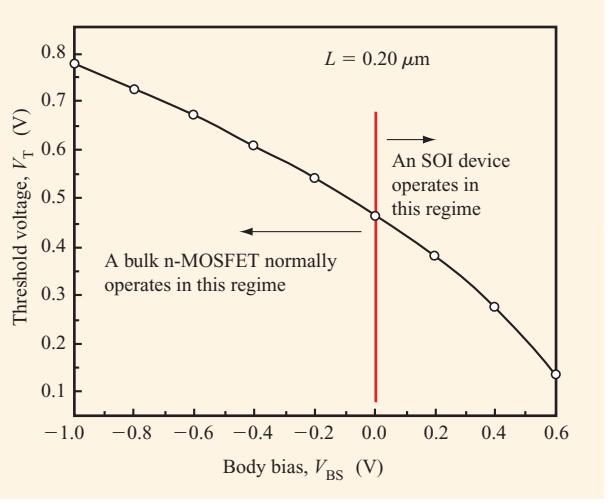

Figure 2

Threshold voltage as a function of body bias (i.e., the "body effect") for an n-FET with channel length $L = 20\ \mu$m.

protection), characterization of floating-body effects, test, and burn-in. However, no product group was committed to using SOI CMOS, because several elements were lacking: firm demonstration of performance, field use of SOI (i.e., an unproven technology), and the existence of a clear and improving bulk CMOS roadmap. On its own initiative, the SOI Technology Development Group modified a PowerPC 604* (a 32-bit PowerPC* microprocessor, and the first product incorporating bulk 0.25-μm CMOS, used by IBM and Apple Computer, Inc.). The SOI version of the PowerPC 604 was fully functional, booted IBM AIX* (the IBM version of the UNIX** operating system) and AppleOS (the Apple Computer operating system), was put through extensive test and reliability stress, and, most significantly, demonstrated the performance gain possible with SOI (27% for PowerPC 604) [3]. By the time SOI demonstrated these capabilities, high-end bulk CMOS technology was moving to a 0.22-μm generation, and SOI development was redirected to developing a 0.22-μm CMOS generation.

IBM server groups have always used the most advanced CMOS technology for their microprocessors. In 1997, they chose 0.22-μm CMOS SOI [4] for their next-generation microprocessors. That effort resulted in the first mainstream application of SOI CMOS, significant circuit learning, and its fabrication in a CMOS manufacturing line. The first SOI product, a 64-bit PowerPC microprocessor, was shipped in the summer of 1998. A limited number (a few tens of thousands) of 32-bit PowerPC 750* microprocessors, using 0.22-μm CMOS SOI, were also shipped to customers. Since then, the development has focused on 0.18-μm [5], 0.13-μm [6], and 0.1-μm [7] CMOS SOI technologies, with a greatly increased customer base.

Device design

The primary feature of MOS in SOI is that the local substrate ("body") of the device floats electrically, and therefore the substrate–source bias voltage, V_{BS}, is not fixed (**Figure 1**). As V_{BS} changes, the device threshold voltage, V_T, will change (**Figure 2**). This "instability" in V_T is what has made SOI device design very challenging. One manifestation of the threshold variation is the "kink effect," or increase in the output conductance of the device near drain-to-source bias, V_{DS}, of 1 V, the bandgap of silicon [1]. This is caused by the impact-ionization-induced increase in V_{BS} with increasing V_{DS}, and the resulting reduction of V_T; when V_{DS} becomes large enough, impact ionization current (holes) flows to the undepleted body, increasing the body charge and V_{BS}, resulting in a decrease in V_T. One method widely used to minimize floating-body effects is to use fully depleted (FD) SOI devices. In FD devices, the SOI film thickness is (much) smaller than the channel depletion width, and therefore the body charge is fixed. Any impact ionization charges (majority carriers) flowing into the depleted body are readily swept to the source because of the much-reduced potential barrier. For this reason, in the early stages of SOI technology development the focus was on FD SOI devices, since the kink effect and other floating-body effects such as dynamic V_T variation were considered serious problems. Initially a further benefit attributed to FD SOI was the improvement with respect to short-channel effects (SCE) [1]. **Figure 3** shows simulated V_T roll-off of bulk Si (top curve) and four SOI n-FETs with different film thicknesses. As the SOI film thickness is reduced, the V_T roll-off does improve, but the

improvement is only caused by the reduction in junction depth. As shown in **Figure 4**, if the junction depth in bulk Si is the same as the SOI film thickness, the V_T roll-off in bulk Si and SOI is the same [2]. It has been shown that for a given channel length, FD SOI devices exhibit increased short-channel effects (SCE) compared to partially depleted SOI unless the silicon film thickness becomes much smaller than the depletion depth [8].

In early work in IBM, it was quickly realized that FD SOI is not manufacturable. In sub-0.25-μm CMOS technologies, the control of SCE is critical. By adopting partially depleted (PD) device design, one can use many of the techniques developed for controlling SCE in bulk-Si CMOS, i.e., retrograde well, halo, and source and drain shallow extension [3]. Also, in PD SOI (because in general $V_{BS} > 0$) the SCE is decreased even in comparison to a bulk technology with identical doping. There are many other benefits associated with PD SOI: It is very difficult to design a high-V_T FD device, because if the film doping is increased in order to raise the V_T, the device is not fully depleted; it must be made thinner, and then the V_T decreases. In other words, it is difficult to have a device design point for FD SOI with reasonable off-state current, I_{OFF}. Nearly all sub-0.25-μm CMOS technologies have multiple V_Ts, which is difficult to achieve on FD SOI. Furthermore, it will be shown that dynamic V_T variation (present on PD SOI) is a significant cause of the performance boost of digital circuits in SOI. Two SOI-unique effects can affect the circuit functionality: "passgate" leakage [9] and the "history" effect [10]. The excess transient passgate leakage is actually higher in FD devices, since passgate leakage is strongly dependent on the bipolar gain of the device, and it is much easier to degrade the bipolar gain on PD SOI. The "history dependence" of propagation delay is larger in PD SOI, but as shown later, this is a manageable effect in most circuits. For the cases in which floating-body effects must be completely eliminated, it is possible to form an electrical body contact in PD SOI, but not in FD [11]. In summary, there are compelling reasons why the PD device design is preferable; these are summarized in **Table 1**.

Circuit design and floating-body effects

The key feature of PD SOI is the variability of V_{BS} and V_T. Under static steady-state conditions, V_{BS} is determined by the balance of currents through the two back-to-back diodes and impact ionization near the drain. Under dynamic switching conditions, V_{BS} depends on the previous electrical switching "history" of the device, as well as on the instantaneous node voltages [10–12]. The history effect (or the resulting delay variation) initially seems to be most daunting. During technology development, considerable effort was spent to minimize this effect (by balancing various leakage currents). The variation of delay as a function of switching history is about 8% per gate (and >20% in some passgate circuits initialized under the right conditions). This uncertainty in delay may seem excessive. However, there also is considerable uncertainty in the delay through a gate in a bulk-Si CMOS design (and the designers have developed techniques to take this

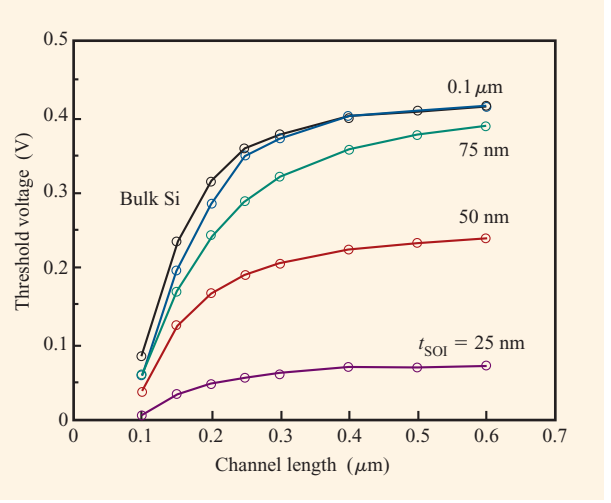

Figure 3

Threshold-voltage roll-off for bulk Si and SOI of various thicknesses, t_{SOI}, at a film doping of 2×10^{17} for 0.1-μm devices with gate-oxide thickness, t_{OX}, of 7 nm. Reproduced with permission from [2]; © 1994 IEEE.

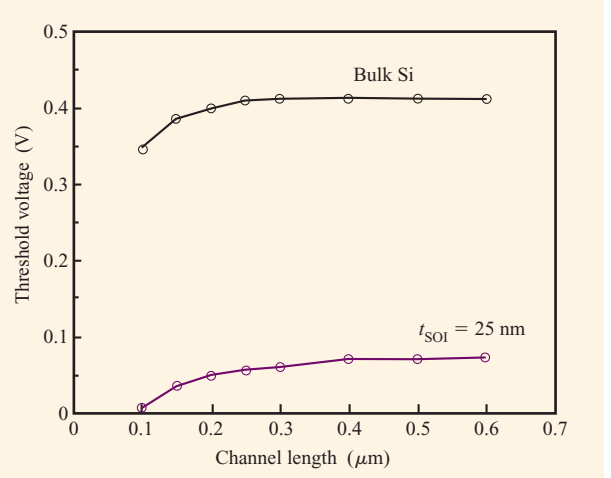

Figure 4

Threshold roll-off for n-FET on bulk Si with 25-nm junction depth and n-FET on 25-nm-thick SOI film, t_{SOI}, both with film doping of 2×10^{17} and gate-oxide thickness, t_{OX}, of 7 nm.

Table 1 Partially depleted SOI vs. fully depleted SOI.

	Partially depleted	*Fully depleted*
Manufacturability	+	
Design point (high V_T)	+	
Multiple V_T	+	
Breakdown voltage	+	
SCE	+	
Kink effect	+	
Body contact	+	
Passgate leakage	+	
History dependence		+

Table 2 Elements of technology and design giving rise to delay uncertainty.

Source of variation	*Variation in delay (%)*
Across-chip polysilicon width variation	15–20
Temperature (25–85°C)	10–20
Top vs. bottom switch in NAND	20–35
Number of 1s on an OR	10–40
On-chip V_{DD} variation (10%)	10
SOI history effect	8

into account). Among the phenomena contributing to uncertainty in a bulk-Si design (**Table 2**) are intrachip process variation (about 15–20% of gate delay); delay variation caused by top vs. bottom switching devices in NAND gates, or the number of simultaneous "1"s arriving at a NOR gate (about 20–35%); on-chip circuit supply voltage (V_{DD}) variations (about 10%), and temperature variations (on-chip and ambient, which can cause 10–20% change in delay). The SOI-induced uncertainty in delay, while not negligible, is no larger than other sources of uncertainty on a chip, and can be managed in a similar fashion [13]. At the microprocessor level, the history effect in SOI is much less noticeable [13–15] than the uncertainty in delay of a simple circuit block in SOI (i.e., an inverter or a NOR stage). The processor cycle time (or the chip frequency) is usually determined by a few cycle-limiting paths (i.e., the longest latch-to-latch delay). By initializing the chip at various conditions and then running the cycle-limiting path, one can measure the history effect at the chip level. This has been done on the IBM 64-bit PowerPC. Less than 3% variation due to the history effect at the chip level has been observed on cycle-limiting paths. More important, the cycle-limiting paths do not change as the chip initial condition is changed. Furthermore, the cycle-limiting paths were the same on the bulk-Si and SOI designs when the bulk-Si design was mapped to SOI. The reduced impact of history at the chip level is expected, since cycle-limiting paths pass through many gates, and some speed up and some slow down as their history is changed.

The other floating-body effect is "passgate leakage," or the reduction in the device V_T when the body charges to very high voltage, as high as V_{DD} [9]. A number of techniques have been developed to minimize the impact of passgate leakage [14, 16, 17], and most of the effort in mapping bulk-Si circuits to SOI is spent on fixing circuits affected by it [13, 15]. The challenge of circuit design in SOI is to properly screen every circuit type for passgate leakage (functionality, noise performance, and timing) [14, 16, 17].

The decision to migrate a design to SOI must balance the expected extra work required to enable the functionality of every circuit on SOI (and taking advantage of SOI) against remaining in bulk CMOS and having lower performance but an easier migration path. The experience has been that the first couple of re-maps of bulk designs to SOI were challenging. For example, the design teams had to discover the hard way the need for body contact on SRAM sense amplifiers and the need for grounding the back of the wafer. Once a design team is experienced, subsequent SOI designs are produced much more smoothly and quickly. An example is the 64-bit PowerPC design, which has been migrated into two generations of design within IBM. The first bulk-to-SOI map, using 0.22-μm CMOS SOI, was difficult. However, migration of the design to 0.18-μm SOI was much smoother and quicker [9, 13]: It required the same number of design revisions as a bulk-Si design that was mapped from one bulk generation to the next bulk generation.

Performance

The performance advantage of SOI over bulk Si is caused by the elimination of area junction capacitance, the lack of a reverse body effect in stacked circuits, and the fact that the SOI body is slightly forward-biased under most operating conditions. Considerable effort during the development of SOI CMOS was devoted to evaluating the performance of SOI against that of the best bulk-Si technology. Bulk Si has been compared to SOI at both the simulation and hardware levels. Simulation allows one to distinguish the various effects that give rise to the SOI performance advantage. At the simulation level, the main challenges have been the creation of a calibrated SOI model and the proper extraction of all of the parasitic components of the circuit that is being simulated.

To evaluate the effect of area junction capacitance on the performance gain, two critical paths (one with and one without global wire loading) were simulated. **Figure 5** shows the result for a number of technologies. The performance gain due to the reduction of the area

junction capacitance, C_J, is between 9% and 25%, and is greater for the path with no wire loading. Because of junction optimization in the 0.35–0.22-μm generations, the effect of C_J reduction has been diminishing. However, as we approach 0.18 μm and beyond, because of the need for high doping concentrations to control SCE in bulk Si, C_J will form a larger portion of the node capacitance.

The other cause of the performance gain of SOI over bulk Si is that the body of the device floats. Thus, not only is V_{BS} never less than zero (i.e., there is no MOS reverse body effect in stacked and passgate circuits), but $V_{BS} > 0$ under most operating conditions. To assess the importance of the body effect, critical paths of the microprocessor were simulated (**Figure 6**) using SOI models (0.22-μm and 0.18-μm CMOS SOI), with body floating (as in the usual operation), and tied to ground (GND) (n-FET) and V_{DD} (p-FET). As expected, leaving the body floating results in higher performance. Simulation was also carried out using the low-V_T models (lower V_T and 10× higher I_{OFF}) to assess the impact of the lower body effect, as in low-V_T devices, which are used in critical paths. As expected, for low-V_T devices, the importance of the body effect is diminished.

Assessing the advantage of SOI over bulk Si using hardware at the microprocessor level turned out to be significantly more complicated. SOI and bulk-Si designs are not identical, since they have different design margins. Moreover, it is very difficult to determine the L_{EFF} for a large microprocessor because of the variation of the polysilicon gate linewidth across the chip. There are also lot-to-lot variations in speed distributions even for the same nominal polysilicon linewidth. As a result, SOI and bulk-Si designs are usually centered at different polysilicon linewidths. In addition, the designs are at different levels of maturity. Nevertheless, for the 0.22-μm CMOS generations, two designs were fabricated in IBM Microelectronics over an extended period of time, allowing comparison of bulk-Si and SOI performance at the microprocessor level. **Figure 7** shows the f_{MAX} (i.e., the module maximum operating frequency) for the PowerPC 750, and **Figure 8** shows the (maximum operating) frequency for "Istar," a 64-bit PowerPC, vs. process sigma.[1] Performance gains of 22–33% were observed. In both cases, the same paths were the cycle-limiting paths (although there were some differences in the dependence of the maximum frequency of microprocessor operation on voltage). Simulation of a wide variety of critical paths indicates that performance gains of 20–35% are feasible when a design is moved from bulk Si to SOI.

[1] Sigma is the standard deviation of the distribution of the channel length (centered at the nominal polysilicon linewidth) of a given technology in the early stages of manufacturing. As technology matures and the distribution of the polysilicon tightens, the nominal polysilicon is shifted by a few (the original) sigma to a shorter polysilicon linewidth to increase the performance of the product.

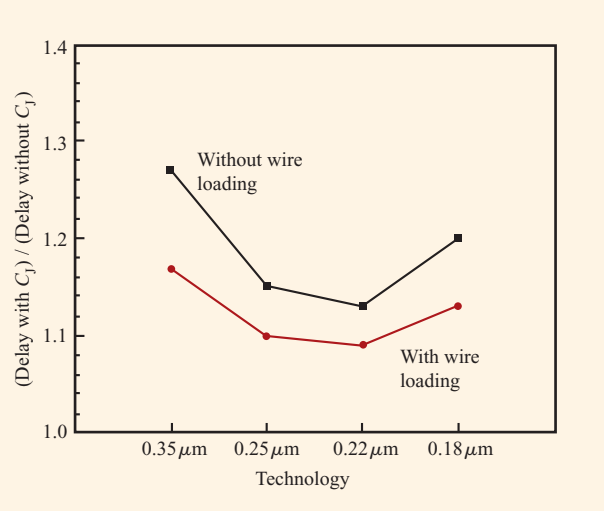

Figure 5

Simulated delay improvement of two critical paths of an IBM server, with and without junction capacitance, in different technologies.

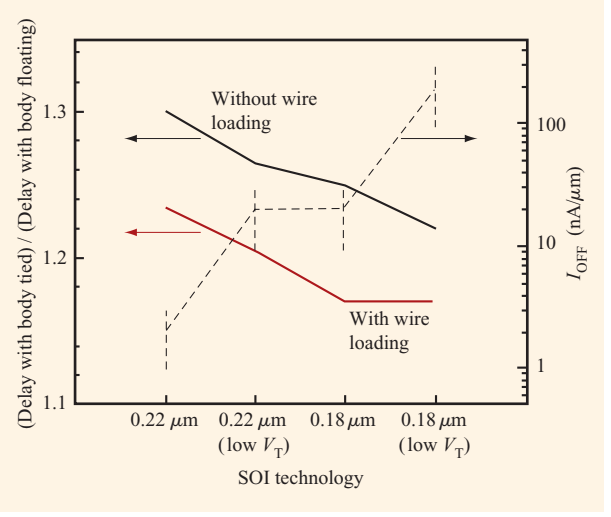

Figure 6

Simulated impact of the floating body on critical path delay for various SOI technologies, and the I_{OFF} for the corresponding technologies.

Extendibility

As SOI is scaled to 0.1 μm and beyond, one of the issues involved is SOI advantage and scalability. As the bulk MOS transistor is scaled, a number of trends are apparent. Channel doping must be increased (by 1.6× per generation) to counter the short-channel effects [18]. The

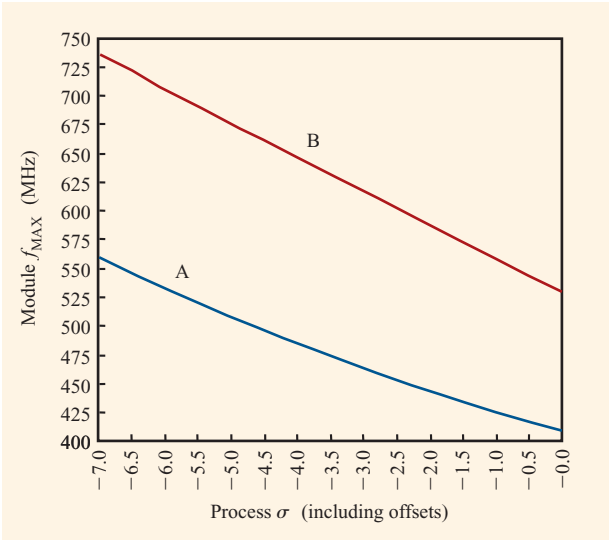

Figure 7

PowerPC 750 in 0.22-μm bulk (curve A) and SOI CMOS (curve B) technologies (2 V, 65°C).

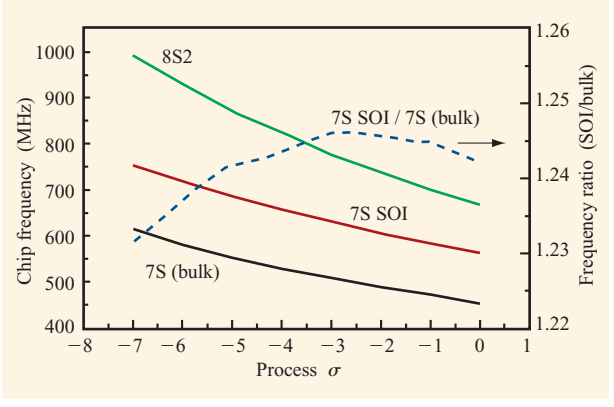

Figure 8

Performance, f_{MAX}, of the 64-bit PowerPC "Istar" in bulk and CMOS 0.22-μm technologies (7S), and 0.18-μm SOI CMOS (8S2) technology. The dashed curve shows the ratio of the performance of 7S SOI to that of 7S bulk.

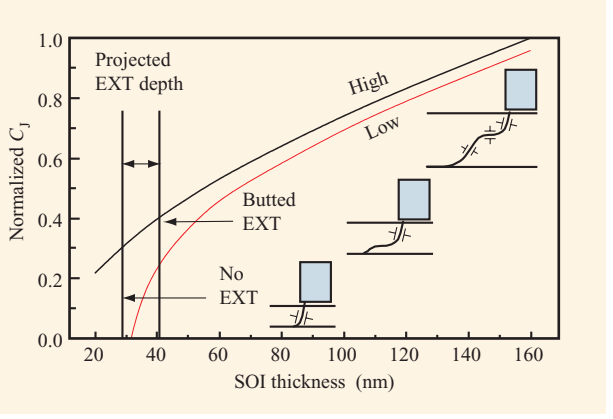

Figure 9

Simulated impact of SOI film thickness on total C_J. EXT = source–drain junction extension (shallow junction used to improve short-channel effect). High/low = range of junction capacitance values. The structures on the right of the figure are gate and junction schematics of a device on SOI. The device includes both the deep junction and the extension (top schematic). As film thickness is reduced, only the extension junction capacitance affects the device. Reproduced with permission from [12]; © 2000 IEEE.

device junction area is reduced by a scaling factor of 0.7×. Thus, the area junction capacitance per unit width, C_J, is expected to remain constant. The drop in C_J observed in the 0.25- to 0.18-μm generations has been due to one-time well optimization. Furthermore, as the microprocessor (and ASIC) chip frequencies are increased over the GHz range, the use of heavily wire-loaded stages will be diminished. Breaking up long wires and using repeaters actually improves the speed and noise performance. The use of larger (i.e., less wire-dominated) buffers amplifies the SOI advantage as it relates to C_J reduction.

As one moves to channel lengths below 0.1 μm, channel doping (and doping in halo devices [19]) must be increased by more than 1.6× per generation [11]. The use of SOI has made it possible to reduce the perimeter junction capacitance (as compared to bulk CMOS) by using thinner SOI (**Figure 9**) [12]. Indeed, using thin SOI films allows one to use higher doping concentrations in the halos, thus reducing SCE [7]. Going to thinner films has a noticeable impact on device performance. This is shown in **Figure 10**, based on the results of our second-generation 0.13-μm SOI CMOS, where the ring performance improves as the film thickness is reduced [20].

Body-effect scaling is more complicated. The increase in channel doping and the drop in V_{DD} worsen the impact of the reverse body effect. However, as the device is scaled, the I_{OFF} target has been increased by about 10× per generation (due to reduction in V_T) to obtain sufficient performance gain (Figure 6). The lowered V_T reduces the SOI advantage as it relates to body effect. The I_{OFF} target for the 0.13-μm generation is in the range of one hundred nA/μm (and for low V_T, I_{OFF} is 10× higher). It remains to be seen whether the rapid increase in I_{OFF} (and lowering V_T) can continue as scaling continues to below 0.1 μm.

By careful balancing of C_J and the channel capacitance, the history effect of SOI will not grow past ~7% [18]. In fact,

Table 3 Characteristics of IBM 64-bit PowerPC on bulk and SOI (0.22-μm and 0.18-μm generations).

	Bulk 0.22 μm	*SOI 0.22 μm*	*SOI 0.18 μm*
Core clock frequency (MHz)	450	550	660
L1 caches	128KB inst. + 128KB data	128KB inst. + 128KB data	128KB inst. + 128KB data
L2 directory	104 × 16K	104 × 16K	146 × 16K
Supply voltage (V)	1.8	1.8	1.5
Transistors	34M	34M	44M
Die size (mm^2)	139	139	128
Power (W)	22	24	18
L_{eff} (n-FET) (μm)	0.12	0.12	0.08
T_{OX} (nm)	3.5	3.5	2.6
Metallization	6 layers Cu	6 layers Cu	7 layers Cu
Contacted M2–M4 pitch (μm)	0.81	0.81	0.63

we do not observe much history-effect degradation of our 0.1-μm CMOS SOI technology at nominal or reduced voltages.

Applications of SOI

The IBM Server Group has been the first design group to adopt SOI technology across their product line. **Table 3** summarizes the first- and second-generation "Star" series 64-bit PowerPCs on SOI [13]. IBM servers using these microprocessors have been in production since 1998. For the next-generation microprocessors, IBM has developed the POWER4, a microprocessor with two cores and a shared L2 array (174 million transistors). It is designed using 0.18-μm CMOS SOI and runs at a frequency of 1.3 GHz [21]. As we move to the 0.13-μm generation and beyond, the use of SOI is spreading to more "commodity" microprocessors and SRAM.

SOI is opening a number of opportunities in the low-power arena. In many applications, from hand-held devices to large servers, power is becoming a limiting factor. One of the attractions of the SOI technology is its low-power behavior. For a given CMOS generation, SOI provides higher performance than a comparable bulk technology. This performance "headroom" allows for operation at lower voltage and lower power (as much as 2–3×) [22]. In other words, the lowest-active-power technology is the highest-performance technology operated at low voltage. This effect is illustrated in **Figure 11**, in which power is plotted against delay for an unloaded NAND3 delay chain for a number of technologies (0.30-μm to 0.13-μm SOI and bulk CMOS). At present,

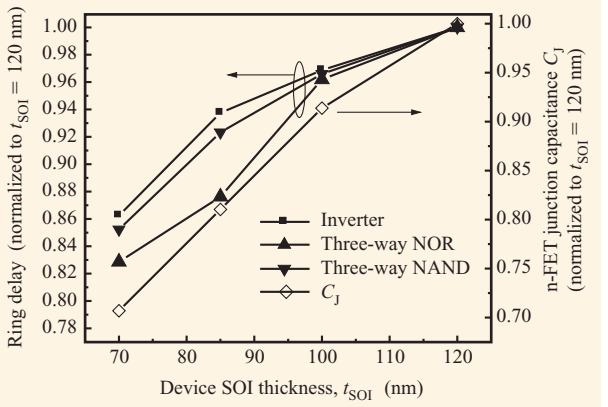

Figure 10

Measured impact of SOI film thickness on total C_J and ring performance. Reproduced with permission from [20]; © 2001 IEEE.

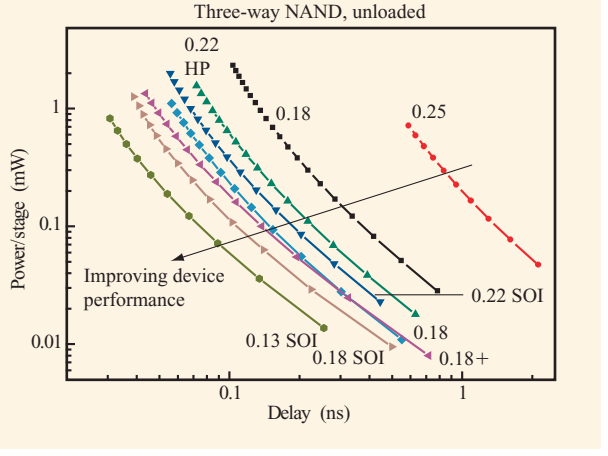

Figure 11

Delay vs. power of CMOS technology generations: As the technology performance improves, so do the power–delay characteristics of the technology. HP = high-performance version of IBM 0.22-μm bulk CMOS.

the technology group is developing a 0.13-μm low-power CMOS SOI technology (optimized down to 0.7 V) for commodity applications, along with the associated ASIC libraries.

One concern expressed about the use of SOI for low power has been its high off-current, I_{OFF}. The design point of SOI technology at a given generation has been set to match the worst-case SOI I_{OFF} to that of the highest-performance bulk at that generation. This usually means matching I_{OFF} at the minimum channel length of the

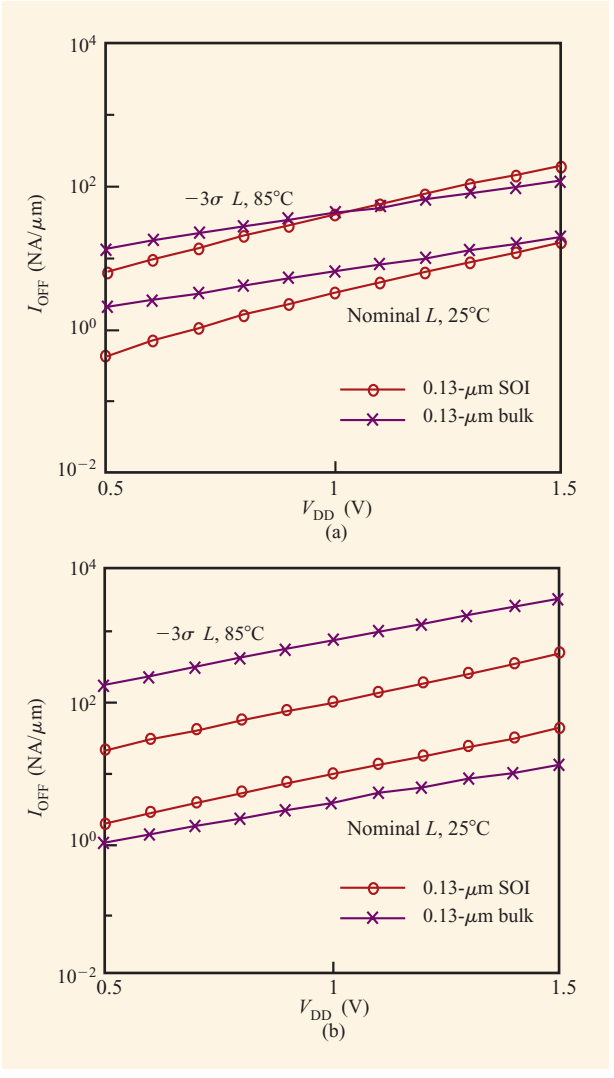

Figure 12

Off-current comparison of 0.13-μm bulk and SOI (a) n-FET and (b) p-FET devices at 25°C and 85°C.

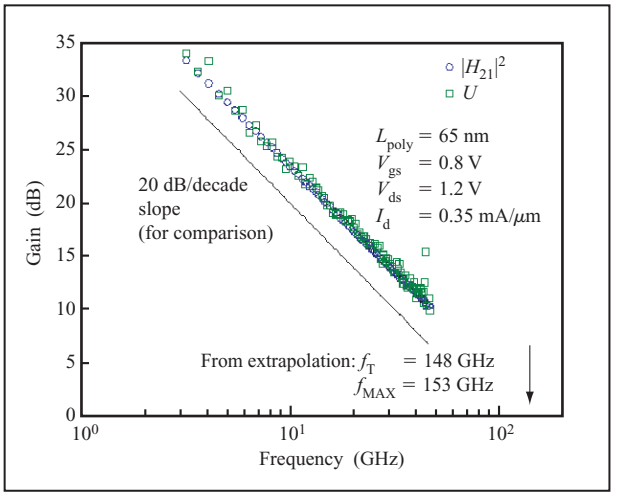

Figure 13

f_T and f_{MAX} of a 0.07-μm effective-channel-length n-FET from 0.13-μm CMOS technology. H_{21} = unilateral gain; U = short-circuit current gain; L_{poly} = device channel length; V_{gs} = gate-to-source bias voltage; V_{ds} = drain-to-source bias voltage; I_d = drain current. Reproduced with permission from [22]; © 2001 IEEE.

technology, L_{MIN}, and at high temperature. Then, at nominal channel length and supply, the I_{OFF} in SOI is about 10× higher than the bulk-Si I_{OFF}. As one reduces the voltage to lower the SOI active power, the SOI I_{OFF} decreases much faster than that of bulk, matching the bulk-Si I_{OFF} at low voltage. As shown in **Figure 12**, a 0.13-μm SOI at low voltage has nearly the same I_{OFF} as a high-performance 0.13-μm bulk technology at low voltage (and >20% higher performance) [23].

With the increased use of wireless technology, SOI CMOS offers some unique opportunities in the mixed-signal and radio-frequency rf circuits. Wireless technologies require high-performance transistors and low-loss passive devices (inductors and capacitors). SOI allows the use of high-resistivity substrate (>2K Ω-cm), which can result in high Q for the passive elements (i.e., inductor Q greater than 30), which would minimize the crosstalk among analog and digital circuits. High device performance is manifested in SOI FETs having unity current and power gain frequencies, f_T and f_{MAX}, of more than 150 GHz (for the 0.13-μm generation), the fastest for any CMOS technology to date (**Figure 13**). The noise figure is 1 dB at 5 GHz. The ability to integrate high-performance rf and low power on the same chip with minimum crosstalk will open the path to many new applications [24].

For communications and low-noise circuits, there are applications that require bipolar transistors. A group in the Research Division has demonstrated a novel vertical SiGe-base bipolar transistor on SOI with depleted collector, on SOI films with the same thickness as that used in CMOS SOI. This transistor can conceptually be integrated with high-performance CMOS on high-resistivity substrates, opening the way to potentially novel applications [25].

One of the requirements for the widespread use of any technology is the capability to build a system-on-a-chip. Most system-on-a-chip manifestations require embedded DRAM (EDRAM), and one of the concerns with implementing DRAM on SOI has been the passgate leakage, which can lead to cell discharge based on the data pattern. One proposed method of implementing

EDRAM on SOI is to use patterned SOI: Build the DRAM cells on bulk Si and all of the other circuits on SOI [26]. **Figure 14** shows an implementation of patterned SOI [27]. In fact, we have built EDRAM macros on such films. **Figure 15** shows retention-time (time for the first memory failure to retain data beyond specific value) plots of array diagnostic monitors (ADMs) produced on a 524Kb macro using bulk and patterned SOI [26]. (The ADM is a macro that specifically tests EDRAM cell functionality in an array environment and its retention characteristics.) The chart shows that the first retention fails of EDRAM bits in either of the patterned SOI masks occur at 128 ms. Early retention fails for bulk EDRAM occur between 128 and 256 ms. The slightly higher single-cell fail count at lower retention times is largely due to the smaller devices present in the high-performance SOI process technology which contribute to higher off-state leakage of the array device (L_{drawn} = 0.22 μm).

Future

As we move forward, SOI CMOS technology development is completed for the 0.13-μm generation [6, 20] and has been initiated for the 0.1-μm generation [7, 27]. These technologies are simply the highest-performance CMOS in production. In terms of power, SOI offers a better power–delay product compared to bulk Si (trading off performance for lower power). SOI offers the opportunity for the use of very-high-resistivity substrates to achieve low-loss passive elements. This capability, along with SOI n-FETs with an f_T of >150 GHz, opens exciting opportunities for low-power rf circuits. The ability to integrate vertical SiGe-base bipolar with SOI CMOS will open new application areas for SOI. The initial step toward integrating EDRAM on SOI has been taken. SOI CMOS is being applied to more mainstream microprocessors and SRAMs, and is being adopted by a number of companies. Bringing SOI into the mainstream of Si technology has been challenging. However, as we move to the 0.1-μm generation and beyond, SOI offers the total solution, and it will be the technology of choice.

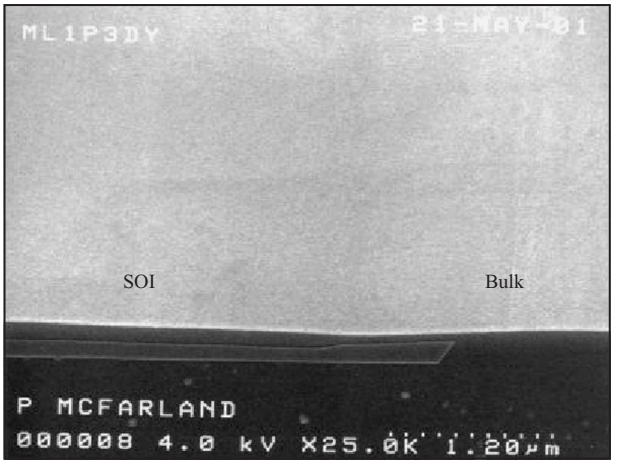

Figure 14

SOI wafer patterned using the SIMOX process. Reproduced with permission from [27]; © 2001 IEEE.

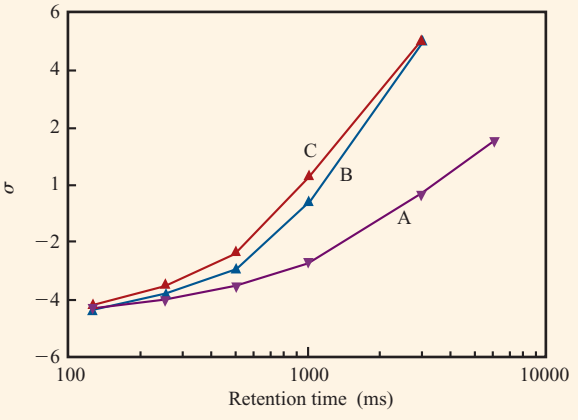

Figure 15

Retention-time plots of array diagnostic monitors produced on bulk Si using (curve A) the industry-standard 0.13-μm process; (curve B) an all-DRAM macro on bulk region of patterned SOI using the 0.13-μm SOI CMOS process; and (curve C) a DRAM cell in bulk Si using the 0.13-μm CMOS process. Sigma is the standard deviation of the distribution of the DRAM cell retention time (centered at 50% fail probability). Reproduced with permission from [27]; © 2001 IEEE.

Acknowledgments

The mainstreaming of the SOI CMOS technology in IBM was made possible by the hard work and total dedication of hundreds of engineers and technicians in a number of IBM divisions (Research, Microelectronics, and Server) over many years, and by the vision of the IBM management who began the work and supported it through its difficult times. In addition to many fellow engineers, the author is grateful to Tak Ning, who initiated the work in the Research Division in 1989; Bijan Davari, who led the SOI development from early days through its most challenging hours to the present time; John Kelly III, who had the vision to move SOI into the Microelectronics Division and enlist the IBM Server Group to use it across their product portfolio; Russ Lange, for his unique insight into SOI device operation; and Michael Polcari and Harry Calhoun, who led the work in the Research Division and in the Advanced Silicon Technology Center, bringing to the SOI project many years of their valuable development leadership.

*Trademark or registered trademark of International Business Machines Corporation.

**Trademark or registered trademark of The Open Group.

References

1. J. P. Colinge, *Silicon-on-Insulator Technology: Materials to VLSI*, Kluwer Academic Publishers, Dordrecht, Netherlands, 1991.
2. G. G. Shahidi, C. A. Anderson, B. A. Chappell, T. I. Chappell, J. H. Comfort, B. Davari, R. H. Dennard, R. L. Franch, P. A. McFarland, J. S. Neely, T. H. Ning, M. R. Polcari, and J. D. Warnock, "A Room Temperature 0.1 μm CMOS on SOI," *IEEE Trans. Electron Devices* **41,** 2405 (1994).
3. D. J. Schepis, F. Assaderaghi, D. S. Yee, W. Rausch, R. J. Bolam, A. C. Ajmera, E. Leobandung, S. B. Kulkarni, R. Flaker, D. Sadana, H. J. Hovel, T. Kebede, C. Schiller, S. Wu, L. F. Wagner, M. J. Saccamango, S. Ratanaphanyarat, J. B. Kuang, M. C. Hsieh, K. A. Tallman, R. M. Martino, D. Fitzpatrick, D. A. Badami, M. Hakey, S. F. Chu, B. Davari, and G. G. Shahidi, "A 0.25 μm CMOS on SOI and Its Application to 4 Mb SRAM," *IEDM Tech. Digest*, p. 345 (1997).
4. A. Ajmera, J. W. Sleight, F. Assaderaghi, R. Bolam, A. Bryant, M. Coffey, H. Hovel, J. Lasky, E. Leobandung, W. Rausch, D. Sadana, D. Schepis, L. F. Wagner, K. Wu, B. Davari, and G. Shahidi, "A 0.22 μm CMOS–SOI Technology with a Cu BEOL," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 1999, p. 15.
5. E. Leobandung, E. Barth, M. Sherony, S.-H. Lo, R. Schulz, W. Chu, M. Khare, D. Sadana, D. Schepis, R. Bolam, J. Sleight, F. White, F. Assaderaghi, D. Moy, G. Biery, R. Goldblatt, T.-C. Chen, B. Davari, and G. Shahidi, "A 0.18 μm CMOS on SOI Technology," *IEDM Tech. Digest*, p. 445 (1997).
6. P. Smeys, V. McGahay, I. Yang, J. Adkisson, K. Beyer, O. Bula, Z. Chen, B. Chu, J. Culp, S. Das, A. Eckert, L. Hadel, M. Hargrove, J. Herman, L. Lin, R. Mann, E. Maciejewski, S. Narasimha, P. O'Neil, S. Rauch, D. Ryan, J. Toomey, L. Tsou, P. Varekamp, R. Wachnik, T. Wagner, S. Wu, C. Yu, P. Agnello, J. Connolly, S. Crowder, C. Davis, R. Ferguson, A. Sekiguchi, L. Su, R. Goldblatt, and T. C. Chen, "A High Performance 0.13 μm SOI CMOS Technology with Cu Interconnects and Low-*k* BEOL Dielectric," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2000, p. 184.
7. S. K. H. Fung, M. Khare, D. Schepis, W. H. Lee, S. H. Ku, H. Park, J. Snare, B. Doris, A. Ajmera, K. P. Muller, P. Agnello, P. Gilbert, and J. Welser, "Gate Length Scaling Accelerated to 30nm Regime Using Ultra-Thin Film PD-SOI Technology," *IEDM Tech. Digest*, p. 629 (2001).
8. L. T. Su, J. B. Jacobs, J. E. Chung, and D. A. Antoniadis, "Deep-Submicrometer Channel Design in Silicon-on-Insulator (SOI) MOSFET's," *IEEE Electron Device Lett.* **15,** 366 (1994).
9. F. Assaderaghi, G. G. Shahidi, L. Wagner, M. Hsieh, M. Pelella, S. Chu, R. H. Dennard, and B. Davari, "Transient Pass-Transistor Leakage Current in SOI MOSFET's," *IEEE Electron Device Lett.* **18,** 241 (1997).
10. F. Assaderaghi, G. G. Shahidi, M. Hargrove, K. Hathorn, H. Hovel, S. Kulkarni, W. Rausch, D. Sadana, D. Schepis, R. Schulz, D. Yee, J. Sun, R. Dennard, and B. Davari, "History Dependence of Non-Fully Depleted (NFD) Digital SOI Circuits," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 1996, p. 122.
11. 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," *ISSCC Tech. Digest*, p. 426 (1999).
12. S. K. H. Fung, N. Zamdmer, P. J. Oldiges, J. Sleight, A. Mocuta, M. Sherony, S.-H. Lo, R. Joshi, C. T. Chuang, I. Yang, S. Crowder, T. C. Chen, F. Assaderaghi, and G. Shahidi, "Controlling Floating-Body Effects for 0.13 μm and 0.10 μm SOI CMOS," *IEDM Tech. Digest*, p. 231 (2000).
13. T. C. Buchholtz, A. G. Aipperspach, D. T. Cox, N. V. Phan, S. N. Storino, J. D. Strom, and R. R. Williams, "A 660 MHz 64b SOI Processor with Cu Interconnects," *ISSCC Tech. Digest*, p. 88 (2000).
14. D. H. Allen, A. G. Aipperspach, D. T. Cox, N. V. Phan, and S. N. Storino, "A 0.2 μm 1.8 V SOI 550 MHz 64b PowerPC Microprocessor with Copper Interconnects," *ISSCC Tech. Digest*, p. 438 (1999).
15. D. Stasiak, J. Tran, F. Mounes-Toussl, and S. Storino, "A 2nd Generation 440 ps SOI 64b Adder," *ISSCC Tech. Digest*, p. 288 (2000).
16. M. Canada, C. Akrout, D. Cawthron, J. Corr, S. Geissler, R. Houle, P. Kartschoke, D. Kramer, P. McCormick, N. Rohrer, and L. Warriner, "A 580 MHz 32 bit PowerPC Microprocessor in 0.12 micron L_{eff} CMOS SOI with Cu Interconnects," *ISSCC Tech. Digest*, p. 430 (1999).
17. J. Eckhardt and P. D. Muench, "0.25 micron 1.8 V 1 GHz PLL for SOI Microprocessors," *ISSCC Tech. Digest*, p. 436 (1999).
18. T. Ghani, K. Mistry, P. Packan, S. Thompson, M. Stettler, S. Tyagi, and M. Bohr, "Scaling Challenges and Device Design Requirements for High Performance Sub-50 nm Gate Length Planar CMOS Transistors," *VLSI Tech. Digest*, p. 174 (2000).
19. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, New York, 1998.
20. J. W. Sleight, P. R. Varekamp, N. Lustig, J. Adkisson, A. Allen, O. Bula, X. Chen, T. Chou, W. Chu, J. Fitzsimmons, A. Gabor, S. Gates, P. Jamison, M. Khare, L. Lai, J. Lee, S. Narasimha, J. Ellis-Monaghan, K. Peterson, S. Rauch, S. Shukla, P. Smeys, T.-C. Su, J. Quinlan, A. Vayshenker, B. Ward, S. Womack, E. Barth, G. Biery, C. Davis, R. Ferguson, R. Goldblatt, E. Leobandung, J. Welser, I. Yang, and P. Agnello, "A High Performance 0.13 μm SOI CMOS Technology with a 70 nm Silicon Film and with a Second Generation Low-*k* Cu BEOL," *IEDM Tech. Digest*, p. 245 (2001).
21. C. J. Anderson, J. Petrovick, J. M. Keaty, J. Warnock, G. Nusbaum, J. M. Tendler, C. Carter, S. Chu, J. Clabes, J. DiLullo, P. Dudley, P. Harvey, B. Krauter, J. LeBlanc, P.-F. Lu, B. McCredie, G. Plum, P. J. Restle, S. Runyon, M. Scheuermann, S. Schmidt, J. Wagoner, R. Weiss, S. Weitzel, and B. Zoric, "Physical Design of a Fourth-Generation POWER GHz Microprocessor," *ISSCC Tech. Digest*, p. 232 (2001).
22. N. Zamdmer, A. Ray, J.-O. Plouchart, L. Wagner, N. Fong, K. A. Jenkins, W. Jin, P. Smeys, I. Yang, G. Shahidi, and F. Assaderaghi, "SOI CMOS for Low Power and RF," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2001, p. 85.
23. N. Zamdmer, J.-O. Plouchart, J. Kim, L.-H. Lu, S. Narasimha, A. Ray, M. Sherony, L. Wagner, N. Fong, and K. A. Jenkins, "Networking Applications of a High-Performance 0.13-μm SOI CMOS Technology," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2002, in press.
24. S. Narasimha, A. Ajmera, H. Park, D. Schepis, N. Zamdmer, K. A. Jenkins, J.-O. Plouchart, W.-H. Lee, J. Mezzapelle, J. Bruley, B. Doris, J. W. Sleight, S. K. Fung, S. H. Ku, A. C. Mocuta, I. Yang, P. V. Gilbert, K. P. Muller, P. Agnello, and J. Welser, "High Performance Sub-40nm CMOS Devices on SOI for the 70nm Technology Node," *IEDM Tech. Digest*, p. 625 (2001).

25. J. Cai, A. Ajmera, C. Ouyang, P. Oldiges, M. Steigerwalt, K. Stein, K. Jenkins, G. Shahidi, and T. Ning, "Fully-Depleted-Collector Polysilicon-Emitter SiGe-Base Vertical Bipolar Transistor on SOI," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2002, in press.
26. R. Hannon, S. S. K. Iyer, D. Sadana, J. P. Rice, H. L. Ho, B. A. Khan, and S. S. Iyer, "0.25 μm Merged Bulk DRAM and SOI Logic Using Patterned SOI," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2000, p. 66.
27. H. L. Ho, M. D. Steigerwalt, B. L. Walsh, T. L. Doney, D. Wildrick, P. A. McFarland, J. Benedict, K. A. Bard, D. Pendleton, J. D. Lee, S. L. Maurer, B. Corrow, and D. K. Sadana, "A 0.13 μm High-Performance SOI Logic Technology with Embedded DRAM for System-On-A-Chip Application," *IEDM Tech. Digest*, p. 503 (2001).

Received October 17, 2001; accepted for publication January 25, 2002

G. G. Shahidi *IBM Microelectronics Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533 (shahidi@us.ibm.com).* Dr. Shahidi received his B.S., M.S., and Ph.D. degrees, all in electrical engineering, from MIT. In 1989 he joined the IBM Thomas J. Watson Research Center, where he initiated the SOI development program. After demonstration of device design and 5-in. SOI material in the IBM Research Division, Dr. Shahidi, along with the SOI development effort, moved to the IBM Microelectronics Division Advanced Silicon Technology Center (ASTC) in Hopewell Junction, New York. Over the following years, Dr. Shahidi led the development of SOI CMOS technology at the ASTC. This work resulted in the development of 8-in. SOI technology infrastructure, demonstration of SOI performance gain, qualification of multiple CMOS SOI technologies and their transfer to manufacturing, establishment of design infrastructure, and the first mainstream use of SOI. Dr. Shahidi is currently the Director of High-Performance Logic Development, and an IBM Fellow.