High-performance CMOS variability in the 65-nm regime and beyond

K. Bernstein
D. J. Frank
A. E. Gattiker
W. Haensch
B. L. Ji
S. R. Nassif
E. J. Nowak
D. J. Pearson
N. J. Rohrer

Recent changes in CMOS device structures and materials motivated by impending atomistic and quantum-mechanical limitations have profoundly influenced the nature of delay and power variability. Variations in process, temperature, power supply, wear-out, and use history continue to strongly influence delay. The manner in which tolerance is specified and accommodated in high-performance design changes dramatically as CMOS technologies scale beyond a 90-nm minimum lithographic linewidth. In this paper, predominant contributors to variability in new CMOS devices are surveyed, and preferred approaches to mitigate their sources of variability are proposed. Process-, device-, and circuit-level responses to systematic and random components of tolerance are considered. Exploratory, novel structures emerging as evolutionary CMOS replacements are likely to change the nature of variability in the coming generations.

Introduction

Variability in the delay and power consumption of CMOS devices, circuits, and chips arises from scaling very large-scale integrated (VLSI) circuit technologies beyond the ability to control specific performance-dependent and power-dependent parameters [1]. This erosion in device and interconnect parameter precision has elevated variability to a first-order limitation to continued technology scaling. This process and device variability challenge to continued scaling [2] exacerbates the already-critical power dissipation problem, and is one of the most urgent problems confronting designers. Attempts to improve parameter precision in the manufacturing process now commonly confront atomistic-level constraints. Below 65 nm, quantum-mechanical limitations will make the achievement of parameter precision exponentially more difficult.

Delay and power variability in CMOS devices is influenced by many contributors. Parameter variation manifests itself in the distributions of process tolerance; it appears in voltage- and temperature-induced tolerance arising from the operating environment both locally to the circuit and across-chip. Variability can be temporal or spatial in nature. Temporally, the variability can occur across nanoseconds (such as in the SOI history effect [3]) to years (such as in process centering); this is shown in **Table 1**. This time dependence may arise from instantaneous changes in circuit performance induced by use, and it is associated with a specific technology. Added delay, such as that needed to discharge residual charge possibly trapped in capacitance between devices in NAND gate stacks, is temporal. The silicon-on-insulator (SOI) history effect and device self-heating are additional application-dependent examples. SOI device body history and charging storage effects have a temporal, structural dependence. Aging-induced variation arising from wear-out mechanisms has a negative impact on performance. Negative-bias temperature instability (NBTI) affecting p-FETs and hot-electron effects affecting n-FETs both elevate device thresholds, degrading device and circuit performance [4]. Electromigration (EM) [5] slowly erodes interconnect admittance, becoming more severe below 65 nm because of higher interconnect current densities. The term *spatial variation* refers to lateral and vertical differences from intended polygon dimensions and film thicknesses [1]. Spatial variation modes exist between

0018-8646/06/$5.00 © 2006 IBM

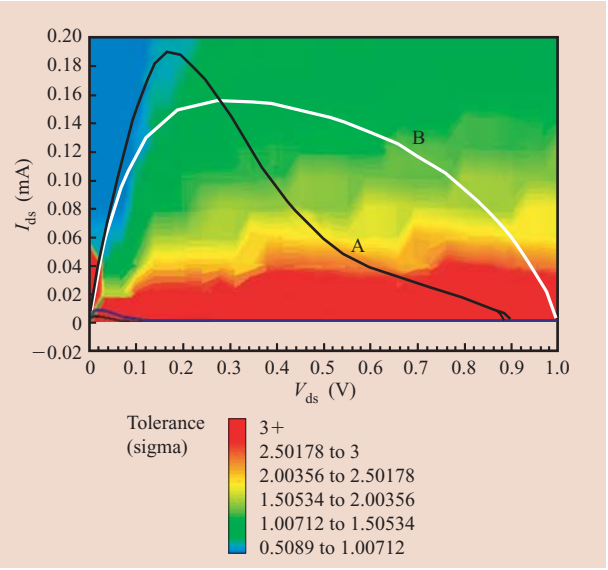

Figure 1

Device trajectory and associated tolerance.

Table 1 Order-of-magnitude variability time domains and estimated delay impact.

Time domain (s)	*Mechanism*	*Delay impact approx. (3 sigma)* (%)
10^{12}	Lithography node	20
10^{9}	Electromigration	5
10^{8}	Hot-electron effect	5
10^{6}	Negative bias temperature instability	15
10^{4}	Chip electrical mean variation	15
10^{-1}	Across-chip L_{poly} variation	15
10^{-4}	Self heating/temperature	12
10^{-8}	SOI history effect	10
10^{-10}	Supply voltage	17
10^{-10}	Line-to-line coupling	10
10^{-11}	Residual source/drain charge	5

devices, between circuits, between chips, and across wafers, lots, and the lifetime of any particular fabrication system.

Parameter tolerance may be deconvolved into random and systematic components. Circuit sensitivity to variation is also a strong function of the specific circuit topology used to implement a given logic function. The examination of the trajectory in I–V space for devices under use conditions provides a strong indication of the delay deviation one may expect. The plot in **Figure 1** shows trajectories of the operation of the n-FETS in the common NAND2. The upper n-FET device is kept at high voltage, and the lower-voltage n-FET device (curve A) is switched. The background color in the figure is indexed to the magnitude of device current variation actually observed in dc hardware characterization of the device, operated at that specific drain–source voltage (V_{DS}), drain–source current (I_{DS}), and implied gate–source voltage (V_{GS}) point on the plot. The red shading indicates regions of the highest device current tolerance, and green shading shows areas of the lowest device current tolerance. Clearly, delay variation in a circuit is higher when the output is gated by the lower of the two devices, highlighted by the larger portion of the transition spent in the high-tolerance (red) region. Informed choices among alternative circuit topologies for a given function below the 65-nm node can be selected using this criterion.

Device variability

Categorizing variability

There are multiple ways of describing device variability; a useful approach is shown in **Table 2**. This particular breakdown is useful because it separates issues requiring different statistical treatments in anticipating their circuit impacts. This also structures our discussion of these effects. Variations are separated into rows according to spatial domain: those that involve the chip mean, those that vary within the chip but have local or chip-to-chip correlation, and those that vary randomly from device to device. The columns identify variations arising from the process used to make the device, or originating from device behavior changes over time. This last category is further divided into reversible and irreversible changes. Examples of sources of variation and/or the parameters which should be monitored are also shown. Temporal, irreversible device variation contributors are associated with aging and device wear-out.

Intrinsic device variability

Intrinsic variations are caused by atomic-level differences between devices that occur even though the devices may have identical layout geometry and environment. These stochastic differences appear in dopant profiles, film thickness variation, and line-edge roughness. An example is shown in **Figure 2**, in which threshold voltages of ~3,500 identical n-MOSFETs laid out in a compact array have been measured. Even though there is no systematic process variation between the FETs, there is still a fairly wide Gaussian distribution of threshold

voltages. Another example is shown in **Figure 3**, in which ~1,500 different FETs have been measured for each of 32 different length × width combinations, again for FETs in compact arrays. The standard deviation, σ_{V_T}, of each of the distributions has been extracted and is plotted to show the dependence on channel area. As can be seen, the smallest FETs can have σ_{V_T} in excess of 30 mV. The majority of the V_T variation is shown to be due to the atomistic nature of the dopants in MOSFETs [3]. The implant and annealing processes result in the placement of a random number of dopants in the channel (described by a Poisson distribution) and in the random positioning of the atoms that are present, as illustrated in **Figure 4(a)**. All of the dopants in a 50-nm n-MOSFET have been positioned by a Monte Carlo procedure, and their positions are plotted in 3D perspective [4]. As shown, the source and drain doping is quite dense, but the channel doping is susceptible to statistical variation. Actually, most of the acceptors present are seen in the quasi-neutral body region. Only a few hundred ionized acceptors in the body of this FET are responsible for setting the threshold voltage. Since these N-ionized dopants are subject to Poisson statistics, the uncertainty in the number of dopants is approximately $N_d = N^{0.5}$, or 5–10% of the total number of dopants for small FETs.

The uncertainty caused by atomistic doping has been the focus of substantial research [3–11]. It has been found that this uncertainty can give rise to significant V_T variation, the details of which depend on the doping profile. In general, doping near the surface and close to the actual channel has the largest effect on V_T, so retrograde doping profiles (which keep the dopant away from the channel) are desirable and have been shown to produce smaller threshold voltage (V_T) variation [9]. Removing the doping from the channel altogether (in SOI devices) could potentially reduce σ_{V_T} even further, but the V_T must then be set by gate-metal workfunctions or by a separately biased back gate [4, 10, 12].

Quantum-mechanical effects in the channel have been shown to increase σ_{V_T} (compared with simulations without quantum mechanics) [3], and doping in the gate polySi also contributes to the σ_{V_T}. In very short FETs, statistical doping effects can cause significant variation in

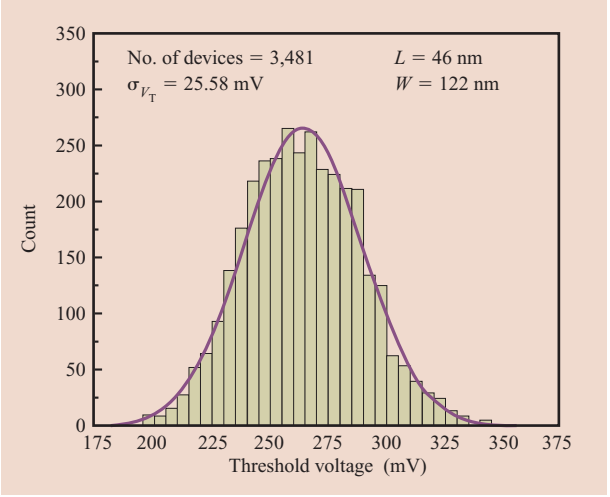

Figure 2

Threshold voltage histogram for FETs in the 90-nm-technology node.

Table 2 Categorization of device variations.

Proximity	*Spatial*	*Temporal*	
		Reversible	*Irreversible*
Variation of chip mean	Parameter means (L_G, V_T, t_{ox})	Environmental operating temperature Activity factor	Hot-electron effect NBTI shift
Within-chip variation	Pattern-density/ layout-induced transconductance	On-die hot spots	Hot-spot-enhanced NBTI
Device-to-device variation	Atomistic dopant variation Line-edge roughness Parameter std. dev.	SOI body history Self heating	$\sigma_{VT\text{-}NBTI}$ (NBTI-induced V_T distribution)

Temporal—dynamic, time-dependent delay variation
Pattern density—variation caused by variation in density of polygons in given area
Hot spots—regions of excessive local heating caused by high power dissipation density
Hot-spot-induced NBTI—Threshold variation caused by excessive local heating
Self heating—Individual device heating caused by extended periods of high device current

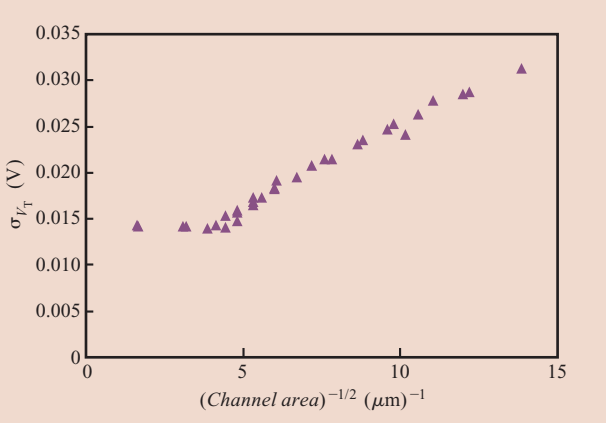

Figure 3

Example of σ_{V_T} vs. (*channel area*)$^{-1/2}$ for n-MOSFETs in an exploratory technology. Each point is a different length × width geometry.

short-channel behavior; a random deficit of doping concentration in the wrong place can create near-punchthrough states. Combining the data from many different simulations, it has been found that the spread in V_T can be approximately expressed as

$$\sigma_{V_T} = 3.19 \times 10^{-8} \left(\frac{t_{ox} N_A^{0.4}}{\sqrt{L_{eff} W_{eff}}} [\mathrm{V}] \right),$$

where N_A, L_{eff}, and W_{eff} are the average channel doping and the effective channel length and width, respectively [3]. Comparing with Figure 3, we observe that the $1/(L_{eff}\ W_{eff})^{0.5}$ dependence is indeed realized in the data.

Atomic-scale fluctuations in doping levels and device feature sizes also cause variation in the source/drain region, affecting the overlap capacitance and the effective source resistance. **Figure 4(b)** shows the randomly placed dopant atoms in a top view of a MOSFET [10]. Though the gate edge is perfectly smooth here, the fluctuations in doping level cause uncertainty in the edge of the source and drain, which translates into source/drain (S/D) capacitance and resistance variations. Line-edge roughness (LER) can be expected to exacerbate this effect.

LER, perhaps the second most significant contributor to variability, arises from statistical variation in the incident photon count during lithography exposure, and the absorption rate, chemical reactivity, and molecular composition of the photoresist [13]. **Figure 5** shows an example of simulating the exposure and development of a small via hole using extreme ultraviolet (EUV) lithography [14]. The randomness of the resulting via hole is very clear. Similar roughness occurs along the gates of MOSFETs, causing variability in the effective gate length as one moves along the width of a FET. The component of σ_{V_T} due to LER should vary as $1/(W_{eff})^{0.5}$, and simulations have generally shown this component to be small compared with the atomistic doping effect [3, 15]. Nevertheless, in devices approaching punchthrough, LER variation could be quite important.

Another source of intrinsic device variability arises from atomic-scale oxide thickness variations. Physical gate oxide thickness is currently down to 1 nm, equivalent to approximately five inter-atomic spacings. Experiments have shown that the oxide thickness actually varies by one or two atomic spacings on a nanometer-length scale [16]. Simulations of this effect have shown that it can give rise to a σ_{V_T} component up to half that of the doping, but since it is uncorrelated with the doping, it adds in quadrature, yielding only a ~10% increase in overall σ_{V_T} [3, 17]. In addition to threshold voltage variation, oxide thickness variations give rise to significant variation in the oxide tunneling current, since the tunneling current varies exponentially with the thickness. Over a whole chip this may result in a substantial increase in average oxide leakage current, but it is difficult to quantify experimentally. Oxide thickness variations are also responsible for the universally observed mobility degradation at elevated transverse field, often thought of as surface scattering. Thickness variation causes potential variation across the MOSFET channel, scattering the carriers and decreasing mobility at high lateral electric field values. Since these effects are atomistic, they must vary randomly from device to device. We should expect them to cause significant variations in nanoscale device mobility. This additional on-current uncertainty is beyond the current tolerance associated with V_T variations.

Extrinsic process variability

Extrinsic variation is due to unintentional shifts in contemporary process conditions. It is typically not associated with fundamental atomistic problems, but rather with the operating dynamics of a modern fabricator.

Extrinsic variability can be present in multiple references: a) from lot to lot, b) from wafer to wafer within a lot, c) across wafers, d) from chip to chip within a reticle in multi-reticle products, and e) across-chip. **Figure 6** provides a breakdown of the amount of variation seen in 90-nm hardware from wafer to wafer, from chip to chip, and within a chip. Each class has distinct contributors within the manufacturing process. Note that significantly more variation occurs chip-to-chip than wafer-to-wafer within a lot. Chip-to-chip variability has its source in both by-wafer and by-reticle process steps. By-wafer processing steps that assert variation include a) rapid thermal anneal, when temperature gradients

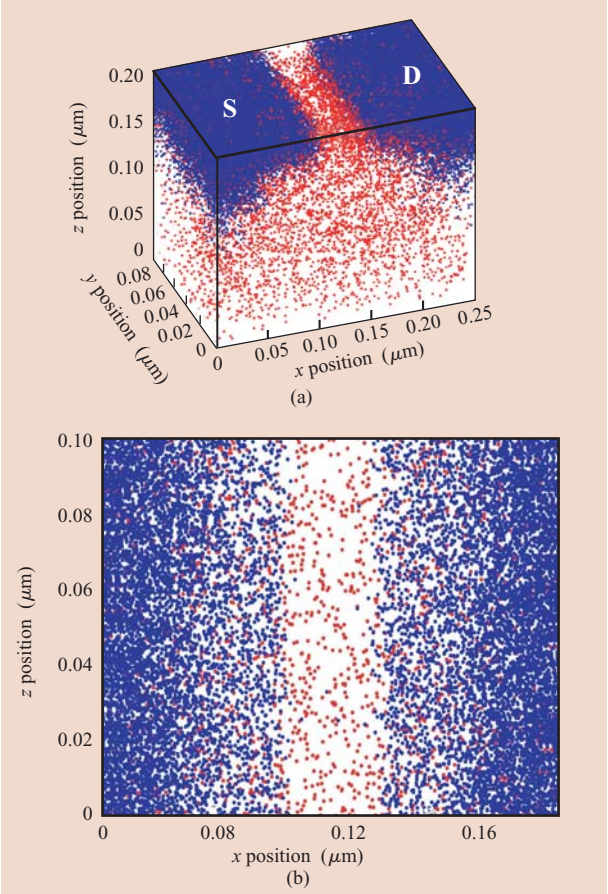

Figure 4

(a) Randomly placed dopants in a 50-nm channel-length MOSFET. Blue dots are donors creating the source and drain. Red dots are acceptors, primarily in the channel. The gate is not shown, but would cover the channel region between source (S) and drain (D). (b) Top view.

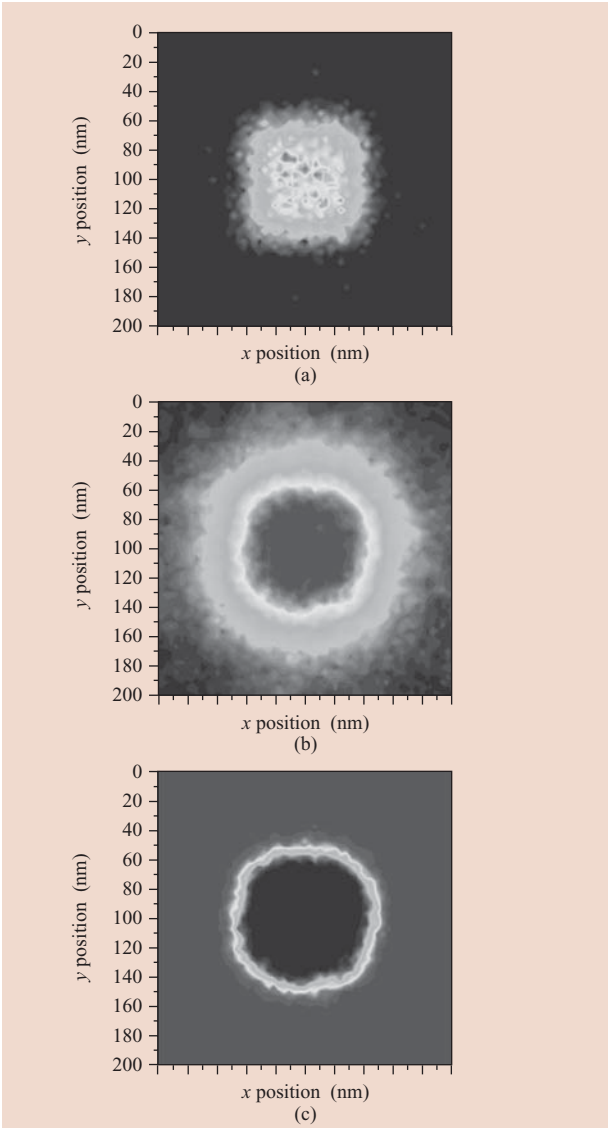

Figure 5

Simulation of atomistic variation in oxide thickness: (a) photons absorbed; (b) deprotected polymer; (c) dissolved polymer. From [14], reproduced with permission; ©2003 SPIE.

appear across the wafer, b) photoresist development, and c) etching. By-reticle, the photolithography process contributes variability if the focus changes as the mask is stepped across the wafer. Focus variation can be caused by exposure tool lens astigmatisms or by wafer/chuck nonplanarity.

Within-chip variability can be separated into similar-structure variability and dissimilar-structure variability. Within-chip similar-structure variability originates in across-wafer variations that each chip intercepts, as well as in across-reticle variations caused by mask or by-reticle photolithography processes. Note that both categories can be influenced by design attributes such as proximity of features and density of polygons. Dissimilar-structure variations have their sources not only in processing steps that differ by structure (such as mask levels devoted to high- or low-V_T transistors only), but also in processing sensitivities to layout variations of structures. Although created simultaneously using identical process steps, different instances of the same structures in different orientations show variations. Varying polygon densities change the local consumption of process chemicals. Photoresist and etch process chemistry are affected by this class of variability contributors. Dissimilar-structure variations can be significant; e.g., the solid curves in Figure 6 represent the distribution of monitors reflecting

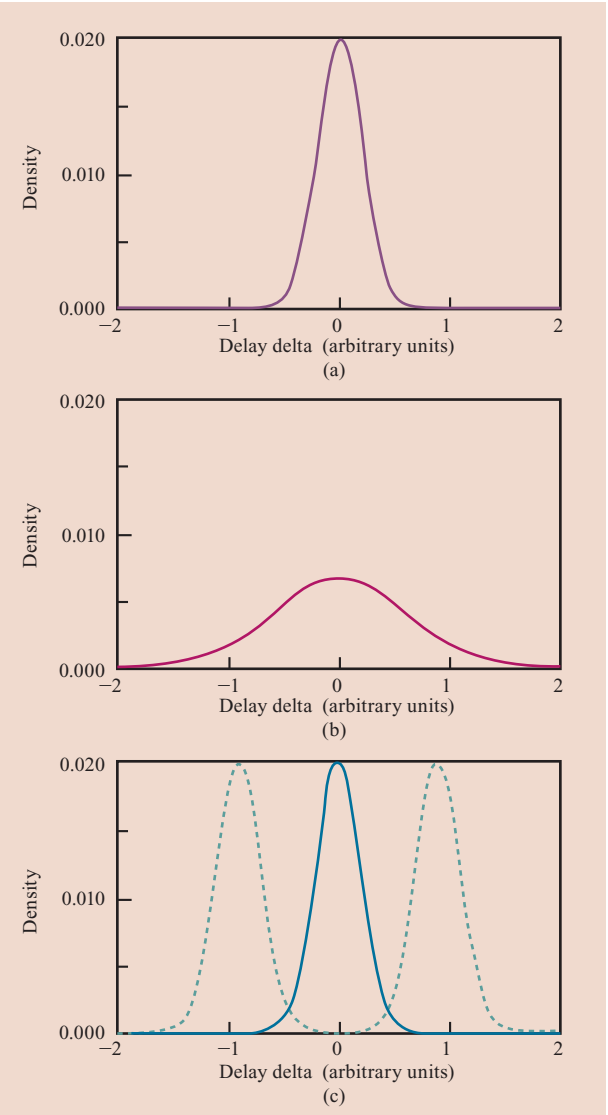

Figure 6

Example 90-nm hardware probability density function data illustrating distribution widths for various categories of delay variation: (a) wafer to wafer; (b) chip to chip; (c) within chip. The *y*-axis on each plot represents the relative density of gates at a given delay delta.

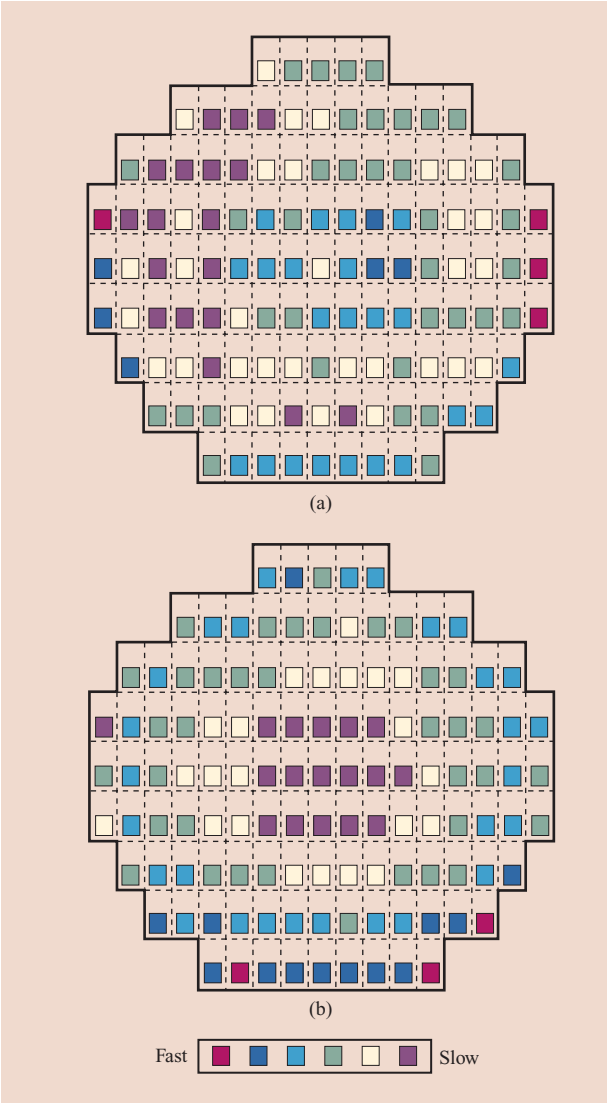

Figure 7

Wafer maps showing indicators of (a) source/drain resistance and (b) overlap capacitance.

similar-structure delays within a chip, while the dashed curves illustrate mean shifts in delays for dissimilar structures within our example hardware. Delays of the structures represented by the dashed curves are normalized to facilitate comparison with the solid-curve structures.

Finally, even same-delay hardware can have different characteristics. For example, **Figure 7** shows across-wafer variability in structures that are indicators of two different transistor attributes: 1) source–drain resistance and 2) gate-to-source and gate-to-drain overlap capacitance. Assuming similar channel lengths and thresholds, one can anticipate a chip coming from the center of the wafer, where indicators suggest favorable source/drain resistance but unfavorable overlap capacitance. This chip can exhibit the same nominal delay as another that is well removed from the center, even though it has very different component transistor parameters. Such differences may in turn cause divergence in circuit response to across-chip voltage and temperature sensitivity, as discussed in the circuits discussion which follows.

Placement-induced device variation

On a chip, placement-related sources of variation can also result in changes in the electrical parameters of active (transistor) and passive (wire) devices. These sources include manufacturing variability, which translates unavoidable spatial fluctuations in the fabrication process into corresponding changes in electrical parameters. Manufacturing variability may be systematic in nature, meaning that there is a well-understood relationship between design instances or layouts and the resulting electrical parameter values. A prime example of a systematic relationship includes the chemical–mechanical polishing (CMP)-induced relationship between the thicknesses of metal or inter-layer dielectric (ILD) and the layout feature density [1].

A key difference between systematic and random variability is in the manner in which it is treated in the circuit design cycle. Systematic phenomena may be modeled, anticipating the impact of the associated variability. Using the example of CMP above, one may analyze the impact of the CMP process on a design and adjust the design layout or timing to mitigate resulting precision problems [18]. Random phenomena, however, require the designer to perform worst-case analysis [19], invariably resulting in additional required design margin. This margin guards against the maximum (worst) timing impact that this random contribution to delay can cause. Understanding the sources, impacts, and dependencies associated with variability can decrease design margins and improve the competitiveness of a design.

Wear-out-induced timing changes

Physical variability also has a temporal component arising from the time dependence of certain aging and wear-out mechanisms. Designers address the timing problems from aging by modeling circuit delay changes when shipped and at end of life (EOL). Satisfaction of the maximum allowable critical path delay must be ensured in both settings. Contemporary CMOS technology asserts three mechanisms which must be anticipated in timing. Negative bias temperature instability (NBTI) reduces the performance of p-channel MOSFETs by slowly increasing the threshold voltage of the device, robbing it of overdrive [20]. NBTI arises from the generation of interface states and positive trapped charge while the device is in operation ($V_{gate} = 0$ V, $V_d = V_s = V_{dd}$). Hot-electron effect (HotE) degrades n-MOSFET on-current by injecting additional charge into the gate oxide which must be overcome in order to turn the device on [21]. HotE occurs when lateral device fields are elevated. Finally, electromigration [22] depletes the interconnect of conductor atoms over an extended period. EM arises from current densities in excess of the reliable limit of the wire. The reader is directed to the references for a more thorough treatment of these phenomena.

Time-dependent variability is a strong function of the capacitive loading and the ratio of p-FET to n-FET device widths (beta ratio), how often and how long the device is on (activity factor), and the chip environmental (voltage and temperature) operating conditions of a given circuit over the lifetime of the product.

Use-induced device variation

An integrated circuit is composed of numerous devices spatially distributed over a relatively small area of silicon. These devices are typically connected to one or more power supplies via a network of wires referred to as the on-chip power grid. With modern high-speed integrated circuits consuming many tens of watts in active and passive (leakage) power, temperature and power-supply variations have emerged as important sources of design variability [23]. It is not uncommon to have power-supply variations create a 10% variation in delivered power to different parts of a design, and that same 10% variation can in turn cause a similar amount of delay variation. Local temperature variations within the die cause variations in device mobility and threshold voltage as well as wire resistivity. These variations lead to changes in the delay of various paths within the die, and are mitigated by the quality of the package and cooling solution chosen. **Figures 8(a)** and **8(b)** [24] respectively show simulated power-supply variations within an application-specific integrated circuit (ASIC) design and examples of measured temperature variation within a microprocessor design. Techniques for estimating these types of environmental variations have existed for some time and have recently become efficient enough to be used for full-chip analysis [25]. Work is ongoing to link these types of variations to chip performance estimation (typically timing) [26, 27].

Circuit response

The static combinatorial CMOS circuit response to variability in process, voltage, and temperature has a strong dependence on specific schematic topology. To measure this dependence, Monte Carlo analyses assessing the robustness of various logic alternatives for a simple NAND and the more complex 16-bit adder functions were completed. For each function, selected electrical parameters were separately subjected to manufacturing process and operating-environment-induced tolerance. Independent parameter contributions to total variability were deconvolved in order to quantify the sensitivity of each circuit to each parameter.

In the first study, variability of delay and power was evaluated for the static, pulsed static, passgate, and dynamic realizations of the two-way NAND function.

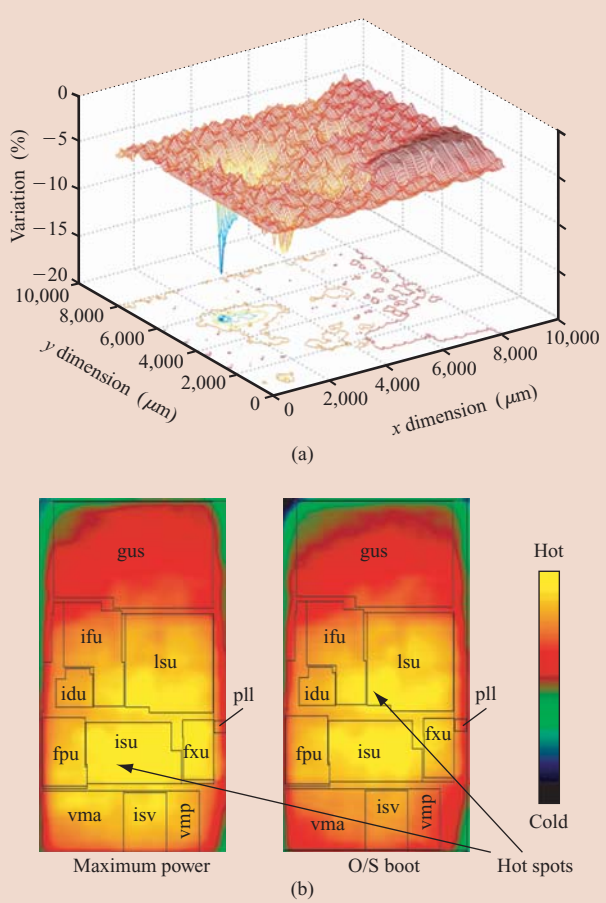

Figure 8

(a) Percentage of V_{DD} variation within an ASIC design (simulated). (b) Temperature variation within a processor running patterns for worst-case power (left) and while booting operating system (right) (measured). Courtesy N. Rohrer, IBM Austin Research Laboratory. From [24], with permission.

A NAND3 chain built in 90-nm partially depleted SOI CMOS technology was modeled from 1,000 statistically independent cases. **Figure 9(a)** provides plots of one sigma/mean of delay and power for each realization, respectively. Static CMOS displays the most well-controlled delay variation levels, with a normalized variability of 6.4%, while passgate-based circuits suffer significantly greater variability at 8.7%. The dynamic and pulsed static styles remain comparable to the static case, with 6.7% and 6.8% delay variability, respectively. While the static CMOS implementation displays a normalized power variability of 4.3%, the passgate-based style exhibits the highest amount at 5.7%. The variability of the dynamic and pulsed static styles remains lower than that of passgate structures, at 4.6% and 5.1%. Eleven 16-bit adders that span a range of circuit architectures and logic-evaluation styles were designed and subjected to 200 Monte Carlo simulations. The three basic architectures are the ripple carry adder with a passgate-based Manchester carry chain (static and dynamic) [29], logarithmic carry-select (static, dynamic, and passgate) [17], and carry-lookahead (Kogge–Stone radix 2 and radix 4 [30], Han–Carlson [31], and Brent–Kung [32]). A fan-out-of-4 (FO4) static inverter loads the critical paths for all adder designs. To conduct an unbiased comparison of the effects of process variability on designs within each

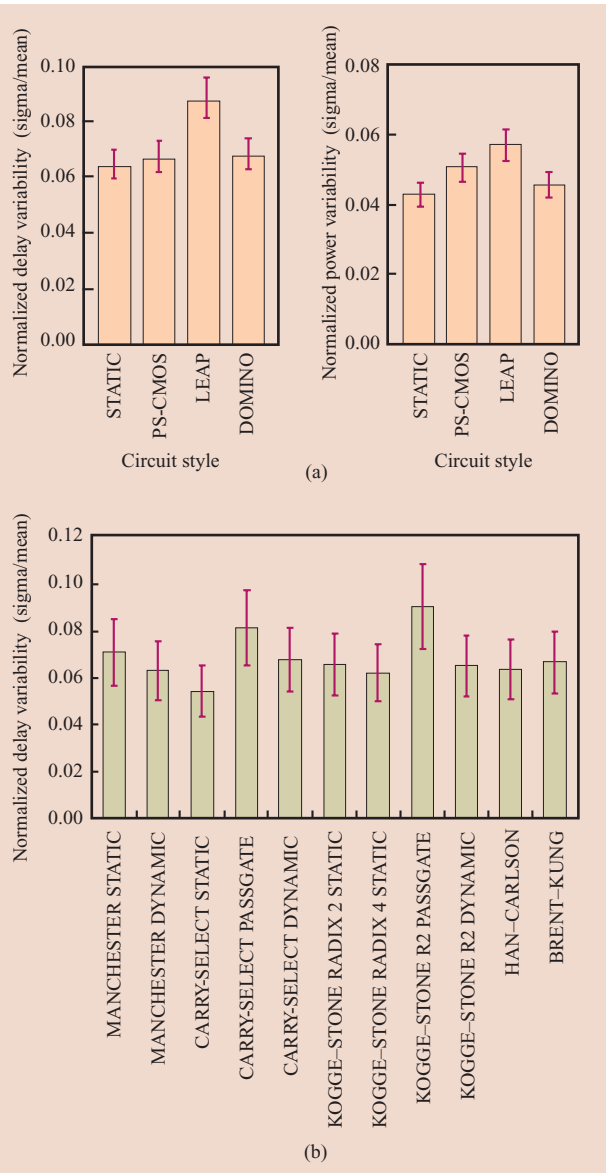

Figure 9

(a) Normalized sigma/mean delay and power for the NAND2. (b) Normalized delay variability of 16-bit adders. From [28], with permission; ©2004 IEEE.

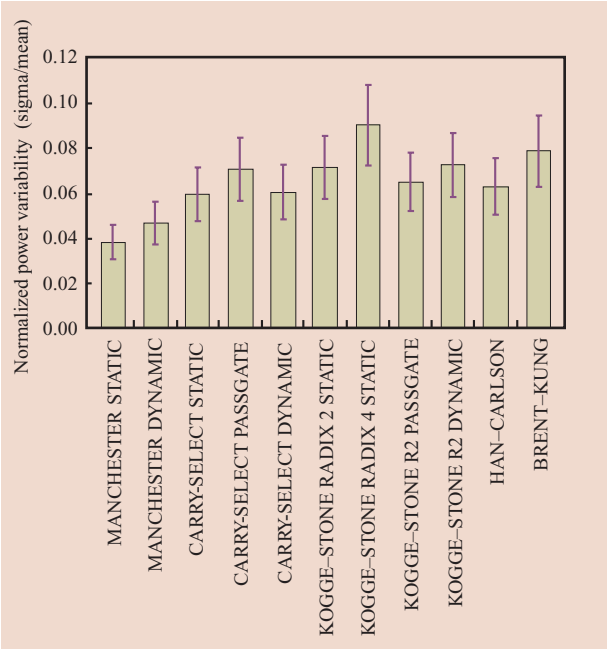

Figure 10

Normalized power variability of various 16-bit adders.

circuit type, transistor sizes were objectively optimized for delay with an in-house software routine that uses a genetic biological solution algorithm.

A substantial portion of the total variability experienced in complex circuits arises from choice of implementation. The static implementation of the carry-select adder is the most resistant to delay variation (5.4%), as shown in **Figure 9(b)**. While variability levels for most other static and dynamic designs fall within 20% of the static carry-select, passgate families clearly exhibit the worst variation control. The three designs with the highest relative delay variation are the static ripple carry adder with passgate-based Manchester carry chain (7.1%), the passgate implementation of the carry-select (8.2%), and the passgate-based radix 2 Kogge–Stone (9.1%).

Trends in adder power variability are shown in **Figure 10**. The static ripple carry adder using the Manchester carry chain displays the most predictable power values (3.8% variability), while the variation in other designs ranges between 22% and 137% higher. The two least robust designs from a power perspective are the static, radix 2 Brent–Kung (7.9%) and static, radix 4 Kogge–Stone (9.1%) adders, each with spreads more than 100% larger. This result may be attributed to the higher relative complexities of these designs, each having large intermediate capacitances along critical path nodes.

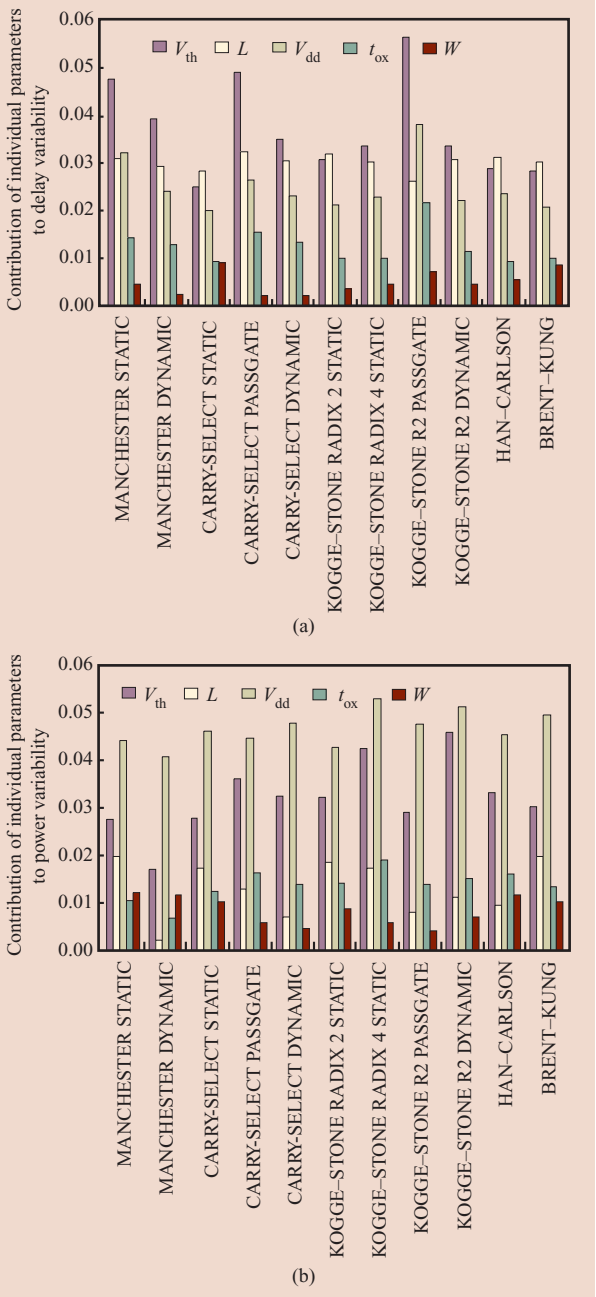

Figure 11

(a) Contributions of individual parameters to delay variability of 16-bit adders. (b) Contributions of individual parameters to power variability of 16-bit adders. Device threshold, length, supply voltage, and gate oxide thickness and width are examined.

Finally, of particular interest is the topology dependence of individual parameter sensitivity. **Figure 11(a)** captures the change in delay caused by moving device width, length, gate oxide, or base

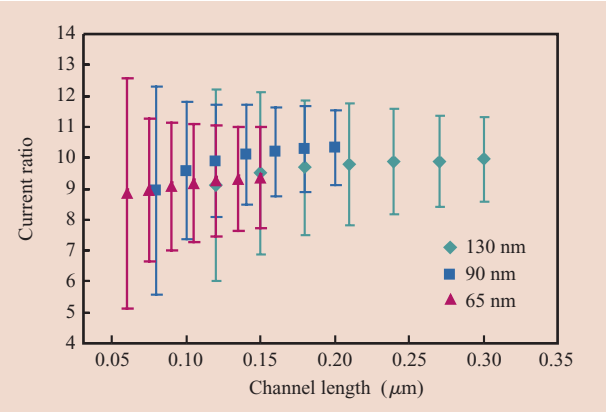

Figure 12

Simulated current-ratio variability of 10:1 current mirror for 130-nm, 90-nm, and 65-nm technologies. The nominal value and ±3 standard deviation ranges from 1,000-case Monte Carlo simulation are plotted as a function of channel length.

threshold independently across its full ±3 sigma process window. For each of these cases, secondary parameter dependencies on the parameter being altered are allowed to occur. Threshold voltage is found to be the most significant parameter in the topologies studied, with an average contribution of 3.7% for the adders.

The designs most sensitive to variations in threshold voltage are the passgate-based styles. The effects of gate length L are nearly as significant as V_{th} contributions, accounting for an average of 3% of the overall variability in both cases. Furthermore, supply-voltage variations account for average contributions of 2.4% (NAND chains) and 3% (adders). The process parameters t_{ox} and W are the least significant, with average respective contributions of 1.4% and 0.3% for the NAND chains, and 1.2% and 0.5% for the adders. Process control of t_{ox} and W is also typically very good. These results quantify the high sensitivity of delay to fluctuations in V_{th}, V_{dd}, and device length L, consistent for NAND chains and the family of adders, across all logic evaluation styles. Clearly, efforts to impose tighter control over these three parameters during manufacturing and design processes would significantly improve the ability to control the range of transistor gate delays.

Variability in active power dissipation is affected by supply tolerance: **Figure 11(b)** shows average V_{dd} power variability contributions of 4.7% for the adders caused by ±5% voltage variation within specifications. Fluctuations in V_{th} also contribute significantly to power variation, accounting for 3.2% power tolerance. Techniques for improving V_{th} control during manufacturing and for reducing V_{dd} noise during circuit operation both improve power dissipation predictability.

Analog circuit variability considerations

Analog circuits with differential operations are affected by "mismatch" between nominally identical components due to the technology and layout variability, long before such variability becomes noticeable for digital circuit designers. The variations affecting analog performance may be mismatches in transistor V_{th}, channel length and width, and mismatches in passive components such as resistors and capacitors. To meet a given performance specification, analog designers overcome unwanted variability with multiple approaches, i.e., using symmetric layout style and dummy devices to ensure that the environmental mismatch is kept to a minimum, using more chip area to put in devices larger than the minimum, and using additional tunable circuits for compensation and correction.

Environmental dummy devices are nonfunctional devices that are used to improve device tracking. They are widely used to improve current tracking in current mirrors and offset voltage tracking in differential circuits such as current mode logic (CML) amplifiers/summers, comparators, latches, and op-amps. For more advanced technology generations, adding dummy devices to the perimeter of the mirror devices also mitigates the stress variation and improves tracking. Even numbers of fingers for the reference and mirror FETs are also recommended to reduce the FET S/D asymmetry resulting from angled implants.

By adhering to these strict layout rules for environmental symmetry, systematic variations are mostly removed so that the analog circuit is subject predominantly to local mismatch due to random variations. Extensive Monte Carlo simulations for local random variations are then used to ensure that the circuit meets the targeted performance metrics over all process, voltage, temperature corners, and variations. **Figure 12** shows an example of Monte Carlo simulation for a 10-to-1-current mirror circuit as a function of channel length in 130-, 90-, and 65-nm technologies. Here the channel widths of the reference and mirror n-FET devices are also scaled with the channel length, and the drain voltage of the mirror n-FET is uniformly distributed from 20% to 40% of V_{dd}. The V_{dd} values are assumed to be 1.2 V, 1.0 V, and 0.9 V respectively for 130-nm, 90-nm, and 65-nm technologies. Figure 12 shows that the variation rises significantly as one approaches the minimum channel length for each technology. For this reason, a channel length of 1.5 to 2 times the minimum is typically chosen for good matching. The overall optimum device size is a balance between variability (decreasing with increasing size), circuit performance (e.g., operation frequency and bandwidth), and chip area. Also from Figure 12, note that one observes a very modest reduction in variability in

migrating from 90-nm- to 65-nm-technology nodes at a given channel length, in marked contrast to the more significant improvement in the total variability window that is seen when migrating from 130-nm to 90-nm technology. The nominal channel length for each technology (the fourth of seven bars in each color) shows virtually identical average variability.

More recently, as industry-standard data rates pass 3 Gb/s and approach the 6+ Gb/s to 11+ Gb/s realm, analog blocks must meet ever more stringent performance metrics (e.g., bandwidth, gain, linearity, jitter, power, and chip area)[1] [33]. To achieve these targets in the face of increasing variability, mismatch-related degradation such as dc offset can sometimes be compensated with correction circuits that provide power-on calibration, continuously adaptive real-time calibration, or both. In differential CML comparators and summers, dc offset is corrected by measuring a tail current from one leg of the outputs and applying an offset current from the current digital-to-analog converter (IDAC) block to achieve a constant, calibrated value. As an example, a typical maximum-range 32-mV dc offset (including that from device mismatch and that from data input) can be canceled to within 2 mV with a 5-bit IDAC. The IDAC area is roughly proportional to the maximum range of the offset cancellation, assuming a fixed resolution. The additional chip area for dc offset cancellation will be directly proportional to the variability, assuming that the IDAC itself is not affected by variability.

Finally, hot-electron/NBTI lifetime stress affects analog circuits in more subtle ways than are observed in digital circuits. A typical analog block may consist of CML circuits (usually consisting of resistor and n-FETs) for highest performance and custom digital CMOS circuits for reduced power and area. These various blocks see different operating points (duty cycles, voltage swings, bias currents, etc.) over their lifetime and may respond differently to hot-electron and NBTI stresses. Determining end-of-life conditions for the ensemble of components and the effect on overall circuit performance is more complex than is the case for standard, generalized inverter CMOS logic circuits. The hot-electron and NBTI degradation for each device is calculated for several cycles under the worst stress condition and extrapolated for lifetime cycles. The resulting circuit netlist with "end-of-life" degraded electrical parameters is then simulated for corners and statistics. **Figure 13** shows simulated ranges for the new and the end-of-life (i.e., after elevated voltage stress screens, burn-in, and 200k-hour life stress) relative delay between the clock and 10-bit data at the analog–digital interface in a receiver [2]. The 10-bit data are from different paths for the data, timing, and edge information

[1]T. Beukema et al., "A 6.4 Gb/s CMOS SerDes Core with Feedforward and Decision-Feedback Equalization," submitted to *J. Solid-State Circuits*.

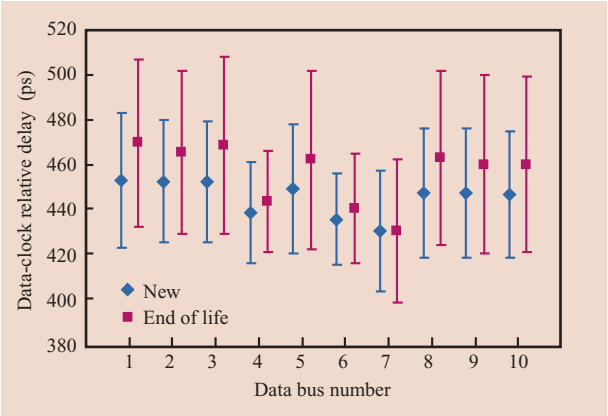

Figure 13

Mixed-signal circuit responses to lifetime HotE/NBTI stress and resulting data path variability. Simulated variation of timing difference between clock and ten data signals at analog–digital interface.

in the decision feedback equalization architecture, and are generated with different clock stages from the final clock. Figure 13 shows that the end-of-life relative timing can be either slower or faster, but always has an increased range, representing additional variability.

Emerging technologies

New device features and architectures

Scaling of CMOS technology over the past twenty years has pushed a number of variability mechanisms out of the "negligible" regime, to the point where today they have become significant factors in circuit design. Two obvious examples are random dopant fluctuations and gate-dielectric leakage. In this section we explore the role that some process technology features, ranging from very recent to exploratory deployment, may play in variability. New features can be categorized as comprising either new structures or new materials (and sometimes both). Of benefits to the technology, an alternative classification comprises improvements to transport mechanisms and enhancements of short-channel behavior. **Table 3** illustrates both classifications as a matrix with some entries for potential new elements. The table addresses new, potentially significant sources of variation in CMOS as these mechanisms might be introduced.

Uniaxial strain is already employed in 90-nm CMOS [34, 35], using process schemes including straining films overlaying the FETs, SiGe regrown source/drain regions of p-FETs, and so-called stress memorization (via transfer of strain from overlaid films). In all of these cases, the mobility of carriers is increased by the

Table 3 Advanced device improvements (ETSOI: extremely thin SOI).

Feature	*Materials*	*Structure*
Transport improvement	Uniaxial strain SGOI Germanium Compound semiconductors	Uniaxial strain Hybrid orientation technology Ballistic transport
Short-channel behavior improvement	High-k gate Metal gate	FinFET/Tri-Gate ETSOI Back-gating

introduction of tensile strain or compressive strain to enhance electron or hole mobility, respectively. Many structural details of each FET, such as placement of adjacent gates, number of vias, distance to other FETs, and proximity to isolation, conspire to vary the effective mobility, threshold voltage, and subthreshold leakage of each FET on a given die [36]. Previously, this mechanism was present in CMOS technology arising from unwanted, residual strain (e.g., in trench fill) in new technologies. This mechanism is being cultivated and enhanced, and thus the magnitude of variation can be much greater. Some of this variability can be predicted with advanced CAD tools and thus effectively reduced to residual errors with respect to the model predictions.

One method of obtaining uniaxial strain, namely the case of SiGe regrown source/drain regions, introduces the potential for added variability in short-channel effects, due to local variations in etch depth of the source/drain regions prior to growth of the SiGe. Junction profiles are modulated by several mechanisms, dependent in magnitude on details of the implementation, and the effective junction depth and halo dose controlling short-channel effects cause variations across a die.

Hybrid orientation technology [37] introduced p-FETs selectively on {110} silicon planes, where hole mobility is considerably higher than in the traditional {100} plane for CMOS channels. The mobility is not isotropic within this plane, however, and current must be in a {100} direction with the plane in order to obtain the full benefit of the technology. Thus, deviations in current direction from gate-to-wafer alignment add a new degree of variability to the p-FET drive current.

Metal-gate effects may depend strongly on which of several structural schemes being pursued is considered. Metal-gate options include pure metals with an intrinsic workfunction, alloys, and doped metal alloys. The use of alloys and/or dopant in metal gates for V_t control may introduce variability from alloy composition, micro-domains, and random dopant fluctuation; hence, such schemes require close scrutiny to determine the quantitative behavior of V_t variation as a function of physical gate size.

The value of the gate workfunction or a given V_t is key to determining whether the device subthreshold conduction is at the surface of the channel, or significantly "buried" away from the gate–dielectric interface. Buried-channel operation (in subthreshold) is relevant to this discussion simply because additional variability is introduced in these circumstances. Subthreshold swing, V_t, and off-current all change with channel depth; variable short-channel effects compound these effects. Hence, any proposal to use workfunction/V_t combinations that result in buried subthreshold operation requires great care to ensure that the burden added from variability does not negate intrinsic gains derived from the new device structures and materials.

High-k is an especially challenging case. New variations might be anticipated on many fronts, including variations in dielectric constant and thickness, as well as variations in fixed charge and surface states. It is known, furthermore, that new aging mechanisms are being introduced, potentially resulting in additional variation with aging.

Thin silicon channels

Several structural advances aim to reduce short-channel effects, such as drain-induced barrier lowering (DIBL) and subthreshold swing, by the use of a very thin silicon channel which is fully depleted of majority carriers during operation. These include extremely thin SOI (ETSOI), double-gate (DG), and back-gate (BG) architectures, which may be achieved by means of many structural proposals. Since several common variability mechanisms are shared by these architectures, all of the structures in which they can be embodied are susceptible. We first review these mechanisms, comparing the intrinsic strengths and weaknesses of each architecture, and then discuss how some recent structural proposals affect these variations.

Two fundamental changes to V_t variation mechanisms are introduced by fully depleted (FD) devices. The first modification arises from the V_t dependence on the first power of body doping exhibited by the FD device in contrast to conventional bulk or PDSOI devices. There, V_t varies as the 0.4 power of the body doping. This is because the compensating factor in partially depleted (PD) FETs, which captures depletion depth change with doping, does not exist in fully depleted devices. As a result, V_t may vary more strongly with doping variation, such as random doping fluctuations (RDFs). Second, an entirely new factor, the body thickness variation, is introduced. Since these devices are fully depleted, changes in body thickness result in changes in the charge in the body (unless some type of self-compensating process

scheme is employed) which, by Gauss' law, result in changes in channel potential, and thus changes in V_t. These two factors must be quantitatively analyzed in any thin-silicon FET to ensure that the additional variations introduced do not overwhelm intrinsic gains delivered by the structure.

Two other variation factors intrinsically accompany the thin-silicon class of FETs: extrinsic (series) resistance (Rext) [38] and, at the limits of scaling, atomic fluctuations in body thickness, similar to the random-dopant fluctuation problem [39]. The Rext problem is driven by the difficulty in forming low-resistance paths to the channel from the contacts to the source and drain. In these FETs the silicon is so thin that raised-source/drain or other similar structures are required for low resistance. These process additions then provide a new source of current and transconductance variation. The limits dictated by atomic fluctuations in body thickness compete with limitations due to variations in FET behavior from confinement effects. For silicon thicknesses comparable to 5 nm or less, the confinement of the inversion layer is small enough to significantly raise the inversion-state energy levels and alter V_t and mobility. However, at these thicknesses the area scale of such FETs can give rise to a few hundred silicon atoms comprising the channel; thus, random fluctuations in silicon thickness are also likely to play a role analogous to that of the RDF in today's state-of-the-art CMOS technology SRAMs.

Thus far we have discussed properties shared by the class of thin-silicon FETs. We next examine structures that are of interest for implementing thin-silicon device architectures.

FDSOI

FDSOI has been dominated by an extension of conventional PDSOI, simply to thinner silicon layers, typically between 5 nm and 20 nm. One notable exception is the so-called "silicon-on-nothing" structure [40], in which a void is selectively formed under the channel to provide a very thin silicon region. An entirely new mechanism introduced by these two structures is a sensitivity to surface states and charge at and below the back silicon interface. Variations in swing, V_t, and mobility may result. Charging from the "antenna effect" during interconnect processes may introduce such degradation and complicate variability immensely. Damage from ionizing radiation may also introduce new variation with age.

DELTA transistor, or FinFET

The DELTA FET [41], popularly known as a FinFET [42], is a promising candidate among double-gate architectures. The body-thickness issues discussed earlier may be most challenging for this structure, since its body thickness is defined by a lateral lithographic process; such techniques typically present poorer tolerances than those of thin-film deposition used for planar devices. Variations in the thickness of the fin height result in variations of FET width. An interesting consequence is that global variations of this thickness result in all FETs changing in width by the same percentage, in complete contrast to the case of a planar architecture, where global changes result in all devices varying by the same absolute dimension. In planar devices, wider transistors are employed because better width tolerance is required; in FinFET technology, wider widths are achieved by increased numbers of fins, and the tolerance remains the same, regardless of width. However, wider (FinFET) devices do suppress one (new) variation—the variation caused by edge fin characteristics vs. non-edge fins within a single FET.

Tri-Gate

The Tri-Gate [43] structure is a short FinFET with thin gate dielectric on top of the fin as well as on the sides. By keeping the aspect ratio in the vicinity of 1:1, the width of the fin can be somewhat large for the same short-channel-effect suppression as in an equivalently tall FinFET. This factor relieves some pressure from the lithographic demands of FinFETs, and thus can reduce sensitivity to lithography-induced body-thickness variations. On the negative side, however, this FET is further subjected to strong V_t dependence on fin height, since this dimension plays an active role in channel potential by design. Thus, the silicon-thickness-induced variation terms now have two degrees of freedom rather than one, and the width tracking of FETs displays some aspects of the planar device (i.e., the top of the fin is conventional in its width dependence on lithography) and some aspects of the FinFET (i.e., the global fin height width dependence).

Furthermore, the Tri-Gate structure is subject to the same surface state and charge sensitivities of the bottom silicon interface as those in ETSOI.

It is not widely appreciated in the literature that simulations of Tri-Gate FETs attribute most of the advantage in performance of this structure to the inherent superior short-channel effects of a corner-geometry channel. This is similar to the parasitic channel that can sometimes be observed in a planar FET with trench isolation. To the degree that this artifact is featured in the device architecture, it also introduces a potentially significant variability mechanism. The V_t and other short-channel characteristics of the corner depend strongly on the shape, or radius of curvature, of this corner. Additionally, the transport of inversion carriers at the corner interface and its dependence on the crystalline orientations are likely to result in corner transconductance variations as well.

Back gate

Back-gate transistors (BGFETs) hold promise for relief from RDF, since V_t can be set by the back-gate potential, reducing dependence on channel doping. However, channel doping, in the form of halo, or pocket, doping from the source and drain areas, has been relied on for many generations of CMOS to flatten V_t as a function of L_{gate}. Thus, BGFETs introduce an increased sensitivity of V_t to variation in L_{gate}. While the mean variation of L_{gate} of a given die can be compensated by suitable control of the back-gate voltage, the variation of L_{gate} within the die (across-chip linewidth variation, ACLV) cannot so easily be "tuned out." Thus, the reduction in RDF-driven V_t variation is offset by additional ACLV-driven V_t variation within a die. A nonplanar structure suggested for the implementation of BGFETs is the split-gate FinFET [44] or FT-FinFET [45]. Such FinFETs are constructed with the two sides of the gate disconnected from each other to provide independent gate control. It is clear that the mechanisms visited in the preceding discussion on FinFETs would apply equally here.

Summary of emerging technologies

A number of new device architectures and structures for achieving these architectures hold promise to enable further progress in CMOS technology improvement. Each of these carries new mechanisms for variability, both systematic and random. Careful quantitative studies of these issues are required to demonstrate that the benefits derived from each structure are not excessively compromised by the added variability.

The outlook for future timing precision

New device structures and materials may allow CMOS to scale further, but variability is unlikely to decrease, since smaller devices contain fewer atoms and consequently exhibit less self-averaging. The situation may be improved by removing most of the doping, which is the largest source of intrinsic variations, but there will still be interfaces, which also exhibit randomness. In aggressively scaled devices, there is always an interface nearby, and this may become the dominant source of variability. On the processing side, variation can be reduced through the learning that goes into steadily improving manufacturing yield, but cost tradeoffs dictate that variability will be reduced no more than is absolutely necessary to keep CMOS processing profitable for its developers.

Conclusions

The inability to scale the tolerance of multiple electrical parameters along with their nominal value has contributed to a virtual crisis in the ability to improve performance and reduce power consumption in new processes. The continued infusion of new materials and structures provides an illusion of conventional scaling, but presents additional idiosyncrasies as well. Anticipation of these mechanisms and their influence on variability is critical. Circuit and architecture design innovation will enable the extension of CMOS technology beyond currently recognized limits. These lessons will be important in addressing the more profound challenges of novel emerging technologies.

Acknowledgments

We gratefully acknowledge contributions by Huifang Qin, Paul Friedberg, Ruth Wang (UC Berkeley) and Ronald Bolam (IBM Burlington, Vermont).

References

1. B. E. Stine, D. S. Boning, and J. E. Chung, "Analysis and Decomposition of Spatial Variation in Integrated Circuit Processes and Devices," *IEEE Trans. Semicond. Manuf.* **10**, No. 1, 24–41 (1997).
2. D. J. Frank, Y. Taur, and H.-S. P. Wong, "Generalized Scale Length for Two-Dimensional Effects in MOSFETs," *IEEE Electron Device Lett.* **19**, No. 10, 385–387 (1998).
3. A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of Intrinsic Parameter Fluctuations in Decananometer and Nanometer-Scale MOSFETs," *IEEE Trans. Electron Devices* **50**, 1837 (2003).
4. D. J. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H.-S. P. Wong, "Device Scaling Limits of Si MOSFETs and Their Application Dependencies," *Proc. IEEE* **89**, 259–288 (2001).
5. T. Mizuno, J. Okamura, and A. Toriumi, "Experimental Study of Threshold Voltage Fluctuation Due to Statistical Variation of Channel Dopant Number in MOSFET's," *IEEE Trans. Electron Devices* **41**, 2216 (1994).
6. P. A. Stolk and D. B. M. Klaassen, "The Effect of Statistical Dopant Fluctuations on MOS Device Performance," *IEDM Tech. Digest*, p. 627 (1996).
7. P. A. Stolk, F. P. Widdershoven, and D. B. M. Klaassen, "Modeling Statistical Dopant Fluctuations in MOS Transistors," *IEEE Trans. Electron Devices* **45**, 1960 (1998).
8. H.-S. P. Wong, Y. Taur, and D. J. Frank, "Discrete Random Dopant Distribution Effects in Nanometer-Scale MOSFETs," *Microelectron. Reliabil.* **38**, 1447–1456 (1998).
9. D. J. Frank, Y. Taur, M. Ieong, and H.-S. P. Wong, "Monte Carlo Modeling of Threshold Variation Due to Dopant Fluctuations," *Symp. VLSI Technol.*, pp. 169–170 (1999).
10. D. J. Frank and H.-S. P. Wong, "Simulation of Stochastic Doping Effects in Si MOSFETs," *Proceedings of the International Workshop on Computational Electronics*, 2000, pp. 2–3.
11. M. Hane, T. Ikezawa, and T. Ezaki, "Atomistic 3D Process/Device Simulation Considering Gate Line-Edge Roughness and Poly-Si Random Crystal Orientation Effects," *IEDM Tech. Digest*, pp. 9.5.1–9.5.4 (2003).
12. H.-S. P. Wong, D. J. Frank, P. M. Solomon, H.-J. Wann, and J. Welser, "Nanoscale CMOS," *Proc. IEEE* **87**, 537–570 (1999).
13. T. Brunner, "Why Optical Lithography Will Live Forever," *J. Vac. Sci. Technol. B* **21**, No. 6, 2632–2637 (2003).

14. J. Cobb, F. Houle, and G. Gallatin, "The Estimated Impact of Shot Noise in Extreme Ultraviolet Lithography," *Proc. SPIE* **5037**, 397 (2003).
15. A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic Parameter Fluctuations in Decananometer MOSFETs Introduced by Gate Line Edge Roughness," *IEEE Trans. Electron Devices* **50**, 1254 (2003).
16. K. A. Bowman, T. Xinghai, J. C. Eble, and J. D. Meindl, "Impact of Extrinsic and Intrinsic Parameter Variations on CMOS System on a Chip Performance," *Proceedings of the Twelfth Annual IEEE International ASIC/SOC Conference*, 1999, pp. 267–271.
17. G. Roy, F. Adamu-Lema, A. R. Brown, S. Roy, and A. Asenov, "Intrinsic Parameter Fluctuations in Conventional MOSFETs Until the End of the ITRS: A Statistical Simulation Study," *Proceedings of the 7th International Conference on New Phenomena in Mesoscopic Systems and the 5th International Conference on Surfaces and Interfaces in Mesoscopic Devices (NPMS/SIMD)*, Maui, HI, 2005, pp. 35–36.
18. V. Mehrotra, S. Nassif, D. Boning, and J. Chung, "Modeling the Effects of Manufacturing Variation on High-Speed Microprocessor Interconnect Performance," *IEDM Tech. Digest*, pp. 767–770 (1998).
19. S. Nassif, "Within-Chip Variability Analysis," *IEDM Tech. Digest*, pp. 283–286 (1998).
20. C. E. Blat, E. H. Nicollian, and E. H. Poindexter, "Mechanism of Negative Bias Temperature Instability," *J. Appl. Phys.* **69**, No. 3, 1712–1720 (1991).
21. E. Takeda, "Hot-Carrier Effects in Submicrometer MOS VLSIs," *IEE Proc.* **131**, 153–162 (1984).
22. D. Young and A. Christou, "Failure Mechanism Models for Electromigration," *IEEE Trans. Reliabil.* **43**, No. 2, 186–192 (1994).
23. H. H. Chen and D. D. Ling, "Power Supply Noise Analysis Methodology for Deep-Submicron VLSI Chip Design," *Proceedings of the IEEE Design Automation Conference*, 1996, pp. 638–643.
24. H. Su, F. Liu, A. Devgan, E. Acar, and S. R. Nassif, "Full Chip Leakage Estimation Considering Power Supply and Temperature Variations," *Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED)*, 2003, pp. 78–83.
25. S. Nassif and J. Kozhaya, "Fast Power Grid Simulation," *Proceedings of the IEEE Design Automation Conference*, 2000, pp. 156–161.
26. S. Nassif, A. Gattiker, C. Long, and R. Dinakar, "Timing Yield Estimation from Static Timing Analysis," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, 2001, pp. 437–442.
27. J. A. G. Jess, K. Kalafala, S. R. Naidu, R. H. J. M. Otten, and C. Visweswariah, "Statistical Timing for Parametric Yield Prediction of Digital Integrated Circuits," *Proceedings of the IEEE Design Automation Conference*, 2003, pp. 932–937.
28. N. J. Rohrer, M. Canada, E. Cohen, M. Ringler, M. Mayfield, P. Sandon, P. Kartschoke, J. Heaslip, J. Allen, P. McCormick, T. Pfluger, J. Zimmerman, C. Lichtenau, T. Werner, G. Salem, M. Ross, D. Appenzeller, and D. Thygesen, "PowerPC* 970 in 130 nm and 90 nm Technologies," *Digest of Technical Papers, IEEE Solid-State Circuits Conference*, 2004, pp. 68–69.
29. J. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*, 2nd ed., Prentice-Hall, Inc., New York, 2003.
30. P. Kogge and H. Stone, "A Parallel Algorithm for the Efficient Solution of a General Class of Recurrence Equations," *IEEE Trans. Computers* **C-22**, 786–793 (1973).
31. T. Han and D. Carlson, "Fast Area-Efficient VLSI Adders," *Proceedings of the 8th Annual Symposium on Computer Arithmetic*, 1982, pp. 49–56.
32. R. Brent and H. Kung, "A Regular Layout for Parallel Adders," *IEEE Trans. Computers* **C-31**, 260–264 (1982).
33. T. Beukema, M. Sorna, K. Selander, S. Zier, B. L. Ji, P. Murfet, J. Mason, W. Rhee, H. Ainspan, B. Parker, and M. Beakes, "A 6.4Gb/s CMOS SerDes Core with Feedforward and Decision-Feedback Equalization," *Digest of Technical Papers, IEEE Solid-State Circuits Conference*, 2005, pp. 2633–2645.
34. V. Chan, R. Rengarajan, N. Rovedo, W. Jin, T. Hook, P. Nguyen, J. Chen, E. Nowak, X.-D. Chen, D. Lea, A. Chakravarti, V. Ku, S. Yang, A. Steegen, C. Baiocco, P. Shafer, H. Ng, S.-F. Huang, and C. Wann, "High Speed 45nm Gate Length CMOSFETs Integrated into a 90nm Bulk Technology Incorporating Strain Engineering," *IEDM Tech. Digest*, pp. 77–80 (2003).
35. S. Thompson, N. Anand, M. Armstrong, C. Auth, B. Arcot, M. Alavi, P. Bai, J. Bielefeld, R. Bigwood, J. Brandenburg, M. Buehler, S. Cea, V. Chikarmane, C. Choi, R. Frankovic, T. Ghani, G. Glass, W. Han, T. Hoffmann, M. Hussein, P. Jacob, A. Jain, C. Jan, S. Joshi, C. Kenyon, J. Klaus, S. Klopcic, J. Luce, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, P. Nguyen, H. Pearson, T. Sandford, R. Schweinfurth, R. Shaheed, S. Sivakumar, M. Taylor, B. Tufts, C. Wallace, P. Wang, C. Weber, and M. Bohr, "A 90 nm Logic Technology Featuring 50 nm Strained Silicon Channel Transistors, 7 Layers of Cu Interconnects, Low k ILD, and 1 μm^2 SRAM Cell," *IEDM Tech. Digest*, pp. 61–64 (2002).
36. S. Eneman, P. Verheyen, R. Rooyackers, F. Nouri, L. Washington, R. Degraeve, B. Kaczer, V. Moroz, A. De Keersgieter, R. Schreutelkamp, M. Kawaguchi, Y. Kim, A. Samoilov, L. Smith, P. P. Absil, K. De Meyer, M. Jurczak, and S. Biesemans, "Layout Impact on the Performance of a Locally Strained PMOSFET," *Symp. VSLI Technol.*, pp. 22–23 (2005).
37. M. Yang, M. Ieong, L. Shi, K. Chan, V. Chan, A. Chou, E. Gusev, K. Jenkins, D. Boyd, Y. Ninomiya, D. Pendleton, Y. Surpris, D. Heenan, J. Ott, K. Guarini, C. D'Emic, M. Cobb, P. Mooney, B. To, N. Rovedo, J. Benedict, R. Mo, and H. Ng, "High Performance CMOS Fabricated on Hybrid Substrate with Different Crystal Orientations," *IEDM Tech. Digest*, pp. 453–456 (2003).
38. M. Ieong, H. S.-P. Wong, E. Nowak, J. Kedzierski, and E. C. Jones, "High Performance Double-Gate Device Technology Challenges and Opportunities," *Proceedings of the International Symposium on Quality Electronic Design*, 2002, pp. 492–495.
39. H. Mahmoodi and S. Mukhopadhyay, "Estimation of Delay Variations Due to Random-Dopant Fluctuations in Nanoscale CMOS Circuits," *IEEE J. Solid-State Circuits* **40**, No. 9, 1787–1796 (2005).
40. T. Sato, H. Nii, M. Hatano, K. Takenaka, H. Hayashi, K. Ishigo, T. Hirano, K. Ida, N. Aoki, T. Ohguto, K. Ino, I. Mizushima, and T. Tsunashima, "SON (Silicon on Nothing) MOSFET Using ESS (Empty Space in Silicon) Technique for SoC Applications," *IEDM Tech. Digest*, pp. 809–812 (2001).
41. D. Hisamoto, T. Kaga, Y. Kawamoto, and E. Takeda, "A Fully Depleted Lean-Channel Transistor (DELTA)—A Novel Vertical Ultra Thin SOI MOSFET," *IEDM Tech. Digest*, pp. 833–836 (1989).
42. X. Huang, W. C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y. K. Choi, K. Asano, V. Subramanian, T. J. King, J. Bokor, and C. Hu, "Sub 50-nm FinFET: PMOS," *IEDM Tech. Digest*, pp. 67–70 (1999).
43. B. Doyle, B. Boyanov, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, R. Rios, and R. Chau, "Tri-Gate Fully Depleted CMOS Transistors: Fabrication, Design, and Layout," *Symp. VLSI Technol.*, pp. 133–134 (2003).
44. D. M. Fried, E. J. Nowak, J. Kedzierski, J. S. Duster, and K. T. Komegay, "A Fin-Type Independent-Double-Gate NFET," *Proceedings of the Device Research Conference*, 2003, pp. 45–46.

45. Y. X. Liu, M. Masahara, K. Ishii, T. Tsutsumi, T. Sekigawa, H. Takashima, H. Yamauchi, and E. Suzuki, "Flexible Threshold Voltage FinFETs with Independent Double Gates and an Ideal Rectangular Cross-Section Si-Fin Channel," *IEDM Tech. Digest*, pp. 986–988 (2003).

Received September 30, 2005; accepted for publication May 1, 2006; Internet publication August 6, 2006

Kerry Bernstein *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (kbernste@us.ibm.com)*. Mr. Bernstein is a Senior Technical Staff Member at the IBM Thomas J. Watson Research Center. He is currently responsible for future product technology definition, performance, and application. He received his B.S. degree in electrical engineering from Washington University in St. Louis, joining IBM in 1978. He holds 50 U.S. patents and is a coauthor of three college textbooks and multiple papers on high-speed and low-power CMOS. Mr. Bernstein is currently interested in the area of high-performance, low-power advanced circuit technologies. He is a Senior Member of the IEEE, and is a staff instructor at RUNN/Marine Biological Laboratories, Woods Hole, Massachusetts.

David J. Frank *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (djf@us.ibm.com)*. Dr. Frank received a B.S. degree from the California Institute of Technology in 1977 and a Ph.D. degree in physics from Harvard University in 1983. Since graduation, he has worked at the IBM Thomas J. Watson Research Center, where he is a Research Staff Member. His studies have included non-equilibrium superconductivity, III–V devices, and exploring the limits of scaling of silicon technology. His recent work includes the modeling of innovative Si devices, analysis of CMOS scaling issues such as power consumption, discrete dopant effects and short-channel effects associated with high-k gate insulators, exploring various nanotechnologies, investigating the usefulness of energy-recovering CMOS logic and reversible computing concepts, and low-power circuit design. Dr. Frank is an IEEE Fellow; he has served as chairman of the Si Nanoelectronics Workshop and is an associate editor of the *IEEE Transactions on Nanotechnology*. He has authored or co-authored more than 90 technical publications and holds nine U.S. patents.

Anne E. Gattiker *IBM Research Division, Austin Research Laboratory, 11501 Burnet Road, Austin, Texas 78758 (gattiker@us.ibm.com)*. Dr. Gattiker holds a Ph.D. degree from Carnegie Mellon University, where she was a National Science Foundation Fellow. Since joining IBM in 1998, she has worked in the IBM Worldwide Test Engineering group in Burlington, Vermont, and at the IBM Austin Research Laboratory in Austin, Texas, where she is a Research Staff Member. Her research interests include design-for-manufacturability and variability characterization, as well as defect-based test, reliability screens, and defect diagnosis. Dr. Gattiker has published more than twenty technical papers, has participated on numerous conference panels, and has been a co-winner of best paper awards at the IEEE International Test Conference and the IEEE International Conference on Microelectronic Test Structures. She is the ITC Technical Program Chair in 2006 and is on the program committees of ITC and the ASM International Symposium on Testing and Failure Analysis.

Wilfried Haensch *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (whaensch@us.ibm.com)*. In 1981, Dr. Haensch received his Ph.D. degree from the Technical University of Berlin, Germany, in the field of theoretical solid-state physics. In 1984 he joined Siemens Corporate Research in Munich to investigate high-field transport in MOSFET devices, and in 1988 he joined the DRAM development team at the Siemens Research Laboratory to investigate new cell concepts. In 1990, he joined the DRAM alliance between IBM and Siemens to develop quarter-micron 64M DRAM. In this capacity, Dr. Haensch was involved with device

characterization of shallow-trench bounded devices and cell-design concerns. In 1996, he moved to a manufacturing facility to build various generations of DRAM. His primary mission was to transfer technologies from development into manufacturing and to guarantee a successful yield ramp of the product. In 2001, Dr. Haensch joined the IBM Thomas J. Watson Research Center to lead a group concerned with novel devices and applications. He is currently responsible for post-45-nm-node device design and its implications for circuit functionality.

Brian L. Ji *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (blji@us.ibm.com)*. Dr. Ji received a B.S. degree from the University of Science and Technology of China, Hefei, in 1984, and a Ph.D. degree in physics from Harvard University in 1991. From 1991 to 1994, he was a research scientist at SUNY at Stony Brook, where he studied nanofabrication, single-electron memory/logic devices, and superconducting devices. He was a visiting scientist in physical sciences at the IBM Thomas J. Watson Research Center in 1995. The following year he joined the IBM Microelectronics Division in Hopewell Junction, New York, where he worked on several projects in VLSI circuit design, test, and product definition, including 256-Mb, 512-Mb, and 1-Gb DRAMs, and the logic-based embedded memory. From 2001 to 2005 Dr. Ji was involved in analog design for high-speed serial link products. Since 2005, he has been a researcher at the IBM Thomas J. Watson Research Center, studying a number of issues in silicon devices, circuits, and systems, including technology variability, SOI SRAM, and exploratory low-power circuits.

Sani R. Nassif *IBM Research Division, Austin Research Laboratory, 11501 Burnet Road, Austin, Texas 78758 (nassif@us.ibm.com)*. Dr. Nassif received a Ph.D. degree from Carnegie Mellon University in the 1980s. He worked for ten years at Bell Laboratories on various aspects of design and technology coupling, including device modeling, parameter extraction, worst-case analysis, design optimization, and circuit simulation. In 1996 Dr. Nassif joined the IBM Austin Research Laboratory, where he currently manages the Tools and Technology Department, which is focused on design/technology coupling and includes activities in model-to-hardware matching, simulation and modeling, physical design, statistical modeling, statistical technology characterization, and similar areas.

Edward J. Nowak *IBM Systems and Technology Group, 1000 River Street, Essex Junction, Vermont 05452 (ejnowak@us.ibm.com)*. Dr. Nowak received his B.S. degree in physics in 1973 from M.I.T., and M.S. and Ph.D. degrees, also in physics, from the University of Maryland in 1975 and 1978, respectively. In 1981, following postdoctoral research at New York University, he joined IBM in Essex Junction, Vermont, to work on DRAM development. Since 1985, Dr. Nowak has worked in high-performance CMOS device design. His current interests include energy-driven device design and FinFET device architectures.

Dale J. Pearson *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (dale_ pearson@us.ibm.com)*. Mr. Pearson received a B.S. degree in chemistry from Texas Lutheran University in 1979 and an M.S. degree from the University of Wisconsin at Madison in 1981. After working in the IBM Microelectronics Division and the General Electric Corporation, in 1984 he joined the IBM Research Division, where he has worked and directed efforts on Cu VLSI wiring, process technology for low-T_c superconducting circuits, communications VLSI circuit design, and product development and circuit techniques to manage and mitigate VLSI process variability. Mr. Pearson is currently the Associate Director for Systems Research at the IBM Thomas J. Watson Research Center.

Norman J. Rohrer *IBM Systems and Technology Group, 1000 River Street, Essex Junction, Vermont 05452 (rohrern@us.ibm.com)*. Dr. Rohrer is a Distinguished Engineer in the IBM Systems and Technology Group, Essex Junction, Vermont. In 1987 he received a bachelor's degree in physics and mathematics from Manchester College, North Manchester, Indiana. He received a master's degree and a Ph.D. degree in electrical engineering from Ohio State University, Columbus, in 1990 and 1992, respectively. Dr. Rohrer has been a lead designer on PowerPC 750 and 970 products for the Apple G3 and G5 chips and the Nintendo GameCube**. His interests lie in the area of high-speed circuit optimization for future technologies. He holds 22 patents and is a coauthor on two books titled *High Speed CMOS Circuit Design Styles* and *SOI Circuit Design Concepts*. Dr. Rohrer has been a Senior Member of the IEEE since 2003.