

Optimization of silicon technology for the IBM System z9

IBM 90-nm silicon-on-insulator (SOI) technology was used for the key chips in the System z9™ processor chipset. Along with system design, optimization of some critical features of this technology enabled the z9™ to achieve double the system performance of the previous generation. These technology improvements included logic and SRAM FET optimization, mask fabrication, lithography and wafer processing, and interconnect technology. Reliability improvements such as SRAM optimization and burn-in reliability screen are also described.

D. J. Poindexter
S. R. Stiffler
P. T. Wu
P. D. Agnello
T. Ivers
S. Narasimha
T. B. Faure
J. H. Rankin
D. A. Grosch
M. D. Knox
D. C. Edelstein
M. Khare
G. B. Bronner
H.-J. Nam
S. A. Butt

Introduction

A key goal of the System z9* was to provide twice the system-level performance of the prior zSeries* 990. To achieve this goal, the chip technology and design would have to provide more than 40% improvement in frequency to 1.72 GHz over the predecessor z990 system, with added logic and on-chip memory and without an increase in overall power consumption. The prior-generation z990 system [1] was designed using the IBM 130-nm silicon-on-insulator (SOI) chip technology with eight levels of metal interconnect. To achieve the z9* objectives, the 90-nm SOI technology with ten levels of Cu metallurgy was used, resulting in a 1.4X nominal performance improvement over the 130-nm technology [2]. Several technical innovations were required to enable this performance gain; the most significant were highly strained tensile n-FET liner, advanced gate-patterning techniques, hierarchical back-end-of-line (BEOL) using a low-*k* insulator, and electrically programmable fuses.

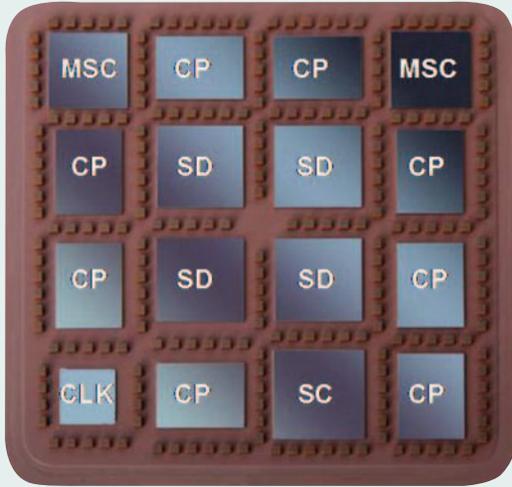
At the center of the z9 system is the multichip module (MCM) containing the processor electronics. The MCM is an advanced 95-mm × 95-mm glass ceramic module that contains sixteen chips interconnected with more than half a kilometer of internal cross-chip wiring. Fifteen of the sixteen chip sites use the 90-nm SOI technology: eight dual-core central processor (CP) chips, four SD chips (L2 data cache, 98 Mb custom), one storage control (SC) chip, and two memory storage control (MSC) chips (L2 to L3 interface). The sixteenth chip is the clock using IBM 130-nm bulk CMOS technology. The z9 system, when fully configured as a 54-way system, comprises four

MCMs, each containing 16 dual-core processor cores running at 1.72 GHz. **Figure 1** is a schematic diagram of the MCM.

The scaling of high-performance process technologies from 180 nm to 130 nm to 90 nm produced increased performance and density but also resulted in increased device leakage currents and stability concerns for the SRAM cell. Indeed, in 90-nm technology, leakage currents and their associated power can be comparable to active switching power in typical designs [3]. Given that a) the multichip module has an upper power limit; b) each chip on the MCM affects the system-level performance uniquely; and c) each chip has its own power/performance relationship, there was a requirement to optimize the entire chipset to allow the highest possible system-level performance at an acceptable power level. **Table 1** shows the percentage of MCM power used by each of the chip types (note that Figure 1 shows multiple instances of each type of chip). The power supply of the system was also optimized for these purposes, allowing tradeoffs to be made among active power, passive power, and cycle time. System-level reliability requirements dictated that these designs undergo an aggressive burn-in, with increased voltage and temperature (producing up to 400 W per chip), prior to system build. Controlling the leakage currents in the chips under these severe conditions also presented a number of issues which are described in the paper.

The above combination of challenges drove an aggressive and systematic approach to process window evaluation. Process window evaluation entails

©Copyright 2007 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.



Chip name	CP	MSC	SD	SC
Function	Central processor	L2 to L3 interface	L2 data cache	L2 storage control
Size (mm × mm)	11.84 × 15.78	14.3 × 14.3	15.4 × 15.6	16.4 × 16.4
Technology	90-nm SOI	90-nm SOI	90-nm SOI	90-nm SOI
Metal levels (4-2-2-2)/ low- <i>k</i>	10 Cu	10 Cu	10 Cu	10 Cu
Nominal voltage (V)	1.2	1.2	1.2	1.2
Frequency (GHz)	1.72	0.85	0.85	0.85
Est. power (W)	81	26	22	28
SRAM (Mb)	12	–	97.5	17.4
Transistors	121M	25M	600M	162M

Figure 1

System z9 multichip module and 90-nm-technology chips.

understanding the impact of process variations within the specification range on the chip at wafer final test, module, and system test. From a power/performance perspective, device design optimization (to improve the power/performance characteristics of the individual devices), semiconductor and mask technology enhancements (to reduce device-to-device variability across the chips), and device threshold voltage experimentation (to optimize leakage currents at the required performance levels for each of the designs) were all of primary importance. This work was tightly coupled to burn-in characterization and system-level verification, ensuring that the chip-level optimizations produced acceptable characteristics at subsequent levels of assembly.

In this paper we first describe the key features of the 90-nm SOI technology and then discuss the logic and SRAM designs, including optimization of the gate oxide,

programmable fuses, and interconnect technology. In the section on optimization of power and performance, we discuss device optimization, including device design, mask design and fabrication, lithography, and wafer processing. Finally we describe the burn-in screen, which ensures high reliability.

90-nm SOI technology

The key 90-nm SOI technology attributes are listed below [4]:

- CMOS technology on p-SOI substrates.
- Two gate-oxide thicknesses.
 - Thin oxide is nitrided SiO₂ 1.12 nm thick.
 - Thick oxide is 2.2 nm thick.
- FET device types include the following:
 - Regular-threshold-voltage (regular- V_T) n-FET and p-FET (n-FET uses highly stressed liner).
 - Low-threshold-voltage (low- V_T) n-FET and p-FET.
 - High-threshold-voltage (high- V_T) n-FET and p-FET.
 - Thick-oxide n-FET and p-FET.
- Low-resistance Co-silicided n⁺ and p⁺ polysilicon diffusions.
- Shallow-trench isolation (STI).
- Precision resistor.
- Tungsten local interconnect level for connecting diffusions and polysilicon wires/gates.
- Ten levels of Cu metallization:
 - 1X (thin wire) 0.25 μm thick, 0.28 μm minimum pitch.
 - 2X, 4X, and 6X (2X = twice the thickness and pitch of the 1X level).
- Planarized passivation and interlevel dielectrics:
 - Four 1X and two 2X wiring levels insulated with SiCOH ($k = 3.0$).
 - Two 4X and two 6X wiring levels insulated with fluorinated SiO₂ ($k = 3.6$).
- Electrically programmable fuses (eFUSEs).
- C4 terminals.

This technology was practiced in a state-of-the-art, 300-mm fabrication facility (fabricator) [5].

Logic device design

The device technology was first described in 2002 [6]; a cross section is shown in **Figure 2(a)**. In the device design, elements were incorporated to maximize performance while simultaneously controlling short-channel effects and the associated leakage issues. The gate dielectric uses a nitrided oxide measuring 1.12 nm thick. The nitrogen content was tuned to minimize gate leakage and achieve

Table 1 MCM power dissipation by chip type.

	Power dissipated (%)
CP	66
SD	36
SC	14
MSC	4

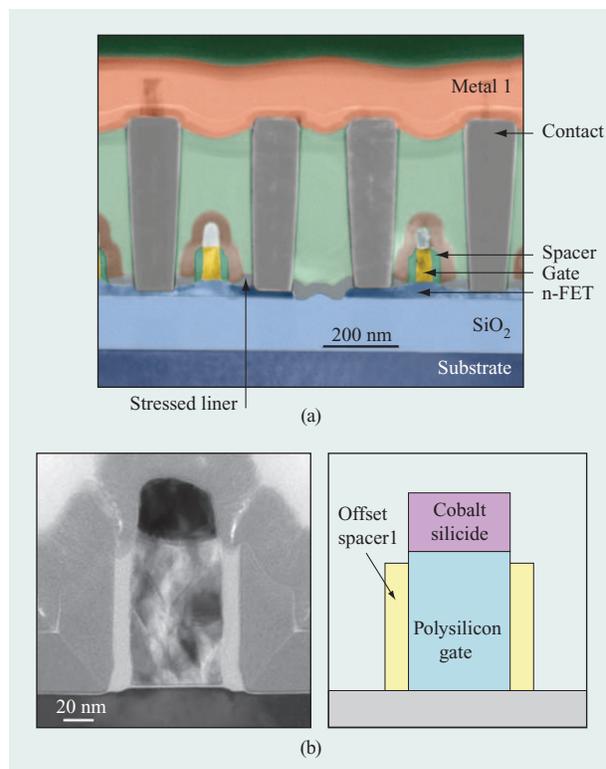
low T_{inv} (electrical equivalent thickness when the device channel is inverted) with minimal degradation of carrier mobility. The gate electrode and source/drain dopings were tuned for optimal gate polysilicon activation and device short-channel control. An extension offset spacer [Figure 2(b)] was implemented to allow for a high-dose extension implant (to minimize the extrinsic resistance of the device) while retaining low gate-to-extension overlap to minimize the parasitic capacitance [6].

Two separate nitride spacer thicknesses were implemented to independently control the source/drain doping and cobalt silicide proximity from the channel for n-FET and p-FET devices (Figure 3). With the implementation of the thin spacer, a substantial reduction in the source/drain resistance was obtained for the n-FET, leading to improved device characteristics. A thicker spacer was retained on the p-FET in order to independently control the short-channel effect.

The logic device menu offered three separate logic device pairs: the high- V_t , regular- V_t , and low- V_t devices. These device pairs had subthreshold leakage currents in the nA/ μm range, tens-of-nA/ μm range, and hundreds-of-nA/ μm range for the high- V_t , regular- V_t , and low- V_t pairs, respectively (all for nominal channel lengths at 1.0 V and 25°C). Associated with the higher leakage, the low- V_t device set offered ~10% more drive current than the regular- V_t device set. Similarly, the high- V_t device set delivered about 15% less drive current than the regular- V_t set. All of the device types were utilized in the designs and allowed the design team to actively trade off power and performance. In general, the low- V_t devices were used very judiciously and inserted only in performance-critical paths. In fact, for the chip with the greatest low- V_t usage (CP), the low- V_t device usage was only 3% of the total device width on the chip. In contrast, the high- V_t devices were used extensively to reduce leakage currents in circuits that were not performance-critical. For example, the CP chip utilized ~30% of the total device width used by high- V_t devices.

SRAM device design

In addition to the logic devices, the technology offered the ability to selectively tune the devices in the SRAM cells.

**Figure 2**

(a) Transmission electron micrograph of cross section of 90-nm front end of line, including n-FETs and contacts. (b) Oxide offset spacer on PC sidewall for short-channel control and optimization of source/drain resistance.

While the technology supported SRAM cells down to 0.99 μm^2 , the cell utilized in System z9, which measured 1.15 μm^2 in area, was chosen for performance, standby power, and stability. Parametric optimization suggested the use of a high- V_t n-FET and an intermediate- V_t p-FET (having a threshold voltage between the regular and high- V_t values) in the SRAM cell that was implemented in the designs. The same six-device-FET SRAM cell is used across the four-z9 90-nm chipset (Figure 1), including the high-performance L1 core arrays on the CP chip and large L2 cache arrays on the SD chip. The CP arrays were designed to operate at twice the frequency of the SD arrays. Array cell stability, performance, and reliability had to be optimized and balanced against one another to meet the z9 application requirements. The 90-nm technology reduced the size of the SRAM cells by ~50% compared with the previous generation. With continued area scaling, SRAM stability has become a critical issue, and the device strengths and threshold voltages were carefully optimized to achieve acceptable stability margins and performance levels. To ensure high

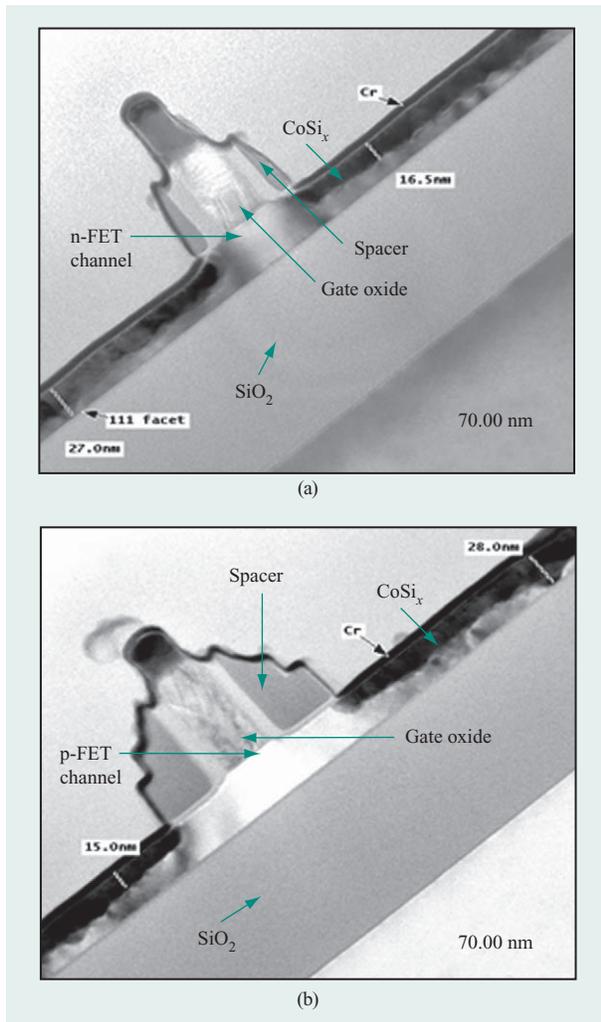


Figure 3

Dual spacer optimization for high-performance 90-nm CMOS technology: (a) Thin spacer and close silicide proximity for n-FETs; (b) wide spacer for p-FETs.

reliability, various monitor structures were stressed; the results showed that the key parameters [time-dependent dielectric breakdown, hot-carrier stress, and negative-bias thermal instability (NBTI)] were within the acceptable limits.

The basic SRAM cell design and reliability of device parameters were also determined to be acceptable. A product-like reliability evaluation was then performed on a 5-Mb SRAM at 1.5 V and 140°C. These initial results revealed a serious V_{\min} degradation even after 2 hr stress. (V_{\min} is the lowest voltage at which an SRAM array still functions; it is a key reliability parameter, since V_{\min} tends to increase over the life of a product due to NBTI [7]). If V_{\min} exceeds the worst-case (low) operating

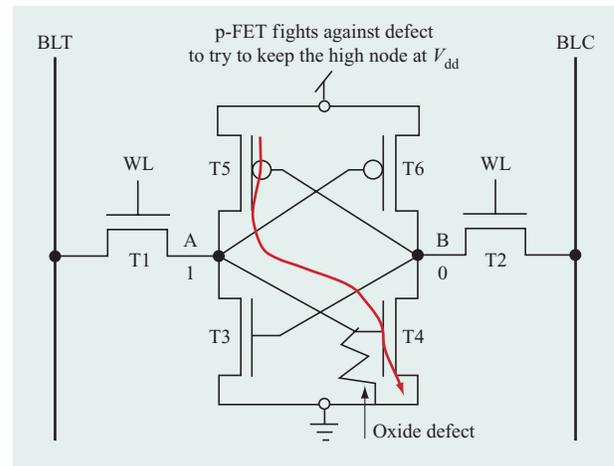


Figure 4

Schematic diagram of six-FET-cell SRAM.

voltage of the chip, a failure will occur in the field. In this study, the mean V_{\min} value increased by more than 170 mV under stress, and standard deviation increased three times its initial value. Further, the initial V_{\min} showed very large shifts, whereas some of the modules with high initial values showed very little shift [Figure 5(a), shown later]. Almost all of these shifts were related to single-cell fails in the module. Since the above results were unacceptable for product reliability, a series of studies were launched in order to understand the root cause of these V_{\min} shifts, to develop process changes to fix them, and to model this behavior in order to predict it in future products.

A series of electrical and physical measurements were performed on failing bit cells. These measurements revealed a clear pattern in which the pull-down n-FET (T3 and T4 in Figure 4) showed an increase of more than two to three orders of magnitude in gate leakage current, indicating that the vast majority of the fails were related to gate oxide breakdown in the pull-down n-FET. Physical analysis on the failing bits did not show any additional defect signatures in the SRAM cell. In Figure 4, we hypothesize a gate defect in T4. The n-FET gate leakage allows node A to drift downward and node B to drift upward, thus affecting the read stability margin of the cell and similarly V_{\min} .

Silicon oxide optimization

Silicon oxide integrity in the SRAM bit cell and SRAM FET strengths were the focus of process improvements. A series of experiments were launched to improve gate oxide integrity in the SRAM cell. The key process changes that produced the greatest reduction of such defects were the

n-FET polysilicon pre-doping implant dose, the gate oxide thickness, and plasma charging damage during the middle-of-the-line reactive ion etch (RIE) process. The SRAM device strength, at fixed leakage, was also improved continuously along with logic device performance, which was described earlier. Each process element was evaluated systematically such that their combined impact on post-stress V_{\min} could be evaluated. Optimization of the gate oxide thickness and n-FET polysilicon pre-doping as well as continued reductions in plasma charging enabled a very significant improvement in V_{\min} behavior. **Figures 5(a)** and **5(b)** respectively show a correlation between initial and post-stress V_{\min} values on chips before and after the process fixes. Initial modules showed poor correlation between initial and post-stress V_{\min} , whereas modules with all of the process fixes showed a very clear correlation between initial and post-stress V_{\min} , with a predictable constant shift of ~ 40 mV expected due to NBTI in the p-FET. This data was verified in volume as chips went through the burn-in screen as described below. With these process fixes implemented, the final SRAM reliability met the highest IBM quality standards.

Programmable fuses

The technology supports an electrically programmable fuse (eFUSE) which operates via an electromigration mechanism, providing high reliability with no additional masks [8]. A programmed fuse is pictured in **Figure 6(a)**. This capability was used extensively, allowing the SRAM arrays to be repaired at various levels in the system build (particularly powerful was repair at the module level following burn-in). In addition to SRAM repair, eFUSES were used to store key chip and multichip module information which was leveraged in system-level verification and debug.

Interconnect technology

The development of a robust, high-performance BEOL interconnect technology was of critical importance in meeting the demanding power, frequency, and reliability requirements of the z9 processors. Copper metallization, pioneered by IBM in the mid-1990s [9], was extended to ten levels in a hierarchical structure. The specific designs discussed here utilized four “thin-wire” or 1X (the terminology corresponds to the multiplier on thin-wire thickness) levels at $0.25\text{-}\mu\text{m}$ thickness and $0.28\text{-}\mu\text{m}$ minimum pitch, two 2X levels (at twice the thin-wire thickness and pitch), two 4X levels, and two 6X levels; see **Figure 6(b)**. The entire wiring hierarchy can be considered to be embedded in low-dielectric-constant insulators. All 1X and 2X levels are in SiCOH ($k = 3.0$) dielectric (an organosilicate glass), and the 4X and 6X levels are in fluorinated silicon dioxide [SiO_2 ($k = 3.6$)].

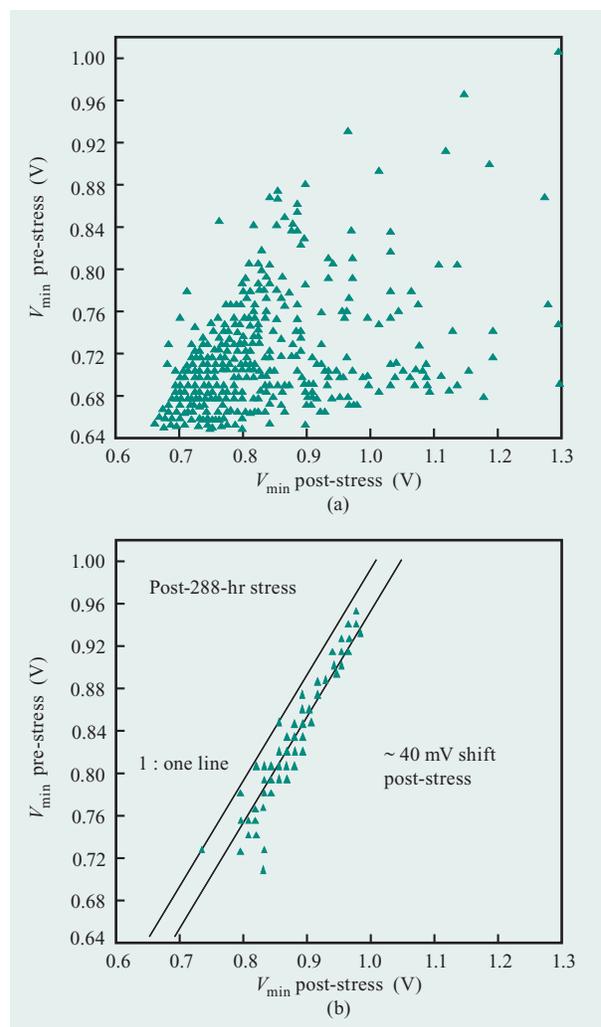


Figure 5

Initial V_{\min} vs. post-stress V_{\min} values (a) before and (b) after process fixes.

Of particular significance was the development of the low- k SiCOH film [10, 11], which has a relative dielectric constant of ~ 3.0 and an integrated “effective” dielectric constant (including interconnect caps) of ~ 3.2 . SiCOH films are known to be tensile and mechanically relatively weak compared with the SiO_2 or fluorinated SiO_2 glass films which they replace; thus, they are more susceptible to brittle fracture from thermomechanical mismatches with the Cu or from chip–package interactions [12]. This risk was addressed during the research phase of this project through a careful choice of the film precursors and plasma deposition techniques enabled by a deep fundamental understanding of the materials set. Further optimization of process parameters in the development and manufacturing phases led to a reduction in the film stress to levels that

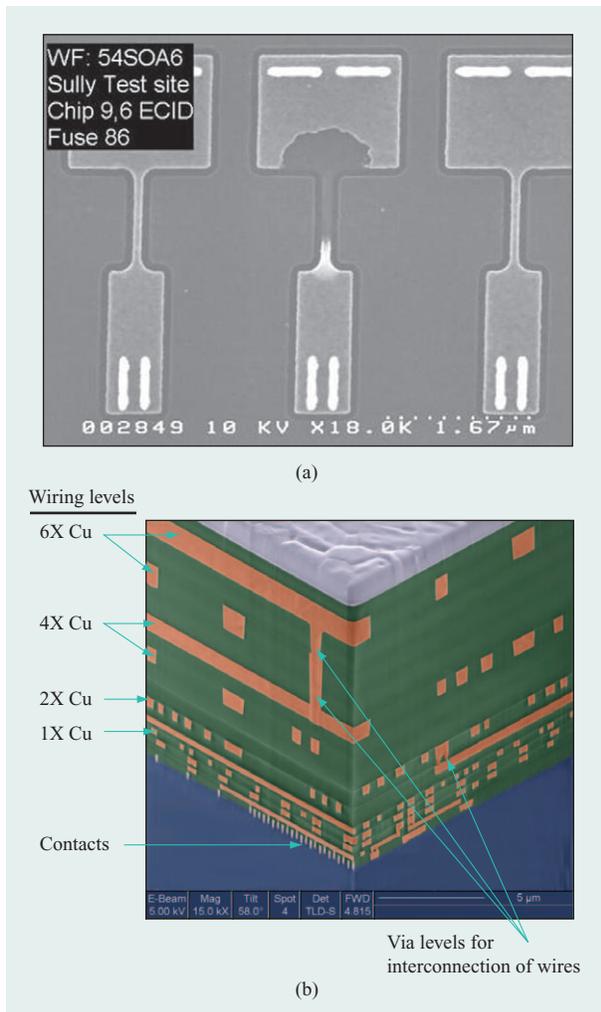


Figure 6

Scanning electron micrographs of (a) gate conductor fuse after programming via an electromigration method; (b) back-end-of-line cross section.

produced a robust BEOL technology immune to stresses induced by chip-to-package interactions.

The inclusion of this low-dielectric-constant material at the 1X and 2X levels led to a ~20% capacitance reduction at those levels without compromising yield or reliability compared with the fluorinated glass dielectric ($k \sim 3.6$) used in the 130-nm technology node. The high level of wafer-level reliability was due in large part to the extreme care that was exercised in analyzing and improving the adhesion properties at all film interfaces.

Optimization of power and performance

In 90-nm high-performance technology, the issue of high standby leakage currents is significant and had to be

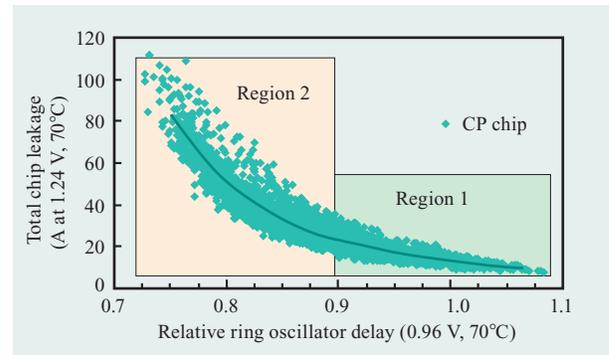


Figure 7

Total chip leakage as a function of the relative delay of a ring oscillator with a fan-out of 3.

addressed on the entire chipset on the multichip module. The two most challenging chips were the central processor (CP) chip, whose frequency ultimately dictated the system performance, and the L2 data cache (SD) chip with its very large number of SRAM cells (97 Mb) and the associated narrow-width devices.

The primary tools used to measure success or failure in this metric are product-like ring oscillators whose performance/power ratio has been shown to correlate well with product measurements [13]. **Figure 7** shows the data of a product-like ring oscillator of inverters with a fan-out of 3. Two regions are shown in this figure. In Region 1, the chip leakage is relatively independent of the ring oscillator delay and is mainly determined by leakage through the gate oxide. In Region 2, the chip leakage rises sharply as the device subthreshold leakage becomes significant and ultimately dominates the chip leakage. Subthreshold leakage is controlled by an intimate coupling between the individual FET device design point (on-current vs. off-current), the optimization of the relative strengths of the n-FET/p-FET device pairs, and the control of these parameters across the chip (and wafer).

Device optimization

The initial device design points were relatively close to those given in [6], where the leakage currents of the various n-FET/p-FET pairs were matched. Previous experience had shown that the nominal conditions might not be optimal at chip level; thus, the initial attempt at power/performance optimization involved a relatively straightforward analysis of the chip response to the regular- V_t device process window experiment (given the process flow, the high- and low- V_t devices follow the regular- V_t devices). It was observed that the cycle time of the chip was a strong function of the n-FET drive current,

while the standby leakage current of the chip was a strong function of the p-FET leakage current. Because of the inherent differences in electron and hole mobility, the n-FET is $\sim 2X$ stronger than the p-FET for a given FET width and length. Thus, for balanced CMOS circuits the p-FETs are $\sim 2X$ wider than the n-FETs, and since subthreshold leakage is a function of transistor width, p-FET leakage across the chip can dominate n-FET leakage. As a result, the process was modified to reduce the n-FET threshold voltages, increasing strength and leakage while simultaneously increasing the p-FET threshold voltages.

In addition, given the relatively strong dependence of chip cycle time to n-FET drive current, there was a strong desire to further improve the n-FET characteristics, and the use of tensile nitride liner was investigated [14]. While a tensile contact liner film enhances electron mobility, it degrades hole mobility, and Ge implantation was used to relax this film over the p-FET devices. **Figure 8(a)** shows the results of the strained liner implementation on the n-FET drive current. As shown, the results were profound and led to rapid implementation of this technology for the 90-nm SOI technology and the z9 chipset.

Results on the CP chip for the device optimizations outlined above are shown in **Figure 8(b)**. Note the substantial improvement in the chip leakage at a fixed ring oscillator delay. Each data point is the average of 50 chips. Under these conditions, the maximum allowable current for the chip is ~ 50 A and, as shown, the device improvements allowed substantially faster ring oscillator speeds ($\sim 6.5\%$) to be consistent with this requirement. Improvements of this sort allow recentering of the device to a faster point and subsequent system-level cycle-time improvements at equivalent parametric yields.

Previously, improvements to the device design point were discussed; this work was absolutely necessary, but it was insufficient to attain the desired MCM power/performance goals for this program. The next problem that was addressed was the control of the device design point across the relatively large chips and across the entire 300-mm wafer. Key parameters here are the gate conductor dimension (which includes both the mask and lithographic components) as well as the dopant profiles that determine the device characteristics.

The critical gate conductor image starts its life as a drawn dimension inside the design tool flow. The next stage in its transformation into an actual, on-wafer device is manipulation through a series of optical proximity correction (OPC) steps, designed to take the design data and manipulate it so that the actual on-wafer dimensions are as close as possible to the designed structures. These sets of manipulations are especially critical on this level, since the final on-wafer dimensions are ~ 40 nm and they are imaged on the wafer surface using 193-nm deep-

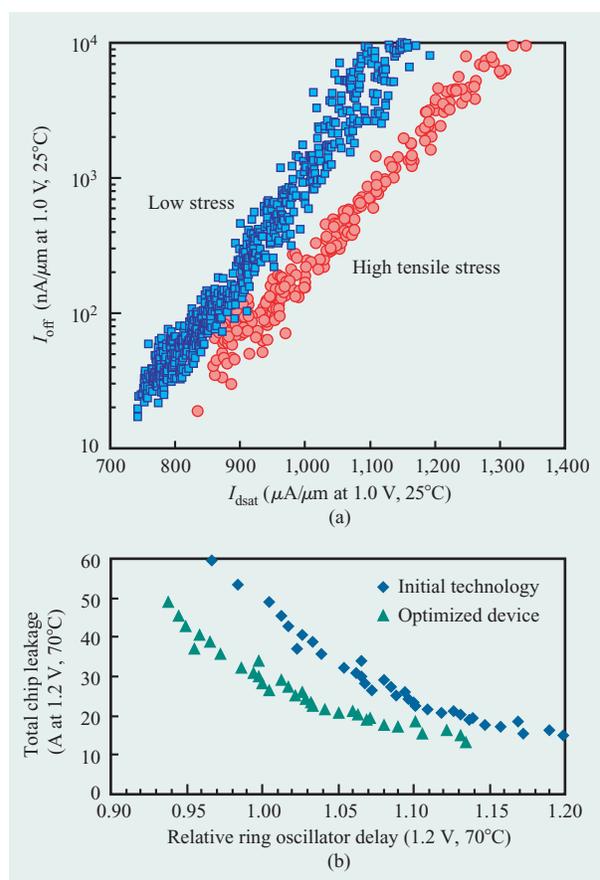


Figure 8

(a) Impact of a highly strained tensile liner on the n-FET drive current. (b) Response of the CP chip to the device optimizations.

ultraviolet light. Issues such as line-end foreshortening and through-pitch variations are minimized through the use of OPC algorithms [15]. Following the OPC steps, the printing behavior of the entire chip is simulated and checked against the initial design input, and deviations are highlighted in a series of steps referred to as model-based verification [16].

Mask design and fabrication

Following a successful exit of the verification steps, the manipulated data is used to generate the lithographic mask that is used to expose the wafers in the fabrication facility. Great care must be taken in this critical step, since even if everything else in the fabrication process is absolutely perfect, the chips still suffer from the linewidth variation present on the mask. The gate conductor images are transferred onto the mask through direct write utilizing an electron beam.

Table 2 Summary of key mask parameters as a function of the mask-writing technology. Note that the mask dimensions are reduced by a factor of 4 when imaged on the wafer.

<i>Parameter</i>	<i>RSG writer</i>	<i>VSB writer</i>
Technology node	130 and 90 nm	90 and 65 nm
Local (1 mm × 1 mm) dimensional uniformity	7 nm	2 nm
Global dimensional uniformity	10 nm	5 nm
<i>X</i> to <i>Y</i> offset	5 nm	3 nm

Through 130-nm technology and into 90-nm technology, critical masks were fabricated using a raster-scanned gaussian beam (RSG) patterning tool. RSG writers form images by continuously raster-scanning an 80-nm round gaussian beam with very high-speed blanking (320 MHz) to form the images [17]. This type of mask-writing tool was utilized in the initial design passes for this program. We note that this type of mask-writing tool is consistent with the specifications outlined by the 2004 International Technology Roadmap for Semiconductors for the 90-nm-technology node. For the newer chip design revisions, a high-speed, vector-shaped-beam (VSB) e-beam system was implemented. VSB systems form images at the appropriate position on the mask and pass the e-beam through several high-precision shaping apertures to form rectangles of various sizes and image them into the resist. The use of shaping apertures on the VSB system resulted in a significant improvement in corner-rounding performance compared with the RSG system.

Another major difference between the two e-beam writers was in the correction scheme used to compensate for the e-beam proximity effect [18]. This effect is caused by backscattering of electrons off the mask substrate surface and non-ideality of the projected beam, resulting in additional undesired electron dose in regions adjacent to other exposed areas up to 25 μm away. If not handled properly, this proximity effect can cause undesirable offsets between sparse and densely written areas of the mask that would subsequently have a deleterious effect on chip performance. The RSG system uses a two-pass proximity-effect correction method known as GHOST [19] which requires the use of a very low-contrast resist. In contrast, the VSB system uses a real-time proximity-effect correction method that adjusts the local e-beam dose on the basis of the local proximity environment caused by adjacent shapes on the mask. Under this scheme, resist images in very dense gratings receive a lower primary e-beam dose than do isolated images in order to compensate for the unintended backscattering.

A summary of the key parameters attainable with each of the mask writers is given in **Table 2**. As shown, masks fabricated with the VSB writer were significantly improved compared with their earlier counterparts fabricated on an RSG system.

Finally, the VSB writer has the capability to support correction of dimensional errors as a function of position across the entire reticle field. Basically, this enables the dramatic reduction or elimination of systematic dimensional variation across the entire mask. Sources of these systematic variations include errors caused by both the e-beam writer and subsequent pattern-transfer processes. These variations have been found to be significant under certain conditions.

Lithography and wafer processing

Once we were satisfied with the tolerance control on the gate conductor mask, the challenge moved to the wafer. As discussed previously, the OPC steps are designed such that the design dimensions are accurately imaged on the wafer surface. Once imaged, the gate conductor features are generated by etching the polysilicon regions that are not protected by the photoresist. In general, the etch bias (the dimensional difference between the final etched feature and the initial feature in photoresist) is not constant across the wafer and can lead to a center-to-edge-of-the-wafer variation in the gate conductor dimension. It should be noted that for relatively large chips, this across-wafer signature can yield a significant across-chip variation in the gate conductor dimension and can adversely affect the performance of a chip at fixed power. The gate conductor etch process was optimized to flatten the across-wafer etch bias signature as much as possible. Modifications to the basic etch chemistry (gas flows, pressures, and wafer temperature) were extremely effective in reducing this effect. Further process optimizations involved detailed characterization of the etch bias of the improved process across the wafer and subsequent adjustments to the lithographic dose at the edge regions of the wafer. The exposure map was relatively complex and was optimized to match the etch chamber.

The process optimizations outlined above all concern control of the gate conductor dimension, which is indeed the primary variable that governs device behavior; however, other variables also play an important role. Two other areas of focus were the control of the spacer films and the control of the optical coupling of the rapid thermal processing steps that control the transient local temperature variations in the junction activation steps. The impact of the spacer films on final device control is conceptually simple—tighter control of this dimension reduces the spread in both the extrinsic device resistance and the short-channel behavior of the device. Control

of the transient temperatures during the rapid thermal processing steps is more subtle and involves control of the local emissivity of the wafer, which in turn is a function of the local pattern density of both isolation and device structures [20, 21]. To achieve more uniform pattern density, dummy shapes are added to the design data with fill algorithms. It was found that modification to our standard fill algorithms (filling both isolation and gate layers prior to mask build) could have a significant impact on the across-chip device control, and these algorithms were adjusted appropriately for that purpose.

As a result of the power/performance optimizations discussed previously (device design improvements, improvements in the mask dimensional control, and improvements in the across-wafer and across-chip parametric device control), substantial improvements were made to the performance of the processor chip at fixed power, as illustrated in **Figure 9**. As shown, a benefit in ring oscillator delay was achieved compared with the initial technology in the leakage range of interest (~50 A). These technology improvements yielded an improvement of >5% performance at the system level at fixed power.

Burn-in reliability screen

To achieve the reliability objectives of the z9 system, the silicon chips utilized on the multichip module are subjected to a burn-in process which is an effective product-defect and reliability screen [22]. An effective burn-in process involves the use of elevated temperature and voltage beyond the application conditions to remove early failures from the population. The relationships among the applied voltage, the applied temperature, and the defect acceleration achieved are well established [23]. **Figure 10** shows an example of the equivalent power-on hours accelerated by increased burn-in temperature and voltage.

Burn-in is typically most effective at accelerating early-life failures. The intent of the burn-in process was to improve the chip reliability, and ultimately the reliability of the z9 system. The use of accelerated conditions on these advanced technologies comes with significant challenges.

The key challenge faced when implementing burn-in for the z9 chipset was power. Burn-in power per chip ranged between 70 and 400 W. In the burn-in processing, a power challenge is synonymous with a thermal challenge [24, 25]. Burn-in power has increased with technology scaling and the power trends of server systems. Burn-in powers have increased by ~400X in the last ten years.

In the system, chip power is generated from both leakage currents and transistor switching currents which are associated with running at the nominal system speed.

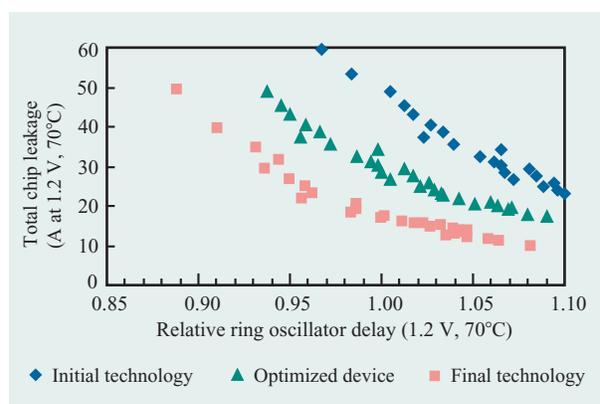


Figure 9

Total chip leakage as a function of the relative delay of a ring oscillator for the initial technology, the optimized device, and the final technology for the CP chip. Each data point is the average of 50 chips.

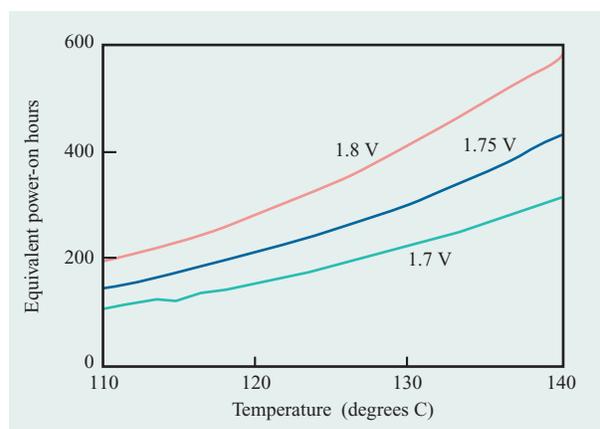


Figure 10

Burn-in equivalent power-on hours vs. temperature by voltage.

In the burn-in environment, the clock speed is reduced, and transistor leakage currents are the primary power driver. Power at burn-in conditions is generally between 2X and 4X of that seen in the system application. What fundamentally drives the high powers are the accelerated application conditions associated with burn-in processing. The relationship between leakage currents and the acceleration factors of burn-in (voltage and temperature) are exponential in nature. Small increases in voltage and/or temperature drive nonlinear increases in leakage currents. Leakage currents are also related to channel length. The short-channel transistors that are

desirable for system and chip performance also have the highest leakage. In summary, to obtain system goals of high reliability and maximum performance from a given technology requires a burn-in solution capable of handling high power.

In addition to the raw power levels associated with burn-in, a number of other dimensions of the power/thermal challenge are in play. The first dimension is the variation of power. At the same accelerated voltage and temperature conditions, each individual silicon chip of the same design may exhibit varied power dissipation. This is driven by inherent process fabrication variation, key elements of which were discussed above. The second dimension of burn-in power that presents additional challenges involves power density and thermal feedback. The areas of higher power density on a die become thermal hot spots which are amplified at burn-in conditions. These hot spots can become the limiting factor with respect to the reliability acceleration conditions that can be achieved. Hot spots tend to have higher concentrations of shorter-channel devices. As hot spots reach burn-in conditions, higher leakages are experienced, which in turn drives more power, and this drives higher local temperatures. This localized self-feedback mechanism can progress to an unstable condition. The confluence of process variation, design, and a slightly variable burn-in thermal environment can interact to create a thermal runaway situation. This runaway condition, although extremely low in probability and frequency, is difficult to predict. The z9 chips and modules were thermally modeled to ensure that the burn-in hardware can handle not only the average power dissipation but also the effects of hot spots.

The z9 burn-in methodology is *in situ* burn-in [26]. This methodology allows for continuous device stress while verifying the expected results from the application of test patterns (e.g., testing). In earlier generations, dynamic burn-in was used, but it had limited verification of functionality. This technique allowed more burn-in escapes (field fails not occurring in burn-in) because chip function was not guaranteed. *In situ* burn-in minimizes burn-in escapes caused by equipment failures, pattern coverage, and product functionality.

The burn-in methodology with respect to tool interface, implementation, and tool control is accomplished by the part number program (PNP). The PNP is the key software component that controls the tool set points (e.g., voltage, temperature, and time), pattern applications, and other features.

The z9 PNP is partitioned into two *in situ* segments. The fundamental purpose of the first segment is reliability defect screening at high-voltage and high-temperature conditions with slower pattern application or clock speeds. The second segment is used as the guard band to

the final system assembly test-and-ship product-quality-level improvement. This segment includes the use of high clock speed and full test coverage at temperature and voltage conditions which are less than burn-in but beyond typical system application conditions.

Up-front engineering includes power modeling, which is then followed by extensive characterization of early silicon. This level of engineering enables the correct burn-in platform selection by power and capability. The burn-in characterization results are merged into the wafer sorting plan to aid in burn-in power management. This combined strategy, along with the burn-in tool flexibility within and across the platforms, optimizes cost and reliability.

The burn-in hardware solution itself is a system which allows more than a hundred chips to be processed in parallel. This tooling is an advance over prior generations in power-delivery capability as well as thermal management and control [27, 28]. It also represents an improvement in the integration level between the various subsystems. At the heart of the hardware solution is the thermal solution. In this tooling, each chip is brought into physical contact with a water-cooled tin-plated copper heat sink. For the z9, the chips are mounted on temporary ceramic carriers allowing for single-chip module-level burn-in. The tooling heat sink directly contacts the back side of the chip. The heat sinks are resident in heat-sink arrays inside the tooling [28]. To optimize thermal performance, special design consideration was given to the heat-sink design, the thermal interface surface, and the interface medium [29] between the chip and the heat sink.

Thermal control is accomplished by modulating the cooling water flow to each individual chip in the tool [27]. A complex control algorithm based on the system thermal model is utilized to accomplish this control. Chip temperatures are monitored by a heat-sink integral back-side-contact temperature sensor and/or through the use of multiple on-chip thermal sensors included in the chip design. Each chip is individually monitored and controlled. The ability to run on-chip at-speed patterns is also supported by a combination of chip design and tooling specification.

A number of design considerations had to be made to accommodate the high currents and powers required by the chips. Additional design considerations were given to the handling of heat generated by joule heating of all current-carrying paths from the supply out to the design under test. For example, the socket design (**Figure 11**) included contacts capable of handling high per-contact currents as well as an integral air-cooled heat sink to assist in the removal of waste heat from the burn-in board and contactor. Additional design considerations were made to accommodate manufacturing and cost

considerations, including surface mount capability and a socket hardware reuse strategy.

The burn-in learning curve for z9 was steep; power management was key. Two major burn-in challenges which required innovation and understanding were the complex thermal control algorithms and the thermal interface. The thermal algorithms included fine-tuning of the imbedded thermal model which controls the chip temperature. The thermal interface challenges included an in-depth study and understanding of the relationships between heat sinks and chips as well as the interface medium [28] itself. Key to good thermal conduction was the flatness of the heat-sink surface which contacts the die surface. The flatness/cost tradeoff had to be understood. Fabricating interface surfaces which are extremely flat can be costly and can adversely affect lead times. Equipment that is not flat enough means yield loss, potential burn-in process cycle-time issues, and less thermal performance. On the road to understanding flatness, other important capabilities and techniques were developed to enable measurement, correlation, and diagnostic capability. Translation from optical flatness to tool diagnostic electrical parameters was the final step for ensuring a level of flatness across the entire toolset.

Developing and fine-tuning the thermal control algorithm of the tool for each die design was another significant challenge. The algorithm had simultaneously to control the temperature to a set point for both high- and low-power devices. The challenge was eased slightly by the data-collection capability of the toolset. By design, the tool can collect voltage, current, temperature, and other parameters for all parts in the run at a given time. After several rounds of data analysis and algorithm adjustment, a robust thermal control solution emerged. This solution had to handle chips of widely varying powers, different die designs, and different package types, as well as the transients associated with temperature, voltage, and clock speed changes during the burn-in cycle.

Due in large part to the success of the burn-in operation, the z9 multichip module was able to meet its reliability objectives.

Summary

The 90-nm SOI technology features used for the z9 multichip module chipset were reviewed. The necessary optimization and enhancement of this technology for z9 were also discussed. Many of the technology challenges faced were required in order to obtain a 2X system performance increase over the previous generation. This drove the technology and design to a 40% increase in frequency for the MCM. On top of the base SOI technology features, later developments, which included process optimization, computational lithographic

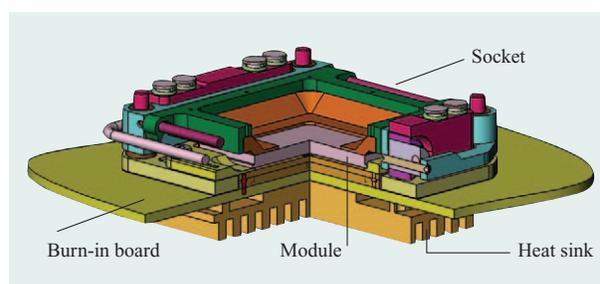


Figure 11

Burn-in socket with module cut away.

advances (OPC), FET device tuning, and mask build optimization, were key to improving delay performance over the nominal technology to meet the performance goals. This increased performance had to be balanced by power considerations, which led to the innovations necessary to successfully burn in the z9 chips. Good defect acceleration in burn-in allowed the z9 to meet its overall reliability goals. Electrically programmable fuses were used after burn-in to replace defective SRAM cells and thus increase yield. Another 90-nm-generation challenge was SRAM reliability. By successful modeling, employing careful experimentation and state-of-the-art failure analysis, the SRAM cells, reduced in area $\sim 50\%$ from the previous technology, were optimized to meet both their performance and reliability goals. Numerous innovations and a systematic approach were required across the spectrum of technology disciplines for the 90-nm chips of the z9 multichip module to ultimately allow the overall system to reach its goals.

Acknowledgments

The authors would like to recognize the many members of the IBM technology manufacturing, development, and research teams, and the z9 chip design, test, and burn-in teams who contributed to this achievement. In particular we would like to acknowledge the following individuals for their outstanding contributions: Richard Rizzolo for his work on system integration bridging back to wafer and module test; Debbie Hamm, Janet Rocque, and Stephen Hennessey for their product engineering work; Jim Crafts and Karre Greene for their work on test development and functional characterization, providing timely feedback to the processing and systems team; Jeff Brown and Ravikumar Ramachandran for process engineering on the early user hardware build; Amy Chan and Raymond Van Roijen for their FET device integration; and Allen Gabor for his work in reducing across-field device variation.

*Trademark, service mark, or registered trademark of International Business Machines Corporation in the United States, other countries, or both.

References

1. B. W. Curran, Y. H. Chan, P. T. Wu, P. J. Camporese, G. A. Northrop, R. F. Hatch, L. B. Lacey, J. P. Eckhardt, D. T. Hui, and H. H. Smith, "IBM eServer* z900 High-Frequency Microprocessor Technology, Circuits, and Design Methodology," *IBM J. Res. & Dev.* **46**, No. 4/5, 631–644 (2002).
2. P. Smeys, V. McGahay, I. Yang, J. Adkisson, K. Beyer, O. Bula, Z. Chen, et al., "A High Performance 0.13 μm SOI CMOS Technology with Cu Interconnects and Low- k BEOL Dielectric," *IEEE Symposium on VLSI Technology, Digest of Technical Papers*, June 13–15, 2000, pp. 184–185.
3. E. J. Nowak, "Maintaining the Benefits of CMOS Scaling When Scaling Bogs Down," *IBM J. Res. & Dev.* **46**, No. 2/3, 169–180 (2002).
4. M. Khare, S. H. Ku, R. A. Donaton, S. Greco, C. Brodsky, X. Chen, A. Chou, et al., "A High Performance 90nm SOI Technology with 0.992 μm^2 6T-SRAM Cell," *IEDM Tech. Digest*, December 8–11, 2002, pp. 407–410. DOI: 10.1109/IEDM.2002.1175865.
5. J. Yario, "2005 Top Fab: IBM," *Semiconductor International*, December 1, 2005 (available online).
6. S.-F. Huang, C.-Y. Lin, Y.-S. Huang, T. Schafbauer, M. Eller, Y.-C. Cheng, S.-M. Cheng, et al., "High Performance 50 nm CMOS Devices for Microprocessor and Embedded Processor Core Applications," *IEDM Tech. Digest*, 2001, pp. 11.1.1–11.1.4. DOI: 10.1109/IEDM.2001.979474.
7. K. Mueller, S. Gupta, S. Pae, M. Agostinelli, and P. Aminzadeh, "T-Cell Circuit Dependent GOX SBD Model for Accurate Prediction of Observed V_{ccmin} Test Voltage Dependency," *Proceedings of the 42nd Annual IEEE International Reliability Physics Symposium*, April 25–29, 2004, pp. 426–429.
8. C. Kothandaraman, S. K. Iyer, and S. S. Iyer, "Electrically Programmable Fuse (eFUSE) Using Electromigration in Silicides," *IEEE Electron Device Lett.* **23**, No. 9, 523–525 (2002). DOI: 10.1109/LED.2002.802657.
9. D. Edelstein, J. Heidenreich, R. Goldblatt, W. Cote, C. Uzhoh, N. Lustig, P. Roper, et al., "Full Copper Wiring in a Sub-0.25 μm CMOS ULSI Technology," *IEDM Tech. Digest*, 1997, pp. 773–776. DOI: 10.1109/IEDM.1997.650496.
10. D. Edelstein, C. Davis, L. Clevenger, M. Yoon, A. Cowley, T. Nogami, H. Rathore, et al., "Reliability, Yield, and Performance of a 90 nm SOI/Cu/SiCOH Technology," *Proceedings of the IEEE International Interconnect Technology Conference*, June 7–9, 2004, p. 214.
11. D. Edelstein, H. Rathore, C. Davis, L. Clevenger, A. Cowley, T. Nogami, B. Agarwala, et al., "Comprehensive Reliability Evaluation of a 90 nm CMOS Technology with Cu/PECVD Low- k BEOL," *Proceedings of the 42nd Annual IEEE International Reliability Physics Symposium*, April 25–29, 2004, pp. 316–319.
12. W. Landers, D. Edelstein, L. Clevenger, C. Das, C.-C. Yang, T. Aoki, F. Beaulieu, et al., "Chip-to-Package Interaction for a 90 nm Cu/PECVD Low- k Technology," *Proceedings of the IEEE International Interconnect Technology Conference*, June 7–9, 2004, pp. 108–110.
13. M. Bhushan, A. Gattiker, M. B. Ketchen, and K. K. Das, "Ring Oscillators for CMOS Process Tuning and Variability Control," *IEEE Trans. Semicond. Manuf.* **19**, No. 1, 10–18 (2006).
14. H. S. Yang, R. Malik, S. Narasimha, Y. Li, R. Divakaruni, P. Agnello, S. Allen, et al., "Dual Stress Liner for High Performance Sub-45nm Gate Length SOI CMOS Manufacturing," *IEDM Tech Digest*, 2004, pp. 1075–1077.
15. L. W. Liebmann, S. M. Mansfield, A. K. Wong, M. A. Lavin, W. C. Leipold, and T. G. Dunham, "TCAD Development for Lithography Resolution Enhancement," *IBM J. Res. & Dev.* **45**, No. 5, 651–665 (2001).
16. J. A. Bruce, E. W. Conrad, G. J. Dick, D. J. Nickel, and J. G. Smolinski, "Model-Based Verification for First Time Right Manufacturing," *Proc. SPIE* **5756**, 198 (2005).
17. T. H. Newman, J. M. Chabala, B. J. Marleau, F. Raymond III, O. Toublan, M. A. Gesley, and F. E. Abboud, "Evaluation of OPC Mask Printing with a Raster Scan Pattern Generator," *Proc. SPIE* **4691**, 1320–1330 (2002).
18. T. Chang, "Proximity Effect in Electron-Beam Lithography," *J. Vac. Sci. Technol.* **12**, 1271–1275 (1975).
19. G. Owen and P. Risman, "Proximity Effect Correction for Electron Beam Lithography by Equalization of Background Dose," *J. Appl. Phys.* **54**, 3573–3581 (1983).
20. I. Ahsan, N. Zamdmer, O. Glushchenkov, R. Logan, E. J. Nowak, H. Kimura, J. Zimmerman, et al., "RTA-Driven Intra-Die Variations in Stage Delay, and Parametric Sensitivities for 65nm Technology," *IEEE Symposium on VLSI Technology, Digest of Technical Papers*, 2006, pp. 170–171.
21. S. Talwar, D. Markle, and M. Thompson, "Junction Scaling Using Lasers for Thermal Annealing," *Solid State Technol.* **46**, 83 (2003).
22. J. VanHorn and P. Nigh, "Towards Achieving Reliability Expectations in a Future of GHz, Nanometers, Amps, Watts and Miles of Copper," *Conference Proceedings of the 10th Annual Known Good Die (KGD) Industry Workshop*, September 2003; see http://www.semtech.org/meetings/archives/other/20001030/07_Test_Screens_Van%20Horn.pdf.
23. JEDEC Solid State Technology Association, "JESD91A Method for Developing Acceleration Models for Electronic Component Failure Mechanisms"; see www.jedec.org.
24. M. Knox, "Facing the Power of the Future—High End Burn-In Directions and Trends," *Conference Proceedings of the IEEE Burn-In and Test Socket Workshop (BiTS)*, March 2001; see <http://www.bitsworkshop.org/archive/archive2001/2001s1.pdf>.
25. M. Knox, "IC Power—The Influence and Impact of Semiconductor Technology," *Conference Proceedings of the IEEE Burn-In and Test Socket Workshop (BiTS)*, March 2004; see <http://www.bitsworkshop.org/archive/archive2004/2004s1.pdf>.
26. B. El-Kareh and W. Tonti, *Chip Reliability*, Topic 1 tutorial presented at the 1997 International Reliability Physics Symposium.
27. H. Hamilton and J. Urbanek, "An Overview of High-Power Burn-In with Test," *Chipscale Rev.*, April 2004, pp. 47–49.
28. H. E. Hamilton, "Thermal Aspects of Burn-In of High Power Semiconductor Devices," *Proceedings of the Eighth Intersociety Conference*, May 30–June 1, 2002, pp. 626–634. DOI: 10.1109/ITHERM.2002.1012513.
29. Y. Cossette, T. Noble, Q. Qiao, and M. Knox, "Thermal Interface Materials: An Implementation Process (from Cradle to Grave)," *Conference Proceedings of the IEEE Burn-In and Test Socket Workshop (BiTS)*, March 2003; see <http://www.bitsworkshop.org/archive/archive2003/2003s5.pdf>.

Received April 6, 2006; accepted for publication October 2, 2006; Internet publication February 13, 2007

Daniel J. Poindexter *IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, New York 12533 (poindext@us.ibm.com)*. Mr. Poindexter is a Senior Technical Staff Member in the IBM 300-mm characterization organization. He received a B.S. degree in mechanical engineering from Carnegie Mellon University in 1982 and joined IBM at the Burlington, Vermont, facility that same year. He received an M.S. degree in electrical engineering from Rensselaer Polytechnic Institute in 1990. Mr. Poindexter received an IBM Outstanding Technical Achievement Award for his work on implementing 90-nm chip technology for zSeries and pSeries* computers; he continues this work for the 65-nm generation.

Scott R. Stiffler *IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, New York 12533 (sstiffle@us.ibm.com)*. Mr. Stiffler is currently a Distinguished Engineer in the IBM Semiconductor Research and Development Center in East Fishkill, New York. He received a Ph.D. degree in materials science from Cornell University and has worked for IBM since 1982 in both the Research and Microelectronics Divisions. Throughout his career, his primary focus has been on semiconductor process development from inception to product qualification.

Philip T. Wu *IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, New York 12533 (ptwu@us.ibm.com)*. Mr. Wu is a Senior Technical Staff Manager in the IBM 300-mm characterization organization. He received a B.S. degree in electrical engineering from the University of Michigan in 1974, an M.S. degree in electrical engineering from Stanford University in 1976, an M.B.A. degree from Rensselaer Polytechnic Institute in 1994, and Project Management Institute certification in 2001. He joined IBM in 1976 and has since worked in advanced VLSI design, test, and characterization. He received four IBM Outstanding Technical Achievement Awards for his work on zSeries technology development. Mr. Wu has received three IBM Invention Achievement Awards; he holds seven U.S. patents.

Paul D. Agnello *IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, New York 12533 (agnello@us.ibm.com)*. Dr. Agnello received his B.S., M.S., and Ph.D. degrees from the Electrical, Computer, and Systems Engineering Department of Rensselaer Polytechnic Institute. He joined the IBM Thomas J. Watson Research Center in 1988 and the IBM Technology Group in 1992. During his career at IBM, he has worked on low-temperature CVD processing of Si and SiGe epi and selective epi, Co and Ti alloy silicide development, 0.25- μ m-node device integration, and 180-nm-node Cu BEOL integration. He managed the device integration group responsible for bringing the IBM 130-nm and 90-nm SOI CMOS technologies to market and was named an IBM Distinguished Engineer in 2003. Currently he manages the high-performance CMOS integration and device project responsible for the 45-nm-node high-performance logic technology. Dr. Agnello is the author or co-author of more than 60 publications; he holds more than 15 U.S. patents.

Thomas Ivers *IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, New York 12533 (iverst@us.ibm.com)*.

Shresh Narasimha *IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, New York 12533 (snaras@us.ibm.com)*.

Thomas B. (Tom) Faure *IBM Systems and Technology Group, 1000 River Street, Essex Junction, Vermont 05452 (tfaure@us.ibm.com)*. After receiving a B.S. degree in chemical engineering from Clarkson University in 1986, Mr. Faure joined IBM at the Burlington, Vermont, facility to work on photomask development. He is currently a Senior Technical Staff Member working on advanced photomask development and manufacturing. He is an author or co-author of 15 patents and six technical papers.

Jed H. Rankin *IBM Systems and Technology Group, 1000 River Street, Essex Junction, Vermont 05452 (rankinh@us.ibm.com)*. Mr. Rankin received a B.S. degree in chemical engineering from Clarkson University in 1995 prior to joining the IBM Burlington, Vermont, Technology Development Group; he recently received a master's degree in business administration from Phoenix University. At IBM he has been responsible for reactive ion etching (RIE) development, optical proximity correction (OPC) generation lithography development, and process integration. He is currently a Senior Engineer focusing on 90-nm integration and infrastructural optimization. Mr. Rankin holds more than 70 patents; he has industry publications on ACLV optimization, process migration, and mask process improvement.

David A. Grosch *IBM Systems and Technology Group, 1000 River Street, Essex Junction, Vermont 05452 (grosch@us.ibm.com)*. Mr. Grosch received a B.S. degree in electrical engineering from the Rochester Institute of Technology in 1985, joining IBM that same year. He is currently an Advisory Engineer working on zSeries burn-in development.

Marc D. Knox *IBM Systems and Technology Group, 1000 River Street, Essex Junction, Vermont 05452 (marcknox@us.ibm.com)*. Mr. Knox received a B.S. degree in electrical engineering from the State University of New York and an M.S. degree in manufacturing systems engineering from Rensselaer Polytechnic Institute. He is a test development engineer working at the IBM facility in Essex Junction, Vermont, involved in multiple aspects of test and burn-in development engineering. Mr. Knox holds a number of patents and patents pending related to test and burn-in.

Daniel C. Edelstein *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (edelstei@us.ibm.com)*. Dr. Edelstein is an IBM Fellow and Manager of BEOL Technology Strategy at the IBM Thomas J. Watson Research Center. He received his B.S., M.S., and Ph.D. degrees in applied physics from Cornell University for work on ultrafast quantum electronics. At IBM, he has worked since 1989 on advanced on-chip (ULSI) interconnect technologies and System-on-Package concepts, spanning the range of interconnect technical and physical aspects. Overall, he has led teams to apply these to the innovation and implementation of Cu and Cu/Low-*k* on-chip interconnect technologies for IBM integrated circuit chips. Dr. Edelstein played a leadership role in the IBM industry-first "Cu chip" technology, and more recently in the introduction of the IBM Cu/SiCOH low-*k* BEOL. He currently leads strategic projects on future enhancements to these technologies. He holds 76 U.S. patents and has received numerous technical, publication, and corporate awards. As coinventor, one of his patents was awarded Inventor of the Year 2006 by the New York State Intellectual Property Law Association. In May 2006, he was appointed an IBM Fellow for his career contributions to ULSI interconnect technologies.

Mukesh Khare *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (kmukesh@us.ibm.com)*. Dr. Khare received the M.Tech. degree from the IIT-Bombay, India, in 1994 and the M.S., M. Phil., and Ph.D. degrees from Yale University in 1995, 1997, and 1999 respectively. In 1998 he joined the IBM Semiconductor Research and Development Center in Hopewell Junction, New York, and has been actively pursuing research and development in CMOS technology since then. As the FEOL integrator, he led the development of 90-nm SOI technology. He is currently a Senior Manager leading 32-nm silicon technology research at the Thomas J. Watson Research Center. Dr. Khare's research interests include the development of advanced devices and structures for CMOS technology, process integration, SRAM and device reliability for acceptable operating voltage, gate dielectric development, and technology/design interaction.

Gary B. Bronner *IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, New York 12533 (gbronner@us.ibm.com)*. Dr. Bronner is a Distinguished Engineer at IBM and Project Leader for the IBM SOI Development Alliance. Since 2004 he has been responsible for the research and development of the IBM 90-nm, 65-nm, and 45-nm SOI technologies, which are used for IBM internal microprocessors and the processors of selected partners. He was previously in charge of the IBM DRAM Development Alliance and led the development of IBM DRAM technology from the 0.25- μm generation through the 0.11- μm generation. The technology he developed continues to be used worldwide by all DRAM vendors using trench capacitor DRAM technology and internally in IBM for embedded DRAM products. Dr. Bronner received a B.S. degree in electrical engineering from Brown University and M.S. and Ph.D. degrees from Stanford University. He is a Fellow of the IEEE. He has reached the 23rd IBM Invention Achievement Plateau, with 64 patents issued.

Hyun-Jang Nam *IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, New York 12533 (hjnam@us.ibm.com)*.

Shahid A. Butt *IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, New York 12533 (sbutt@us.ibm.com)*. After receiving his M.S. degree in materials science and engineering from the Rochester Institute of Technology, Mr. Butt joined Infineon Technologies Incorporated, working in the area of lithography development. In 2003, he joined the 300-mm lithography group at IBM and worked on the introduction and subsequent support of the lithography process for the 90-nm technology. He is currently working on 45-nm logic process development.