Volume 12 Issue 02 Published June 17, 2008 ISSN 1535-864X DOI: 10.1535/itj.1202

Intel® Technology Journal

Intel's 45nm CMOS Technology

Intel Technology Journal Q2'08 (Volume 12, Issue 2) focuses on Intel® 45nm high-k metal gate silicon technology. To quote Gordon Moore, co-founder of Intel, "this is the biggest change in transistor technology in 40 years." In this journal are seven papers that cover the details of transistors and interconnects, variation and design for manufacturability, and packaging and reliability.

Inside you'll find the following articles:

More information, including current and past issues of Intel Technology Journal, can be found at:
http://developer.intel.com/technology/itj/index.htm

Volume 12 Issue 02 Published June 17, 2008 ISSN 1535-864X DOI: 10.1535/itj.1202

Intel® Technology Journal
Intel's 45nm CMOS Technology

Articles

THIS PAGE INTENTIONALLY LEFT BLANK

Preface

By Lin Chao
Editor and Publisher, *Intel Technology Journal*

Hello. This Q2'08 *Intel Technology Journal* (Vol. 12, Issue 2) covers Intel's 45nm CMOS technology which proved pivotal to the advancement in silicon processes technology by extending Moore's Law.

On a personal note, it has been a privilege for the past 12 years to serve as your publisher and editor for this *Intel Technology Journal* (ITJ). This is my last issue as publisher and editor. My thanks to the many of you who have read ITJ since it went online in 1997. ITJ was a pioneer and was among the first technical journals to publish on the Web, and use the Web as its sole publishing means. I remember in 1996 making the decision to publish the ITJ solely online and at the time thinking the Web is risky, but worth trying. Time has shown that the Web was the right decision!

In future, this Journal will be integrated into Intel Press publications.

ITJ has been a repository of Intel technical advancements. ITJ started in 1979 and we are making preparations to make some of the earliest papers available for sharing. I wish to thank the authors who shared with a world-wide community their knowledge and insights about how the technology came into being; after all, who knows better than the actual people who worked on it? Indeed, the authors are also the people who worked directly on the technology. To the authors, I am proud and grateful to have gotten to know you and shared your passions and the heights and lows of advancing the forefront of technology. Thank you to this journal's dedicated team, to Intel's management for your support, and the world-wide community of readers and authors; I am very proud of what we created together. Thank you.

This *Intel Technology Journal* (Vol. 12, Issue 2) focuses on Intel® 45nm high-k metal gate silicon technology. To quote Gordon Moore, co-founder of Intel Corporation, "this is the biggest change in transistor technology in 40 years." In this journal there are seven papers that give in-depth coverage to three key aspects of the technology's transistors and interconnects, variation and design for manufacturability, and packaging and reliability. The new process also produces Intel's first completely lead-free microprocessor products.

Two papers look at transistors and interconnects. The first paper reviews, for 45nm technology, the details of the high-k+metal gate transistors that have been introduced for the first time into high-volume manufacturing. The second paper introduces the issues associated with on-die interconnects and describes how they are addressed on Intel's 45nm high-performance logic process technology. The on-die interconnect stack for Intel's 45nm process generation delivers a 2X higher area density, a 10% lower average capacitance, and improved power distribution.

The next three papers look at manufacturing issues related to variations and designs. The third paper states that process variation is not an insurmountable barrier to Moore's law, but is simply another challenge to be overcome. This is illustrated with data from the 45nm process generation where process variation is shown to be at least equivalent to (and in many cases better than) process variation in the 65nm- and 90nm-process generations.

The fourth paper looks at 45nm SRAM as a lead test vehicle. Every new generation of process technology at Intel is developed and certified using an SRAM-based "X-chip." X6 is the technology lead vehicle used for the 45nm technology serving as a platform for the co-optimization of circuit design and process technology for SRAMs as well as critical design evidence for products. The fifth paper reviews the complex dependencies of design and process to be able to manufacture into microprocessor products. Co-optimization between design

and process is required for a highly manufacturable process technology. This paper discusses how cooperation between design and process meets the challenges for maintaining Moore's Law while delivering fast ramp-up and high yields. The variation, density, and yields on the 45nm process show the success of this Design for Manufacturing (DFM) methodology.

The final two papers look at packaging and reliability. The sixth paper reviews in depth the reliability of Intel's 45nm HK+MG transistors demonstrating that these devices deliver reliability comparable to conventional SiO_2 devices at ~30% higher operating fields with negligible stress-induced leakage current (SILC) degradation.

In the seventh and final paper, we shared some of the key challenges associated with the development of a high-volume manufacturing compatible assembly process for packaging Intel's 45nm, completely lead-free devices. Key technical challenges were addressed through development of novel FLI solders, fluxing material, and process solutions. Small form-factor packaging challenges were overcome by a series of innovative materials and process changes to achieve a reduction in form factor while meeting the technology reliability goals. This enables Intel's continuing leadership in thin and light notebooks and smart phone devices.

We are very proud to have eliminated the use of lead in this technology which also makes Intel a leader in achieving environmentally green products.

Foreword

By Bill Holt

Senior Vice President, General Manager, Technology and Manufacturing Group

In 1965, as the first integrated circuit products entered the marketplace, Intel co-founder Gordon Moore made the now famous observation that the number of transistors on a chip doubled periodically. Not only would Moore's dictum prove true, its longevity has surprised everyone, including Moore himself. The most powerful chips today have over two billion transistors each, up from less than a hundred components per chip when Moore made his observation. This steady, relentless increase in transistor density has made computing powerful, affordable, and ubiquitous. It has enabled the internet age.

While Intel's 45nm technology represents the next step in this continual evolution, lurking under the hood are some revolutionary changes in materials and structures. As the transistor was scaled, fundamental limits of some of the building blocks were reached. One component of the transistor, the gate insulator, had been thinned successively to the point that in the prior generations it was a scant 5 atomic layers thick and could be scaled no further. In anticipation of this issue, Intel began a program in the 1990s to find a new material to replace silicon dioxide, the material used for the gate insulator for over 40 years. The result was a new insulator material based on the element hafnium. This new high k (high dielectric constant) gate insulator needed to be coupled with a change to another component of the transistor, the gate electrode, which was changed from silicon to a proprietary set of metals. Intel's 45nm CMOS technology is the first in the world to feature these new materials, in what Gordon Moore has called "the biggest change in transistor technology since the introduction of the polysilicon gate MOS transistors in the late 1960s." Interestingly, Intel led the move to manufacturing polysilicon gate electrodes in the 1960s and now is leading the move back to metal gate electrodes. Intel's ability to lead the industry in delivering this exciting new technology is a reflection of a strong research and development pipeline, and the dedication, discipline, and hard work of hundreds of talented engineers.

Beyond the significant changes to the transistors, this technology also features improved interconnects, the wires that link the transistors together on a chip. Ideally, these wires transit signals across the chip with rapid speed and low power consumption. 45nm interconnects are, on average, 10% faster than those of the previous generation. The technology also needs to be affordable. Intel's unique low-cost patterning techniques allowed the extension of proven 193nm dry lithography to 45nm at a lower cost than immersion lithography. The technology also supports Intel's green computing initiative. In addition to the lower power consumption levels that result from the new transistors and interconnects, Intel's 45nm technology features another first: a lead-free process and enablement of lead-free packaging.

A key strength that allows Intel to bring new technologies to market quickly lies in its advantage as an integrated device manufacturer to co-optimize the product design and the manufacturing process. Design for manufacturability techniques, methods for dealing with variation, and the ability to prototype circuits early in the development cycle all lead to a technology that is ready to enter the market early and ramp to high volumes quickly. The details of these aspects of Intel's groundbreaking 45nm CMOS technology are covered in this issue of the *Intel Technology Journal.*

THIS PAGE INTENTIONALLY LEFT BLANK

Technical Reviewers for Q2 2008 ITJ

Mohsen Alavi, Technology and Manufacturing Group
Sridhar Balakrishnan, Technology and Manufacturing Group
Ken Brown, Technology and Manufacturing Group
Udi Gavish, Technology and Manufacturing Group
Michael L. Hattendorf, Technology and Manufacturing Group
Debendra Mallik, Technology and Manufacturing Group
Kaizad Mistry, Technology and Manufacturing Group
Gunjan H. Pandya, Digital Enterprise Group
Stephen M. Ramey, Technology and Manufacturing Group
Sunit Rikhi, Technology and Manufacturing Group
Chris Rumer, Technology and Manufacturing Group
Michael J. Smith, Digital Enterprise Group
Krishnamurthy Soumyanath, Corporate Technology Group
Michael R. Stark, Technology and Manufacturing Group
Andrew W.H. Yeoh, Technology and Manufacturing Group

THIS PAGE INTENTIONALLY LEFT BLANK

45nm High-k+Metal Gate Strain-Enhanced Transistors

Chris Auth, Technology and Manufacturing Group, Intel Corporation
Mark Buehler, Technology and Manufacturing Group, Intel Corporation
Annalisa Cappellani, Technology and Manufacturing Group, Intel Corporation
Chi-hing Choi, Technology and Manufacturing Group, Intel Corporation
Gary Ding, Technology and Manufacturing Group, Intel Corporation
Weimin Han, Technology and Manufacturing Group, Intel Corporation
Subhash Joshi, Technology and Manufacturing Group, Intel Corporation
Brian McIntyre, Technology and Manufacturing Group, Intel Corporation
Matt Prince, Technology and Manufacturing Group, Intel Corporation
Pushkar Ranade, Technology and Manufacturing Group, Intel Corporation
Justin Sandford, Technology and Manufacturing Group, Intel Corporation
Christopher Thomas, Technology and Manufacturing Group, Intel Corporation

Index words: CMOS transistor, logic technology, high-k gate dielectric, metal gate

ABSTRACT

For the 45nm technology node, high-k+metal gate transistors have been introduced for the first time in a high-volume manufacturing process [1]. The introduction of a high-k gate dielectric enabled a 0.7x reduction in Tox while reducing gate leakage 1000x for the PMOS and 25x for the NMOS transistors. Dual-band edge workfunction metal gates were introduced, eliminating polysilicon gate depletion and providing compatibility with the high-k gate dielectric.

In addition to the high-k+metal gate, the 35nm gate length CMOS transistors have been integrated with a third generation of strained silicon and have demonstrated the highest drive currents to date for both NMOS and PMOS. An SRAM cell size of $0.346\mu^2$ has been achieved while using 193nm dry lithography. High yield and reliability has been demonstrated on multiple single-, dual-, quad-, and six-core microprocessors.

INTRODUCTION

One of the key methods to enable transistor gate length scaling over the past several generations has been to scale the gate oxide. This improves the control of the gate electrode over the channel, enabling both shorter channel lengths and higher performance. As the gate oxide was scaled the gate leakage increased; this increase in gate leakage was insignificant until the 90nm technology node (Figure 1). At the 90nm and 65nm nodes, the scaling of the gate oxide slowed as a result of the power limitations from the increase in gate leakage. In order to overcome this at the 45nm technology, a gate dielectric with a higher dielectric constant (high-k) has been introduced. This enabled a >25x gate leakage reduction while scaling the Tox by 0.7x.

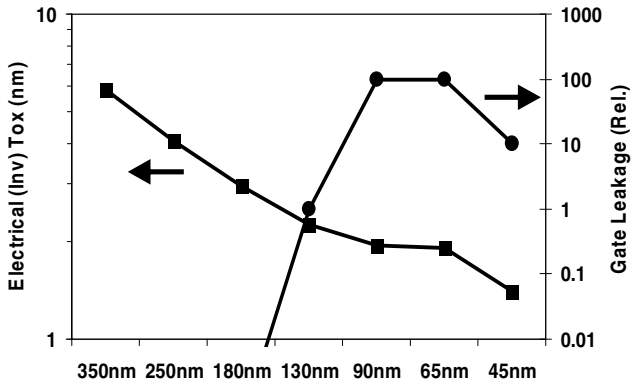


Figure 1: Trend of inversion Tox and gate leakage vs. Intel technology node

The introduction of high-k gate dielectrics has been slowed by several issues [2–4]. The first was the interaction by the high-k material with the existing polysilicon gates. This interaction led to high trap densities at the interface that pinned the Vt of the transistor to an undesirable value. The second was the

degradation of the channel mobility in the presence of high-k dielectrics. The third issue was the poor reliability of the high-k dielectric.

The gate electrode effectiveness has also been increasingly impacted by poly depletion effects. This has led to lower drive currents when the transistor is turned on. By selecting a compatible metal gate electrode with the high-k gate dielectric, both the poly depletion effects and the Vt pinning at the high-k/polysilicon interface can be eliminated while providing higher channel mobilities [5].

In introducing high-k+metal gate transistors for the 45nm generation, these significant challenges needed to be overcome. First, we had to determine which material to use for the high-k dielectric and find dual-band edge metals that were compatible with that high-k dielectric. Second, an integrated CMOS flow needed to be developed that matched the channel mobility of SiO_2 while meeting the reliability requirements for the technology. The development of this CMOS flow was complicated by the need to mesh the process requirements of the metal gate process with both the thermal limitations of the junction formation steps and the uniaxial strain-inducing steps, both of which have become central to the transistor architecture.

Along with the above-mentioned improvements in performance and gate leakage with high-k+metal gate, a key requirement of the technology node was an increased packing density for the transistors. For each node, an ~50% area scaling is expected, and this technology continues that trend. A key challenge to overcome in this scaling is the loss of performance due to scaling of the stress-inducing features of the technology. Use of 193nm dry lithography for critical layers at the 45nm technology node was preferred over moving to 193nm immersion lithography, due to lower cost and greater maturity of the toolset. In order to achieve the tight 160nm gate and contact pitch requirements, unique gate and contact patterning process flows were developed and implemented.

TRANSISTOR PROCESS FLOW

The two common methods for introducing a metal gate to the standard CMOS flow include either a "gate-first" or "gate-last" process. Most comparisons of these two process flows focus on the ability to select the appropriate workfunction metals, the ease of integration, or the ability to scale; however, these comparisons typically fail to comprehend the interaction of the process flows with the strain-inducing techniques. By use of a high-k first and metal gate-last flow, it is possible to maximize the benefits of the stress-inducing steps and high temperature junction formation, while minimizing the thermal processing of the workfunction metals.

In the metal gate-first flow (Table 1), the high-k dielectric and dual-metal processing are completed prior to the polysilicon gate deposition. The dual metal gates are then subtractively etched along with the poly gates prior to Source/Drain (S/D) formation. In contrast, for the high-k first and metal gate-last flow used in this work, a standard polysilicon gate is deposited after the hafnium-based, high-k gate dielectric deposition (Figure 2). This is followed by a standard polysilicon processing flow through the salicide formation steps.

Table 1: Comparison of unique steps in gate-first and high-k first, metal gate-last process flows. Key differences are highlighted in bold.

Gate-First	High-k first, Gate-Last
Isolation	Isolation
High-k gate deposition	High-k gate deposition
Dual Metal-Gate Deposition	Poly-Silicon gate deposition/patterning
Poly-Silicon gate deposition **Poly-Silicon/metal etch**	S/D formation
S/D formation	Salicide/Contact etch stop
Salicide/Contact etch stop	**Poly-Silicon gate removal**
1st ILD deposition/polish	**Dual-Metal Gate deposition**
Contact formation	Contact formation

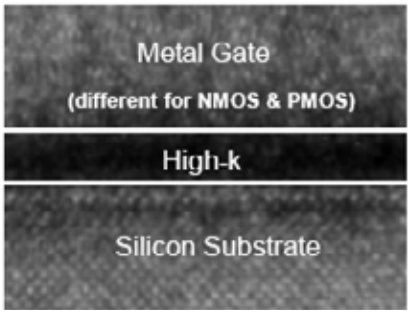


Figure 2: TEM of high-k/metal gate stack

After deposition of the contact etch stop and the first Interlayer Dielectric (ILD) films, a polish step is used to expose the poly gates and enable removal of the dummy poly. The PMOS workfunction metal is then deposited. A patterning step removes the PMOS metal from the NMOS

area. The NMOS workfunction metal is deposited, and the gate trenches are filled with Al for low gate resistance. By using novel gap-fill techniques, robust gate resistance is enabled to sub-30nm gate lengths (Figure 3). A metal polish step is used to remove the excess metal and planarize the gate trenches. The flow then continues with the contact and interconnect processing steps.

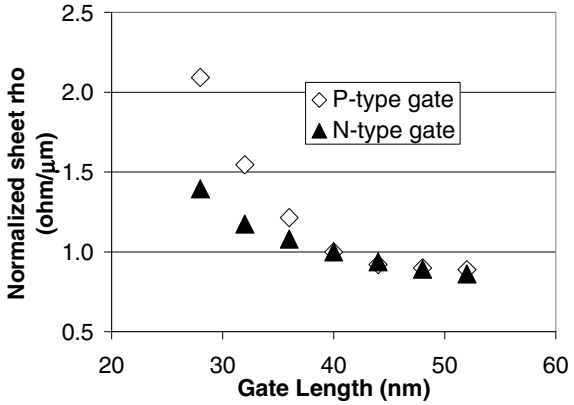


Figure 3: Gate sheet rho versus gate length showing scalability of gate fill process

Figure 4 shows a TEM of the high-k/metal gate NMOS and PMOS transistors with the embedded SiGe S/D strain layer on the PMOS and Ni salicide. The strained silicon techniques that Intel first introduced at the 90nm and 65nm nodes were further enhanced in this generation. The Ge concentration of the embedded SiGe S/D was increased to 30% from the previous generations of 23% in Intel's 65nm technology [6] and 17% in the 90nm technology [7].

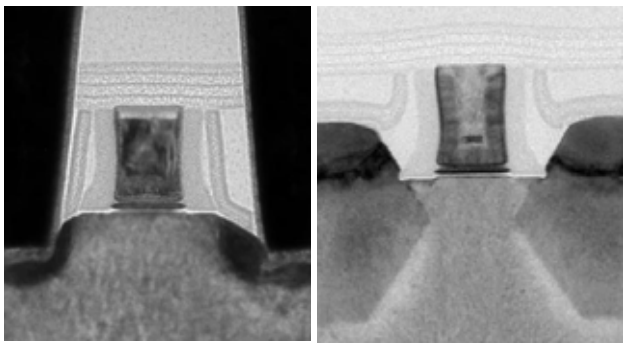


Figure 4: TEMs of high-k+metal gate NMOS and PMOS transistors

DESIGN RULES AND 193NM DRY PATTERNING

Contacted gate pitch is a key measure of front-end density, and the scaling to 160nm maintains the 0.7x scaling trend (Figure 5). This is the most aggressive contacted gate pitch reported to date for a 45nm high-performance logic technology. The contact process has also been modified, with trench contacts replacing conventional contacts for lower series resistance. Trench-contact-based local routing improves layout density, especially for cross-coupled inverter pairs that are very common in microprocessor SRAM and register file arrays. Tight pitches and trench contacts allow SRAM cell size to be scaled to 0.346μm^2 (Figure 6).

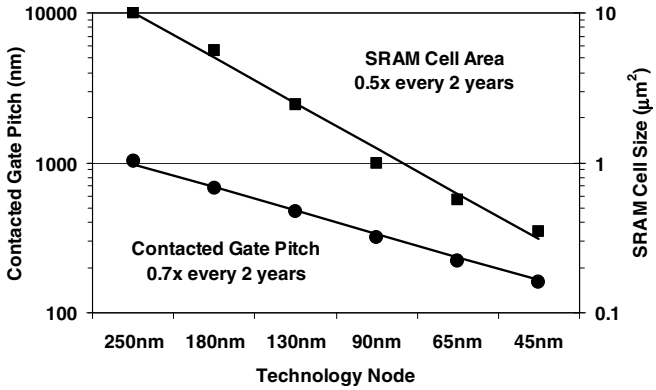


Figure 5: Contacted gate pitch and SRAM cell size scaling trend for Intel's technology nodes

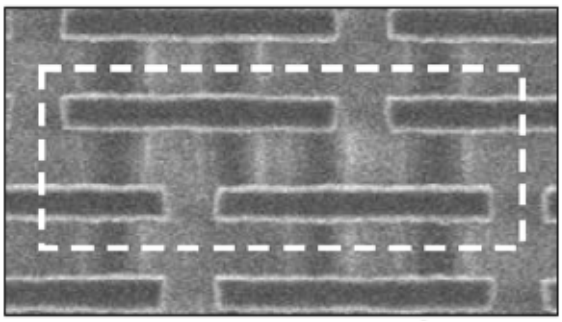


Figure 6: Diffusion and poly layers for 0.346 μm^2 6-T SRAM cell

In order to enable these tight pitches by use of low-cost 0.92NA 193nm dry patterning, innovative processes were developed to produce robust patterning. This is demonstrated by the fidelity of the poly lines in Figure 6. The gate patterning process uses a double patterning scheme. Initially the gate stack is deposited including the polysilicon and hardmask deposition. The first lithography step patterns a series of parallel, continuous lines. Only discrete pitches are allowed, with the smallest at 160nm, to assist in the patterning. A second masking step is then used to define the cuts in the lines. The two-step process enables abrupt poly endcap regions, devoid of rounding that allows for tight contact-to-gate design rules (Figure 7). There are no additional masking steps from this process, since the 65nm generation also used two reticles for poly patterning.

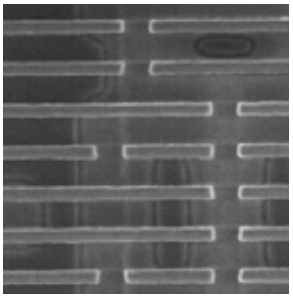


Figure 7: Top-down SEM post-poly-patterning process showing 160nm poly pitch with minimum gate length lines. Note the square poly ends, devoid of rounding.

The contact patterning process uses a similar pitch restriction to facilitate lithography. Trench diffusion contacts run parallel to the gates with discrete pitches, while trench gate contacts run orthogonal to the gates. Use of trench contacts has the added benefits of lowering the contact resistance by >50% and allowing their use as a local interconnect, which improves SRAM/logic density by up to 10%.

TRANSISTOR RESULTS

The introduction of the high-k gate dielectric delivered a dramatic gate leakage reduction relative to 65nm transistors of >25X for NMOS and 1000X for PMOS (Figure 8).

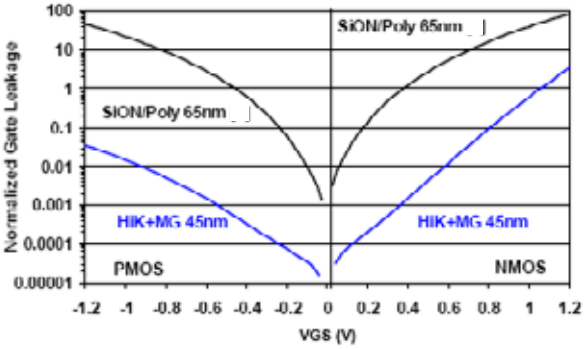


Figure 8: Gate leakage reduction of 25-1000x with use of high-k+metal gate relative to 65nm technology

The high-k+metal gate transistors exhibit excellent short channel characteristics due to the combination of Tox scaling and the optimal workfunction metal gates (Figures 9 and 10). The excellent gate control is also illustrated in the well-behaved subthreshold characteristics (Figure 11).

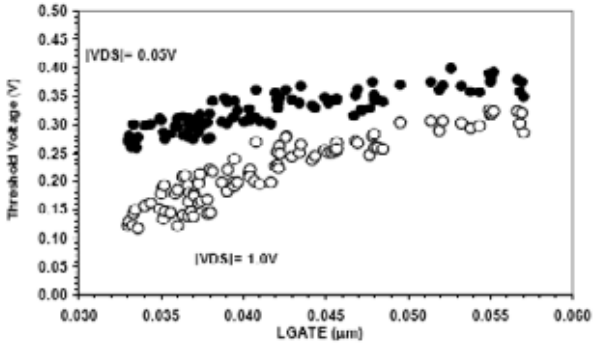


Figure 9: NMOS Vt vs. Lg shows excellent SCE and DIBL

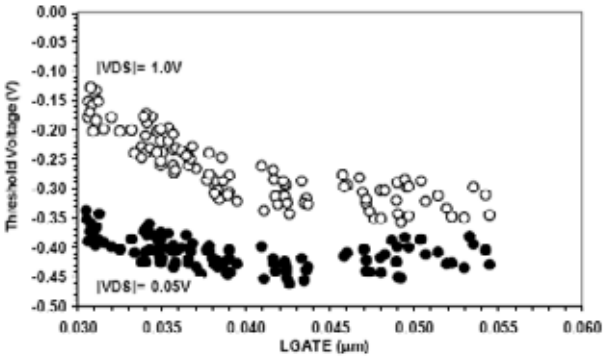


Figure 10: PMOS Vt vs. Lg shows excellent SCE and DIBL

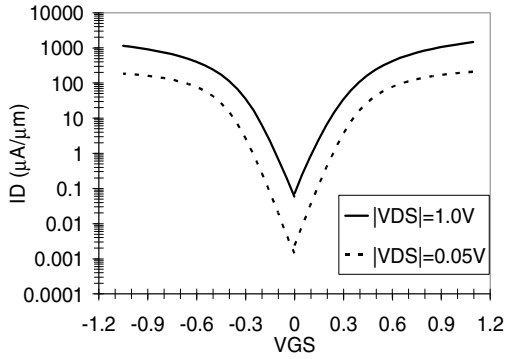


Figure 11: Subthreshold Id-Vgs for both NMOS and PMOS transistors

PMOS performance is improved by using high-k+metal gate as well as by the enhancements to the embedded SiGe processing. The PMOS drive current (Figure 12) of 1.07 mA/µm is a marked 51% improvement over 65nm [8]. NMOS drive current (Figure 13) is 1.36mA/µm, 12% better than the previous-generation, 65nm transistors. The average drive current improvement versus 65nm is 32% at the same voltage and Ioff, despite scaled transistor pitch. The linear drive currents show similar enhancements with PMOS (Figure 14) at 0.178mA/µm and NMOS (Figure 15) at 0.192mA/µm. These drive currents are benchmarked at 1.0V, a low 100nA/µm Ioff and at 160nm

contacted gate pitch. Both the saturated and the linear drive currents represent the best drive currents reported to date for a 45nm technology at low Ioff. Figure 16 shows the transistor performance vs. gate pitch for this generation illustrating that both density and performance are improved with this transistor flow.

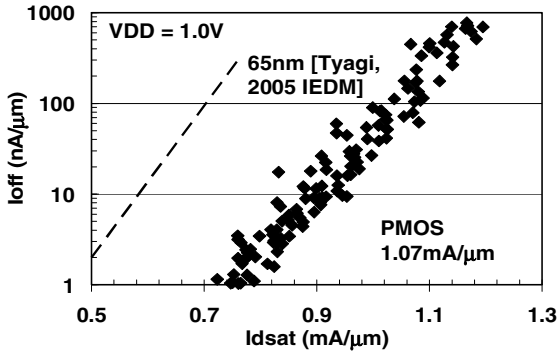


Figure 12: PMOS Idsat of 1.07mA/μm at 100nA/μm Ioff and Vdd =1.0V

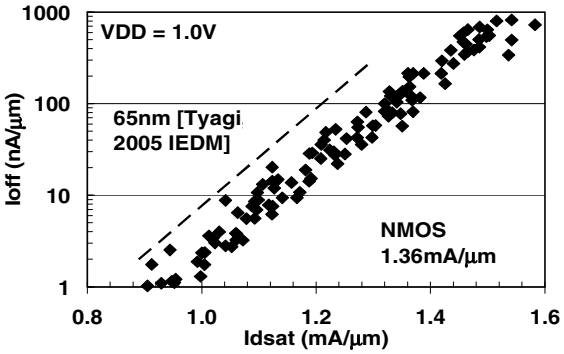


Figure 13: NMOS Idsat of 1.36mA/μm at 100nA/μm Ioff and Vdd =1.0V

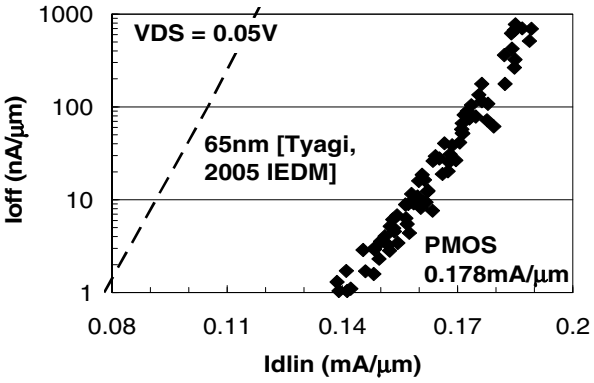


Figure 14: PMOS Idlin of 0.178mA/μm at 100nA/μm Ioff and Vds=0.05V

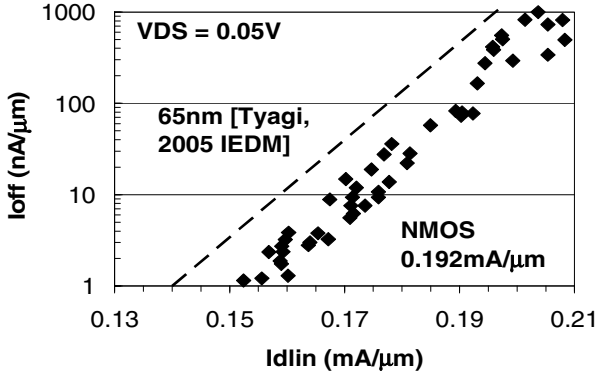


Figure 15: NMOS Idlin of 0.192mA/μm at 100nA/μm Ioff and Vds=0.05V

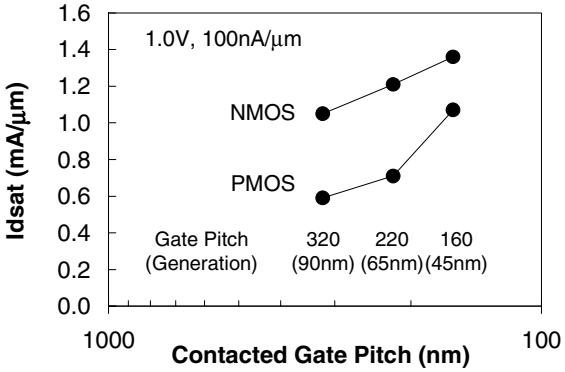


Figure 16: Performance vs. gate pitch for 90, 65, and 45nm generations

STRESS ENHANCEMENT IN A METAL GATE FLOW

Since its introduction at the 90nm node, strain has become a central performance enhancement element for the standard CMOS flow. The most commonly used techniques for implementing strain in the transistors include embedded SiGe in the PMOS S/D, stress memorization for the NMOS, and a nitride stress-capping layer for NMOS and PMOS devices (Table 2).

Table 2: Comparison of stress enhancement methods for 65nm and 45nm nodes. New features are highlighted in bold.

65nm Method	45nm Method
PMOS	**PMOS**
Embedded SiGe S/D	Embedded SiGe S/D **with higher %Ge**
	Poly Gate Removal Enhancement
NMOS	**NMOS**
Tensile Nitride Cap	**Tensile Trench Contacts**
Gate Stress Memorization + S/D Stress Memorization	**Metal Gate Stress (MGS)** + S/D Stress Memorization

A key benefit of using a gate-last flow comes from removing the poly gate from the transistor after the stress-enhancement techniques are in place. It has been shown that the stress benefit from the embedded S/D SiGe process is enhanced through this removal of the poly gate, since the poly gate acts as a buffer counteracting the effect of the embedded S/D SiGe [9]. This benefit can be illustrated in simulation with an estimated 50% increase in lateral compressive stress by removal of the polysilicon gate (Figure 17). The combined impact of the increased Ge fraction in the embedded S/D and the strain enhancement from the gate-last process allow for a 1.5x higher hole mobility compared to 65nm, despite the scaling of the transistor pitch from 220nm to 160nm.

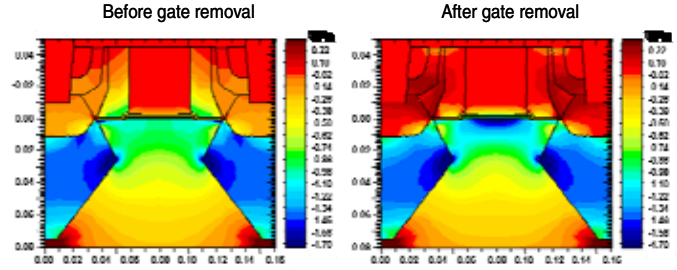


Figure 17: Stress contours in the PMOS transistor before and after the removal of the polysilicon dummy gate. Stress in the channel is shown to increase 50% from 0.8GPa to >1.2 GPa.

For the NMOS device, two methods of stress enhancement have been employed in this technology. First, the loss of the nitride stress layer benefit, due to scaling the pitch from the 65nm technology node, has been overcome by the introduction of trench contacts and by tailoring the contact fill material to induce a tensile stress in the channel. The NMOS response to tensile (control) vs. compressive contact fill materials is shown in Figure 18. The stress impact of the trench contact fill material on the PMOS device is mitigated by use of the raised S/D inherent in the embedded SiGe S/D process.

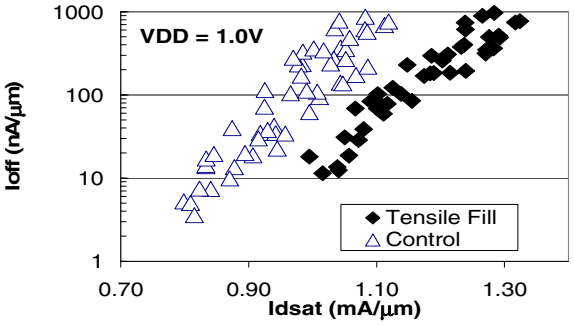


Figure 18: Ion-Ioff benefit of tensile contact fill showing a 10% NMOS Idsat benefit. Contact resistance is matched for the two fill materials.

For NMOS stress memorization, there are two primary methods commonly used, one is memorization of stress in the (S/D) of the device and the other is memorization in the poly gate [10]. The metal gate-last flow is compatible with the S/D method while the poly gate component would be compromised. To compensate for this, the poly gate component is replaced by Metal Gate Stress (MGS). i.e., modifying the metal-gate fill material to directly induce stress in the channel [11]. By introducing a gate fill material with a compressive stress, the performance of the NMOS device is enhanced and adds to the contact fill technique (Figure 19). By use of a dual-metal process with PMOS first, the stress of the NMOS gate is decoupled from the PMOS gate through optimization of the PMOS gate stack to buffer the stress.

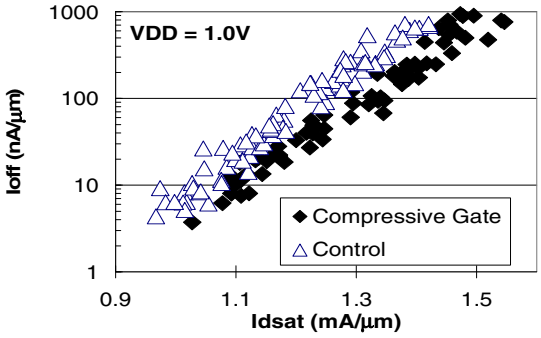


Figure 19: Ion-Ioff benefit of compressive gate stress showing a 6% NMOS Idsat gain. Tensile contact fill is used on both sets of data.

RING OSCILLATORS

The transistor performance gains are reflected in the ring oscillator performance data. Gate delay data from ring oscillators with a fanout of 2 is benchmarked at 100nA/µm Ioff each for NMOS and PMOS, at 1.2V for 65nm and a lower 1.1V for 45nm. The ring oscillators use the minimum contacted gate pitch (220nm and 160nm) for

each technology. Despite the scaling of both voltage and gate pitch, FO=2 gate delay is reduced from 6.65pS (65nm) to 5.1pS (45nm), for a gain of 23% (Figure 20). Table 3 breaks out the RO gains between NMOS/PMOS Idsat, Idlin, and the gate and junction capacitances, illustrating the marked impact of the PMOS performance gains on the ring oscillators.

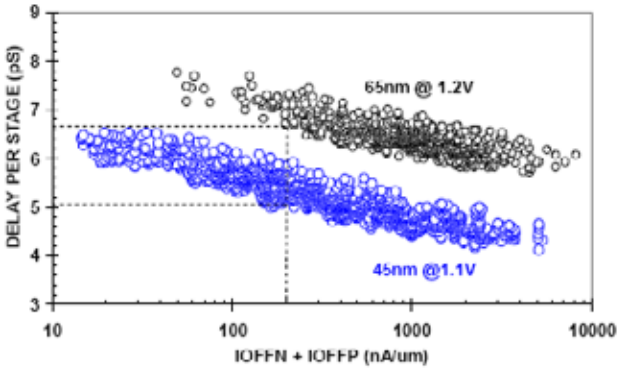


Figure 20: Ring oscillator delay vs. leakage for fanout=2. Comparison of delay for 65nm vs. 45nm is at 1.2 and 1.1V, respectively.

Table 3: Breakdown of RO gains vs. 65nm results. The voltage scaling term accounts for the reduction in VDD from 1.2V (65nm) to 1.1V (45nm).

Component	Benefit (%)
PMOS Idsat	+13
PMOS Idlin	+18
NMOS Idsat	+3
NMOS Idlin	+2
Cjunction	+2
Cgate/Cov	-8
Voltage Scaling	-7
Total	+23

CONCLUSION

High-k+metal gate transistors have been integrated into a manufacturable 45nm process for the first time. Selection of the metal-gate flow (high-k first, metal-gate last) was made to maximize the benefit from the strained silicon steps. Novel stress techniques were also developed to replace the stress methods that are compromised due to scaling and the metal gate flow. The scaling of the transistor density was achieved through development of new poly and contact patterning schemes The resultant transistors provide record drive current at low leakage and at tight contacted gate pitch achieving both performance and density benefits. This is demonstrated in the ring oscillators with a 23% gate delay reduction compared to 65nm at the same Ioff and 10% lower VDD.

ACKNOWLEDGMENTS

We gratefully acknowledge the collaborative efforts of our colleagues in the Portland Technology Development Group, the Technology Computer Aided Design Group, and in the Corporate Quality and Reliability Group.

REFERENCES

[1] K. Mistry, et al., "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging." *IEDM Technical Digest*, 2007, pp. 247–250.

[2] V. Misra, G. Lucovsky, & G. Parsons, "Issues in High-k Gate Stack Interfaces." *MRS Bull., Vol. 27, No. 3*, 2001, pp. 212–216.

[3] C. Hobbs et al., "Fermi Level Pinning at the Poly-Si/Metal Oxide Interface." *Symposium on VLSI Technology Digest*, 2003, pp. 9–10.

[4] G. Ribes et al., "Review on High-k Dielectrics Reliability." *IEEE Transactions on Device and Materials Reliability, Vol. 5, No. 1*, 2005, pp. 5–19.

[5] R. Kotlyar et al., "Inversion Mobility and Gate Leakage in High-k/Metal Gate MOSFETs." *IEDM Technical Digest*, 2004, page 391.

[6] P. Bai et al., "A 65nm Logic Technology Featuring 35nm Gate Lengths, Enhanced Channel Strain, 8 Cu Interconnect Layers, Low-k ILD and 0.57 um2 SRAM Cell." *IEDM Technical Digest*, 2004, pp. 657–660.

[7] T. Ghani et al., "A 90nm high-volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors." *IEDM Technical Digest*, 2003, pp. 978–980.

[8] S. Tyagi, et al., "An advanced low power, high performance, strained channel 65nm technology." *IEDM Technical Digest,* 2005, pp. 1070–1072.

[9] J. Wang, et al., "Novel Channel-Stress Enhancement Technology with eSiGe S/D and Recessed Channel on Damascene Gate Process." *Symposium on VLSI Technology Digest*, 2007, pp. 46–47.

[10] A. Wei, et al., "Multiple Stress Memorization in Advanced SOI CMOS Technologies." *Symposium on VLSI Technology Digest,* 2007, pp. 216–217.

[11] C. Kang, et al., "A Novel Electrode-Induced Strain Engineering for High Performance SOI FinFET utilizing Si (110) Channel for Both N and PMOSFETs." *IEDM Technical Digest,* 2006, pp. 885–888.

AUTHORS' BIOGRAPHIES

Chris Auth joined Intel in 1997 after completing his Ph.D. degree in Electrical Engineering from Stanford University. He initially developed the NOR flash cell for Intel's 0.18µm node. After joining Portland Technology Development in 2000, he developed the industry's first use of strain for transistor enhancement at the 90nm and 65nm nodes. At the 45nm node, he led the process development and introduction of the industry-first, high-k/metal-gate process. He has received two IAAs, holds five US Patents, and has authored/co-authored 18 publications. His e-mail is chris.auth at intel.com.

Mark Buehler received his B.S. degree in Chemical Engineering from the University of California at Berkeley. He then completed his M.S. and Ph.D. in Chemical Engineering at the University of Washington. In 1995, he joined the Portland Technology Department of Intel Corporation. He has worked in both metal and oxide CMP process development since Intel's 0.25µm technology node. He is currently working on metal CMP for Intel's 32nm and 22nm nodes. His e-mail is mark.buehler at intel.com.

Annalisa Cappellani is a Senior Integration Engineer in the Portland Technology Development Department. She joined Intel in 2001 after receiving a Ph.D. degree from the University of Newcastle (UK). She conducted all her Ph.D. work on Poly/metal gate for advanced RF CMOS and DRAM in Germany at Siemens AG research labs (Munich) and at Infineon AG development fab (Dresden). As a member of PTD group at Intel she has been responsible first for the introduction of NiSi in the 90nm and then of the Metal Gate integration in 45nm logic technology. Annalisa is currently working on the pathfinding for the 22nm logic technology node. Her e-mail is annalisa.cappellani at intel.com.

Chi-hing Choi joined Intel in 1991 after completing a Ph.D. degree in Materials Science and Engineering at Northwestern University. He has worked on developing metal gate deposition technologies since 2003. His e-mail is Chi-hing.Choi at intel.com.

Gary Ding joined Intel in 1994 after completing his Ph.D. degree in Materials Science and Engineering at Cornell University. He has worked on metal interconnect, salicide, electroplating, and planarization projects. He is currently a planarization group leader in Portland Technology Development responsible for front end planarization. His e-mail is gary.ding at intel.com.

Weimin Han joined Intel in 1991 after completing a Ph.D. degree in Physics at Oregon State University. He is a Diffusion Group Leader in Intel's Logic Technology Development organization and worked on diffusion process development since Intel's 0.8µm process technology. He is currently responsible for developing an advanced ALD/CVD batch deposition process for the 32/25nm CMOS node. His e-mail is Weimin.c.Han at intel.com.

Subhash M. Joshi joined Intel in 2000 after completing a Ph.D. degree in Materials Science at Duke University. He worked on integrating low-k dielectrics with flip-chip packaging and improving C4 bump reliability on Intel's 90nm technology, together with early development of the lead-free bump and redistribution layer modules. Since 2004, he has worked on front-end process integration, including development of shallow trench isolation, dual metal gate, and trench contact modules for Intel's 45nm technology. His e-mail is subhash.m.joshi at intel.com.

Brian McIntyre is an Etch Group Leader in Intel's Logic Technology Development organization. He joined Intel in 1994 and has worked on etch process development since Intel's 0.35µm process technology. He is currently responsible for developing patterning processes for the 22nm CMOS node. He received his Ph.D. degree from the University of California, Berkeley in 1994. His e-mail is brian.mcintyre at intel.com.

Matt Prince joined Intel in 1988 after graduating from Clarkson University. He has developed CMP processes, consumables, and equipment for Intel's 0.8µm to 45nm technology nodes. He has developed STI, ILD, W, Cu, and poly CMP processes for Intel's logic technologies.

Pushkar Ranade joined Intel in 2003 after graduating with a Ph.D. degree from the University of California, Berkeley. At Berkeley, his research was in the area of sub-70nm CMOS transistor design and involved the integration of novel gate stack materials and ultra-shallow junctions. Since joining Intel, he has worked on transistor integration and development of Intel's 65nm and 45nm technology nodes, and he is currently working on the 22nm technology node. He has authored or co-authored over 35 technical publications and holds 4 U.S. patents. His e-mail is pushkar.ranade at intel.com.

Justin Sandford joined Intel in 1994 after graduating from Arizona State University. He has worked on the front-end process integration of Intel's silicon process technology beginning at the 0.25µm node. Justin is currently working on the development of Intel's 22nm Logic Technology. His e-mail address is Justin.S.Sandford at intel.com.

Christopher Thomas joined Intel in 1997 after completing his Ph.D. degree in Physics from Northwestern University. He supported thin film technology development for the 0.18µm, 90nm, and 45nm nodes. He holds nine US patents. He is currently the High-k Deposition Group Leader, and for the past five years has been responsible for developing the gate oxide deposition process. His e-mail is christopher.d.thomas at intel.com.

THIS PAGE INTENTIONALLY LEFT BLANK

Process and Electrical Results for the On-die Interconnect Stack for Intel's 45nm Process Generation

Peter Moon, Technology Manufacturing Group, Intel Corporation
Vinay Chikarmane, Technology Manufacturing Group, Intel Corporation
Kevin Fischer, Technology Manufacturing Group, Intel Corporation
Rohit Grover, Technology Manufacturing Group, Intel Corporation
Tarek A. Ibrahim, Technology Manufacturing Group, Intel Corporation
Doug Ingerly, Technology Manufacturing Group, Intel Corporation
Kevin J. Lee, Technology Manufacturing Group, Intel Corporation
Chris Litteken, Technology Manufacturing Group, Intel Corporation
Tony Mule, Technology Manufacturing Group, Intel Corporation
Sarah Williams, Technology Manufacturing Group, Intel Corporation

Index words: 45nm, process technology, copper

ABSTRACT

This paper introduces the issues associated with on-die interconnects and describes how they are addressed on Intel's 45nm high-performance, logic process technology. The 45nm process generation uses carbon-doped oxide low-k dielectric and aggressive scaling of the silicon-carbide-nitride (SiCN) etchstop film to achieve a 10% average capacitance reduction at the MT1-MT8 layers, relative to the previous 65nm generation. A thick MT9 layer is added to provide a low resistance path for power routing. MT1-MT8 interconnect RC performance benchmarking and the process and reliability concerns associated with the MT9 redistribution layer are discussed. The combined MT1-MT9 interconnect stack provides high performance and high reliability and enables a completely lead-free (Pb-free) product.

INTRODUCTION

The density and performance of the on-die interconnect stack must continue to improve with each process generation to support increasingly powerful and compact microprocessors. At the same time, process costs must be strictly controlled. The on-die interconnect stack for Intel's 45nm process generation extends the dual damascene processing scheme used on previous generations and adds several innovations for increased density and performance [1].

A new, low-cost and low-resistance Metal-9 layer (MT9) is formed on top of the conventional interconnect stack to support improved power distribution between the C4 bumps and the other metal layers. This MT9 redistribution layer is formed by copper plate-up between sacrificial photoresist lines in a similar fashion to C4 bump formation. The MT9 layer is covered by a silicon nitride passivation layer, a polymer dielectric, and finally, the copper C4 bumps.

The Metal-1 to Metal-3 (MT1–MT3) layers use 193nm dry lithography with 160nm pitch to achieve high-density wiring with lower process cost than immersion lithography. These metal layers also achieve low capacitance due to aggressive scaling of the SiCN etchstop film along with use of carbon doped oxide (CDO) low-k dielectric. The median RC performance for the MT2 layer is 0.20fF/um capacitance and 3.3ohm/um resistance at 160um pitch, as measured on the high-volume manufacturing process with electromigration performance that is consistent with the requirements of high-performance microprocessors.

Overall, the on-die interconnect stack for the 45nm process generation delivers 2X higher area density, 10% lower average capacitance, improved power distribution, and a completely Pb-free product.

Process Discussion

The on-die MT1–MT7 interconnects are formed by dual damascene patterning with highly manufacturable low-k

CDO dielectrics. The lower layer metal pitches are 160nm, while upper layer metal pitches increase progressively to optimize density and performance as shown in Table 1 and Figure 1. MT8 is also formed by dual damascene patterning, but it uses PECVD SiO_2 as the dielectric film, and the MT8 layer is covered with a thick PECVD silicon nitride film.

Table 1: Layer material, pitch, thickness and aspect ratio

Layer	Dielectric Material	Pitch (nm)	Thick (nm)	Aspect Ratio
Metal 1	Low k	160	144	1.8
Metal 2	Low k	160	144	1.8
Metal 3	Low k	160	144	1.8
Metal 4	Low k	240	216	1.8
Metal 5	Low k	280	252	1.8
Metal 6	Low k	360	324	1.8
Metal 7	Low k	560	504	1.8
Metal 8	SiO_2	810	720	1.8
Metal 9	Polymer	30.5µm	7µm	0.4

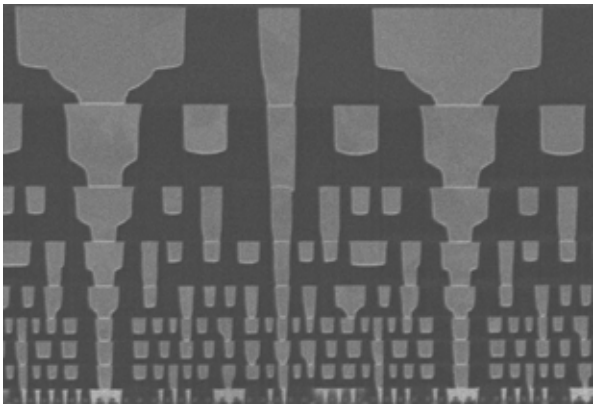


Figure 1: SEM image of interconnect stack up to MT8

The MT9 redistribution layer is formed using a plate-up process. It begins by depositing a blanket Barrier/Seed (Ti/Cu) layer over the entire wafer. The MT9 line/space pattern is created using a thick photoresist followed by standard Cu electroplating to form the MT9 lines. The resist is then stripped and the Barrier/Seed between MT9 lines is removed. The finished MT9 wires are capped with a 400nm PECVD silicon nitride film for improved isolation.

The MT9 layer is covered with 16um of a spin-on polymer dielectric that is patterned by standard lithography techniques and cured in a furnace. From this point onwards the processing is very similar to that used to form the Cu bumps utilized in the 65nm process [2, 3].

Figure 2 shows a cross-section view of the MT9, VA9, and Cu bump layers.

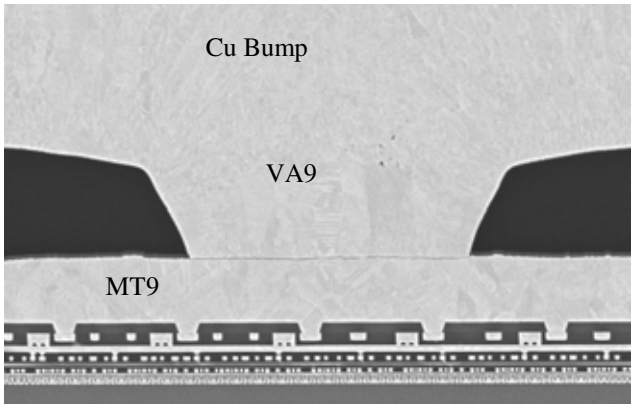


Figure 2: SEM image detailing the MT9 and VA9 layers

After Cu bump formation, the wafers are tested, diced into individual die, and assembled into packaged units. The electrical connection between die and package uses a Pb-free solder as shown in Figure 3. Details of this process are described elsewhere [4].

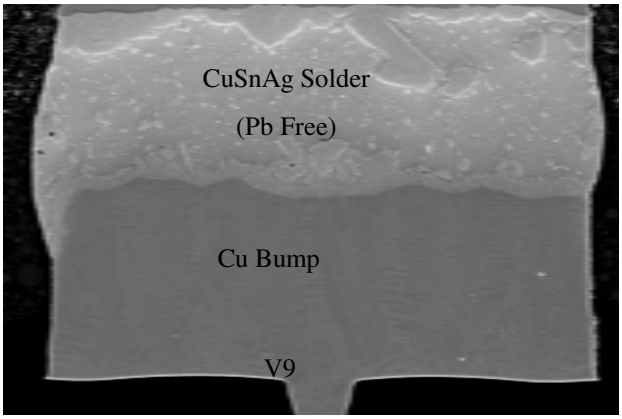


Figure 3: SEM image showing a Bump post chip attach

RESULTS

The 45nm process delivers reduced capacitance relative to the previous 65nm process generation through a combination of techniques. The newer process replaces SiO_2 dielectric with low-k CDO dielectrics at MT1, MT6, and MT7 for a >20% capacitance improvement at those layers as shown in Figure 4. Also, at the lower metal layers, the SiCN etch stop layer is aggressively scaled for an additional 5% capacitance reduction. Overall, the newer process achieves a 10% average capacitance improvement while re-using the robust CDO dielectric film and capital equipment from the previous process generation.

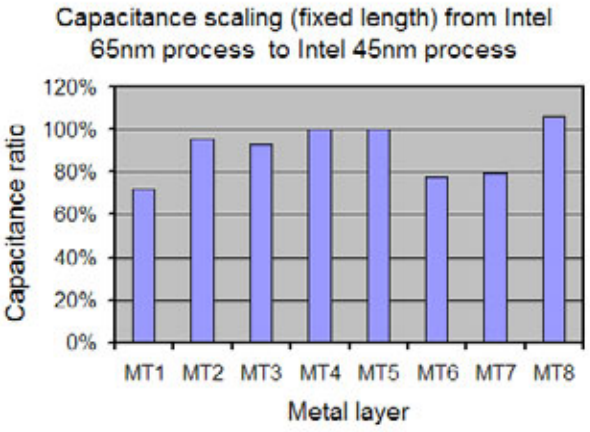
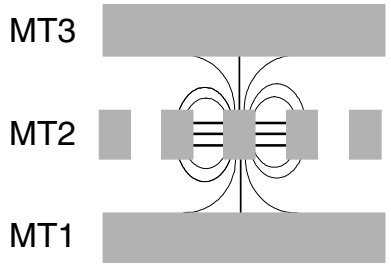


Figure 4: Capacitance comparison between the 45nm process and 65nm process

It is important to accurately benchmark interconnect RC performance. Interconnect capacitance and resistance are measured for minimum pitch lines with 50% dense, minimum-pitch metal patterns directly above and below the measured feature. Total capacitance is the sum of line-line capacitance and layer-layer capacitance as indicated in Figure 5.



$$C_{TOTAL} = 2 \times C_{LINE\text{-}LINE} + 2 \times C_{LAYER\text{-}LAYER}$$

Figure 5: Components of capacitance measurement

Measured capacitance and resistance values for the 45nm process are shown in Figures 6 and 7 at the MT2 and MT6 layers, respectively. The MT2 layer delivers median values of 0.20 fF/um total capacitance and 3.3 ohm/um resistance at 160nm pitch. Resistance at lower metal layers includes a 10% resistance penalty to enable high current density without electromigration failures for high-performance logic products. The MT6 process delivers median values of 0.21 fF/um total capacitance and 0.38 ohm/um resistance at 360nm pitch.

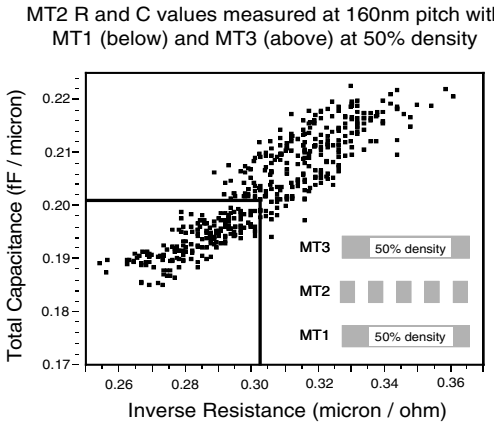


Figure 6: R and C values for MT2 at 160nm pitch

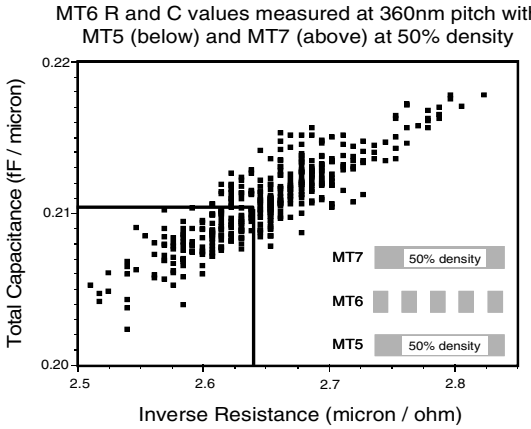


Figure 7: R and C values for MT6 at 360nm pitch

The thick MT9 lines allow for neighboring bumps to be connected by a low-resistance path that enables current redistribution upon initial electromigration (EM) damage at individual bump/solder joints. The large cross-section area of MT9 and its excellent electromigration resistance achieves redundancy of neighboring bumps without concern for electromigration in MT9 itself. Figure 8 shows an electromigration Time-to-Failure comparison between a single daisy chain of C4 bumps vs. two daisy chains of bumps tested in parallel, connected through MT9 with twice the amount of current; and both tested under accelerated conditions of higher temperature and higher current. As the plot indicates, MT9 redundant layout increases bump EM performance by at least 1.65X. Due to long testing periods, the stress ended without any redundant links failing. These data show that MT9 delivered the desired bump EM improvement.

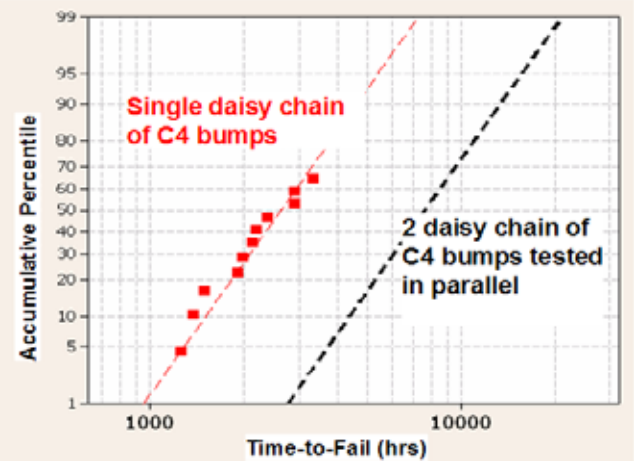


Figure 8: EM fail rate for redundant Cu bumps

Figure 9 shows a Confocal Scanning Acoustic Microscope (CSAM) image of a product unit (full stack) taken after packaging with Sn-Ag-Cu solder. Areas of cracking or interface delamination will show up as white or black spots apart form the contrast variation of the underlying pattern. As seen in Figure 9, no contrast areas are observed, which shows that the unit is free of cracking or interface delamination. These results have been demonstrated on thousands of units and across the expected process variations.

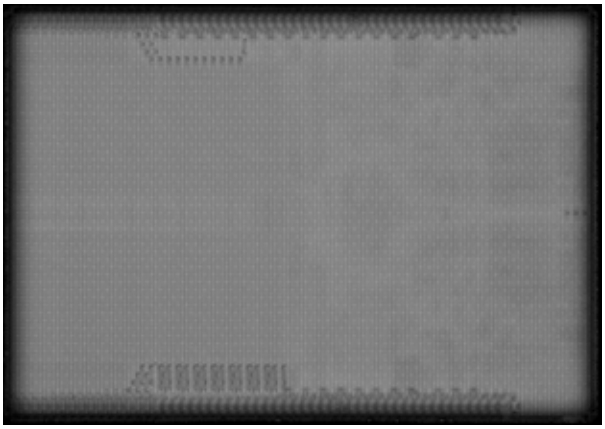


Figure 9: CSAM image of a production unit post packaging

It is critical that this thick MT9 layer also be reliable in addition to achieving electrical benefits. Figure 10 shows a CSAM of the MT9/VA9 layer after 25 hours of HAST stress; the image is clean of any delamination or cracking.

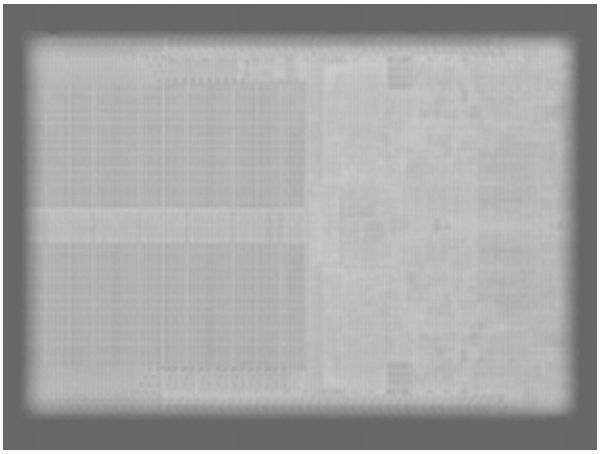


Figure 10: CSAM image of a production unit after 25 hours of HAST

DISCUSSION

Interconnect design is a compromise between density, RC performance, and cost: narrow wires deliver high-density but relatively poor RC performance, while wide wires have better RC performance but reduced density. Extra metal layers can improve either density or RC performance, but they also add process cost. High-performance logic products are designed with fine pitch wires at the lower layers where density is critical, and with wide/thick wires at the upper layers where RC performance is critical.

Intel's 45nm interconnect process uses a unique solution to the problem of interaction between density, RC performance, and cost, by adding a low-cost, very-low-resistance layer to improve the power distribution network at the upper metal layers. The reduced requirement for power distribution in the MT6, MT7, and MT8 layers and the move to low-k dielectric at MT6 and MT7 enables these layers to allocate more chip area to signal wires, resulting in improved RC performance, without adding process cost.

The lower metal layers use 160nm pitch with dry 193nm lithography to achieve high density with acceptable RC performance and relatively low process cost. The tungsten-filled local interconnect layer beneath MT1 helps to achieve high density. The re-use of robust dielectric materials with optimized thickness improves RC performance at the lower metal layers while maintaining high reliability and controlling process cost.

The new process exceeds the reliability requirements for high-volume manufacturing. The 45nm process with its seven layers of low-k dielectrics has significant intrinsic margins to film cracking and/or interface delamination even with the increased stress of a Pb-free process. The MT9 redistribution layer has excellent EM performance

on its own and enables redundancy between Cu bumps, which effectively improves their EM resistance as well. The complete interconnect stack is capable of withstanding temperature shock, HAST, and PreCon stresses, all of which exceed end-of-life goals [5, 6, 7].

CONCLUSION

The on-die interconnect stack for Intel's 45nm process generation delivers a 2X higher area density, a 10% lower average capacitance, and improved power distribution relative to the previous process generation. The new process also enables Intel's first completely Pb-free microprocessor products.

ACKNOWLEDGMENTS

We gratefully acknowledge the Portland Technology Development Group and the TD Quality and Reliability Engineering organizations for their contributions to this work.

REFERENCES

[1] K. Mistry et al., "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning and 100% Pb-free Packaging." *IEDM Technology Digest*, 2007, pps. 247–250.

[2] P. Bai et al., "A 65nm Logic Technology Featuring 35nm Gate Lengths, Enhanced Channel Strain, 8 Cu Interconnect Layers, Low-K ILD, and .57um SRAM Cell." *IEDM Technology Digest*, 2004, pps. 657–666.

[3] A. Yeoh et al., "Copper Die Bumps (first level Interconnect) and Low-K Dielectric in 65nm High Volume Manufacturing." In *ECTC Proceedings,* 2006.

[4] M. Renavikar et al., "Materials Technology for Environmentally Green Micro-electronic Packaging." *Intel Technology Journal*, *Volume 12, Issue 1*, 2008.

[5] "Temperature Cycling." *JESD22-A104-C,* May 2005.

[6] "Accelerated Moisture Resistance—unbiased HAST." *JESD22-A118,* December 2000.

[7] "Highly Accelerated Temperature and Humidity Stress (HAST)." *JESD22-A110-B,* February 1999.

AUTHORS' BIOGRAPHIES

Peter Moon joined Intel in 1988 after completing his Ph.D. degree in Materials Science at MIT. He has worked on process integration for Intel's 0.8um, 0.35um, 130nm, and 45nm silicon process technologies including Intel's first use of shallow trench isolation (0.35um), copper interconnects (130nm), and Pb-free microprocessor products (45nm). Peter is currently leading the development of Intel's interconnect process for the 22nm process generation. His e-mail is peter.moon at intel.com.

Vinay Chikarmane completed his B.Tech. degree in Chemical Engineering from IIT, Bombay and his Ph.D. degree in Materials Science from the University of Texas at Austin. He is currently a Principal Engineer in PTD, leads Copper Interconnect Development at the 32nm node, and is the Engineering Group Leader in the Thin Films area. He previously led Copper Interconnect Development at the 45nm, 65nm, and 90nm nodes, and he worked on engineering development of Intel's first Copper Technology (130nm) and the last two Aluminum Interconnect Technology nodes. His e-mail is vinay.b.chikarmane at intel.com.

Kevin Fischer is a Senior Individual Contributor in PTD. He received his Ph.D. degree in Electrical Engineering from the University of Wisconsin at Madison. He has worked at Intel for seven years in process integration. Beginning on the 90nm process technology, he has worked on four different generations of back-end technology and is currently working on back-end process development for Intel's 22nm technology. His e-mail is kevin.j.fischer at intel.com.

Rohit Grover completed his B. Tech. degree in Engineering Physics from IIT Bombay in 1997, and he received M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Maryland, College Park in 1999 and 2004, respectively. Since 2003, he has been working at Intel as a Senior Process Integration Engineer in the Far Back-end Group of Portland Technology Development. Rohit's team works on enabling lead-free, first-layer interconnects, managing packaging-induced low-k ILD stresses, and first-layer interconnect materials that are compatible with next-generation microprocessor architectures. His e-mail is rohit.grover at intel.com.

Doug Ingerly joined Intel in 2000. He holds a Ph.D. degree in Material Science and Engineering from the University of Wisconsin, Madison and a B.S. degree in Ceramic Engineering from the University of Missouri, Rolla. He has worked on process integration for Intel's 130nm and 65nm interconnect processes and led the far backend integration for the 45nm process. Currently, he is working on Intel's 22nm interconnect process. His e-mail is doug.b.ingerly at intel.com.

Tarek A. Ibrahim joined Intel Corporation in 2004 after completing his Ph.D. degree in Electrical Engineering at the University of Maryland, College Park. He is a member

of the Portland Technology Development integration group and worked on the interconnect process integration for the 45nm node. He is currently leading the focus team responsible for the interconnect patterning process of the 22nm process generation. His e-mail is tarek.a.ibrahim at intel.com.

Kevin J. Lee is a Principal Engineer in Intel's Portland Technology Development Group, specializing in far back-end silicon processing. He was the technical lead for the thick MT9 redistribution layer process development. Prior to that he was Intel's Pb-free Bump Program Director, conducted Intel's initial copper damascene electroplating research and PbSn bump electroplating module development, and developed the buffer coat processes for Intel's 0.8um, 0.6um, and 0.35um technologies. He earned a B.A. degree from Hamline University in 1986 and a Ph.D. degree from the University of Illinois at Urbana-Champaign in 1991, both in Chemistry. His e-mail is kevin.j.lee at intel.com.

Chris Litteken joined Intel's Logic Technology Development organization in 2004 after completing his Ph.D. degree in Material Science from Stanford University. His research interests include the development of fracture mechanics methodologies to improve cracking/adhesion within thin-film structures. As a Senior Quality and Reliability Engineer, he develops predictive models for the thermo-mechanical reliability of 22nm, 45nm, and 65nm microprocessors and chipsets. His e-mail is christopher.s.litteken at intel.com.

Tony Mule is a Senior Process Integration Engineer in the PTD BE Integration group. He received his M.S./Ph.D. degrees in Electrical Engineering from the Georgia Institute of Technology in 2004. Upon joining Intel, he has focused primarily on low-k dielectric patterning development for both the 45nm and 22nm technology generations. His e-mail is tony.v.mule at intel.com.

Sarah Williams is an etch group leader in Intel's Logic Technology Development organization. She joined Intel in 2000 after completing her Ph.D. in Physical Chemistry at the University of Wisconsin-Madison. She is currently working on the development of interconnect processes for the 32nm logic technology. Her e-mail is Sarah.A.Williams at intel.com.

Managing Process Variation in Intel's 45nm CMOS Technology

Kelin Kuhn, Technology and Manufacturing Group, Intel Corporation
Chris Kenyon, Technology and Manufacturing Group, Intel Corporation
Avner Kornfeld, Technology and Manufacturing Group, Intel Corporation
Mark Liu, Technology and Manufacturing Group, Intel Corporation
Atul Maheshwari, Technology and Manufacturing Group, Intel Corporation
Wei-kai Shih, Technology and Manufacturing Group, Intel Corporation
Sam Sivakumar, Technology and Manufacturing Group, Intel Corporation
Greg Taylor, Technology and Manufacturing Group, Intel Corporation
Peter VanDerVoorn, Technology and Manufacturing Group, Intel Corporation
Keith Zawadzki, Technology and Manufacturing Group, Intel Corporation

Index words: 45nm, process variation, DFM

ABSTRACT

The key message of this paper is that process variation is not an insurmountable barrier to Moore's Law, but is simply another challenge to be overcome. This message is illustrated with data from the 45nm process generation where process variation is shown to be at least equivalent to (and in many cases better than) process variation in the 65nm- and 90nm-process generations.

We begin this paper with an introduction and historical overview of process variation. Although there has been a trend in recent years to convey process variation as a new challenge associated with advanced CMOS technologies, process variation has been a continuing theme throughout the history of semiconductor process engineering.

We continue with a review of critical sources of variation specific to the 45nm generation, including highly random effects (random dopant fluctuation, line-edge and line-width roughness), variation associated with the gate dielectric (oxide thickness, fixed charge, defects and traps), patterning proximity effects (classical, and those based on optical proximity correction (OPC)), variation associated with polish (shallow-trench isolation, gate, and interconnect), variation associated with strain (wafer-level biaxial, high-stress capping layers, and embedded silicon-germanium (SiGe)), and variation associated with implants and anneals (implant tool-based, implant profile, rapid-thermal anneal, and implant variation associated with poly-grain boundaries).

We then explore the variety of process, design, and layout techniques used in the 45nm generation to mitigate the impact of variation. Pure process mitigation techniques include targeting key transistor properties to reduce random dopant fluctuation, reducing traps at the high-*k* metal-gate (HiK+MG) interface to reduce random charge variation, improving patterning techniques to reduce line-edge roughness and endcap variation, and improving polishing technologies to reduce systematic cross-wafer variation. Combination design-process techniques include optimizing topology, using OPC to reduce random and systematic variation, and adding dummy features to reduce systematic variation. Pure design techniques include chopping techniques to compensate for random variation and common-centroid layout techniques to compensate for systematic variation.

We move on to illustrate the success of these mitigation techniques by reviewing detailed data characterizing variation in the 45nm generation. Three different types of measurements are presented to illustrate various variation mechanisms. The first is in-fab measurement of variation, used to characterize gate dimensional variation for the 45nm versus 65nm and 90nm generations. The second is low-frequency electrical measurement of matched transistor pairs, used to extract random variation for 45nm versus 65nm transistors. The third is measurements of product ring oscillators, used to determine both systematic and random within-wafer and within-die variation for 45nm versus 65nm products.

Finally, we reinforce the key message that variation does not pose an insurmountable barrier to Moore's law, but is simply another challenge to be overcome.

INTRODUCTION AND HISTORICAL OVERVIEW

Moore's-Law-driven technology scaling has improved VLSI performance by five orders of magnitude in the last four decades. As advanced technologies continue the pursuit of Moore's Law, a variety of challenges will need to be overcome. One of these challenges is management of process variation [1, 2].

Although there has been a trend in the CMOS literature in recent years to convey process variation as a new challenge associated with advanced CMOS technologies, that viewpoint does not effectively capture the history of process variation. Process variation has always been a critical aspect of semiconductor fabrication.

The first discussion of random variation in semiconductor devices was Shockley's 1961 analysis of random fluctuations in junction breakdown [3]. Shockley's concepts of random variation were extended to MOS devices by Keyes [4] in 1975 when he modeled the effect of random fluctuations in the number of impurity atoms in the depletion layer of a field-effect transistor (FET). Systematic variation in MOS devices was first addressed formally in 1974 by Schemmert and Zimmer [5] when they computed the sensitivity of ion-implanted MOS threshold voltages as a function of the implantation energy and the oxide thickness. A more extensive analysis of threshold voltage sensitivity using a closed-forum numerical simulation was presented by Yokoyama et al. in 1980 [6] with a Monte Carlo approach developed by Alvarez in the same year [7]. Interconnect variation has also received significant attention over the years, with Lin et al. presenting a detailed treatment in 1998 [8] that was expanded by many authors in the early 2000s [37, 40–43].

While the continued decrease in the ratio of feature sizes to fundamental dimensions (such as atomic dimensions and light wavelengths) means that management of variation will play a significant role in future technology scaling, the evidence shows that process variation has been a continuing theme throughout semiconductor history.

CRITICAL SOURCES OF VARIATION IN THE 45NM GENERATION

45nm technology is subject to a number of variation effects that are well documented in the literature [9–63]. Examples include highly random effects (random dopant fluctuation (RDF) [9–17], line-edge and line-width roughness, line-edge and line-width roughness (LER) and (LWR), respectively [18–21]), variations in the gate dielectric (oxide thickness variations [22–26], fixed charge [27], and defects and traps [28–34]), patterning proximity effects (classical, and those associated with OPC [35]), variation associated with polish (shallow trench isolation (STI) [36, 40], gate [37–38], and interconnect [39,42-44]), variation associated with strain (wafer-level biaxial 46–49, 57], high-stress capping layers [50–52], and embedded silicon-germanium (SiGe) [53–56]), and variation associated with implants and anneals (tool-based [58], pocket implants [59–60], rapid-thermal anneal RTA [61] and variation associated with poly grains [62–63]).

Random Dopant Fluctuation (RDF)

MOS threshold voltage variation due to random fluctuations in the number and location of dopant atoms is an increasingly significant effect in sub-micron CMOS technologies (see Figure 1 and [9–17]). As the number of dopant atoms in the channel decreases with scaled dimensions, the impact of the variation associated with the atoms increases. Figure 2 illustrates the decreasing average number of dopant atoms in the channel as a function of the technology node. Note the change from the 1µm technology node (with many thousands of dopant atoms in the channel) to the 32nm technology node (with less than 100 atoms in the channel).

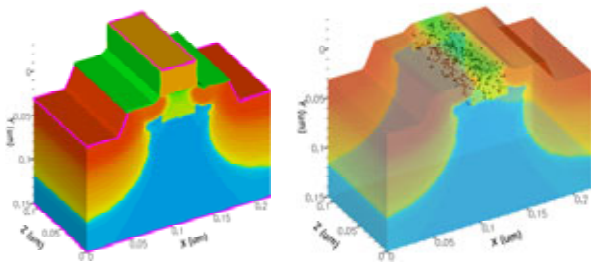


Figure 1: Random dopant fluctuations (RDF) are an important effect in sub-micron CMOS technologies

RDF is assumed to be the major contributor to device mismatch of identical adjacent devices and is frequently represented by Stolk's formulation (Equation 1)

$$\sigma V_{Tran} = \left(\frac{\sqrt[4]{4q^3 \varepsilon_{si} \phi_B}}{2} \right) \cdot \frac{T_{ox}}{\varepsilon_{ox}} \cdot \left(\frac{\sqrt[4]{N}}{\sqrt{Weff \cdot Leff}} \right) = \frac{1}{\sqrt{2}} \left(\frac{c_2}{\sqrt{Weff \cdot Leff}} \right) \quad (1)$$

illustrating that matching improves with decreases in channel doping (*N*) and gate oxide thickness (*Tox*), and it degrades when device area decreases [12].

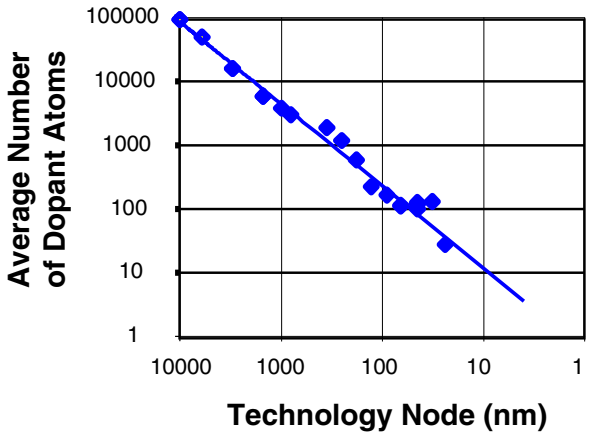


Figure 2: Average number of dopant atoms in the channel as a function of technology node

While Equation (1) assumes that the only contribution to random variation between two adjacent matched devices is random dopant fluctuation, in practice it is known that additional effects also contribute to the measured variation [14]. Identifying the magnitude and root cause for these additional effects is important in facilitating the development of mitigation techniques. Many groups have attempted to estimate the size of these additional effects by comparing measured data to simulation [15–16]. As an example, we reported the results of such a study [17] where we compared simulation results to 65nm silicon data and showed that simulated RDF is ~65% of the total NMOS σV_T. Similar results were obtained when we compared 45nm simulation results to data where the simulated RDF is ~60% of the total PMOS σV_T.

Line-edge and Line-width Roughness (LER and LWR)

While random fluctuations in patterned lines occur in both the front-end and the back-end of the process, the primary concern with LER/LWR is variations in poly-gate patterning (see Figures 3 and 4). For poly-gate patterning, LER and LWR are associated with increases in the sub-threshold current [18, 19] as well as degradation in the threshold voltage (V_T) characteristics [20, 21].

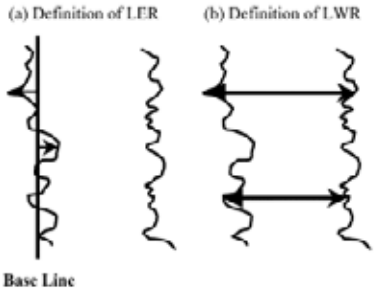


Figure 3: LER/LWR definitions [19]

Diaz et al. [18] quantified the impact of LER on transistor performance by comparing devices from a 130nm technology (80nm nominal gate lengths and 17Å oxide) that were patterned with a 193nm binary solution (9.3nm LER) and 248nm alternating phase shift mask (APSM) solution (6.5nm LER). LER reduction from 9.3nm to 6.5nm translated into measured improvement of 1.5X for a nominal device. For the subnominal 70nm device, a 2X improvement was observed.

In a similar experiment, Kim et al. [19] evaluated the impact of LER and LWR on device performance using a set of 80nm node single nMOS transistors from low-power SRAM devices fabricated with various combinations of gate length, gate width, LWR, and LER. The amount of LER and LWR was controlled by applying different resist materials, defocus, and overetch time. Their experimental data showed that LER effects began when the gate length was less than 85nm. They observed a four-order of magnitude increase in the standard deviation of the subthreshold current for the smallest gate lengths in the study.

Fukutome et al. [20] were able to use scanning tunneling microscopy (STM) to directly assess the impact of LER on the carrier profiles of source-drain extensions in sub-micron MOSFETs. They observed that the roughness of extension edges induced by gate LER depended on the implanted dose, halos (pockets), and various co-implantations. They showed an improvement of 4nm in V_T roll-off with a decrease in the average LER, and they confirmed that co-implants induced a degradation of 5mV in the standard deviation of V_T.

Asenov et al. [21] studied the combined effect of LER and random discrete dopants on current fluctuations. They were able to demonstrate that the two sources of fluctuations act in a statistically independent manner when taken into account simultaneously in the simulations. They also showed that the LER-induced current fluctuations have a much stronger channel length dependence and, as devices are scaled to shorter dimensions, LER is expected to supplant RDFs as the dominant variation source.

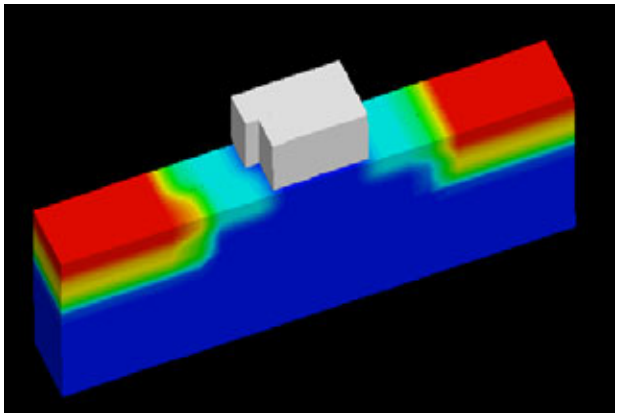


Figure 4: LER/LWR of poly gates has been modeled by a number of researchers [18–21]

Variations in the Gate Dielectric

The high-*k* metal-gate (HiK+MG) devices used in the 45nm generation are subject to a number of variation effects in the gate dielectric [22–34]. These include variations in oxide thickness, fixed charge, and interface traps. These physical changes in the dielectric result in parametric variations in drive current, gate tunneling current, or threshold voltage.

Oxide Thickness

Asenov et al. [22] have studied the intrinsic threshold voltage fluctuations introduced in the atomic scale roughness of the gate interfaces in deep submicrometer MOSFETs through carefully designed simulation experiments. Their simulations show that intrinsic threshold voltage fluctuations induced by local oxide thickness variations become comparable to voltage fluctuations introduced by RDF for conventional MOS devices with dimensions 30nm and below.

Koh et al. [26] have evaluated gate-tunneling leakage current both experimentally and theoretically for MOSFETs with 1.2- to 2.8nm-thick conventional SiO_2 gate oxides. They showed that the statistical distribution of gate-tunnel leakage current causes significant fluctuations in V_T when the gate oxide tunnel resistance becomes comparable to the gate poly-Si resistance. They set the scaling limits (when using a low-resistive silicide gate, and with a conventional gate oxide) at an 0.8nm gate oxide thickness.

Fixed Charge

The presence of fixed charge in the high-*k* layer can affect the mobility and the threshold voltage. As a consequence, variation in the fixed charge may affect the uniformity of the threshold voltages on devices. Kaushik et al. [27] have studied this effect and estimated the fixed charge in high-*k* dielectric films based on a slant-etched SiO_2 layer that allows a thickness series on a single wafer.

Defects and Traps

Electron mobility degradation and V_T instability due to fast transient charging (FTC) in electron traps is a continuing concern in high-*k* dielectrics.

Lucovsky [28] has extensively investigated defects in HfO_2 gate dielectrics through the combination of spectroscopic measurements with electrical detection of defect states. Two types of defects have been proposed, those associated with grain boundaries in the nanocrystalline HfO_2 and those associated with different charge states of the O-atom vacancy. Similar conclusions are reached by other researchers [29–31].

Wen et al. [32] have investigated the effects of FTC by studying the impact of metal gate electrodes on mobility degradation. Their studies suggest that the increase of FTC in $HfSi_x$ may be attributed to higher density of the O vacancies in the high-*k* dielectric caused by the $HfSi_x$-induced O scavenging process.

Optimization of HfO_2 processing such as N incorporation [33] or use of $HfSiO_xN_y$ [34] has also been shown to reduce the charge-trapping effects.

Patterning Proximity Effects

The general lithography expression for the minimum resolvable critical dimension (CD)—assumes equal line/space—is given in Equation (2),

$$\mathrm{CD}_{\min} = k_1 \frac{\lambda}{\mathrm{NA}} \qquad (2)$$

where k_1 is a measure of lithographic aggressiveness (small is aggressive) and includes illumination conditions, resist materials/chemistry, OPC and other resolution enhancement techniques (RETs). The continued decrease in k_1 for more advanced technologies is illustrated in Figure 5.

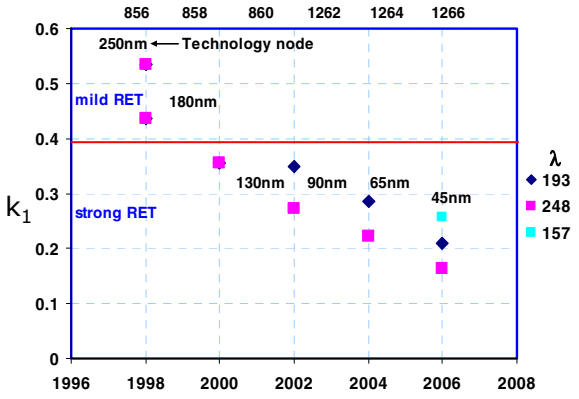


Figure 5: Generational trend in k_1

A variety of techniques can be applied to layers with low k_1 to improve the lithographic patterning and reduce the variation. One of the most powerful of these is OPC [35].

OPC pre-distorts the mask data following specific algorithms in order to achieve a desired pattern on the wafer. OPC is based on a highly phenomenological process model that incorporates lumped optics, resist, wafer stack, and mask effects. This model generates a mask-to-wafer optical transfer function, and an OPC algorithm is written to invert the transfer function. An OPC recipe is developed using an iterative algorithm that modifies the starting database in order to achieve the desired pattern on the mask. An example of the power of OPC is shown in Figure 6, which compares patterning with and without OPC applied.

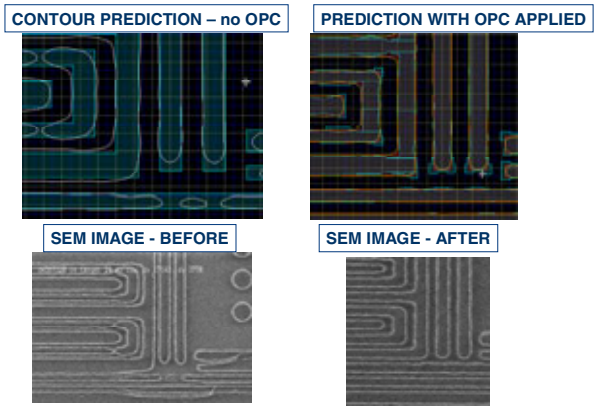


Figure 6: OPC pre-distorts the mask data in order to achieve a desired pattern on the wafer

Polish

Chemical mechanical polish (CMP) is a critical process step in advanced semiconductor technologies. In the front end, CMP has been used for polishing STI [36], and more recently for polishing gate-in, gate-last, metal gate processes [37]. In the back-end, CMP is used for polishing dielectrics in a conventional process and metals in a damascene process [37].

In a traditional STI process [36], shallow trenches are etched into silicon using a nitride hard mask followed by oxide deposition to fill the trenches. A CMP step removes the excess oxide on top of the nitride and partially polishes the nitride layer. The remaining nitride is stripped to expose the active regions where subsequent processing forms the transistors. Subsequent process steps (poly patterning, spacer, silicide formation, etc.) are sensitive to variations in the height of the oxide "steps" between the edge of the STI and diffusion produced by CMP variation.

In a high-*k*, first gate-last process [37], SiO_2 growth is replaced by high-*k* gate dielectric formation. After interlayer dielectric ILD deposition, a poly polish step exposes poly gates, and a gate trench is formed by removal of the dummy poly. Workfunction and conduction metals are deposited in the gate trench and then planarized using a metal polish step. The gate-fill step is sensitive to variable height gates produced by the poly gate CMP variation, and subsequent process steps are sensitive to both height and recess variation from the combination of poly-gate and metal-gate CMP variation [38].

In the back-end [39] the traditional subtractive process uses a metal etch to pattern and remove titanium and aluminum. The subtractive metal process is followed by ILD, a CMP planarization step, and tungsten via fabrication step. In the subtractive process, the sensitivity is to ILD variation produced by the dielectric CMP planarization. Damascene-copper reverses the process, by etching troughs and vias into an insulator, depositing a copper diffusion barrier and copper into the troughs, and using CMP to remove excess copper and barrier material. An ILD is added after the Cu-CMP. In the damascene process, the sensitivity is to Cu and ILD variation produced by the metal CMP planarization.

One commonly applied method for improvement of variation in any CMP process is the addition of dummy-features. Tian et al. [40] review some of the historical approaches to dummy-feature placement and modeling and present a time-dependent relation between post-CMP topography and layout pattern density for CMP in STI.

In the back end, much recent literature has been devoted to the topic of modeling interconnect variation produced by CMP. For example, Yu et al. [41] characterize the smoothing and planarization effects of ILD polishing by a polynomial equation with a small number of fitted parameters. Choi et al. [42] combine a set of scripts and commercial tools to incorporate Cu-interconnect CMP effects in a full-chip static timing analysis. Soumyanath et al. [43] present a nonintrusive time-domain technique to characterize interconnect performance on a 0.25um, 1.8V process. The technique is based on simple time-delay measurements from a repetitive waveform. Finally, Mehrotra et al. [44] analyze interconnect timing performance in a high-speed microprocessor by using timing analysis in conjunction with a post-extraction net adjustment.

Strain

Prior to the 130nm process generation, classic "Dennard" transistor scaling [45] was sufficient to support the 0.7X delay reduction per generation required by Moore's Law. For the 90nm generation and beyond, additional enhancements have been required. Primary among these enhancements is the use of strain.

During the 1980s, researchers began to explore channel strain approaches for transistor enhancement where thin Si layers were grown on relaxed SiGe substrates such that the thin Si layer would take the larger lattice constant of the SiGe and create biaxial tensile stress in the channel [46–49].

In the early 2000s, a new class of transistor strain approaches was developed that used process features external to the transistor (rather than strain in the channel itself as with the biaxial approaches) to strain the transistor. Among these approaches were high-stress capping layers [50–52] and the use of embedded SiGe in the PMOS source-drain regions [53–56].

Process strain creates a number of new variation challenges, both random and systematic. Researchers are beginning to focus both theoretically and experimentally on quantifying the magnitude of strain-induced variation. In Tsang et al. [57] for example, an analytical model was developed to predict threshold variation as a function of Ge fraction, layer thickness, channel length, and doping profile. This model was verified with simulations and experimental data for n- and p-MOSFETs in both single- and dual-channel architectures.

Implant and Anneal

In addition to the fundamental variation mechanism of random dopant fluctuation (discussed earlier), there are also a number of variation sources associated with the physical implant and anneal processes.

The implant tool conditions are a significant source of transistor variation. Al-Bayati et al. [58] have studied the device sensitivity of ultra-shallow junction processes to tool-related implant and annealing parameters. In their work, NMOS and PMOS devices were studied to quantify variation as a function of the accuracy of dose, purity of dose, spike anneal peak temperature, and the ramp-up and cool-down rates.

The architecture of the pocket (halo) and extension (tip) implants is also critical for variation management. Tanaka et al. [59–60] have investigated the statistical V_T distribution for a variety of pocket (halo) implantation conditions through both experimental measurements and device simulation. They showed that the increase in V_T asymmetry caused by the pocket profile degrades the total fluctuation of V_T by greater than 15%.

The advent of advanced RTA processes has introduced new variation sources. Ahsan et al. [61] investigated the impact of RTA anneal on process variation and noted that most of the observed variation can be accounted for by lamp annealing-driven variations in R_{ext} and V_T. They also showed that the variation correlates with the calculated reflectivity for the lamp RTA spectrum and is dependent on the local, mm-scale pattern density.

An additional variation mechanism related to implant technology arises from the poly-crystalline nature of conventional gates. Enhanced diffusion, variations in dopant activation, and implant channeling along grain boundaries can all cause increased variation. Fukutome et al. [62] have investigated the effect of randomly oriented and rotated poly-Si gate grains on lateral carrier profiles of extension regions in sub-50nm MOSFETs by direct observations and electrical measurements. By optimizing the grain boundary they were able to demonstrate a 26% reduction in threshold voltage variation. Brown et al. [63] developed a coherent 3-D statistical simulation study of the impact of poly-Si granularity on the variability in CMOS transistors and concluded that for realistically scaled bulk MOSFETs, the poly-Si and random dopant-induced variations compete at 35nm and 25nm channel lengths. They further concluded that if LER does not scale by the International Technology Roadmap for Semiconductors (ITRS), fluctuation due to poly-grain boundaries becomes the dominant source of variability for channel lengths below ~25nm.

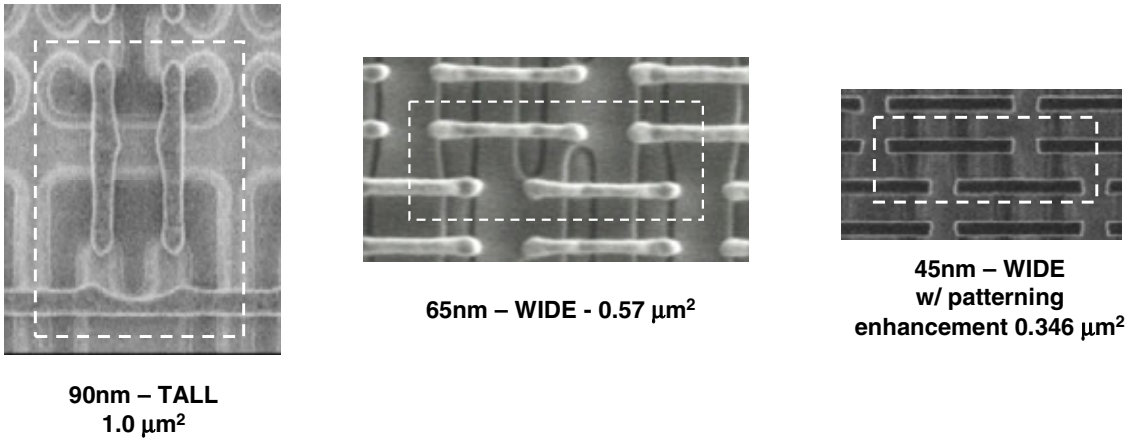


Figure 7: Cell topology enhancements for mismatch improvement

PROCESS, DESIGN AND LAYOUT TECHNIQUES USED IN THE 45NM GENERATION TO MITIGATE THE IMPACT OF VARIATION

Many techniques were applied in the 45nm generation to mitigate the impact of process variation. These techniques can be characterized as pure process techniques (i.e., techniques transparent to design), combination process-design techniques (i.e., techniques that exercise tight cooperation between process and design), and pure design techniques (i.e., techniques transparent to process). Examples of pure process mitigation techniques include targeting key transistor properties to reduce random dopant fluctuation, reducing traps at the HiK+MG interface to reduce random charge variation, improving patterning techniques to reduce LER and endcap variation, and improving polishing technologies to reduce systematic cross-wafer variation. Examples of combination design-process techniques include optimizing topology, using optical proximity correction to reduce random and systematic variation, and adding dummy features to reduce systematic variation. Pure design techniques include chopping and autozeroing to compensate for random variation and common-centroid layout to compensate for systematic variation.

Process Mitigation Techniques

Pure process mitigation techniques are techniques implemented by the process and transparent to design. As an example, recall that RDF is a major contributor to random variation and is frequently represented by Stolk's formulation (Equation 2)

$$\sigma V_{Tran} = \left(\frac{\sqrt[4]{4q^3 \varepsilon_{si} \phi_B}}{2} \right) \cdot \frac{T_{ox}}{\varepsilon_{ox}} \cdot \left(\frac{\sqrt[4]{N}}{\sqrt{Weff \cdot Leff}} \right) = \frac{1}{\sqrt{2}} \left(\frac{c_2}{\sqrt{Weff \cdot Leff}} \right) \quad (2)$$

illustrating that matching improves with decreases in channel doping (*N*) and gate oxide thickness (*Tox*), and degrades when device area decreases [12].

Historical scaling (which reduces gate oxide thickness) suggests a continued improvement in the random variation coefficient (*C2*). However, as illustrated in Figure 8, the historical improvement trend in *C2* slowed when gate leakage concerns limited gate oxide scaling with conventional gate oxides at the 65nm generation. The introduction of 45nm HiK+MG, which restored historical gate oxide scaling due to reduction in gate oxide leakage, was a pure process technique that mitigated the impact of RDF and enabled a return to an historical scaling trend.

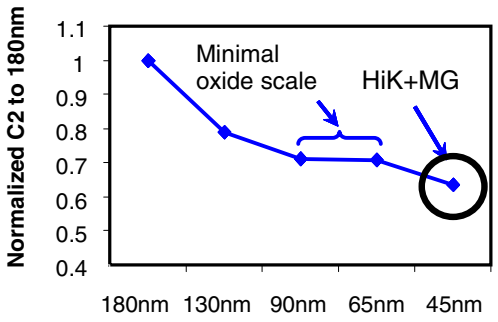


Figure 8: HiK+MG enables a return to an historical scaling trend with associated improvement in C2

Traps in the HiK+MG dielectric are another source of random variation. A number of process improvements were incorporated in the 45nm process to reduce the impact of traps. Figure 9 shows pulsed IV characteristics for early versions of the HiK process vs. a later improved version. Initial HiK+MG material showed a large hysteresis effect—as well as high-bias-temperature instability (BTI) degradation in direct-current (DC) stress. Later versions of the process (incorporating a variety of

improvements) showed negligible hysteresis demonstrating that traps were virtually eliminated.

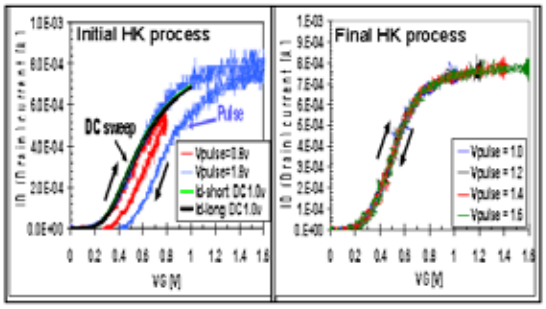


Figure 9: Dielectric trap improvement in the 45nm generation as measured with pulsed IV [64]

LER and LWR are key contributors to random variation in advanced technologies. A variety of advanced patterning techniques were applied in the 45nm generation to improve the patterning and reduce the LER (see Figure 10).

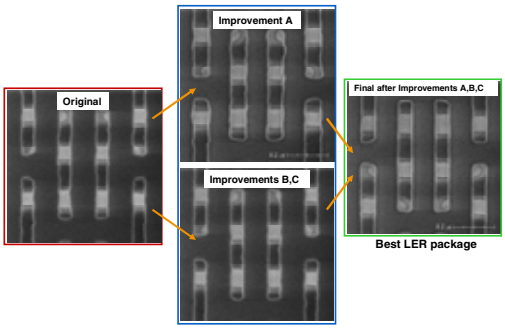


Figure 10: A variety of techniques were applied in the 45nm generation to improve the patterning and reduce the LER

Another lithographic variation improvement incorporated in the 45nm generation was to change the poly-patterning process so that the poly endcaps are square rather than rounded (see Figure 11 and 7). Square endcaps eliminate the systematic variation associated with "dogbone" and "icicle" endcaps.

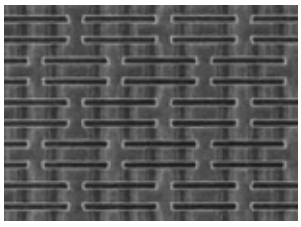


Figure 11: Square poly endcaps implemented in 45nm technology to eliminate the variation of "dogbone" and "icicle" endcaps

A number of modules in the 45nm generation were able to incorporate significant process improvements to reduce systematic variation. One of many examples is shown in Figure 12, which illustrates the improvement in 45nm MT1 within-wafer (WIW) resistance uniformity over the 65nm generation due to improvements in Cu CMP.

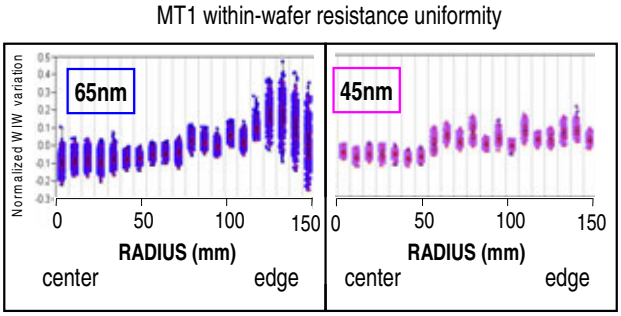


Figure 12: Improvement in 45nm MT1 within-wafer resistance uniformity due to improvements in Cu CMP

Combination Process—Design Mitigation Techniques

Combination process-design mitigation techniques are techniques that exercise tight cooperation between process and design. An example of a combination process-design mitigation strategy is to change the topology of the SRAM from a "tall" design to a "wide" design (see Figure 7 and Ref. [19]). The wide design improves CD control and variation by aligning the poly in a single direction, eliminating diffusion corners, and relaxing some patterning constraints on other critical layers.

Combination design-process improvements resulting from optimization between reticle enhancement techniques and the lithography process were widely used in the 45nm generation. An example is shown in Figure 13, which shows the improvement resulting from an OPC/RET update that resolved the issue of a poor resist profile causing variation in metal pattern after etch.

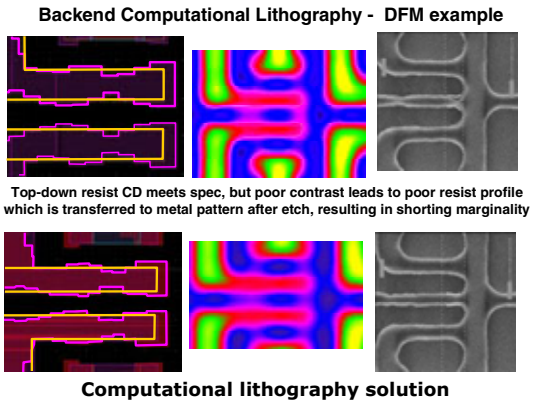


Figure 13: An example of a design-process OPC/RET update in the 45nm generation that resolved the issue of a poor resist profile causing variation in metal pattern after etch

Combination process-design strategies such as dummification and fill techniques at diffusion, poly, and in the back-end have been used historically to reduce systematic variation induced by the lithography, etch, and polish modules. The 45nm generation continued this improvement trend by extending the dummification and fill methodologies of past generations. Figure 14 illustrates this improvement by comparing poly dummification between 65nm and 45nm test vehicles.

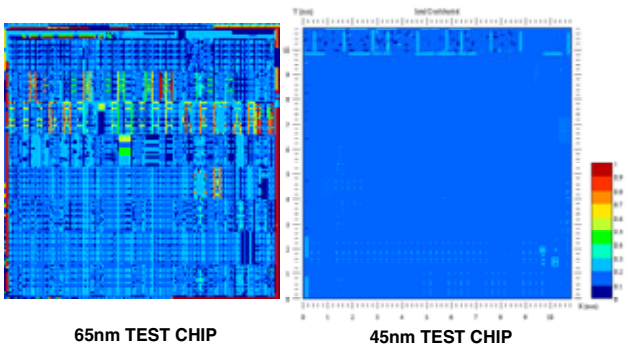


Figure 14: Poly dummification improvements between 65nm and 45nm generations as demonstrated on a test vehicle

Recent generations have seen the impact of non-uniformities at the poly layer extend beyond lithography, etch, and polish into modules such as RTA anneal [60]. Figure 15 shows an example of a combination design-process mitigation strategy where dummy features were incorporated to improve poly density and thus improve RTA temperature uniformity to reduce systematic transistor variation.

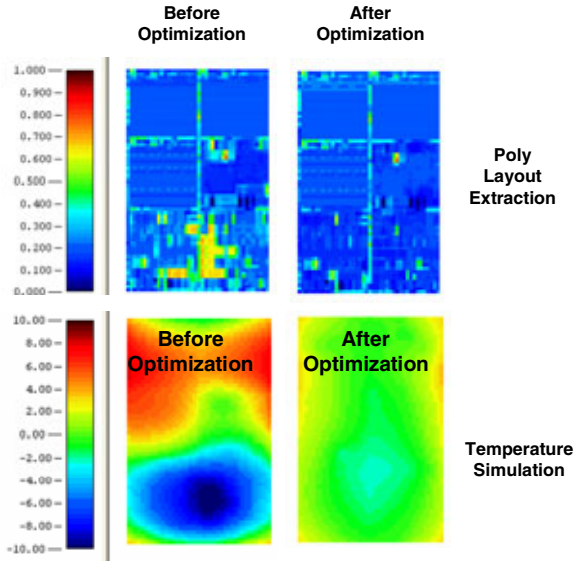


Figure 15: Dummy features were incorporated in 45nm generation to improve poly density and thus RTA temperature uniformity

Design Mitigation Techniques

Pure design mitigation techniques are techniques implemented by design and transparent to process. An example of a pure design mitigation technique used in the 45nm generation to reduce the impact of random variation is chopping, as shown in Figure 16. In chopping, the inputs to a differential amplifier are swapped, or chopped, under the control of a clock signal. The same clock signal is used to swap the outputs, and then the results are low-pass filtered.

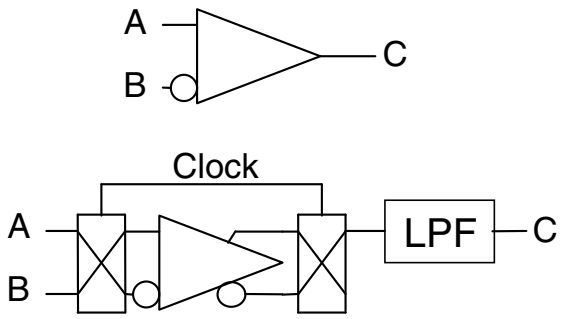


Figure 16: Chopping was used in the 45nm generation as a pure design technique to mitigate the impact of random variation

A more sophisticated design approach used in the 45nm generation to reduce the Vccmin impact of mismatch due to random variation in the SRAM was the incorporation of dynamic forward body bias (FBB) as shown in Figure 17.

In this approach, the Nwell is partially discharged 1-cycle before the word-line by a programmable pulse. The Nwell then remains at lower bias during back-to-back access to minimize switching power. The SRAM PMOS FBB circuitry is integrated along the 8-column boundary and consumes less than 2% area overhead.

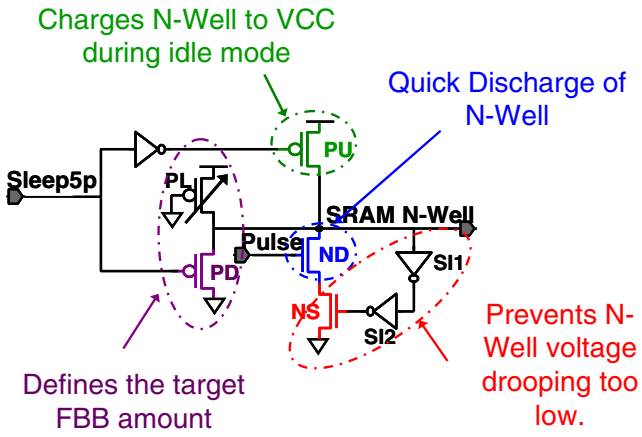


Figure 17: Use of SRAM PMOS FBB to reduce the Vccmin impact of mismatch due to random variation

Systematic variation is best minimized through the use of good layout techniques. One of the design techniques used in the 45nm generation was to lay out matched devices so that they have the same centroid or center of gravity (see Figure 18); then, any device effect that manifests itself as a gradient across the layout will impact each set of matched devices equally.

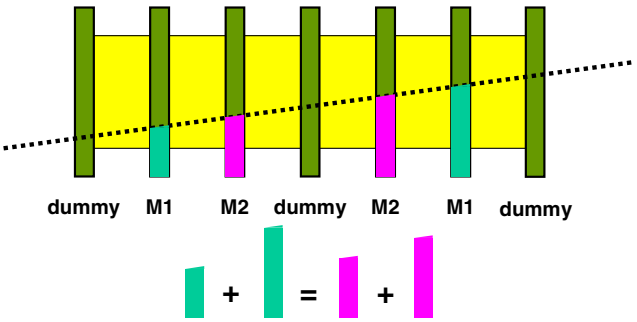


Figure 18: Common centroid layouts were used in the 45nm generation as pure design techniques to mitigate the impact of systematic variation

CHARACTERIZATION OF VARIATION IN THE 45NM GENERATION

We illustrate the success of these mitigation techniques by reviewing detailed data characterizing variation in the 45nm generation. Three different types of measurements are presented to illustrate various variation mechanisms. The first is an in-fab measurement of variation, used to characterize CD variation for 45nm versus 65nm and 90nm generations. The second is DC electrical measurement of matched transistor pairs, used to extract random variation for 45nm versus 65nm transistors. The third is frequency measurements of product ring oscillators, used to determine both systematic and random WIW and within-die (WID) variation for 45nm versus 65nm products.

In-Fab Characterization of Critical Dimension (CD)

Maintaining poly-gate control is critical for managing variation between process generations. Figure 19 presents summary data from in-line measurements of gate CD across four generations that show that the 45nm technology generation was able to maintain a 0.7X scaling to prior generations for WID, WIW, and total variation.

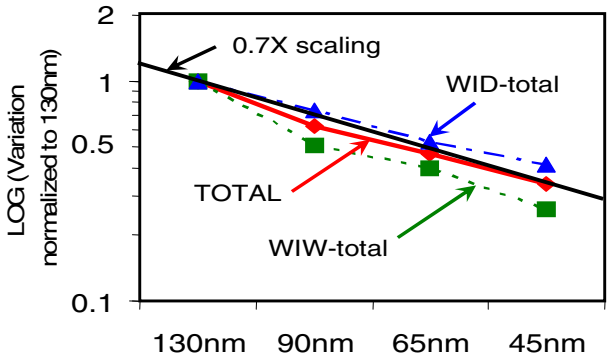


Figure 19: The 45nm generation continued the historical scaling trend of 0.7X in poly-gate variation control

DC Measurement of Matched Transistor Pairs

DC electrical measurement of matched transistor pairs is a basic technique used to extract random variation for 45nm versus 65nm transistors. Figure 20 illustrates the random variation for both 65nm and 45nm generations as extracted from matched transistor pairs. Note an ~20% improvement in intrinsic random variation from the 65nm to 45nm generations.

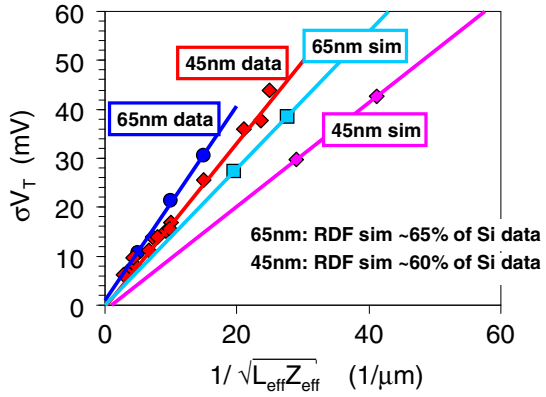


Figure 20: Pelgrom plot illustrating the improvement of the 45nm over 65nm generations

Ring Oscillator Measurements

A powerful tool for assessing process variation is locating ring oscillators (see Figures 21 and 22) routinely in all product designs. The detailed ring-oscillator data can be used to identify areas of concern for process teams to resolve.

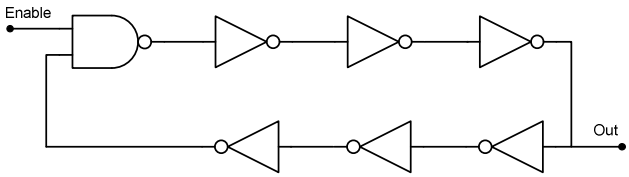


Figure 21: Ring oscillators can be located in product die

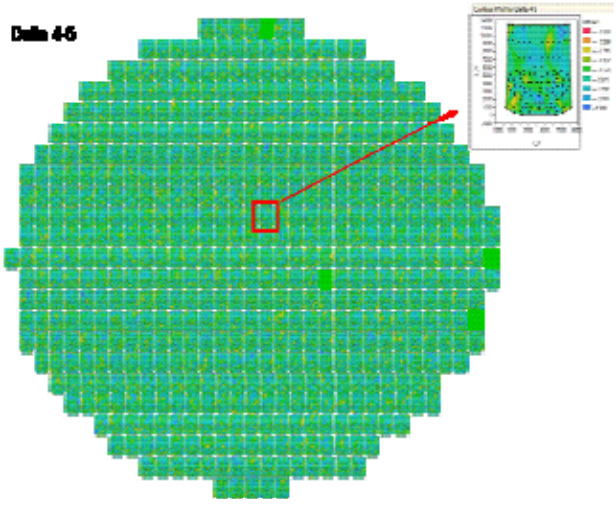


Figure 22: Ring oscillators were added to 45nm products to provide detailed within-die and within-wafer variation data on revenue product material

Figures 23 and 24 show examples of the use of ring-oscillator frequency to determine systematic and random WIW variation across generations. Systematic WIW variation data from ring oscillators (Figure 23) on microprocessor product material illustrates that systematic variation has remained essentially constant across the last four generations. Random WIW variation data from ring oscillators (Figure 24) on microprocessor product material illustrates the ~50% improvement in random variation between the 65nm and 45nm generations enabled by HiK+MG. Note also that the 45nm random WIW variation is comparable to the 130nm process generation.

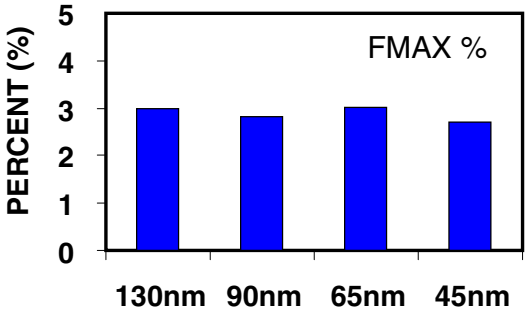


Figure 23: Systematic within-wafer variation data from ring oscillators on microprocessor product material illustrating that systematic variation has remained constant across the last four generations

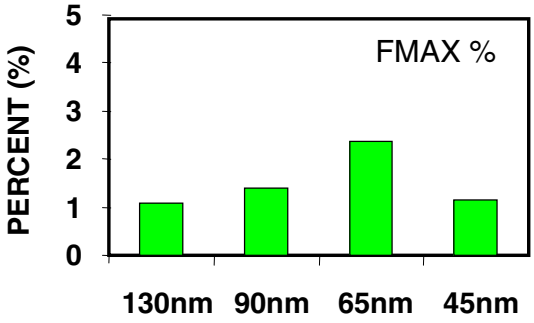
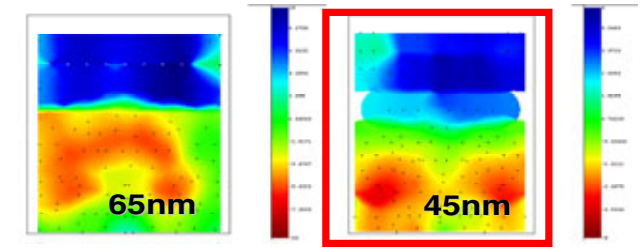
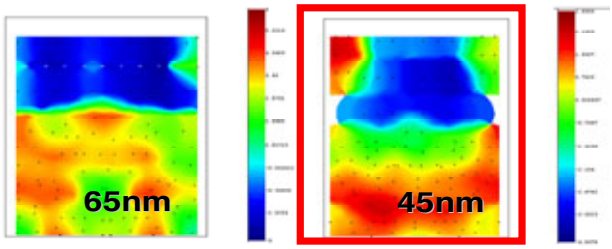


Figure 24: Random within-wafer variation data from ring oscillators on microprocessor product material illustrating the ~50% improvement in random variation enabled by HiK+MG between the 65nm and 45nm generations

Figure 25 gives an example of ring-oscillator data (used in conjunction with a calibration structure) to extract average systematic WID V_T variation comparisons between 65nm and 45nm generations. Note that NMOS has improved 45% (from 20mV to 11mV) and PMOS has improved 22% (from 9mV to 7mV) between the 65nm and 45nm generations.



VTN variation (Mean die VTN – VTN)
Range: 20mV for 65nm → 11mV for 45nm



VTP variation (Mean die VTP – VTP)
Range: 9mV for 65nm → 7mV for 45nm

Figure 25: Ring-oscillator data (used in conjunction with a calibration structure) to extract 65nm to 45nm systematic within-die NMOS and PMOS V_T variation illustrating the improvement is enabled by HiK+MG between the 65nm and 45nm generations

CONCLUSION

Although there has been a trend in the CMOS literature in recent years to convey process variation as a new challenge, process variation has always been a critical element in semiconductor fabrication. From the first discussion of random variation by Shockley in 1961 [3] to the most recent 45nm results [14, 17], understanding and mitigating process variation has been a continuing theme throughout semiconductor history.

While management of process variation is likely to play an increasingly important role in technology scaling, a variety of process, design, and layout techniques can be applied to mitigate the impact of this variation. Examples of pure process mitigation techniques used in the 45nm technology include targeting key transistor properties to reduce RDF, reducing traps at the HiK+MG interface to reduce random charge variation, improving patterning techniques to reduce LER and endcap variation, and improving polishing technologies to reduce systematic cross-wafer variation. Examples of combination design-process techniques used in the 45nm generation include optimizing topology, using OPC to reduce random and systematic variation and adding dummy features to reduce systematic variation. Examples of pure design techniques used in the 45nm generation include chopping techniques to compensate for random variation and common-centroid techniques to compensate for systematic variation.

The success of the 45nm process variation mitigation techniques is well illustrated by 45nm data. In-line measurements of gate CD across four generations show that the 45nm technology generation was able to maintain a 0.7X scaling to prior generations for WID, WIW, and total variation. Intrinsic random variation extracted from matched transistor pairs shows an ~20% improvement in intrinsic random variation from the 65nm to 45nm generations. Systematic WIW variation data from ring oscillators on microprocessor product material illustrates that systematic variation has remained essentially constant across the last four generations. Random WIW variation data from ring oscillators on microprocessor product material illustrates an ~50% improvement in random variation between the 65nm and 45nm generations enabled by HiK+MG. Ring oscillator data (used in conjunction with a calibration structure) shows that NMOS average systematic WID V_T variation has improved 45% (from 20mV to 11mV) and PMOS has improved 22% (from 9mV to 7mV) between the 65nm and 45nm generations.

The key message of this paper is that process variation is not an insurmountable barrier to Moore's Law, but is simply another challenge to be overcome. This message is illustrated with data from the 45nm process generation where process variation is shown to be at least equivalent to (and in many cases better than) process variation in the 65nm and 90nm process generations.

ACKNOWLEDGMENTS

We acknowledge the contributions of the 45nm process and design teams without whom this work would have been impossible. In addition, we acknowledge key contributions from Robert Bigwood, Tao Chen, Martin Giles, Mingwei-Huang, Sean Ma, Peter Moon, Karin Wells-Kilpatrick, and Kevin Zhang as well as advice and support from Mark Bohr and Kaizad Mistry.

REFERENCES

[1] S. Borkar, "Designing reliable systems from unreliable components: the challenges of transistor variability and degradation." *IEEE Microelectronics, Volume 26, Issue 6,* November–December 2005, pp. 10–16.

[2] S. Springer et al., "Modeling of Variation in Submicrometer CMOS ULSI Technologies." *IEEE Transactions on Electron Devices*, *Volume 53, Issue 9*, September 2006, pp. 2168–78.

[3] W. Shockley, "Problems related to p-n junctions in silicon." *Solid-State Electronics*, *Volume 2*, January 1961, pp. 35–67.

[4] R. W. Keyes, "Effect of randomness in the distribution of impurity ions on FET thresholds in integrated electronics." *IEEE Journal Solid-State Circuits*, *Volume SSC-10, Issue 4*, August 1975, pp. 245–247.

[5] W. Schemmert, G. Zimmer, "Threshold-voltage sensitivity of ion-implanted m.o.s. transistors due to process variations." *Electronics Letters*, *Volume 10*, *Issue 9*, May 2, 1974, pp. 151–152.

[6] Yokoyama, K.; Yoshii, A.; Horiguchi, S., "Threshold-Sensitivity Minimization of Short-Channel MOSFET's by Computer Simulation." *IEEE Journal of Solid-State Circuits, Volume 15, Issue 4*, August 1980, pp. 574–579.

[7] Alvarez, A.R.; Akers, L.A.; "Monte Carlo analysis of sensitivity of threshold voltage in small geometry MOSFETs." *Electronics Letters*, *Volume 18, Issue 1*, January 7, 1982, pp. 42–43.

[8] Z, Lin, C.J. Spanos, L.S. Milor, and Y.T. Lin, "Circuit sensitivity to interconnect variation." *IEEE Transactions on Semiconductor Manufacturing, Volume 11, Issue 4*, November 1998, pp. 557–568.

[9] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterisation and modeling of mismatch in MOS transistors for precision analogue design." *IEEE Journal Solid-State Circuits, Volume SSC-21, Issue 6*, December 1986, pp. 1057–1066.

[10] T. Mizuno, J. Okamura, and A. Toriumi, "Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFETs." *IEEE Transactions on Electron Devices, Volume 41, Issue. 11*, November 1994, pp. 2216–2221.

[11] J. T. Horstmann, U. Hilleringmann, and K. F. Goser, "Matching analysis of deposition defined 50-nm MOSFETs." *IEEE Transactions on Electron Devices*, *Volume 45, Issue 1*, January 1997, pp. 299–306.

[12] P. Stolk, F. Widdershoven and D. Klaassen, **"**Modeling statistical dopant fluctuations in MOS transistors." *IEEE Transactions on Electron Devices, Volume 45, Issue 9*, September 1998, pp. 1960–1971.

[13] H. Mahmoodi, S. Mukhopadhyay and K. Roy, "Estimation of delay variations due to random-dopant fluctuations in nanoscale CMOS circuits." *IEEE Journal of Solid State Circuits*, *Volume 40, Issue 9*, September 2005, pp. 1787–1795.

[14] Takeuchi, K. et al., "Understanding Random Threshold Voltage Fluctuation by Comparing Multiple Fabs and Technologies." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2007, pp. 467–470.

[15] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 µm MOSFET's: A 3-D "atomistic simulation study." *IEEE Transactions on Electron Devices*, *Volume 45, Issue 12*, December 1998, pp. 2505–2513.

[16] H. Yang; Macary et al., "Current mismatch due to local dopant fluctuations in MOSFET channel." *IEEE Transactions on Electron Devices*, *Volume 50, Issue 11*, November 2003, pp. 2248–2254.

[17] Kuhn, Kelin J., "Reducing Variation in Advanced Logic Technologies: Approaches to Process and Design for Manufacturability of Nanoscale CMOS." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2007, pp. 471–474.

[18] C. H. Diaz, H.-J. Tao, Y.-C. Ku, A. Yen, and K. Young, "An experimentally validated analytical model for gate line edge roughness (LER) effects on technology scaling*." IEEE Electron Device Letters*, *Volume 22, Issue 6*, June 2001, pp. 287–289.

[19] H.-W. Kim, J.-Y. Lee, J. Shin, S.-G. Woo, H.-K. Cho, and J.-T. Moon, "Experimental investigation of the effect of LWR on sub-100-nm device performance." *IEEE Transactions on Electron Devices*, *Volume 51, Issue 12*, December 2004, pp. 1984–1988.

[20] Fukutome, H., Momiyama, Y., Kubo, T., Tagawa, Y., Aoyama, T., and Arimoto, H., "Direct Evaluation of Gate Line Edge Roughness Impact on Extension Profiles." *IEEE Transactions on Electron Devices, Volume 53, Issue 11*, November 2006, pp. 2755–2763.

[21] Asenov, A., Brown, A.R., Davies, J.H., Kaya, S., and Slavcheva, G., "Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness." *IEEE Transactions on Electron Devices*, *Volume 50, Issue 5*, May 2003, pp. 1254–1260.

[22] A. Asenov, S. Kaya, and J.H. Davies, **"**Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations." *IEEE Transactions on Electron Devices, Volume 49, Issue 1*, January 2002, pp. 112–119.

[23] P. Andrei and I. Mayergoyz, "Quantum mechanical effects on random oxide thickness and random dopant induced fluctuations in ultrasmall semiconductor devices." *Journal of Applied Physics*, *Volume 94, Issue 11*, December 2003 pp. 7163–7172.

[24] E. Cassan, P. Dollfus, S. Galdin, and P. Hesto, "Calculation of direct tunneling gate current through ultrathin oxide and oxide/nitride stacks in MOSFETs and H-MOSFETs." *Microelectronics. Reliability*, *Volume 40*, 2000, pp. 585–588.

[25] A. Asenov and S. Kaya, "Effect of oxide roughness on the threshold voltage fluctuations in decanano MOSFETs with ultrathin gate oxide." In *Proceedings SISPAD*, 2000, pp. 135–138.

[26] M. Koh et al., "Limit of gate oxide thickness scaling in MOSFETs due to apparent threshold voltage fluctuation introduced by tunnelling leakage current." *IEEE Transactions on Electron Devices*, *Volume 48*, *Issue 1,* January 2001, pp. 259–264.

[27] Kaushik, V.S. et al., "Estimation of fixed charge densities in hafnium-silicate gate dielectrics." *IEEE Transactions on Electron Devices*, *Volume 53*, *Issue 10*, October 2006, pp. 2627–2633.

[28] Lucovsky, G., "Intrinsic limitations on the performance and reliability of high-k gate dielectrics for advanced silicon devices." *IEEE International Integrated Reliability Workshop Final Report*, October 17–20, 2005, page 13.

[29] G. Bersuker et al., "Effect of structural properties of Hf-based gate stack on transistor performance." In *Proceedings Materials Research Society Symposium 2004*, *Volume 811*, page 31.

[30] J. L. Gavartin, A. L. Shluger, A. S. Foster, and G. Bersuker, "The role of nitrogen-related defects in high-*K* dielectric oxides: Density-functional studies." *Journal of Applied Physics*, *Volume 97, Issue 5*, March 2005, pp. 053704-1–053704-12.

[31] P. D. Kirsch et al., "Mobility enhancement of high-*k* gate stacks through reduced transient charging." In *Proceedings 35th ESSDERC*, 2005, pp. 367–370.

[32] Wen, H. C. et al., "On Oxygen Deficiency and Fast Transient Charge-Trapping Effects in High-k Dielectrics." *IEEE Electron Device Letters*, *Volume 27, Issue 12*, December 2006, pp. 984–987.

[33] M. Koike et al., "Effect of Hf–N bond on properties of thermally stable amorphous HfSiON and applicability of this material to sub-50nm technology node LSIs." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2003*,* pp. 4.7.1–4.7.4.

[34] M. A. Quevedo-Lopez et al., "High performance gate first HfSiON dielectric satisfying 45 nm node requirements." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2005, pp. 17.1–17.4.

[35] L. Capodieci, "From optical proximity correction to lithography-driven physical design (1996-2006): 10 years of resolution enhancement technology and the roadmap enablers for the next decade." In *Proceedings SPIE Volume. 6154*, 615401.

[36] Nag, S., Chatterjee, A., Taylor, K., Ali, I., O'Brien, S., Aur, S., Luttmer, J.D., and Chen, I.C., "Comparative evaluation of gap-fill dielectrics in shallow trench isolation for sub-0.25 μm technologies." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 1996, pp. 841–845.

[37] Mistry, K. et al., "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2007, pp. 247–250.

[38] Khare, Mukesh; "High-K/Metal Gate Technology: A New Horizon." *IEEE Custom Integrated Circuits Conference, CICC '07*, September 16–19, 2007 pp. 417–420.

[39] Stamper, A.K., McDevitt, T.L., and Luce, S.L., "Sub-0.25-micron interconnection scaling: damascene copper versus subtractive aluminum." *IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop,* September 23–25, 1998, pp. 337–346.

[40] Ruiqi Tian, Xiaoping Tang, and Wong, M.D.F., "Dummy-feature placement for chemical-mechanical polishing uniformity in a shallow-trench isolation process." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *Volume 21, Issue 1*, January 2002, pp. 63–71.

[41] T.K. Yu et al., "A two-dimensional low pass filter model for die-level topography variation resulting from chemical mechanical polishing of ILD films." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 1999, pp. 909–912.

[42] Munkang Choi and Milor, L., "Impact on circuit performance of deterministic within-die variation in

nanoscale semiconductor manufacturing." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *Volume 25, Issue 7*, July 2006, pp. 1350–1367.

[43] K Soumyanath, S. Borkar, Z. Chunyan, and B. Bloechel, "Accurate on-chip interconnect evaluation: a time-domain technique." *IEEE Journal of Solid-State Circuits*, *Volume 34 Issue,* May 1999, pp. 623–631.

[44] V. Mehrotra, "Modeling the effects of manufacturing variation on high-speed microprocessor interconnect performance." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 1998, pp.767–770.

[45] R. H. Dennard et al., "Design of ion-implanted MOSFET'S with very small physical dimensions." *IEEE Journal of Solid-State Circuits, Volume 10, SC-9*, October 1974, pp. 256–268.

[46] Rim, K.; Hoyt, J.L.; and Gibbons, J.F., "Fabrication and analysis of deep submicron strained-Si n-MOSFET's." *IEEE Transactions on Electron Devices, Volume 47, Issue 7*, July 2000, pp. 1406–1415.

[47] Rim, K. et al., "Strained Si NMOSFETs for high performance CMOS technology." *2001 Symposium on VLSI Technology, Digest of Technical Papers*, June, 12–14, 2001, pp. 59–60.

[48] Rim, K. et al., Fabrication and mobility characteristics of ultra-thin strained Si directly on insulator (SSDOI) MOSFETs." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2003, pp. 3.1.1–3.1.4.

[49] Meikei Ieong et al., "Scaling beyond conventional CMOS device." In Proceedings of the 2004 7th International Conference on Solid-State and Integrated Circuits Technology, Volume 1, October 18–21, 2004, pp. 31–34.

[50] Shimizu, A. et al.,"Local mechanical-stress control (LMC): A new technique for CMOS-performance enhancement." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2001, pp. 19.4.1–19.4.4.

[51] Ito, S. et al., "Mechanical stress effect of etch-stop nitride and its impact on deep submicron transistor design." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2000, pp. 247–250.

[52] S. Pidin et al., "A novel strain enhanced CMOS architecture using selectively deposited high tensile and high compressive silicon nitride films." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2004, pp. 213–216.

[53] Thompson, S. et al., "A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low k ILD, and 1 /spl mu/m/sup 2/ SRAM cell." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2002, pp. 61–64.

[54] Ghani, T. et al., "A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2003, pp. 11.6.1–11.6.

[55] Mistry, K. et al., "Delaying forever: Uniaxial strained silicon transistors in a 90nm CMOS technology." *2004 Symposium on VLSI Technology, Digest of Technical Papers*, June 15–17, 2004, pp. 50–51.

[56] Harris, H. et al., "Band-Engineered Low PMOS VT with High-K/Metal Gates Featured in a Dual Channel CMOS Integration Scheme." *2007 IEEE Symposium on VLSI Technology,* June 12–14, 2007, pp. 154–155.

[57] Tsang, Y.L. et al., "Modeling of the Threshold Voltage in Strained Si/Si1-x GexSi1–yGex CMOS Architectures." *IEEE Transactions on Electron Devices, Volume 54, Issue 11*, November 2007, pp. 3040–3048.

[58] Al-Bayati et al., "Advanced CMOS device sensitivity to USJ processes and the required accuracy of doping and activation." In *Proceedings of the 14th International Conference on Ion Implantation Technology,* September 22–27, 2002, pp. 185–188.

[59] Tanaka, T. et al., "Vth fluctuation induced by statistical variation of pocket dopant profile," *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2000, pp. 271–274.

[60] Tanaka, T.; Usuki, T.; Momiyama, Y.; Sugii, T., "Direct measurement of Vth fluctuation caused by impurity positioning." *2000 Symposium on VLSI Technology, Digest of Technical Papers*, June 13–15 2000, pp. 136–137.

[61] Ahsan, I. et al.,"RTA-Driven Intra-Die Variations in Stage Delay, and Parametric Sensitivities for 65nm Technology." *2006 Symposium on VLSI Technology*, *Digest of Technical Papers*, June 2006, pp. 170–171.

[62] H. Fukutome, Y. Momiyama, T. Kubo, E. Yoshida, H. Morioka, M. Tajima, and T. Aoyama, "Suppression of poly-gate-induced fluctuations in carrier profiles of sub-50 nm MOSFETs." *IEEE International Electron Devices Meeting, IEDM Technical Digest*, December 2006, pp. 1–4.

[63] Brown, A.R., Roy, G., and Asenov, A., "Poly-Si-Gate-Related Variability in Decananometer MOSFETs with conventional architecture, *IEEE Transactions on Electron Devices, Volume 54, Issue 11*, November 2007, pp. 3056–3063.

[64] S. Pae et al., "BTI Reliability of 45nm High-K + Metal-Gate Process Technology." *2008 IEEE International Reliability Physics Symposium*, to appear in *IRPS'08 Proceedings*, April 27–May 1, 2008.

AUTHORS' BIOGRAPHIES

Kelin J. Kuhn is an Intel Fellow in the Technology and Manufacturing Group and Director of Logic Device Technology. Dr. Kuhn received a B.S. degree in Electrical Engineering from the University of Washington in 1980, and M.S. and Ph.D. degrees in Electrical Engineering from Stanford University in 1985. Dr. Kuhn is presently the device lead of the pathfinding team responsible for the transistor architecture for Intel's 22nm process technology and was the device lead for Intel's 45nm process. She joined Intel Corporation in 1997, and since then she has held a variety of technical positions on the 0.35um, 130nm, and 90nm nodes. Previously, Kuhn was a tenured faculty member in the Department of Electrical and Computer Engineering at the University of Washington. Dr. Kuhn is the author of numerous technical publications on the optical and electronic properties of semiconductor devices as well as the author of the textbook *Laser Engineering*. She received the 2006 Intel Achievement award for her work in enabling Intel's 45nm node High-*k* metal gate transistor process. Her e-mail is kelin.ptd.kuhn at intel.com.

Chris Kenyon is a lithography group leader in Intel's Logic Technology Development organization. He joined Intel in 1996 and has worked primarily on Intel's gate-patterning process since that time. He is currently responsible for developing the gate-patterning process for the 32 and 22nm nodes. He received his B.A. degree from Princeton University in 1990 and his Ph.D. degree from Caltech in 1996 in Physical Chemistry. His e-mail is Chris.Kenyon at intel.com.

Avner Kornfeld is a Principal Engineer in the Technology and Manufacturing Group working on analog design, variation, and on the interaction between design and process. Prior to his current role he worked on power and frequency optimization and projection for Intel mobile microprocessors in 130nm, 90nm, 65nm, and 45nm nodes. He joined Intel in 1992 and worked for the Intel Communication and the Mobile Microprocessor Groups. He received B.Sc., M.Sc., and D.Sc. degrees in Electrical Engineering from the Technion–Israel Institute of Technology in 1980, 1982, and 1986, respectively. Dr. Kornfeld was a lecturer in the Department of Electrical Engineering–Technion, from 1986 to 1991 working on cryogenic focal plane signal processing and device flicker noise. He received the 1997 Intel Achievement Award for his work on Intel's 100/10 Mbps fast Ethernet transceiver design. His e-mail is avner.kornfeld at intel.com.

Sam Sivakumar is an Intel Fellow and Director of Lithography in Intel's Portland Technology Development Group in Oregon. He is responsible for the definition, development, and deployment of Intel's next-generation lithography processes. Sivakumar joined Intel in 1990 and throughout his career with the company has worked in the lithography area on photoresists, patterning equipment, and process development. He has contributed to lithography development, characterization, and transfer to high-volume manufacturing of every submicron process technology generation at Intel since 1990.

Greg Taylor is an Intel Fellow in the Technology and Manufacturing Group and Director of Mixed Signal Circuit Technology. He is responsible for the delivery of the analog circuits needed by future microprocessors. Prior to his current role he worked on the development of eight generations of Intel microprocessors since joining the company in 1991, with a focus on the mixed signal portions of these chips. Before joining Intel, Dr. Taylor worked at Bipolar Integrated Technology where he worked on ECL floating point units and two RISC microprocessors. He received the B.S., M.S., and D.Eng. degrees in Computer and Systems Engineering from Rensselaer Polytechnic Institute in 1981, 1983, and 1985, respectively, where he was supported by a Fannie and John Hertz Foundation Fellowship. His e-mail is greg.design.taylor at intel.com.

Peter VanDerVoorn is a Device Engineer in the Technology and Manufacturing Group. Dr. VanDerVoorn received the B.S. and Ph.D. degrees in Electrical Engineering from Cornell University in 1992 and 1998. He is presently the Device Engineer responsible for 45nm process enhancement and has been involved in the 45nm High-K/metal gate program from the pathfinding phase. Dr. VanDerVoorn joined Intel Corporation in 1998 and has since held Process Integration and Device Engineering positions in the 130nm, 90nm, and 45nm nodes. His e-mail is peter.j.VanDerVoorn at intel.com.

Keith Zawadzki is a Device Engineer in Intel's Logic Technology Development organization. He joined Intel in

1998 and is currently responsible for developing 22nm process technology. He received a B.Sc. degree in Microelectronic Engineering from the Rochester Institute of Technology in 1996 and an EEMS degree from the University of Texas, Austin in 1998. His e-mail is Keith.E.Zawadzki at intel.com.

Mark Y. Liu is a Process Integration Engineer in Intel's Logic Technology Development organization. He joined Intel in 1995 as a Process Engineer in the ion implantation group and later held positions as Ion Implantation Group Leader and Advanced Thermal Annealing Group Leader. He received his Ph.D. degree in Electrical Engineering from the University of Minnesota and his B.S. degree in Physics from Beijing University in China. His e-mail is mark.y.liu at intel.com.

Atul Maheshwari is a Senior Design Engineer in the Logic Technology Development Department of the Technology and Manufacturing Group. He is responsible for the delivery of performance and process variation probing circuits for 45nm and 32nm process technology. He received his B.E degree from Gujarat University, India in 1998 and M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Massachusetts, Amherst in 2001 and 2004, respectively. His e-mail is atul.maheshwari at intel.com.

Wei-Kai Shih received a Ph.D. degree in Electrical Engineering from the University of Texas at Austin in 1997, with a focus on device physics and TCAD simulation. He joined Intel TCAD in 1997 to develop a full-chip thermal reliability methodology and related simulation tools, which later became part of RV-TPRSIM. He has also worked extensively in the area of compact transistor modeling, specifically gate leakage modeling and analog-RF modeling. Wei-kai Shih is currently a Group Lead in the DTS Core CAD Technology Department, managing a group of 15 CAD developers to deliver interconnect, process variation, and reliability models and simulation tools, as well as their application to technology and design co-optimization.

THIS PAGE INTENTIONALLY LEFT BLANK

45nm SRAM Technology Development and Technology Lead Vehicle

Uddalak Bhattacharya, Technology and Manufacturing Group, Intel Corporation
Yih Wang, Technology and Manufacturing Group, Intel Corporation
Fatih Hamzaoglu, Technology and Manufacturing Group, Intel Corporation
Yong-Gee Ng, Technology and Manufacturing Group, Intel Corporation
Liqiong Wei, Technology and Manufacturing Group, Intel Corporation
Zhanping Chen, Technology and Manufacturing Group, Intel Corporation
Joe Rohlman, Technology and Manufacturing Group, Intel Corporation
Ian Young, Technology and Manufacturing Group, Intel Corporation
Kevin Zhang, Technology and Manufacturing Group, Intel Corporation

Index words: SRAM, sleep transistor, forward body bias, ECC, raster, GTL, PLL, DTS, DLL, VCO, QPI

ABSTRACT

Every new generation of process technology at Intel is developed and certified using an SRAM-based "X-chip." X6 is the technology lead vehicle used for the 45nm technology serving as a platform for the co-optimization of circuit design and process technology for SRAMs as well as critical design collaterals for products.

As the workhorse of the embedded memory, SRAMs play an essential role in all Intel products in achieving power-performance goals. SRAMs are also ideally suited for process-defect sensitivity and detection. The SRAMs on X6 had featured several different SRAM designs and register files that were individually optimized to take advantage of the Hi-K metal gate process for various product applications. Intel's revolutionary 45nm technology was instrumental for aggressive SRAM scaling. The tileable SRAM array in X6 was architected to directly support product applications. The X6 also served as the vehicle for several critical memory circuit technology developments, including second-generation dynamic sleep control and dynamic Forward Body Bias (FBB). To support process and design learning, X6 includes an infrastructure of advanced test features: for example, an Error Correction Code (ECC) emulator is designed to quantify the benefit of error corrections and a Programmable Built-in Self Test (PBIST) for high-speed testing with raster capability. Fine-granularity In-Die-Variation (IDV) oscillators track process variation. Critical circuits such as electrically programmable Fuse, Phase Lock Loop (PLL), Digital Thermal Sensor (DTS), and advanced I/Os allowed technologists and designers to work closely to optimize the process and circuits earlier.

The X6 testchip successfully met the goals of process and critical collateral certification to support both process and product development needs and played an essential role in Intel's rapid product ramp at the 45nm technology node.

INTRODUCTION

"X-chips" have been serving the needs of process technology development at Intel for many generations. They have always been based on SRAM building blocks. SRAM has the unique advantages of stressing the most critical design rules of a process technology, providing large area coverage that increases defect sensitivity, and of providing fine addressable granularity to enable fault isolation and analysis.

In this paper we focus on X6, the testchip that was used to certify the Hi-K metal gate process technology at the 45nm node. Over the generations, the X-chips have evolved beyond the original goal of providing test data needed to identify process defects, to a platform for the co-optimization of process technology and critical design collaterals for the products. In the following section, we provide an overview of the X-chip.

Testability is a key concern with X-chips, and X6 is no exception. The testchip must be sensitive to process defects and must exercise the critical design rules. A process should not get certified on the testchip and encounter yield or performance issues when a product is

subsequently ramped on it. The complexity of modern semiconductor processes results in significant lead time from wafer starts to end of line, while the total time available to certify a process remains, at best, the same. For the process engineer this means only a limited number of "info-turns" during the development phase. To reduce this information turnaround time, X6 includes a test infrastructure that allows the collection of relevant data with reduced test time and features that aid fault isolation. We discuss this aspect further in the test features section.

We now provide descriptions of the critical design collateral content of X6 SRAMs and some key mixed-signal circuits that have high process sensitivities in the context of product design. Therefore, these classes of circuits are obvious choices to be included in the testchip. We discuss a product-ready and tileable common SRAM subarray design along with performance results. In addition to 6T SRAM arrays, X6 also included multi-port SRAM, a.k.a., register file (RF) memories, for the first time to broaden the process and product co-learning. New fuse technology is developed with X6 to provide for expanded product requirements. A significant number of analog elements such as different kinds of Phase Lock Loop (PLL), Digital Thermal Sensors (DTS), and advanced I/O circuits were also part of the X6 testchip. In the following sections these collaterals are described along with the key learnings for technology development and circuit optimization.

X-CHIP OVERVIEW

The X-chip series started many generations ago with the X1 testchip that was the certification vehicle for the 180nm node process technology. From X1 to X6 the role and content of the testchip has grown significantly. However, in the midst of these changes some elements and challenges remain the same.

The memory cell still remains the most critical collateral, the design of which requires careful consideration. The footprint of the memory cell often sets the design rule limit at a given technology node, and it has a large impact on the product die-size, given the tens of megabyte-sized caches of modern CPUs. The stability of the cell at lower voltages often determines the low voltage operation of the products and hence the power consumption. Any systematic process defect mechanism will have a huge impact on product yield due to the tiling of millions of memory bits in product caches. Due to these reasons it has been a product design requirement for several generations to copy exactly the memory bits developed on the X-chips.

The SRAM bits are organized into addressable units called subarrays that have a rectangular matrix-like structure with rows and columns. During a read or write operation from the memory subarray, a specific row and specific columns are activated depending on the address, and a group of bits called a Word are read or written. A subarray is designed to be very compact since it is tiled many times to form an on-die cache in a real product. The compactness can be achieved by exercising the tightest design rules. Once a subarray design is complete, large portions of the X-chip area can be tiled with these subarrays with minimal additional effort. Thus, the SRAM on the X-chips has always proved to be sensitive to many different kinds of process defects. The fine granular addressability of the compact subarrays allows rapid isolation of the defect location.

On the test infrastructure side X-chips have always carried an I/O interface that can communicate at the desired speeds with all the different tester platforms. A Programmable Built-in Self Test (PBIST) has also been an integral part of X-chips to allow high-speed, on-die testing and burn-in tests. Among the mixed-signal collaterals, X-chips have always included the PLL. This also serves as a clock multiplier for on-die, high-speed SRAM testing.

X-chip design faces some unique challenges compared to product design. All the key electrical parameters of the transistors and interconnects will evolve significantly from the beginning to the final goal of the technology, when the process reaches its maturity. Yet, if the design is not robust enough to comprehend this evolution and still be largely functional, it will not be able to provide robust silicon data for continuous process learning. Testchip circuits are designed largely based on engineering judgment with limited simulation data available to the designers when the design is finalized. The concept of "correct-by-construction" is essential in ensuring the robust design. Layout design rules get defined almost simultaneously with the testchip design process. This creates uncertainties and the inevitable last-minute changes for testchip designers. The impact of design rule uncertainties is mitigated by working closely with the rule-definition teams. Another reason for tight coupling with the design-rule-development activity is that the testchip often throws up scenarios not considered at the time of design rule definition. The testchip provides the first reality check for the process design rule set and validation flows. The design time allowed for the testchip is rather short. The testchip design schedule must be synchronized with the process integration schedule so that it does not become a bottleneck for technology development. Process learning accelerates significantly with the arrival of the first testchip silicon.

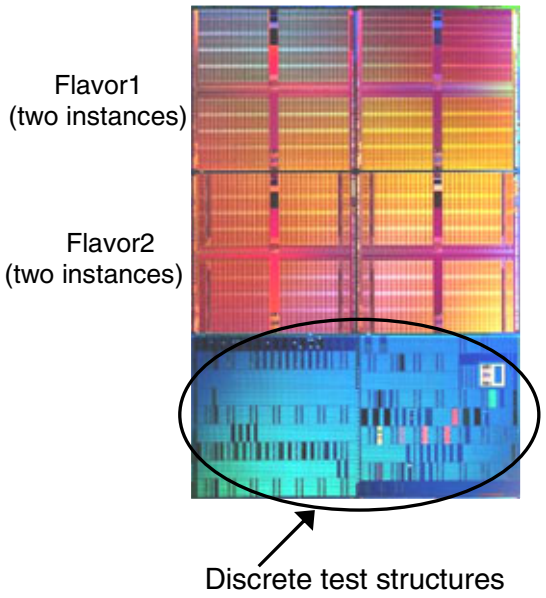


Figure 1: X6 reticle

Figure 1 shows a diagram of the X6 reticle. There are two different flavors of die on this reticle, and each die is instanced twice. Thus we have four SRAM-based dice on the reticle, and two dice are allocated for discrete test structures. We focus on the SRAM-based X6 testchip. The two flavors of the X6-SRAM testchip have similar architecture and content. However, having two different flavors allows designers to test different flavors of the same circuits and compare directly the merits of each. While the same can be done on a single testchip design, having two die flavors makes it convenient by adding parallelism to the design process, which then shortens the design time.

X6 Test Features

The main SRAM chip-level tileable unit is called the Raster Unit (RU). An RU is a portion of SRAM convenient for generating raster maps. It is significantly bigger than a small subarray but small enough to be tested in a reasonable time. All RUs can be tested in parallel, which saves test time. The RU-based architecture of the testchip allows easy integration of different flavors of memory bits and subarray designs while ensuring that each flavor is statistically significant in size.

The chip-level inputs are collected from the I/Os located at two edges of the die. The major busses are located along a center spine of the die connecting the RUs to the I/O and other control blocks. Another center spine in the orthogonal direction contains the major non-SRAM blocks such as the PBIST, Test Access Port (TAP), Fuse unit, and PLL. The synchronous circuits are all located within the RU with the exception of PBIST. There are two clock domains in each RU. One of them is a high-frequency clock and the other a low-frequency clock. The high-frequency clock is contained entirely within the RU, and no synchronous paths communicate between the RUs. This feature makes the chip design very tolerant of clock skew.

The PBIST on X6 can be programmed to generate the test patterns required for SRAM testing. This is useful in cases where the tester cannot support complex test patterns, as in the burn-in test platform and also for high-frequency testing where expensive high-speed testers and complex I/O circuitry can be avoided by on-die testing. The scheme for high-frequency testing is illustrated in Figure 2. The PBIST is part of the low-frequency clock domain. The advantage is that the PBIST design is unlikely to become the bottleneck for frequency and hence is tolerant to the process being off-target in terms of performance, during the development stage. A Parallel-to-Serial (PS) converter converts the low-frequency instructions received as parallel instructions to high-speed serial instructions [1]. An on-die compare circuit compares the “read” data from the SRAM, and the “expect” data from the PBIST pattern to determine pass/fail.

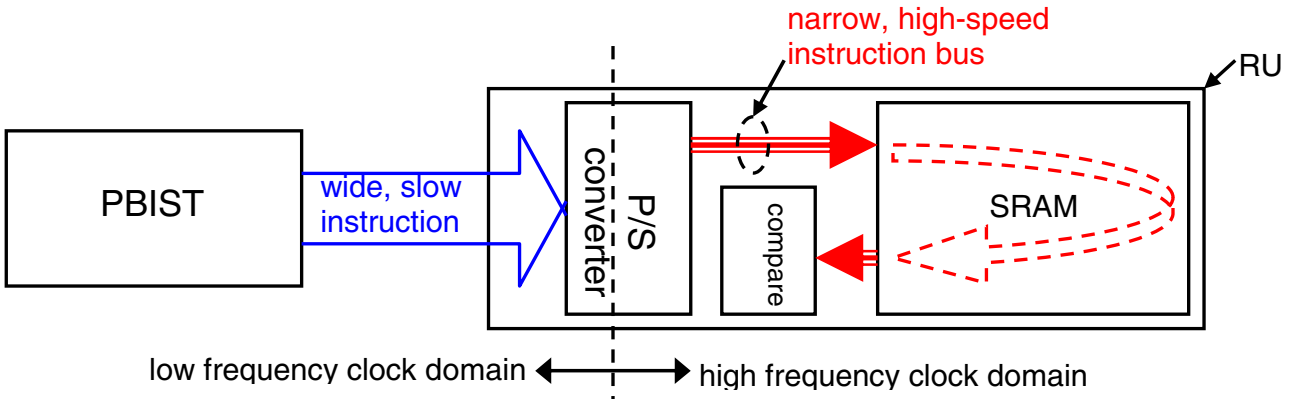


Figure 2: High-frequency testing scheme

The high-frequency testing scheme described above is adequate for determining the SRAM maximum frequency of operation (Fmax). However, for SRAM design validation and process learnings, it is also quite useful to have raster data at Fmax to determine what limits Fmax—especially for products. The high-frequency raster feature will stall the PBIST when a failure occurs. The faulty addresses will be recorded in on-die storage elements and later can be scanned out to identify the location of the faulty bits.

Error Correction Codes (ECC) are becoming very important in modern SRAM design. ECC is an essential part of the large array for data integrity under all product conditions. Since X6 is used to optimize the SRAM memory bit and subarray design, it is important to quantify the impacts of having ECC. X6 has a simple ECC emulator that can quantify the benefit of ECC.

X6 has a TAP that controls the scan function. There are many scan chains dedicated to a specific purpose. Two major roles of the scan chains for process learning are to physically isolate a defect and to support process variation tracking through the use of In-Die-Variation (IDV) oscillators. When a process defect occurs between two stages in pipelined implementation of synchronous circuits, it is possible to feed the logic cone with the defect with various input combinations by using a scan chain and also by recording the output of the logic cone the same way. By analyzing the logic cone output for different input combinations you can, in a large number of cases, pinpoint the failing device. The majority of the circuitry above the subarray level in the RUs is pipelined, and scan provides a useful method of failure isolation and process learning. IDV oscillators are well known for their ability to track process variation of different process parameters ranging from transistor performance to various kinds of leakage, interconnect delay, and device mismatch [2]. The easiest way to control and extract data from IDV oscillators is through scan chains. X6 provides a robust set of scan chains for this purpose that allow a fine spatial granularity of IDV oscillators on-chip, which otherwise would be impossible to manage using I/O bumps alone.

X6 Memory Design: SRAM and Register File

CMOS technology scaling continues to drive the increase of on-die memory density to meet performance needs in applications such as microprocessors. Meanwhile, the device variation and leakage are increasing as the miniaturization of the transistor continues. As a result, it has become increasingly challenging to design SRAM with an adequate stability margin for low-voltage operation while at the same time keeping the power consumption low enough to meet system-level power requirements. The X6 testchip is a 153Mb SRAM design that is optimized for a 45nm Hi-K Metal-Gate technology [3].

Figure 3 shows the SEM top-down view of the 0.346 μm^2 6T-SRAM cell fabricated in the 45nm Hi-K Metal-Gate CMOS technology [3]. The cell design takes full advantage of the new technology features such as trench-contact in achieving high-density and low-voltage operation. The Hi-K Metal Gate transistor technology essentially eliminates the gate leakage in the bitcell. The overall measured cell leakage is reduced by 10x, as shown in Figure 3. This leakage reduction provides enormous benefits in power-constrained applications.

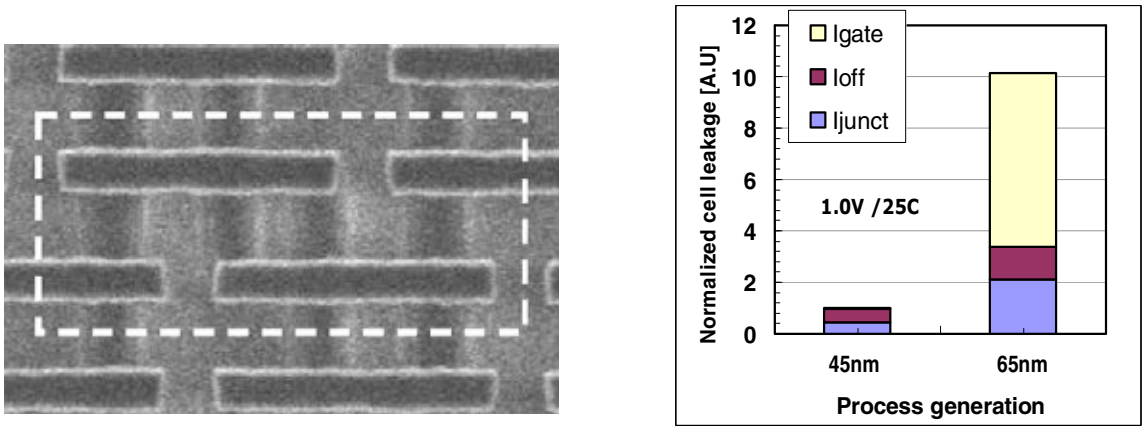


Figure 3: SEM top-down view of 0.346 mm^2 SRAM cell in 45nm technology (left); SRAM cell leakage comparison between 65nm and 45nm technologies (right)

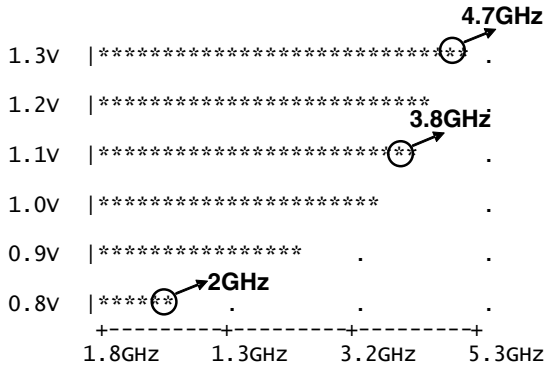


Figure 4: Measured voltage-frequency schmoo

Figure 5 describes the array architecture and configuration along with critical timing of the array. A 16KB subarray with a 256-row bitline and a 128-column wordline is first constructed to achieve optimal array efficiency while meeting an aggressive frequency target. The subarray has both built-in column and row redundancies for yield improvement. The terminology in Figure 5 for rows and columns is (M+N) where M is the base number and N is the redundant value. To meet today's microprocessor bandwidth requirements, the subarray is designed to support 64-bit wide Read and Write. Write is performed with a 64-bit data stream within a single cycle. Read data comes out of the subarray in two consecutive cycles with two 32-bit "chunks," in order to minimize the global routing congestion. Both Sleep and Forward Body Bias (FBB) have independent control circuits along the 256x256 sector boundary, which provides the fine granularity to achieve balanced design between switching power and area efficiency. The design has achieved up to 3.8GHz operating frequency at 1.1V power supply as shown in Figure 4.

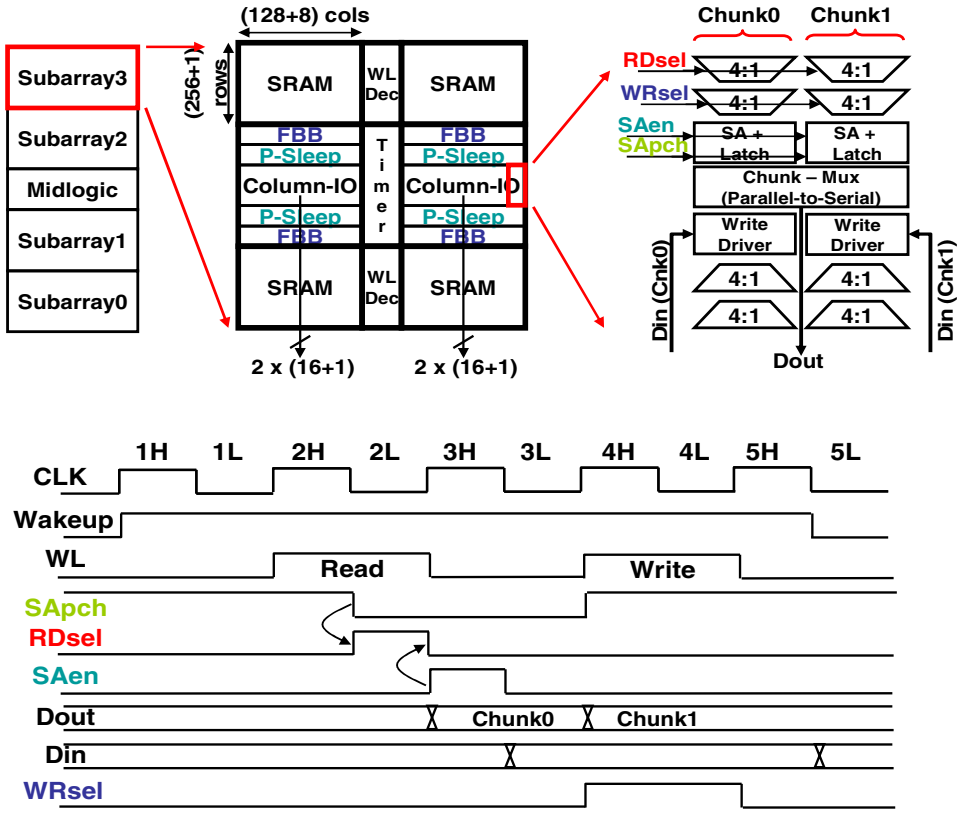


Figure 5: 16-KB subarray configuration and critical signals' timing diagram

The PMOS strength in the 6T SRAM cell is essential to maintain the cell stability during the active mode for low-voltage operation. Using lower threshold voltage PMOS is often prohibitive due to excessive transistor leakage. In X6, a dynamic FBB for the PMOS in the SRAM cell is developed to improve the robustness of low-voltage operation while meeting stringent product and manufacturing requirements at minimum design overhead. Figure 6 shows the critical circuits for the dynamic FBB design.

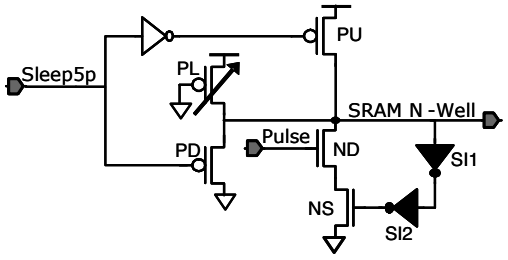


Figure 6: Dynamic SRAM PMOS FBB circuit

The amount of FBB is determined by the ratio of two PMOS devices (PL and PD), and these devices have a built-in programming control. To meet fast and dynamic requirements, an NMOS pull-down path, formed by transistors ND and NS and controlled by a pulse signal, is employed to achieve the fast voltage transition at the N-well. A feedback or shutoff mechanism is also used to prevent the N-well voltage from dropping too low and causing excessive junction leakage. The trip points of the inverters, SI1 and SI2, are optimized to meet this need. The pull-down signal pulse is programmable and generated off the wakeup signal. It starts the discharge of the N-well voltage one cycle before the WL is turned on in order to ensure that the N-well voltage level has reached the intended static level. By applying the FBB selectively to the activated portions of the large array, the overall power impact from FBB is kept to a minimum. Test results demonstrate that a wide range of FBB strength can be achieved under high-frequency operation. The stronger PMOS under FBB can improve the minimum operating voltage up to 75mV without increasing the overall SRAM leakage power dissipation.

The use of dynamic sleep design to lower SRAM power supply has been proven to be effective in reducing the static power consumption by reducing leakage [4]. The SRAM voltage control during the sleep state is critical to maintain the integrity of the data stored in the array. High-volume manufacturing requires programming capability [5]. A scheme with active feedback control on SRAM VCC has been proposed to improve Process-Voltage-

Temperature (PVT) variation [6] requiring off-chip voltage reference. In this design, an on-die programmable voltage generator is designed with N-well-based precision resistors. It provides low design overhead as well as insensitivity to different PVT conditions. A simplified Op-Amp further reduces the overall area overhead down to less than 2%. The integrated new design is shown in Figure 7. SRAM VCC is more sensitive to temperature, as it is dependent on subthreshold leakage, which is the dominant leakage source in Hi-K metal gate technology where gate leakage has been eliminated. The use of the active control along with the temperature-insensitive on-die reference voltage generator provides a much tighter SRAM VCC distribution compared to passive control. This translates into about 100mV lower standby voltage during sleep state and better leakage reduction.

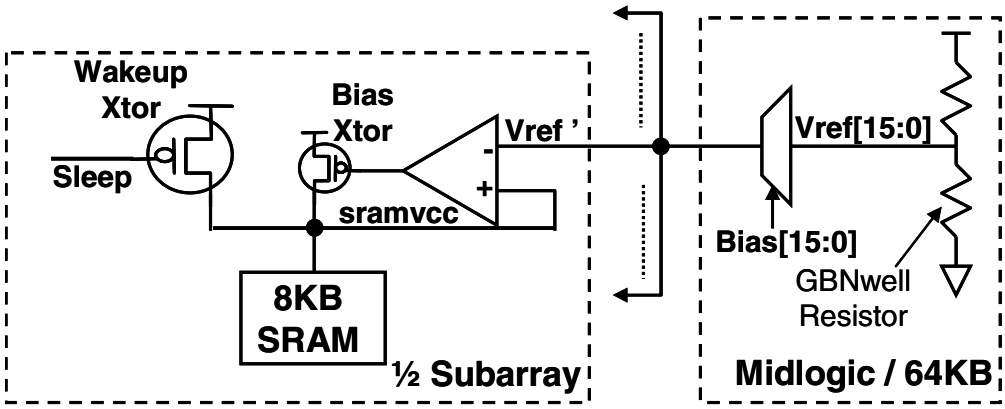


Figure 7: Active SRAM VCC control with integrated on-die programmable reference voltage generation

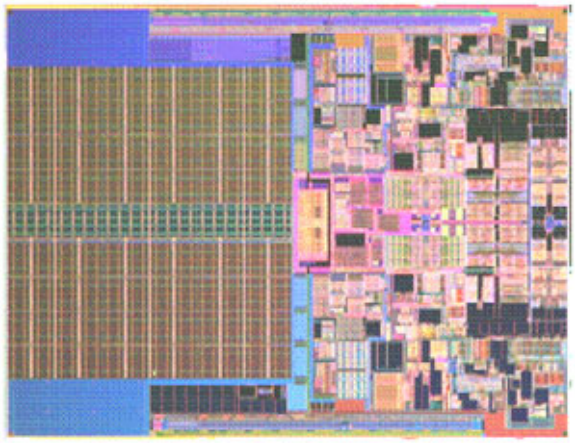


Figure 8: 45nm Intel® Core™2 Processor with 6MB L2 cache that is formed with the common 16KB subarray

The modular architecture of the X6 SRAM design described above has enabled the 16KB subarray to be used directly as the building block for a 6MB L2 cache in the next-generation Intel® Core™2 processor-based CPU [7]. The die photo of this product with the L2 cache highlighted is shown in Figure 8. This silicon-verified and process-optimized subarray was one of the key contributors to the successful production ramp of this product. Thus, the utility of using a common SRAM design for similar applications across different products to reduce manufacturing risk is clearly established. A detailed description of the SRAM design can be found in [8].

Multi-port register file memories are very common and important in CPU and other logic products. Due to unique circuit topologies used in this kind of memory, the sensitivity to process variation, e.g., NMOS and PMOS strength ratio, has increased as technology scaling continues. The X6 testchip contained several important RF arrays that are directly used by the lead CPU products at Intel. They have proved to be very effective in optimizing the transistor parameters as well as providing better circuit-modeling enhancement for various designs.

X6 FUSE

Fuse-based electrically programmable read-only memories (PROM) have been widely used in microprocessors for a variety of important applications. These include improving microprocessor yield by repairing defective SRAM bits using row/column redundancy, permanently storing die identification, and effectively trimming devices used in analog circuits such as thermal sensors and I/O. Fuse designs in X6 were optimized for the Hi-K metal gate process. In this generation we developed an array-based design for high-density fuses that significantly improved the area efficiency of the fuse block for increased product applications.

X6 Mixed Signal and Analog Circuits

The X6 technology lead vehicle includes a group of mixed signal and analog collateral circuits. CPU products require increasing amounts of analog circuit content. The analog technology dependencies differ from digital dependencies by device and component type, and they have a stronger focus on gain and operating range metrics as opposed to digital drive current and capacitive load. A technology that produces high-yielding, high-performing SRAM and digital logic does not ensure analog circuit functionality and performance. Including analog circuits on the X6 technology lead vehicle enabled evaluation of unique analog technology requirements.

A set of circuits targeted at the lead 45nm product were incorporated in the design to allow greater technology flexibility. They include PLL, Gunning Transceiver Logic (GTL) I/O buffers, and DTS. The PLL design also includes multiple types of Voltage Controlled Oscillators (VCO) for design optimization. The GTL I/O buffer contains programmable termination resistors for design tuning. All these circuits have been directly used in the lead product design. Figure 9 shows a waveform captured from the product I/O.

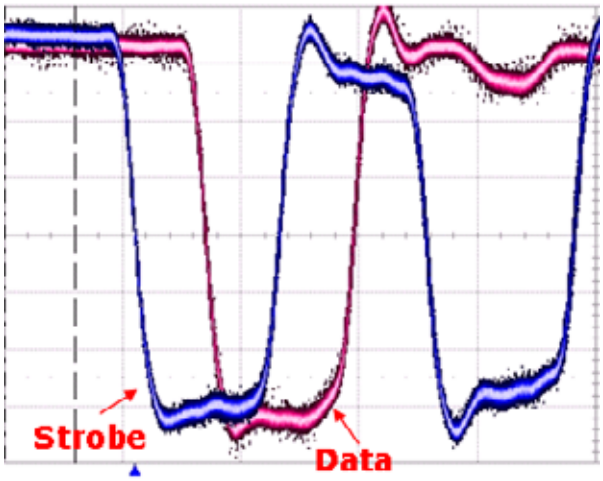


Figure 9: GTL I/O waveform (X-axis 1ns/div)

As CPUs demands more data bandwidth, the future generation of Intel CPUs will feature new high-speed serial I/Os, also called Quick Path Interconnect (QPI) circuits. Many circuits used in the physical implementation of QPI circuits have strong process technology dependency. Several critical components of QPI circuits, including differential transceiver, Delay Locked Loop (DLL), duty cycle correction circuits, and low-jitter LC-PLL are incorporated in X6 and provide significant early silicon-based learning. A capability of 6.4GT/s has been demonstrated as shown by the eye-diagram in Figure 10.

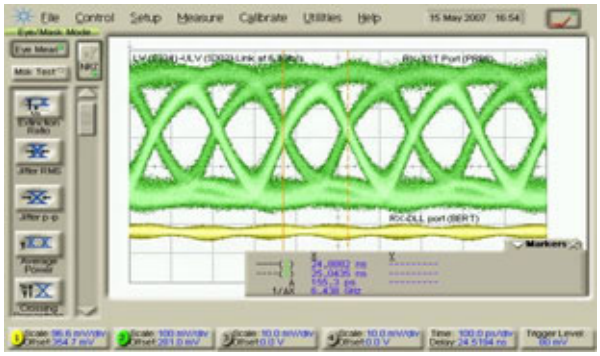


Figure 10: 6.4GT/s operation shows clean eye-diagram and error-free with BIST

SUMMARY

The role of the X6 testchip in the 45nm Hi-K metal gate process development is discussed in this paper. X6 played a critical role in supporting process development by providing a platform for process yield and performance enhancement. In parallel it also served as a test bed for co-optimization of critical design collaterals with the process technology. There exists a significant advantage in having the process technology and the critical design collaterals ready simultaneously for product use. Design considerations and performance of some of the collaterals such as the SRAM memory bit and subarray, PLL, DTS, and high-speed I/Os has been outlined. The early effort on X6 was a key contributor to the success of the lead product ramp in the 45nm Hi-K metal gate technology.

ACKNOWLEDGMENTS

We are grateful to many members of the PTD technical staff including members of advanced memory design, mixed signal circuit collateral development, advanced circuits and technology integration, and process technology teams for their contributions to this work.

REFERENCES

[1] W. Chen and U. Bhattacharya, "Method for bi-directional data synchronization between different clock frequencies." *US Patent 6,956,918.*

[2] M. Hatizalambrou, A. Neureuther, and C. Spanos, "Ring oscillator sensitivity to spatial process variation." *1st International workshop on Statistical Metrology (IWSM)*, June 1996.

[3] K. Mistry, et al., "A 45nm Logic Technology with higk-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging." *IEDM Technical Digest*, December 2007, pp. 247–250.

[4] K.-S. Min, K. Kanda, and T. Sakurai, "Row-by-row dynamic source-line voltage control (RRDSV) scheme for two orders of magnitude leakage current reduction of sub-1-V-VDD SRAM's." In *Proceedings IEEE*

International Symposium Low Power Electronics and Design (ISLPED), August 2003, pp. 66–71.

[5] K. Zhang, et al., "SRAM design on 65-nm CMOS technology with dynamic sleep transistor for leakage reduction." *IEEE J. Solid-State Circuits, vol. 40*, April 2005, pp. 895–901.

[6] M. Khellah, et al., "A 256-Kb Dual-VCC SRAM Building Block in 65-nm CMOS Process with Actively Clamped Sleep Transistor." *IEEE J. Solid-State Circuits, vol.42*, January 2007, pp. 233–242.

[7] V. George, et al., "Penryn: 45-nm Next Generation Intel® Core™2 Processor." *ASSCC Technical Digest*, November 2007, pp. 14–17.

[8] F. Hamzaoglu, et al., "A 153Mb-SRAM design with dynamic stability enhancement and leakage reduction in 45nm High-K Metal-Gate CMOS technology." *ISSCC Technical Digest*, February 2008, pp. 376–377.

AUTHORS' BIOGRAPHIES

Uddalak Bhattacharya is a Principal Engineer in the Logic Technology Development Organization at Intel Corporation. His interests include SRAM test circuit design, testing, measurement, and yield analysis. He received his Ph.D. degree in Electrical Engineering from the University of California, Santa Barbara in 1996. His e-mail is uddalak.bhattacharya at intel.com.

Yih Wang is a Senior Design Engineer in Intel's Logic Technology Development Organization. He joined Intel in 2001 after receiving his Ph.D. degree from the University of Florida. He has worked on SRAM developments for a variety of Intel logic process technologies. Currently, he is responsible for developing the low-power and high-speed SRAM designs for future logic processes. His e-mail is yih.wang at intel.com.

Fatih Hamzaoglu received his Ph.D. degree from the University of Virginia, Charlottesville, in 2002, in Electrical Engineering. After his Ph.D., he joined the Logic Technology Development Group at Intel Corporation. Since then, he has been working on low-power, high-performance SRAM cache design. He is the author/co-author of 19 papers and holds four issued and 6 pending patents. His e-mail is fatih.hamzaoglu at intel.com.

Yong-Gee Ng is a Senior Design Engineer at the Portland Technology Development Group. His interests are in low-power, high-performance SRAM cache designs. He received B.A. and M.S. degrees in Electrical Engineering from Arizona State University, Tempe, in 1989 and 1991, respectively. His e-mail is yong-gee.ng at intel.com.

Liqiong Wei is a Component Design Engineer in the Logic Technology Development Group at Intel. Her technical interests include low-power, high-performance memory design and validation. She received her B.S. and M.S. degrees in Electrical Engineering from Peking University, Beijing, and a Ph.D. degree in Electrical and Computer Engineering from Purdue University. She is a member of the IEEE. Her e-mail is liqiong.wei at intel.com.

Zhanping Chen received a B.S. degree in Computer Science and Technology from Peking University, Beijing, China, and a Ph.D. degree in Electrical and Computer Engineering from Purdue University, West Lafayette, IN, in 1991 and 1999, respectively. He joined Intel Corporation in June 1999 and currently leads the development of programmable read-only memory (PROM). His research interests include CAD tool development and circuit design/optimization for low power and high performance. Dr. Chen received the Best Paper Award at the 2000 IEEE International Symposium on Quality Electronic Design, San Jose, CA. His e-mail is zhanping.chen at intel.com.

Joe Rohlman is an Engineering Manager in the Advanced Design Group in the Logic Technology Development Organization. He leads a team of engineers focused on mixed signal circuit design on Intel's next-generation process technology. Prior to joining the Logic Technology Development Organization, Joe held management and design roles on Microprocessor and Ethernet product teams. Joe received an MSEE degree from the University of Michigan. His e-mail is joseph.f.rohlman at intel.com.

Ian Young is a Senior Fellow and Director of Advanced Circuits and Technology Integration in the Technology and Manufacturing Group at Intel. He received his B.A. and M.S. degrees in Electrical Engineering from the University of Melbourne, Australia. He received his Ph.D. in Electrical Engineering from the University of California, Berkeley. He is currently directing the development of high-speed serial I/O and mixed-signal analog circuits in 32nm logic and SOC technologies and doing research for optical I/O. His e-mail is ian.young at intel.com.

Kevin Zhang is an Intel Fellow and Director of Advanced Memory Circuits and Technology Integration in the Logic Technology Development Group at Intel Corporation. He is responsible for embedded memory technology development and has led the design and validation of technology vehicles from 90nm to 45nm technology generations, which played a key role in both process technology development and product design. He holds a Ph.D. degree in Electrical Engineering from Duke University. His e-mail is kevin.zhang at intel.com.

45nm Design for Manufacturing

Clair Webb, Technical and Manufacturing Group, Intel Corporation

Index words: DFM, 45nm, design rules

ABSTRACT

Co-optimization between design and process is required for a highly manufacturable process technology. This paper discusses this co-optimization and how it meets the challenges for maintaining Moore's Law while delivering new processes and designs capable of fast ramp to high yields. Poly is one of the most critical layers for control of variation, and it needs the most restrictive rules. We show the change in poly rules over the last few processes to illustrate how rules have changed to meet manufacturing requirements. The variation, density, and yields on the 45nm process show the success of this Design for Manufacturing (DFM) methodology.

INTRODUCTION

The difficulties in continuing Moore's Law with the lack of improvement in lithography resolution are well known [1, 2, 3]. Design rules have to change and Design for Manufacturing (DFM) methodology has to continue to improve to enable Moore's Law scaling. In this paper we discuss our approach to DFM through co-optimization across design and process. We define the design rules for a new technology early in the definition process well before the technology development is complete. This early definition of design rules allows the design to start in parallel with the technology development. Early accurate modeling of the design rules and layout is a key to making this process successful. The design rules must meet the requirements of a highly manufacturable process at the beginning of the production ramp for the first product.

Our DFM goals and methodology are different from those of some other manufacturers. The basic rules for drawing transistors, other layers, and DFM requirements are not separate in our definition process. We have a few guidelines such as suggesting that designers use redundant vias where possible, but most of our DFM requirements are included as required rules that all designs must meet for all layout. Some other companies have simple basic layout rules and provide other rules that are guidelines or suggestions for changes to layout or design that would improve manufacturability or reduce variation. Designers make tradeoffs for area and cost to decide if they will implement these guidelines. Some of their DFM changes are available only after analysis of the initial layout. Another difference in our methodology is that we tend toward adding rules to prevent something in layout that might affect design, instead of depending upon modeling of product layout to find problems. We build the requirements for manufacturability and low variation into the basic design rules as hard design rule requirements. Product design starts in parallel with technology development. Design rules must not be changed significantly after design work starts. Our DFM methodology depends upon modeling to define the rules very early in technology development. Therefore, our methodology ensures that products are ready for ramp-up in multiple fabs to high-volume and high-yield manufacturing without changes. This may make our design rules more complicated than those of some of our competitors, but this methodology ensures that all our products are capable of high yield when they tape in their first stepping.

Co-optimization across process and design is required to ensure we understand and balance all requirements. By co-optimizing design and process early in the development cycle, we arrived at a set of design rules that met 45nm process and design requirements. The thoroughness of this early work resulted in these rules being stable through the development cycle, which led to the successful insertion of 45nm technology in high-volume manufacturing, ensuring the continued march of Moore's Law.

The poly layer is the most critical layer for control of variation. Due to this need to control variation, the poly layer was the first layer on which restrictive design rules were used. We need to build the requirements for minimizing variation into the rules. How the rules have changed for the poly layer shows how designs have changed. The analysis of the impact of changes in the poly rule shows how we consider design and process needs in defining design rules.

DESIGN FOR MANUFACTURING GOALS

The first requirement for a new technology is the continuation of Moore's Law. A new technology should have twice the number of transistors. This is increasingly difficult with the lack of improvement in wave length for lithography and in transistor scaling issues. Keeping manufacturing variation in check as we improve transistor density is a scaling challenge as discussed in many papers and conferences [4, 5]. Transistor channel length is a major focus for controlling manufacturing variation. The variation must scale with the poly pitch for the new process.

The second requirement of a new technology is the ability to ramp very quickly to high volume with multiple designs in multiple fabs. The design and process have to be manufacturable at the beginning of the ramp. Design rules have to be defined early in the process development work to allow product design to be done in parallel with the process development. There must be no major changes to design rules late in process development or during manufacturing ramp. Predictive modeling of the rules must be done well before process development is complete. The design rules may be conservative, to ensure that the design is very robust, but they should not be too conservative, to ensure that we derive the maximum benefit from Moore's Law scaling. The challenge is in defining this optimum robustness.

The third requirement of a new technology is for its yields to be as good as, or better than, the previous-generation technology; and for its learning curve to be as fast as, or faster than, the previous-generation technology. Judgment is required to define the technology to strike the right balance between the difficulty of the technology and the impact of the DFM requirements on the design. Some breakthroughs may be required to meet design requirements, but these cannot be so difficult that they slow the yield learning goal.

CO-OPTIMIZATON

Lithography, optical proximity correction process, and design requirements all need to be understood in defining a new technology. Co-optimization to balance requirements for all of these areas is needed to define a very manufacturable process and design. The different needs of each area must be balanced in the context of different constraints. The design groups want small die size for low cost, but they also want design work to be easy. Flexible design rules are needed to allow designers to optimize the product with minimum effort and area. Lithography engineers also want simple design rules but cannot allow unlimited flexibility. OPC engineers want layout to be very predictable so there are no hotspots that are caused by the use of unexpected combinations of rules. In the definition process, all of these different goals are considered to find an acceptable solution. The ability to consider all aspects of the problem from design to high-volume manufacturing, within an affordable envelope of multi-dimensional constraints, is the key to good DFM solutions.

Co-optimization is done from the beginning with a multi-disciplinary team that includes experts in design, lithography, OPC, and processing. We have design experts in the process development organization who closely interact with the process development engineers, and we also include representatives from the lead product designs and CAD tools' engineers in the definition process. The discussions between these experts start about four years before the technology begins production. There was a significant increase in the co-optimization work for the 45nm technology. More layout studies were done early in the definition process, and there was an increase in the modeling of design rules.

The goal of co-optimization is balancing the risk and difficulty between design and manufacturing. It may appear to designers that our solutions are not balanced, since design is more difficult on each new process. The reality is that the process and patterning choices are often limited. The lack of improvement in resolution of the lithography tools limits patterning options. DFM definition often has to make the choice that will have the least impact upon design. A solution with no impact is not possible, however. The need for low costs may result in changes in design rules for smaller die size or better yields to offset concerns about increased design effort. Judgment about manufacturability will take priority over concerns about increased work for designers.

DESIGN RULE DEFINITION

Modeling is the cornerstone of today's design rule definition process. Building predictive models for the new technology is one of the challenges for a new technology. The rule definition work is front-loaded, with the rules defined before technology development is complete. Figure 1 shows the elements of the design rule definition process.

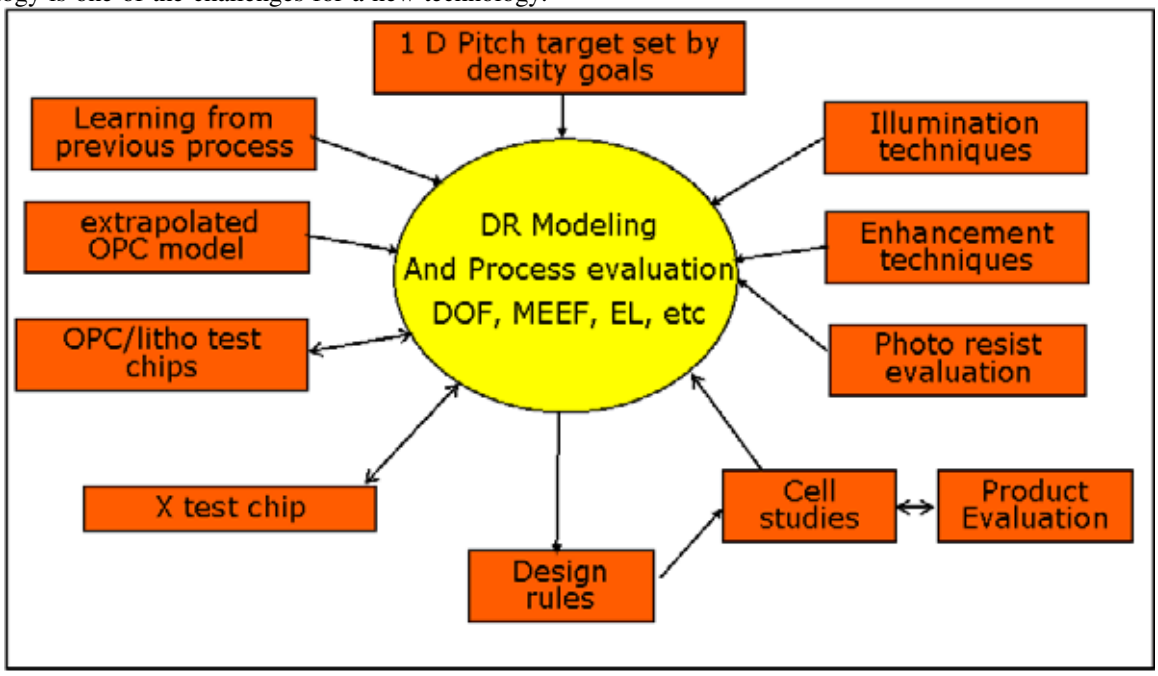


Figure 1: Design rule definition

The transistor density scaling goal drives the 1D pitch requirement for the key layers like poly and Metal-1 (M1). This 1D pitch must scale by 0.7X per process generation. 2D rules such as line end-to-end space or minimum line length can be more difficult to scale than the 1D rules. Isolated and wide lines may have scaling problems. The techniques needed to meet the 1D scaling requirements may make scaling the 2D rules more difficult. A change in illumination technology to get good minimum line width and minimum space may not allow the same scaling of rules for wider lines, or it may make scaling of rules different for X and Y directions.

Design rules are not changed after the beginning of manufacturing ramp. Learning about difficult process issues feeds forward into the design rule definition of the new process. Some design guidelines on previous technology may become hard rules on the new technology. Some structures that caused significant process problems and/or required significant process changes may be eliminated by new rules. Device or interconnect models may be simplified by the elimination of parametric variables caused by simplified design rules.

Modeling of design rules starts with extrapolation of OPC models from the previous technology. New lithography tools, illumination techniques, enhancement techniques, and resists are evaluated to determine the best method for 1D scaling, and to understand the changes needed in other rules. Learning about the capabilities of the new tools is a continuous process during rule definition. Typical and worst-case layout topologies are analyzed to evaluate process issues like Mask Error Enhancement Factor (MEEF) and depth of focus. Test reticules are created to calibrate the models.

Cell studies are done using the rules generated from the modeling studies. Data are extracted from designs on older technologies to understand the requirements for critical layouts and how rule changes might affect design. Standard logic cells, register files, SRAM bits, and metal routing are all included in the layout studies. As rules mature, product groups are included in the evaluation. Evaluation of the patterning capabilities and the impact on layout is a continuous and iterative process until rules are final. Important layout topologies identified by design are analyzed by using the models, and they are included in new test reticules. Design rules that limit meeting the transistor density goals are evaluated with the models to understand if rules need to be changed or improved.

Figure 2 shows a typical simulation from a study of a design rule. This simulation studied the process margin for line end-to-end space as a function of line length. A high MEEF is an indication of poor process control. Some line lengths have insufficient control of the end-of-line space.

This can create line-to-line shorts. One option to fix this would be to change the illumination as shown by the different lines on the graph. Another option would be to create a design rule that does not allow line lengths of 0.2 to 0.3u. If a rule change is proposed, cell layouts are done with the new rules to understand the impact on transistor design. A study may be done on data from designs on previous processes to determine if the design rule being changed is commonly used. If the rule change is shown to cause a significant change in area, we would consider other pattering solutions.

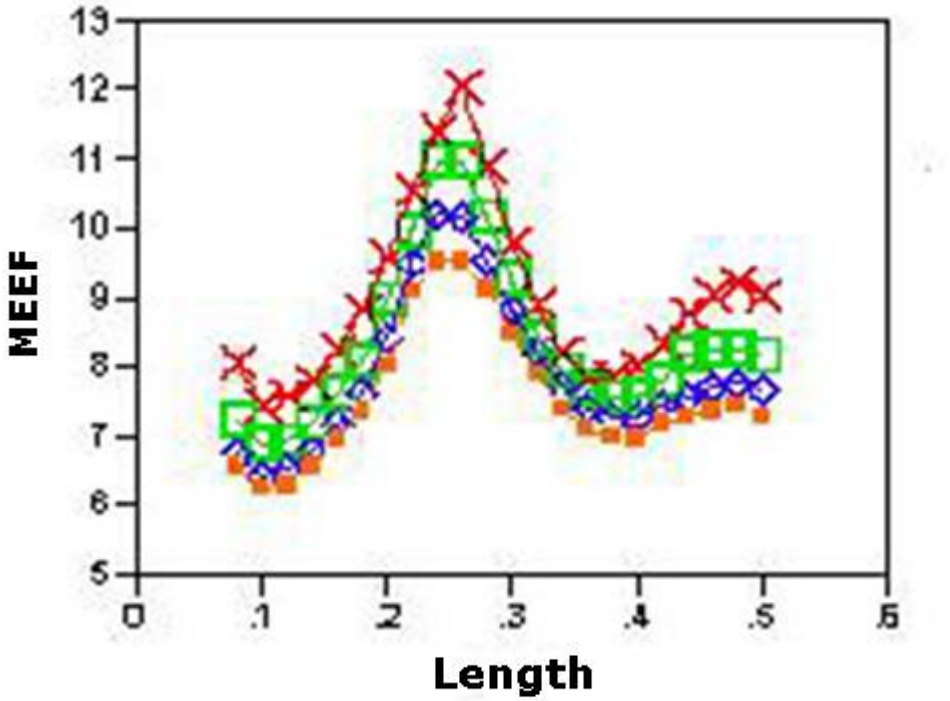


Figure 2: Length vs. MEEF

The main process development vehicle is the X-chip test vehicle that includes large SRAM designs, process and design test structures, and process-sensitive circuits. There are earlier mask sets that include some of the rules and features of the new process, but the X-chip is the first mask set that has large circuit blocks with all of the rules. The processing of the test chip is used to validate the rules, not define the rules. There may be a few changes in rules depending on what we learn from the test chip. The number of rules that are added or changed have to be limited, because product designs will have started before the test chip is processed. As the development process continues, the justification for design rule changes becomes more difficult. By the time production starts, a design rule problem has to be fixed by process unless it is impossible to fix that way.

Figure 3 shows our trend in the number of rules over time for the 45nm process. The timeline is relative to tape-out of first design on the technology. In addition to added new rules, there were some minor changes for better and worse rule values. The number of rules increased during the design and layout of the test chip as the modeling work continued and OPC flows were developed. There was a small increase in the number of rules after processing of the test chip. These changes were not all due to patterning issues. New understanding of the transistor of metal processing can create the need for new rules. By the time masks were created for the first product, the rules were stable. There were only a few changes in design rules after the first design started. The 45nm rule stability was good, but there were more changes than desired. On the 32nm process we had fewer changes during the test chip design and after the start of the first product. Process development continued during debug of the first design, but the rules did not change during this time. Process improvement continues through the life of the process to reduce defect density, and cost, but this is done without changing design rules.

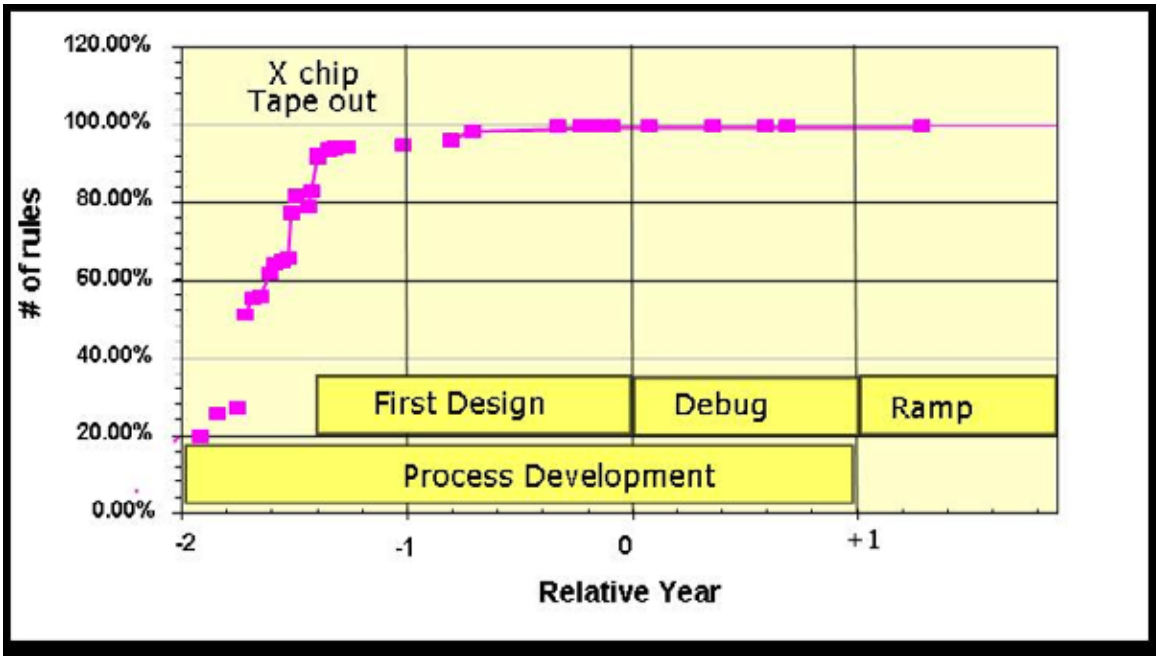


Figure 3: Changes in number of poly rules

EVOLUTION OF POLY RULES

The poly layer is one of the most difficult layers to pattern and process. Control of poly CD is one the most critical requirements in the process, due to its affect on transistor performance and variation. Poly CD control must scale for the new technology to keep the percentage variation of the channel length constant. Contacted gate pitch is a big factor in SRAM cell area and logic transistor density. These critical requirements make poly the first layer to need new pattering solutions and design rules. In the following sections, we show how poly design rules have changed over the last few processes and how DFM methodology has evolved and has been used in the definition work.

130nm Process

Our 130nm process had simple rules. There was limited early modeling of layout. Layout had random combinations of poly widths, spaces, and device orientation. Products had some issues with poly corner rounding that affected small devices. This issue was helped by DFM guidelines for layout of small devices, but the guidelines came after initial design of lead products. There was limited involvement from product design engineers in the early design-rule definition. Rule definition was primarily simple scaling of rules from previous technology generations.

90nm Process

The 90nm process included more restrictions for poly layout, and the number of poly rules increased by 47%. All devices had to be in the same orientation except for the memory bits. The difference in printing of shapes parallel and orthogonal to the scan direction was one reason for this change. Poly-over-field-routing was allowed in the X or Y orientation. The memory bit shown in Figure 4 is a unique topology with transistors in both orientations that could be modeled and characterized to account for any difference due to device orientation. Poly corner rounding and modeling were analyzed and modeled early in the definition process.

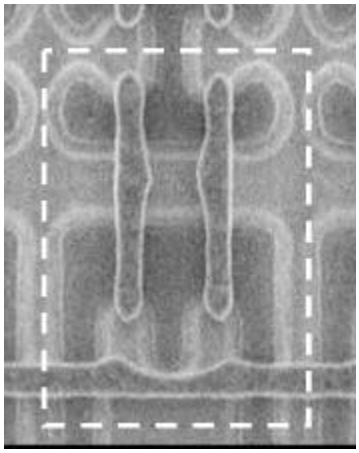


Figure 4: 90nm SRAM bit layout

The main design impact of the one-device orientation is that a block of layout cannot be placed in two orientations, rotated 90°. I/O buffers on the top and right edges for the die must have unique layouts, even if circuits are identical. This added some layout effort, but had no affect on die size. Since there was no die size impact, the rule change was a better solution for random layout than trying to model differences due to device orientation, as was done for the memory bit.

65nm Process

The number of poly rules increased by 65% for the 65nm process, and rules were more complicated. Rules changed to allow the use of phase shift masks. All devices including the memory bit had devices in one orientation as shown in Figure 5. This layout is almost ideal for patterning with simple rectangular shapes on all layers.

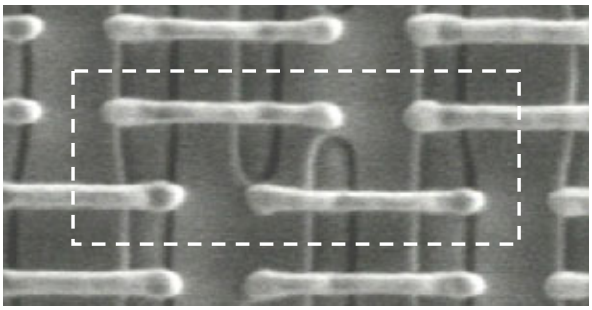


Figure 5: 65nm SRAM

There were differences in poly layout rules based upon pitch and poly space and orientation. The complex rules had little effect on transistor density. Many of the new rules affected special cases that did not occur often or were easy to fix. For example, layout of a minimum width device has some new rules for end-cap, but design did not use a lot of minimum width devices. Where they were used, the end cap rules were usually easy to meet due to area being limited by other rules.

Cell layout studies and modeling of layout increased for this generation for worst-case structures. The layout was random and it was difficult to determine all worst-case layouts.

45nm Process

Meeting the transistor density and process requirements of the 45nm technology required significant changes in design rules. The high-K metal gate transistors on this technology are the biggest change in transistors in 40 years [6]. 193nm patterning tools were needed to minimize cost and risk. Poly had to be printed using 193nm dry tools and still had to meet our need for 0.7X scaling of the pitch. Variation could not increase. The early modeling work increased significantly from the 65nm process. This included earlier involvement of the OPC experts in the design-rule definition and evaluation phases. Our goal was a more comprehensive evaluation of rules and layout topologies through modeling.

There were some changes in poly layout on the 90nm and 65nm process, but the layout has remained very random. Design could have used any poly pitch >= minimum pitch, and any channel length >=minimum was allowed. The different channel lengths could be randomly mixed. Transistors were in one orientation, but poly routing could be in either orientation. Corner rounding of poly close to devices could impact transistor performance.

Early in the definition work we asked if poly patterning in logic could be similar to the poly patterning in the SRAM bit introduced in the 65nm generation. One of the big concerns for making poly layout more regular was limiting the channel layout choices available to design. Design had always had few restrictions on the channel lengths. We wondered if this freedom was necessary. Figure 6 shows the channel lengths used in one 65nm design. Most of the devices at 0.10u and 0.11u are in the SRAM bits. The few devices at longer channel lengths were primarily in analog and I/O circuits. 99% of the devices in random logic had minimum channel length or minimum + .01u. The main reason for using longer channel length in logic was to reduce device leakage. Channel lengths can be limited in logic as long as there are options for low leakage devices. Device leakage is strongly dependent upon Le, so a very small change in Le can reduce leakage by 3X. Higher voltage can also be used to reduce leakage. There were some circuits where a longer Le device had to be replaced by two or more minimum channel length devices. Based on analysis of data like this and other layout studies, the design was partitioned into three groups: logic, analog, and SRAM. Analog and SRAM are treated as special cases.

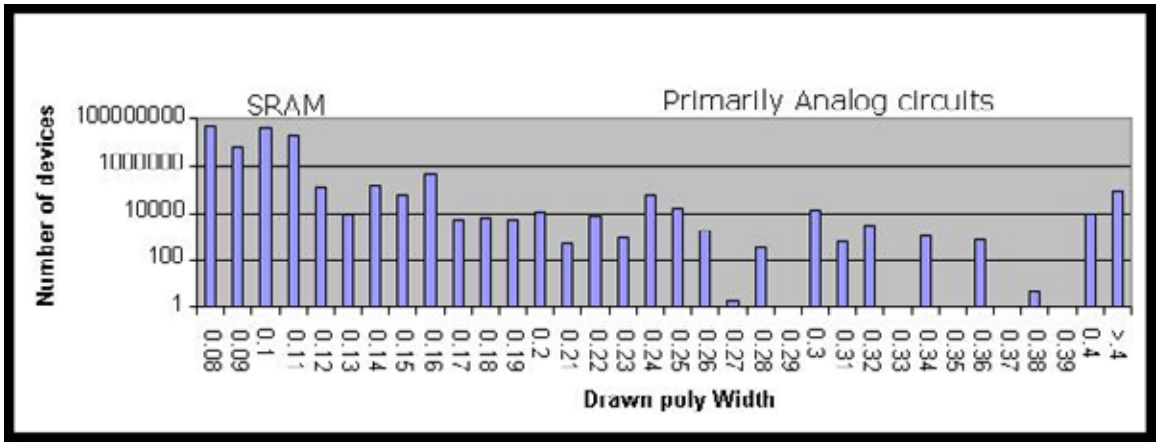


Figure 6: Product channel lengths on 65nm design

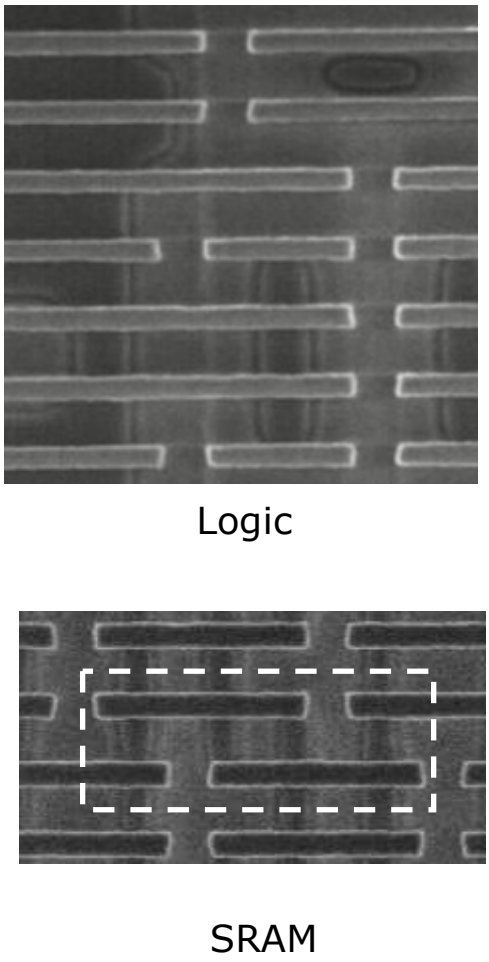


Figure 7: 45nm SRAM and logic poly layout

The 45nm process has trench contacted-based local routing. This eliminated the need to use poly routing orthogonal to the transistor gate and the wide poly used for poly contacts. Layout studies were done to understand if one poly pitch was possible and how this affected rules for other layers. The local routing and limitations on transistor channel length had allowed logic poly layout to be one pitch and one direction as shown in Figure 7. The number of layout rules for logic layout was reduced by 37%. This reduction is not large, as the simple layout might indicate, because the poly rules include rules for poly spacing to other layers, and there are several rules for end-caps and poly end-to-end.

OTHER LAYERS

The regular layout for poly also simplified patterning of contacts and M1. The contact pitch is the same as the gate pitch. M1 parallel to the gates has the same minimum pitch as poly.

Very restrictive rules similar to poly rules were not needed to meet process or patterning requirements for other layers, but the rules for some other layers are complex. The metal pitch choices do not need to be as restrictive as the poly layer pitches, since metal variation requirement is not as tight. The metal variation has less impact upon path delay than the poly CD variation. Metal layer design rules have some types of design rule restrictions used on poly layers on previous technologies. Printing of isolated metal lines was one of the issues on 45nm technology. Rules were added to restrict the use of isolated lines. Product layout uses several metal widths and a range of spaces. It is difficult to limit the width and pitch choices for metal, due to the need for wide lines and spaces to optimize RC delays, capacitance, and power delivery. CAD tools must change to support the changes in metal rules. One of the learnings from the 45nm technology was that the CAD tool work needs to start earlier.

45NM MANUFACTURABILITY

Standard Cell Area Scaling

The first goal for any new process is to maintain the transistor density scaling trend. Figure 8 shows the standard cell area scaling trend. This analysis used a large standard cell library from our microprocessor designs and is weighted by typical cell usage. Transistor density scaling has followed the 0.5X density improvement per process generation. The 45nm process meets this trend despite the change to more complicated and restrictive design rules.

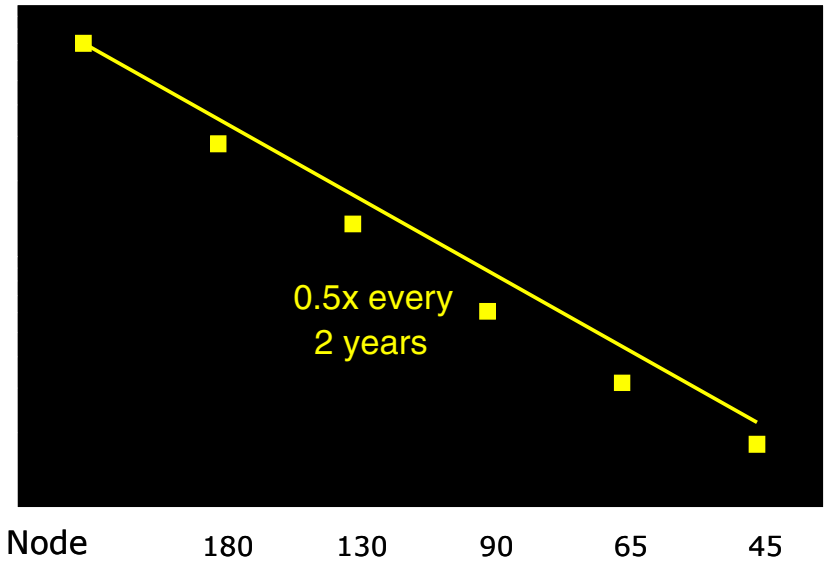


Figure 8: Standard cell density

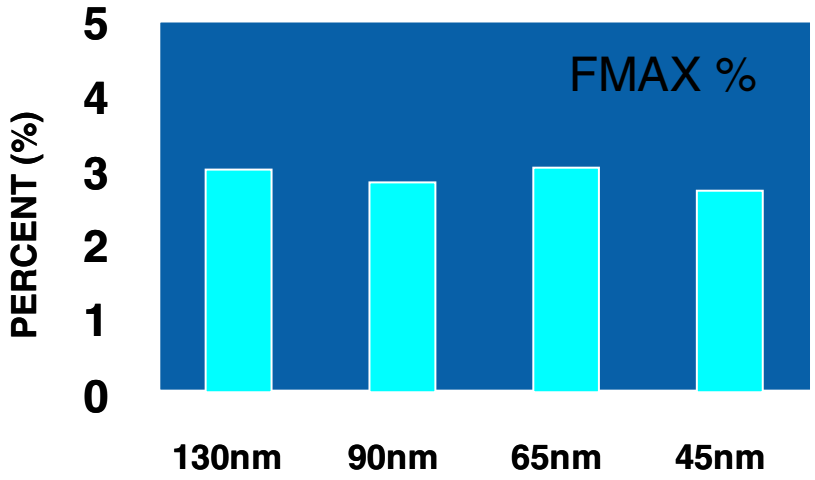


Figure 9: Within wafer variation of oscillator frequency for the 130nm through 45nm technology generations

Variation

One of the requirements for poly patterning is scaling of variation. Figure 9 shows the within wafer variation of oscillator frequency for the 130nm through 45nm generations. The variation in frequency has remained at less than 3%. This frequency variation includes the affects of poly CD, VT, and other sources of device variation. Variation has scaled with the decreasing poly pitch, despite the changes to new transistors, changes in poly patterning, and device sizes.

Yield Learning

The final measure of success of DFM is process yield. Our yield learning trend history is shown in Figure 10. The yield learning rate on 45nm technology is as fast as previous processes and is trending toward lower defect density than the 65nm process. There is no way to measure how the DFM rules contributed to this learning. Design rules, process definition, process tools, and many other things affect the yield learning rate and final high-volume manufacturing yields. All of these things are necessary to meet the manufacturing goals.

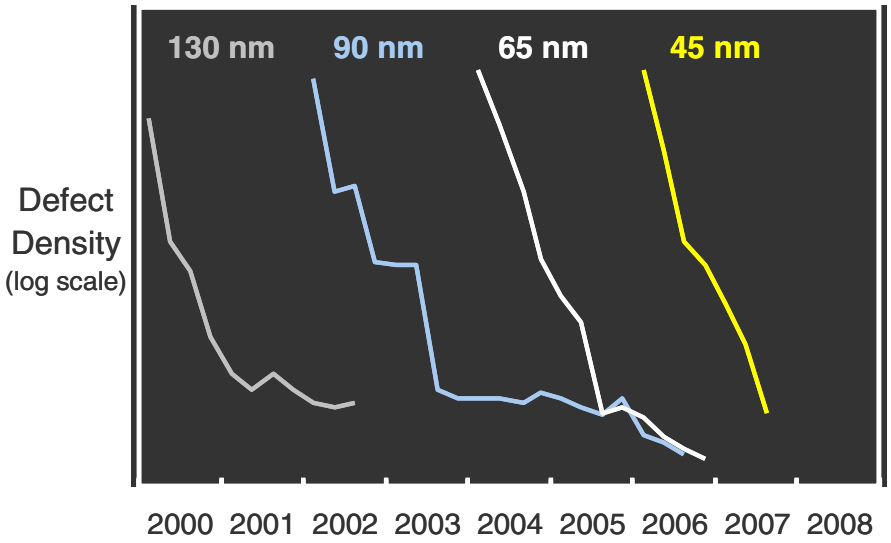


Figure 10: Yield trend

SUMMARY

Our 45nm technology maintains Moore's Law on a technology with a new transistor technology and dry 193nm lithography. Transistor density is on the 2x trend line and variation scaled. Poly rules were changed to allow simple one-pitch poly patterning with no impact to transistor density or product performance. Extensive modeling of the rules provided design rules that were stable before first product tape-out and robust enough to allow a fast high-volume technology ramp.

REFERENCES

[1] G.E. Moore, "Lithography and the Future of Moore's Law." In *Proceedings SPIE, vol. 2440*, 1995, pp. 2–17.

[2] Borodovsky, Yan, "Marching to the Beat of Moore's Law." In *Proceedings SPIE*, *vol. 6153, Advances in Resist Technology and Processing XXIII,* March 2006.

[3] Alfred K. Wong, "Microlithography: Trends, Challenges, Solutions, and their Impact on Design." *IEEE Micro*, *vol. 23, no. 2*, March/April 2003, pp. 12–21.

[4] Shekhar Borkar, "Parameter Variations and Impact on Circuits and Microarchitecture." *40th Design Automation Conference, 338,* (*DAC'03*), 2003.

[5] Kuhn, Kelin J., "Reducing Variation in Advanced Logic Technologies: Approaches to Process and Design for Manufacturability of Nanoscale Cmos." *IEDM* 2007, pp. 471–474.

[6] Mistry, Kaizad A, "45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging." *IEDM* 2007, pp. 247–250.

ACKNOWLEDGMENTS

Barbara Chappell, Sam Sivakumar, Charles Wallace, Mark Bork, and Robert Bigwood.

AUTHOR'S BIOGRAPHY

Clair Webb is an Intel Fellow. He has been with Intel for 31 years. He has worked on DRAM, SRAM, and cache designs. For the last few process generations he has worked on design rule definition and design issues for

new process technologies. His e-mail is clair.webb at intel.com.

45nm Transistor Reliability

Jeffrey Hicks, Logic Technology Development Quality and Reliability, Intel Corporation
Daniel Bergstrom, Logic Technology Development, Intel Corporation
Mike Hattendorf, Logic Technology Development, Intel Corporation
Jason Jopling, Logic Technology Development Quality and Reliability, Intel Corporation
Jose Maiz, Logic Technology Development Quality and Reliability, Intel Corporation
Sangwoo Pae, Logic Technology Development Quality and Reliability, Intel Corporation
Chetan Prasad, Logic Technology Development Quality and Reliability, Intel Corporation
Jami Wiedemer, Logic Technology Development, Intel Corporation

Index words: 45nm, High-K + Metal Gate, Bias Temperature Instability, Time Dependent Dielectric Breakdown (TDDB)

ABSTRACT

It has been clear for a number of years that increasing transistor gate leakage with device scaling would ultimately necessitate an alternative to traditional SiON dielectrics with polysilicon gates. Material systems providing higher dielectric constants, and therefore allowing physically thicker dielectrics, have been the object of extensive research. Such high-k dielectrics, when combined with metal gate electrodes, have emerged as the leading alternative, demonstrating good transistor performance and offering reductions in gate leakage of 25X-100X. Achieving the required reliability, particularly at the high operating electric fields at which the performance advantages are realized, however, proved much more difficult.

Intel strove to overcome the reliability obstacles by introducing high-k dielectrics combined with metal gate electrodes (HK+MG) transistors in its 45nm logic process, as it judged the transition to this technology would provide compelling performance advantages. In this paper we discuss the general considerations for the reliability of HK+MG transistors and specifically we discuss what was achieved with Intel's 45nm process technology.

A particularly extensive effort was undertaken to characterize the reliability physics of this revolutionary new transistor and to gather the data to ensure accurate modeling of failure rates. This entailed accelerated testing and fully integrated test vehicles, representing over seven orders of magnitude in the transistor area, at a variety of stress conditions, some of which lasted over three months.

The intrinsic transistor reliability fail modes addressed in this paper fall into two basic classes. First we have the integrity of the transistor dielectric itself, which in the course of operation, can fail, a phenomenon typically referred to as Time Dependent Dielectric Breakdown (TDDB). The transistor must be engineered to ensure that components don't wear out within their operating lifetimes.

Second, in addition to abrupt failure of the dielectric, transistors can also experience progressive parametric degradation. The primary parametric reliability mode for traditional SiON-based transistors is a slowdown of the PMOS devices due to progressive trapping of charge, typically referred to as Bias Temp Instability (BTI). For HK-based dielectrics, at their higher target operating fields, similar degradation is observed on NMOS transistors as well as potentially significant increases in gate leakage with stress, known as Stress Induced Leakage Current (SILC).

We discuss these reliability phenomena and illustrate that while they pose large reliability challenges for HK+MG, these challenges can be overcome through refinement of process architecture and optimization of processing conditions. Intel's 45nm technology is shown to achieve intrinsic TDDB and aggregate (N+P) BTI performance equivalent to its 65nm predecessor with negligible SILC at its 30% higher operating electric fields.

PROCESS BACKGROUND

The 45nm high-k dielectrics combined with metal gate electrodes (HK+MG) transistors studied in this work have a Hafnium-based gate dielectric and dual workfunction

metal gate electrodes for NMOS and PMOS. The transistor fabrication utilizes a HK first and MG last process as detailed in [1]. In this flow, HK is deposited using an Atomic Layer Deposition (ALD) process, and polysilicon is used for the gate patterning. After the Interconnect Dielectric deposition, a polish step exposes poly gates, and the dummy poly is removed. Then, workfunction metal electrodes are deposited followed by a gate fill process. The SiO_2 equivalent oxide thickness (EOT) of the HK plus the Interface Layer (IL) that forms between the HK and the silicon is ~1.0nm. Figure 1 describes the gate stack, with its SiO_2-like interface layer (IL) and the HK dielectric proper.

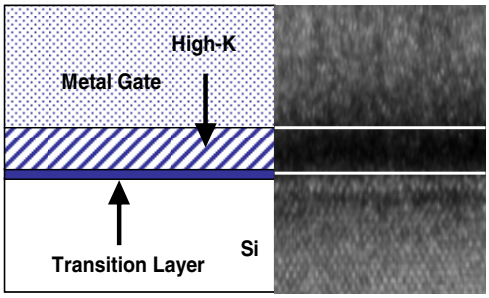
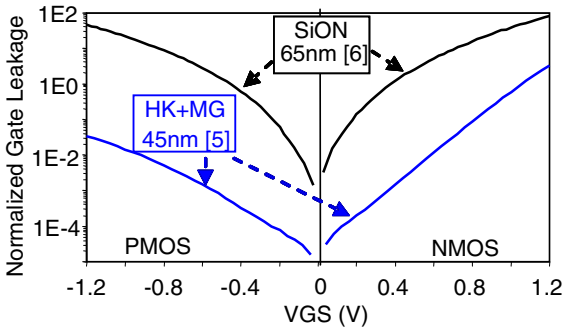


Figure 1: TEM of High-K and Metal gate (HK+MG) transistor stack

Figure 2(a) shows the typical gate leakage behavior of these 45nm HK+MG transistors measured in inversion, compared to that of 65nm Poly/SiON transistors [2]. Substantial reduction (25-1000X) is achieved with the physically thicker HK film employed while still enabling inversion Tox scaling consistent with the historical trend of ~0.7X per technology generation. Figure 2(b) shows the appropriate band diagrams for these HK+MG transistors. Lower inversion gate leakage current on the HK+MG PMOS is due to the larger band offset than for the NMOS (as indicated by the direction of arrows of the tunneling current).

(a)



(b)

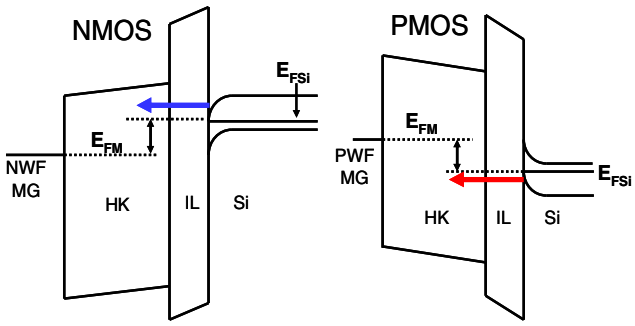


Figure 2: (a) 45nm HK+MG leakage comparison to that of 65nm Poly/SiON transistors. HK+MG process enables 25-1000X gate leakage reduction. (b) Band diagrams of HK+MG transistors.

Despite the reduced gate leakage current due to use of physically thicker HK dielectric, the effective E-field in the operating regime increases substantially (~30%). Figure 3 shows the gate dielectric E-field vs. technology node. The increase in the E-field provides higher transistor performance, but this must be achieved with no degradation in reliability.

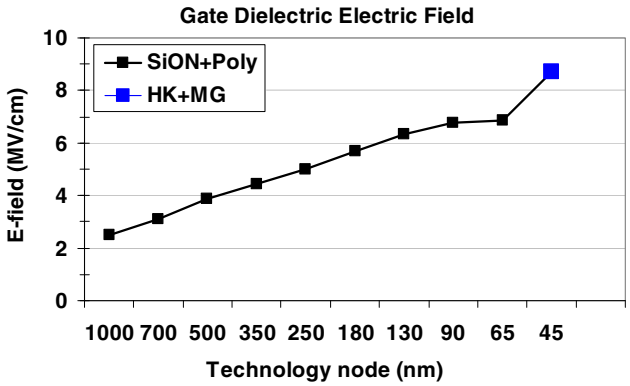


Figure 3: Gate Electric field increase vs. technology nodes. The average E-field for the 45nm technology is 30% higher than in the 65nm generation.

The large increase in E-field with HK+MG contributes to the reliability challenges of this major technology transition. In the next section we provide background on the key reliability mechanisms that need to be considered in this transition.

TRANSISTOR RELIABILITY FUNDAMENTALS

Transistor Bias Temperature Instability (BTI) Degradation

When subject to operating bias, transistors exhibit changes in transistor characteristics over time, an effect termed Bias Temperature Instability or BTI. Typically, transistor thresholds $|(V_T)|$ increase, and other electrical parameters, such as drive current (I_D) and transconductance (G_m), are also affected. At typical operating fields of SiO_2 transistors, BTI is only significant for PMOS transistors with negative gate bias (NBTI).

BTI results from the creation of both interface states (D_{it}) and oxide trapped charges (D_{ot}), and the mechanism is accelerated by both voltage and temperature. As the name implies, the PMOS channel must be inverted for NBTI to occur. NBTI degradation does not require a large amount of tunneling current and can be significant even at a very low bias. Characterization of true NBTI degradation is very challenging, due to the recovery of trapping that occurs in stress upon removal of the bias. Several fast measurement techniques have been developed to minimize the recovery influence including On-The-Fly (OTF) measurements, Ultra-fast V_T measurements on the order of micro-seconds, and Pulse-IV measurements [3, 4].

Figure 4 shows a diagram of a typical PMOS NBTI degradation and recovery process well reported in the literature. There is no consensus on the exact physical mechanism, but one of the leading models for PMOS NBTI recovery is the back diffusion of Hydrogen near the substrate/dielectric interface [5].

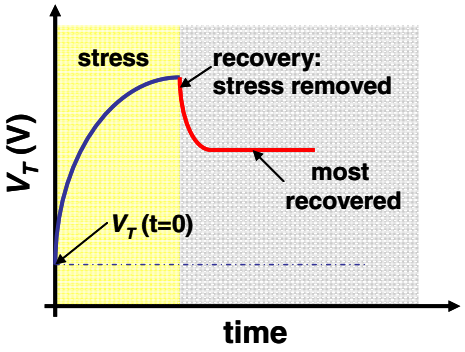


Figure 4: PMOS NBTI vs. Time illustrating both degradation and recovery

PMOS NBTI is recognized in the industry as a major reliability mechanism in advanced logic technologies. Degradation of maximum operating Frequency (Fmax) and circuit margin, in particular at Minimum Operating Voltage (Vmin), must be addressed within product design and testing to ensure an adequate margin to specifications over operating lifetimes.

A particularly important circuit case is SRAM memory. Transistors within the SRAM cells are typically amongst the smallest within a technology, and the SRAM Static Noise Margin (SNM) is highly sensitive to device mismatch. The scaling of SRAM memory arrays has increased the sensitivity to NBTI-induced transistor V_T mismatch, which can degrade Vmin characteristics over time. 6T SRAM cell area has traditionally reduced 2X every two years, as shown in Figure 5, which means bit counts are also increasing at a corresponding rate. In addition to design and layout approaches to improve Vmin margin, error correction techniques are often leveraged in cache designs [6].

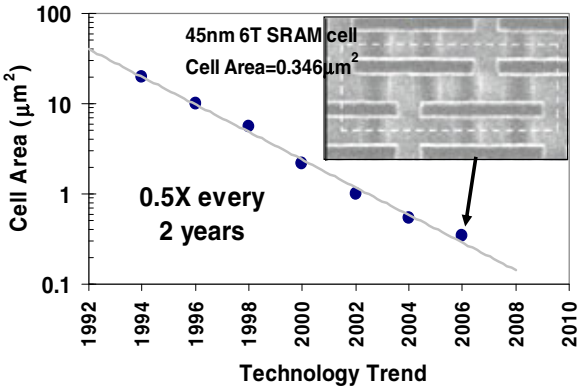


Figure 5: 6T SRAM cell size scaling trend showing 2X cell area scaling every two years [7]

Figure 6 shows an example of Vmin dependence on the SRAM cache array size. Both the magnitude of Vmin and the Vmin spread increase with cache array size due to transistor variations. Thus, understanding device variability at both time 0 and over time, given BTI effects, has become increasingly important with cache cell/array size scaling.

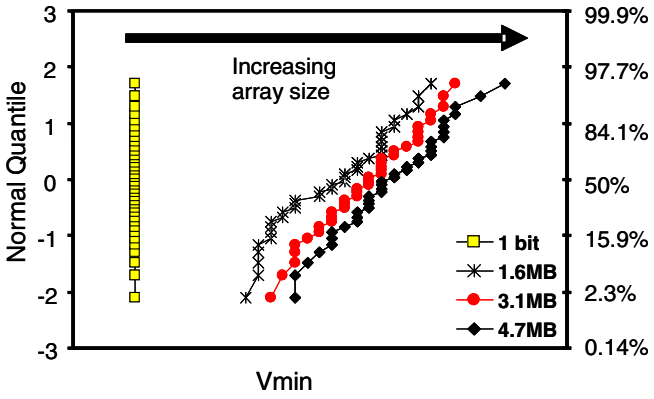


Figure 6: Vmin dependence on cache array size for 6-T SRAM [7]. Transistor aging due to PMOS NBTI will further degrade Vmin characteristics.

PMOS NBTI remains a key concern of HK+MG CMOS transistors, and at the higher operating fields of these transistors compared to SiO_2, NMOS BTI degradation under inversion (PBTI) can also be significant [8]. Unacceptably high levels of BTI degradation for HK+MG on both NMOS and PMOS have been reported. In addition, fast charge trapping/de-trapping has also been reported on HK materials, which complicates characterization and calls the validity of conventional DC stress into question. BTI reliability degradation has been shown to be modulated by processing and integration changes such as thermal treatments, adding dopants, and nitridation techniques [9, 10]. Better charge trapping properties and improved reliability results on transistors fabricated with HK silicate dielectrics rather than HK oxides have also been reported [11].

Extensive BTI experimental data collection was undertaken in the development of Intel's 45nm HK+MG transistor technology to support reliability modeling and process optimization work. A summary of these results and a discussion of the mechanisms responsible for BTI in HK+MG transistors are presented below. We demonstrate that, with appropriate transistor architecture and processing, net BTI degradation that is comparable to, or better than, that observed with traditional SiON dielectrics, can be achieved for HK+MG dielectrics operating at ~30% higher E-fields. The optimized HK film stack used in Intel's 45nm HK+MG process also shows negligible hysteresis and transient trapping associated with fast carriers.

Gate Dielectric Breakdown, Time-Dependent Dielectric Breakdown (TDDB) and Stress-Induced Leakage Current (SILC)

The transistor gate dielectric provides isolation of the gate electrode from the conducting channel, providing the high input impedance of CMOS transistors. The reliability of the gate is, therefore, of primary importance in transistor reliability. Multiple evaluation techniques exist for assessment of gate dielectric integrity, with Time Dependent Dielectric Breakdown (TDDB) testing being the standard methodology for developing operating lifetime reliability projections. TDDB characterization is performed with elevated voltage and temperature, with either constant voltage (CVS) or constant current (CCS) on transistors or capacitors, until a failure is observed. Failure is typically based on an increase in gate current Ig, but definitions vary and can significantly impact projections.

TDDB can occur on NMOS and PMOS under all operating bias conditions (inversion, accumulation); however, the rate of dielectric damage is very strongly modulated by the band structure of the material system and, traditionally, NMOS in inversion mode tends to be the limiter for TDDB lifetime.

Although there is no rigid consensus in the literature on the exact physical mechanisms that dominate gate dielectric breakdown, it is generally attributed to a combination of several mechanisms—charge injection, interface, bulk trap state generation, and trap-assisted conduction. During operation, the electric field across the gate dielectric causes the generation of electrical defects or "traps." These traps modify the local electric field and enhance leakage current in the dielectric through various hopping and tunneling processes. With cumulative stress, more trap states are created and, consequently, a gradual increase of the gate current is observed: this is known as Stress Induced Leakage Current (SILC) degradation. Eventually, a point is reached where a conductive "chain" of traps is established between the cathode and the anode as depicted in Figure 7. The statistical theory that describes this process is called the Percolation Theory [12]. The completion of this chain results in a large increase in current flow and potentially collateral damage to the device, which may critically impact the circuit.

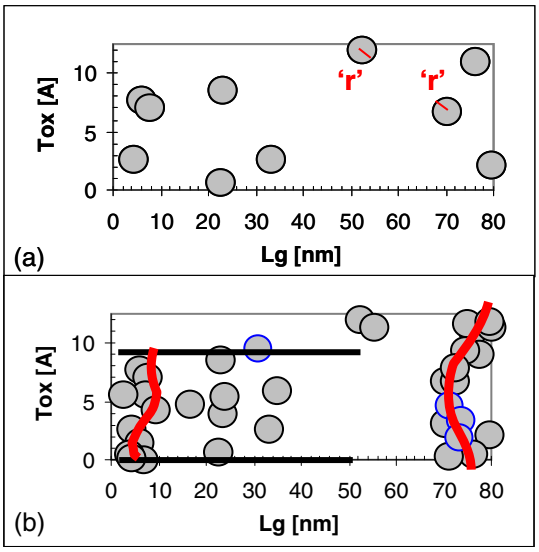


Figure 7: Percolation Theory describes traps as spheres of radius "r. When several of them form a complete chain from anode to cathode, breakdown (BD) occurs. The thinner the dielectric, the fewer the traps needed to cause BD [12].

The use of HK+MG stacks to overcome scaling limitations of conventional SiO_2 dielectrics introduces additional complexities in the form of materials, band structure, and interfaces that can significantly impact the TDDB mechanics and performance. Problematic HK dielectric lifetimes, SILC degradation, and interface and bulk-trap densities have been reported extensively in the literature [13, 14], and these issues need to be overcome to match the high reliability standards that have been established with conventional dielectrics.

Unless specified otherwise, TDDB stresses reported in this work were carried out with DC CVS. Monitoring of the gate leakage was performed by interrupting the stress with negligible measurement delay between stress and measurement. Care was taken to ensure that the measurement phase did not result in additional trap creation or degradation. The monitor measurements were conducted at two bias conditions corresponding to nominal and low voltage of operating conditions of products. The results focus on reliability of the optimized process flow, referred to as *Final,* but the affects of process optimizations are illustrated with results of material from early unoptimized process architectures and flows, referred to as *Initial.* Results for Intel's 65nm process are also referenced as benchmarks for mature ultra-thin SiON+PolySi devices [2].

The devices evaluated in this work are single transistors as well as arrays of transistors tied together electrically in parallel to generate large gate area structures with realistic transistor-like layouts. Each leg of the transistor arrays has a drawn gate length of 40nm while the electrical length is much smaller. The SRAM cache data reported here were collected on a fully integrated 4.5Mbit cache array. Acceleration factors were extracted through such testing to understand the sensitivity of the TDDB lifetimes to voltage and temperature. To minimize the extrapolation uncertainties in TDDB models, large sample sizes were accumulated at multiple stress condition combinations on test structures with a gate area range of over seven decades.

It will be demonstrated that, with an optimized transistor architecture and process flow, dielectric reliability comparable to that obtained on traditional SiON dielectrics can be achieved for HK+MG dielectrics operating at ~30% higher E-fields with negligible SILC prior to breakdown.

In Process Charging

It is well understood in the industry that dielectric quality as well as transistor parametric characteristics can be degraded due to process-charging induced damage from plasma processes within the fabrication flow. The charge that may accumulate on interconnect 'antennae' connected to transistor gates in the course of these processes can result in sufficiently high stress to induce unrecoverable changes to the transistor characteristics, or in extreme cases, even catastrophic device damage. The standard approach to protect against such plasma-induced damage is to provide a discharge path in the form of diodes or transistors. The protection needs are a function of the specific antennae connected to a device as well as the intrinsic leakage of the transistor and the charging characteristics of the fabrication processes. Design rules are defined to ensure sufficient protection to prevent any transistor damage during processing.

Process charging is one concern that has benefited from traditional dielectric scaling; increases in gate oxide leakage have made ultra thin SiO_2 dielectrics less susceptible to damage. With the large reduction in gate leakage that HK dielectrics provide, the charging rules must therefore be tightened to more historical levels.

RESULTS

Introduction to Reliability Results

The results of reliability characterizations of Intel's 45nm HK+MG process are presented below. The first two sections address degradation of the gate dielectric considering progressive increases in leakage current or SILC and dielectric breakdown. The third section considers degradation in transistor operating characteristics due to charge trapping within the bulk of

the dielectric or at interfaces within the stack. The results reported here are for devices with the process minimum 'drawn' channel lengths of ~40nm, but devices spanning the permitted layout range were evaluated with similar results. Unless stated otherwise, the data below are for devices on the optimized 45nm HK+MG process flow. Some results from 45nm development process flows as well as for a mature 65nm process are shown for comparative purposes.

Dielectric Reliability

Transistor dielectric reliability was assessed for a wide range of transistor structures ranging from single cache bitcell transistors to 4.5Mb SRAM cache test vehicles. Conventional CVS was employed unless otherwise indicated.

Dielectric Reliability—TDDB

TDDB is associated with a substantial increase in current through the transistor dielectric. In the data below, the definition of failure for extraction of TDDB lifetimes is the point where an abrupt increase in dielectric current is observed or a "hard breakdown" (HBD) occurs. This increase in current due to HBD may be sufficient to, but won't necessarily, result in circuit failure.

Figure 8 compares matched Electric field TDDB on SiON+PolySi vs. the optimized Intel 45nm HK+MG gate [15].

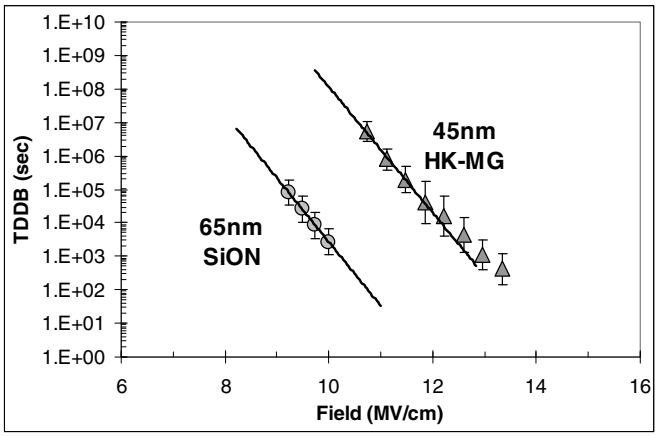


Figure 8: TDDB vs. Electric field comparison of HK+MG [1] and SiON [2] showing that an optimized HK+MG dielectric can support 30% higher field than a mature SiON-based technology at matched TDDB lifetime

Figure 9 shows the median TDDB lifetime data collected on the *Final* optimized process over a much greater range: more than 6 MV/cm and more than seven orders of magnitude in time. Note, the data exhibit a clear transition in the acceleration behavior that occurs at ~12MV/cm.

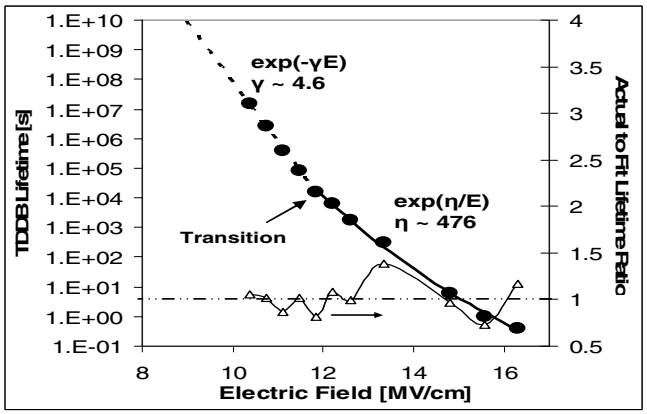


Figure 9: Long-term TDDB measured on the *Final* process flow shows a change in the acceleration slope at ~12MV/cm. Empirical fits to the data show good fitting with fit delta of 1±0.3.

Figure 10 summarizes the scaling of TDDB lifetime with gate area over a range of nearly eight decades. Using a conventional Weibull model, the area-scaling data corresponds to a Beta value of approximately 1.4.

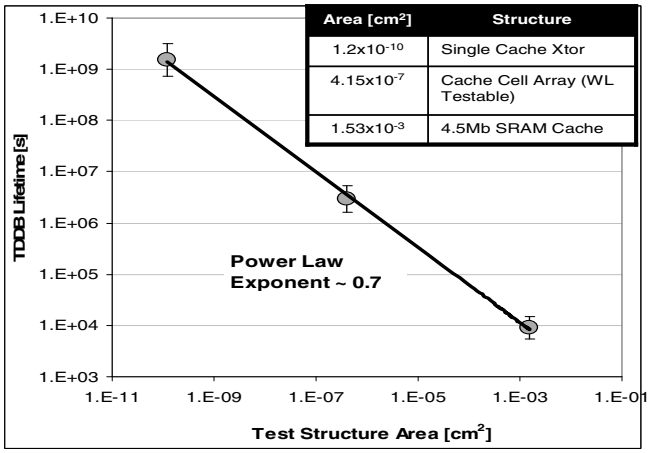


Figure 10: Area scaling of TDDB lifetime measured on single cache bitcell transistors, array cells and 4.5Mb SRAM cache test vehicles demonstrating consistent TDDB scaling with gate area over a range of eight decades

Dielectric Reliability—Stress Induced Leakage (SILC)

Transistor dielectrics can exhibit intrinsic increases in gate leakage prior to dielectric breakdown that may be large enough in aggregate to noticeably degrade the static power of an IC even where there is no impact on circuit functionality or performance. As discussed previously, this SILC effect has been reported to be a major concern for HK dielectrics.

Figure 11 compares early time evolution of Ig under inversion stress of NMOS devices fabricated with an early, unoptimized 45nm HK+MG process flow to that of material from the *Final* optimized process flow and for a conventional SiON stack.

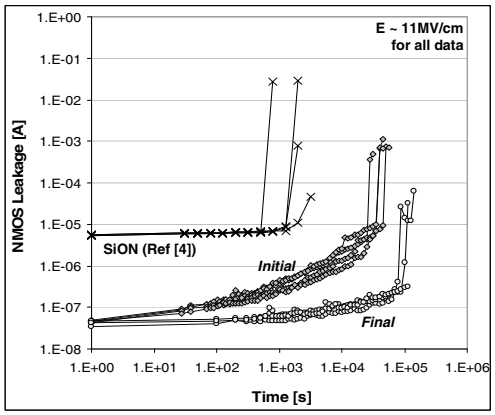


Figure 11. SILC degradation on Initial vs. Final process flows demonstrates improvement through process optimization. The SiO_2 case is shown for reference and exhibits relatively negligible SILC degradation as expected. The Final process exhibits similar levels of SILC degradation as observed on the SiO_2 reference.

Transistor Degradation—Bias Temp Instability (BTI)

Bias Temp Instability (BTI) degradation of 45nm HK+MG transistors was studied over a range of bias and temperature conditions to allow models to be generated and to provide insight into the physical mechanisms and processing interactions.

Unless stated otherwise, the data below are for devices stressed statically (DC) and with fixed delay between completion of stress and device characterization for each stress interval to ensure a consistent level of any recovery due to charge detrapping. During stress, the transistor gates were biased at either positive or negative polarity while all other terminals were grounded. V_T was measured using 50mV on the drain.

The time evolution of degradation in transistor drive (Idsat) for 45nm HK+MG NMOS and PMOS devices at accelerated inversion stress conditions is shown in Figure 12. Both show power law time dependencies: $\Delta Idsat \sim t^n$, with a somewhat lower time slope of .17 observed on the NMOS compared to ~.2 for the PMOS [16].

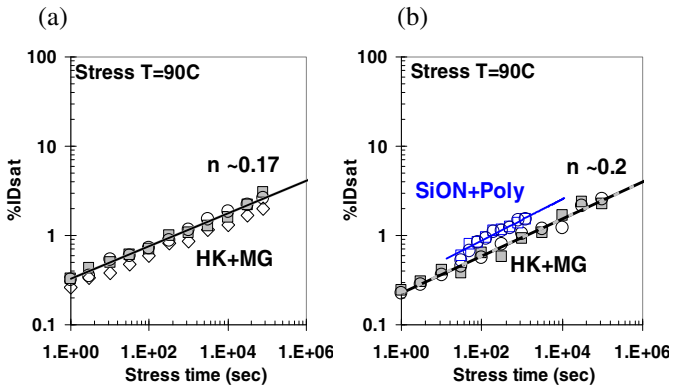


Figure 12: (a) NMOS PBTI time dependence. (b) PMOS NBTI time dependence of HK+MG transistors. Transistors are drawn at W/L=0.9um/0.04um for HK+MG process. Poly/SiON have W/L=1um/0.04um drawn dimensions.

Characterization of transistor degradation under DC stresses over a range of temperatures from 60°C to 110°C shows Arrhenius time dependence as shown in Figure 13:

$$\text{Time to given TTF}(\Delta V_T) \sim e^{Ea/kT}$$

with Ea ~ .7eV for PMOS; NMOS shows a lower temperature dependence of ~.46 eV.

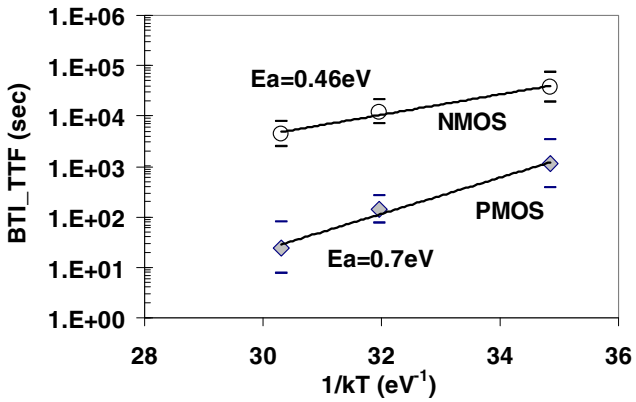


Figure 13: NMOS PBTI and PMOS NBTI activation energy (Ea)

Figures 14 and 15 show the dependence of Bias temp V_T degradation on the applied electrical field for PMOS and NMOS devices, respectively, as well as comparative results for Intel's 65nm SiO_2 technology.

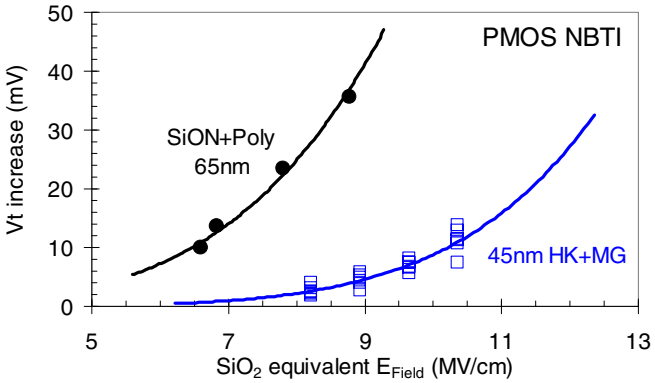


Figure 14: PMOS NBTI V_T shift vs. Electric field

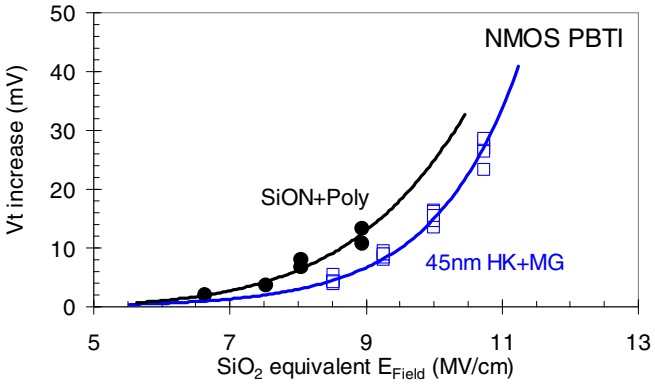


Figure 15: NMOS PBTI V_T shift vs. Electric field

Transistor Tranconductance (G_m) can provide useful insights into the nature of the BTI degradation mechanisms. Figure 16 shows the correlation of degradation in G_m to that of V_T, comparing an un-optimized HK film stack from the early development stage *Initial* and the optimized *Final* process. Note that V_T shift is well correlated to the % G_m degradation for both NMOS and PMOS for the *Initial* HK process, while it only correlates well to the PMOS on the *Final* HK process.

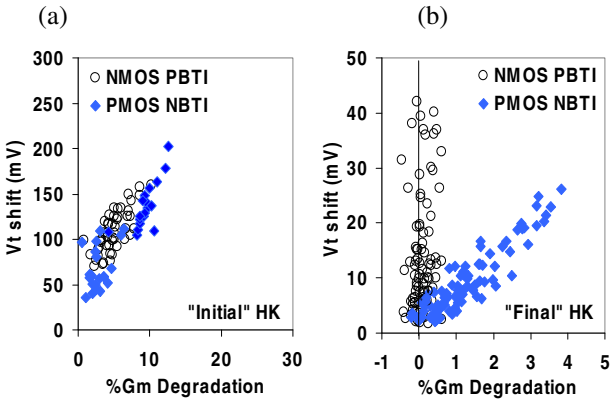
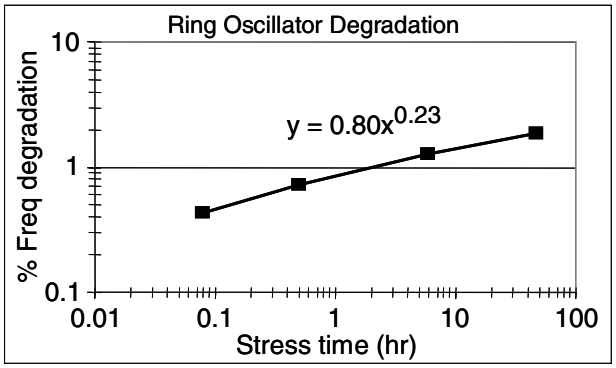


Figure 16: NMOS and PMOS BTI V_T shift vs. % G_m degradation on (a) Initial process vs. (b) Final 45nm HK+MG process.

In addition to discrete transistor test structures, circuits of various complexities are used to validate transistor degradation models. Ring Oscillators (ROs) are particularly useful for validating BTI impact on performance over variations in configuration and layout. Figure 17(a) shows representative 45nm HK+MG RO degradation data with a time slope of ~.2 as expected from device-level BTI. Figure 17(b) shows the RO stress data vs. discrete transistor results showing very good agreement.

(a)



(b)

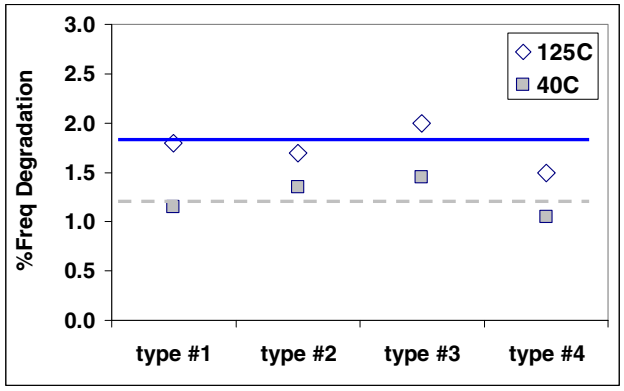


Figure 17: (a) Ring Oscillator (RO) degradation (b) RO stress data against BTI model based on discrete transistors

Transistor Degradation—Fast Traps

Another manifestation of degradation in transistor characteristics due to charge trapping is rapid shifting in device thresholds upon application of bias. This effect is of particular concern for HK dielectrics. Intel's optimized 45nm HK+MG process has negligible fast trapping, but the following comparison to early development process revisions illustrate that this is a potential area of concern: Figure 18 compares NMOS BTI V_T shift over time for the *Initial* and *Final* HK+MG processes. Note the very large, increase in V_T <1 sec. and the relative lack of dependence of V_T degradation on stress voltage for the *Initial* process compared to the mature *Final* process.

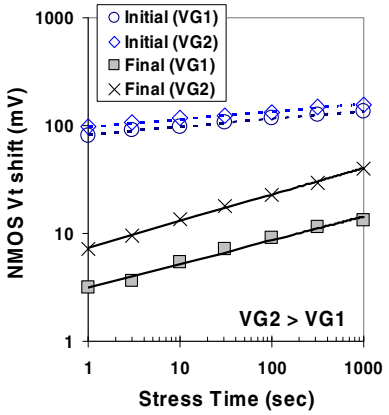


Figure 18: NMOS PBTI V_T shift for Initial and Final HK+MG process

Transistor Degradation—Hot Carrier

Injection of hot carriers can result in degradation in V_T and G_m. The low conduction band offset for Hf-based HK dielectric in contact with a silicon substrate results in a reduced barrier, and it has been suggested in the literature that this poses a potentially increased risk for the hot carrier injection [17]. Figure 19 shows NMOS hot electron reliability comparison between Intel's 45nm HK+MG and 65nm Poly/SiON transistors [2]. The 45nm HK+MG transistors actually show a large improvement in lifetime (>7X at same Isub) relative to 65nm. TTF to impact ionization slope is very similar between the two technologies.

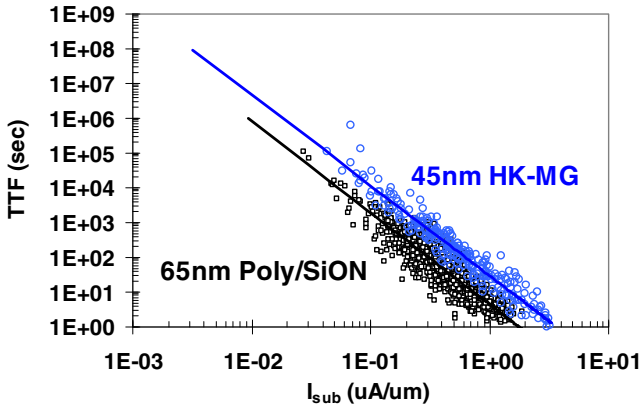


Figure 19: NMOS hot electron performance of 45nm HK+MG and 65nm Poly/SiON transistors

Process Charging

Assessment of process charging damage was conducted using specialized test structures with antennae connected to the gate of 'victim' transistors with varying levels of charging protection. The complete structure set extends beyond that allowed within the process design rules.

Transistor parametric characteristics at the completion of processing are monitored on these structures, and transistor reliability stresses are conducted and compared to reference devices with no antennae. These results confirm no charging degradation on devices compliant to design rule protection requirements.

DISCUSSION

TDDB results show that Intel's 45nm HK+MG transistor delivers equivalent dielectric lifetimes for NMOS devices at a 30% higher electric field. Due to the band offsets, PMOS TDDB has much higher margins. Conventional Gate Oxide (GOX) modeling formulations are largely applicable to these devices; however, the dependence of TTF on the applied electric field is not a constant exponential: the high field regime, above ~12MV/cm, exhibits a shallower acceleration slope. This acceleration factor change is observed to occur at an e-field value similar to the point of transition in gate leakage from a direct tunneling (DT) regime to a Fowler-Nordheim tunneling (FN) regime, as shown in Figure 20.

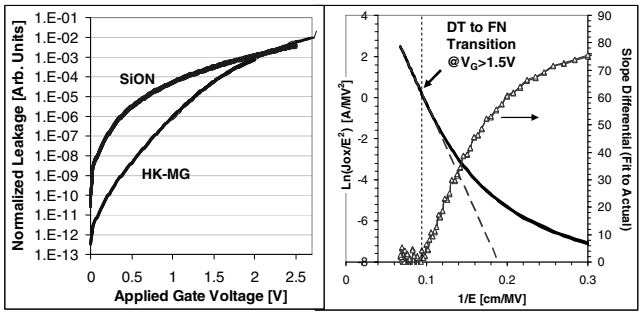


Figure 20. Left plot shows gate leakage vs. bias for HK+MG [1] vs. SiON [2]. At higher V_G values, DT leakage through SiON dominates the net leakage, but at lower biases, leakage through the HK+MG stack drops off rapidly due to DT through the HK layer. Right plot shows DT to FN transition occurs in the HK+MG stack at V_G>1.5V (triangles are delta from fit to actual slope; transition extracted at delta • 0).

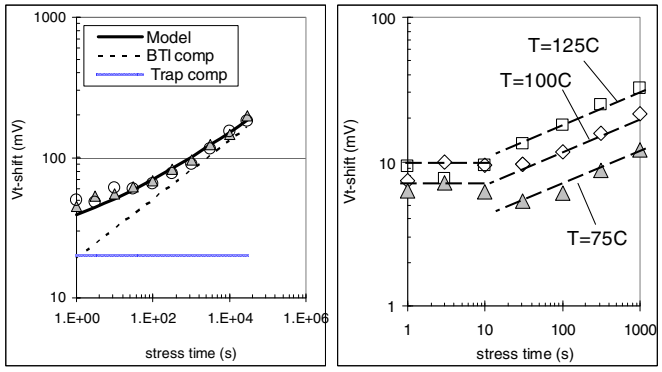
Such a change in the tunneling mechanism is expected for bilayer dielectric stacks and has been predicted for HK+MG by Dunga et al. [18]. At lower biases, direct tunneling through the entire HK+MG stack causes gate fluence to drop rapidly—and Degraeve et al. [19] have used a fluence-driven model to predict a strong increase in TDDB lifetimes at voltages near the operating range. However, very limited data exists in the literature for such low voltage TDDB data on HK+MG stacks. For the Intel HK+MG process, CVS data were collected spanning the low- and high-field regimes—with total stress times exceeding three months duration to provide good resolution. The field dependence was found to follow an exp(E) relation in the low-field regime, and an exp(1/E) relation in the high-field regime, consistent with the results of Hu et al. and McPherson et al. [20, 21]. This is not unique to HK+MG stacks; a similar transition is observed for SiO_2 data at high E-fields. However proper characterization across this range assumes fundamental importance for HK+MG where the typical operating and transistor characterization field span this transition.

The TDDB area scaling for Intel's 45nm HK+MG follows the conventional Weibull formulation. The β is in line with expectations for the thickness of the gate stack and the good agreement of the model out to product-like areas demonstrates an absence of 'defect' TDDB concerns on the *Final* optimized process flow.

The BTI results show that with sufficient optimization, aggregate NMOS+PMOS degradation levels equivalent to those achieved on SiO2 can be achieved on HK+MG operating at 30% higher E-fields. For the final process, PMOS BTI is matched at 50% higher E-field, and the PMOS degradation behavior on Intel's 45nm HK+MG process is found to be very similar to that observed on conventional SiON with equivalent acceleration and time dependence. The correlation of G_m and V_T shifts, the polarity of the V_T shift and charge pumping, and recovery data (not shown) all support that PMOS BTI is similarly driven by positive charge trapping near the Si/dielectric interface (D_{it}). P_b (Si-H dangling bonds at the interface) defects have been proposed for the PMOS IL degradation in NBTI stress [22].

On the other hand, the data show that NMOS BTI is largely driven by electron trapping within the HK bulk on the optimized process with additional contribution from interface traps on the *Initial* process. This explains the lack of NMOS G_m shift on the optimized process (the trap generation is further away from the interface) and the lower observed Ea on NMOS PBTI due to the direct tunneling of electrons from the substrate into the HK bulk [23, 24]. NMOS SILC degradation is also attributed to bulk traps, and a strong correlation is observed to trap density measured with charge pumping. Electron trapping in Hf-based oxides has been attributed to the presence of Oxygen vacancies [25].

The very large initial V_T shifts during BTI stresses observed on the initial HK process are due to high densities of pre-existing fast traps associated with D_{it} generation and/or hole trapping [4, 26]. These traps are very shallow, explaining the low observed temperature dependence, as shown in Figure 21(b).



(a) (b)

Figure 21: (a) BTI model with existing traps for "initial" process. (b) Fast trap component shows very weak temperature effect.

This effect is essentially eliminated with reduction in trap densities on the optimized HK process. This is further illustrated by pulsed IV sweeps on the *Initial* and *Final* processes that show large hysteresis on the former (shown in Figure 22). Fast traps in HK can lead to large instabilities resulting in underestimation of degradation with DC measurements.

(a) (b)

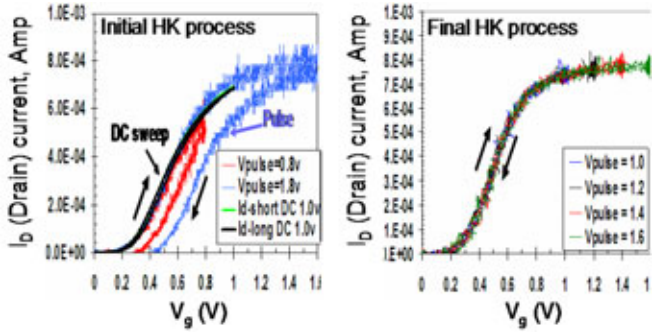


Figure 22: Pulse IV characterization of material with (a) poor reliability and (b) good reliability. Pulse t_r/t_f=100µsec, width=350µs used.

The rate of hot carrier injection for Intel's 45nm HK+MG transistor has been shown to be substantially lower than for conventional devices at the 65nm node. The data suggest that the effect continues to be dominated by that IL trapping with similar dependence on impact ionization induced substrate current. The reduced degradation is likely explained by the lower injection rate of the generated charge for the physically thicker dielectric. Figure 23 shows high temperature AC stressed RO data at several different stress voltages. The power-law time dependence slope increased with stress voltage implying that hot carrier degradation is contributing to BTI degradation during RO stress only at biases well in excess of real application conditions, and therefore hot carrier degradation is not a significant reliability consideration at any realistic use condition for this 45nm technology.

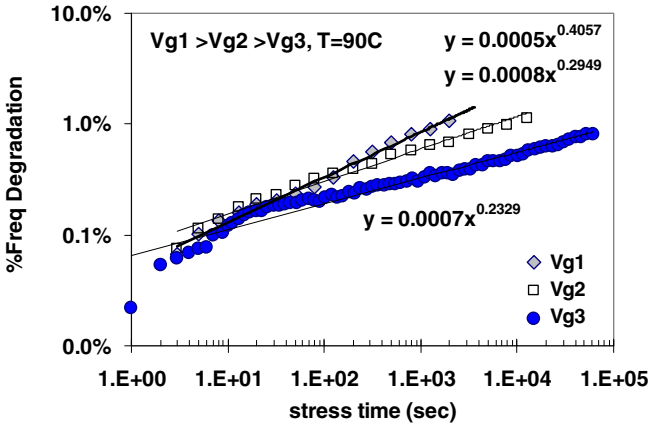


Figure 23: 45nm HK+MG Ring Oscillator (RO) stressed in AC at high temperature

CONCLUSIONS

This paper provides an overview of the reliability of Intel's 45nm HK+MG transistors demonstrating that these devices deliver reliability comparable to conventional SiO_2 devices at ~30% higher operating fields with negligible SILC degradation.

All of the major transistor reliability modes have been discussed with particular focus on those of greatest potential concern for HK+MG devices. The data for non-optimized material from early stages of technology development demonstrate that the reliability concerns for HK transistors raised in the literature are well founded. TDDB and BTI, the two modes which traditionally constrain transistor reliability, required dramatic improvement from the early experimental process revisions, for e.g., a>3X reduction in BTI V_T shifts. Similarly SILC and fast trapping are legitimate concerns but can be engineered out.

The primary HK reliability impact for TDDB, BTI, fast trapping, and SILC are all shown to relate to the potential for high trap densities in the HK stack. Controlling the trap densities within the HK and at the interfaces within the stack is critical to achieving the level of reliability demonstrated here.

The data also illustrate some challenges and potential pitfalls in characterizing the reliability of HK+MG transistors. In particular, accurate modeling of TDDB requires data collection over longer durations to properly calibrate the field dependence, and the presence of fast traps can result in underestimation of degradation using conventional DC measurements.

The overriding message is that although the revolutionary change represented by HK+MG poses daunting reliability challenges, these can all be addressed with appropriate transistor architecture and process optimizations, delivering the same reliability as conventional SiO_2 devices with compelling power/performance advantages.

ACKNOWLEDGMENTS

M. Agostinelli, M. Brazier, R. Chau, G. Dewey, T. Ghani, J. Kavalieros, K. Kuhn, M. Kuhn, M. Metz, K. Mistry, S. Ramey, A. Roskowski, J. Sandford, C. Thomas, J. Thomas, C. Wiegand, and J. Wiedemer.

REFERENCES

[1] K. Mistry et al., "A 45nm Logic Technology with High-k + Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging." In *IEDM Technology Digest*, 2007.

[2] P. Bai et al., "A 65 nm Logic Technology Featuring 35nm Gate Lengths, Enhanced Channel Strain, 8 Cu Interconnect Layers, Low-k ILD, and 0.57um^2 SRAM Cell." In *IEDM Technology Digest*, 2004, pp. 657–660.

[3] J. Mitard et al., "Large-Scale Time Characterization and Analysis of PBTI in HfO_2/Metal Gate Stacks." In *Proceedings of International Reliability Physics Symposium*, 2006, pp. 174–178.

[4] H. Reisinger et al., "A Comparison of Very Fast to Very Slow Components in Degradation and Recovery Due to NBTI and Bulk Hole Trapping to Existing Physical Models." *IEEE Transactions on Device and Materials Reliability*, *vol. 7, no. 1*, 2007, pp. 119–129.

[5] M. Alam et al., "A Critical Examination of the Mechanics of Dynamic NBTI for PMOSFETs." In *IEDM Technology Digest* 2003, pp. 345–348.

[6] A. Agarwal et al., "A Process-Tolerant Cache Architecture for Improved Yield in Nanoscale Technologies." *IEEE Transactions on VLSI*, *vol. 13, no. 1*, January 2005, pp. 27–38.

[7] J. Maiz, "Reliability Challenges: Preventing Them from Becoming Limiters to Technology Scaling." In *Proceedings of International Integrated Reliability Workshop*, Invited talk, 2006.

[8] G. Bersuker et al., "Mechanism of Electron Trapping and Characteristics of Traps in HfO_2 Gate Stacks." *IEEE Transactions on Device and Materials Reliability, vol. 7, no. 1*, 2007, pp. 138–145.

[9] M. Houssa et al., "Detrimental Impact of Hydrogen on Negative Bias Temperature Instabilities in HfO_2-based pMOSFETs." In *Symposium on VLSI Technology Digest* 2004, pp. 212–213.

[10] G. Wilk et al., "High-k gate dielectrics: Current status and materials properties considerations," *Journal of Applied Physics*, *vol. 89, no. 10*, 2001, pp. 5243–5275.

[11] A. Shanware et al., "Characterization and Comparison of the Charge Trapping in HfSiON and HfO_2 Gate Dielectrics." In *IEDM Technology Digest*, pp. 939–942, 2003.

[12] J. H. Stathis, "Percolation models for gate oxide breakdown." *Journal of Applied Physics*, *vol. 86, no. 10*, 1999, pp. 5757–5766.

[13] S. C. Song et al., "Highly Manufacturable 45nm LSTP CMOSFETs Using Novel Dual High-k and Dual Metal Gate CMOS Integration." In *Symposium on VLSI Technology Digest*, 2006, pp. 13–14.

[14] G. Ribes et al., "Review on High-k Dielectrics Reliability Issues." *IEEE Transactions on Device and Materials Reliability, vol. 5, no. 1*, 2005, pp. 5–19.

[15] C. Prasad et al., "Dielectric Breakdown in a 45 nm High-K/Metal Gate Process Technology." In Press: in *Proceedings of International Reliability Physics Symposium*, 2008.

[16] S. Pae et al., "BTI Reliability of 45 nm High-K + Metal-Gate Process Technology." In Press: in *Proceedings of International Reliability Physics Symposium*, 2008.

[17] H. Park et al., "Decoupling of Cold-Carrier Effects in Hot-Carrier Reliability Assessment of bHfO_2 Gated nMOSFETs." *IEEE Electron Device Letters, vol. 27, no. 8*, 2006, pp. 662–664.

[18] M. V. Dunga et al., "Modeling of Direct Tunneling Current in Multi-layer Gate Stacks." *Nanotechnology 2003*, *vol. 2*, pp. 306–309.

[19] R. Degraeve et al., "Stress Polarity Dependence of Degradation and Breakdown of SiO2/High-K Stacks." In *Proceedings of International Reliability Physics Symposium*, 2003, pp. 23–28.

[20] C. Hu et al., "A Unified Gate Oxide Reliability Model." in *Proceedings of International Reliability Physics Symposium*, 1999, pp. 47–51.

[21] J. W. McPherson et al., "Trends in the Ultimate Breakdown Strength of High Dielectric-Constant Materials." *IEEE Transactions On Electron Devices*, *vol. 50, no. 8*, 2003, pp. 1771–1778.

[22] M. Houssa et el., "Reaction-dispersive proton transport model for negative bias temperature

instabilities." *Applied Physics Letters*, *vol. 86, 093506*, 2005.

[23] G. Ribes et al., "Review on High-k Dielectrics Reliability Issues." *IEEE Transactions on Device and Materials Reliability, vol. 5, no. 1*, 2005, pp. 5–19.

[24] S. Kalpat et al., "BTI Characteristics and Mechanisms of Metal Gated HfO_2 Films with Enhanced Interface/Bulk Process Treatments." *IEEE Transactions on Device and Materials Reliability*, *vol. 5, no. 1*, 2005, pp. 26–35.

[25] A. Kang et al., "Electron spin resonance observation of trapped electron centers in atomic-layer-deposited hafnium oxide on Si." *Applied Physics Letters*, *vol. 83, no. 16*, 2003, pp. 3407–3409.

[26] V. Huard et al., "New characterization and modeling approach for NBTI degradation from transistor to product level." In *IEDM Technology Digest*, 2007.

AUTHORS' BIOGRAPHIES

Jeffrey Hicks is a Senior Principle Engineer in Intel's Technology Development Quality and Reliability Group and was the Q&R Program Manager for Intel's 45nm HK+MG technology. He received his B.S. degree in Applied Physics from Caltech in 1980 and joined Intel's Technology Development Group in Santa Clara CA in that year as a Reliability Engineer working on Bipolar and EPROM non-volatile memory technologies. In his years at Intel to date Mr. Hicks has served in numerous Quality and Reliability functions spanning Technology Development, Manufacturing, Product, and Customer Quality and Reliability. He has several patents received or pending and has published a number of technical papers including Best Paper at the 1999 International Reliability Physics Symposium. He is the recipient of three Intel Achievement Awards. His e-mail is Jeffrey.Hicks at intel.com.

Daniel Bergstrom is a Group Leader for front-end metals deposition in Intel's Portland Technology Development Group. He has been with Intel since 1997 and has worked on metal deposition technologies stretching from 0.25µm down to his current work on 32nm technology. He received a B.S. degree in Ceramic Engineering from the University of Illinois in 1992 and a Ph.D. degree in Materials Science from the University of Illinois in 1997. His e-mail is daniel.bergstrom at intel.com.

Michael Hattendorf is a Process Integration Engineer working in Intel's Portland Technology Development Group. He has been with Intel since 2002 and has focused on 45nm transistor yield, performance, and reliability for the last four years. He received B.S., M.S., and Ph.D. degrees in Electrical Engineering from the University of Illinois at Urbana-Champaign. His e-mail is michael.l.hattendorf at intel.com.

Jason Jopling is a Quality and Reliability Engineer in Intel's Logic Technology Development Group working primarily on memory and transistor reliability. His interests include process- and device-level impacts on quality/reliability of integrated memories, circuits, and products. He holds dual B.S. degrees in Biomedical and Electrical Engineering from the University of Southern California and M.S. and Ph.D. degrees in Electrical Engineering from Duke University. His e-mail is jason.jopling at intel.com.

Jose Maiz is an Intel Fellow and Director of Logic Technology Quality and Reliability. He joined Intel in 1983 and became a Fellow in 2002. He is presently responsible for identification of silicon reliability limiters to scaling and their resolution for Intel's next-generation silicon technologies and logic products. Dr. Maiz graduated with a degree in Physics from the University of Navarra in San Sebastian in 1976. He then moved to the U.S., graduating with M.S. and Ph.D. degrees in Electrical Engineering from Ohio State University in 1980 and 1983, respectively. In 1983 he joined Intel Corporation where he has contributed in various management and individual contributor roles to more than 10 silicon technology generations in the areas of quality and reliability. Maiz holds 9 patents and has 13 more pending. He has authored or co-authored over 35 publications and conference presentations, a number of them invited including keynote addresses at the 2007 IRW and MRQW conferences. He is a recipient of an Intel Achievement Award and numerous divisional recognition awards. He was also a Fulbright Scholar in 1978. He is a senior member of the IEEE. His e-mail is Jose.Maiz at intel.com.

Sangwoo Pae received a B.S.E.E. degree from the Chung-Ang University in Korea (1994) and M.S. and Ph.D. degrees in 1996 and 1999, respectively, in ECE from Purdue University in West Lafayette, IN. He has been with Intel since 2000 in the Logic Technology and Development Quality and Reliability Group in Oregon. He has contributed to the development and reliability of Intel's 130nm, 90nm, and 45nm technology and recently has worked on 32nm logic technology nodes. He is currently a Sr. Staff QRE. He has authored/co-authored over 20 publications and holds 2 patents with 9 more pending. His e-mail is Sangwoo.pae at intel.com.

Chetan Prasad received his B.E. degree from the University of Mumbai in 1997 with an emphasis on Electronics and Telecommunications. He received his M.S. degree in 1999 and his Ph.D. degree in 2003 from Arizona State University, where his research focused on non-equilibrium carrier dynamics and relaxation in III-V semiconductor nanostructures. He joined Intel

Corporation in 2003 and is currently a Senior Reliability Engineer with Intel's Technology Development Quality and Reliability Group. His responsibilities include research in transistor reliability and dielectric wear-out mechanisms for Intel's advanced process technologies. He has contributed to Intel's 65nm and 45nm technologies and is currently working on enhancement activities on the 45nm technology node. Chetan has authored/co-authored over 30 papers published in peer-reviewed journals and conference proceedings and has 2 patents pending. He has received an Intel Achievement Award and an LTD Divisional Award. His e-mail is chetan.prasad at intel.com.

Jami Wiedemer is a Device Engineer in Intel's Portland Technology Development Group. She joined Intel in 2005 and has focused on electrical properties of high-k gate dielectrics and their impact on 45nm transistor performance. She is currently working on pathfinding for 22nm process technology. She received M.S. and Ph.D. degrees in Electrical Engineering from Cornell University and a B.S. degree in Electrical Engineering from the University of Notre Dame. Her e-mail is jami.a.wiedemer at intel.com.

Flip-Chip Packaging Technology for Enabling 45nm Products

Neha M. Patel, Technology and Manufacturing Group, Intel Corporation
Vijay Wakharkar, Technology and Manufacturing Group, Intel Corporation
Sairam Agrahram, Technology and Manufacturing Group, Intel Corporation
Nitin Deshpande, Technology and Manufacturing Group, Intel Corporation
Mengzhi Pang, Technology and Manufacturing Group, Intel Corporation
Ravindra Tanikella, Technology and Manufacturing Group, Intel Corporation
Rahul Manepalli, Technology and Manufacturing Group, Intel Corporation
Pat Stover, Technology and Manufacturing Group, Intel Corporation
James Jackson, Technology and Manufacturing Group, Intel Corporation
Ravi Mahajan, Technology and Manufacturing Group, Intel Corporation
Prabhat Tiwari, Technology and Manufacturing Group, Intel Corporation

Index words: lead-free, flip chip, 45nm, packaging, form factor management

ABSTRACT

Intel's packaging team has been working on developing completely Pb-free packages that can be utilized in a variety of products and market segments including the newly emerging mobile Internet devices. These technologies have been introduced into high-volume manufacturing to enable packaging of 45nm silicon devices in 2007. In order to hit this milestone, a significant number of engineering challenges had to be overcome to select and integrate the new assembly materials into the complex interconnect structure of Intel's 45nm process technology. The new solder alloys for first-level interconnect resulted in significantly higher stress on the silicon, and the Intel team reengineered many aspects of the assembly materials and process technology to resolve the crucial problem and deliver this innovative technology. The change to lead-free (Pb-free) solder alloys necessitated the development of alternate flux materials to clean off the more tenacious tin oxides from the solder surface. The new flux material had to be stable at high process temperatures as well as cleanable following the chip attach process to allow strong adhesion between the underfill, the bump metallurgy, and the die passivation. The new Pb-free, first-level interconnect architecture is superior to the older materials in many aspects, including higher current-carrying capability, and it is more reliable and environmentally friendly, being 100% Pb-free. The development of the high-k 45nm devices also enables Intel to introduce high-performance microprocessors with very low power consumption. These devices enable development of fully functional personal-computer-like features in a hand-held device. However, in order to successfully integrate a 45nm silicon chip in a hand-held device, significant reduction in the form factor of the flip-chip package was essential. In order to achieve this goal, significant technology challenges were overcome through the introduction of new underfill material and process technologies. Further reduction in the z-height required die-thinning and introduction of thinner substrates. In this paper key technical challenges associated with Pb-free interconnect and form-factor reduction are discussed.

INTRODUCTION

Intel's 45nm technology portfolio includes the Penryn family of processors, which build on the success of the revolutionary core microarchitecture as well as the family of devices, based on Intel® Centrino® Atom™ processor technology, targeted for mobile Internet device applications. The Penryn family of processors features new dual-core desktop processors, quad-core desktop processors, quad-core server processors, and dual-core mobile processors. These new 45nm processors include features to improve performance at any given frequency:

they have up to 50% larger L2 caches and expanded power-management capabilities for new levels of energy efficiency [1]. The Intel® Atom™ processors have been developed from the ground up with the aim of minimizing the power consumption and yet enabling a high-performance fully functional chip for the mobile Internet device segment [2].

The introduction of 45nm products also marks the move to 100% Pb-free packages to meet Intel's environmental performance goals. Lead has been used generously in the past as the primary component of the metal alloy (a mixture of tin and lead) used to electrically connect the silicon processor to the motherboard via an organic package. The interconnect hierarchy is shown in Figure 1.

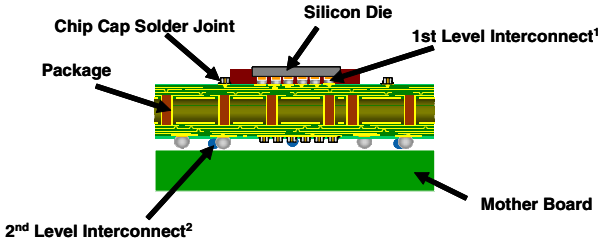


Figure 1: Schematic of a flip-chip ball grid array (FCBGA) pack

Due to the harmful impact of lead on the environment, however, Intel's engineers engineered a solution to get every last milligram of lead out of the package. The entire First-Level Interconnect (FLI) architecture was reengineered to be 100% Pb-free in a phased approach. In the first phase, Cu bumps were incorporated as part of the 65nm process technology CPUs [3] in place of the more compliant high *Pb*-bumps on the silicon die. This was followed by tin-silver-copper (SAC) solder in place of eutectic *Pb-Sn,* as part of the 45nm technology. The new architecture not only meets the stringent quality and reliability requirements of Intel products but also significantly improves bump cracking, bump electromigration, and solder fatigue performance, all of which are crucial to enabling reliable, high-performance microprocessors. Both *Cu*-bumps and SAC solder on the substrate are much stiffer than their leaded counterparts and impart significantly higher thermomechanical stress on the mechanically-weak, low-dielectric constant materials on the silicon die [4]. The use of a higher number of low-*k*-based metal layers in the 45nm products for improved interconnect performance further exacerbates the stress-management challenge. These challenges were resolved with an optimization of far back-end architecture, design, materials, and processes in both fab and assembly.

The move towards producing thin and light laptops and the need for improving the processing power in smart phones and mobile Internet devices has led to increased focus on the development of small form-factor, flip-chip package technologies that can accommodate full function, high I/O CPUs. In these market segments, the focus is on reducing the overall form factor of the package. This reduction in space is essential to enable system-level reduction in the motherboard size used in hand-held, mobile Internet device applications [2]. The drive towards smaller form factors places tremendous pressure on minimizing the area occupied by the flip-chip package and requires significant re-engineering of the flip-chip packaging technology.

In this paper we focus on the novel materials and processes that were developed in order to overcome these challenges to achieve 100% Pb-free interconnect as well as a significant reduction in the form factor for flip-chip packages.

FIRST-LEVEL LEAD-FREE INTERCONNECT

The 45nm process incorporates high-K+metal gate (HiK+MG) transistors for the first time along with third-generation strained silicon, nine copper interconnect layers, 193nm dry patterning, and 100% Pb-free packaging. A complete overview of the 45nm silicon process technology is given in [5]. The die far back-end architecture makes use of a very thick metal 9 (TM9) layer and a polymer dielectric, as shown in Figure 2. The integration of Pb-free solder with the novel TM9 architecture that employed a new dielectric/passivation material posed new challenges for material compatibility and M9 stack reliability, in addition to low-*k* dielectric material cracking/delamination. These issues were fully resolved through an iterative optimization of the metal interconnect, package design, materials, and processes within the fab and assembly.

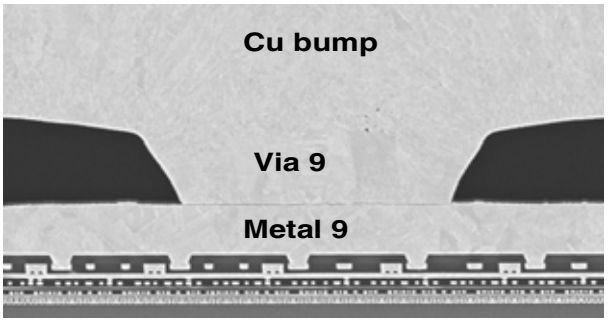


Figure 2: SEM depicting the M9 and V9 layers: M9/V9 is significantly thicker than all the other die metal layers

Challenges of First-level, Lead-free Interconnect

Compared to their leaded counterparts, tin-rich, Pb-free solder materials possess physical, metallurgical, and mechanical properties that pose significant challenges to flip-chip assembly and reliability. Their higher melting point (~30°C higher than SnPb) leads to an increase in the thermal expansion mismatch between silicon die and organic packaging and induces higher stress in the FLI solder joint compared to previous generations. The intrinsically higher mechanical stiffness of Pb-free solder and dramatically reduced mechanical strength of low-k dielectric materials of the silicon backend structures led to significant assembly challenges such as die Interlayer Dielectrics (ILD) cracking and occasional solder joint interfacial delamination (Figure 3). The inferior wettabililty of the Pb-free solder results in a reduced solder wicking with die copper column during chip attachment (Figure 4), and it considerably increases sensitivity to assembly-interaction-related failures, such as solder joint interconnect opens and non-wets, which require more stringent solder bump dimensional control. To achieve a healthy assembly yield for our Pb-free SAC process, tolerances, and sensitivities several new manufacturing process parameters had to be thoroughly studied, e.g., bump height variation and bump-level defects. Solder joint voiding is also of concern, as the outgassing of soldering flux residues becomes retarded due to higher surface tension and is accelerated by the ease of the tin oxide formation of the liquid solder.

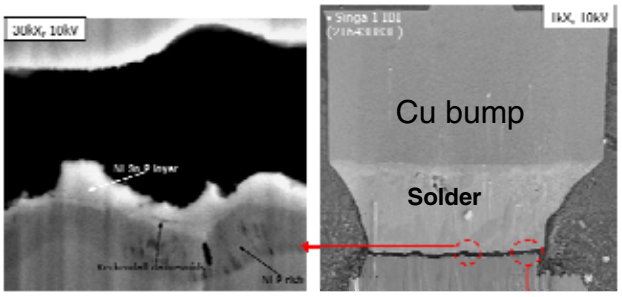


Figure 3: Solder joint delamination along solder/surface finish interface

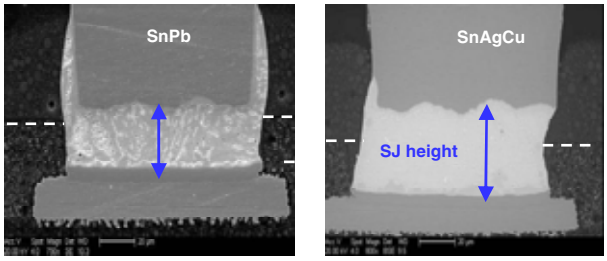


Figure 4: Reduced wettability of *SnAgCu* with die *Cu*-bump leads to a reduced solder joint collapse

The use of existing chip-attach fluxes developed for leaded solder alloys can cause significant defects, such as solder voids and interfacial nonwets, in the Pb-free FLI joints, due to the higher thermodynamic stability of the oxides in Pb-free solders. Furthermore, the need for higher peak temperature and Time Above Liquidus (TAL) for Pb-free solder can cause significantly higher flux residue to remain on the packages. Both these factors necessitated the development of a new flux that needed to be highly active in order to remove the thermodynamically stable tin oxides and improve wettability of solder to *Cu*. These new fluxes also had to have significantly reduced outgassing at reflow peak temperature and the TAL in order to reduce FLI voiding and in order to leave a residue that was easily cleanable by the hot water during the deflux process. In addition, the new flux formulations had to be compatible with the substrate solder resist material and the new die passivation material, have sufficient tackiness to prevent die misalignment and substrate solder bridging, and be capable of being printed or sprayed on the substrate.

The chip-attach process for flip-chip packaging follows these main steps: flux application, chip placement, reflow of chip joints, and cleaning of flux residue. Intel has been printing the flux in CPU packaging for many years. This process involves printing flux through an aperture opening in a stencil on the substrate bump area. This flux application process has worked very well for high-volume applications up to this point, but it has reached its limit for larger die sizes and multiple dies on a package. Using a flux printing process for larger and multiple dies can lead

to scraping of substrate bumps that could potentially cause reliability problems such as non-wets, limited performance because of decreased maximum current-carrying capacity (referred to as IMAX), and reduced yield.

An out-of-the-box approach was pursued to enable a spray flux application process (SPRINT) by using a print flux material that was designed to be highly viscous at room temperature so it does not flow.

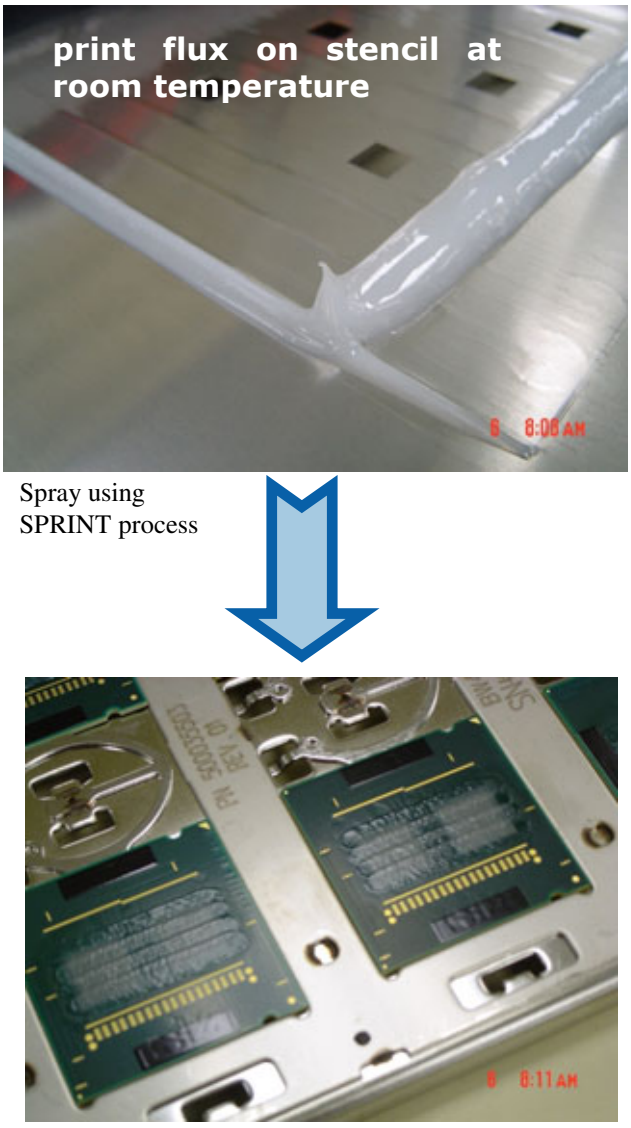


Figure 5: Conceptual description of "SPRINT" process—spraying a Print flux material using a dispenser

The print flux material is a non-Newtonian fluid with shear thinning properties. Several equipment and process changes had to be made to get the required dispense characteristics (Figure 5). An optimum fluid path heater design equipped with sufficient flux hold-up volume and heated to the target temperature prior to dispensing (Figure 6) was needed to enable a stable and capable dispensing process. Significant process characterization work had to be carried out to optimize different dispense parameters such as fluid and atomizing coaxial air-pressure, dispense temperature, dispense height and width, and line speed. The goal was to get just the right amount of flux on the substrate bumps: too little flux led to poor quality joints between substrate and die bumps, and too much flux caused die misalignment during the reflow process. All the process characterization work led to the fundamental understanding of the impact of flux dispense parameters on die misalignment yield and FLI joint quality.

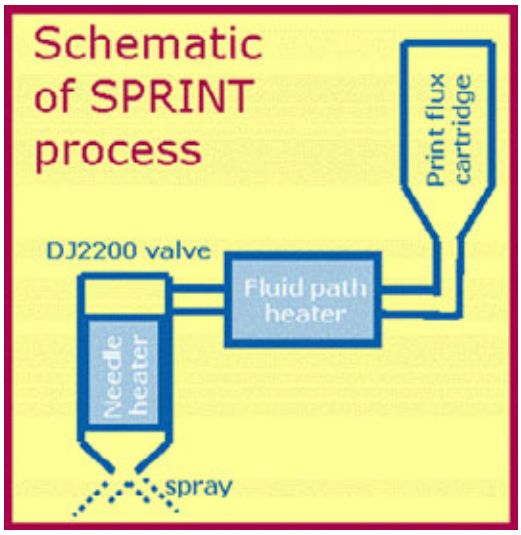


Figure 6: Schematic of "SPRINT" (spraying the print flux) process

Results and Discussions

To overcome the challenges of integrating Pb-free, flip-chip packaging with 45nm silicon, significant work was carried out from design, materials, processing, and metrology perspectives. Based on over five years of research experience on Pb-free materials, Intel selected *Sn-Ag-Cu* solder metallurgy as the flip-chip die attachment material. This material has showed significant improvement in solder-joint quality compared to the other more commonly used Pb-free alloys in our case.

Enabling the SPRINT process to high-volume manufacturing involved solving several manufacturability issues specific to ease and repeatability of tool maintenance.

The SPRINT process has met yield and reliability goals both during the development of Pb-free packages and during high-volume manufacturing ramp in Intel factories. At the end of the development cycle, the SPRINT process

is at 99.5+% in three different Intel sites, and the samples used during the development phase show no die misalignment.

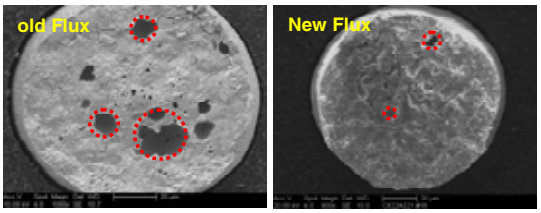


Figure 7: Comparison of interfacial defects between old and new fluxes: new flux is significantly better

Figures 7, 8, and 9 show the comparison of FLI joint interfacial defects, solder voiding, and electromigration performance between the new flux and previous-generation fluxes. The new flux provides significantly better performance in all three attributes. Figure 10 shows the comparison between leaded (65nm process technology) and Pb-free electromigration performance with optimized materials/process.

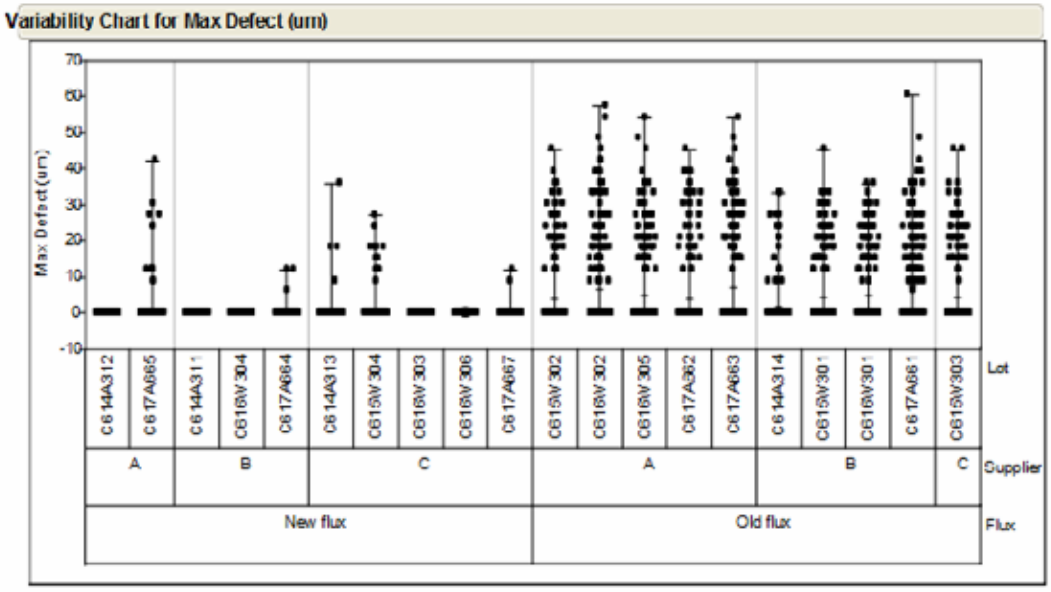


Figure 8: Comparison of FLI solder voids between new (left) and old (right) fluxes: new flux is significantly better

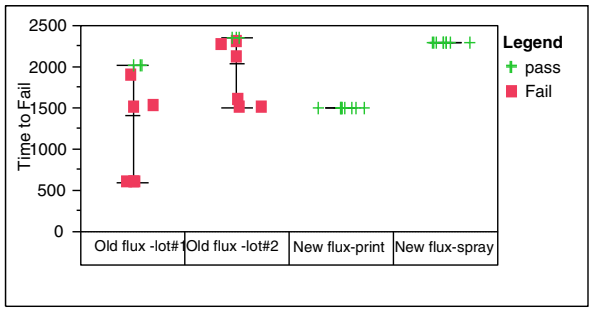


Figure 9: Comparison of electromigration performance between old flux, new flux (print formulation) and new flux (spray formulation): new flux is significantly better

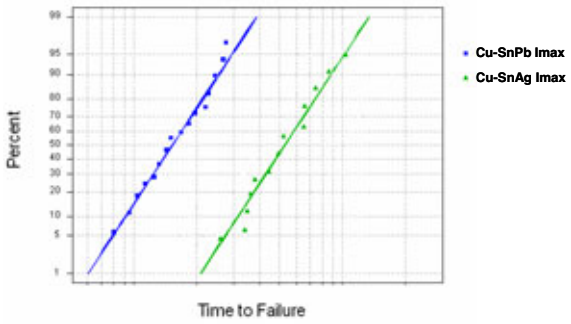


Figure 10: Comparison of bump electromigration performance of *Cu-Sn/Pb* and 100% *Pb-free* FLI

Figure 11 shows C-mode scanning acoustic microscope (CSAM) images of product units (complete stack-up). Cracking or interface delamination in the ILD or TM9 stack will show up as white or black spots apart form the

contrast variation of the underlying pattern. The image on the left was taken after packaging: no contrast areas are observed, showing that the unit is free of cracking and/or interface delamination. These results have been reproduced in high volume, at process extremes, and have been proven to have significant margin through execution of well-designed hammer tests that impart significantly higher stress on the die than typically observed in manufacturing. The image on the right shows that the parts are clean of any issues even after reliability stresses.

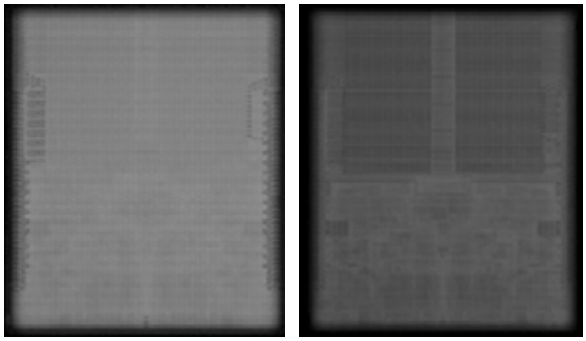


Figure 11: C-SAM units of production units: image on the left represents a unit post packaging and the image on the right is a unit post 25 hrs of HAST

These results indicate that the 45nm technology FLI architecture meets the stringent quality and reliability criteria despite the higher stress induced by Pb-free solders. Intel's FLI architecture with *Cu*-bumps is very unique and provides significantly better electromigration performance and power distribution performance than typical 100% solder-based FLI interconnects, as shown in Figure 12.

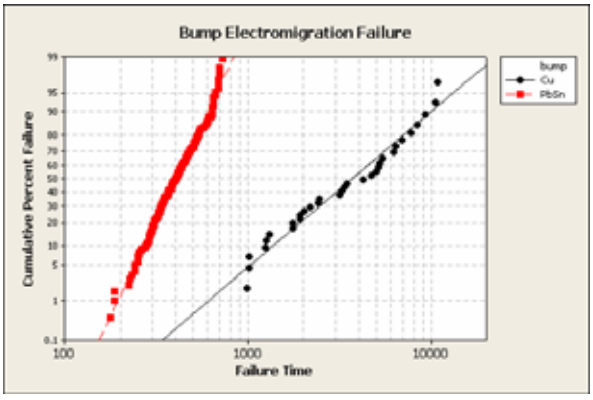


Figure 12: Comparison of bump electromigration performance of high *Pb*-bumps and Pb-free FLI

Sn-Ag-Cu solders have better wettability to metal pads, leading to reduced solder voiding, bumping, and assembly solder joint open yield loss. They also possess increased solder joint strength due to the suppression of under-bump nickel barrier layer diffusion and intermetallic growth. SAC solders also possess enhanced electromigration resistance arising from the synergistic reactions of reduced metal diffusion and interfacial defects.

As discussed earlier, the loss of solder joint collapse margins with the advent of Pb-free materials requires more stringent solder bump dimensional control. To achieve this control, solder bumping, flux material, and the reflow process were improved. The substrate C4 solder bump metrology was also modified significantly to screen out bump-level defects at high through-put speeds. Additionally, substrate packaging materials and design were also optimized to accommodate the higher reflow temperature of Pb-free solder to mitigate the associated reliability issues.

FORM FACTOR MANAGEMENT

Development of small form-factor packaging technologies required significant engineering effort in addition to the challenges involved with the integration of Pb-free, 45nm silicon technology. Key technical challenges include package size reduction in XY dimension and a reduction in the thickness of the package itself, by reducing die thickness and substrate thickness. Thin die challenges include the thinning process, stress-induced electrical property changes, and package warpage concerns. In addition, wafer thinning would require ensuring relief of residual stresses caused by the thinning process and protecting the die from chipping and cracking during wafer thinning and die singulation. Handling of the significantly warped thin wafer and die is another key challenge. A thin package with thin die is prone to warpage due to its low stiffness. Decreases in die thickness can make this problem worse, impacting the board assembly yield. Making the problem even more challenging is the need for a more coplanar BGA ball field to enable board assembly with finer BGA pitch. Solutions for these challenges required significant innovation and re-engineering in packaging material and process technologies. The approach, along with a few of the examples, is discussed below.

PACKAGE SIZE REDUCTION IN XY

The top surface of a flip-chip package contains the silicon die attached to the organic substrate and is underfilled with an epoxy material. The underfill material's primary function is to protect the flip-chip solder joints from failing due to stresses induced by the CTE mismatch between the die and the package during processing and use. The underfill material can occupy a significant area on the package as shown in Figure 13.

Figure 13: Example of underfill spread on a flip-chip CPU package

One of the key factors that controls the underfill spread on the package are the flow characteristics of the underfill material. Underfill flow under the die is driven by capillary flow. A line of underfill material is dispensed next to the die at high temperature, and capillary pressure in the material pulls the material under the die. In order to minimize the underfill spread away from the die, it is critical to increase the flow speed of the underfill material under the die, as schematically illustrated in Figure 14.

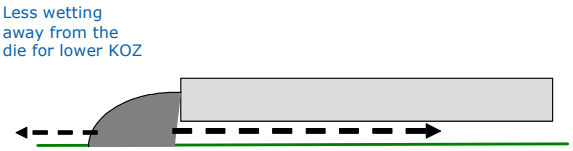


Figure 14: Competing forces for underfill spread under the die vs. away from die

In an idealized case, underfill flow can be modeled as flow between parallel plates [6], where the time to fill for an underfill between parallel plates can be calculated using this equation:

$$\text{Flow Time} = \frac{3\mu L^2}{h\gamma\cos\theta}$$

where γ is the surface tension of the underfill material, μ is the viscosity of the material, h is the gap between the plates, θ is the angle of wetting between the underfill and two surfaces, and L is the flow distance. Based on this elementary flow model, minimizing the μ/γ ratio can improve the flow speed of the underfill material.

Results and Discussion

A series of formulations with surface tension of the underfill material varying from ~20 dynes/cm^2 to 40 dynes/cm^2 at 110°C, and viscosities varying from 0.5 poise to 1.6 poise, were tested in a simple parallel plate flow experiment. Figure 15 shows the time to flow versus distance results for the formulations tested.

Based on the data from this simple flow test, it was observed that over a distance of ~15mm, significant reduction in flow time could be observed. These formulations were then tested on a package, leading to significant improvement on the flow speed of the material and therefore a reduction in the underfill spread on the package as shown in Figure 16.

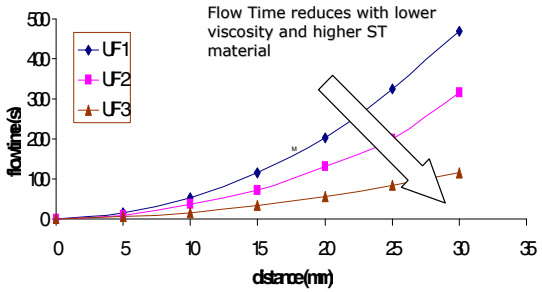


Figure 15: Flow time vs. distance for various underfill materials

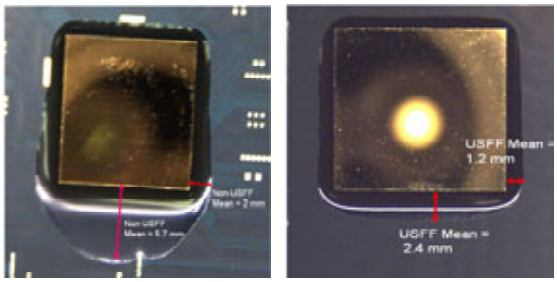


Figure 16: Underfill epoxy spread on the die with normal UF and optimized UF

The underfill material is a highly engineered epoxy-silica composite material with optimum thermo-mechanical properties (*Tg*, CTE, Modulus, etc.) to minimize stress transfer to die, and to prevent bump fatigue and die-cracking. In addition, the underfill material is also designed for toughness and adhesion to various interfaces (solder resist on substrate, passivation on die, *Cu*-bump, FLI solder, and silicon) under Pb-free, reflow conditions. These requirements lead to a very specific choice of epoxy resin chemistries that can be used in underfill material development. Adding the challenge of improved flow with higher μ/γ ratio required careful study of the surface tension/viscosity of each of the resin components in the material and to the selection of the right combination of components in order to provide good flow, and yet not compromise the reliability of the underfill. As shown in Figure 16, optimization of the μ/γ ratio of the

underfill is very influential in minimizing the underfill spread on the package. However, it is also critical to note that there is a limit to how much the underfill viscosity can be reduced and to how much the surface tension of the underfill can be increased beyond which other effects take over. These can lead to increased spread of the underfill on the sides (due to very low viscosity of the material) and de-wetting of the underfill from the substrate, due to too high a surface tension of the material. Using this approach, new underfill formulations have been developed and deployed in high volume for enabling small form-factor, flip-chip packages. Utilization of this technology in combination with other novel process and design optimization approaches have led to significant reduction (up to 60%) in the package size for small, form-factor packages.

Z-HEIGHT REDUCTION

Industry benchmark trends indicate that overall package height in high-end smart phones are typically less than 1mm. While most of these packages to date are wirebonded and overmolded, there is an increased need to develop high *I/O* flip-chip packages that have *z*-heights of less than 1mm. In order to achieve significant package height reduction we discuss die thinning and the use of thinner/coreless substrates.

Results and Discussion

The typical thickness of a 12 inch wafer is ~750–800um. Since the active silicon needed for functioning of the device is less than 20-30um, a significant reduction in *z* height can be achieved by die thinning. Wafer backgrinding is common for devices used in stacked die, wirebond packaging technology, where die thinning down to 50-75um is routinely employed in high-volume manufacturing. However, die thinning of a high-density bumped flip-chip die is a significant challenge. Figure 17 shows the SEM image of a bumped wafer with *Cu*-bumps with the bump height nominally ~50 um.

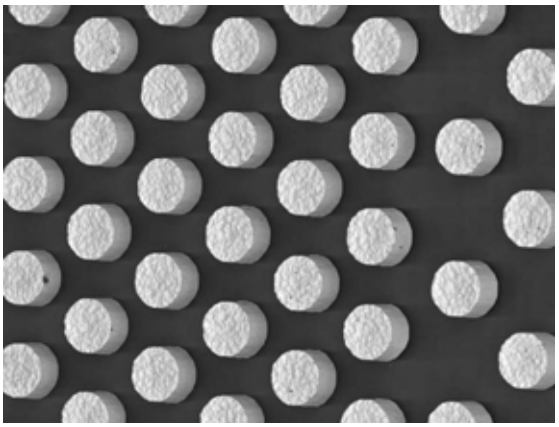


Figure 17: SEM image of a Cu bumped wafer

In order for the backgrinding process to work effectively, it is critical for the backgrinding tape to completely encapsulate the bumped wafer without any voids. This is essential since the presence of voids or the inability of the tape to completely encapsulate the bumps can lead to cracking of the wafer or non-uniform backgrinding. In addition to the encapsulation, it is also critical that the tape be removed without leaving any residue or damaging the *Cu*-bump/BLM during the tape peel step. In order to achieve this critical balance of lamination, uniform backgrinding and clean removal post backgrinding, a multi-layer adhesive, UV, curable tape with a soft backing layer was developed. The backgrind tape is laminated to the wafer, and the hardness of the multilayer adhesive layer contacting the wafer surface is well controlled to allow complete encapsulation of the bumps. After the backgrinding process, the tape is exposed to UV cure, leading to cross-linking and hardening of the multi-layer adhesive. This allows for easy detaping of the tape from the backgrinded wafer, since the harder *x*-linked adhesive can easily delaminate from the wafer/bump surface. Using this technology, Intel has been able to develop a high-volume manufacturing process for wafer thinning 12-inch, high-density bumped wafers down to 75um.

In addition to wafer thinning, the other element that can enable further z height reduction is minimizing the substrate thickness. In the case of the organic substrate, a standard build-up core of ~800um is used for typical packages. This core thickness in the substrate is typically selected in order to balance the electrical requirements of the package and its ability to go through the assembly line and SMT processing. In order to decrease the overall thickness of the package, one of the approaches is to reduce the thickness of the core layer in the package. Depending on the need for the overall package thickness, either die thinning or substrate thinning, or both, can be employed to produce the overall desirable thickness. With the combination of die thinning and substrate thinning, Intel now has the capability to produce flip-chip packages with package heights that are 33% less than standard packages (Figure 18).

Figure 18: Comparison of a standard package z-height with a thin core-thin die package

Board Level Interconnect

Intel's 45nm technology is useful for a wide variety of products from servers to mobile devices. In order to meet the size requirements of ultra mobile personal computers and mobile Internet devices, the 45nm packaging

technology had to scale the physical size of the packages to meet the demands of the ultra mobile products.

Scaling down the size of these packages while maintaining the product functionality of Intel Architecture requires the board-level interconnect to scale as well. In the case of the Intel Atom processor, the ball pitch scaled down to 0.6mm from a 0.8 to 1.27mm pitch for PCs. This scaling creates two key challenges: routing the signals on the board and ensuring a robust solder joint.

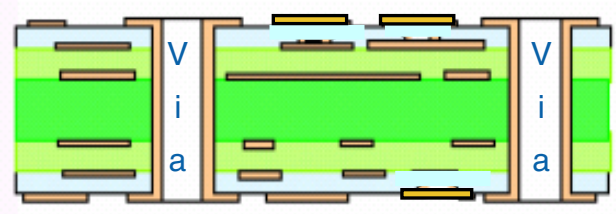


Figure 19: Cross section of a Type 3 motherboard

Traditional PCs utilize a Type 3 motherboard that is characterized by mechanically drilled vias as shown in Figure 19. In mobile stripline routing, the signals that come out of the package have to travel through these vias on their way out of the package. Unfortunately, these mechanically drilled vias do not scale well and limit the package ball pitch to approximately 0.8mm. The routing issue is solved by using High Density Interconnect (HDI) motherboards shown in Figure 20. HDI boards contain one or more layers that are connected to other layers through microvias. As the name implies, the laser drilled microvias are significantly smaller than the mechanically drilled vias and allow the signals to break out from the package.

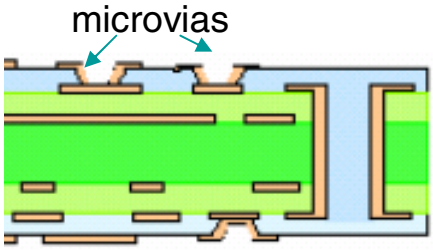


Figure 20: Cross section of an HDI motherboard

Solder joint reliability is the other board-level challenge with SFF packages. By the very nature of their use, mobile products are subject to drops. The smaller solder joints of SFF packages have less mass to handle these mechanical stresses. Gluing the corners of the packages provides the additional strength and mechanical margin for the mobile drop condition as shown in Figure 21.

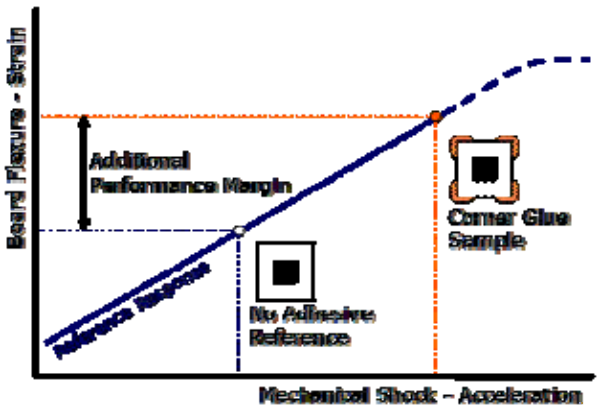


Figure 21: Corner glue provides mechanical margin for mobile drop condition

CONCLUSION

A series of iterative optimizations of design, fab, assembly materials, and process led to a high-yielding, 100% Pb-free, 45nm packaging process. The on-time and on-target delivery of this technology led to Intel's continued one-generation-ahead lead in the 45nm product, underlining Intel's assurance that it will develop a new process technology with enhanced microarchitecture or an entirely new microarchitecture every year. The elimination of lead in this technology also makes Intel the leader in achieving environmentally green products

In this paper we shared some of the key challenges associated with the development of a high-volume manufacturing compatible assembly process for packaging Intel's 45nm, completely Pb-free devices. Key technical challenges were addressed through development of novel FLI solders, fluxing material, and process solutions. In addition, stress transfer to silicon and its impact to low-k ILD integrity were reduced, by the use of novel interconnect designs. Small form-factor packaging challenges were overcome by a series of innovative materials and process changes to achieve a reduction in form factor while meeting the technology reliability goals. This enables Intel's continuing leadership in thin and light notebooks and smart phone devices.

ACKNOWLEDGMENTS

We acknowledge the materials, process and platform teams that supported the 45nm packaging technology and small form-factor packaging.

REFERENCES

[1]"Introducing the 45nm next generation Intel core microarchitecture." Appeared in print on http://www.intel.com/technology/architecture-silicon/intel64/

[2] Intel Press Release http://www.intel.com/consumer/learn/mids/centrino-atom-detail.htm?iid=learn_mids+whatsnew

[3] Mark Bohr "Intel's 65nm logic technology." *65 nm Press Release* August 2004.

[4] B. Chandran, R. Mahajan, M. Bohr, and Q. Vu, "The Mechanical Side of Ultra-low k: Can it Take the Strain?" *FUTURE FAB International*, 2004, pp. 121–124.

[5] K. Mistry et al., "A 45nm logic technology with high K+Metal gate transistors, Strained Silicon, 9 Cu Interconnect layers, 193nm dry patterning and 100% Pb-free packaging." *IEDM Technical Digest*, 2007, pp. 247–250.

[6] M.K. Schweibert and W. L.K. Leong, "Underfill flow as viscous flow between parallel plates driven by capillary action." *IEEE transactions on CPMT - Part C. Volume 19, No. 2*, 1996.

AUTHORS' BIOGRAPHIES

Neha Patel is a Senior Materials Engineer in the Assembly Test Technology Development (ATTD) Group. She has been with Intel for three years and has worked extensively in the area of first-level interconnect polymer materials and thermal interface materials. She obtained her B.A. degree in Physics from St. Xaviers College, Bombay and M.S. and Ph.D. degrees in Physics from Case Western Reserve University, Cleveland, OH. Her e-mail is neha.m.patel at intel.com.

Vijay Wakharkar is a Principal Engineer in the ATTD Group and has worked at Intel for 17 years. Vijay received his B.A. degree from the College of Engineering, Poona in 1982 and his Ph.D. degree in Materials Science and Engineering in 1988. Prior to joining Intel, he spent two years as a Post Doctoral Fellow at IBM Almaden Research Center in the Polymer Materials Division. Currently, Vijay is in ATTD's Path-finding Group working on the 1272 Package path-finding programs. Prior to joining the ATTD Path-finding Group in March this year, Vijay was responsible for managing the development and pathfinding of Materials Technologies that end up in ATTD's packaging platforms. He is one of the founding editors of the *Intel Assembly Technology and Test Journal* (IATTJ) and also serves on Intel's Thermo-mechanical Patent Committee. His e-mail is vijay.s.wakharkar at intel.com.

Sairam Agrahram is a Group Leader of the Technology Integration Group in ATTD. He has been with Intel for seven years has worked extensively on identifying, understanding, and resolving die package interactions and assembly integration issues. He obtained his bachelors degree in Chemical Engineering from Osmania University, India and a Ph.D. degree in Chemical Engineering from Georgia Institute of Technology, Atlanta, USA.

Nitin Deshpande is a Senior Process Engineer in the ATTD Group. He has been with Intel for seven years and has worked extensively in the area of process engineering on chip-attach and material-handling processes. He obtained his Bachelors degree in Mechanical Engineering from the Indian Institute of Technology, Bombay and an M.S. degree in Aeronautics from the California Institute of Technology, Pasadena, CA. His e-mail is nitin.a.deshpande at intel.com.

Mengzhi Pang is a Staff Packaging Engineer in the Materials Technology Development Group and has been with Intel for over three years. She has worked extensively in the areas of lead-free solder metallurgy and solder bumping technology of first-level-interconnects. Mengzhi received her Ph.D. degree in Materials Science from Washington State University and then spent three years as a postdoctoral researcher in Materials Science at Cornell University. She has more than ten US patents awarded/filed and more than ten peer-reviewed journal publications. Her e-mail is mengzhi.pang at intel.com.

Ravindra Tanikella received his B. Tech. degree in Chemical Engineering from Osmania University, India, and his Ph.D. degree in Chemical Engineering from the Georgia Institute of Technology. He is currently employed in the ATTD Group at Intel Corporation in Chandler, Arizona. His research interests include advanced materials and processes for microelectronics packaging. His e-mail is ravindra.v.tanikella at intel.com.

Rahul Manepalli is a Technology Development Manager in the ATTD Group. He currently manages the materials development efforts for large form-factor, flip-chip packages (server/graphics products), and he also manages the assembly materials path-finding team. Rahul has been with ATTD for over eight years and has worked extensively in development, integration, and certification of several generations of flip-chip and stacked die-packaging materials. Rahul obtained his B. Tech. degree in Chemical Engineering from Osmania University, Hyderabad, India and his Ph.D. degree in Chemical Engineering from the Georgia Institute of Technology, Atlanta, GA. His e-mail is rahul.n.manepalli at intel.com.

Pat Stover is the Ultra-Small Form Factor Packaging Program Manager in the ATTD Group, focusing on assembly packaging, process, and material development for small, fine-pitch FC-BGA products. Pat has been with Intel, and ATTD, for twelve years where he has worked in many process engineering and technology development project management roles. Pat obtained his Bachelors

degree in Chemical Engineering from the University of Illinois in 1996. His e-mail is patrick.n.stover at intel.com.

James D. Jackson has worked at Intel for 19 years in pathfinding, development and high-volume manufacturing across fab, packaging, and boards. He is currently working in ATTD as a platform-level physical interconnect architect and as the key contact for all Apple packaging issues. His Chemical Engineering degrees are from the University of Texas at Austin and the University of Washington. His e-mail is james.d.jackson at intel.com.

Ravi Mahajan is a Senior Principal Engineer in the ATTD Pathfinding Group responsible for packaging technology envelopes for the 22nm technology node. He also manages university research for the ATTD Group. Ravi has a Ph.D. degree from Lehigh University, an M.S. degree from the University of Houston, and a B.S. degree from the University of Bombay, all in Mechanical Engineering. He is an ASME Fellow and an IEEE Senior Member. His e-mail is ravi.v.mahajan at intel.com.

Prabhat Tiwari is the 1266 Platform Manager in the ATTD Group. He has extensive experience in the area of copper interconnects and packaging technologies. He has been in Intel for eight years and has worked in PTD and FSM organizations as a backend integration manager. He obtained his Bachelors and Masters degrees from the India Institute of Technology, Kanpur, India and his Ph.D. degree in Materials Science and Engineering from N.C. State University, Raleigh, NC. His e-mail is Prabhat.tiwari at intel.com**.**

THIS PAGE INTENTIONALLY LEFT BLANK

For further information visit:

developer.intel.com/technology/itj/index.htm