12.1

Reliability Challenges for 45nm and Beyond

J.W. McPherson, Ph.D., TI Senior Fellow
Texas Instruments, Inc.
13560 North Central Expressway, MS 366, Dallas, Texas 75243
j-mcpherson@ti.com

ABSTRACT

Scaling, for enhanced performance and cost reduction, has pushed existing CMOS materials much closer to their intrinsic reliability limits. This will require that designers will have to be very careful with: high current densities, voltage overshoots, localized hot spots on the chip, high duty-cycle applications, and high thermal-resistance packaging. In addition to the reliability issues, interconnect RC time-delay will worsen with scaling because Cu resistivity is expected to increase due to surface and grain boundary scattering in very narrow interconnects. Also, the low-k interconnect-dielectric introduction rate has been much slower than ITRS roadmap forecasts.

Categories and Subject Descriptors

B.8.1 [Performance and Reliability]: Reliability, Testing, and Fault-Tolerance.

General Terms: Design, Performance, Reliability

Keywords: CMOS, reliability, scaling, design, materials

1. Introduction

The dominance of CMOS technology over the last three decades has permitted the industry to focus its efforts primarily on scaling and this has led to remarkable CMOS success in terms of performance, functionality, and cost per function. During this same time period, tremendous improvements in chip reliability have also been accomplished through extensive reliability-physics learning and proactive reliability-engineering efforts.[1-3] With nearly 30+ years of continual CMOS scaling, existing CMOS materials have now been pushed to their physical and reliability limits. Presently, at the 65nm node, gate oxide thickness is ~ 1.2nm [4] and with a leakage of ~ 100 a/cm^2 at 1.0V [5]. Interconnect low-k dielectric minimum spacing can be 70-80nm, similar to gate oxide thicknesses ~ 20 years ago; however, these low-k dielectrics are far from the quality of gate oxides in terms of electrical [6] and mechanical strength [7]. This paper proceeds with a look at the design challenges as one continues to scale CMOS, from the front-end to the back-end, with the primary focus on reliability issues.

2. Impact of Scaling on Transistor Performance

The fundamental MOSFET structure is shown in Fig.1. One will note that changes to the structure and materials over the last 20 years could be described as more "evolutionary" rather than "revolutionary". Silicide was added to lower the sheet resistance of the poly electrode and source/drain regions but the silicide process had to be developed under the boundary condition of little/no adverse impact on existing gate oxide or junction quality.

Sidewall spacers were also added to help create lightly-doped drain (LDD) regions to improve transistor robustness to hot-carrier injection (HCI) but, again, the LDD had to be developed under the boundary condition of no adverse impact to transistor.

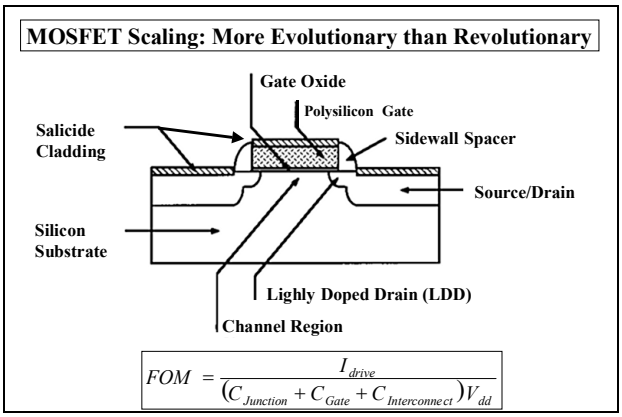

Fig.1 -- Standard MOSFET scaling structure. An expression for relative transistor performance figure-of-merit (FOM) is also shown.

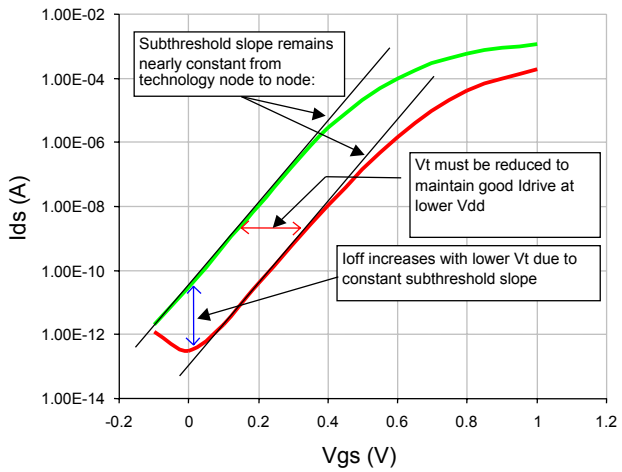

Fig. 2 -- Transistor I_{drive} versus I_{off} tradeoffs.

In order to gain higher levels of circuit performance, scaling the transistor to smaller gate lengths (in order to reduce gate-delay) is required from technology node to node. Since the power supply voltage (Vdd) has also tended to reduce from node to node, improvements in transistor drive-current (I_{drive}) are required with each new technology node. This has been accomplished primarily

DAC 2006, July 24–28, 2006, San Francisco, California, USA.

by using thinner gate oxides and lower V_t MOSFETs. Since the sub-threshold current for the typical MOSFET has remained roughly constant (~100 mV/decade-of-current) from node to node [see Fig. 2], this generally translates to higher transistor off-state leakage (I_{off}). I_{off} increases serve to reduce I_{on}/I_{off} operating margin, increases circuit-level standby power, and reduces the margin for I_{ddq} -type defect detection.

Strained silicon is presently touted as a process enhancement for increasing the I_{on} of the transistor, with little/no increase in I_{off}.[2] By stressing n-channel devices in a tensile mode and p-channel devices in a compressive mode, I_{on} improvements of ~ 40% can be achieved, as is shown in Fig. 3.

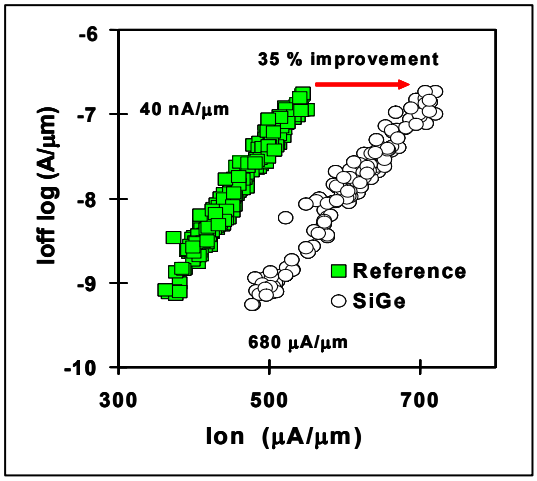

Fig. 3 -- Strain can improve the I_{on} without a degradation in I_{off}

The fact that the basic MOSFET structure has not changed dramatically during the last 20 years has thus facilitated scaling and allowed for continuous improvements in both device and reliability performance. Scaling has been primarily limited only by the industry's ability to pattern and etch as opposed to having to develop radically different materials with radically different properties and interfaces which could impact MOSFET reliability.

3. Impact of Scaling on Gate-Oxide

Today, at the 65nm technology node, gate oxide thickness is ~ 1.2nm. [4] Without nitridation of the gate oxide, the leakage current at 1-volt can be ~ 1000 a/cm^2 [5]. With nitridation, the leakage can be reduced by about an order of magnitude (see Fig. 4). At 1.2nm physical thickness (only a few monolayers of Si-O bonds), defect density can be a concern. From the processing side, extensive TDDB testing is required with hard breakdown, soft-breakdown and stress-induced leakage current precisely recorded. From the design side, over-voltage conditions in designs and use conditions must be avoided, or their impact on reliability fully comprehended. Assuming that the high gate leakage can be tolerated, scaling the electrical-equivalent gate oxide thickness below 1.2nm, will likely require higher levels of nitrogen in the film so as to keep a reasonable physical thickness for acceptable defect density.

If the high gate leakage cannot be tolerated, e.g., portable designs where battery lifetime is a design priority, then high-k gate dielectrics may be needed. Shown in Fig. 5 is the reduced leakage that can be achieved with high-k gate dielectrics while keeping a very low electrical-equivalent oxide thickness (EOT). [5] The EOT for a high-k film is given by:

$$EOT = \left(\frac{k_{SiO_2}}{k_{high-k}}\right) \bullet (thickness)_{high-k} \quad (2)$$

One of the leading high-k gate dielectrics is HfSiON [5], and the reduced leakage is clearly shown in Fig. 5.

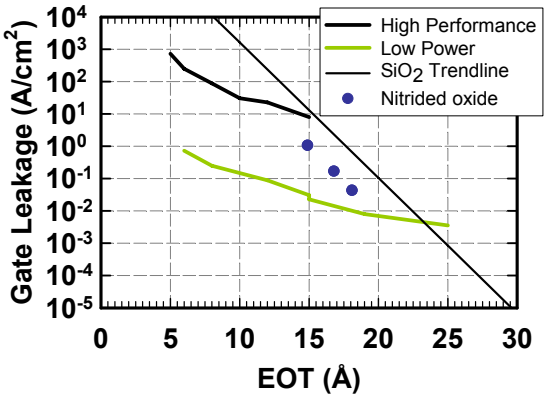

Fig. 4 -- Silica gate-dielectric leakage with dielectric scaling.

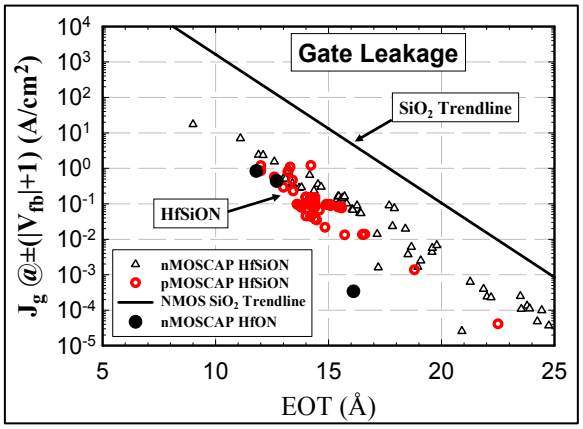

Fig. 5 -- Reduced leakage with HfSiON gate dielectric.

4. Impact of Scaling on NBTI

The electrical impact of negative-bias temperature instability (NBTI) on p-MOSFETs is well established. [8] NBTI causes a shift in the p-channel threshold voltage and a decrease in the mobility of the inversion channel. Apparently, the Si-H bonds (which are needed for interface-state passivation) can become broken under use conditions (field, current, temperature). The key features of NBTI-induced p-MOS degradation are illustrated in Fig. 6.

Even though NBTI is an old mechanism, it is more important today because: electric fields in the gate oxide are higher, devices are operating at higher temperatures due to higher power dissipation, and voltage headroom (difference between V_{Gate} and V_T) is much smaller than in the past. As is shown in Fig. 7, the NBTI impact on the ring oscillators is most easily observed at the lower operating voltages.[8] Thus, the NBTI impact on Vmin circuit operation can be quite significant and an important reliability concern.

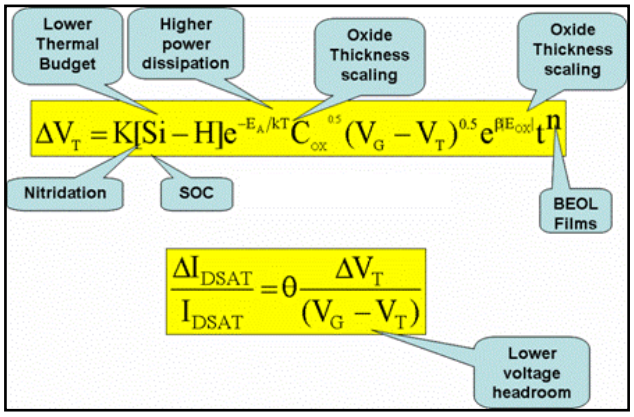

Fig. 6 -- Key Features of NBTI-Induced Degradation

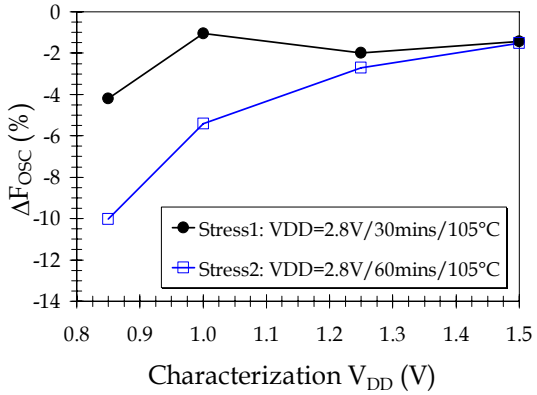

Fig. 7-- Impact of NBTI on ring oscillator is very evident at lower voltages and thus tends to impact Vmin circuit operation significantly.

5. Impact of Scaling on HCI

Hot-carrier injection into the gate oxide (like NBTI) is certainly not a new issue.[9] Carriers, as they are accelerated along the channel can become energetic enough that, through scattering and/or impact ionization, can be injected into the gate oxide causing interface-state generation. Strong factors helping to reduce HCI effects include: LDD implants and voltage reduction with scaling, as is shown in Fig. 8. However, one will also note that for gate lengths less than 50nm, no significant reduction in gate voltage is planned for additional scaling. This will certainly drive an increased emphasis on HCI effects as one continues to scale.

6. Impact of Scaling on Silicided Junctions

Scaling has forced many changes in the silicides used for simultaneous silicidation of poly and shallow n+ and p+ source/drain diffusions: $TiSi_2$ was used at the 250nm technology node, $COSi_2$ at the 180 and 130nm technology nodes, and NiSi has been the favorite for 90nm and below. A monosilicide such as NiSi consumes less of the silicon in the sallow junctions (see Fig. 9) and generally translates to less junction leakage. [10]

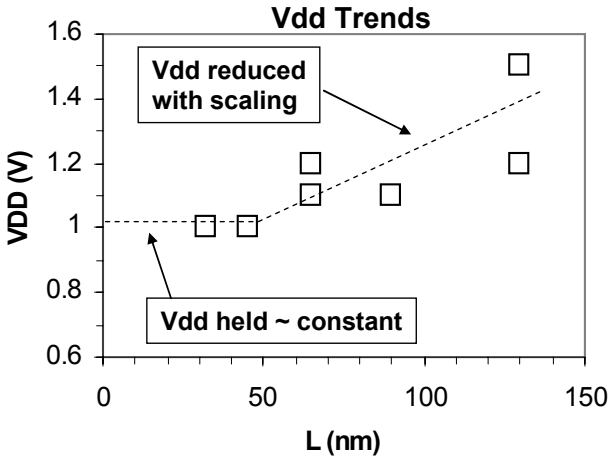

Fig. 8 -- As the gate length (L) was reduced, generally there was a corresponding reduction in operating voltage (Vdd). However, little/no reduction in gate voltage is expected for continued scaling below 50nm.

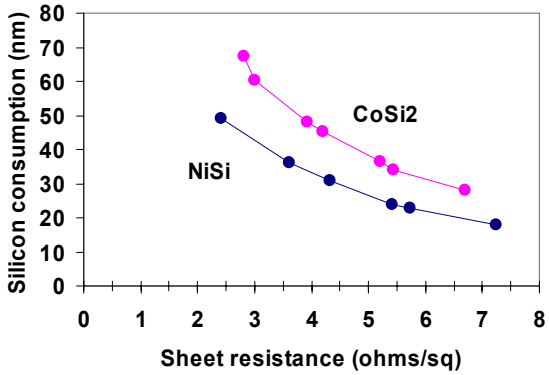

Fig. 9 -- A monosilicide such as NiSi consumes less of the silicon junctions (versus $CoSi_2$) and generally translates to less junction leakage.

As we continue to scale further, the choice of silicides is rather limited. While silicides such as Pd_2Si do consume less silicon than NiSi, there is little/no evidence of a resistivity advantage as is shown in Table 1.

Table 1: Selected silicide resistivities

Silicide :	NiSi	PtSi	Pd_2Si
Resistivity: (μohm-cm)	20-30	25-35	25-35

7. Impact of scaling on ESD

The electrostatic discharge (ESD) protection for CMOS devices seems to never get easier with scaling. Generally scaling has introduced either new materials or weaknesses in the old materials. Fig.10 indicates a general downward trend in ESD robustness with scaling.

Feature Size	IC Process Parameter	Impact on ESD robustness	
<1 μm	Silicide	↓	Poor thermal resistance
<0.25 um	Leff	↓	Local channel heating
65-90 nm	Tox <40 A	↓	Oxide stress
32 nm	FinFET SOI	↓	Metal current density Ch. Self-heating

Fig. 10 – Technology advancements can have an adverse impact on ESD design for CMOS circuits.

8. Impact of Scaling on RC Time Delay

Interconnect RC time delay is becoming increasingly important with scaling. Shown in Fig.11 is a fully embedded Cu lead (with a low-k_1 intra-level dielectric and a low-k_2 inter-level dielectric). The estimated RC time-delay for this fully embedded Cu lead is given by

$$RC = \frac{2\rho_{Cu}k_1\varepsilon_o L^2}{w \bullet s}\left[1 + \frac{k_2}{k_1}\left(\frac{w \bullet s}{t \bullet t_2}\right)\right] \qquad (1)$$

As for continued scaling, the Cu resistivity ρ_{Cu} is expected to increase due to interfacial and grain boundary scattering. While lower-k dielectrics hold some promise for further RC reduction, their introduction rate into CMOS technologies has been much slower than the ITRS roadmap predictions (see Fig 12) due to both mechanical and electrical weaknesses. [6,7]

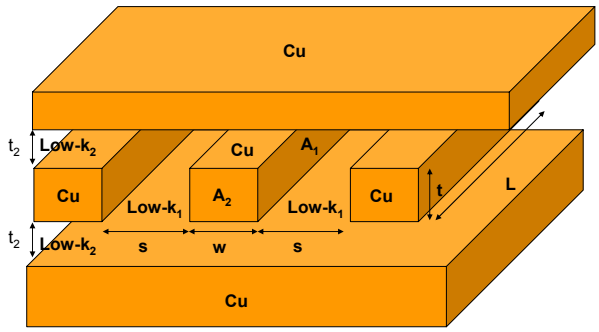

Fig. 11 -- Fully embedded Cu lead with low-k_1 intra-level dielectric and low-k_2 inter-metal dielectric.

9. Impact of Scaling on Electromigration

Cu electromigration(EM) transport generally occurs along interfaces with the Cu / capping-layer interface generally being the expressway for greatest mass transport.[11] For the same current density, the EM performance is expected to degrade with scaling due to the relative increase in Cu-interface area versus volume. Furthermore, the lower-modulus low-k dielectrics will serve to create less back-flow stress (reduced Blech effect) when the Cu tries to accumulate toward the anode. Some type of Cu surface-cladding (as illustrated in Fig. 13) will likely be required.[11]

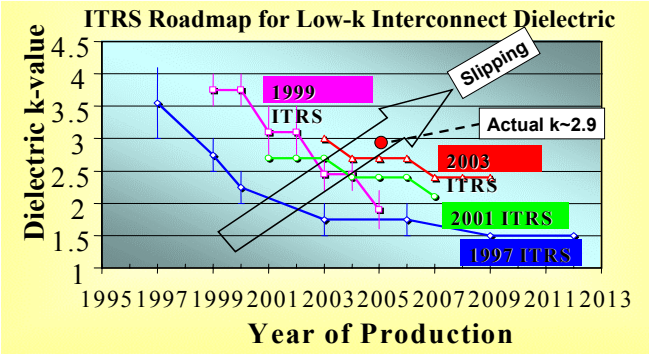

Fig. 12 -- According to 1997 ITRS Roadmap, 90 and 65nm technology in 2006 should be using low-k = 1.8. In reality, the industry is generally using low-k ~ 2.9 in 2006. Obviously, the introduction rate for new low-k materials has been slipping.

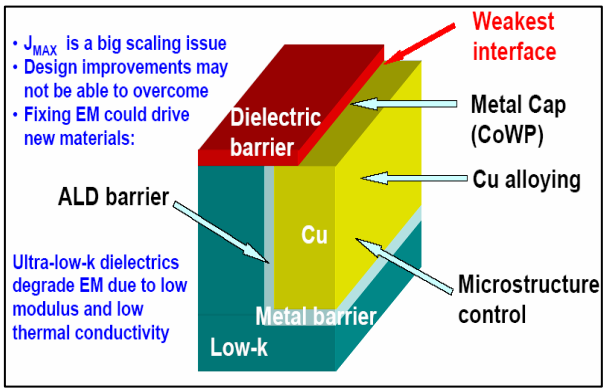

Fig. 13 -- Cu surface-cladding with CoWP can have a very positive impact on Cu electromigration performance.

10. Impact of Scaling on Stress Migration

Stress Migration (SM), in fully constrained damascene Cu, can result in void formation under and in vias for advanced interconnect systems and can be accelerated by high temperature baking (150-200°C). [12] Generally, this has primarily been an issue when a single minimum-size via is contacted to a wide Cu lead. However, more recently, SM has also been reported for narrow metal leads.[13] With continued scaling, the via-voiding with narrow metal leads could become an issue simply because fewer vacancies will be required to cause an unacceptable via resistance rise. The Cu-cladding, as is shown in Fig. 13, might also be beneficial for SM.

11. Impact of Scaling on Low-k Dielectrics

Presently, at the 65nm technology node, nominal minimum metal-widths are ~ 100nm with similar adjacent intra-metal spacing. With process variations, the spacing can easily be ~ 70-80nm, approximately gate oxide thickness some 20 years ago. However, the mechanical strength [7] (as shown in Fig.14) and electrical breakdown strength [6] (as shown in Fig. 15) are much lower than for silica gate-dielectric. For silica-based low-k interconnect dielectrics, the dielectric constant k is normally reduced by replacing normal Si-O network bonding with terminated Si-CH_3 bonds. This replacement has a two fold impact on reducing k: the electronic polarizability contribution of the oxygen-ion is reduced as well as the Si-O dipolar contribution. However, this Si-CH_3 terminated

bond reduces the mechanical strength of the amorphous network is shown in Fig. 14.

The impact on silica electrical-breakdown strength (Ebd) and on time-dependent dielectric breakdown (TDDB) due to the replacement of Si-O bonds with Si-CH_3 bonds is also shown in Fig. 15. We see that both the Ebd and TDDB tend to reduce in low-k silica-based materials as the dielectric constant decreases. The reduction in both the mechanical strength and electrical strength raises obvious concerns about the reliability of low-k film scalability.

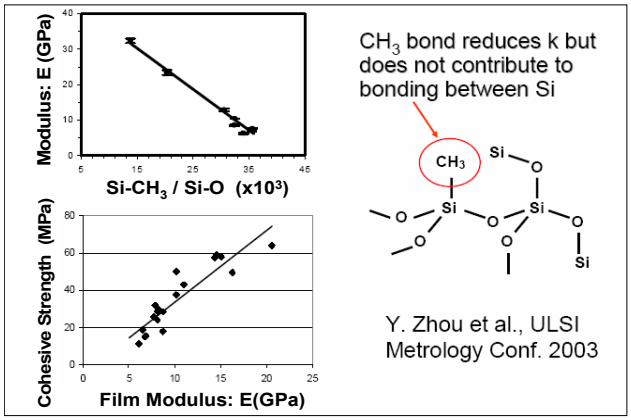

Fig. 14 -- For silica-based low-k dielectrics, modulus and cohesive strength tend to decrease with higher concentration of Si-CH_3 bonding.

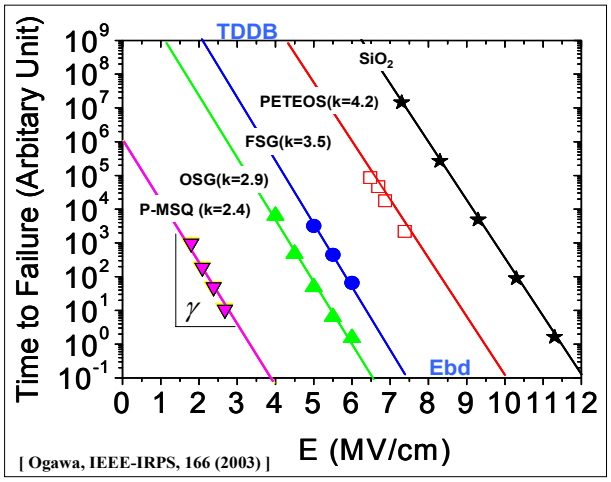

Fig. 15 -- Trends in silica-based low-k dielectric-breakdown strength (Ebd) and time-dependent dielectric breakdown (TDDB).

12. Impact of Scaling on Joule Heating

With scaling, the impact of Joule-heating in metal leads will become even more of an issue because the low-k dielectric films, generally with higher degree of porosity, will have poorer thermal conductivity characteristics as is shown in Table 2.

Table 2: Low-k Thermal Conductivities

Material	Dielectric Constant	Thermal Conductivity (mW/°C-cm)
PE-TEOS	4.2	~ 12
FSG	3.6	~ 8
OSG	2.8	~ 5

In Fig.16 is shown the Joule heating impact in metal leads due to the use of low-k interconnect dielectrics. The temperature rise is most severe for higher levels of metal.

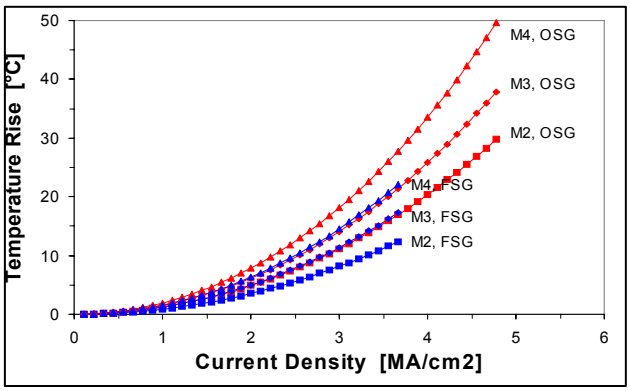

Fig. 16 -- Impact of low-k on Joule heating versus metal level.

13. Impact of Scaling on Defectivity

As illustrated in Fig. 17, for many years, the industry has been printing feature sizes smaller than the wavelength of light used for pattern exposure. This was accomplished by significant "tricks" used to make this happen: attenuated phase shift, model-based optical proximity correction (OPC), restricted design rules, etc. [2] However, with the tricks, the feature size may not be exactly as drawn (see Fig. 18). Immersion lithography is expected to be the next significant photolithography trick. Here the index of refraction of air is replaced with higher index such as water or oil.

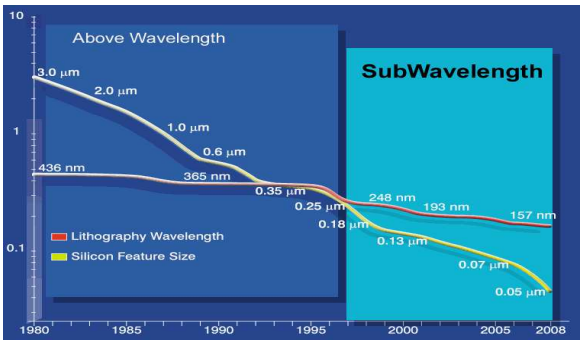

Fig. 17 -- Beginning ~ mid 90's, the industry started printing sub wavelength feature sizes.

With continued scaling, killing defects will become even smaller in size. In Fig. 19, we show a "killing" metal-defect (historically taken as one-half the metal space).[2] We also show this killing metal-defect both at the 130nm node and at the 32nm node. One can see that the killing defect is becoming vanishingly small. In fact, shown in Fig. 20 is the impact of increasing the time-window between Cu CMP-clean and dielectric capping-layer deposition on

the Cu. One can easily see that the dielectric strength of the low-k material degrades, even though no visible defects can be observed.

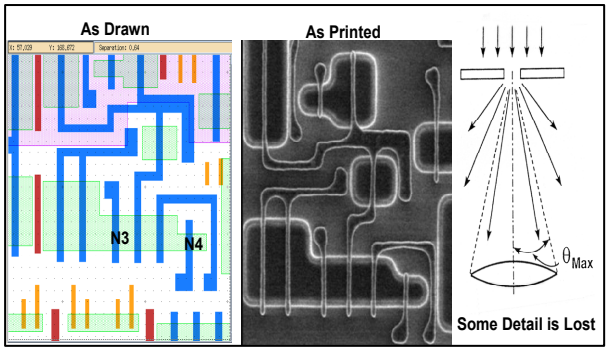

Fig. 18 -- Photolithography issues with continued scaling. What you draw may not be exactly what you get.

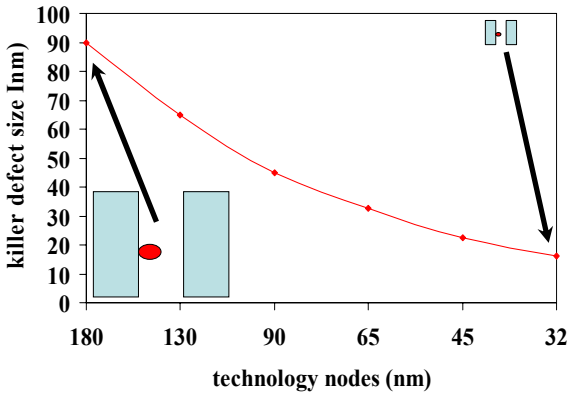

Fig 19 -- Killing metal-defect (taken as one-half the metal space) shown at both the 130nm and 32nm technology nodes.

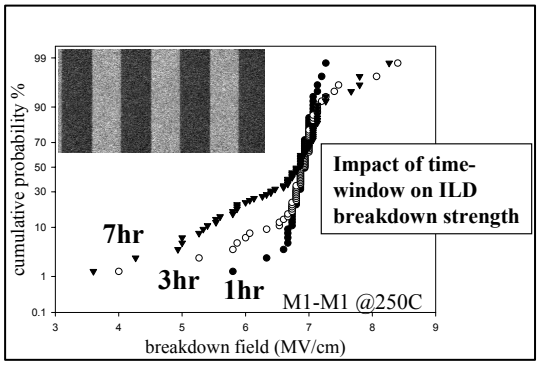

Fig. 20 -- Simply increasing the time-window between Cu CMP clean and capping-layer deposition can decrease the low-k breakdown strength with no visible defects observed.

14. Conclusions

Moore's law has benefited greatly from the fact that the basic CMOS materials have not undergone revolutionary changes during the many years of scaling. This has facilitated continual scaling and a lowering of chip failure rates even though device complexity has increased dramatically. However, the industry is now undergoing a shift in the basic CMOS materials used (e.g., high-k gate dielectrics, metal gates, ultra-low interconnect dielectrics, etc.) and will no longer enjoy the more than two decades of scaling and reliability experience with the older materials. Whether the introduction of these new materials will produce a discontinuity in the downward trend for chip failure rates will depend largely upon our ability to accelerate our reliability physics understanding for these new materials and to quickly develop high confidence-level design rules for them. This will mean that all legacy-related reliability design rules (based on the older materials) must be thoroughly reviewed, questioned and reevaluated when using the newer materials.

15. Acknowledgments

Many thanks to many people: Ennis Ogawa, Ajit Shanware, Vijay Reddy, Charvaka Duvvury, Honglin Guo, Homi Mogul, Gaddi Haase, Jinyoung Kim, Anand Krishnan, Srikanth Krishnan, Randy McKee, and Nagaraj Sarvithri.

16. References

[1] J. McPherson, "Scaling-induced reductions in CMOS reliability margins" IEEE-ISQED Proceedings, 123 (2001).

[2] J. McPherson, "Yield and reliability challenges for 32nm and beyond", IEEE-IEDM Shortcourse (2005).

[3] J. McPherson, "Reliability challenges associated with interconnect scaling, MRS Tutorial (2006).

[4] Y-H Lee "Prediction of logic product failure due to thin gate oxide breakdown", IEEE-IRPS Proceedings, 18 (2006).

[5] A. Rotondaro, et al., "Advanced CMOS transistors with novel HfSiON gate dielectric", VLSI Technology Digest of Technical Papers, 148 (2002).

[6] E. Ogawa, et al., "Leakage, breakdown and TDDB characteristics of porous low-k silica", IEEE-IRPS Proceedings, 166 (2003)

[7] Y. Zhou, et al., "Correlation of surface and film chemistry with mechanical properties". ULSI Metrology Conf., (2003).

[8] V. Reddy, et al., " Impact of NBTI instability on digital circuit reliability", IEEE-IRPS Proceedings, 248 (2002).

[9] E. Takeda, et al., "AC hot-carrier effects in scaled MOS devices", IEEE-IRPS Proceedings, 118 (1991).

[10] A. Lauwers, et al., "Formation of NiSi-silicide pn shallow junctions" J. Vac. Sci. & Tech. B, 2026 (2001).

[11] C-K Hu, "Effects of overlayers on electromigration reliability improvement for Cu/Low-k", IEEE-IRPS Proceedings, 222 (2004).

[12] E. Ogawa, et al., "Sress-induced voiding under vias connected to wide Cu leads", IEEE-IRPS Proceedings, 312 (2002).

[13] T. Kouno, et al., "Stress-induce voiding under vias connected to narrow Cu lines", IEEE-IEDM Tech. Digest, 195 (2005).