Leakage Current in Low Standby Power and High Performance Devices: Trends and Challenges

(Invited Paper)

Geoffrey C-F Yeap
Motorola Inc., DigitalDNA Laboratories,
3501 Ed Bluestein Blvd., MD: K10,
Austin, TX 78721
1-512-933-8093
geoffrey.yeap@motorola.com

ABSTRACT

IC technology is continuing to scale according to Moore's Law, with the overall chip circuit requirements driving the MOSFET device and process integration requirements and optimal choices. In the 2001 International Technology Roadmap for Semiconductors (ITRS) [1] the driver for the high performance logic is maximizing MOSFET intrinsic speed, while the driver for low standby power logic is minimizing MOSFET leakage current. Total leakage current of a MOSFET consists of three components: off-state sub-threshold leakage current, gate direct tunneling leakage current and source/drain junction leakage current. In this paper, trends and challenges for each leakage current component in low standby power and high performance devices are discussed from the perspective of the 2001 ITRS and recently reported literatures.

Categories and Subject Descriptors

B.7.1 [**Integrated Circuits**]: Types and Design Styles–*advanced technologies, VLSI (very large scale integration).*

General Terms: Performance, Design and Reliability.

Keywords: Leakage current, off-state sub-threshold leakage, gate tunneling leakage, low standby power, high performance, CMOS technology, system-on-a-ship (SoC).

ISPD'02, April 7-10, 2002, San Diego, California, USA.

1. INTRODUCTION

In 1960, Kahng and Atalla [2] announced the "silicon dioxide field surface device" which is often regarded as the forerunner of the modern MOSFET. Today, silicon MOSFET-based integrated circuits have become the dominant technology of the semiconductor industry. There are literally thousands of MOS-transistor circuits in production today, ranging from simple logic gates to complex system-on-a-chip (SoC) designs using logic, memory and signal processing functions on the same silicon chip. Since the mid-1960s, the history and progress in MOS-IC technology has been dominated by the scaling of the device feature size. Within a period of 35 years, the minimum feature size in a MOS technology generation, following Moore's Law [3], has been reduced by a factor of 200, from about 20 to 0.1 μm. For both memory and logic chips, the result has been exponential increases in speed and functional density versus time combined with exponential decreases in power dissipation and cost per function versus time. In particular, the number of logic transistors and memory bits per chip have been quadrupling every three to four years, while the speed of microprocessors, for example, has been more than doubling every three years, increasing from about 2 MHz for the Intel® 8080 in the mid-1970's to well over 1 GHz for current leading-edge chips. Design rule scaling has led to reductions in the design rules from about 8 um in 1972 to the current 90 nm node leading-edge technology [4-7]. This works out to a reduction by a factor of about 0.87 per year, or by a factor of ~ 0.7 in a time interval of between two and three years. The IC industry is running into increasing difficulties in continuing to scale at this rate, owing to the small dimensions in current IC technology and certain key device, material, and process limits that are approaching.

The International Technology Roadmap for Semi-conductors (ITRS) projects the progression of the CMOS technology over the next 15 years deep into the sub-100nm generations as shown in Table 1. There are two main application categories for logic chips: high

Table 1 High Performance (HP) and Low Standby Power (LSP) Logic Technology Requirements. Data from 2001 ITRS [1]

		Near Term							Long Term		
Calendar Year		2001	2002	2003	2004	2005	2006	2007	2010	2013	2016
Technology Node		130nm			90nm			65nm	45nm	32nm	22nm
HP Physical Gate Length	nm	65	53	45	37	32	28	25	18	13	9
Nominal HP Power Supply	V	1.2	1.1	1.0	1.0	0.9	0.9	0.7	0.6	0.5	0.4
HP Equi. Oxide Thickness	nm	1.3-1.6	1.2-1.5	1.1-1.6	0.9-1.4	0.8-1.3	0.7-1.2	0.6-1.1	0.5-0.8	0.4-0.6	0.4-0.5
Nominal HP Ioff	µA/µm	0.01	0.03	0.07	0.1	0.3	0.7	1	3	7	10
Nominal HP Idsat	µA/µm	900	900	900	900	900	900	900	1200	1500	1500
HP NMOS Gate Delay	ps	1.65	1.35	1.13	0.99	0.83	0.76	0.68	0.39	0.22	0.15
LSP Physical Gate Length	nm	90	80	65	53	45	37	32	22	16	11
Nom. LSP Power Supply	V	1.2	1.2	1.2	1.2	1.2	1.2	1.1	1	0.9	0.9
LSP Equi. Oxide Thickness	nm	2.4-2.8	2.2-2.6	2.0-2.4	1.8-2.2	1.6-2.0	1.4-1.8	1.2-1.6	0.9-1.3	0.8-1.2	0.7-1.1
Nominal LSP Ioff	pA/µm	1	1	1	1	1	1	1	3	7	10
Nominal LSP Idsat	µA/µm	300	300	300	300	400	400	400	500	600	700
LSP NMOS Gate Delay	ps	4.61	4.41	3.95	3.57	2.51	2.32	2.26	1.43	0.91	0.66

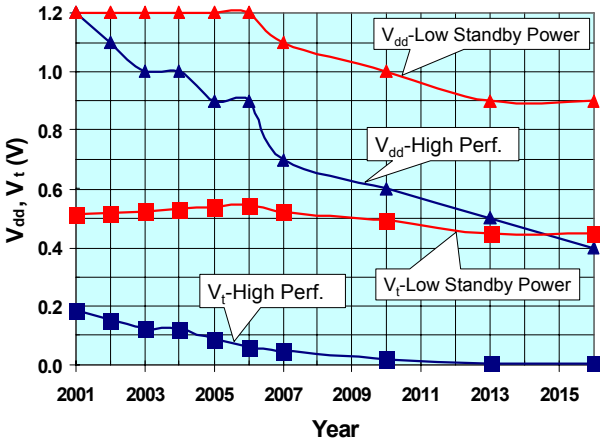

Figure 1. 2001 ITRS projections of Vdd and Vt Scaling

performance (HP) and low standby power (LSP) logic. Since each of these application areas has different overall chip requirements, the scaling goals are different. HP logic is used mainly for high-end microprocessor and network processors, where the main goal is maximum chip speed. The scaling and the device design are aimed at maximizing transistor speed, with a trade-off of relatively high MOSFET off-state leakage current. In contrast, LSP logic is used mainly in mobile systems, where the main goal is preserving battery life by minimizing the chip power dissipation, particularly the static power dissipation. Hence, the device design and scaling are aimed at minimizing the MOSFET off-state leakage current, with a tradeoff of reduced MOSFET speed. Figure 1 shows the 2001 ITRS projections of Vdd and Vt scaling for both HP and LSP devices.

In this paper, the trends and challenges of leakage current, which consists of off-state sub-threshold leakage, gate direct tunneling leakage and source/drain junction leakage currents, in HP and LSP devices will be the discussed. The discussion is limited to logic technology since many of the key technology issues with scaling are driven by the logic technology.

2. Trend and Challenges of Off-state Sub-threshold Leakage Current

Figure 2 shows 2001 ITRS projected scaling of MOSFET frequency (f_i) and off-state sub-threshold leakage ($I_{off\text{-}sub}$) for HP and LSP devices. For HP logic, the target is an average 17% per year increase in frequency, to match the historic rate of improvement in device performance. Ion

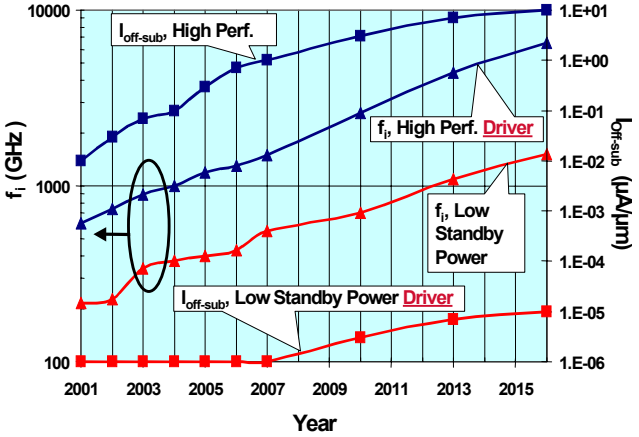

Figure 2. 2001 ITRS projected scaling of MOSFET frequency (f_i) and off-state sub-threshold leakage ($I_{off\text{-}sub}$) for high performance and low standby power devices

must stay relatively high to maximize fi. Hence, as shown in the table, the NMOSFET Ion stays constant at 900 μA/μm until the 65 nm technology node in 2007, and then it increases somewhat in the later years. (The PMOSFET Ion is 40 – 50% of the NMOSFET Ion.) Ion is strongly dependent on the overdrive, (Vdd–Vt), where Vt is the effective threshold voltage, and as shown in the table, Vdd is decreasing sharply with device scaling. Hence, Vt must be reduced along with Vdd to keep Ion at the specified values. But (1/Ioff-sub) is exponentially dependent on Vt, where Ioff-sub is the off-state subthreshold leakage current of the MOSFET. Hence, Ioff-sub increases rapidly as Vt decreases, and becomes particularly large, at greater than 1μA/μm, for the long-term years. As a result of the increasing Ioff-sub, the static power dissipation per device increases with device scaling despite the reduction in Vdd and device dimensions with scaling.

For LSP logic, the ITRS sets targets for maximum transistor leakage current aimed at giving reasonable battery life for mobile applications. LSP chips are used for consumer type mobile applications such as cellular telephones, where the performance is lower than HP chips. In the 2001 ITRS, the maximum Ioff-sub for LSP logic starts at 1 pA/μm in 2001 and stays at that level until 2007, after which it rises very slowly. Note that Lg and Tox scale down sharply with succeeding years, but the actual values of Lg lag behind those used in high performance logic by two years (i.e., Lg is 65 nm for high performance logic in 2001, but it is 65 nm for LSP logic in 2003). The resultant average improvement in the LSP device performance, fi, is about 14% per year for LSP. As shown in Figure 2 and Table 1, Ioff-sub, Ion, and fi are lower for LSP, and higher for HP logic, with Ioff-sub for HP logic ranging from four to six orders of magnitude larger than for LSP. One key issue is the relatively slow scaling of Vdd for the LSP transistors. This is because Vt must scale slowly to meet the low targets for Ioff-sub. Vdd must follow Vt in scaling slowly for two reasons: to obtain reasonable device performance the overdrive, (Vdd–Vt), must remain relatively large, and for adequate circuit switching noise margins, Vdd must be larger than at least 2 × Vt. Since dynamic power dissipation is proportional to $(Vdd)^2$, the dynamic power dissipation is higher for LSP and lower for high performance logic. However, the static power dissipation follows the Isub-off values and is larger by far for the HP and smaller for the LSP logic devices. Also, since the lateral electric field ~ (Vdd/Lg), this field increases sharply with device scaling, and this sharp increase will result in difficulty in controlling short channel effects and possibly in significant reliability problems in the long-term years.

As CMOS technology scales down to 90nm technology node and beyond, a new paradigm of device and circuit co-design [8] such as well bias and power down/reduction is required to meet the competing requirements of high performance and low standby/dynamic power. The rapid increase in Ioff-sub with scaling must be dealt with to keep chip power dissipation within tolerable limits so that the low cost plastic package and general ceramic package may be used for SoC products [5]. An increasingly common approach of using multi Vt and multi gate oxide integrated transistors on the same chip was reported in a 90nm node modular CMOS technology platform that enables device and circuit co-design techniques (e.g., well biasing and power down/reduction) for low standby power, high performance, and RF/Analog SoC applications [7]. These multi-Vt and multi gate oxide integrated transistors have various performance levels, off-state leakage differentiation from tens of pA/μm to tens of nA/μm, and Vdd to best meet diverse SoC product requirements on performance, static and dynamic powers. Figure 3 shows that LSP

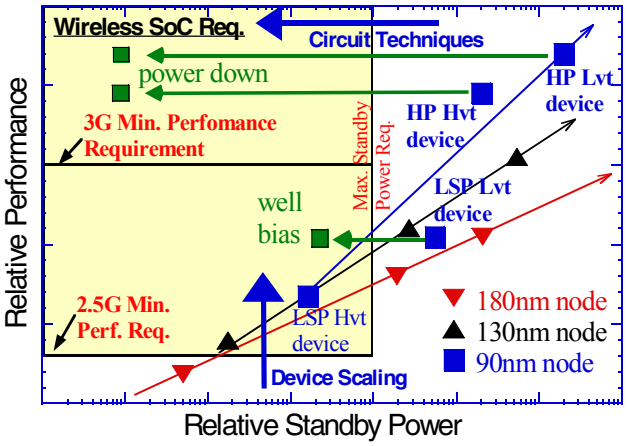

Figure 3. Device scaling and well bias scheme enable LSP Hvt and Lvt transistors to meet 2.5G wireless SoC requirements. The 3G wireless data rich SoC requirement met by power down technique enabled by HP/LSP transistors in the triple gate oxide SoC flow

high Vt devices may support die cost reduction and ~15% performance improvement due to device scaling for a high volume consumer wireless product migrating from 130nm to 90nm technology node. However, further performance improvement with the same standby power requires LSP low Vt devices with higher drive currents while relying on well bias to minimize standby power. HP devices are required to meet the minimum performance requirement for the third generation (3G) data-rich wireless SoC applications. Due to lower body factor, well biasing on these low Vt and high Ioff transistors are not effective (<2x) in Ioff-sub reduction. Thus, low leakage (high Vt) LSP transistors in the triple gate oxide process are needed as cut-off devices to implement power down/reduction circuit techniques to minimize standby power for 3G wireless applications. Hence, a realistic picture of scaled HP/LSP ICs is that the standby power dissipation due to Ioff-sub be

controlled by utilizing more than one type of transistor and by utilizing device/design/architectural techniques.

The need of controlling short channel effects and the push for more drive current through reduced source/drain resistance for aggressively scaled HP and LSP MOSFETs require higher channel doping and heavily doped abrupt source/drain extensions. Figure 4 shows measured gated edge junction leakage vs. channel doping for NMOS device [9]. Source/drain junction leakage may become a significant portion of total Ioff for LSP devices, which have lower Ioff-sub targets than HP devices, for 65nm technology node or beyond.

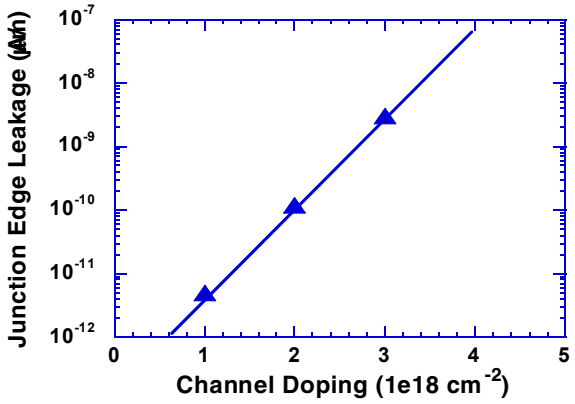

Figure 4. Measured gated edge junction leakage vs. channel doping for MOS devices. Data from [9].

3. Trend and Challenges of Gate Direct Tunneling Leakage Current

Gate direct tunneling leakage current (Jg) isone of the crucial component of the total Ioff. Jg increases exponentially with thickness and Vdd. For every 0.2nm reduction Tox causes 10x increase in Jg. For HP devices,

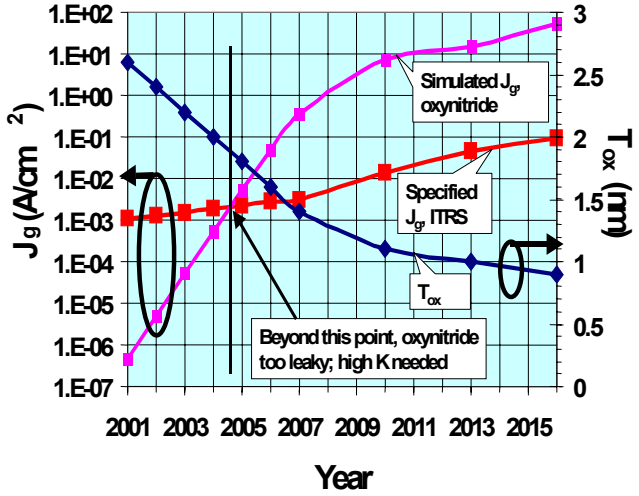

Figure 5. 2001 ITRS projections versus simulations of direct tunneling gate leakage current density for low standby power logic. High K gate dielectric with reduced gate tunneling current is projected to be required in 2005 for LSP devices.

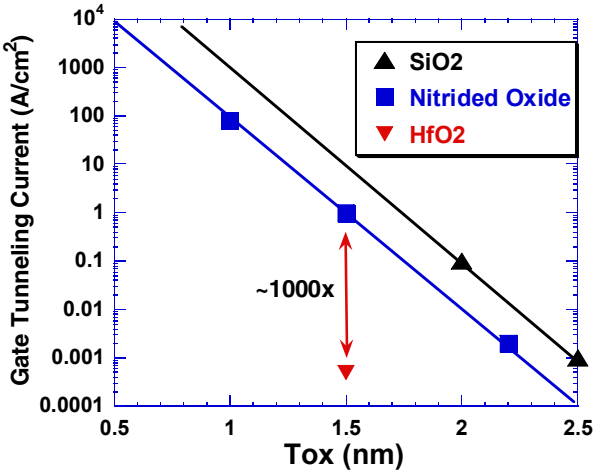

Figure 6. Gate direct tunneling current dependence on equivalent oxide thickness for pure, nitrided and metal oxides.

the 2001 ITRS projected that heavily nitrided oxide will be sufficient to meet the gate tunneling leakage requirements until the end of the roadmap due large increases in the allowable gate leakage current with HP device scaling. However for LSP logic, given the Tox and Vdd in 2005, the maximum gate tunneling leakage current target cannot be met using oxynitride because of direct tunneling. Hence, high K metal oxide gate dielectric will be required in 2005 for LSP devices as shown in Figure 5. Excessive gate tunneling leakage has reported to cause performance degradation due to inversion charge lose [10] and induce significant fluctuations in device characteristics such as threshold voltage and transconductance [11]. Recent study [12] also suggests that static logic is able to tolerate high Jg while dynamic logic and analog circuits are more significantly affected by excessive Jg under low Vdd operation.

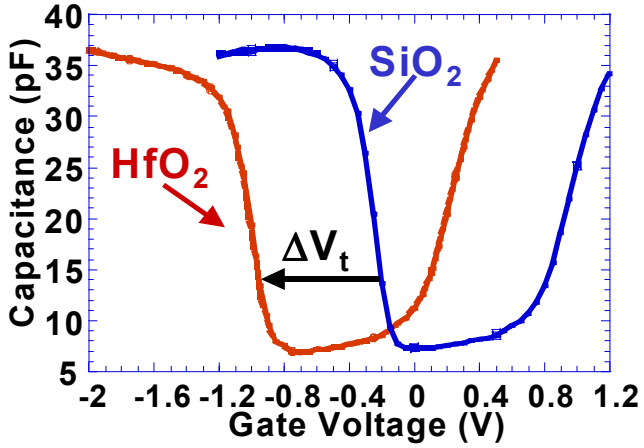

Figure 7. Large Vt shift seen in HfO2 devices as compared to SiO_2 devices even though HfO2 devices yield ~1000x reduction in gate tunneling current [14]

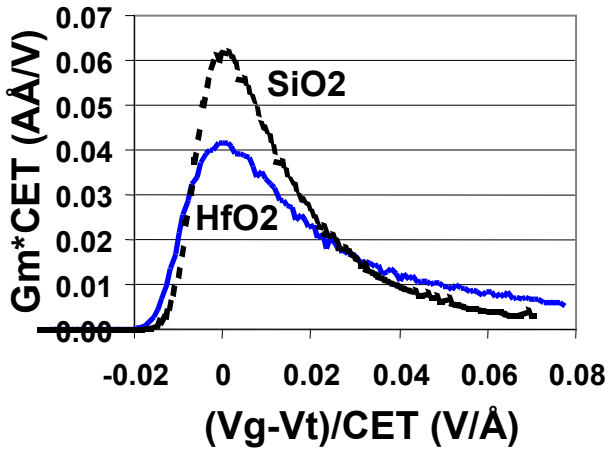

Figure 8. Normalized transconductance of NMOS HfO2 devices is ~70% of the SiO_2 devices [14]

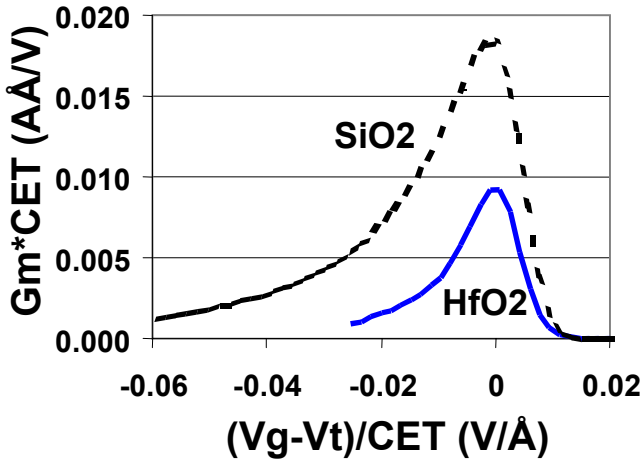

Figure 9. Normalized transconductance of PMOS HfO2 devices is ~50% of the SiO_2 devices [14]

Figure 6 shows gate tunneling current (Jg) dependence on equivalent oxide thickness for pure, nitrided and metal oxide. ~10x reduction in Jg is possible at the same equivalent oxide thickness for optimized nitrided oxides. Extensive research in high K gate dielectrics as a replacement for the IC industry workhouse of pure and nitrided oxides has been conducted in the last five years [13]. Popular high-K gate dielectrics are zirconium dioxide, hafnium dioxide and their silicates. Recent study on 80nm poly-Si gate CMOS with HfO_2 gate dielectrics [14] shows that hafnium dioxide with 1.5nm equivalent oxide thickness is able to yield ~1000x lower gate tunneling current than nitrided and pure oxides as shown in Figure 6. High-K material integration and process need to be optimized to reduce defects such that the same gate tunneling current reduction may be obtained in both small gate area transistors and large gate area transistors.

Even though well behaved transistor characteristics such as inversion Tox, sub-threshold slope and Vt are obtained with high-K metal oxide devices, these devices suffer large Vt shift especially in PMOS devices due to high fixed charge as shown in Figure 7, raising reliability and performance concerns. Mobility degradation is observed on HfO_2 devices. Figure 8 and Figure 9 show that normalized transconductances of HfO_2 NMOS and PMOS are respectively ~70% and ~50% of that of the SiO_2 devices [14]. Much more work is needed to improve high-K material processing, interface layer quality and integration with CMOS flow to eliminate or minimize mobility degradation and high fixed charge. Both NMOS and PMOS transistor mobility and performance may be enhanced with the use of strained silicon independent of device scaling [15].

4. Conclusions

IC technology is continuing to scale according to Moore's Law, with the overall chip circuit requirements driving the MOSFET device and process integration requirements and optimal choices. In the 2001 ITRS the driver for the high performance logic is maximizing MOSFET intrinsic speed, while the driver for low power logic is minimizing MOSFET leakage current. The rapid increase in off state sub-threshold leakage and gate tunneling currents due to device scaling for more performance may be controlled by utilizing multi Vt and multi gate oxide transistors and by utilizing device/design/architectural techniques such as well biasing and power down/reduction. High-K dielectrics has been shown to reduce gate tunneling leakage current by ~1000X; however, much more work is needed to improve high-K material processing, interface layer quality and integration with CMOS flow to eliminate or minimize mobility degradation and high fixed charge.

5. ACKNOWLEDGMENTS

The author acknowledges the data from 2001 ITRS Process Integration, Device, and Structures (PIDS) Technical Working Group, and the discussions with Peter Zeitzoff and Jim Chung. The author also thanks management support from Michael Woo and Bob Yeargain.

6. REFERENCES

[1] Semiconductor Industry Association (SIA), International Roadmap for Semiconductors 2001 edition, Austin, TX: International SEMATECH, 2001, Process Integration, Devices, and Structures Chapter. (This is available for printing and viewing from the internet, with the following URL: http://public.itrs.net.)

[2] D. Kahng and M. M. Atalla, "Silicon-Silicon Dioxide Field Induced Surface Devices", *IRE Solid-State Device Research Conference*, 1960.

[3] G. E. Moore, "Lithography and the Future of Moore's Law," *Proceedings of VIIIth Optical/Microlithography Conference,* SPIE Vol. **2440**, pp. 2 – 17, February 1995.

[4] S. Parihar et al., "A High Density 0.10μm CMOS Technology Using Low K Dielectric and Copper Interconnect," *Tech. Digest of 2001 IEDM*, pp.249 - 252, December 2001.

[5] A. Ono et al., "A 100 nm Node CMOS Technology for Practical SOC Application requirements", *Tech. Digest of 2001 IEDM*, pp. 511 - 514, December 2001.

[6] K. Miyashita et al., "A High Performance 100nm Generation SOC Technology [CMOS IV] for High Density Embedded Memory and Mixed Signal LSIs", *Tech. Digest of 2001 Symposium on VLSI Technology*, 2001.

[7] G. C-F Yeap et al., "A 100nm Copper/Low-K CMOS Technology with Multi Vt and Multi Gate Oxide Integrated Transistors for Low Standby Power, High Performance and RF/Analog System on Chip Applications", *Tech. Digest of 2002 Symposium on VLSI Technology*, 2002.

[8] M. Fukuma et al., "New Frontiers of Sub-100nm VLSI Technology – Moving toward Device and Circuit Co-design", *Tech. Digest of 2000 Symposium on VLSI Technology*, pp. 4-7, 2000.

[9] T. Ghani et al., "Scaling Challenges and Device Design Requirements for High Performance Sub-50nm Gate Length Planar CMOS Transistors", *Tech. Digest of 2000 Symposium on VLSI Technology*, pp. 174-175, 2000.

[10] B. Yu et al., "Limits of Gate Oxide Scaling in Nano-Transitors", *Tech. Digest of 2000 Symposium on VLSI Technology*, pp. 90-91, 2000.

[11] M. Koh et al., "Limit of Gate Oxide Thickness Scaling in MOSFETs due to Apparent Threshold Voltage Fluctuation Induced by Tunnel Leakage Current", *IEEE Trans. Electron Devices*, VOL. 48, NO.12, pp. 2823-2829, Dec. 2001.

[12] C-H Choi et al., "Impact of Gate Direct Tunneling Current on Circuit Performance: A Simulation Study", *IEEE Trans. Electron Devices*, VOL. 48, NO.12, pp. 2823-2829, Dec. 2001.

[13] Please see papers in High K gate dielectric sessions in the *Tech. Digest of IEDM* and *Tech. Digest of Symposium on VLSI Technology* in the last five years.

[14] C. Hobbs et al., "80nm Poly-Si Gate CMOS with HfO2 Gate Dielectric," *Tech. Digest of 2001 IEDM*, pp. 937-939, December 2001.

[15] K. Rim et al., "Characteristics and Device Design of Sub-100nm Strained Si N- and PMOSFETs", *Tech. Digest of 2002 Symposium on VLSI Technology*, 2002.