On the Gate Oxide Scaling of Sub-100nm CMOS Transistors

Seungheon Song, Jihye Yi, Woosik Kim
Kazuyuki Fujihara, Ho-Kyu Kang, Joo-Tae Moon, and Moon-Yong Lee

Abstract — **Gate oxide scaling for sub-100㎚ CMOS devices has been studied. Issues on the gate oxide scaling are reviewed, which are boron penetration, reliability, and direct tunneling leakage currents. Reliability of Sub-2.0㎚ oxides and the device performance degradation due to boron penetration are investigated. Especially, the effect of gate leakage currents on the transistor characteristics is studied. As a result, it is proposed that thinner oxides than previous expectations may be usable as scaling proceeds. Based on the gate oxide thickness optimization process we have established, high performance CMOS transistors of L_{gate}=70㎚ and T_{ox}=1.4㎚ were fabricated, which showed excellent current drives of 860㎂/㎛ (NMOS) and 350㎂/㎛ (PMOS) at I_{off}=10㎁/㎛ and V_{dd}=1.2V, and CV/I of 1.60ps (NMOS) and 3.32ps (PMOS).**

Index Terms — **Gate oxide, CMOS, scaling, gate leakage, boron penetration**

I. Introduction

For higher integration density and higher performance, scaling down the length scale and operation voltage of CMOS is inevitable. Table 1 is a summary of high performance logic technology requirements of SIA's ITRS (International Technology Roadmap for Semiconductor) published in 1999. Based on this prospect, the gate length is reduced by approximately 11% every year and the same drive current density is maintained with reduced operation voltage, which is possible with reduced gate dielectric thickness. Gate and junction capacitances of a transistor are also reduced due to the lateral size reduction even though thinner gate dielectrics are employed, therefore, the gate delay is expected to be reduced by 10% every year.

However, scaling down CMOS devices beyond 100㎚ with satisfactory performance is not easy and many issues should be solved, which are summarized in Table 2. To overcome these issues, virtually all parts of the transistor should be carefully modified and optimized, which includes modifications of process conditions, physical structures, and materials [1] ,[2] ,[7].

In this paper, we will focus on the issues on gate oxide scaling, which are reliability, boron penetration, and gate leakage current. Specifically, we will show experimental results on the reliability of sub-2.0㎚ oxide, the effect of boron penetration and a novel technology to suppress it, and the effect of gate leakage current on the transistor characteristics. Finally, scaling limit of the silicon oxide as a gate dielectric is discussed.

II. Experimental

For the fabrication of the MOS capacitors and the transistors, we employed conventional CMOS technologies and several novel technologies including epitaxial growth. After well and channel implantations, undoped-Si was grown using LPCVD selective epitaxial growth (SEG). For the gate dielectric, oxynitrides (at RTP or furnace) or stack structures of oxynitride and LPCVD nitride were used with T_{ox}=1.3~1.7㎚. To

Manuscript received June 10, 2001; revised June 25, 2001.
The authors are with Samsung Electronics Co., San #24, Nongseo-lee, kiheung- Eup, Yongin-city, Kyungki-Do, Korea.
(e-mail : Shsong1@Samsung.co.kr) Tel : +82-31-209-3878

Table 1. SIA's 1999 ITRS. High performance logic technology requirements.

Year	2001	2003	2005	2008	2011	2014
Technology Generation	150nm	120nm	100nm	70nm	50nm	35nm
Gate Length (nm)	100	80	65	45	32	22
Vdd (V)	1.5	1.5	1.2	0.9	0.6	0.6
Tox (nm)	1.9	1.9	1.5	1.2	0.8	0.6
Nominal Ion (μA/μm)	750/350	750/350	750/350	750/350	750/350	750/350
Maximum Ioff (nA/μm)	8	13	20	40	80	160
Gate Delay CV/I (ps)	8.6	6.9	5.7	3.7	2.6	2.4
Gate Rs (Ω /□)	4-6	4-6	4-6	4-6	4-6	4-6
Ig @ 100℃ (nA/μm)	8	13	20	40	80	160
Active Poly Doping	3.1E20	3.1E20	4.6E20	5.4E20	7.3E20	1.2E21
Drain Ext. Xj (nm)	30-50	24-40	20-33	16-26	11-19	8-13
Drain Ext. Rs (Ω /□)	280-730	240-675	200-625	150-525	120-450	100-400

Table 2. Issues of the CMOS scaling beyond 100nm

Issue	Due to
Short Channel Effect	Reduced channel length with less reduced junction depth and broadness.
Gate Leakage Current	Direct tunneling through ultra-thin oxide.
Vth Fluctuations	Gate length fluctuation. Dopants density fluctuations.
Gate Poly Depletion	Solid solubility limit. Increased vertical field. Boron penetration.
Junction Capacitance	Higher doping and abrupt junction.
Mobility Degradation	Increased channel doping. Increased vertical field. Boron penetration.
Junction Leakage	Shallow junctions with silicide metallization.
S/D Resistance	Shallow junctions.
Gate Sheet Resistance	Narrow gate length.

reduce gate poly depletion, additional implantations were applied before gate patterning. Sub-100nm gate length was defined by KrF photolithography with size reduction using SiON hard mask etching technique. High dose low energy (<3keV) S/D extension implantations and optimized pocket implantations were performed to reduce short channel effect (SCE). NiSi were used for metallization. Fig. 1 is a SEM image of a 65nm transistor fabricated using these processes and Fig. 2 is a TEM image of an 1.4nm oxynitride gate dielectric.

III. Results And Discussion

A. *Issues on Gate Oxide Scaling*

For a scaled transistor, thinner gate dielectric increases the gate capacitance, as a result, increases inversion charge density in the channel at the same or reduced operation voltage, and increases drain current. High gate capacitance also enhances the controllability of the channel charge, therefore, suppresses short channel effect (SCE). For these reasons, gate oxide thickness

scaling is the most important part of the transistor scaling. Fig. 3 shows V_{th} roll-off behavior of transistors with several different values of gate oxide thickness. However, scaling down the thickness of the silicon oxide based gate dielectrics below 2.0㎚ induces several serious issues.

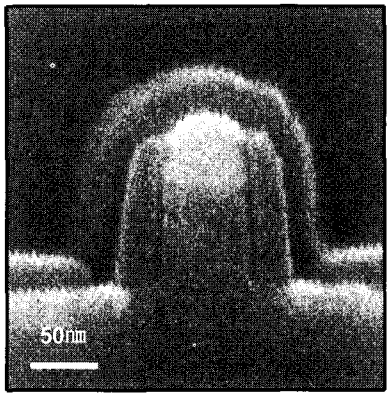


Fig. 1. SEM image of a transistor with L_{gate}=65㎚.

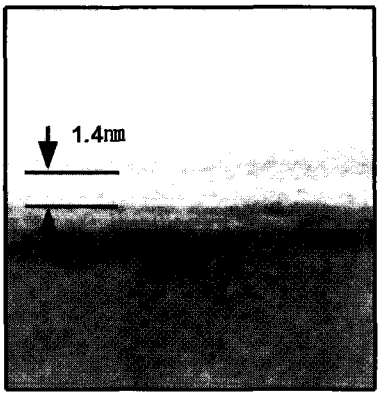


Fig. 2. TEM Image of an oxynitride gate dielectric of 1.4㎚.

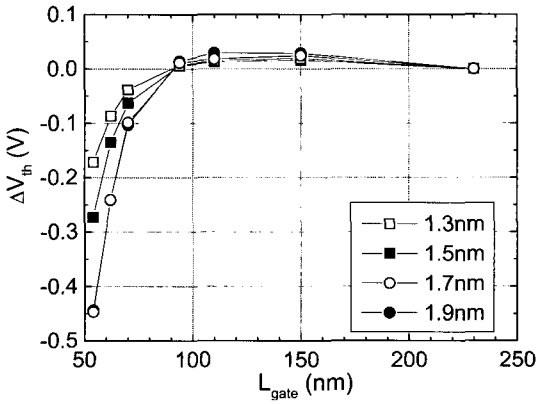


Fig. 3. Threshold voltage roll-off behavior of NMOS ransistors with T_{ox}=1.3~1.9㎚.

First of all, ultra-thin gate oxide has large gate leakage current density due to the direct tunneling. Fig. 4 shows gate leakage current density versus gate voltage. Increased gate leakage results in higher I_{off} and progressively destroys transistor operation characteristics, and also increases standby power consumption. Therefore, 1.0A/㎠ has been regarded as the limit of the leakage currents for high performance CMOS. If this is true, the limit of the oxide thickness would be 1.4~1.5㎚.

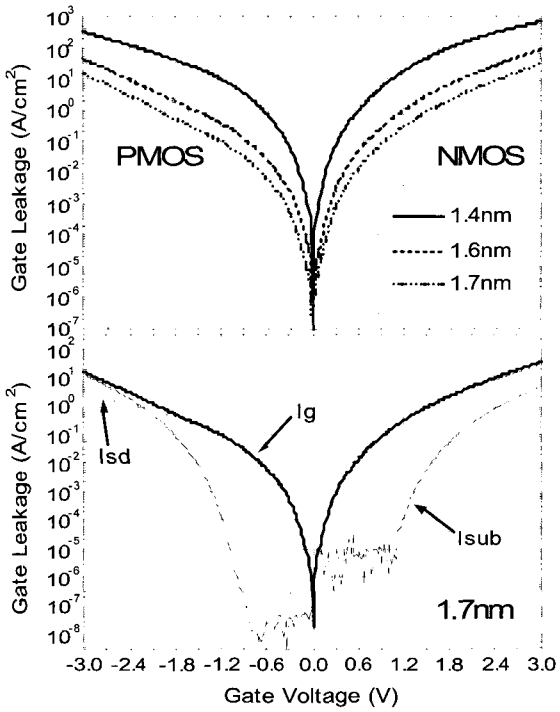


Fig. 4. (Top) Gate leakage current density of MOS capacitors with T_{ox}=1.4~1.7㎚. (Bottom) Gate leakage current components into substrate and source/drain.

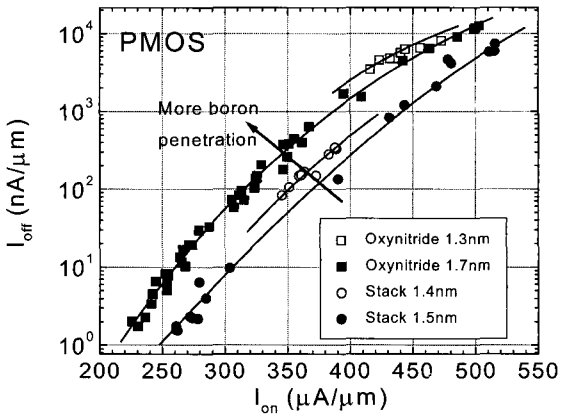


Fig. 5. I_{on}-I_{off} relations of PMOS transistors with 1.3~1.7㎚ thick oxynitrides and stack gate dielectrics.

Second issue is the boron penetration of the dual gate PMOS transistors. Boron penetration induces ploy depletion, threshold voltage shift and fluctuation and drive current degradation. Oxynitrides have been used widely to suppress boron penetration, but oxynitrides also meet the limit for scaled devices with T_{ox}<1.5㎚ and higher boron doses. Fig. 5 shows a plot of I_{off} versus Ion for transistors with different gate dielectrics (oxynitrides and the stacks of an oxynitride and a LPCVD nitride) and different equivalent oxide thicknesses. When a high BF_2 implantation dose was applied, I_{on} at a fixed I_{off} showed strong dependence on the degree of boron penetration instead of T_{ox}. It has been also found that stack gate dielectrics of an oxynitride and a LPCVD silicon nitride effectively suppress boron penetration.

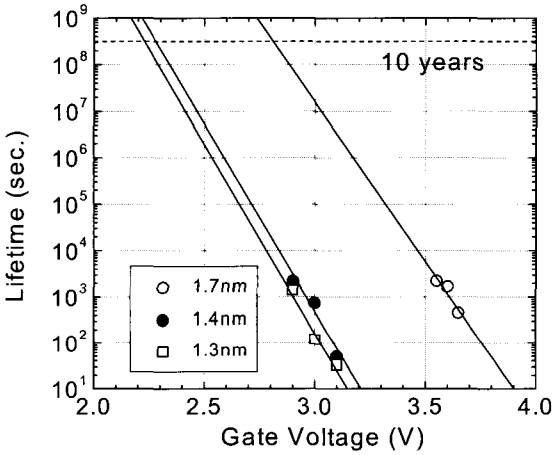


Fig. 6. TDDB of ultra-thin gate oxides of T_{ox}=1.3~1.7㎚.

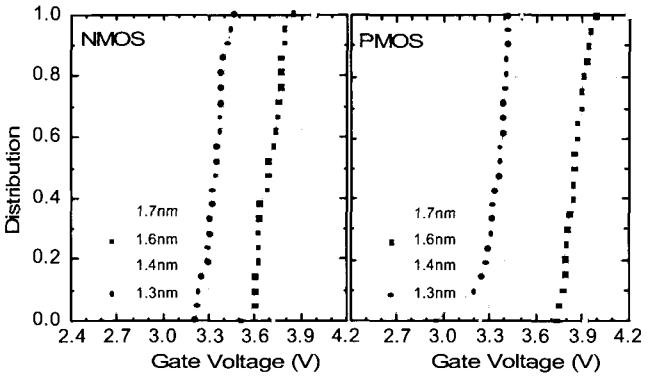


Fig. 7. Breakdown voltage distributions of 1.3~1.7㎚ oxynitrides.

Third issue is the reliability. It is very important to figure out the low-limit of the gate oxide scaling in terms of the reliability for the volume production. In this work, we used ultra-thin catalytic wet oxide. Fig. 6 and Fig. 7 are TDDB measurements and breakdown voltage distributions for oxides of T_{ox}=1.3~1.7㎚ measured from 18.3㎛ × 18.3㎛ MOS capacitors, respectively, which show good characteristics.

To solve these major issues, it is believed that a high-k dielectric is needed and ITRS proposes that it should be before the year of 2005. However no high-k material has met all the conditions the gate dielectric should have. Aluminum, Hafnium, or Zirconium oxides or silicates are strong candidates, but, they have problems of fixed charges or/and mobility degradation. Therefore, it is unclear that whether we could implement a high-k material as a gate dielectric soon and it becomes very important to figure out how far we could go with silicon oxide based gate dielectrics.

B. *Exploring Gate Oxide Scaling Limit*

Among those three major issues on gate oxide scaling, gate leakage has been regarded as the most urgent and serious problem not only for low-power application but also for high performance applications. On the other hand, the boron penetration problem has a trade-off relation with the source/drain engineering, therefore, has more process flexibility. Also, in the previous section, we have shown that the reliability of ultra-thin oxide may not be a big problem. In this section, gate oxide scalability is discussed in the viewpoint of the gate leakage current.

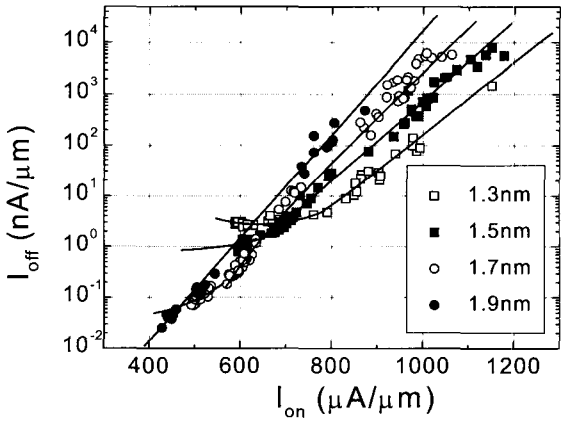


Fig. 8. I_{on}-I_{off} relations of NMOS transistors with T_{ox}= 1.3~1.9㎚ at 1.2V.

Fig. 8 shows NMOS I_{off} versus I_{on} for transistors with T_{ox}=1.3~1.9㎚. Normally, Ion and log(I_{off}) has a linear relationship, and, higher Ion for a fixed I_{off} is found for thinner oxide. However as shown in Fig. 8, it is not true for the ultra-thin oxide where I_{off} is not decreasing below a limiting value, which is due to the gate leakage current, and is determined by the gate oxide thickness. I_{on} at a fixed I_{off} for a thinner oxide is not always higher than that for a thicker oxide if I_{off} for a thinner oxide is dominantly determined by gate leakage current. Therefore, for a given technology node, the optimum value of T_{ox} instead of minimum possible T_{ox} should be used. The optimum value of T_{ox} is related to the maximum allowed I_{off} and the channel and the source/drain design. Below, we will describe this behavior more in depth.

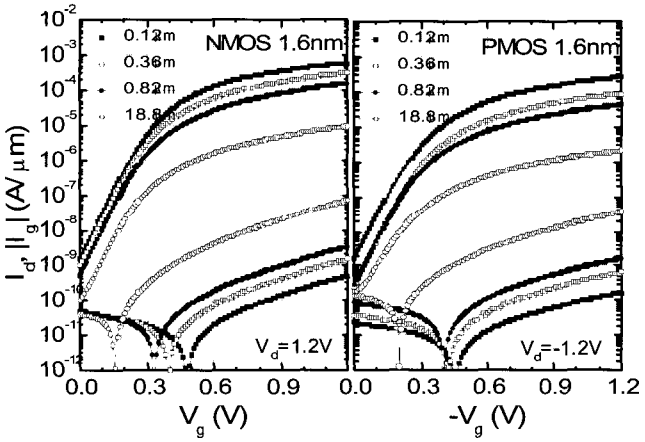


Fig. 9. I_d, $|I_g|$ versus V_g for various L_{gate} and T_{ox}=1.6㎚.

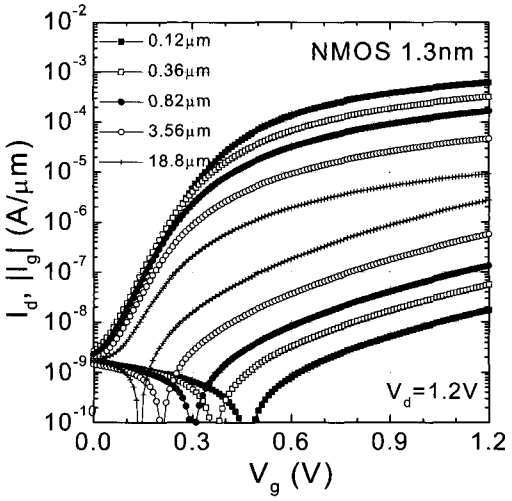


Fig. 10. NMOS I_d and $|I_g|$ versus V_g for various L_{gate} and T_{ox}=1.3㎚.

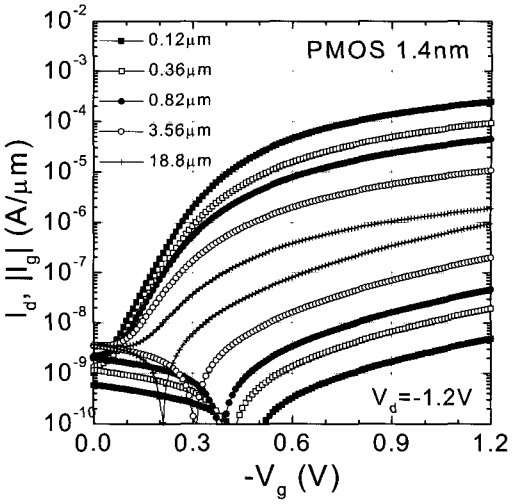


Fig. 11. PMOS I_d and $|I_g|$ versus V_g for various L_{gate} and T_{ox}=1.4㎚.

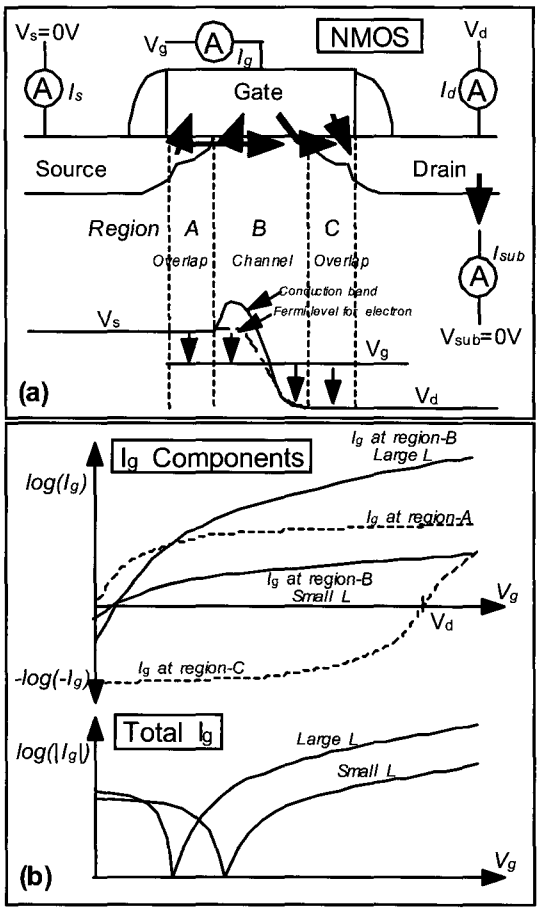


Fig. 12. (a) Modeling of I_g for NMOS and corresponding band diagram. (b) Modeling of L_{gate} dependent I_g components and total I_g.

Fig. 9~11 show drain currents (I_d) and gate leakage currents (I_g) versus gate voltage (V_g) for both NMOS

and PMOS transistors of various Lgate with gate oxide thicknesses of 1.3~1.6㎚. Fig. 12 shows schematics of the modeling of the gate leakage current to understand Fig. 9~11. Gate leakage can be divided into three components as shown in Fig. 12(a). Currents from gate-to-S/D overlap (region A, C) do not depend on L_{gate}, but depend on the overlap length. On the other hand, currents from the region-B depend on L_{gate} and the channel doping profile. For thicker gate oxides, smaller I_g does not affect transistor operation much as shown in Fig. 9. However, for thinner gate oxides, I_g degrades sub-threshold characteristics and sets the lower-limit of I_{off} for a given L_{gate} as shown in Fig. 10~11. For NMOS (see, Fig. 10), it is found that I_g at V_g=0 is mainly from the region-C (see, Fig.12) and does not depend on L_{gate}. It can be reduced by proper optimization of the overlap length together with the tradeoff relation between gate-to-S/D capacitance and S/D parasitic sheet resistance. For PMOS, as shown in Fig. 11, the value of I_{off} lower-limit is found to be decreased as L_{gate} is decreased, which implies leakage from region-B is also important (see, Fig.12). The different behavior between NMOS and PMOS may be due to the different channel and S/D extension profile.

In summary, as device size (L_{gate} and gate-to-S/D overlap) scales down further, thinner gate oxides can be used even though high gate leakage currents exist. Also, an optimum value of the oxide thickness exists for a given technology node with a given acceptable I_{off} limit. In addition, based on the SIA's roadmap (Table 1), acceptable I_{off} is higher for a smaller device, therefore, it is important to note that the high-k dielectrics may not be needed in the near future.

C. *High Performance CMOS*

To obtain the best performance with conventional CMOS technologies, we have fabricated transistors with optimized process conditions. L_{gate} was 70㎚, and 1.4㎚-thick oxynitride gate dielectric was used. To reduce boron penetration, boron and BF_2 doses were carefully controlled and minimum amount of heat budget was applied. Nickel salicide was used to enhance PMOS performance. As a result, outstanding current drives of 860㎂/㎛ (NMOS) and 350㎂/㎛ (PMOS) and CV/I of 1.60ps (NMOS) and 3.32ps (PMOS) were achieved at I_{off}=10㎁/㎛ and V_{dd}=1.2V. Table 3 shows summary of characteristics of these transistors with other high performance CMOS transistors from references.

Table. 3. Characteristics of high performance transistors.

	This Work	Ref. [3]	Ref. [4]	Ref. [5]	Ref. [6]
Lgate (㎚)	**70**	100	75	50	80
Phy. Tox (㎚)	**1.4**	2.0	≤2.0	-	1.5
Inv. Tox [N] (㎚)	**2.1**	3.0	2.7	3.2	2.54
Inv. Tox [P] (㎚)	**2.5**	3.0	3.0	3.2	
Vdd (V)	**1.2**	1.2	1.2	1.2	1.2
Ion [N] (㎂/㎛)	**860**	820	800	690	836
Ion [P] (㎂/㎛)	**350**	340	349	305	376
Ioff [N] (㎁/㎛)	**10**	10	10	10	6.5
Ioff [P] (㎁/㎛)	**10**	10	10	10	21
CV/I [N] (ps)	**1.60**	1.68	1.44	1.06	1.56
CV/I [P] (ps)	**3.32**	4.10	2.97	2.16	3.47

IV. CONCLUSION

Gate oxide scaling for sub-100㎚ CMOS devices was studied. Issues on gate dielectric scaling with silicon oxide based dielectric were reviewed. The effect of boron penetration was investigated and novel stack gate dielectrics of the silicon oxynitride and the LPCVD nitride were developed. Reliability of ultra-thin silicon oxynitride showed good characteristics in the TDDB lifetime and the breakdown voltage distribution.

Gate oxide scalability was studied in the viewpoint of gate leakage current. Modeling and experimental results showed that thinner oxides than previous believed limit may be usable as scaling proceeds, therefore, high-k gate dielectrics may not be needed in the near future as ITRS proposes. Also, in the regime of large gate leakage current, gate oxide with optimized thickness should be used for the best performance rather than available minimum thickness.

High performance CMOS transistors of L_{gate}=70㎚ and T_{ox}=1.4㎚ were fabricated, which showed excellent current drives of 860㎂/㎛ (NMOS) and 350㎂/㎛ (PMOS) and CV/I of 1.60ps (NMOS) and 3.32ps (PMOS) at I_{off}=10㎁/㎛ and V_{dd}=1.2V.

REFERENCES

[1] S. Song, W. S. Kim, J. M. Ha, G. G. Lee, J.-H. Ku, H. S.

Kim, C. S. Kim, C. J. Choi, T. H. Choe, J. Y. Yoo, W. S. Song, J. W. Park, S. H. Jeong, D. H. Baek, K. Fujihara, H. K. Kang, S. I. Lee, and M. Y. Lee, "High Performance Transistors with State-of-the-Art CMOS Technologies," *IEDM Tech. Dig.*, pp. 427-430, 1999.

[2] S. Song, W. S. Kim, J. S. Lee, T. H. Choe, J. H. Choi, M. S. Kang, U. I. Chung, N. I. Lee, K. Fujihara, H. K. Kang, S. I. Lee, and M. Y. Lee, "Design of Sub-100nm CMOSFETs: Gate Dielectrics and Channel Engineering," *Symp. on VLSI Technology Dig. of Tech. Papers*, pp. 190-191, 2000.

[3] T. Ghani, S. Ahmed, P. Aminzadeh, J. Bielefeld, P. Charvat, C. Chu, M. Harper, P. Jacob, C. Jan, J. Kavalieros, C. Kenyon, R. Nagisetty, P. Packan, J. Sebastian, M. Taylor, J. Tsai, S. Tyagi, S. Yang, and M. Bohr, "100nm Gate Length High Performance / Low Power CMOS Transistor Structure," *IEDM Tech. Dig.*, pp. 415-418, 1999.

[4] M. Mehrotra, J. C. Hu, A. Jain, W. Shiau, S. Hattangady, V. Reddy, S. Aur, and M. Rodder, "A 1.2V, Sub-0.09 μm Gate Length CMOS Technology," *IEDM Tech. Dig.*, pp. 419-422, 1999.

[5] I. Y. Yang, K. Chen, P. Smeys, J. Sleight, L. Lin, M. Leong, E. Nowak, S. Fung, E. Maciejewski, P. Varekamp, W. Chu, H. Park, P. Agnello, S. Crowder, F. Assaderaghi, and L. Su, "Sub-60nm Physical Gate Length SOI CMOS," *IEDM Tech. Dig.*, pp. 431-434, 1999.

[6] A. Ono, K. Fukasaku, T. Matsuda, T. Fukai, N. Ikezawa, K. Imai, and T. Horiuchi, "A 70nm Gate Length CMOS Technology with 1.0V Operation," *Symp. on VLSI Technology Dig. of Tech. Papers*, pp. 14-15, 2000.

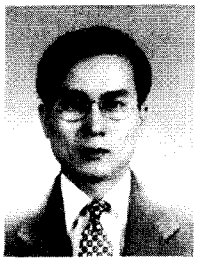
[7] S. Song, J. H. Yi, W. S. Kim, J. S. Lee, K. Fujihara, H. K. Kang, J. T. Moon, and M. Y. Lee, "CMOS Device Scaling Beyond 100nm," *IEDM Tech. Dig.*, pp. 235-238, 2000.

Seungheon Song was born in Korea on September 21, 1970. He received the B.S. degree in Physics from Korea Advanced Institute of Science and Technology (KAIST), Taejon, Korea in 1991 and the Ph.D degree in Physics from Massachusetts Institute of Technology (MIT), MA, USA in the field of synchrotron x-ray scattering and surface science in 1996. In 1996, he joined Semiconductor R&D Center of Samsung Electronics Co., Ltd., Kiheung, Korea. Between 1996 and 1998, he was involved in the process control and the failure analysis of the Alpha-microprocessors, FLASH memory devices, and SRAM devices. Since 1999, he has been working on the development of sub-100nm high performance CMOS transistors, 3D CMOS integration, and a new memory device called Scalable Two-Transistor Memory (STTM). His current research interests include CMOS scaling, Nano-scale electron devices, and next generation memory devices. He has published over 20 technical papers in international journals and conference proceedings in the field of condensed matter physics and CMOS devices. He was an invited speaker of several conference meetings including American Physical Society (APS) March Meeting in 1995, San Jose, CA, International Electron Device Meeting (IEDM) in 2000, San Francisco, CA, and Korean Conference on Semiconductors (KCS) in 2001, Seoul, Korea. He is a senior engineer of Samsung Electronics Co., Ltd., and a member of IEEE.

Jihye Lee was born in Korea on May 19, 1976. She received the B.S. degree in Physics Education and M.S. degree in Physics from Seoul National University (SNU), Seoul, Korea in 1998 and 2000, respectively. In 2000, she joined Semiconductor R&D Center of Samsung Electronics Co., Ltd. where she is involved in the development of 3D CMOS integration and the process and architecture of a new memory device.

Woosik Kim was born in Korea on December 27, 1968. He received the B.S. degree, the M.S. degree, and the Ph.D degree in Physics from Seoul National University (SNU), Seoul, Korea in 1991, 1993, and 1998, respectively. In 1998, he joined Semiconductor R&D Center of Samsung Electronics Co., Ltd., Kiheung, Korea where he has been involved in the development of high performance CMOS transistors, CMOS with metal gate electrode, transistor reliability, and the process technologies for a next generation memory device. Currently, he is a senior engineer of Samsung Electronics Co., Ltd.

Kazuyuki Fujihara was born in Japan on November 6, 1959. He received the B.S. degree and the M.S. degree in Electronics Engineering from Yamaguchi University, Ube, Japan, in 1982 and 1984, respectively. In 1984, he joined Fujitsu Ltd., Kawasaki, Japan, where he was involved in LPCVD and metal CVD process for CMOS device fabrication. In 1993, he joined Samsung Electronics Co., Ltd., Kiheung, Korea. His current major activity is focused on the front-end process of CMOS technology.



Ho-Kyu Kang received the B.S. degree in Metallurgical engineering from Hanyang University, Seoul, Korea in 1983, and the M.S. degree in Material Science and Engineering from Korea Advanced Institute of Science and Technology in 1985. In 1993, he received the Ph. D degree in Material Science and Engineering from Stanford University, CA, USA. In 1983, he joined Semiconductor Division of Samsung Electronics Co., Ltd. where he had been involved in the MOS transistor and SRAM process integration. Since 1993, he has been involved in the development of DRAM capacitors, CMOS transistor technologies, inter-layer dielectrics, Cu interconnections, and isolation technologies. Currently, he is a manager of the process development team of Semiconductor R&D center of Samsung Electronics Co., Ltd. He is an author of more than 35 technical papers and holds 15 patents related to semiconductor technologies.

Joo-Tae Moon was born in June 20, 1962. He received the B.S. degree in Metallurgical engineering from Seoul National University (SNU) in 1984, and the M.S. degree and the Ph.D degree in Material Science from Korea Advanced Institute of Science and Technology in 1986 and 1989, respectively. In 1989, he joined Samsung Electronics Co., Ltd. Kiheung, Korea, where he was involved in the development of process technology for memory devices including 64Mb, 256Mb, and 1Gb DRAMs. Currently, he is a director of Process Development Team, Semiconductor R&D Center of Samsung Electronics Co., Ltd.

Moon-Yong Lee was born in Korea on May 28, 1952. He received the B.S. degree in material science from Seoul National University (SNU), Seoul, Korea in 1975, and the M.S. degree in material science from Korea Advanced Institute of Science and Technology (KAIST), Taejon, Korea in 1977. In 1986, he received the Ph.D degree in material science from Ohio State University, Ohio, USA. Between 1976 and 1980, he was with Hyundai Engine Industry. In 1986, he joined Samsung Electronics Co., Ltd. where he has been involved in the development of 16Mb~1Gb DRAMs. Currently, he is an executive vice president and general manager of Semiconductor R&D Center of Samsung Electronics Co., Ltd.