A Cost Effective 32nm High-K/ Metal Gate CMOS Technology for Low Power Applications with Single-Metal/Gate-First Process

X. Chen, S. Samavedam[1], V. Narayanan, K. Stein, C. Hobbs[1], C. Baiocco, W. Li, D. Jaeger, M. Zaleski[1], H. S. Yang, N. Kim[2], Y. Lee, D. Zhang[1], L. Kang[1], J. Chen[2] H. Zhuang[3], A. Sheikh, J. Wallner, M. Aquilino, J. Han[3], Z. Jin, J. Li, G. Massey, S. Kalpat, R. Jha, N. Moumen, R. Mo, S. Kirshnan, X. Wang, M. Chudzik, M. Chowdhury[1], D. Nair, C. Reddy[1], Y. W. Teh[2], C. Kothandaraman, D. Coolbaugh, S. Pandey[2], D. Tekleab[2], A. Thean[1], M. Sherony, C. Lage[1], J. Sudijono[2], R. Lindsay[3], J. H. Ku[4], M. Khare, A. Steegen
IBM Semiconductor Research and Development Center (SRDC), Hopewell Junction, NY 12533
[1]Freescale Semiconductor, [2]Chartered Semiconductor Manufacturing, [3]Infineon Technologies AG, [4]Samsung Electronics Co., Ltd
e-mail: chenxian@us.ibm.com

Abstract

For the first time, we have demonstrated a 32nm high-k/metal gate (HK-MG) low power CMOS platform technology with low standby leakage transistors and functional high-density SRAM with a cell size of 0.157 μm^2. Record NMOS/PMOS drive currents of 1000/575 μA/μm, respectively, have been achieved at 1 nA/μm off-current and 1.1V V_{dd} with a low cost process. With this high performance transistor, V_{dd} can be further scaled to 1.0V for active power reduction. Through aggressive EOT scaling and band-edge work-function metal gate stacks, appropriate Vts and superior short channel control has been achieved for both NMOS and PMOS at L_{gate}=30nm. Compared to SiON-Poly, 30% RO delay reduction has been demonstrated with HK-MG devices. 40% Vt mismatch reduction has been shown with the Tinv scaling. Furthermore, it has been shown that the 1/f noise and transistor reliability exceed the technology requirements.

Introduction

Fast growing mobile applications require low power CMOS technology that can offer enhanced performance, low standby power and active power without increasing the cost [1]. High-k metal gate is considered to be one of the most promising technology enabler for low power application to meet the above requirements [2]. One concern for HK-MG in low power application is the cost from dual metal gates [3, 4]. In this paper, we present a cost-effective low power 32nm technology with single-metal/gate-first process[5]. The use of Hf based high-k gate dielectrics allows us to maintain a low-gate leakage of <0.1A/cm^2 while continuing to provide substantial room for EOT scaling. The dramatic short-channel improvements with EOT-scaling enable us to scale L_{gate} down to 30nm. Moreover, the strong performance offered by HK-MG allow us to trade off some performance for a cost-balanced low-power technology, by foregoing embedded SiGe S/D stressor. V_{dd} of the core LSTP (low stand-by power) transistors has been scaled down to 1.0V for active power reduction. Furthermore, a low operating power (LOP) transistor is offered with the same gate stack as the LSTP transistor at reduced 0.8V V_{dd} (Table1). A HK-MG compatible programmable element, eFUSE, is offered for redundancy and chip-id applications.

Process Integration

LSTP and LOP core transistors are fabricated on the same chip with thick oxide I/O transistors using a dual-oxide process. As shown from TEM (Fig.1), 30nm transistors were fabricated with gate first high-k/metal gate process employing Hf-based high-k gate dielectrics and a thermally stable single metal with conventional Si capping layer. After gate stack patterning and etch, conventional spacer process and self-aligned implants were used. Less than 3% total process cost is added with HK-MG compared with Poly-SiON gate stack. Advanced millisecond-based RTA annealing has been implemented for shallow junction engineering to scale gate length to 30nm without any negative impact to the gate stack. Dopant optimization enabled eFUSE elements with high post-programmed resistance. A 1.2 NA/193 nm wet immersion lithography was used for critical levels to enable a 70% scaling of key ground rules from 45nm. Table 2 shows the key FEOL and BEOL ground rules and scaling factors from 45nm. BEOL is composed of 6 to 11 layers of copper metallization in ultra low-K (k=2.4) dielectric (Fig.2). Due to the increased effects of scaling on 1X (100nm pitch) wires, additional use of hierarchy is employed by the addition of 1.3X pitch intermediate wiring levels for lower resistance.

Device Characteristics and SRAM Cell

With high-k/metal gate, Tinv is reduced by more than 1 nm with same gate leakage as the SiON-poly gate stack (Fig.3). Leading edge NFET and PFET performance has been achieved in this 32nm technology due to Tinv scaling and optimized high-k/Si interface. High drive currents of 1000/575 μA/μm (N/P) have been achieved at I_{off} =1nA/μm and V_{dd}=1.1V without additional stress engineering complexity and cost, such as e-SiGe process (Figs. 4, 5). With high drive capability, V_{dd} can be scaled down to 1V for the 32nm technology. One concern for 45nm to 32nm node scaling is the performance degradation from series resistance and stress loss due to gate pitch scaling [6]. With an optimized process, <3% (N) and 0% (P) performance degradation was observed with gate pitch scaling from 252 nm to 126 nm (Fig.6). Good subthreshold characteristics have been shown for both NFET and PFET with HK-MG stack (Fig.7). With effective band-edge work-function gate stacks, appropriate Vt and short channel control was achieved (Figs. 8, 9). Most importantly, EOT scaling has enabled L_{gate} scaling down to 30nm. Compared with SiON-poly, HK-MG devices shows less DIBL at same L_{gate}. The transistor performance at 0.8V operation (LOP) is also shown in Fig.10. Without an additional gate oxide [7] to maintain process simplicity, good LOP transistor performance (Ion = 700/340 μA/μm @ 100nA/μm off-current) was achieved at V_{dd}=0.8V. HK-MG ac performance is shown in Fig.11. Compared with SiON-poly, 30% delay reduction has been demonstrated with HK-MG due to improved transistor DC performance and the L_{gate} scaling from 40nm (SiON-poly control) to 30nm (HK-MG).

Top-down SEM of a functional 0.157 μm^2 SRAM cell is shown in Fig.12. Device widths and lengths have been scaled from the 45nm technology to achieve a high density SRAM cell. With a conventional SiON-gate stack, Vt mismatch would increase and degrade SRAM stability. However, due to aggressive Tinv reduction in this 32nm HK-MG low power technology, the mismatch is reduced 40% compared with SiON-poly gate stack (Fig.13), significantly improving SRAM operating voltage range. The butterfly curve of the 0.157μm^2 dense SRAM cell is shown in Fig. 14 and a static noise margin (SNM) of 250 mV is achieved.

Potential 1/f noise and reliability degradation have been critical concerns for the HK-MG devices. Impacts of HK-MG gate stack on 1/f noise have been examined and no degradation was observed compared with poly-SiON gate stack from previous node (Fig. 15). NBTI and PBTI were assessed for LTSP transistors. Both NFETs and PFETs meet the 10 year lifetime at V_{dd}=1.2V after stress at 125 °C (Fig. 16).

Conclusion

We have successfully demonstrated a cost-effective 32nm LSTP and LOP technology platform which uses a gate-first and single-metal HK-MG process with an effective band-edge work-function solution for both NFETs and PFETs. EOT scaling with HK-MG enables superior transistor performance with aggressive gate length scaling and a dense SRAM cell with good SNM. Good noise performance and reliability of HK-MG devices was also confirmed.

This work was performed at the IBM Microelectronics Div., Semiconductor Research & Development Center, Hopewell Junction, NY 12533. The authors would like to thank J. Jagannathan, P. Gilbert, G. Patton and T-C Chen for their support.

References:

[1] T. Hook, IEDM 2006, p.701

[2] ITRS 2005.

[3] K. Mistry, IEDM2007, P. 247

[4] H. Huang, IEDM2007, P. 285

[5] M. Chudzik, VLSI2007, P.194

[6] S. Wu, IEDM 2007, P. 263

[7] E. Josse, IEDM 2006, p.693

Trans	LSTP	LOP	I/O
Vdd (V)	1.0	0.8	1.5/ 1.8
Lgate (nm)	30	30	150

Table 1: Key 32nm Transistors.

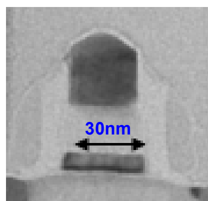


Fig.1: TEM of the Low power transistor with Lgate=30nm.

Rule	Pitch (nm)	Scaling factor from 45nm
Contacted Gate Pitch	126	70%
CA pitch	100	63%
N+/P+	56 (space)	70%
1X metal	100	71%
1.3X metal	130	New level
2X metal	200	71%

Table 2: Key Ground rule of this 32nm CMOS foundry technology.

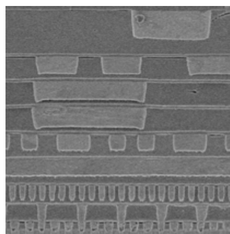


Fig.2: SEM cross-section through BEOL wiring of SRAM. First 9 layers of copper metal are shown where 7 are in ULK dielectric.

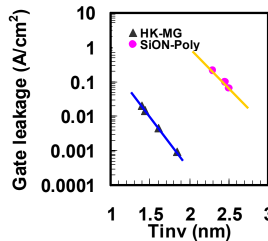


Fig.3: Gate leakage versus Tinv which shows >10nm Tinv reduction with HK-MG at same gate leakage.

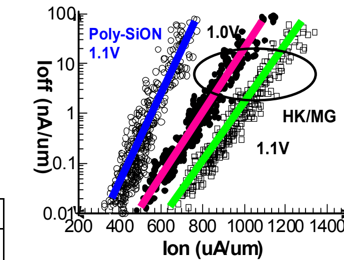


Fig.4: LSTP NFET I_{off}-I_{on} with 1000uA/um I_{on} at 1nA/um I_{off} and 1.1V V_{dd}.

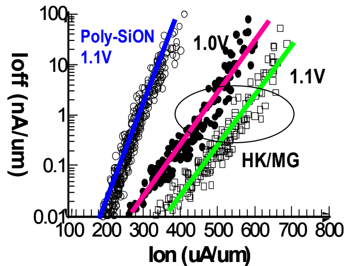


Fig.5: LSTP PFET I_{off}-I_{on} with 575uA/um I_{on} at 1nA/um I_{off} and 1.1V V_{dd}.

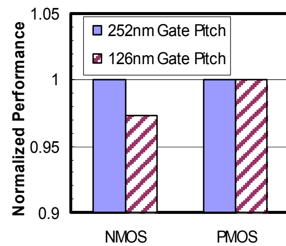


Fig.6: Transistor performance at different gate pitch.

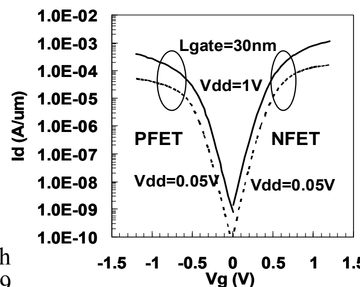


Fig.7: NFET and PFET I_d-V_g curve. SS is 90mV/dec for NFET and PFET.

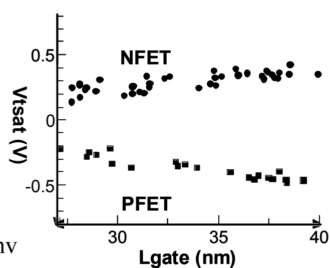


Fig.8: V_t roll-off characteristics of NFET and PFET.

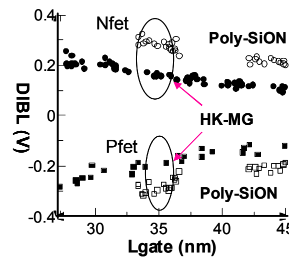


Fig.9: Comparison of DIBL between HK-MG and Poly-SiON.

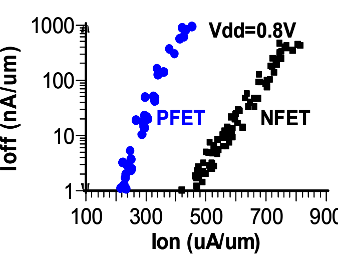


Fig.10: I_{off}-I_{on} of the LOP transistors at 0.8V.

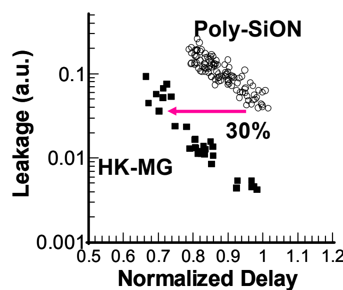


Fig.11: 30% RO delay reduction with HK-MG at same Vdd and leakage.

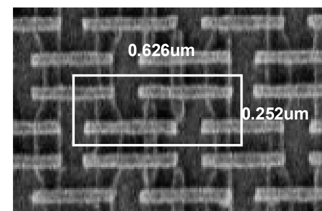


Fig.12: Top-down SEM of a 0.157 um^2 SRAM bit-cell.

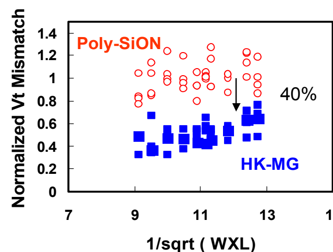


Fig.13: Vt mismatch is reduced 40% with HK-MG due to Tinv scaling.

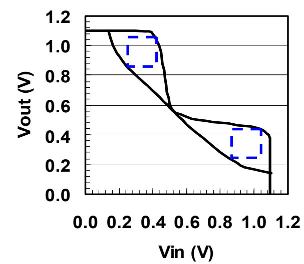
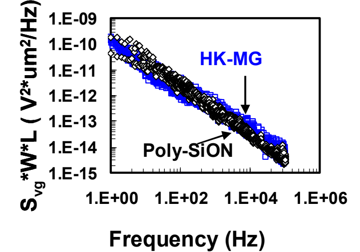


Fig.14: Measured output for the 0.157um² SRAM bit-cell with SNM 250mV.



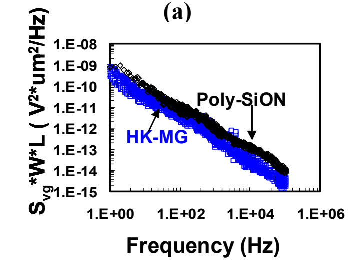
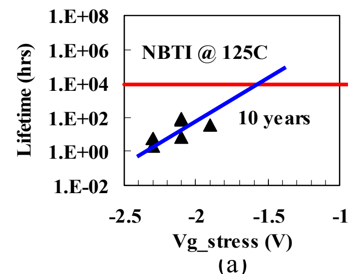


Fig.15: Normalized gate input refered voltage noise for 32nm HK-MG technology compared with Poly-SiON gate stack for NFET (a) and PFET (b).



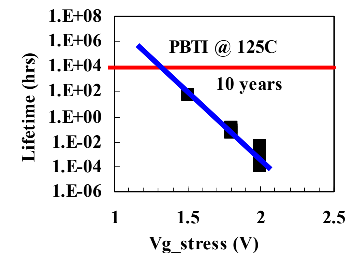


Fig.16: NBTI (a) and PBTI (b) measurements of HK-MG wafer meet 10 year lifetime requirement.