Scaling of 32nm Low Power SRAM with High-K Metal Gate

H. S. Yang, R. Wong, R. Hasumi[1], Y. Gao[2], N.S. Kim[2], D. H. Lee[3], S. Badrudduza[4], D. Nair
M. Ostermayr[5], H. Kang[3], H. Zhuang[5], J. Li, L. Kang[4], X. Chen, A. Thean[4], F. Arnaud[6], L. Zhuang
C. Schiller, D. P. Sun[2], Y. W. Teh[2], J. Wallner, Y. Takasu[1], K. Stein, S. Samavedam[4], D. Jaeger
C.V. Baiocco, M. Sherony, M. Khare, C.Lage[4], J. Pape, J.Sudijono[2], A.L. Steegen, S. Stiffler

IBM Microelectronics Semiconductor Research and Development Center (SRDC), [1]Toshiba, [2]Chartered Semiconductor,
[3]Samsung Electronics, [4]Freescale Semiconductor, [5]Infineon Technologies, [6]STMicroelectronics
2070 Route 52, Zip 32A, Hopewell Junction, NY 12533 E-mail: haining@us.ibm.com Phone: (845)892-1759

Abstract

This paper describes SRAM scaling for 32nm low power bulk technology, enabled by high-K metal gate process, down to 0.149μm^2 and 0.124μm^2. SRAM access stability and write margin are significantly improved through a 50% Vt mismatch reduction, thanks to HK-MG T_{inv} scaling. Cell read current is increased by 70% over Poly-SiON process. Ultra dense cell process window is expanded with optimized contact process. A dual-ground write assist option can additionally enable ultra dense 0.124μm^2 cell to meet low power application requirements.

Introductions

SRAM soft fail rate increased sharply over the past few technology nodes due to worsened device variations, overtaking hard defect fail rate [1]. Random dopant fluctuation is a major contributor to the increased device variations [2]. Vt mismatch between two nearby transistors is a function of device area and gate dielectric thickness. Shrinking device area is required for denser SRAM scaling. However, gate dielectric thickness has not kept pace with ground rule scaling for gate leakage concern. These effects make the continuation of the traditional SRAM scaling unsustainable.

Fig. 1 shows SRAM pull-down transistor Vt mismatch increases since 90nm technology. As a result of deteriorated Vt mismatch, SRAM soft fail rate increases, as illustrated in Fig. 2.

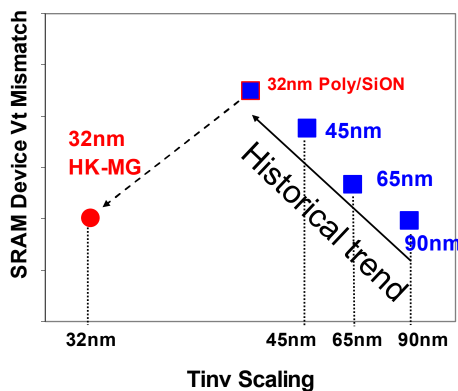

Fig. 1 SRAM Vt mismatch increases as device dimension shrinks. For the first time, the trend is reversed for 32nm node by adopting HK-MG.

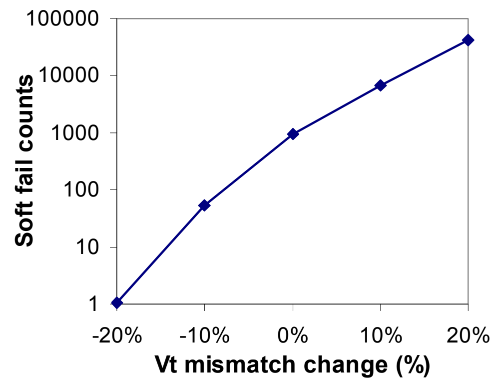

Fig. 2 SRAM soft fails increase with degraded mismatch. Soft fails include access disturb fails and write fails.

32nm SRAM Scaling with HK-MG

For the first time, the trend is reversed for 32nm technology with the adoption of a gate-first hafnium based high-K process (Fig. 1). Vt mismatch is reduced by 50% from increased gate control to the channel from T_{inv} scaling, as shown in Fig. 3. Consequently, VDD_{min} of the 32nm dense SRAM is expected to improve by 180mV (Fig.4).

With the adoption of the HK-MG process, 50% SRAM area scaling is realized for 32nm low power technology (Fig. 5) with improved cell read current at lower standby leakage.

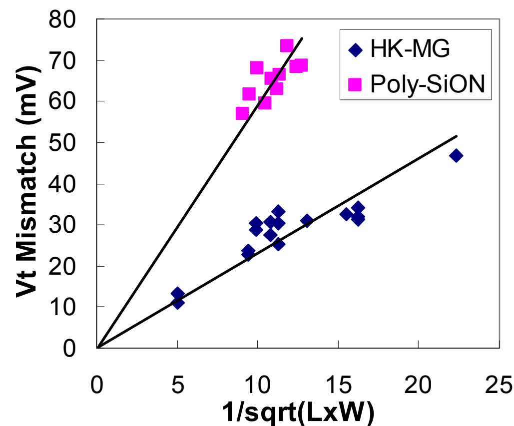

Fig. 3 50% Vt mismatch reduction is realized by using 32nm HK-MG process over 45nm Poly-SiON.

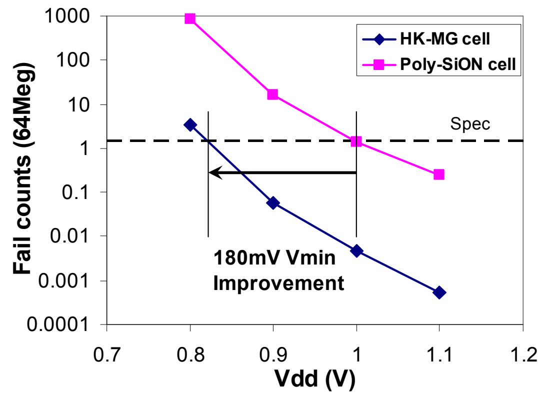

Fig 4 180mV VDD_{min} improvement on 0.149μm² cell with HK-MG over Poly-SiON.

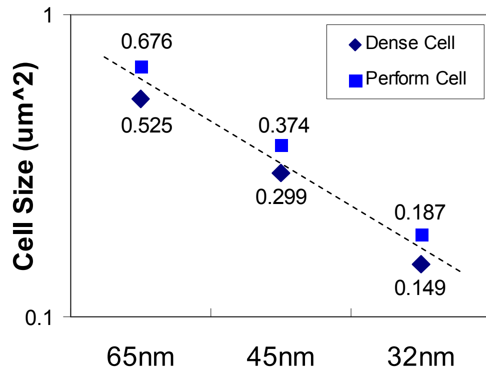

Fig. 5 SRAM cell scaling trend showing continued 50% area reduction for 32nm node.

SRAM Fabrication and Electrical Performance

SRAM cells of 0.149μm², 0.139μm², and 0.124μm² patterned with high NA/193nm immersion lithography are fabricated in a 64Meg SRAM test chip. Header, footer, and wordline driver isolation devices are used to separate the leakage contributions from each cell type. A TEM image of the 0.149μm² cell is shown in Fig. 6 displaying HK-MG gate stack and contacts. Contacts are split to two masks to improve lithographic process window for dense and ultra dense cells (Fig. 7). OPC optimization is performed to achieve a best contact pattern split strategy. An SEM top-down view of the 0.124μm² cell is displayed in Fig. 8. Double patterning technique is used to reduce gate tip to tip spacing.

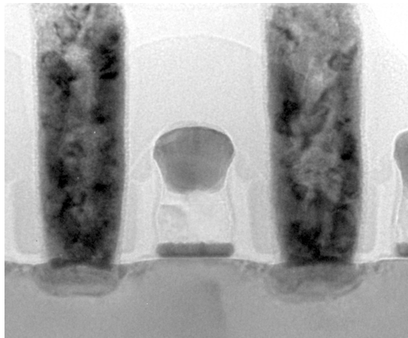

Fig. 6 Cross-sectional TEM of 0.149μm² cell showing HKMG NFET gate stack and contacts.

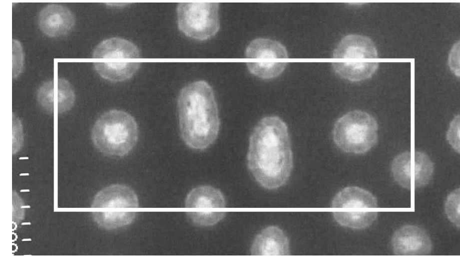

Fig. 7 SEM of 0.139μm² cell showing improved contact process window using double patterning technique.

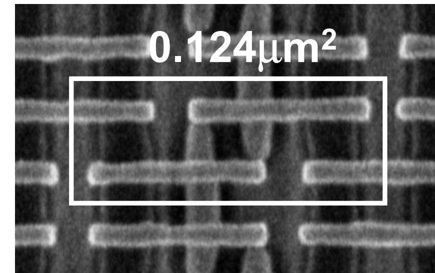

Fig. 8 Top-down SEM of 0.124μm² cell showing improved gate tip to tip spacing using a cut mask.

SRAM read current is improved by 70% over 45nm Poly-SiON process on a same SRAM cell, as shown in Fig 9. Standby leakage is reduced by optimizing gate stack, salicide, and contact process. Fig. 10 indicates gate leakage of HK-MG device is further lowered at low VDD region, comparing to that of poly-SiON device, which reduces standby power of SRAM in retention mode.

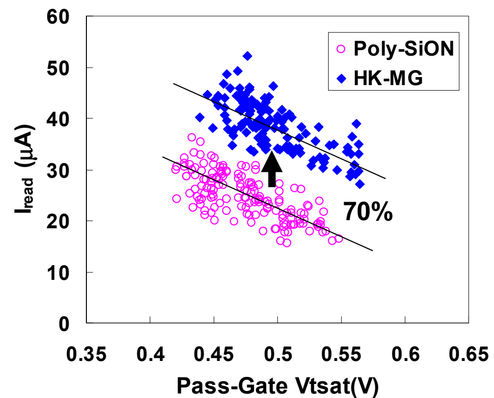

Fig. 9 SRAM read current increases by 70% from HK-MG over Poly-SiON, as demonstrated on 0.299 μm² cell.

Read N-curves of 0.124μm² cell at various VDDs are shown in Fig. 11, where the current is measured as the voltage sweeps on the cell storage node while WL and BL are at VDD. The peak current represents the minimum disturbing current necessary to change the state of the cell, indicating access disturb stability. Corresponding butterfly curves measured on the same set of cells are shown in Fig. 12. Write margin is depicted using write N-curves measured under write bias condition, as in Fig. 13. The minimum

current at the flat region represents cell writability. The methodology is described in [3]. Read and write current of the optimized ultra dense cell are shown in Fig. 14-15. The sensitivity of write margin to device characteristics is measured using Write Margin Macro that contains a 512bit array, as in Fig. 16. Sufficient access stability and write margin are demonstrated for performance and dense cells. However, for the ultra dense cell, write assist is required to meet VDD_{min} requirements.

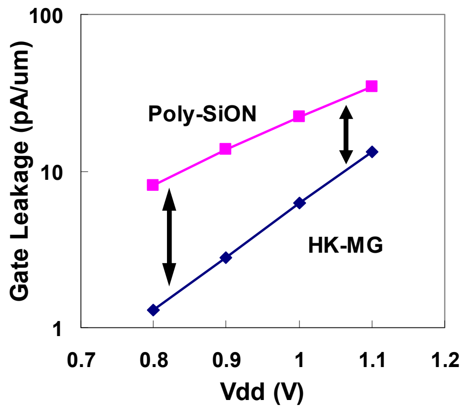

Fig 10 Gate leakage reduction from HK-MG showing low voltage operation benefits.

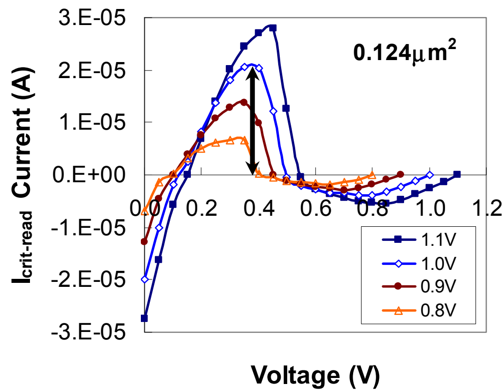

Fig. 11 Read N-curves measured on $0.124\mu m^2$ cells at various VDDs. The peak current represents cell stability.

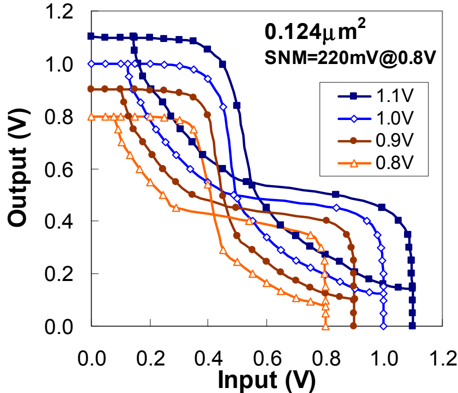

Fig. 12 Corresponding butterfly curves measured on $0.124\mu m^2$. SNM of 220mV is observed at 0.8V VDD.

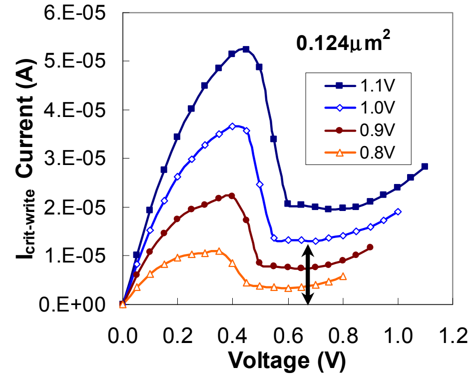

Fig. 13 Write N-curves measured on $0.124\mu m^2$ cells. The minimum current at the flat region indicates write margin.

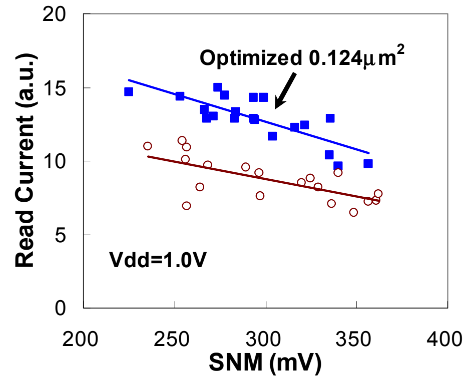

Fig. 14 Optimized $0.124\mu m^2$ cell showing improved read current at the same SNM.

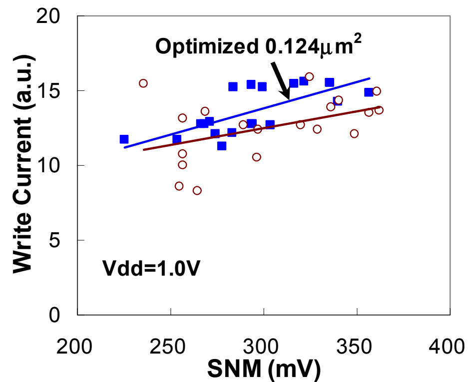

Fig. 15 Write current of the optimized $0.124\mu m^2$ cell is improved at the same SNM.

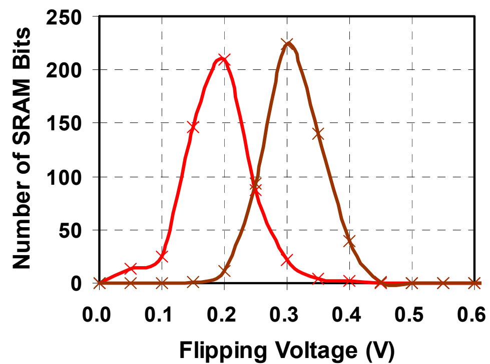

Fig. 16 512bit SRAM Write Margin Macro showing cell write voltage as a function of device gamma ratio.

Dual-ground Write Assist

Multiple write assist schemes have been proposed, including lowered cell VDD, boosted wordline, and a negative bitline ground during write operation. The implementation of lowered cell VDD is limited by the risk of data retention [1]. The boosted WL VDD sacrifices stability of half-selected cells. The negative BL approach requires expensive triple-well process or level shifters with considerable periphery area penalty.

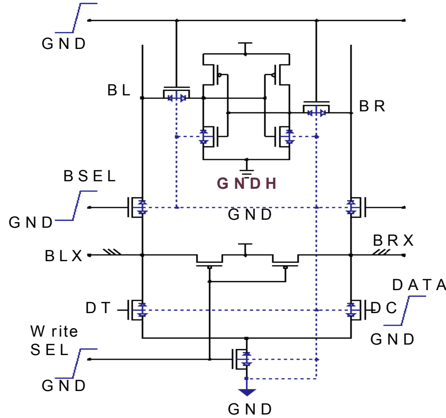

Fig. 17 Dual-ground write assist circuit schematic with a constant cell GNDH > GND bias (no voltage bump during read or write process).

A dual-ground scheme can be implemented to mitigate these limitations. The SRAM cell is tuned to have a strong PFET to be sufficiently stable. During write operation, one of the writing BLs is pulled to ground voltage GND, while SRAM cell ground voltage stays at an elevated ground voltage GNDH, which is about 150mV higher than GND. Fig. 17 illustrates circuit schematic featuring dual-ground write assist scheme. Simulated waveforms of the write operation with and without the write assist are shown in Fig. 18. Note the level of the low storage node is brought to a negative voltage relative to cell ground for a strong write of "0". Fig. 19 shows the effectiveness of the write assist. Write margin of the ultra dense cell is increased by 3σ, corresponding to 300mV VDD_{min} improvement. Fig. 20 indicates fail count reduces as write assist voltage increases.

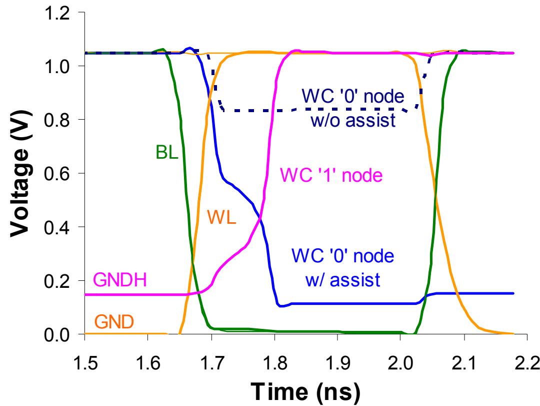

Fig. 18 Simulated waveforms of 0.124μm^2 cell with or without dual-ground write assist, at a worse case (WC) condition.

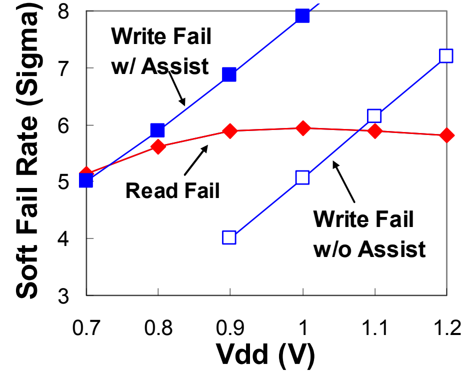

Fig. 19 Simulation showing improved write margin with dual-ground write assist solution on 0.124μm^2 SRAM cell.

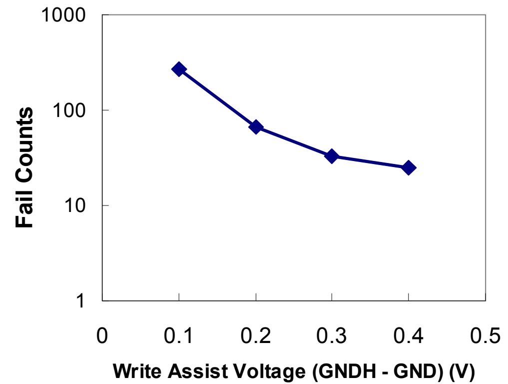

Fig. 20 Measured fail counts as a function of write assist voltage (negative BL GND voltage relative to cell GNDH).

Conclusion

The superior Vt mismatch and gate leakage of HK-MG devices have enabled a continued 50% SRAM area scaling for 32nm low power technology with improved cell characteristics for low voltage and low standby leakage applications. A dual-ground write assist option can further improve operating voltage range of the 0.124μm^2 ultra dense cell.

Acknowledgements

This work was performed at the IBM Microelectronics Div. Semiconductor Research & Development Center, Hopewell Junction, NY, 12533. The authors thank O. Heymann, A. Roberts, and F. Towler for design support.

References

[1] H. Pilo et al., IEEE J. Solid-State Circuits, vol.42, No.4, 4/07, p.813
[2] A. Agarwal et al., IEEE 2004 Custom Integrated Circuits Conf., p.353
[3] C. Wann et al., 2005 IEEE VLSI-TSA International Symposium, p. 21