A 32nm Logic Technology Featuring 2nd-Generation High-k + Metal-Gate Transistors, Enhanced Channel Strain and 0.171μm² SRAM Cell Size in a 291Mb Array

S. Natarajan, M. Armstrong, M. Bost, R. Brain, M. Brazier, C-H Chang, V. Chikarmane, M. Childs, H. Deshpande, K. Dev, G. Ding, T. Ghani, O. Golonzka, W. Han, J. He*, R. Heussner, R. James, I. Jin, C. Kenyon, S. Klopcic, S-H. Lee, M. Liu, S. Lodha, B. McFadden, A. Murthy, L. Neiberg, J. Neirynck, P. Packan, S. Pae*, C. Parker, C. Pelto, L. Pipes, J. Sebastian, J. Seiple, B. Sell, S. Sivakumar, B. Song, K. Tone, T. Troeger, C. Weber**, M. Yang, A. Yeoh, K. Zhang
Logic Technology Development, * Quality and Reliability Engineering, ** TCAD, Intel Corporation.
Phone: (503) 613-8029; Email: sanjay.natarajan@intel.com

ABSTRACT

A 32nm generation logic technology is described incorporating 2nd-generation high-k + metal-gate technology, 193nm immersion lithography for critical patterning layers, and enhanced channel strain techniques. The transistors feature 9Å EOT high-k gate dielectric, dual band-edge workfunction metal gates, and 4th-generation strained silicon, resulting in the highest drive currents yet reported for NMOS and PMOS. Process yield, performance and reliability are demonstrated on a 291 Mbit SRAM test vehicle, with 0.171μm² cell size, containing >1.9 billion transistors.

INTRODUCTION

For the past 40 years, relentless focus on transistor scaling and Moore's Law led to ever-higher transistor performance and density, translating into tremendous increases in microprocessor functionality and performance. Traditional device scaling led to a steady increase in leakages which threatened the continuance of Moore's Law. The introduction of high-k + metal gate transistors on the 45nm generation[1] broke through some of these scaling barriers. This paper reports on a 32nm process technology, including a 2nd generation high-k + metal-gate, which provides industry-leading transistor performance and density.

KEY DESIGN RULES & TECHNOLOGY FEATURES

Table 1 summarizes key design rules and layer thicknesses for the 32nm node. Contacted gate pitch, a key measure of transistor density, is scaled to 112.5nm (drawn at 116nm pitch and shrunk 3% to 112.5nm pitch), maintaining 0.7x scaling trend (Fig. 1). This is the most aggressive contacted gate pitch reported for 32nm high-performance logic technologies. 193nm immersion lithography is used at critical patterning layers to enable aggressive design rule scaling. Channel-strain techniques are used to increase drive currents, including optimizing the strain enhancement benefit of the replacement metal gate (RMG) process flow[2] (Fig. 2).

TRANSISTOR PERFORMANCE

Physical gate length scaling, from 35nm in the 45nm generation[1] to 30nm in the 32nm generation, is enabled by high-k dielectric scaling and shallow junction engineering. NMOS and PMOS Idsat/Ioff curves are shown in Fig. 3. At 1.0V Vdd, 100nA/um Ioff, and 112.5nm contacted gate pitch, NMOS and PMOS saturated drive currents are 1.55mA/um and 1.21mA/um, respectively (drive currents are higher at larger gate pitch). Idsat is improved ~14% over 45nm[1] for both NMOS and PMOS, and these are the best drive currents reported to-date for 32nm technology. NMOS and PMOS linear drive currents, shown in Fig. 4, are 0.228mA/um and 0.198mA/um, respectively, at 100nA/um Ioff, Vgs=1.0V, and Vds=50mV. These represent an improvement of 19% for NMOS and 11% for PMOS over 45nm[2]. 112.5nm pitch transistor I-V and sub-threshold characteristics are shown in Figs. 5 and 6. Sub-threshold slopes are maintained at ~100mV/decade. Fig. 7 shows NMOS and PMOS Vt vs Lgate, showing good Vt roll-off and DIBL. Cgate is reduced due to gate length scaling, and CV/I gate delay is improved 22% compared to 45nm[1] at the same leakage and Vcc.

RELIABILITY

Optimization of the high-k + metal-gate stack yields excellent reliability characteristics. Figs. 8 and 9 show PMOS NBTI and NMOS PBTI, respectively, compared to 45nm[1]. NBTI is matched to 45nm, while PBTI is improved relative to 45nm. Fig. 10 shows superior oxide breakdown relative to 45nm[1], supporting 10-15% higher E-field.

INTERCONNECTS

9 layers of Cu interconnect are employed, using low-k CDO dielectrics on the first 7 layers. The interconnect stack up to Metal-8 is shown in Fig. 11. Lower-layer metal pitches are matched to contacted-gate pitch while upper layer pitches increase progressively to provide optimal density and performance. Fig. 12 shows the M9, Via9, and C4 layers. Thick M9 is used for improved on-die power distribution. Compared to [1], interconnect capacitance is reduced by etch-stop layer scaling and lower-k CDO. Fig. 13a shows measured capacitance and resistance values at the M2 layer, with median values of 0.2fF/um and 8 ohm/um at 112.5nm pitch. Resistance at lower metal layers includes a penalty to enable high electromigration (EM) current density required for high performance products. Fig. 13b shows healthy M2 EM performance for multiple wafers.

SRAM

The 32nm yield learning vehicle is a 291Mbit SRAM featuring a 0.171μm² SRAM cell, >1.9B transistors, and an array density of 4.2Mb/mm². The first functional 32nm 291Mb SRAM was reported in September 2007 and is the largest functional SRAM reported to-date on any 32nm technology. Fig. 14 shows the 0.171μm² cell and a die photo (taken at upper-level metal layer) of the 291Mb SRAM die. The SRAM demonstrates 3.8GHz operating frequency at 1.1V (Fig. 15), and Fig. 16 shows excellent active Vmin characteristics for 3.25Mb sub-arrays (770mV median Vmin) to enable healthy low-voltage performance.

CONCLUSIONS

We have demonstrated an industry-leading 32nm CMOS technology for high-performance microprocessors with excellent transistor and interconnect performance and aggressive design rule scaling. We have shown a high-performance, high-density 291Mb SRAM test vehicle featuring 0.171μm² cell size fabricated using all 32nm process features. We have shown the highest drive currents and largest functional SRAM array reported to date. This process is on track for high-volume manufacturing in 2009.

REFERENCES

1. K. Mistry, et al., *IEDM Tech. Dig.*, p.247, 2007.
2. C. Auth, et al., *Symp. VLSI Tech. Dig.*, p. 128, 2008.
3. D. Ingerly, et al., *Intl. Interconnect Tech. Conf.*, p. 216, 2008.

Layer	Pitch (nm)	Thick (nm)	Aspect Ratio
Isolation	140.0	200	-
Contacted Gate Pitch	112.5	35	-
Metal 1	112.5	95	1.7
Metal 2	112.5	95	1.7
Metal 3	112.5	95	1.7
Metal 4	168.8	151	1.8
Metal 5	225.0	204	1.8
Metal 6	337.6	303	1.8
Metal 7	450.1	388	1.7
Metal 8	566.5	504	1.8
Metal 9	19.4µm	8µm	1.5
Bump	145.9µm	25.5µm	-

Table 1: Layer pitch, thickness and aspect ratio

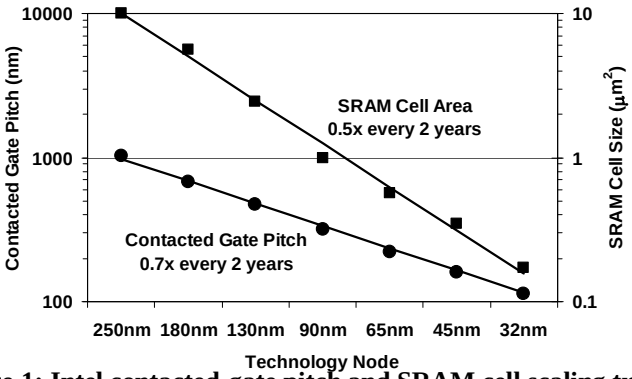

Figure 1: Intel contacted-gate pitch and SRAM cell scaling trends

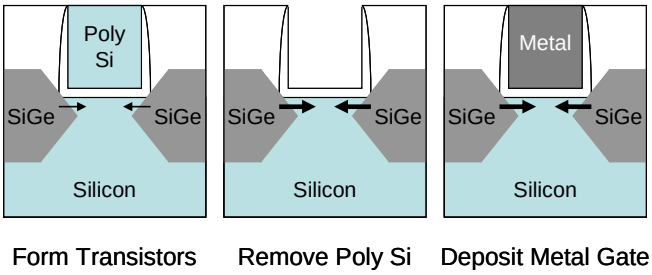

Figure 2: Increased channel strain due to RMG process flow

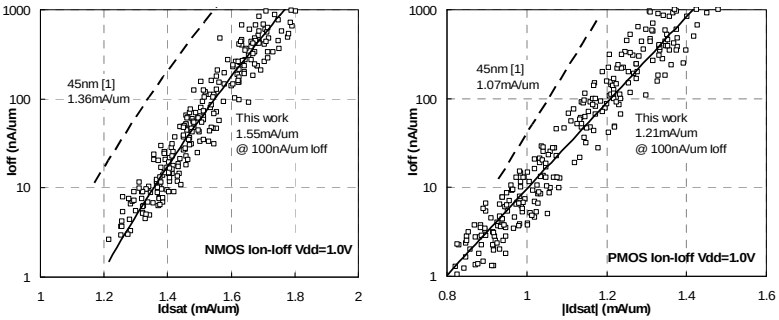

Figure 3: NMOS and PMOS Idsat vs Ioff at 1.0V

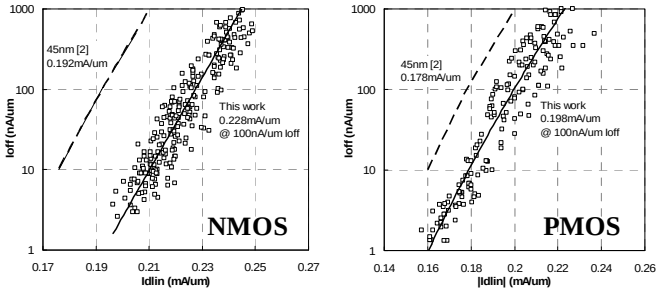

Figure 4: NMOS and PMOS Idlin vs Ioff

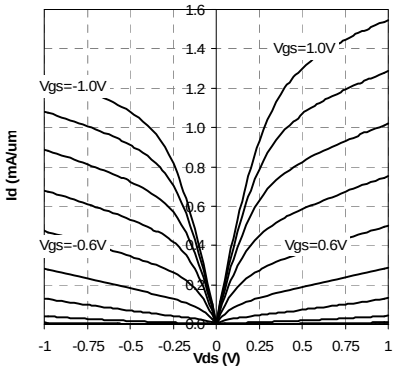

Figure 5: Transistor I-V curves

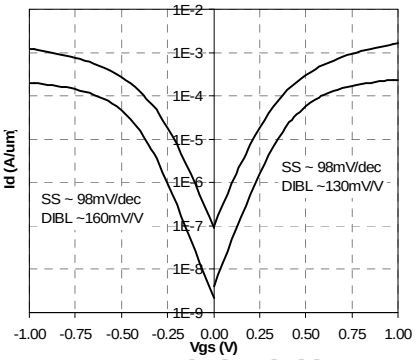

Figure 6: Sub-threshold curves

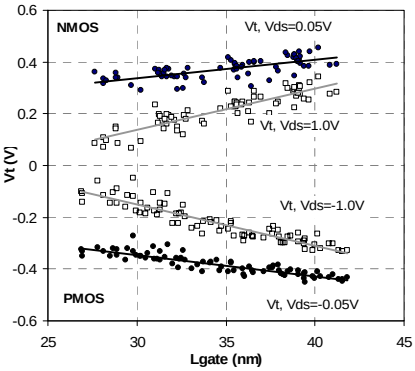

Figure 7: Vt vs Lgate

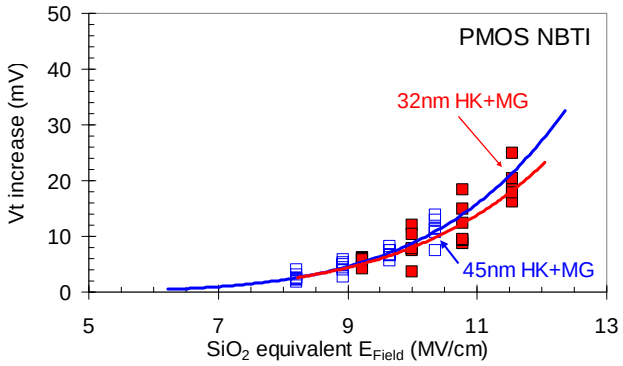

Figure 8: PMOS NBTI Vt shift vs electric field

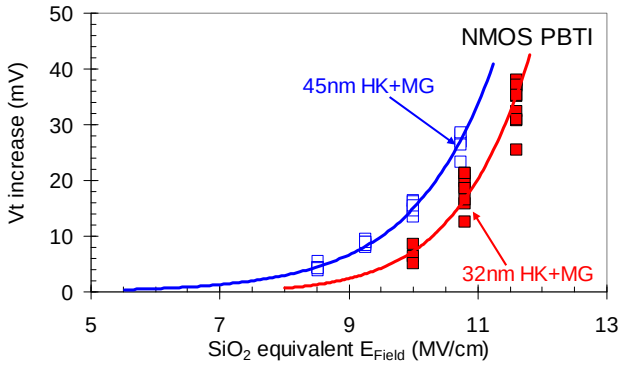

Figure 9: NMOS PBTI Vt shift vs electric field

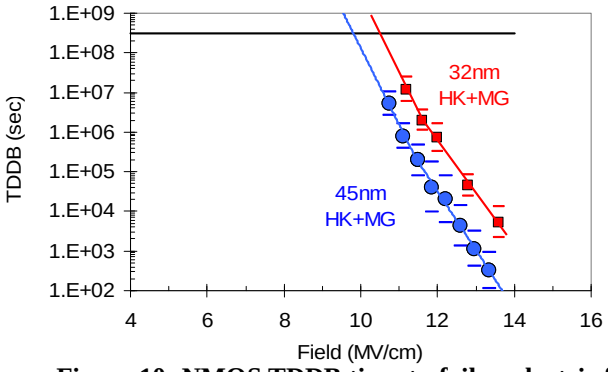

Figure 10: NMOS TDDB time-to-fail vs electric field

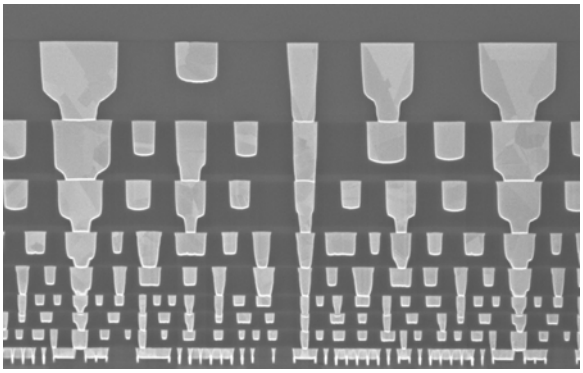

Figure 11: Cross-section of interconnect stack (8 layers)

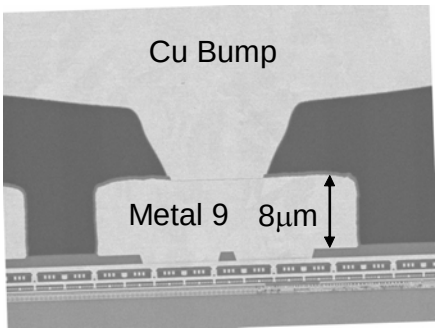

Figure 12: SEM image detailing Metal 9 and Cu Bump layers

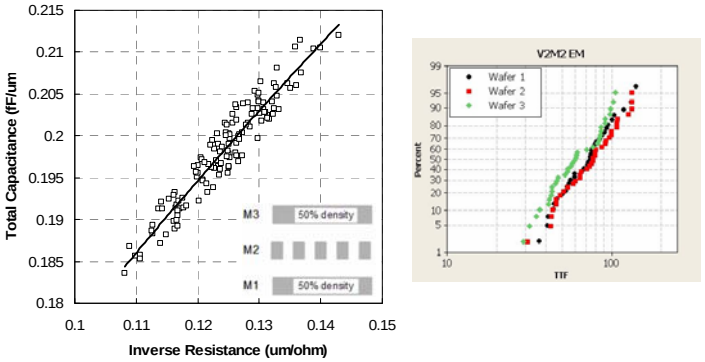

Figure 13: (a) Metal-2 min-pitch R and C values and (b) electromigration fail-rate

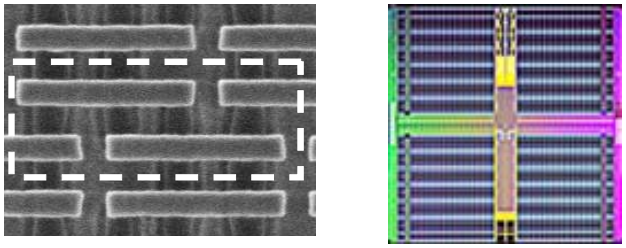

Figure 14: (a) diffusion and poly layers of 0.171µm^2 SRAM and (b) 291Mb SRAM die photo

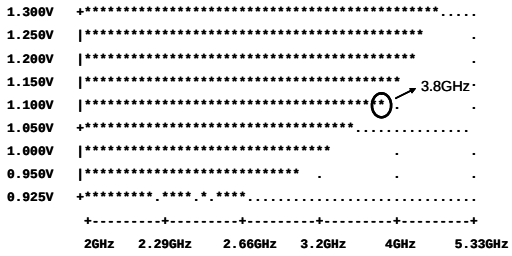

Figure 15: Voltage-frequency shmoo for 291Mb SRAM

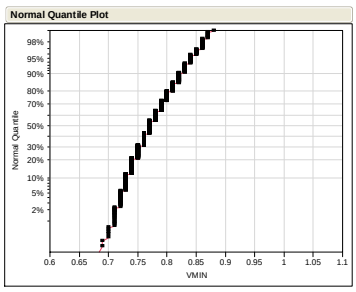

Figure 16: Vmin distribution for 3.25Mb sub-arrays