High Performance 32nm Logic Technology Featuring 2nd Generation High-k + Metal Gate Transistors

P. Packan, S. Akbar, M. Armstrong, D. Bergstrom, M. Brazier, H. Deshpande, K. Dev, G. Ding, T. Ghani, O. Golonzka, W. Han, J. He*, R. Heussner, R. James, J. Jopling, C. Kenyon, S-H. Lee, M. Liu, S. Lodha, B. Mattis, A. Murthy, L. Neiberg, J. Neirynck, S. Pae*, C. Parker, L. Pipes, J. Sebastian, J. Seiple, B. Sell, A. Sharma, S. Sivakumar, B. Song, A. St. Amour, K. Tone, T. Troeger, C. Weber**, K. Zhang, Y. Luo, S. Natarajan

Logic Technology Development, * Quality and Reliability Engineering, ** TCAD, Intel Corporation.
Intel Corporation, RA3-353, 2501 NW 229th Ave., Hillsboro, OR 97124
Phone: (503) 613-9869; Fax: (503) 613-8950; e-mail: paul.a.packan@.intel.com

Abstract:

A 32nm logic technology for high performance microprocessors is described. 2nd generation high-k + metal gate transistors provide record drive currents at the tightest gate pitch reported for any 32nm or 28nm logic technology. NMOS drive currents are 1.62mA/um Idsat and 0.231mA/um Idlin at 1.0V and 100nA/um Ioff. PMOS drive currents are 1.37mA/um Idsat and 0.240mA/um Idlin at 1.0V and 100nA/um Ioff. The impact of SRAM cell and array size on Vccmin is reported.

Technology Overview:

Continuing Moore's law to the 32nm technology node requires difficult trade-offs in gate length, S/D contact area and contact-to-gate margins. As dimensions are reduced, less area is available for contacting S/D regions leading to potential Rext increases as well as less area for introducing strain for mobility enhancement to improve device performance. To continue the historical trends of both area and performance improvement requires novel solutions. Figure 1 shows the 32 nm node is continuing the historic trend in gate pitch.

In addition to intrinsic device performance improvement, the ability to operate at low Vcc is becoming even more critical for low power products. This paper presents a high performance, 112.5nm pitch high-k + metal gate strain enhanced technology that continues Moore's law to the 32nm technology node and enables low Vccmin operation.

Figure 2 shows the layer pitch, thickness and aspect ratio for the 32nm technology. The reduction of source-drain area requires further improvements in strain and contact technologies for mobility enhancement and Rext reduction. Figure 3 schematically demonstrates one of the scaling issues. For given short channel characteristics and constant Ioff, as gate lengths are decreased the threshold voltage must be increased. The improvement in drive current due to the shorter effective channel length is offset by the reduction in overdrive (Vg-Vt). For small channel lengths and low Vcc, the reduction in overdrive dominates and device drive currents are reduced. Figure 4 shows a 9% Idlin increase at fixed Ioff for a 36nm device with lower Vt compared to a 32nm device, demonstrating the trade-off between Vt and Lgate. This shows that even higher Id than reported in this paper could have been achieved with longer Lgate. However, the density and reduced capacitance benefits of the shorter gate lengths are preferred.

Layer	Pitch (nm)	Thick (nm)	Aspect Ratio
Isolation	140	200	--
Contacted Gate	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	19.4um	8um	1.5

Figure 2 – Layer pitch, thickness and aspect ratio for the 32nm technology.

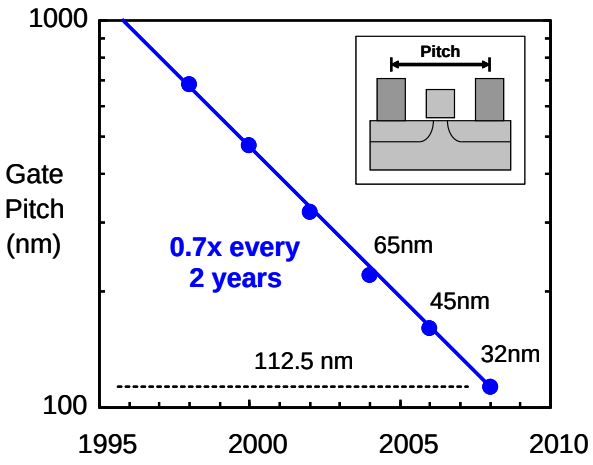

Figure 1 – Gate pitch as a function of time and technology nodes.

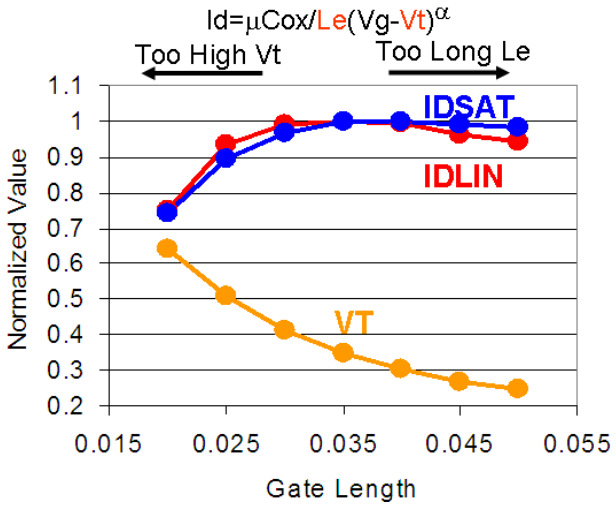

Figure 3 – Simulated drive current and Vt vs. gate length at constant Ioff.

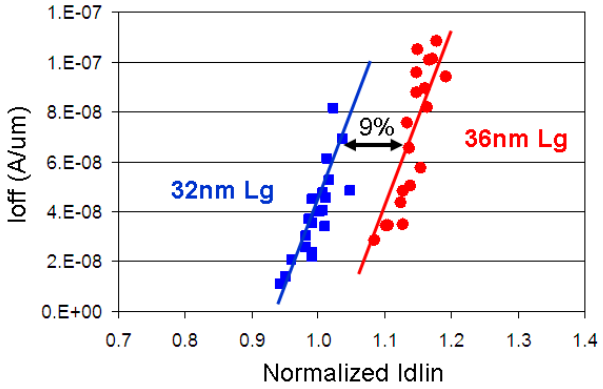

Figure 4 - Ioff vs. normalized Idlin showing a 9% Idlin increase for the 36nm device with lower Vt compared to the 32nm device.

Transistors:

This 32nm technology uses 4th generation strained silicon and 2nd generation high-k and replacement metal gate flow [1-3]. Using a replacement metal gate flow enables stress enhancement techniques to be in place before removing the poly gate from the transistor. It has been shown that this can further enhance strain and is a key benefit of this flow [2, 4]. A cross section of the NMOS and PMOS devices are shown in Fig. 5. The introduction of raised NMOS source and drain regions enables reduced device resistance helping to mitigate the pitch scaling issues discussed above. The proximity of the PMOS SiGe region to the channel continues to be reduced for enhanced channel strain. The combination of 4th generation strained silicon and 2nd generation high-k + metal gate results in PMOS saturated (Vds=1.0V) and linear (Vds=0.05V) drive currents of 1.37mA/um and 0.240mA/um at 1.0V and 100nA/um Ioff (Fig. 6). These represent a 28% improvement in Idsat and a 35% improvement in Idlin over the 45nm technology [2,3] and are the highest reported drive currents for any 32nm or 28nm technology. Furthermore, this is the first report of PMOS linear drive current exceeding NMOS and is the result of 4 generations of PMOS strain engineering enhancements. NMOS saturated and linear drive currents are 1.62mA/um and 0.231mA/um at 1.0V and 100nA/um Ioff (Fig. 7). This is a 19% increase in Idsat and a 20% increase in Idlin over the 45nm technology [2,3]. Figure 8 shows 32nm transistors continue the historic trend of increased drive current while reducing gate pitch. These drive currents are the highest reported currents at the smallest gate pitch of any 32nm or 28nm technology. Device I-V and subthreshold characteristics are shown in Fig. 9 and 10. Subthreshold slopes are maintained at ~100mV/decade. Figure 11 shows good Vt roll-off and DIBL characteristics.

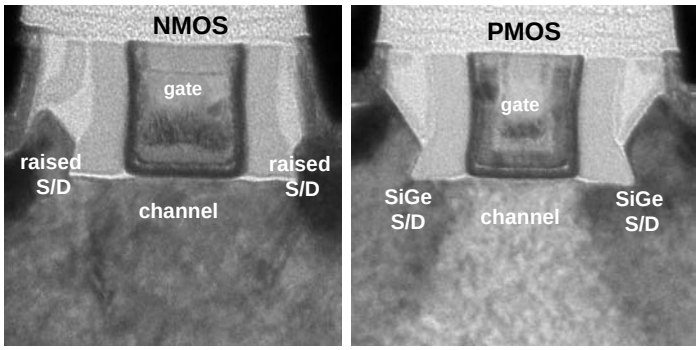

Fig. 5 - Cross section of NMOS and PMOS devices showing raised S/D regions for reduced parasitic resistance.

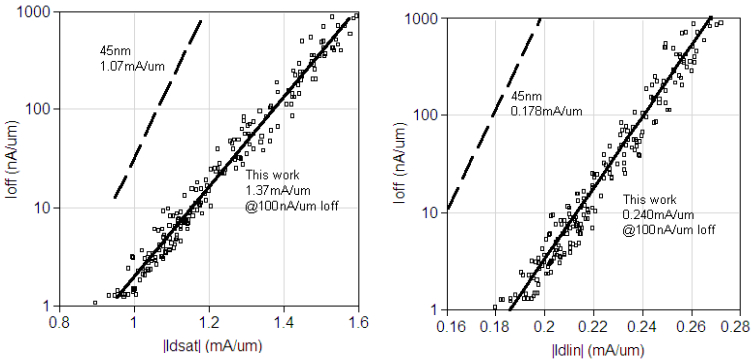

Figure 6 – PMOS Idsat and Idlin vs. Ioff at Vcc=1.0 V showing a 28% improvement in Idsat and 35% improvement in Idlin over the 45nm technology and are the highest reported drive currents for any 32nm or 28nm technology.

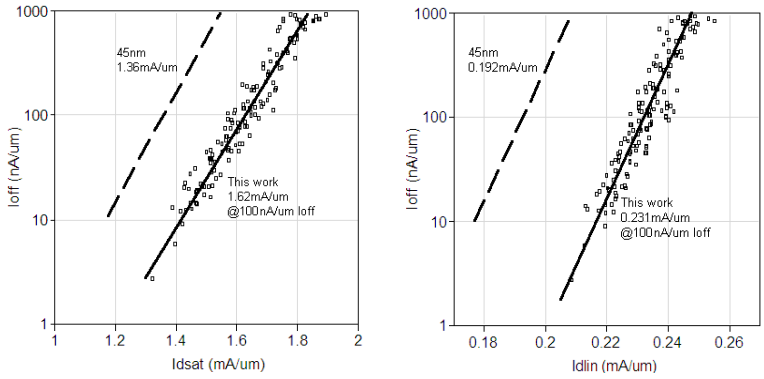

Figure 7 – NMOS Idsat and Idlin vs. Ioff at Vcc=1.0 V showing a 19% improvement in Idsat and 20% improvement in Idlin over the 45nm technology and are the highest reported drive currents for any 32nm or 28nm technology.

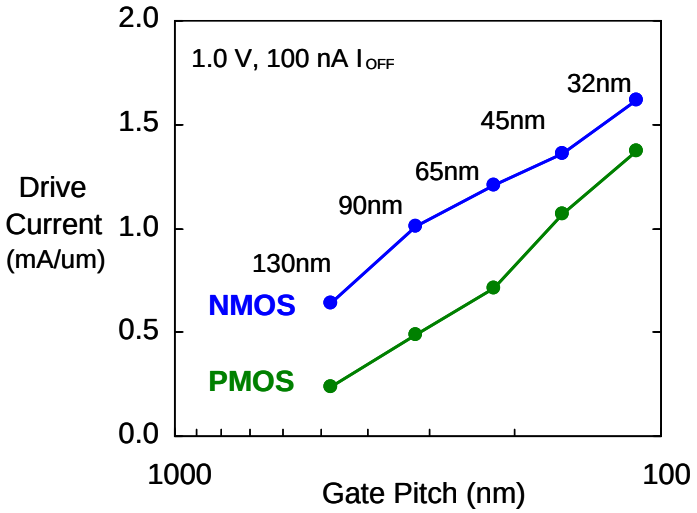

Fig. 8 – Drive current as a function of gate pitch and technology node demonstrating 32nm technology maintains the historic scaling trends.

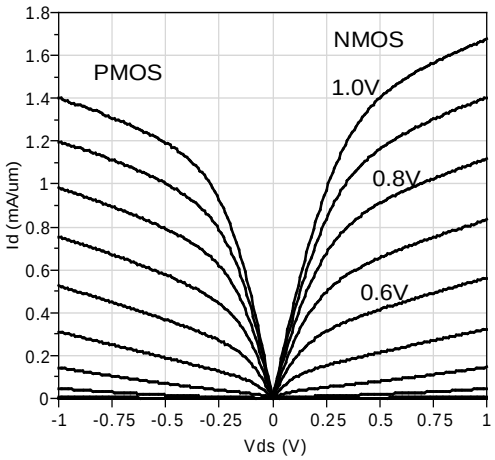

Fig. 9 – I-V characteristics for PMOS and NMOS devices.

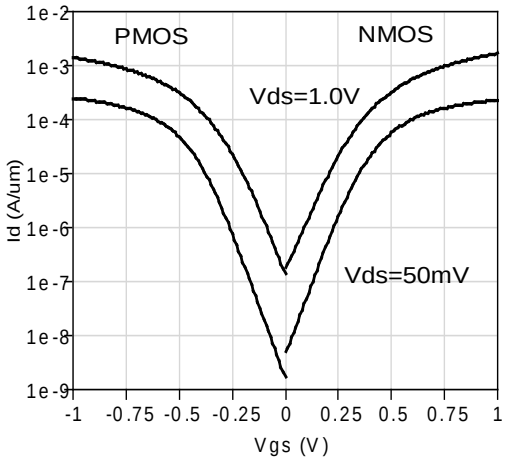

Fig. 10 - Subthreshold characteristics of PMOS and NMOS devices showing ~100mV/decade slope at Vds= 1.0V and 50mV.

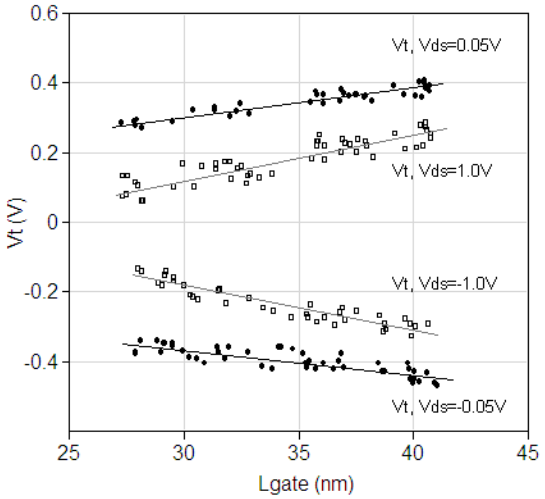

Fig. 11 – PMOS and NMOS Vt vs. gate length at Vcc=1.0V for Vds=1.0V and 50mV showing good Vt roll-off and DIBL characteristics.

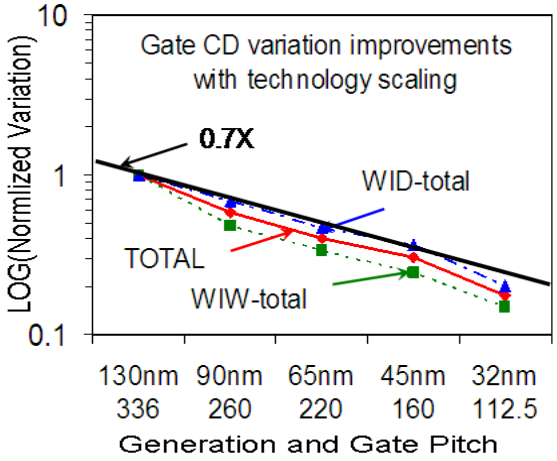

Fig. 12 – Gate length variation trend as a function of technology node showing a reduction in variation of 0.7x/generation.

SRAM:

The ability to operate at low Vcc is critical for low power applications. Device variation plays a key role in determining the minimum operating voltage (Vccmin) for SRAM and register file (RF) circuits. Figures 12 and 13 show the 32nm technology continues the trend of 0.7x reduction in gate length variation and the magnitude of both systematic and random within wafer variation does not increase for the 32nm technology. In addition to device variation, cell size, array size and intrinsic distribution determine Vccmin for memory cells. Although it is relatively easy to produce a few small SRAM arrays that operate at low Vcc, the important goal is to produce large arrays that operate at low Vcc. Figure 14 shows the Vccmin distribution for a 3.25Mb array for cell sizes of 0.171um^2, 0.199um^2 and 0.256 um^2. As expected, the larger cell sizes support smaller Vccmin due to reduced random variation for larger devices. Vccmin can easily differ by 150 mV depending on the distribution percentage reported. Figure 15 shows the effect of array size on Vccmin for array sizes of 3.25Mb and 91Mb for a 0.199um^2 cell. The shift of 70mV is due to the statistics from the larger number of cells. The Vccmin data above is for active memory operation. For standby mode where the memory cell must only hold the bit, Vccmin can be more than 200mV lower (Fig. 16).

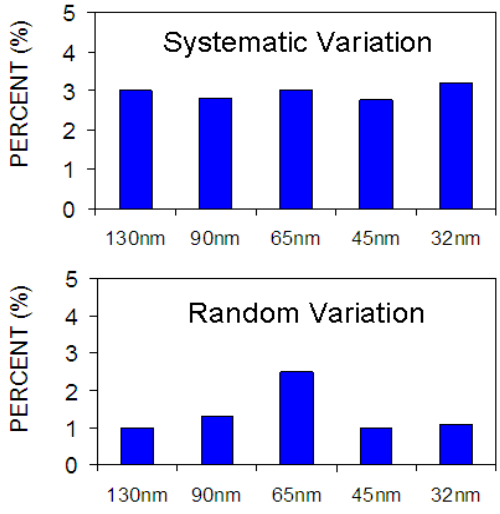

Fig. 13 – Within wafer systematic and random variation based on ring oscillator structures vs. technology node. The 32nm node has similar variation compared to the 45nm node.

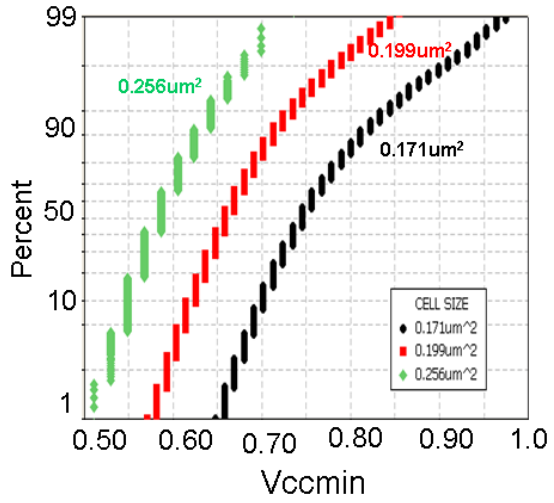

Fig. 14 – Vccmin distribution for a 3.25Mb array for cell sizes of 0.171um^2, 0.199um^2 and 0.256um^2. As expected, the larger cell sizes support smaller Vccmin.

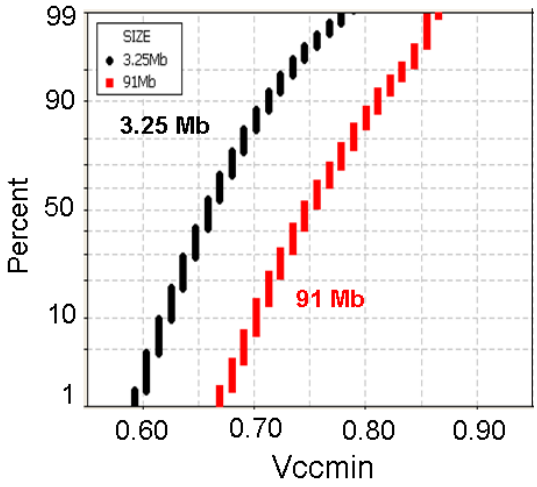

Fig. 15 – Vccmin distribution for array sizes of 3.25Mb and 91Mb for the 0.199um^2 cell. The shift of 70mV is due to the statistics from the larger number of cells.

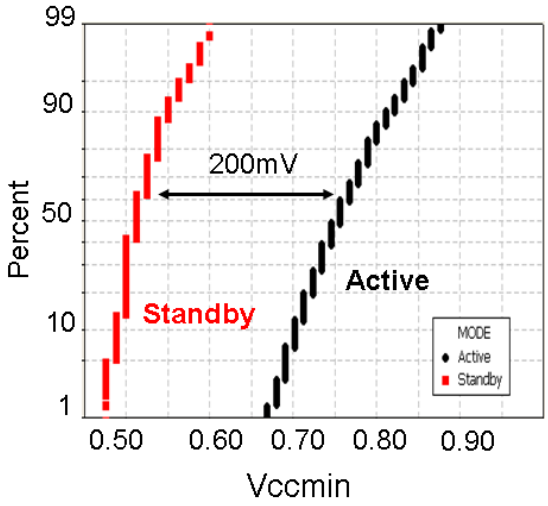

Fig. 16 – Vccmin distribution comparing active and standby modes. Standby mode can have a 200mV lower Vccmin.

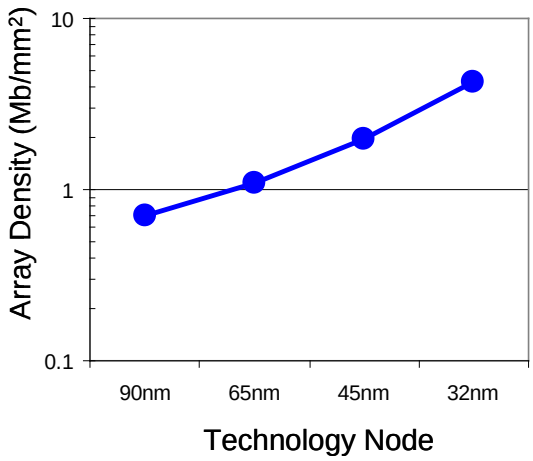

Fig. 17 – Array density vs. technology node demonstrating excellent scaling and the highest reported array density for 32nm or 28nm technologies.

When evaluating and comparing low voltage memory operation, cell size, Vccmin distribution, and array size must all be considered. Array density, which includes memory cells, sense amps and control circuitry, is another important SRAM metric to report. Figure 17 shows array density versus technology node, demonstrating the highest reported array density for a 32nm or 28nm technology of 4.2Mbit/mm^2 which is 2x the density of the previous 45nm generation.

Conclusion:

A high performance 32nm logic technology with 2nd generation high-k + metal gate transistors is presented. Record NMOS and PMOS drive currents are reported, along with the tightest contacted gate pitch for a 32nm or 28nm technology. Variation for the 32nm technology was shown to be the same as the 45nm technology. Excellent Vccmin and the highest reported SRAM array density for 32nm or 28nm technology was reported. Fully functional 32nm processors using this technology were demonstrated in systems in Jan 2009 and are on track for volume production in 2H 2009. This 32nm technology maintains the historical scaling trends of both area and performance and continues Moore's law.

References:

[1] S.Natarajan et al., IEDM Tech. Dig., p. 941, (2008).
[2] C.Auth et al., Symp. VLSI Dig., p 128, (2008).
[3] K.Mistry et al., IEDM Tech. Dig., p.247, (2007).
[4] J.Wang et al., VLSI Sym. Dig, p 46, (2007).