A 22nm High Performance and Low-Power CMOS Technology Featuring Fully-Depleted Tri-Gate Transistors, Self-Aligned Contacts and High Density MIM Capacitors

C. Auth, C. Allen, A. Blattner, D. Bergstrom, M. Brazier, M. Bost, M. Buehler, V. Chikarmane, T. Ghani, T. Glassman, R. Grover, W. Han, D. Hanken, M. Hattendorf, P. Hentges, R. Heussner, J. Hicks*, D. Ingerly, P. Jain, S. Jaloviar, R. James, D. Jones, J. Jopling*, S. Joshi, C. Kenyon, H. Liu, R. McFadden, B. McIntyre, J. Neirynck, C. Parker, L. Pipes, I. Post, S. Pradhan, M. Prince, S. Ramey*, T. Reynolds, J. Roesler, J. Sandford, J. Seiple, P. Smith, C. Thomas, D. Towner, T. Troeger, C. Weber**, P. Yashar, K. Zawadzki, K. Mistry
Logic Technology Development, *Quality and Reliability Engineering, **TCAD, Intel Corp., Hillsboro, OR, U.S.A.
Contact: email chris.auth@intel.com

Abstract

A 22nm generation logic technology is described incorporating fully-depleted tri-gate transistors for the first time. These transistors feature a 3rd-generation high-k + metal-gate technology and a 5th generation of channel strain techniques resulting in the highest drive currents yet reported for NMOS and PMOS. The use of tri-gate transistors provides steep subthreshold slopes (~70mV/dec) and very low DIBL (~50mV/V). Self-aligned contacts are implemented to eliminate restrictive contact to gate registration requirements. Interconnects feature 9 metal layers with ultra-low-k dielectrics throughout the interconnect stack. High density MIM capacitors using a hafnium based high-k dielectric are provided. The technology is in high volume manufacturing.

Introduction

In the era of classical scaling, transistor performance improved primarily as a result of dimensional scaling. In the past decade, performance has progressed through introduction of transistor architecture innovations, including strained silicon [1] and high-k/metal-gate technologies [2]. Multi-gate devices have long held the promise of improved transistor electro-statics, offering improved performance at lower supply voltages and significantly reduced short channel effects [3, 4]. While the FINFET featured gate control on two sides of a fin, the Tri-Gate transistor extended gate control to three sides of the fin [5]. This 22nm process technology is the first to exploit fin based Tri-Gate devices and combine their benefits with strained silicon and high-k/metal-gate. The standard scaling requirements for the strained silicon components and for the gate and contact pitches also needs to be addressed at the 22nm node. 193nm immersion lithography is used extensively to achieve the 8nm channel fin widths along with the tight gate, contact and interconnect pitches. Use of self-aligned contacts enables aggressive scaling of the gate pitch while optimizing the gate length, spacer width and contact width for device considerations.

Design Rules and Technology Features

Table I summarizes the key design rules. Contacted gate pitch is scaled to 90nm and the SRAM cell size is reduced to 0.092μm^2, maintaining traditional scaling trends (Fig. 1). Optimization of fin width and height is a key consideration in the tri-gate device. The fin width must be sufficiently thin to be fully-depleted and an even narrower fin width is desired for improved short-channel effects (SCE). A fin width of 8nm is used to balance SCE and Rext (Fig. 2). To reduce the Rext impact of the narrow fins further, in-situ doped raised S/Ds are used. A fin height of 34nm is chosen to balance drive current vs. capacitance.

For the PMOS device, SiGe S/Ds are employed to induce compressive stress in the channel (Fig. 3). The Ge concentration of the SiGe S/D was increased from 40% in our 32nm technology [6] to 55% in 22nm. The PMOS also benefits from the higher mobility of the <110> sidewalls and the strain enhancement from the gate last process. Stress enhancement has also been incorporated in the NMOS device to improve performance.

Transistor Performance & Reliability

To support low power SoC integration with high performance microprocessors, three transistor types are offered. HP devices are targeted for high performance with leakage in the 20-100nA range, based on product need. MP and SP transistors offer lower leakage to enable SoC product power-performance optimization.

	HP	MP	SP
$T_{OX,E}$ (nm)	0.9	0.9	0.9
L_{GATE} (nm)	30	34	34
I_{OFF} (nA/um)	20-100	5-20	1-5

The improved gate control of the tri-gate structure can be seen in the steep subthreshold slopes (~70 mV/dec.) and very low DIBL (~50 mV/V) for minimum Lgate devices (Fig. 4). These enhanced transistor characteristics can be used to reduce leakage, improve performance and/or enable lower operating voltage for reduced active power. The 22nm generation tri-gate transistors can support threshold voltages more than 100 mV lower than the previous 32nm planar technology (Fig. 5). The lower threshold voltage combined with the strain enhancement leads to an increase of 13% Idsat on NMOS at 0.8V and 10nA/μm compared to 32nm planar transistors. The lower threshold voltage impact is more apparent in Ieff (Vdd=0.8V) [7] where a 46% drive current improvement is demonstrated (Fig. 6). PMOS shows similar drive current improvements of 27% Idsat and 40% Ieff (Fig. 7).

Use of the tri-gate structure introduces new corners and sidewall orientations that need careful optimization of the high-k + metal-gate stack to produce excellent reliability. Both NMOS and PMOS oxide breakdown is improved compared to 32nm (Fig. 8). NMOS PBTI is also improved compared to 32nm (Fig. 9) and the net BT shift for combined N and PBTI is matched to 32nm.

Self-Aligned Contacts, Interconnects and MIM Capacitor

The self-aligned contact (SAC) process is enabled by recessing the gate metal after it has been planarized. After the gate has been recessed, a silicon nitride etch stop is deposited and planarized. This is followed by a capping oxide prior to contact patterning. The contacts are etched selectively to the silicon nitride protecting the gates. A TEM of contacts that have been intentionally overlaid on the gates, illustrates the ability of the selective etch to avoid shorting to the gates (Fig 10). The 9 layers of copper interconnect, including two layers with 80nm minimum pitch using cost-effective single patterning, are used in conjunction with a low-k or an ultra-low-k dielectric at all layers providing 13-18% lower capacitance than 32nm (Fig. 11) [8]. A MIM decoupling capacitor utilizing a Hi-K dielectric is introduced between MT8 and MT9 for power and signal applications (Fig. 12).

SRAM/Microprocessors & Conclusion

The 22nm yield learning vehicle was a 380Mb SRAM featuring a three SRAM cells, a High Density (HD) 0.092μm^2 cell, a Low Voltage (LV) 0.108μm^2 cell and a High Performance (HP) 0.130μm^2 cell. The SRAM demonstrates 4.6GHz operating frequency at 1V. High Yield has also been demonstrated on several microprocessors

Tri-Gate transistors have been integrated with High-k gate oxide, metal gates and advanced strain techniques into a manufacturable 22nm CMOS process with 8nm minimum features. The technology also features an ultra-low K dielectric and MIM capacitor, and is now in volume manufacturing in multiple factories.

References

1. T. Ghani, et al, IEDM Technical Digest, p. 978-980 (2003)
2. K. Mistry, et al, IEDM Technical Digest, p. 247-250 (2007)

References Continued

3. D. Hisamoto et al, IEDM Technical Digest, p.833-836 (1989)
4. X. Huang et al, IEDM Technical Digest, p. 67-70 (1999)
5. R. Chau et al, Solid-State Devices and Mat., p. 68-69 (2002)
6. P. Packan et al, IEDM Technical Digest, p. 659-662 (2009)
7. M. H. Na, et al, IEDM Technical Digest, p. 121-124, (2002)
8. D. Ingerly, Submitted 2012 IITC

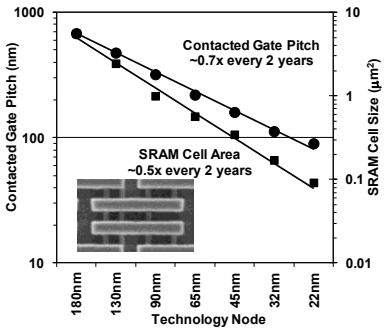

Fig.1 Contacted Gate Pitch and SRAM cell size scaling trends

Layer	Pitch (nm)
Fin	60
Contacted Gate	90
Metal 1	90
Metal 2	80
Metal 3	80
Metal 4	112
Metal 5	160
Metal 6	240
Metal 7	320
Metal 8	360
Metal 9	14μm

Table I: Layer Pitches

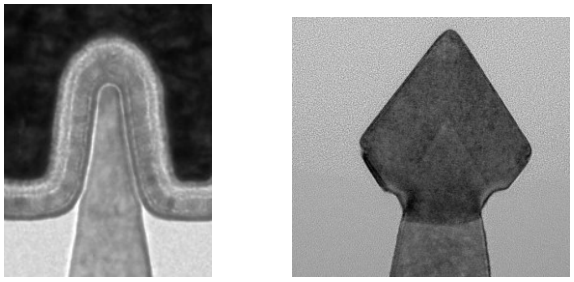

Fig.2 TEMs of the PMOS channel under the gate (left) and in the S/D region (right) showing the SiGe epitaxy in the S/D region.

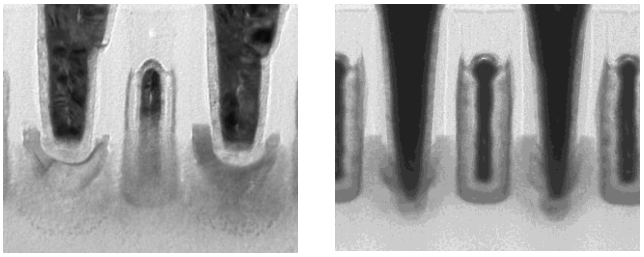

Fig.3 TEMs of tri-gate NMOS (left) and PMOS (right) transistors. Wrap-around of the contacts on S/D leads to illusion of contacts penetrating S/D on PMOS TEM.

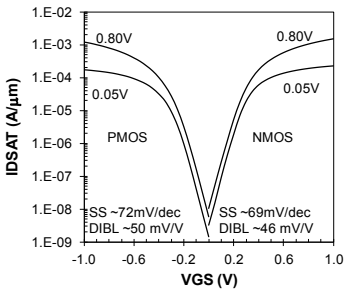

Fig.4 Subthreshold Id-Vgs for both NMOS and PMOS transistors.

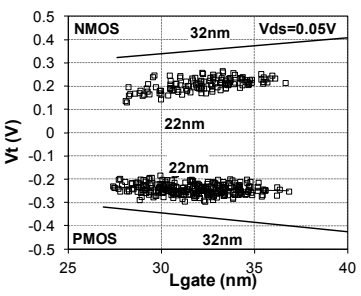

Fig.5 Linear threshold voltage vs. gate length for NMOS and PMOS transistors.

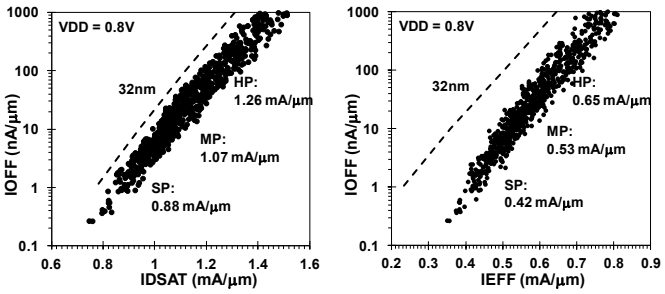

Fig.6 NMOS Ion-Ioff showing 13% Idsat improvement and 46% Ieff improvement relative to 32nm [6] at 0.8V and 10nA. HP, MP and LP devices are benchmarked at 100nA, 10nA and 1nA Ioff respectively.

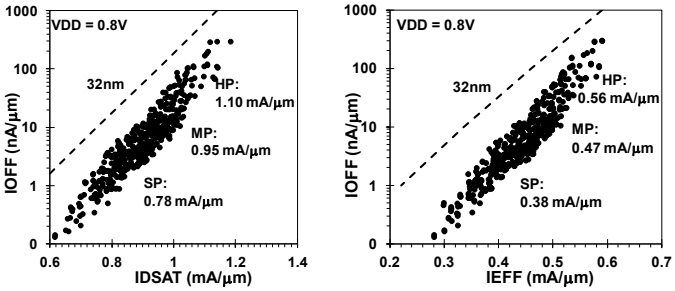

Fig.7 PMOS Ion-Ioff showing 27% Idsat improvement and 40% Ieff improvement relative to 32nm [6] at 10nA and 0.8V.

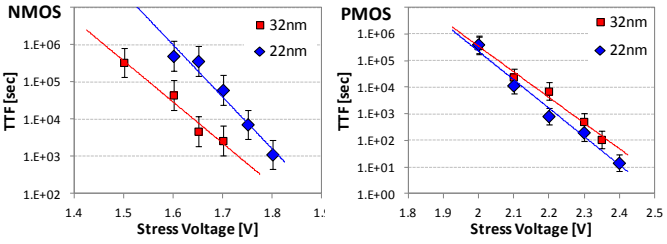

Fig. 8 TDDB for NMOS (left) and PMOS (right) compared to 32nm [6].

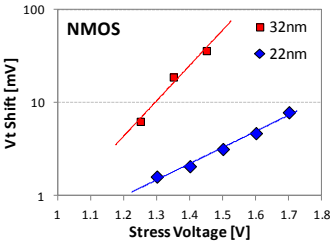

Fig. 9 NMOS PBTI comparison to 32nm [6].

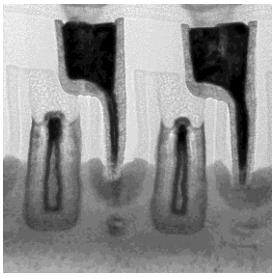

Fig. 10 TEM showing mis-aligned contacts avoiding the gate and contacting the S/D regions.

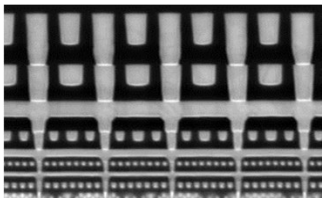

Fig. 11 Metal-1 to Metal-8 in cross-section, showing via and trench profiles.

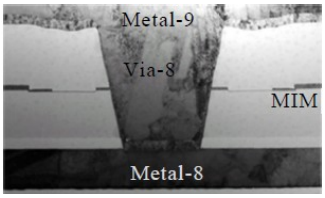

Fig. 12 TEM of Via-8, showing a MIM capacitor embedded in silicon nitride, and sidewall connections to the electrode.