A 10nm High Performance and Low-Power CMOS Technology Featuring 3rd Generation FinFET Transistors, Self-Aligned Quad Patterning, Contact over Active Gate and Cobalt Local Interconnects

C. Auth, A. Aliyarukunju, M. Asoro, D. Bergstrom, V. Bhagwat, J. Birdsall, N. Bisnik, M. Buehler, V. Chikarmane, G. Ding, Q. Fu, H. Gomez, W. Han, D. Hanken, M. Haran, M. Hattendorf, R. Heussner, H. Hiramatsu, B. Ho, S. Jaloviar, I. Jin, S. Joshi, S. Kirby, S. Kosaraju, H. Kothari, G. Leatherman*, K. Lee, J. Leib, A. Madhavan, K. Marla, H. Meyer, T. Mule, C. Parker, S. Parthasarathy, C. Pelto, L. Pipes, I. Post, M. Prince, A. Rahman*, S. Rajamani, A. Saha, J. Dacuna Santos*, M. Sharma, V. Sharma, J. Shin, P. Sinha, P. Smith, M. Sprinkle, A. St. Amour, C. Staus, R. Suri, D. Towner, A. Tripathi, A. Tura, C. Ward, A. Yeoh
Logic Technology Development, *Quality and Reliability Engineering, Intel Corporation
Contact: email chris.auth@intel.com

ABSTRACT

A 10nm logic technology using 3rd-generation FinFET transistors with Self-Aligned Quad Patterning (SAQP) for critical patterning layers, and cobalt local interconnects at three local interconnect layers is described. For high density, a novel self-aligned contact over active gate process and elimination of the dummy gate at cell boundaries are introduced. The transistors feature rectangular fins with 7nm fin width and 46nm fin height, 5th generation high-k metal gate, and 7th-generation strained silicon. Four or six workfunction metal stacks are used to enable undoped fins for low Vt, standard Vt and optional high Vt devices. Interconnects feature 12 metal layers with ultra-low-k dielectrics throughout the interconnect stack. The highest drive currents with the highest cell densities are reported for a 10nm technology.

INTRODUCTION

In the past decade, innovations in transistor architecture have accelerated transistor performance from the introduction of strained silicon [1] to high-k/metal-gate [2] to the introduction of multi-gate, FinFET devices [3]. Despite the slowing of lithography scaling, transistor density scaling has accelerated in Intel's 14nm [4] and 10nm technologies. The traditional node notation has lost its ability to adequately measure the transistor density benefit between technologies. By using a transistor density metric that combines the NAND and scan flip-flop densities (Fig. 1), the acceleration in transistor density starting with a 2.5x scaling at the 14nm generation can be seen. This 10nm technology continues that acceleration at 2.7x from the 14nm generation through use of SAQP w/193nm immersion lithography, improved transistor matching to enable fewer fins in the standard cell library and novel process features to enable tighter layout (Fig 2).

KEY DESIGN RULES & TECHNOLOGY FEATURES

Table I summarizes the key design rules. Contacted gate pitch is scaled to 54nm and the minimum 6-T SRAM cell size is reduced to 0.0312μm^2, maintaining traditional scaling trends (Fig. 3). SAQP is introduced at the diffusion, metal-0 and metal-1 layers to achieve tight fin pitches down to 34nm and metal pitches of 36nm with 193nm immersion lithography. Scaling of density critical interconnect layers is up to 0.51x vs. the traditional 0.7x.

To maximize the density scaling at the 10nm generation, several additional architecture features have been added. The first eliminates the dummy gate that has been present at the cell boundaries. By introducing a minimum isolation step at the boundary, the two neighboring cell transistors are isolated by the width of a single gate (Fig. 4).

The next feature is to place the contact to the gate over the active area of the device (Fig. 5). This improves the density by eliminating the need to extend the gate over the isolation to make the contact.

In order to place the gate contact over the active area, both the diffusion contact and gate contacts are formed with a novel self-aligned contact process (Fig. 6). Self-aligned diffusion contacts have been used in high volume manufacturing since the 22nm generation to enable the tight contact to gate overlay requirements. Self-aligned diffusion contacts are formed by recessing the gate metal and then depositing a silicon nitride etch-stop on top of the metal to prevent the diffusion contact from shorting to the gate. With this technology an additional recess of the diffusion contact is added and a silicon carbide etch-stop layer is deposited to prevent the gate contact from shorting to the diffusion contact.

With the aggressive scaling of the contact space, contact resistance (Rext) is a key consideration in optimization of the transistor. In this technology, the contact metal stack has introduced two features to lower the contact resistance. The first is replacing the tungsten contact metal with cobalt. This provides a 60% reduction in contact line resistance. The second is adding a conformal titanium layer to wraparound the source/drain diffusion regions to lower the spreading resistance. A thin NiSi layer has also been added to lower the PMOS contact resistance.

The fin patterning uses SAQP to achieve a 7nm fin width at a 34nm pitch (Fig. 7). In-situ doped, raised S/Ds are used to provide low Rext along with improved mobility from strain enhancement. To balance drive current vs. capacitance,

a fin height of 46nm has been used. A low-k gate spacer has been introduced to reduce the parasitic contact-gate capacitance by 10%. Four or six workfunction metals are used depending on product need. This enables undoped fins for low Vt, standard Vt and optional high Vt devices providing improved mobility, short channel effects and transistor matching for all transistors. The improved transistor matching enables aggressive reduction in fin usage, improving transistor density.

TRANSISTOR PERFORMANCE AND VARIATION

The 3rd generation FINFETs show the characteristic steep subthreshold slopes (~70 mV/dec.) and very low DIBL (~70 mV/V) for minimum Lgate devices (Fig. 8). The transistor performance is enhanced by the use of stress enhancement techniques from in-situ doped strained epitaxy in the S/D. A novel ILD0 stress through the gate, orthogonal to the fin, is used to enhance the NMOS drive by an additional 5%. The combination of the stress enhancement along with the contact resistance reduction techniques and the reduced fin pitch leads to a NMOS Idsat of 1.78mA/μm at 0.7V and 10nA/μm (Fig 9). This is an increase of 71% Idsat on NMOS compared to 14nm FINFET transistors. Similarly, NMOS Idlin is 0.475mA/μm, an increase of 100% compared to 14nm FINFET transistors. PMOS shows similar drive current improvements of 35% Idsat and 55% Idlin (Fig. 10). Transistor I-V curves are shown in figure 11.

RELIABILITY

Optimization of the high-k + metal-gate stack yields excellent reliability characteristics. Fig. 12 shows NMOS and PMOS TDDB, compared to 14nm. Both show a clear improvement relative to 14nm.

INTERCONNECTS

The interconnect stack with 12 layers of interconnect is shown in Fig. 13. Cobalt is introduced at the lowest two interconnect layers providing a 5-10x improvement in electromigration and a 2x reduction in via resistance. SAQP is used at the lowest two metal layers to achieve a 40nm metal-0 pitch and a 36nm metal 1 pitch. Self-aligned Double patterning (SADP) is used at Metal 2-Metal 5 to enable a 44nm metal pitch with scaling of up to 0.51x compared to 14nm. A cobalt cladding layer is utilitzed at Metal 2 – Metal 5 to improve electromigration. Low-k CDO dielectrics are used on 11 layers.

SRAM, PRODUCT AND YIELD

The 10nm yield learning vehicle was a 204Mb SRAM featuring a three SRAM cells, a High Density (HD) 0.0312μm^2 cell, a Low Voltage (LV) 0.0367μm^2 cell and a High Performance (HP) 0.0441μm^2 cell. Yield has been demonstrated on an SRAM test vehicle and on microprocessors.

CONCLUSIONS

This paper presents an industry-leading 10nm CMOS technology with excellent transistor and interconnect performance and aggressive design rule scaling. This is the densest CMOS technology demonstrated to date. The process features 3rd generation FinFETs with optimized fin profiles at 34nm Fin pitch using SAQP and 54nm Gate pitch using SADP. The interconnect stack features self-aligned contacts/vias with a 36nm minimum metal pitch formed with SAQP, cobalt interconnects at critical layers, and aggressive design rule scaling from the 14nm generation. We have shown a high-performance, high-density SRAM featuring 0.0312μm^2 cell size fabricated using all 10nm process features.

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)
3. C. Auth, et al., Symp. VLSI Techical Digest, p. 131-132 (2012)
4. S. Natarajan, et al., IEDM Technical Digest, p. 71-73 (2014)

Layer	Pitch (nm)	Scaling
Fin	34	0.8x
Contacted Gate	54	0.77
Metal 0	40	0.71
Metal 1	36	0.51
Metal 2	44	0.85
Metal 3	44	0.79
Metal 4	44	0.55
Metal 5	52	0.51
Metal 6	84	0.53
Metal 7	112	0.70
Metal 8	112	0.70
Metal 9	160	0.63
Metal 10	160	0.63
TM0	1080	1.0
TM1	11μm	0.78

Table 1: Layer Pitches

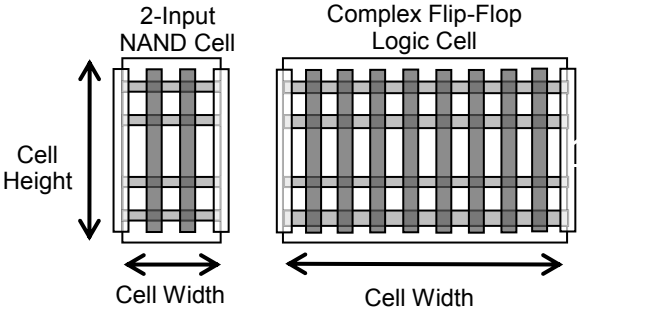

Figure 1: Definition of transistor density metric

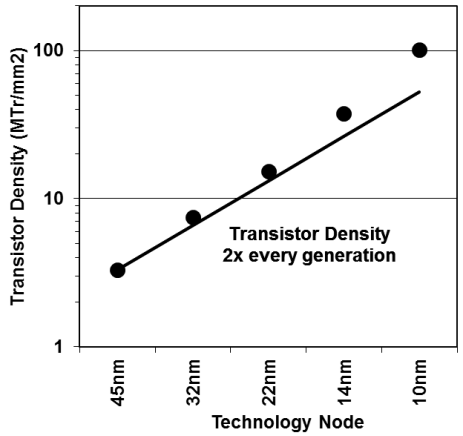

Figure 2: Logic Transistor Density Multi-Generation Scaling Trend

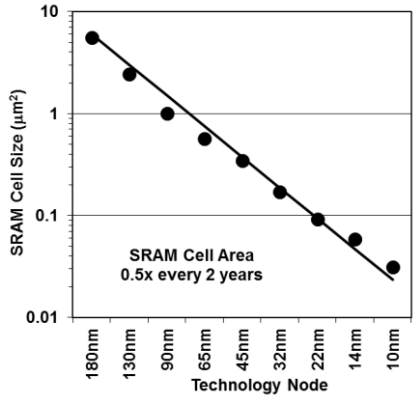

Figure 3: SRAM Scaling Trend

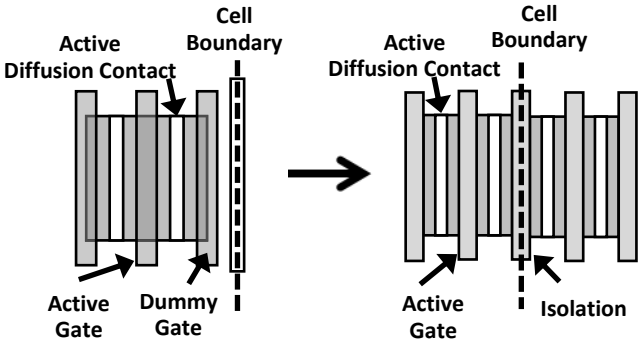

Figure 4: Dummy Gate elimination layout comparison

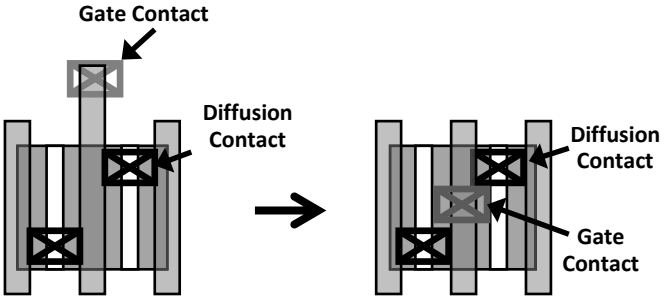

Figure 5: Contact over active gate layout comparison

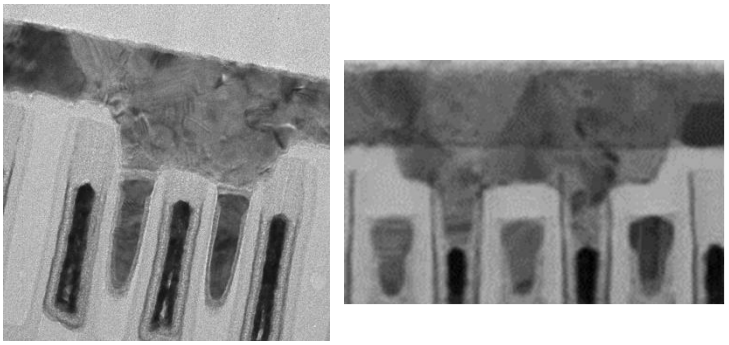

Figure 6: Self-aligned Diffusion and Gate contact images (not same scale)

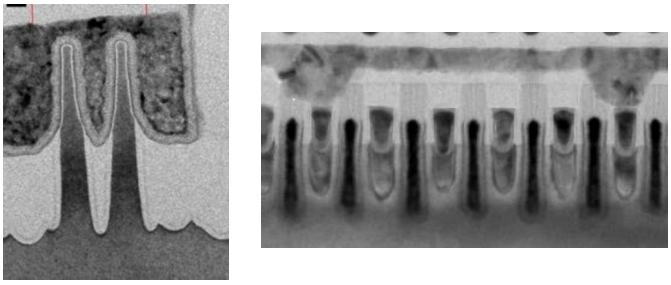

Figure 7: Transistor Fin and Gate-Cut Images (not same scale)

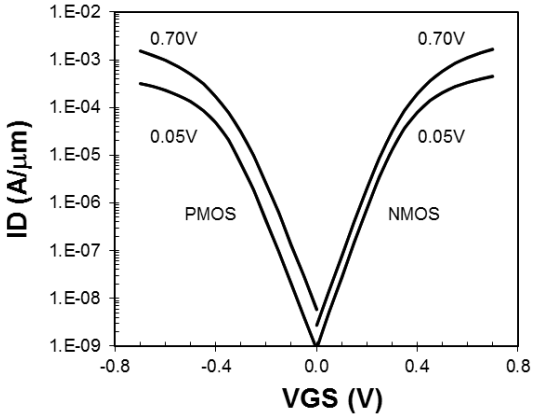

Figure 8: Subthreshold Curves

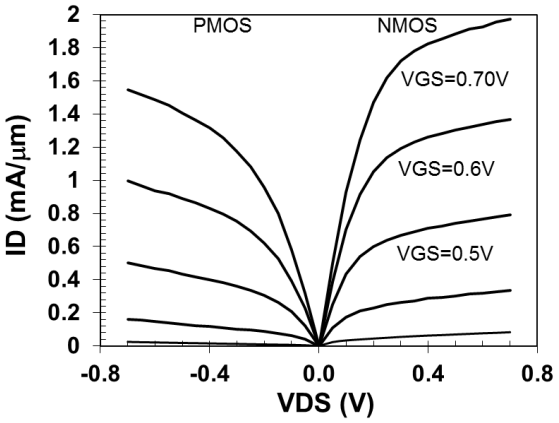

Figure 11: Transistor I-V Curves

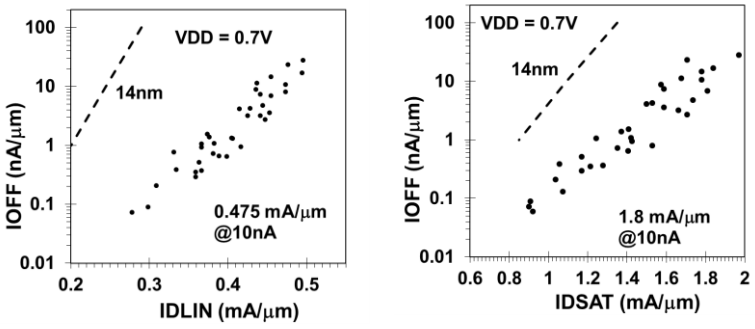

Figure 9: NMOS Idsat and Idlin curves

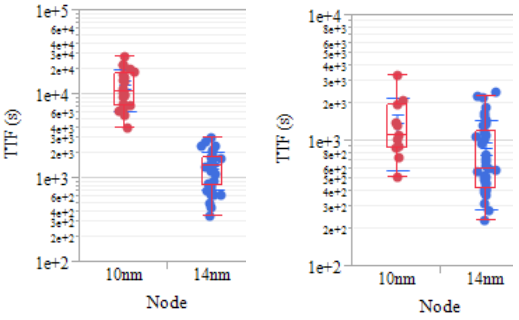

Figure 12: PMOS (left) and NMOS (right) TDDB

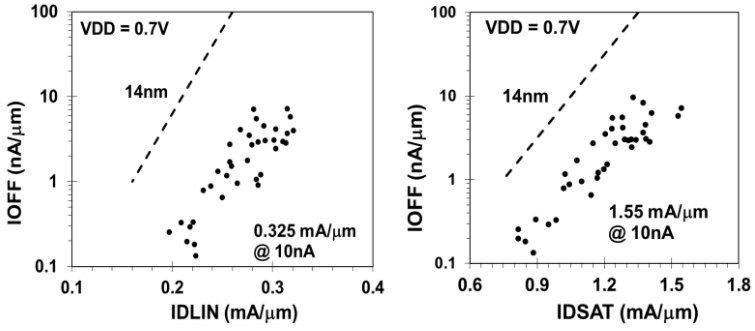

Figure 10: PMOS Idsat and Idlin curves

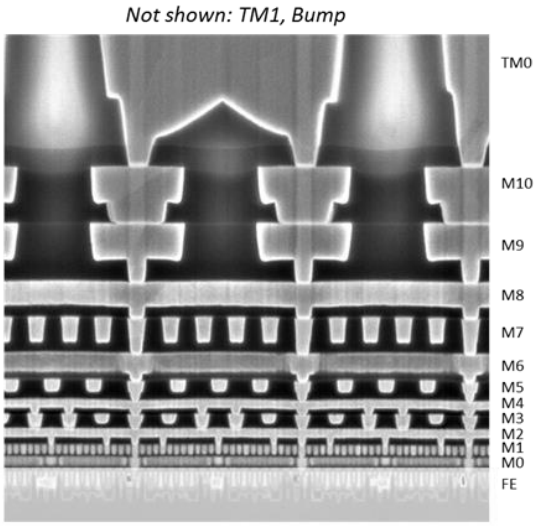

Figure 13: Interconnect Stack