Simplistic Simulation-Based Device-VT-Targeting Technique to Determine Technology High-Density LELE-Gate-Patterned FinFET SRAM in Sub-10 nm Era

Sushil Sudam Sakhare, *Member, IEEE*, Kenichi Miyaguchi, *Member, IEEE*, Praveen Raghavan, *Member, IEEE*, and Abdelkarim Mercha

***Abstract*— For the first time, we present complete device threshold voltage (VT)-targeting methodology for FinFET SRAM in 10-nm technology, considering capacitance due to metal patterning and device variability to set target read current for different variants of SRAM architecture to determine technology high-density (HD) SRAM cell. The VT-targeting methodology brings into play the worst case read and write margins available for SRAM cell to determine nominal device VT by tuning the work function of metal gate. Analysis shows that for minimum leakage current, 112 SRAM cell is optimum, whereas for the same area of 0.0546 μm^2 with 50% higher leakage, 122 SRAM outperform by 5% and 20% improved read and write margins, respectively. The 122 SRAM as HD cell reduces the cost of the technology by sharing P-channel field effect transistor (PFET) and N-channel field effect transistor (NFET) VT mask with the high threshold voltage logic devices, whereas the 112 SRAM device shares only NFET VT mask. The 111 SRAM can achieve target performance at lesser area of 0.048 μm^2 by compromising read stability, which will result in lower yield. At 64-nm pitch, litho-etch litho-etch (LELE) double-patterned gate impacts device performance and alleviates variability; hence the read margin of SRAM cell should consider an additional $1\sigma_{rsnm}$ margin to retain the same yield in 10-nm-technology era.**

***Index Terms*— 10-nm technology, design methodology, design technology co-optimization (DTCO), double patterning gate, FinFET SRAM design, high density (HD), self-aligned double patterning (SADP), self-aligned quadruple patterning (SAQP), SRAM threshold voltage (VT) targeting, technology SRAM.**

I. Introduction

ENABLING designs using 193i lithography restricts patterning of technology components. Fin patterning and formation on bulk with 36-nm fin pitch is realized using pitch-quadrupling technique, where fin width is determined by the second sidewall thickness of self-aligned quadruple patterning (SAQP) [1], whereas metal patterning with 48-nm metal pitch is realized using pitch-splitting technique, where spacing between the two metal patterns is determined by the sidewall of mandrel [self-aligned double patterning (SADP)] [1], [2]. A discrete number of FinFET for a device and uniformly spaced metal limits SRAM design freedom that impacts an area gain and performance. Scaling of technology demands an area gain to achieve twice the density of SRAM to reduce cost per bit. With the increasing usage of high-density (HD) SRAM cell in system on chip (SOC), it is essential to achieve an area gain sustaining at least same yield and performance.

Increasing the area of device reduces threshold voltage (VT) variation ($\sigma \Delta V$th) of the device [3], [8], but due to scaling, effective area of device per fin reduces, which in turn increases $\sigma \Delta V$th. To compensate for variability, the number of fins for an SRAM should increase, but that will result in higher cell area and a potential disadvantage for the technology. To reduce $\sigma \Delta V$th, increasing the length of gate (Lg) is a common practice used in SRAM design, but at 64-nm gate pitch processing gate, using litho-etch litho-etch (LELE) increases the variability in device and Source/Drain (S/D) contact space. Relaxing gate pitch is the only option to bring down variations, but it comes at the cost of higher area for SRAM cell.

Owing to increasing complexity of processing and effect of patterning quantization over design freedom demands systematic approach to design SRAM in 10-nm era. In FinFET technology, pull-up ratio (PR) and cell ratio (CR) of SRAM cell are decided by the architecture of a cell. These ratios are quantized, which leads to reduced margins for read and write operations of a cell architecture. Hence maximum read static noise margin (RSNM) [4], write trip point (WTP), and read current (I_{read}) at lowest leakage current (I_{leak}) and area are achieved by optimizing VT of SRAM devices. In [9], the ratio of N-curve [23] metric I_{crit} to its variation σI_{crit} is defined as access disturb margin to design SRAM cell, which will decide device strength. An approach of VT-targeting using WF tuning for SRAM cell is considered in [10] for system on insulator (SOI) SRAM for low-voltage operation by selecting sets of VTs. A simulation-based approach to design SRAM is elaborated in [11] to show the sensitivity of VTs on SRAM parameters. In nanoscale CMOS technologies, innovative process and design solutions will help to achieve optimum performance for SRAM [12]. In this process, one important parameter to achieve the desired performance is work function (WF) tuning of metal gate [14]–[17]. Hence we present a systematic way of device VT-targeting for

Manuscript received May 30, 2014; revised September 9, 2014; accepted October 28, 2014. Date of publication December 9, 2014; date of current version May 18, 2015. The review of this paper was arranged by Editor S. S. C. Song.

The authors are with the imec, Leuven 3001, Belgium (e-mail: sushilnet@gmail.com; kenichi.miyaguchi@imec.be; praveen.raghavan@imec.be).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2014.2368358

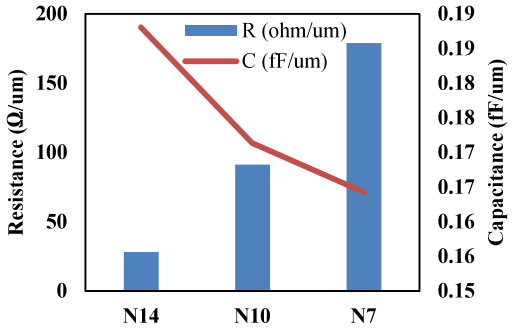


Fig. 1. Assuming that the depth of metal equals metal pitch, capacitance per unit length remains similar compared with resistance seen for minimum critical dimension. Resistance per unit length is growing almost double for scaled technology from N14 to N7 [5].

SRAM architectures to yield SRAM in variability-prone 10-nm-technology era.

In this paper, we analyze and target VT for three SRAM architectures: 1) 111 (pull-up, pass-gate, pull-down); 2) 112 (2 pull-down); and 3) 122 (2 pass-gate, 2 pull-down) on the basis of the number of fins per device. Device read current variability and the impact on minimum I_{read} due to metal pattering choice for bit-line [bit-line/bit-line-bar (BL/BLB)] signals are discussed in Section II. Section III elaborates on device VT-targeting method to achieve 6σ stability and read current. Section IV compares the characteristics of architectures based on targeted VTs and summarizes about HD SRAM cell for 10-nm node. It is essential to use Lg larger than logic device to achieve desirable characteristics for SRAM cell. LELE is the option to pattern gate at 64-nm pitch, and thus Section V shows the impact of LELE for gate and explains added margins required to retain the same yield.

II. Target Read Current

A. Minimum Read Current Estimation

Choice of gate and metal pitch results in area shrinkage of SRAM cell. In 193 immersion technology, 48-nm metal pitch can be processed using LELELE (LE3) and SADP [1], [2]. SADP provides uniform metal lines, whereas LE3 suffers from overlay issue. Relative variation seen among three patterns in LE3 is reduced by good process control, which is difficult to achieve in sub-10-nm technology [2]. Metal pattern complexity and variability can be reduced by restricting metals to have unidirectional patterns using SAQP and SADP [2].

SAQP and SADP restricts spacing between the two metal patterns to minimum critical dimension (CD) [1], which results in uniform capacitance for metal patterns in subnanometer technologies (Fig. 1). In general, metal patterns in SRAM have minimum critical dimension to achieve desired area scaling. Minimum width of metal increases effective resistance for SRAM signals, whereas worst SRAM cell performance is determined by bit-line capacitance during read operation. A minimum read current supplied by worst SRAM cell should be sufficient to discharge BL capacitance in the expected read time [6]. Number of devices, fin-pitch, and effective device dimension (including local interconnects) decide the height of the SRAM cell as well as resultant metal widths of BL/BLB signals. The width of FinFET SRAM cell is twice the gate pitch and it is smaller than the height of SRAM cell [Fig. 2(a)]; thus routing BL and BLB along the width of SRAM cell results in lower capacitance. In a column of SRAM array shown in Fig. 2(c), BL sees additional drain capacitance due to unselected devices (pass-gate) connected, which results in effective higher capacitance. Hence, total capacitance of BL signal is the summation of metal capacitance and drain capacitance of connected devices. If HD SRAM cell of the technology is expected to run at a frequency of 500 MHz, then worst SRAM cell should generate differential voltage of 10% of the supply between BL and BLB within 1 ns (assuming 50% duty cycle) [6]. Hence the minimum read current is calculated as

$$I_{\min} = \{nC_b + (n-1)C_{\text{drain}}\} \times \Delta V/\Delta t$$

where

- ΔV voltage swing on bit-line (70 mV);
- Δt word-line activation time of 1 ns, for 50% duty cycle at 500 MHz frequency;
- C_{bl} capacitance of metal BL/BLB pattern per cell;
- C_{drain} drain capacitance of unselected devices connected to BL per cell;
- n number of cells in a column connected to BL/BLB.

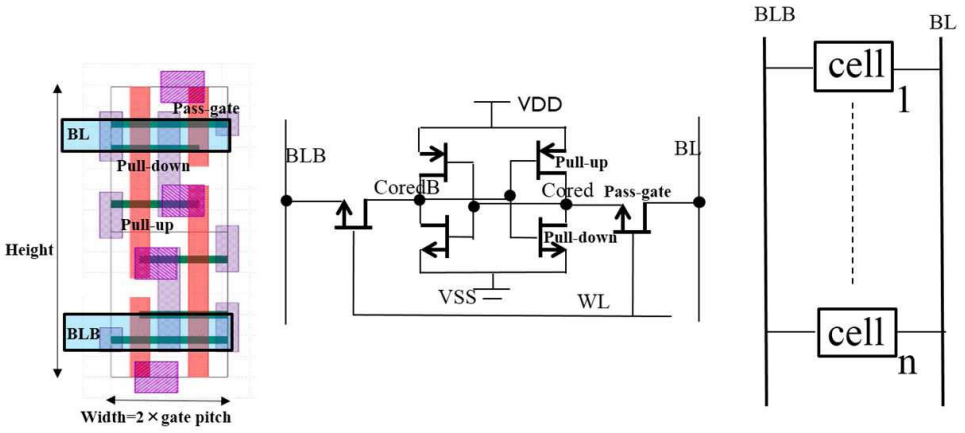
For 128 cells in a column, resultant $I_{\min}$ is tabulated in Table I.

B. Target Read Current Estimation

Scaling of technology should outperform previous technology node in performance and area. The area gain is achieved by scaling gate pitch and performance gain by scaling Lg and device dimension. For 10-nm node, 64-nm gate pitch and Lg = 20 nm achieve area scaling and relative performance gain from previous technology node. Double patterning is must to pattern gate at the pitch of 64 nm using 193i technology. If SADP is used for gate patterning, then it will force same Lg for logic and SRAM. The SADP for gate ensures minimum CD variation, as spacers are uniform across wafer [1]. Effective source- and drain-contacting area of device is the space available between two neighboring gates. As SADP uses single lithography and etching step, it has minimum overlay issue between two neighboring gates, ensures uniform source and drain contact space, and reduces pattern variability. It is a well-known practice to have larger Lg for SRAM devices in CMOS technologies [18], [20], [21]. Increasing Lg reduces effective contact space and thus increases contact resistance of device, but reduces leakage and device variability [3]. Thus double-pattern gate technologies will have two choices for SRAM design.

1) Lg same as logic device (20 nm), using SADP + cut.
2) Lg greater than logic device, using LELE + cut.

In Fig. 6, read current and leakage of 111, 112, 122 SRAM cells are analyzed; increasing Lg results in drastic decrease in read and leakage current until Lg = 24 nm. When Lg > 24 nm, effective contact resistance of device starts dominating to show decrement in read current but not



(a) (b) (c)

Fig. 2. (a) Typical layout and (b) schematic of FinFET SRAM cell. (c) Arrangement of SRAM cell in a column having n rows sharing same metal BL/BLB signal.

TABLE I

I_{MIN} OF 111, 112, AND 122 SRAM CELLS FOR 128 ROWS PER COLUMN

SRAM architecture	I_{min}(μA) for an array of 128
111	0.55
112	0.73
122	1.12

TABLE II

SIMULATED 6(σ I_{read}) AND MINIMUM READ CURRENT (I_{min}) SETS TARGET READ CURRENT FOR HD APPLICATION

SRAM	I_{min} (μA)	σI_{read} (μA) Lg=20nm	σI_{read} (μA) Lg=24nm	$Target_{Iread}$ (μA) Lg=20nm $=6\sigma I_{read}+I_{min}$	$Target_{Iread}$ (μA) Lg=24nm $=6\sigma I_{read}+I_{min}$
111	0.55	1.886	1.84	11.87	11.59
112	0.73	3.29	2.98	20.47	18.61
122	1.12	3.2	2.925	20.32	18.67

in leakage current, thus Lg = 24 nm is selected as optimum for SRAM devices under LELE assumption for gate.

In Table II, simulated σI_{read} for Lg = 20 nm and Lg = 24 nm is tabulated, Lg = 24 nm reduces effective variation in read current (σI_{read}) compared to Lg = 20 nm (nominal). SRAM can achieve maximum yield of 6σ when I_{min} of worst SRAM cell is $>6\sigma I_{read}$; thus the read current target can be defined as $6\sigma I_{read} + I_{min}$. Table II tabulates the required target read current based on simulation and I_{min} to qualify architecture as technology HD SRAM cell.

III. METHODOLOGY FOR VT-TARGETING USING WORK FUNCTION TUNING

Quantization of devices in FinFET technology restricts design freedom of sizing. The only option to design a stable SRAM cell is to tune the threshold voltage of SRAM devices by tuning metal gate WF. For selected device dimension and targeted subthreshold slope, nominal performance is modeled in device model file (Fig. 14). VT variation measurements are added in device model to define corners. Modeled VT and

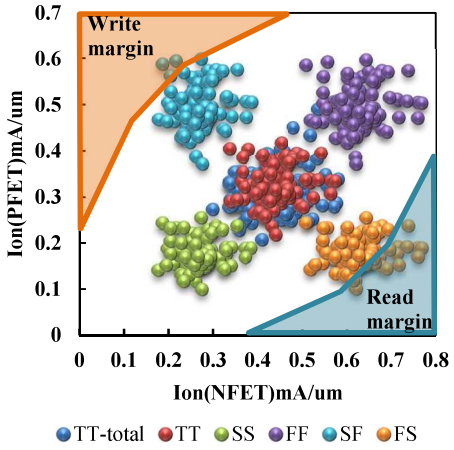


Fig. 3. Process corners definitions based on intra and interdie variation, Ion NFET versus Ion PFET of 10-nm technology node device.

β variations help in optimizing best possible threshold voltage for SRAM architecture for maximum yield and optimum performance. Fig. 3 shows saturation current of P-channel field effect transistor (PFET) and N-channel field effect transistor (NFET), where intra and inter die variations collectively defines device corners.

RSNM and WTP show margins available for read and write operations of SRAM cell [4], [7]. Maximum RSNM at poor WTP is unacceptable, as both read and write operations are equally important for SRAM. Design freedom to achieve equalized RSNM and WTP is difficult because of quantized CR and PR in FinFET technology. The supply sensitivity of read margins in Fig. 4 shows lowest RSNM for 3σ fast NFET and 3σ slower PFET, which is a FS corner of technology device. Hence the observed RSNM at FS corner 125 °C is considered for analyzing worst case RSNM of a SRAM cell. Similarly, during write operation, the lowest WTP is observed for 3σ slow NFET and 3σ fast PFET in Fig. 5 at 125 °C. Thus, WTP observed at SF corner is considered for analyzing worst case WTP of a SRAM cell.

At 125 °C, internal storage nodes Cored and CoredB [Fig. 2(b)] are under the influence of thermal noise. If SRAM

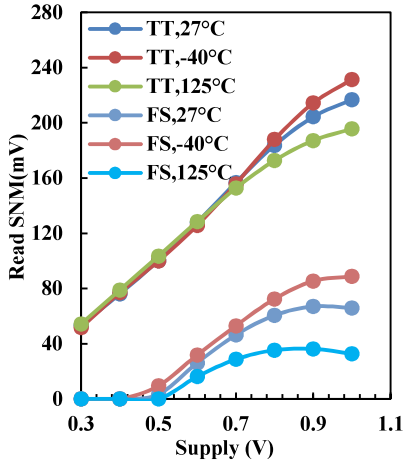


Fig. 4. Read SNM sensitivity to supply voltage of SRAM cell. FS corner shows minimum RSNM for SRAM.

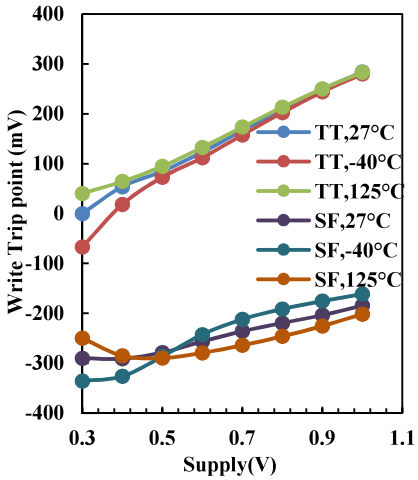


Fig. 5. WTP sensitivity to supply voltage of SRAM cell. SF corner shows minimum WTP for SRAM.

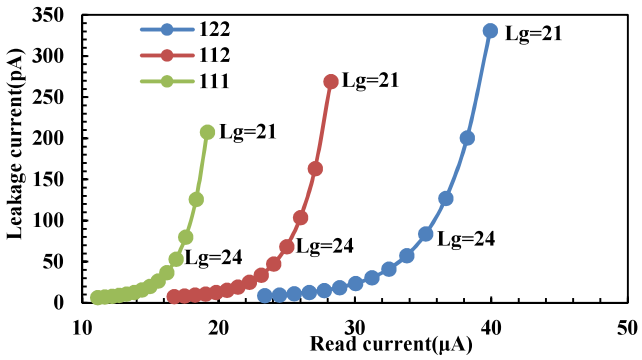


Fig. 6. Lg sensitivity to Leakage and Read current for 111, 112, 122 SRAM cell. SRAM shows sensitivity to Lg and effective S/D contact space. Lg = 24 nm shows optimum S/D contact space at acceptable degradation of performance.

cell operating at 125 °C has RSNM below thermal noise, then during read operation, due to read disturb cell can flip, which is undesirable. To ensure 6σ yield, RSNM of worst case SRAM should lie above thermal noise seen at 125 °C and that defines RSNM criteria as $\mu_{\text{rsnm}} - 6\sigma_{\text{rsnm}} \geq$ thermal noise shown in Fig. 15 [19], where μ_{rsnm} is mean RSNM and σ_{rsnm} is sigma of RSNM.

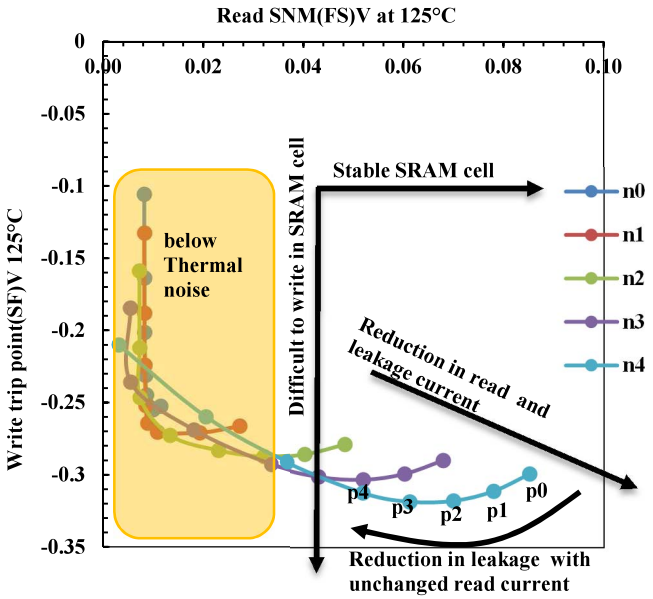


Fig. 7. From n0 to n4 VT, NFET increases and from p0 to p4 VT, PFET increases. From n0 to n4 RSNM increases, but I_{read}, I_{leak} and WTP decreases. For selected NFET, from p0 to p4, leakage and RSNM decreases, whereas WTP improves.

TABLE III
OPTIMUM RANGE OF WORK FUNCTIONS ARE CONSIDERED FOR TARGETING SRAM CELL DEVICE

NFET name	Work Function (eV)	PFET name	Work Function (eV)
n0	4.56	p0	4.63
n1	4.59	p1	4.60
n2	4.62	p2	4.57
n3	4.65	p3	4.54
n4	4.68	p4	4.51

As shown in Fig. 7, WF from n0 to n4 and p0 to p4 increases VT of NFET and PFET, respectively. Table III shows the set of WFs considered for analyzing optimum VT of PFET and NFET device. Proposed method of work function tuning brings into play the worst case RSNM and WTP to obtain optimum nominal VT for SRAM devices. At nominal supply voltage of 0.7 V, worst case RSNM at FS-125 °C is analyzed against worst case WTP at SF-125 °C.

From Fig. 7, we observed that RSNM increases as NFET VT increases from n0 to n4, whereas WTP, read current, and leakage current decreases. For the highest VT of n4 for NFET, cell has lowest leakage and WTP, but minimum read current. If n4 NFET gives target read current, then, from p0 to p4, leakage of SRAM cell reduces with reduction in RSNM but improvement in WTP. Lower VT of n0 results in very small RSNM, which lies below thermal noise; hence all points below thermal noise are discarded as they represents unstable cell. For the given architecture of FinFET SRAM cell, best VTs of PFET and NFET should result in maximum read current, minimum leakage, maximum RSNM, and maximum WTP. Selecting WF p3 for PFET gives limiting RSNM and minimum leakage for selected n3 NEFT; thus p3 PFET and n3 NFET provide first optimum VTs for given SRAM at

TABLE IV

122 SRAM CELL ACHIEVES MAXIMUM WRITE MARGIN AT LOWEST RSNM. READ AND WRITE MARGINS ARE IMPROVED FOR Lg = 24 nm, BUT OPTIMUM LEAKAGE IS ACHIEVED FOR 112 SRAM CELL WITH ACCEPTABLE READ AND WRITE MARGINS; 111 SRAM IS RETARGETED TO ACHIEVE READ CURRENT TARGET AT THE COST OF READ STABILITY TO GAIN AREA

SRAM (Lg=24nm)	μ_{wtp} (mV)	μ_{rsnm} (mV)	σ_{rsnm} (mV)	$\frac{\mu_{rsnm}}{\sigma_{rsnm}}$	$\frac{\mu_{wtp}}{\mu_{rsnm}}$	μ_{leak} (pA)	μ_{Iread} (μA)	σ_{Iread} (μA)	$\frac{\mu_{Iread}}{\mu_{Ileak}}$ ($\times 10^6$)	μ_{Iread} - target$_{Iread}$ - 127*μ_{leak} (μA)	$\frac{\mu_{Iread} - I_{min}}{6\sigma_{Iread(required)}}$
111	55.53	141.7	19.99	7.10	0.39	48.79	7.89	1.84	0.16	-3.7	1.22
112	147	133	17.76	7.48	*1.11*	*50.27*	16.5	*2.98*	*0.33*	-2.11	*2.63*
122	174.6	125.7	15.87	7.92	*1.39*	74.46	19.9	2.925	0.27	*1.22*	3.13
111(retargeted)	156	114.9	22	5.22	*1.36*	49.38	11.25	1.84	0.23	-0.34	1.78

TABLE V

ONLY 122 ACHIEVES TARGET READ CURRENT FOR Lg = 20 nm; 122 SRAM CELL LAGS BY 2.92 μA FROM TARGET READ CURRENT WITH OPTIMUM WRITE MARGIN 0.147 V SATISFYING RSNM CRITERIA

SRAM (Lg=20nm)	μ_{wtp} (mV)	μ_{rsnm} (mV)	σ_{rsnm} (mV)	$\frac{\mu_{rsnm}}{\sigma_{rsnm}}$	$\frac{\mu_{wtp}}{\mu_{rsnm}}$	μ_{leak} (pA)	μ_{Iread} (μA)	σ_{Iread} (μA)	$\frac{\mu_{Iread}}{\mu_{Ileak}}$ ($\times 10^6$)	μ_{Iread} - target$_{Iread}$ - 127* μ_{leak} (μA)	$\frac{\mu_{Iread} - I_{min}}{6\sigma_{Iread(required)}}$
111	54.11	145.8	20.85	6.99	0.37	55.83	5.87	1.886	0.11	-6.00	0.89
112	102.4	142.8	18.54	7.7	0.72	58.9	14.9	3.29	*0.25*	-5.57	2.36
122	147.0	132.3	16.8	7.9	*1.11*	86.54	17.4	3.2	0.20	-2.93	2.71

10-nm node. To achieve $6\sigma_{\text{rsnm}}$ yield, combination p3n3 is analyzed for RSNM criteria using Monte Carlo simulation. If $(\mu_{\text{rsnm}} - 6\sigma_{\text{rsnm}})$ is greater than thermal noise, then p3n3 is optimum VT combination for SRAM devices; otherwise reduce PFET VT or increase NFET VT to buy more read margin and redo Monte-Carlo until the criterion shown in Fig. 15 is satisfied [19].

Owing to the lack of a back-gate bias option in fully depleted FinFET devices, complicated WF engineering is required to achieve multiple V_{th} solution for undopped FinFET [14], [16], [17]. Ion implantation into metal gate-stack of TiN/HfO2 provides suitable VT for devices keeping channel undopped. Another approach to achieve desired VT for device is by doping channel [15], which will definitely change A_{vt} and hence corners of the device. Repeating SRAM device VT-targeting methodology with updated corners will result in desired WFs for gate-stack of SRAM devices.

IV. DETERMINATION OF TECHNOLOGY HD SRAM CELL

For Lg = 20 nm, all variants in Table V show poor read current performance. The 122 SRAM cell lags by 2.93 μA with nominal read current of 17.4 μA; ratio of $\mu_{\text{wtp}}/\mu_{\text{rsnm}} > 1$ shows better write margin for read stable cell. For SADP gate pattern, if σ_{Iread} is reduced from 3.2 to 2.71 μA, then at higher leakage of 86.53 pA, 122 SRAM can be a HD cell at 10-nm node. The 111 SRAM cell that occupies smallest area of 0.046 μm^2 gives poor read current of 7.89 and 5.87 μA for Lg = 24 and 20 nm, respectively (Tables IV and V), whereas 122 SRAM (area of 0.0546 μm^2) cell is the only cell exceeding target read current for nominal read current of 19.9 μA for Lg = 24 nm (Table IV). It also shows maximum $\mu_{\text{wtp}}/\mu_{\text{rsnm}}$ of 1.39 signifying maximum write margin for read stable SRAM cell, but at the cost of higher leakage (74.46 pA) (Fig. 9). Maximum $\mu_{\text{Iread}}/\mu_{\text{leak}}$ of 0.33×10^6 for

TABLE VI

DEVICE VT TARGETED FOR SRAM CELLS AT LG = 24NM

SRAM (Lg=24nm)	NFET (μ) (mV)	NFET (σ) (mV)	PFET (μ) (mV)	PFET (σ) (mV)
111	538	35.89	-458	37.58
112	477.2	25.38	-516	37.58
122	507.5	25.38	-458	37.58
111(retargeted)	477.2	35.89	-458	37.58

Lg = 24 nm 112 SRAM cell (0.0546 μm^2) has minimum leakage of 50.27 pA at 16.5 μA read current, which is lagging by 2.11 μA (Table IV) from target read current. If σ_{Iread} of 112 SRAM cell is reduced from 2.98 μA to 2.63 μA, then 112 SRAM is the best possible SRAM with maximum yield performance at optimum leakage. Thus 112 with Lg = 24 nm can be a technology HD SRAM cell for 10 nm.

In published FinFET [20], [21] and FDSOI [18] technologies, SRAM devices are targeting 30 pA/um OFF current (I_{off}) at nominal supply voltage. It is explicitly mentioned that logic high threshold voltage (HVT) transistor of FinFET technology have same device as SRAM device [20], [21], similar to SOI technology [18], where HVT transistors are used for designing SRAM cell with larger Lg. For reported VTs for SRAM devices in Table VI at Lg = 24 nm, If 112 is selected as technology HD SRAM cell, then NFET VT mask can be shared with logic HVT device to reduce the cost of the technology, but PFET VT mask cannot be shared because an additional HVT PFET mask for WF tuning is required for logic device as shown in Fig. 13. In the case of 122 SRAM cell, both PFET and NFET mask can be shared with HVT logic device VT mask to reduce the cost of the technology; thereby selecting 122 as technology SRAM cell will reduce cost of the technology at highest performance and desired yield. In [18], [20], and [21], SRAM devices are targeted at 30 pA/um I_{off},

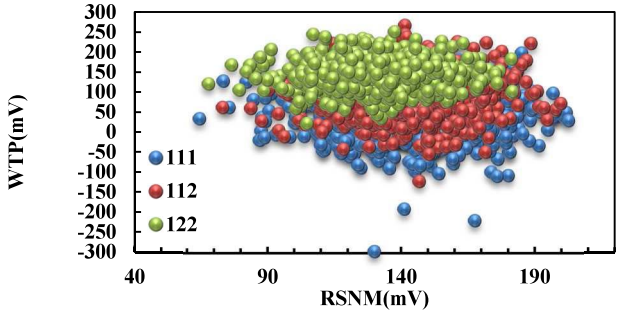


Fig. 8. From 111 to 122, WTP improves at reduced RSNM for Lg = 20 nm (SADP gate).

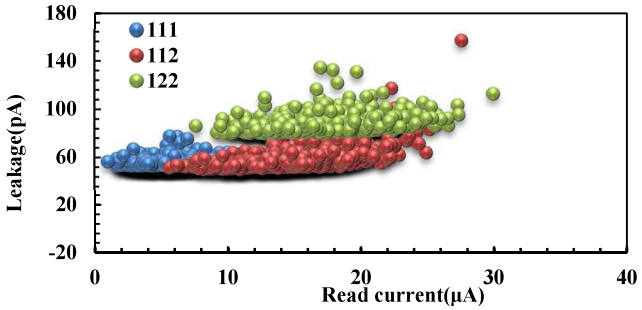


Fig. 9. Leakage is increasing at lower read current for Lg = 20 nm (SADP gate).

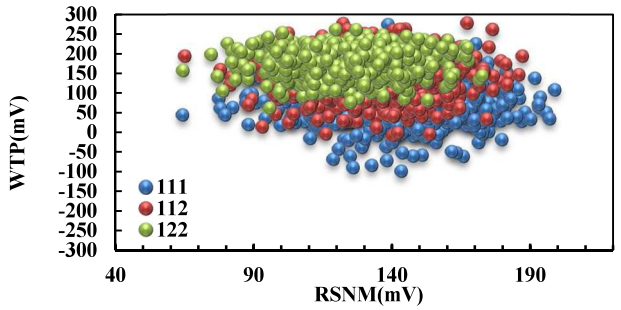


Fig. 10. WTP spread is reduced by 10%, whereas required RSNM is reduced from Lg = 20 nm to Lg = 24 nm (LELE gate).

which is in close proximity of I_{off} achieved by 122 SRAM cell devices as seen in Fig. 13. If 30 pA/um target I_{off} is applied to 112 SRAM cell devices, then it will mitigate read current performance due to increased VT of NEFT device.

At 10-nm node, to gain in area by 19%, 111 SRAM cell can be retargeted to achieve read current target instead of RSNM target. The retargeted new VTs (Tables IV and VI) show that by compromising RSNM at increased leakage of 49.38 pA, cell can give nominal read current of 11.25 μA. The effective read margin available is reduced to $5.22\sigma_{\text{rsnm}}$ from $7.1\sigma_{\text{rsnm}}$ (includes thermal noise) thus for 111 SRAM cell target read current performance can be achieved by compromising read stability of cell.

We observed in Figs. 8 and 10 that effective spread and nominal RSNM are decreased by 6% and 5%, respectively. The WTP spread is reduced by 10% and WTP nominal is increased by 19% for Lg = 24 nm; it shows that SRAM with

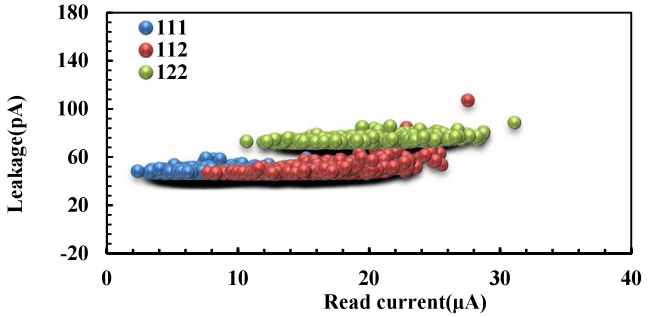


Fig. 11. Leakage spread is reduced by 50% from Lg = 20 nm to Lg = 24 nm (LELE gate) for increased read current.

Lg = 24 nm achieves desired yield at lower RSNM along with higher write margins. Similarly, in Figs. 9 and 11, we see that the nominal leakage is decreased by 14% and spread decreased by two times for Lg = 24 nm as well as read current spread decreased by 9%, whereas nominal read current is increased by 14% for Lg = 24 nm. Thus having Lg greater than nominal helps in achieving better characteristics for any SRAM cell.

V. LELE Double-Pattern Gate Impact on SRAM Cell Performance

At 10-nm technology, to enable SRAM design with Lg = 24 nm and logic design with Lg = 20 nm at 64 nm gate pitch requires LELE for gate patterning. At logic nominal Lg = 20 nm, gate spacer thickness of 8 nm gives 28-nm wide active local interconnect space for source and drain contact. In the case of Lg = 24 nm, the contact space reduces to 24 nm, which increases effective contact resistance. The LELE processing option suffers from CD variation and overlay that can lead to asymmetric contacting space for a device [Fig. 12(b) and (c)]. Introduction of LELE for gate adds more variability to technology, which will affect logic and SRAM design. The SADP process for gate gives uniform CD with minimum CD variation than LELE, but constrains same Lg for SRAM and logic design. A standard layout of a FinFET SRAM shown in Fig. 12(a) suffers from variability due to LELE patterning for gate. An overlay of 3 nm and 3σ CD variation of 3 nm can lead to asymmetric SRAM cell. As shown in Fig. 12(a)–(c), two consecutive gate patterns form a SRAM cell in FinFET technology; thus under CD variation, Gate pattern A can have CD of 27 nm ($\mu + 3\sigma$) and second Gate pattern B can have CD of 21 nm ($\mu - 3\sigma$). The PU1, PD1, and PG0 devices in SRAM share same Gate pattern A, whereas PU0, PD0, and PG1 devices share Gate pattern B. If Pattern B suffers from overlay issue with respective Pattern A, then two asymmetric layouts are possible, which represents extreme cases.

Case I: Pattern B displaced closer to Pattern A by 3 nm [Fig. 12(b)].

Case II: Pattern B displaced away from Pattern A by 3 nm [Fig. 12(c)].

Displacement of gate patterns results in asymmetric contact spacing for all signals of SRAM cell. In Fig. 12(b),

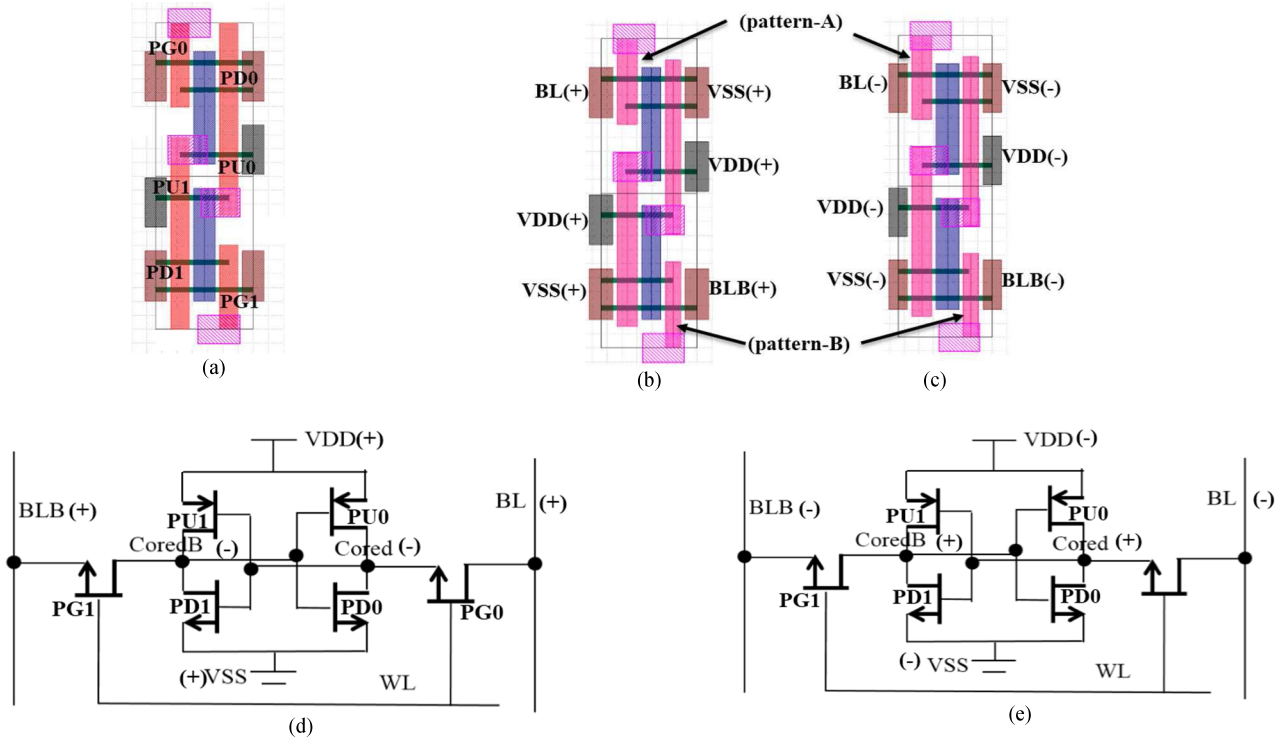


Fig. 12. Lg = 24 nm nominal layout of FinFET SRAM cell. (b) Gate pattern A Lg = 27 nm ($\mu + 3\sigma$) and Gate pattern B Lg =21 nm ($\mu - 3\sigma$), Pattern B displaced by 3 nm (overlay) toward Pattern A. (c) Gate pattern A Lg = 27 nm ($\mu + 3\sigma$) and Gate pattern B Lg = 21 nm ($\mu - 3\sigma$), Pattern B displaced by 3 nm (overlay) away from Pattern A. Schematic (d) and (e) captures layout effects in (b) and (c) respectively.

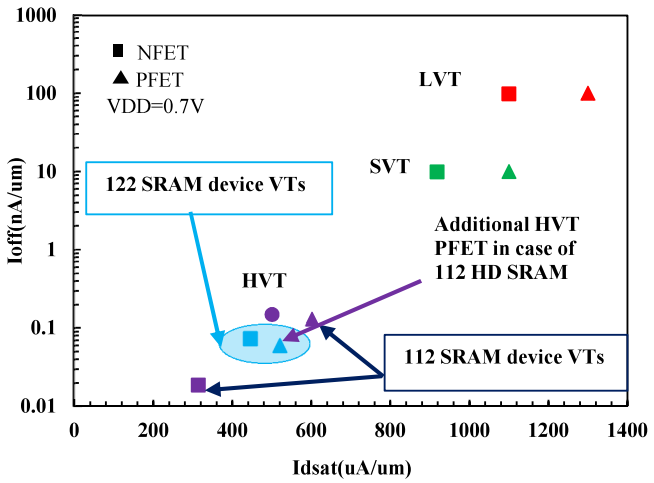


Fig. 13. 112 NFET VT mask can be shared with HVT device of technology where Lg of SRAM device is 24 nm, but defining HVT PFET of technology needs additional mask. In the case of 122 SRAM, both PFET and NFET VT mask can be shared with mask of HVT device of technology.

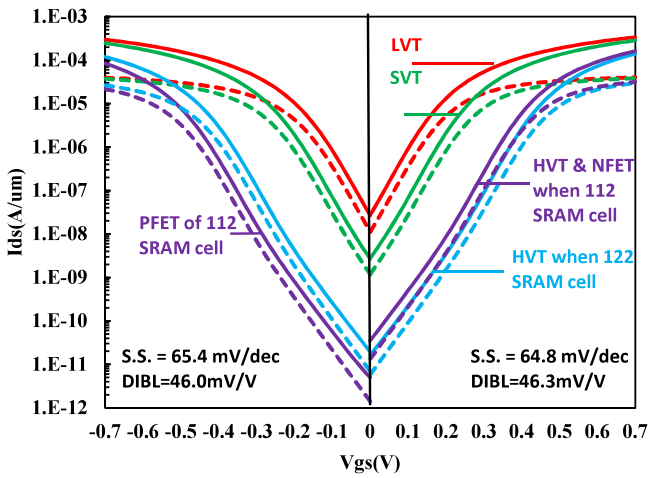


Fig. 14. NFET and PFET Id–Vg characteristics of multiVT devices with Lg = 20 nm for LVT, SVT, and SRAM VT.

resultant layout reduces effective contact spacing for Cored and CoredB signals, but increases contact spacing for BL, BLB, VDD, and VSS signals. Similarly in Fig. 12(c), contact spacing is reduced for BL, BLB, VDD, and VSS signals, but increased for Cored and CoredB signals. Fig. 12(d) and (e) captures layout effect shown in Fig. 12(b) and (c), respectively.

Case II presents worst case for RSNM, where unmatched devices for pass-gate and pull-down along with increased contact resistance of BL and VSS degrades read margin by 19 mV equivalent to $1\sigma_{\text{rsnm}}$ (Fig. 15), whereas in Case I, unmatched devices for pass-gate and pull-up along with increased gate contact Cored and CoredB signal resistance reduces write margin by 4 mV (Fig. 16). Nominal leakage of cell increased by 14%, whereas spread increased by 3 times (Fig. 17), but impact on read current is very minimum (170 pA) (Fig. 18). The leakage is increased by Lg variation, whereas read current sees minimum penalty due to differential current between two sets of device stack having strong pass-gate, weak pull-down, and vice-versa. For 128 cells in a column, read time in Fig. 19 shows increment by negligible 5 ps (50% of $\sigma_{\text{read-time}}$) due to increase in drain

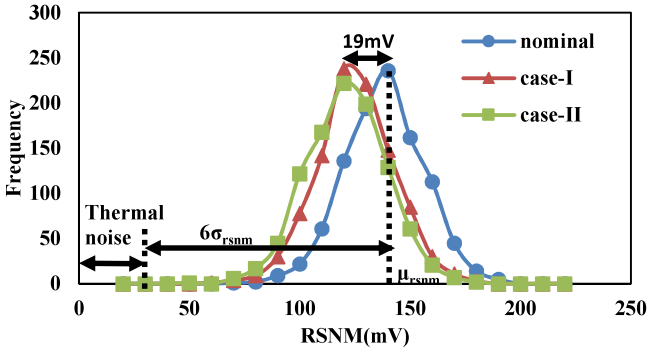


Fig. 15. RSNM criteria considered for targeting SRAM cell VT is μ_{rsnm}-$6\sigma r_{snm} \geq$ thermal noise [19]. Due to LELE variation, RSNM is degraded by 19 mv that is equivalent to 1σ.

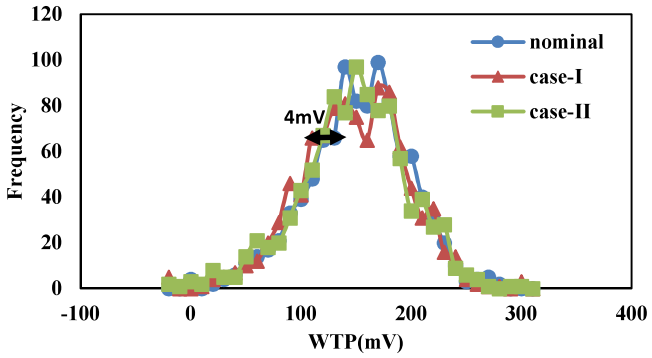


Fig. 16. WTP is decreased by 4 mV due to LELE for gate patterning.

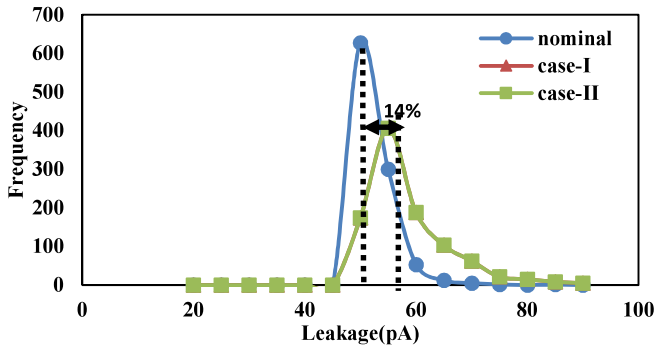


Fig. 17. Nominal leakage is increased by 14% due to gate patterning variability.

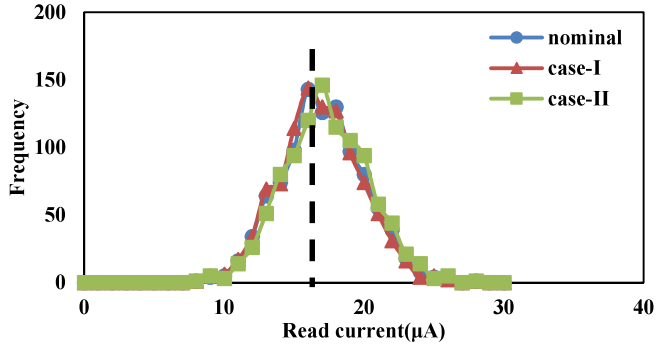


Fig. 18. Read current is unaffected as it is the difference between two stacks of NFET, where weak pass gate compensate strong pull down and vice versa.

contact capacitance of BL/BLB signal. The most impacted parameters due to LELE gate pattern are leakage and RSNM. The reduction in RSNM can be considered while targeting VT

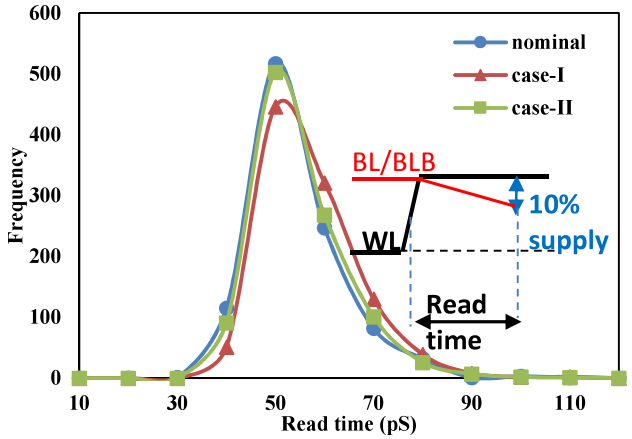


Fig. 19. Increased drain contact area in Case I increases the capacitance of BL, which shows an increase in read-time for 128 cell in a column.

of SRAM device, but that will reduce read current and hence the performance.

VI. Conclusion

Observed variations of nominal device defines corners for technology logic devices that can be used to target SRAM device VTs by assessing worst case RSNM and WTP for architectures of FinFET SRAM to target technology HD SRAM cell architecture and to obtain optimum VTs at a minimum number of simulations. RSNM, Iread, and cost can select 122 SRAM with Lg = 24 nm, but at the cost of higher leakage, whereas 112 SRAM architecture lags target read current by 2.11 μA, but have minimum leakage for same area of 0.0546 μm^2 at 10-nm node. To increase area gain, 111 SRAM can be retargeted, but at lower yield of 4.26σ. Selecting 122 as technology HD SRAM cell will reduce cost of technology by sharing VT mask of both PFET and NFET with technology HVT devices, whereas in the case of 112 SRAM cell, only NFET VT mask can be shared. In double pattern gate technology due to overlay and CD variation, stability of SRAM cell is reduced by $1\sigma_{rsnm}$, which will result in poor yield for targeted devices. To ensure maximum yield, loss of $1\sigma_{rsnm}$ should be considered while targeting VTs of SRAM devices at the cost of additional performance penalty.

Acknowledgment

The authors would like to thank for the helpful discussions with Qualcomm, and Global Foundries. They would also like to thank for the support from J. Ryckaert, M. G. Bardon, M. Dehan, D. Verkest, N. Horiguchi, and A. Thean of Process & Technology Group at imec.

References

[1] P. Xu *et al.*, "Sidewall spacer quadruple patterning for 15 nm half-pitch," *Proc. SPIE*, vol. 7973, p. 79731Q, Feb. 2011.

[2] B. Vandewalle, B. Chava, S. Sakhare, J. Ryckaert, and M. Dusa, "Design technology co-optimization for a robust 10 nm metal1 solution for logic design and SRAM," *Proc. SPIE*, vol. 9053, pp. 90530Q-1–90530Q-13, Mar. 2014.

[3] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[4] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. 22, no. 5, pp. 748–754, Oct. 1987.

[5] *International Technology Roadmap for Semiconductors*. [Online]. Available: http://www.itrs.net/Links/2012ITRS/Home2012.htm, accessed May 20, 2013.

[6] A. Karandikar and K. K. Parhi, "Low power SRAM design using hierarchical divided bit-line approach," in *Proc. Int. Conf., Comput. Design, VLSI Comput. Process. (ICCD)*, Oct. 1998, pp. 82–88.

[7] A. J. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 658–665, Apr. 2001.

[8] T. B. Hook, J. B. Johnson, and J. Shah, "Anomalous dependence of threshold voltage mismatch of short-channel transistors," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2805–2807, Jan. 2011.

[9] C. Wann *et al.*, "SRAM cell design for stability methodology," in *Proc. IEEE VLSI-TSA Int. Symp. VLSI Technol. (VLSI-TSA-Tech)*, Apr. 2005, pp. 21–22.

[10] V. P.-H. Hu, M.-L. Fan, S. Pin, and C.-T. Chuang, "Threshold voltage design of UTB SOI SRAM with improved stability/variability for ultralow voltage near subthreshold operation," *IEEE Trans. Nanotechnol.*, vol. 12, no. 4, pp. 524–531, Jul. 2013.

[11] A. Asenov *et al.*, "Simulation based transistor-SRAM co-design in the presence of statistical variability and reliability," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2013, pp. 33.1.1–33.1.4.

[12] K. Zhang, E. Karl, and Y. Wang, "SRAM design in nano-scale CMOS technologies (Invited)," in *Proc. Symp. VLSI Technol. (VLSIT)*, Jun. 2012, pp. 85–86.

[13] T. Matsukawa *et al.*, "Comprehensive analysis of variability sources of FinFET characteristics," in *Proc. Symp. VLSI Technol.*, Jun. 2009, pp. 118–119.

[14] A. Veloso *et al.*, "Flexible and robust capping-metal gate integration technology enabling multiple-VT CMOS in MuGFETs," in *Proc. Symp. VLSI Technol.*, Jun. 2008, pp. 14–15.

[15] C.-H. Lin *et al.*, "Channel doping impact on FinFETs for 22 nm and beyond," in *Proc. Symp. VLSI Technol.*, Jun. 2012, pp. 15–16.

[16] R. Singanamalla *et al.*, "Multi-VT engineering in highly scaled CMOS bulk and FinFET devices through Ion implantation into the metal gate stack featuring a 1.0 nm EOT high-*k* oxide," in *Proc. Int. Symp. VLSI Technol. Syst. Appl. (VLSI-TSA)*, Apr. 2010, pp. 112–113.

[17] L. Witters *et al.*, "Treshold voltage modulation in FinFET devices through arsenic ion implantation into TiN/HfSiON gate stack," in *Proc. IEEE Int. SOI Conf.*, Oct. 2007, pp. 31–32.

[18] C. Fenouillet-Beranger *et al.*, "Hybrid FDSOI/bulk high-*k*/metal gate platform for low power (LP) multimedia technology," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2009, pp. 1–4.

[19] P. A. Stolk *et al.*, "CMOS device optimization for mixed-signal technologies," in *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, Dec. 2001, pp. 10.2.1–10.2.4.

[20] C.-H. Jan *et al.*, "A 22 nm SoC platform technology featuring 3-D tri-gate and high-*k*/metal gate, optimized for ultra low power, high performance and high density SoC applications," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2012, pp. 3.1.1–3.1.4.

[21] W. Shien-Yang *et al.*, "A 16 nm FinFET CMOS technology for mobile SoC and computing applications," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2013, pp. 9.1.1–9.1.4.

[22] J. Petry *et al.*, "A VFB tunable single metal single dielectric approach using as I/I into TiN/HfO_2 for 32 nm node and beyond," in *Proc. Int. Symp. VLSI Technol., Syst., Appl., (VLSI-TSA)*, Apr. 2009, pp. 57–58.

[23] E. Grossar, M. Stucchi, K. Maex, and W. Dehaene, "Read stability and write-ability analysis of SRAM cells for nanometer technologies," *IEEE J. Solid-State Circuits*, vol. 41, no. 11, pp. 2577–2588, Nov. 2006.

Sushil Sudam Sakhare (M'14) received the bachelor's degree in electronics engineering from the Sardar Patel College of Engineering, Mumbai, India, and the M.Tech. degree in electrical engineering from IIT Bombay, Mumbai, in 2007.

He joined imec, Leuven, Belgium, as a Senior Researcher in 2013, where he is involved in advanced CMOS technology SRAM/DRAM and emerging memories.

Kenichi Miyaguchi (M'99) received the B.E. and M.E. degrees in electrical and electronic engineering from Kyoto University, Kyoto, Japan, in 1997 and 1999, respectively.

He was with Texas Instruments Incorporated, Dallas, TX, USA, from 2007 to 2012. Since 2012, he has been involved in research on advanced CMOS technology with imec, Leuven, Belgium.

Praveen Raghavan (M'08) received the bachelor's degree in electrical engineering from Regional Engineering College, Tiruchirappalli, India, the master's degree in electrical engineering from Arizona State University, Tempe, AZ, USA, and the Ph.D. degree from Katholieke Universiteit Leuven, Leuven, Belgium, in 2009.

He is currently a Principal Scientist with the Design Technology Enablement Group, imec, Leuven.

Abdelkarim Mercha received the M.Sc. degree in electrical engineering from ENSI de Caen, Caen, France, the master's degree in business management from IAE Caen, Caen, and the Ph.D. degree in microelectronics from the University of Caen, Caen, in 1997 and 2000, respectively.

He joined imec, Leuven, Belgium, in 2001, where he is currently the Program Manager of the Integrated Solutions for Technology Exploration Program.