Analysis of different FinFET design techniques used for Low-Power Robust SRAMs

Prashanth Rajan
North Carolina State University
Raleigh, NC-27606, USA
prajan@ncsu.edu

***Abstract*—Memories take up nearly 80% of the die area in high performance processors and as a result there is a need for high performance, low-leakage and highly robust SRAMs. In this paper, different design techniques to optimize FinFET devices for robust and low-power SRAMs have been analyzed. Firstly, the impact of surface orientation on stability, performance and power of 6-T and 8-T FinFET SRAMs has been analyzed. Secondly, gate sidewall offset spacer thickness has been optimized to reduce leakage power and improve stability under process variations. Finally, we look at how introducing an asymmetric drain spacer extension can improve performance of a SRAM compared to conventional SRAM.**

Keywords—FinFET, SRAM, surface orientation, spacer

I. INTRODUCTION

THE PAST few decades has seen the emergence of the complimentary metal-oxide semiconductor (CMOS) technology coupled with transistor sizing. However, aggressive device scaling has led to variations in device parameters and increased short-channel effects (SCEs) [1] [2]. Thinner gate-oxides have led to improve SCEs but also lead to increased gate leakage. As a result, different transistor structures have been analyzed to replace MOSFET devices and among them FinFETs have emerged as a favorite for various reasons. Among general double gate structures, FinFETs are drawing significant interest in the semiconductor industry due to ease of fabrication compared to a planar double gate structure. FinFETs also provide reduced short channel effects, lower leakage current, higher "ON" current and better scalability.

Along with the scaling advantages of FinFETs, some major challenges that need to be overcome when designing circuits with FinFET devices is the issue of width quantization. The width of a FinFET device can increase only in quanta of silicon fin height and the effect of width quantization becomes more critical in circuits where proper sizing is required to ensure correct functionality. This problem is further compounded by the effect of process variations. Hence due to all these reasons it is important to explore different design considerations for FinFET based SRAMs.

In this paper, four different methodologies have been explored for the design of robust low-power FinFET SRAMs. Firstly, we look at how optimizing the orientation of crystal surface results in better performance for FinFET SRAMs[3]. The effect of fin-orientation on cell stability (read/write/hold), performance and leakage for 6-T and 8-T SRAM cell configurations. It has been observed that orientation optimization can improve read static noise margin (SNM), Write margin and access time of a 6-T cell. Similarly, for a 8-T cell, multi-orientation improves the write stability with no impact on read stability and cell performance.

Secondly, we look at a device-optimization technique for FinFETs to reduce leakage and improve stability in an SRAM cell. The gate sidewall spacer thickness of FinFET devices has been optimized [4] to reduce leakage and improve robustness of the SRAM cell. A device simulator named Taurus [5] was used for the modeling and analysis based on which a methodology for optimizing the gate sidewall spacer thickness to improve the drain capacitance to on-current ratio in a FinFET has been developed. The proposed optimization not only reduced SCE, but also subthreshold current, and gate-edge direct-tunneling leakage, cell leakage . The optimization also improved static-nose margin (SNM), and reduced cell access time considering the word-line delay. Moreover, the analysis showed that the proposed optimization reduced the sensitivity of the device threshold voltage (V_t) to L_g (printed gate length) and T_{si} (silicon thickness) variations which helps to improve the cell robustness (read, write, hold, and access failure reduces).

Finally, we look at how introducing an additional offset spacer [6] only on the drain side to introduce drain-side underlap in FinFET devices results in defined drain and source terminals. This device is called asymmetric drain spacer extension (ADSE) FinFET. Bidirectional currents in these transistors are not the same, and there is asymmetry in the magnitudes of currents for positive and negative drain-to-source voltages (V_{DS}). This asymmetry in current is used to achieve improvement in both read and write stability of 6T SRAM bit cells. ADSE FinFETs exhibit lower short channel effects, lower drain-induced barrier lowering (DIBL), and reduced subthreshold leakage.

Other papers such as Moradi et al[7] propose asymmetry in the device by unequally doping the drain and source terminals of FinFETs [asymmetrically doped (AD) FinFETs] have shown the device to exhibit improved read and write margins in scaled technologies, helping extenuate read-write conflict. In addition, AD FinFETs have been found to exhibit improved

short-channel characteristics (lower DIBL, subthreshold current), which results in significant reduction in leakage of AD-FinFET SRAM cell. AD-FinFET SRAMs also show improvement in hold stability and write time at the cost of increase in access time with no area penalty.

The rest of the paper looks at each of these methodologies in detail, looking at the process of how it was achieved and their effects on different performance metrics.

II. OPTIMIZING SURFACE ORIENTATION IN FINFETS [3]

A SRAM designed with 32nm FinFET devices (Tsi(silicon thickness)=7nm, EOT(equivalent oxide thickness)=1.73nm with oxynitride, H_{FIN}=36nm,V_{DD}=0.8) was modeled using mixed-mode device simulations [8].

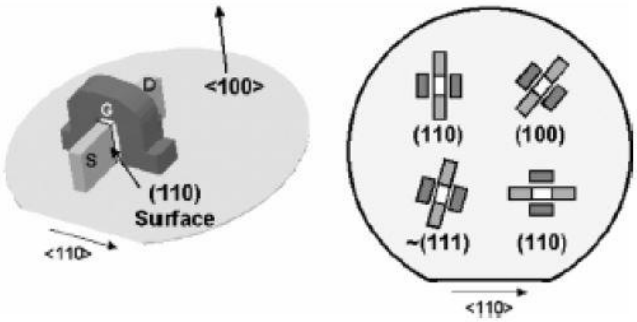

Fig 1 : Modification of surface orientation in FinFET

Maximum mobility for NMOS was observed along (100) surface whereas that for PMOS along (110) surface (Fig. 2). This dependence of mobility on surface orientation is due to the change in surface density of atoms and effective hole and electron masses at different orientation (Table 1). These properties affected the various scattering phenomena such as columbic, phonon and surface roughness scattering and thus resulted in the observed trend in mobility (Fig 2(a)). However, due to velocity saturation, the change in current is lower than the change in mobility (Fig. 2(b)). [3]

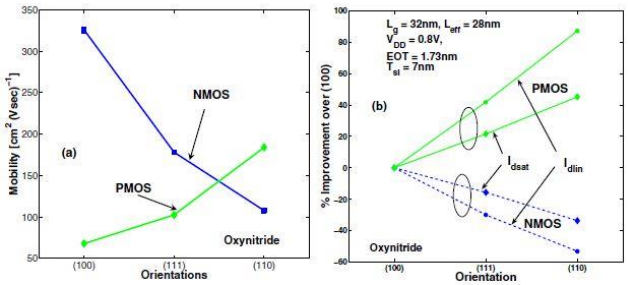

Fig 2: Effect on surface orientation on (a) Mobility (b) Current

Due to the differences in the interface trap density, oxide charge and quantum mechanical effects, there is a shift in device threshold voltage (V_t). Threshold voltage is the voltage at which the conduction band goes below the Fermi level. However, when quantum mechanical effects are considered, due to energy quantization, the lowest level that electrons can occupy is not the bottom of the conduction band; rather it is the first energy level which is slightly higher. Quantum mechanical effects tend to marginally increase the device V_t (~26mV). Among different surface orientations, V_t change depends on both the sub-band energy levels and the density of state effective masses. Higher the effective density of states for the lowest sub-band, lower the threshold voltage [3].

Surface orientation has also been shown to have negligible impact on gate leakage [9]. Thus an optimal orientation giving better stability and performance of the SRAM cell comes at no added cost of the leakage current [3].

In FinFET SRAM, sizing opportunities are limited as device widths are quantized (in quanta of fin height). Hence, modification of device mobility, which could be obtained by rotation of the fins, was used to optimize β-ratio in FinFET SRAM. β-ratio is defined as the (W/L) of nMOS pull down transistors of inverter to the (W/L) of nMOS pass transistors. A 6-transistor SRAM cell designed with all single fin devices (Fig 3) was considered for the analysis [3].

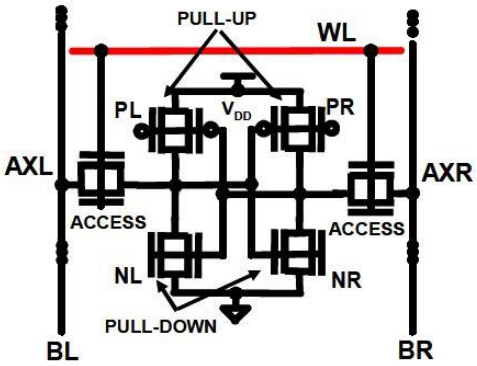

Fig 3: Conventional FinFET based 6-T SRAM

A. *Effect of fin orientation on stability and performance*

Read Stability: Read Static Noise Margin (SNM) is used as a measure to determine the read stability. It is defined as the difference between V_{TRIPRD} and V_{READ}. V_{READ} is the voltage to which a node storing '0' rises during a read operation. Trippoint associated with the inverter storing a '1' is V_{TRIPRD}. Since read stability depends on (V_{TRIPRD}- V_{READ}), high V_{TRIPRD} and low V_{READ} improves stability. A higher mobility of pull-down (PD) NMOS and a lower mobility of access (AX) NMOS help reduce V_{READ}. While reading a '0', PD NMOS is in linear and AX is in saturation region. Due to velocity saturation, the linear current depends more strongly on orientation compared to the saturation current (Fig 2b). Hence (100) PD & (100) AX give lower V_{READ} compared to (110) PD & (110) AX (Fig 4a) [3].

Lowest V_{READ} was achieved with (110) AX and (100) PD devices since NMOS mobility is lowest at (110) and highest at (100). To increase V_{TRIPRD}, higher mobility of PMOS pull-up (PUP) compared to PD NMOS is required (Fig. 4b). This is achieved at (110) surface for both PUP and PD NMOS, since on (110), PMOS mobility is highest and NMOS mobility is lowest. Read stability depends on the difference of the two voltages, and was observed to be maximum at (110) PUP, (100) PD and (110) AX (43% larger compared to all (110) devices, Fig. 5) [3].

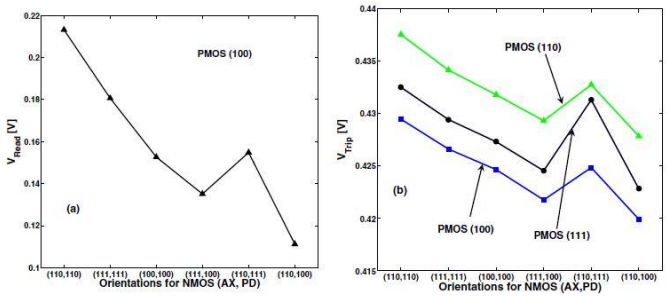

Fig. 4: Effect of surface orientation on (a) Read-Voltage (b) Trip-point.

Write Stability: Write Stability of the cell is measured using write margin which is the maximum voltage on a bit-line that allows writing to the cell while the other bit-line is at V_{DD}. Higher the write margin, greater is the stability. Use of a weaker PUP and a stronger AX helps the node storing '1' to discharge faster, thus facilitating a quicker write of '0'. Hence, the write margin improves with a strong NMOS AX (100) and a weak PUP (100) (Fig 6). Write margin can also be improved if the trip-point of the inverter storing a '0' is increased. Strong PUP (110) and weak PD NMOS (110) improve the trip-point of the inverter. Due to these effects it was observed that, the cells with all (100) devices have maximum write margin (~16% higher than all (110) devices, Fig. 6) [3].

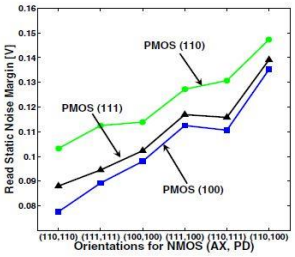

Fig. 5: Read SNM

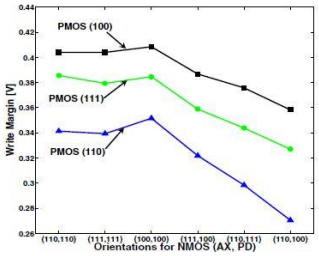

Fig. 6: Write Margin

Hold Stability: Hold stability of the cell is measured using the Hold Static Noise Margin (SNM). It is a measure of how strongly the node storing '1' and the node storing '0' are coupled to V_{DD} and V_{SS} respectively. Stronger coupling and thus higher stability can be achieved by having strong PUP (110) and strong PD NMOS (100). Different orientations were observed to have a weak impact on hold SNM (Fig. 7) [3].

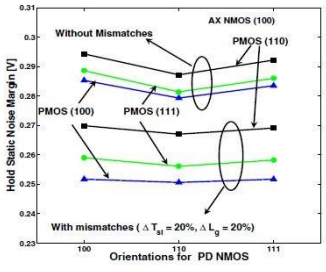

Fig 7: Hold static noise margin

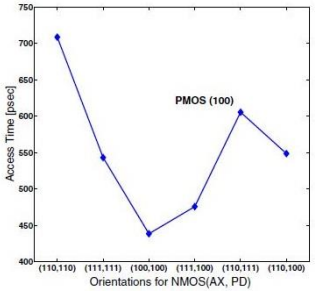

Fig 8 :Read SNM

Access Time: Read current is used as a measure to determine the Access time. Higher read current suggests faster access during a read operation. Strong AX NMOS and PD NMOS allow higher read current, thus taking a lower access time and giving better performance. (Fig. 8). Access time was observed to be the lowest for (100) AX & (100) PD (~38% higher than all (110) devices) (Fig 8) [3].

B. Effect of process variations

The cell stability under worst case process variation conditions was analyzed. Simulations considering worst-case device mismatches (±20% change in T_{si} and L_{gate}) showed that the multi-oriented cells have lower degradation in Read SNM, Write Margin and Hold SNM due to process variation (Fig. 7, 9). Fig 9 compares the conventional design with all (110) devices with optimized multi-oriented cell (PUP (100), AX NMOS (110), PD NMOS (100)). Optimization of fin orientation was found to improve the cell robustness under the process variation [3].

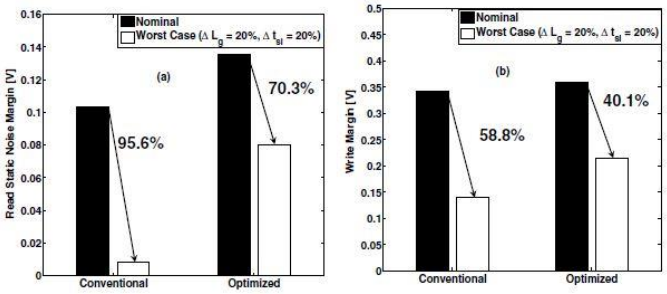

Fig. 9: (a) Read SNM (b) Write Margin under process variations

C. Multi-oriented 8-T FinFET SRAM cell

If the read and write operations are decoupled by having separate read and write lines, as in the case of an 8-transistor SRAM cell (Fig. 10), cell stability can be further improved. While reading a "0" in this case, since the Word line (WL) is turned off, the AX NMOS is in cut-off. Thus V_{READ} is 0V and hence an improved Read SNM. However, 8-T SRAM configuration does not give any improvement in write margin over that obtained from a 6-T SRAM cell. The impact of fin-orientation for 8-T FinFET SRAM was also analyzed. It was observed that the default orientation, i.e. (110) PUP, (110) PD, (110) AX and (110) Read NMOS, gave the maximum Read SNM for 8-T cell. This is due to the fact that, read SNM of 8-T cell depends primarily on V_{TRIP} (as V_{READ}~0) which is improved by using (110) PMOS. However, use of (110) AX and (110) PUP significantly degrades write margin (Fig. 6). As V_{READ}=0, a higher β-ratio for AX to PD does not impact read stability. Similarly, the cell performance was improved by having stronger NMOS transistors for the read operation. It was also found that, if (100) orientation is chosen for the two read NMOS transistors, the read access time reduces by almost 38% but there is a significantly large area overhead. The possible alternative to fin orientation was increasing the strength of the transistors by going for two fins instead of one (double-fins for AX NMOS in place of one would increase the Write margin). This, however, increased the gate capacitance of the access NMOS, and hence the word-line capacitance, thereby increasing the power consumption. This analysis shows that, effective use of the vertical structure of the FinFET technology (better write stability using multi-orientated cells) coupled with the circuit configuration of 8-T

cell (better read stability), helps to design robust SRAM cell[3].

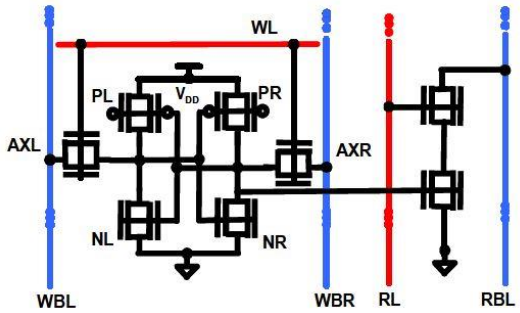

Fig. 10: FinFET based 8-T SRAM cell with single fin pull down

III. DEVICE OPTIMIZATION TECHNIQUE (GATE SIDEWALL OFFSET SPACER OPTIMIZATION) [4]

Device design for an SRAM is governed by the stability, power consumption, and access time of an SRAM cell. Due to the large size and low activity, the power of an SRAM array is dominated by the leakage power of the individual cells. The stability of an SRAM cell is determined by the SNM of the cell. Hence, to design a low-power and robust cell, the total cell leakage needs to be reduced, and the SNM needs to be improved. This section looks at how the gate sidewall spacer thickness can be optimized to reduce leakage and increase robustness of the SRAM cell.

A. Leakage power

$$I_{leak} = I_{sub} + I_{gate} \qquad (1)$$

Because of low activity, SRAM is a relatively cooler section of a chip. At low temperatures, subthreshold leakage (I_{sub}) is not very high, and hence, the contribution of gate leakage (I_{gate}) to the total leakage (I_{leak}) becomes significant [4].

B. SRAM stability

With technology scaling, process variations have made the prediction of SRAM characteristics increasingly difficult. Process variations can result in device mismatch in a cell increasing failure probability. Besides process variations, inherent to technology scaling is the increase in SCE, i.e., drain induced barrier lowering (DIBL). Increasing DIBL can also result in device mismatch in a cell resulting in an increase in failure probability. While reading the cell shown in Fig. 11 [VL(Q)= 1 and VR(QB) = 0], due to the voltage divider action between right access transistor (AXR) and pull down transistor (NR), the voltage at node R (VR) increases to a positive value V_{READ}. If V_{READ} is higher than the trip point of the left inverter PL–NL (V_{TRIPRD}), then the cell flips while reading the cell. This represents a read-failure event. Increase in V_{READ} will increase the drain-to-source (V_{ds}) voltage of NR and decrease the V_{ds} of AXR. Since, nominally, V_{READ} is less than $V_{dd}/2$, the V_t of NR is higher than AXR, making AXR stronger. [4]

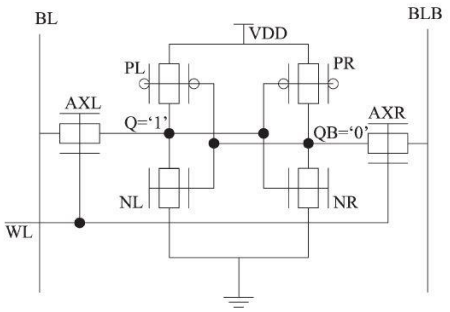

Fig. 11. Double-gate SRAM cell.

Fig. 12 shows the butterfly curves for two SRAM cells designed using devices with very low DIBL and very high DIBL. It can be seen that the SNM of a cell, designed using devices with higher DIBL, reduces because of the increase in strength of access transistor. Reduced SNM results in increased read cell failure.

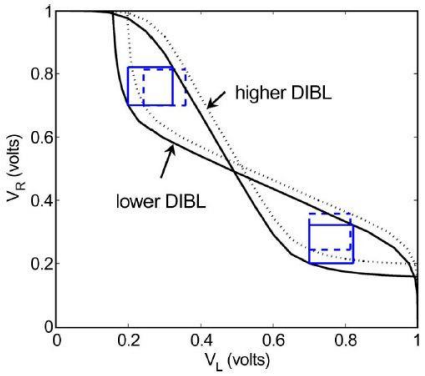

Fig. 12. Butterfly plots to compute SNM.

Fig. 13 shows variation in V_{READ} with increase in DIBL. V_{READ} and, hence, the read-failure probability increases with the increase in DIBL. Therefore, one important parameter to consider while designing devices for a robust SRAM is DIBL. Increase in spacer thickness and, hence the channel length, reduces DIBL resulting in increased stability. It is also observed that reduced SCE (i.e., reduced DIBL) reduces the sensitivity of the threshold voltage to the process induced variations in device geometry. This further improves the stability of an SRAM cell. [4]

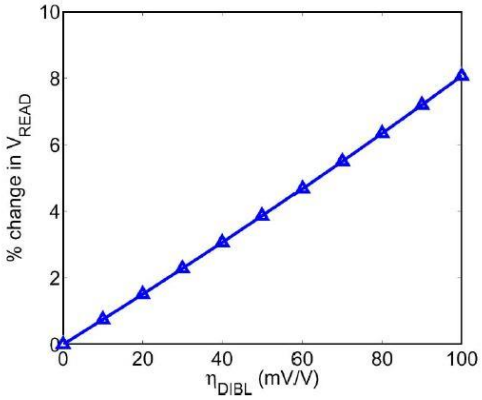

Fig. 13. Variation in *V*READ with DIBL.

C. Device optimization

Fig. 14 shows the effect of increasing the spacer thickness

on device geometry. The source/ drain (S/D) doping peaks were controlled at the edges of the spacers. As the spacer thickness increases, the channel length of the device increases while keeping the bottom dimension of the gate constant.

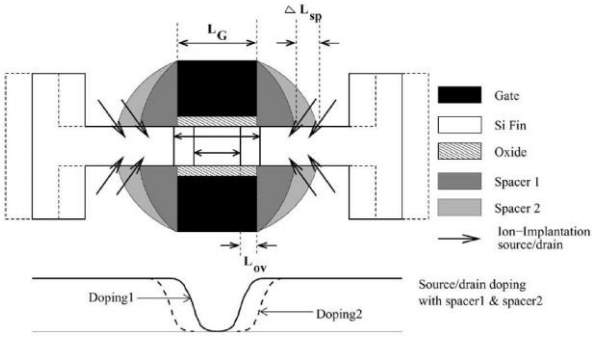

Fig. 14. FinFET schematic showing the effects of increase in spacer thickness.

The effect of increasing spacer thickness (ΔL_{sp}) was analyzed. Leakage currents in nanoscale FinFETs consist of subthreshold current (I_{sub}), gate-to-channel tunneling current ($I_{g,ON}$), and edge direct-tunneling current (I_{edt}) through gate-to-S/D overlap region. In the OFF-state (V_{gs} = 0 V, V_{ds} = V_{dd}), I_{sub} and I_{edt} dominate the total leakage current. Increasing ΔL_{sp} increases L_{ch}, thereby, exponentially reducing I_{sub}. Increase in ΔL_{sp} also reduces I_{edt} because of the reduced gate-S/D overlap length (L_{ov})[10]. The performance of a device is governed by its on-current, gate capacitance (C_g), and drain capacitance (C_d). Since increase in ΔL_{sp} results in higher effective channel length and threshold voltage, on-current reduces with an increase in ΔL_{sp} (Fig. 15). [4]

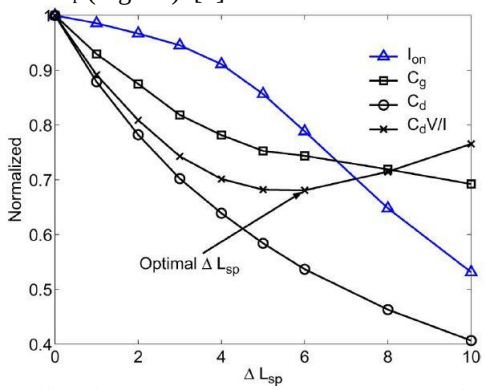

Fig. 15. Effect of Δ*L*sp (spacer thickness) on the device performance.

Sidewall thickness was optimized to minimize OFF-state device leakage (I_{sub} + I_{edt}) and to minimize the ratio of drain capacitance to ON-current (C_dV/I_{ON}). The optimized device showed 82% reduction in I_{sub} and 70x reduction in I_{edt}. The above optimization resulted in 20% reduction in on-current and 47% reduction in C_d (C_dV/I reduces by 30%). Due to better SCE, the optimum device also shows lower DIBL (about 58%) and better subthreshold slope (about 10%). A schematic of how spacers are formed are shown in Fig 16. Ion-implantation of S/D are governed by the outer edges of the spacers. As the spacer thickness increases, the peaks of the S/D move away from each other resulting in increased channel length. Spacer thickness is determined by the thickness of the deposition layer [Fig. 16(a)]. As shown in Fig. 16(b), spacers are formed by vertical reactive ion etching (RIE) of the spacer material.

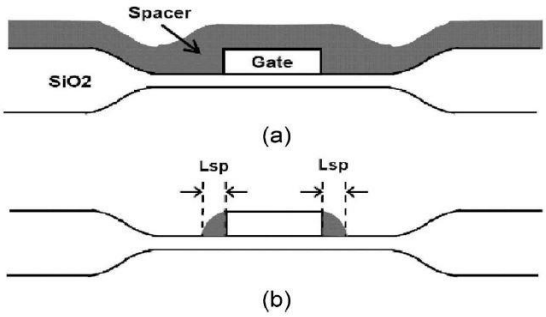

Fig. 16. Spacer thickness (*L*sp) is determined by the thickness of deposition layer. (a) After deposition of spacer material. (b) After vertical RIE.

The assumption here was that the increase in the spacer thickness resulted in moving away of the S/D contacts because of the design constraints restricting the angle of ion-implantation of S/D and contact resistance.

D. Effect of spacer length on SRAM

ΔL_{sp} optimization reduces the I_{sub} (by 82%) and I_{edt} (by 70x) components of the cell leakage (Fig. 17).

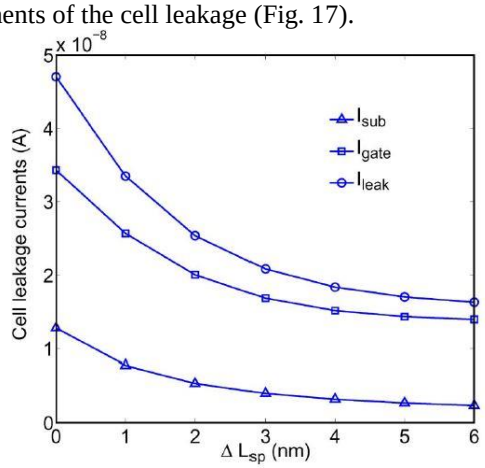

Fig. 17. Effect of Δ*L*sp on SRAM cell leakage.

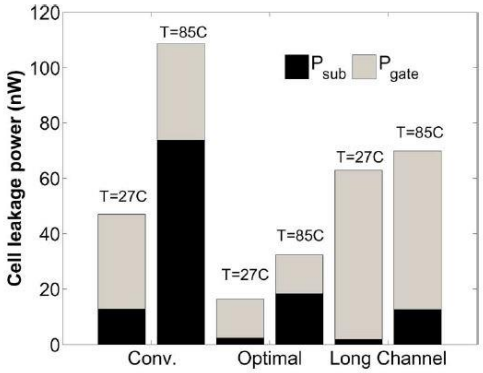

Fig. 18. Normalized subthreshold, gate and total leakage in an SRAM cell at T = 27, 85 ◦C.

However, as $I_{g,ON}$ is insensitive to ΔL_{sp}, the overall reduction in the cell leakage is 65% at T = 27 ◦C and 70% at T = 85˚C (Fig. 17). In longer gate length FinFET (longchannel) SRAM

cell, subthreshold leakage reduces; however, the gate leakage I_{edt} does not change and $I_{g,ON}$ increases because of the increased gate area. This results in over 2*x* increase in leakage-power dissipation in the long-channel SRAM cell compared to the optimal case. [4]

With the scaling of technology, process imperfection is becoming a major concern in maintaining the reliability of an SRAM cell. The major sources of parameter variations in FinFET are T_{si} and L_g [11] and gate shift because of misplaced spacers. [4]

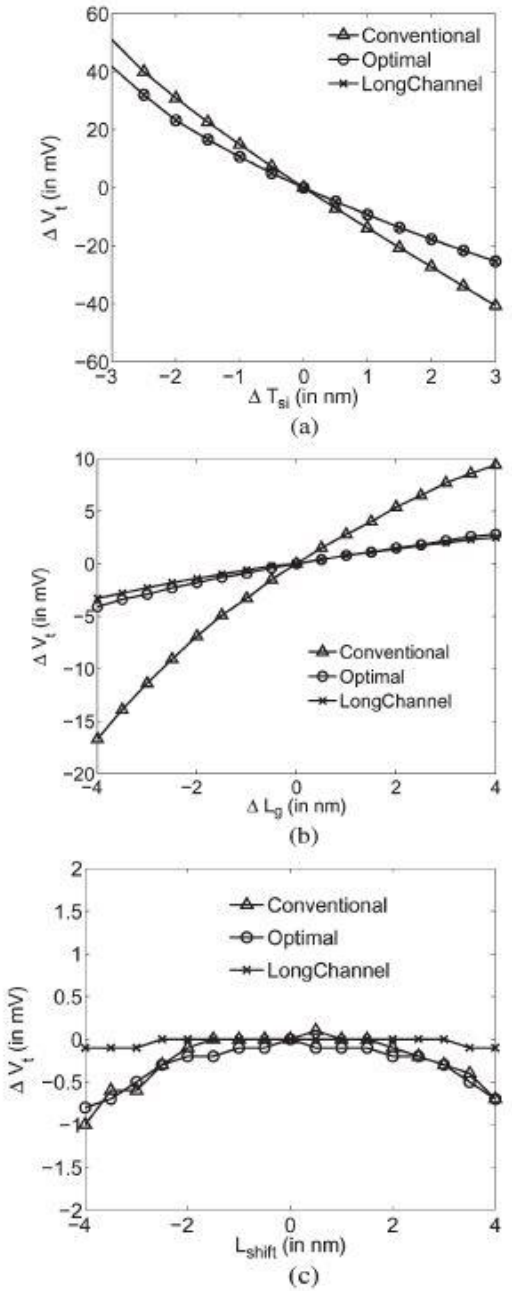

Fig. 19. *Vt* variations in conventional, optimized, and long-channel (by increasing printed gate length) devices with process variations in (a) *T*si, (b)*Lg*, and (c)*L*shift.

Fig. 19 shows the effect of process variations on the threshold voltages (V_t) of the conventional, optimal, and long-channel (by increasing printed gate length) devices. Sensitivity of V_t to process variation is almost the same for optimal and long channel devices because of reduced SCE. We observe that, in optimal device (ΔL_{sp} = 6 nm), the variations in V_t due to variations in L_g and T_{si} reduces. This is attributed to the lower SCE in the optimal device due to the longer effective channel length. We also observe that the effect of gate shift L_{shift} on V_t is minimal. Therefore, for read-failure analysis in SRAMs, we only consider the effect of T_{si} and L_g variations. In the following section, we will look at the read-failure, write-time, and data retention voltage (DRV) analyses for nominal and worst case SRAM cells designed using conventional and optimal devices. [4]

E. Read Failure Analysis

The read-failure analysis was performed on a 6-T SRAM cell designed using conventional and optimal devices considering random variations in *T*si and *Lg* in cell transistors. The variation in *T*si and *Lg* resulted in statistical variation in SNM of a cell. If the SNM of a cell is less than zero, the cell flips during reading. Hence, the read-failure probability can be given by $\Pr\{\text{SNM} < 0\}$. The distribution in SNM for each transistor considering its *T*si and *Lg* variation, and overall SNM distribution was obtained. [4]

It can be seen that 1) the mean of the distribution increases because of the increase in *Vt* and 2) the standard deviation decreases because of the reduced sensitivity to process variations. [4]

Fig. 20 shows the overall SNM distribution of the cell with conventional and optimal devices. It can be observed that the tail of the SNM distribution in case of the optimal device is shifted to the right which results in a significant reduction in read-failure probability. [4]

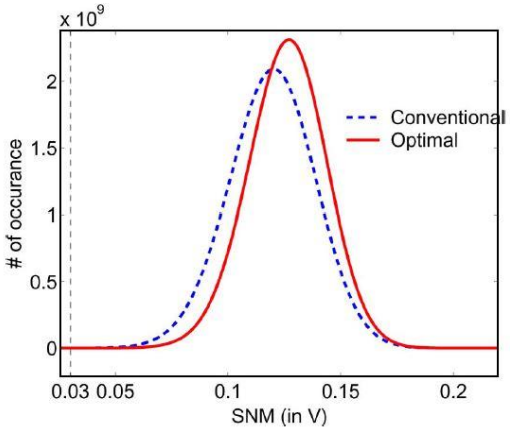

Fig. 20. Overall SNM distribution.

F. Write Time Analysis

Write time is the time required to flip the cell, i.e., left and right charge storing nodes VQ and VQB (Fig. 11), respectively. Table I shows the write times in nominal and worst cases for three SRAM cells. In the worst case, for node storing "0," the pull-down transistor is strong $(-dVt)$, and the

pull-up and access transistors are weak $(+dV_t)$. Similarly, for node storing "1," the pull-up transistor is strong $(-dV_t)$, and the pull-down and access transistors are weak $(+dV_t)$. Write time reduces with the reduction in total capacitance at the nodes storing data. [4]

	Parameter	Value		
		Conventional	Optimal	Long-channel
Read Stability Analysis	SNM (mV)	129	132	137
Write Time (ps)	nominal	25.4	25.7	23.8
	worst case	54.2	42.3	52.9
DRV	worst case (V)	0.22	0.16	0.15

Table 1 : SRAM ARRAY STABILITY (TALL CELL LAYOUT)

In the optimal devices, effective gate and drain capacitances reduced resulting in reduced capacitance at the charge-storing node. However, lower on-current adversely affected the write time. In the nominal case, all the three SRAM cells have similar write times. In the worst case, write time of optimal device is comparatively lower because of reduced sensitivity of V_t to process variations. [4]

G. Data Retention Voltage (DRV)

DRV is the minimum supply voltage at which the stored data in a cell is still preserved . DRV has been calculated for the SRAM cells designed using three devices. Worst case variations in V_t (due to process variations in L_g and T_{si}) resulting in worst SNM have been considered to compute maximum DRV[4].Table I shows the DRV values for three cases. It can be seen that DRV for optimal devices is lower than conventional because of increased robustness.

IV. ASYMMETRIC DRAIN SPACER FINFETS [6]

This approach is different from the approach in the previous section in the sense that sidewall spacer thickness (L_{sp}) is increased by (ΔL_{sp}) on the drain side only. The source side overlap remains the same as conventional FinFETs. The gate length (L_g) is not changed. This structure (Fig 21) can be fabricated using the same procedure for a standard FinFET, with an additional step for introducing underlap on the drain side. One method of introducing the asymmetric spacer on the drain side is by depositing the spacer material with a mask on the source side to prevent any material from depositing on the source side. Then, reactive ion etching (RIE) can be used to vertically etch the spacer material. Once the extra spacer is formed on the source drain side, S/D doping can be performed to achieve the required doping profile.

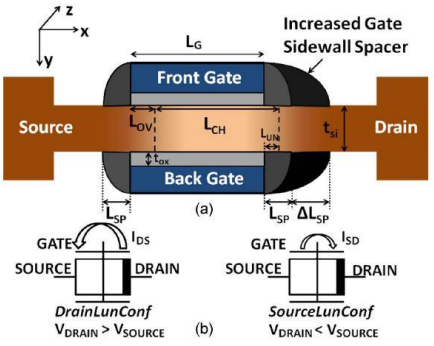

Fig. 21. (a) Structure of ADSE FinFET showing gate–drain underlap (*L*UN). (b) ADSE-FinFET based pass transistor showing bidirectional current flow. The underlapped terminal is indicated by a thick line.

Two-dimensional simulations were carried out for parameters that correspond to the 32-nm technology node using the drift-diffusion-based Taurus device simulator [5]. L_G = 32 nm, t_{si} (silicon body thickness) = 9 nm and t_{ox} (gate oxide thickness) = 1.6 nm were used with intrinsic silicon body and S/D doping of 10_{20} cm^{-3}. The work function difference between the gate and the silicon body was −0.1 eV for NMOS and 0.11 eV for p-channel metal–oxide semiconductor (PMOS). L_{SP} was chosen so that the overlap between the gate and the source (L_{OV}) was 3.5 nm. [6]

The proposed device structure does not have any asymmetry with respect to the front and back gates (i.e., *t*ox, gate work function, and gate voltage for the front and back gates are equal). The asymmetry that we discuss in this section comes only from the increase in the spacer thickness introduced on the drain side. This condition results in a gate–drain underlap and an asymmetry in drain current (I_D) for positive and negative V_{DS}. As shown in Fig. 21(a), the terminal with underlap has been defined to be the drain terminal. Hence, positive V_{DS} for NMOS (and negative V_{DS} for PMOS) implies that the terminal with underlap has higher voltage than the other terminal and hence acts as a drain for the electrons. This configuration (i.e., $V_{DS} > 0$ for NMOS and $V_{DS} < 0$ for PMOS) was called *DrainLunConf*. Negative V_{DS} for NMOS (and positive V_{DS} for PMOS) means that the terminal with underlap has lower voltage and hence acts as a source for electrons. This configuration ($V_{DS} < 0$ for NMOS and $V_{DS} > 0$ for PMOS) was called *SourceLunConf* for the purposes of analysis. [6]

A. Effect on Underlap on surface potential, conduction band profile, drain current in subthreshold region, subthreshold characteristics and gate current in ADSE FinFETs

We start the discussion for the case when $V_{DS} = 0$. Hence, the discussion is valid for both SourceLunConf and DrainLunConf. Subsequently, we extend the discussion for nonzero V_{DS} and separately consider the two configurations. Fig. 22(a) shows the conduction band profile for ΔL_{SP} = 15 nm at V_{DS} = 0 V for different V_{GS}. It can be observed that, for high V_{GS} (greater than the threshold voltage V_{th} of the device), there is an extra potential barrier (EPB) for the electrons in the underlapped part of the device. The height of the barrier increases with increasing V_{GS}. EPB does not exist for low V_{GS}. [6]

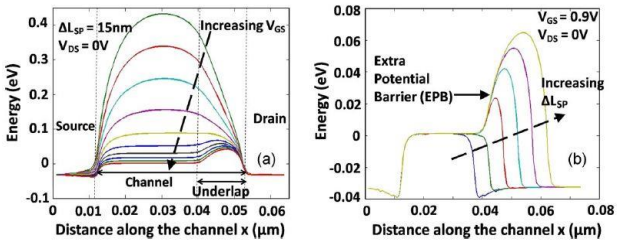

Fig. 22. Conduction band profile for (a) different values of *V*GS at *V*DS = 0 V and Δ*L*SP = 15 nm. EPB appears at high *V*GS for (b) different values of Δ*L*SP at *V*GS = 0.9 V and *V*DS = 0 V.

Let us now look at the effect of V_{DS} on EPB. For D*rainLunConf*, the drain voltage pulls down the conduction band in the vicinity of the drain. This condition reduces the height of EPB [Fig. 23(a)]. For *SourceLunConf*, higher voltage is applied at the non-underlapped terminal. Hence, the effect of increasing |*V*DS| on EPB in the underlapped region is not significant [Fig. 23(b)]. To sum up, the effect of underlap on the electrostatics of the device is the emergence of an EPB at high *V*GS. EPB is not significant for low values of underlap.

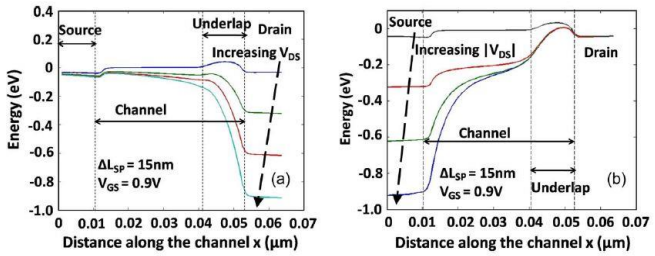

Fig. 23. Conduction band profile for different values of VDS at VGS = 0.9 V and ΔLSP = 15 nm for (a) DrainLunConf and (b) SourceLunConf.

The drain current (for $V_{GS} > V_{th}$) in ADSE FinFETs is affected by underlap in the following two ways: 1) effective channel length increases and 2) device electrostatics change; in particular, EPB is introduced. [6]

An increase in the effective channel length of the devices due to gate–drain underlap results in the reduction of SCEs. The subthreshold leakage current (I_{SUB}), SS, and DIBL decrease with increasing underlap [see Fig. 24(a) and (b)]. In Fig. 24(a), I_{OFF} is I_{SUB} at V_{GS} = 0 and $V_{DS} = V_{DD}$). [6]

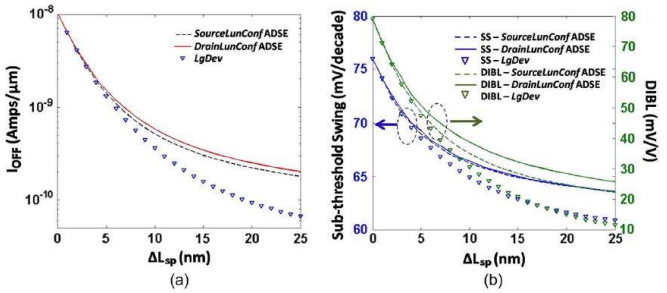

Fig. 24. Comparison of (a) OFF-currents and (b) SS and DIBL versus ΔLSP for SourceLunConf, DrainLunConf,

The gate leakage current (I_{GATE}) is composed of the following two primary components: 1) current due to direct tunneling of the carriers from the channel to the gate and 2) edge tunneling current from the S/D to the gate in the gate–S/D overlap portion. For low V_{GS}, carrier concentration in the channel is less, which makes direct tunneling from the channel to the gate negligible. [6]

For *DrainLunConf*, the underlap on the drain terminal decreases the effect of drain voltage on the channel, which results in lower surface potential in the nonunderlapped part of the channel compared to the nominal device ($\Delta L_{SP} = 0$). The electric field in the gate dielectric increases, which results in increasing I_{GATE} with ΔL_{SP} [Fig. 25(b)]. For *SourceLunConf*, the underlap is on the terminal with zero voltage bias, and it decreases the effect of zero bias on the surface potential in the non-underlapped part of the channel. [6]

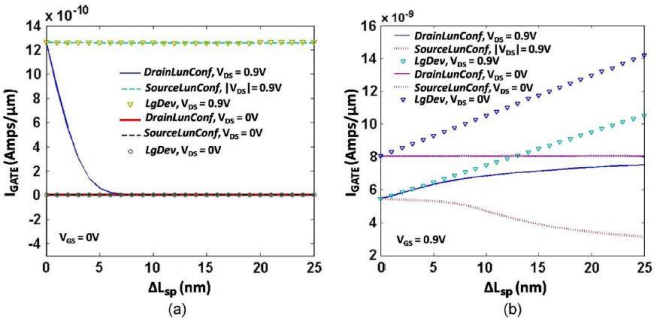

Fig. 25. Gate leakage current versus Δ*L*SP for *SourceLunConf*, *DrainLunConf* at *V*DS = 0 V and 0.9 V and *V*GS = (a) 0 V and (b) 0.9 V.

B. ADSE FinFET based 6T SRAM Cell

The device design for SRAM is governed by the stability (read and write), power consumption, access time, and area of an SRAM cell. The stability of an SRAM cell can be estimated by the read static-noise margin (SNM) and the WM of the cell. During the read operation, bit-lines BL and BLB are precharged to V_{DD}, and word line (WL) is asserted (Fig. 26). The voltage at node Q (V_Q) rises to some positive value (V_{read}), depending on the resistive divider action of access transistor (AXR) and pull-down transistor (NR) and can cause a flip in the stored data. A weaker access transistor and, hence, lower V_Q implies larger read stability. During the write operation, BL is charged to V_{DD} and BLB is discharged to *GND*. The voltage at node QB (V_{QB}) is pulled down to a value, depending on the relative strengths of pull-up transistor (PL) and access transistor (AXL). A stronger access transistor and, hence, lower V_{QB} implies larger write stability. The write stability of a cell can be characterized using WM. [6]

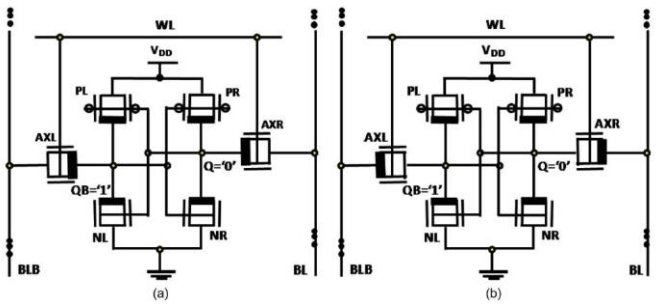

Fig. 26. Two configurations of the ADSE 6T SRAM schematic based on the connection of access transistors. (a) *CUS* configuration. (b) *CUB* configuration

However, the conventional definition of WM [12] does not capture the dependence of write stability on the access transistor [13], which is particularly important for characterizing ADSE-FinFET based 6T SRAM cells. Therefore, the definition proposed in [13] is used to estimate the write stability of the cell. While implementing SRAM, designers use transistors with gate length longer than the minimum gate length provided by the technology [14]. This approach is done to reduce the SCE and improve leakage. Thus, in this paper, we also implement SRAM with *LgDev* devices and compare those with ADSE-based SRAM. *LgDev* devices are those devices having the same channel length as

ADSE FinFETs but without any underlap. Fig. 26 shows the two possible configurations in which an ADSE device SRAM bit cell can be implemented. The configurations are called *ConnectUnderlapStorage (CUS)* and *ConnectUnderlapBitline (CUB)*, respectively. The number of fins used is 1, 1, and 2 for the pull-up (PL and PR), access (AXL and AXR), and pull-down (NL and NR) transistors, respectively, in both configurations. [6]

C. CUS SRAM

In the *CUS* SRAM, the access transistors are connected such that the terminal with underlap is connected to the storage nodes. In this configuration, the asymmetry of the current flow in the access transistors (AXR and AXL) to improve both read and write stability is made use of. Looking at the read operation. As aforementioned, when we perform a read operation, node Q is at V_{read}, and BL is at V_{DD}. Because V_{DS}(AXR) < 0, reverse drain current (I_{DR}) flows through AXR during a read operation. The EPB (extra potential barrier) on the source side reduces I_{DR}, thus rendering AXR weak compared to NR. This condition results in a lower V_{read} appearing on node Q during the read operation. As we increase Δ*Lsp*, AXR becomes much weaker, and V_{read} reduces. This condition helps in improving the read stability of the SRAM bit cell. Fig. 27(a) shows the variation of read SNM with Δ*Lsp*. It is shown that we get a significant improvement in read SNM with an increase in Δ*Lsp*. [6]

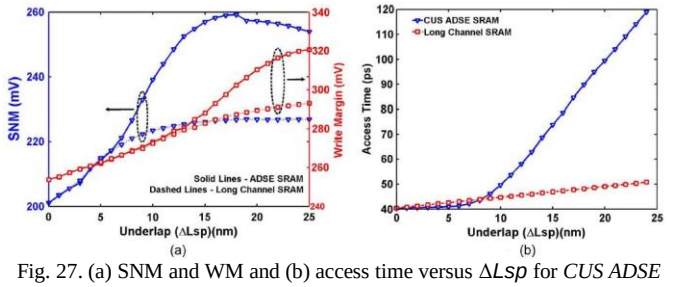

Fig. 27. (a) SNM and WM and (b) access time versus Δ*Lsp* for *CUS ADSE* and long-channel-based SRAM.

We now discuss how the improvement in the read SNM is obtained not only due to an increase in the channel length but also because of the effect of EPB on the source side of the access transistor. Fig. 27(a) shows the plot of read SNM versus Δ*Lsp* for *LgDev*-based SRAM. It is shown that, for Δ*Lsp* > 7 nm, there is a significant improvement in the read SNM for ADSE SRAM compared with the *LgDev*-based SRAM. This is due to the effect of the EPB, which plays a key role in decreasing the strength of AXR compared to NR. This condition leads to a larger improvement in read SNM. This effect is not observed in *LgDev*-based SRAM. Thus, read SNM saturates after some increase in gate length and does not further improve.

Let us now consider the write operation in CUS SRAM. Let us assume that the cell stores a "0," and we need to write "1" to it. To perform this operation, BL will be charged to V_{DD}, and BLB will be discharged to GND. Because V_{DS}(AXL) > 0, I_{DF} flows through AXL during a write operation. We know that the effect of EPB reduces with increasing V_{DS}. This effect results in a higher I_{DF} through AXL. On the other hand, V_{DS} for PL is low, due to which the EPB reduces its strength. This condition results in an increase in the relative strength of AXL. Fig. 27(a) shows the plot of WM versus Δ*Lsp* for ADSE and *LgDev*-based SRAM. It is shown that WM increases with Δ*Lsp*. ADSE based SRAM shows larger improvement in WM compared to *LgDev*-based SRAM. This case indicates a useful role of the EPB in improving the WM.

Fig. 27(b) shows the dependence of access time on Δ*Lsp* for ADSE and *LgDev* SRAM. 256 bit cells in one column, and the bitline capacitance is taken as 0.2 fF/um [15]. It is shown that access time degrades more rapidly in ADSE-based SRAM compared to *LgDev* SRAM. As seen previously, the access transistor is rendered weak during the read operation (because of the EPB), and this results in higher access time. The access time also increases for *LgDev* due to a reduction in I_{ON} with an increase in gate length, but this increase is much less compared to ADSE SRAM. [6]

D. CUB SRAM

In the *CUB* SRAM configuration, access transistors are connected so that the terminal with Δ*Lsp* is connected to the bitlines (BL and BLB). This case results in the reduction of drain capacitance of the access transistors connected to the bitline in the column [5]. Fig. 28(a) shows the change in access time with an increase in Δ*Lsp* for different values of bitline capacitance. We observe that, if the bitline capacitance dominates the total capacitance (C_{BL} = 0.2 fF/um), then the access time increases because of the reduction in I_{ON}-current. However, if the drain capacitance dominates the total capacitance (C_{BL} = 0.02 fF/um), then the access time decreases with an increase in Δ*Lsp*. In present-day technologies, the total capacitance is dominated by bitline capacitance [15]. This is also expected to be true in future technologies. Fig. 28(b) compares the read SNM and WM of *CUB* SRAM and *LgDev*-based SRAM. As it can be observed, both read SNM and WM for *CUB* SRAM are worse compared to *LgDev* SRAM. This is because, in this configuration, we are not able to take advantage of the asymmetry in the current flow that was possible in the *CUS* SRAM configuration. Thus, *CUB* SRAM does not provide us with much benefit compared to the *LgDev*-based SRAM. [6]

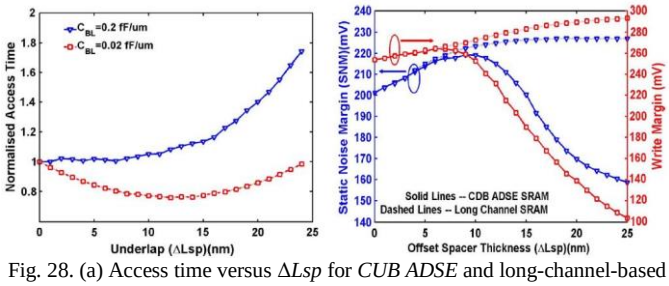

Fig. 28. (a) Access time versus Δ*Lsp* for *CUB ADSE* and long-channel-based SRAM for different bitline capacitance levels. (b) SNM and WM versus Δ*Lsp* for *CUB ADSE* and long-channel-based SRAM.

Having analyzed both *CUB* and *CUS* configurations, the *CUS* configuration helps in improving both read and write stability, at the cost of increased access time and area. However, *CUB* does not provide any improvement in read SNM, WM, and access time. Thus the performance metrics are derived for the *CUS* configuration of ADSE SRAM. [6]

E. Comparison of performance metrics with conventional SRAMs

Compared to the conventional device (ΔLsp = 0), the ADSE optimum device results in 57% reduction in cell leakage, 11% improvement in SNM, and 6% improvement in WM (Fig 29). This result comes at the expense of 7% increase in access time and area. Even when compared to the optimum *LgDev* device, using the optimum ADSE device gives 29% improvement in cell leakage and 3% improvement in SNM with similar WM and access time. [6]

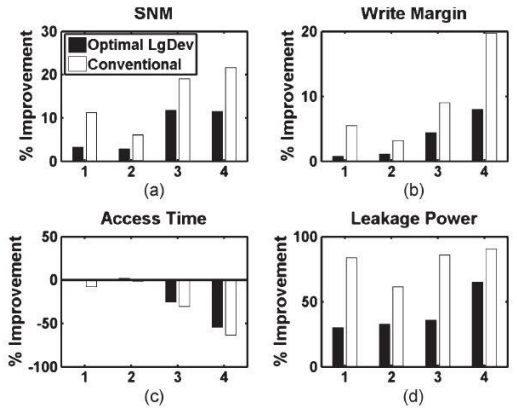

Fig. 29. Percentage improvement of *CUS* ADSE SRAM in (a) SNM, (b) WM, (c) TACC, and (d) leakage power for four cases: 1) *EqWt*; 2) *high performance*; 3) *low leakage*; and 4) *high stability*.

For a high-performance application, the optimal device has very small ΔL_{sp} because as the spacer length increases, the access time increases and it is better to use a device with channel length close to a conventional device. For high stability, optimal devices have very high $\Delta L_{sp,}$ this is because with increase in channel length, the SCE effects are mitigated and this in turn increases the robustness of the SRAM cell. In ADSE SRAMs, asymmetry in access transistors in addition to longer channel length is responsible for significant improvement in stability.

V. CONCLUSIONS

The first analysis showed at the same leakage power, cells with optimized fin-orientation showed higher cell stability and performance. The second technique where the spacer technique is optimized for reducing the leakage current and minimizing the effect of performance degradation reduces the sensitivity of the threshold voltage to fluctuations in process parameters such as silicon thickness and gate length. A 65% reduction in gate leakage power and 200x reduction in read-failure probability with only a marginal increase in cell area was observed. Finally, the technique of introducing a drain spacer extension was used to achieve a 11% improvement in read SNM, along with 6% improvement in write margin in 6T SRAM cells. There was also a 57% reduction in leakage power compared to conventional SRAM cells.

ACKNOWLEDGEMENT

The author wishes to thank Dr. Paul Franzon, NCSU for initiating the study and exploration of FinFET based SRAM design. The author also wishes to thank NCSU libraries for access to IEEE articles, journals and references which were helpful in the production of this paper.

REFERENCES

[1] *International Technology Roadmap for Semiconductors*, 2006.

[2] K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, D. J. Pearson, and N. J. Rohrer, "High performance CMOS variability in the 65-nm regime and beyond," *IBM J. Res. Develop.*, vol. 50, no. 4/5, pp. 433–449, Jul.–Sep. 2006.

[3] Gangwal, S.; Mukhopadhyay, S.; Roy, K., "Optimization of Surface Orientation for High-Performance, Low-Power and Robust FinFET SRAM," *Custom Integrated Circuits Conference, 2006. CICC '06. IEEE* , vol., no., pp.433,436, 10-13 Sept. 2006.

[4] Bansal, A.; Mukhopadhyay, S.; Roy, K., "Device-Optimization Technique for Robust and Low-Power FinFET SRAM Design in NanoScale Era," *Electron Devices, IEEE Transactions on* , vol.54, no.6, pp.1409,1419, June 2007.

[5] Taurus Device Simulator, Synopsys Inc., v W-2004.09.

[6] Goel, A.; Gupta, S.K.; Roy, K., "Asymmetric Drain Spacer Extension (ADSE) FinFETs for Low-Power and Robust SRAMs," *Electron Devices, IEEE Transactions on* , vol.58, no.2, pp.296,308, Feb. 2011.

[7] Moradi, F.; Gupta, S.K.; Panagopoulos, G.; Wisland, D.T.; Mahmoodi, H.; Roy, K., "Asymmetrically Doped FinFETs for Low-Power Robust SRAMs," *Electron Devices, IEEE Transactions on* , vol.58, no.12, pp.4241,4249, Dec. 2011.

[8] MEDICI Version V-2003.12-0, Synopsys.

[9] L. Chang, et. al. IEEE TED, pp. 1621-27, Oct. 2004.

[10] Bansal, A.; Paul, B.C.; Roy, K., "Impact of gate underlap on gate capacitance and gate tunneling current in 16 nm DGMOS devices," SOI Conference, 2004. Proceedings. 2004 IEEE International , vol., no., pp.94,95, 4-7 Oct. 2004.

[11] S. Xiong and J. Bokor, "Sensitivity of double-gate and FinFET Devices to process variations," *IEEE Trans. Electron Devices*, vol. 50, no. 11, pp. 2255–2261, Nov. 2003.

[12] W. J. Dally and J.W. Poulton, *Digital Systems Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[13] K. Takeda, H. Ikeda, Y. Hagihara, M. Nomura, and H. Kobatake, "Redefinition of write margin for next-generation SRAM and write margin monitoring circuit," in *Proc. IEEE Int. Solid State Circuits Conf.*, Feb. 2006, pp. 2602–2611.

[14] S. Mukhopadhyay, H. Mahmoodi, and K. Roy, "Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscale CMOS," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 12, pp. 1859–1880, Dec. 2005.

[15] ITRS, 2008. [Online]. Available: http://www.itrs.net.