Simulation-based Study of Super-steep Retrograde Doped Bulk FinFET Technology and 6T-SRAM Yield

Xi Zhang

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2016-3
http://www.eecs.berkeley.edu/Pubs/TechRpts/2016/EECS-2016-3.html

January 5, 2016

Acknowledgement

To my wonderful parents, for your unbounded love and support.
To Yashu, you are my wonder of wonder.

Simulation-based Study of Super-steep Retrograde Doped Bulk FinFET Technology and 6T-SRAM Yield

by Xi Zhang

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

Tsu-Jae King Liu
Research Advisor

Date

* * * * * *

Borivoje Nikolić
Second Reader

Date

Abstract

Simulation-based Study of Super-steep Retrograde Doped Bulk FinFET Technology and 6T-SRAM Yield

by

Xi Zhang

Master of Science in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Tsu Jae King Liu, Chair

FinFET technology already has been adopted by the semiconductor industry beginning at the 22 nm node of complementary metal-oxide-semiconductor (CMOS) field-effect transistor (FET) technology, due to the superior electrostatic integrity of multi-gate transistor structures. Although silicon-on-insulator (SOI) wafers are ideal substrates for the manufacture of FinFETs with low off-state leakage current, they are more expensive than conventional bulk-silicon wafers. If a bulk-silicon wafer is used as the substrate for FinFET fabrication, heavy "punch-through stopper" doping is needed at the base of the fins to suppress off-state leakage current. A conventional doping process results in dopants within the fin (channel region), however, which degrades transistor on-state performance. The benefits of a super-steep retrograde fin doping profile (such that the channel region is lightly doped while the base region is heavily doped), which can be achieved using oxygen insertion technology, are quantified through three-dimensional (3-D) simulations, for low-power FinFET technology at the 8/7nm node.

As the transistor gate length is scaled down, variability in transistor threshold-voltage (V_t) increases; this hinders reductions in operating voltage, particularly for large arrays of six-transistor (6T) static memory (SRAM) cells. 3-D device simulations are performed using Sentaurus Device to investigate systematic sources and random sources of V_{t} variation, for both SSR FinFET and control FinFET devices. A compact analytical current-voltage (I-V) model calibrated to the 3-D device simulations is used to estimate 6T-SRAM cell yield and minimum operating voltage, to quantify the benefit of oxygen insertion technology for voltage scaling.

To my wonderful parents, for your unbounded love and support.

To Yashu, you are my wonder of wonder.

Contents

List of Figures

List of Tables

Chapter 1

Introduction: Background and Motivation

The continued miniaturization of the metal-oxide-semiconductor field-effect transistor (MOSFET) to gate lengths below 25 nm requires the adoption of multi-gate structures in order to maintain good gate control of the electric potential in the channel region, *i.e.* good electrostatic integrity. For this reason, a three-dimensional (fin-shaped) channel structure straddled on three sides by the gate electrode, referred to as either the "tri-gate" or "FinFET" design, was adopted by Intel Corporation for mass production of complementary MOS (CMOS) integrated-circuit (IC) products, beginning at the 22 nm technology node [1]. The scalability of the FinFET design to sub-25 nm gate lengths was first demonstrated at the University of California, Berkeley[2].

Fig. 1.1 shows how the increasing latency of dynamic random access memory (DRAM) and process-memory performance gap with advancements in CMOS technology. This trend has driven growth in cache memory (static RAM, or SRAM) capacity on-chip (monolithically integrated with the processor). As transistors are scaled down in size, however, process-induced variations in threshold voltage (V_t) increase [3] and pose a serious challenge for achieving sufficiently high yield for large SRAM arrays. The superior electrostatic integrity of the FinFET not only provides for higher transistor on-state current for a given operating voltage - or lower operating voltage for a given level of circuit performance - but also mitigates the short-channel effect (SCE) and drain-induced barrier lowering (DIBL) for reduced performance sensitivity to process-induced variations, to overcome this challenge.

Although silicon-on-insulator (SOI) wafers are ideal substrates for the manufacture of FinFETs

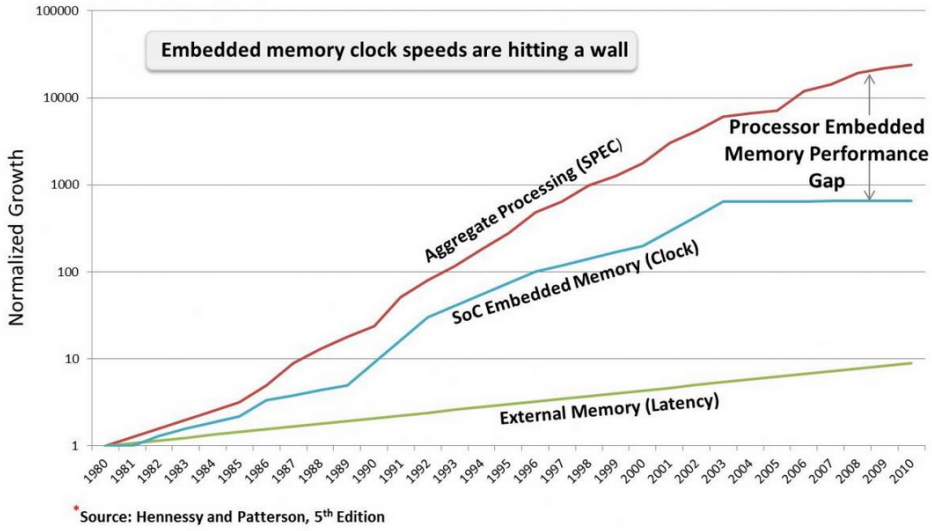

Figure 1.1: Increasing DRAM latency and processor-memory performance gap [4].

with low off-state leakage current [5], they are much more expensive than conventional bulk-silicon wafers. If a bulk-silicon wafer is used as the substrate, heavy doping is needed at the base of the fins to suppress off-state leakage current (I_{off}). However, a conventional doping process results in dopants within the fin (channel region), on the order of 1×10^{18} cm^{-3}. Such heavy channel doping degrades the on-state current (I_{on}) of bulk-silicon FinFETs, due to Coulombic scattering, so that they have lower $I_{\text{on}}/I_{\text{off}}$ ratio relative to SOI FinFETs [5]. Also, heavy channel doping results in random dopant fluctuation (RDF) induced V_{t} variation (σV_t) which is proportional to $1/\sqrt{\text{WL}}$[6], where W is the channel width and L is the channel length. oxygen insertion technology can provide for super-steep retrograde doping profiles [7], *i.e.* lightly doped channel regions together with heavily doped base regions, to overcome these challenges for bulk-silicon FinFETs, and thereby improve $I_{\text{on}}/I_{\text{off}}$ and reduce V_{t} variation to facilitate reductions in operating voltage for power savings.

1.1 Bulk FinFET Structures

Fig. 1.2 shows perspective views of the three-dimensional (3-D) FinFET structures simulated using Sentaurus Device [8] in this study. Note that the channel region of the control FinFET is heavily

doped, whereas the channel region of the SSR FinFET is lightly doped. The substrate doping, source/drain doping profiles and geometrical design parameters are identical for both devices.

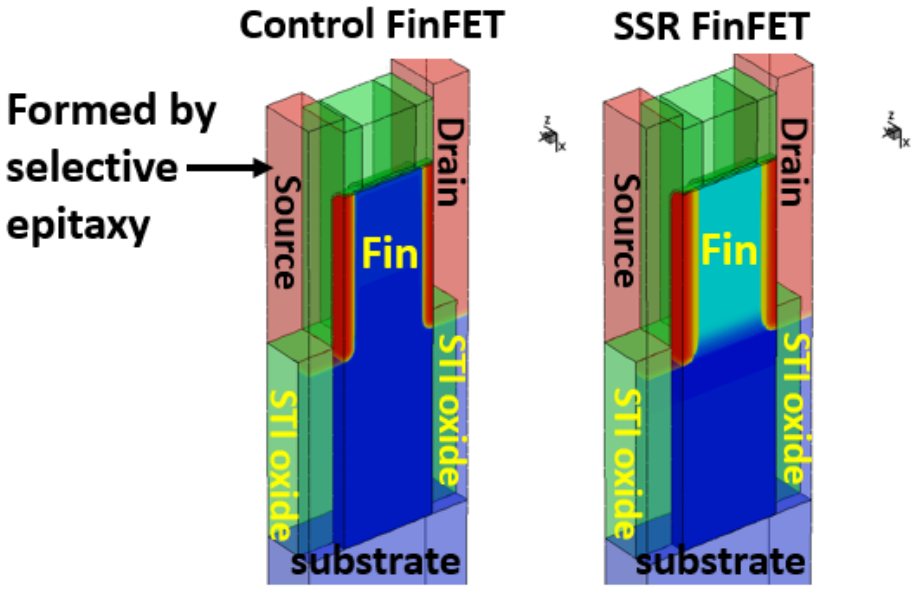

Figure 1.2: Simulated 3-D n-channel FinFET structures. The net dopant concentration is represented in color using a hyperbolic arcsine scale.

Nominal Design Parameter Values The FinFET gate length is 15 nm, which corresponds to the 8/7 nm technology node in the International Technology Roadmap for Semiconductors (ITRS), 2013 Edition [9]. The equivalent gate-oxide thickness is 0.7 nm. The fin height (H_{Si}) is 40nm and the fin width (W_{Si}) is 8nm, so that the fin aspect ratio is 5. The gate work function is assumed to be tunable to achieve an off-state leakage current specification of 100 pA/μm, as required for low-power applications. The current is normalized to the effective width of FinFET (W_{eff}), which is defined as the peripheral length of the silicon fin region, *i.e.*the sum of the lengths of the top portion of the fin and the sidewall portions of the fin. The fin shape is rectangular with rounded corners (with 1 nm radius of curvature), as in Intel's 14 nm FinFET technology [10]. Therefore, the effective width of FinFET used to normalize the current in this study is $W_{\text{eff}} = 2 \times (H_{\text{Si}} - 1nm) + W_{\text{Si}} - 2nm + \pi \times 1nm = 87.1416nm$. The thickness of the shallow-trench isolation (STI) oxide is 50 nm.

The FinFET structures each comprise heavily doped raised source/drain (S/D) contact regions

formed by selective epitaxial growth (SEG)[11], for reduced parasitic resistance. Abrupt source/drain doping profiles are advantageous for reduced parasitic series resistance, for higher I_{on} [12]. As shown in Fig. 1.3, doping gradients as steep as 1.4 nm/dec and 1.8 nm/dec have been experimentally achieved for epitaxially grown silicon doped with phosphorus and carbon (for n-channel FETs) and epitaxially grown silicon-germanium doped with boron (for p-channel FETs), respectively. In this work, the S/D junctions are assumed to have a Gaussian doping profile with 2 nm/dec gradient and peak concentration 2×10^{20} cm^{-3}. Fig. 1.4 shows the simulated vertical doping profile in the source/drain regions. The SEG S/D regions comprise doped silicon for n-channel FETs, and silicon-germanium (SiGe) with germanium concentration of 50% [1] for p-channel FETs, with parameter values taken from [13]. Ohmic contacts (specific contact resistivity $3 \times 10^{-9}\ \Omega \cdot \text{cm}^2$) are made to the top surfaces of the SEG S/D regions.

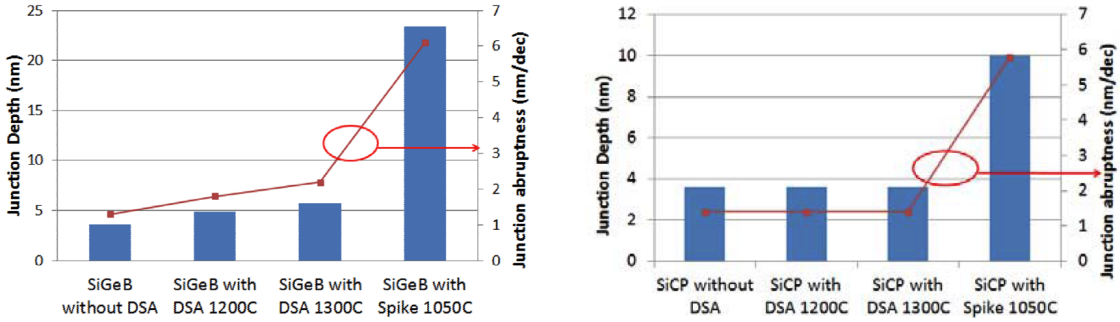

Figure 1.3: Experimentally measured source/drain junction depth and abruptness of SiGeB splits (left) and SiCP splits (right) from [14].

Table 1.1 summarizes the nominal values of the various design parameters for the SSR FinFETs and control FinFETs in this study.

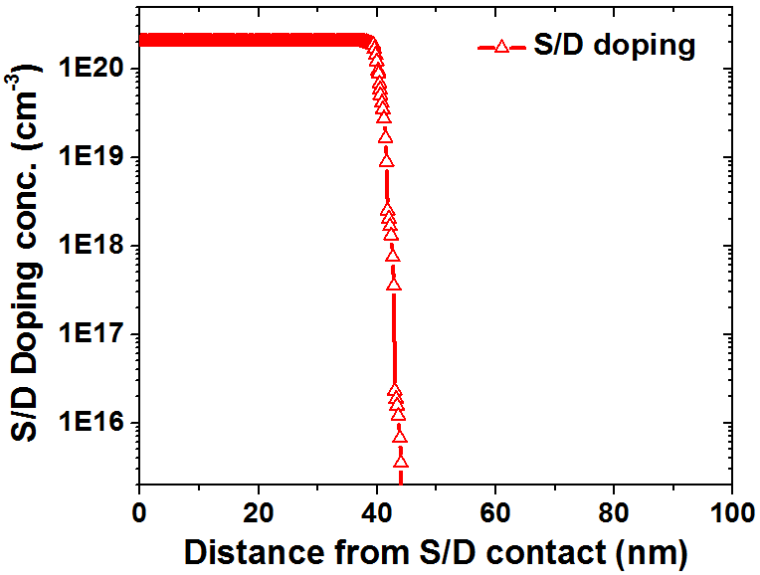

Figure 1.4: Vertical source/drain doping profile, with a gradient of 2 nm/dec at the base of the fin.

	control FinFET		SSR FinFET	
	N-channel	P-channel	N-channel	P-channel
L_{gate} (nm)	15	15	15	15
EOT (nm)	0.7	0.7	0.7	0.7
I_{off} spec (pA/μm)	100	100	100	100
H_{Si} (nm)	40	40	40	40
W_{Si} (nm)	8	8	8	8
N_{fin} (cm^{-3})	1×10^{18}	1×10^{18}	1×10^{16}	1×10^{16}
$N_{\text{fin,peak}}$ (cm^{-3})	2.5×10^{18}	2.5×10^{18}	5×10^{18}	5×10^{18}
N_{sub} (cm^{-3})	2.5×10^{18}	2.5×10^{18}	2.5×10^{18}	2.5×10^{18}
SD Doping Gradient (nm/dec)	2	2	2	2
N_{SD} (cm^{-3})	2×10^{20}	2×10^{20}	2×10^{20}	2×10^{20}
fin doping gradient (nm/dec)	>40	>40	3.3	6.9

Table 1.1: Device Structure Parameters for Nominal FinFET Design

Chapter 2

Comparative Study of SSR FinFET $vs.$ control FinFET

2.1 Introduction

Planar ultra-thin-body FET structures provide for superior gate control and eliminate the need for in the channel/body region, and therefore can be adopted to reduce process-induced variations in transistor performance[15]. However, they require SOI wafer substrates which cost much more than conventional bulk-silicon wafers. Therefore, the leading semiconductor companies (namely Intel Corporation and Samsung Electronics) have adopted the "FinFET" three-dimensional (3-D) thin-body FET structure for high-volume CMOS production on bulk-silicon wafers [1] [10]. To suppress unwanted "punch-through" current flow in the silicon below the fin (out of the control of the gate) between the source and drain regions, heavy doping is needed at the base of the fin. Ideally there should be negligible doping in the fin channel region, *i.e.* a super-steep retrograde (SSR) fin doping profile, to maximize charge-carrier mobility and to minimize threshold voltage (V_t) variations due to random dopant fluctuation (RDF) effects. In this chapter, the benefits of SSR fin channel doping, which can be achieved using oxygen insertion technology, for improved bulk-silicon FinFET performance are investigated via 3-D semiconductor device simulations. A compact model is then calibrated to the simulated characteristics of the optimized FinFET designs,

to enable fast estimation of SRAM cell performance and yield.

2.2 Device Simulation and Design Optimization

2.2.1 Device Simulation Approach

The technology computer aided design (TCAD) software package Sentaurus Device [8] was used to simulate FinFET performance using the drift-diffusion transport model [16] calibrated to ballistic Monte Carlo simulations, the Philips unified model for carrier mobility, bandgap narrowing model, density gradient quantization model, and nonlocal-path trap-assisted tunneling model [17]. The fin sidewall surfaces (along which the transistor current flows) are assumed to be along $\{110\}$ crystallographic planes, with transistor current flow in a $<110>$ direction. To boost transistor on-state current, 2 GPa (tensile) uniaxial stress is induced in the fin channel region for n-channel devices, whereas -2 GPa (compressive) uniaxial stress is induced in the fin channel region for p-channel devices. Fig. 2.1 shows cross-sectional views of the n-channel SSR FinFET. The fin corner is rounded (with 1 nm radius of curvature) for reduced gate leakage and enhanced gate control [18].

Drift-diffusion Transport Model The drift-diffusion model [16] describes carrier transport considering spatial variations in the electrostatic potential, the electron affinity χ, the band gap energy, carrier concentrations, and carrier effective masses m_N and m_p [19]. The electron and hole current densities are given by:

$$\overrightarrow{J_n} = \mu_n(n\nabla E_C - 1.5nkT\nabla \ln m_n) + D_n(\nabla n - n\nabla \ln \gamma_n) \tag{2.1}$$

$$\overrightarrow{J_p} = \mu_p(p\nabla E_v + 1.5pkT\nabla \ln m_p) - D_p(\nabla p - p\nabla \ln \gamma_p) \tag{2.2}$$

where μ_n and μ_p are electron mobility and hole mobility, respectively, D_n and D_p are electron diffusivity and hole diffusivity, respectively. γ_n and γ_p are given by Fermi statistics:

$$n = \gamma_n N_C \exp(\frac{E_{F,n} - E_C}{kT}) \tag{2.3}$$

$$p = \gamma_p N_V \exp(\frac{E_V - E_{F,p}}{kT}) \tag{2.4}$$

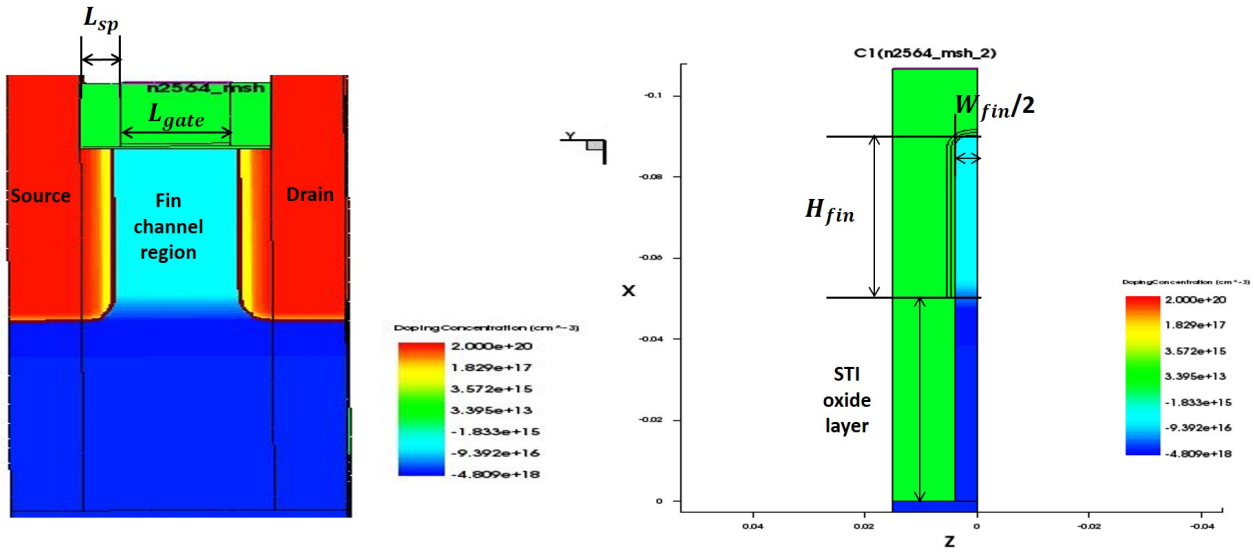

Figure 2.1: Cross-sectional views of the n-channel SSR FinFET structure. The net dopant concentration is represented in color using a hyperbolic arcsine scale. The fin aspect ratio is $\frac{H_{\text{fin}}}{W_{\text{fin}}} = 5$; the fin shape is rectangular as in Intel's 14 nm FinFET technology [10]; the fin corner radius of curvature is 1 nm.

For Boltzmann statistics, $\gamma_n = 1$ and $\gamma_p = 1$.

Philips Unified Mobility Model The Philips unified mobility model proposed by Klassen[20] unifies the descriptions of majority carrier and minority carrier mobilities. This model considers the different scattering mechanisms of lattice scattering, donor scattering, acceptor scattering, and electron-hole pair scattering. (Electron-electron scattering and hole-hole scattering are neglected because they have secondary-order effect.) Temperature dependence of mobility, electron-hole pair scattering, screening of ionized impurity atoms by charge carriers, and clustering of impurity atoms at high dopant concentrations also are taken into account.

Density-gradient Quantization Model For the density gradient model described in [21] [22], a generalized diffusion-drift transport equation was derived by making the electron gas equation of state density-gradient dependent, to describe quantum-mechanical behavior in the strong inversion regime.

Nonlocal Path Trap-assisted Tunneling Model The nonlocal path trap-assisted tunneling model uses the Wentzel-Kramers-Brillouin (WKB) approximation to calculate the transmission coefficient using the tunneling energy barrier together with Kane's model parameters A and B from [17].

2.2.2 Device Design Optimization

The effective channel length (L_{eff}) and the peak location of the punch-through stopper (PTS) doping profile ($X_{\text{fin,peak}}$) for the SSR FinFETs are separately optimized to maximize the on-state drive current $I_{\text{ds,sat}}$ while meeting the same off-state current specification ($I_{\text{off}} = 100\text{pA}/\mu\text{m}$).

L_{eff} Optimization The electrical channel length L_{eff} is defined as the lateral distance between the points where the source/drain dopant concentration falls to 2×10^{19} cm^{-3}. Previous work [23] showed that L_{eff} could be tuned to adjust the tradeoff between series resistance and short channel effect. In practice, L_{eff} can be tuned by adjusting the gate-sidewall spacer length (L_{sp}) or the source/drain junction doping gradient. In this work, the source/drain junction doping gradient is steep (2 nm/dec) to provide for low parasitic source/drain series resistance. Therefore, L_{eff} is tuned by adjusting L_{sp} for each device structure to maximize I_{on}.

Fig. 2.2 shows how $I_{\text{ds,sat}}$ varies with L_{eff} for n-channel FinFETs and p-channel FinFETs. The optimal values of L_{eff} are 20 nm for the n-channel control FinFET and 21 nm for the n-channel SSR FinFET. The optimal value of L_{eff} is larger for the SSR FinFET because it has a lightly doped channel region and hence is more susceptible to the short-channel effect.

Because their SEG source/drain regions comprise SiGe which has a smaller band gap energy, p-channel FinFETs have larger gate-induced drain leakage (GIDL) due to band-to-band tunneling (BTBT), which can dominate off-state leakage current. Therefore, L_{eff} must be larger to adequately suppress GIDL. The optimal values of L_{eff} are 23 nm for the p-channel SSR FinFET and 24 nm for the p-channel control FinFET. (Shorter values of L_{eff} are not included in Fig. 2.2 because I_{off} exceeds the specification.) The optimal value of L_{eff} is larger for the control FinFET because it has a heavily doped channel region and hence is more susceptible to GIDL. The cross-sectional diagrams in Fig. 2.3 compare the BTBT rate distributions within the p-channel FinFETs in the off state. Direct and indirect band-to-band-tunneling rates are calculated using Kane's model [24][19].

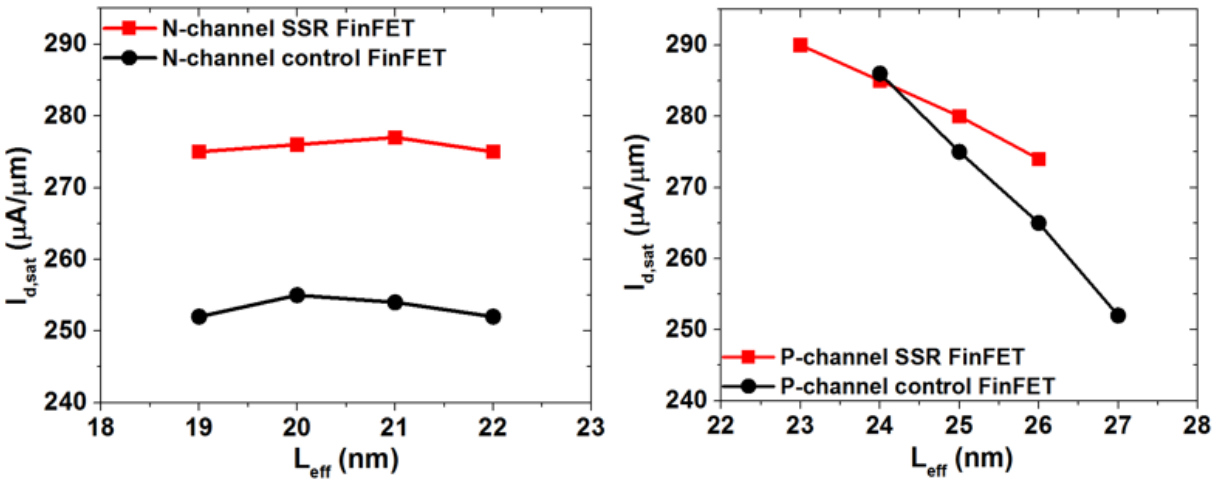

Figure 2.2: On-state drive current ($I_{\text{ds,sat}}$) normalized to W_{eff} versus electrical channel length (L_{eff}). Longer values of L_{eff} are required for the p-channel devices in order to suppress GIDL.

The BTBT generation rate per unit volume G is given by:

$$G = A \cdot (\frac{F}{F_0})^P \cdot exp(-\frac{B}{F}) \tag{2.5}$$

where the values of A and B are taken from [17]; F_0 is 1 V/cm; $P = 2$ & 2.5 for direct & indirect band-to-band-tunneling, respectively; F is the electric field and B_{dir} & B_{ind} are each proportional to $E_g^{3/2}$.

Fig. 2.4 shows the net dopant concentration profiles along the channel direction, from the source region to the drain region, for each of the optimized FinFET designs. The p-channel control FinFET has the longest L_{eff} and therefore the smallest amount of drain-induced barrier lowering (DIBL).

Punch-through Stopper (PTS) Doping Optimization Previous work [7] showed that partial monolayers of oxygen within silicon cause boron atoms to pile up, effectively blocking boron diffusion; thus, the insertion of partial monolayers of oxygen at a depth corresponding to the base of the silicon fin would facilitate the formation of a super-steep retrograde (SSR) fin doping profile. Fig. 2.5 compares the optimized fin doping profiles for control FinFET and SSR FinFET devices. The retrograde doping gradient for the n-channel SSR FinFET is 3.3 nm/dec, whereas for the p-

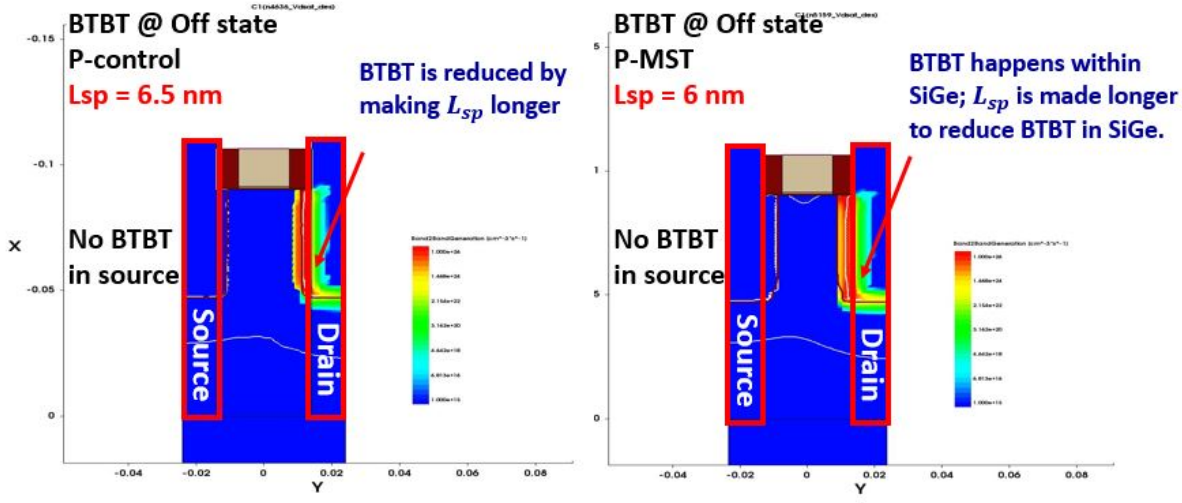

Figure 2.3: Band-to-band-tunneling rate contour plots for p-channel FinFETs in the off state ($V_{GS} = 0V, V_{DS} = V_{DD}$).

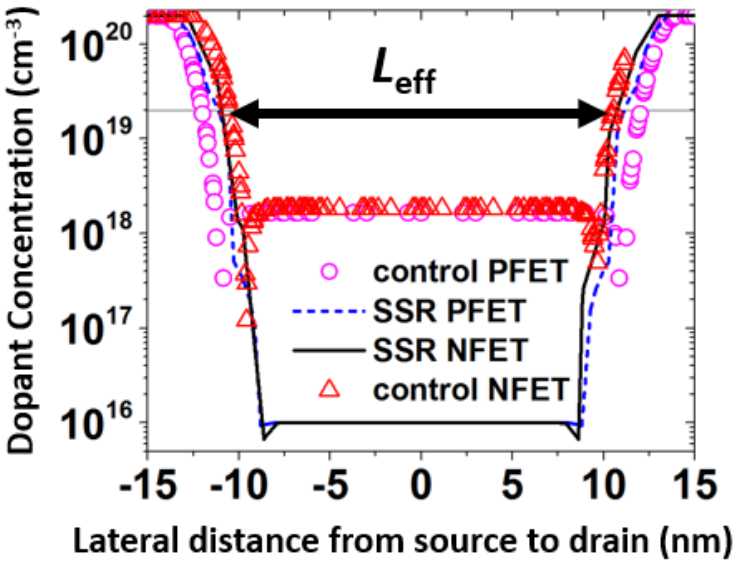

Figure 2.4: Net dopant concentration profiles along the channel direction, from the source region to the drain region, for the optimized FinFET designs.

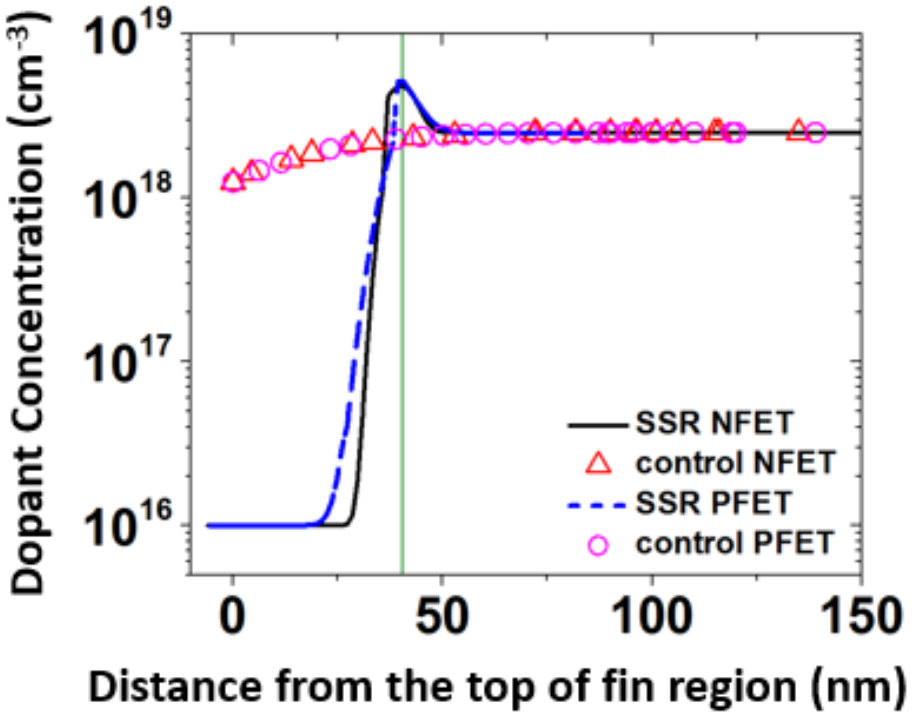

Figure 2.5: Optimized fin doping depth profiles for control FinFETs and SSR FinFETs.

channel SSR FinFET it is 6.9 nm/dec. This is because the dopant diffusion blocking effect is greater for boron than for n-type dopants [7]. Both the peak dopant concentration ($N_{\text{fin,peak}}$) and its depth location ($X_{\text{fin,peak}}$) were separately optimized for n-channel SSR and p-channel SSR FinFETs to maximize $I_{\text{ds,sat}}$: ($N_{\text{fin,peak}}$ tunes the tradeoff between GIDL and punch-through current, while $X_{\text{fin,peak}}$ tunes the tradeoff between effective fin height and punch-through stopper effectiveness.) Fig. 2.5 shows the optimized fin channel doping profiles. The optimal value of $X_{\text{fin,peak}}$ is 46 nm and the optimal value of $N_{\text{fin,peak}}$ is 5×10^{18} cm^{-3} for both n-channel and p-channel SSR FinFETs.

Nominal Device Performance Table 2.1 summarizes key device performance parameters for the optimized FinFET designs, obtained from 3-D device simulations. Threshold voltages $V_{\text{t,sat}}$, $V_{\text{t,lin}}$ are defined as the gate-source voltage V_{gs} corresponding to a constant current $I_{\text{sub}} = 100\text{nA} \times \frac{W_{\text{eff}}}{L_{\text{gate}}}$ at drain-source voltages V_{ds} 0.80 V, 50 mV. For a maximum operating voltage V_{DD} = 0.8 V, oxygen insertion technology provides for 9% and 2% improvement in $I_{\text{d,sat}}$ for n-channel and p-channel FinFETs, respectively. The benefit of higher carrier mobility is greater for operation in the linear

regime ($V_{gs} = 0.8\ V$, and $V_{ds} = 50$ mV): oxygen insertion technology provides for 10% and 5% improvement in $I_{d,lin}$ for n-channel and p-channel FinFETs, respectively. The p-channel FinFETs have higher drive current in comparison with the n-channel FinFETs because the fins have {110} sidewall surfaces and uniaxial stress is more effective for boosting hole mobility [25].

Fig. 2.6 shows the simulated FinFET transfer characteristics (drain current I_{ds} on a logarithmic scale as a function of gate-to-source voltage V_{gs}). Note that GIDL current is lower for SSR FinFETs as compared with control FinFETs.

	control FinFET		SSR FinFET	
	N-channel	P-channel	N-channel	P-channel
V_{dd} (V)	0.8	0.8	0.8	0.8
$V_{ds,lin}$ (V)	0.05	0.05	0.05	0.05
L_{eff} (nm)	21	24	22	23
$V_{t,sat}$ (V)	0.384	-0.374	0.370	-0.373
$V_{t,lin}$ (V)	0.435	-0.412	0.412	-0.409
$I_{d,sat}$ (μA/μm)	255	286	277	291
$I_{d,lin}$ (μA/μm)	62	75	68	79
SSwing (mV/dec)	73	71	72	70
DIBL (mV/V)	69	52	57	48

Table 2.1: Summary of key performance parameters for the optimized FinFET designs.

2.3 Compact Model Calibration

In this study a compact (analytical) model for transistor current as a function of applied voltages is used to estimate six-transistor (6T) SRAM cell performance and yield, following the methodology established in [26]. The model, which is presented in detail in the Appendix, is based on the I-V equations for a short-channel MOSFET, which account for channel length modulation (CLM), velocity saturation, and bulk charge effects. Since the compact model does not account for GIDL,

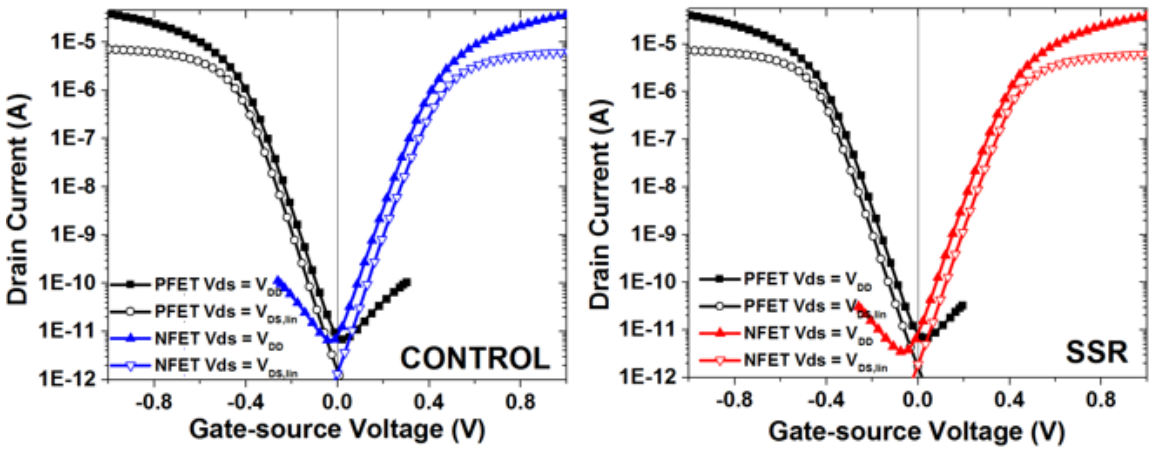

Figure 2.6: Simulated transfer characteristics for the optimized FinFET designs.

3-D device simulations without GIDL are used to calibrate the compact model, as shown in Fig. 2.7 and Fig. 2.8.

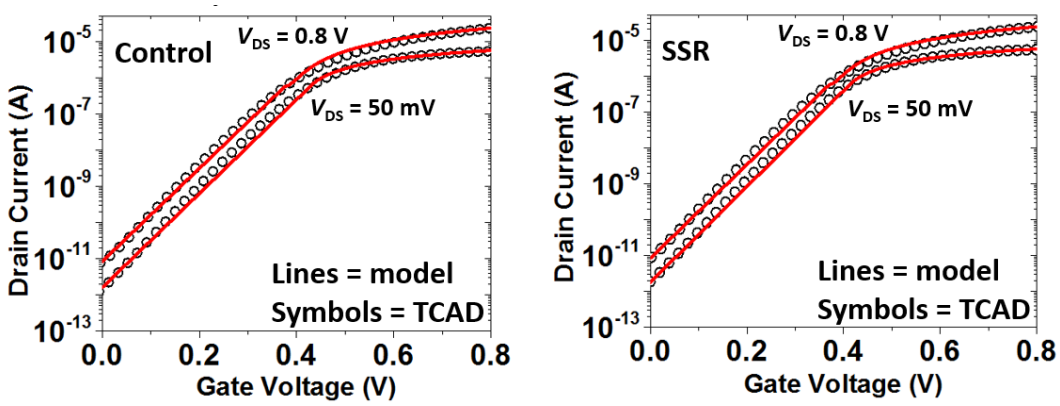

Figure 2.7: Comparison of the calibrated compact model (lines) and simulated I-V characteristics (symbols) for n-channel FinFETs.

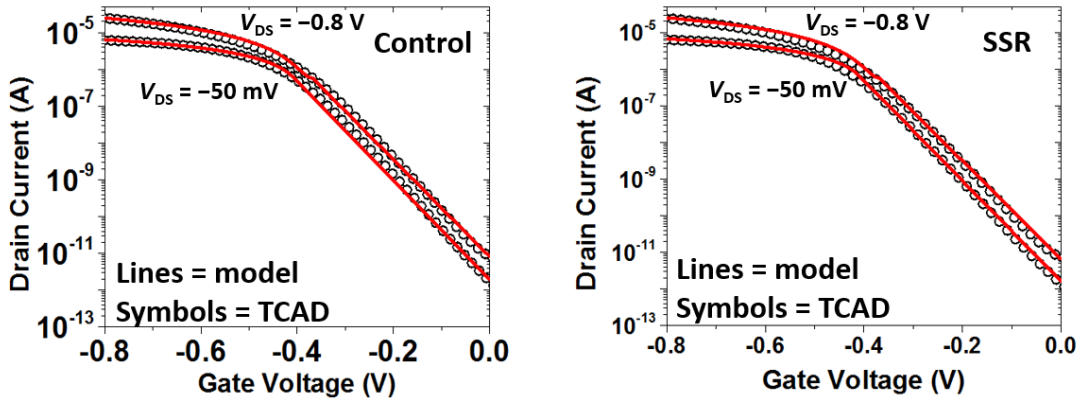

Figure 2.8: Comparison of the calibrated compact model (lines) and simulated I-V characteristics (symbols) for p-channel FinFETs.

Chapter 3

Bulk-Si FinFET Variability Study and 6T-SRAM Yield Estimation

3.1 Introduction

In this chapter, the benefit of SSR fin channel doping for reducing the sensitivity of FinFET performance and threshold voltage (V_t) on process-induced variations is first investigated. The effects of systematic variations on transistor threshold voltage, on-state current and off-state leakage are shown to be accurately predicted by the compact model. The impact of intrinsic variations such as random dopant fluctuation, gate work function variation, and gate line-edge-roughness (LER) [27] [31] are considered and quantified using the noise-like impedance field method (IFM) in TCAD device simulations. Finally, an analysis of six-transistor (6T) SRAM cell performance and estimation of 6T-SRAM cell yield as a function of cell operating voltage is performed using the calibrated compact model, to quantify the benefit of oxygen insertion technology for facilitating voltage scaling.

3.2 Variability Study

Sources of variation can be categorized as either systematic, caused by process variations, or random, caused by intrinsic variations [28].

3.2.1 Systematic Sources of Variation

Variations in physical device dimensions are caused by variations in the fabrication process, and systematically result in variations in transistor performance. In this study, process-induced variations in L_{gate}, W_{Si} and EOT are assumed to have Gaussian distributions with $\pm 10\%$ variation corresponding to three standard deviations away from the mean (nominal) value.

Fig. 3.1 show the dependences of transistor threshold voltage (V_{t}) and off-state leakage current (I_{off}), respectively, on the physical gate length (L_{gate}). The calibrated compact model can be seen to accurately predict these systematic variations. The saturation threshold voltage ($V_{\text{t,sat}}$) and linear threshold voltage ($V_{\text{t,lin}}$) each decrease with decreasing L_{gate} due to the short-channel effect, and I_{off} correspondingly increases with decreasing L_{gate}. SSR FinFETs show slightly greater sensitivity of I_{off} to changes in L_{gate} since $\log I_{\text{off}} \propto \frac{-V_t}{\text{SSwing}}$ and SSR FinFETs have steeper (smaller) subthreshold swing (SSwing).

Fig. 3.2 show the dependences of V_{t} and I_{off}, respectively, on the physical fin width (W_{fin}). The calibrated compact model can be seen to accurately predict these systematic variations. Electrostatic integrity (*i.e.* gate control) improves with decreasing W_{fin}, so that I_{off} decreases with decreasing W_{fin}. However, the series resistance of the fin source/drain regions increases with decreasing W_{fin}, resulting in degraded I_{on}.

Fig. 3.3 show the dependences of V_{t} and I_{off}, respectively, on the equivalent oxide thickness (EOT) of the gate dielectric stack. The calibrated compact model can be seen to accurately predict these systematic variations. Decreasing EOT increases capacitive coupling between the gate and the channel region for improved electrostatic integrity and hence lower I_{off}. Note that the impact of EOT variation on V_{t} and I_{off} is negligible compared to the impact of L_{gate} variation and the impact of W_{fin} variation.

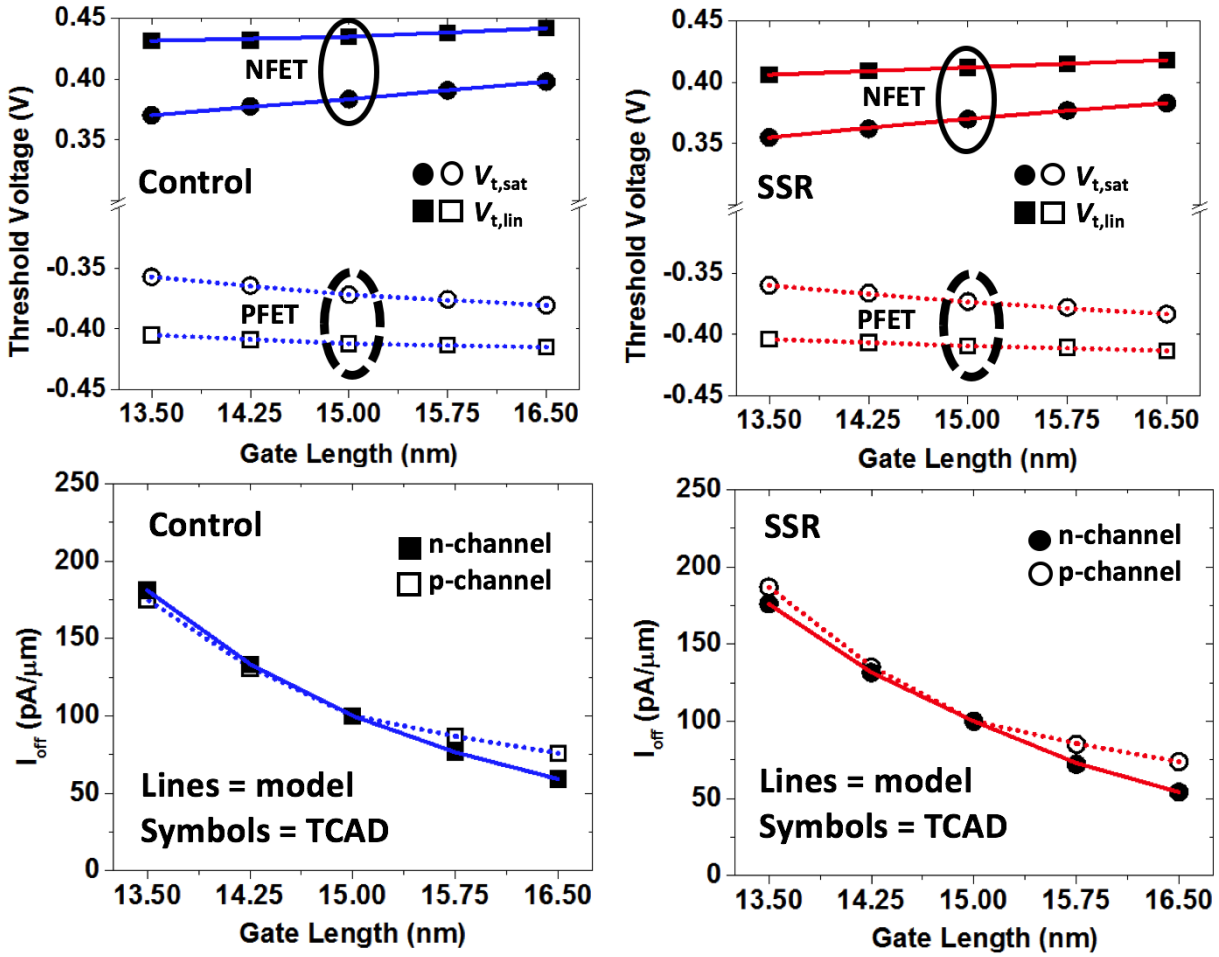

Figure 3.1: FinFET threshold voltage (V_t) and off-state current $vs.$ gate length, from TCAD 3-D device simulations (symbols) and from the calibrated compact model (lines). The solid lines correspond to n-channel devices, while the dotted lines correspond to p-channel devices.

3.2.2 Random Sources of Variation

Random sources of variation become dominant as transistors are scaled down toward atomic dimensions, and can limit IC manufacturing yield [29]. These sources include random dopant fluctuations (RDF) and gate work function variation (WFV) [27] [31]. In particular, WFV has been identified as the dominant contributor to $V_{\rm t}$ variation for FinFET technology[29]. Depending on the average dopant concentration, variations in $V_{\rm t}$ and $I_{\rm off}$ due to RDF can become significant as the volume of the fin channel region shrinks [32]. Besides WFV and RDF, gate line-edge-roughness (LER) due to stochastic variations in lithography and etch processes can affect $V_{\rm t}$ due to the short-

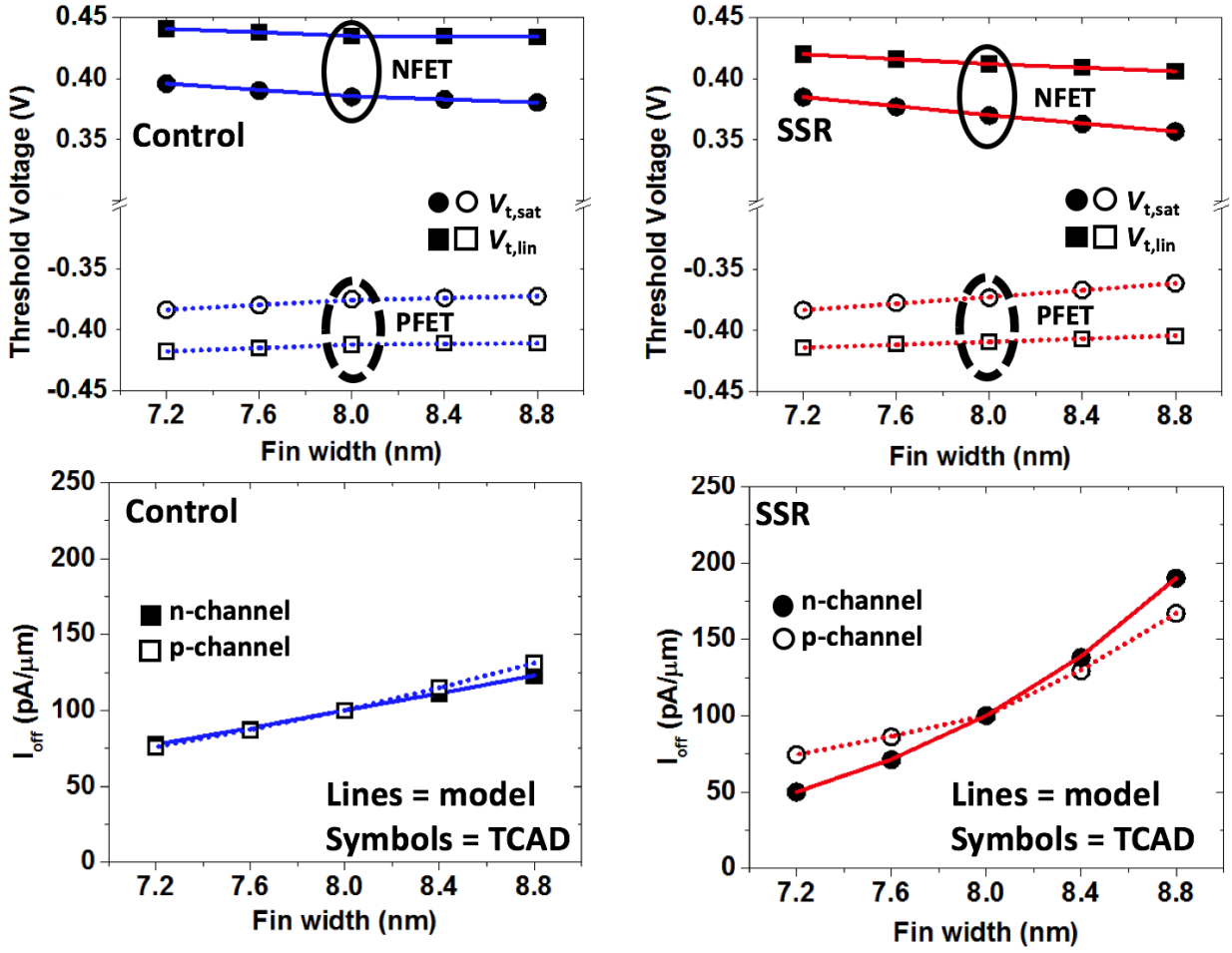

Figure 3.2: FinFET threshold voltage (V_t) and off-state current $vs.$ fin width, from TCAD 3-D device simulations (symbols) and from the calibrated compact model (lines). The solid lines correspond to n-channel devices, while the dotted lines correspond to p-channel devices.

channel effect [33]. The use of spacer lithography (also known as self-aligned double patterning, SADP) [34] to define nanometer-scale critical dimensions (gate length and fin width) is becoming prevalent, however, so that LER is not expected to be a significant source of random variability in FinFET performance. (In a SADP process, the critical dimension is defined by the thickness of a deposited film, which is locally very uniform.) Also, a previous 3-D device simulation study showed that, among the random sources of variation, LER is negligible compared with WFV and RDF for FinFET technology [35].

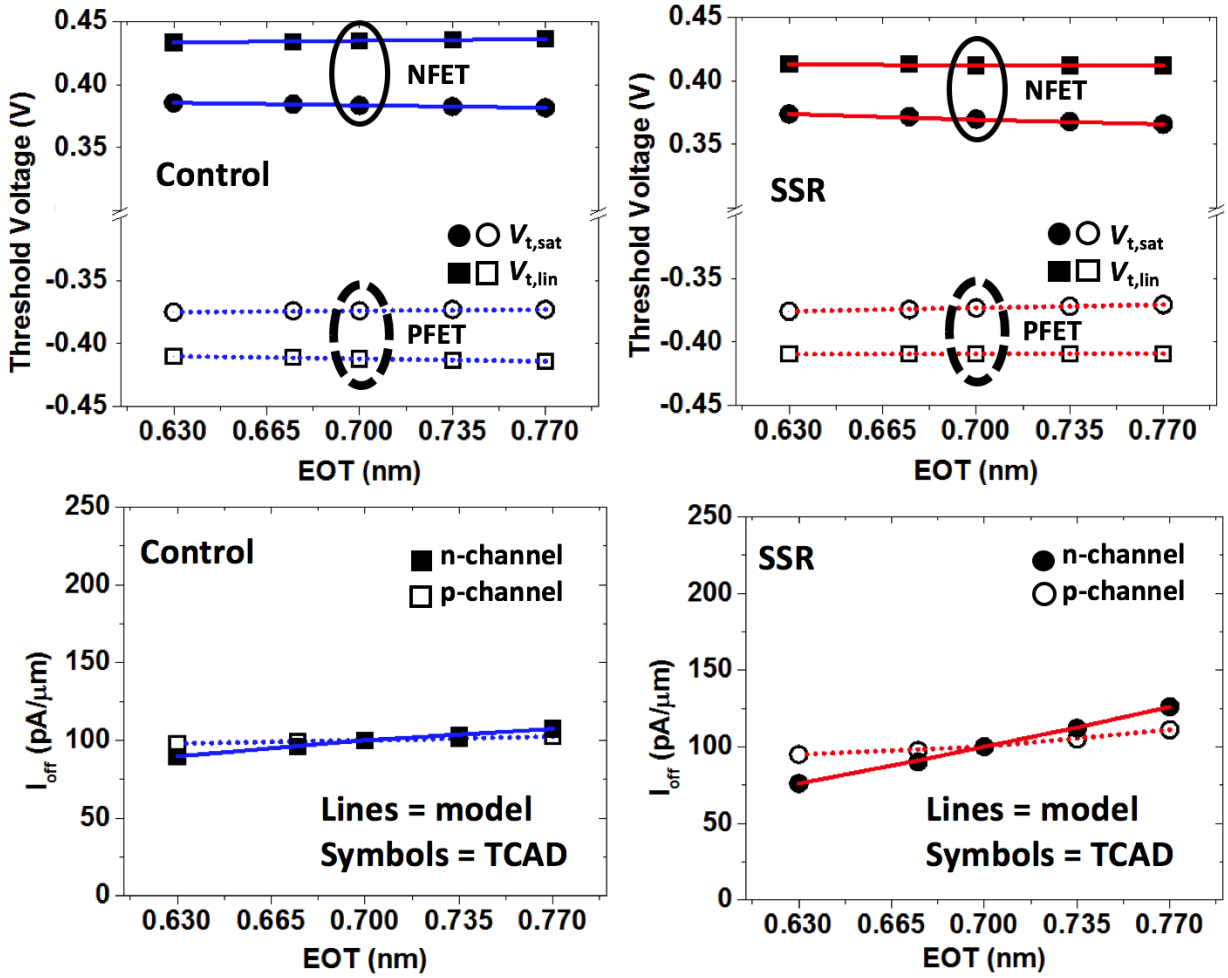

Figure 3.3: FinFET threshold voltage (V_t) and off-state current $vs.$ equivalent oxide thickness, from TCAD 3-D device simulations (symbols) and from the calibrated compact model (lines). The solid lines correspond to n-channel devices, while the dotted lines correspond to p-channel devices.

Random Dopant Fluctuations (RDF) Random dopant fluctuations have been shown to result in significant threshold voltage (V_t) variability for planar bulk-silicon MOSFETs with gate lengths smaller than 0.1μm, due to the very small volume of the depletion region in this regime resulting in a relatively small number of dopant atoms which determine $V_{\rm t}$ [32]. The standard deviation of variation is proportional to $\sqrt{N_a/W \times L}$, where N_a is the average dopant concentration in the depletion region, W is the channel width, and L is the channel length. In this study, RDF-induced variability is determined using the noise-like impedance field method [19], [36]. The results summarized in Table 3.1 show that SSR FinFETs have good immunity to RDF, since they have relatively light dopant concentration within the (fully depleted) fin channel region so that the

depletion charge negligibly affects the threshold voltage.

Work Function Variation (WFV) The equivalent oxide thickness (EOT) of the gate dielectric layer(s) must be reduced as the transistor gate length is scaled down, in order to maintain good gate control and thereby suppress V_{t} variability due to SCE and DIBL. To avoid excessive gate leakage due to direct tunneling through an ultra-thin dielectric, a high-permittivity (high-k) dielectric material should be used to achieve EOT below 1 nm. Due to Fermi-level pinning at the interface of doped polycrystalline-silicon (poly-Si) and high-k material [37], which undesirably affects the effective work function of the poly-Si, and remote soft optical phonon scattering which degrades inversion-layer mobility [38], a metal gate material must be used in conjunction with high-k gate dielectric material. High-k/metal gate stacks have been used in high-volume CMOS production since the introduction of Intel's 45 nm technology [39].

The work function (WF, in eV) is defined as the minimum energy required to remove an electron from the solid material, which is the sum of the bulk chemical potential (due to electron-electron correlation and exchange effects) and surface dipole potential. The bulk chemical potential is a fixed material property, whereas the surface dipole potential is dependent on crystalline orientation. The stochastic nature of the metal layer deposition process results in local variations in the WF of a polycrystalline metal gate electrode.

Assuming that the random sources of variation (WFV and RDF) are independent, threshold-voltage variability due to all random sources is:

$$\sigma V_{\mathrm{t,total}} = \sqrt{(\sigma V_{\mathrm{t,WFV}})^2 + (\sigma V_{\mathrm{t,RDF}})^2} \tag{3.1}$$

Table 3.1 summarizes FinFET off-state leakage variability and threshold voltage (V_t) variability due to random sources, obtained via TCAD simulations. Note that WFV has a dominant effect. These results indicate that SSR FinFETs are less susceptible to random sources of variation and therefore are promising for achieving higher IC manufacturing yield.

		control FinFET		SSR FinFET	
		N-channel	P-channel	N-channel	P-channel
σI_{off} (μA/μm)	WFV only	43.12	36.95	50.12	38.00
	RDF only	27.16	23.78	7.37	6.42
	WFV & RDF	51.28	43.94	50.12	38.58
$\sigma V_{\text{t,sat}}$ (mV)	WFV only	23.2	23.4	23.7	22.9
	RDF only	16	13.4	21	25
	WFV & RDF	28.2	27	23.8	23

Table 3.1: Variability in FinFET off-state current and saturation threshold voltage (V_t) due to random sources of variation.

3.3 6T-SRAM Cell Yield Estimation

In this section, the architecture and operation of a 6T-SRAM cell is reviewed to introduce two key SRAM performance metrics, the read static noise margin (SNM) and the writability current I_W. The calibrated compact model [26] is used in lieu of mixed-mode TCAD 3-D device simulation (which is computationally expensive) to accurately calculate these metrics. Then the sensitivities of SNM and I_W to variations in device parameter X_I, $\frac{\partial \text{SNM}}{\partial X_I}$ and $\frac{\partial I_W}{\partial X_I}$, are calculated to determine the minimum variability that results in cell operation failure ("cell sigma") [26] as a function of the cell operating voltage, to find the minimum cell operating voltage ($V_{\text{DD,min}}$).

3.3.1 SRAM Cell Architectures

An SRAM array comprises many cells arranged in rows and columns, with each cell storing one bit of information. The most common cell architecture (illustrated by the circuit diagram in Fig. 3.4) utilizes six transistors arranged as two cross-coupled inverters, each comprising one p-channel "pull-up" (PU) transistor with its source tied to V_{DD} and one n-channel "pull-down" (PD) transistor with its source tied to ground, and two n-channel "pass-gate" transistors used to connect the left and right internal storage nodes **CH** and **CL** (which store complementary voltage signals) to the left and right bit lines **BL** and $\overline{\textbf{BL}}$ (shared by every cell in the same column of the SRAM array),

respectively. The gates of the pass-gate (PG) transistors are connected to a single wordline (shared by every cell in the same row of the SRAM array), **WL**.

Alternative SRAM cell architectures are shown in Fig. 3.5. The eight-transistor cell design provides for larger operating voltage margins and hence higher cell yield, but at the cost of larger layout area (lower storage density) [40]. The four-transistor cell design is much more compact but has higher static power dissipation [41].

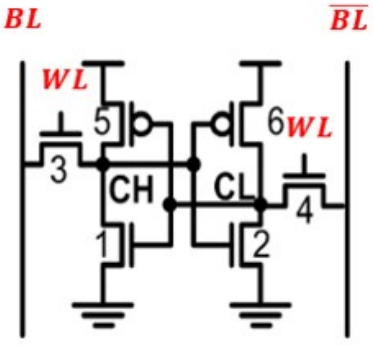

Figure 3.4: Circuit diagram of the 6T-SRAM cell design, which comprises two cross-coupled inverters and two n-channel pass-gate devices from [26].

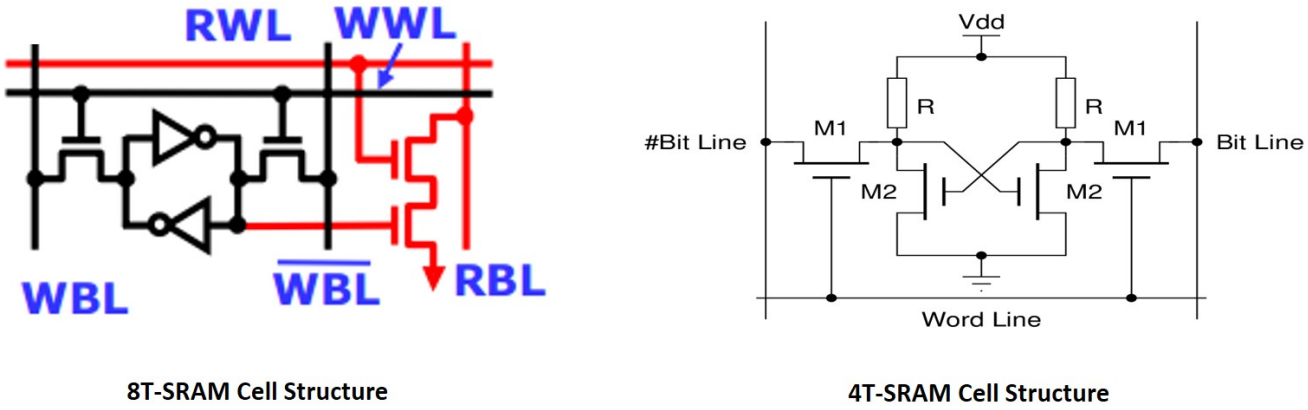

Figure 3.5: Circuit diagrams for alternative SRAM cell designs from [40], [41].

3.3.2 6T-SRAM Cell Operation

Read Operation To read out the information stored in a 6T-SRAM cell, both bit lines (**BL** and $\overline{\textbf{BL}}$) are first "pre-charged" to a high voltage, V_{DD}. Then the word line (**WL**) is pulsed with a high

voltage to turn on the PG transistors and thereby connect the bit lines to the internal storage nodes, so that the bit line connected to the internal storage node with a low voltage (*i.e.* storing a logic "0") is discharged through its corresponding PG transistor and PD transistor, as illustrated in Fig. 3.6. The process of discharging that bit line (so that its voltage reaches ground potential) is sped up by a positive feedback effect of the cross-coupled inverter as well as a sense amplifier connected to the bit lines at the periphery of the cell array.

During a read operation, the voltage at the internal storage node (**CH** in Fig. 3.6) rises due to the voltage divider effect (*i.e.* current flowing through the PG transistor in series with the PD transistor results in a non-zero voltage drop across the PD transistor). A Read disturb error occurs if this voltage rises above the tipping point of the opposite inverter, so that the positive feedback effect causes the storage node to flip states (*i.e.* to a logic "1"). For good read stability, the PD transistor should be stronger (*i.e.* have less on-state resistance) than the PG transistor. The cell **beta ratio** is defined as the ratio of PD transistor drive (on-state) current to PG transistor drive current. For planar MOSFET technology, this ratio can be finely tuned by adjusting the drawn channel widths of the PD and PG transistors. For FinFET technology, it can only be practically tuned coarsely by adjusting the number of fins (connected in parallel between the source and drain regions) in each device. Better writability is achieved by using more fins for the PD devices than for the PG devices.

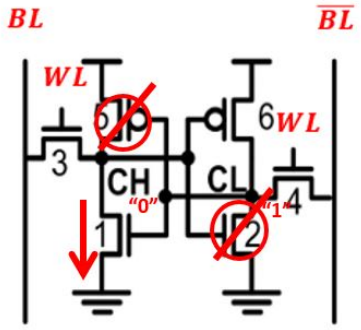

Figure 3.6: Illustration of current flowing in a 6T-SRAM cell during a Read operation. "0" is stored at the internal storage node **CH** so that the bit line **BL** discharges through PG transistor 3 and PD transistor 1. If the voltage at **CH** rises above the tipping point of the opposite inverter (so that PG transistor 2 turns on), it can flip to the "1" state erroneously from [26].

Write Operation To write information from the bit lines (one at low voltage corresponding to logic "0" and the other at high voltage corresponding to logic "1") into a 6T-SRAM cell, the word line (**WL**) is pulsed with a high voltage to turn on the PG transistors and thereby connect the bit lines to the internal storage nodes, so that the bit line at low voltage will discharge the internal storage node through its corresponding PG transistor, as illustrated in Fig. 3.7. This PG transistor must be stronger than the PU transistor which acts to maintain a high voltage on the internal storage node. Once the tipping point of the opposite inverter is reached, the process of discharging the internal storage node (so that its voltage reaches ground potential) is sped up by a positive feedback effect of the cross-coupled inverter.

For immunity against write failure, the PG transistor should be stronger (*i.e.* have less on-state resistance) than the PU transistor. The cell **gamma ratio** is defined as the ratio of PG transistor drive (on-state) current to PU transistor drive current. For planar MOSFET technology, this ratio can be finely tuned by adjusting the drawn channel widths of the PG and PU transistors. For FinFET technology, it can only be practically tuned coarsely by adjusting the number of fins (connected in parallel between the source and drain regions) in each device. Better immunity to read disturbance is achieved by using more fins for the PG devices than for the PU devices.

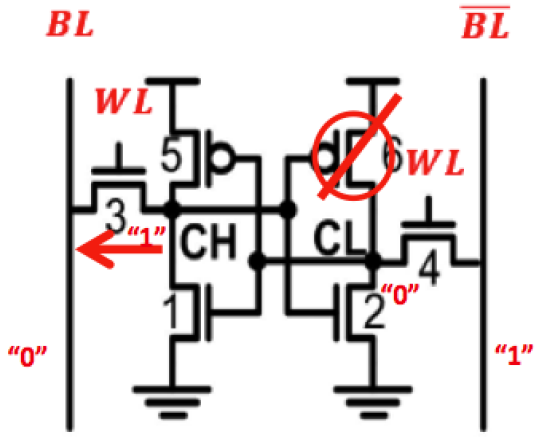

Figure 3.7: Illustration of current flowing in a 6T-SRAM cell during a Write operation. The internal node **CH** is initially at a high voltage (storing a "1") and must be discharged through PG transistor 3; this is resisted by the PU transistor 5. A write failure occurs if the PU transistor is stronger than the PG transistor from [26].

3.3.3 Read Static Noise Margin (SNM)

The Read static noise margin (SNM) is a quantitative measure of the robustness of a SRAM cell to read disturb error. It is the minimum amount of noise (change in voltage) required to change the state of the cell, and is extracted from a composite "butterfly plot" of the voltage transfer curves (VTCs, $V_{\mathrm{CH}}\ vs.\ V_{\mathrm{CL}}$ and $V_{\mathrm{CL}}\ vs.\ V_{\mathrm{CH}}$) for the cross-coupled inverters during a Read operation, *i.e.* with both PG transistors turned on and both bit lines pre-charged to V_{DD}. As illustrated in Fig. 3.8, the Read SNM corresponds to the length of the largest square that fits within the smaller "lobe" of the butterfly plot.

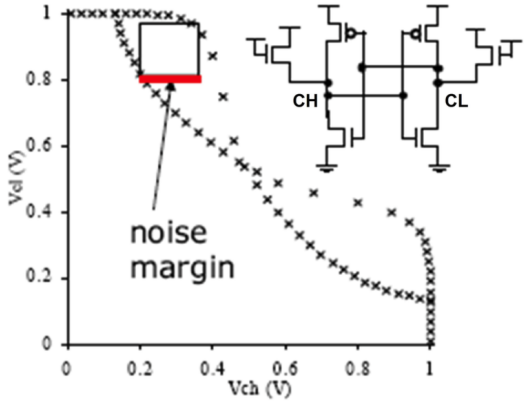

Figure 3.8: The Read Static Noise Margin (SNM) of a 6T-SRAM cell is determined from a butterfly plot comprising the voltage transfer curves for each of the cross-coupled inverters from [26].

3.3.4 Write-ability Current (I_W)

The writeability current was proposed by C. Wann of IBM Corp. [3] as a quantitative measure of the robustness of a SRAM cell to write error. It is the minimum amount of current flowing out of the internal storage node as it is discharged from V_{DD} toward ground potential, as illustrated in Fig. 3.9.

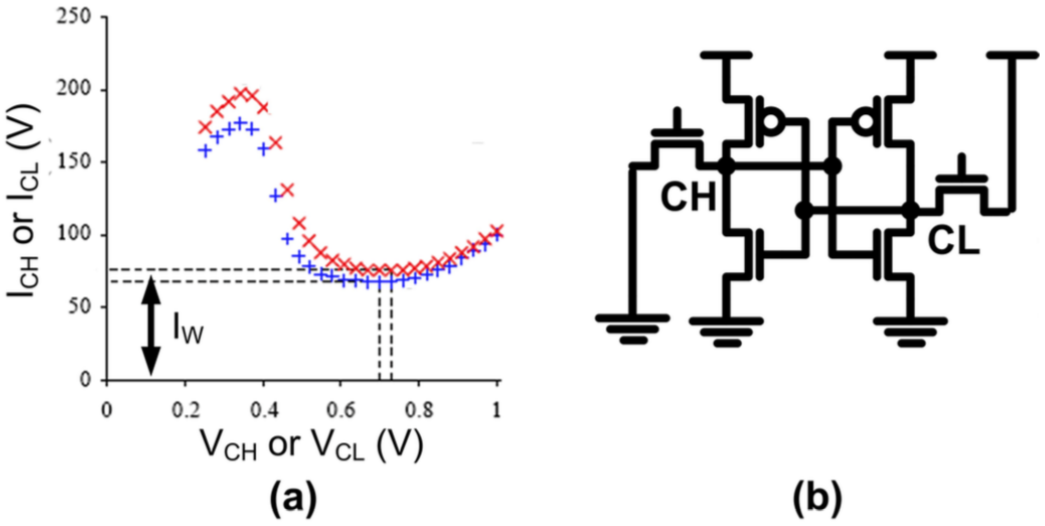

Figure 3.9: The writeability current of a 6T-SRAM cell corresponds to the local minimum of either of the "write-N" curves I_{CH} $vs.$ V_{CH} and I_{CL} $vs.$ V_{CL}. These curves are generated by sweeping V_{CH} or V_{CL} and measuring the nodal current at **CH** or **CL**, respectively, during a write operation from [26].

3.3.5 Compact Modeling of Read SNM & I_W

Methodology In this study, the calibrated compact model [26] is used to calculate 6T-SRAM cell performance metrics and to determine the sensitivities of these metrics to variations in device design parameter values. Read SNM is obtained from the inverter VTCs during a Read operation, which are obtained by applying Kirchhoff's Current Law, for example:

$$\begin{aligned} I_{\text{D1}}(V_{\text{GS}} = V_{\text{CL}}, V_{\text{DS}} = V_{\text{CH}}) = I_{\text{D3}}(V_{\text{GS}} = V_{\text{WL}} - V_{\text{CH}}, V_{\text{DS}} = V_{\text{BL}} - V_{\text{CH}}) \\ + I_{\text{D5}}(V_{\text{GS}} = V_{\text{CL}} - V_{\text{DD}}, V_{\text{DS}} = V_{\text{CH}} - V_{\text{DD}}) \end{aligned} \tag{3.2}$$

where I_{Dx} is the drain current through device x. During a read operation, V_{WL} is biased at V_{DD}. Note that Eqn. 3.2 assumes that there is no current other than between the source and drain regions of a transistor in the 6T-SRAM cell, *i.e.* gate and junction leakage currents are assumed to be negligible.

I_W is determined by calculating the I_{CH} $vs.$ V_{CH} curve and I_{CL} $vs.$ V_{CL} curve corresponding to

write "0" and write "1" operations, respectively. The current is found by iteration, for example:

$$I_{\mathrm{CH}} = I_{\mathrm{D3}}(V_{\mathrm{GS}} = V_{\mathrm{WL}}, V_{\mathrm{DS}} = V_{\mathrm{CH}}) - I_{\mathrm{D5}}(V_{\mathrm{GS}} = V_{\mathrm{CL}} - V_{\mathrm{DD}}, V_{\mathrm{DS}} = V_{\mathrm{CH}} - V_{\mathrm{DD}}) + I_{\mathrm{D1}}(V_{\mathrm{GS}} = V_{\mathrm{CL}}, V_{\mathrm{DS}} = V_{\mathrm{CH}}) \tag{3.3}$$

FinFET-based 6T-SRAM Cell Designs As alluded in the previous chapter, p-channel FinFETs have higher drive current in comparison with n-channel FinFETs because the fin sidewall surfaces and current flow direction are optimal for hole transport and because uniaxial stress is more effective for boosting hole mobility [10]. A consequence of this is that the number of fins in the PG transistors should be greater than that in the PU transistors to ensure good write-ability (**gamma ratio** greater than 1). In turn, the number of fins in the PD transistors should be greater than that in the PG transistors to ensure good read stability. Various 6T-SRAM cell designs have been investigated by the semiconductor industry, therefore, for 16 nm FinFET technology [42]. These are named according to the number of fins in the PU, PD and PG devices: 1-fin PU, 1-fin PD, 1-fin PG (1-1-1 cell); 1-fin PU, 2-fin PD, 1-fin PG (1-2-1 cell); 1-fin PU, 2-fin PD, 2-fin PG (1-2-2 cell); and 1-fin PU, 3-fin PD, 2-fin PU (1-3-2 cell). Table 3.2 summarizes the read SNM and write-ability current values for the various 6T-SRAM cell designs. It can be seen that the 1-2-1 cell has the best read stability while the 1-2-2 cell has the best write-ability.

	control FinFET				SSR FinFET			
Cell design	1-1-1	1-2-1	1-2-2	1-3-2	1-1-1	1-2-1	1-2-2	1-3-2
Read SNM (mV)	157	208	147	174	156	207	144	173
I_{w} (μA)	11.4	6.80	33.0	30.1	13.10	8.61	36.10	33.6

Table 3.2: Comparison of FinFET-based 6T-SRAM cell performance metrics.

6T-SRAM cells comprised of SSR FinFETs show significantly better write-ability in comparison with cells of the same design comprised of control FinFETs, because the **gamma ratio** for SSR FinFET technology is larger than that for the control FinFET technology. The read SNM for SSR FinFET technology is comparable to that for control FinFET technology, since the cell **beta ratio** is determined by the ratio of the number of fins in the PD transistor to the number of fins in the PG transistor, which is the same for both technologies.

3.3.6 6T-SRAM Cell Yield Modeling

Variability in transistor performance due to systematic and random sources of variation can result in a 6T-SRAM cell with Read SNM < 0 V or $I_W < 0$ which does not function properly, *i.e.* cell failure. Cell sigma [26] is defined as the minimum total number of standard deviations from the nominal value, for any combination of 18 device parameters (gate length, fin width, and threshold voltage for each of the 6 transistors in a 6T-SRAM cell), that causes a read disturb error or a write failure. By assuming that 3σ deviation from the mean (nominal) value corresponds to $\pm 10\%$ variation in L_{gate} and W_{fin}, and accounting for random V_{t} variations due to WFV and RDF, the cell sigma is modeled in a multi-dimensional variation space. In this variation space, each dimension corresponds to one device parameter, and the probability of occurence decreases with increasing deviation from the nominal value (*i.e.* variation). To illustrate the concept of cell sigma, Fig. 3.10 shows a simple example of modeling the read SNM in a 2-dimensional variation space. There is a region corresponding to combinations of PG1 threshold-voltage variation and PD1 threshold-voltage variation which result in read disturb error, referred to as the **surface of failure**; the read SNM cell sigma is the minimum distance from the origin to this **surface of failure**.

3.3.7 Estimation of 6T-SRAM Cell Minimum Operating Voltage

Lower operating voltage is beneficial for reducing power consumption, but can result in lower nominal values of read SNM and/or I_W and therefore higher probability of cell failure due to variations. The minimum cell operating voltage ($V_{\text{DD,min}}$) is defined as the lowest operating voltage for which the cell meets the six-sigma yield requirement (for SRAM arrays with greater than 256 Mb capacity) for both read SNM and I_W.

The methodology established by [26] is used herein to determine the read SNM and I_W cell sigma values, based on the sensitivities of these SRAM metrics to each device parameter X_I ($\frac{\partial \text{SNM}}{\partial X_I}$ and $\frac{\partial I_W}{\partial X_I}$). Table 3.3 summarizes $V_{\text{DD,min}}$ for the various 6T-SRAM cell designs. It can be seen that oxygen insertion technology facilitates voltage scaling to below 0.50 V, with the lowest $V_{\text{DD,min}}$ of 0.44 V for the 1-3-2 cell design. The 1-2-1 cells have the largest $V_{\text{DD,min}}$ due to relatively poor write-ability as shown in Fig. 3.14 and Fig. 3.15. Fig. 3.13 directly compares cell sigmas for

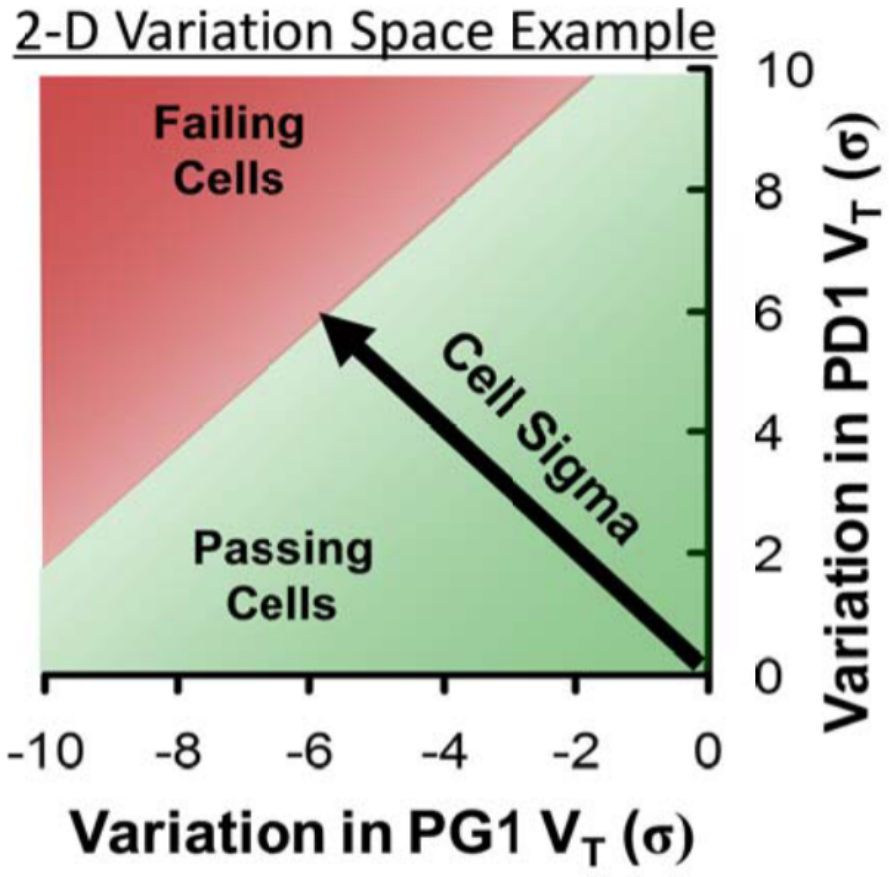

Figure 3.10: Illustration of read SNM cell sigma for a 2-dimensional variation space from [26].

SSR FinFET technology $vs.$ control FinFET technology, for the 1-3-2 cell design. For the control FinFET technology, $V_{\text{DD,min}}$ is limited by I_W cell sigma, which degrades faster with decreasing V_{DD} than does the read SNM cell sigma of SSR FinFET technology. These results indicate that SSR FinFET technology should be less susceptible to degradation in 6T-SRAM write-ability due to decreased **gamma ratio** [10].

These results indicate that SSR FinFET technology can provide for 60-70 mV reduction in 6T-SRAM cell operating voltage. It also reveals that the 1-2-1 cell design has the worst $V_{\text{DD,min}}$ despite having the best read SNM as shown in Table 3.2. This is because $V_{\text{DD,min}}$ is limited by I_W cell sigma since p-channel FinFETs are stronger than n-channel FinFETs, for the same number of fins. Both the 1-2-2 and 1-3-2 cell designs are projected to be able to scale to operating voltages below 0.5 V.

	control FinFET				SSR FinFET			
Cell design	1-1-1	1-2-1	1-2-2	1-3-2	1-1-1	1-2-1	1-2-2	1-3-2
$V_{DD,min}$ (V)	0.56	0.63	0.52	0.50	0.49	0.57	0.45	0.44

Table 3.3: Comparison of minimum operating voltage ($V_{DD,min}$) for various FinFET-based 6T-SRAM cell designs.

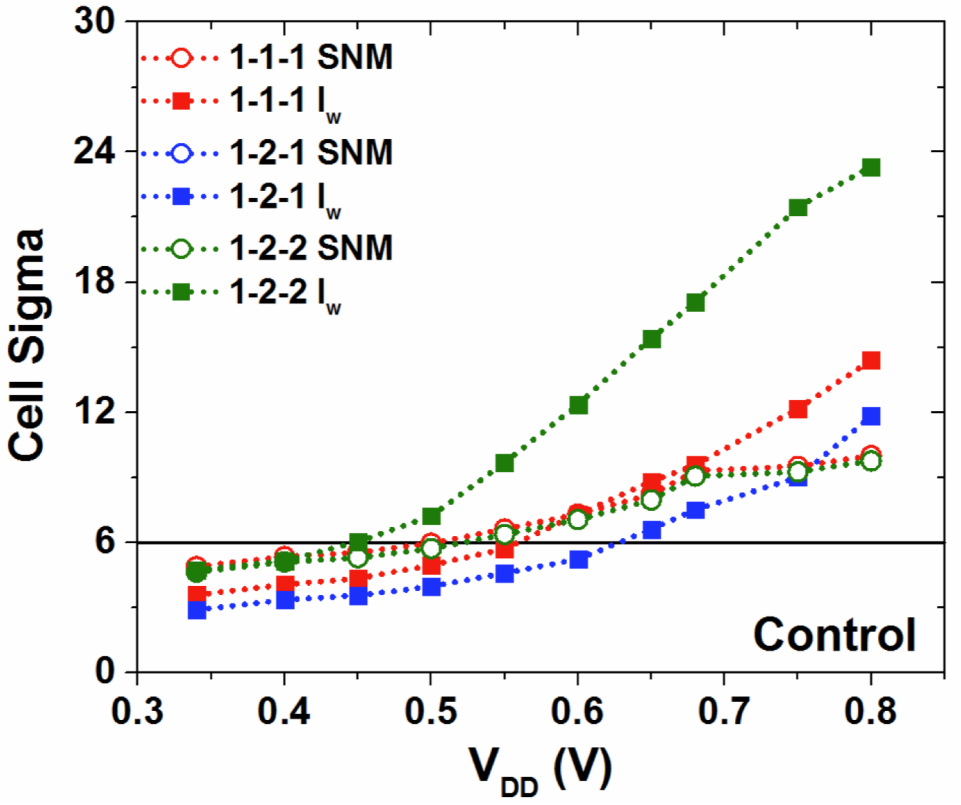

Figure 3.11: Read SNM (open circles) cell sigma and I_W (solid squares) $vs.$ V_{DD} for various SRAM cell designs implemented with control FinFET technology.

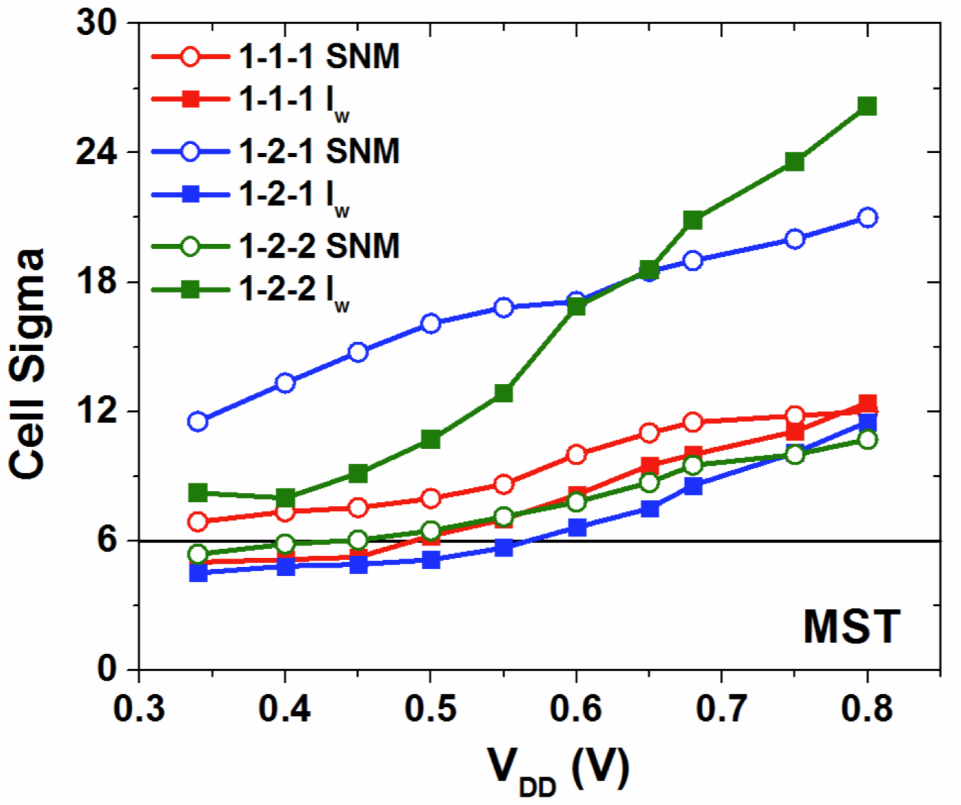

Figure 3.12: Read SNM (open circles) cell sigma and I_W (solid squares) $vs.$ V_{DD} for various SRAM cell designs implemented with SSR FinFET technology.

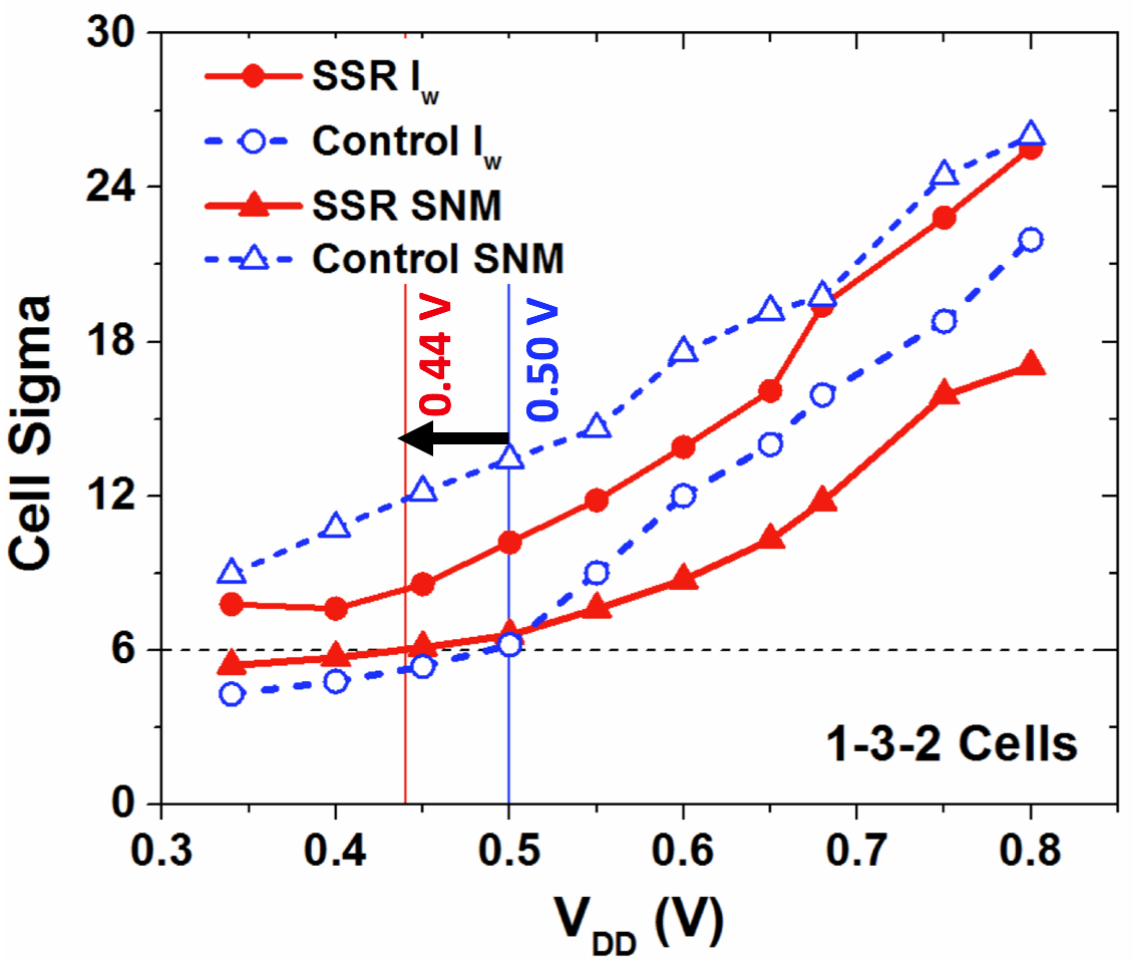

Figure 3.13: Read SNM (open circles) cell sigma and I_W (solid squares) $vs.$ V_{DD} for the 1-3-2 cell design.

Chapter 4

Conclusion

Low-power bulk-silicon FinFET technology at the 7/8nm node [9] has been investigated via TCAD 3-D device simulation and compact modeling in this study. A super-steep retrograde (SSR) fin channel doping profile enabled by oxygen insertion technology is beneficial for improving device performance (particularly $I_{\mathrm{d,lin}}$ by 10% for NMOS and by 5% for PMOS) and for reducing sensitivity of device performance to process-induced variations. These benefits provide for superior write-ability of 6T-SRAM cells, and are projected to facilitate reductions in the minimium cell operating voltage (by 60-70 mV as compared with conventional FinFET technology), to below 0.50 V. This work shows that bulk-silicon FinFET technology can extend CMOS scaling beyond the 10 nm node.

Future work

Chapter 5

Appendix: FinFET I-V Compact Model

The analytical model used in this study for fast estimation of 6T-SRAM performance metrics and cell yield is adopted from [26]. It is based on the short-channel MOSFET I-V equations, which include channel length modulation (CLM), velocity saturation, and bulk charge effects:

$$I_{ds} = \begin{cases} I_{\text{sub}}(1 - e^{-\frac{V_{\text{DS}}}{V_{\text{th}}}})e^{\frac{V_{\text{GS}}-V_t}{SS}} & V_{\text{GS}} \leq V_t, \\ I_{\text{sub}}(1 - e^{-\frac{V_{\text{DS}}}{V_{\text{th}}}}) + \mu_l C_{ox}\frac{W}{L}\frac{V_{\text{DS}}(V_{\text{GS}}-V_t-mV_{\text{DS}}/V_0)}{1+\frac{V_{\text{GS}}-V_t}{\text{E}_{\text{sat}}L}}(1+\lambda V_{\text{DS}}) & V_{\text{GS}} > V_t \text{ and } V_{\text{DS}} \leq \frac{V_{\text{GS}}-V_t}{m}, \\ I_{\text{sub}}(1 - e^{-\frac{V_{\text{DS}}}{V_{\text{th}}}}) + \mu_{\text{S}} C_{ox}\frac{W}{L}\frac{(V_{\text{GS}}-V_t)^2}{1+\frac{V_{\text{GS}}-V_t}{\text{E}_{\text{sat}}L}}(1+\lambda V_{\text{DS}}) & V_{\text{GS}} > V_t \text{ and } V_{\text{DS}} > \frac{V_{\text{GS}}-V_t}{m}. \end{cases} \tag{5.1}$$

where C_{ox} is the gate oxide capacitance per unit area, W is the effective width of the fin channel region, L is the gate length, V_{GS} is the gate-source voltage, V_{DS} is the drain-source voltage, V_{th} is the thermal voltage (26 mV at 300K), and I_{sub} is the current level that defines the threshold voltage, *i.e.* $100\text{nA} \times \frac{W_{\text{eff}}}{L_{\text{gate}}}$. $V_{\text{t,sat}}$ and $V_{\text{t,lin}}$ are threshold voltage (V_t) extrapolated at the constant current level I_{sub} from I_{ds} $vs.$ V_{gs} curves. Accounting for the DIBL effect:

$$V_{\text{t}} = V_{T0} - D \cdot V_{\text{DS}} \tag{5.2}$$

To ensure continuity between different MOSFET operating regimes, the parameter V_0 is given by:

$$V_0 = \frac{1}{1 - \frac{\mu_{\text{S}}}{2\mu_l}} \tag{5.3}$$

The values of the remaining fitting parameters are calculated based on seven I-V data points obtained from TCAD 3-D device simulations. Table 5.1 lists these seven operating points:

Parameter	V_{GS} (V)	V_{DS} (V)
$I_{D,lin}$	V_{DD} (0.80 V)	$V_{DS,lin}$ (50 mV)
$I_{D,linlo}$	$V_{GS,linlo}$ (higher than $V_{t,lin}$)	$V_{DS,lin}$
$I_{D,sat}$	V_{DD}	V_{DD}
$I_{D,lo}$	V_{DD}/2	V_{DD}
I_{OFF}	0 V	V_{DD}
$I_{D,hi}$	V_{DD}/2	V_{DD}/2
$V_{t,lin}$		$V_{DS,lin}$
$V_{t,sat}$		V_{DD}

Table 5.1: FinFET operating points used to calibrate the compact model.

The equations used to calculate the fitting parameters are given below:

$$D = |(V_{t,lin} - V_{t,sat})/(V_{DD} - V_{ds,lin})| \tag{5.4}$$

$$V_{T0} = V_{t,sat} + D \cdot V_{DD} \tag{5.5}$$

$$SS = |V_{t,sat}/(ln\frac{I_{off}}{I_{sub}})| \tag{5.6}$$

$$E_{sat}L = \frac{(V_{DD}/2 - V_{t,sat})(I_{D,lo} - I_{sub})(\frac{V_{DD}-V_{t,sat}}{V_{DD}/2-V_{t,sat}})^2 - (V_{DD} - V_{t,sat})(I_{d,sat} - I_{sub})}{I_{d,sat} - I_{sub} - (I_{D,lo} - I_{sub})(\frac{V_{DD}-V_{t,sat}}{V_{DD}/2-V_{t,sat}})^2} \tag{5.7}$$

$$\lambda = \frac{\frac{I_{d,sat}-I_{sub}}{I_{D,hi}-I_{sub}} - \frac{(V_{DD}-V_{t,sat})^2}{(V_{DD}+(\frac{V_{DD}}{2}-V_{ds,lin})DIBL-V_{t,lin})^2}\frac{E_{sat}L+V_{DD}+(\frac{V_{DD}}{2}-V_{ds,lin})DIBL-V_{t,lin}}{E_{sat}L+V_{DD}-V_{t,sat}}}{\frac{(V_{DD}-V_{t,sat})^2(E_{sat}L+V_{DD}+(\frac{V_{DD}}{2}-V_{ds,lin})DIBL-V_{t,lin})}{(V_{DD}+(\frac{V_{DD}}{2}-V_{ds,lin})DIBL-V_{t,lin})^2(E_{sat}L+V_{DD}-V_{t,sat})} - \frac{I_{d,sat}-I_{sub}}{2(I_{D,hi}-I_{sub})}} \tag{5.8}$$

$$\mu_S = \frac{2m(I_{d,sat} - I_{sub})(1 + \frac{(V_{DD}-V_{t,sat})}{E_{sat}L}}{(V_{DD} - V_{t,sat})^2(1 + \lambda V_{DD})} \tag{5.9}$$

To increase the accuracy for modeling low-field mobility, low-field mobility μ_l is a function of gate-source voltage:

$$\mu_l = \mu'_L(\frac{V_{GS} + V_t(V_{DS})}{V_{DD} + V_{t,lin}})^\alpha \tag{5.10}$$

where μ'_L is the low-field mobility at $V_{GS} = V_{DD}$, and μ_{lO} is the low-field mobility at $V_{GS} = V_{gs,linlo}$. ($V_{DD} > V_{gs,linlo} > V_{t,lin}$).

μ_{lO} and μ'_L are given by:

$$\mu_{lO} = \frac{1}{C_{ox}W_{eff}L_g}\frac{[I_{ds,lin} - I_{sub}(1 - e^{\frac{V_{ds,lin}}{V_{th}}})]\frac{1+\frac{V_{DD}-V_{t,lin}}{E_{sat}L}}{V_{ds,lin}(1+\lambda V_{ds,lin})}}{V_{DD} - V_{t,lin} - mV_{ds,lin}V_{0,lo}} \tag{5.11}$$

where, $V_{0,\text{lo}}$ is given by:

$$V_{0,\text{lo}} = \frac{1 - \frac{\mu_S(V_{gs,\text{linlo}} - V_{t,\text{lin}})}{2\times[I_{d,\text{linlo}} - I_{sub}(1 - e^{\frac{V_{ds,\text{lin}}}{V_{th}}})]\frac{1+\frac{V_{gs,\text{linlo}} - V_{t,\text{lin}}}{E_{sat}L}}{V_{ds,\text{lin}}(1+\lambda V_{ds,\text{lin}})}}}{1 - \frac{V_{ds,\text{lin}} m \times \mu_S}{2\times[I_{d,\text{linlo}} - I_{sub}(1 - e^{\frac{V_{ds,\text{lin}}}{V_{th}}})]\frac{1+\frac{V_{gs,\text{linlo}} - V_{t,\text{lin}}}{E_{sat}L}}{V_{ds,\text{lin}}(1+\lambda V_{ds,\text{lin}})}}} \tag{5.12}$$

$$\mu'_L = \frac{1}{C_{ox}W_{eff}L_g}\frac{[I_{ds,\text{lin}} - I_{sub}(1 - e^{\frac{V_{ds,\text{lin}}}{V_{th}}})]\frac{1+\frac{V_{DD} - V_{t,\text{lin}}}{E_{sat}L}}{V_{ds,\text{lin}}(1+\lambda V_{ds,\text{lin}})}}{V_{DD} - V_{t,\text{lin}} - mV_{ds,\text{lin}}V_{0,\text{lin}}} \tag{5.13}$$

and $V_{0,\text{lin}}$ is given by:

$$V_{0,\text{lin}} = \frac{1 - \frac{\mu_S(V_{DD} - V_{t,\text{lin}})}{2\times[I_{ds,\text{lin}} - I_{sub}(1 - e^{\frac{V_{ds,\text{lin}}}{V_{th}}})]\frac{1+\frac{V_{DD} - V_{t,\text{lin}}}{E_{sat}L}}{V_{ds,\text{lin}}(1+\lambda V_{ds,\text{lin}})}}}{1 - \frac{V_{ds,\text{lin}} m \times \mu_S}{2\times[I_{ds,\text{lin}} - I_{sub}(1 - e^{\frac{V_{ds,\text{lin}}}{V_{th}}})]\frac{1+\frac{V_{DD} - V_{t,\text{lin}}}{E_{sat}L}}{V_{ds,\text{lin}}(1+\lambda V_{ds,\text{lin}})}}} \tag{5.14}$$

Bibliography

[1] C. Auth, C. Allen, A. Blattner, D. Bergstrom, M. Brazier, M. Bost, M. Buehler, V. Chikarmane, T. Ghani, T. Glassman, R. Grover, W. Han, D. Hanken, M. Hattendorf, P. Hentges, R. Heussner, J. Hicks, D. Ingerly, P. Jain, S. Jaloviar, R. James, D. Jones, J. Jopling, S. Joshi, C. Kenyon, H. Liu, R. McFadden, B. Mcintyre, J. Neirynck, C. Parker, L. Pipes, I. Post, S. Pradhan, M. Prince, S. Ramey, T. Reynolds, J. Roesler, J. Sandford, J. Seiple, P. Smith, C. Thomas, D. Towner, T. Troeger, C. Weber, P. Yashar, K. Zawadzki, and K. Mistry. A 22nm high performance and low-power cmos technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density mim capacitors. In *VLSI Technology (VLSIT), 2012 Symposium on*, pages 131–132, June 2012.

[2] D. Hisamoto, Wen-Chin Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, Erik Anderson, Tsu-Jae King, J. Bokor, and Chenming Hu. Finfet-a self-aligned double-gate mosfet scalable to 20 nm. *Electron Devices, IEEE Transactions on*, 47(12):2320–2325, Dec 2000.

[3] Clement Wann, Robert Wong, DJ Frank, R Mann, Shang-Bin Ko, P Croce, Dallas Lea, Dennis Hoyniak, Yoo-Mi Lee, James Toomey, et al. Sram cell design for stability methodology. In *VLSI Technology, 2005.(VLSI-TSA-Tech). 2005 IEEE VLSI-TSA International Symposium on*, pages 21–22. IEEE, 2005.

[4] John L Hennessy and David A Patterson. *Computer architecture: a quantitative approach*. Elsevier, 2011.

[5] Huiming Bu. Finfet technology a substrate perspective. In *SOI Conference (SOI), 2011 IEEE International*, pages 1–27. IEEE, 2011.

[6] Asen Asenov and Subhash Saini. Suppression of random dopant-induced threshold voltage

fluctuations in sub-0.1-μm mosfet's with epitaxial and δ-doped channels. *Electron Devices, IEEE Transactions on*, 46(8):1718–1724, 1999.

[7] N Xu, N Damrongplasit, H Takeuchi, RJ Stephenson, NW Cody, A Yiptong, X Huang, M Hytha, RJ Mears, and Tsu-Jae King Liu. Mosfet performance and scalability enhancement by insertion of oxygen layers. *Proc. IEEE Electron Devices Lett*, pages 318–320, 2012.

[8] Sentaurus Device User Guide and E Version. Synopsys inc. *Mountain View, CA*, page 384, 2010.

[9] Linda Wilson. International technology roadmap for semiconductors (itrs). *Semiconductor Industry Association*, 2013.

[10] S Natarajan, M Agostinelli, S Akbar, M Bost, A Bowonder, V Chikarmane, S Chouksey, A Dasgupta, K Fischer, Q Fu, et al. A 14nm logic technology featuring 2 nd-generation finfet, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm 2 sram cell size. In *Electron Devices Meeting (IEDM), 2014 IEEE International*, pages 3–7. IEEE, 2014.

[11] A Hokazono, K Ohuchi, K Miyano, I Mizushima, Y Tsunashima, and Y Toyoshima. Source/drain engineering for sub-100 nm cmos using selective epitaxial growth technique. In *Electron Devices Meeting, 2000. IEDM'00. Technical Digest. International*, pages 243–246. IEEE, 2000.

[12] Jakub Kedzierski, Meikei Ieong, Edward Nowak, Thomas S Kanarsky, Ying Zhang, Ronnen Roy, Diane Boyd, David Fried, and H-SP Wong. Extension and source/drain design for high-performance finfet devices. *Electron Devices, IEEE Transactions on*, 50(4):952–958, 2003.

[13] Max V Fischetti and Steven E Laux. Band structure, deformation potentials, and carrier mobility in strained si, ge, and sige alloys. *Journal of Applied Physics*, 80(4):2234–2252, 1996.

[14] Shiyu Sun, Shankar Muthukrishnan, Ben Ng, Stephen Nagy, Alan Zojaji, Rubi Lapena, Yihwan Kim, Vivek Rao, Dimitry Kouzminov, Naushad Variam, et al. Enable abrupt junction and advanced salicide formation with dynamic surface annealing. *physica status solidi (c)*, 9(12):2436–2439, 2012.

[15] Tsu-Jae King Liu, Changhwan Shin, Min Hee Cho, Xin Sun, B Nikolic, and Bich-yen Nguyen. Sram cell design considerations for soi technology. In *SOI Conference, 2009 IEEE International*, pages 1–4. IEEE, 2009.

[16] M. G. Ancona. Macroscopic physics of the silicon inversion layer. *Physical Review B*, 35(15):7959–7965, 1987.

[17] Kuo-Hsing Kao, Anne S Verhulst, William G Vandenberghe, Bart Soree, Guido Groeseneken, and Kristin De Meyer. Direct and indirect band-to-band tunneling in germanium-based tfets. *Electron Devices, IEEE Transactions on*, 59(2):292–301, 2012.

[18] Weize Xiong, Gabriel Gebara, Joyti Zaman, Michael Gostkowski, Billy Nguyen, Greg Smith, David Lewis, C Rinn Cleavelin, Rick Wise, Shaofeng Yu, et al. Improvement of finfet electrical characteristics by hydrogen annealing. *Electron Device Letters, IEEE*, 25(8):541–543, 2004.

[19] Synopsys Mountain View CA U.S.A. Sentaurus user's manual. 2014.

[20] DBM Klaassen. A unified mobility model for device simulation—i. model equations and concentration dependence. *Solid-State Electronics*, 35(7):953–959, 1992.

[21] M. G. Ancona. Quantum correction to the equation of state of an electron gas in a semiconductor. *Physical Review B*, 39(13):9536–9540, 1989.

[22] MG Ancona and HF Tiersten. Macroscopic physics of the silicon inversion layer. *Physical Review B*, 35(15):7959, 1987.

[23] Varadarajan Vidya. *Thin-Body Silicon FET Devices and Technology*. PhD thesis, Citeseer, 2007.

[24] EO Kane. Zener tunneling in semiconductors. *Journal of Physics and Chemistry of Solids*, 12(2):181–188, 1960.

[25] Lee Smith, Victor Moroz, Geert Eneman, Peter Verheyen, Faran Nouri, Lori Washington, Malgorzata Jurczak, Oleg Penzin, Dipu Pramanik, and Kristin De Meyer. Exploring the limits of stress-enhanced hole mobility. *Electron Device Letters, IEEE*, 26(9):652–654, 2005.

[26] Andrew Evert Carlson. *Device and Circuit Techniques for Reducing Variation in Nanoscale SRAM*. ProQuest, 2008.

[27] Xingsheng Wang, Andrew R Brown, Binjie Cheng, and Asen Asenov. Statistical variability and reliability in nanoscale finfets. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 5–4. IEEE, 2011.

[28] Kelin J Kuhn. Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale cmos. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 471–474. IEEE, 2007.

[29] T Matsukawa, S O'uchi, K Endo, Y Ishikawa, H Yamauchi, YX Liu, J Tsukada, K Sakamoto, and M Masahara. Comprehensive analysis of variability sources of finfet characteristics. In *VLSI Technology, 2009 Symposium on*, pages 118–119. IEEE, 2009.

[30] Azeez J Bhavnagarwala, Xinghai Tang, and James D Meindl. The impact of intrinsic device fluctuations on cmos sram cell stability. *Solid-State Circuits, IEEE Journal of*, 36(4):658–665, 2001.

[31] Hamed F Dadgour, Kazuhiko Endo, Vivek K De, and Kaustav Banerjee. Grain-orientation induced work function variation in nanoscale metal-gate transistors—part i: Modeling, analysis, and experimental validation. *Electron Devices, IEEE Transactions on*, 57(10):2504–2514, 2010.

[32] Asen Asenov. Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm mosfet's: A 3-d "atomistic" simulation study. *Electron Devices, IEEE Transactions on*, 45(12):2505–2513, 1998.

[33] A. Dixit, K.G. Anil, E. Baravelli, P. Roussel, A. Mercha, C. Gustin, M. Bamal, E. Grossar, R. Rooyackers, E. Augendre, M. Jurczak, S. Biesemans, and K. De Meyer. Impact of stochastic mismatch on measured sram performance of finfets with resist/spacer-defined fins: Role of line-edge-roughness. In *Electron Devices Meeting, 2006. IEDM '06. International*, pages 1–4, Dec 2006.

[34] Chris Bencher, Yongmei Chen, Huixiong Dai, Warren Montgomery, and Lior Huli. 22nm half-pitch patterning by cvd spacer self alignment double patterning (sadp). In *SPIE Ad-*

vanced Lithography, pages 69244E–69244E. International Society for Optics and Photonics, 2008.

[35] Peng Zheng, Y-B Liao, Nattapol Damrongplasit, M-H Chiang, and T-JK Liu. Variation-aware comparative study of 10-nm gaa versus finfet 6-t sram performance and yield. 2014.

[36] Andreas Wettstein, Oleg Penzin, Eugeny Lyumkis, and Wolfgang Fichtner. Random dopant fluctuation modelling with the impedance field method. In *Proceedings of International Conference on Simulation of Semiconductor Devices and Processes*, pages 91–94, 2003.

[37] Christopher C Hobbs, Leonardo RC Fonseca, Andrey Knizhnik, Veeraraghavan Dhandapani, Srikanth B Samavedam, William J Taylor, John M Grant, LG Dip, Dina H Triyoso, Rama I Hegde, et al. Fermi-level pinning at the polysilicon/metal-oxide interface-part ii. *Electron Devices, IEEE Transactions on*, 51(6):978–984, 2004.

[38] Wenjuan Zhu, Jin-Ping Han, and TP Ma. Mobility measurement and degradation mechanisms of mosfets made with ultrathin high-k dielectrics. *Electron Devices, IEEE Transactions on*, 51(1):98–105, 2004.

[39] Kaizad Mistry, C Allen, C Auth, B Beattie, D Bergstrom, M Bost, M Brazier, M Buehler, A Cappellani, R Chau, et al. A 45nm logic technology with high-k+ metal gate transistors, strained silicon, 9 cu interconnect layers, 193nm dry patterning, and 100% pb-free packaging. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, pages 247–250. IEEE, 2007.

[40] Leland Chang, David M Fried, Jack Hergenrother, Jeffrey W Sleight, Robert H Dennard, Robert K Montoye, Lidija Sekaric, Sharee J McNab, Anna W Topol, Charlotte D Adams, et al. Stable sram cell design for the 32 nm node and beyond. In *VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on*, pages 128–129. IEEE, 2005.

[41] Bastien Giraud, Amara Amara, and Andrei Vladimirescu. A comparative study of 6t and 4t sram cells in double-gate cmos with statistical variation. In *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, pages 3022–3025. IEEE, 2007.

[42] Makoto Yabuuchi, Masao Morimoto, Yasumasa Tsukamoto, Shinji Tanaka, Koji Tanaka, Miki Tanaka, and Koji Nii. 16 nm finfet high-k/metal-gate 256-kbit 6t sram macros with

wordline overdriven assist. In *Electron Devices Meeting (IEDM), 2014 IEEE International*, pages 3–3. IEEE, 2014.