COPPE
Instituto Alberto Luiz Coimbra de
Pós-Graduação e Pesquisa de Engenharia
UFRJ

STUDY AND DEVELOPMENT OF LOW POWER CONSUMPTION SRAMs ON 28 nm FD-SOI CMOS PROCESS

Luis Fabián Olivera Mederos

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Antonio Petraglia

Rio de Janeiro
Setembro de 2017

STUDY AND DEVELOPMENT OF LOW POWER CONSUMPTION SRAMs ON 28 nm FD-SOI CMOS PROCESS

Luis Fabián Olivera Mederos

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Antonio Petraglia, Ph.D.

Prof. José Gabriel Rodríguez Carneiro Gomes, Ph.D.

Prof. João Baptista de Oliveira e Souza Filho, D.Sc.

Prof. Felipe Maia Galvão França, Ph.D.

Carlos

Prof. Carlos Galup Montoro, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2017

Mederos, Luis Fabián Olivera

Study and Development of Low Power Consumption SRAMs on 28 nm FD-SOI CMOS Process/Luis Fabián Olivera Mederos. – Rio de Janeiro: UFRJ/COPPE, 2017.

XXI, 97 p.: il.; 29,7cm.

Orientador: Antonio Petraglia

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2017.

Referências Bibliográficas: p. 82 – 92.

1. CMOS memory circuits. 2. leakage current. 3. low-power design. 4. process variations. 5. SRAM. 6. static noise margin. 7. sub-threshold operation. 8. write margin. 9. yield. I. Petraglia, Antonio. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

I dedicate this thesis to:
the memory of my father,
L. A. Olivera, who inspired me to
follow this way. And to Gabriela,
Cachita and Ana for their constant,
unconditional love and support.

Acknowledgments

Agradeço ao meu orientador, o professor Antonio Petraglia, por sua orientação e conhecimentos transmitidos ao longo destes 5 anos de mestrado e doutorado, o qual foi além da simples atividade de pesquisa. Sua constante disposição e suporte possibilitaram o desenvolvimento desta tese de doutorado.

A mi madre Gabriela y mis abuelas Cacha e Irma, que gracias a las largas conversaciones diarias, me transmiten el apoyo necesario para mantenerme motivado y con un propósito en la vida; a Julio y Jhona por ser parte de la familia y cuidar de mi madre y mis abuelas; a mis padrinos Eduardo e Lidia; y a mi primo Federico, Helena e mi ahijado Enzito por los buenos momentos vividos en esta etapa.

Aos colegas e amigos do Laboratório de Processamento Analógico e Digital de Sinais (PADS) pelo agradável ambiente de trabalho, que contribuíram direta ou indiretamente neste trabalho: Fernanda, Thiago, Genildo, Odair, David, João, Eduardo, Motta, Gabriela, Gustavo, Felipe e Estevão. Aos professores Gabriel, Baruqui, Mariane e Teodósio.

A los grandes amigos de Uruguay que con ayuda de la tecnología siempre están presentes a pesar de la distancia, para proporcionar aquellos momentos de desestrés que uno precisa para continuar enfocado: Javier O., Juan, Rodrigo, Javier S., Sebastian, Emilio, Ari, Marcelo y Federico. Y también a los amigos del grupo de microelectrónica de la UDELAR.

À família Italiano, que sempre proporcionou carinho e apoio, além de me fazer sentir parte da mesma.

Y por último, a la persona que comparte conmigo los buenos y malos momentos, siendo muy importante para mi motivación diaria, Ana Italiano.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

ESTUDO E DESENVOLVIMENTO DE SRAMs PARA BAIXO CONSUMO DE ENERGIA EM PROCESSO 28 nm FD-SOI CMOS

Luis Fabián Olivera Mederos

Setembro/2017

Orientador: Antonio Petraglia

Programa: Engenharia Elétrica

O projeto de circuitos analógicos em processos nanométricos CMOS ($<$ 90 nm) pode ser substancialmente afetado pelas variações do processo de fabricação, sendo cada vez mais desafiador para os projetistas alcançar soluções eficientes no desempenho dos circuitos mediante o uso de modelos analíticos. Simulações extensas com alto custo computacional são normalmente requeridas para providenciar um correto funcionamento do circuito. Por outro lado, devido ao fato que a estrutura *bulk*-CMOS esta alcançando seus limites de escala ($<$ 32 nm), outros transistores foram desenvolvidos como sucessores, tais como o *fully depleted silicon-on-insulator* (FD-SOI), *Multigate* MOSFET, entre outros, surgindo novas técnicas de projeto que utilizam as características aprimoradas destes dispositivos. Dessa forma, esta tese de doutorado se foca no desenvolvimento de modelos analíticos dos parâmetros mais importantes do cache SRAM implementado em processo CMOS FD-SOI de 28 nm, principalmente para explorar as dimensões dos transistores com baixo custo computacional, e assim produzir soluções eficientes em termos de consumo de energia, velocidade e rendimento. Aproveitando o baixo custo computacional e a alta concordância dos modelos analíticos, nesta tese fomos capazes de propor um dimensionamento não tradicional para a célula de memória 6T-SRAM, em que diferentemente do clássico dimensionamento "*thin-cell*", os comprimentos dos transistores são utilizados como variável de projeto com o fim de reduzir o consumo estático de corrente. A estrutura *single-P-well* (SPW), combinada com a técnica *reverse-body-biasing* (RBB) foram utilizadas para alcançar um melhor balanço entre as correntes especificas dos transistores do tipo P e N. Como resultado, foi implementada uma cache SRAM de 128 kb, e as simulações mostraram que o circuito pode consumir uma energia média por operação de 0.604 pJ/word-access (64 I/O bits) utilizando uma fonte de alimentação de 0.45 V e frequência de operação 40 MHz. O cache SRAM ocupou uma área no chip de 0.060 mm^2.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

STUDY AND DEVELOPMENT OF LOW POWER CONSUMPTION SRAMs ON 28 nm FD-SOI CMOS PROCESS

Luis Fabián Olivera Mederos

September/2017

Advisor: Antonio Petraglia

Department: Electrical Engineering

Since analog circuit designs in CMOS nanometer ($<$ 90 nm) nodes can be substantially affected by manufacturing process variations, circuit performance becomes more challenging to achieve efficient solutions by using analytical models. Extensive simulations are thus commonly required to provide a high yield. On the other hand, due to the fact that the classical bulk MOS structure is reaching scaling limits ($<$ 32 nm), alternative approaches are being developed as successors, such as fully depleted silicon-on-insulator (FD-SOI), Multigate MOSFET, FinFETs, among others, and new design techniques emerge by taking advantage of the improved features of these devices. This thesis focused on the development of analytical expressions for the major performance parameters of the SRAM cache implemented in 28 nm FD-SOI CMOS, mainly to explore the transistor dimensions at low computational cost, thereby producing efficient designs in terms of energy consumption, speed and yield. By taking advantage of both low computational cost and close agreement results of the developed models, in this thesis we were able to propose a non-traditional sizing procedure for the simple 6T-SRAM cell, that unlike the traditional thin-cell design, transistor lengths are used as a design variable in order to reduce the static leakage. The single-P-well (SPW) structure in combination with reverse-body-biasing (RBB) technique were used to achieve a better balance between P-type and N-type transistors. As a result, we developed a 128 kB SRAM cache, whose post-layout simulations show that the circuit consumes an average energy per operation of 0.604 pJ/word-access (64 I/O bits) at supply voltage of 0.45 V and operation frequency of 40 MHz. The total chip area of the 128 kB SRAM cache is 0.060 mm^2.

Contents

List of Figures

List of Tables

List de Symbols

η	Body biasing effect of MOS transistor
η_n	Body biasing effect of NMOS transistor
η_p	Body biasing effect of PMOS transistor
λ	DIBL effect of MOS transistor
λ_n	DIBL effect of NMOS transistor
λ_p	DIBL effect of PMOS transistor
σ	Standard deviation
$\sigma_{V_{To}}$	Standard deviation of MOS transistor threshold voltage
$\sigma_{V_{Ton}}$	Standard deviation of NMOS transistor threshold voltage
$\sigma_{V_{Top}}$	Standard deviation of PMOS transistor threshold voltage
A_{VTo}	Technology constant for threshold voltage mismatch variations
A_V	Maximum small-signal gain of the inverter
C_{BL}	Bit-line capacitance
C_{WL}	Word-line capacitance
E_D	SRAM dynamic energy per operation
E_S	SRAM static energy per operation
E_T	SRAM total energy per operation
F_{CLK}	SRAM operation frequency
G_m	Total transconductance of inverter
I_{DS}	Drain-to-source current of MOS transistor

I_{DSn}	Drain-to-source current of NMOS transistor
I_{GND}	Current drained to the ground by the SRAM
$I_{Leak,AT}$	Leakage current of the access transistor
I_{Leak}	SRAM leakage current
I_{Read}	Read current produced by the access transistor
I_{SDp}	Source-to-drain current of PMOS transistor
I_{So}	Specific current of MOS transistor
I_{Son}	Specific current of NMOS transistor
I_{Sop}	Specific current of PMOS transistor
L	Length of MOS transistor
L_{at}	Access transistor length
L_{pd}	Pull-down transistor length
L_{pu}	Pull-up transistor length
L_n	Length of NMOS transistor
L_p	Length of PMOS transistor
M	Number of word-line rows of the SRAM
N	Number of bit-line columns of the SRAM
n	Slope factor of MOS transistor
n_n	Slope factor of NMOS transistor
n_p	Slope factor of PMOS transistor
P_{Read}	Probability of SRAM read operation
P_{Write}	Probability of SRAM write operation
r_{at}	Access transistor ratio
r_{pd}	Pull-down transistor ratio
r_{pu}	Pull-up transistor ratio

T_{CLK}	SRAM cycle time
T_{PC}	Pre-charge time of the SRAM
T_{Read}	Read operation cycle time of the SRAM
T_{SA}	SA delay
T_{SAEN}	Time to enable the SA
T_{WD}	Write-driver time of the SRAM
T_{Write}	Write operation cycle time of the SRAM
U_T	Thermal voltage
V_{BB}	Body bias voltage
V_{BL}	Bit-line voltage
V_{Bn}	Substrate-to-ground voltage of NMOS transistor
V_{Bp}	Substrate-to-ground voltage of PMOS transistor
V_{BS}	Substrate-to-source voltage of MOS transistor
$V_{DD,min}$	Minimum supply voltage
V_{DD}	Supply voltage
V_{DS}	Drain-to-source voltage of MOS transistor
V_{GS}	Gate-to-source voltage of MOS transistor
V_{IL}	Low nominal output voltage of the inverter
V_{in}	Input voltage
V_{OH}	High nominal output voltage of the inverter
V_{OS}	Maximum input-offset of the SA
V_{out}	Output voltage
V_S	Metastability voltage of inverter
V_{SG}	Source-to-gate voltage of MOS transistor
V_{SS}	Ground voltage

$V_{T,\mathrm{eff}}$	Effective threshold voltage of MOS transistor
V_{To}	Threshold voltage of MOS transistor
V_{Ton}	Threshold voltage of NMOS transistor
V_{Top}	Threshold voltage of PMOS transistor
V_{WL}	Word-line voltage
W	MOS transistor width
W_{at}	Access transistor width
W_{pd}	Pull-down transistor width
W_{pu}	Pull-up transistor width
W_n	NMOS transistor width
W_p	PMOS transistor width

List of Abbreviations

6T-SRAM	Six-Transistor Static Random Access Memory
7T-LTSA	Seven-Transistor Latch-Type Sense Amplifier
BL	Bit-Line
CMOS	Complementary Metal-Oxide-Semiconductor
DCI	Differential Charge Injection
DIBL	Drain-Induced Barrier Lowering
FD-SOI	Fully Depleted Silicon-On-Insulator
FF	N-Fast-P-Fast corner
FFA	N-Fast-P-Fast larger spread corners
FinFET	Fin Field-Effect Transistor
FOM	Figure Of Merit
FS	N-Fast-P-Slow corner
FSA	N-Fast-P-Slow larger spread corners
HSNM	Hold Static Noise Margin
IoT	Internet of Things
ITRS	International Technology Roadmap for Semiconductors
LVT	Low Threshold Voltage
MEP	Minimum Energy Point
MOSFET	Metal-Oxide Semiconductor Field-Effect Transistors
NMOS	N-channel Metal-Oxide-Semiconductor

PMOS	P-channel Metal-Oxide-Semiconductor
RBB	Reverse-Body-Biasing
RD	Rectangular-Diffusion
RDF	Random Dopant Fluctuations
RSNM	Read Static Noise Margin
RVT	Regular Threshold Voltage
SA	Sense Amplifier
SF	N-Slow-P-Flow corner
SFA	N-Slow-P-Flow larger spread corners
SNM	Static Noise Margin
SoC	System-on-Chip
SOI	Silicon-On-Insulator
SPW	Single-P-Well
SRAM	Static Random Access Memory
SS	N-Slow-P-Slow corner
SSA	N-Slow-P-Slow larger spread corners
TT	N-Typical-P-Typical corner
UICM	Unified Current-Control Model
ULP	Ultra Low Power
ULV	Ultra Low Voltage
VTC	Voltage Transfer Curve
WL	Word-Line
WSNM	Write Static Noise Margin

Chapter 1

Introduction

1.1 Motivation: Low-Power and Robust SRAM Design

Historically, static random access memories (SRAMs) have been designed to fill two needs: direct interface with the central processor unit (CPU) at speeds faster than dynamic random access memories (DRAMs), and to replace DRAMs in systems that require low power consumption. Due to the growth of internet of things (IoT) industry applications, ultra low power (ULP) has received increasing attention in system-on-chip (SoC) designs owing to the emergence of several circuits that require prolonged battery life, such as mobile phones, personal computers, sensor networks and biomedical implantable circuits [1–4].

According to the International Technology Roadmap for Semiconductors (ITRS) forecast shown in Fig. 1.1 [5, 6], SRAMs currently should occupy above 90% of the SoC area, and their leakage currents are one of the major reasons for static power dissipation. An effective way to reduce this static dissipation is by decreasing the supply voltage at the minimum operation value, since it has an exponential influence on the static leakage [7–10]. However, as can be seen in Fig. 1.2, the scaling of CMOS technology below 90 nm nodes creates a crucial challenge for designers, since the manufacturing process variability is not scaling at the same rate as those of transistor dimensions [11–13], and hence as the technology node decreases so does the static noise margin (SNM). Usually, the smallest possible dimensions are used to realize high-density SRAM cells, and consequently the large variability of these transistors may have a dramatic impact on the minimum supply voltage that would ensure a proper cell operation, such that stored data would not be lost. One of the most common metrics to quantify the minimum supply voltage is the SNM [14–22], since their positive values allow safe cell operation. In addition to the manufacturing process variations, short channel effects have a strong impact on SNM when minimum dimensions are used, and such effects should be taken into account on the design.

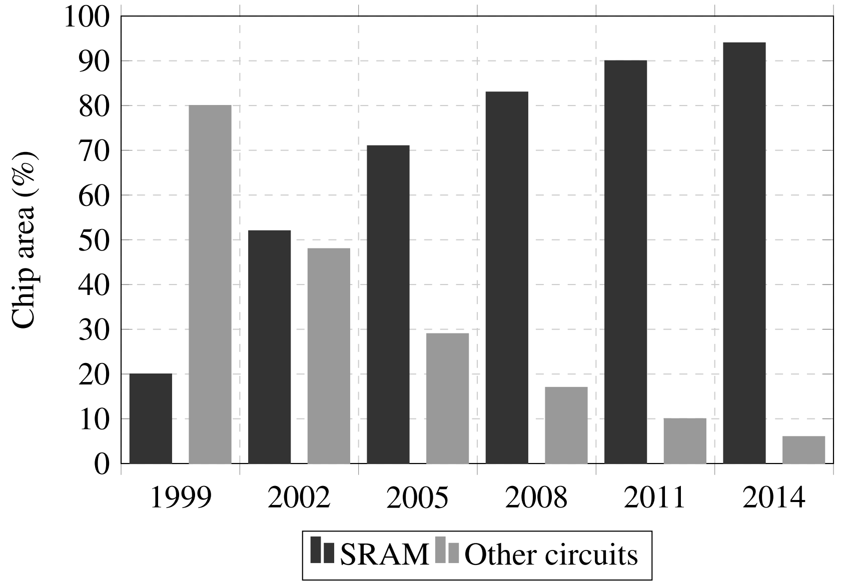


Figure 1.1: Chip area percentage of the SRAM and other circuits *versus* technology scaling [5, 6].

There are four kinds of failure mechanisms caused by process variability on the SRAM [23–25]: hold failure (data stored by the cell is lost due to supply voltage reduction when the memory array operates in standby mode), write failure (new data is not written in the bit-cell), read failure (data stored by the cell is destroyed during read operation), access failure (data stored by the cell is not correctly read). The read failure is the dominant failure mechanism at normal operating conditions, and hence arises the need to design a circuit able to properly read the stored data, which is well-known in the literature as sense amplifier (SA) [26]. The purpose of the SA is to read the contents of many memory cells since it can achieve a fast decision time due to its strong positive feedback. However, in ultra-low-power applications, when the supply voltage must be decreased, the SA decision speed also is decreased [27]. From the point of view of SRAM design, the SA input-offset voltage is a crucial specification since it affects the SRAM yield and speed [27–30].

In order to deal with the large variability of the nanometer nodes, robust operation of SRAM cache has been widely studied in the literature [6, 13, 31]. The most common way to estimate SRAM failure probability is the standard Monte Carlo method, which often requires high computational cost simulations to provide proper circuit operation under manufacturing process variations. To avoid this large number of circuit simulations, and hence to reduce the design time, other techniques were proposed, some of them based on more efficient simulations [32–34], and others on analytical models [23, 24, 35], which predict the circuit performance and produce initial transistor sizing. In this way, as showed in details in Section 1.3, this thesis focuses on the development of analytical expressions for the major performance parameters of the SRAM cache, specially concerning the transistor sizing with low computational cost, so that efficient solutions in terms of energy

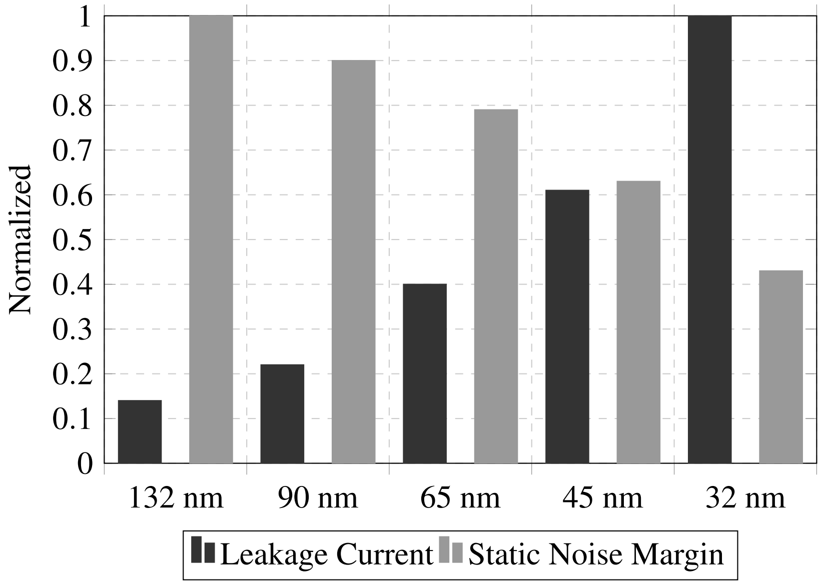


Figure 1.2: SNM compared with leakage current *versus* technology nodes scaling [6].

consumption, speed and yield are produced.

1.2 SRAM Cache

A classical SRAM cache is formed by a memory cell array, row and column decoders, sense amplifiers (SA), write drivers and bit-line pre-charge circuits, as illustrated in Fig. 1.3. The memory array is made by connecting bit-cells through N bit-line columns and M word-line rows. In order to have read and write access to a memory cell, the input address is loaded on the decoders, and thus word-line and bit-line are selected. It is important to note that M usually represents a huge number of cells connected to each column, producing a large load capacitance on the bit-lines. Hence, pre-charge circuits are necessary to deal with their effects on write and read operations. SRAM caches currently reported in the literature achieve 128 Mb [36–38]. Thus, these are 2^{27}= 134.217.728 bit-cells, and hence any small design improvement yields great impact on the entire array either in terms of occupied area or energy consumption.

1.2.1 SRAM Bit-cell

A memory bit-cell is a circuit that can store a single digital data: “0” or “1”. The standard six transistor cell (6T-SRAM) depicted in Fig. 1.4 is widely used in view of its speed, robustness and circuit simplicity [39]. It consists of two cross-coupled CMOS inverters (M_1-M_3 and M_2-M_4) and two n-type access transistors (M_5 and M_6). The inverter pairs are responsible for holding the binary information by their mutual positive feedback, and the access transistors allow read/write access and isolate the cell in hold mode.

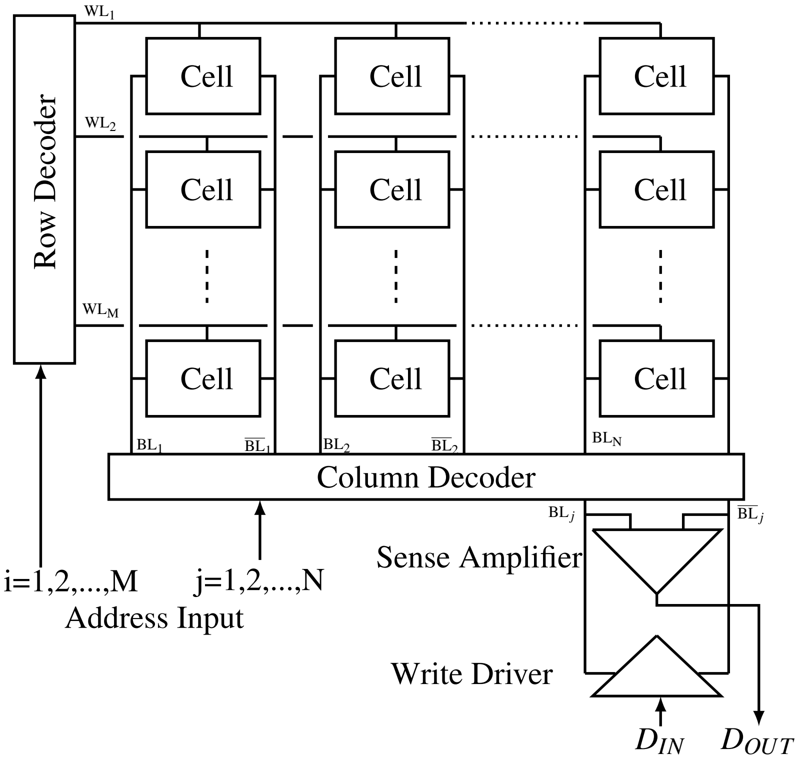


Figure 1.3: General architecture of the SRAM cache.

1.2.2 Sense Amplifier

One of the major peripheral circuits in the SRAM cache is the SA. The primary function of a SA is to amplify a small differential signal from the bit-line during read operation. The bit-lines are pre-charged before the read operation, and weakly discharged only on one side in read operation when the bit-cell is connected through their access transistors. Once the access transistors discharge one side of bit-lines until achieving a voltage difference higher than the input-offset of the SA, the SA is activated to provide an output decision. Usually, the seven-transistor latch-type sense amplifier (7T-LTSA) shown in Fig. 1.5 is used in view of its simplicity, eventhough specific and improved topologies are often reported in the literature [40, 41]. The 7T-LTSA contains two isolating transistors (M_5 and M_6), two cross-coupled inverters (M_1-M_3 and M_2-M_4) and a footer transistor (M_7). When a voltage difference is established between BL and $\overline{\text{BL}}$, the sense amplifier is disconnected from the bit-lines as the *sense enable* (SE) signal is activated. At the same time, helped by their positive feedbacks, the cross-coupled inverters are turned on through the footer transistor to rapidly generate the proper output.

1.2.3 Pre-charge Circuit

Each column of the SRAM array has a bit-line pair (BL and $\overline{\text{BL}}$) which is connected to a pre-charge circuit. Basically, this circuit must pull-up the large capacitance of the voltage

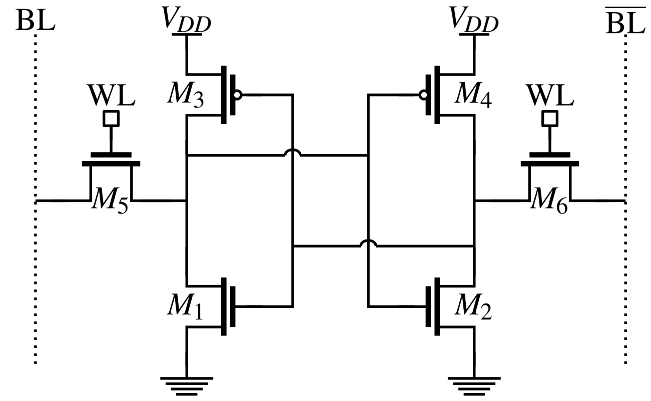


Figure 1.4: Schematic of the 6T-SRAM cell.

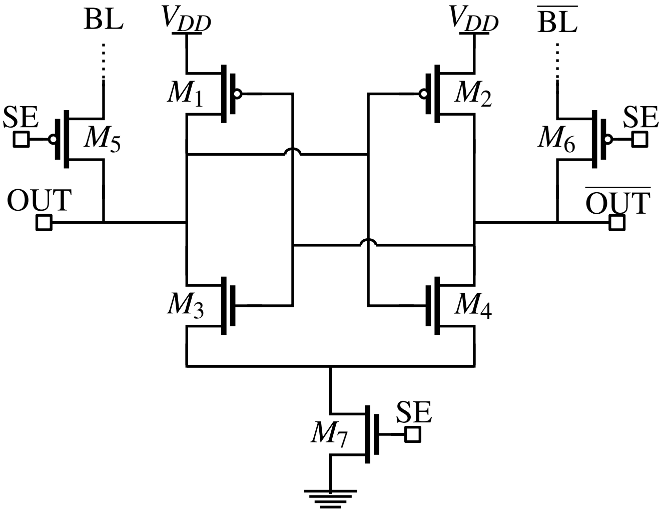


Figure 1.5: Schematic of the 7T latch-type sense amplifier.

nodes BL and $\overline{\text{BL}}$ to the pre-charge level, which is usually equal to the supply voltage. The basic pre-charge circuit is shown in Fig. 1.6. Note that when $\overline{\text{PC}}$ is low, P-type transistors M_1, M_2 and M_3 are tuned on, thereby connecting the bit-lines BL and $\overline{\text{BL}}$ to V_{DD}. The only function of transistor M_3 is to eliminate the offset between bit-lines.

1.2.4 Write Driver

Similar to the case of pre-charge circuit, each bit-line pair (BL and $\overline{\text{BL}}$) is connected to a write driver circuit. In write operation, after bit-lines are pre-charged, the driver must pull-down one side of the bit-lines (BL or $\overline{\text{BL}}$) to ground in order to produce a differential voltage for beating the positive feedback of the memory cell, thereby writing the logic information properly. It can be implemented by the circuit of Fig. 1.7, which activates the access transistors M_3 and M_4 through the write enabled (WRT) signal and only one bit-line (BL or $\overline{\text{BL}}$) by turning on either M_1 or M_2, depending on the input data (DIN) value to be “0” or “1”, respectively.

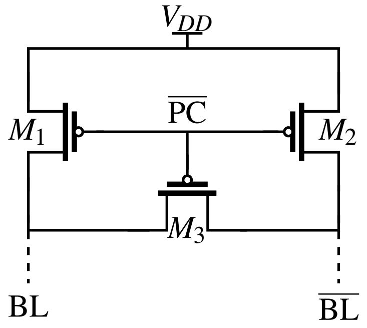


Figure 1.6: Schematic of the pre-charge circuit.

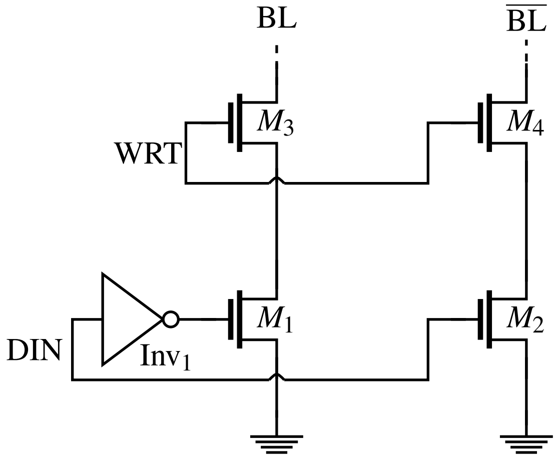


Figure 1.7: Schematic of the write driver circuit.

1.2.5 Mode of Operations

The SRAM cache can have three states of operation: standby (idle mode), read (data is read from the addressed cell) and write (data is written to the addressed cell).

1.2.5.1 Standby Mode

The standby mode holds the logical information when the cell is not being used. This mode is an efficient low power solution in view of the fact that the static leakage can be considerably reduced by decreasing the supply voltage down to the minimum as possible (see Fig. 1.8). This minimum is determined by the hold SNM (HSNM), since their positive values ensure the proper bit-cell operation.

1.2.5.2 Read Operation

The basic read operation is illustrated by the timing diagram in Fig. 1.9, which illustrates the following steps:

R.1: The read cycle starts with both bit-lines (BL's) pre-charged to the pre-charge level

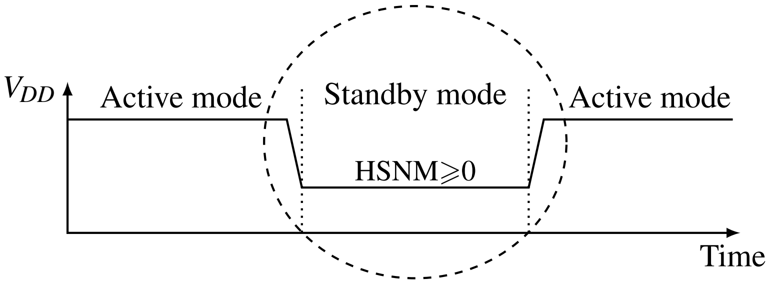


Figure 1.8: Timing diagram of standby mode technique.

by the pre-charger circuit through activating $\overline{\text{PC}}$= “0”.

R.2: The word-line (WL) signal is activated and one of the bit-lines starts to discharge through the bit-cell access transistor. At the same time the SA is connected to the bit-lines by establishing SE=“0”.

R.3: When the differential voltage between bit-lines achieves the maximum input-offset of the SA (V_{OS}), the SA is disconnected from the bit-lines by applying SE=“1” in order that a fast decision is taken at their outputs.

R.4: The read data is available at the output (OUT) of the SA in the next rising edge of clock (CLK).

1.2.5.3 Write Operation

On the other hand, the basic write operation is also detailed in the timing diagram of Fig. 1.9, following the next steps:

W.1: The write cycle starts with both bit-lines pre-charged to the pre-charge level by the pre-charger circuit through activating $\overline{\text{PC}}$= “0”.

W.2: The word-line signal is activated and one side of bit-lines starts to discharge through the write driver, either BL when DIN=“0” or $\overline{\text{BL}}$ when DIN=“1”.

W.3: The differential voltage between bit-lines ΔV_{BL} beats the feedback of bit-cell voltage nodes Q and $\overline{\text{Q}}$, thereby writing the new data.

1.2.6 SRAM Performance

In ULP SRAM cache designs, speed and energy consumption are the major specification parameters. However, other features are important, depending on each application, such as the memory size, the number of accessed bits per operation, minimum supply voltage and chip area. Table 1.1 shows performance comparisons among some reported ULP SRAM designs. Observe that the energy per operation is commonly specified in either

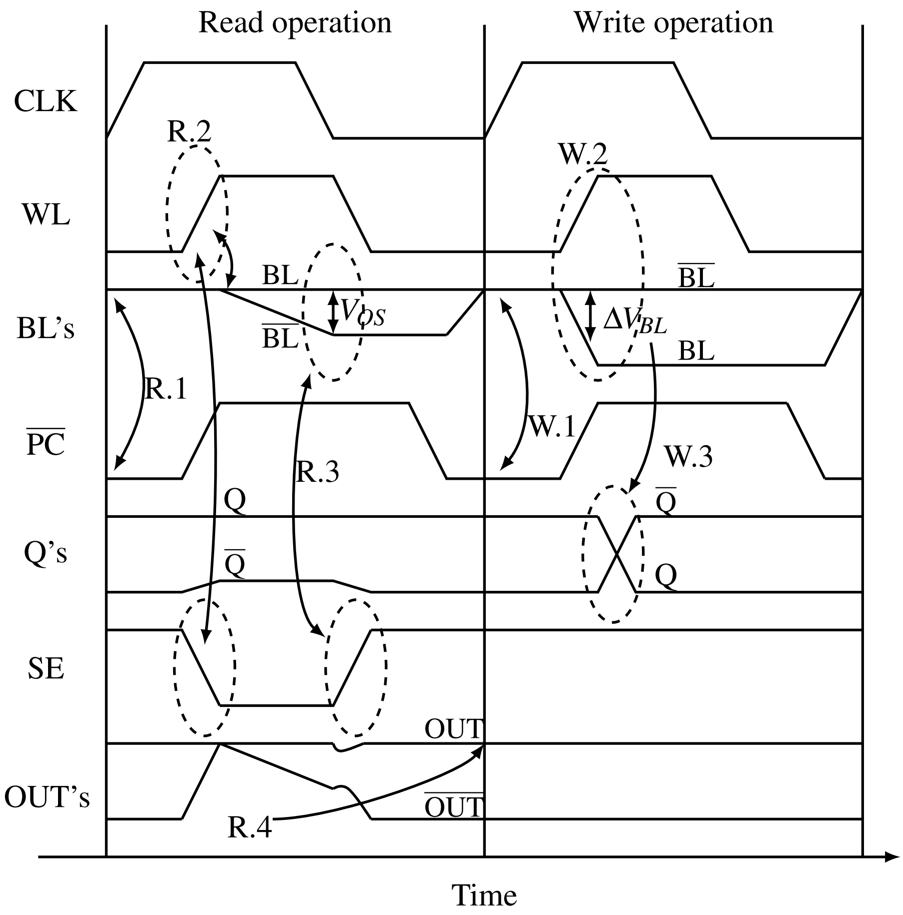


Figure 1.9: Read and write operations timing diagram.

pJ/word-access or fJ/bit-access units, which refers to the average energy consumption per accessed word or bit, respectively, in active mode (write or read operations).

Table 1.1: Performance comparisons among reported ULP SRAM designs.

Parameter	[42]	[43]	[44]
Technology	28 nm FD-SOI		20 nm Bulk
Minimum supply voltage (V)	0.36	0.35	0.6
Bit-cell area (μm^2)	0.232	0.384	N/A
Memory Size (kb)	128	64	128
Word size (bits)	64	32	32
Freq.(MHz) @ $V_{DD,min}$	9	13	N/A
Energy (pJ/word-access)	3.36 @0.45V	1.15 @0.35V	2.15 @0.6V
Energy (pJ/word-access)*	3.36 @0.45V	4.60 @0.35V	4.30 @0.6V
Energy (fJ/bit-access)*	52.5 @0.45V	71.8 @0.35V	67.2 @0.6V

* Normalized to memory size 128 Kb and word size 64.

1.3 Thesis Contributions

Since analog circuit designs in CMOS nanometer (< 90 nm) nodes can be substantially affected by manufacturing process variations, circuit performance becomes more challenging to achieve efficient solutions by using analytical models. Extensive simulations are thus commonly required to provide a high yield. On the other hand, due to the fact that the classical bulk MOS structure is reaching scaling limits (< 32 nm), other transistors are being developed as successors, such as fully depleted silicon-on-insulator (FD-SOI), Multigate MOSFET, FinFETs, among others, and new design techniques emerge by taking advantage of the improved features on devices. This improvement gives us the chance to evaluate simpler and well-known circuits that were used in micrometer processes (> 90 nm), when the short channel effects were not an obstacle, such as the 7T-LTSA and 6T-SRAM cells. In that way, this thesis focused on the development of analytical expressions for the major performance parameters of the SRAM cache implemented in 28 nm FD-SOI CMOS, mainly to explore the transistor dimensions at low computational cost, so that producing efficient designs in terms of energy consumption, speed and yield. By taking advantage of both low computational cost and close agreement results of the developed models, in this thesis we were able to propose a non-traditional sizing for the simple rectangular-diffusion (RD) 6T cell. As a result of this sizing exploration, we obtained an excellent energy per operation consumption compared with the state-of-the-art reported in the literature. Therefore, the contributions of this thesis can be separated in the following topics:

Hold static noise margin: In [17], a model for the HSNM margin was reported assuming that DIBL effect was negligible, and in [18] a procedure to include DIBL effect was described. In addition to what was reported in [17] and [18], in Section 3.1 and also published in [45], we propose a model that also considers body bias effect, which mainly improves the model for nanometer technologies.

Write and read static noise margins: Many analytical models that predict the statistical behavior of HSNM and RSNM have been proposed in the literature [17, 18, 20, 45]. However, to the best of our knowledge, no analytical solution for the write SNM (WSNM) has so far been reported in the literature. A common model for analyzing the tradeoffs between read stability and write ability is the N-curve [46], which provides information on both operation modes without having an analytical expression for WSNM. In Section 3.3 we present an analytical expression for the WSNM of 6T-SRAM cells at sub-threshold operation. Consequently, by observing the WSNM and the RSNM expressions, in Section 3.3.3 we propose an alternative 6T-SRAM design parameter Γ, whose role is to control the trade-offs between read and write cell margins. In addition, this parameter includes information on both well-known pull-up and cell ratios [14, 46, 47], and have opposite logarithm dependency between write and read margins. In this way, a fast statistical model, composed by the proposed WSNM and a preview for RSNM, is carried out in order to dimension 6T-SRAM cells for high yield under global and local process variations.

Energy-efficient sizing for 6T-SRAM cells: In Chapter 4 we present a complete analytical model to explore the SRAM cache performance. The major system metrics, such as the static and dynamic energies per operation, static noise margins, operation frequency and leakage currents are carefully modeled. By taking advantage of this low computational cost models, we explore the performance of a 128 kb SRAM cache implemented with the simple 6T rectangular-diffusion (RD) cell. The proposed 6T-SRAM cell sizing of Section 3.3.6.1, in which the cell transistors are designed with the same width (W_{pu}=W_{pd}=W_{at}=W_x), the same pull-up and pull-down lengths (L_{pu}=L_{pd}=L_x) and minimum access length (L_{at}=$L_{\min}$= 30 nm), in combination with the single-P-well (SPW) structure at reverse-body-biasing (RBB), achieve a minimum energy point (MEP) of 604 fJ/word-access at V_{DD}= 450 mV and operation frequency of 40 MHz. As shown in Table 1.1, our approach can reduce by 80% the energy per operation reported in [42], mainly for the reason that [42] uses the traditional "thin-cell" [48], in which the leakage current increases by increasing transistor widths to compensate the pull-up and pull-down imbalance. Thus, as also suggested in [49] for saving more that 50% of the energy consumption of digital circuits in 28 nm FD-SOI CMOS process, the increasing of transistor lengths considerably reduces the static leakage of the cell. However, this approach cannot be applied to the access transistor since it must drive high bit-lines capacitances. For this reason, we propose an energy-efficient and simple sizing that only depend on the variables L_x and W_x.

Sub-threshold operation of the sense amplifier (SA): In Chapter 5 we present analytical expressions to estimate input-offset and yield of the classical 7T-LTSA at subthreshold operation. According to what has been published in [29, 50], our contribution lies mainly on the simplification of the input-offset expression for subthreshold operation (see Eq. (5.14)), which allows the prediction of the SA yield at low computational cost.

1.4 Thesis Outline

In Chapter 2 is presented the used 28 nm FD-SOI CMOS process as well as an appropriate transistor modeling to deal with their nanometer issues, such as short and narrow channel effects. Chapter 3 develops expressions of the hold, read and write SNMs, and as consequence, proposes a statistical model to analyze the behavior of the 6T-SRAM cell under manufacturing process variations. Chapter 4 presents a complete analytical modeling of the SRAM cache, such as the bit-line and word-line capacitances, maximum frequency of operation, static leakage, access transistor leakage and static and dynamic energies. By taking advantage of this models and the other ones presented in Chapter 3 for the noise margins, we explore both cell transistor dimensions and cache architecture to optimize the performance, mainly aiming at low-power applications without losing focus on stability and speed. In Chapter 5 are presented models to estimate the sense amplifier input offset and yield for a 7T-LTSA topology operating at sub-threshold operation. Chapter 6 shows both layout implementation and simulation results of the 128 kb SRAM cache. And finally, in Chapter 7 we present the conclusions and future directions of this thesis.

Chapter 2

28 nm FD-SOI CMOS process

2.1 Advantages of SOI CMOS

Since the bulk CMOS is running into physical limits, other devices have been developed, as well as silicon-on-insulator (SOI) technology, which provides a promising low-power solution to chip implementation due to the fact that it has low capacitance and enables high-speed operation at lower supply voltage [51]. The box layer provides many advantages for SOI CMOS over the bulk CMOS technology. Therefore, regarding our objective, one the major advantages is the reduced drain-to-substrate capacitance, which helps to improve the switching speed of CMOS devices. This can be observed in Fig. 2.1, which shows the power consumption and the access time for a 4 Mb SRAM. Note that bulk CMOS can consume up to three times more than that SOI CMOS for the same speed. The latest generation of SOI technology is the fully-depleted (FD-SOI), and according to the ITRS, these are necessary to preserve the scaling law. FD-SOI devices have entered

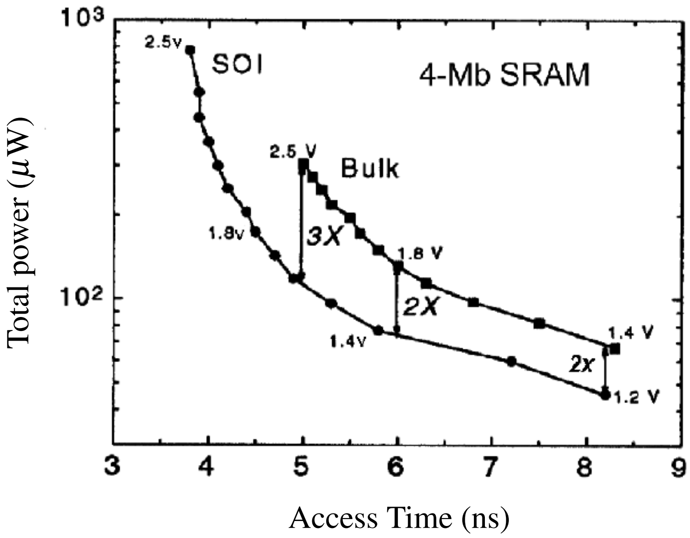


Figure 2.1: Speed and power consumption of bulk and SOI CMOS [51].

in production at the 28 nm node as an alternative to bulk CMOS, by reporting excellent short-channel electrostatic control, low leakage currents, and reduced random dopant fluctuations (RDF) [52]. Beside these performance improvements, the FD-SOI process has two important features: substrate biasing with a wide voltage range and flipped well transistors (see Fig. 2.2). By taking advantage of the flipped well features it is possible to combine N-type and P-type transistors on the same well, and hence decrease the area that is usually lost between minimum distances in the technology layout rules. In addition, flipped-well transistors have a threshold voltage that is lower than that of the standard ones, so that P-type transistors with higher specific current can be obtained at the same dimension.

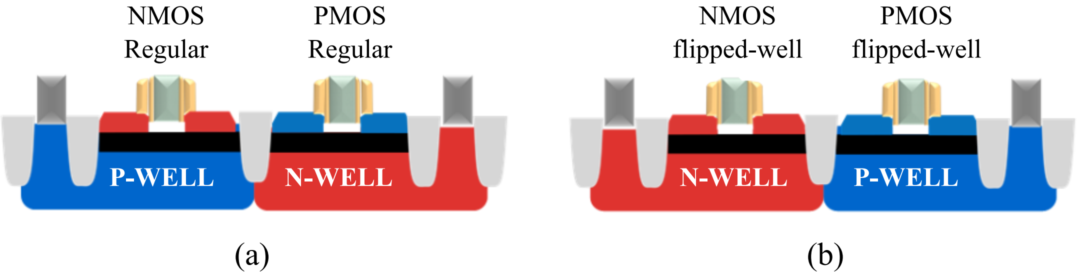


Figure 2.2: Illustration of the FDSOI structure [53]: a) standard; b) flipped well.

2.2 Transistor Models

As previously mentioned, we are interested in low power applications, in which low supply voltages are required to avoid static power dissipation. Therefore, we have studied to use similar models as we are habituated with bulk CMOS transistor for low power applications, such as [54] and [55].

In most sections of this document we used the sub-threshold transistor model. A better extraction of transistor parameters can be made by also considering the near-threshold region [56–58]. The following expression can be used to approximate the drain to source current in sub/near-threshold region [59]:

$$I_{DS} = I_{so}\frac{W}{L}\ln^2\left[1+\exp\frac{V_{GS}-V_{T,\mathrm{eff}}}{2nU_T}\right] \tag{2.1}$$

where

$$V_{T,\mathrm{eff}} = V_{To} - \lambda V_{DS} - \eta V_{BS} \tag{2.2}$$

is the effective threshold voltage, I_{so} is the specific current, V_{To} is the threshold voltage, λ is the drain-induced barrier lowering effect (DIBL), η represent the substrate biasing characteristic, n is the slope factor, U_T is the thermal voltage, and V_{GS}, V_{DS} and V_{BS} denote, respectively, the gate-to-source, drain-to-source and substrate-to-source voltages.

Parameters I_{so}, n, λ, V_{To} and η strongly depend on the width (W) and length (L) due to the influences of narrow and short channel effects. Therefore, we extracted these parameters for different values of W and L. Figs. 2.3 and 2.4 show the results of these extractions for

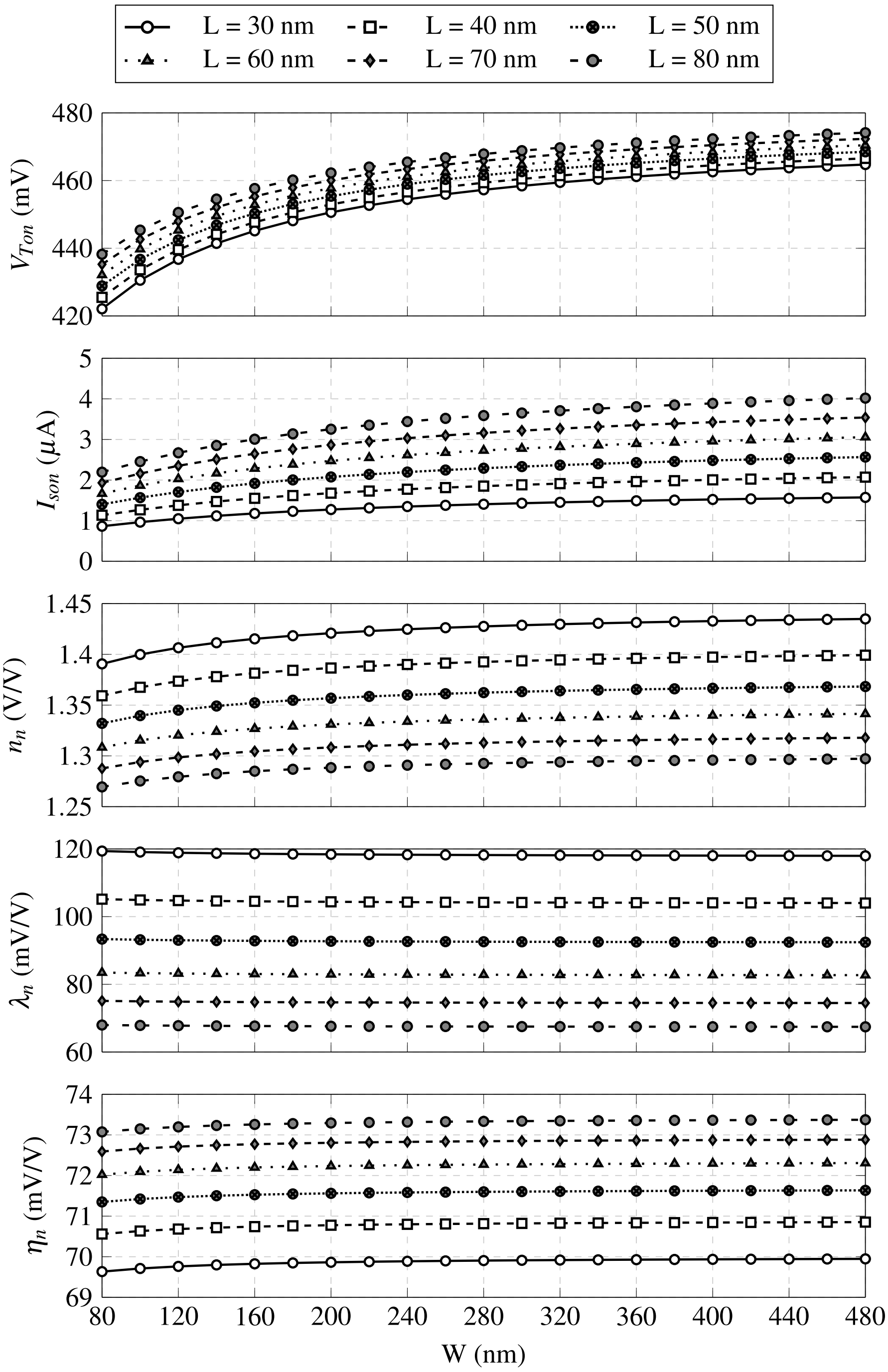


Figure 2.3: Extracted curves for regular NMOS.

regular NMOS and flipped well PMOS, respectively. The combination of regular NMOS and flipped well PMOS is used for two reasons: area savings, in view of the fact that they are fabricated on the same well, and that flipped-well PMOS has a higher gain that regular PMOS since it has a lower threshold voltage. When transistors operate in the

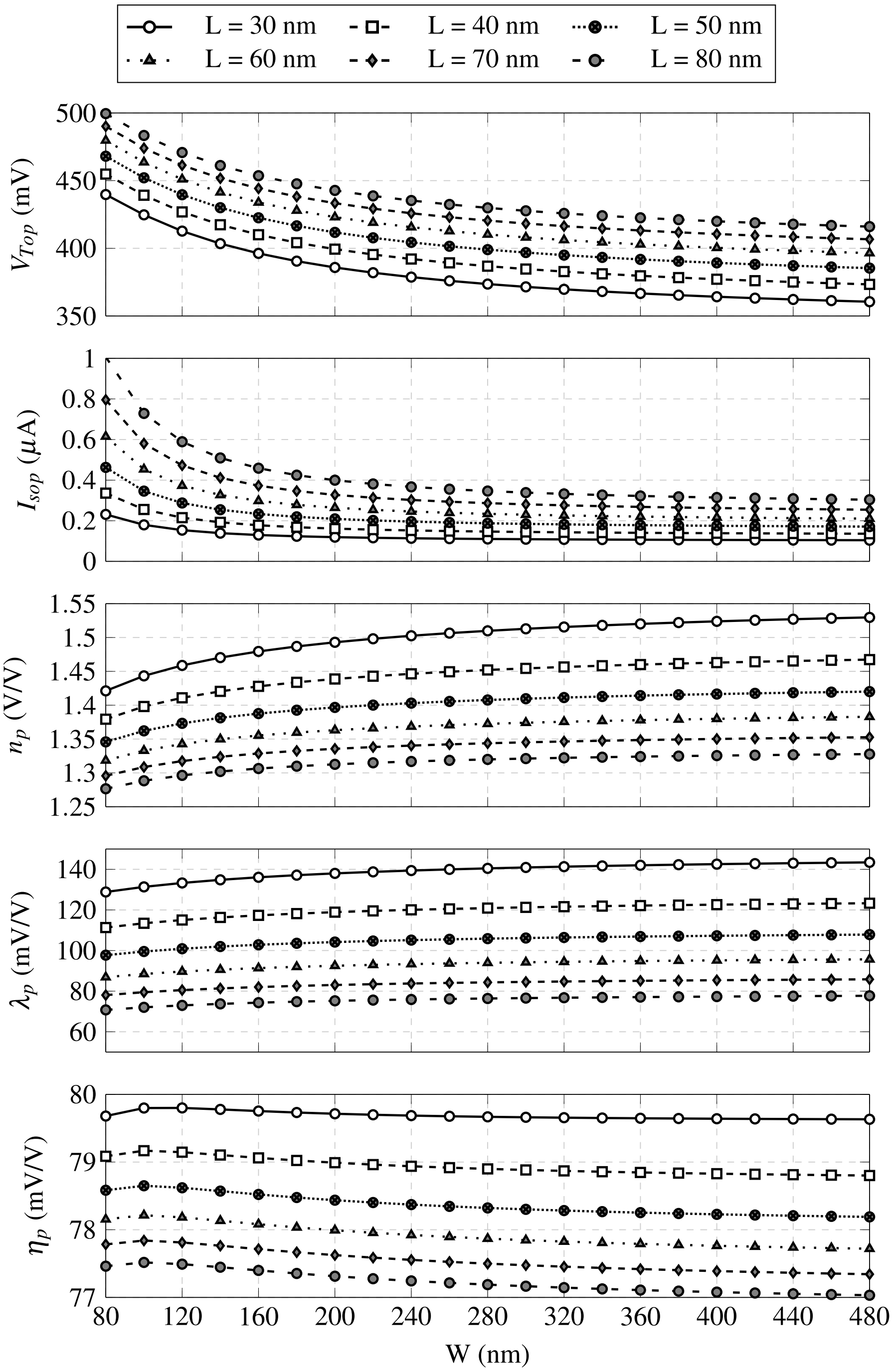
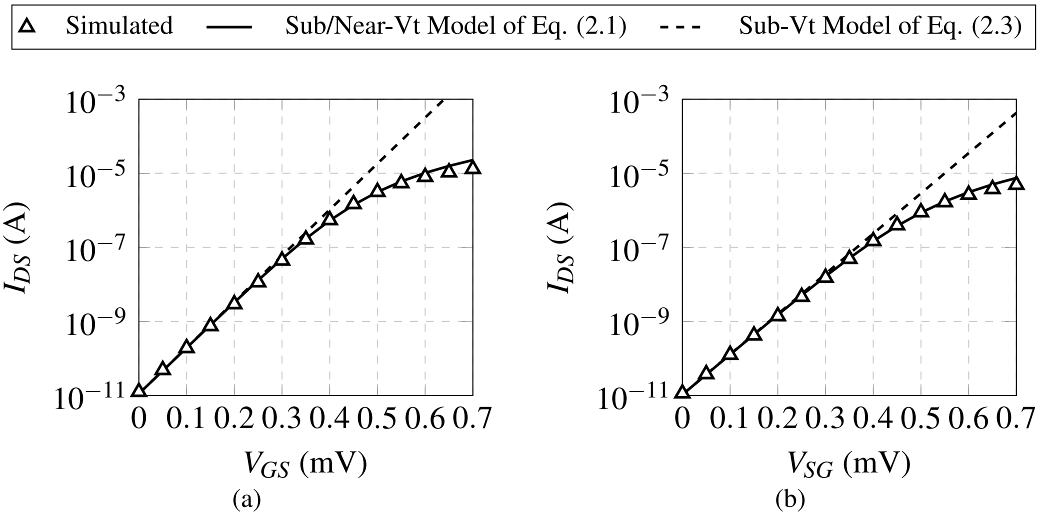


Figure 2.4: Extracted curves for flipped-well PMOS.

sub-threshold region, the drain-to-source current is given by [17, 60–62]

$$I_{DS} = I_{so}\frac{W}{L}\exp\left(\frac{-V_{To}}{nU_T}\right)\exp\left(\frac{V_{GS}+\lambda V_{DS}+\eta V_{BS}}{nU_T}\right)\left[1-\exp\left(\frac{-V_{DS}}{U_T}\right)\right]. \tag{2.3}$$

In order to show the effects of using the sub-threshold model in Eq. (2.3), Figs. 2.5(a) and 2.5(b) present the currents produced by sub/near and sub-threshold models for both regular NMOS and flipped well PMOS transistors, respectively.



(a)

(b)

Figure 2.5: Simulated drain current in comparison with both sub and near threshold models: (a) Regular NMOS (b) flipped-well PMOS.

2.2.1 Global Process Variations

Process variations, also known as global variations, affect all devices on chip in the same way [63, 64]. Usually, to predict their influences, these variations are set on the simulator at their worst-case corners, which are denoted as SS (N-slow-P-slow), FF (N-fast-P-fast), FS (N-fast-P-slow) and SF (N-slow-P-fast). In order to include this information in our analytical models, we extracted parameters I_{so}, n, λ, V_{To} and η for all corner cases by varying W and L. For instance, the extraction results when transistors are used at their minimal dimension are shown in Table 2.1. As expected, V_{To} and I_{so} suffer the major process variations, whereas V_{To} variations are the most important in view of the fact V_{To} has an exponential dependence on the sub-threshold current given by Eq. (2.3).

Table 2.1: Extraction results from process corner variations for transistor dimensions W_n=W_p= 80 nm and L_n=L_p= 30 nm.

Parameter	TT	FS	SF	FF	SS
V_{Ton} (mV)	422.3	391.9	453.0	404.5	437.4
I_{son} (μA)	2.3	1.89	2.9	2.7	1.93
n_n (V/V)	1.38	1.37	1.40	1.44	1.36
λ_n (mV/V)	118.7	118.8	118.6	135.2	111.5
η_n (mV/V)	69.7	68.8	70.4	66.9	71.7
V_{Top} (mV)	440.4	470.0	410.7	415.8	473.1
I_{sop} (μA)	0.62	0.72	0.54	0.61	0.73
n_p (V/V)	1.42	1.42	1.41	1.47	1.37
λ_p (mV/V)	128.0	127.6	128.5	153.5	107.2
η_p (mV/V)	80.0	80.3	78.9	78.2	82.2

2.2.2 Local Mismatch Variations

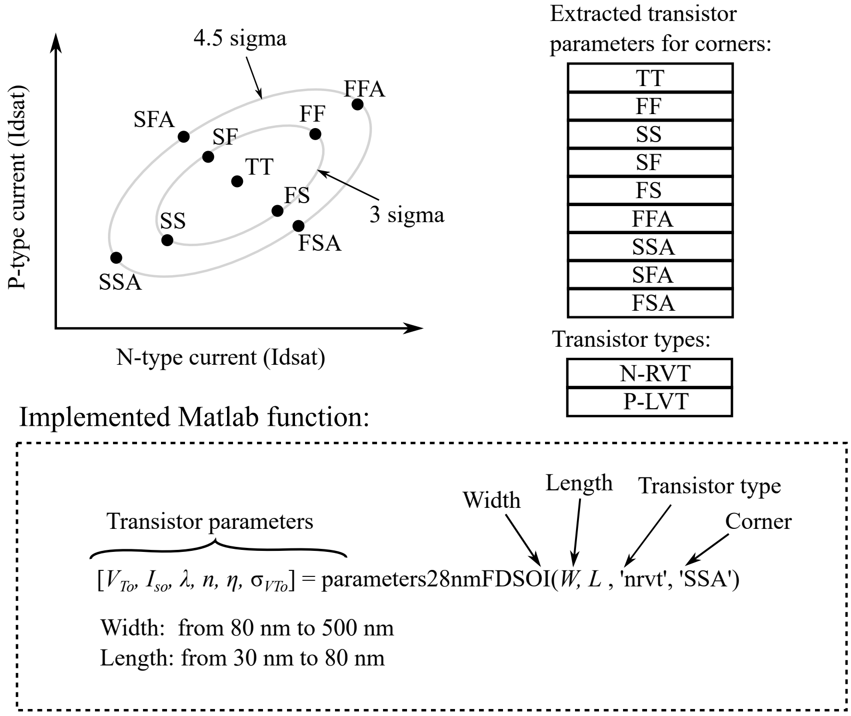
In most sections of this thesis we applied the Pelgrom's model [65, 66] to estimate the threshold voltage variations due to the influences of mismatching. This model defines the standard deviation of the threshold voltage as

$$\sigma_{V_{To}} = \frac{A_{VTo}}{\sqrt{WL}} \tag{2.4}$$

where A_{VTo} is a technology dependent parameter. In this technology, parameters $A_{VTo,n}$ and $A_{VTo,p}$ are approximately 1.23 mV·μm for both N-type and P-type transistors.

2.2.3 Matlab Function for Transistor Parameters

In addition to the classical corners (FF, SS, FS and SF), the 28 nm FD-SOI CMOS technology provides larger spread corners (FFA, SSA, FSA and SFA), which are obtained at 4.5σ. As depicted in Fig. 2.6, we implemented a Matlab function that uses the extracted results, previously showed in Fig. 2.3, and provides for our scripts the interpolated values of transistor parameters V_{To}, I_{So}, λ, η and $\sigma_{V_{To}}$ for the desired dimension within the extracted range (30 nm<L<80 nm and 80 nm<W<500 nm). In addition, the function called as "*parameters28nmFDSOI*" can be set for any corner point (FF, SS, FS, SF, FFA, SSA, FSA and SFA) for both N-RVT (NMOS regular V_T) and P-LVT (PMOS low V_T).



```
[V_To, I_so, λ, n, η, σ_VTo] = parameters28nmFDSOI(W, L , 'nrvt', 'SSA')
```

Width: from 80 nm to 500 nm
Length: from 30 nm to 80 nm

Figure 2.6: Diagram of the implemented Matlab function that provides the transistor parameters.

Chapter 3

SNM of 6T-SRAM Cells at Sub-threshold Operation

The usual metrics to quantify SRAM cell robustness at low supply voltage is the static noise margin (SNM) [14, 15], since its positive values allow safe cell operation. CMOS scaling down below 90 nm nodes creates a major challenge for designers, as process variability and short channel effects have a strong impact on the minimum supply voltage for proper cell operation [67]. Hence, extensive simulations on the design stage are required to ensure high yield of SRAM cells. Design methodologies based on either analytical models with low computational cost [15] or more efficient simulations [33, 34] are very useful in improving time to market. Many analytical models that predict the statistical behavior of hold and read SNM (HSNM and RSNM) have been proposed in the literature [20, 45]. However, to the best of our knowledge, no analytical solution for the write SNM (WSNM) has so far been reported in the literature. A common model for analyzing the tradeoff between read stability and write ability is the N-curve [46], which provides information on both operation modes without having an analytical expression for WSNM. Therefore, Section 3.1 presents a simple and yet highly accurate model for HSNM that considers DIBL and body bias effects, which is useful for both bulk CMOS and FD-SOI CMOS technologies. In [17], a model for the HSNM margin was presented assuming that DIBL effect was negligible, and in [18] a procedure to include DIBL effect was described. In addition to what was reported in [17] and [18], we propose a model that also considers body bias effect [45], which is crucial for FD-SOI CMOS processes. In Section 3.3, we present an analytical expression for the WSNM of 6T-SRAM cells at sub-threshold operation. Consequently, by observing the WSNM and RSNM expressions, we propose an alternative 6T-SRAM design parameter Γ, which includes information of both pull-up ratio [46] and cell ratio [14], and have opposite logarithm dependency between write and read margins. In addition, a fast statistical model, composed by the proposed WSNM and a preview for RSNM, is carried out in order to dimension 6T-SRAM cells for high yield under global and local process variations.

3.1 HSNM Modeling

In hold operation, the cell is only retaining the digital data through the positive feedback of their cross-coupled inverters. In this case, the effect of access transistors can be neglected, and the cell is reduced to analyze the static behavior of the inverters.

3.1.1 Static Behavior of the Inverter

By considering that the inverter circuit of Fig. 3.1(a) operates in sub-threshold region, and according to Eq. (2.3), the current through N and P type transistors are given, respectively, by

$$I_{DSn} = I_{Sn}\exp\left(\frac{V_{in} - V_{Ton} + \lambda_n V_{out} + \eta_n V_{Bn}}{n_n U_T}\right)\left(1-\exp\left(\frac{-V_{out}}{U_T}\right)\right) \tag{3.1}$$

$$\begin{aligned} I_{SDp} = I_{Sp}\exp&\left(\frac{V_{DD} - V_{in} - |V_{Top}| + \lambda_p(V_{DD} - V_{out}) + \eta_p(V_{DD} - V_{Bp})}{n_p U_T}\right) \\ &\times\left(1-\exp\left(\frac{V_{out} - V_{DD}}{U_T}\right)\right) \end{aligned} \tag{3.2}$$

where $I_{Sn} = I_{son}W_n/L_n$, $I_{Sp} = I_{sop}W_p/L_p$, and V_{Bn} and V_{Bp} are the substrate-to-source voltages of N-type and P-type transistors, respectively. The static analysis can be made when there is no output current [59], and hence by equaling Eq. (3.1) to (3.2). Therefore, the voltage transfer curve (VTC) of the inverter circuit illustrated in Fig. 3.1(b), which takes into account DIBL and body bias effects, can be determined by solving I_{DSn}=I_{SDp}, yielding

$$V_{in} = V_{OFF} + \frac{\bar{n}U_T}{2}\log\left[\frac{1-\exp\left(\frac{V_{out}-V_{DD}}{U_T}\right)}{1-\exp\left(\frac{-V_{out}}{U_T}\right)}\right] + \bar{\lambda}\frac{V_{DD}}{2} - \bar{\lambda}V_{out} \tag{3.3}$$

where

$$\begin{aligned} V_{OFF} =& \frac{\bar{n}V_{DD}}{2}\left[\frac{2+2\eta_p+\lambda_p}{2n_p} - \frac{\lambda_n}{2n_n}\right] + \frac{\bar{n}U_T}{2}\log\left(\frac{I_{Sp}}{I_{Sn}}\right) \\ &+ \frac{\bar{n}V_{Ton}}{2n_n} - \frac{\bar{n}|V_{Top}|}{2n_p} - \frac{\bar{n}\eta_p V_{Bp}}{2n_p} - \frac{\bar{n}\eta_n V_{Bn}}{2n_n} \end{aligned} \tag{3.4}$$

is the input voltage that produces half the supply voltage at the output, and

$$\bar{n} = 2\frac{n_p n_n}{n_p + n_n} \tag{3.5}$$

$$\bar{\lambda} = \frac{\bar{n}}{2}\left(\frac{\lambda_p}{n_p} + \frac{\lambda_n}{n_n}\right)\,. \tag{3.6}$$

Fig. 3.2 shows VTC simulation results compared with the analytical ones in Eq. (3.3) by setting V_{Bn}=V_{Bp}=V_{BB} in three cases: ground, half supply voltage and supply voltage. Note in Fig. 3.2 that by decreasing the substrate voltage V_{BB} we can compensate the VTC curve due to the fact that stronger P-type transistors are obtained.

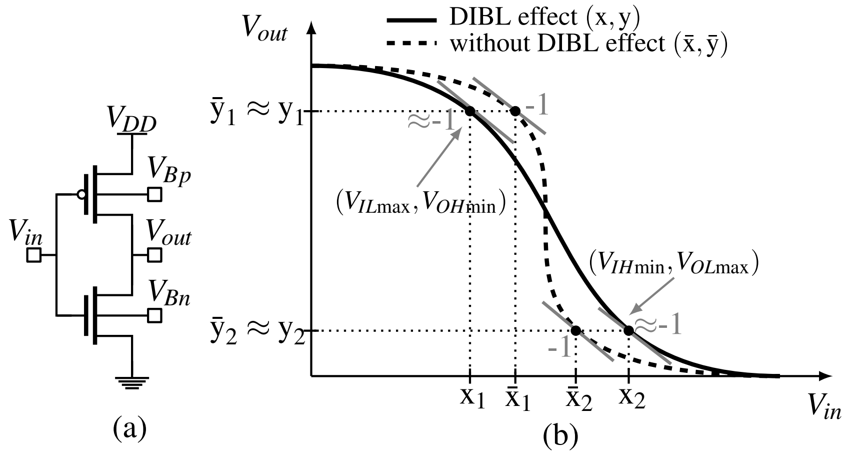


Figure 3.1: Schematic (a) and VTC illustrative diagram (b) of the inverter circuit.

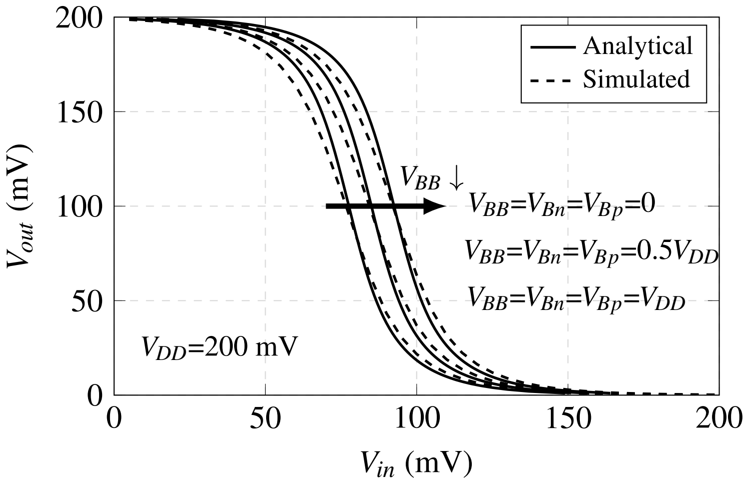


Figure 3.2: Simulation results of VTC curves *versus* predicted ones by Eq. (3.3) for various substrate voltages.

3.1.2 HSNM Considering DIBL and Body Biasing effects

In order to improve the HSNM accuracy prediction in comparison with other models presented in the literature, we include the DIBL and body biasing effects. According to [17], the low and high noise margins are defined, respectively, as

$$\mathrm{NM_L} = V_{IL\mathrm{max}} - V_{OL\mathrm{max}} \tag{3.7}$$

$$\mathrm{NM_H} = V_{OH\mathrm{min}} - V_{IH\mathrm{min}} \tag{3.8}$$

where, V_{OH} and V_{IL} denote, respectively, the high and low nominal output voltages obtained with full-swing input, whereas the pairs ($V_{IL\text{max}}$,$V_{OH\text{min}}$) and ($V_{IH\text{min}}$,$V_{OL\text{max}}$) are the unity-gain points, as illustrated in Fig. 3.1(b). The voltage $V_{OH\text{min}}$ is found by equaling to -1 the derivative of Eq. (3.3) with respect to V_{out} and assuming $\exp\left(-V_{out}/U_T\right) \ll 1$ for high values of V_{out}, yielding

$$V_{OH\text{min}} = \text{y}_1 = V_{DD} - U_T \log\left(\frac{2\bar{\lambda} - 2 - \bar{n}}{2\bar{\lambda} - 2}\right) . \tag{3.9}$$

Similarly, by equaling to -1 the derivative of Eq. (3.3) and assuming $\exp\left((V_{DD} - V_{out})/U_T\right) \gg 1$ for low values of V_{out}, we obtain

$$V_{OL\text{max}} = \text{y}_2 = U_T \log\left(\frac{2\bar{\lambda} - 2 - \bar{n}}{2\bar{\lambda} - 2}\right) . \tag{3.10}$$

The values of V_{out} concerning the unity-gain points (y_1 and y_2) can be approximated by the other ones without DIBL effects ($\bar{\text{y}}_1$ and $\bar{\text{y}}_2$), since the errors defined as

$$\varepsilon_{\text{y}_1} = \text{y}_1 - \bar{\text{y}}_1 = -U_T \log\left(\frac{2+\bar{n}}{2+\bar{n}-2\bar{\lambda}}\right) \tag{3.11}$$

$$\varepsilon_{\text{y}_2} = \text{y}_2 - \bar{\text{y}}_2 = U_T \log\left(\frac{2+\bar{n}}{2+\bar{n}-2\bar{\lambda}}\right) \tag{3.12}$$

are small enough for realistic values of $\bar{\lambda}$ and $\bar{n}$ (see Fig. 3.3). Therefore,

$$V_{OH\text{min}} \approx \bar{\text{y}}_1 = V_{DD} - U_T \log\left(\frac{2+\bar{n}}{2}\right) \tag{3.13}$$

$$V_{OL\text{max}} \approx \bar{\text{y}}_2 = U_T \log\left(\frac{2+\bar{n}}{2}\right) \tag{3.14}$$

which agree with [17]. Now, by replacing the approximation of Eq. (3.13) with the VTC in Eq. (3.3) (V_{out} = $V_{OH\text{min}}$), thereby accounting for all effects, we find

$$V_{IL\text{max}} = \text{x}_1 = \frac{\bar{n}U_T}{2} \log\left(\frac{\bar{n}}{\bar{n}+2}\right) - \bar{\lambda} U_T \log\left(\frac{2}{\bar{n}+2}\right) + V_{OFF} - \frac{\bar{\lambda} V_{DD}}{2} . \tag{3.15}$$

In addition, replacing Eq. (3.14) with (3.3) yields

$$V_{IH\text{min}} = \text{x}_2 = -\frac{\bar{n}U_T}{2} \log\left(\frac{\bar{n}}{\bar{n}+2}\right) + \bar{\lambda} U_T \log\left(\frac{2}{\bar{n}+2}\right) + V_{OFF} + \frac{\bar{\lambda} V_{DD}}{2} . \tag{3.16}$$

Finally, the $\mathrm{NM_L}$ and $\mathrm{NM_H}$ defined by Eqs. (3.7) and (3.8) are given, respectively, by

$$\mathrm{NM_L} = \frac{\bar{n}U_T}{2}\log\left(\frac{\bar{n}}{\bar{n}+2}\right) + (1-\bar{\lambda})U_T\log\left(\frac{2}{\bar{n}+2}\right) + V_{OFF} - \frac{\bar{\lambda}V_{DD}}{2} \tag{3.17}$$

$$\mathrm{NM_H} = \frac{\bar{n}U_T}{2}\log\left(\frac{\bar{n}}{\bar{n}+2}\right) + (1-\bar{\lambda})U_T\log\left(\frac{2}{\bar{n}+2}\right) + V_{DD} - V_{OFF} - \frac{\bar{\lambda}V_{DD}}{2}\,. \tag{3.18}$$

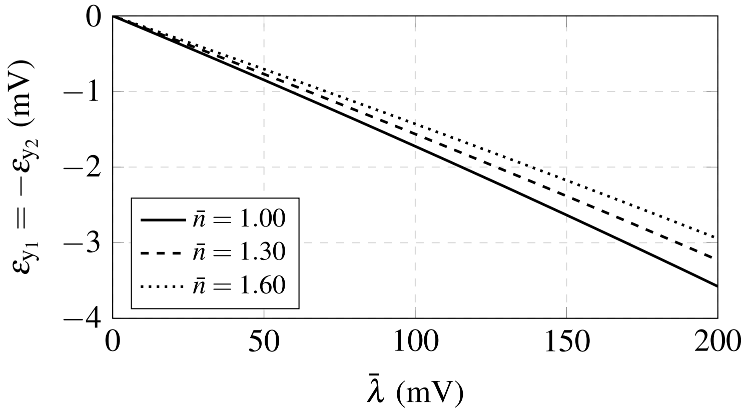


Figure 3.3: Approximation error of Eqs. (3.11) and (3.12).

The errors produced by the approximations made in Eqs. (3.13) and (3.14) on the $\mathrm{NM_L}$ and $\mathrm{NM_H}$, respectively, can be analyzed by assuming that small errors on the vertical coordinates (ε_y) generate an opposite error on the horizontal coordinates ($\varepsilon_x \approx -\varepsilon_y$), since the derivative values are close to -1. Therefore, from Eqs. (3.7) and (3.8), it follows that the absolute errors of the low and high noise margins can be approximated by

$$\varepsilon_{\mathrm{NM_L}} = \varepsilon_{x_1} - \varepsilon_{y_2} \approx -\varepsilon_{y_1} - \varepsilon_{y_2} \tag{3.19}$$

$$\varepsilon_{\mathrm{NM_H}} = \varepsilon_{y_1} - \varepsilon_{x_2} \approx \varepsilon_{y_1} + \varepsilon_{y_2} \tag{3.20}$$

which are approximately zero in view of Eqs. (3.11) and (3.12).

3.1.3 Global and Local Process Variations

The global corners FS (fast-N-type slow-P-type) and SF (slow-P-type fast-P-type) in combination with local variations are determinant for ensuring a positive noise margin under manufacturing process variations [68]. As illustrated in Fig. 3.4, the positive values of Eqs. (3.17) and (3.18) determine a region where the threshold voltage variations produce a feasible design with which the inverter operates safety. It should also be observed that the limits established by Eqs. (3.17) and (3.18) are closer to FS and SF corner points, respectively, than to the other corner points (FF ans SS).

Assuming that process variations only affect V_{To}, the local mismatch of the noise margin is

$$\Delta\mathrm{NM} = \Delta\mathrm{NM_L} = -\Delta\mathrm{NM_H} = \frac{\bar{n}}{2}\left(\frac{\Delta V_{Ton}}{n_n} - \frac{\Delta V_{Top}}{n_p}\right) \tag{3.21}$$

whose variance is

$$\sigma_{\mathrm{NM}}^2 = \frac{\bar{n}^2}{4}\left(\frac{\sigma_{V_{Ton}}^2}{n_n^2} + \frac{\sigma_{V_{Top}}^2}{n_p^2}\right) \tag{3.22}$$

where $\sigma_{V_{Ton}}$ and $\sigma_{V_{Top}}$ are the standard deviations of the threshold voltages for N-type and P-type transistors, respectively. In order to provide a robust cell operation under the process variations, we estimate the worst-case of the HSNM by combining the corner points and their local variations as

$$\mathrm{HSNM_{WC}} = \min(\mathrm{NM_{L,FS}} - 3\sigma_{\mathrm{NM}}, \mathrm{NM_{H,SF}} - 3\sigma_{\mathrm{NM}}) \tag{3.23}$$

where $\mathrm{NM_{L,FS}}$ and $\mathrm{NM_{H,SF}}$ are given by Eq. (3.17) at the FS corner and by Eq. (3.18) at the SF corner, respectively. Note that Eq. (3.23) can be used to estimate the worst-case only by knowing the global parameters at both FS and SF corners, and the local mismatch which depends on the transistor area. Hence, low computational cost estimation can be achieved in order to avoid the extensive simulations that are usually required to provide high-yield designs.

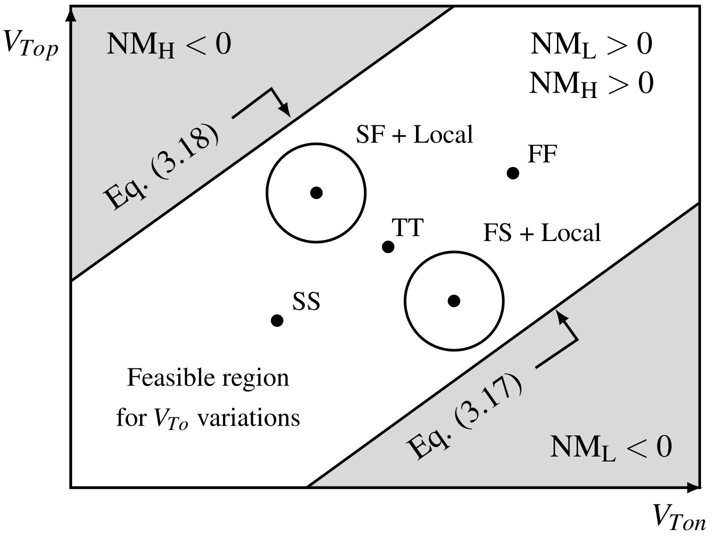
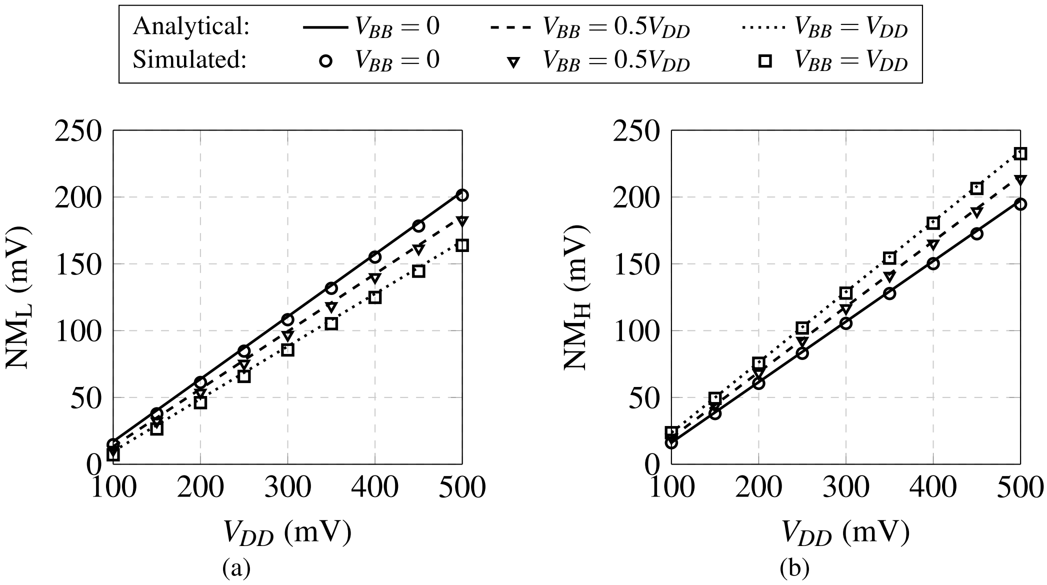


Figure 3.4: White region limited by straight lines (Eqs. (3.17) and (3.18)) pertaining to feasible designs under V_{To} variations.

3.1.4 Model Results in 28 nm FD-SOI CMOS

A design example was carried out in 28 nm FD-SOI CMOS process and validated through HSPICE simulations. As seen in Section 3.3.6, the single-P-well (SPW) structure of

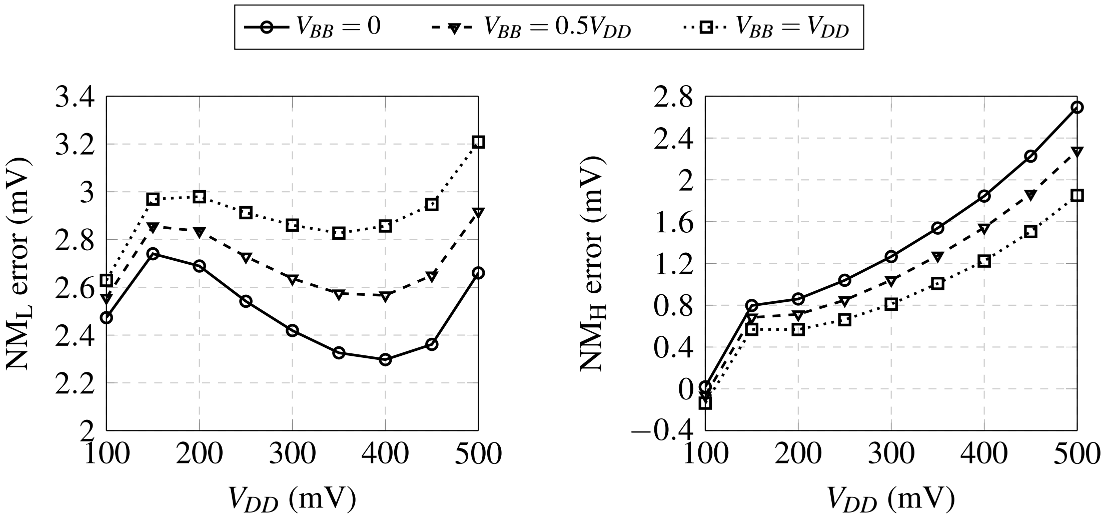
Fig. 3.16 was used to implement the circuit. Figs. 3.5(a) and 3.5(b) present the nominal simulation results of the low and high noise margins, respectively, compared with the predicted ones shown in Eqs. (3.17) and (3.18). The circuit was simulated by setting the substrate voltage (V_{BB}) at three values: ground, half supply voltage, and supply voltage. The absolute error results (Figs. 3.6(a) and 3.6(b)) are less than 3.2 mV along the supply voltage range, showing that the results produced by the analytical model are in close agreement with the simulation results. In order to validate the analytical worst-case noise margin in Eq. (3.23), in Fig. 3.7, we compared it with the worst-case obtained by 1000 Monte Carlo runs including global and local variations. The minimum supply voltage that provides a positive noise margin is achieved in the case that V_{BB}=0, for which the analytical value is 212.98 mV compared to 214.92 mV obtained by simulations. The runtime of the worst-case estimation obtained from Eq. (3.23) takes approximately 0.7 seconds, whereas Monte Carlo simulations in HSPICE require 4.5 minutes to conclude.



(a)

(b)

Figure 3.5: Simulation results of $\mathrm{NM_L}$ and $\mathrm{NM_H}$ compared with the predicted ones obtained from the respective analytical model in Eqs. (3.17) and (3.18): (a) nominal $\mathrm{NM_L}$ (b) nominal $\mathrm{NM_H}$.



(a) (b)

Figure 3.6: Absolute error of the analytical models of $\mathrm{NM_L}$ and $\mathrm{NM_H}$ compared with simulations: (a) absolute error of $\mathrm{NM_L}$ (b) absolute error of $\mathrm{NM_H}$. The absolute error is calculated as the difference between analytical and simulated results.

3.2 RSNM Modeling

Fig. 3.8 shows the 6T-SRAM circuit during the read mode. A complete analytical model for RSNM at sub-threshold operation was reported in [20] as

$$\mathrm{RSNM}=\frac{\begin{pmatrix} n_5(n_1n_2-n_1n_4\lambda_2+n_{ro}(n_2+n_4))V_{DD}-n_{ro}n_1(n_2+n_4)V_{WL} \\ +n_{ro}(n_2+n_4)(n_1V_{To5}-n_5V_{To1})+n_5n_1(n_4V_{To2}-n_2|V_{To4}|) \\ +n_5(n_1n_4(\lambda_2-n_2)-n_{ro}(n_2+n_4))\log(2)U_T \\ +n_{ro}n_5n_1(n_2+n_4)\log\left(\dfrac{I_{S1}}{I_{S5}}\right)U_T \\ +n_5n_1n_2n_4\log\left(\dfrac{I_{S4}}{I_{S2}}\right)U_T \end{pmatrix}}{n_5(n_{ro}+n_1)(n_2+n_4)} \tag{3.24}$$

where

$$n_{ro}=\frac{n_1n_5}{n_1(\lambda_5+\eta_5+1)+n_5\lambda_1} \tag{3.25}$$

and I_{Si}=$I_{so,i}W_i/L_i$, for i= 1,2,4,5. Work [20] also analyzed the statistical behavior of RSNM by considering that threshold voltages are independent random variables. Hence the mean

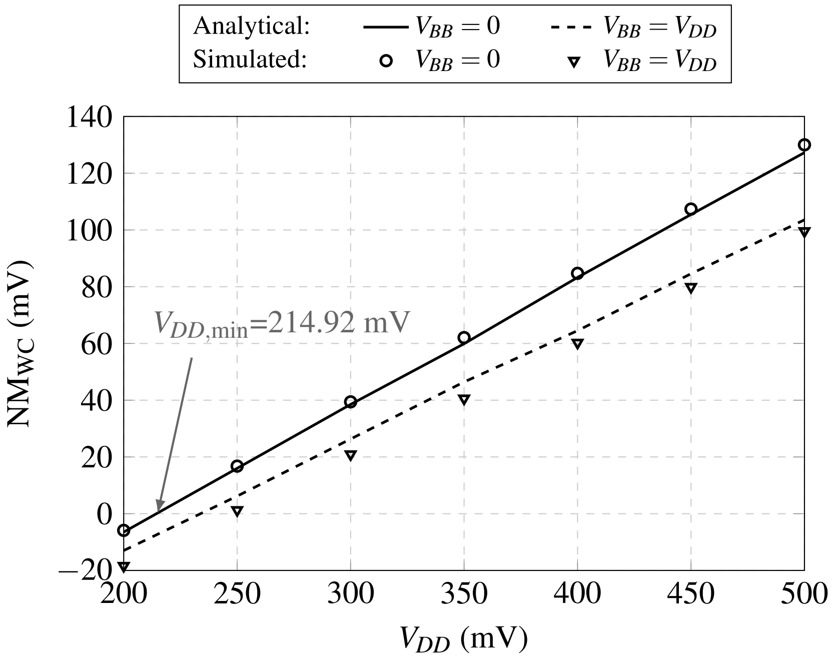


Figure 3.7: Noise margin obtained as the worst-case of 1000 samples of Monte Carlo simulations compared with the other ones predicted by Eq. (3.23).

and standard deviation of RSNM can be derived, respectively, as

$$\mu_{\mathrm{RSNM}} = \mathrm{RSNM} + \sum_{i=1}^{6} \left(\frac{1}{2} \frac{\partial \mathrm{RSNM}^2}{\partial V_{Toi}^2} \right) \sigma_{V_{Toi}}^2 \tag{3.26}$$

$$\sigma_{\mathrm{RSNM}} = \sqrt{\sum_{i=1}^{6} \left(\frac{\partial \mathrm{RSNM}}{\partial V_{Toi}} \sigma_{V_{Toi}} \right)^2} \tag{3.27}$$

which, considering linear dependency on threshold voltages, result

$$\mu_{\mathrm{RSNM}} = \mathrm{RSNM} \tag{3.28}$$

$$\sigma_{\mathrm{RSNM}} = \frac{\sqrt{n_{ro}^2 n_5^2 (n_2+n_4)^2 \sigma_{V_{To1}}^2 + n_1^2 n_4^2 n_5^2 \sigma_{V_{To2}}^2 + n_1^2 n_2^2 n_5^2 \sigma_{V_{To4}}^2 + n_{ro}^2 n_1^2 (n_2+n_4)^2 \sigma_{V_{To5}}^2}}{n_5 (n_{ro}+n_1)(n_2+n_4)}. \tag{3.29}$$

This analytical modeling of RSNM [20] was an important motivation to our thesis, since to the best of our knowledge, no analytical solution for WSNM at sub-threshold operation was so far been reported in the literature. Therefore, in Section 3.3 we present our approach for modeling the WSNM.

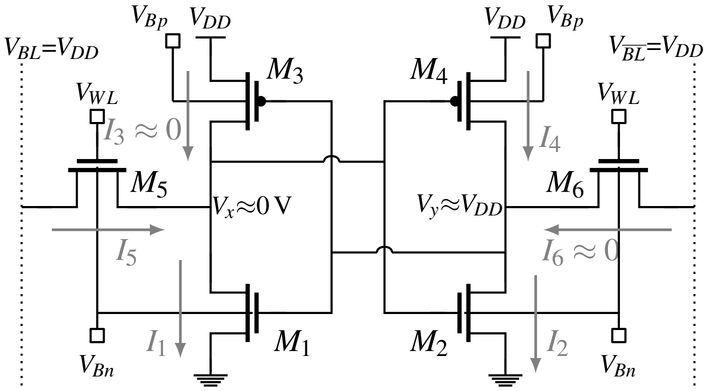


Figure 3.8: 6T-SRAM cell circuit during the read operation.

3.3 WSNM Modeling

3.3.1 WSNM Formulation

In order to simplify the handling of the equations, the drain-to-source current of CMOS transistors at sub-threshold operation, presented in Eq. (2.3), can be rewritten as

$$I_{DS} = \frac{W}{L} I_{So} \exp\big(a_n(V_{GS} - V_{To} + \lambda V_{DS} + \eta V_{BS})\big) D(V_{DS}) \tag{3.30}$$

where

$$a_n = \frac{a}{n} \tag{3.31}$$

$$a = \frac{1}{U_T} \tag{3.32}$$

$$D(X) = 1 - \exp(-aX). \tag{3.33}$$

Fig. 3.9 shows the 6T-SRAM cell circuit during the write operation, in which it is initially assumed that the internal nodes retain the voltages V_x=0 V and V_y=V_{DD}, the bit-lines are pre-charged at V_{BL}=V_{DD} and $V_{\overline{BL}}$=$V_{DD}-\Delta V_{BL}$, respectively, and the word-line is activated by V_{WL}=V_{DD}. The DC analysis of this circuit can be decomposed into two inverters (M_3-M_1 and M_4-M_2) loaded by their access transistors (M_5 and M_6). Then the voltage transfer curves $H_1(V_x, V_y)$ and $H_2(V_x, V_y)$ are determined, respectively, as

$$H_1(V_x, V_y) : I_3 + I_5 - I_1 = 0 \tag{3.34}$$

$$H_2(V_x, V_y) : I_4 - I_6 - I_2 = 0 \tag{3.35}$$

where the currents are

$$I_1 = A_1 \exp\left(a_{n1}(V_y + \lambda_1 V_x)\right) D(V_x) \tag{3.36}$$

$$I_2 = A_2 \exp\left(a_{n2}(V_x + \lambda_2 V_y)\right) D(V_y) \tag{3.37}$$

$$I_3 = A_3 \exp\left(-a_{n3}(V_y + \lambda_3 V_x)\right) D(V_{DD} - V_x) \tag{3.38}$$

$$I_4 = A_4 \exp\left(-a_{n4}(V_x + \lambda_4 V_y)\right) D(V_{DD} - V_y) \tag{3.39}$$

$$I_5 = A_5 \exp\left(-a_{n5}(1 + \lambda_5 + \eta_5)V_x\right) D(V_{DD} - V_x) \tag{3.40}$$

$$I_6 = A_6 \exp\left(a_{n6}\lambda_6 V_y\right) D(V_y - V_{\overline{BL}}) \tag{3.41}$$

and

$$A_1 = I_{S1} \exp\left(a_{n1}(\eta_1 V_{Bn} - V_{To1})\right) \tag{3.42}$$

$$A_2 = I_{S2} \exp\left(a_{n2}(\eta_2 V_{Bn} - V_{To2})\right) \tag{3.43}$$

$$A_3 = I_{S3} \exp\left(a_{n3}(-\eta_3 V_{Bp} + V_{DD}(1 + \lambda_3 + \eta_3) - |V_{To3}|)\right) \tag{3.44}$$

$$A_4 = I_{S4} \exp\left(a_{n4}(-\eta_4 V_{Bp} + V_{DD}(1 + \lambda_4 + \eta_4) - |V_{To4}|)\right) \tag{3.45}$$

$$A_5 = I_{S5} \exp\left(a_{n5}(\eta_5 V_{Bn} + (1 + \lambda_5)V_{DD} - V_{To5})\right) \tag{3.46}$$

$$A_6 = I_{S6} \exp\left(a_{n6}(\eta_6 V_{Bn} - (1 + \eta_6 + \lambda_6)V_{\overline{BL}} + V_{DD} - V_{To6})\right) \tag{3.47}$$

where I_{Si}=$(W_i/L_i)I_{Soi}$, for i= 1,2,..,6. The conventional WSNM [69] is defined as the width of the smallest embedded square between the curves H_1 and H_2 as illustrated in Fig. 3.10(a). The square width value can be found once we achieve an analytical solution for the coinciding point (P_0) between H_1 and H_2 (see Fig. 3.10(b)) as a function of the transistors parameters, since, according to [70], the DC noise voltage can be modeled on the transistors threshold voltages. The analytical solution of the coinciding point is a modeling challenge and some approximations must be made in Eqs. (3.34) and (3.35) around the initial conditions of the write operation thereby reducing the problem complexity.

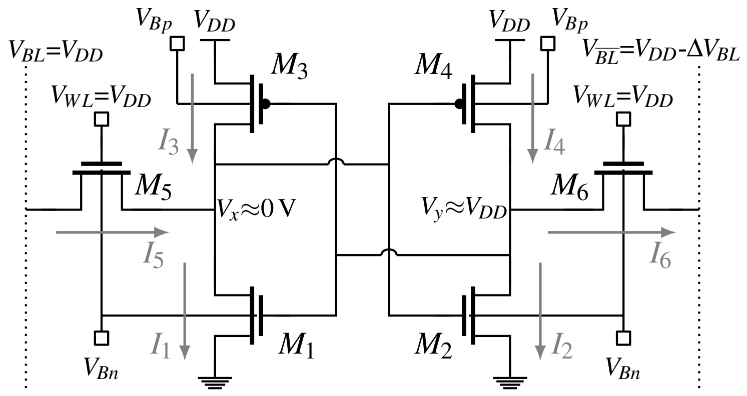


Figure 3.9: 6T-SRAM cell circuit during the write operation.

Accordingly, by considering that the cell retains V_x=0 V and V_y=V_{DD} before starting the

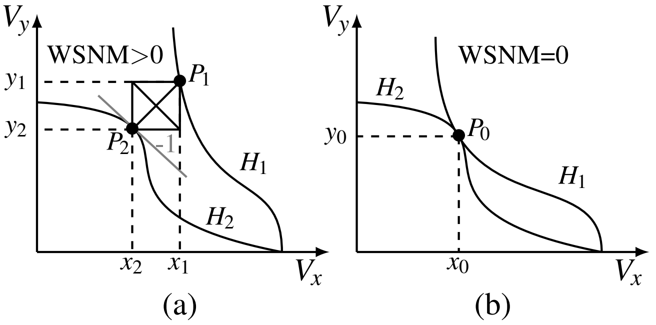


Figure 3.10: Conventional write static noise margin: (a) smallest embedded square (b) coinciding point.

write operation, current I_3 can be neglected, and as long as Δ_{BL} is slightly higher than $4U_T$ [71], the current through M_6 dominates the one through M_2, such that I_2 can also be neglected. In addition, the factors $D(V_{DD}-V_x)$ and $D(V_y-V_{\overline{BL}})$ in Eqs. (3.40) and (3.41), respectively, assume values close to 1. Therefore, using the proposed approximations and rearranging Eqs. (3.34) and (3.35), we obtain

$$V_y = \frac{1}{a_{n1}}\left[\log\left(\frac{A_5}{A_1(1-\exp(-aV_x))}\right) - a_0 V_x\right] \tag{3.48}$$

$$V_x = \frac{1}{a_{n4}}\left[\log\left(\frac{A_4(1-\exp(-aV_{DD}+aV_y))}{A_6}\right) - a_1 V_y\right] \tag{3.49}$$

where

$$a_0 = a_{n1}\lambda_1 + a_{n5}(1+\eta_5+\lambda_5) \tag{3.50}$$

$$a_1 = a_{n4}\lambda_4 + a_{n6}\lambda_6 \quad . \tag{3.51}$$

The behavior of Eqs. (3.48) and (3.49) are shown in Fig. 3.11 which also includes simulation results. It should be observed that the analytical curves are in close agreement with the simulated ones around the region of interest, in which V_x and V_y have low and high voltage values, respectively. Furthermore, Eq. (3.48) can be simplified around P_0 as

$$V_y = \frac{1}{a_{n1}}\left[\log\left(\frac{A_5}{A_1}\right) - a_0 V_x\right] . \tag{3.52}$$

Now, assuming that P_0 lies on the unity-gain point of H_2 (P_2), we observe that the coordinates x_0 and y_0 (see Fig 3.10(b)) can be obtained by equaling the derivative of Eq. (3.49) to -1, yielding

$$x_0 = \frac{1}{a_{n4}}\left[\log\left(\frac{A_4}{A_6}\right) + \log\left(\frac{n_4}{n_4+1}\right) - a_1 V_{DD}\right] \tag{3.53}$$

$$y_0 = V_{DD} - U_T \log(1+n_4) \quad . \tag{3.54}$$

Finally, substituting Eqs. (3.53) and (3.54) into (3.52), we obtain the condition for the coinciding point as a function of the transistor parameters and the supply voltage as

$$\begin{aligned}\frac{1}{a_0}\left[\log\left(\frac{A_5}{A_1}\right) - a_{n1}V_{DD} + \frac{a_{n1}}{a}\log(1+n_4)\right] = \\ \frac{1}{a_{n4}}\left[a_1V_{DD} - \log\left(\frac{A_4}{A_6}\right) - \log\left(\frac{n_4}{1+n_4}\right)\right].\end{aligned} \tag{3.55}$$

The WSNM can be quickly formulated from Eq. (3.55) by modeling the DC noise margin on the threshold voltages [70] of transistors M_1 and M_4 as "$V_{To1}-$WSNM" and "$|V_{To4}|-$WSNM", respectively. Therefore, the final expression for the WSNM gives

$$\begin{aligned}\mathrm{WSNM} = (a_0+a_{n1})^{-1}\left[\log\left(\frac{A_5}{A_1}\right) + \frac{a_0}{a_{n4}}\log\left(\frac{A_6}{A_4}\right) - \frac{a_0}{a_{n4}}\log\left(\frac{n_4}{n_4+1}\right)\right. \\ \left. + \frac{a_{n1}}{a}\log(n_4+1) - V_{DD}\left(a_{n1} - \frac{a_0a_1}{a_{n4}}\right)\right].\end{aligned} \tag{3.56}$$

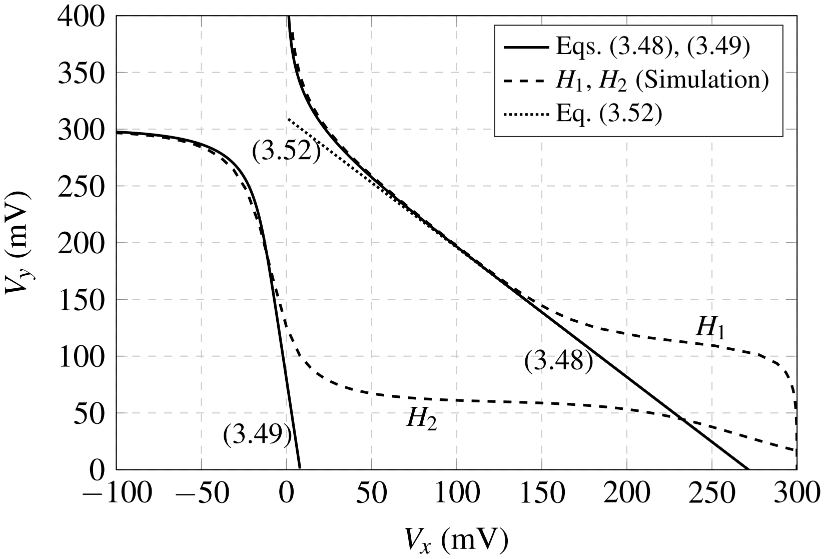


Figure 3.11: Behavior of the analytical equations presented in Eqs. (3.48), (3.49) and (3.52) compared with simulation results of the voltage transfer curves H_1 and H_2.

3.3.2 Dependency on Supply and Bit-line Voltages

Fig. 3.12 shows the WSNM as a function of the supply voltage for various ΔV_{BL} values. The analytical curves are in close agreement with the simulated ones. However, it can be observed that the model slightly underestimates the results obtained by simulation, owing to the fact that we approximated Eq. (3.48) as (3.52).

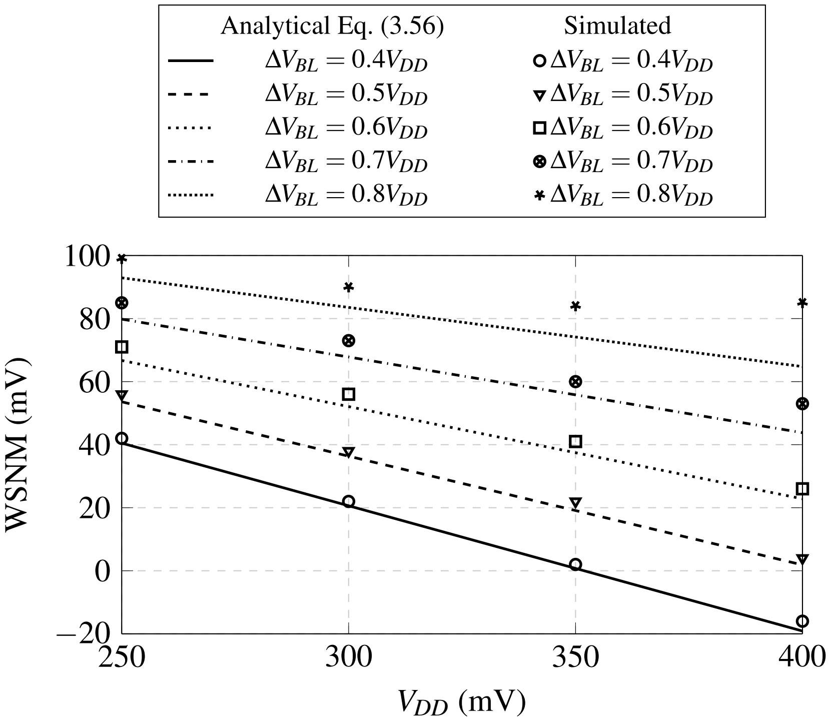


Figure 3.12: WSNM as a function of the supply voltage.

3.3.3 Dependency on 6T-SRAM Sizing

The design of SRAM cells requires a balance between write and read static noise margins. The influences of transistor dimensions on these margins can be appreciated more clearly by considering the ideal case in which all transistors have the same slope factor n, and that DIBL and body bias effects are negligible, so that Eq. (3.56) can be approximated as

$$\begin{aligned}\mathrm{WSNM} = &\frac{\Delta V_{BL} - V_{DD}}{2} + \frac{V_{To1} + |V_{To4}| - V_{To5} - V_{To6}}{2} \\ &+ \frac{nU_T}{2}\left[\log\left(\frac{I_{So6}I_{So5}}{I_{So4}I_{So1}}\right) + \frac{\log(n+1)}{n} + \log\left(\frac{n+1}{n}\right)\right] \\ &+ \frac{nU_T}{2}\log\left(\frac{W_6/L_6}{W_4/L_4}\frac{W_5/L_5}{W_1/L_1}\right)\end{aligned} \tag{3.57}$$

and the RSNM of Eq. (3.24) as

$$\begin{aligned}\mathrm{RSNM} = &\frac{3V_{DD}}{4} - \frac{V_{WL}}{2} + \frac{2V_{To5} - 2V_{To1} + V_{To2} - |V_{To4}|}{4} \\ &+ \frac{nU_T}{2}\left[\log\left(\frac{I_{So1}}{I_{So5}}\sqrt{\frac{I_{So4}}{I_{So2}}}\right) - \frac{(n+2)\log(2)}{2}\right] \\ &+ \frac{nU_T}{2}\log\left(\frac{W_1/L_1}{W_5/L_5}\sqrt{\frac{W_4/L_4}{W_2/L_2}}\right)\end{aligned} \tag{3.58}$$

which show the noise margins dependency on voltage between bit-lines (ΔV_{BL}), supply voltage, transistor aspect ratios and transistor parameters V_{To}, I_{So} and n. Moreover, by assuming symmetry (M_{pd}=M_1=M_2, M_{pu}=M_3=M_4 and M_{at}=M_5=M_6), Eqs. (3.57) and (3.58) can be rewritten as

$$\text{WSNM} = A_w \log(\Gamma) + B_w \tag{3.59}$$

$$\text{RSNM} = A_r \log\left(\frac{1}{\sqrt{\Gamma}}\right) + B_r = -\frac{A_r}{2}\log(\Gamma) + B_r \tag{3.60}$$

where

$$\Gamma = \frac{r_{at}^2}{r_{pu} r_{pd}} = \frac{W_{at}^2}{W_{pu} W_{pd}} \frac{L_{pu} L_{pd}}{L_{at}^2} \tag{3.61}$$

in which r_{at}, r_{pu} and r_{pd} are the aspect ratios of the access, pull-up and pull-down transistors, respectively, and A_w, A_r, B_w and B_r are parameters that have low dependency on sizing. The parameter Γ includes information about the 6T-SRAM design parameters, such as the cell ratio r_{pd}/r_{at} [14], and the pull-up ratio r_{pu}/r_{at} [46]. Fig. 3.13 presents the WSNM and RSNM as functions of Γ. This result confirms the opposite "log(Γ)" dependency observed in Eqs. (3.59) and (3.60). For instance, in the simulation result shown in Fig. 3.13, if we consider Γ slightly larger than 1, it is possible to provide positives margins in both read and write operations, simultaneously.

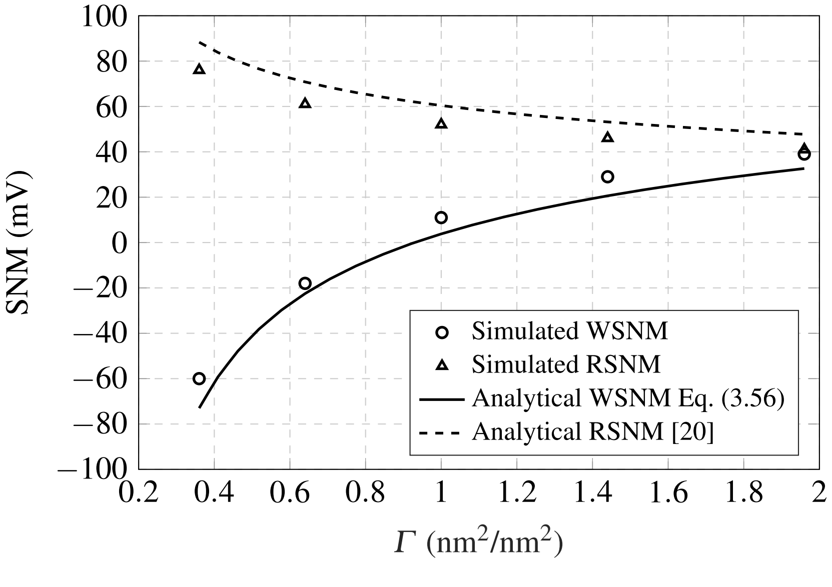


Figure 3.13: Static noise margins as a function of parameter Γ.

3.3.4 Dependency on Temperature

Fig. 3.14 shows the behavior of WSNM along a temperature range from -40 to 120 °C, for $V_{DD} = 0.20$ V and $V_{DD} = 0.35$ V. As expected, the WSNM increases linearly due to the fact that the thermal voltage U_T (see Eq. (3.57)) is linearly related to temperature.

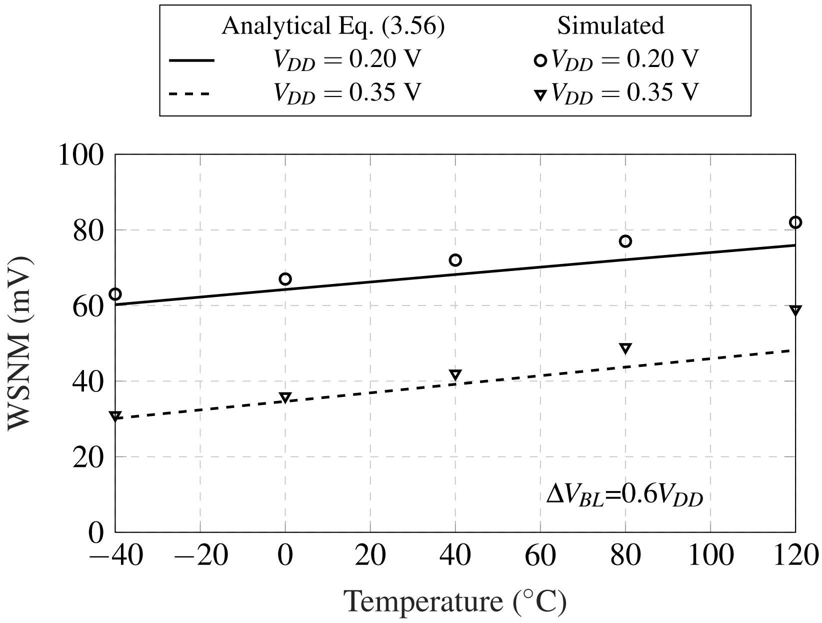


Figure 3.14: WSNM as a function of temperature.

3.3.5 WSNM Distribution

Assuming that threshold voltages are independent random variables, the mean and the standard deviation of WSNM can be estimated, respectively, as

$$\mu_{\text{WSNM}} = \text{WSNM} + \sum_{i=1}^{6} \left(\frac{1}{2} \frac{\partial \text{WSNM}^2}{\partial V_{Toi}^2} \right) \sigma_{V_{Toi}}^2 \tag{3.62}$$

$$\sigma_{\text{WSNM}} = \sqrt{\sum_{i=1}^{6} \left(\frac{\partial \text{WSNM}}{\partial V_{Toi}} \sigma_{V_{Toi}} \right)^2} \tag{3.63}$$

which, considering linear dependency on threshold voltages, yield

$$\mu_{\text{WSNM}} = \text{WSNM} \tag{3.64}$$

$$\sigma_{\text{WSNM}} = \frac{\sqrt{n_5^2 \sigma_{V_{To1}}^2 + n_o^2 \sigma_{V_{To4}}^2 + n_1^2 \sigma_{V_{To5}}^2 + (n_0^2 n_4^2 / n_6^2) \sigma_{V_{To6}}^2}}{n_0 + n_5} \tag{3.65}$$

where $n_0 = n_1 \eta_5 + n_1 \lambda_5 + n_5 \lambda_1 + n_1$, $\sigma_{V_{Toi}} = A_{V_{Toi}} / \sqrt{(W_i L_i)}$, for i= 1,2,..,6, and $A_{V_{Toi}}$ is a technology dependent parameter. Fig. 3.15 presents distributions produced by Monte Carlo simulations of WSNM at supply voltages $V_{DD} = 0.25$ V and $V_{DD} = 0.40$ V. As can be observed, the simulation results are in close agreement with the analytical Gaussian distributions predicted by Eqs. (3.64) and (3.65). In fact, the relative errors (simulation-analytical)/simulation of the mean and standard deviation are, in the worst-cases, 3.84% and 0.75%, respectively.

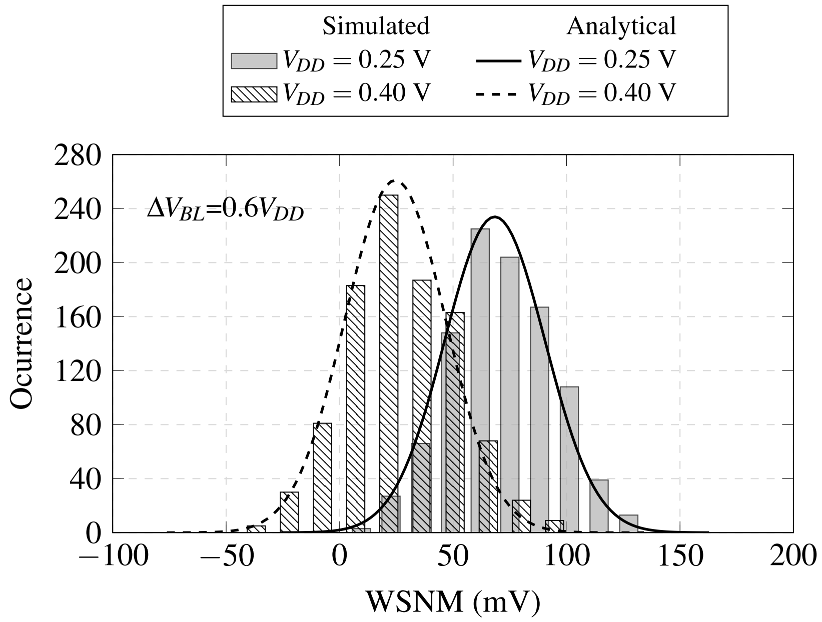


Figure 3.15: WSNM Monte Carlo (1000 samples) simulation results compared with the analytical probability distribution of Eqs. (3.64) and (3.65).

3.3.6 Proposed Model for High Yield SRAM

An analytical model to predict the worst-case of HSNM including global and local process variations was reported in [45]. Such approach can be extended to estimate the worst-cases of write and read static noise margins, that is,

$$\text{WSNM}_{\text{WC}} = \mu\{\text{WSNM}_{\text{SF}}\} - N\sigma\{\text{WSNM}_{\text{SF}}\} \tag{3.66}$$

$$\text{RSNM}_{\text{WC}} = \mu\{\text{RSNM}_{\text{FS}}\} - N\sigma\{\text{RSNM}_{\text{FS}}\} \tag{3.67}$$

respectively, where FS and SF denote the N-Fast-P-Slow and N-Slow-P-Fast corners, respectively, and N is a positive number that defines a confidence interval for local mismatch variations. The WSNM distribution presented in Eqs. (3.64) and (3.65) can be used to estimate Eq. (3.66), and the RSNM distribution reported in [20] to estimate Eq. (3.67). In addition, to control the stability balance, we define the worst-case SNM (SNM$_{\text{WC}}$) as the minimum value between WSNM$_{\text{WC}}$ and RSNM$_{\text{WC}}$ of Eqs. (3.66) and (3.67), respectively.

3.3.6.1 Cell Sizing Aided by the Proposed Model

A sizing exploration for the 6T-SRAM cell was carried out in a 28 nm FD-SOI CMOS process and validated through HSPICE simulations. The single-p-well (SPW) structure [72] of Fig. 3.16 was used to implement the cell, owing to the fact that N-RVT and P-LVT share the same well, and hence body biasing techniques can be employed to compensate for the usually imbalance between N-type and P-type devices. The rectangular-diffusion cell [73] cell was chosen to implement the 6T-SRAM. The transistors were de-

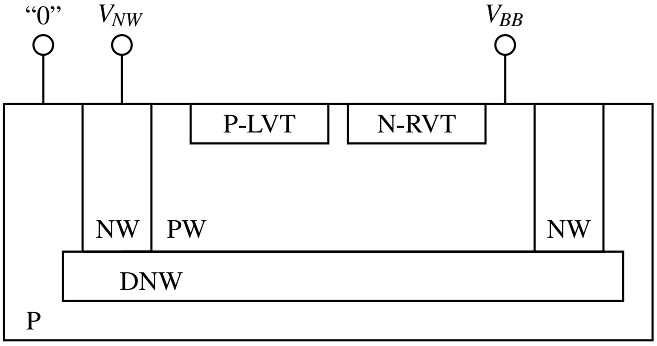


Figure 3.16: SPW structure in 28 nm FD-SOI CMOS process.

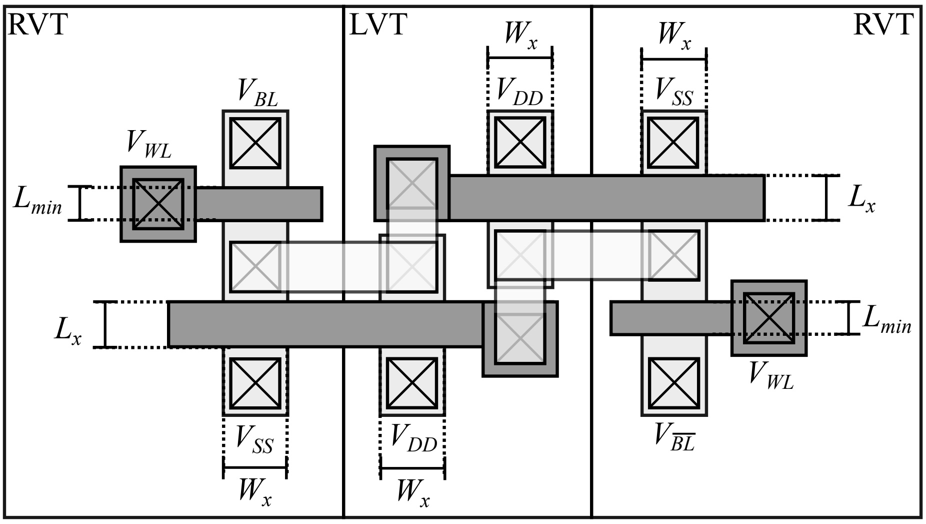


Figure 3.17: Proposed layout for the 6T-SRAM cell.

signed with the same width (W_{pu}=W_{pd}=W_{at}=W_x), the same pull-up and pull-down lengths (L_{pu}=L_{pd}=L_x) and minimum access length (L_{at}=L_{min}= 30 nm), as depicted in the proposed layout of Fig. 3.17. In this case, Γ=L_x^2/L_{min}^2, so that by varying Γ through L_x, we can explore an optimal balance between WSNM and RSNM, since as mentioned before, Eqs. (3.59) and (3.60) have opposite behaviors with respect to Γ. As long as our objective is to develop a high yield cell design, the influence of W_x on local mismatch variations must also be considered. Therefore, Fig. 3.18 shows level curves of WSNM$_{\mathrm{WC}}$, RSNM$_{\mathrm{WC}}$ and SNM$_{\mathrm{WC}}$ as functions of the dimensions L_x and W_x under the following simulation conditions: V_{DD}= 0.4 V, V_{Bn}=V_{Bp}=0 V, ΔV_{BL}=0.8V_{DD} in write mode and V_{WL}=0.8V_{DD} in read mode. The simulated results were obtained from the worst-cases of 500 Monte Carlo runs for a set of 42 points of (L_x,W_x), spending a total time of 8750 seconds, whereas the predicted ones obtained from Eqs. (3.66) and (3.67) take 10.73 seconds for a set of 143 points of (L_x,W_x). As expected, by increasing L_x we increase WSNM and decrease RSNM, both with logarithmic dependency. On the other hand, by increasing W_x we can improve the worst-case of both margins, and it is more noticeable for RSNM. A high-yield 6T-SRAM design having approximately 10 mV on the worst-case of both read and

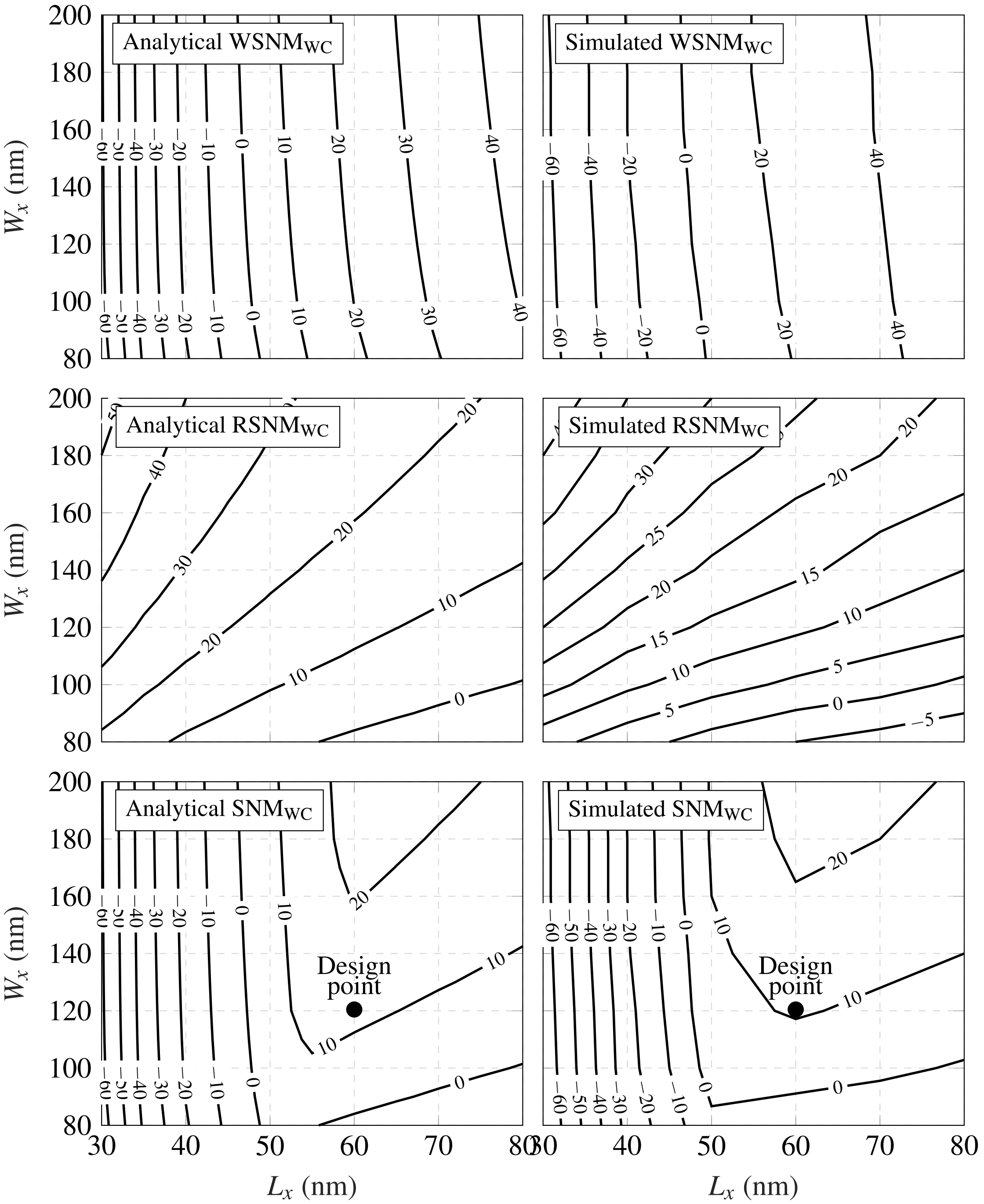


Figure 3.18: Simulation and analytical results for the worst-case static noise margins.

write static noise margins can be obtained by choosing a design point of L_x= 60 nm and W_x= 120 nm, as shown in Fig. 3.18. Fig. 3.19 shows voltage transfer curves produced by Monte Carlo simulation for the obtained design point, in both write and read operation modes. As can be observed, the SNMs are positive under global and local variability for 500 simulated samples, which is in agreement with Fig. 3.18.

Typically, for traditional 6T-SRAM design, a cell ratio (r_{pd}/r_{at}) higher than 1.2 is

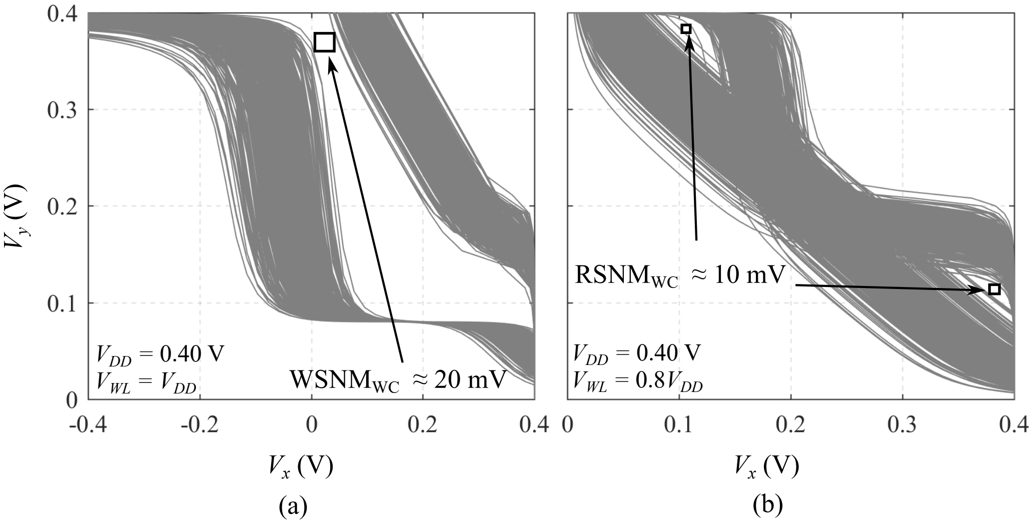


Figure 3.19: Monte Carlo simulations, including global and local variability, for the voltage transfer curves: (a) write operation; (b) read operation.

required to provide a correct read operation, and a pull-up ratio (r_{pu}/r_{at}) lower than 1.8 is needed to maintain proper write-ability [47]. The proposed design point dimensions (L_x= 60 nm and W_x= 120 nm) produces values of 0.5 for both ratios, which as indicated in [47], is correct for write operation only. However, the reduction of the word-line voltage (V_{WL}) to 0.8V_{DD}, in read operation, reduces the access transistor specific gain with respect to the pull-down transistor, thereby increasing the effective cell ratio as well as the RSNM (see Eq. (3.58)). Hence, considering that the voltage V_x stores "0" in read operation, the effective cell ratio can be estimated by

$$\overline{\beta} \approx \frac{(W_1/L_1)I_{So1}\exp(a_{n1}V_{DD})}{(W_5/L_5)I_{So5}\exp(a_{n5}V_{WL})} \tag{3.68}$$

which yields $\overline{\beta}$= 2.29 when V_{WL}= 0.8V_{DD} and V_{DD}= 0.4 V for the proposed design point. This reduction of the word-line voltage level is one of the read-assist approaches to improve the read margin of 6T-SRAMs [74, 75].

3.4 Summary

An analytical model for HSNM that includes DIBL and body biasing effects was proposed in Section 3.1, improving the model for nanometer technologies compared with preview models on the literature. In Section 3.3, we developed an analytical expression for the WSNM at sub-threshold operation. By taking advantage of this model, we proposed an alternative 6T-SRAM design parameter Γ, whose role is to control the well-known trade-off between read and write cell margins. By relating Γ to pull-up and pull-down transistor lengths (L_x), and considering the influences of the widths (W_x) on SNM, we carried out a non-traditional cell sizing implemented in 28 nm FD-SOI CMOS process (see Fig 3.18). The SPW structure combined with reverse body-biasing (V_{BB}=V_{Bn}=V_{Bp}=0) technique allowed design points that have the same lengths on pull-up and pull-down transistors, since P-type transistors with higher specific current are obtained. The implemented Matlab function used to estimate HSNM, RSNM and WSNM of the SRAM cell is presented in Appendix A.2.

Chapter 4

Low-Power SRAM Cache Design Approach

Energy per operation consumption and cell stability are the major performance metrics that must be improved in SRAM cache designs [76]. However, it is not obvious to optimize these metrics simultaneously, since high-cost Monte Carlo simulations are required for every dimensioned design. Hence analytical models to estimate energy, stability, and yield are very useful, in reducing the design time and providing initial transistor sizing. This chapter presents a complete analytical modeling of the SRAM cache, such as the bit-line and word-line capacitances, maximum frequency of operation, static leakage, access transistor leakage and static and dynamic energies. By taking advantage of these models and the ones presented in Chapter 3 for the noise margins, we explore both cell transistor dimensions and cache architecture to design an energy-efficient SRAM cache, without losing focus on stability and speed. As a result of this design exploration, we concluded that efficient solutions in terms of energy consumption and static noise margins can be achieved by dimensioning the simple RD cell [73] through the sizing proposed in Section 3.3.6.1, since by increasing L_x we also reduce the leakage currents of pull-up and pull-down transistors. Hence as we show in this chapter, a reduction of at least 62% on the static leakage can be obtained in view of the fact that pull-up and pull-down leakages become negligible in comparison with the access transistor leakage. Nevertheless, as mentioned in Section 3.3.6.1, the transistor widths (W_x) were also included on the sizing exploration since these have strong influence on both static noise margin and operation frequency of the SRAM cache.

4.1 Cache Architecture Discussion

Many works have analyzed the minimum energy point (MEP) and demonstrated that it is achieved at sub/near-threshold operation [77, 78]. Nevertheless, these approaches follow the traditional SRAM practice, whose architecture has more rows than columns [79]. Recently it has been demonstrated that the optimum energy consumption at low supply voltage operation is not achieved at the maximum number of rows [80] (see Fig. 4.1), because at low supply voltages the static energy achieves similar order than that of the dynamic one. Currently, low-power SRAMs [42] use the general architecture depicted

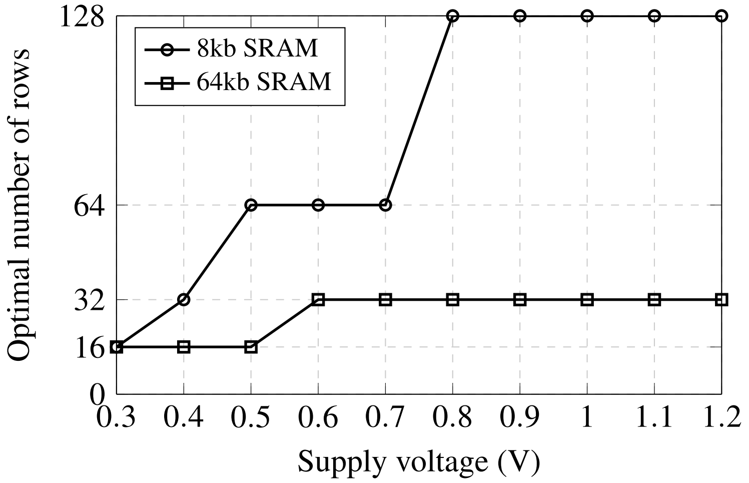


Figure 4.1: Optimal numbers of rows versus supply voltage for energy-efficient SRAM arrays in 65 nm CMOS process technology [80].

in Fig. 4.2, where M_b and N_b are the numbers of rows and columns, respectively, of the SRAM block sub-array. Then this blocks are organized in a main array of size i_x x j_x. In [42], for instance, a 128 kb SRAM is formed by M_b= 32, N_b= 64, i_x= 16 and j_x= 4.

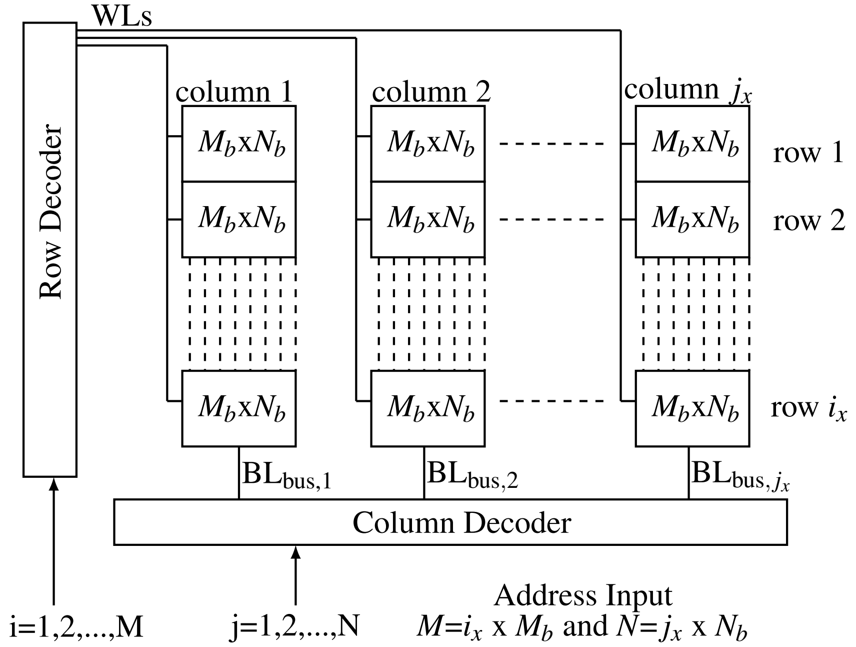
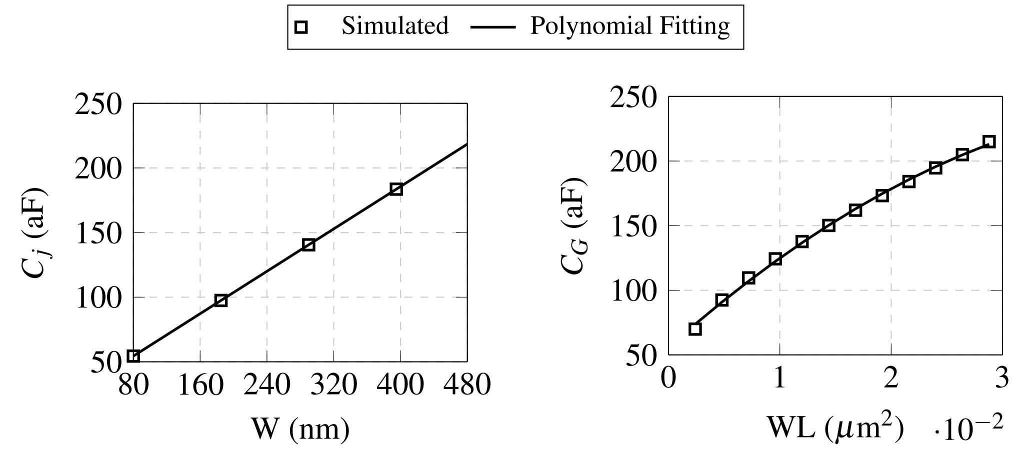


Figure 4.2: MxN SRAM arquitecture divided in sub-arrays of M_bxN_b bit-cells.

4.2 Bit-Line and Word-Line Capacitances

Two major capacitances that have strong influence on the SRAM power consumption are the bit-line and the word-line capacitances. The bit-line capacitance is mainly determined by the total junction capacitance of the access transistor connected to the column of the SRAM array. As can be seen in Fig. 4.3(a), the junction capacitance (C_j) can be assumed linearly dependent on the access transistor width. On the other hand, the word-line capacitance is mainly composed by the total gate capacitance of the access transistors connected to one row of the SRAM array. As shown in Fig. 4.3(b), the total gate capacitance (C_g) can be modeled by a second-order polynomial which is a function of the gate area WL.



(a) C_j= 410e-12W+22e-18

(b) C_G=-77e9(WL)^2+7.7e-3(WL)+0.6e-18

Figure 4.3: Extracted capacitances for a N-type transistor on 28 nm FDSOI CMOS process: (a) junction capacitance (b) gate capacitance.

4.3 Static Leakage Current

One of the major causes for SRAM static power dissipation is the leakage current of the cells that are not being used and are only retaining the logic information, as shown in Fig. 4.4. Therefore, as analyzed in [81], the leakage current of the 6T-SRAM cell can be estimated by adding the sub-threshold leakage currents of transistors M_1, M_4 and M_6. Hence we can derive the total leakage current of the SRAM array (M rows and N columns) as

$$I_{Leak} = \sum_{m=1}^{M\times N} (I_{1,\mathrm{m}} + I_{4,\mathrm{m}} + I_{6,\mathrm{m}}) \tag{4.1}$$

where

$$I_{\mathrm{k,m}} = I_{Sok}\frac{W_k}{L_k}\exp\left(\frac{-V_{Tok}}{n_k U_T}\right)\exp\left(\frac{\lambda_k V_{DD}}{n_k U_T}\right)\left[1-\exp\left(\frac{-V_{DD}}{U_T}\right)\right], \tag{4.2}$$

k is the transistor number of the 6T-SRAM cell, for k = 1,4,6, and m is the cell number of the SRAM array, for m= 1,...,$M \times N$. According to [11], the mismatch variations of

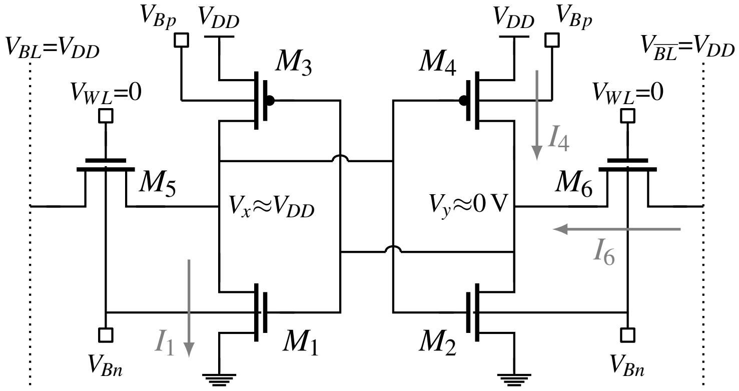


Figure 4.4: 6T-SRAM cell circuit during hold operation.

$I_{\mathrm{k,m}}$ are given by a log-normal distribution, which only considers the threshold voltage variations, and their expected value and variance can be derived, respectively, as

$$\mu_{k,m} = I_{k,m}\exp\left(\frac{\sigma^2_{V_{Tok}}}{2n_k^2U_T^2}\right) \tag{4.3}$$

$$\sigma^2_{k,m} = I^2_{k,m}\left[\exp\left(\frac{\sigma^2_{V_{Tok}}}{n_k^2U_T^2}\right)-1\right]\exp\left(\frac{\sigma^2_{V_{Tok}}}{n_k^2U_T^2}\right) \tag{4.4}$$

where, $\sigma_{V_{Tok}}$ is the standard deviation of the threshold voltage of transistor M_k, for k= 1,4,6. By taking advantage of the large number of memory cells ($M \times N$) on the array, and considering that the transistor variations are independent, we can use the central limit theorem to estimate the mean and variance of the SRAM cache leakage current as

$$\mu_{Leak} = \sum_{m=1}^{M\cdot N}(\mu_{1,m}+\mu_{4,m}+\mu_{6,m}) \tag{4.5}$$

$$\sigma^2_{Leak} = \sum_{m=1}^{M\cdot N}(\sigma^2_{1,m}+\sigma^2_{4,m}+\sigma^2_{6,m}) \tag{4.6}$$

The leakage simulation results compared with the predicted ones by Eqs. (4.5) and (4.6) are shown in Table 4.1, which result confirms that by only considering the threshold voltage variation effects, it is possible to obtain a good estimative of the static leakage current.

Table 4.1: Static leakage current at V_{DD}= 0.3 V of a 2 kb SRAM sub-array (32 rows and 64 columns) formed by the cell designed in Section 3.3.6.1.

	Monte Carlo (100 samples)	Predicted by Eqs. (4.5) and (4.6)
μ_{Leak}	190.7 nA	152.9 nA
σ_{Leak}	2.93 nA	2.13 nA

4.4 Access Transistor Leakage

When a bit-cell is being read, there is a large number of other cells isolated by their access transistors on the bit-line. The leakage currents of these transistors must be low enough to provide a correct read operation. As depicted in Fig. 4.5, the worst-case is achieved when the SA tries to read a bit-cell that stores “0” (“1”) and all other cells in the column are storing “1” (“0”) [13]. In this case, the leakage currents on the other side of the bit-line wrongly discharge the bit-line capacitance, and hence the input difference being read by the sense amplifier is deteriorated. Therefore, the total leakage current of M-1 access transistors connected on the same bit-line is given by

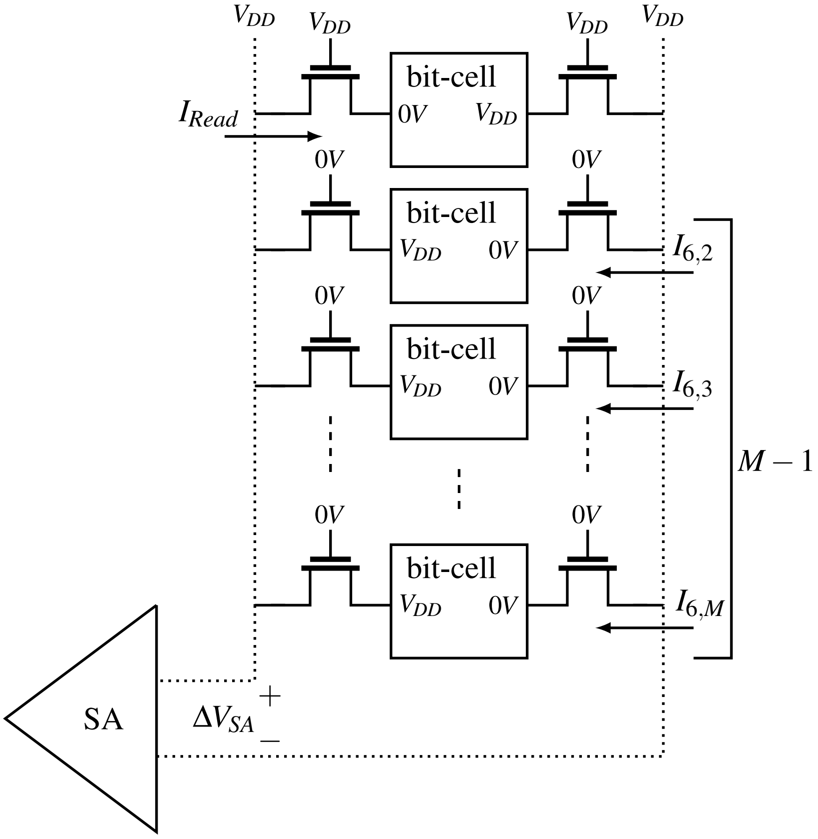


Figure 4.5: Access transistor leakage currents from adjacent bit-cells on the same column reduces the effective read current [13].

$$I_{Leak,AT} = \sum_{m=1}^{M-1} I_{6,\mathrm{m}} \tag{4.7}$$

and the read current as

$$I_{Read} = I_{6,\mathrm{m}} \exp\left(\frac{V_{DD}}{n_6 U_T}\right). \tag{4.8}$$

Since M-1 is usually a large number of cells, we apply the central limit theorem, and hence, the expected value and variance of the total access transistor leakages, respectively, can be estimated by

$$\mu_{Leak,AT} = (M-1)\cdot I_{6,\mathrm{m}} \exp\left(\frac{\sigma^2_{V_{To6}}}{2n_6^2U_T^2}\right) \tag{4.9}$$

$$\sigma^2_{Leak,AT} = (M-1)\cdot I^2_{6,\mathrm{m}} \left[\exp\left(\frac{\sigma^2_{V_{To6}}}{n_6^2U_T^2}\right)-1\right]\cdot \exp\left(\frac{\sigma^2_{V_{To6}}}{n_6^2U_T^2}\right). \tag{4.10}$$

On the other hand, a good estimate of the minimum read current under mismatch variations on the threshold voltage can be obtained by evaluating Eq. (4.8) at the 3σ worst-case, that is

$$I_{Read,\mathrm{min}} = I_{6,\mathrm{m}} \exp\left(\frac{V_{DD}}{n_6U_T}\right)\exp\left(\frac{-3\sigma_{V_{To6}}}{n_6U_T}\right). \tag{4.11}$$

Table 4.2 shows the simulation results of $\mu_{Leak,AT}$, $\sigma_{Leak,AT}$ and $I_{Read,\mathrm{min}}$ compared with their respective predictions in Eqs. (4.9), (4.10) and (4.11). As can be observed, by using V_{DD} higher than 0.35 V, we can obtain a minimum read current that is approximately 10 times higher than that of the mean value of leakage current, so that providing a correct read operation in the worst case of process variability. Accordingly, we define the current ratio between the access transistor leakage and read current as

$$R_{ATvsRead} = \frac{I_{Read,min}}{\mu_{Leak,AT}} \tag{4.12}$$

Table 4.2: Simulated *versus* analytical results for the access transistor leakage and minimum read currents. The bit-line has M = 8×32, which is formed by 8 stacked SRAM sub-arrays.

	V_{DD} (V)	$\mu_{Leak,AT}$ (nA)	$\sigma_{Leak,AT}$ (pA)	$I_{Read,\mathrm{min}}$ (nA)	$R_{ATvsRead}$
Monte Carlo	0.30	19.67	845	56.83	2.88
Simulation	0.35	22.32	958	233.7	10.47
(100 samples)	0.40	23.13	1080	835.1	36.13
Predicted by	0.30	18.74	752	43.45	2.32
Eqs. (4.9),	0.35	22.07	887	204.7	9.27
(4.10) and (4.11)	0.40	26.00	1045	964.1	37.10

4.5 Minimum Cycle Time of SRAM Operation

According to [82], the minimum cycle time of the SRAM can be estimated as

$$T = \max(T_{Read}, T_{Write}) \tag{4.13}$$

where T_{Read} and T_{Write} are the required time to complete the read and write operations, respectively. Since the write operation time can be controlled by the transistor widths of write driver, we considered that the read operation determines the critical time. Therefore, as depicted in the timing diagram of Fig. 4.6, the required time to execute the read operation can be approximated by

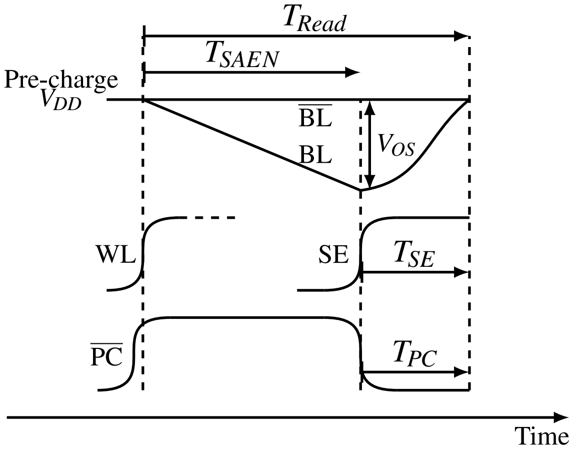


Figure 4.6: Read operation timing diagram [13].

$$T_{Read} = T_{SAEN} + \max(T_{PC}, T_{SE}) \tag{4.14}$$

$$T_{SAEN} = \frac{C_{BL}V_{OS}}{I_{Read}} \tag{4.15}$$

where T_{SAEN} is the time to enable the sense amplifier, T_{PC} is the pre-charging time, T_{SE} is the sense amplifier (SA) delay, V_{OS} is the maximum input offset of the SA, C_{BL} the bit-line capacitance and I_{Read} is the read current of Eq. (4.8).

4.6 Energy per Operation Modeling of SRAM Cache

The energy per operation consumption of the SRAM cache can be separated into dynamic and static components. Basically, the dynamic energy is determined by the capacitance switching on the read and write operations, and the static one by the sub-threshold leakage currents of the cells [80, 82]. Therefore, the SRAM total energy per operation can be expressed as

$$E_{\mathrm{T}} = E_{\mathrm{S}} + E_{\mathrm{D}} \tag{4.16}$$

where E_{T}, E_{S} and E_{D} denote the total, static and dynamic energies per operation, respectively. The static energy of SRAM array is given by

$$E_{\mathrm{S}} = V_{DD}I_{Leak}T \tag{4.17}$$

where I_{Leak} is the leakage current of Eq. (4.1) and T is the SRAM operation period of Eq. (4.13). On the other hand, the dynamic energy can be expressed as

$$E_{\text{D}} = P_{Read}E_{Read} + P_{Write}E_{Write} \tag{4.18}$$

$$E_{Read} = C_{WL,r}V_{WL,read}^2 + C_{BL,r}V_{DD}^2 \tag{4.19}$$

$$E_{Write} = C_{WL,w}V_{WL,write}^2 + C_{BL,w}V_{DD}^2 \tag{4.20}$$

where P_{Read} and P_{Write} are the probabilities of read and write operations, respectively, $C_{WL,r}$ and $C_{WL,w}$ are the effective word-line capacitances of read and write operations, respectively, $C_{BL,r}$ and $C_{BL,w}$ are the effective bit-line capacitances of read and write operations, respectively, and $V_{WL,read}$ and $V_{WL,write}$ are the word-line voltages of read and write operations, respectively. The implemented Matlab function used to estimate the energy consumption of the SRAM cache is presented in Appendix A.1.

4.7 Performance Modeling of SRAM Cache

In order to simplify the energy analysis of the SRAM array and compare it with other works, we choose an architecture of 128 kb formed by M_b= 32, N_b= 64, i_x= 16 and j_x= 4 (see Fig. 4.2). In this case, we can approximate the capacitances of Eq. (4.18) as $C_{WL,r}$=$C_{WL,w}$=$(2\times N_b)C_G$ (two access transistors per cell) and $C_{BL,r}$=$C_{BL,w}$=$(16\times M_b)C_j$, and assume the same probabilities for read and write operations, that is, P_{read}=P_{write}= 0.5. Fig. 4.7 shows level curves of static and dynamic energies from Eqs. (4.17) and (4.18), respectively, by using the sizing proposed in Section 3.3.6.1 for the RD cell in Fig. 3.17. As can be observed in Fig. 4.7(a), when pull-up and pull-down transistor lengths (L_x) are increased, the static energy is reduced in view of the fact that the leakage currents given by Eq. (4.2) are also reduced. As indicated in Fig. 4.7(b), the dynamic energy only depends on W_x, since capacitances C_G and C_j of the access transistors are determined by W_x. Therefore, as can be seen in Fig. 4.8, efficient values for the RD cell can be achieved by considering L_x= 60 nm and W_x= 120 nm, resulting in a total energy per operation of 897.3 pJ/word-access at V_{DD}= 0.42 V.

It is important to note that in Figs. 4.7 and 4.8 the cycle time T is not fixed, which was determined by the worst case regarding the read operation time presented in Eq. (4.14). Fig. 4.9(a) shows level curves of the operation frequency (inverse of the cycle time). This frequency is mainly determined by the bit-line discharging in read operation, which is proportional to the access transistor ratio $W_x/L_{\min}$ through the read current (see Eq. (4.14)). On the other hand, the worst-case static noise margin of Fig. 4.9(b), discussed in detail on Section 3.3.6, yields 3.2 mV at V_{DD}= 0.42 V, for which global and local variations of the manufacturing process was considered. Once we fixed the transistor dimensions of the RD cell at L_x= 60 nm and W_x= 120 nm, another important analysis can be made by

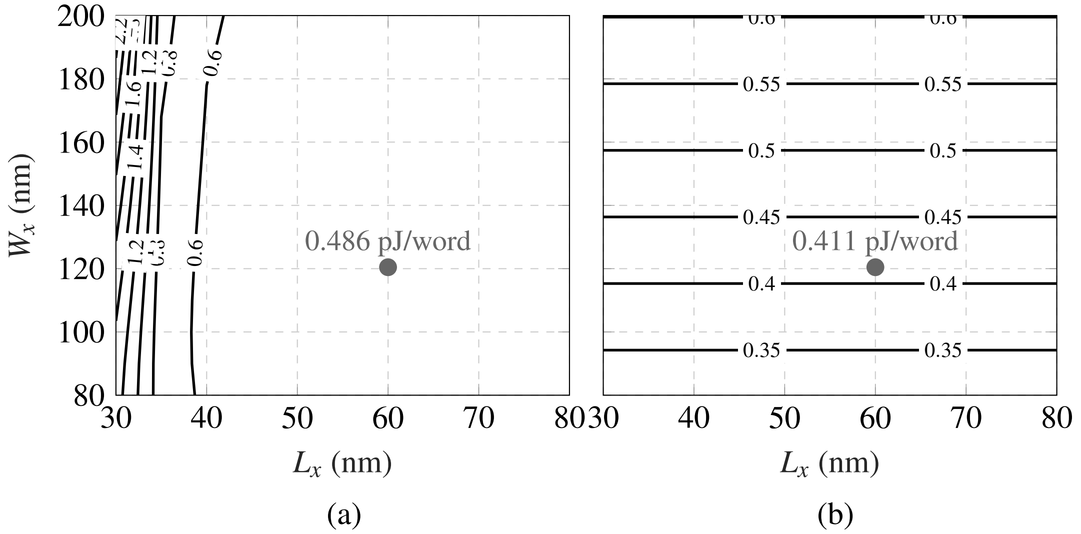


Figure 4.7: Sizing exploration produced by our analytical models for a 128 kb SRAM cache formed by the RD cell (see Fig. 3.17) at V_{DD}= 0.42 V and V_{BB}=0 V: (a) static energy (pJ/word-access); (b) dynamic energy (pJ/word-access).

varying the supply voltage. Fig. 4.10 shows SNM, energy, frequency of operation and the ratio $R_{ATvsRead}$ (see Eq. (4.12)) as a function of the supply voltage. As also shown before in Fig. 4.9(b), positive values of SNM can be obtained at V_{DD}= 420 mV. However, as can be seen in Fig. 4.10, a minimum energy point (MEP) [78] of 0.738 pJ/word is achieved at approximately V_{DD}= 450 mV, with SNM= 10 mV, frequency= 23.17 MHz and $R_{ATvsRead}$= 18.05 A/A. Table 4.3 shows the performance obtained by our analytical modeling compared with both schematic circuit simulations and other works reported in literature for the same 28 nm FD-SOI CMOS technology. At first appearance, the energy per operation seems extremely lower than those of the other works, especially compared with [42], which used the same architecture formed by M_b= 32, N_b= 64, i_x= 16 and j_x= 4. However, since the proposed sizing uses the lengths of pull-up and pull-down transistors (L_x) as one of their variables, the currents I_1 and I_4 of Eq. (4.1) can be considerably reduced when L_x is increased, achieving improved results in terms of static energy consumption as seen in Fig. 4.7(a). This energy reduction also shows the importance of the proposed design parameter $\mathit{\Gamma}$ introduced in Section 3.3.6.1, since as shown in Fig. 4.11, it

Table 4.3: Performance comparison of our analytical modeling with both schematic simulation results and other works reported in literature.

Parameter	This work		Literature	
	Analytical	Simulation	[42]	[43]
Supply voltage (V)	0.45	0.45	0.45	0.35
Total energy (pJ/word-access)	0.738	0.725	3.360	4.600
Frequency (MHz)	23	23	40	10

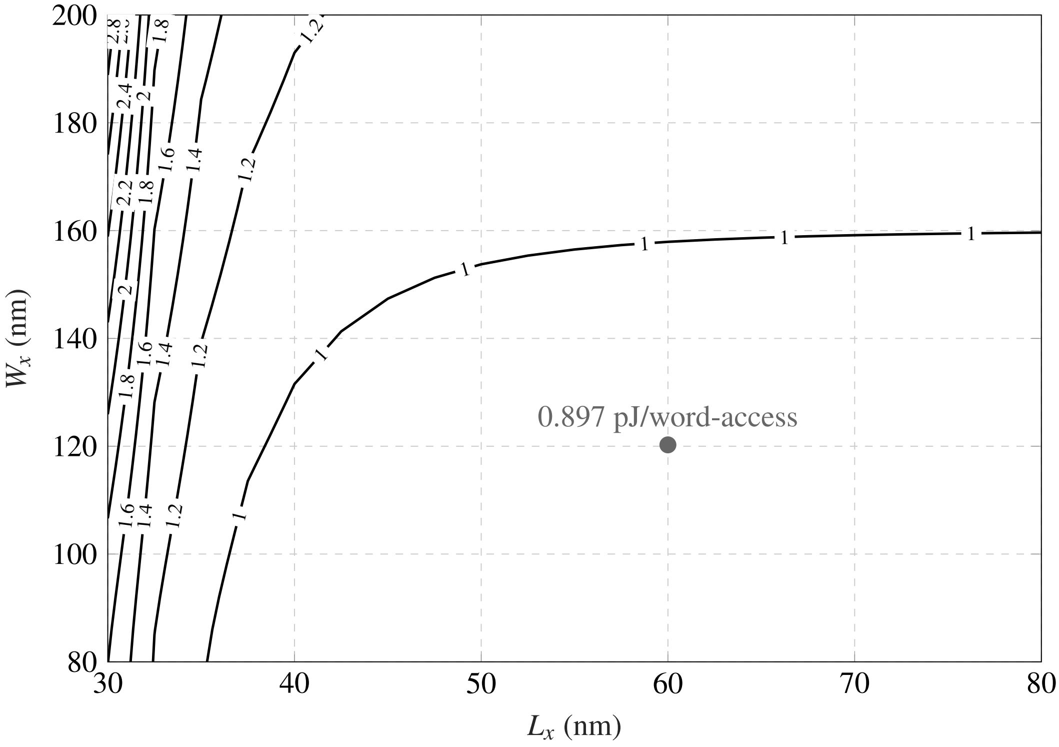
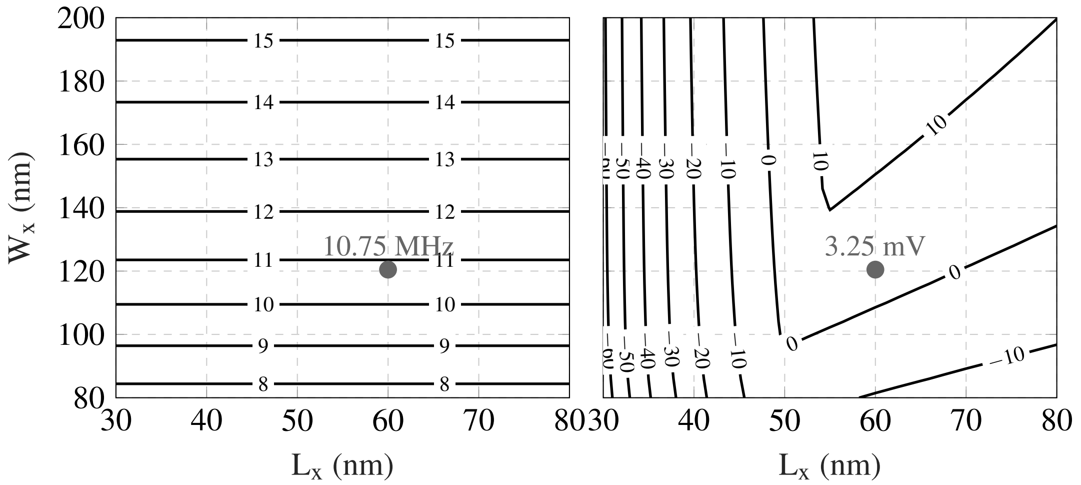


Figure 4.8: Sizing exploration of the total energy (pJ/word-access) produced by our analytical models for a 128 kb SRAM cache formed by the RD cell of Fig. 3.17 at V_{DD}= 0.42 V and V_{BB}= 0 V.

controls the balance between WSNM and RSNM, and reduces the static energy by 62%. Then a reduction of 37% can achieved for the total energy compared with the case in which L_x=L_{min}. Fig. 4.12(a) shows static leakage current simulation results for the traditional "thin-cell" [48] sizing, in which the pull-down and access transistor widths are usually increased to improve the cell margins that enable the proper cell operation using minimum lengths. The proposed sizing (see Fig. 4.12(b)) can reduce the leakage current level by more than 62%. Therefore it is not surprising that the proposed design point (L_x= 60 nm and W_x= 120 nm) can reduce the static energy by more than 70% in comparison with what was reported in [42, 43]. Designs based on increasing transistor lengths has been suggested as an efficient procedure for low power digital circuits [49].

In terms of speed, our approach ensures the correct operation at almost half the frequency reported in [42]. Whereas in this analysis we consider that the bit-lines are discharging completely in each cycle, and the operation frequency can be increased by decreasing V_{OS} in the SA design (see Eq. (4.14)). In Chapter 5 the SA design is developed, and in Chapter 6 final test simulations and performance of the SRAM cache are presented.



(a) (b)

Figure 4.9: Sizing exploration produced by our analytical models for a 128 kb SRAM cache formed by the RD cell (see Fig. 3.17) at V_{DD}= 0.42 V and V_{BB}= 0 V: (a) operation frequency (MHz); (b) worst-case static noise margin (mV).

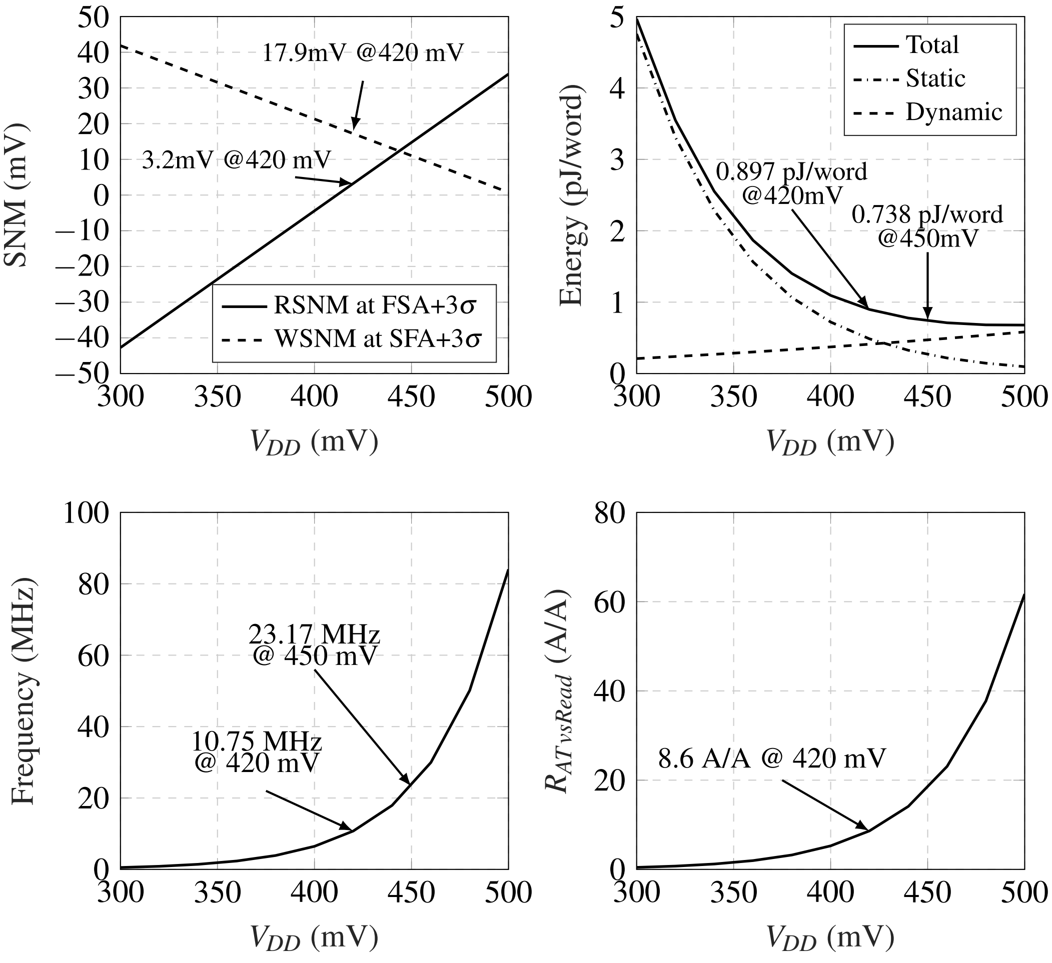


Figure 4.10: 128 kb SRAM cache performance produced by our analytical models as function of supply voltage, using the RD cell (see Fig. 3.17) at L_x= 60 nm and W_x= 120 nm.

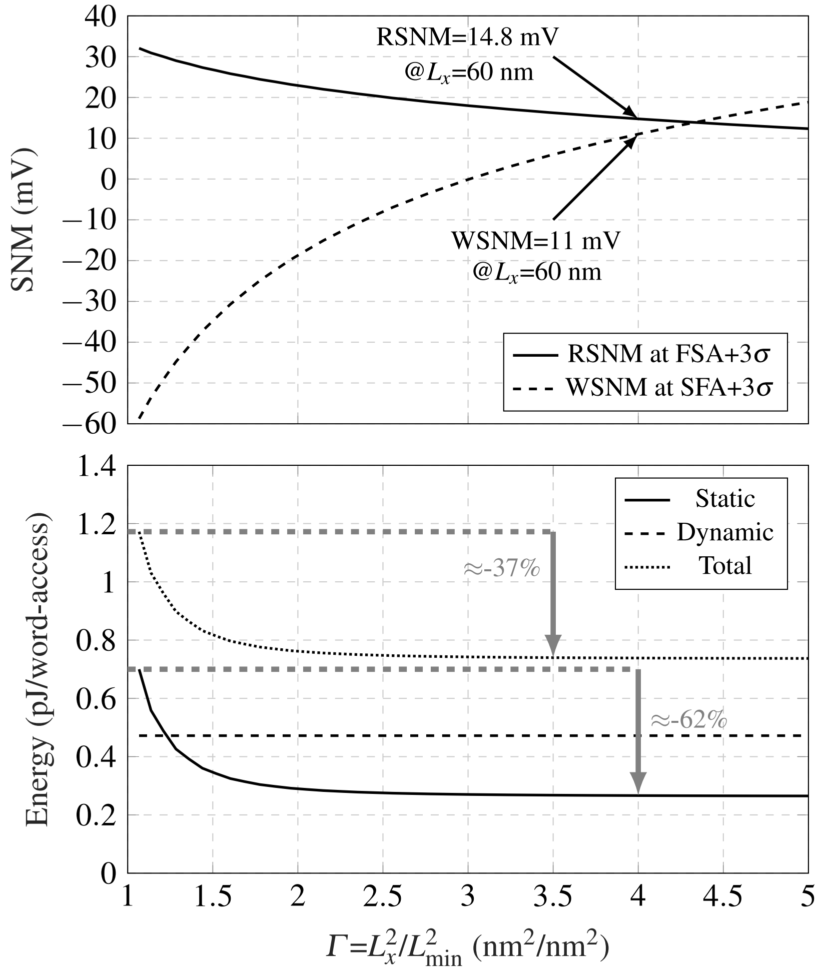
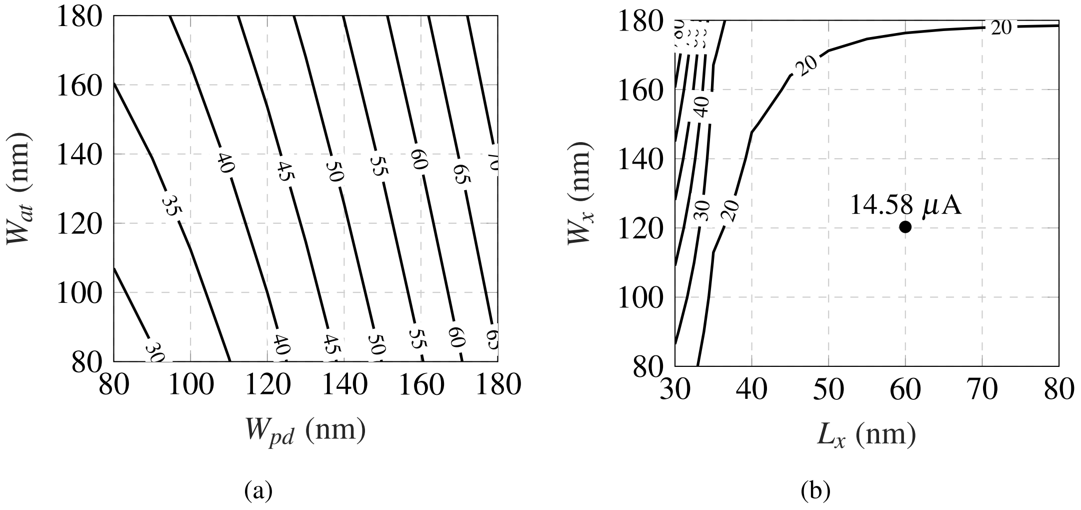


Figure 4.11: Static noise margin and energy consumption as a function of the parameter Γ by varying L_x at V_{DD}= 450 mV, V_{BB}=V_{Bn}=V_{Bp}= 0 V, W_x= 120 nm and $L_{\min}$= 30 nm.



(a)

(b)

Figure 4.12: Static leakage current (μA) simulation of 128 kB 6T-SRAM cells as function of transistor dimensions at V_{DD}= 450 mV: (a) traditional sizing by using W_{pu}= 80 nm and L_{pu}=L_{pd}=L_{at}= 30 nm; (b) proposed sizing.

4.8 Summary

In this chapter we presented a complete analytical model to assess the SRAM cache performance. The major system metrics, such as the static and dynamic energies per operation, static noise margins, operation frequency and leakage currents were carefully modeled. By taking advantage of these low computational cost models, we explored the performance of a 128 kb SRAM cache implemented with the simple 6T rectangular-diffusion (RD) cell. The proposed energy-efficient sizing of Section 3.3.6.1, in which the cell transistors are designed with the same width (W_{pu}=W_{pd}=W_{at}=W_x), the same pull-up and pull-down lengths (L_{pu}=L_{pd}=L_x) and minimum access transistor length (L_{at}=$L_{\min}$= 30 nm), in combination with the SPW structure at reverse-body-biasing (RBB) (V_{BB}=V_{Bn}=V_{Bp}=0 V), achieved a MEP of 0.738 pJ/word-access at V_{DD}= 450 mV and operation frequency of 23 MHz. As shown in Table 4.3, our approach can reduce by 75% the energy per operation compared with [42], mainly for the reason that [42] uses the traditional "thin-cell" [48], in which the leakage current increases by increasing transistor widths to adjust the cell margins that enable the proper cell operation. Thus, as also currently suggested in [49] for saving more that 50% of the energy in digital circuits, the increase of transistor lengths considerably reduces the cell static leakage. However, this cannot be applied to the access transistor since it must discharge higher bit-line capacitances, and for this reason we proposed an energy-efficient sizing by using the aforementioned variables L_x and W_x. The results presented in this chapter were not simulated with extracted components (parasitic capacitances and resistances). In Section 6, on the other hand, we present the final simulations that include all parasitic effects extracted from layout, so that we can verify the properly operation of the SRAM under the manufacturing process variations.

Chapter 5

Sense Amplifier

The bit-line capacitance is an important issue in low-power SRAMs. The design key is to avoid large discharging time of these capacitances, which is usually required in read operation. The discharging time can be decreased by reducing the SA input offset. Therefore, the lower is the SA input offset, the lower are the energy consumption and required time to complete the read operation [83]. In this chapter, we present analytical models to estimate the sense amplifier input offset and yield for the latch type sense amplifier (LTSA) topology operating at sub-threshold operation.

5.1 Input Offset

5.1.1 Bit-line Pre-charging Near the Metastability Voltage

Fig. 5.1 shows the 7T-LTSA circuit. During the read operation, once the signal SE is activated, the sensor is instantaneously isolated from the bit-lines though their access transistors ($M_{AT,1}$ and $M_{AT,2}$), and the footer transistor ($M_{n,3}$) is turned on. At this moment the SA can be approximated as the latch circuit of Fig. 5.2(a). Consequently, the small signal model of this circuit is shown in Fig. 5.2(b), in which $G_{m,1}$ and $G_{m,2}$ denote the total transconductance of each cross-coupled inverter ($M_{n,1}$-$M_{p,1}$ and $M_{n,2}$-$M_{p,2}$), and C_1 and C_2 are the total capacitances of the outputs $V_{o,1}$ and $V_{o,2}$, respectively. It is important to note that capacitances C_1 and C_2 do not include the bit-line capacitances, since as mentioned before, SA is disconnected from the bit-line when enabled. According to what is reported in [29, 50] the latch input offset can be derived as

$$V_{OS} = \alpha\left(V_{o,2}\big|_{t=0} - V_{S,2}\right) - V_{o,1}\big|_{t=0} + V_{S,1} \tag{5.1}$$

$$\alpha = \sqrt{\frac{G_{m,2}}{G_{m,1}}\frac{C_2}{C_1}} \tag{5.2}$$

where $V_{S,1}$ and $V_{S,2}$ are the metastability voltages.

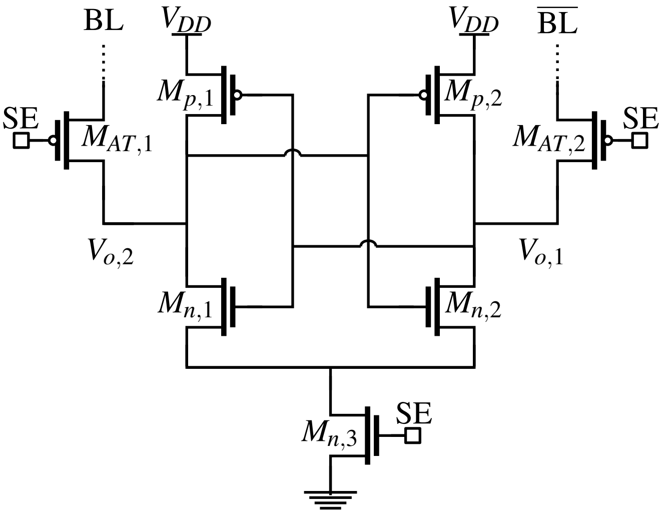


Figure 5.1: 7T latch type sense amplifier (LTSA) circuit.

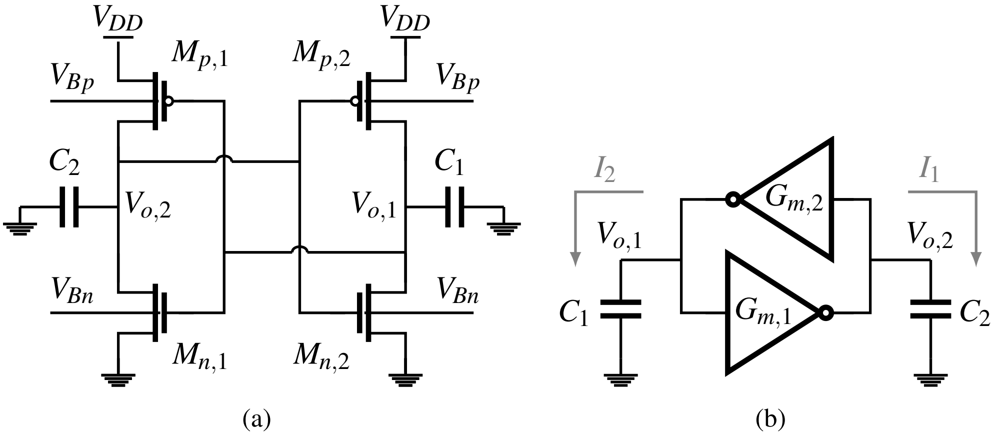


Figure 5.2: Simplified latch-type sense amplifier: (a) schematic circuit; (b) small-signal model.

As long as the analysis is made in sub-threshold region, the transconductance over drain current ratio (G_m/I_D) of transistors can be approximated by $1/(\bar{n}U_T)$ [84]. Therefore, the total transconductance of the cross-coupled inverters can be written as

$$G_{m,1} \approx \frac{I_1}{\bar{n}U_T} = \frac{1}{\bar{n}U_T}\left[\beta_{p,1}\exp\left(\frac{V_{DD}-V_{o,1}}{\bar{n}U_T}\right)+\beta_{n,1}\exp\left(\frac{V_{o,1}}{\bar{n}U_T}\right)\right] \tag{5.3}$$

$$G_{m,2} \approx \frac{I_2}{\bar{n}U_T} = \frac{1}{\bar{n}U_T}\left[\beta_{p,2}\exp\left(\frac{V_{DD}-V_{o,2}}{\bar{n}U_T}\right)+\beta_{n,2}\exp\left(\frac{V_{o,2}}{\bar{n}U_T}\right)\right] \tag{5.4}$$

where

$$\beta_{n,i} = I_{son,i}\exp\left(\frac{-V_{Ton,i}+\eta_n V_{Bn}}{n_{n,i}U_T}\right) \tag{5.5}$$

$$\beta_{p,i} = I_{sop,i}\exp\left(\frac{-|V_{Top,i}|+\eta_p(V_{DD}-V_{Bp})}{n_{p,i}U_T}\right) \tag{5.6}$$

for *i*= 1,2. Furthermore, we can derive the effective values of Eqs. (5.3) and (5.4) regarding the input voltage range from 0 to V_{DD}, respectively, as

$$\overline{G_{m,1}} = \frac{1}{V_{DD}} \int_0^{V_{DD}} G_{m,1}(V_{o,1})\mathrm{d}V_{o,1} = \frac{(\beta_{p,1}+\beta_{n,1})\left[\exp\left(\frac{V_{DD}}{\bar{n}U_T}\right)-1\right]}{V_{DD}} \tag{5.7}$$

$$\overline{G_{m,2}} = \frac{1}{V_{DD}} \int_0^{V_{DD}} G_{m,2}(V_{o,2})\mathrm{d}V_{o,2} = \frac{(\beta_{p,2}+\beta_{n,2})\left[\exp\left(\frac{V_{DD}}{\bar{n}U_T}\right)-1\right]}{V_{DD}}. \tag{5.8}$$

On the other hand, as illustrated in Fig. 5.3, we propose to approximate the metastability point by the intersection of the tangent curves of the VTC centers $V_{OFF,1}$ and $V_{OFF,2}$ (see Eq. (3.3)), which follow the definition in Eq. (3.4). Then the approximated metastability voltages can be expressed as

$$\overline{V_{S,1}} = \frac{A_V^2 V_{OFF,1} - A_V V_{OFF,2} + (V_{DD}/2)(A_V-1)}{A_V^2-1} \tag{5.9}$$

$$\overline{V_{S,2}} = \frac{A_V^2 V_{OFF,2} - A_V V_{OFF,1} + (V_{DD}/2)(A_V-1)}{A_V^2-1} \tag{5.10}$$

where

$$A_V = \left| \frac{\Delta V_o}{\Delta V_i}\bigg|_{V_i=V_{OFF}} \right| = \left(\bar{\lambda} + \frac{\bar{n}}{\exp\left(\frac{V_{DD}}{2U_T}\right)-1} \right)^{-1} \tag{5.11}$$

is the small-signal gain of the inverters at their centers. The simulation results of the metastability voltage as a function of supply voltage compared with the one predicted by Eq. (5.9) is shown in Fig. 5.4. In Fig. 5.5 the simulated small-signal gain A_v is compared to the analytical of Eq. (5.11) for an inverter implemented in 28 nm FD-SOI CMOS. Observe that A_v can be approximated by $1/\bar{\lambda}$ at supply voltages higher than 250 mV, and

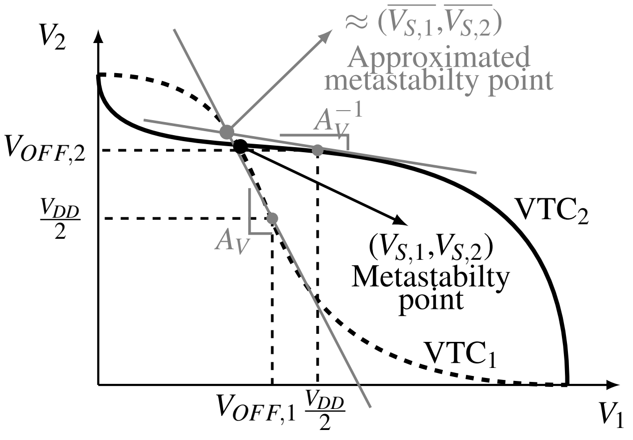
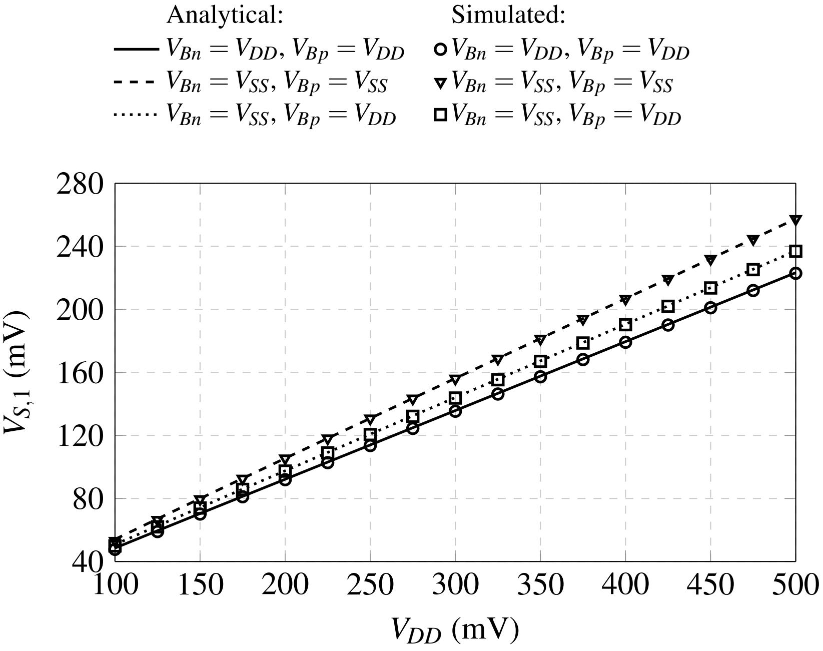


Figure 5.3: Graphical illustration of the metastability point approximation.



(a)

Figure 5.4: Meta-stable state simulation results as a function of supply voltage compared with the predicted ones by Eq. (5.9).

hence the resulting metastability voltages become

$$\overline{V_{S,1}} = \frac{V_{DD}}{2}(\bar{\lambda}+1) + \frac{\bar{n}U_T}{2}\log\left(\frac{\beta_{p,1}}{\beta_{n,1}}\right) - \bar{\lambda}\frac{\bar{n}U_T}{2}\log\left(\frac{\beta_{p,2}}{\beta_{n,2}}\right) \tag{5.12}$$

$$\overline{V_{S,2}} = \frac{V_{DD}}{2}(\bar{\lambda}+1) + \frac{\bar{n}U_T}{2}\log\left(\frac{\beta_{p,2}}{\beta_{n,2}}\right) - \bar{\lambda}\frac{\bar{n}U_T}{2}\log\left(\frac{\beta_{p,1}}{\beta_{n,1}}\right). \tag{5.13}$$

Note that in the ideal case where there is no DIBL effect ($\bar{\lambda}$=0) and the inverters are compensated (β_p=β_n), the metastability center will be at half the supply voltage.

Now, considering that the initial values of $V_{o,1}$ and $V_{o,2}$ in Eq. (5.1) are equal to V_{DC}, and substituting Eqs. (5.7), (5.8), (5.12), and (5.13) into (5.1), we can estimate the SA input offset at sub-threshold operation as

$$V_{OS} = \sqrt{\frac{C_2}{C_1}}\sqrt{\frac{\beta_{n,2}+\beta_{p,2}}{\beta_{n,1}+\beta_{p,1}}}(V_{DC}-\overline{V_{S,2}}) - V_{DC} + \overline{V_{S,1}} \tag{5.14}$$

Eq. (5.14) is valid when the four transistors of the latch are operating in the saturation region. This condition is not true when V_{DC} assumes initial values near of V_{DD} or 0, since P-type and N-type transistors operate in linear region.

As we have been proposing in this document, we are interested in developing statistical equations in order to obtain approximated worst-case values of performance parameters at a low computational cost. Therefore, we can derive the mismatch variations of

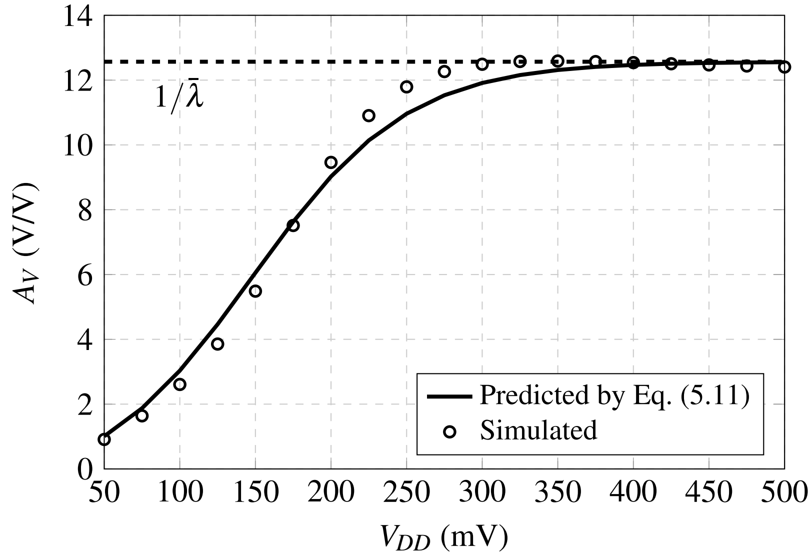


Figure 5.5: Maximum small-signal gain of the inverter circuit in 28 nm FD-SOI CMOS.

Eq. (5.14) by using the differential method as

$$\Delta V_{OS} = \frac{\partial V_{OS}}{\partial V_{Ton1}}\Delta V_{Ton1} + \frac{\partial V_{OS}}{\partial V_{Ton2}}\Delta V_{Ton2} + \frac{\partial V_{OS}}{\partial V_{Top1}}\Delta V_{Top1} + \frac{\partial V_{OS}}{\partial V_{Top2}}\Delta V_{Top2} \tag{5.15}$$

where the partial derivatives yield

$$\frac{\partial V_{OS}}{\partial V_{Ton1}} = \frac{1}{2}\left(\beta_n H_{os} - \bar{\lambda} - 1\right) \tag{5.16}$$

$$\frac{\partial V_{OS}}{\partial V_{Ton2}} = -\frac{1}{2}\left(\beta_n H_{os} - \bar{\lambda} - 1\right) \tag{5.17}$$

$$\frac{\partial V_{OS}}{\partial V_{Top1}} = \frac{1}{2}\left(\beta_p H_{os} + \bar{\lambda} + 1\right) \tag{5.18}$$

$$\frac{\partial V_{OS}}{\partial V_{Top2}} = -\frac{1}{2}\left(\beta_p H_{os} + \bar{\lambda} + 1\right) \tag{5.19}$$

where

$$H_{os} = \frac{\frac{V_{DD}}{2}(1+\bar{\lambda}) + \frac{\bar{n}U_T}{2}\log\left(\frac{\beta_p}{\beta_n}\right)(1-\bar{\lambda}) - V_{DC}}{\bar{n}U_T(\beta_n + \beta_p)}. \tag{5.20}$$

Now, by deriving Eq. (5.15) in terms of variances, and considering that the same type transistors have equal standard deviation, we obtain

$$\sigma_{OS}^2 = \frac{1}{2}\left(\beta_n H_{os} - \bar{\lambda} - 1\right)^2 \sigma_{V_{Ton}}^2 + \frac{1}{2}\left(\beta_p H_{os} + \bar{\lambda} + 1\right)^2 \sigma_{V_{Top}}^2. \tag{5.21}$$

Observe that the capacitances C_1 and C_2 were considered identical and without variability.

5.1.1.1 Yield

According to [27], once we have the standard deviation of the input offset (σ_{OS}), the SA yield (Yield_{SA}) can be evaluated as

$$\text{Yield}_{\text{SA}}(\Delta V_{IN}) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\Delta V_{IN}}{\sigma_{OS}\sqrt{2}}\right)\right] \tag{5.22}$$

where

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp^{-y^2} dy. \tag{5.23}$$

Fig. 5.6 shows the yield results predicted by using Eqs. (5.21) and (5.22) compared with simulations of the 7T-LTSA, by varying the differential input voltage for various initial values at the output ($0.35V_{DD}$, $0.50V_{DD}$ and $0.75V_{DD}$). The supply voltage was set to 300 mV to ensure sub-threshold operation. The results show that the analytical models are in close agreement with simulations, making the proposed model appropriate to predict the SA input offset at a low computational cost.

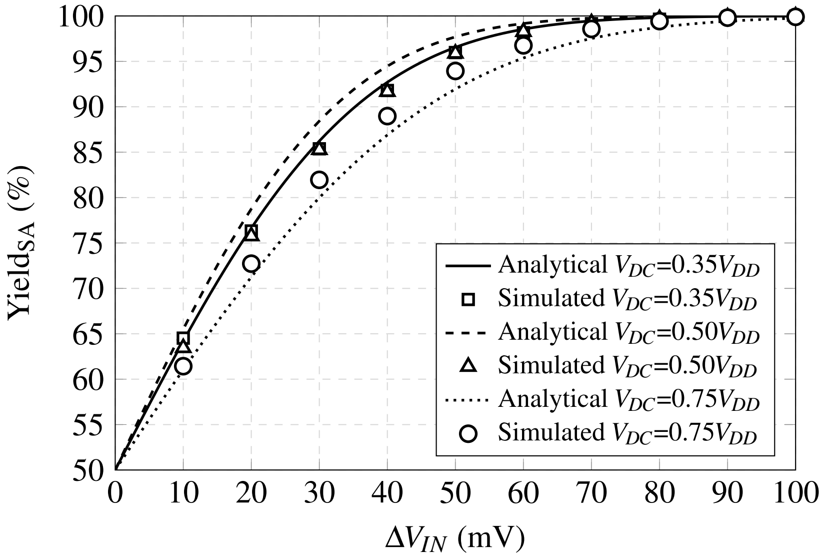


Figure 5.6: Analytical results compared with the ones obtained by 10.000 samples of Monte Carlo simulations; V_{DD}= 0.3 V, W_p= 120 nm, W_n= 80 nm and L_p=L_n= 30 nm.

5.1.2 Bit-lines Pre-charged at Supply Voltage

An analytical model for the input-offset estimation of the 7T-LTSA sense amplifier, when the bit-lines are pre-charged to V_{DD}, was reported in [28]. That model proposes the estimation of the standard deviation as

$$\sigma_{OS} = \frac{\sqrt{2}\sigma_{V_{Ton}}}{1 - \text{DCI} - \text{DIBL}} \tag{5.24}$$

where DCI represents the differential charge injection produced by the access transistor switching, and DIBL is the drain-induced barrier lowering effect. Fig. 5.7 shows the yield simulation results compared with analytical ones from Eqs. (5.24) and (5.22), by neglecting DCI effects. As can be observed, analytical and simulation results are in close agreement. Observe that the simulated SA, which has dimensions W_p/L_p= 140/30 nm/nm, W_n/L_n= 150/50 nm/nm and W_{AT}/L_{AT}= 120/30 nm/nm, can achieve a yield of 99% at V_{IN}= 50 mV. On the other hand, Fig. 5.8 shows Monte Carlo simulations of the SA transient behavior and, as can be observed, the obtained worst-case delay is 3.85 ns.

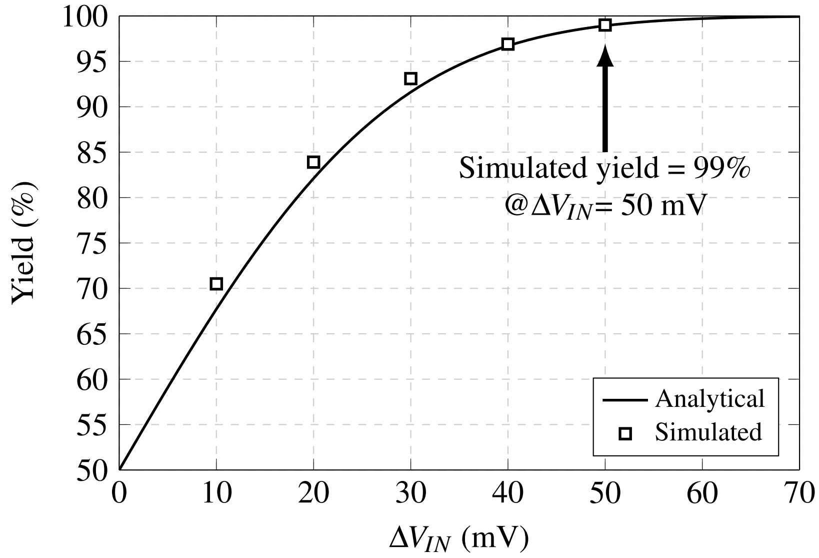


Figure 5.7: Analytical yield results compared with the ones obtained by 1000 samples of Monte Carlo simulation. Where V_{DD}= 0.45 V, W_p/L_p= 140/30 nm/nm, W_n/L_n= 150/50 nm/nm and W_{AT}/L_{AT}= 120/30 nm/nm.

5.2 Summary

In this chapter analytical models to estimate input-offset and yield of the classical 7T-LTSA at sub-threshold operation were presented. According to [29, 50], we contributed mainly by simplifying the latch input-offset expression at sub-threshold operation (see Eq. (5.14)), which allows the prediction of the SA yield. It is important to observe that Eq. (5.14) loses the importance when there are no mismatch variations, since in that case V_{OS} is equal to zero. However, we are interested in its variability, and hence by differentiating Eq. (5.14) we obtained (5.21), which predicts the V_{OS} standard deviation.

Regarding the SRAM cache optimal frequency of 23 MHz and supply voltage of 450 mV obtained in Section 4, the simulated SA in Section 5.1.2 (W_p/L_p= 140/30 nm/nm, W_n/L_n= 150/50 nm/nm and W_{AT}/L_{AT}= 120/30 nm/nm) provides a worst-case delay of 3.85 nS (see Fig. 5.8) at V_{DD}= 450 mV, which is an acceptable result in view of the fact

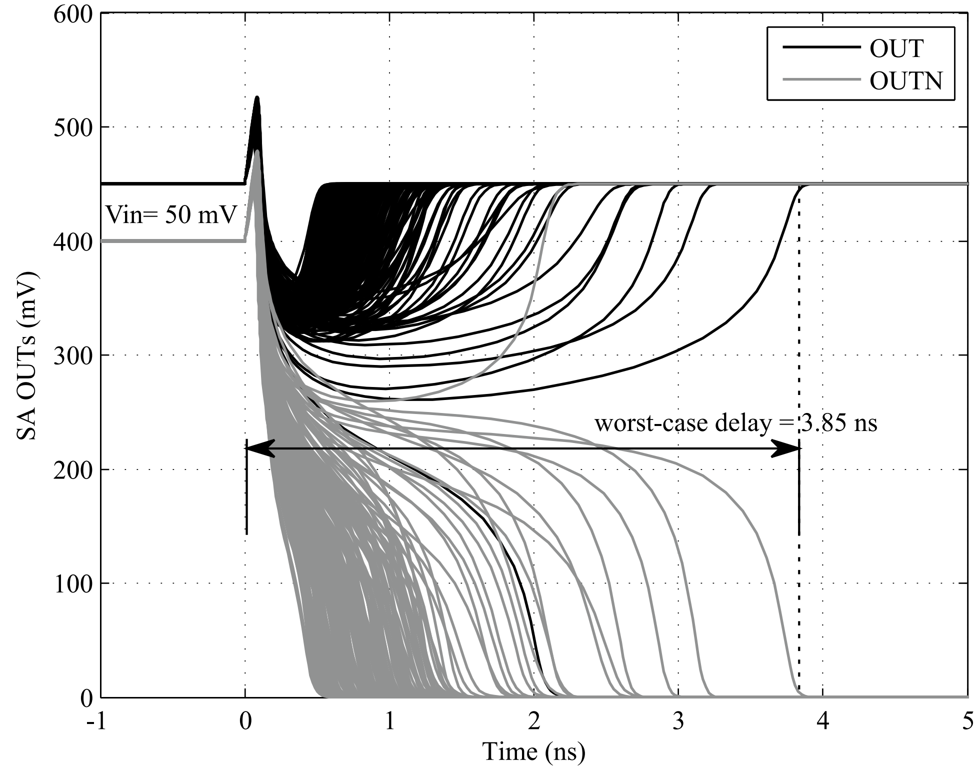


Figure 5.8: Monte Carlo simulation (200 samples) results, including global and local variations, of the sense amplifier transient behavior.

that it is approximately 11 times faster than 1/(23 MHz). Moreover, the SA has a maximum input-offset of 50 mV with yield of 99% as seen in Fig. 5.7.

Chapter 6

SRAM Layout and Simulations

6.1 SRAM Cache Layout

The performance of the SRAM array can be improved, in terms of speed and power consumption, by shrinking the dimensions of their transistors since extrinsic and parasitic capacitances can be reduced. Nevertheless, these shrunk dimensions can degrade the noise margins in view of the fact that minimum transistor dimensions have several issues, such as small differences between devices and imperfections, due to the manufacturing process variations [85]. In consequence, many works reported in literature propose different layout topologies for the 6T-SRAM cell in nanometer nodes [19, 73, 85, 86].

6.1.1 Rectangular-Diffusion 6T-SRAM

A common topology for high-density SRAMs is the rectangular-diffusion (RD) [19], in which both access and pull-down N-type transistors share a straight and long rectangular shape. This reduces the fabrication process variability [73], thereby improving the SNM and making the layout topology appropriate for low voltage operation. Fig. 6.1 shows the implemented layout for the RD 6T-SRAM cell. This layout was designed with minimal standard distance rules to minimize the wasted area between diffusions, poly-silicons, metals, as other factors. In addition, the cell needs to be symmetric with respect to both vertical and horizontal axis for two reasons: to reduce systematic variations [87] and to connect the cell with others in any direction, hence allowing a high-density SRAM. Therefore, the active area of the cell of Fig. 6.1, when it is used in a high-density array, occupies 0.434 μm^2 (348 nm x 1248 nm). Fig. 6.2(a) shows how the cell is connected with other ones to produce a 32x64 sub-array of Fig. 6.2(b).

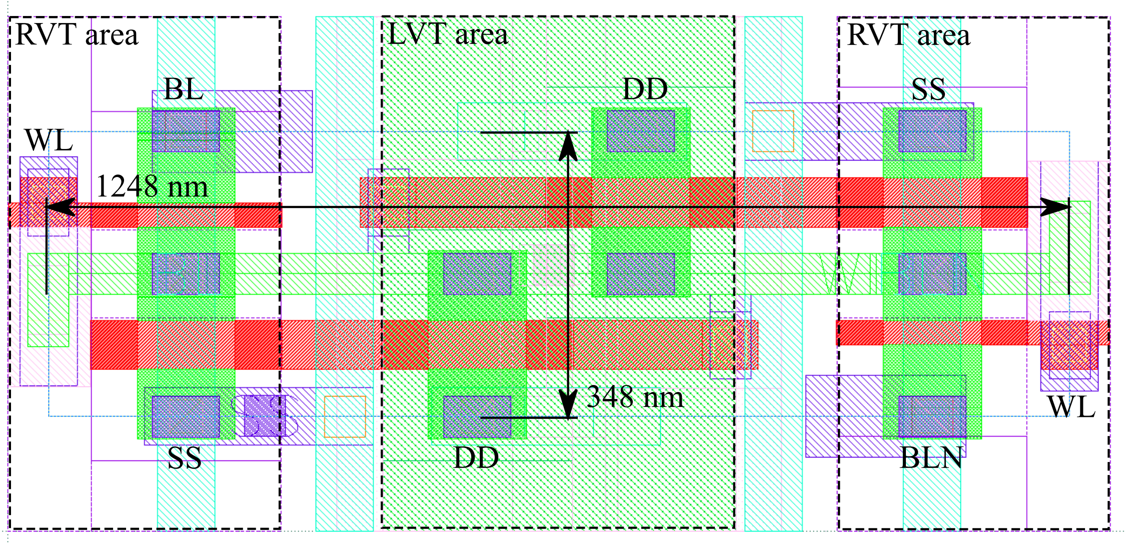
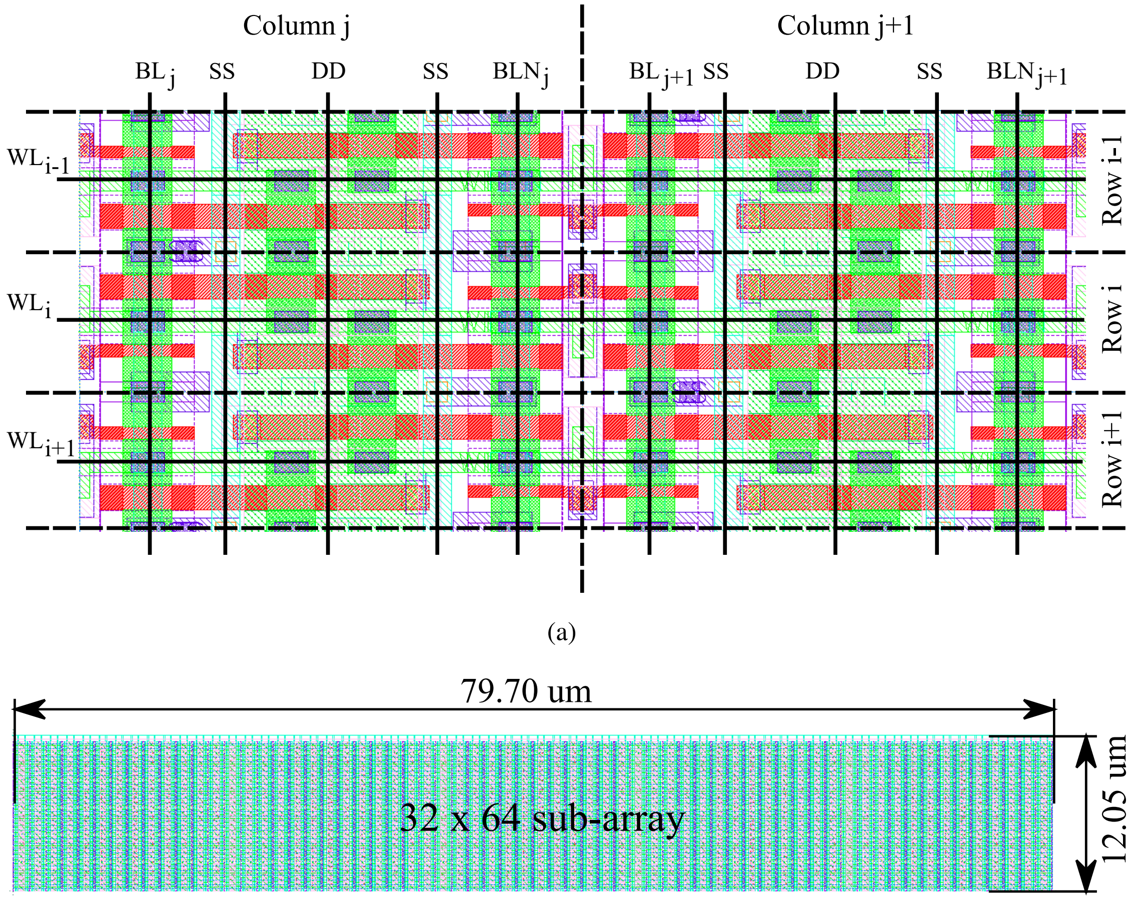


Figure 6.1: 6T-SRAM cell layout.



(b)

Figure 6.2: Layout of 2 kb SRAM sub-array block: (a) 6T-SRAM cell arrangement for area saving; (b) 32x64 (2 kb) sub-array block.

6.1.2 7T Latch-Type Sense Amplifier

Fig. 6.3 shows the 7T sense amplifier layout, which is symmetric about vertical axis in view of the fact that any small imbalance can produce undesirable parasitic resistances and/or capacitances, and hence increasing both SA input offset and delay. Observe that the layout width was implemented equal to the memory cell in order to arrange it at the top of the SRAM column.

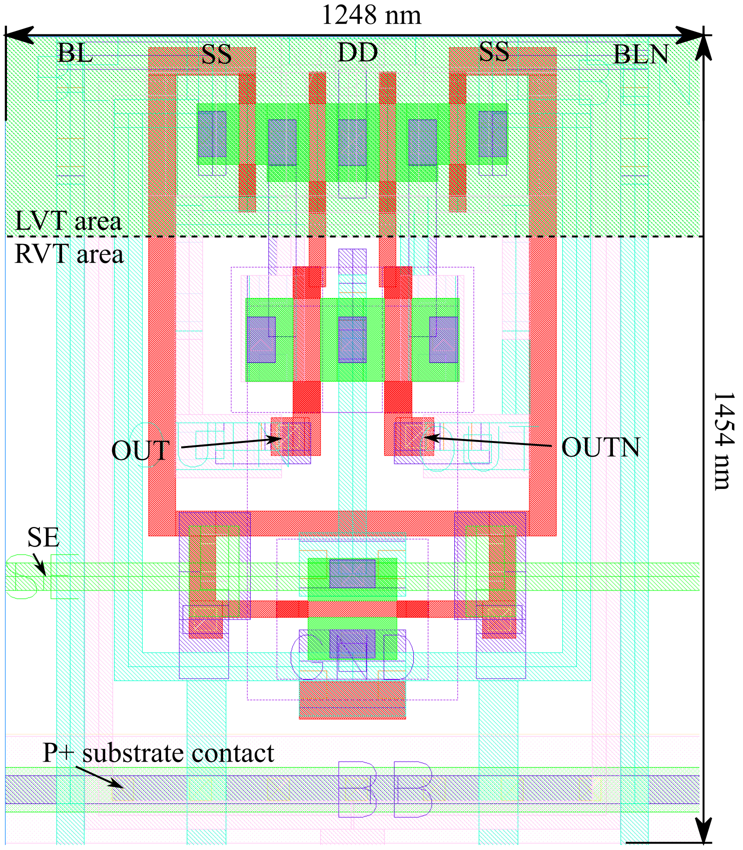


Figure 6.3: 7T latch-type sense amplifier layout.

6.1.3 Pre-charge and Write Drivers

The pre-charge and write driver circuits must be designed, respectively, to charge and discharge the bit-line capacitance (C_{BL}) in a desired lapse of time. In order to have a proper operation under process variations, we imposed that this time must be 10 times faster than the SRAM operation frequency, when SS process corner is used. The needed currents of approximately 3.5 μA, at SS corner, were implemented by dimensioning the pre-charge and write driver as shown in Figs. 6.4(a) and 6.4(b), respectively. Their implemented layouts can be seen in Figs. 6.5 and 6.6. Observe that the layout widths are also equal to that of the memory cell in order to arrange them at the bottom of the SRAM column.

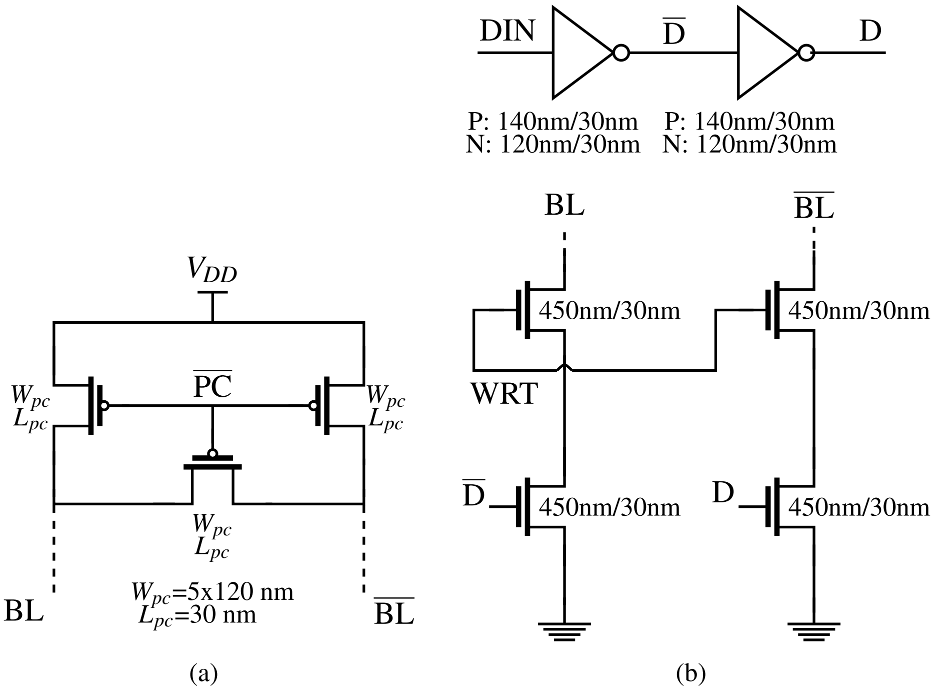


Figure 6.4: Schematic circuits of the pre-charge (a) and write driver (b).

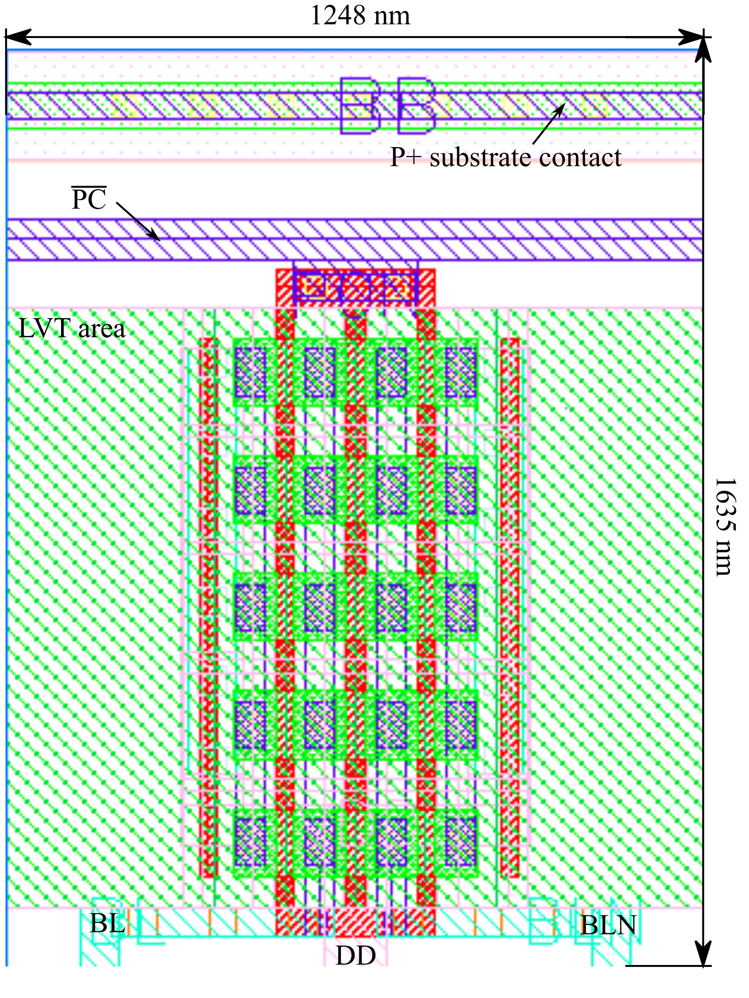


Figure 6.5: Pre-charge circuit layout.

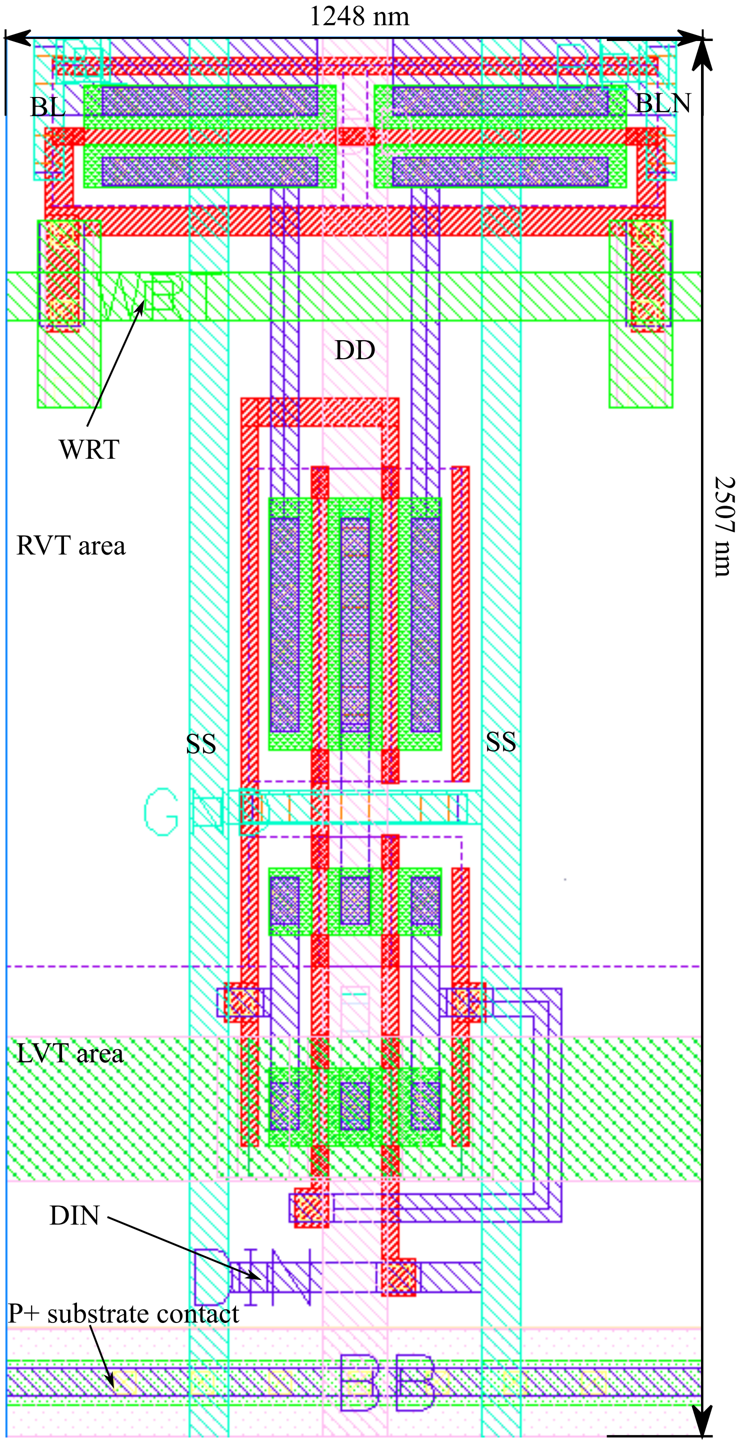


Figure 6.6: Write driver circuit layout.

6.1.4 Layout Arrangement of the SRAM Cache

A 32 kb SRAM macro with 64 I/O word-access can be formed by stacking 16 sub-arrays of 2 kb, 64 pre-charge circuits, 64 sense amplifiers and 64 write drivers, as shown in Fig. 6.7. The dimensions of the total 32 kb SRAM macro are 0.194 mm and 0.080 mm of height and width, respectively, occupying a chip area of 0.015 mm^2. In this way, we implement the 128 kb SRAM cache by using 4 macros of 32 kb.

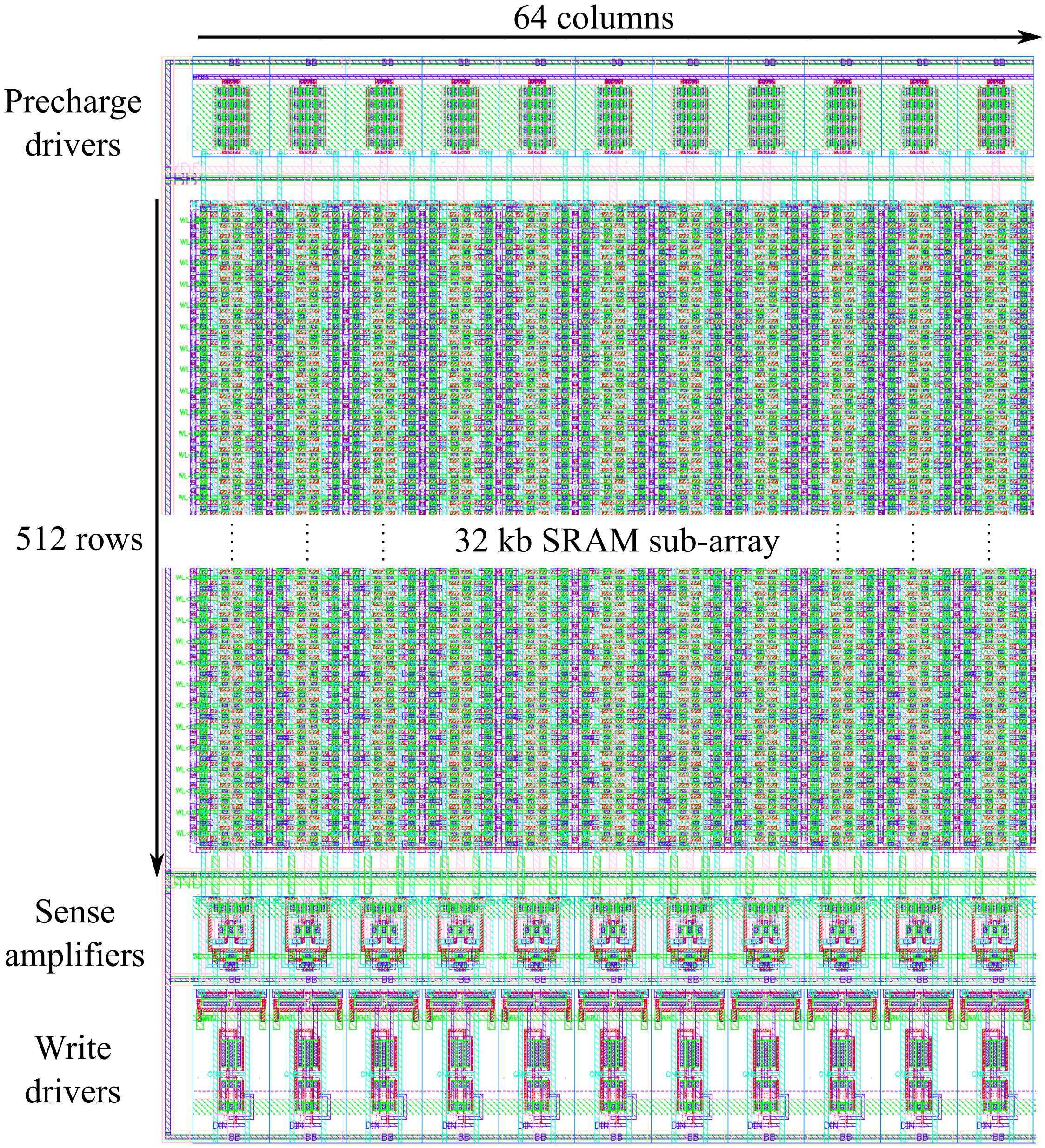


Figure 6.7: Layout arrangement of the 32 kb SRAM macro.

6.2 Standby Mode Simulations

In view of the fact that the SRAM cache is not used all the time in active mode (read and write operations), the standby is an essential mode of operation to save energy consumption since the static leakage can be reduced. Fig. 6.8 shows the simulation results of the 128 kb SRAM leakage current as a function of the supply voltage. Observe that by considering the typical corner (TT) and decreasing the supply voltage from 450 mV to 250 mV, the leakage current is reduced from 14.75 μA to 8.85 μA, which represents a relative reduction of 40% on the static leakage. The minimum as possible supply voltage is determined by the HSNM, and their simulation results are shown in Figs. 6.9(a) and 6.9(b) for V_{DD}= 450 mV and V_{DD}= 250 mV, respectively. In order to produce high yield under the manufacturing process variations, we choose V_{DD}= 250 mV as the minimum supply since the simulated worse-case HSNM resulted 10 mV. Another important aspect of this standby mode, besides the power consumption, is the transition time from the standby to the active mode, since the large capacitances that are involved in the SRAM cache increase this transition time. A long transition time may cause a significant speed overhead and this could prevent the use of the standby mode [10]. Fig. 6.10 shows the leakage current transient simulation of the 128 kb SRAM in both active and standby transitions. In this case, the minimum transition time to provide a peak current lower than 36% with

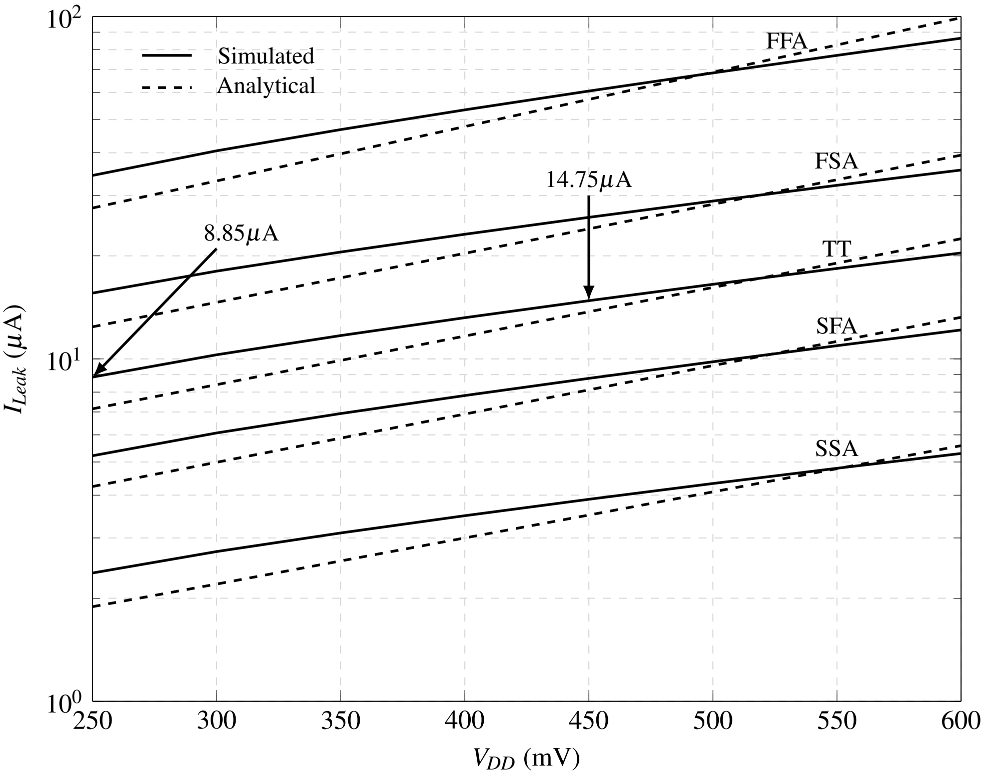
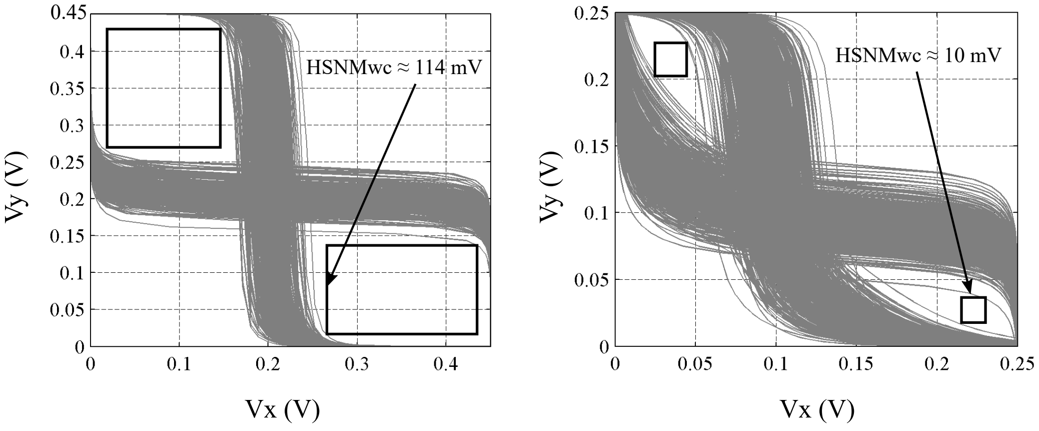


Figure 6.8: Simulated leakage current of the 128 kb SRAM as a function of the supply voltage.



(a) (b)

Figure 6.9: Voltage transfer curve simulations of the 6T-SRAM cell in hold mode: (a) V_{DD}= 0.45 V; (b) V_{DD}= 0.25 V

respect to the difference between the active and standby currents, is approximately 6 μs, which is 140 times slower than the operation frequency of 23 MHz.

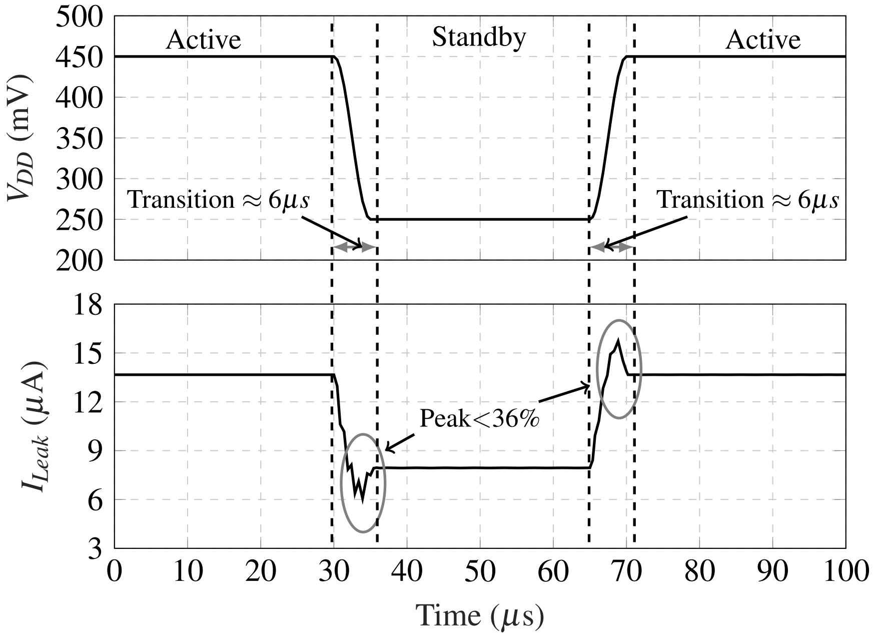


Figure 6.10: Transient simulation of the leakage current between transitions of active and standby modes.

6.3 Active Mode Simulations

6.3.1 Test Bench Considerations

It is necessary to simulate the logical information “0” and “1” with the same probability in write and read operations. This can be easily implemented by generating a word of 64 bits using the following Matlab line for each word to be written:

```
wordBits = (randn(1,64)>0)*VDD;
```

The resulting vector *wordBits*, in combination with the independent source “$\mathrm{V_{pwlf}}$” of Spectre[1] simulator, can be used to create the input signals in the desired time range.

6.3.2 Transient Behavior of the SRAM Cache

Due to the high computational cost that the 128 kb SRAM requires to simulate its transient behavior, we separate these simulations in macros of 32 kb (64 columns and 512 rows), which requires approximately 23 hours to complete 10 operations (5 reads and 5 writes) when its extracted[2] circuit is used. It is a good estimative since only one macro operates in active mode at the same time, while the other three operate in hold mode separated by a word-line decoder (see Fig. 4.2). On the other hand, the simulations need to include the mismatching effects since it is crucial to determine the SRAM proper operation and power consumption, and for this reason all transient simulations presented in this section include one run of mismatch variations.

6.3.2.1 32 kB SRAM Macro at Supply Voltage 0.45 V

As shown in the VTC Monte Carlo simulations of Fig. 6.11, for the designed 6T-cell, we chose V_{WL}= 0.85V_{DD} in both read and write operations, thereby providing positive static noise margins at supply voltage 450 mV. Fig. 6.12 shows the transient simulation results for the 32 kb SRAM macro at TT corner by using only schematic circuit, in which 5 write and 5 read operations are executed. The 64 bits of each bus DIN (data input), BL’s (BL and BLN), Q’s (internal nodes of the memory cells Q and QN) and OUT (read outputs by the SA) are overlapped on the same plot in order to emphasize the mismatch variability. As can be observed, there are no errors during write and read operations. However, the differential voltages between bit-lines in read operation discharge much faster than does the SA maximum input offset of 50 mV (see Fig. 5.7). This result indicates that the needed time to activate the SA can be decreased, and hence the operation frequency increased. Fig. 6.13 shows the bit-line voltages by varying the process corners (see Fig. 2.6) in cases:

[1] Circuit simulator of Cadence software (https://www.cadence.com).

[2] Extracted circuit is generated from the implemented layout, and includes parasitic components, such as resistances and capacitors.

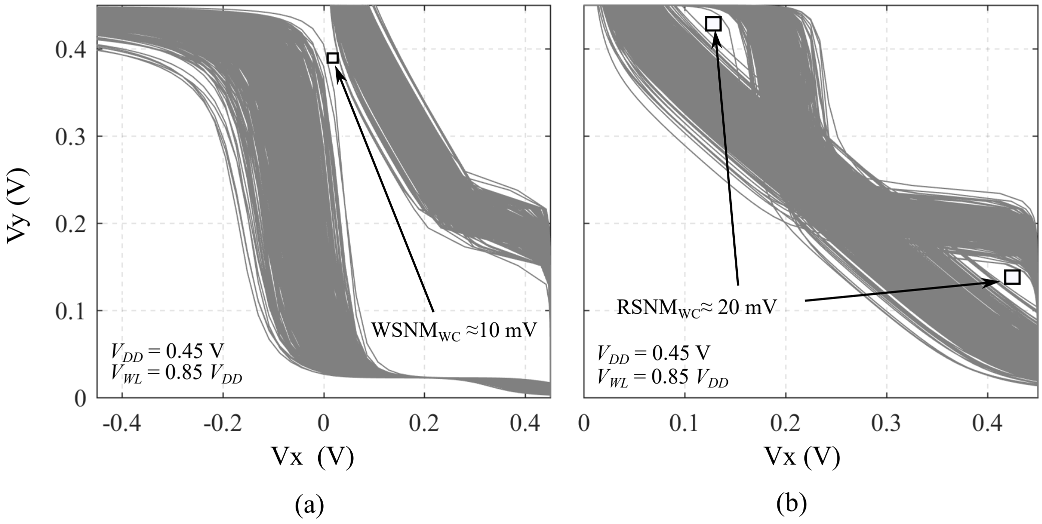


Figure 6.11: Voltage transfer curve Monte Carlo simulations, including global and local variability, of the 6T-SRAM cell: (a) write operation; (b) read operation.

SSA, SFA, TT, FSA, FFA and TT+EXT (TT corner including extracted components). As expected, the slow-slow (SSA) corner is critical for the system speed in view of the fact that the times T_{PC} (pre-charge time), T_{SAEN} (time to enable the SA), T_{SA} (SA delay) and T_{WD} (write driver time) achieve their maximums. In order to develop an estimative of the nominal frequency that the SRAM can operate properly at 450 mV, we found the maximum of T_{PC}, T_{SAEN} and T_{WD} when extracted components are included and TT process corner is used, which is denoted as TT+RC in Fig. 6.13. Hence the minimum cycle time, also analyzed previously in Eq. (4.13), is determined by

$$T_{CLK} = \max[T_{WD} + T_{PC}, T_{SAEN} + \max(T_{PC}, T_{SA})] \approx 25 \text{ ns} \tag{6.1}$$

where T_{WD}≈ 10 ns, T_{PC}≈ 7 ns, T_{SAEN}≈ 16 ns, T_{SA}≈ 3 ns and we add an additional time of 2 ns to cover the rise and fall of the signal. The time T_{SAEN} was determined as the maximum time that the differential voltage between bit-lines achieves 50 mV (V_{OS}) as indicated in Fig. 6.13. Therefore, Fig. 6.14 shows the transient simulation of the 32 kB SRAM macro at SSA corner by increasing the frequency to 1/T_{CLK}= 40 MHz.

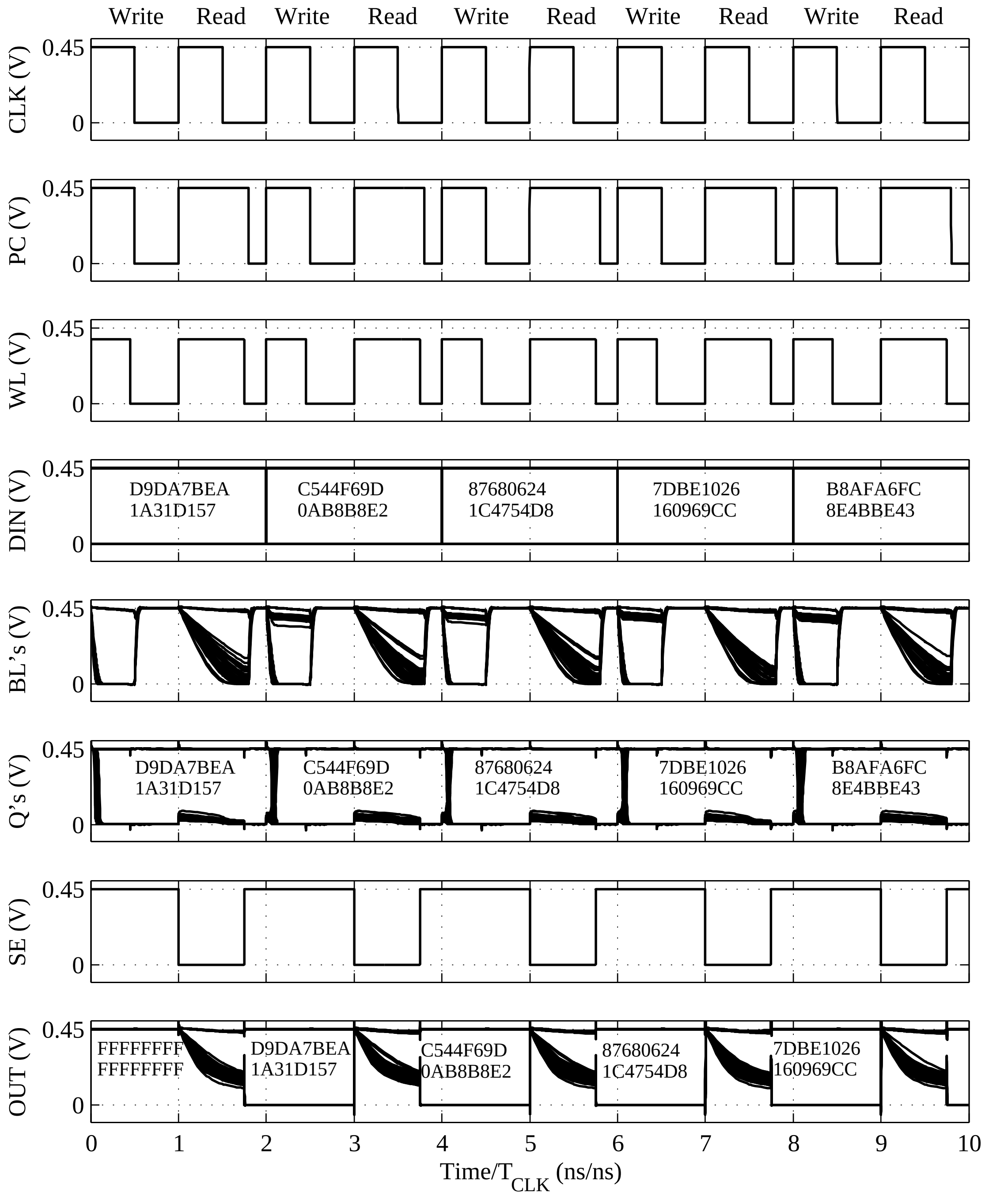


Figure 6.12: Transient simulation for 32 kb SRAM macro at V_{DD}= 0.45 V, T_{CLK}= 43.16 ns, TT process corner and one run of mismatch variability.

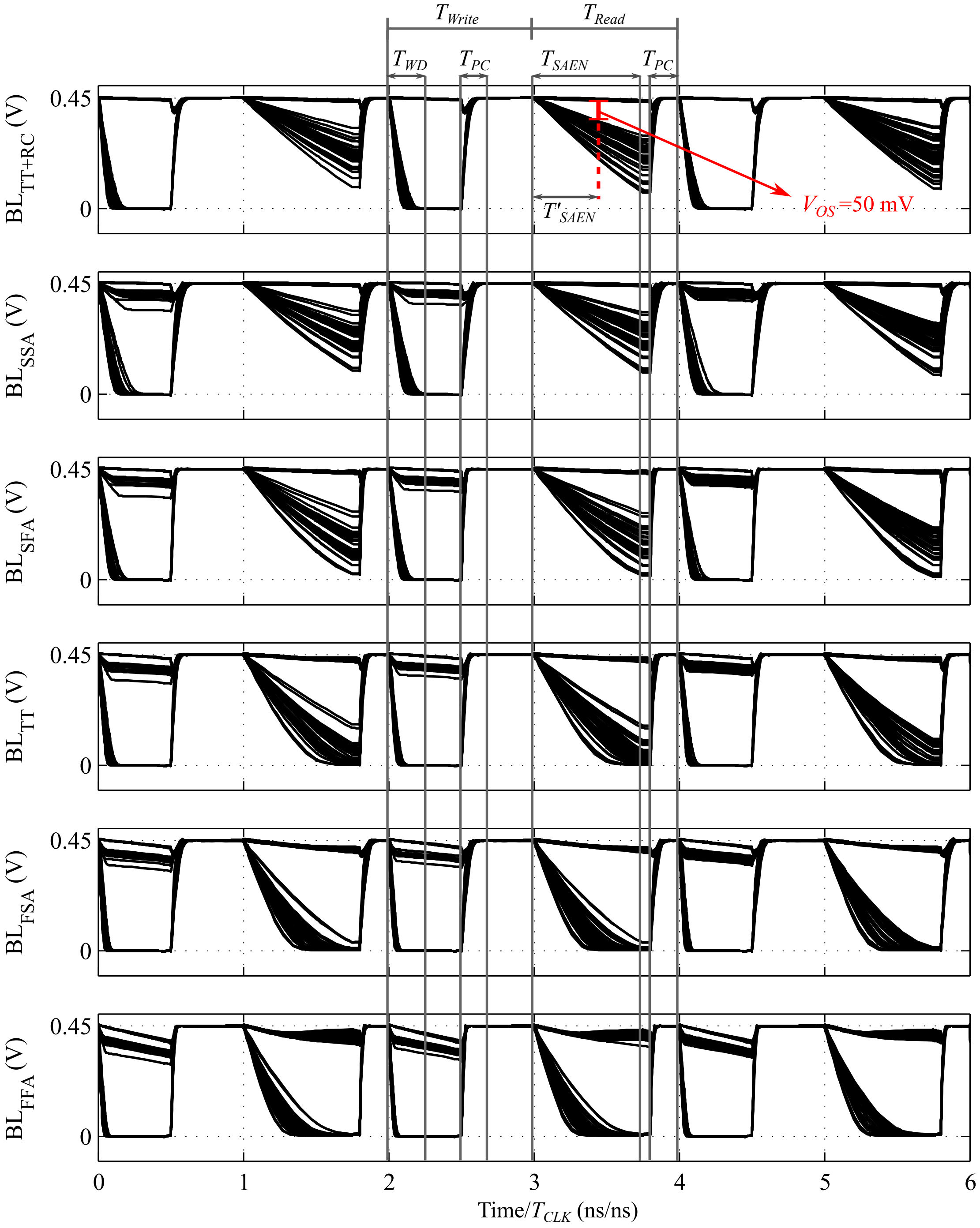


Figure 6.13: Bit-lines (64 bits overlapped) transient behavior at all process corners for 32 kb SRAM macro at V_{DD}= 0.45 V, T_{CLK}= 43.16 ns (23 MHz operation), and one run of mismatch variability.

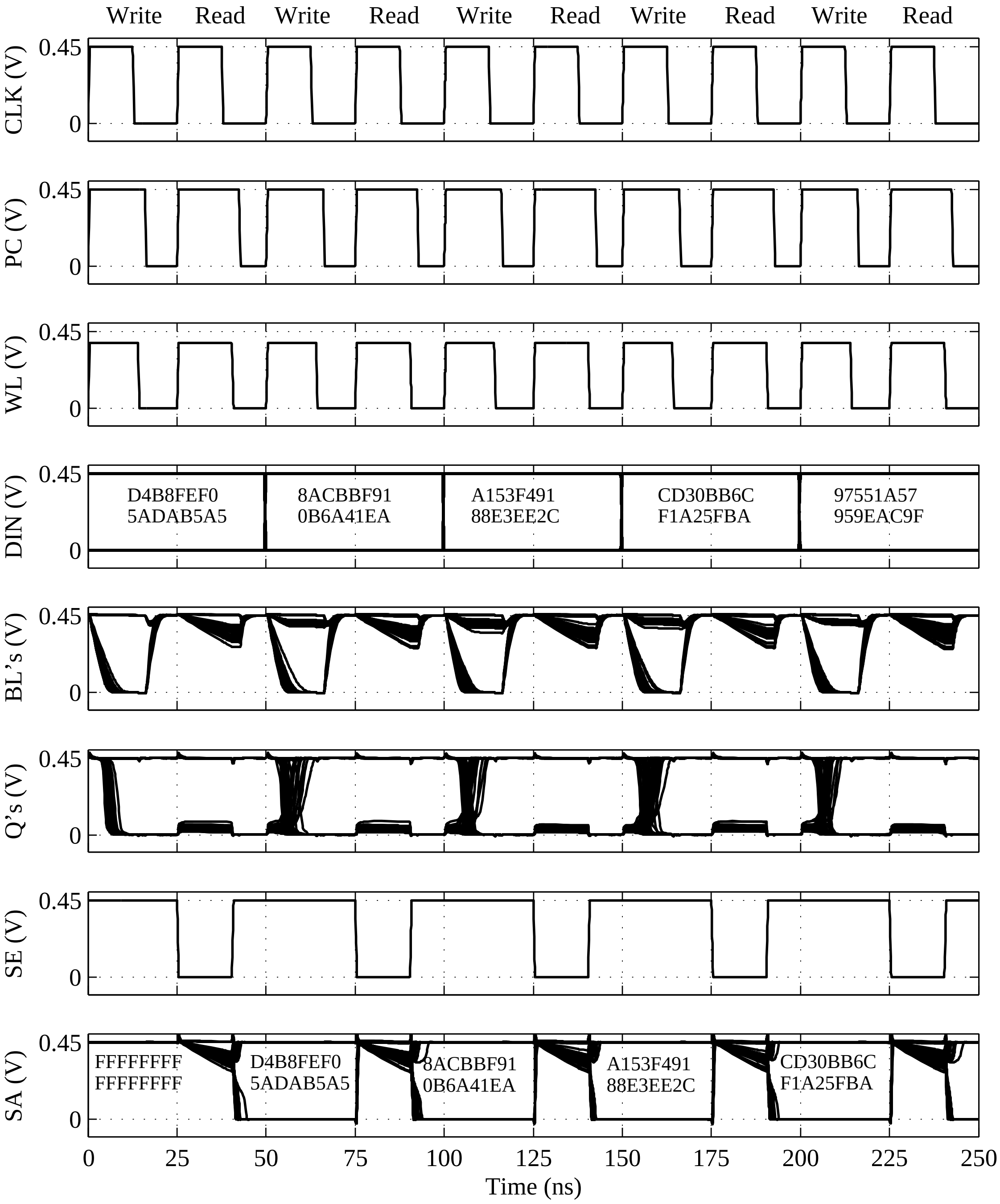


Figure 6.14: Transient simulation for 32 kb SRAM macro at V_{DD}= 0.45 V, operation frequency 40 MHz, SSA process corner and one run of mismatch variability.

6.3.3 Energy per Operation of the 128 kb SRAM

In view of the fact that many voltage sources are used to control the SRAM in our test bench schematic, we estimate the power consumption by simulating the current drained to the ground (I_{GND}) of the SRAM as shown in Fig. 6.15. Hence by multiplying this current by the supply voltage we obtain

$$\text{Power}(t) = V_{DD}(t) I_{GND}(t) \tag{6.2}$$

in μW. However, we are interested in estimating the average energy per operation, also analyzed in Section 4.6, which can be obtained by integrating the power consumption during a fixed number of operation periods (N_O) as

$$E_T = \frac{1}{N_O} \int_0^{N_O T_{CLK}} \text{Power}(t)\mathrm{d}t \tag{6.3}$$

in pJ/word-access. Usually, this energy is specified as an average of write and read operations, since as can be seen in the simulation result of Fig. 6.16, these operations requires different power consumptions. Observe that the pre-charge procedure does not have important influence on the energy, which is 5.3% and 3.41% of the total energies of write and read operation, respectively. On the other hand, the energy in write operation is higher than that of the read energy, mainly owing to fact that in write operation the bit-lines are discharged completely. Table 6.1 shows the simulated active (1 macro of 32 kb) and static (3 macros of 32 kb) energies per operation for the 128 kb SRAM cache operating at 23 MHz and 0.45 V. The obtained total energy of 795.6 fJ/word-access at TT corner is in close agreement with the one of 738 fJ/word-access predicted by our analytical models in Chapter 4. These results can also be observed in the histogram of Fig. 6.17, which shows how the energy is affected by the process corner variations. Note that static energy is the most affected in view of the fact that it has an exponential influence on the threshold volt-

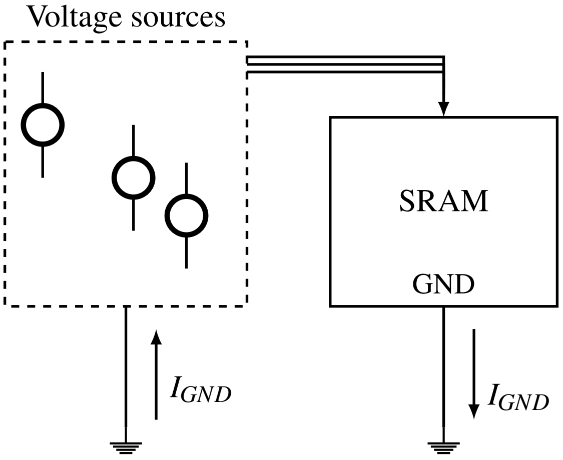


Figure 6.15: Power consumption test bench.

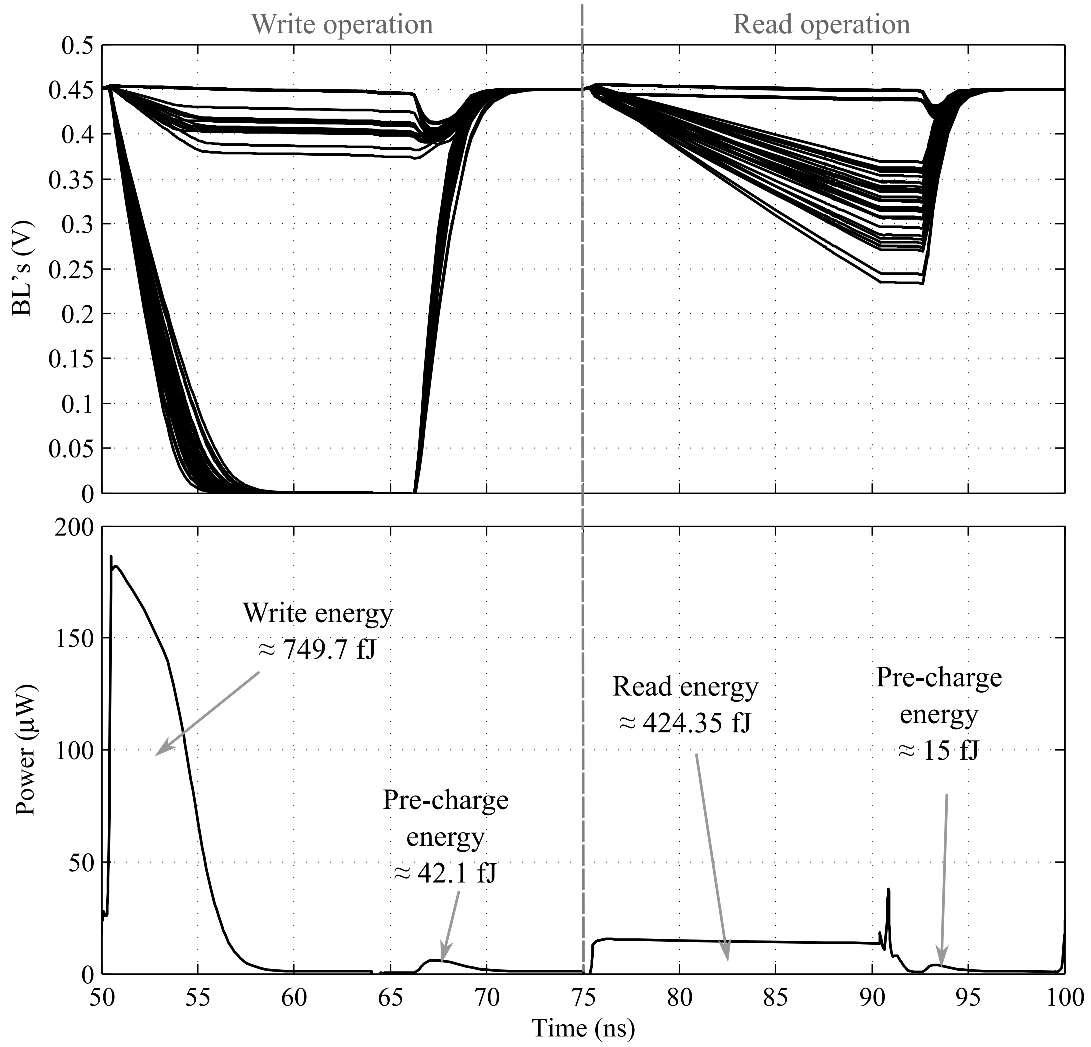


Figure 6.16: Bit-lines voltages and total power consumption of the 32 kb SRAM macro in active mode, at operation frequency of 40 MHz and supply voltage of 0.45 V.

age. However, the parasitic capacitances and resistances of layout implementation have an important influence in SRAM design, and in Table 6.1 we also show the energy when these effects are included. As expected, the active component increases 11.66% mainly by the reason that higher capacitances on bit-lines are included. On the other hand, both metal and transistor parasitic resistances produce a static component decrease of 36.2%.

As we determined in Eq. (6.1), the frequency of operation at supply voltage 0.45 V can be increased up to 40 MHz. Therefore, in Table 6.2 are shown the energy per operation results when the frequency is 40 MHz, and in Fig. 6.18 an histogram of these energies as a function of the process corner is presented. Hence two important observations can be made compared with the operation case of 23 MHz: the active energy decreases due to the fact that the bit-lines in read operation are less discharged, and on the other hand the static energy decreases since the period of the operation was decreased (see Eq. (4.17)). In this way, the post-layout simulations show that the implemented 128 kB SRAM cache

Table 6.1: Energy per operation simulations of the 128 kb SRAM cache at V_{DD}= 0.45 V, $V_{WL,read}$=$V_{WL,write}$= 0.85V_{DD}, V_{BB}=V_{Bp}=V_{Bp}=0 V, operation frequency of 23 MHz and using one run of mismatching variations.

Corner	Simulation time (hours)	Active energy 1x32 kb macro (fJ/word-access)	Static energy 3x32 kb macros (fJ/word-access)	Total energy 128 kb cache (fJ/word-access)
Using only schematic circuit				
TT	≈ 8	541.0	254.6	795.6
FSA	≈ 8	610.1	445.5	1055.6
SFA	≈ 8	468.0	152.4	620.4
FFA	≈ 8	785.2	1029.7	1814.9
SSA	≈ 8	389.6	68.15	457.8
Including extracted RC components from layout				
TT+RC	≈ 20	604.8	162.5	767.3

can operate at 40 MHz with a supply voltage of 0.45 mV, consuming an average energy per operation of 603.6 fJ/word-access.

Table 6.2: Energy per operation simulations of the 128 kb SRAM cache at V_{DD}= 0.45 V, $V_{WL,read}$=$V_{WL,write}$= 0.85V_{DD}, V_{BB}=V_{Bp}=V_{Bp}=0 V, operation frequency of 40 MHz and using one run of mismatching variations.

Corner	Simulation time (hours)	Active energy 1x32 kb macro (fJ/word-access)	Static energy 3x32 kb macros (fJ/word-access)	Total energy 128 kb cache (fJ/word-access)
Using only schematic circuit				
TT	≈ 8	427.8	146.5	574.3
FSA	≈ 8	506.5	258.1	764.6
SFA	≈ 8	370.1	88.3	458.4
FFA	≈ 8	638.3	596.4	1234.7
SSA	≈ 8	319.6	39.48	359.1
Including extracted RC components from layout				
TT+RC	≈ 23	509.5	94.12	603.6

6.3.4 Frequency *versus* Supply Voltage

The minimum SRAM period of operation, which was determined by Eq. (6.1) for 450 mV, can also be estimated for a range of supply voltages. Hence Table 6.3 shows the worst-case times of pre-charge, write, read and sense operations obtained at supply voltages from 400 mV to 600 mV. Consequently, we estimate the minimum period of operation T_{CLK}, which defines the maximum operation frequency of the SRAM as F_{CLK} = 1/T_{CLK}.

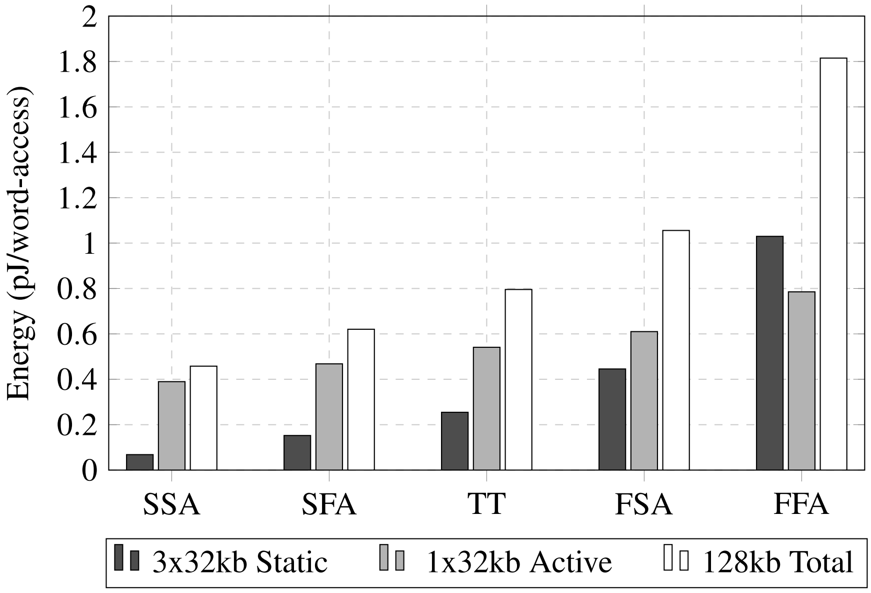


Figure 6.17: Energy per operation simulation results at 23 MHz and 0.45 V.

Table 6.3: Worst-case times of the SRAM operation obtained from simulations, by using TT process corner, mismatch variations and extracted RC component from layout.

V_{DD} (mV)	T_{PC} (ns)	T_{WD} (ns)	T_{SAEN} (ns)	T_{SA} (ns)	T_{CLK} (ns)
400	12	30.5	51	11	66.7
450	7	10	16	3	25
500	3	5	6	1	11
550	2.5	3.5	3.2	0.9	8
600	1.7	2.5	2.4	0.5	5

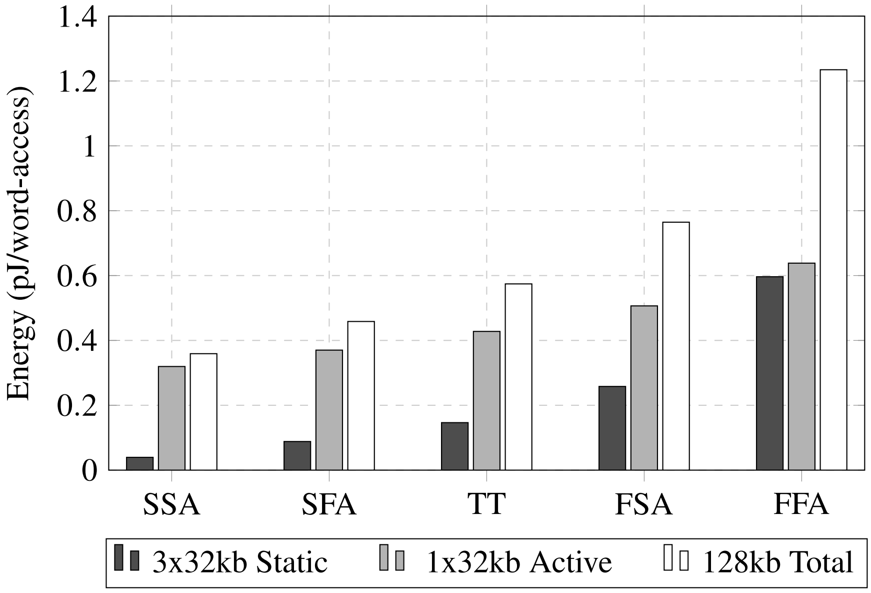


Figure 6.18: Energy per operation simulation results at 40 MHz and 0.45 V.

6.4 Summary of the SRAM Performance

Fig. 6.19 shows the maximum frequency and energy per operation versus the supply voltage obtained by post-layout simulations, using mismatch variations and TT corner. Fig. 6.19 includes the major performance parameters for our SRAM application, since the ratio between energy and frequency is the popular figure of merit (FOM) when the interest is low power consumption [88, 89]. Observe that a MEP of 603.6 fJ/word-access is achieved at supply voltage of 450 mV and operation frequency of 40 MHz. In order to compare this work with others reported in the literature, we define the FOM as similar as [88, 89], that is,

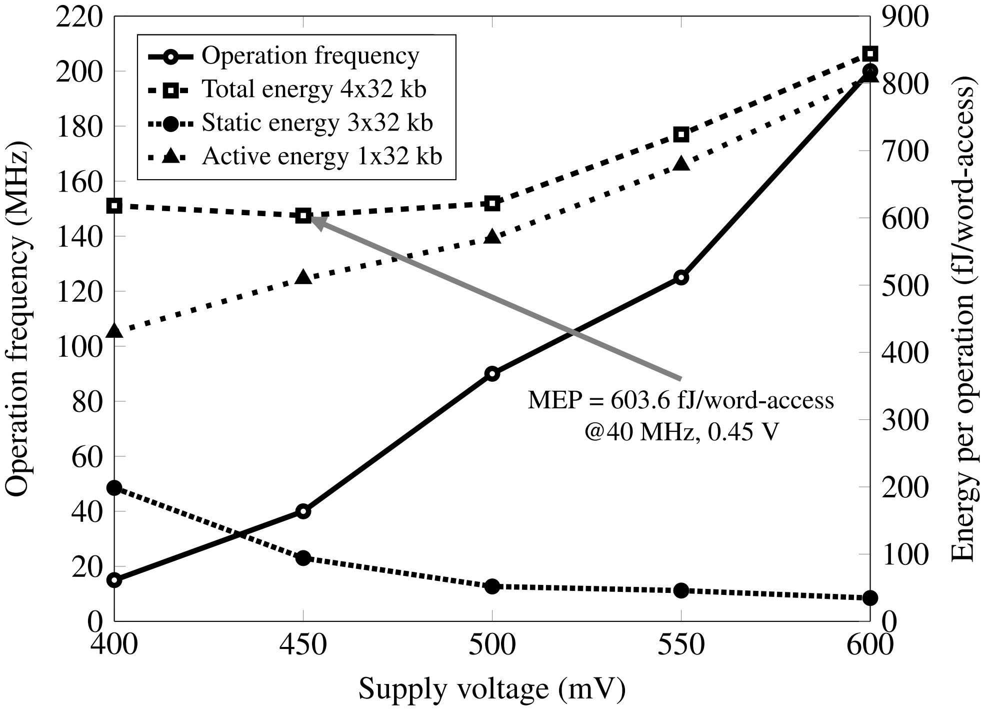


Figure 6.19: Maximum operation frequency and energy per operation *versus* supply voltage, by using the same simulation conditions of Table 6.3.

$$\mathrm{FOM} = \frac{F_{CLK}}{E_T A_{BC}} \quad \left[\frac{\mathrm{MHz}}{(\mathrm{pJ/word\text{-}access}) \cdot \mu\mathrm{m}^2}\right] \quad , \tag{6.4}$$

where E_T is the average energy per operation and A_{BC} is the bit-cell area. Table 6.4 shows the simulated performance of our work compared with other reported ULP SRAM designs. The increasing of pull-up and pull-down transistor lengths (see Fig. 6.1) produces a higher bit-cell area compared with other works. This increasing helped to reduce considerably the static power as well as the average energy per operation consumption. On the other hand, the obtained frequency of operation achieve similar orders compared with other works. The trade-offs involving energy, bit-cell area and frequency can be well appreciated by the FOM results in Table 6.4 obtained from Eq. (6.4). The excellent FOM obtained by simulation results shows that the implemented circuit can achieve a significant improvement compared with the state-of-the-art ULP SRAMs. It is important to note that the other works achieve lower minimum supply voltages than that obtained in our work, but these are achieved by using some additional techniques to improve noise margins, such as dynamic forward body-biasing (DFBB) [42]. However, our approach improves the FOM without any additional techniques, only by using an energy-efficient sizing of the 6T-SRAM cell and the SPW structure at fixed reverse body-biasing voltage of 0 V.

Table 6.4: Performance comparisons with other reported ULP SRAM designs.

Parameter	This work	[42]	[43]	[88]	[44]
Technology	28 nm FD-SOI				20 nm Bulk
Bit-cell structure	6T	6T	10T	7T	N/A
Minimum supply voltage (V)	0.40	0.36	0.35	0.24	0.6
Bit-cell area (μm^2)$*^1$	0.434	0.232	0.384	0.261	N/A
Memory Size (kb)	128	128	64	128	128
Word size (bits)	64	64	32	64	32
Freq. (MHz)$*^2$	15	9	13	15	N/A
Energy (pJ/word-access)$*^3$	0.603	3.36	4.60	2.30	4.30
FOM $\left(\frac{\text{MHz}}{(\text{pJ/word-access})\mu\text{m}^2}\right)*^2$	55.9	11.5	7.35	25.0	N/A

$*^1$ Using standard logic rules $*^2$ Evaluated at minimum supply voltage

$*^3$ Normalized to memory size 128 kb and word size 64

Chapter 7

Conclusions

This thesis was mainly focused on the development of analytical expressions for the major performance parameters of the SRAM cache implemented in 28 nm FD-SOI CMOS, to explore the transistor dimensions at low computational cost, thereby producing an energy-efficient SRAM cache, without losing focus on stability and speed. This chapter summarizes the contributions of this research and discusses future directions.

7.1 Parameter extraction for 28 nm FD-SOI CMOS

As shown in Chapter 2, a function that provides the transistor parameters (I_{so}, V_{To}, λ, η, n and $\sigma_{V_{To}}$) by choosing the transistor dimensions (W and L) and process corner (TT, FS, SF, FF and SS) was implemented for both regular N-type and flipped-well P-type transistors (see Fig. 2.6). In view of the fact that this thesis focused on the development of analytical models, the implemented function was useful to evaluate our expressions, thereby enabling transistor sizing design at low computational cost. For instance, in the SRAM cell sizing of Fig. 3.18, simulations spent a total time of 2.43 hours, whereas the predicted ones obtained by Matlab functions required only 10.73 seconds.

7.2 Analytical Modeling of SNM

In order to analyze the 6T-SRAM stability, in Chapter 3 we presented expressions to predict HSNM, RSNM and WSNM at sub-threshold operation, in which the variability of this margins under the manufacturing process was considered. Matlab routines to explore their values at low computational cost were implemented and can be seen in Appendix A.2.

Close agreement expressions useful to predict the HSNM that include DIBL and body biasing effects were proposed in Section 3.1 and published in [45]. These mainly contributed improvements of the model for nanometer technologies compared with previous models reported in the literature [17, 18]. In this way, in [68] we also published an analytical region idea (see Fig. 3.4), whose purpose is to determine boundaries inside which the supply voltage produce a feasible operation of the 6T-SRAM cells under manufacturing process variations.

Many analytical models that predict the statistical behavior of HSNM and RSNM have been well studied in the literature [20]. However, to the best of our knowledge, no analytical solution for the conventional definition of WSNM has so far been reported in the literature. In Section 3.3 we developed an analytical expression for WSNM at sub-thrseshold operation. Consequently, having the explicit expressions for WSNM and RSNM, we proposed an alternative 6T-SRAM design parameter Γ, whose role is to control the well-known trade-off between read and write cell margins. By relating Γ to pull-up and pull-down transistor lengths, and also considering the influences of the widths on the local variability, we carried out a non-traditional sizing procedure (see Fig 3.18) for 6T-SRAM cells implemented in 28 nm FD-SOI CMOS process.

7.3 Energy-efficient 6T-SRAM sizing

In Chapter 4 we presented a complete analytical model to assess the SRAM cache performance. The major system metrics, such as the static and dynamic energies per operation, static noise margins, operation frequency and leakage currents were carefully modeled. By taking advantage of these low computational cost models, we explored the performance metrics for a 128 kb SRAM cache implemented with the simple 6T-SRAM cell. In this way, an energy-efficient cell sizing was proposed, in which the transistors are designed with the same width (W_{pu}=W_{pd}=W_{at}=W_x), the same pull-up and pull-down lengths (L_{pu}=L_{pd}=L_x) and minimum access length (L_{at}=$L_{\min}$=30 nm) by using the SPW structure at reverse-body-biasing (RBB) (V_{BB}=V_{Bn}=V_{Bp}=0 V), and a reduced word-line voltage of approximately 0.85V_{DD} to provide a proper cell ratio in read operation. The post-layout simulation results presented in Chapter 6 showed that the 128 kb SRAM cache can achieve a MEP of 0.604 pJ/word-access at V_{DD}=450 mV and operation frequency of 40 MHz. As shown in Table 6.4, our approach can reduce by more than 74% the energy per operation compared with [42, 43, 88], mainly for the reason that these works employ the traditional “thin-cell” sizing [48], in which the leakage current increases by increasing transistor widths to adjust the cell margins that enable proper cell operation. Thus, as also currently suggested in [49] for saving more that 50% of the energy in digital circuits, the increase of transistor lengths can considerably reduce the cell static leakage. However, this cannot be applied to the access transistor since it must discharge higher bit-line capacitances, and for this reason we proposed an energy-efficient sizing by using the aforementioned variables L_x and W_x. On the other hand both area and minimum supply voltage increase, in comparison with what was reported in [42, 43, 88], as also shown in Table 6.4. However, the trade-offs involving area, frequency and energy per operation are the major metrics for ULP SRAMs, and can be analyzed by the FOM in Eq. (6.4), which resulted in an improvement of at least 55.3% with respect to designs reported in [42, 43, 88].

A Matlab routine to explore the SRAM performance at low computational cost were implemented, as can be seen in Appendix A.1, which provides the values of static and dynamic energies per operation, the current ratio between the access transistor leakage and read currents (see Eq. (4.12)), and the cycle operation time.

7.4 SA input-offset at sub-threshold operation

Analytical models to estimate input-offset and yield of the classical 7T-LTSA at sub-threshold operation were developed in Chapter 5. According to [29, 50], we contributed mainly by simplifying the latch input-offset expression in sub-threshold operation (see Eq. (5.14)), which allows the prediction of the SA yield. As a result of this analysis, the metastability voltages were carefully modeled, and as can be seen in Fig. 5.3, an approximation was proposed based on the intersection of the tangent curves at the inverter centers, for determining these voltages with excellent agreement (see Fig. 5.4).

7.5 Future Directions

In view of the fact that the proposed ULP SRAM improves the energy per operation mainly by the use of an energy-efficient 6T-SRAM cell sizing, more complex techniques and circuits could be used to further improve the SRAM figure of merit, such as dynamic body-biasing (DBB) [42, 90], improved sense amplifier topologies to reduce the input-offset [26, 40] thereby increasing the frequency of operation, write and read assist techniques [91, 92] to reduce the power consumption of operations, data prediction in the read path to save bit-line switching power [42], among other solutions.

The resulting Matlab routines of this work, could be used to implement a CAD (computer-aided design) interface to assist on the SRAM design. In this way, the parameter extractions for flipped-well NMOS and regular PMOS should be included to our routines in order to analyze other techniques and structure combinations in the 28 nm FD-SOI CMOS process.

Energy-efficient processors for wearable sensor nodes and biological signals processing techniques have been reported [93–96] as a direct application of ULP SRAMs. The study of these kinds of applications can be a valuable continuity of this research, since it requires a collaboration team composed by students and researchers in many micro-electronics/electronics areas, such as digital and analog circuit design, signal processing, algorithm development, etc.

Bibliography

[1] ROSSI, D., PULLINI, A., LOI, I., et al. "A 60 GOPS/W, -1.8 v to 0.9 v body bias ULP cluster in 28 nm UTBB FD-SOI technology", *Solid. State. Electron.*, v. 117, pp. 170–184, 2016. ISSN: 00381101. doi: 10.1016/j.sse.2015.11.015. Disponível em: <`http://dx.doi.org/10.1016/j.sse.2015.11.015`>.

[2] MAYR, C., PARTZSCH, J., NOACK, M., et al. "A Biological-Realtime Neuromorphic System in 28 nm CMOS Using Low-Leakage Switched Capacitor Circuits", *IEEE Trans. Biomed. Circuits Syst.*, v. 10, n. 1, pp. 243–254, feb 2016. ISSN: 1932-4545. doi: 10.1109/TBCAS.2014.2379294. Disponível em: <`http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7038235`>.

[3] ANDERSSON, O., CHON, K., SORNMO, L., et al. "A 290 mV Sub- VT ASIC for Real-Time Atrial Fibrillation Detection", *IEEE Trans. Biomed. Circuits Syst.*, v. 9, n. 3, pp. 377–386, June 2015. ISSN: 1932-4545. doi: 10.1109/TBCAS.2014.2354054.

[4] CHEN, Y.-P., JEON, D., LEE, Y., et al. "An Injectable 64 nW ECG Mixed-Signal SoC in 65 nm for Arrhythmia Monitoring", *IEEE J. Solid-State Circuits*, v. 50, n. 1, pp. 375–390, Jan 2015. ISSN: 0018-9200. doi: 10.1109/JSSC.2014.2364036.

[5] ZORIAN, Y. "Embedded memory test and repair: infrastructure IP for SOC yield". In: *Test Conference, 2002. Proceedings. International*, pp. 340–349, 2002. doi: 10.1109/TEST.2002.1041777.

[6] SINGH, J., MOHANTY, S. P., PRADHAN, D. K. *Robust SRAM Designs and Analysis*. 1 ed. , Springer-Verlag New York, 2013. ISBN: 978-1-4614-0817-8.

[7] QIN, H., CAO, Y., MARKOVIC, D., et al. "Standby supply voltage minimization for deep sub-micron SRAM". In: *Microelectron. J.*, v. 36, pp. 789–800, 2005. doi: 10.1016/j.mejo.2005.03.003.

[8] NOURIVAND, A., AL-KHALILI, A. J., SAVARIA, Y. "Postsilicon tuning of standby supply voltage in srams to reduce yield losses due to parametric data-retention

failures", *IEEE Trans. Very Large Scale Integr. Syst.*, v. 20, n. 1, pp. 29–41, 2012. ISSN: 10638210. doi: 10.1109/TVLSI.2010.2093938.

[9] HUANG, G., QIAN, L., SAIBUA, S., et al. "An efficient optimization based method to evaluate the DRV of SRAM cells", *IEEE Trans. Circuits Syst. I Regul. Pap.*, v. 60, n. 6, pp. 1511–1520, 2013. ISSN: 15498328. doi: 10.1109/TCSI.2012.2226504.

[10] SIVA G. NARENDRA, A. C. A. *Leakage in Nanometer CMOS Technologies*. Series on Integrated Circuits and Systems. 1 ed. , Springer US, 2006. ISBN: 978-0-387-25737-2,978-0-387-28133-9.

[11] MORIFUJI, E., PATIL, D., HOROWITZ, M., et al. "Power Optimization for SRAM and Its Scaling", *IEEE Trans. Electron Devices*, v. 54, n. 4, pp. 715–722, apr 2007. ISSN: 0018-9383. doi: 10.1109/TED.2007.891869. Disponível em: <`http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4142890`>.

[12] ISLAM, A., HASAN, M. "A technique to mitigate impact of process, voltage and temperature variations on design metrics of SRAM Cell", *Microelectron. Reliab.*, v. 52, n. 2, pp. 405–411, 2012. ISSN: 00262714. doi: 10.1016/j.microrel.2011.09.034. Disponível em: <`http://dx.doi.org/10.1016/j.microrel.2011.09.034`>.

[13] ABU-RAHMA, M. H., ANIS, M. *Nanometer Variation-Tolerant SRAM*. New York, Springer New York, 2013. ISBN: 978-1-4614-1748-4.

[14] SEEVINCK, E., LIST, F., LOHSTROH, J. "Static-noise margin analysis of MOS SRAM cells", *IEEE J. Solid-State Circuits*, v. 22, n. 5, pp. 748–754, 1987. ISSN: 0018-9200. doi: 10.1109/JSSC.1987.1052809.

[15] CALHOUN, B. H., CHANDRAKASAN, A. P. "Static noise margin variation for sub-threshold SRAM in 65-nm CMOS", *IEEE J. Solid-State Circuits*, v. 41, n. 7, pp. 1673–1679, 2006. ISSN: 00189200. doi: 10.1109/JSSC.2006.873215.

[16] ONICIUC, L., ANDREI, P. "Sensitivity of static noise margins to random dopant variations in 6-T SRAM cells", *Solid. State. Electron.*, v. 52, n. 10, pp. 1542–1549, 2008. ISSN: 00381101. doi: 10.1016/j.sse.2008.06.029.

[17] ALIOTO, M. "Understanding DC behavior of subthreshold CMOS logic through closed-form analysis", *IEEE Trans. Circuits Syst. I Regul. Pap.*, v. 57, n. 7, pp. 1597–1607, 2010. ISSN: 15498328. doi: 10.1109/TCSI.2009.2034233.

[18] TAJALLI, A., LEBLEBICI, Y. "Design Trade-offs in Ultra-Low-Power Digital Nanoscale CMOS", *IEEE Trans. Circuits Syst. I Regul. Pap.*, v. 58, n. 9, pp. 2189–2200, Sept 2011. ISSN: 1549-8328. doi: 10.1109/TCSI.2011.2112595.

[19] ALORDA, B., TORRENS, G., BOTA, S., et al. "Adaptive static and dynamic noise margin improvement in minimum-sized 6T-SRAM cells", *Microelectron. Reliab.*, v. 54, n. 11, pp. 2613–2620, 2014. ISSN: 00262714. doi: 10.1016/j.microrel.2014.05.009. Disponível em: <`http://dx.doi.org/10.1016/j.microrel.2014.05.009`>.

[20] SAEIDI, R., SHARIFKHANI, M., HAJSADEGHI, K. "Statistical analysis of read static noise margin for near/sub-threshold SRAM cell", *IEEE Trans. Circuits Syst. I Regul. Pap.*, v. 61, n. 12, pp. 3386–3393, 2014. ISSN: 15498328. doi: 10.1109/TCSI.2014.2327334.

[21] JEONG, H., YANG, Y., LEE, J., et al. "One-sided static noise margin and gaussian-tail-fitting method for SRAM", *IEEE Trans. Very Large Scale Integr. Syst.*, v. 22, n. 6, pp. 1262–1269, 2014. ISSN: 10638210. doi: 10.1109/TVLSI.2013.2268543.

[22] TANAKA, C., SAITOH, M., OTA, K., et al. "Analysis of static noise margin improvement for low voltage SRAM composed of nano-scale MOSFETs with ideal subthreshold factor and small variability", *Solid. State. Electron.*, v. 109, pp. 58–62, 2015. ISSN: 00381101. doi: 10.1016/j.sse.2015.03.013. Disponível em: <`http://dx.doi.org/10.1016/j.sse.2015.03.013`>.

[23] MUKHOPADHYAY, S., MAHMOODI, H., ROY, K. "Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscaled CMOS", *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, v. 24, n. 12, pp. 1859–1879, 2005. ISSN: 02780070. doi: 10.1109/TCAD.2005.852295.

[24] MUKHOPADHYAY, S., MAHMOODI, H., ROY, K. "Reduction of Parametric Failures in Sub-100-nm SRAM Array Using Body Bias", *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, v. 27, n. 1, pp. 174–183, jan 2008. ISSN: 0278-0070. doi: 10.1109/TCAD.2007.906995. Disponível em: <`http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4358298`>.

[25] VATAJELU, E. I., BOSIO, A., DILILLO, L., et al. "Analyzing the effect of concurrent variability in the core cells and sense amplifiers on SRAM read access failures". In: *Design Technology of Integrated Systems in Nanoscale Era*

(DTIS), 2013 8th International Conference on, pp. 39–44, March 2013. doi: 10.1109/DTIS.2013.6527775.

[26] NA, T., WOO, S. H., KIM, J., et al. "Comparative study of various latch-type sense amplifiers", *IEEE Trans. Very Large Scale Integr. Syst.*, v. 22, n. 2, pp. 425–429, 2014. ISSN: 10638210. doi: 10.1109/TVLSI.2013.2239320.

[27] WICHT, B., NIRSCHL, T., SCHMITT-LANDSIEDEL, D. "Yield and speed optimization of a latch-type voltage sense amplifier", *IEEE J. Solid-State Circuits*, v. 39, n. 7, pp. 1148–1158, July 2004. ISSN: 0018-9200. doi: 10.1109/JSSC.2004.829399.

[28] WOO, S.-H., KANG, H., PARK, K., et al. "Offset voltage estimation model for latch-type sense amplifiers", *IET Circuits, Devices & Systems*, v. 4, n. 6, pp. 503, 2010. ISSN: 1751858X. doi: 10.1049/iet-cds.2010.0092. Disponível em: <http://digital-library.theiet.org/content/journals/10.1049/iet-cds.2010.0092>.

[29] DO, A. T., KONG, Z. H., YEO, K. S. "Criterion to evaluate input-Offset voltage of a latch-Type sense amplifier", *IEEE Trans. Circuits Syst. I Regul. Pap.*, v. 57, n. 1, pp. 83–92, 2010. ISSN: 15498328. doi: 10.1109/TCSI.2009.2016182.

[30] FAN, M.-L., HU, V. P.-H., CHEN, Y.-N., et al. "Variability Analysis of Sense Amplifier for Subthreshold Ultra-Thin-Body SOI SRAM Applications", *IEEE Trans. Circuits Syst. II Express Briefs*, v. 59, n. 12, pp. 878–882, 2012.

[31] MOHAMMAD, B. S., SALEH, H., ISMAIL, M. "Design Methodologies for Yield Enhancement and Power Efficiency in SRAM-Based SoCs", *IEEE Trans. Very Large Scale Integr. Syst.*, v. 23, n. 10, pp. 2054–2064, 2015. ISSN: 10638210. doi: 10.1109/TVLSI.2014.2360319.

[32] KHALIL, D., KHELLAH, M., KIM, N. S., et al. "Accurate estimation of SRAM dynamic stability", *IEEE Trans. Very Large Scale Integr. Syst.*, v. 16, n. 12, pp. 1639–1647, 2008. ISSN: 10638210. doi: 10.1109/TVLSI.2008.2001941.

[33] KHALIL, D., KHELLAH, M., KIM, N. S., et al. "SRAM dynamic stability estimation using MPFP and its applications", *Microelectron. J.*, v. 40, n. 11, pp. 1523–1530, 2009. ISSN: 00262692. doi: 10.1016/j.mejo.2009.01.015.

[34] SINGHEE, A., RUTENBAR, R. A. "Statistical blockade: Very fast statistical simulation and modeling of rare circuit events and its application to memory design", *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, v. 28, n. 8, pp. 1176–1189, 2009. ISSN: 02780070. doi: 10.1109/TCAD.2009.2020721.

[35] AGHABABA, H., EBRAHIMI, B., AFZALI-KUSHA, A., et al. "Probability calculation of read failures in nano-scaled SRAM cells under process variations", *Microelectron. Reliab.*, v. 52, n. 11, pp. 2805–2811, 2012. ISSN: 00262714. doi: 10.1016/j.microrel.2012.04.022. Disponível em: <`http://dx.doi.org/10.1016/j.microrel.2012.04.022`>.

[36] CHEN, Y. H., CHAN, W. M., WU, W. C., et al. "A 16 nm 128 Mb SRAM in High- κ Metal-Gate FinFET Technology With Write-Assist Circuitry for Low-VMIN Applications", *IEEE J. Solid-State Circuits*, v. 50, n. 1, pp. 170–177, Jan 2015. ISSN: 0018-9200. doi: 10.1109/JSSC.2014.2349977.

[37] SONG, T., RIM, W., JUNG, J., et al. "A 14 nm FinFET 128 Mb SRAM With V_{MIN} Enhancement Techniques for Low-Power Applications", *IEEE J. Solid-State Circuits*, v. 50, n. 1, pp. 158–169, Jan 2015. ISSN: 0018-9200. doi: 10.1109/JSSC.2014.2362842.

[38] SONG, T., RIM, W., PARK, S., et al. "A 10 nm FinFET 128 Mb SRAM With Assist Adjustment System for Power, Performance, and Area Optimization", *IEEE Journal of Solid-State Circuits*, v. 52, n. 1, pp. 240–249, Jan 2017. ISSN: 0018-9200. doi: 10.1109/JSSC.2016.2609386.

[39] WICHT, B. *Current Sense Amplifiers*. Springer Berlin Heidelberg, 2003. ISBN: 978-3-642-05557-7.

[40] VERMA, N., A.P. CHANDRAKASAN. "A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy", *IEEE J. Solid-State Circuits*, v. 43, n. 1, pp. 141–149, 2008. ISSN: 0018-9200. doi: 10.1109/JSSC.2007.908005.

[41] JEONG, H., KIM, T., KANG, K., et al. "Switching pMOS Sense Amplifier for High-Density Low-Voltage Single-Ended SRAM", *IEEE Trans. Circuits Syst. I Regul. Pap.*, v. 62, n. 6, pp. 1555–1563, jun 2015. ISSN: 1549-8328. doi: 10.1109/TCSI.2015.2415171. Disponível em: <`http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7112585`>.

[42] BISWAS, A., CHANDRAKASAN, A. P. "Dynamic Body-Biasing and Output Data Prediction in 28nm FDSOI", v. 1, n. c, pp. 433–436, 2016.

[43] ABOUZEID, F., BIENFAIT, A., AKYEL, K. C., et al. "Scalable 0.35 V to 1.2 V SRAM Bitcell Design From 65 nm CMOS to 28 nm FDSOI", *IEEE J. Solid-State Circuits*, v. 49, n. 7, pp. 1499–1505, July 2014. ISSN: 0018-9200. doi: 10.1109/JSSC.2014.2316219.

[44] FUJIWARA, H., YABUUCHI, M., MORIMOTO, M., et al. "A 20nm 0.6V 2.1 μW/MHz 128kb SRAM with no half select issue by interleave wordline and hierarchical bitline scheme". In: *2013 Symp. on VLSI Technology*, pp. C118–C119, June 2013.

[45] OLIVERA, F., PETRAGLIA, A. "Analytic Modeling of Static Noise Margin Considering DIBL and Body Bias Effects". In: *IEEE Int. Symp. on Circuits and Syst. (ISCAS)*, pp. 489–492, May 2017. doi: 978-1-4673-6853-7/17/$31.00.

[46] GROSSAR, E., STUCCHI, M., MAEX, K., et al. "Read stability and writeability analysis of SRAM cells for nanometer technologies", *IEEE J. Solid-State Circuits*, v. 41, n. 11, pp. 2577–2588, 2006. ISSN: 00189200. doi: 10.1109/JSSC.2006.883344.

[47] ISLAM, A., HASAN, M. "Variability aware low leakage reliable SRAM cell design technique", *Microelectronics Reliability*, v. 52, n. 6, pp. 1247–1252, 2012. ISSN: 00262714. doi: 10.1016/j.microrel.2012.01.003. Disponível em: <http://dx.doi.org/10.1016/j.microrel.2012.01.003>.

[48] YAMAOKA, M., TSUCHIYA, R., KAWAHARA, T. "SRAM Circuit with Expanded Operating Margin and Reduced Stand-by Leakage Current Using Thin-BOX FD-SOI Transistors". In: *2005 IEEE Asian Solid-State Circuits Conference*, pp. 109–112, Nov 2005. doi: 10.1109/ASSCC.2005.251677.

[49] VEIRANO, F., SILVEIRA, F., NAVINER, L. "Asymmetrical length biasing for energy efficient digital circuits". In: *2017 IEEE 8th Latin American Symposium on Circuits Systems (LASCAS)*, pp. 1–4, Feb 2017. doi: 10.1109/LASCAS.2017.7948060.

[50] NIKOOZADEH, A., MURMANN, B. "An analysis of latch comparator offset due to load capacitor mismatch", *IEEE Trans. Circuits Syst. II Express Briefs*, v. 53, n. 12, pp. 1398–1402, 2006. ISSN: 10577130. doi: 10.1109/TCSII.2006.883204.

[51] SAKURAI, T., MATSUZAWA, A., DOUSEKI, T. *Fully-Depleted SOI CMOS Circuits and Technology for Ultralow-Power Applications*. 1 ed. , Springer US, 2006. ISBN: 978-0-387-29217-5.

[52] PLANES, N., WEBER, O., BARRAL, V., et al. "28nm FDSOI technology platform for high-speed low-voltage digital applications". In: *VLSI Technology (VLSIT), 2012 Symposium on*, pp. 133–134, June 2012. doi: 10.1109/VLSIT.2012.6242497.

[53] NIKOLIC, B., BLAGOJEVIC, M., THOMAS, O., et al. "Circuit design in nanoscale FDSOI technologies", *Proceedings of the International Conference on Microelectronics, ICM*, , n. Miel, pp. 3–6, 2014. doi: 10.1109/MIEL.2014.6842076.

[54] ENZ, C. C., VITTOZ, E. A. "An Analytical MOS Transistor Model Valid in All Regions of Operation and Dedicated to Low-Voltage and Low-Current Applications", v. 14, 1995.

[55] SCHNEIDER, M. C., GALUP-MONTORO, C. *CMOS Analog Design Using All-Region MOSFET Modeling*. NY, USA, Cambridge University Press, 2010. ISBN: 978-0-521-11036-5.

[56] JESPERS, P. In: *The Gm/ID Methodology, a Sizing Tool for Low-voltage Analog CMOS Circuits: The Semi-empirical and Compact Model Approaches*, Springer Publishing Company, Incorporated, 2012. ISBN: 1461425050, 9781461425052.

[57] GALUP-MONTORO, C., SCHNEIDER, M. In: *Mosfet Modeling for Circuit Analysis and Design*, World Scientific, 2007. ISBN: 978-981-256-810-6.

[58] GALUP-MONTORO, C., SCHNEIDER, M. C. "MOSFET Parameter Extraction". In: *Mosfet Modeling for Circuit Analysis and Design*, World Scientific Publishing Co. Pte. Ltd., cap. 11, River Edge, NJ, USA, 2007.

[59] VITTOZ, E. A. "Weak Inversion for Ultimate Low-Power Logic". In: Piguet, C. (Ed.), *Low-Power Electronics Design*, CRC Press, 2004.

[60] ALIOTO, M. "Ultra-low power VLSI circuit design demystified and explained: A tutorial", *IEEE Trans. Circuits Syst. I Regul. Pap.*, v. 59, n. 1, pp. 3–29, 2012. ISSN: 15498328. doi: 10.1109/TCSI.2011.2177004.

[61] GALUP-MONTORO, C., SCHNEIDER, M. C., MACHADO, M. B. "Ultra-low-voltage operation of CMOS analog circuits: Amplifiers, oscillators, and rectifiers", *IEEE Trans. Circuits Syst. II Express Briefs*, v. 59, n. 12, pp. 932–936, 2012. ISSN: 15497747. doi: 10.1109/TCSII.2012.2231042.

[62] TSIVIDIS, Y. *Operation and Modeling of the MOS Transistor*. 2 ed. , Oxford University Press, USA, 2003.

[63] SPRINGER, S. K., LEE, S., LU, N., et al. "Modeling of Variation in Submicrometer CMOS ULSI Technologies", *IEEE Trans. Electron Devices*, v. 53, n. 9, pp. 2168–2178, Sept 2006. ISSN: 0018-9383. doi: 10.1109/TED.2006.880165.

[64] NEUBERGER, G., WIRTH, G., REIS, R. "Process Variability". In: *Protecting Chips Against Hold Time Violations Due to Variability*, pp. 5–14, Dordrecht, Springer Netherlands, 2014. ISBN: 978-94-007-2427-3. doi: 10.1007/978-94-007-2427-3_2. Disponível em: <`http://dx.doi.org/10.1007/978-94-007-2427-3_2`>.

[65] PELGROM, M. J. M., TUINHOUT, H. P., VERTREGT, M. "Transistor matching in analog CMOS applications". In: *International Electron Devices Meeting 1998. Technical Digest (Cat. No.98CH36217)*, pp. 915–918, Dec 1998. doi: 10.1109/IEDM.1998.746503.

[66] YUAN, X., SHIMIZU, T., MAHALINGAM, U., et al. "Transistor Mismatch Properties in Deep-Submicrometer CMOS Technologies", *IEEE Trans. Electron Devices*, v. 58, n. 2, pp. 335–342, Feb 2011. ISSN: 0018-9383. doi: 10.1109/TED.2010.2090159.

[67] KURUDE, S., MITTAL, S., GANGULY, U. "Statistical Variability Analysis of SRAM Cell for Emerging Transistor Technologies", *IEEE Trans. Electron Devices*, v. 63, n. 9, pp. 3514–3520, 2016. ISSN: 00189383. doi: 10.1109/TED.2016.2590433.

[68] OLIVERA, F., PETRAGLIA, A. "Analytic Boundaries for 6T-SRAM Design in Standby Mode". In: *2016 29th Symposium on Integrated Circuits and Systems Design (SBCCI)*. IEEE, 2016.

[69] MAKINO, H., NAKATA, S., SUZUKI, H., et al. "Reexamination of SRAM cell write margin definitions in view of predicting the distribution", *IEEE Trans. Circuits Syst. II Express Briefs*, v. 58, n. 4, pp. 230–234, 2011. ISSN: 15497747. doi: 10.1109/TCSII.2011.2124531.

[70] SHARIFKHANI, M., SACHDEV, M. "SRAM Cell Stability: A Dynamic Perspective", *IEEE J. Solid-State Circuits*, v. 44, n. 2, pp. 609–619, 2009. ISSN: 0018-9200. doi: 10.1109/JSSC.2008.2010818.

[71] ALICE WANG, BENTON H. CALHOUN, A. P. C. In: *Sub-threshold Design for Ultra Low-Power Systems*, Springer, 2006. ISBN: 9780387335155, 9780387345017.

[72] CORSONELLO, P., FRUSTACI, F., PERRI, S. "Low-leakage SRAM wordline drivers for the 28-nm UTBB FDSOI technology", *IEEE Trans. Very Large Scale Integr. Syst.*, v. 23, n. 12, pp. 3133–3137, 2015. ISSN: 10638210. doi: 10.1109/TVLSI.2014.2384007.

[73] YAMAOKA, M., OSADA, K., ISHIBASHI, K. "0.4-V logic-library-friendly SRAM array using rectangular-diffusion cell and delta-boosted-array voltage scheme", *IEEE J. Solid-State Circuits*, v. 39, n. 6, pp. 934–940, 2004. ISSN: 00189200. doi: 10.1109/JSSC.2004.827796.

[74] MORADI, F., PANAGOPOULOS, G., KARAKONSTANTIS, G., et al. "Multi-level wordline driver for low power SRAMs in nano-scale CMOS technology". In: *2011 IEEE 29th International Conference on Computer Design (ICCD)*, pp. 326–331, Oct 2011. doi: 10.1109/ICCD.2011.6081419.

[75] YABUUCHI, M., TSUKAMOTO, Y., MORIMOTO, M., et al. "13.3 20nm High-density single-port and dual-port SRAMs with wordline-voltage-adjustment system for read/write assists". In: *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 234–235, Feb 2014. doi: 10.1109/ISSCC.2014.6757414.

[76] NAYAK, D., PRIYABRATA, D., MAHAPATRA, K. "Current Starving the SRAM Cell : A Strategy to Improve Cell Stability and Power", *Circuits Syst. Signal Process.*, v. 36, n. 8, pp. 3047–3070, 2017. ISSN: 1531-5878. doi: 10.1007/s00034-016-0466-5. Disponível em: <`"http://dx.doi.org/10.1007/s00034-016-0466-5`>.

[77] CALHOUN, B. H., CHANDRAKASAN, A. P. "A 256-kb 65-nm Sub-threshold SRAM Design for Ultra-Low-Voltage Operation", *IEEE Journal of Solid-State Circuits*, v. 42, n. 3, pp. 680–688, March 2007. ISSN: 0018-9200. doi: 10.1109/JSSC.2006.891726.

[78] CALHOUN, B. H., WANG, A., CHANDRAKASAN, A. "Modeling and sizing for minimum energy operation in subthreshold circuits", *IEEE J. Solid-State Circuits*, v. 40, n. 9, pp. 1778–1785, 2005. ISSN: 00189200. doi: 10.1109/JSSC.2005.852162.

[79] FRANZON, P. D. "Energy Consumption Modeling and Optimization for SRAM's", *IEEE J. Solid-State Circuits*, v. 30, n. 5, pp. 571–579, 1995. ISSN: 1558173X. doi: 10.1109/4.384170.

[80] GARG, A., KIM, T. T. H. "SRAM array structures for energy efficiency enhancement", *IEEE Trans. Circuits Syst. II Express Briefs*, v. 60, n. 6, pp. 351–355, 2013. ISSN: 15497747. doi: 10.1109/TCSII.2013.2258247.

[81] QIN, H. *Deep Sub-Micron SRAM Design for Ultra-Low Leakage Standby Operation*. Tese de Doutorado, EECS Department, University of California,

Berkeley, May 2007. Disponível em: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-74.html>.

[82] WANG, B., ZHOU, J., KIM, T. T.-H. "SRAM devices and circuits optimization toward energy efficiency in multi-Vth CMOS", *Microelectronics Journal*, v. 46, n. 3, pp. 265 – 272, 2015. ISSN: 0026-2692. doi: http://dx.doi.org/10.1016/j.mejo.2014.12.003. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0026269214003528>.

[83] SINGH, R., BHAT, N. "An offset compensation technique for latch type sense amplifiers in high-speed low-power SRAMs", *IEEE Trans. Very Large Scale Integr. Syst.*, v. 12, n. 6, pp. 652–657, June 2004. ISSN: 1063-8210. doi: 10.1109/TVLSI.2004.827566.

[84] SILVEIRA, F., FLANDRE, D., JESPERS, P. G. A. "A gm/ID based methodology for the design of CMOS analog circuits and its application to the synthesis of a silicon-on-insulator micropower OTA", *IEEE Journal of Solid-State Circuits*, v. 31, n. 9, pp. 1314–1319, Sep 1996. ISSN: 0018-9200. doi: 10.1109/4.535416.

[85] AMAT, E., AMATLLÉ, E., GÓMEZ, S., et al. "Systematic and random variability analysis of two different 6T-SRAM layout topologies", *Microelectron. J.*, v. 44, n. 9, pp. 787–793, 2013. ISSN: 0026-2692. doi: 10.1016/j.mejo.2013.06.010. Disponível em: <http://dx.doi.org/10.1016/j.mejo.2013.06.010>.

[86] MANN, R. W., CALHOUN, B. H. "New category of ultra-thin notchless 6T SRAM cell layout topologies for sub-22nm". In: *2011 12th International Symp. on Quality Electronic Design*, pp. 1–6, March 2011. doi: 10.1109/ISQED.2011.5770761.

[87] SAINT, C., SAINT, J. *IC Mask Design: Essential Layout Techniques*. McGraw-Hill, 2002. ISBN: 9780071389969.

[88] MOHAMMADI, B., ANDERSSON, O., NGUYEN, J., et al. "A 128 kb single-bitline 8.4 fJ/bit 90MHz at 0.3V 7T sense-amplifierless SRAM in 28 nm FD-SOI". In: *ESSCIRC Conference 2016: 42nd European Solid-State Circuits Conference*, pp. 429–432, Sept 2016. doi: 10.1109/ESSCIRC.2016.7598333.

[89] BRUCE JACOB, SPENCER NG, D. W. *Memory Systems: Cache, DRAM, Disk*. MA, USA, Elsevier Inc., 2008. ISBN: 978-0-12-379751-3.

[90] PUSCHINI, D., RODAS, J., BEIGNE, E., et al. "Body Bias usage in UTBB FD-SOI designs: A parametric exploration approach", *Solid-State Electronics*, v. 117, pp. 138–145, 2016. ISSN: 00381101. doi: 10.1016/j.sse.2015.11.019. Disponível em: <`http://dx.doi.org/10.1016/j.sse.2015.11.019`>.

[91] MANN, R. W., WANG, J., NALAM, S., et al. "Impact of circuit assist methods on margin and performance in 6T SRAM", *Solid-State Electronics*, v. 54, n. 11, pp. 1398–1407, 2010. ISSN: 00381101. doi: 10.1016/j.sse.2010.06.009. Disponível em: <`http://dx.doi.org/10.1016/j.sse.2010.06.009`>.

[92] ZIMMER, B., TOH, S. O., VO, H., et al. "SRAM assist techniques for operation in a wide voltage range in 28-nm CMOS", *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 59, n. 12, pp. 853–857, 2012. ISSN: 15497747. doi: 10.1109/TCSII.2012.2231015.

[93] ICKES, N., FINCHELSTEIN, D., CHANDRAKASAN, A. P. "A 10-pj/instruction, 4-MIPS micropower DSP for sensor applications", *Proceedings of 2008 IEEE Asian Solid-State Circuits Conference, A-SSCC 2008*, pp. 289–292, 2008. doi: 10.1109/ASSCC.2008.4708784.

[94] KWONG, J., CHANDRAKASAN, A. P. "An energy-efficient biomedical signal processing platform", *IEEE J. Solid-State Circuits*, v. 46, n. 7, pp. 1742–1753, 2011. ISSN: 00189200. doi: 10.1109/JSSC.2011.2144450.

[95] "Low-power processor architecture exploration for online biomedical signal analysis", *IET Circuits, Devices & Systems*, v. 6, n. 5, pp. 279, 2012. ISSN: 1751858X. doi: 10.1049/iet-cds.2012.0011. Disponível em: <`http://digital-library.theiet.org/content/journals/10.1049/iet-cds.2012.0011`>.

[96] MONTAGNA, F., BENATTI, S., ROSSI, D. "Flexible, Scalable and Energy Efficient Bio-Signals Processing on the PULP Platform: A Case Study on Seizure Detection", *Journal of Low Power Electronics and Applications*, v. 7, n. 2, pp. 16, 2017. ISSN: 2079-9268. doi: 10.3390/jlpea7020016. Disponível em: <`http://www.mdpi.com/2079-9268/7/2/16`>.

Appendix A

Matlab Routines

A.1 Energy Function

```
function [StaticE,DynamicE,TotalE,IATvsIRead,Tcicle] ...
    = energy(size,VDD,Vbn,Vbp,LocalSigma,corner,pdk)
%% Cell transistor dimensions
%PUP (pull-up p)
Wp = size(1)*1e-9; Lp = size(2)*1e-9;
%PDN (pull-down n)
Wn = size(3)*1e-9; Ln = size(4)*1e-9;
%ATN (access-transistor n)
Wat = size(5)*1e-9; Lat = size(6)*1e-9;

%% Transistor parameters
temp = 25;
q = 1.602e-19;
kB = 1.3806504e-23;
UTo = kB.*(temp+273)./q;%Thermal Voltage
[B5,Vt5,Is5,lam5,n5,nu5,sigmaVt5] = ...
    parameters28nmFDSOI(Wat,Lat,'nrvt',corner,pdk);
[B1,Vt1,Is1,lam1,n1,nu1,sigmaVt1] = ...
    parameters28nmFDSOI(Wn,Ln,'nrvt',corner,pdk);
[B4,Vt4,Is4,lam4,n4,nu4,sigmaVt4] = ...
    parameters28nmFDSOI(Wp,Lp,'plvt',corner,pdk);

%% Memory array definition
Vwlread = 0.85*VDD; Vwlwrite = 0.85*VDD;
%Sub-Block size
```

```
Mb = 32; Nb = 64; bitAccess = Nb;
%Array-size
ix = 16; jx = 4;
M = Mb*ix; N = Nb*jx; MemorySize = M*N;

%% Static leakage
%Static Current Leak nominal
%I4read=B4*exp(nu4*(VDD-Vbp)/n4/UTo)*exp(lam4*VDD/n4/UTo)*...
%(1-exp(-VDD/UTo));
%I1read=B1*exp(nu1*(Vbn)/n1/UTo)*exp(lam1*VDD/n1/UTo)...
%*(1-exp(-VDD/UTo));
%I5read=B5*exp(nu5*(Vbn)/n5/UTo)*exp(lam5*VDD/n5/UTo)...
%*(1-exp(-VDD/UTo));
%IleakCell = (I4read + I1read + I5read);
%Static Current leak statistical
I4read = B4*exp(nu4*(VDD-Vbp)/n4/UTo)*exp(lam4*VDD/n4/UTo)...
    *(1-exp(-VDD/UTo))*exp(sigmaVt4^2/2/n4^2/UTo^2);
I1read = B1*exp(nu1*(Vbn)/n1/UTo)*exp(lam1*VDD/n1/UTo)...
    *(1-exp(-VDD/UTo))*exp(sigmaVt1^2/2/n1^2/UTo^2);
I5read = B5*exp(nu5*(Vbn)/n5/UTo)*exp(lam5*VDD/n5/UTo)...
    *(1-exp(-VDD/UTo))*exp(sigmaVt5^2/2/n5^2/UTo^2);
IleakCell = (I4read + I1read + I5read);

%% Bitline capacitance
%Bitline capacitance, mainly diffusion capacitance
CBLo = 410.1e-12*Wat + 21.6e-18;
%Cap of 1 bitcell, as two access transistors
CWLo = 2*(-7.69e10*(Wat*Lat)^2 + 0.007675*(Wat*Lat) + 5.577e-17);

%% Access Transistor Leakage vs Iread
%weak/moderate inversion
%IleakAT=(M-1)*Is5*log(1+exp((-Vt5+nu5*Vbn+lam5*VDD)/n5/UTo/2))...
%      .^2.*exp(sigmaVt5^2/2/n5^2/UTo^2);
%IreadMin=Is5*log(1+exp((Vwlread-Vt5-LocalSigma*sigmaVt5+lam5*VDD...
%      +nu5*Vbn)/n5/UTo/2)).^2;
%weak oinversion
IleakAT= (M-1)*B5*exp(nu5*(Vbn)/n5/UTo)*exp(lam5*VDD/n5/UTo)...
    *(1-exp(-VDD/UTo))*exp(sigmaVt5^2/2/n5^2/UTo^2);
```

```
IreadMin= Is5*exp((-Vt5-LocalSigma*sigmaVt5+nu5*Vbn)/n5/UTo)...
    *exp((lam5+1)*Vwlread/n5/UTo);
IATvsIRead = IreadMin/IleakAT;

%% Energy results
VOS = 100e-3;
%to achieve value "Vos" at 3/4 Tcicle
Tcicle = M*CBLo/IreadMin*VOS*(3/4);
StaticE = (Mb*Nb)*(ix*jx)*IleakCell*Tcicle*VDD;
DynamicE = 0.5*(Nb*CWLo*Vwlwrite^2 + MemorySize/jx*CBLo*VDD^2)+...
           0.5*(Nb*CWLo*Vwlread^2  + MemorySize/jx*CBLo*VDD^2);
TotalE = StaticE + DynamicE;
```

A.2 Static Noise Margins Function

```
function [HSNM,RSNM,WSNM] ...
    = noiseMargins(size,VDD,Vbn,Vbp,LocalSigma,corner,pdk)
%PUP (pull-up p)
Wp = size(1)*1e-9; Lp = size(2)*1e-9;
%PDN (pull-down n)
Wn = size(3)*1e-9; Ln = size(4)*1e-9;
%ATN (access-transistor n)
Wat = size(5)*1e-9; Lat = size(6)*1e-9;
%Only write setup
VWLwrite = VDD; DeltaVBL = VDD*0.85;
VBLwrite = VDD - DeltaVBL;
%Only read setup
Vwlread = VDD*0.85;
%% Transistor parameters
[B5,Vt5,Is5,lam5,n5,nu5,sigmaVt5] = ...
    parameters28nmFDSOI(Wat,Lat,'nrvt',corner,pdk);
[B6,Vt6,Is6,lam6,n6,nu6,sigmaVt6] = ...
    parameters28nmFDSOI(Wat,Lat,'nrvt',corner,pdk);
[B1,Vt1,Is1,lam1,n1,nu1,sigmaVt1] = ...
    parameters28nmFDSOI(Wn,Ln,'nrvt',corner,pdk);
[B2,Vt2,Is2,lam2,n2,nu2,sigmaVt2] = ...
```

```
    parameters28nmFDSOI(Wn,Ln,'nrvt',corner,pdk);
[B4,Vt4,Is4,lam4,n4,nu4,sigmaVt4] = ...
    parameters28nmFDSOI(Wp,Lp,'plvt',corner,pdk);
[B3,Vt3,Is3,lam3,n3,nu3,sigmaVt3] = ...
    parameters28nmFDSOI(Wp,Lp,'plvt',corner,pdk);

%% Write static noise margin
UTo = 0.02568;
a = 1/UTo; an1 = a/n1;an2 = a/n2;an4 = a/n4;
an5 = a/n5; an6 = a/n6; ao = an1*lam1+an5*(1+nu5+lam5);
a1 = an4*lam4+an6*lam6; C = exp(-a*VDD);
A1W = Is1*exp(-an1*Vt1)*exp(an1*nu1*Vbn);
A4W = Is4*exp(an4*(1+lam4)*VDD-an4*Vt4)*exp(an4*nu4*(VDD-Vbp));
A5W = Is5*exp(an5*(VWLwrite+lam5*VDD)-an5*Vt5)*exp(an5*nu5*Vbn);
A6W = Is6*exp(an6*(VWLwrite-VBLwrite) -an6*lam6*VBLwrite ...
      -an6*nu6*VBLwrite -an6*Vt6)*exp(an6*nu6*Vbn);
%conventional definition
WSNM=-(-VDD*a*a1*ao+VDD*a*an1*an4+log(A1W/A5W)*a*an4+a*ao*...
log(A4W*a/(A6W*(a+an4)))+log(an4/(a+an4))*an1*an4)/(a*an4*(ao+an1));
DWSNM2=(n5/(lam1*n5+lam5*n1+n1*nu5+n1+n5))^2....
*(sigmaVt1^2+sigmaVt5^2)+(n4*(lam1*n5+lam5*n1+n1*nu5+n1)...
/(n1*(lam1*n5+lam5*n1+n1*nu5+n1+n5)))^2*(sigmaVt4^2+sigmaVt6^2);
WSNM = WSNM - LocalSigma*sqrt(DWSNM2);%mismatching influence
%% Read static noise margin
no = n1*n5/(n1*(lam5+nu5+1)+n5*lam1);
Vt5x = (Vt5 - nu5*Vbn);
Vt1x = (Vt1 - nu1*Vbn);
Vt2x = (Vt2 - nu2*Vbn);
Vt4x = -(Vt4 - nu4*(VDD-Vbp));
RSNM = ( n5*(n1*n2-n1*n4*lam2+no*(n2+n4))*VDD ...
       - no*n1*(n2+n4)*Vwlread ...
       + no*(n2+n4)*(n1*Vt5x-n5*Vt1x) + n5*n1*(n4*Vt2x+n2*Vt4x) ...
       + n5*(n1*n4*(lam2-n2) - no*(n2+n4))*log(2)*UTo ...
       + no*n5*n1*(n2+n4)*log(Is1/Is5)*UTo ...
       + n5*n1*n2*n4*log(Is4/Is2)*UTo )/n5/(no+n1)/(n2+n4);
n = 2*n1*n2/(n1+n2);
DRSNM2 = (1/2)^2*sigmaVt1^2 + (1/4)^2*sigmaVt4^2 ...
         +(1/2)^2*sigmaVt5^2 + (1/4)^2*sigmaVt2^2;
```

```
RSNM = RSNM - LocalSigma*sqrt(DRSNM2);%mismatching influence

%% Hold static noise margin
n = 2*n3*n1/(n3+n1);
lam = (lam1*n3+lam3*n1)/(n1+n3);
Vino = n*VDD/2*((2+2*nu3+lam3)/2/n3-lam1/2/n1) ...
      + n*UTo/2*log(Is3/Is1)+n/2/n1*Vt1 - n/2/n3*Vt3 ...
      - n/n3*nu3/2*Vbp - n/n1*nu1/2*Vbn;
%Derivation of the SNM
NML = n*UTo/2*log(n/(n+2)) + (1-lam)*log(2/(n+2))*UTo +...
    Vino - lam*VDD/2;
NMH = n*UTo/2*log(n/(n+2)) + (1-lam)*log(2/(n+2))*UTo +...
    VDD - Vino - lam*VDD/2;
DHSNM = sqrt((1/2)^2*sigmaVt3^2+(1/2)^2*sigmaVt1^2);
HSNM = min(NML,NMH)-LocalSigma*DHSNM;
```