A Combined DOE-ILP Based Power and Read Stability Optimization in Nano-CMOS SRAM

Garima Thakral[1], Saraju P. Mohanty[2], Dhruva Ghai[3], and Dhiraj K. Pradhan[4]
Department of Computer Science and Engineering, University of North Texas, USA.[1,2,3]
Department of Computer Science, University of Bristol, UK.[4]
Email-ID: saraju.mohanty@unt.edu[2] and pradhan@compsci.bristol.ac.uk[4].

Abstract

A novel design approach for simultaneous power and stability (static noise margin, SNM) optimization of nano-CMOS static random access memory (SRAM) is presented. A $45nm$ single-ended seven transistor SRAM is used as a case study. The SRAM is subjected to a dual-V_{Th} assignment using a novel combined Design of Experiments and Integer Linear Programming (DOE-ILP) algorithm, resulting in 50.6% power reduction (including leakage) and 43.9% increase in the read SNM. The process variation analysis of the optimal SRAM carried out considering twelve device parameters shows the robustness of the design.

1 Introduction

The memory subsystem consumes a substantial portion of the total-power budget of a system-on-a-chip (SoC) [17]. Reducing memory power dissipation will improve power-efficiency and reliability of the SoC. Stability of SRAM becomes the major concern when the nano-CMOS is used for its fabrication due to process variations. Variations in the device parameters translate into variations in SRAM attributes, such as power and stability. Under adverse operating conditions such SRAMs may inadvertently corrupt the stored data. It is challenging to maintain an acceptable SNM in embedded SRAMs while scaling the minimum feature sizes and supply voltages of the SoC [13, 9].

The current literature is rich in variants of SRAM. A nine-transistor SRAM is proposed in [9, 8] that has higher stability and low power consumption. A Schmitt-trigger based SRAM in [6] provides better read stability and better write ability. A ten transistor SRAM at a low voltage and faster readout operation is proposed in [12]. A sub V_{Th} SRAM is presented in [16]. A methodology is proposed in [1] to analyze the stability of an SRAM considering device parameter fluctuations. In [2, 7, 3], a dual-V_{Th} and dual-T_{ox} assignment method is presented for low-power SRAM.

The rest of the paper is organized as follows: Section 2 discusses the new design flow. The baseline $45nm$ SRAM design is discussed in Section 3. Section 4 presents the combined DOE-ILP based algorithm. Section 5 studies the effect of variability in device parameters on the optimal SRAM, followed by conclusions in Section 6.

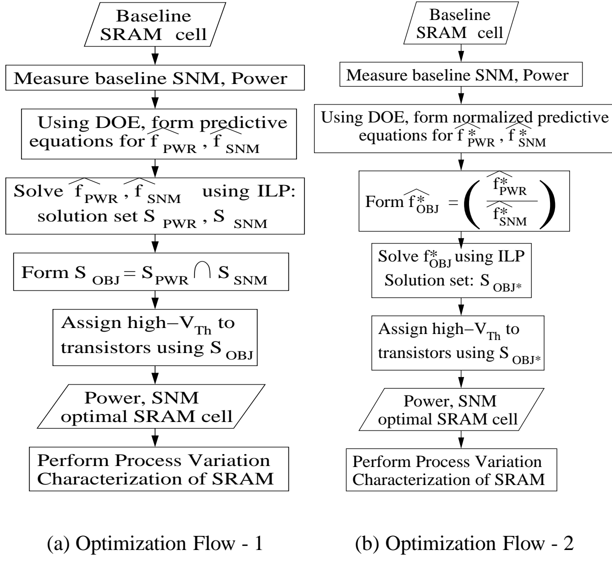
2 Proposed Optimal SRAM Design Flows

The two flows investigated are shown in Fig. 1. The input to each of the flows is a baseline SRAM which refers to the design with nominal sized transistors for a specific technology. In the embedded SRAM, it is increasingly challenging to maintain the read SNM while reducing power consumption. To reduce the power consumption this paper investigates the process-level technique, called dual-V_{Th}. For the nano-CMOS node (e.g. $45nm$) under consideration, leakage is a major component of total power [10]. Hence its reduction through dual-V_{Th} reduces total power.

What is *important at this step is the selection of appropriate transistors for high-V_{Th} assignment* so that performance of SRAM is not degraded. To address this important problem of choosing transistors for high-V_{Th} assignment we propose combined DOE-ILP algorithms. The combined approach reduces the optimization search space and convergence solutions faster (due to DOE) while maintaining the accuracy of ILP. The approach thus can handle large circuits for optimization in reasonable time for optimal solutions.

In optimal-design flow 1 (Fig. 1(a)), predictive equations are formulated for power ($\widehat{f_{PWR}}$), and SNM ($\widehat{f_{SNM}}$) based on the experiments performed on the V_{Th} state (high or nominal) of each transistor. These predictive equations ($\widehat{f_{PWR}}$ and $\widehat{f_{SNM}}$), and the constraints are assumed to be linear. Each of the solution variables is restricted to be either 0 (nominal V_{Th}) or 1 (high V_{Th}). In essence the linear

[0]This research is supported in part by NSF award numbers CCF-0702361 and CNS-0854182.



(a) Optimization Flow - 1 (b) Optimization Flow - 2

Figure 1. Proposed Design flows for SRAM.

objective function is optimized subjected to linear equality and linear inequality constraints. Thus, ILP is the fittest way to solve the optimization problem. The solution set for power minimization is called S_{PWR}, and the solution set for SNM maximization is called S_{SNM}. In order to obtain power minimization and SNM maximization simultaneously, the overall objective set S_{OBJ} is formulated as $S_{PWR} \cap S_{SNM}$ ($\cap$ refers intersection).

In the optimal-design flow 2 (Fig. 1(b)), the predictive equations for power ($\widehat{f_{PWR}*}$), and SNM ($\widehat{f_{SNM}*}$) are normalized based on the experiments performed on the V_{Th} state (high or nominal) of each transistor. Normalization of the two different dimensional attributes, such power and SNM allows formulation of a combined objective function for their simultaneous optimization. The objective function to be solved: $\widehat{f_{OBJ}*}$ is formed as the division of $\widehat{f_{PWR}*}$ and $\widehat{f_{SNM}*}$. The minimization of $\widehat{f_{OBJ}*}$ leads to simultaneous power minimization (numerator) and SNM maximization (denominator). The solution set is called $S_{OBJ}*$.

After the optimal dual-V_{Th} assignment is obtained from either of the two ways (i.e. Fig. 1(a) or Fig. 1(b)), the new SRAM configuration is re-simulated for power and SNM. For nano-CMOS SRAM it is important that they perform under several process variations, thus the statistical variability is studied subjected to twelve device parameters.

A 7-transistor SRAM topology which is suitable for the low voltage regime and tolerant to read failure is used [14] as a case study for the proposed methodology. However, the proposed methodology is applicable to any SRAM.

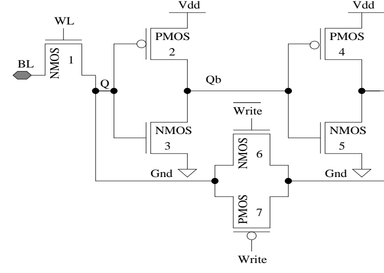


Figure 2. A single-ended 7-transistor SRAM cell [14]; load transistors - 2, 4; driver transistors - 3, 5; access transistors - 1, 6, 7.

3 Single-Ended 7 Transistor SRAM Design

3.1 Design of the SRAM for $45nm$ CMOS

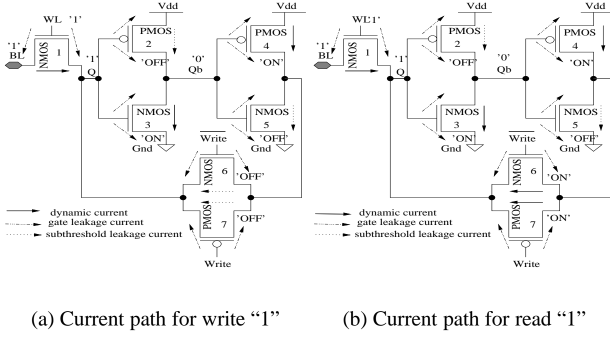
The single-ended 7-transistor SRAM cell is shown in Fig. 2 is composed of a read and write access transistor (transistor 1), two inverters (transistors 2, 3, 4 and 5) connected back to back to store the 1-bit information and a transmission gate (transistors 6 and 7). The transmission gate opens the feedback connection between inverters during the write operation. The cell operates on a single bit-line, instead of two bit-lines as in standard six transistor SRAM cell. Both reading and writing operations are performed over the single bit-line. The word-line (WL) is asserted high prior to write/read operation. When the cell is in hold mode, WL is low and a strong feedback is provided to the cross-coupled inverters by the transmission gate.

3.2 Static Noise Margin Measurement

SNM is defined as the maximum amount of noise that can be tolerated at the cell nodes just before flipping the states. SNM is expressed by the following [13]:

$$SNM_{sram} = V_{Th} - \left(\frac{1}{k+1}\right) \times \left[\left(\frac{V_{dd} - \left(\frac{2r+1}{r+1}\right) V_{Th}}{1 + \left(\frac{r}{k(r+1)}\right)} \right) - \left(\frac{V_{dd} - 2V_{Th}}{1 + \left(\frac{r}{q}k\right) + \sqrt{\left(\frac{r}{q}\right)\left(1 + 2k + \frac{r}{q}k^2\right)}} \right) \right] \quad (1)$$

Where r is the ratio of driver to access transistor sizes, q is the ratio of load to access transistor sizes, k is defined as $\left[\left(\frac{r}{r+1}\right)\left(\sqrt{\frac{r+1}{r+1-V_s^2/V_r^2}} - 1\right)\right]$, V_s is $(V_{dd} - V_{Th})$ and V_r is $\left[V_s - \left(\frac{r}{r+1} V_{Th}\right)\right]$. Thus, *SNM is dependent on the threshold voltage V_{Th}*. For measurement, SNM is defined



(a) Current path for write "1" (b) Current path for read "1"

Figure 3. Current paths in the SRAM cell.

Algorithm 1 : Simultaneous power/SNM optimization

1: **Input:** Baseline circuit, Nominal/High-V_{Th} models.
2: **Output:** Objective set S_{OBJ} = $[f_{PWR}, f_{SNM}]$ with transistors identified for high V_{Th} assignment.
3: Setup experiment for transistors of SRAM cell using 2-Level Taguchi L-8 array, where the factors are the transistors and the responses are average P_{sram} and read SNM_{sram}.
4: **for** Each 1:8 experiments of 2-Level Taguchi L-8 array **do**
5: Perform simulations and record P_{sram} and SNM_{sram}.
6: **end for**
7: Form predictive equations: $\widehat{f_{PWR}}$ for power, $\widehat{f_{SNM}}$ for SNM.
8: Solve $\widehat{f_{PWR}}$ using ILP. Solution set: S_{PWR}.
9: Solve $\widehat{f_{SNM}}$ using ILP. Solution set: S_{SNM}.
10: Form $S_{OBJ} = S_{PWR} \cap S_{SNM}$.
11: Assign high V_{Th} to transistors based on S_{OBJ}.

as the length of the side of the largest square that can be fitted inside the lobes of the butterfly curve [7, 13].

3.3 Power and Leakage Measurement

The total power of a nano-CMOS SRAM circuit is:

$$P_{sram} = P_{dyn} + P_{sub} + P_{gate}, \quad (2)$$

where P_{dyn} is dynamic power, P_{sub} is subthreshold leakage, and P_{gate} is gate leakage. SRAM cells retain data for some duration of time as they cannot be shut off and also leakage is a prominent component of power [16, 10]. So, minimizing of leakage is necessary. One of the major components of power, the subthreshold leakage is:

$$I_{sub} = Cexp\left(\frac{V_{gs} - V_{Th}}{Sv_{therm}}\right)\left(1 - exp\left(\frac{-V_{ds}}{v_{therm}}\right)\right), \quad (3)$$

where $C = \left(\mu_0\left(\frac{\epsilon_{ox}W}{T_{ox}L_{eff}}\right)v_{therm}^2 e^{1.8}\right)$. Thus, subthreshold leakage is *exponentially dependent on* V_{Th} [10].

In a nano-CMOS SRAM circuit, the current flow in each device depends of the location the device in the circuit as well the operation being performed. Thus, for accurate measurement of power it is important that the currents are identified. Fig. 3 shows the current paths for various read and write operations for a SRAM cell. When the transistor is in ON state it has active current along with the gate leakage [5]. When the transistor is in OFF state, it has gate-oxide leakage current and subthreshold leakage current [5].

For clarity, let us discuss Fig. 3(a) which shows the current path for write "1" operation. In this case, bit line and WL are precharged to level "1" which form a path for Q, thus Q will be at level "1". Hence transistor 1 is ON and carries both active current and gate leakage current. PMOS transistor of the first inverter (transistor 2) will be OFF and NMOS (transistor 3) are ON. The active current flows in NMOS along with the gate-leakage current, whereas PMOS carries gate leakage and subthreshold leakage. In the case of second inverter, NMOS (transistor 5) is in OFF state which has gate leakage and subthreshold leakage, whereas PMOS (transistor 4) which is in ON state carries active current and gate-leakage current. The transistors of the transmission gate (transistor 6 and 7) are OFF while the SRAM cell performs write function hence subthreshold and gate-leakage current flow through them.

In summary, both power and SNM are affected by the threshold voltage V_{Th} and dual-V_{Th} technique is promising for their optimization in nano-CMOS SRAM design.

The SRAM cell has been simulated using $45nm$ CMOS PTM model [18], with minimum sized transistors and a V_{dd} of $0.7V$. The power and SNM results for the baseline design is presented in Table 1 and butterfly curve in Fig. 5(a).

4 Combined DOE-ILP Based Algorithms

Algorithm 1 and 2 presents two combined DOE-ILP optimization algorithms. The two versions differ in the way the power and SNM objectives are simultaneously tackled.

In Algorithm 1, experimental analysis is performed for the transistors of the SRAM using 2-Level Taguchi L-8 array [11]. Simulations are run for each 1:8 experiments of 2-Level Taguchi L-8 array and the values for both power and SNM are recorded. Using DOE, the linear predictive equations are formulated. 2-Level Taguchi L-8 array approach of DOE is a better choice compared to the other techniques as it is fast and efficient. For example, a *full factorial experiment* will take $2^7 = 128$ runs, whereas a 2-Level Taguchi L-8 array resulted in 8 only runs.

In Algorithm 2, normalized equations for power ($\widehat{f_{PWR}*}$) and SNM ($\widehat{f_{SNM}*}$) are obtained. The objective function ($\widehat{f_{OBJ}*}$) is formed as the division of ($\widehat{f_{PWR}*}$) and ($\widehat{f_{SNM}*}$), as minimization of this would lead to simultane-

Algorithm 2 : Simultaneous power/SNM optimization

1: **Input:** Baseline circuit, Nominal/High-V_{Th} models.
2: **Output:** Objective set $S_{OBJ}* = [f_{PWR}*, f_{SNM}*]$ with transistors identified for high V_{Th} assignment.
3: Setup experiment for transistors of SRAM cell using 2-Level Taguchi L-8 array, where the factors are the transistors and the responses are average P_{sram} and read SNM_{sram}.
4: **for** Each 1:8 experiments of 2-Level Taguchi L-8 array **do**
5: Perform simulations and record P_{sram} and SNM_{sram}.
6: **end for**
7: Form normalized predictive equations: $\widehat{f_{PWR}*}$ and $\widehat{f_{SNM}*}$.
8: Form $f_{OBJ}* = \left(\frac{\widehat{f_{PWR}*}}{\widehat{f_{SNM}*}}\right)$.
9: Solve $\widehat{f_{OBJ}*}$ using ILP. Solution set: $S_{OBJ}*$.
10: Assign high V_{Th} to transistors based on $S_{OBJ}*$.

ous power minimization and SNM maximization. This objective function is then solved to get the solution set S_{OBJ*}.

The factors of DOE are the 7 transistors of the SRAM cell, and the responses are the average-power consumption ($\widehat{f_{PWR}}$) and SNM ($\widehat{f_{SNM}}$) of the cell. Each factor can take a high V_{Th} state (1) or a nominal V_{Th} state (0). The experiments are run, and the half-effects are recorded. The predictive equations of are obtained from the pareto plots of half-effects of transistors.

4.1 Power Minimization: S_{PWR}

The predictive equation for average power is:

$$\begin{aligned} \widehat{f_{PWR}}(nW) &= 118.2075 - 5.975 \times x_1 - 28.955 \times x_2 \\ &- 23.1625 \times x_3 - 10.995 \times x_4 - 10.6375 \times x_5 \\ &- 12.1425 \times x_6 + 6.475 \times x_7. \end{aligned} \quad (4)$$

Where, x_1 represents the V_{Th}-state of transistor 1, x_2 represents the V_{Th}-state of transistor 2, and so on. The ILP for average power minimization is formulated as:

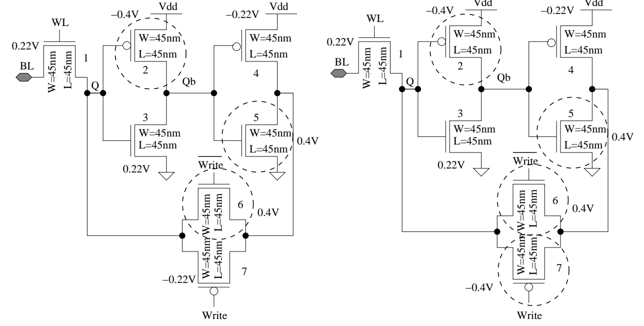
$$\begin{array}{ll} \min & \widehat{f_{PWR}} \\ \text{s.t.} & f_{SNM} > \tau_{SNM} \text{ and } x_i \ \forall\, i \ 1 \rightarrow 7 \text{ either 0 or 1}, \end{array} \quad (5)$$

where τ_{SNM} is a designer defined constraint on SNM. Solving the ILP problem, the optimal solution is obtained as follows: $S_{PWR} = [x_1 = 1, x_2 = 1, x_3 = 1, x_4 = 1, x_5 = 1, x_6 = 1, x_7 = 0]$. This is interpreted as transistors 1, 2, 3, 4, 5, 6 are of high V_{Th}, and transistor 7 is of nominal.

4.2 SNM Maximization: S_{SNM}

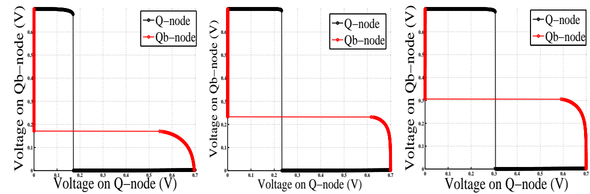
The predictive equation for read SNM is expressed as:

$$\begin{aligned} \widehat{f_{SNM}}(mV) &= 156.675 - 44.025 \times x_1 + 58.725 \times x_2 \\ &- 53.925 \times x_3 - 6.425 \times x_4 + 32.575 \times x_5 \\ &+ 19.375 \times x_6 - 19.625 \times x_7. \end{aligned} \quad (6)$$



(a) Configuration for Flow-1 (b) Configuration for Flow-2

Figure 4. Dual-V_{Th} configurations of SRAM.



(a) For Baseline (b) For S_{PWR} (c) For S_{OBJ}

Figure 5. Butterfly curve of three alternatives.

The ILP formulation for SNM maximization is obtained as:

$$\begin{array}{ll} \max & \widehat{f_{SNM}} \\ \text{s.t.} & f_{PWR} < \tau_{PWR} \text{ and } x_i \ \forall\, i \ 1 \rightarrow 7 \text{ either 0 or 1}, \end{array} \quad (7)$$

where τ_{PWR} is the designer defined power constraint. ILP yields the optimal solution as: $S_{SNM} = [x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 0, x_5 = 1, x_6 = 1, x_7 = 0]$.

4.3 Combined Power / SNM Optimization

4.3.1 Approach-1

The objective set S_{OBJ} for simultaneous optimization of power and SNM is formed as:

$$S_{OBJ} = S_{PWR} \cap S_{SNM}, \quad (8)$$

where $\cap$ is the intersection of two solution sets S_{PWR} and S_{SNM}. To obtain low-power and high-SNM SRAM, we use the set intersection operator to achieve S_{OBJ} which has the set values of $S_{PWR} \cap S_{SNM}$. In other words, we pick

devices which are part of low-power and high-SNM solution sets. The constraints are same as the above ILP formulations. The ILP results in the following: $S_{OBJ} = [x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 0, x_5 = 1, x_6 = 1, x_7 = 0]$, which leads to a configuration of Fig. 4(a) and results in Table 1.

4.3.2 Approach-2

In this, normalized forms of $\widehat{f_{PWR}}$ and $\widehat{f_{SNM}}$ denoted as $\widehat{f_{PWR}*}$ and $\widehat{f_{SNM}*}$ are used. The normalized is performed by division of each data by the maximum value of data. Normalization enables directly accommodating different units, while forming the objective function as:

$$\begin{aligned}\widehat{f_{PWR}*} &= 0.58 - 0.03 \times x_1 - 0.14 \times x_2 \\ &- 0.11 \times x_3 - 0.05 \times x_4 - 0.05 \times x_5 \\ &- 0.06 \times x_6 + 0.03 \times x_7. \end{aligned} \tag{9}$$

$$\begin{aligned}\widehat{f_{SNM}*} &= 0.52 - 0.15 \times x_1 + 0.19 \times x_2 \\ &- 0.18 \times x_3 - 0.02 \times x_4 + 0.11 \times x_5 \\ &+ 0.06 \times x_6 - 0.06 \times x_7. \end{aligned} \tag{10}$$

The combined objective function is formed as follows:

$$\begin{aligned}\widehat{f_{OBJ}*} &= \left(\frac{\widehat{f_{PWR}*}}{\widehat{f_{SNM}*}}\right), \\ &= 0.18 \times x_3 - 0.02 \times x_4 + 0.11 \times x_5 \\ &+ 0.06 \times x_6 - 0.06 \times x_7, \end{aligned} \tag{11}$$

Eqn.11 is obtained from the division of normalized values of eqn.9 and normalized values of eqn.10. Through normalization, we eliminate the condition of different units of power and SNM and hence we get quotient as 11. The ILP formulation for this combined method is obtained as:

$$\begin{aligned}&\min \ \widehat{f_{OBJ}*} \\ &\text{s.t.} \ \ f_{PWR} < \tau_{PWR}, f_{SNM} > \tau_{SNM}, x_i \ \forall\, i \in \ 0 \text{ or } 1.\end{aligned} \tag{12}$$

For this, optimal solution is obtained as: $S_{OBJ}* = [x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 0, x_5 = 1, x_6 = 1, x_7 = 1]$, whose SRAM configuration is shown in Fig. 4(b) and results in Table 1.

To study the power and SNM of the optimal SRAM, simulations are performed for various V_{dd} as shown in Fig. 6. It is observed that both power and SNM increases with increase in V_{dd}. For the $V_{dd} = 0.7V$, the power has reduced by 44.2% and SNM has increased by 43.9% compared to the baseline design using approach 1, and the power has reduced by 50.6% and SNM is increased by 43.9% compared to the baseline design using approach 2.

From the experimental data, it is observed that approach 2 is more effective in achieving reduced power and high SNM, compared to the baseline design.

Table 1. Results for different objectives

Design Alternative	Parameter	Value	Change
Baseline	P_{sram}	$203.6\, nW$	–
	SNM_{sram}	$170\, mV$	–
S_{PWR}	P_{sram}	$26.34\, nW$	87.1% decrease
	SNM_{sram}	$231.9\, mV$	26.7% increase
S_{SNM}	P_{sram}	$113.6\, nW$	44.2% decrease
	SNM_{sram}	$303.3\, mV$	43.9% increase
S_{OBJ} Approach 1	P_{sram}	$113.6\, nW$	44.2% decrease
	SNM_{sram}	$303.3\, mV$	43.9% increase
$S_{OBJ}*$ Approach 2	P_{sram}	$100.5\, nW$	50.6% decrease
	SNM_{sram}	$303.3\, mV$	43.9% increase

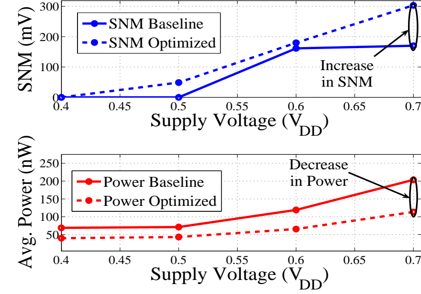


Figure 6. Power / SNM comparison of SRAM.

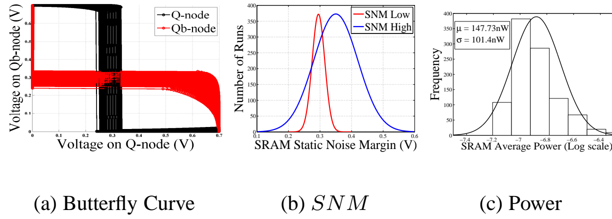
5 Variability Analysis of the SRAM

Threshold voltage variation (standard deviation) is [15]:

$$\sigma_{V_{Th}} = \left(\frac{\sqrt[4]{4 \times q^3 \times \epsilon_{Si} \times \phi_B}}{2}\right) \times \left(\frac{T_{ox}}{\epsilon_{ox}}\right) \times \left(\frac{\sqrt[4]{N_{ch}}}{\sqrt{W \times L}}\right) \tag{13}$$

where T_{ox} - oxide thickness, N_{ch} - channel dopant concentration, L - length, W - width, $\phi_B = [2 \times \kappa_B \times T \times \ln(N_{ch}/n_i)]$ (with κ_B Boltzmann's constant, T temperature, n_i intrinsic carrier concentration, q elementary charge), and ϵ_{ox} and ϵ_{Si} are permittivity of oxide and silicon. Since V_{Th} affects power and SNM (Eqn. (3) and Eqn. (1)), these parameters affect them also. Thus, *twelve process parameters* are selected for statistical variability study: NMOS/PMOS channel length (T_{oxn}, T_{oxp}), NMOS/PMOS channel doping concentration (N_{chn}, N_{chp}), access-transistor length and width (L_{na}, L_{pa}, W_{na}, W_{pa}), driver-transistor length and width (L_{nd}, W_{nd}), load-transistor length and width (L_{pl}, W_{pl}). Some of the parameters are independent and some are correlated which is taken into account during simulation for realistic study.

The SNM is exhaustively evaluated through Monte Carlo simulations to ensure there is no process variation induced failure in the SRAM. Monte Carlo simulation is an efficient approach because it does not require relating the output to input which otherwise would have been cumbersome for the large number of parameters [4]. A correlation coefficient of 0.9 between T_{oxn} and T_{oxp} is assumed.



(a) Butterfly Curve (b) SNM (c) Power

Figure 7. Process variation study of SRAM from flow-1 using 1000 Monte Carlo runs.

Table 2. Statistical Process Variation Effects.

Optimization	Parameter	μ	σ
S_{PWR}	P_{sram}	$28.91nW$	$8.26nW$
	SNM_{sram}	$180mV$	$30mV$
S_{SNM}	P_{sram}	$147.73nW$	$101.4nW$
	SNM_{sram}	$295mV$	$28mV$
S_{OBJ}: Approach 1	P_{sram}	$147.73nW$	$101.4nW$
	SNM_{sram}	$295mV$	$28mV$
$S_{OBJ}*$: Approach 2	P_{sram}	$135.24nW$	$101.85nW$
	SNM_{sram}	$295mV$	$28mV$

Each of the process parameters is assumed to have a Gaussian distribution with mean (μ) taken as the nominal values specified in the PTM [18] and standard deviation (σ) as 5% of the mean. Fig. 7 and Table 2 present the results.

6 Conclusions

A methodology for simultaneous optimization of SRAM power and read stability is presented. A $45nm$ single-ended seven transistor SRAM was subjected to the proposed methodology leading to 50.6% power reduction and 43.9% increase in read stability (read SNM). A novel DOE-ILP algorithm is used for power minimization and read SNM maximization. The effect of process variation of twelve process parameters on the SRAM is evaluated, and it is found to be process variation tolerant. A 8×8 array has been constructed using the optimized cells whose average power consumption is $4.5\mu W$. For a broad comparative perspective, in [2, 3] only leakage is considered and dynamic power is not accounted in optimization, contrary to the current paper that accounts all components (dynamic, subthreshold, gate leakages). In [2, 3], a combined dual-V_{Th} and dual-T_{ox} assignment is used where the leakage power reduction is 53.5% and SNM increase is 43.8%. However, our methodology which considers only dual-V_{Th} has resulted in power reduction (accounting all components) of 50.6% and increase in read SNM of 43.9%. Assuming that the dual-V_{Th} and dual-T_{ox} will need more number of masks compared to only dual-V_{Th} for fabrication of the SRAM chip, our SRAM will need half of that of [2, 3] for comparable performance. Future research will involve SRAM-array optimization where variability will be accounted in flow.

References

[1] K. Agarwal and S. Nassif. Statistical Analysis of SRAM Cell Stability. In *Proc. of DAC*, pp. 57–62, 2006.

[2] B. Amelifard, F. Fallah, and M. Pedram. Reducing the Sub-threshold and Gate-tunneling Leakage of SRAM Cells using Dual-Vt and Dual-Tox Assignment. In *Proc. Design Automation and Test in Europe*, pp. 1–6, 2006.

[3] B. Amelifard, F. Fallah, and M. Pedram. Leakage Minimization of SRAM Cells in a Dual-V_t and Dual-T_{ox} Technology. *IEEE Trans. VLSI Systems*, 16(7):851–860, 2008.

[4] R. Keerthi and C.-H. Chen. Stability and Static Noise Margin Analysis of Low-Power SRAM. In *Proc. Instrumentation and Measurement Technology Conf.*, pp. 1681–1684, 2008.

[5] E. Kougianos and S. P. Mohanty. Metrics to Quantify Steady and Transient Gate Leakage in Nanoscale Transistors: NMOS Vs PMOS Perspective. In *Proc. 20th IEEE International Conference on VLSI Design*, pp. 195–200, 2007.

[6] J. P. Kulkani, K. Kim, S. P. Park, and K. Roy. Process Variation Tolerant SRAM Array for Ultra Low Voltage Applications. In *Proc. Design Automation Conf.*, pp. 108–113, 2008.

[7] J. Lee and A. Davoodi. Comparison of Dual-V_t Configurations of SRAM Cell Considering Process-Induced V_t Variations. In *Proc. of ISCAS*, pp. 3018–3021, 2007.

[8] S. Lin, Y. B. Kim, and F. Lombardi. A low leakage 9t sram cell for ultra-low power operation. In *Proc. ACM Great Lakes symposium on VLSI*, pp. 123–126, 2008.

[9] Z. Liu and V. Kursun. Characterization of a novel nine-transistor sram cell. *IEEE Trans. VLSI Systems*, 16(4):488492, April 2008.

[10] S. P. Mohanty and E. Kougianos. Simultaneous Power Fluctuation and Average Power Minimization during Nano-CMOS Behavioural Synthesis. In *Proc. 20th IEEE International Conference on VLSI Design*, pp. 577–582, 2007.

[11] D. C. Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, Inc., 6th edition, 2005.

[12] S. Okumura and et al. A 0.56-V 128kb 10T SRAM Using Column Line Assist (CLA) Scheme. In *Proc. International Sympo. Quality Electronic Design*, pp. 659–663, 2009.

[13] E. Seevinck, et al. Static noise margin analysis of MOS SRAM cells. *IEEE J. Solid-State Circuits*, 22(5), 1987.

[14] J. Singh, J. Mathew, D. K. Pradhan, and S. P. Mohanty. A Subthreshold Single Ended I/O SRAM Cell Design for Nanometer CMOS Technologies. In *Proc. International SOC Conference*, pp. 243–246, 2008.

[15] P. A. Stolk, F. P. Widdershoven, and D. B. M. Klaassen. Modeling Statistical Dopant Fluctuations in MOS Transistors. *IEEE Trans. Electron Devices*, 45(9):1960–1971, 1998.

[16] N. Verma and A. P. Chandrakasan. A $256kb$ $65nm$ 8T Sub-threshold SRAM Employing Sense-Amplifier Redundancy. *IEEE J. Solid-State Circuits*, 43(1):141–149, Jan 2008.

[17] N. Yoshinobu and et al. Review and future prospects of low voltage RAM circuits. *IBM J. Research and Development*, 47(5/6):525–552, 2003.

[18] W. Zhao and Y. Cao. New Generation of Predictive Technology Model for sub-$45nm$ Design Exploration. In *Proc. of ISQED*, pp. 585–590, 2006.