Statistical DOE-ILP Based Power-Performance-Process (P3) Optimization of Nano-CMOS SRAM

Saraju P. Mohanty[1], Jawar Singh[2], Elias Kougianos[3], and Dhiraj K. Pradhan[4]

NanoSystem Design Laboratory (NSDL), University of North Texas, USA.[1,3]
Department of Electronics and Communication Engineering,
Jaypee University of Engineering and Technology, India.[2]
Department of Computer Science, University of Bristol, UK.[4]

Abstract

As technology continues to scale, maintaining important figures of merit of Static Random Access Memories (SRAMs), such as power dissipation and an acceptable Static Noise Margin (SNM), becomes increasingly challenging. In this paper, we address SRAM instability and power (leakage) dissipation in scaled-down technologies by presenting a novel design flow for simultaneous Power minimization, Performance maximization and Process variation tolerance (P3) optimization of nano-CMOS circuits. 45 nm and 32 nm technology node standard 6-Transistor (6T) and 8T SRAM cells are used as example circuits for demonstration of the effectiveness of the flow. Thereafter, the SRAM cell is subjected to a dual threshold voltage (dual-V_{Th}) assignment based on a novel statistical Design of Experiments-Integer Linear Programming (DOE-ILP) approach. Experimental results show 61% leakage power reduction and 13% increase in the read SNM. In addition, process variation analysis of the optimized cell is conducted considering the variability effect in twelve device parameters. To the best of the authors' knowledge, this is the first study which makes use of statistical DOE-ILP for optimization of conflicting targets of stability and power in the presence of process variations in SRAMs.

Key words: Nanoscale CMOS, Process-variation aware design, Low-Power design, Static random access memory, Design of Experiments, Integer Linear Programming.

Email address: `saraju.mohanty@unt.edu`[1], `jawar.singh@jiet.ac.in`[2], `elias.kougianos@unt.edu`[3], `pradhan@compsci.bristol.ac.uk`[4] (Dhiraj K. Pradhan[4]).

1 Introduction and contributions

SRAM is a volatile memory that retains data bits as long as power is being supplied. It provides fast access to data and is very reliable. Degraded bitcell currents and leakages, and poor SRAM bitcell noise margins, when a large number of devices are integrated into a single die, result in process and design variability which in turn leads to a great loss of parametric yield [1]. A sufficiently large Static Noise Margin (SNM), reduced power consumption and a process variation tolerant circuit are needed in order to prevent substantial loss of parametric yield caused by the technology scaling induced side effects. Thus, the operations of SRAM have become very critical with the advancement of CMOS technology. In this section, we discuss the importance of the factors that have been considered for optimization, and present the motivation behind the research presented in this paper. By reducing the power consumption significantly, and maximizing the static noise margin we can increase the efficiency and reliability of the SRAM cell. However, the SRAM cell becomes susceptible to process variation at lower supply voltages which in turns decreases its noise handling capacity.

SRAM arrays are widely used as cache memory in microprocessors and Application-Specific Integrated Circuits (ASICs) and occupy a large portion of the die area. Large arrays of fast SRAM help to improve the performance of the system. Thus, balancing these requirements is driving the effort to minimize the footprint of SRAM cells [1].

Power dissipation: Embedded systems, particularly those targeted towards low duty cycles and portable applications (e.g. mobile phones), require extremely low energy dissipation as they are typically battery powered. In such systems, a significant amount of power is consumed during memory accesses, which affects the battery life. Hence, efficient active and leakage power saving SRAM designs need to be explored for higher reliability and longer operation of battery powered systems. Different design methods have been proposed, such as decrease in supply voltage, which reduces the dynamic power quadratically and reduces the leakage power linearly [2]. However, with technology scaling, leakage current increases exponentially and reliability is affected significantly due to poor stability noise margins and process variation. These technology scaling-induced side effects are further exacerbated by reduced supply voltage introduced in order to achieve energy efficiency. Figure 1 shows the comparison of normalized read Static Noise Margin (SNM) and leakage current of a 6T SRAM cell for different technology nodes. The minimum feature sized devices with cell ratio (β =2) is used for simulation using the Predictive Technology Model (PTM) [3]. It can be seen from Figure 1 that the read SNM of a 6T SRAM cell is gradually decreasing with technology scaling, while the leakage current is exponentially increasing. Moving from the 132 nm to the 32 nm technology node, there is 55% reduction in the read SNM while there is 86% increase in leakage current. Therefore, alternative cell topologies or optimization methodologies are needed for nano-regime technologies that provide low standby power (leakage) and higher stability margins (SNM). Along this line, several SRAM cell topologies have been proposed in the recent past to address the ultra-low power requirements [4–8]. Hence, in this paper, standard 6T and 8T SRAM [6,7] cells are used as baseline circuits for optimization.

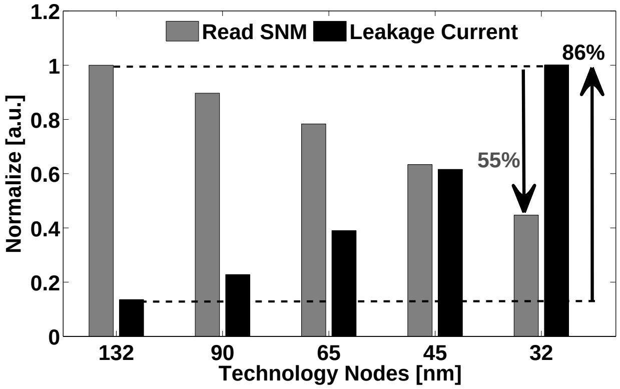


Fig. 1. Comparison of read SNM and leakage current of standard 6T SRAM bitcell for different technology nodes.

Performance: SNM can serve as a figure of merit in stability evaluation of SRAM cells. The read SNM is defined as the minimum DC noise voltage which is required to flip the state of the SRAM cell [9] during the read operation. It is measured as the length of the side of the largest square that fits inside the lobes of the butterfly curve of the SRAM. Thus, in this paper we treat the SNM as a measure of performance. The SNM of even defect-free cells is gradually declining with technology scaling, as shown in Figure 1. SRAM cells with compromised stability can limit the reliability of on-chip data storage making it more sensitive to transistor parameter shift with aging, voltage fluctuations and ionizing radiation [1]. Detection and correction/repair of such cells in modern scaled-down SRAMs becomes a necessity.

Process Variation: Millions of minimum-size SRAM cells are tightly packed making SRAM arrays the densest circuitry on a chip. Such areas on the chip can be especially susceptible and sensitive to manufacturing defects and process variations [1]. 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. In SRAMs, it is observed that as the supply voltage is reduced, the sensitivity of the circuit parameters to the process variation increases [10]. For system integration, SRAM must be compatible with subthreshold combinational logic operating at ultra-low voltages. However, this leads to increase in sensitivity to parameter variability. This problem will worsen in nanometer technologies with ultra-low voltage operation and makes SRAM design and stability analysis more challenging. The variations in threshold voltage (V_{Th}) of SRAM cell transistors due to random dopant fluctuations is the principal reason for parametric failures. The threshold voltage variation is related to the device geometry (length, width and oxide thickness) and doping profile. Equation 1 shows how the threshold voltage standard deviation ($\sigma_{V_{Th}}$) varies with the gate oxide thickness (T_{ox}), the channel dopant concentration (N_{ch}) and the channel length (L) and width (W) [11]:

$$\sigma_{V_{Th}} = \left(\frac{\sqrt[4]{4q^3 \epsilon_{Si} \phi_B}}{2} \right) \left(\frac{T_{ox}}{\epsilon_{ox}} \right) \left(\frac{\sqrt[4]{N_{ch}}}{\sqrt{WL}} \right), \tag{1}$$

where $\phi_B = 2\ \kappa_B T \ln(N_{ch}/n_i)$ with N_{ch} the channel dopant concentration, κ_B Boltzmann's constant, T the absolute temperature, n_i the intrinsic carrier concentration, q the elementary charge, and ϵ_{ox} and ϵ_{Si} the permittivity of oxide and silicon, respectively. The above expression is consistent with observations that $\sigma_{V_{Th}}$ is inversely proportional to the square root of the device area.

In order to address the above issues, we propose a methodology involving power and performance optimization in the presence of process variations in SRAM cells. However, it is a non-trivial task to simultaneously maintain reduced power dissipation, improved performance (which is SNM in this paper) and process variation tolerance. The **distinct contributions** of this research are as follows:

(1) A novel design flow for simultaneous Power-Performance-Process variation (P3) optimization in nanoscale SRAMs is introduced.
(2) 45 nm standard 6T and 8T SRAM cells are subjected to the proposed methodology.
(3) For P3 optimization of the 6T and 8T SRAM cells, we propose a novel statistical Design of Experiments (DOE) - Integer Linear Programming (ILP) based approach. It achieved 61% power reduction and 13% SNM increase.
(4) Process variation analysis of the optimal SRAM is conducted considering twelve device parameters and demonstrates the robustness of the design.
(5) The proposed methodology for P3 optimization and DOE-ILP approach is also tested on the 32 nm technology node based 6T and 8T SRAM cells.

The notations and definitions used in this paper are given in Table 1. The rest of the paper is organized in the following manner: Related prior research is discussed in section 2. Section 3 presents the proposed P3 design flow for SRAM cell optimization. The baseline SRAM design and its operation, are discussed in section 4. Section 5 highlights the statistical DOE-ILP step of the P2 design flow. This is followed by conclusions and future research in section 6.

2 Related Prior Research in SRAM

Several design and optimization methodologies have been presented in the current literature addressing the nanoscale challenges of SRAM circuits. A high-level overview of a selected subset relevant to this work is presented in Table 2.

Table 1
Notation and definitions used in this paper.

DOE	: Design of Experiments
ILP	: Integer Linear Programming
P2	: power and performance
P3	: power, performance and process variation
SNM	: Read static noise margin
V_{Th}	: threshold voltage
μ_{PWR}, σ_{PWR}	: mean and standard deviation of power of SRAM cell
μ_{SNM}, σ_{SNM}	: mean and standard deviation of SNM of SRAM cell
τ_{PWR}	: designer defined constraint for power
τ_{SNM}	: designer defined constraint for SNM
$S_{\mu_{PWR}}$, $S_{\sigma_{PWR}}$	: solution sets for mean and standard deviation of power
$S_{\mu_{SNM}}$, $S_{\sigma_{SNM}}$	: solution sets for mean and standard deviation of SNM
S_{obj}	: final objective set
S_{PWR}	: solution set for powr consumption of SRAM cell
S_{SNM}	: solution set for SNM of SRAM cell
$\cap$	: set intersection operator
V_N	: static noise voltage source
V_{DD}	: supply voltage
V_{aux}	: auxiliary function
μ, σ	: mean and standard deviation (Gaussian distribution values)
$\mu_{baseline}$, $\sigma_{baseline}$	: Gaussian mean and standard deviation for baseline desgin
P_{dyn}	: dynamic power consumption
P_{sub}	: subthreshold power
P_{gate}	: gate-oxide power
P_{total}	: total power consumption
I_{dyn}	: dynamic leakage
I_{sub}	: subthreshold leakage
I_{gate}	: gate-oxide leakage
I_{total}	: total current

Table 2
Comparison of related research in SRAM

SRAM	Power		SNM		Tech.	Research
Research	Value	% Reduction	Value	% Increase	Node	Techniques
Agrawal [12]	–		160 mV (approx.)		65 nm	Modeling based approach
Liu [13]	31.9 nW (leakage)		300 mV		65 nm	Separate data access mechanism
Kulkarni [10]	0.11 μW (leakage)		78 mV		130 nm	Schmitt Trigger
Lin [2]	4.95 nW (standby)		310 mV		32 nm	Separate read mechanism
Bollapalli [14]	10 mW (total)		–		45 nm	Separate word line groups
Azam [15]	63.9 μW vs 44.4 μW	44 % (total)	299 mV		45 nm	Separate read/write assist circuitry
Singh [16]	–	28 % (total)	–	53-61 %	65 nm	Two-port 6T-SRAM and multiport capabilities
Thakral [17]	100.5 nW	50.6 %	303.3 mV	43.9 %	45 nm	DOE-ILP
Nalam [18]	–	10-15 % (leakage)	–	10-15 %	45 nm	Two-phase Write and Split Bitline Sensing
Amelifard [9]	–	53.5 %	–	43.8 %	65 nm	Dual V_{Th} and V_{Tox}
Singh [19]	–		305 mV	65.9 %	65 nm	Subthreshold 7T-SRAM
This Paper	1.64 nW (leakage)	60 %	143.9 mV	4 %	6T, 45 nm	Statistical DOE-ILP
	2.85 nW	61 %	318.2 mV	13 %	8T, 45 nm	
	1.81 nW	53 %	81.4 mV	13 %	6T, 32 nm	
	2.34 nW	55 %	222.4 mV	12.7 %	8T, 32 nm	

The stability of the SRAM cell in the presence of random fluctuations is analyzed using a modeling based approach in [12]. In [14] the authors quote only the reduced power dissipation. In [10], a Schmitt-trigger based SRAM is proposed which provides better read stability, write ability and process variation tolerance compared to the standard 6T SRAM cell. A 9-transistor SRAM cell is proposed in [2], which increases the stability and reduces power consumption compared to the traditional 6T SRAM. A method is presented in [9,20], based on dual-V_{Th} and dual-T_{ox} assignments, for low power design of SRAM while maintaining performance. In [21] a compact model of critical charge of a 6T SRAM cell is presented for estimating the effects of process variations on its soft error susceptibility. In [16] the authors have presented a different design methodology of two-port 6T SRAM with multiport capabilities. In [18] the authors have explored power (only leakage) and SNM parameters using two phase write and split bitline differential sensing. In [17], a DOE-ILP based methodology is proposed for dual-V_{Th} assignment without accounting for process variations, which is important for nanoscale CMOS. In [15] an SNM enhancement technique is presented that results in undisturbed storage nodes but this achievement comes at the expense of additional transistors. In [22], the effect on performance and yield of the SRAM cell has been presented from BEOL (Back-end-of-line design) lithography effects, which is important in terms of manufacturing of the SRAM chip. The authors in [19] have presented a 7T SRAM topology, which is suitable for low voltage applications and it is also tolerant to read failures.

This archival journal paper is based on our conference publication [23]. The journal paper includes considerable additional material, such as functional simulation analysis of standard 6T and 8T SRAM cells (different than the previously published one) for different nano-CMOS technology nodes.

3 The Proposed Methodology for P3-Optimal Nano-CMOS SRAM

The proposed design flow to achieve P3-optimal design of both 6T and 8T SRAM circuits is shown in Algorithm 1 in pseudo-code form.

Algorithm 1 P3-optimal design methodology for nano-CMOS SRAM

```
Input: SRAM topologies (6T and 8T cells) and technology nodes (45 nm and 32 nm).
Output: P3 optimized (power minimization, performance maximization and process variation tolerant) SRAM cell.
Perform the baseline design of the SRAM cells.
Measure power and performance of baseline SRAM cells.
Goto Algorithm 2 for optimizing baseline SRAMs.
Re-simulate SRAM cells to obtain P2 (power minimization and performance maximization) SRAM cells.
Perform process variation characterization of SRAM cell using device parameters (in this case 12 device parameters).
Obtain P3 optimal SRAM cells.
Construct SRAM array to observe the feasibility of the SRAM cells.
```

The input to the proposed design flow is baseline SRAM cells which refer to the 6T and 8T SRAM circuits with nominal sized transistors for a specified technology. Maintaining an acceptable SNM as well as reduced power consumption embedded SRAMs, while scaling the minimum feature size and supply voltages of system-on-a-chip (SoC) is a very challenging task. There are various ongoing research works which discuss techniques to reduce power consumption such as dual-V_{Th}, dual-V_{DD}, etc. In this paper, we adopt the process-level technique called dual-V_{Th}. Thus, in order to achieve the optimized nano-CMOS circuit we have measured power and SNM values simultaneously using Design of Experiments (DOE). The idea is that leakage is a major component of the total power for the nano-CMOS.

Hence, by reducing power through the dual-V_{Th} technique we achieve reduction of total power along with noticeable improvement in performance.

The research problem here is defined as the selection of transistors for high V_{Th} assignment. Further, the assignment is done in such a way that along with the power reduction, the performance metric (i.e. SNM) should not be compromised. To address this research problem of choosing the correct transistors for high-V_{Th} assignment we propose a *novel statistical Design of Experiments-Integer Linear Programming (DOE-ILP) methodology* (Algorithm 2). Design of experiments or experimental design is the concept of purposeful changes of the inputs in order to study the corresponding changes in the output. A complete full factorial design matrix with two level settings per parameter (low and high voltage threshold) for n transistors would require 2^n total runs (2^6 for the 6T cell and 2^8 for the 8T cell). In order to expand the applicability of this approach to large circuits, we followed a Taguchi screening methodology, instead [24]. Taguchi designs are orthogonal with respect to the main effects (in this case the threshold voltages) but contain aliased second order interactions. Since we are subsequently applying ILP techniques, this is not a serious limitation. The implementation of a 2-Level Taguchi design matrix helps in substantially faster optimization time while maintaining good accuracy of the results. Further, ILP combined with DOE is useful for optimizing the linear objective function subject to constraints and to obtain a bound on the optimal value to solve the predictive equations that are formed using DOE. This combined approach has the potential to handle large circuits for optimization in reasonable time.

Once we obtain the P2 optimized SRAM circuit we perform process variation, where variability is considered in 12 device parameters. Detailed discussion is provided in section 5. After successfully performing the above steps we achieve the target, that is a P3 optimal SRAM cell.

Let us discuss the theory behind the ILP formulations presented in this paper (figure 2). The idea is that the baseline mean ($\mu_{baseline}$) of the quantity (power or SNM) under consideration needs to be shifted left or right depending on whether it should be minimized ($\mu_{minimized}$) or maximized ($\mu_{maximized}$). Also, the baseline standard deviation ($\sigma_{baseline}$) of the quantity (which is a measure of the spread) needs to be minimized to $\sigma_{minimized}$.

4 Design and Modeling of Baseline SRAM Circuits

A typical SRAM cell uses two cross-coupled inverters forming a latch and access transistors. The access transistors enable access to the cell during read and write operations and provide cell isolation during the not-accessed state. An SRAM cell is designed to provide non-destructive read access, successful write capability and data storage (or data retention) for as long as the cell is powered.

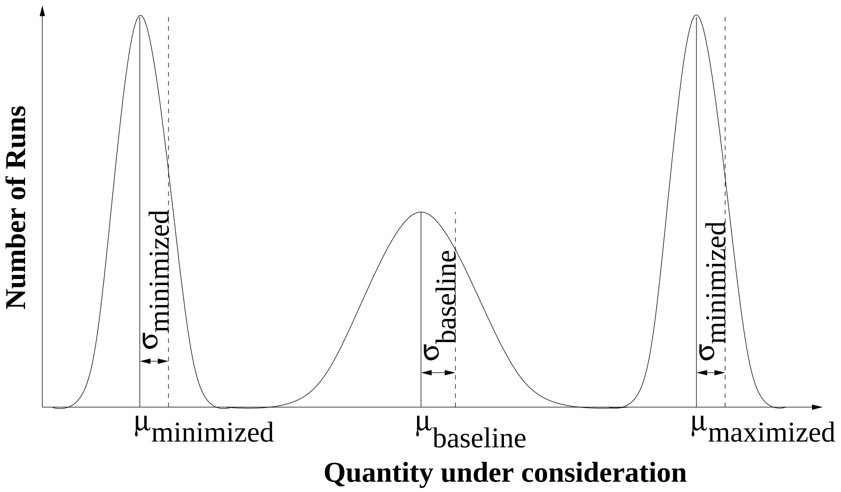
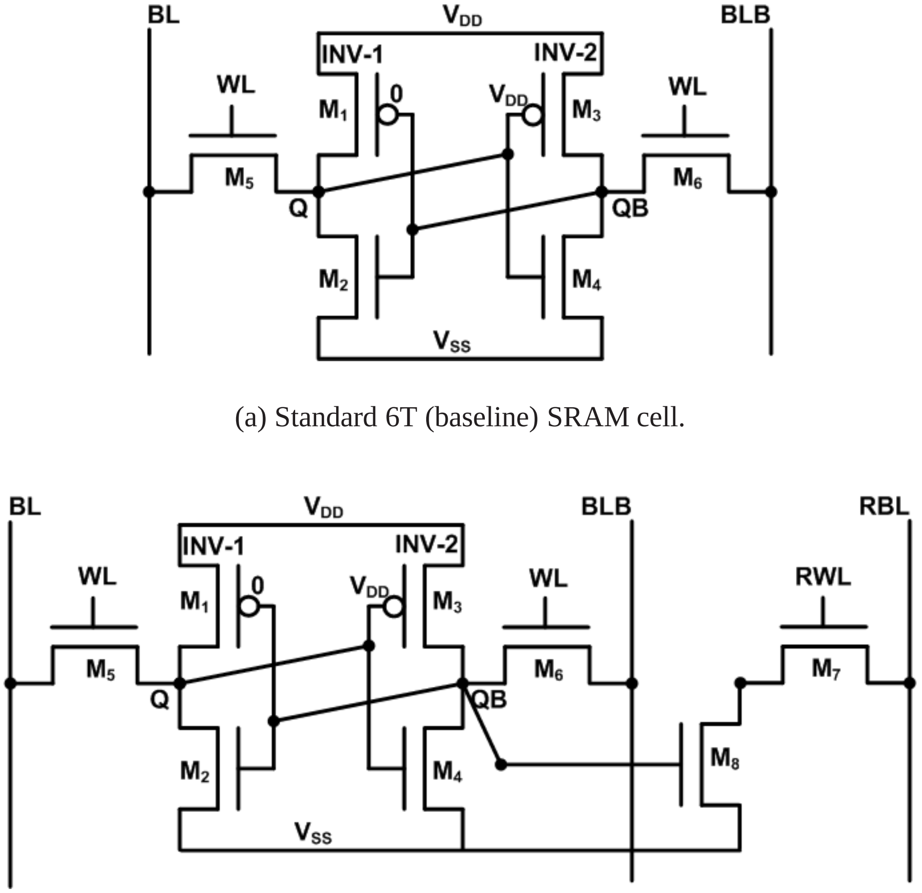


Fig. 2. Variation tolerant optimization of the SRAM.

4.1 Baseline SRAM Design for 45 nm and 32 nm CMOS

In general, the cell design must strike a balance between cell area, robustness, speed, leakage and yield [1]. Smaller cells result in a smaller array area and hence smaller bit line and word line capacitances, which in turn helps to improve the access speed performance. Reducing the transistor dimensions is the most effective means to achieve a smaller cell area. However, transistor dimensions cannot be reduced indefinitely without compromising the other parameters. For instance, smaller transistors can compromise the cell stability. Often, performance and stability objectives restrict arbitrary reduction in cell transistor sizes. Similarly, cell area can be traded off for special features such as improved radiation hardening or multi-port cell access.

The baseline standard 6T and 8T cells are shown in figure 3 (a) and (b), respectively. The standard 6T cell topology has been most commonly used in the industry, while 8T has received great attention in the recent past, as low-power substitute with significant improvement in the read the SNM as compared to the 6T cell[6,7]. In a standard 6T cell, both read and write operations are performed via the same pass gate access transistors (i.e. M_5 and M_6) as shown in figure 3 (a). As a result, there is always a conflicting read and write requirement, since, we can not simultaneously optimize both devices for read and write operations. Hence, the standard 6T cell has low read SNM which further diminishes with voltage scaling. In order to address this conflicting requirement and poor read noise margin problem, isolated read and write operation based SRAM cells are proposed. In the 8T cell, both read and write operations are isolated. The write operation is performed via pass gate access transistors (i.e. M_5 and M_6), while the read operation is performed via a separate read port which is comprised of transistor M_7 and M_8, as shown in figure 3 (b). The isolated read port provides significant improvement in read SNM, since we can optimize the SRAM cell independently for both operations. The SRAM cells have been designed at the 45 nm technology node with the supply voltage, V_{DD} = 0.9 V. The sizes of all the transistors are estimated with pull up ratio α=1 and cell ratio, β=2.



(a) Standard 6T (baseline) SRAM cell.

(b) Read SNM free 8T (baseline) SRAM cell.

Fig. 3. The standard 6T and 8T SRAM cells as baseline circuits for P3 optimization.

The power consumption and SNM of the baseline cells are measured from functional simulations and are tabulated as shown in Table 3. τ_{PWR} and τ_{SNM} are designer defined constraints in the optimization methodology. In this paper, we have taken the parameters τ_{PWR} and τ_{SNM} as baseline values which are shown in Table 3. We discuss each of the modes of operation of the 6T and 8T cells in detail in the following section.

Table 3
Leakage power and SNM for baseline SRAM cells.

Parameters	45 nm		32 nm	
	6T	8T	6T	8T
τ_{PWR}	5.70 nW	5.81 nW	5.29 nW	5.35 nW
τ_{SNM}	141.94 mV	281.44 mV	76.28 mV	197.78 mV

4.2 Modes of operation for the 6T and 8T cells

4.2.1 Read operation

Prior to initiating a read operation, the bit lines (BL and BLB) are precharged to V_{DD}. The read operation is initiated by enabling the word line (WL) and connecting the precharged bit lines to the internal nodes of the cell via access transistors (i.e. M_5 and M_6), as shown in Figure 3 (a). During read access, BLB starts discharging via node QB, and as a result there will be a potential difference between BL and BLB. This potential difference is sensed by the sense amplifier and information is read out. In order to ensure a non-destructive read operation the sizes of the transistors must be chosen carefully. For example, M_2 and M_4 must be stronger than M_5 and M_6 to keep the node voltage lower than the trip voltage of the inverters. Similarly, for a successful write operation M_5 and M_6 must be stronger than M_1 and M_3.

However, the read operation of the 8T cell is entirely different from the standard 6T cell, as shown in Figure 3 (b). In the 8T cell, the read bitline (RBL) is precharged to V_{DD} before commencing the read operation. During read access, the precharged bitline starts discharging if the node QB holds ‘0’, otherwise RBL remains high. The status of RBL is sensed by the sense amplifier to read out the information. In the 8T cell there are separate read and write ports. Therefore, the sizing requirements are relaxed and each port can be sized according to the read/write requirement.

4.2.2 Write operation

The write operation of standard 6T and 8T cells is identical. In both cells, the write operation begins with precharging the bit lines (BL and BLB). During write access, the word line (WL) is enabled connecting both access transistors to the internal data storage nodes (Q and QB). In order to flip the state of the cell as shown in figure 3, the write driver pulls down the bitline BL, which is connected to node Q, while keeping the BLB high.

4.2.3 Hold operation

The hold operation has its own significance, particularly for data retention. During hold mode, word lines (WL and RWL) are disabled and the cross coupled inverters are tightly connected to each other for longer data retention. However, hold SNM of the 6T cell is usually higher than the read SNM. In the 8T M cell, the hold SNM is almost equal to the read SNM because of the separate read port.

4.3 Leakage Measurement

Leakage power plays a vital role in the nano-regime and in certain SoC applications it dominates the dynamic power. This section deals with different leakage power measurements of standard 6T and 8T cells under the idle sate.

4.3.1 Power Model

The major sources of power dissipation for a nano-CMOS circuit are due to capacitive switching, subthreshold leakage, and gate leakage. Both dynamic and static power are significant fractions of total power dissipation. Each one of them has several forms and origins; they flow between different terminals and in different operating conditions of a transistor. It is essential to study the power consumption profile of SRAMs in order to estimate and minimize their power consumption, especially when they are made of nanoscale CMOS transistors. An SRAM consumes dynamic power only when the bitline or wordline are switching their level from low-to-high or high-to-low for Write or Read operations. On the other hand, including the hold (idle) state, power dissipation happens continuously in the form of gate oxide leakage and subthreshold leakage. In general, SRAM contributes to the major portion of the total leakage power in a modern processor during idle states.

4.3.2 Leakage Model

The leakage model consists of subthreshold leakage current and gate oxide leakage current. We discuss each of them in brief. The subthreshold leakage is modeled as follows [1]:

$$I_{sub} = I_S \exp\left(\frac{V_{gs} - V_{Th}}{nv_t}\right)\left(1 - \exp\left(\frac{-V_{ds}}{v_t}\right)\right), \qquad (2)$$

where $n = \left(1 + \frac{C_d}{C_{ox}}\right)$, $v_t = \left(\frac{kT}{q}\right)$ is the thermal voltage, V_{Th} is the threshold voltage, I_S is the current when V_{gs} equals V_{Th}, V_{gs} is the gate-to-source voltage, and V_{ds} is drain-to-source voltage.

The gate oxide leakage current is modeled using the following expression [25]:

$$I_g = AWL\left(\frac{T_{oxref}}{t_{ox}}\right)^{ntox}\left(\frac{V_g\,V_{aux}}{t_{ox}^2}\right)e^{-B(\alpha-\beta|V_{ox}|)(1+\gamma|V_{ox}|)t_{ox}}, \tag{3}$$

where $A = \left(\frac{q^2}{8\Pi h\phi_b}\right)$, $B = \left(\frac{8\pi\sqrt{2qm_{ox}\phi_b^{3/2}}}{3h}\right)$, m_{ox} is the effective carrier mass in the oxide, ϕ_b is the tunneling barrier height, t_{ox} is the oxide thickness, T_{oxref} is the reference oxide thickness at which all parameters are extracted, $ntox$ is a fitting parameter, V_{aux} is an auxiliary function which approximates the density of tunneling carriers as well as available states, and α, β and γ are the controlling parameters for electron tunneling.

In addition, leakages consists of diode leakage flowing in the transistors of the cell. The diodes are formed between the diffusion region of the transistor and the substrate consumes power in the form of reverse bias current which is drawn from the power supply.

4.3.3 Leakage Current Paths in the Hold State

The current flow in each transistor of the cell depends on its location and the operation being performed. The current paths for hold (idle) state are shown in figure 4 for the 6T cell. The solid arrows shown in the figure are for the subthreshold current. The dashed arrows represent gate oxide leakage current which is present in the transistor when they are in the “OFF” state. Essentially, when the transistor is in the “ON” state it carries dynamic current along with the gate oxide leakage current and when the transistor is in the “OFF” state it will have gate oxide leakage current as well as subthreshold leakage current.

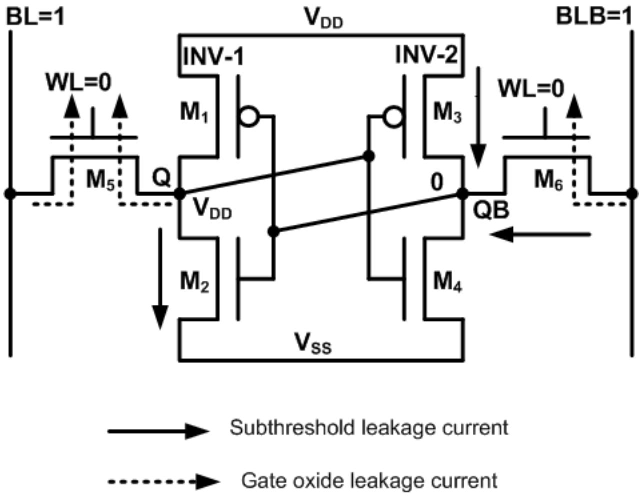


Fig. 4. Leakage current paths during the hold state for the 6T (baseline) cell.

We discuss the hold state current paths in detail, as shown in Figures 4 and 5, for 6T and 8T cells. In the hold state, the word line is disabled (WL = ‘0’) and the bit lines

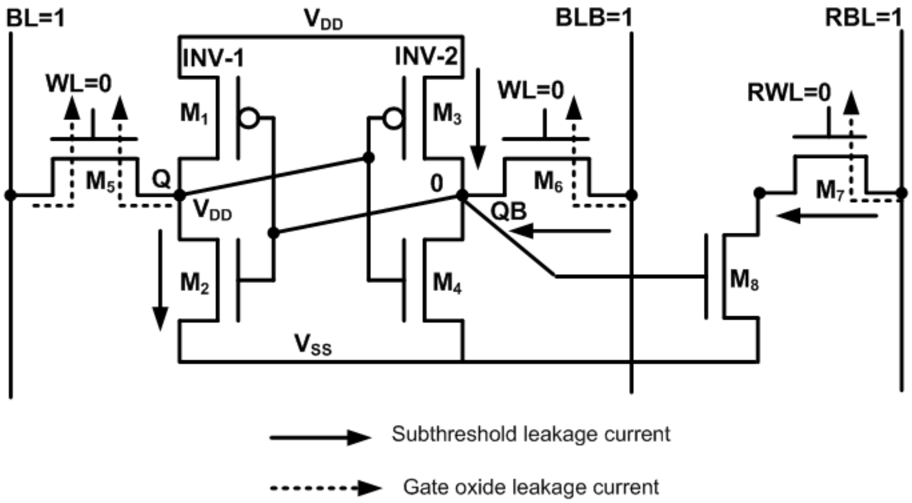


Fig. 5. Leakage current paths during the hold state for the 8T (baseline) cell.

(BL and BLB) are tied to ‘l’. Under this state, transistor M_5 and M_6 are in cut-off, carrying gate oxide leakage current. On the other hand, transistor M_2 and M_3 carry subthreshold leakage current and preserve the cell state (i.e. node Q = V_{DD} and node QB = ‘0’). However, in the 8T cell the read-port (comprised of transistor M_7 and M_8) adds two more leakage current components and increases overall leakage power, as shown in Figure 5. Leakage power in both cells is measured as the power supplied by V$_{DD}$, when all word line and bit lines are connected appropriately and data storage nodes (Q and QB) are maintained appropriately for sufficient time to complete the operation under study.

4.4 SNM Model and Measurement

SNM can serve as a figure of merit in stability evaluation of SRAM cells. The SNM measurement model is described in this section. The SNM of even defect-free cells is declining with technology scaling, as discussed in previous sections. SRAM cells with compromised stability can limit the reliability of on-chip data storage making them more sensitive to transistor parameter shift with aging, voltage fluctuations and ionizing radiation. Detection and correction/repair of such cells in modern scaled-down SRAMs becomes a necessity. Figure 6 (a) shows the simulation setup for the 6T cell SNM measurement, consisting of the two inverters (INV-1 and INV-2) in feedback and voltage sources V_N. The same SNM simulation setup can easily be extended for the 8T cell. In other words, the hold SNM setup is equivalent to the hold and read SNM setup of the 8T cell. The two voltage sources are static noise sources. A static noise source can be defined as DC disturbance and mismatch due to variations and processing in the operating conditions of the cell [26]. The two DC voltage sources V_N are placed in adverse direction to the input of the inverters of the SRAM circuit in order to obtain the worst case SNM. The SNM is the maximum amount of noise that can be tolerated at the cell nodes just before flipping the states. In order to obtain the butterfly curve shown in Figure 6 (b), the voltages are varied to and from nodes Q and QB alternatively. The SRAM cell is simulated for 45 nm CMOS technology using the PTM model [3] with supply voltage V_{DD} of 0.9 V and with minimum sized transistors. The worst case SNM obtained from the butterfly curves are also shown in dotted lines in Figure 6 (b) and marked with a small circle.

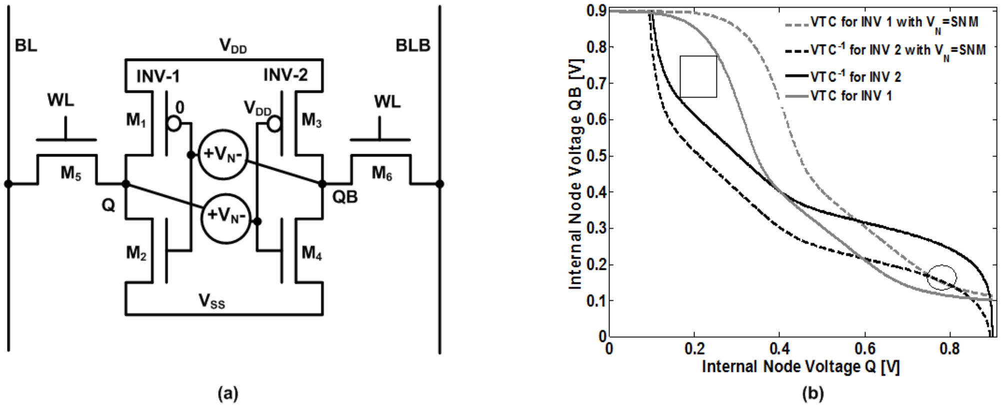


Fig. 6. Simulation set-up for SNM measurement.

Table 3 shows leakage power and SNM results for the baseline design (6T and 8T cells). The PVT condition is nominal process voltage variation and temperature is taken as room temperature or 27^{o}C.

It may be noted that SRAM circuits have many other figures of merit, including read delay and hold SNM which can be considered for optimization. However, this particular paper is inspired by our earlier publication which demonstrates that read SNM is a very important figure of merit [27]. The current paper emphasizes mainly two figures of merit, power consumption and read SNM.

5 Statistical DOE-ILP Algorithm for P2 Optimization

This section discusses in detail the implementation of the statistical Design of Experiments (DOE)-Integer Linear Programming (ILP) algorithm, which is at the heart of the P3 optimization design flow.

5.1 The Optimization Algorithm

As shown in Algorithm 2, the baseline SRAM cells are taken as the input along with the baseline model file and high threshold model file. The PVT condition is nominal process values for all devices, nominal power supply and the temperature is taken as room temperature or 27^oC. We subject the baseline 6T and 8T cells to a DOE [28] based approach using a 2-Level Taguchi L_8 array. The factors are the V_{Th} states of the different transistors of the SRAM cells (figure 3). Each factor can take a high V_{Th} state (1) or a nominal V_{Th} state (0). The L_8 array provides different experimental runs for 6T and 8T cells. Monte Carlo simulations for N runs are performed for each experiment trial. The mean (μ) and standard deviation (σ) values of the resulting probability density function (approximated by a histogram) are recorded for average power and performance (SNM) of the SRAM cell. Thereafter, using DOE, predictive equations are formed for μ and σ and are denoted by $\widehat{\mu_{PWR}}$, $\widehat{\sigma_{PWR}}$ for power and for SNM as $\widehat{\mu_{SNM}}$, $\widehat{\sigma_{SNM}}$. These predictive equations $\widehat{\mu_{PWR}}$, $\widehat{\sigma_{PWR}}$, $\widehat{\mu_{SNM}}$, $\widehat{\sigma_{SNM}}$ are considered to be linear equations with the constraints being high V_{Th} (or state 1) and low V_{Th} (or state 0). Each of these linear equations is then solved using integer linear programming (ILP), depending on whether the quantity under consideration is to be maximized or minimized. The complexity of the algorithm otherwise would be $O(2^n)$ where n is the transistor number.

We obtain the solution sets for mean and standard deviation of power as $S_{\mu PWR}$, $S_{\sigma PWR}$ and the solution sets for mean and standard deviation for SNM as $S_{\mu SNM}$, $S_{\sigma SNM}$. Since we are interested in power minimization and SNM maximization, we form our final objective S_{obj} as $S_{\mu PWR} \cap S_{\sigma PWR} \cap S_{\mu SNM} \cap S_{\sigma SNM}$ ($\cap$ is defined as the intersection of the sets $S_{\mu PWR}$, $S_{\sigma PWR}$, $S_{\mu SNM}$ and $S_{\sigma SNM}$). This is the *strength of the proposed algorithm:* it allows seamless simultaneous optimization of diverse and conflicting objectives. In the case of different objectives the optimization results in a set of transistors, not a specific value in terms of power or SNM. The sets are then combined depending on the multiple objectives targeted for optimization.

Based on S_{obj}, we assign high V_{Th} to the transistors of the cell, and re-simulate to obtain a P3 optimal design. The design flow achieves power reduction and read stability increase. Using this optimized cell, the design flow constructs the SRAM array. However, the scope of this paper has been kept at cell-level optimization.

Monte Carlo simulations of 1000 runs are performed for each experiment. Therefore, we have a total of 6K (for 6T SRAM cell) and 8K (for 8T SRAM cell) Monte Carlo runs, taking 12 parameters in account. The 12 process parameters considered are as follows: (1) T_{oxn}: NMOS gate oxide thickness (nm), (2) T_{oxp}: PMOS gate oxide thickness (nm), (3) L_{na}: NMOS access transistor channel length (nm), (4) L_{pa}: PMOS access transistor channel length (nm), (5) W_{na}: NMOS access transistor channel width (nm), (6) W_{pa}: PMOS access transistor channel width (nm), (7) L_{nd}: NMOS driver transistor channel length (nm), (8) W_{nd}: NMOS driver transistor channel width (nm), (9) L_{pl}: PMOS load transistor channel length (nm), (10) W_{pl}: PMOS load transistor channel width (nm), (11) N_{chn}: NMOS channel doping concentration (cm^{-3}), (12) N_{chp}: PMOS channel doping concentration (cm^{-3}). It may be noted that statistical information about these parameters may not be provided by the foundry. However, they are identified based on various published works [29].

Algorithm 2 P2 optimization in nano-CMOS SRAM

```
Input: Baseline PWR and SNM of the SRAM cell, baseline model file, high-threshold model file.
Output: Optimized objective set $f_{obj}$ = $[f_{PWR},\ f_{SNM}]$ optimal SRAM cell with transistors identified for high $V_{Th}$ assignment.
Setup experiment for transistors of SRAM cell using 2-Level Taguchi L-8 array, where the factors are the $V_{Th}$ states of transistors of SRAM cell, the response for average power consumption is $\widehat{\mu_{PWR}}$, $\widehat{\sigma_{PWR}}$ and the response for read SNM is $\widehat{\mu_{SNM}}$, $\widehat{\sigma_{SNM}}$.
for Each 1:8 experiments of 2-Level Taguchi L-8 array do
    Perform $N$ Monte Carlo runs
    Record $\mu_{PWR}$, $\sigma_{PWR}$ and $\mu_{SNM}$, $\sigma_{SNM}$
end for
Form linear predictive equations
$\widehat{\mu_{PWR}}$, $\widehat{\sigma_{PWR}}$ for power
$\widehat{\mu_{SNM}}$, $\widehat{\sigma_{SNM}}$ for SNM.
Solve $\widehat{\mu_{PWR}}$ using ILP: Solution set $S_{\mu PWR}$.
Solve $\widehat{\sigma_{PWR}}$ using ILP: Solution set $S_{\sigma PWR}$.
Solve $\widehat{\mu_{SNM}}$ using ILP: Solution set $S_{\mu SNM}$.
Solve $\widehat{\sigma_{SNM}}$ using ILP: Solution set $S_{\sigma SNM}$.
Form $S_{obj}$ = $S_{\mu PWR} \cap S_{\sigma PWR} \cap S_{\mu SNM} \cap S_{\sigma SNM}$.
Assign high $V_{Th}$ to transistors based on $S_{obj}$.
Re-simulate SRAM cell to obtain optimized objective set.
```

The objective is to make the data characterization as accurate as possible for the current technology. Each of these process parameters is considered to have a Gaussian distribution with mean (μ) taken as the nominal values specified in the PTM [3] and standard deviation (σ) as 10% of the mean. Amongst these parameters some are independent and others are correlated which is considered during the simulation. A correlation coefficient of 0.9 between T_{oxn} and T_{oxp} is assumed. The responses under consideration are the mean μ_{PWR} and standard deviation σ_{PWR} of the average power consumption and also the mean μ_{SNM} and standard deviation σ_{SNM} of the read SNM of the cell.

The experiments are performed and the half effects are recorded using the following expression:

$$\frac{\Delta(n)}{2} = \left(\frac{avg(1) - avg(0)}{2}\right), \tag{4}$$

where $\left[\frac{\Delta(n)}{2}\right]$ is the half-effect of the n-th transistor, $avg(1)$ is the average value of power when transistor n is in the high-V_{Th} state, and $avg(0)$ is the average value of power when transistor n is in the nominal V_{Th} state.

We have taken normalized predictive equations in order to eliminate the effect of two different units, that is, nW for power and mV for SNM. The normalized pre-

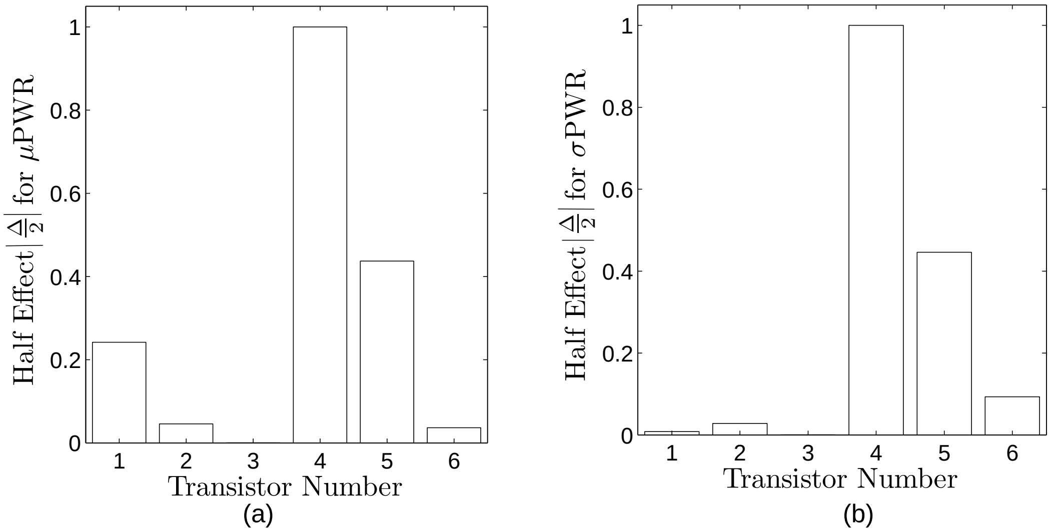


Fig. 7. Pareto plot of 6T SRAM cell for (a) mean leakage power (μPWR) and (b) standard deviation of leakage power (σPWR).

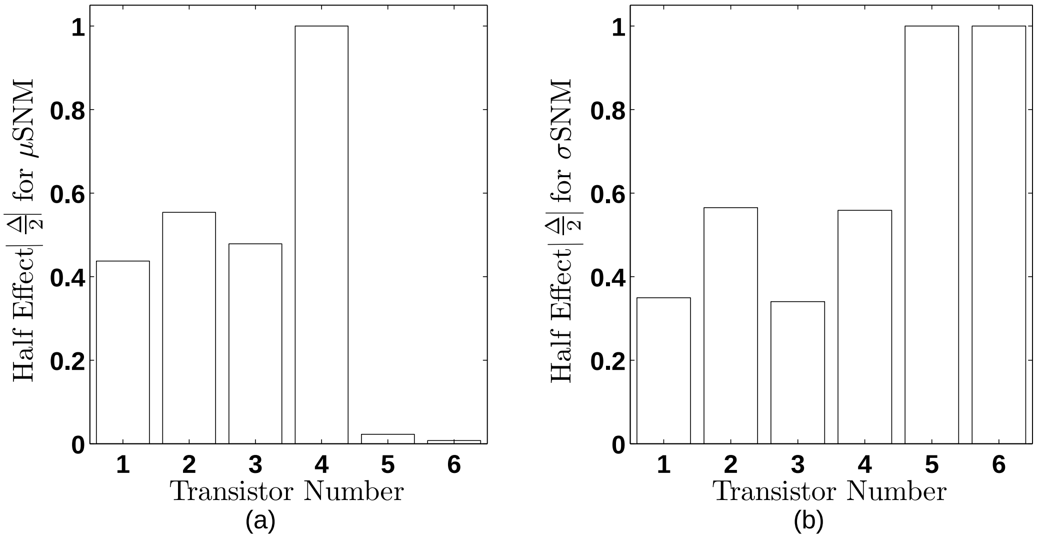


Fig. 8. Pareto plot of 6T SRAM cell for (a) mean read SNM (μSNM) and (b) standard deviation of read SNM (σSNM).

dictive equations are:

$$\hat{f} = \bar{f} + \sum_{n=1}^{6or8} \left(\frac{\Delta(n)}{2} x_n \right), \tag{5}$$

where $\hat{f}$ is the predicted response, $\bar{f}$ is the average of the responses, $\left[\frac{\Delta(n)}{2}\right]$ is the half effect of the n-th transistor, and x_n is the V_{Th} state of the n-th transistor.

The predictive equation for the mean of the average power consumption of the 6T cell is:

$$\begin{aligned}\widehat{\mu_{PWR_{6T}}} = & \, 0.29 - 0.24x_1 + 0.05x_2 \\ & -1.0x_4 + 0.43x_5 + 0.03x_6. \end{aligned} \tag{6}$$

Here, x_i represents the V_{Th} state of transistor i (M_i in figure 3 (a)). Figure 7 (a) shows Pareto plots of the half-effects of the 6T transistors for $\mu_{PWR_{6T}}$. From this, we formulate an ILP problem:

$$\begin{aligned} \min \; & \widehat{\mu_{PWR6T}} \\ \text{s.t.} \; & x_n \in \{0,1\} \, \forall n \\ & \mu_{SNM} > \tau_{SNM}. \end{aligned}$$

As we wish to minimize power consumption, we minimize $\widehat{\mu_{PWR_{6T}}}$. The constraints ‘1’ and ‘0’ represent coded values for high V_{Th} and nominal V_{Th} states, respectively. ILP has been used for small circuits, but the methodology is automated, and hence can be used for larger circuits. Solving the ILP problem, we obtain the optimal solution as: $S_{\mu PWR_{6T}}$ = [x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 0, and x_6 = 1]. This can also be interpreted as transistors 1, 4, and 6 are high V_{Th} transistors, and transistor 2, 3, and 5 are minimal V_{Th} transistors.

The Pareto plot of the half-effects for $\sigma_{PWR_{6T}}$ of 6T SRAM cell is shown in figure 7 (b). Similarly, equation 7 shows the predictive equation for the standard deviation of the leakage power consumption of the SRAM cell:

$$\begin{aligned}\widehat{\sigma_{PWR_{6T}}} = & \, 0.26 + 0.03x_2 + 1.0x_4 \\ & -0.453x_5 + 0.09x_6. \end{aligned} \tag{7}$$

From this, we formulate an ILP problem:

$$\begin{aligned} \min \; & \widehat{\sigma_{PWR_{6T}}} \\ \text{s.t.} \; & x_n \in \{0,1\} \, \forall n \\ & \mu_{SNM} > \tau_{SNM}. \end{aligned}$$

Since we seek to minimize the standard deviation of leakage power consumption, we minimize $\widehat{\sigma_{PWR_{6T}}}$. Solving the ILP problem, we obtain the optimal solution as:

$S_{\sigma PWR_{6T}}$ = [x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 1, and x_6 = 0]. This can also be interpreted as transistor 5 is high V_{Th} and transistors 1, 2, 3, 4, and 6 are nominal V_{Th} transistors.

The predictive equation for $\mu_{SNM_{6T}}$ for the 6T cell is:

$$\begin{aligned}\widehat{\mu_{SNM}}_{6T} = &0.42 + 0.44x_1 + 0.55x_2\\ &+0.48x_3 + 1.0x_4 - 0.02x_5\\ &+0.01x_6,\end{aligned} \tag{8}$$

Figure 8 (a) shows the Pareto plot of the half-effects of the transistors for $\mu_{SNM_{6T}}$ for the 6T cell.

Equation 8 shows the predictive equation for mean of the read SNM of the 6T cell. From this, we formulate an ILP problem:

$$\begin{aligned}\max\ &\widehat{\mu_{SNM}}_{6T}\\ \text{s.t.}\ \ &x_n \in \{0,1\}\,\forall n\\ &\mu_{PWR} < \tau_{PWR}.\end{aligned}$$

Since we want to maximize SNM , we maximize $\widehat{\mu_{SNM}}_{6T}$. Solving the ILP problem, we obtain the optimal solution as: $S_{\mu SNM_{6T}}$ = [x_1 = 1, x_2 = 1, x_3 = 1, x_4 = 1, x_5 = 0, and x_6 = 1]. This can also be interpreted as transistors 1, 2, 3, 4 and 6 are high V_{Th} transistors, and transistor 5, is nominal V_{Th} transistor.

Figure 8 (b) show the Pareto plot of the half-effects of the transistors for σ_{SNM}. The predictive equation for σ_{SNM} is formed as shown in equation 9. Next, we compute the standard deviation of the read SNM for 6T SRAM cell:

$$\begin{aligned}\widehat{\sigma_{SNM}}_{6T} = &0.64 - 0.35x_1 + 0.57x_2\\ &+0.34x_3 + 0.56x_4 + 1.0x_5\\ &-1.0x_6.\end{aligned} \tag{9}$$

From this, we formulate an ILP problem for the 6T cell as follows:

$$\begin{aligned}\min\ &\widehat{\sigma_{SNM_{6T}}}\\ \text{s.t.}\ \ &x_n \in \{0,1\}\,\forall n\\ &\mu_{PWR} < \tau_{PWR}.\end{aligned}$$

As we want to minimize the standard deviation (which is an indication of the spread) of read SNM, we minimize $\widehat{\sigma_{SNM}}$. Solving the ILP problem, we obtain the optimal solution as: $S_{\sigma SNM_{6T}}$ = [x_1 = 1, x_2 = 0, x_3 =0, x_4 = 1, x_5 = 0, and x_6 = 1]. This can also be interpreted as transistors 1, 4 and 6 are high V_{Th} transistors, and transistor 2, 3, and 5 are nominal V_{Th} transistors.

Our final objective function $S_{obj_{6T}}$ is formed as follows:

$$S_{obj_{6T}} = S_{\mu PWR_{6T}} \cap S_{\sigma PWR_{6T}} \cap S_{\mu SNM_{6T}} \cap S_{\sigma SNM_{6T}}, \tag{10}$$

where $\cap$ is interpreted as the set intersection operator. In other words, we pick devices which are part of low-power and high-SNM solution sets. We form normalized equations for power and SNM so that there is no unit interference. We obtain, $S_{obj_{6T}}$ = [x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 0, and x_6 = 1], i.e., transistors 1, 4, and 6 are high V_{Th} transistors, and transistors 2, 3, and 5 are nominal V_{Th} transistors. Figure 9 (a) shows the P3 optimized standard 6T SRAM cell having high V_{Th} transistors are hatched.

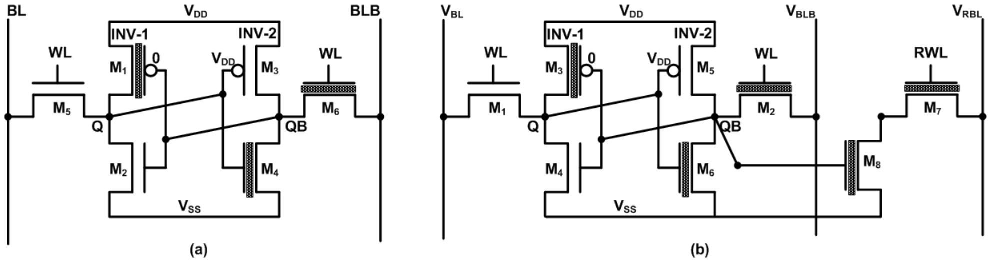


Fig. 9. P3 optimized (a) standard 6T and (b) read SNM free 8T SRAM cells; with hatched transistors having high V_{Th} .

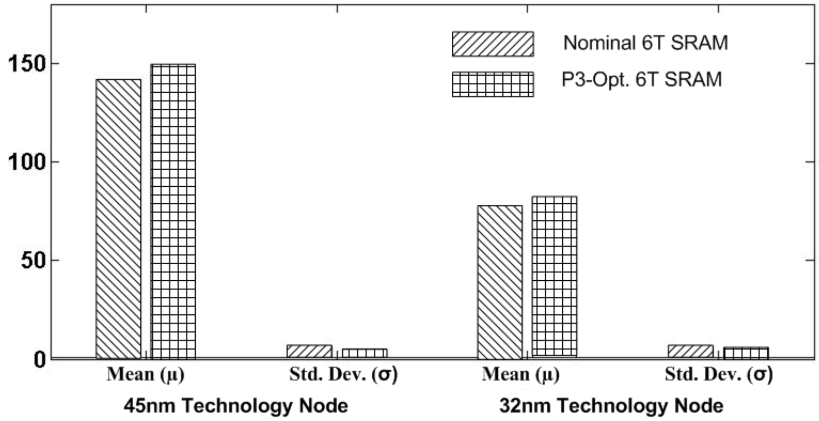


Fig. 10. Statistical mean and standard deviation of read SNM of a nominal and P3 optimized 6T SRAM cell for 45nm and 32nm technology node.

In order to demonstrate the effectiveness of the proposed algorithm (DOE-ILP P3-Optimization), we simulated the 6T and 8T cells for different technology nodes (45 nm and 32 nm). Figures 10 and 11 show the DOE-ILP based dual-V_{Th} assignment results of standard 6T SRAM cell. There is a marginal increase in the read SNM of the 45 nm and 32 nm nodes, while there is a significant reduction (60%) in the mean

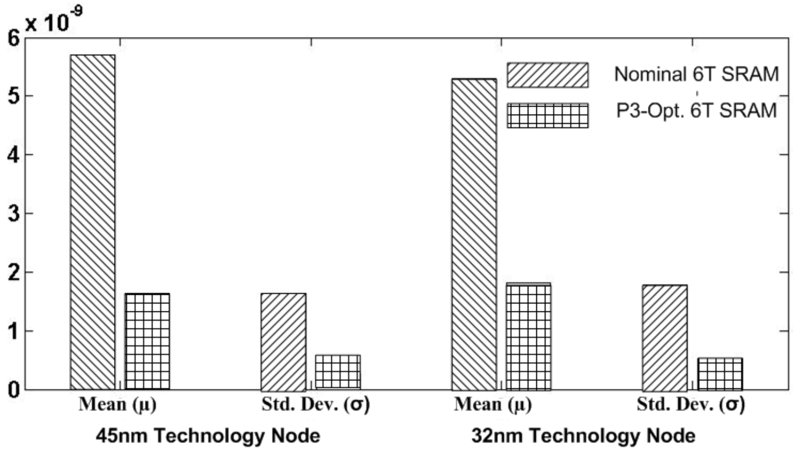
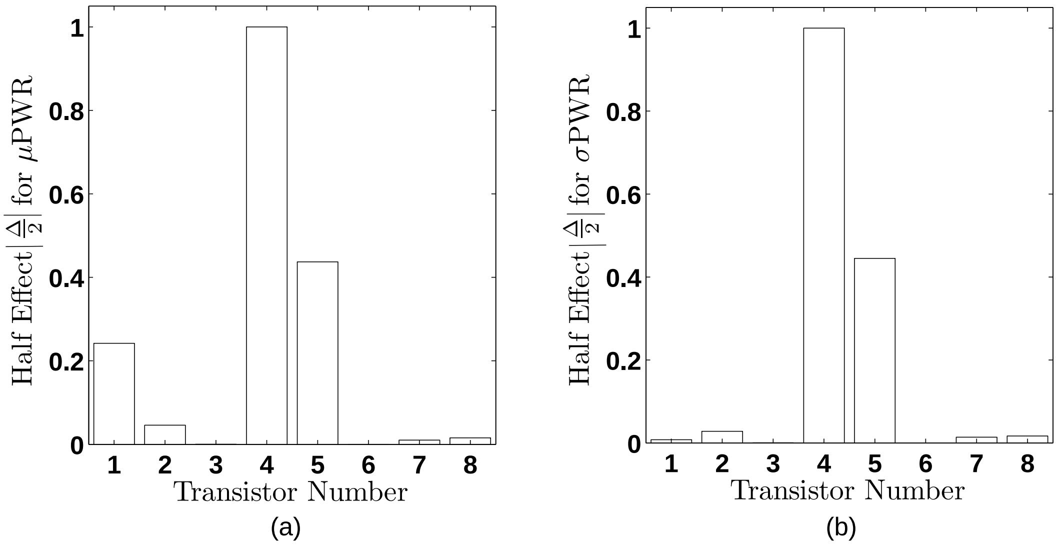


Fig. 11. Statistical mean and standard deviation of leakage power of a nominal and P3 optimized 6T SRAM cell for 45nm and 32nm technology node.

leakage power under P3 optimized approach. However, the small increase in read SNM of the 6T cell is mainly due to the very strict optimization space available. These results are comparable to previous approaches which did not account for process variations [17].



(a)

(b)

Fig. 12. Pareto plot of 8T SRAM cell for (a) mean leakage power (μPWR) and (b) standard deviation of leakage power (σPWR).

5.3 P3 Optimization of the 8T cell

The predictive equations for the mean and standard deviation of leakage power consumption of the 8T cell are:

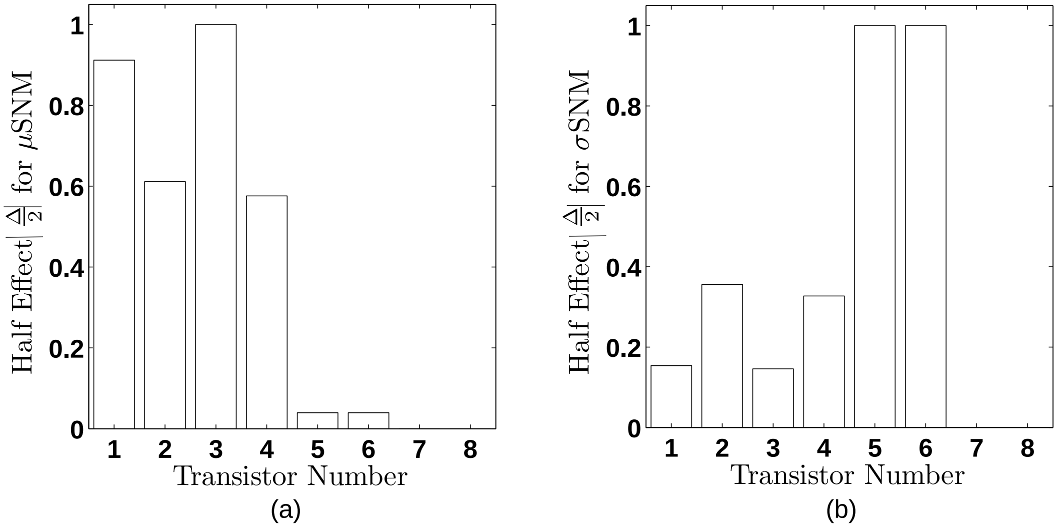


Fig. 13. Pareto plot of 8T SRAM cell for (a) mean read SNM (μSNM) and (b) standard deviation of read SNM (σSNM).

$$\begin{aligned}\widehat{\mu_{PWR_{8T}}} = & 0.3 + 0.24x_1 + 0.06x_2 \\ & +1.0x_4 + 0.43x_5 + 0.01x_7 + 0.02x_8. \end{aligned} \quad (11)$$

$$\begin{aligned}\widehat{\sigma_{PWR_{8T}}} = & 0.10 + 0.01x_1 + 0.03x_2 \\ & +1.0x_4 + 0.44x_5 + 0.01x_7 + 0.01x_8. \end{aligned} \quad (12)$$

Figures 12 (a) and (b) show the Pareto plots of the half-effects of the transistors for $\mu_{PWR_{8T}}$ and $\sigma_{PWR_{8T}}$, respectively. From this, we formulate the ILP problem for minimization of $\mu_{PWR_{8T}}$ and $\sigma_{PWR_{8T}}$:

$$\begin{aligned} & \min\ \widehat{\mu_{PWR_{8T}}}\ \ and \\ & \min\ \widehat{\sigma_{PWR_{8T}}} \\ & \text{s.t. }\ x_n \in \{0,1\}\,\forall n \\ & \qquad \mu_{SNM} > \tau_{SNM}. \end{aligned}$$

Since we wish to minimize the leakage power consumption, we minimize $\widehat{\mu_{PWR_{8T}}}$ and $\widehat{\sigma_{PWR_{8T}}}$. Solving the above formulated ILP problem, we obtain the optimal solution as: $S_{\mu PWR_{8T}}$ = [x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 0, x_6 = 1, x_7 = 1 and x_8 = 1]. This can be interpreted as transistors 1, 4, 6, 7 and 8 are high V_{Th} transistors, and transistor 2, 3, and 5 are nominal V_{Th} transistor. Similarly for $S_{\sigma PWR_{8T}}$ = [x_1 = 0, x_2 = 0, x_3 = 0, x_4 =1, x_5 = 1, x_6 = 0, x_7 = 1 and x_8 = 1]. This can also be interpreted as transistors 4, 5, 7 and 8 are high V_{Th} transistors, and transistor 1, 2, 3 and 5 are nominal V_{Th} transistor.

Pareto plots of the half-effects of the transistors for $\mu_{SNM_{8T}}$ and $\sigma_{SNM_{8T}}$, respectively, for the 8T cell are shown in Figure 13 (a) and (b). Equations 13 and 14 show the derived predictive equation for mean and standard deviation of the read SNM of the 8T cell:

$$\begin{aligned}\widehat{\mu_{SNM}}_{8T} = & 0.40 + 0.91x_1 + 0.03x_2\\ & + 1.0x_3 + 0.58x_4 - 0.04x_5\\ & + 0.4x_6,\end{aligned} \tag{13}$$

$$\begin{aligned}\widehat{\sigma_{SNM}}_{8T} = & 0.37 + 0.15x_1 + 0.35x_2\\ & + 0.15x_3 - 0.33x_4 + 1.0x_5\\ & + 1.0x_6.\end{aligned} \tag{14}$$

In order to maximize the predictive equations formed above for $\widehat{\mu_{SNM_{8T}}}$ and $\widehat{\sigma_{SNM_{8T}}}$, we formulate an ILP problem:

$$\begin{aligned}\max\ & \widehat{\mu_{SNM_{8T}}}\ \ and\\ \min\ & \widehat{\sigma_{SNM_{8T}}}\\ \text{s.t.}\ & x_n \in \{0,1\}\, \forall n\\ & \mu_{PWR} < \tau_{PWR}.\end{aligned}$$

As we want to maximize SNM, we maximize $\widehat{\mu_{SNM_{8T}}}$ and $\widehat{\sigma_{SNM_{8T}}}$. Solving the ILP problem, we obtain the optimal solution as: $S_{\mu SNM_{8T}}$ = [x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 0, x_6 = 1, x_7 = 1 and x_8 = 1]. This can also be interpreted as transistors 2, 3 and 5 are nominal V_{Th} transistors, and transistors 1, 4, 6, 7 and 8 are high V_{Th} transistor. Similarly, for $S_{\sigma SNM_{8T}}$ = [x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 1, x_6 = 1, x_7 = 1 and x_8 = 1]. This can also be interpreted as transistors 1, 4, 5, 7 and 8 are high V_{Th} transistors, and transistors 2 and 3 are nominal V_{Th} transistor.

Our final objective function $S_{obj_{8T}}$ is formed as follows:

$$S_{obj_{8T}} = S_{\mu PWR_{8T}} \cap S_{\sigma PWR_{8T}} \cap S_{\mu SNM_{8T}} \cap S_{\sigma SNM_{8T}}, \tag{15}$$

where $\cap$ is interpreted as the set intersection operator. In other words, we pick devices which are part of low-power and high-SNM solution sets. We form normalized equations for power and SNM so that there is no unit interference because we wish to achieve a low power and high stability in our proposed design. We obtain, $S_{obj_{8T}}$ = [x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 0, x_6 = 1, x_7 = 1 and x_8 = 1], i.e., transistors 1, 4, 6, 7 and 8 are high V_{Th} transistors, and transistors 2, 3, and 5 are nominal V_{Th} transistors. Figure 9 (b) shows the P3 optimized 8T SRAM cell with the high V_{Th} transistors hatched.

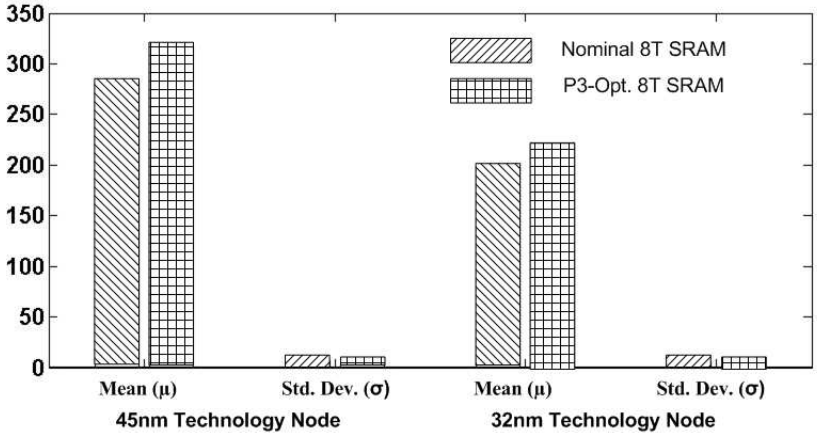


Fig. 14. Statistical mean and standard deviation of read SNM of a nominal and P3 optimized 8T cell for the 45 nm and 32 nm technology nodes.

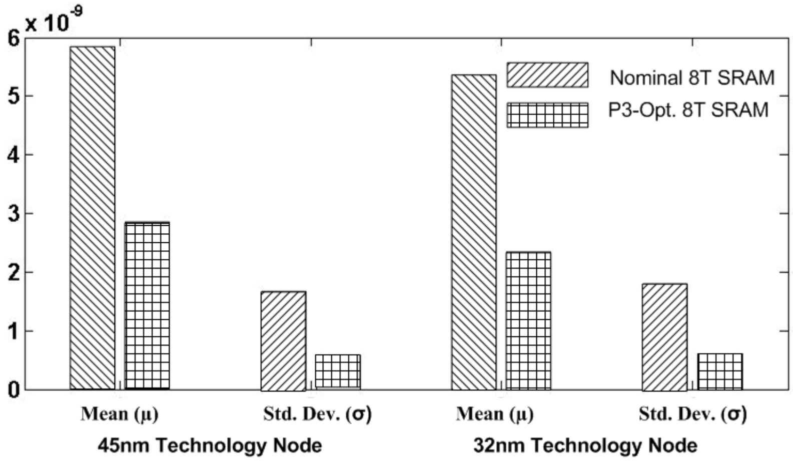
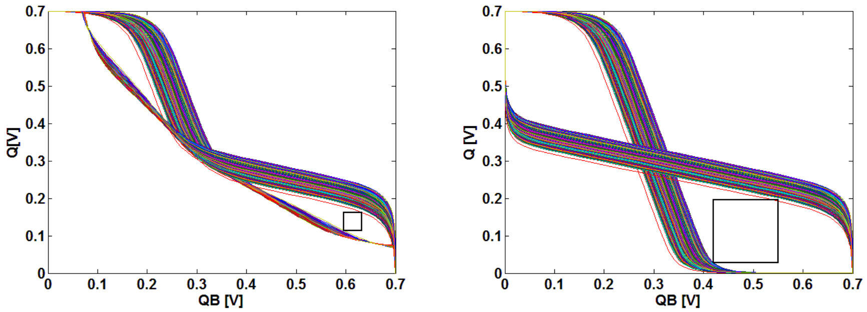


Fig. 15. Statistical mean and standard deviation of leakage power of a nominal and P3 optimized 8T cell for the 45 nm and 32 nm technology node.

Figures 14 and 15 show the DOE-ILP based dual-V_{Th} assignment results obtained from the P3 optimized 8T cell, shown in Figure 9 (b). The absolute value of the read SNM of the 8T cell is 2× higher than the 6T cell. However, there is a 13% increase in the read SNM of the 45 nm and 32 nm nodes with the P3 optimization approach, while the standard deviation of read SNM is almost unchanged. A significant leakage power reduction (51%) under P3 optimized approach is observed with marginal reduction in the standard deviation of the leakage power. These results are very promising and the proposed approach is more suitable for the read SNM free SRAM cells, such as 8T, 9T and 10T [30–35,6,7]. A 13% increase in read SNM of the 8T cell is almost equivalent to 30% of the total read SNM of the standard 6T cell as can be observed from Figures 10 and Figure 14. Figures 16 (a) and (b) show the butterfly curves for the P3 optimized 6T and 8T cells simulated for the 32 nm node. The squares embedded inside the butterfly curves are a measure of the read SNM under process variation. It can be observed that the read SNM of the 8T cell is better than that of the 6T Scell.



(a) Butterfly curves of the 6T SRAM cell.

(b) Butterfly curves of the 8T SRAM cell.

Fig. 16. The standard 6T and 8T SRAM cells as baseline circuits for P3 optimization.

5.4 Comparative Analysis of the Results

In order to obtain a broad perspective of performance for the current algorithms, we compare with some indirectly related work here. The method presented in [9,20] is based on dual-V_{Th} and dual-T_{ox} assignment for low power design while maintaining performance. In [9], a combined dual-V_{Th} and dual-T_{ox} assignment is presented which improves power (only leakage is considered) by 53.5% and SNM by 43.8%. The desired results are obtained by using *both* dual-V_{Th} and dual-T_{ox} assignments, which requires a larger number of masks and lithography steps during fabrication. In the current paper, we have taken into account subthreshold and gate-oxide leakage power which results in total improvement in leakage power by 60% for the 6T cell. For the 8T cell total improvement in leakage power by 61% and SNM by 13% is obtained. This is achieved by considering *only* dual-V_{Th}, thus significantly reducing manufacturing costs as well.

6T and 8T SRAM cells presented in the literature were chosen to experiment with the proposed optimization methodology. It may be noted that the improvement of the power and SNM comes from the identification of the right transistors for proper V_{th} assignment, not from sizing of the transistors. We anticipate that further sizing of the transistors along with V_{th} assignment will further improve the results. However, the proposed optimization methodology is also applicable to other variants present in the literature. Our research is in full swing in SRAM circuit optimization [17,36]. The proposed algorithm and many similar algorithms are being investigated in our research. For example, a high-κ/metal-gate based 10-transistor SRAM circuit is investigated for 32 nm technology in [36]. From the diverse experiments it is observed that the proposed algorithms are independent of SRAM circuit topology, CMOS technology node, and sizes.

6 Conclusions and Future Research

A statistical DOE-ILP approach has been presented in this paper for simultaneous P3 (power-performance-process) optimization of 6T and 8T SRAM cells simulated in 45 nm and 32 nm technology nodes. The read SNM has been treated as the performance metric. The optimization has been performed at cell level. For this, both SRAM cells of 45 nm and 32 nm have been subjected to the proposed approach which leads to 60% leakage power reduction and 13% increase in performance (read SNM). In order to achieve this objective, the novel statistical DOE-ILP approach is used for power minimization and SNM maximization. For process variation effect, 12 design and technology parameters are considered. As part of extension of this research, we plan to propose a P4 optimal methodology (where the 4th "P" is parasitics and the "T" is thermal effects) will be incorporated in this study. Further future work of this research involves array-level optimization of SRAM where mismatch and process variation will be considered as part of the design flow.

Acknowledgments

This research is supported in part by NSF awards numbers CNS-0854182 and DUE-0942629. The authors would like to acknowledge the inputs of Dhruva Ghai and Garima Thakral, graduates of the University of North Texas.

References

[1] A. Pavlov, M. Sachdev, CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies, Springer, New York, 2008.

[2] S. Lin, Y. B. Kim, F. Lombardi, A low leakage 9T SRAM cell for ultra-low power operation, in: Proceedings of the ACM Great Lakes symposium on VLSI, 2008, pp. 123–126.

[3] W. Zhao, Y. Cao, New Generation of Predictive Technology Model for sub-45 nm Design Exploration, in: Proceedings of the International Symposium on Quality Electronic Design, 2006, pp. 585–590.

[4] N. Azizi, F. Najm, A. Moshovos, Low-leakage asymmetric-cell SRAM, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 11 (4) (2003) 701–715.

[5] A. Moshovos, B. Falsafi, F. Najm, N. Azizi, A case for asymmetric-cell cache memories, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 13 (7) (2005) 877–881.

[6] N. Verma, A. P. Chandrakasan, A 256KB 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy, IEEE Journal of Solid-State Circuits 43 (1) (2008) 141–149.

[7] L. Chang, R. Montoye, Y. Nakamura, K. Batson, R. Eickemeyer, R. Dennard, W. Haensch, D. Jamsek, An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches, IEEE Journal of Solid-State Circuits 43 (4) (2008) 956–963.

[8] J. Singh, D. K. Pradhan, S. Hollis, S. P. Mohanty, J. Mathew, Single ended 6T SRAM with isolated read-port for low-power embedded systems, Design, Automation & Test in Europe Conference & Exhibition (2009) 917–922.

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

[10] J. Kulkarni, K. Kim, S. Park, K. Roy, Process variation tolerant SRAM array for ultra low voltage applications, in: Proceedings of the Design Automation Conference, 2008, pp. 108–113.

[11] P. A. Stolk, F. P. Widdershoven, D. B. M. Klaassen, Modeling Statistical Dopant Fluctuations in MOS Transistors, IEEE Transactions on Electron Devices 45 (9) (1998) 1960–1971.

[12] K. Agarwal, S. Nassif, Statistical Analysis of SRAM Cell Stability, in: Proceedings of the Design Automation Conference, 2006, pp. 57–62.

[13] Z. Liu, V. Kursun, High Read Stability and Low Leakage Cache Memory Cell, in: Proceedings of the International Symposium on Circuits and Systems, 2007, pp. 2774–2777.

[14] K. Bollapalli, R. Garg, K. Gulati, S. Khatri, Low power and high performance sram design using bank-based selective forward body bias, in: Proceedings of the 19th ACM Great Lakes symposium on VLSI, 2009, pp. 441–444.

[15] T. Azam, B. Cheng, D. Cumming, Variability Resilient Low-power 7T-SRAM Design for nano-Scaled Technologies, in: Proceedings of the International Symposium on Quality Electronic Design, 2010, pp. 9–14.

[16] J. Singh, D. S. Aswar, S. P. Mohanty, D. K. Pradhan, A 2-Port 6T SRAM Bitcell Design with Multi-Port Capabilities at Reduced Area Overhead, in: Proceedings of the International Symposium on Quality Electronic Design, 2010, pp. 131–138.

[17] G. Thakral, S. P. Mohanty, D. Ghai, D. K. Pradhan, A Combined DOE-ILP Based Power and Read Stability Optimization in Nano-CMOS SRAM, in: Proceedings of the 23rd IEEE International Conference on VLSI Design (ICVD), 2010, pp. 45–50.

[18] S. Nalam, V. Chandra, C. Pietrzyk, R. C. Aitken, B. Calhoun, Asymmetric 6T SRAM with Two-phase Write and Split Bitline Differential Sensing for Low Voltage Operation, in: Proceedings of the International Symposium on Quality Electronic Design, 2010, pp. 139–146.

[19] J. Singh, J. Mathew, D. K. Pradhan, S. P. Mohanty, A Subthreshold Single Ended I/O SRAM Cell Design for Nanometer CMOS Technologies, in: Proceedings of the International SOC Conference, 2008, pp. 243–246.

[20] J. Lee, A. Davoodi, Comparison of Dual-V$_t$ Configurations of SRAM Cell Considering Process-Induced V$_t$ Variations, in: Proceedings of the International Symposium on Circuits and Systems, 2007, pp. 3018–3021.

[21] S. Jahinuzzaman, M. Sharifkhani, M. Sachdev, Investigation of Process Impact on Soft Error Susceptibility of Nanometric SRAMs Using a Compact Critical Charge Model, in: Proceedings of the International Symposium on Quality Electronic Design., 2008, pp. 207–212.

[22] Y. Zhou, R. Kanj, K. Agrawal, Z. Li, R. Joshi, S. Nassif, W. Shi, The impact of BEOL lithography effects on the SRAM cell performance and yield, in: Proceedings of the International Symposium on Quality Electronic Design, 2009, pp. 607–612.

[23] G. Thakral, S. P. Mohanty, D. Ghai, D. K. Pradhan, P3 (Power-Performance-Process) Optimization of Nano-CMOS SRAM using Statistical DOE-ILP, in: Proceedings of the International Symposium on Quality Electronic Design, 2010, pp. 176–183.

[24] D. C. Montgomery, Design and Analysis of Experiments, 7th Edition, John WIley & Sons, Inc., Hoboken, NJ, 2009.

[25] K. Cao, W.-C. Lee, W. Liu, X. Jin, P. Su, S. Fung, J. An, B. Yu, C. Hu, BSIM4 gate leakage model including source-drain partition, in: Electron Devices Meeting, 2000. IEDM Technical Digest. International, 2000, pp. 815–818.

[26] E. Seevinck, F. J. List, J. Lohstroh, Static noise margin analysis of MOS SRAM cells, IEEE Journal of Solid-State Circuits 22 (5) (1987) 748754.

[27] J. Singh, J. Mathew, S. P. Mohanty, D. K. Pradhan, A Nano-CMOS Process Variation Induced Read Failure Tolerant SRAM Cell, in: Proceedings of the International Symposium on Circuits and Systems, 2008, pp. 3334–3337.

[28] D. Ghai, S. P. Mohanty, E. Kougianos, Variability-aware optimization of nano-CMOS Active Pixel Sensors using design and analysis of Monte Carlo experiments, in: Proceedings of the International Symposium on Quality Electronic Design, 2009, pp. 172–178.

[29] T. Mizuno, J. Okamura, A. Toriumi, Experimental Study of Threshold Voltage Fluctuation Due to Statistical Variation of Channel Dopant Number in MOSFETs, IEEE Transactions on Electron Devices 41 (11) (1994) 2216–2221.

[30] A. Wang, A. Chandrakasan, A 180 mv fft processor using sub-threshold circuit techniques, in: Proc.IEEE ISSCC Dig. Tech. Papers, 2004, pp. 229–293.

[31] L. Chang, D. Fried, J. Hergenrother, J. Sleight, R. Dennard, R. Montoye, L. Sekaric, S. McNab, A. Topol, C. Adams, K. Guarini, W. Haensch, Stable sram cell design for the 32 nm node and beyond, VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on (14-16 June 2005) 128–129.

[32] K. Takeda, Y. Hagihara, Y. Aimoto, M. Nomura, Y. Nakazawa, T. Ishii, H. Kobatake, A read-static-noise-margin-free sram cell for low-vdd and high-speed applications, IEEE Journal of Solid-State Circuits 41 (1) (2006) 113–121.

[33] L. Chang, Y. Nakamura, R. Montoye, J. Sawada, A. Martin, K. Kinoshita, F. Gebara, K. Agarwal, D. Acharyya, W. Haensch, K. Hosokawa, D. Jamsek, A 5.3GHz 8T-SRAM with Operation Down to 0.41V in 65nm CMOS, VLSI Circuits, 2007 IEEE Symposium on (2007) 252–253.

[34] B. H. Calhoun, A. P. Chandrakasan, A 256-KB 65-nm Sub-threshold SRAM Design for Ultra-Low-Voltage Operation, IEEE Journal of Solid-State Circuits 42 (3) (2007) 680–688.

[35] Z. Liu, V. Kursun, Characterization of a Novel Nine-Transistor SRAM Cell, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 16 (4) (2008) 488–492.

[36] G. Thakral, S. P. Mohanty, D. Ghai, D. K. Pradhan, A DOE-ILP Assisted Conjugate-Gradient Approach for Power and Stability Optimization in High-k/Metal-Gate SRAM, in: Proceedings of the 20th ACM/IEEE Great Lakes Symposium on VLSI (GLSVLSI), 2010, pp. 323–328.