Ultra-Low Power and Reliable SRAM for System-on-Chip

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

by

Harsh Naranbhai Patel

December 2017

APPROVAL SHEET

This Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author Signature: *Harsh N. Patel*

This Dissertation has been read and approved by the examining committee:

Advisor: Prof. Benton H. Calhoun

Committee Member: Prof. John Lach

Committee Member: Prof. Mircea R. Stan

Committee Member: Prof. Kevin Skadron

Committee Member: Prof. Scott T. Acton

Committee Member: _____

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

December 2017

Abstract

The number of ubiquitous sensors has increased to more than double the human population and is expected to continue growing in the future. The pervasive use of sensors for applications such as personal healthcare and the Internet of Things (IoT) presents a growing sustainability challenge concerning the availability and accessibility of power sources. With 50 billion devices expected to be connected to the Internet by 2020, recharging or replacing batteries at a regular interval will result in enormous time and cost overhead due to the restricted growth of batteries. Therefore, the need for the self-powered systems is being researched as an alternative solution to operate at scaled supply voltage below its threshold voltage to minimize an active power (CV_{DD}^2). The important metrics for such applications change from the traditional performance-driven to low-energy (i.e., longer battery life) and reliability. To address these challenges, it has become critical to re-evaluate the different design decisions made under high-performance requirements and to revalidate any different trade-offs among the metrics under consideration.

While analyzing the major contributors for power dissipation in present state-of-the-art low-power ICs, it has been found that Static Random Access Memory (SRAM) consumes almost 65% of the total chip power. Therefore, designing an SRAM to operate in the subthreshold region of operation provides an opportunity for reducing the overall power dissipation in such low-power ICs. However, the SRAM functionality in the sub-threshold operation becomes extremely sensitive to the process and temperature variations, and other environmental conditions such as radiation-induced soft errors. As a result of these factors, the power reduction and robustness become major challenges for the subthreshold design targeting ULP platforms.

In this work, we present various circuits and architectural techniques to enable ULP and reliable SRAMs to optimize energy per operation for a complex System-on-Chip (SoC). We investigate various subthreshold SRAM design trade-offs to achieve a sub-uW battery-

less BSN system. First, we explore different design knobs for an optimal SRAM design targeting ULP applications. These knobs include leveraging advanced fabrication technology, optimal device selection, circuit design techniques, and architectural techniques. Next, we evaluate the efficacy of different read and write peripheral assist techniques for improving reliability and energy efficiency of subthreshold SRAMs and provide stability, power, and performance trade-offs for system level optimization. The result of various optimization approaches resulted in a leakage and energy optimized SRAM operating over with leakage power reduced to 1.5 pW/bit while consuming only 6.24 pJ/access energy that is 40% more efficient than the previous version. To address a wide range of IoT applications with energy optimization, we propose a Canary sensor based minimum supply voltage (V_{MIN}) tracking with optimal selection of the peripheral assist techniques. Later, we address the reliability issue at subthreshold operation using a process, voltage, and temperature (PVT) variation mitigation controller. Finally, the impact of the radiation-induced soft error is evaluated to provide a comprehensive study on reliability for the ULP SRAM at scaled V_{DD}.

Dedicating what all I learn in my life to

'Gurus' in my life — for enlightening my life with the knowledge,
My Parents — for always giving me the courage to achieve impossible,
My Wife - Pooja — for her unwavering love and motivation

ॐ असतो मा सद् गमय।
तमसो मा ज्योतिर्गमय।
मृत्योर्माऽमृतम् गमय ॥

Meaning:
Lead me from the unreal to the real;
from darkness(ignorance) to light(knowledge);
and from death to immortality.

Acknowledgments

कायेन वाचा मनसेंद्रियैर्वा ।
बुद्ध्यात्मना वा प्रकृतिस्वभावात् ।
करोमि यद्यत् सकलं परस्मै ।
नारायणायेति समर्पयामि ॥ ॥

Meaning:

Whatever I perform with my body, speech, mind, limbs, intellect, or my inner self either intentionally or unintentionally, I dedicate it all to that Supreme Lord Vishnu...

I can still remember the day when I received an email from UVA about my acceptance for Ph.D. I cannot thank god ever enough in my life to give this opportunity. I would like to acknowledge many people throughout my journey of seeking for the knowledge without whom I would not be able to achieve what I have done till date. I hope I will able to remember everyone those have helped me along the way of this journey of learning. I will always remember the days I spent at UVA.

First and foremost, I would like to thank my Guru/adviser - Professor Ben Calhoun. This journey is only possible due to his ability to see me achieving what I am till date. I clearly remember my first 'most sloppy ever' plot of results that I shared with him. Since then, his advice made me like a polished diamond from a raw carbon over the period. His highly influential teaching skill made a significant impression on my mind about 'how to teach.' I am thankful to him for learning not just technical skills but many more non-technical skills those help me (and will be helping me in future!) throughout my course of Ph.D. Thank you, Ben, for everything you did to support and your incredible guidance. I hope my gratitude to you can be seen in my commitment to becoming the best version of myself,and in my determination to strive for perfection.

I am also grateful to have Professor John Lach, Professor Kevin Skadron, Professor Mircea Stan, and Professor Scott Acton as my committee members. I thank all committee members for their continuous support and providing insightful feedback to my research - Thank you all!

A major part of my learning came from working with awesome friends and colleagues at Bengroup. I consider myself very lucky to work in Bengroup with amazing students. I like to thank Jim Boley for his patience during my early days at UVA with lots of questions. I also thank Bengroup senior students: Alicia Klinefelter, Yousef Shakhsheer, Aatmesh Shrivastava, and Kyle Craig, for their support. I am very privileged to work in close collaboration with Jim Boley, Farah Yahya, Arijit Banerjee, and Ningxi Liu as the SRAM team. Especially, I appreciate and thank Farah Yahya, with whom I worked a lot throughout my Ph.D., for her support and readiness to exchange ideas. I am glad to have an opportunity to work with Abhishek Roy and Divya Akella for the continuous exchange of ideas we had throughout the work. As a part of Rice Hall Lab 322, I enjoyed the company of Seyi Ayorinde, Chris Lukas, He Qi, Ningxi Lui, Kevin Leach, Henry Bishop, and Daniel Truesdell. I would also like to thank Kevin Leach for providing his technical expertise on my research work and technical papers. I would like to give my special thanks to Terry Tigner for her 'Dove' love. Her tremendous support made us possible to test our chips timely and a hassle-free conference journey!

I want to thank all the sponsors who have supported my research. Special thanks to Tom Gray - my manager, and Brian Zimmer - my mentor, at NVIDIA for giving me an opportunity to be a part of a cutting edge research work.

Besides my life at the lab doing the research work, I immensely enjoyed my stay in Charlottesville. I am very much grateful to my neighbors for their pleasant company during my stay. I am also grateful to the members of Swadhyay Parivar for their selfless love.

Last but never the least, I would like to thank my family members. I am very grateful for my parents, Naranbhai and Induben, for their continuous support and motivation for

pursuing higher education. Especially, I thank my mother for always keeping me in her prayer while my father for giving me the courage to achieve whatever seems impossible to me. I thank my sister - Rinku - for her sisterly love. I cannot encompass my gratitude to my wife, Pooja, with the words for her support. She always been a lovely wife with continuous support. I am also indebted to almighty God for his blessing of two beautiful children, Samyak and Swara — my best stress busters! I cannot image myself completing my Ph.D. without the help of all these three - my wife and my kids — Thank you so much!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

With the growing need for electronic devices in health care, remote sensing, entertainment, security, communication, and many other fields, power management has become a major concern. Advancements in battery technology are unable to deliver required kW/cm^2 for the present state-of-the-art System-on-Chip (SoC) with a smaller form factor [7]. Therefore, designing low-power systems poses a critical requirement for the applications operating under battery-less or self-harvesting conditions. Additionally, with an increase in life-expectancy, and a desire to bring the ease of quality healthcare to every person, research in the area of Body sensor Nodes (BSNs) is increasing to provide devices for unobtrusive and precise monitoring [8]. In addition to healthcare, remote sensing can be used in a variety of environmental modalities such as temperature change of a remote location, water level degradation, humidity, and geo-position based update [8]. With the continuously changing electronic device requirements, the design specification for different applications varies widely. For example, the Internet of Things (IoT) and BSN applications prioritize energy efficiency and functional reliability over performance, while for general purpose and graphics (GPU) processors, servers, and other high-end applications, energy is sacrificed for higher performance. Typically, IoT devices

operate in the range of few KHz to a few MHz depending on the application, but energy is an important concern in all of these applications [9]. Addressing energy consumption in devices is critical to continuing scale with increased demand for more complex capabilities and performance, whether for low-power IoT devices or high-performance GPU applications.

The authors in [1] and [2] present state-of-the-art of BSN application with ECG as one of the health modalities monitoring with Atrial fibrillation (AFib) detection. The power dissipation among various components in such Ultra-Low Power(ULP) applications shows that the Static Random Access Memory (SRAM) is a major contributor to the power dissipation in the ULP Systems-on-Chip (SoCs) (Figure 1.1). Also, the demand for a larger embedded memory (mostly SRAMs) in such highly integrated SoCs has increased drastically to support a wide range of capabilities [10]. The higher density requirement further tightens the design constraints on power, performance, and reliability. In this research, we address several low-power challenges in SRAM — ranging from the fabrication technology to the system integration. Energy consumption consists of two main components: active energy and leakage energy. The active energy (CV_{DD}^2) dominates at higher voltages, while the leakage energy ($V_{DD} \times I_{Leak}$) dominates at subthreshold voltages. Thus, an optimal V_{DD} exists that minimizes the total energy of design, and usually that V_{DD} lies in the subthreshold region [11]. Since this optimum point changes for different performance needs and designs, it is essential to explore the design space while varying different knobs to determine the optimal combinations of knobs that minimize the overall energy consumption of the design for each application.
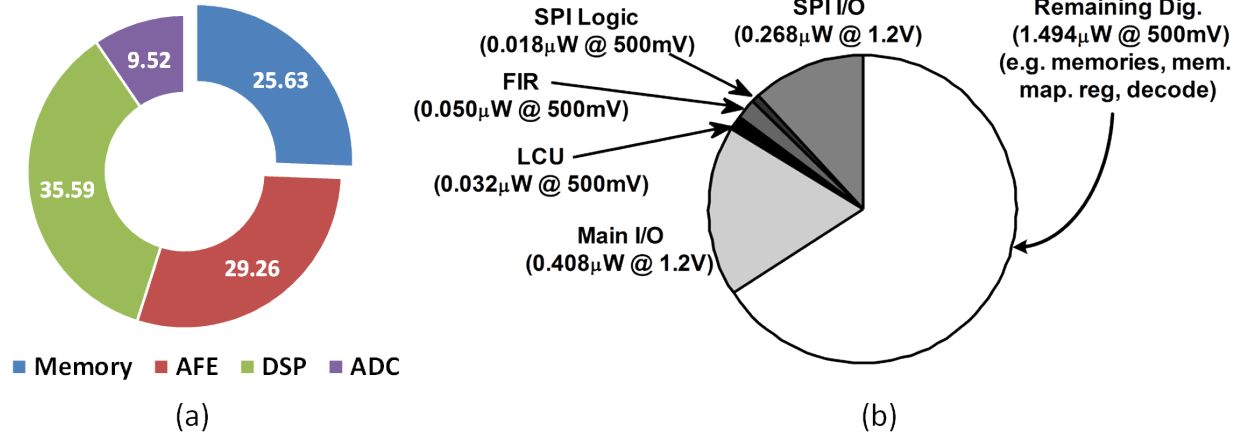
Figure 1.1: Power consumption of present state of the art ULP SoCs (a) [1] (b) [2]

At scaled V_{DD}, ON current (I_{ON}) of the device is reduced, significantly limiting stability and performance. However, advancements in new fabrication technologies and improved device structure allow higher I_{ON} to be achieved at the same V_{DD} while reducing the leakage current (I_{OFF}). Therefore, advancements in the fabrication process enables power and performance benefits. As the SRAM bitcell is a ratio-ed circuit, the logic states are decided based on the relative strengths of its devices. Therefore, the choice of device V_T can be used as one of the design metrics to achieve an application-specific bitcell for required power-performance demand. To reduce the active power, various circuits and architectural level solutions are implemented in the literature. In an effort to reduce the leakage current during the steady state, we use higher threshold voltage (V_T) devices in the SRAM bitcell. High-V_T devices offer leakage reduction but also reduces the performance. Low-power applications like BSNs requires lower performance, but leakage reduction—implementing an SRAM with high-V_T devices improves leakage power significantly.

The active power reduction requires operating devices at lower V_{DD}. As a result, the reduced I_{ON}/I_{OFF} at scaled V_{DD} degrades write and read operations in SRAM bitcells. Therefore, we implement peripheral assist techniques to facilitate the write and read operations. However, it becomes vital to evaluate different assist techniques for subthreshold design based on a metric of interest targeting low-power requirements, as for the ULP application such as

BSN, the metric of interest changes from stability and energy in subthreshold compared to the performance in super-threshold operation. Therefore, finding an optimal assist technique for the subthreshold operations relies on the trade-off between minimum operating supply voltage V_{MIN}, power dissipation, stability, and performance.

$$I_D = I_0 \left(\frac{W}{L}\right) \exp\left(\frac{V_{\text{GS}} - V_{\text{T}} - \eta V_{\text{DS}}}{n\frac{kT}{q}}\right) \left(1 - \exp\left(\frac{-V_{\text{DS}}}{\frac{kT}{q}}\right)\right) \tag{1.1}$$

In the subthreshold region, I_{ON} follows an exponential relation (Equation (1.1)) with the device V_T and temperature (T). Therefore, minor variation in device V_T and temperature drastically impact the I_{ON} and I_{OFF} of the SRAM functionality. The sources of V_T variation can be the process (intra-die and inter-die) or temperature fluctuation. To compensate the effect of such external parameters, the SRAM has to be designed with a larger guard band (margins): the higher the variation, the larger the guard band requirement. Increasing the guard bands results in an inefficient SRAM design with power, performance, and area (PPA) degradation. One of the approaches to reduce the guard band is to design a self-calibrated SRAM that can adjust to process and temperature changes. Reducing the guard band requirement drastically improves the power efficiency. In addition to process and temperature variations, the supply scaling makes SRAM vulnerable to radiation-induced soft errors due to alpha particles. The critical charge (Q_{crit}) reduces linearly with supply voltage, but the change of nonlinear drain capacitance from super-threshold to subthreshold voltages increases the rate of Single Event Upset (SER).

1.2 Thesis Statement

The requirement for increasingly ubiquitous sensors around human life demands self-harvesting Ultra-Low Power platforms such as BSNs. The system-level goal of power reduction can be achieved by optimizing the SRAM power. With supply scaling to subthreshold voltage, the active energy reduces quadratically while leakage energy dominates contemporary power

consumption. Additionally, the reduced drive strength and exponential dependency of V_T on the process, voltage, and temperature (PVT) variation in devices can challenge the reliability of SRAM. The contribution of this research is to enable low-power applications with a reliable and energy-efficient SRAM design with a detailed exploration of reliability - energy - performance trade-offs for a constraint-driven low-power platform.

1.3   Dissertation Organization

1.3.1   Background

Chapter 2 provides the basic structure and behavioral representation of a conventional SRAM and read, write, and hold functional requirements. This section covers various peripheral assist techniques those are used later in this research to enable a reliable SRAM operations in the subthreshold region. Various subthreshold SRAM design challenges are briefed along with design knobs to reduce system power from the technology to architecture level are discussed.

1.3.2   Subthreshold SRAM for ULP Body Sensor Node

Chapter 3 provides a thorough analysis and design consideration to achieve a sub-μW ULP BSN SoC. To optimize the leakage-dominated subthreshold SRAM, the chapter considers four approaches: 1) subthreshold optimized fabrication technology to reduce the leakage and variation, 2) choice of optimal SRAM bitcell for leakage reduction, 3) optimal assist technique for the subthreshold operation considering power-performance-reliable trade-offs, and 4) architectural techniques. An Ultra-Low Leakage (ULL) 55nm Deeply Depleted Channel (DDC) process technology devices are used to achieve 67% reduction in threshold (V_T) variation due to Random Dopant Fluctuation (RDF). While circuit techniques such as subthreshold operation and reverse body biasing (RBB) are co-designed with the technology to maximize the energy/power saving that resulted in a 6T SRAM array operating reliably

down to 200mV with a reduced leakage power of 7nW/kb. Later, a test-chip with six different 8T SRAM bitcells provides comparison across different design space requirements — such as reliability and low power/energy — for IoT applications. Once selecting the optimal bitcell, we evaluate the impact of different peripheral write assist techniques on the reliability and energy efficiency of SRAMs. Finally, a complete 2KB SRAM macro fabricated using 130nm provides measurements showing enabling sub-μW ULP BSN.

1.3.3 SRAM V_{MIN} Tracking using Canary and Optimal Assist Selection

The IoT applications cover a wide range of operation while still considering energy/operation as a critical metric. Chapter 4 proposes a canary sensor-based failure detection and dynamic assist selection based architecture. The proposed adaptive, closed loop memory system that leverages combinations of bias-based peripheral assists for both read and write to expand the operating range of a 256kb 6T SRAM to cover from 1.2V down to 0.38V. Assists are used in reverse to tune canary bitcells that allow a closed loop control of the V_{DD} to track the minimum operating voltage (V_{MIN}) at a desired operating frequency. A test-chip in 130nm is demonstrated with an optimized SRAM macro achieving the low-power requirement across a wide range of the operating frequencies using Dynamic Voltage Scaling (DVS).

1.3.4 Reliability at Ultra-Low Voltage Operation

Chapter 5 addresses reliability challenges at ultra-low V_{DD} operation. With V_{DD} scaled to the subthreshold, the I_{ON} experiences exponential dependency on V_T (Equation (1.1)). Therefore, a small change in V_T due to process, voltage, and temperature (PVT) variations may result in functional failures. A testchip in 130nm demonstrates a digital controller to improve design reliability and energy in PVT variations. Additionally, an aggressive technology and supply voltage scaling has led to increasing concern for reliability. Optimizing power and

energy with sub-V_T operation increases the occurrences of both, static and dynamic, failures exponentially. We explore the impact of radiation-induced soft-errors on the ULP application operating in the subthreshold. We also demonstrate an exponential reduction in the critical charge (Q_{Crit}) of a storage node with supply in near- and sub-V_T design, resulting in a major design consideration for the low-power applications.

1.3.5 Conclusions

Chapter 6 provides a summary of this research work, highlights the contributions and the broader impact, and open research question for the future work.

Chapter 2

Background

In state of the art SoCs, a significant amount of area is occupied by the memory element to store data. One of the ways to store data is in back-to-back connected CMOS inverters creating a bi-stable storage element. The Static Random Access Memory (SRAM) is the volatile type of the data storage because of the required power supply (V_{DD}) to be operational. Figure 2.1 shows two types of conventional bitcells used for the SRAM. The storage nodes (Q/QB) can be accessed for read or write using NMOS pass gates (PGs). Figure 2.1a shows a 6T (six transistors) bitcell. Figure 2.1b represents another design of an SRAM bitcell where the read operation is separated from the write access path. Similar to an 8T bitcell, various bitcell structures (e.g., 7T [12], 9T [13], 10T [14], 14T [15]) have been proposed to optimize different metric of interest for the targeted application. In this dissertation, we constrain the bitcell exploration to 6T and 8T because increasing device size exponentially increases overall SRAM area in an SoC. Besides the SRAM bitcell, an SRAM macro implements peripheral circuitry that includes a row and a column decoder, a sense amplifier (SA), read bitline pre-charge, timing and control signal generation, and assist techniques.
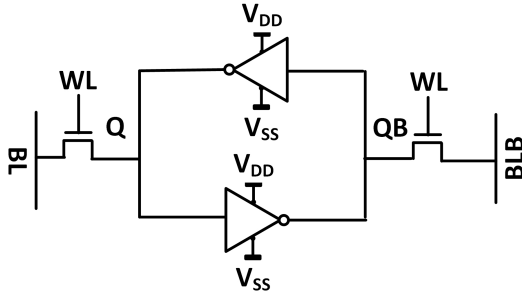
(a) Conventional 6T SRAM cell with inverters connected back-to-back
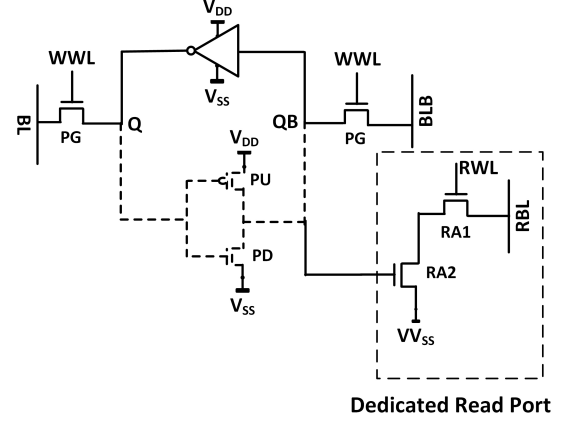
(b) 8T Bitcell with separated Read Port

Figure 2.1: Conventional SRAM Bitcells

2.1 Basic Operations

2.1.1 Write Operations

The basic functionality of the SRAM includes writing, reading, and holding data. Figure 2.2 shows the write and the read operations. During the write operation (Figure 2.2a), the corresponding bit values are set on bitline (BL) and bitline bar (BLB) to be written to the bitcell. Once the BL/BLB bits are stable, PGs are turned ON by applying V_{DD} as a wordline (WL) pulse. Once the PGs are ON, the node storing logical value '1' discharges through the PG and BL/BLB holding the '0'. To ensure a successful write operation, the PG is designed "stronger" compared to the PMOS Pull-up (PU) device and hence allowing discharging current from PG to "win" charging current from the PU. Therefore, a careful design of PU and PG devices is required to ensure the write operation. The ratio of the size of PMOS PU device to the size of NMOS PG device is defined as *pull-up ratio* (PR).

2.1.2 Read Operations

During the read (Figure 2.2b), BL and BLB are precharged to V_{DD} and then left unaccessed. Once the control circuitry decodes the row and column address, a differential voltage between BL and BLB are generated due to the discharge of BL/BLB to the corresponding node holding "0." The SA detects the differential voltage and generates the output according to the data read from the bitcell. Similar to the *pull-up ratio* in the write operation, a successful read operation should allow the BL/BLB discharge through PGs while not letting storage node flipping. Therefore, the resistance of the PGs must be larger than NMOS pull-down (PD) devices. The size of PD device to PG device is defined as *Cell Ratio* (CR).

(a) Write Operation   (b) Read Operation
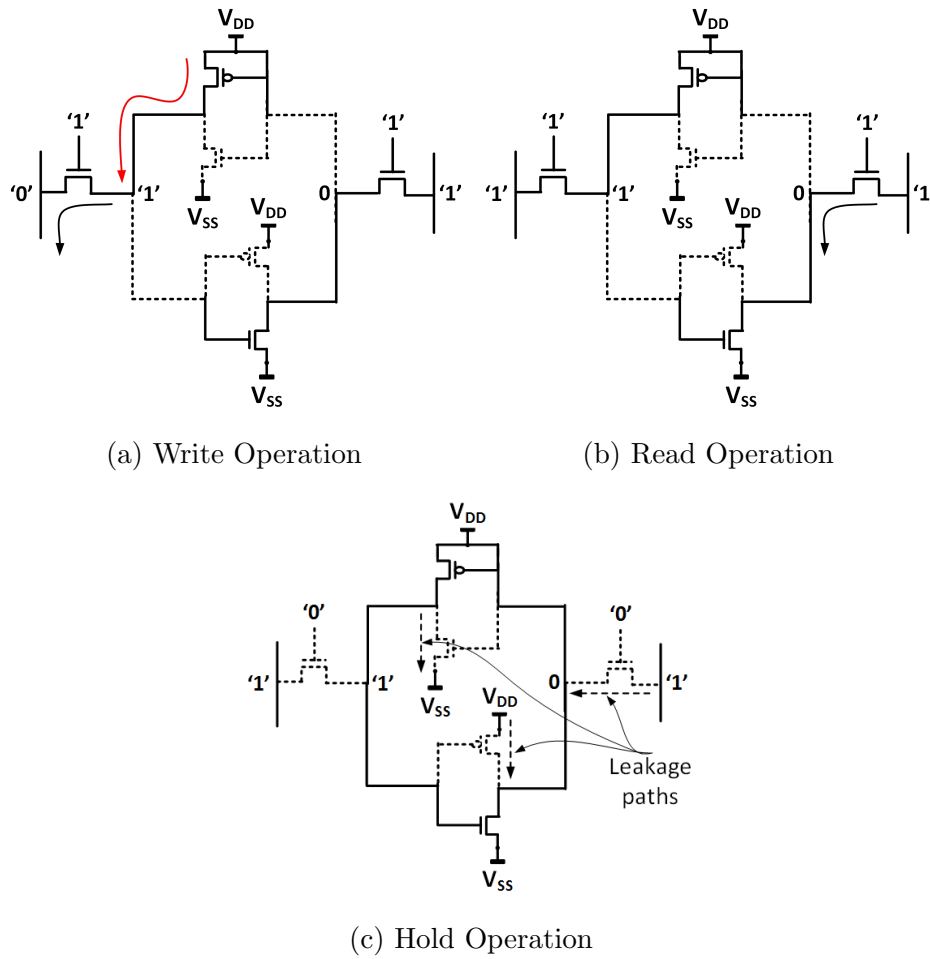
(c) Hold Operation

Figure 2.2: SRAM Operations

2.1.3 Hold Operations

When not being accessed, the SRAM preserves the data until the supply voltage (V_{DD}) is available (Figure 2.2c). During the hold state, the PGs are turned OFF by applying WL=0, and hence the back-to-back connected inverters hold the data. Ideally, an SRAM should be able to hold the data until the power supply is disabled. However, at lower operating V_{DD}s, leakage from the OFF devices (PGs, PU, and PD) can flip the value as shown in Figure 2.2c.

2.2 Assist Techniques: An Overview

The huge drive towards the Internet of Things (IoT) has led to the development of ULP SoCs that are capable of operating on harvested energy [2] [16]. The circuits within such SoCs (e.g., SRAM) must operate reliably under varying process, voltage, and temperature (PVT) conditions, and thus their energy and power consumption must be kept to a minimum. One way to guarantee low power and energy consumption is to scale down the supply voltage to the subthreshold region [17]. However, the reduced on-to-off current ratio (I_{ON}/I_{OFF}) and the exponential dependence of current on the V_T in the subthreshold region introduces many challenges, especially in ratioed circuits such as a conventional 6T SRAM. Additionally, the increased impact of PVT variation in subthreshold causes write and half select (HS) failures[1] in SRAM array. Many approaches were introduced in the literature to address the different challenges facing SRAMs. Some alternative bit-cell topologies have been used in the literature [13, 15, 18], including the 8T bit-cell [18] (shown in Figure 2.1b) with decouple read and write ports to improve the stability for subthreshold operation.

To improve read and write stability at lower supply voltages, different peripheral assist techniques [19, 20] were also used. Figure 2.3 shows the graphical representation of four different peripheral assist techniques: a) V_{DD} Lowering, b) V_{SS} Raising, c) WL Boosting,

---

[1]Half select failures result from an unaccessed bitcell undergoing a dummy read during the write operation in a bitcell from the same row.

(a) V_{DD} Lowering

(b) V_{SS} Raising

(c) Wordline (WL) Boosting

(d) Negative bitline (NegBL)
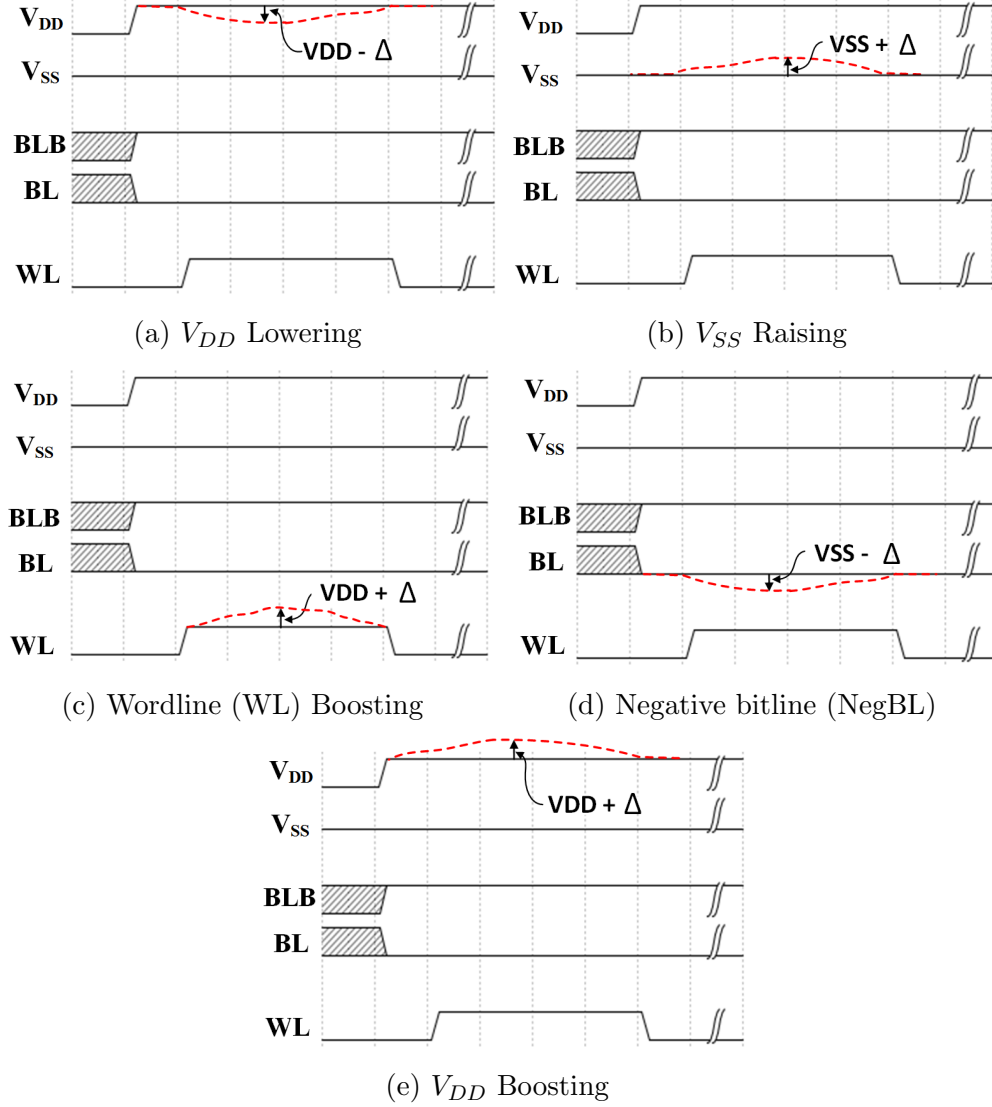
(e) V_{DD} Boosting

Figure 2.3: Write Assist Techniques: a) V_{DD} Lowering, b) V_{SS} Raising, c) Wordline (WL) Boosting and d) Negative bitline (NegBL); Read Assist Technique: e) V_{DD} Boosting

d) Negative Bitlince (NegBL), and e) V_{DD} Boosting. With V_{DD} lowering assist [21], the column/core V_{DD} is reduced to $(V_{DD} - \Delta)$ reducing the $|V_{GS}|$ and thus the strength of the pull-up PMOS (PU) (Figure 2.1b). In the V_{SS} raising assist, the row/core V_{SS} is increased from 0 to Δ, which weakens the PMOS device by reducing its $|V_{GS}|$ due to an increase in gate voltage. Similar to the PMOS assist techniques, WL boosting [21] strengthens the pass-transistor (PGs) by increasing the WL voltage from V_{DD} to $(V_{DD} + \Delta)$. The increase in V_{GS} and hence I_{ON} aids the cell to flip its value. In NegBL, V_{GS} of the pass transistor

writing a '0' into the cell is increased by under-driving the voltage of the bitline (BL) holding '0'. In [22] and [23], the authors evaluated assist techniques for a 6T cell using different technology nodes and metrics for the super-threshold operations.

Similar to the write failure, the reduced I_{ON}/I_{OFF} deteriorate the read operation. There are mainly three types of read failures: 1) Read Disturb failures, 2) Half-Select(HS) failures, and 3) Read Failures. As described in the subsection 2.1.2, careful sizing of PG and PD devices (Figure 2.1) is required to prevent any read-disturb failures [2]. This occurs when a larger potential is developed across the PD transistor that increases the Q node potential above the trip point of the right inverter. To increase read stability, the PD transistor is made stronger than the pass-gate. This ensures that the voltage drop across PD is not sufficient to turn on the another inverter. During the write operation, the un-accessed bitcells undergo a dummy read operation where PG devices turn on due to WL. This results in a Half-Select (HS) failures. During the dummy read, the leakage current from BL/BLB to bitcell result into a wrong Q/QB increase the Q/QB node potential and hence flip the bitcell. Finally, during the read operation, a wrong bit value is read as a result of the read failures [3]. However, in a 6T bitcell, it is challenging to differentiate between HS and read failures in lack of further observability. Similar to the write assist techniques, different PMOS and NMOS-biasing based techniques are used. Figure 2.3e) shows V_{DD} Boosting as one of the read assist techniques.

2.3   Subthreshold SRAM Design Challenges

To allow quadratic active power[4] saving, the supply is scaled to the subthreshold region. However, as mentioned in Equation (1.1), subthreshold I_D is inversely exponentially related to

---

[2]failures that can flip the content of the bitcell either during the read operation

[3]when the sufficient potential difference between BL and BLB is not able to detected by a Sense-Amplifier (SA)

[4]Power dissipated during the switching of the device from ON-to-OFF or vice verse. The active power is defined as $C_L*V_{DD}^2*f$; where f is switching frequency

gate-to-source voltage. The subthreshold SRAM design faces five main challenges compared to the super-threshold design:

1. Reduced I_{ON}/I_{OFF},

2. Impact of process variation,

3. Reduced static margins,

4. Soft-error, and

5. Leakage dominated power.

Subthreshold SRAM faces additional challenges compared to super-threshold SRAM. In [24], the authors showed the significant reduction in I_{ON}-to-I_{OFF} ratio and higher variation across process corners that led to stability and performance degradation of an SRAM. Figure 2.4 shows 1000X I_{ON}/I_{OFF} reduction across different process corners. For a ratio-ed design like SRAM, the functionality of the circuit relies heavily on the relative strength of the devices. Therefore, the I_{ON} variation significantly increases write and read failures at lower V_{DD}. At the same time, an increase in leakage current (I_{OFF}) impacts the power and reliability of an SRAM macro. Additionally, due to the exponential dependence on the threshold voltage, a small variation in V_T results in a huge current change in subthreshold. Figure 2.4 highlights the 100X higher I_{ON}/I_{OFF} variation across various process corners compared to the nominal voltage.

Figure 2.5 maps the challenge of process variation showing the SRAM V_{MIN}. The figure shows the different static (write margin — WM, Read Static Noise Margin — RSNM, and Hold Static Noise Margin — HSNM), and dynamic (write delay — WD and Half-Select — HS) metrics and how they change with process corners. For a typical corner — TT (Typical NMOS; Typical PMOS), write delay (WD) limits V_{MIN} (=0.65V). On the other hand, half-select (HS) failures — measured by the RSNM (static metric) or HS (dynamic metric) — limit
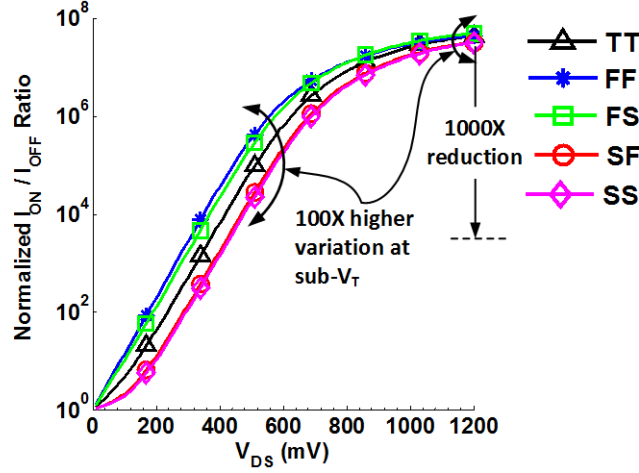
Figure 2.4: Reduced I_{ON}/I_{OFF} ratio and higher process variation at Sub-V_T
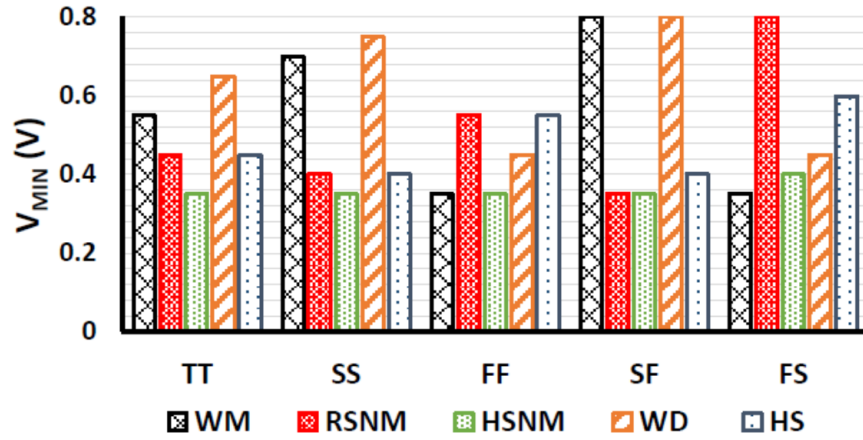
Figure 2.5: Impact of process variation on SRAM V_{MIN} scaling

V_{MIN} for FS (Fast NMOS; Slow PMOS) and FF (Fast NMOS; Fast PMOS) corners. While HS is seldom a concern for nominal operation, it becomes a critical concern for subthreshold designs operating closer to V_{MIN}.

The exponential reduction in I_{ON} current also impact the noise margin — reliability of the design. Figure 2.6 shows the worst case SRAM static noise margin (SNM) under the intra-die variation (10,000 point Monte Carlo) — Write SNM, Hold SNM (HSNM), and Read HSNM (RSNM) — are plotted across V_{DD}s. The figure shows how the read, write and hold noise margin reduces with V_{DD}. Reduced noise margin indicates poor reliability. As shown in the figure, the worst case write SNM limits the functionality, hence the SRAM V_{MIN}, at
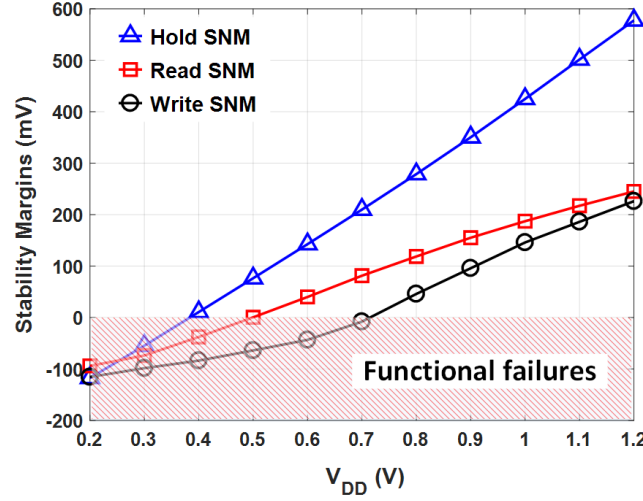
Figure 2.6: Reduced stability margins at scaled V_{DD}

0.7V while the read SNM limits the V_{MIN} at 0.5V as a result of half-select. Most of the time, the reliability is improved at the cost of area or power.

In addition to technology scaling, V_{DD} has been scaled down significantly to minimize the active energy (CV_{DD}^2) [17] for ULP IoTs applications. With excessive V_{DD} scaling to the subthreshold voltage range, the circuit node charge responsible for holding the state reduces resulting in a smaller particle strike causing the flip of the logic state. Likewise, the frequency of soft-errors also increases significantly with altitude. This exaggerates the scope and the potential risk of device failure modes due to soft-errors based on the location and usage. Authors in [25] and [26] presented a comprehensive analysis of terrestrial cosmic radiation as a function of the altitude and places. With growing demand of bio-sensing and other Body Sensor Network (BSN) applications; those require an extremely low power, the reliability issue due to soft-error in subthreshold become a very critical problem to be addressed.

With the subthreshold operation, the supply voltage scaling allows quadratic active energy savings. However, lowering V_{DD} also degrades the switching speed. Consequently, the integration time of the leakage-currents also increases, raising the leakage-energy [3]. Also, the "partially ON" devices increase the leakage power over the time. Figure 2.7 shows a relative increase in leakage energy with V_{DD} scaling. For an SRAM, the unaccessed bitcells

always leak and contribute a larger in the total power in the SoC.
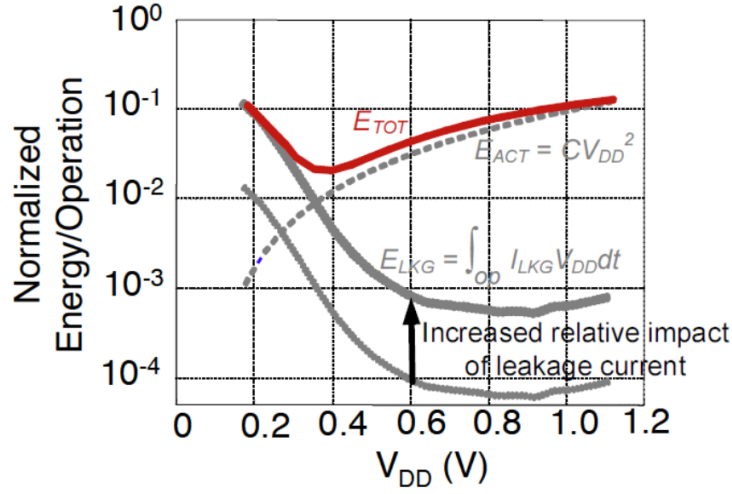
Figure 2.7: Relative increase in leakage power dominance in sub-theshold: An SRAM contribution [3]
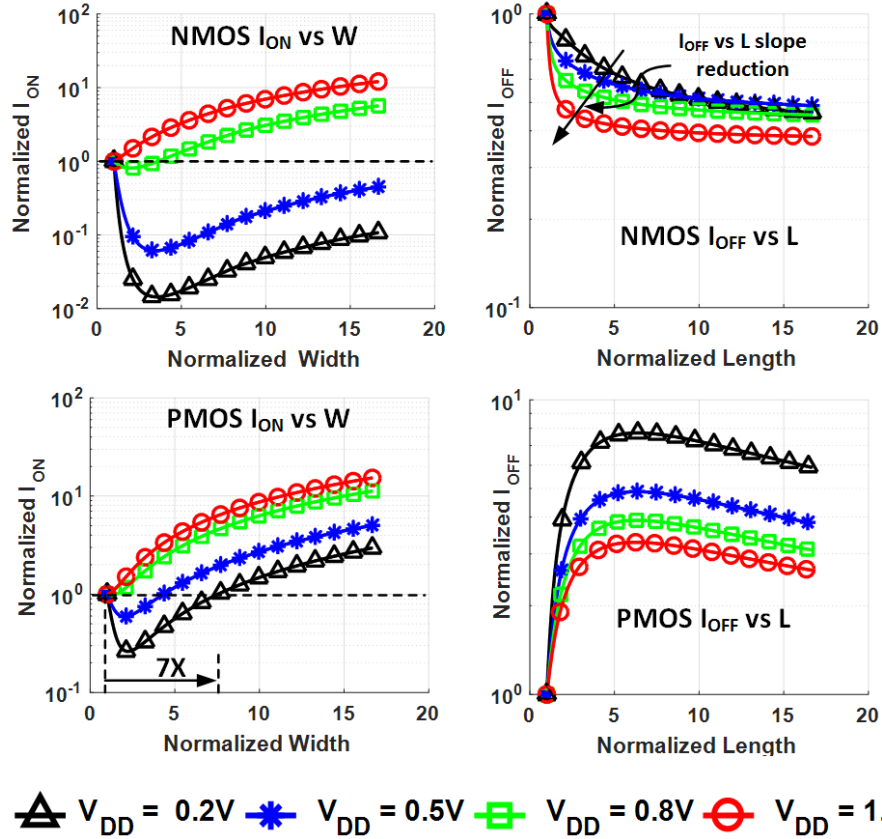
Figure 2.8: Device sizing becomes a less effective knob for increasing I_{ON} in subthreshold (130nm technology)

Additionally, the design knobs such as device sizing used at nominal V_{DD} to optimize the energy and leakage may not be suitable in the subthreshold region. At nominal V_{DD}, the device sizing is controlled to ensure the functionality. However, at lower V_{DD}, I_{ON} does not increase linearly with the width of the transistors due to Inverse Narrow Width Effect (INWE) [27] as shown in Figure 2.8 for the selected technology. Also, increasing in device size also increase the device capacitance that results in higher energy dissipation in the energy-constrained ULP applications at subthreshold. Therefore, at subthreshold voltages, sizing is a weak knob to control I_{ON}. The figure highlights that at $V_{DD} = 1.2$V I_{ON} can be increased linearly with sizing (increasing W) for the selected technology. In contrary, a similar increase in I_{ON} at $V_{DD} = 0.2$V requires more than 7X the minimum size for PMOS while more than 20X for NMOS. Similarly, using a larger channel length (L) for high-V_T NMOS devices results in a reduction in I_{OFF}, whereas, high-V_T PMOS devices experience higher I_{OFF} with small increases in L and will require significantly large L to reduce I_{OFF}. The I_{OFF} vs. L slope reduces with V_{DD}, resulting in a larger area trade-off for a fixed I_{OFF} reduction. Therefore, the subthreshold SRAM design requires new design knobs to reduce the leakage.

Chapter 3

Energy Efficient Subthreshold SRAM Design for an ULP BSN

3.1   Introduction

Body sensor nodes (BSN) promise to provide significant benefits to the health care domain by enabling continuous monitoring, actuation, and logging of patient bio-signal data, which can help medical personnel to diagnose, prevent, and respond to various illnesses such as diabetes, asthma, and heart attacks [28]. The basic functionality of the node is to sense a physical signal (such as temperature, heart rate, pressure, etc.), convert that signal into digital data, process the data on-chip, and transmit the results back to the user. Though they show great potential to influence human life by improving healthcare domain, BSNs present many design and engineering challenges that impede their widespread adoption including node operating lifetime, the small form factor for wearability, and affordable cost. In many applications, limited battery lifetimes severely undermine the deployment of body sensor nodes since the required node operating lifetime is effectively indefinite. While larger battery sizes extend the operating time, the resulting inconvenient form factor prevents its practical use. In addition, batteries require frequent charging or replacement, limiting the application space to which

they apply. To eliminate battery charging and changing bottleneck, we study a BSN system to operate solely from energy harvesting instead of using a battery. Any BSN can reliably operate on a harvested energy source if it consumes less energy than the amount harvested. To optimize power consumption, we study the distribution of power dissipation among different circuit components (Figure 1.1). While the high power dissipating components such as the transmitter are optimized by heavily duty-cycling, the leakage minimization in an SRAM poses a larger challenge for a longer lifetime of a BSN.
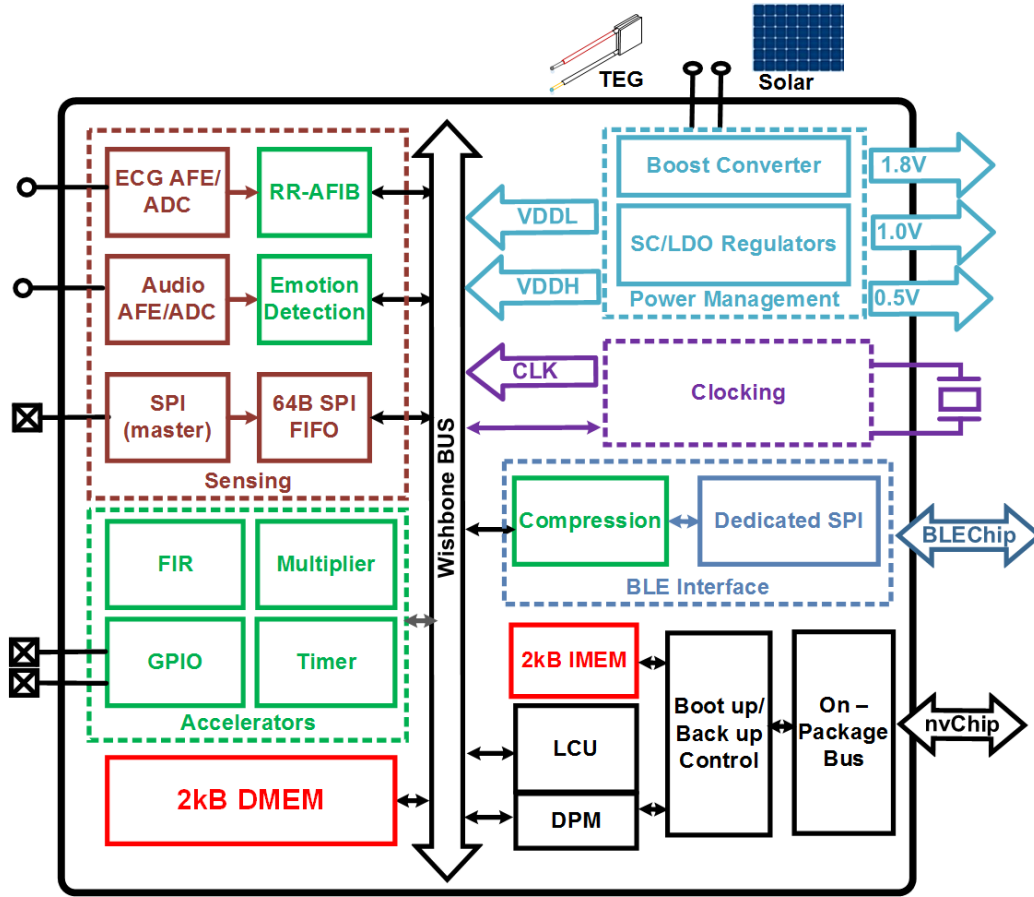
Figure 3.1: Block diagram of SoC showing harvesting system, memories, sensing modalities, accelerators, cold-boot management, and radio interface [4].

Figure 3.1 shows a block diagram of the SoC with main building blocks, and interfaces to communicate across various blocks. The present BSN SoC operates on harvested energy from Thermo-electric generator (TEG) or solar. The required amount of memory (Data Memory — DMEM or Instruction Memory — IMEM in Figure 3.1) for a biomedical system is highly

dependent on the target application. A flexible platform with various types of biosignal data acquisition and processing demand higher frequency requirement in both IMEM and DMEM. However, increasing memory capacity impacts the power requirement of the application. In the presented BSN architecture, the system is optimized for Atrial Fibrillation (AFib)[1] detection and communication over the radios with 2KB of IMEM and DMEM to limit the leakage contribution to the power demand. Depending on the set of applications, the SoC might need to cater to programs with high compression ratios and low storage requirements while at other times accommodating high throughput applications needing large amounts of storage. This creates a design challenge for flexible and ULP BSN design.

To achieve reliable operation at lower V_{DD}, we propose using an 8-transistor (8T) bitcell (Figure 2.1b) instead of a conventional 6T bitcell. Similarly, we implement transistors with a high threshold voltage (high-V_T) in SRAM macro. This is a conventional way of reducing leakage current during standby. In the rest of this chapter, we discuss various approaches to address the subthreshold SRAM challenges as discussed in Chapter 2.3, optimizing memory design from a device selection to the system-level perspective.

3.2   Research Approach

An optimization of leakage power and energy/operation of an SRAM can be achieved at different layers of abstraction, from the technological to the architectural level. Use of advanced technology can offer a better figure of merits with stronger design parameters (e.g., reduced variation, higher I_{ON}, etc.). By leveraging technology benefits, we can maximize the circuit-level benefits. The effectiveness of a system with SRAM can be further optimized by selecting an optimal device in a bitcell. For power-constrained or self-harvesting applications, such as BSNs, an SRAM with leakage minimization can drastically improve battery life and system performance.

---

[1]Atrial fibrillation (also called AFib or AF) is a quivering or irregular heartbeat (arrhythmia) that can lead to blood clots, stroke, heart failure and other heart-related complications. [29]

To achieve the goal of low-power SRAM, we divide our approach into three categories.

1. Enabling a low-power SRAM design using advanced fabrication technology:

   We explore the possibility of combining low-power technology and circuit techniques for energy efficient IoTs using 55nm Ultra-Low-Leakage (ULL) Deeply Depleted Channel (DDC) technology. The subthreshold SRAM design utilizes the benefits of technology such as reduced V_T variation, increased I_{ON}-to-I_{OFF} ratio, and higher stability, while the low-power technique such as Reverse Body Biasing (RBB) further reduces the leakage of the device. This work is aim to demonstrate the first implementation of a complete SRAM macro in DDC technology with optimized architecture.

2. Optimizing SRAM bitcell energy and reliability:

   After considering a technology as one of the low-power design knobs, we compare stability metrics and energy consumption of different 8T bitcells for given design constraints. The optimal choice of transistor threshold voltage within a bitcell varies significantly based on the selection of the metric of interest. We compare the selection of different device selection under various sources of variation (e.g., intra-die, inter-die, and temperature) in subthreshold SRAM and their impact on various SRAM metrics. In this work, we explore new design space consideration for an energy-constrained application targeting battery-less applications.

   For energy-constrained applications such as [2], operating in subthreshold presents major concerns for guaranteeing functionality. As introduced in the previous chapter, the peripheral write and read assists are implemented. To quantify the impact of each assist technique on the reliability of the write operation, we use the write margin (WM) metric. We also study the trends of WM across supply voltages for different assist techniques with varying degrees (percentage of V_{DD}) of assist applied.

3. Ultra-low power SRAM macro:

   After exploring various power reduction using the technology and circuit design knobs,

further minimize the leakage power dissipation can be achieved by implementing micro architecture for a sub-micro watt BSNs applications. With high-V_T devices, the array leakage power reduces significantly, the peripheral circuits start dominating the leakage power. Additionally, we propose architectural changes to co-design with low-power controller achieving system level power reduction.

3.3   Technology Consideration: Deeply Depleted Channel

[2] Ultra-low power consumption and energy-efficient operation are the key requirements for systems catering to IoT applications such as embedded wireless sensors, wearable health monitoring devices, and other similar BSN applications. In such applications, the power consumed by SRAM can dominate the total power of the SoC [30]. Scaling down the supply voltage (V_{DD}) to subthreshold voltage levels reduces the active power, but reduced on-current and variations in device threshold voltage (V_T) due to the Random Dopant Fluctuation (RDF) limits V_{DD} scaling and circuit functionality [31].

Process technology optimization is one of the promising paths enabling ULP operation. In [32], the authors demonstrated a 32nm High-K/Metal Gate (HK-MG) technology for low-power applications. The technology provides higher drive current with reduced off-current. However, it limits V_{DD} to 1.0V or above. Similarly, a 45nm HK-MG process also targets high-performance applications [33,34]. The authors in [35] addressed the limitation of voltage scaling in bulk-CMOS by using extremely thin Silicon-On-Insulator SOI (ETSOI) for low-power applications. The ETSOI [35] and Tri-gate FET [36] structures with selectively grown epitaxial channels after shallow trench isolation (STI) improve performance but do not address V_T variation due to RDF [37,38]. None of these technological advancements allow a 6T SRAM to operate in the subthreshold region or address subthreshold challenges

---

[2]This section is based on the published papers [30]([HNP6]) and [HNP5]

stated in [2] and [24]. In this work, we consider a 55nm Deeply Depleted Channel (DDC) technology with Ultra-Low-Leakage (ULL) devices that are optimized for ULP subthreshold operation due to higher drive strength, reduced variation, and support for V_{DD} scaling for the SRAM and logic. We combine technology and circuit solutions for energy efficient application needs. The proposed 55nm DDC ULL devices reduce V_T variation by fine-tuned control over the channel length while enabling Reverse Body Biasing (RBB), which controls the V_T of the devices based on the state of the SRAM. For example, when SRAM is in idle mode, performing no read/write operation, leakage power is reduced by increasing the V_T of the devices using RBB. Therefore, reduced variation, RBB, and supply scaling are combined with Ultra-Low Leakage (ULL) DDC devices to minimize leakage, power, and energy for a 6T SRAM. We fabricated a testchip with a 1Kb 6T SRAM (Figure 3.2) to demonstrate the technology-circuit technique co-design to achieve the subthreshold requirements.
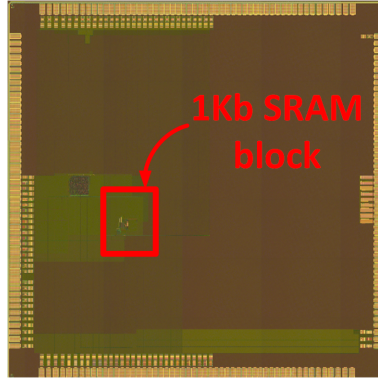
Figure 3.2: Fabricated testchip with 1Kb SRAM

3.3.1   DDC Technology Advantages and Low-Power Optimization

In the subthreshold region, leakage energy dominates the dynamic energy. The total leakage current of a device consists of subthreshold, gate, and junction leakage. Increasing the dosage of impurities in the channel raises V_T and lowers the subthreshold current. Unlike

dopant changes, an increase in the impurities worsens RDF and increases junction leakage. Researchers proposed a DDC technology for 65nm in [33] to optimize the trade-off between V_T variation and subthreshold leakage. We use new ULL devices using the 55nm DDC technology that targets total leakage current reduction with RBB. Once subthreshold leakage is sufficiently reduced, gate leakage dominates the total leakage. The gate leakage strongly depends on the thickness of the gate dielectric (T_{OX}). However, thicker T_{OX} leads to a larger V_T variation and results in 1) higher RDF, and 2) more V_T mismatch between devices. With ULL DDC devices, the V_T degradation with a thicker gate dielectric is relaxed by 60% compared to the conventional device at the same gate dielectric thickness [30].
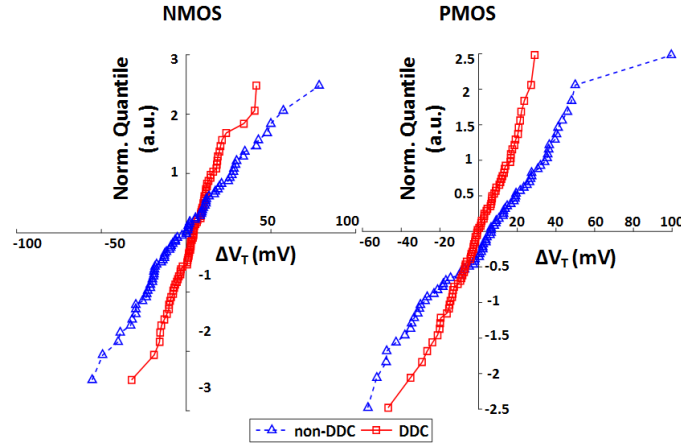
(a) V_T variation spread comparison of DDC and conventional (non-DDC) devices (Lg = 60nm)
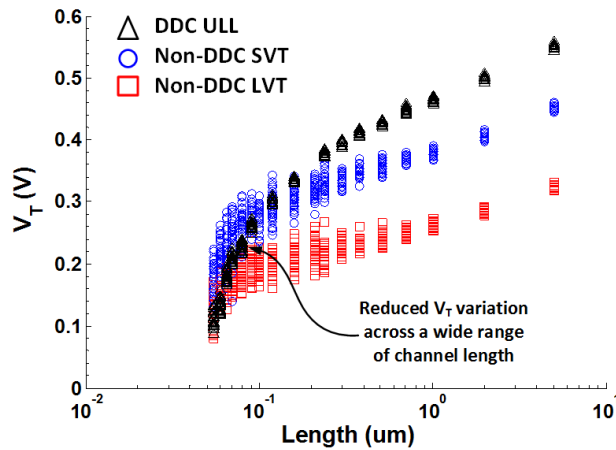
(b) V_T roll-off comparison between DDC and non-DDC devices (W=1μm, V_{DS}=0.9V).

Figure 3.3: Reduced V_T variation in ULL DDC devices helps V_{MIN} scaling
The reduction in V_T variation provides more stability for ratio-ed circuits such as SRAM

and offers better leakage control for dynamic circuits such as DRAM. Higher local and global variation disturbs the circuit functionality in subthreshold due to the exponential dependence of current on V_T. Figure 3.3a shows that the measured 55nm DDC V_T variability is much less than that of a non-DDC technology for NMOS and PMOS devices while Figure 3.3b shows V_T roll-off for a ULL device in the DDC technology compared to conventional standard (SVT) and Low V_T (LVT) devices in a non-DDC technology. ULL DDC shows a strong control over V_T across a wide range of channel lengths. Reduced V_T variation enables further supply scaling without an over-margined design overhead. Therefore, ULL DCC provides an attractive technology to address the two most pressing challenges for sub-μW systems: 1) reduced drive strength, and 2) lower yield due to V_T variation.

3.3.2 Body Biasing: Leakage Minimization and Reliability Challenge

The battery life in present state-of-the-art ULP applications depends on the amount of the leakage current in the memory cell. Figure 3.4 shows various leakage components contributing to the total leakage of the device [39]. In the previous section, we discussed how the DDC technology minimizes the gate leakage (I3) by optimal selection of T_{OX} while reducing the V_T variation (Figure 3.3). The Gate Induced Drain Leakage (GIDL) (I4) increases total leakage at higher gate voltages and therefore has an insignificant contribution to subthreshold current ($V_{GS} \leq V_T$) [40]. In subthreshold, where subthreshold leakage (I_1) dominates the total leakage, controlling threshold voltage using body (called body-biasing) reduces total leakage current significantly. The triple well structure in DDC allows RBB to accentuate the inherent benefits of ULL devices for extra power savings at low V_{DD}. As shown in Equation (1.1), the source-to-body biasing (V_{SB}) controls the V_T [38,41]. The device is reverse body biased (RBB) by applying a negative voltage to the bulk in the case of NMOS and applying $> V_{DD}$ voltage to the bulk of the PMOS to increase the V_T. Equations (3.1) and (1.1) represents

how the change in V_T controls the subthreshold current [39].
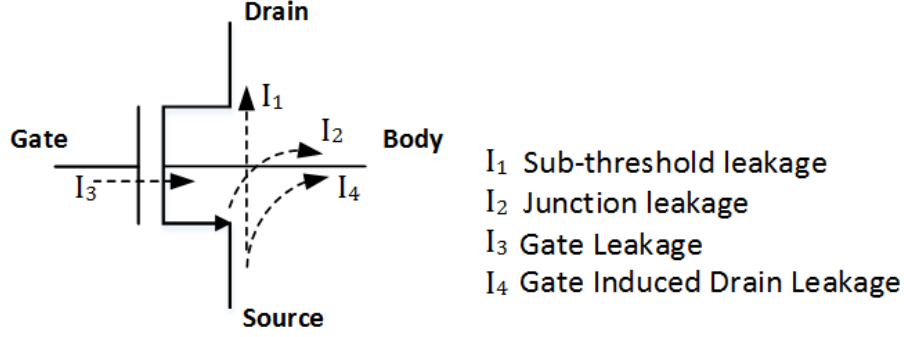
Figure 3.4: Major contributing leakage currents

$$V_T = V_{T0} + \gamma \left(\sqrt{(|V_{SB} + 2\phi_B|)} - \sqrt{(|2\phi_B|)} \right) \tag{3.1}$$

Where,

V_{T0} = threshold voltage when source is connected to the bulk ($V_{SB} = 0$)

ϕ_B = Fermi potential

While RBB provides an effective knob to reduce subthreshold leakage (Equation (1.1)), it also reduces ON current and increases junction leakage current. However, the I_{ON} versus V_{GS} trend for DDC ULL devices indicates that there is insignificant degradation in I_{ON} at higher V_{GS} due to the RBB. While at lower V_{GS} (i.e., < 0.5V), I_{ON} degradation remains much lower compared to other technologies [30]. The higher degree of RBB also increases the junction leakage current (I_2) across reverse-biased substrate-to-source/drain junction of the device. However, the lightly doped n and p regions in DDC ULL devices help to attenuate band-to-band tunneling (BTBT) dominating the PN-junction leakage. Figure 3.5 shows measured subthreshold leakage and junction leakage at varying degrees of RBB. An increase in the junction current (I_{junc}) and a decrease in subthreshold leakage (I_{sub}) is observed at higher degrees of RBB. Across V_{DD}, varying degrees of RBB results in 100X total leakage reduction ($I_{sub} - I_{junc}$).
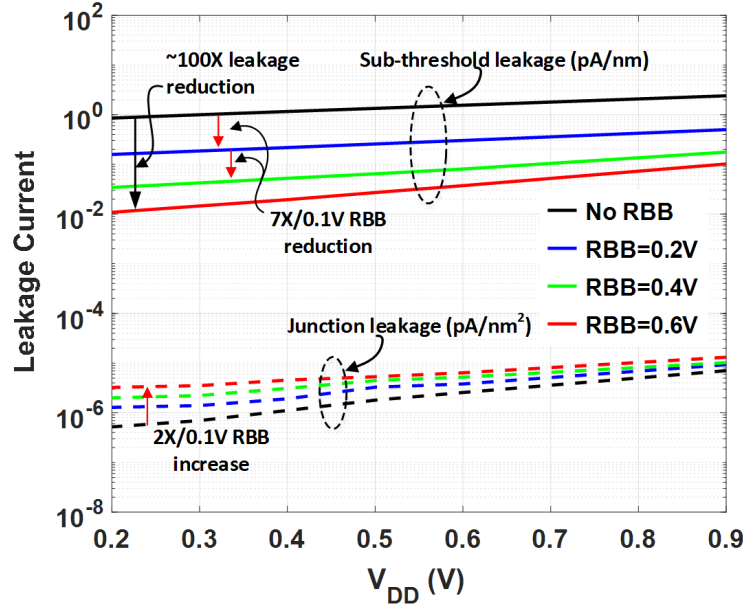
Figure 3.5: Effective leakage reduction using RBB — Increasing degree of RBB reduces subthreshold leakage by 100X while increases junction current increases by less than 10X across supply voltages
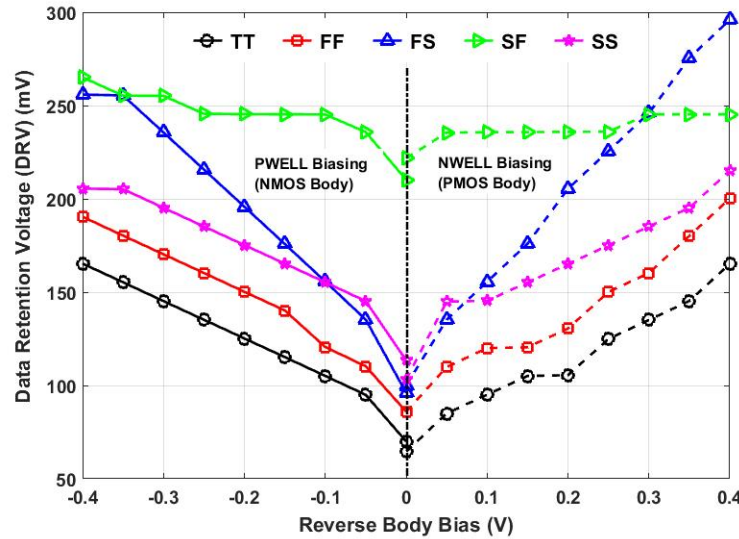
Figure 3.6: Reduced ON current increases DRV that leads towards a higher standby current

Reduced drive I_{ON} due to increased RBB also impacts the reliability and noise margins. Figure 3.6 shows increasing Data Retention Voltage (DRV) of the SRAM with increased RBB. With reduced ON current, the storage voltages move towards the metastable point and therefore the susceptibility to external noise reduces. Unreliability issues in SRAM
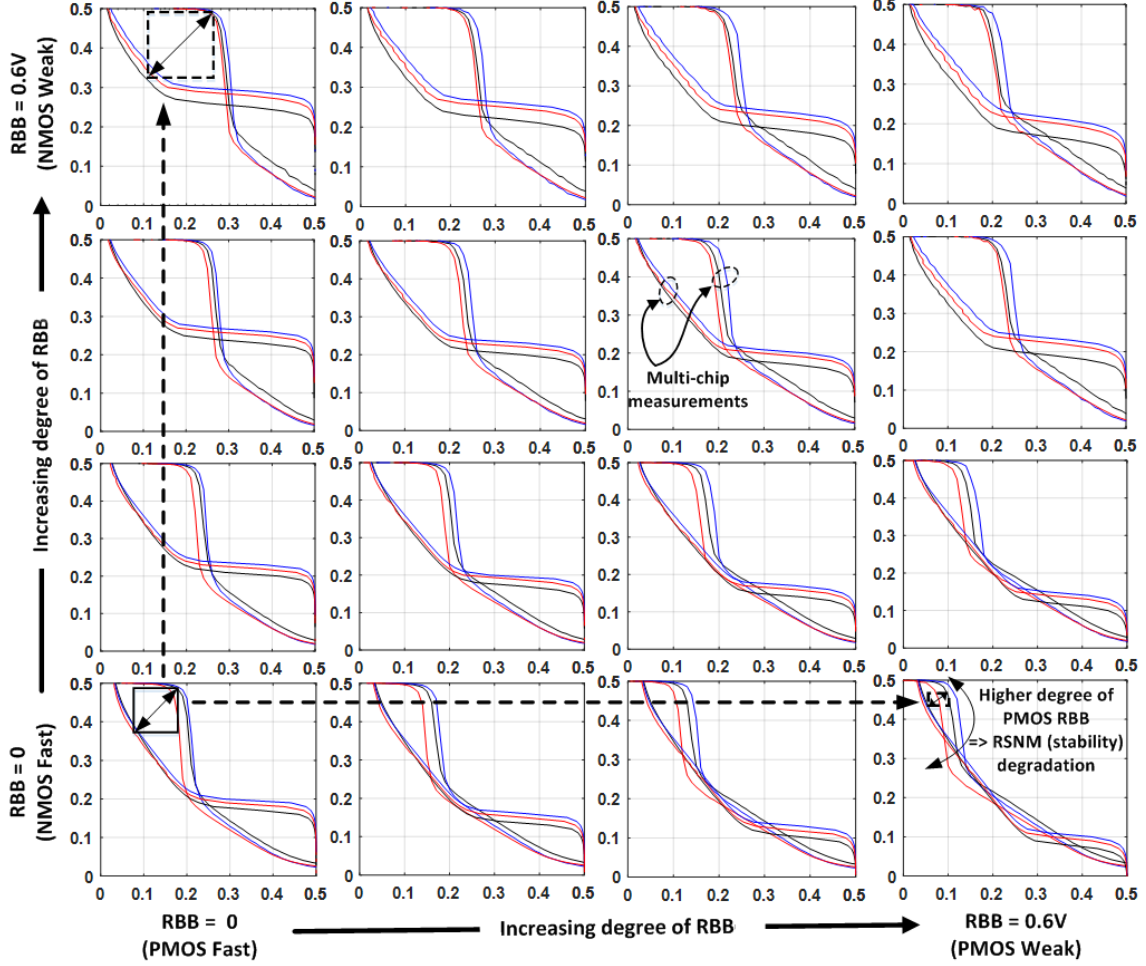
Figure 3.7: Read Static Noise Margin (RSNM) of the 6T bitcell at different combinations of PMOS and NMOS body biasing

are addressed by increasing V_{MIN} to a higher V_{DD}. Similar to DRV, the Read Static Noise Margin (RSNM) of a SRAM depends on the ON current [24]. Figure 3.7 shows measured RSNM degradation at different degrees of RBB for NMOS and PMOS devices. The smaller RSNM margin results in half-select failures. These data show the need to provide separate RBB knobs for NMOS and PMOS devices and to optimize these bias voltages to achieve cell stability at low V_{DD}s. When applying RBB for NMOS devices in the 6T SRAM bitcell, V_T of the two pass transistors, PG1 and PG2 (Figure 2.1), increases. As a result, the leakage from Bitline (BL or BLB) to the storage node reduces. Hence, applying a higher degree of RBB for NMOS improves RSNM. On the other hand, increasing V_T of PMOS devices degrades

RSNM. The greater degree of RBB reduces the drive strength of the ON devices (PMOS PU1 and NMOS PD2 in Figure 2.1) of the 6T SRAM bitcell, resulting in increased RSNM.

3.3.3 Results

To demonstrate the benefits of ULL DDC technology, we fabricated a 1Kb SRAM using a compact (0.865 x 0.492 μm^2) 6T bitcell from ULL devices. We considered an external voltage source for the body biasing. However, similar biasing can be generated with minimal area overhead [37]. Recall that ULL DDC devices are optimized to reduce the leakage current while maintaining sufficient I_{ON}. Figure 3.8 shows the 6T bitcell leakage with applied RBB for the ULL devices and compares it to the leakage in a 6T with LVT devices. The ULL cells enable a higher degree of RBB that results in 75X leakage reduction over LVT. The LVT devices limit the higher degree of RBB as a consequence of the significantly increased junction current, whereas ULL devices reduce total leakage by controlling gate and junction leakage (Figure 3.5).

One of the biggest challenges for subthreshold SRAM operation is the Read-Half Select issue that limits V_{DD} scaling [42]. Figure 3.9 shows the half-select stability (read SNM) of our fabricated 6T SRAM bitcell. The ULL 6T bitcell allows stable read operations at $V_{DD} = 0.2$V, compared to >0.4V for non-DDC devices. Most subthreshold SRAM bitcells use much larger non-6T topologies (e.g., 8T, 9T, 10T, 14T, etc.) due to small margins, so this stable 6T cell enables a much more compact solution for a lower V_{MIN} memory. In the subthreshold region, leakage energy dominates active energy. Thus leakage reduction is critical. Figure 3.10 shows 98% standby leakage reduction for our 1Kb SRAM array with RBB at 0.2V as compared to no RBB. The leakage reduction will allow a 6T SRAM array to minimize the total energy using RBB in the subthreshold region.
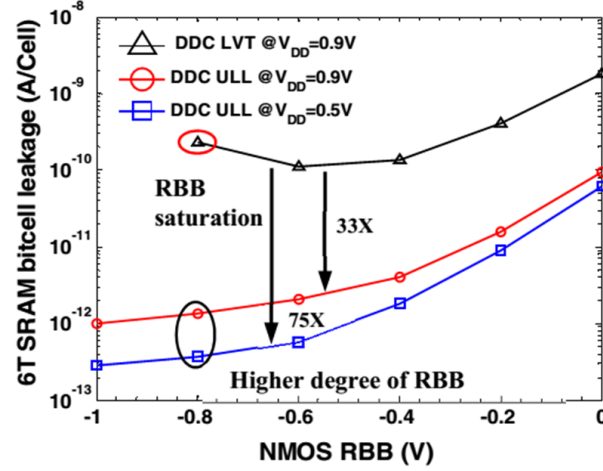
Figure 3.8: 75X 6T bitcell leakage minimization using ULL devices that allow a higher degree of RBB
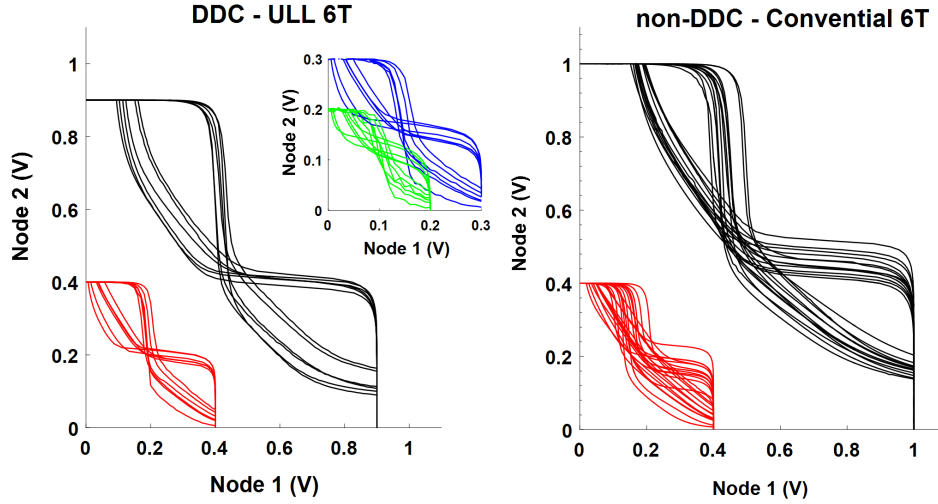
Figure 3.9: Butterfly curves for SRAM 6T bitcell: DDC ULL vs. non-DDC (conventional) bitcell.

Body biasing provides an efficient knob to achieve lower functional V_{DD} (V_{MIN}). To perform successful write and read operations at lower V_{DD}, SRAM requires various peripheral assist techniques [35]. Here, we demonstrate the effectiveness of BB for the write operation (Figure 3.11). There are various peripheral assist techniques [36] to ensure the write/read functionality. In the proposed approach, the pass-transistor (NMOS) is made stronger than the pull-up transistor (PMOS) during the write operation. Applying forward body biasing (FBB) ($V_{SB} < 0$) to NMOS and RBB ($|V_{SB}| > 0$) to PMOS helps to perform a successful write
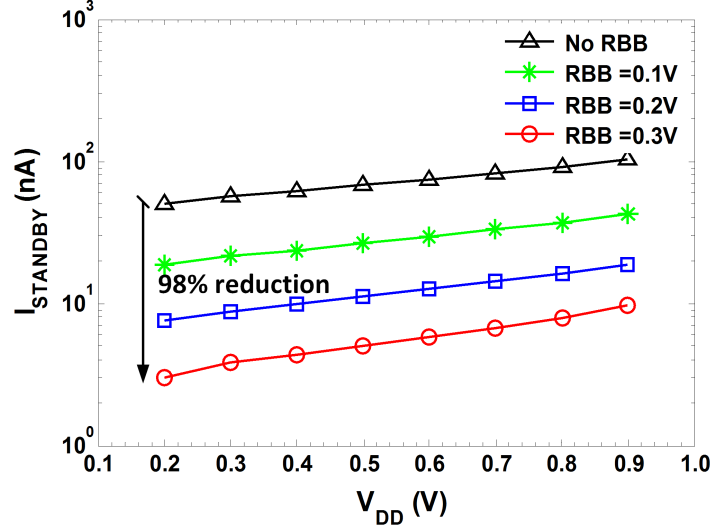
Figure 3.10: Standby leakage of fabricated 1kb SRAM macro reduction with different degrees of reverse body-biasing

operation at lower V_{DD}. To avoid the overhead of two different biasing voltage generation for NMOS and PMOS, we apply the same degree of body biasing to both, NMOS (FBB) and PMOS (RBB), devices. Figure 3.12 shows the improvement in the functional write V_{MIN} using the write margin (WM) as a static measure and maps the functional failure to the yield loss. Figure 3.12b shows the simulation results of WM across V_{DD}s with different degree of applied BB. Here, 0.1V BB represents a combination of 0.1V of RBB to PMOS devices and 0.1V FBB to NMOS devices. The combined BB (FBB for NMOS; RBB for PMOS) improves write yield by up to 80% with 0.1V BB and a further 70% improvement at 0.2V BB (Figure 3.12b). Figure 3.12 shows write V_{MIN} scaling from 0.9V (without assist) to 0.3V (BB=0.3V).

Figure 3.13 shows the measured active energy and performance of the SRAM with different degrees of applied RBB. Since the increase in applied RBB significantly reduces the array leakage current, the array achieves greater energy savings at subthreshold voltages (where leakage dominates) compared to a nominal V_{DD}. The optimized DDC ULL devices allow a higher degree of leakage reduction while maintaining sufficient I_{ON} in the subthreshold region. The use of the leakage-optimized ULL devices with RBB provides substantial power-energy
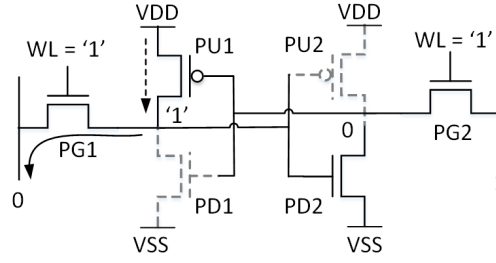
Figure 3.11: 6T SRAM bitcell: Write-0 operation
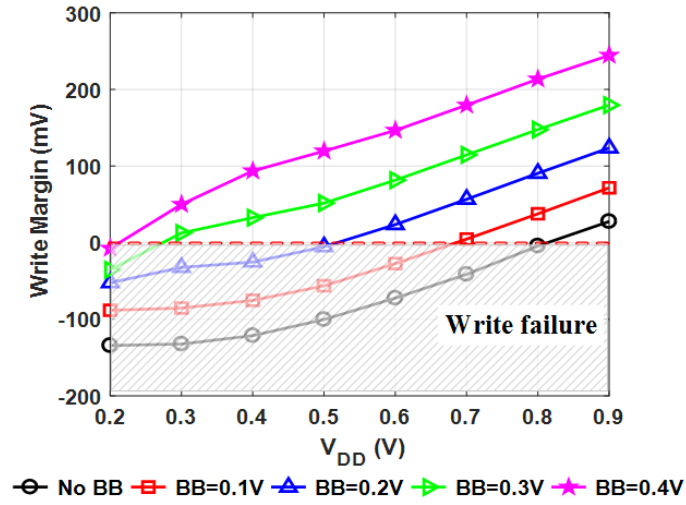
(a) Impact of BB on write-ability (write-worst corner: slow NMOS; fast PMOS): write margin (write-ability) improves by 50% with each 0.1V of BB assist.
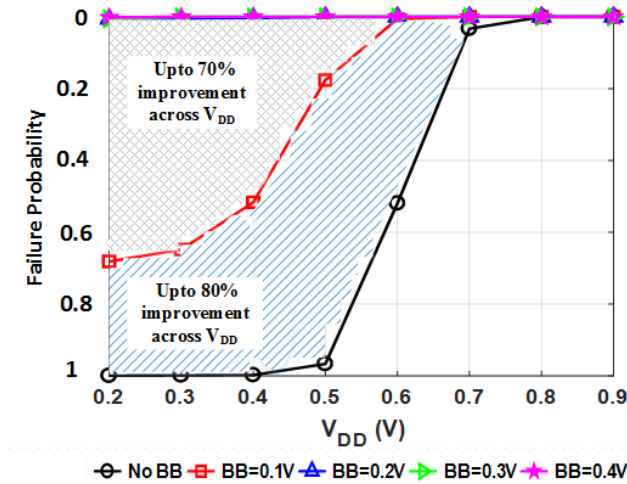
(b) Write yield improvement by different percentage across V_{DD} with applied BB as an assist technique (simulated 10000 points Monte Carlo)

Figure 3.12: Impact of RBB on functional yield improvement
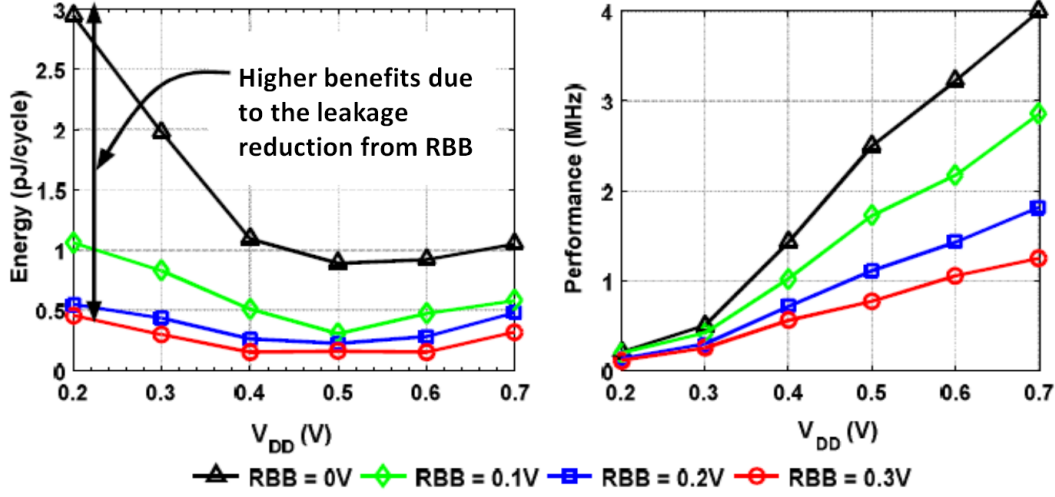
Figure 3.13: SRAM energy and performance optimization using DDC ULL devices and RBB.

benefits. The energy consumption also depends on word size.

3.4   Optimizing SRAM Bitcell Energy and Reliability

[3] After considering an optimal technology for the subthreshold design, we evaluate a higher level of abstraction the to address various subthreshold SRAM design challenges. In [43], the authors performed a similar analysis of exploring the impact of V_T at nominal V_{DD} with a focus on performance. In that study, the authors evaluated the performance of different bitcells at nominal voltages. Thus, subthreshold effects such as the higher impact of variation on SRAM performance were not examined. In addition to that, previous research exploration targets nominal voltage did not evaluate energy, an important metric for battery-less and other energy constrained IoT applications. Therefore, in this work, we explore the impact of variation in the stability and energy consumption of different bitcells at subthreshold voltages. For low-power applications, the conventional 6T bitcell does not allow V_{DD} scaling to the

---
[3]This section is based on the published papers [24]([HNP7])
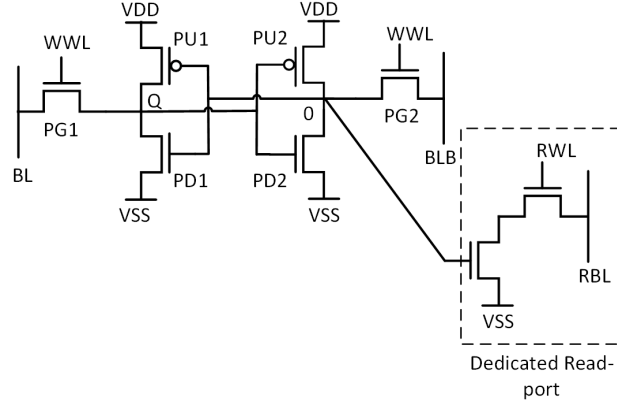
Figure 3.14: 8T SRAM bitcell

Table 3.1: Different Bitcells with device type mapping

Bitcell	Device Usage			
	PU	**PD**	**PG**	**RAs**
HVT	high-V_T	high-V_T	high-V_T	high-V_T
SVT	standard-V_T	standard-V_T	standard-V_T	standard-V_T
MVT1	high-V_T	high-V_T	standard-V_T	standard-V_T
MVT2	standard-V_T	standard-V_T	high-V_T	high-V_T
MVT3	high-V_T	high-V_T	high-V_T	standard-V_T
MVT4	high-V_T	standard-V_T	standard-V_T	standard-V_T

subthreshold region because of write and read-disturb failures [44]. Instead, the 8T bitcell
(Figure 3.14) is widely used in subthreshold SRAMs to enable the independent design of the
read and write ports. Thus, we will focus on the 8T bitcell in our analysis.

In our next approach, we choose the V_T of the devices within the SRAM bitcells to
evaluate various targeted metric improvement. For example, weakening the pull-up devices
(PU1/PU2) or strengthening pass transistors (PG1/PG2) enhances the write functionality.
Therefore, a bitcell with high-V_T pull-up devices and standard-V_T pass transistors improve
write margin and write delay metrics. On the other hand, using high-V_T pass transistors
reduces leakage energy in the bitcell. Similarly, the choice of V_T in the read port (RA1/RA2)

either improves the read speed or reduces the leakage energy. Table 3.1 shows the different bitcells studied in this work. Different combinations of high-V_T and standard-V_T devices are used within these bitcells to improve various metrics like Write Margin (WM), write delay, Data Retention Voltage (DRV), etc. Based on the characteristics of standard-V_T and high-V_T devices, a low-leakage bitcell (HVT) uses all high-V_T transistors while a high-performance bitcell (SVT) uses all standard-V_T devices. A multi-V_T bitcell, MVT3, is a derivative of HVT cell but uses standard-V_T devices in the read port to improve the read performance.

3.4.1   Comparison of Evaluation Metrics

In this section, we compare the six proposed bitcells shown in Table 3.1 with different device optimized for a particular metric improvement. The main evaluation metrics for the bitcells vary based on the targeted application. For example, an application operating at subthreshold voltages can trade-off performance to guarantee read/write functionality at scaled supply voltages to achieve lower energy. In another set of applications, the system-level power consumption might limit the V_{DD} at which the SRAM must operate to a sub-optimal voltage. In such applications, guaranteeing functionality is the primary concern. For this reason, we divide evaluation metrics into two categories: reliability and dynamic metrics. The reliability metrics ensure the static functionality (e.g., DRV, WM, Hold Static Noise Margin (HSNM), and Read Static Noise Margin (RSNM)) while the dynamic category includes metrics such as leakage, read/write energy, and operating speed. Table 3.2 shows the evaluation metrics under each category.

Table 3.2: Evaluation metrics categories

Category	Evaluation Metrics		
Reliability	DRV	HSNM / RSNM	WM
Dynamic	Leakage Power	Read/Write Energy (Power-Delay product)	Maximum operating frequency vs V_{DD}

A. Reliability

In this subsection, we compare the reliability of an SRAM bitcell for various metrics defined before. To consider the impact of process mismatch, we simulate each design parameter by running 1000-point Monte Carlo (MC) simulations to address the within-die (intra-die) variation, while the robustness against across process corners (inter-die) variation and temperature will be discussed in the later part of this section. We present a quick definition of the stability metrics — DRV, HSNM, RSNM, and WM — before discussing the results in detail. The minimum V_{DD} below which the storage nodes (Q-QB) flip when the bitcell is un-accessed (WWL/RWL=0, BL=BLB=1) is defined as the DRV. In the DRV test, the V_{DD} of an un-accessed bitcell is reduced until storage nodes (Q-QB) flip. The HSNM quantifies the ability of an un-accessed bitcell (WWL/RWL=0, BL=BLB=1) to reject DC noise. The RSNM is defined as the ability of a half-selected cell to maintain its state during a pseudo-read operation (WWL=1, BL=BLB=1). The techniques introduced in [45] are used to measure the HSNM and RSNM.

Figures 3.15a and 3.16 show the distribution of DRV, HSNM, and RSNM evaluated at T=25°C with the worst case (min/max) and $\pm 3\sigma$ variation results for each metric. The plots highlight the best and worst choice of bitcell. Since we target ULP BSN applications operating in the near or subthreshold region, RSNM and HSNM metrics were evaluated assuming a supply voltage of 0.5V. The distribution of DRV for different bitcells shows how much intra-die variation change the effective DRV limit. To understand the rationale behind

the DRV variation across different bitcells, we characterize the effect of different devices on DRV by varying their V_T and measuring the sensitivity of the DRV to this change [46]. Figure 3.15b shows the change in DRV as a function of the change in V_T. At low V_{DD}s, the reduced I_{ON}-to-I_{OFF} ratio cause the state to flip even when the bitcell is un-accessed. However, as shown in Figure 3.15b, when the I_{ON} is increased by stronger PU1 and PD2, and when I_{OFF} is reduced by weaker PG2, the DRV can be lowered. As MVT4 provides lower ON current from PU devices and higher OFF current from PDs and PGs, it provides the worst DRV among all other cells.

From Figure 3.16 a), HVT, MVT3, and MVT1 show higher HSNM than other bitcells while MVT4 has a wider distribution of HSNM values. Similar to DRV, the sensitivity of the HSNM to V_T changes in the bitcell transistors affects the behavior of different bitcells. The important parameter controlling the RSNM is cell's β ratio (i.e., the relative strength of PDs-to-PGs). As standard-V_T devices provide higher strength over high-V_T devices, the bitcells with standard-V_T PDs devices and high-V_T PG devices (e.g., MVT2) provide the highest RSNM compared to other bitcells; whereas the contrary (e.g., MVT1) provides the lowest RSNM. Therefore, IoTs applications not employing these solutions for low V_{DD}s

(a) DRV distribution of different bitcells

(b) Impact of device choice on DRV mapped as change in V_T (bitcell with Q node holding 1)
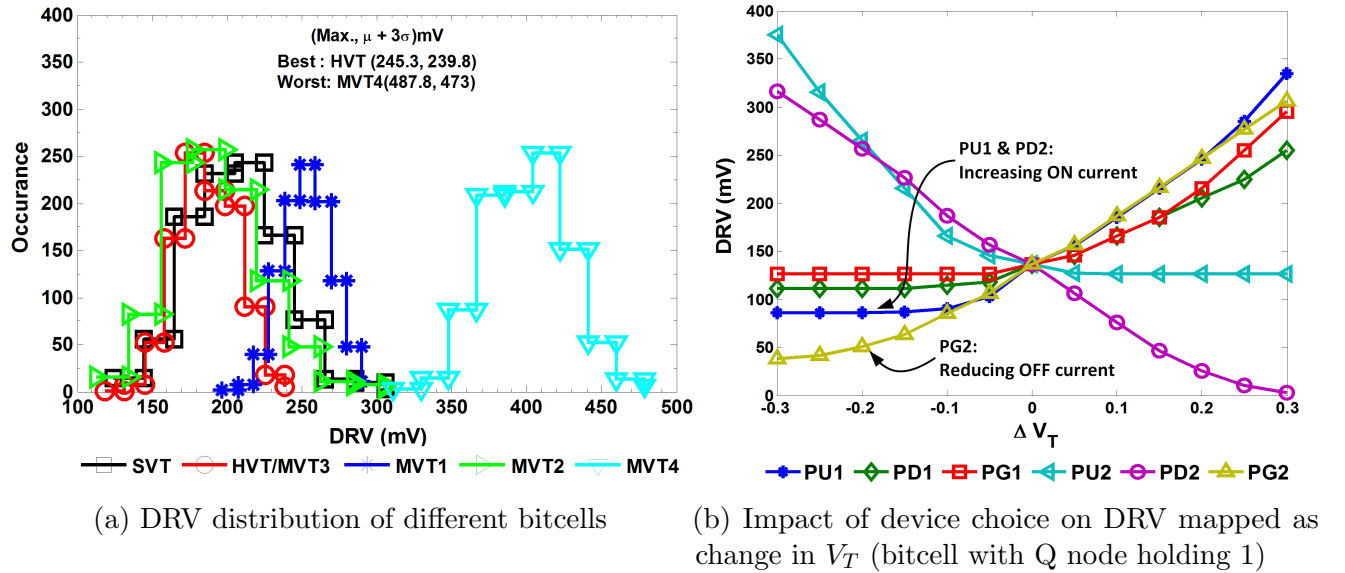
Figure 3.15: Comparison of the different Bitcells for DRV
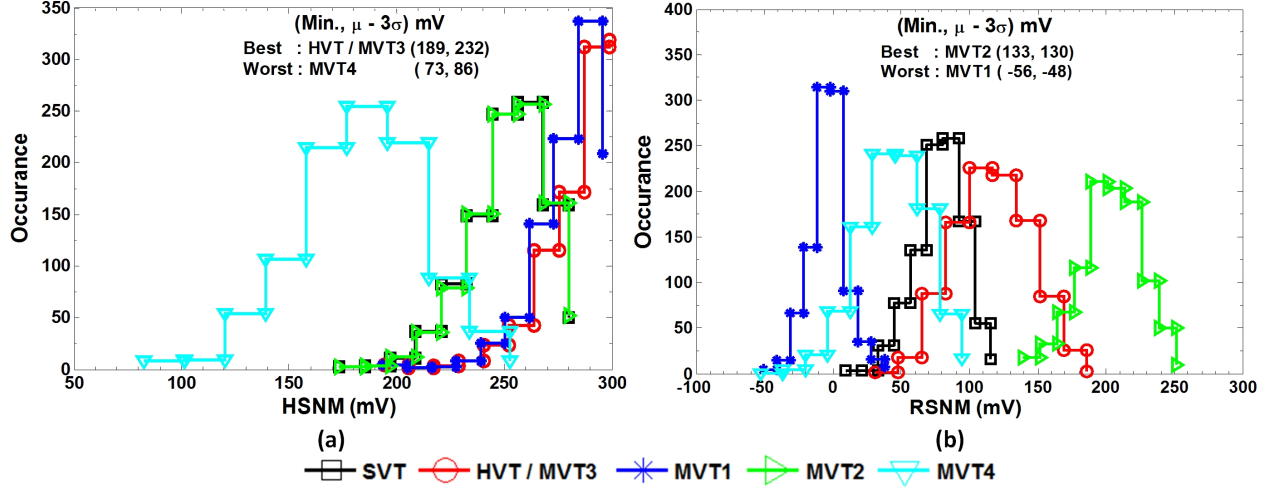
Figure 3.16: Noise margin comparison: a) HSNM, b) RSNM - Distribution under local variation and optimal choice
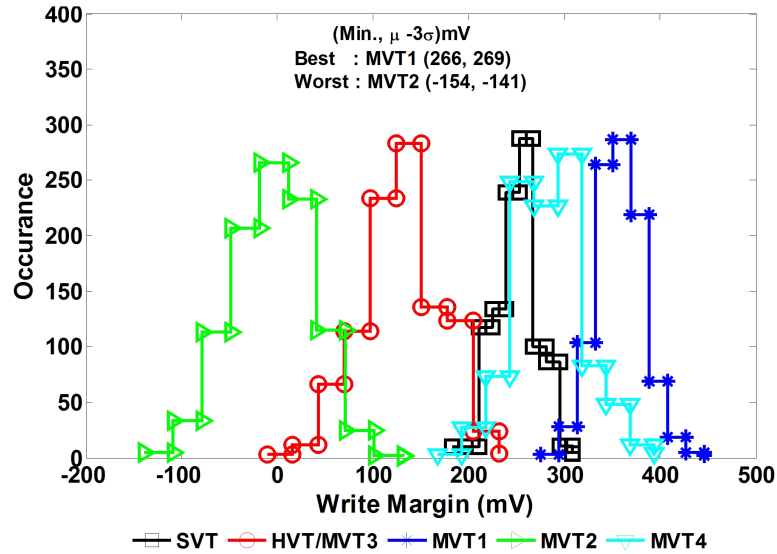
Figure 3.17: WM distribution of different bitcells for V_{DD}=0.5V

operation require an SRAM with smaller DRV (HVT/MVT3), higher HSNM (HVT/MVT3), and higher RSNM (MVT2). Figure 3.17 evaluates the WM of the different bitcells for $V_{DD} = 0.5$V as the subthreshold supply voltage. A successful write operation depends on the relative strength of PU and PG devices. To flip a bitcell, the PGs have to be stronger than the PUs of the cell. Therefore, the bitcell with high-V_T PU devices and standard-V_T PG devices (i.e., MVT1) gives an optimal choice for the write operation (Figure 3.17).
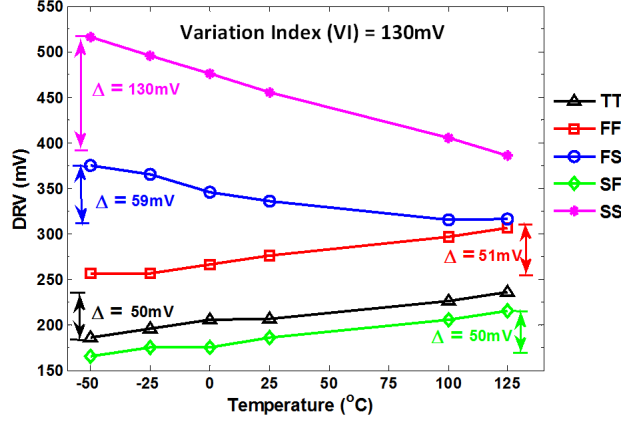
(a) Impact of variation on DRV across corners and temperatures.
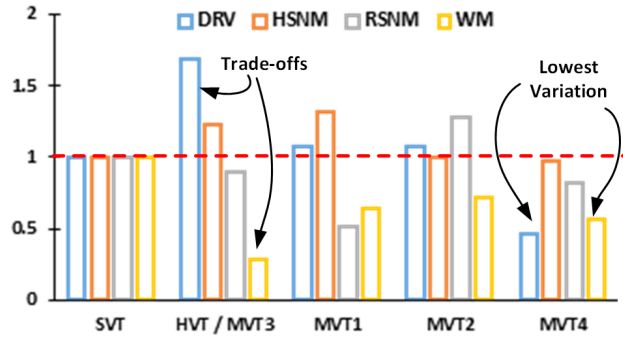
(b) Comparison of VI for different metrics across bitcells

Figure 3.18: Comparison of the different Bitcells under the variation

After addressing the intra-die variation measured across 1000-point MC simulations for different metrics; we evaluate the impact of process and temperature variation. Because I_{ON} has an exponential dependency on V_T [24], marginal variation in V_T (process variation and mismatch) disturbs the functionality of a ratio-ed design such as an SRAM. Similarly, V_T has a linear dependence on temperature [10]. Due to these facts, to create a robust design, we study the susceptibility of the cell against V_T (intra-die and inter-die) and temperature variations for a robust design. We consider five different process corners (TT, FF, FS, SF, SS — NMOS/PMOS) to study the inter-die variation. The robustness against temperature variation is measured by considering a wide temperature range of [-50, 125]^oC.

Figure 3.18a shows a plot of the worst-case DRV of the SVT bitcell across temperature and corners. Here, the worst-case point at a given temperature and corner includes intra-die variation calculated by running 1000-point Monte Carlo simulation and taking the worst measurement. In Figure 3.18a, the SVT bitcell is shown as an example to define the Variation Index metric. We define the Variation Index (VI) as the maximum deviation in a metric that chip, fabricated at any corner, can experience due to temperature variation. For example, an SVT bitcell experiences maximum VI (worst-case variation impact) of 130mV for the selected range of temperatures across different corners. Figure 3.18b normalizes the VIs of the stability metrics for different bitcells. Here, it is important to note that these values represent the variation, not the actual values for the metric. Based on Figure 3.18b, MVT4 shows an optimal choice for a temperature variation resilient (less variation) design despite being considered as the worst case for the DRV and HSNM (Figure 3.15a and Figure 3.16a) while HVT/MVT3 cells experience different trade-offs for the different metrics when compared to SVT bitcell.

The choice of the threshold voltage for each of the bitcell transistors (e.g., PGs, PUs, or PDs) has a different impact on each of the selected metrics, thus resulting in different VIs for different bitcells. This is due to the characteristics of each device at a given corner and temperature. To better understand these results, Figure 3.19 explores the normalized I_{ON}/I_{OFF} ratio of the different devices (high-V_T and standard-V_T) used within the bitcells across temperature and process corners for V_{DD}=0.5V. While these device characteristics depend on the technology and the foundry; they provide an excellent insight into the impact of threshold voltage choice on stability metrics. As shown in Figure 3.19, high-V_T devices (NMOS and PMOS) exhibit higher variation across corners and little variation across temperatures; whereas, the standard-V_T devices show an opposite trend. Since HSNM and DRV metric depends on I_{ON}/I_{OFF} characteristics of the devices used, these trends help explain the results in Figure 3.18b. The bitcell with more standard-V_T devices (e.g., SVT) experience higher variations in DRV/HSNM due to temperature variation than high-V_T devices. Similarly, the

bitcell with more high-V_T devices (e.g., HVT) faces higher variations in DRV/HSNM due to process variation than standard-V_T devices. The combination of these devices results in wider variation characteristics (Figure 3.18b) that require device physics knowledge. Therefore, we propose VI to be considered as one of the most influential metrics for a robust SRAM design for IoTs applications operating in the subthreshold region — where variation is a major concern — and under a wide range of environmental conditions such as temperature.
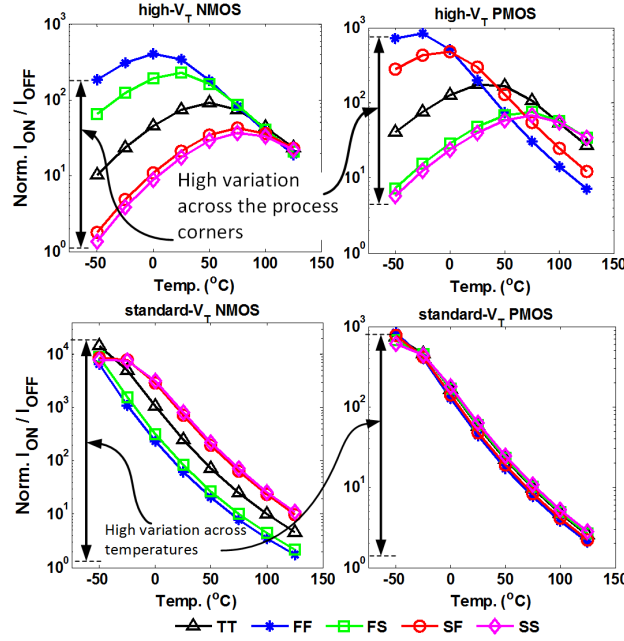
Figure 3.19: Normalized I_{ON}/I_{OFF} characteristics for different devices across process corners and temperatures
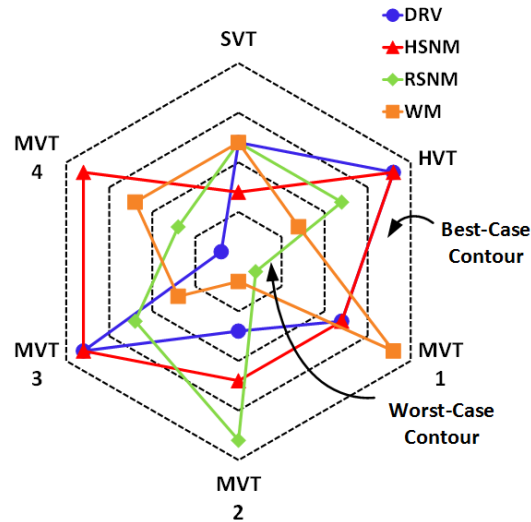
Figure 3.20: Optimal Bitcell selection based on static metrics

Figure 3.20 summarizes the optimal SRAM bitcells for each static metric at low-V_{DD}s and also provides the trade-offs between different static metrics and orders the bitcells according to their effectiveness in improving the corresponding metric. Each edge represents bitcell while the contours with different colors correspond to the metric. As shown in the figure, from outer to inner contour represents the choice of the bitcell from the best to the worst case option.

B. Dynamic Metrics

After evaluating static metrics for robustness, we explore the dynamic metrics in this subsection. We consider 1KB array to calculate leakage power, write/read energy, and operating frequency at low V_{DD}s. Figure 3.21 shows the leakage power of different bitcells while accounting for process and temperature variations. Each plot represents the worst-case leakage power calculated by running 1000-point Monte Carlo simulation. The results clearly distinguish the HVT and MVT3 bitcells as optimal choices for low-leakage application with 370% reduction in leakage compared to SVT bitcell at V_{DD}=0.5V at a nominal temperature (25°C). The percentage reduction in leakage power varies across temperature from 755% (-50°C) to 142% (125°C). Many IoTs applications with less activity factor and higher standby time will significantly benefit from the lower leakage of the HVT/MVT3 bitcells.
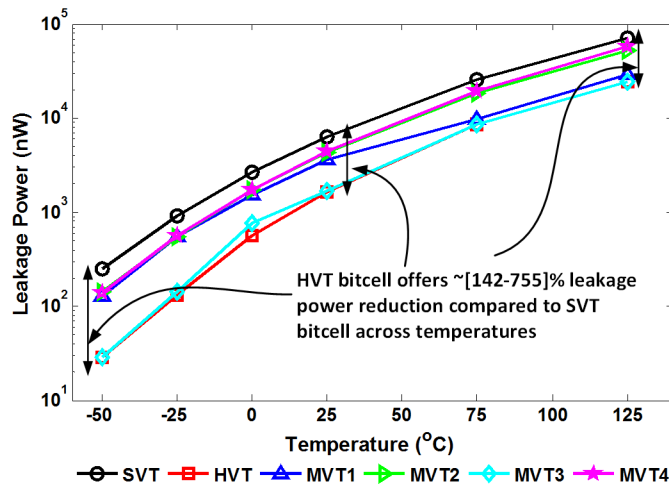
Figure 3.21: Leakage power of 1KB array across temperature for different bitcells

Figure 3.22 shows a comparison of write/read energy across V_{DD}s at their respective worst-case corners. We calculate energy as a power-delay product for an array without the drivers assuming constant peripheral power for different bitcells. The write energy plot shows that bitcells with standard-V_T devices (SVT and MVT1) provide lower active write energy in the subthreshold region because of their faster operation compared to the bitcells with high-V_T devices. At higher V_{DD}s, the standard-V_T transistors offer only marginal improvement in delay over the high-V_T transistors, and thus, high-V_T bitcells have lower write energy. Similarly, the read energy in Figure 3.22 shows how the power-delay product (energy) of different bitcells impacts the optimal choice of bitcell. Figure 3.23 shows the consolidated results of the dynamic metrics for the bitcells with an optimal bitcell selection order in a spider plot.
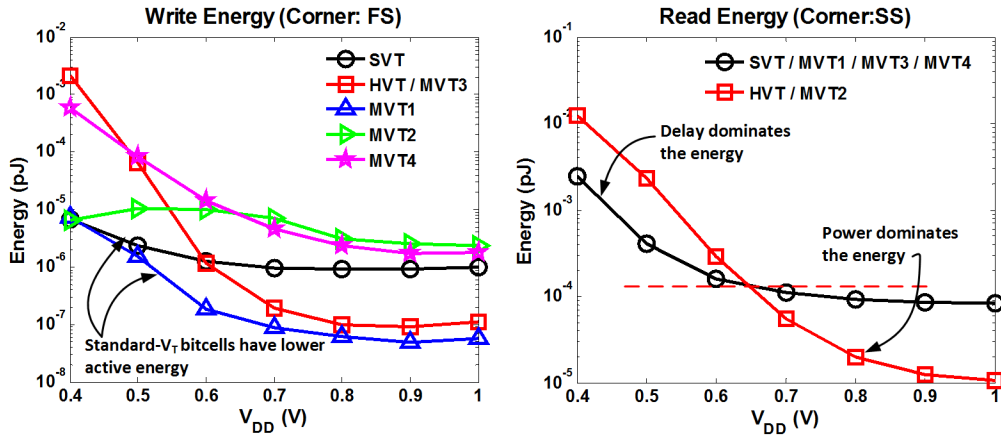
Figure 3.22: Write and Read energy comparison of the different bitcells across the V_{DD}s
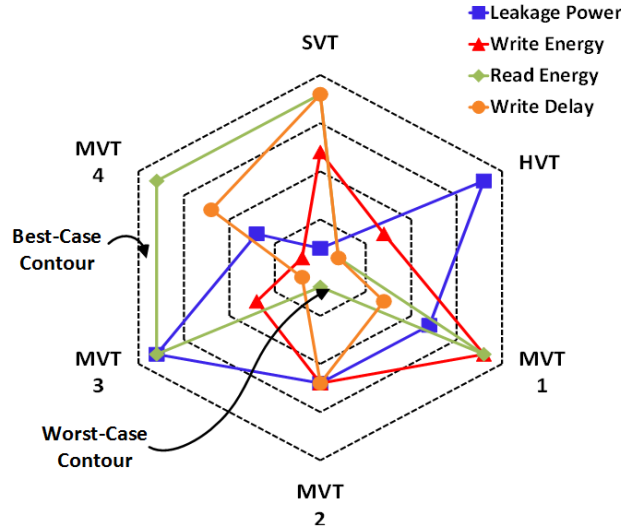
Figure 3.23: Optimal Bitcell choice for dynamic metrics

3.4.2 Results

Figure 3.21 and Figure 3.22 show that the applications being in 'Standby' mode for a long time require low-leakage bitcells, while 'Always ON' applications require lower active energy bitcells. Since many IoTs applications fall under the category of 'Standby — mostly' applications, we fabricated a 2KB array consisting of a 1KB HVT bank as the leakage optimal choice and a 1KB MVT3 bank as the lower active read energy with leakage minimization. Figure 3.24 shows a die micrograph of the fabricated chip in a commercial 130nm technology.
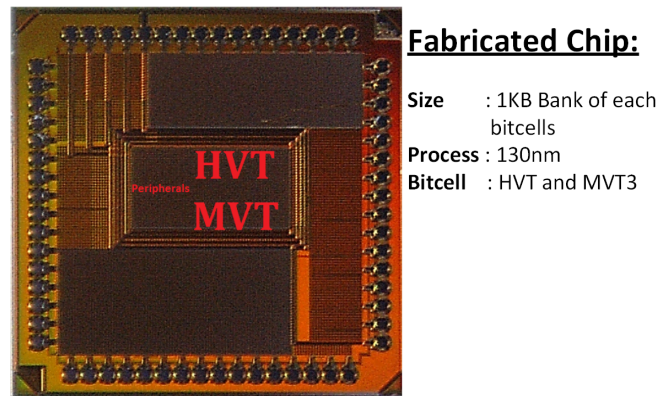
Figure 3.24: Die micrograph of chip showing HVT and MVT banks

(a) Leakage measurement (V_{DD}=0.5V)

(b) Write and Read energy across V_{DD}s

(c) Subthreshold performance with variation and across the V_{DD}s
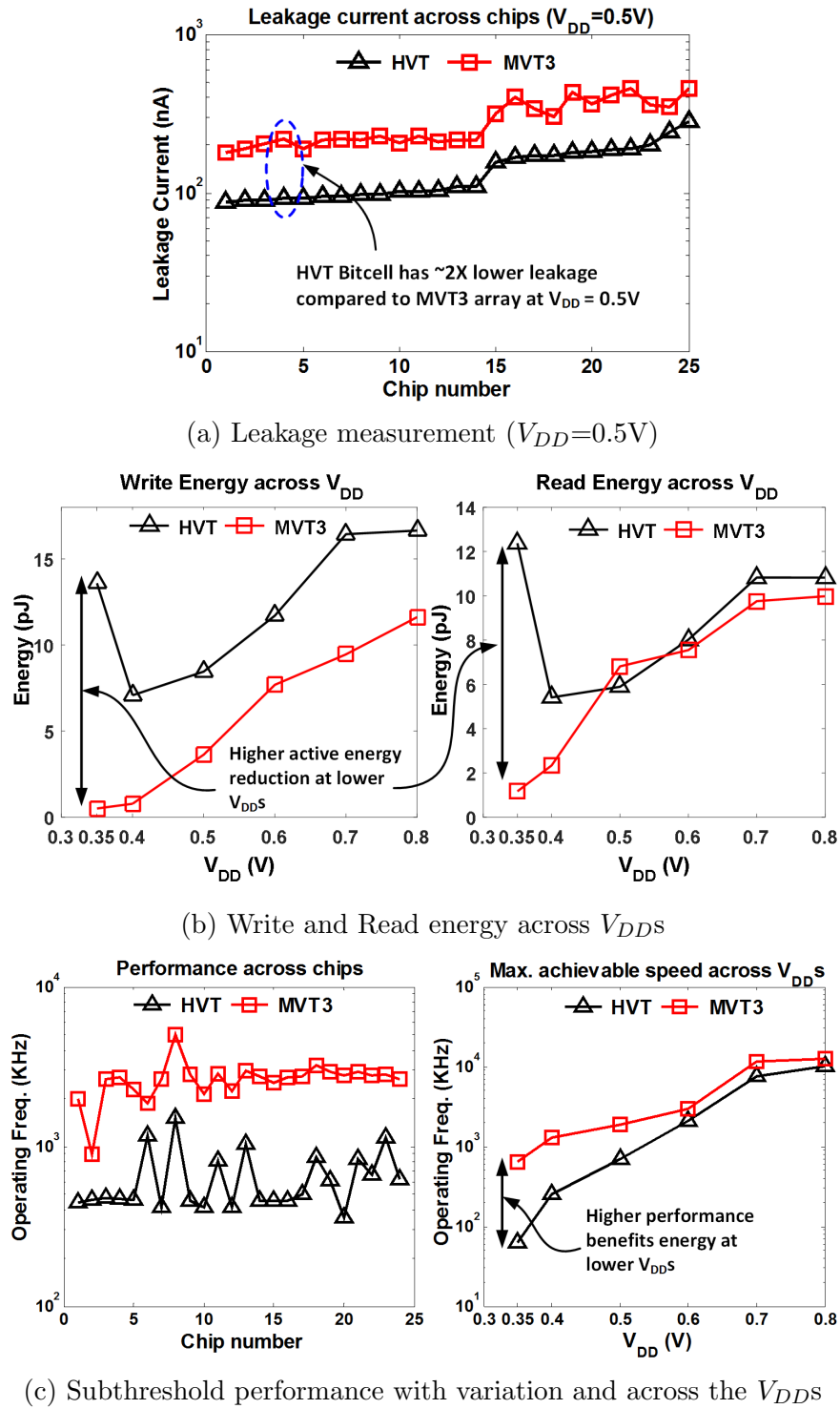
Figure 3.25: Measurement result comparison between HVT and MVT3 bitcells from 24 chips

Figure 3.25 shows the measured leakage current, read/write energy, and performance comparison between the HVT and MVT bitcells. The leakage current observed across 24 chips shows a steady reduction of leakage current by ~2X at V_{DD}=0.5V for the HVT array. The higher leakage power in MVT3 is due to bitline to bulk leakage from the read pass transistor (RA2). We measure the write and read energies as an average of write/read '1' and write/read '0' operations. The MVT3 bank shows a ~2X reduction in the measured energy numbers as compared to the HVT bank. The decrease in read energy is due to the higher read speed of the MVT3 bitcell. Since the fabricated arrays employ read-before-write to address half-select failures during a write operation, the increased read speed will also result in lower write energy. The measured results show deviation from the simulation-based results (Figure 3.22) because additional peripheral circuits were implemented to make the array functional. The results indicate that the MVT3 bitcell reduces active energy compared to HVT bitcell while providing lowered leakage compared to other bitcells.

3.5   Enabling the Next Generation BSNs SoC

[4] In addition to the SRAM array power reduction using high-V_T devices, we further reduce leakage power by various circuits and architecture-level techniques. For example, selectively shutting off the part of the SRAM that is in idle mode further reduces the leakage. Here, the aim is to enable the self-harvesting BSN chip [2] that is highly duty-cycled to achieve sub-μW power. As we highlighted in Figure 2.7, the leakage energy dominates in subthreshold. We discussed various techniques such as the use of low-leakage devices (e.g., ULL in DDC or High-V_T) and smaller bank size to reduce the numbers of bitcells per bitline [47] to minimize the leakage. After optimizing the SRAM bicell array leakage, the peripheral circuit components require leakage mitigation techniques.

---

[4]This section is based on the published papers [HNP1], [HNP3], and [HNP8]

In this work, we present an ULP 1KB SRAM array designed to minimize the leakage power of the SRAM in battery-less BSN applications. The proposed array is validated using a fabricated test-chip in a commercial 130nm process. The proposed SRAM architecture can be easily expanded to extend the capability of the BSN node. Table 3.3 summarizes the specifications and different features of the proposed ULP SRAM macro. The array uses high-threshold (high-V_T) devices and aggressive power gating to reduce the power consumption. A read burst mode is implemented to address the read half-select failures at subthreshold. An 8T bit-cell and a read-before-write implementation are used to address half-select failures. Read and write assist techniques are introduced to ensure correct read and write functionality in the subthreshold regime. The array can operate reliably down to 350 mV and can retain data down to 320 mV. Leakage power is minimized to 12.29 nW/KB at data retention voltage and 1.09 nW/KB without data retention. The proposed SRAM design makes use of multiple circuits, architecture, and assist methods in a unique combination that optimizes SRAM for the targeted ULP BSN application.

Table 3.3: Design specifications and features of the proposed SRAM macro

Tech.	Commercial 130 nm Complementary Metal Oxide Semiconductor (CMOS)
Cell	8T static random access memory (SRAM) cell
Size	1 Kbyte (64 × 128), 16-bit/word
Voltage	350–700 mV
Leakage Power	12.29 nW @ 320 mV (standby) 1.09 nW @ 320 mV (shutdown)
E/access	6.24 pJ/access @ 400 mV
Special features	1. High-threshold (high-V_T) devices 2. Full-swing read 3. Read burst mode 4. Read wordline (RWL) boosting to improve read stability 5. Read-before-write for half-select instability 6. Write wordline (WWL) boosting to improve write stability 7. Aggressive power gating for low power standby and shutdown modes

3.5.1 SRAM Architecture

As shown in 3.1, the data memory (DMEM) and instruction memory (IMEM) are each allocated 2KB. Figure 3.26 shows the overall structure of the array with various block diagrams within the SRAM macro. The 1KB array consists of 64x128 8T bit-cells with row (RDx) and column (CDx) drivers, a row decoder, a read/write control unit with a burst control unit (BCU), and a data management unit (DMU). Due to the challenges of operating the conventional 6T cell at subthreshold voltages, an 8T bitcell with decoupled read and write ports is used. High-V_T devices are used within the bit-cells to reduce their leakage currents and thus the standby power consumption of the array. However, since high-V_T devices have reduced ON current, the read and write margins are significantly degraded, necessitating the use of assist techniques to guarantee correct operation. Additionally, the peripheral circuit implements regular-V_T devices to meet the timing signal generation.
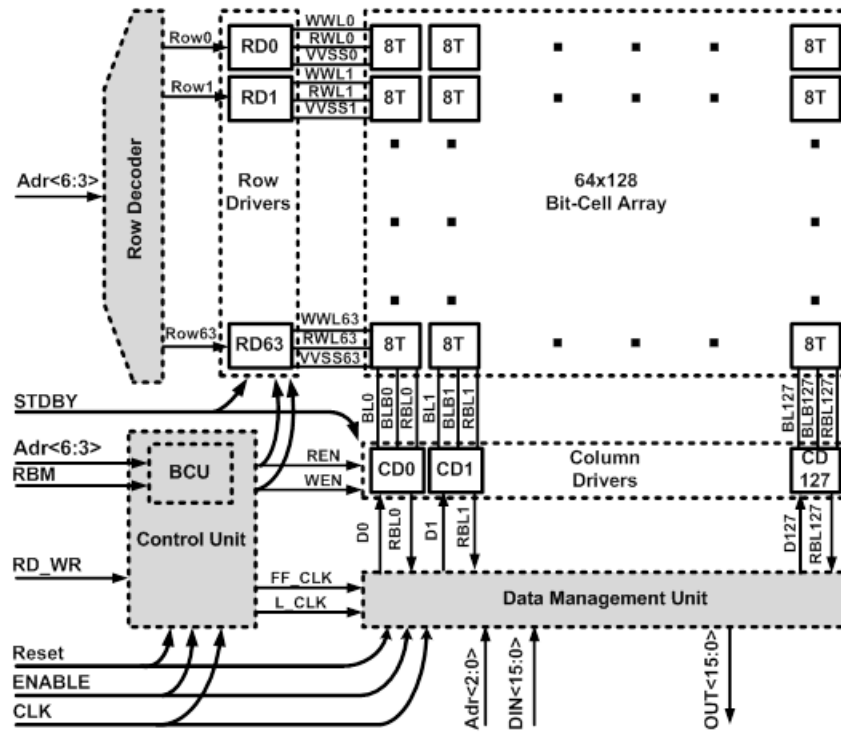
Figure 3.26: Architecture of a low-power 2KB SRAM macro with various power management schemes

3.5.2 Results

To address the HS failures during the write operation in unaccessed bitcells, we implement read-before-write (RBW) where a complete row of the accessed word is read and modified write with new data for the intended word. Here, a successful read of the row is ensured by read assist technique. In addition, to reduce the read power/energy, the array employs an RBM feature which makes use of the fact that when RWL is asserted, the complete row experiences a read operation. Thus, when consecutive addresses in the same row should be read, it is enough to perform the read operation once and save the data in latches for the sequential reads. Accessing the latches will consume significantly lower energy than performing a regular read thus reducing the overall read energy. The Burst Control Unit (BCU) implementing RBM has a negligible impact on the power ($<0.7\%$), performance (0%) and area ($<1\%$) of the system and the potential savings it offers is significantly higher than the cost of implementing it. In Figure 3.26, the BCU implements the RBW and other control logic for the SRAM. Figure 3.27 shows up to 22% read energy minimization using RBM (V_{DD}=0.4V).

The architecture implementation of RBW and RBM reduce the active energy of the SRAM array. We implement three different modes to reduce further active and leakage power:

a) Hold Mode: During this mode, we reduce active power by minimizing clock activity. To enable the Hold Mode, ENABLE signal (Figure 3.26) is applied to perform clock-gating. During Hold mode, SRAM data is maintained while keeping peripheral circuitry ready to access the SRAM. Because of clock-gating, the active power is reduced. However, the efficacy of Hold mode reduces at lower V_{DD}s where leakage power dominates active power (as shown in Figure 3.28).

b) Standby Mode: In a computational system, the SRAM is only being accessed during a fetch (read) and store (write) cycles, and not during execution and computational cycles. Therefore, during the idle state, the Standby mode is applied by asserting STDBY signal

(Figure 3.26). Standby mode provides power to the SRAM array while power-gate the peripheral circuitry to reduce the leakage power from it. Figure 3.28 shows a tenfold reduction in power compared to the Hold mode. It is important to note that the Standby mode requires longer recovery time (i.e., from the idle mode to functional mode) than the Hold mode. Therefore, the effectiveness of the Standby depends on system operations — the longer the idle time for SRAM, the better the recovery time-to-power saving ratio.

c) Shutdown Mode: The system can shut down the SRAM when the stored data is no longer needed. As shown in Figure 3.28, Shutdown mode saves twelve times more power than standby mode.
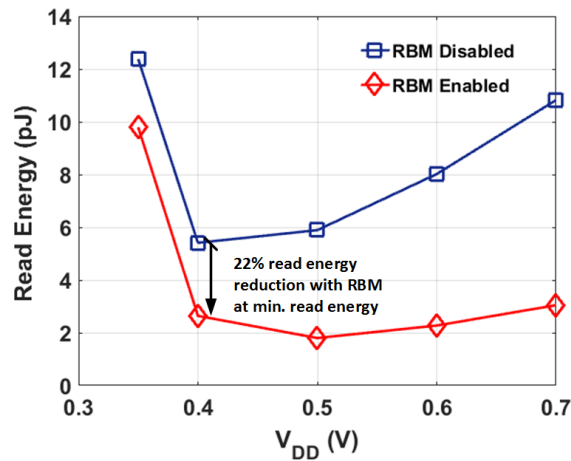
Figure 3.27: Energy minimization using RBM
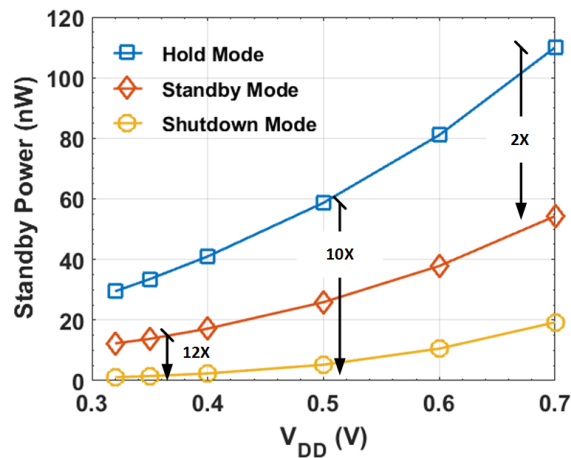
Figure 3.28: Leakage power reduction using different modes: 2X leakage power reduction when data retention is required (standby mode) and 10X power saving with data loss (shutdown mode)
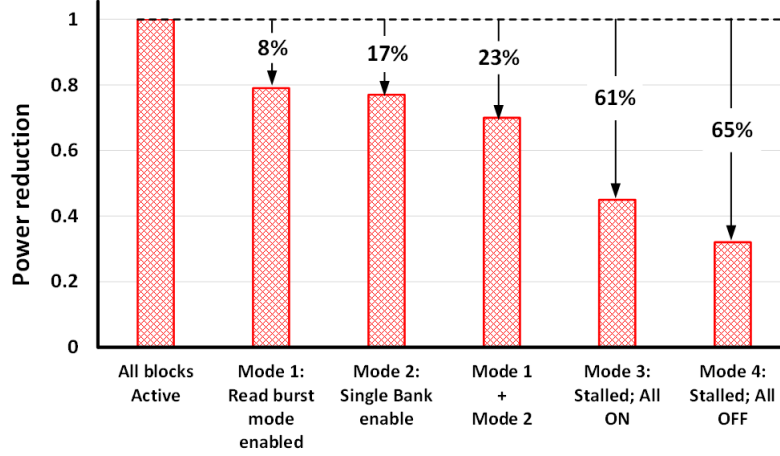
Figure 3.29: System-level effectiveness of various power/energy saving techniques implementation in SRAM

We used proposed SRAM in BSN architecture (Figure 3.1) to quantify the system level benefits from the leakage power and active energy reduction using variation optimization techniques implementation. The SRAM and low-power controller (LPC) are tightly coupled to create various power saving modes. The LPC takes full advantage of these SRAM features resulting in significant (up to 66%) measured power savings (Figure 3.29).

After optimizing the active and leakage power by applying various circuit to architecture-level technique, we successfully reduce the leakage dominance from the SRAM array. However, once the array leakage is reduced, the peripheral circuits start dominating the total power dissipation in the SRAM macro. Figure 3.30 shows the percentage power distribution in the main peripheral components. While analyzing the total distribution of power in peripheral circuitry (Figure 3.30), we found that the read-before-write (RBW)[5] consumes 86% of total peripheral power. Therefore, reducing RBW power will drastically reduce the peripheral leakage power. While RBW limits the leakage power, the lack of a sense amplifier (SA) limits performance. For a BSN-like application where the performance requirement is very low, implementing SA will not help in lowering the total power.

Therefore, we further optimize the peripheral circuits to reduce the leakage and active

---

[5]The architecture technique to address half-select issue at lower V_{DD}s
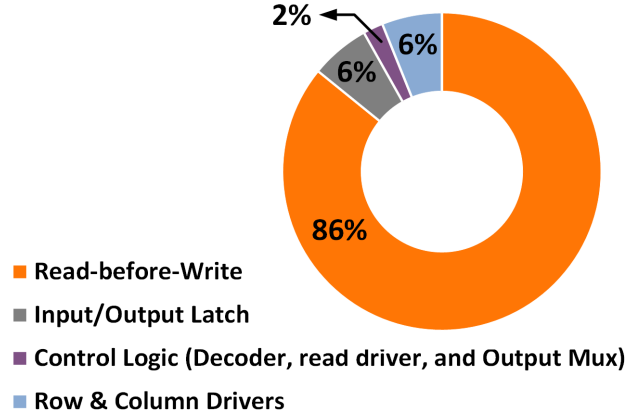
Figure 3.30: Power distribution in 2KB SRAM Peripheral blocks

power by implementing Stacked powergating for the peripheral circuits leakage minimization. Figure 3.31 show the impact of an increase in device stacking on I_{ON} and I_{OFF} currents. As Figure shows, with 2-device stacking, the leakage current (I_{OFF}) decreases significantly (by 65%). This reduction in leakage current comes as a result of an increase in OFF resistance between the path from V_{DD} and ground. The stacking of devices increases the threshold of the device due to body-biasing[6] on the top and hence further reduces the leakage current. However, increase in a number of stacking device does not follow super exponential I_{OFF} reduction. Three device stacking results in only 28% further reduction in I_{OFF}. The percentage benefits on I_{OFF} reduction reduces with increase in stack devices. Adversely, the stacking of the devices also reduce the I_{ON} currents. Figure 3.31 shows I_{ON} reduction with a number of stacked devices. Figure 3.31 highlights that more than 2-stacked devices results in less benefit in I_{OFF} reduction while penalized with higher I_{ON} reduction that impacts the wake-up time from the sleep mode. Therefore, we implemented 2-stacked devices for the peripheral circuits to reduce the leakage during the sleep mode (discussed before) in a newer version of the SRAM design. We fabricated proposed optimization in a 130nm technology, and testing is pending to share the measurement data.

---

[6]In a stacked device circuit, the source of the top device is connected to the drain of the bottom device while the conventional bulk is connected to the ground for the NMOS devices. As a result of the potential difference between bulk and source of the top device, V_T increases.
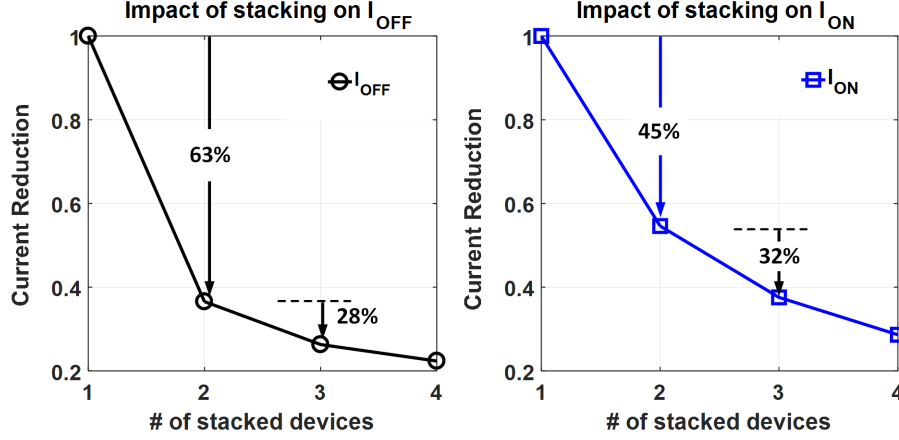
Figure 3.31: Impact of increase in device stacking on I_{ON} and I_{OFF} currents

3.6   Conclusions

In this chapter, we address the subthreshold design challenge at three different abstraction layers of design — technology-circuit co-design, optimal bitcell selection for leakage reduction, and architecture-level techniques — to enable an ULP BSN application.

We demonstrated the benefits of 55nm ULL DDC technology to reduce total leakage current for the subthreshold operation (Isub, Ijunc, and Igate) using RBB as a design knob. We proposed body-biasing as a write assist technique for the SRAM to scale V_{MIN} and to improve yield (by 80% with 0.1V of RBB) in subthreshold. Use of ULL for subthreshold circuits with RBB reduces the leakage of the 6T SRAM by 98% and energy/cycle of SRAM by 83%. We also explored limitations in applying a higher degree of RBB that causes a reliability issue such as an increase in DRV and reduced noise margins due to reduced I_{ON} with reverse body biasing. However, use of selective BB (NMOS RBB only) demonstrated significant improvement in reliability (RSNM).

Our implemented design showed 460% leakage power reduction compared to the available similar technology node design [18]. An SRAM array implemented with high-V_T devices limits maximum leakage reduction from the array. Therefore, we explore the possibility of further

power reduction from the peripheral power. To address the half-select failure, every write operation is ensured by a successful read operation of the complete row, and the read data is written back. Although the Read-Before-Write (RBW) addresses the half-select problem for the subthreshold operation in the SRAM, it increases the power dissipation and limits the maximum write frequency. In the absence of an SA, the read is detected by a full-swing discharge of the read-bitline (RBL). Addressing these two limitations will enable sub-micro watt BSNs application with increased power/energy efficiency.

With the bitcell evaluation study, we addressed the reliability and energy challenge of an SRAM targeted for IoT applications with the transistor threshold voltage as a design knob. Six different bitcells are evaluated across process corners and temperatures, and the trade-offs between the different metrics were studied. This work highlighted an in-depth study of the effect of variations on the different static metrics required for low power applications.

Finally, we fabricated a 1KB SRAM chip fabricated in 130nm CMOS that operates reliably between 350mV and 700mV for ULP subthreshold operation for the targeted BSN application. A read-before-write approach is implemented to address half-select instability. A read burst mode is implemented to reduce read energy when consecutive addresses are accessed and saves 22% active read energy. Aggressive power gating reduces the power consumption down to 12.29nW with retention and 1.09nW when data is not needed (at the data retention voltage: 320mV). Compared to the state-of-the-art ULP SRAMs, the proposed design gives the lowest full array leakage power per bit at 1.5pW/bit for an 8T bit-cell array.

Chapter 4

SRAM V_{MIN} Tracking and Selection of an Energy Optimal Assist Technique

4.1   Introduction

Since battery-powered or energy harvested Internet of Things (IoT) devices operate over a wide range of frequencies (i.e., around 10kHz to 10MHz) [28, 48, 49], there is a need for a reliable SRAM operating over a wide range of supply voltage and frequency. While 8T bitcell requires 10-15% area overhead [30], we consider a conventional 6T bitcell in this chapter and address reliability challenges at lower operating voltages while performance enhancement at nominal V_{DD}. Reducing minimum operating voltage (V_{MIN}) for an SRAM has become increasingly challenging under various parametric variation, as discussed in Chapter 3. With an increasing need for devices operating at lower supply voltages (V_{DD}), a reliable operation of an SRAM is challenging because of the variation having an exponential dependency on I_{ON}. Margining a design for the worst case condition leads to higher operating voltages relative to the typical case, leading to higher energy. In this example, margining for the worst case write

corner (SF — Slow NMOS; Fast PMOS) results in a 220mV increase in V_{MIN} relative to the typical corner (TT — Typical NMOS; Typical PMOS) (Figure 2.5). Because the circuit is not always operating in the worst case process, voltage, and temperature (PVT) condition, there is a potential to regain some of this lost energy. Notably, targeted IoT applications where a reliable SRAM operation is sought over a wide range of supply voltage/frequency operations. Therefore, a system adapting based on failure prediction across different PVT conditions can widely enable the energy-constrained IoT platform.

To allow a failure detection in an SRAM bitcell before the actual failure, Canary bitcell[1] has been proposed in [50, 51]. To help an SRAM to perform read/write operations, "supportive" circuits are used. These circuits are known as peripheral assist techniques. The assist techniques that help the write operation are known as 'Write Assist' (WA) techniques while the assist techniques ensuring a successful read are called 'Read Assist' (RA) techniques. In this work, we present an adaptive, closed loop memory system that leverages combinations of bias-based peripheral assists for both read and write to expand the operating range of a 256kb 6T SRAM to cover from 1.2V down to 0.38V.

Assists can be applied in reverse to tune Canary bitcells that allow a closed loop control of the V_{DD} to track the minimum operating voltage (V_{MIN}) at a desired operating frequency. The design uses peripheral assist techniques together with Canary-based V_{MIN} tracking to maximize the operating range that is compatible with the subthreshold logic as 6T SRAM usually has higher V_{MIN} than logic circuits under PVT variations [52–55] and to minimize guard-banding. The design is thereby optimized for meeting the low power, and varying frequency needs of highly variable Internet of Everything (IoE) applications while retaining the density of 6T cells. Bias-based assist techniques can lower SRAM V_{MIN} [52, 53, 55], but the selection of an optimal assist technique changes with V_{DD} and can affect the power/performance trade-off. Therefore, the chapter first evaluates an optimal assist technique at a given operating voltage. Later, we implement a digital controller to

---

[1]An additional bitcell placed along with the SRAM array that has similar functional characteristics with weakening write and read control to SRAM bitcell

selectively apply the optimal assist technique to achieve optimal energy-per-operation.

4.2 Research Approach

We divide our approach to address the challenge of a wide range frequency operation in two parts.

1. Improving reliability and Energy Requirements of subthreshold SRAM using Assist Techniques:

   Different write assist techniques can improve the writeability of an 8T SRAM memory based on operating V_{DD}. The assist techniques proven to be an optimal choice in super-threshold cannot be directly applied to subthreshold because of the change in the optimal metric selection for assist consideration. For example, the performance is the metric of assist evaluation at super-threshold voltages while reliability is the metric for subthreshold voltages. Therefore, evaluating different assist techniques in the subthreshold region can provide various design trade-offs and hence allow better design decisions. Similar to the reliability metric, improvement in speed using assist can reduce the energy-per-operation.

   Therefore, we characterize four different peripheral write assist techniques for the subthreshold operation across different metrics such as total array write energy, write stability, and achievable lowest possible operating V_{DD} (i.e., V_{MIN}). With this approach, we aim to provide strategies for write assist selection based on system need to reduce system V_{MIN} or SRAM energy-per-operation while guaranteeing reliable functionality.

2. Adaptively tunable wide range SRAM system for a wide operating IoE applications:

   After evaluating the optimal choice of assist technique for the subthreshold operation, we propose a smart and efficient way of controlling different assist techniques dynamically (at runtime) that can allow an SRAM to operate across a wide range of V_{DD}s with

targeted metrics such as low power (lowering an SRAM V_{MIN}), low energy (finding energy optimal V_{DD}) or higher performance (operating at maximum performance) under PVT variations. With the use of such smart controller, we aim to achieve a self-tune SRAM system operating over a wide range of frequencies.

4.3 Improving Reliability and Energy Requirements of subthreshold SRAM using Assist Techniques

[2] In Chapter 2, we discuss the theory of four different write assist (WA) techniques. In [56], the authors show the possibility of scaling V_{DD} using an 8T cell compared to a 6T cell as having the write port independent of the read port that eliminates read disturb faults (RDF) [57]. However, as we showed in Chapter 3, the write failure still limits the SRAM V_{DD} scaling for an 8T bitcell. Different write assist techniques like V_{DD} lowering, V_{SS} raising, Wordline (WL) boosting, and negative bitline (NegBL) can improve write functionality and allow reliable write operations at lower V_{DD}s. Since scaling the supply further increases leakage that accumulates over the much longer SRAM access times [58], optimizing an SRAM is not straightforward at low V_{DD}. This trend clearly highlights the challenge of finding the energy optimal supply voltage for SRAM arrays with the applied assist technique to enable energy-constrained systems to operate at scaled voltages.

The previously published work [2, 21, 23] evaluated assist trend for a super-threshold range that cannot be applied to the low-power system operating in subthreshold. Also, an evaluation of assist techniques must consider the impact on the energy consumption of the full SRAM array and not just the bitcell. Therefore, we show the difference in the evaluation metrics' trends of write assist techniques between subthreshold and super-threshold regions.

---

[2]This section is based on the published papers [42]([HNP9]), [HNP10] and [HNP11]

We also address how factors outside the accessed bitcells, such as half select limitations, can dictate the best solution. Our results show that for energy-constraint systems operating at subthreshold voltage, the choice of V_{DD} and write assist combination is very different than that for super-threshold high-performance systems where assist techniques were previously studied.

4.3.1 Reliability Improvement

As discussed, an energy-constrained application operating at subthreshold faces a major challenge of a reliable operation due to reduced noise margins. Introducing write assist techniques at lower V_{DD}s can improve the functional yield. To quantify the impact of each assist technique on the reliability of the write operation, we use the WM metric. The WM is calculated using the Word-Line (WL) criterion defined in [59]. The V_{DD} and the degree of applied assist are used as design knobs to evaluate assist techniques for energy optimal SRAM design. To provide a fair comparison between the different assist techniques at different supply voltages, we apply a bias voltage (e.g., Δ) equal to a percentage of the supply voltage: 10, 20, 30, and 40% for each assist technique. Sweeping the degree to which each assist is applied helps reveal the lower limit of V_{MIN} and the minimum total array energy. Table 4.1 presents the mapping of percentage based assist application to the absolute biasing voltage with V_{DD}=0.5V as an example.

To decide the allowable degree of assist to achieve lower V_{DD}, we explore the upper and lower bounds for the design knobs (supply voltages and applied degree of assist) for different assist techniques using WM as the evaluation metric. We used TASE [60] for a commercial 130nm node to calculate the WM. To consider the impact of local variation, we determine the worst case WM captured by a 1000 point Monte Carlo simulation at 27°C. Figure 4.1 shows the trends of WM across supply voltages for different assist techniques with 10%, 20%, 30%, and 40% of applied assist. With no assist, SRAM V_{DD} can only be scaled down to 0.7V

with a reliable write operation. Increasing the percentage (30% or more) of applied assist can bring the write V_{MIN} down to 0.3V.

Table 4.1: Applied percentage of assist vs. biasing voltage mapping

Assist Techniques	Applied Percentage Assist (V_{DD}=0.5V case)				
	0%	10%	20%	30%	40%
V_{DD} Lowering	V_{DD} = 0.5V	0.45V	0.4V	0.35V	0.3V
V_{SS} Raising	V_{SS} = 0V	0.05V	0.1V	0.15V	0.2V
WL Boosting	WL = 0.5V	0.55V	0.6V	0.65V	0.7V
NegBL	BL/BLB = 0V	-0.05V	-0.01V	-0.15V	-0.2V
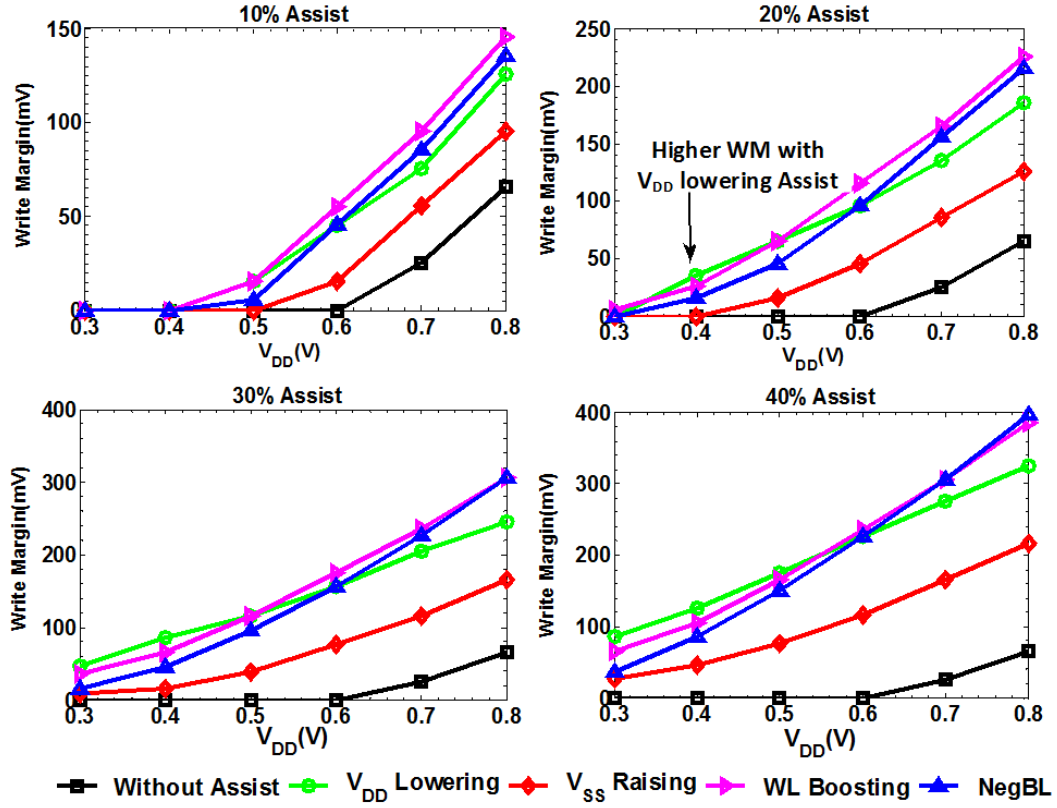
Figure 4.1: Impact of different assist techniques: WM metric

This work shows for the first time that the assist that provides the best WM changes depending on the operating voltage and the degree to which the assist is applied. WL boosting provides the largest improvement in WM among the assist techniques for supply voltages above 0.4V except when 40% of NegBL is applied at V_{DD}=0.8V, where NegBL

shows slightly higher WM. For supply voltages below 0.5V, minimizing V_{DD} gives the highest WM. This change is due to the change in the sensitivity of WM to changes in the threshold voltages of PMOS pull-up (PU) and NMOS pass-gate (PG) devices when V_{DD} is lowered. Weakening the PUs through V_{DD} lowering has a greater impact than making the PGs stronger using WL boosting at low V_{DD}s. The trend shows that the optimal assist technique changes according to the region of operation the subthreshold ($V_{DD} \leq 0.5$V) versus the super-threshold ($V_{DD} \geq 0.6$V). While WL boosting and NegBL are the most reliable at super-threshold voltages, V_{DD} lowering becomes the optimal assist at subthreshold voltages. These conclusions complement the results shown in [22, 23] for the super-threshold designs. Therefore, for low-power applications such as BSNs, the choice of assist technique changes widely for a given supply voltage and degree of applied assist(i.e., higher write/read stability of the bitcell).

4.3.1.1 HSNM degradation due to applied assist techniques

Since the WM metric does not consider the impact of assist techniques on the Hold/Read Static Noise Margin (HSNM/RSNM) of column and row half selected cells in the array, it cannot be used as the sole measure of assist success. Therefore, we study the impact of write assist techniques on the half-selected (HS) cell stability by looking at the HSNM and RSNM of those cells.

Figure 4.2 shows an NxM array with half-selected cells (Rn and Cm) in the same column and row as the write assisted cell (Selected Cell/Word). During the write operation, the WL of the accessed cell will be high. At the same time, the cells with BL=BLB and WL=1 or BL≠BLB and WL=0 are considered as row and column half-selected (HS) cells, respectively. As shown in Figure 4.2, the column HS cells (R1 to Rn) experience the impact of column-based assist techniques (V_{DD} lowering and NegBL), while the row HS cells (C1 to Cm) experience the impact of row-based assist techniques (V_{SS} raising and WL boosting). The hold margin of the column HS cells (Rn) is defined as "Write HSNM." Similarly, the read margin for the

HS cells (Cm) is defined as "Read HSNM." In Figure 4.3, we capture Write HSNM and Read HSNM (RSNM) at different V_{DD}s and write assist degrees. Here, the maximum of the HS V_{MIN} and write V_{MIN} limits the array V_{MIN}.
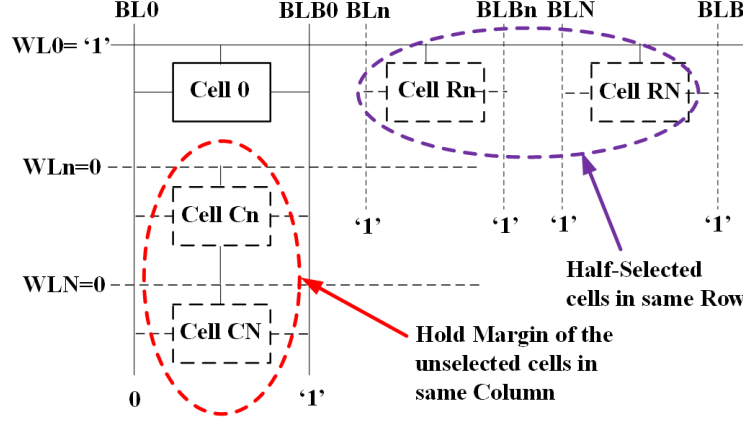
Figure 4.2: Representation of an NxM array with row and column half-selected (HS) cells
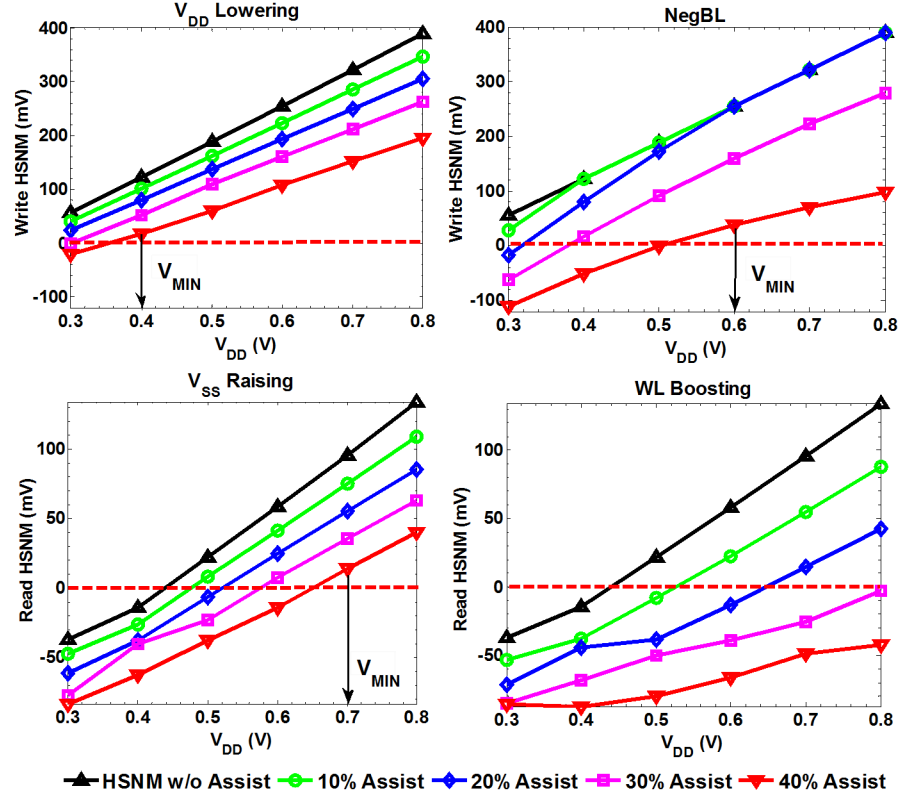
Figure 4.3: Impact of aggressive assist on Read and Write HSNM of HS cells

With V_{DD} lowering, the DRV of column HS cells constrains the degree of V_{DD} lowering assist and limits the array V_{MIN} to 40% at 0.4V. However, the write V_{MIN} (Figure 4.1) is

lowered to 0.3V. Similarly, NegBL increases the V_{GS} differential of column HS cells, limiting the array V_{MIN} to 0.6V for 40% applied NegBL. WL boosting degrades the RSNM of the row HS cells, limiting the degree of applied assist to 10% and the V_{MIN} to 0.6V. Increasing the degree of applied WL boosting raises the HS V_{MIN} above the write V_{MIN}. Similarly, V_{SS} raising, another row-based assist, faces RSNM degradation of row HS cells that limit the array V_{MIN} to 0.7V when 40% assist is applied. Figure 4.1 shows how aggressive assist application can achieve the lowest write V_{MIN} (down to 0.3V), while Figure 4.3 shows the limit on how much assist can be applied due to HS HSNM/RSNM degradation.

Table 4.2: Achievable V_{MIN} for different assist techniques

Assist Technique	Write V_{MIN} at 40% assist (V)	Full Array V_{MIN} at 40% assist (V) (limiting factor)	Lowest Array V_{MIN} (V) (achievable degree of assist)
V_{DD} Lowering	0.3	0.5 (RSNM)	$V_{MIN} = 0.5V$ (10-40%)
V_{SS} Raising	0.3	0.7 (Write-HSNM)	$V_{MIN} = 0.6V$ (10-30%)
WL Boosting	0.3	>0.8 (RSNM)	$V_{MIN} = 0.6V$ (10%)
NegBL	0.3	0.6 (Write-HSNM)	$V_{MIN} = 0.5V$ (10-30%)

Table 4.2 summarizes the combined results from Figure 4.1 and Figure 4.3. For all assists, the write V_{MIN} reaches 0.3V with 40% assist application. However, due to the HS failures,

Table 4.3: Range of V_{DD}s and applied assist for a reliable SRAM operation for different assist techniques

Supply Voltage (V)	Achievable range of applied assist(% of supply)			
	V_{DD} Lowering	V_{SS} Raising	WL Boosting	NegBL
0.5	10 – 40%	X	X	10 – 30%
0.6	10 – 40%	10 -30%	10%	10 – 40%
0.7	0 – 40%	0- 40%	0 – 20%	0 – 40%
0.8	0 – 40%	0 -40%	0 – 20%	0 – 40%

using this degree of assist limits the array level V_{MIN} to a higher value for every assist type. The rightmost column shows the lowest achievable array level V_{MIN} for each assist technique, considering both the WM of the accessed cells and the HSNM/RSNM of HS cells. Table 4.3 summarize the range of V_{DD} where SRAM array can operate reliably with the required degree of assist for different assist techniques.

4.3.2   Energy Consideration

After evaluating the reliability limits on the array V_{MIN} for subthreshold operation, it is important to explore how the array energy dissipation can be minimized using assist techniques to design energy efficient SRAMs for BSN systems. The SoCs can use two types of power domain implementation for different blocks (Figure 4.4): a shared supply voltage between all blocks or split supply voltages between SRAMs and other blocks. For a shared supply, minimizing SRAM array energy may or may not minimize the full system energy, since other components outside the memory may dominate the system power. In this case, minimizing the SRAM V_{MIN} extends the operating range of SRAM and allows more flexibility for co-optimizing SRAM with a non-SRAM block for the system energy minimization. Analysis from the last section shows the V_{DD} range using assist, and we will continue in this section to account for energy consumption in SRAM at low V_{DD}, even if the SRAM energy is not minimized. This provides data for computing system energy in the shared V_{DD} case.

For split supply implementation, reducing the SRAM energy (not necessarily V_{MIN}) will reduce the system energy, since the SRAM is on an independent V_{DD}. Hence, the choice for
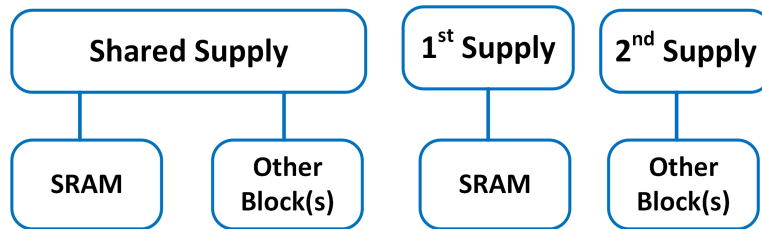
Figure 4.4: System level supply voltage configuration: I) Shared supply II) Split supply

the best assist technique to minimize system energy depends on the voltage allocation method. In this study, we calculate the energy for an array of 1KB (64x128), which can represent a repeatable block for larger capacity memories in present SoCs. The energy is calculated as the Power-Delay-Product (PDP) across supply voltages for different assist techniques using TASE [60]. The write delay for a successful write is measured as the time between 50% of the WL pulse to Q/QB reaching 90% of V_{DD}. Array level power accuracy is achieved by including the WL driver (without decoder), pre-charge circuit, and data driver circuits driving the data to and from 64x128 SRAM array. We divide total array energy into five components as:

Array energy components (write operation):

$$E_{Array} = E_{BL} + E_{WL} + E_{CF} + E_{HS} + E_{Leak} \tag{4.1}$$

Where

E_{BL} = Selected cell BL driver/pre-charge energy with parasitic on BL,

E_{WL} = WL driver energy with parasitic on WL,

E_{CF} = Energy drawn from the cell V_{DD} during the storage node flipping,

E_{HS} = Energy dissipated in HS cells (dummy read through HS cells),

E_{Leak} = Standby cell leakage energy.

We use Equation (4.1) for the write energy to find how different components of the array contribute to the total energy for different assist techniques. Figure 4.5 shows an array level energy break down during a write operation for various assist techniques at the lowest achievable array V_{MIN} (Table 4.2) as a reference. The purpose of this figure is to show the energy dissipation for different assist techniques. The data shows that the total write energy is dominated by half-select energy (E_{HS}) for all the different assist techniques. The contribution of HS energy across supply voltages shows a similar trend. The contribution from E_{WL} is comparatively low for all the assist techniques, even with WL boosting. With V_{DD} lowering and WL boosting assist techniques, E_{CF} is lower than E_{BL} and E_{HS}, while

V_{SS} raising and NegBL show a significant contribution from E_{CF}. This is because of the higher steady-state leakage current from the cell to the BL due to the applied assist biasing.
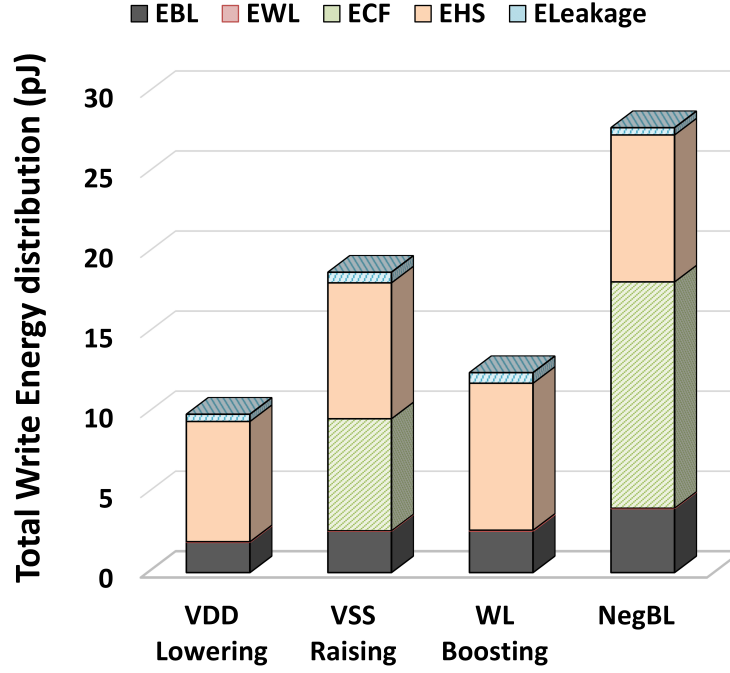
Figure 4.5: Total write energy distribution across assist techniques (at array V_{MIN})

4.3.3 Results

We evaluate the effectiveness of write assist techniques for improving robustness of write functionality and its impact on total energy as 1) WM vs. total write energy (i.e., the total energy consumed by an array during a reliable write operation, i.e., higher WM), and 2) total write energy vs. write delay (i.e., the energy dissipated to achieve a given performance requirement from the system (energy/op)) based on the data summarized in Table 4.3.

4.3.3.1 Reliability – Energy Trade-off

After evaluating the upper and lower bounds of the design knobs (supply voltages and applied assist), we assess how V_{DD} and the degree of assist application affect the write reliability (WM) versus total energy trade-off in this section. For the total write energy calculation, we

use Equation (4.1). To find the optimal V_{DD}, we define the Energy Optimal Contour (EOC) as the Pareto curve showing the best WM vs. write energy for the available design knobs. Table 4.3 provides the degree of assist that can be applied at given supply voltage for the considered assist technique (from Figure 4.1 and Figure 4.3). Figure 4.6 shows the EOCs of different assist techniques connecting WM-energy optimal points. With no assist, write V_{MIN} is limited to 0.7V as further scaling causes write failures. By looking at the total write energy number, the non-assisted array can also provide an energy optimal choice but with higher V_{MIN} requirement.

For the V_{DD} lowering assist, 40% assist can be applied to achieve a V_{MIN} of 0.5V that also provides the optimal energy for a given WM at a corresponding V_{DD}. For V_{SS} raising, increasing the applied assist results in higher write energy. As mentioned before, an increase in the applied assist bias ($V_{SS} + \Delta$) causes higher leakage if WL is kept high after the internal nodes flip. This leakage can be reduced by tightly controlling the WL pulse width to track the write delay that is sufficient for the flipping the cell. Therefore for V_{SS} raising, E_{CF} and E_{HS} increase with aggressive assist application causing total write energy to increase. For
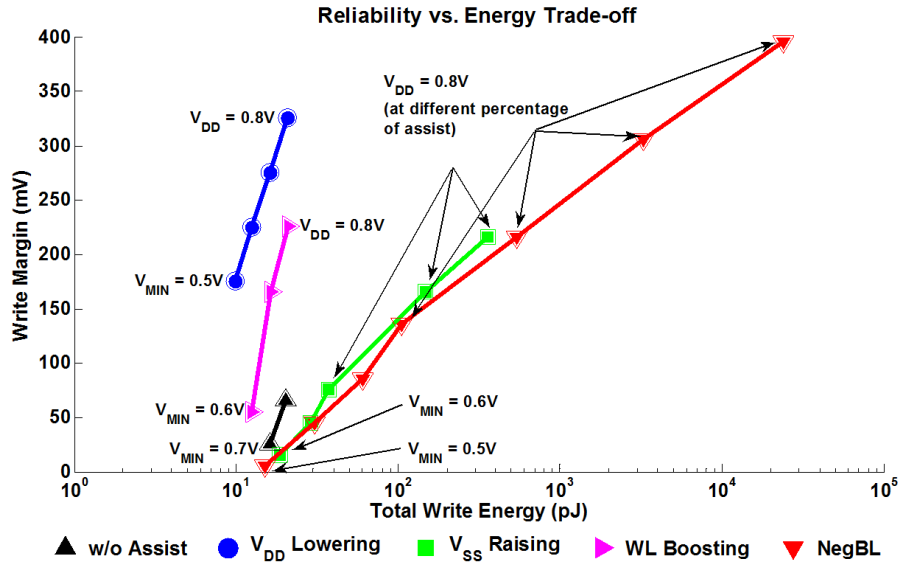
Figure 4.6: WM (reliability) vs. total write energy optimal contours (EOCs) for different assist techniques at achievable V_{MIN}

WL boosting, WL driver energy increases with increased applied assist, but it still represents a very small fraction of the total energy. As WL boosting is limited by RSNM of row HS cells, its achievable V_{MIN} is higher (0.6V). Similar to V_{SS} raising, write energy with NegBL is also constrained by aggressive degrees of assist. In addition to E_{BL}, E_{CF} also goes high due to excess leakage when the cell flips and WL is kept high. With NegBL, an array V_{MIN} of 0.5V can be achieved with 10% applied assist.

Figure 4.6 suggests different strategies for trading off energy, margin, and V_{MIN} for a given system level energy optimal V_{DD}. For shared V_{DD} systems where reducing V_{MIN} is paramount, V_{DD} lowering and NegBL can offer the optimal solution (V_{MIN}=0.5V) with different WM (write stability), and energy trade offs. To minimize energy at higher supply voltages, WL boosting provides the energy optimal choice with only 10% of assist applied compared to V_{DD} lowering, which requires higher assist to achieve the same energy reduction and write margin.

4.3.3.2 Performance – Energy Trade-off

To evaluate the energy requirement for a required performance, we explore the total write energy versus write delay (performance). Figure 4.7 shows the optimal energy for an achievable performance (speed) at a given V_{DD} and applied assist. Even though V_{DD} lowering reduces V_{MIN} down to 0.5V and minimizes the total energy (E_{MIN}), it incurs a significant delay penalty. As highlighted in Figure 4.7, at 0.5V array V_{MIN}, V_{DD} lowering is 121X slower compared to NegBL (at the same array V_{MIN}) and approximately 800X slower than WL boosting at V_{DD} of 0.6V. The delay increases because the reduction in V_{DD} makes the Q/QB node transition from low to high slow. V_{DD} lowering provides an optimal technique for minimizing energy with a marginal difference from NegBL when the delay is not a prime concern. WL boosting provides the optimal energy and speed at 0.6V with 1.5X higher energy and 800X faster compared to the lowest energy point at 0.5V using the V_{DD} lowering assist. Also, WL boosting provides the optimal energy/access choice for systems that run at
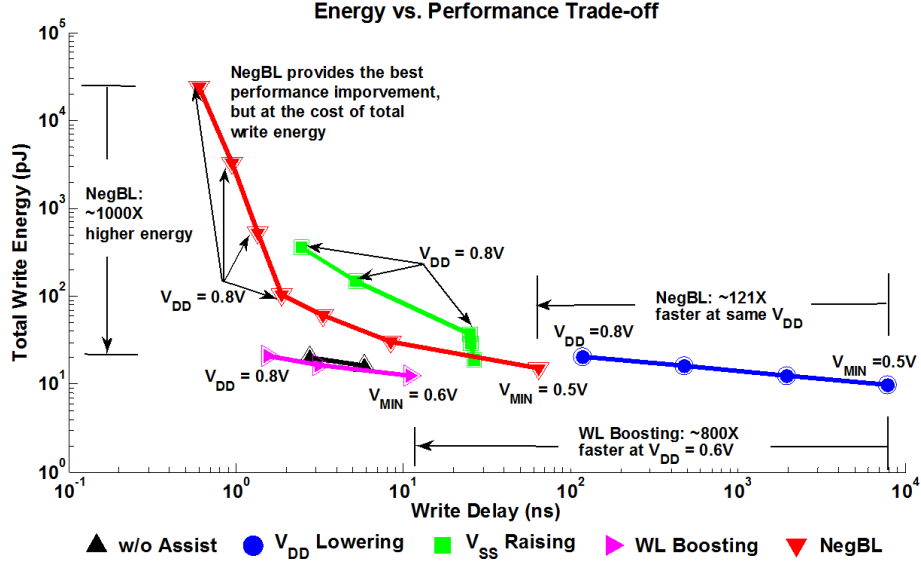
Figure 4.7: Total Write Energy vs. write delay (performance) contours for different assist techniques with achievable array V_{MIN}

higher V_{DD}s (up to 0.8V). On the other hand, NegBL assist provides the largest improvement in delay compared to other assist techniques but incurs an energy penalty due to excessive leakage with the aggressive assist.

The system requirements play a major role in the selection of the appropriate assist technique. Figure 4.8 provides the trade-offs between write stability (reliability), array energy, and write delay (performance) at the lowest achievable array V_{MIN} for each assist techniques. V_{DD} lowering provides an energy optimal solution for ULP application (e.g., BSN) that requires low frequency ($<$10-100's of KHz or below) sensing (e.g., environmental sensing, glucose, SpO_2, skin temperature, etc.) while NegBL provides performance improvement at the same supply but robustness. Figure 4.9 provides the guidelines for choosing the array V_{DD} (V_{MIN}) and assist technique when given an application requirement of E_{MIN}. Depending on the design constraints from the application, users can choose the appropriate assist type, the degree of assist, and V_{DD} based on the information in Figure 4.9.
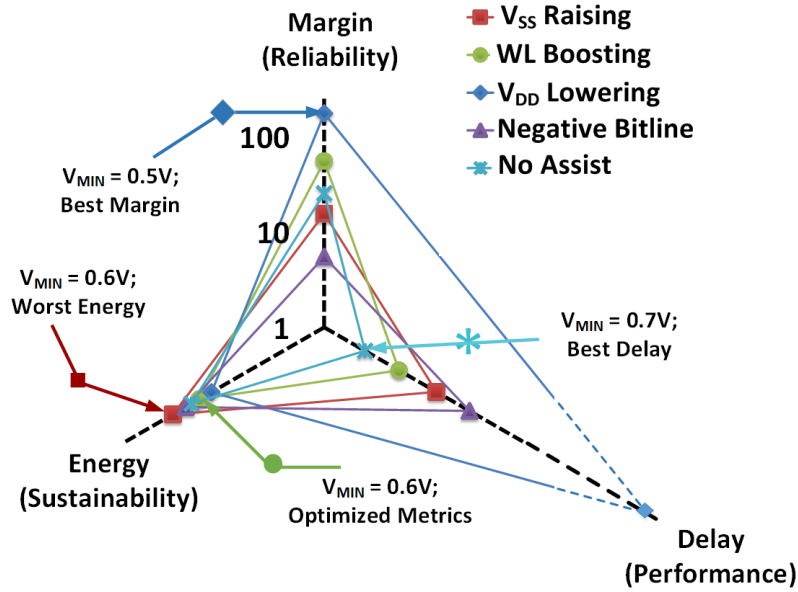
Figure 4.8: Write margin - write delay - array write energy trade-offs at the lowest achievable V_{MIN}
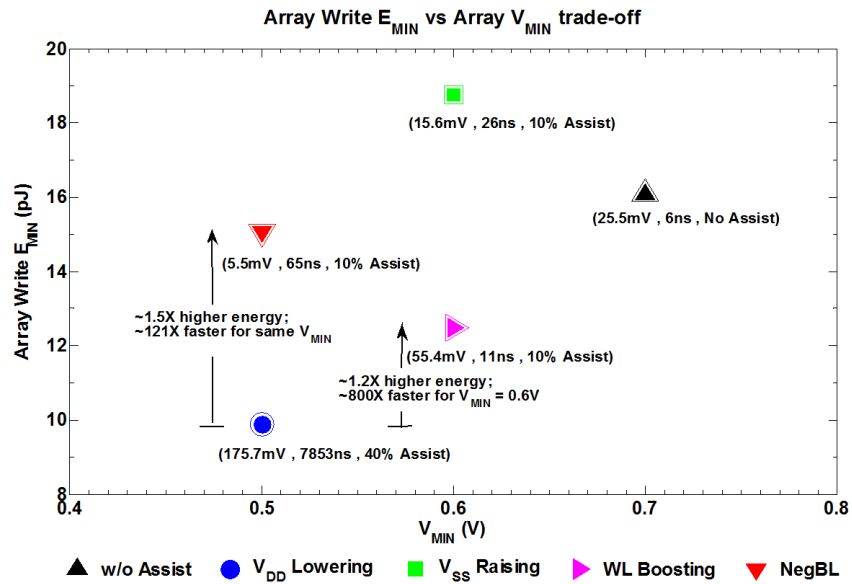
Figure 4.9: Array write E_{MIN} vs. Array V_{MIN} for different assist techniques with achievable (WM, delay, and assist)

4.4 Self-Tunable Wide-Range SRAM using Assist Controller

[3] Previous work distinguishes the optimal assist technique(s) for the sub/near-threshold. In this section, we use the results from the write assist characteristics in sub/near-threshold (0.3V–0.8V) for the high-V_T 6T bitcell. We extend the study to the super-threshold range (0.9V–1.2V) for performance (speed) metrics. Together, we propose a complete SRAM macro that adaptively selects either an optimal single assist technique or a combination of assist techniques and V_{DD} based on the requested system performance. This self-calibration for the change in the frequency is achieved by a Canary sensor that predicts failure mechanism in an SRAM bitcell prior to the actual failures. Figure 4.10 shows the architectural diagram of the complete system with the highlighted Frequency-to-Digital Converter (FDC) and Assist Controller (ASC). The proposed design is validated to operate from 1.2V down to 0.35V with a maximum 250MHz at 1.2V using 130nm technology. The functionality at lower V_{DD} is guaranteed by various write/read assist techniques.

4.4.1 System Architecture

Figure 4.10 shows our full SRAM system comprising a 256kb SRAM in 4 sub-arrays (mats) each with 4 banks of 128x128 6T bitcells and 1 row of 128 Canary bitcells per bank (2kb Canary bitcells total), an assist controller (ASC), a frequency-to-digital converter (FDC), and a built-in self-test (BIST) block for the core SRAM and the Canary bitcells (CBIST). The Canary cells share the peripheral circuits (e.g., write drivers, sense amplifiers, precharge circuits, etc.) with the SRAM array but have dedicated reverse assist (RA) controls [61] that tune write-ability and readability of the canaries by degrading the Canary WL signal using

---

[3]This section is based on the published papers [HNP4]

This work is done in collaboration with Arijit Banerjee and Ningxi Liu
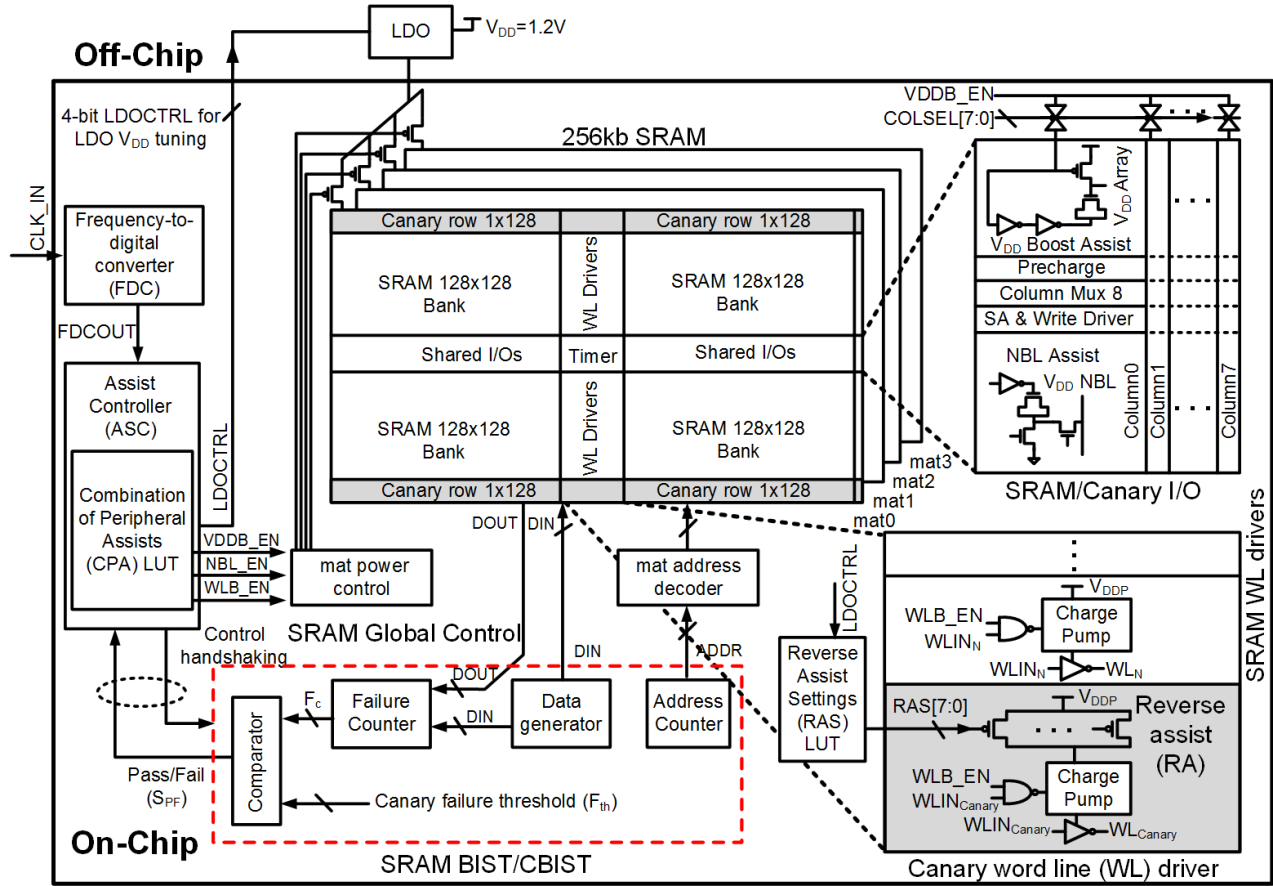
Figure 4.10: Block diagram of an SRAM sub-system with adaptively tunable assist controller and other blocks

eight programmable settings.

Figure 4.11 presents the self-tuning strategy for Canary-based SRAM V_{MIN} tracking and dynamic control over assists, and V_{DD} selection. When tuning is enabled (TRACK = 1), the FDC converts the input clock (CLK_IN frequency to a 16-bit digitized output (FDCOUT) and initializes an (off-chip) Low-Dropout (LDO) regulator to an initial V_{DD} for the given frequency. Then, the ASC chooses an assist configuration for the current V_{DD} from a look-up table (LUT) for flexibly optimizing assist selection based on measured characterization across V_{DD}. The ASC then iterates to find the target V_{MIN} for the given frequency based on the Canary outputs. The CBIST executes Canary write and read operations across all Canary addresses, calculates the number of Canary failures (Fc), then compares Fc with a Canary
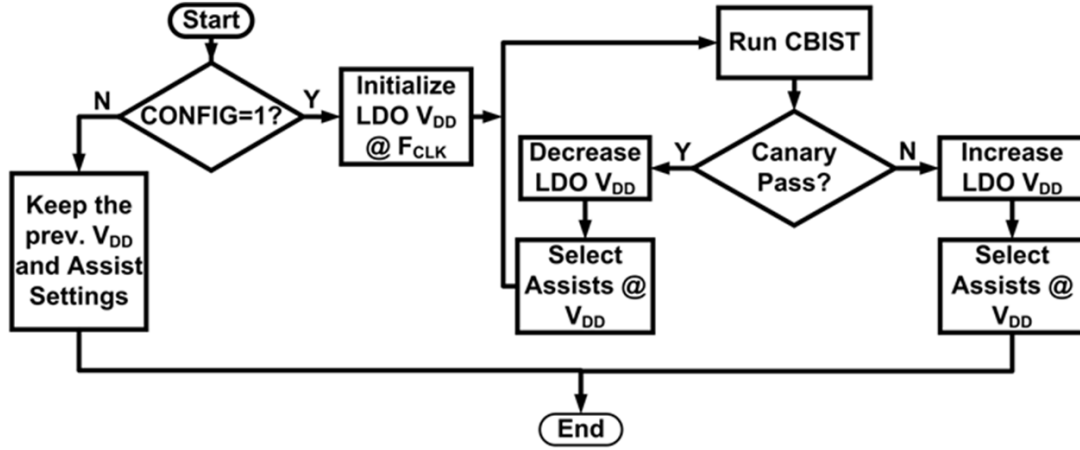
(a) Flowchart of the adaptive system for optimizing assist selection and V_{DD} based on Canary failure detection.
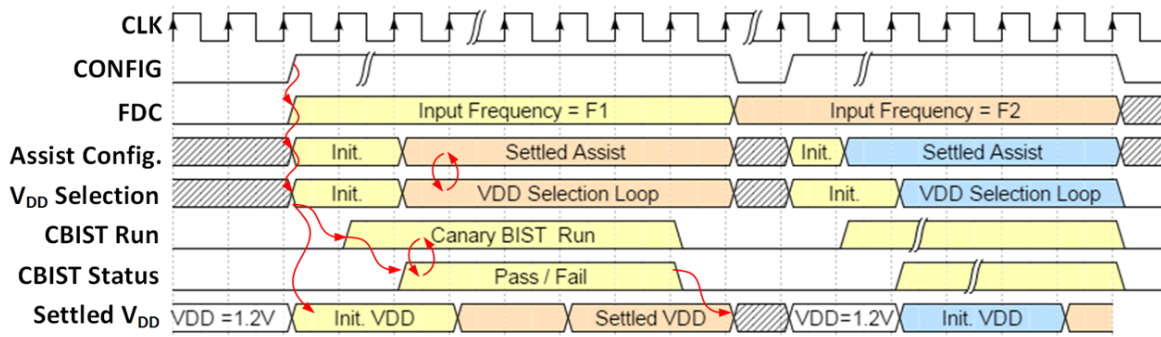
(b) Waveform of an assist controller (ASC) highlighting signal flow based on assist selection decision during the configuration.

Figure 4.11: Flowchart and corresponding system waveforms of the SRAM tracking and assist selection using Canary SRAM.

failure threshold value (Fth) to generate a pass/fail signal (SPF). If the CBIST passes, the ASC reduces V_{DD} by changing a 4-bit signal (LDOCTRL) controlling the off-chip LDO. The ASC repeats this process until the CBIST fails, then it raises V_{DD} to the last operational V_{DD}, which completes the closed-loop tracking for V_{MIN}. The SRAM retains its data through this process, and tuning can be rerun when the frequency changes or to periodically adjust for temperature changes.

4.4.2 Assist Controller

A digital controller, ASC, is designed to implement a Look-up Table (LUT) for the various assist techniques implementation based on simulation to achieve an energy/operation for a wide range of operation. After studying the combined assist effect on V_{MIN} scaling in the previous section, we characterize the possible combinations of assist techniques at super-threshold voltages (0.6V–1.2V) for performance enhancement.
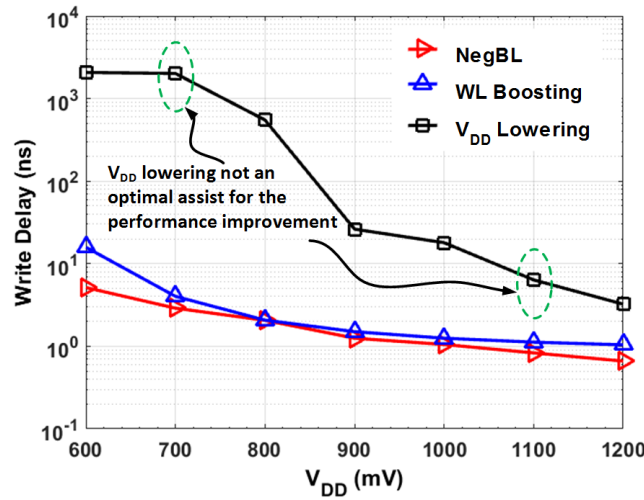
(a) Assist technique evaluation for performance metric: Single assist
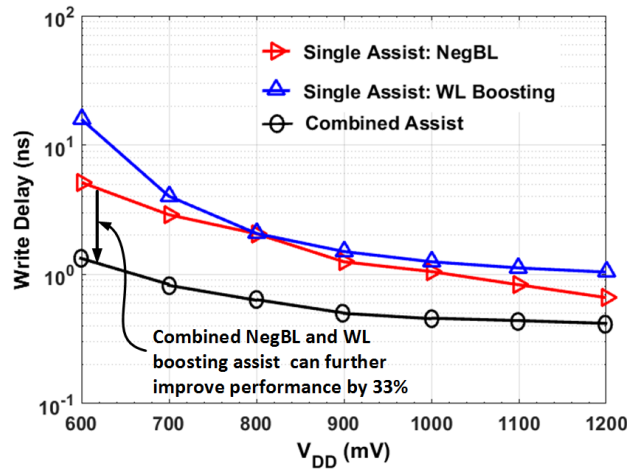
(b) Performance improvement using a combination of write assist techniques

Figure 4.12: Evaluating optimal assist technique for the performance enhancement at super-threshold

Figure 4.12 shows the performance (write delay) improvement using three assist techniques,

shown to be efficient in the previous section, over the super-V_T range of supply voltages. The plot distinguishes WL boosting and NegBL as optimal assist techniques for improving speed. Therefore, the combination of WL boosting and NegBL assist techniques with different degree of assist will increase multifold the performance gain. The results show that there can as much as 30X speed improvement at V_{DD}=0.6V while this performance (speed) benefits reduce as V_{DD} increases from 0.6V to 1.2V. In summary, applying the combination of WL boosting and NegBL at the super-threshold range can increase the performance drastically compared to a single assist. Table 4.4 summarize the V_{DD} tuning algorithm by controlling the LDO inputs and corresponding assist selection to achieve an optimal energy per operation for the SRAM.

Table 4.4: Selection algorithm for the supply voltage tuning (using LDO) and corresponding assist selection.

Frequency requirements	V_{DD} Control	Assist Selection		
		V_{DD} Boosting	WL Boosting	NegBL
High-Perf Application	Increase until reliable operation or max. limit	X	√	√
Low-Power Application	Decrease until failures or min.	√	√	X
Intermediate range of frequency	Increase of decrease until canary failure detection	√* *Only enabled in sub-V_T	√* *Progressive selection to achie performance requirement	

In the configuration mode, based on the requested frequency, ASC initializes the closest V_{DD} and optimal assist techniques where the system can operate without failure. With optimal assist selection, a system can always be pushed for an energy-optimal operation. During a lower performance mode, ASC reduces the V_{DD} by controlling the off-chip LDO inputs. To ensure functionality at reduced V_{DD}, assist techniques are enabled. Once the initial V_{DD} and assist techniques are set, ASC enables BIST in Canary-mode that decides the probability of SRAM failure based on Canary failures. If the pre-set configuration causes

functional failures, the ASC increases the V_{DD} by issuing corresponding signals to the LDO. Here, the change in V_{DD} is achieved by a commercial LDO that provides V_{DD} between 0.4V to 1.2V based on the digital output from the ASC. The settling time (time required by the LDO to provide a stable output voltage based on a change in inputs) of LDO is incorporated by the ASC before enabling the BIST for testing at new set V_{DD}/assist configuration. During functional mode (i.e., not in configuration mode), ASC will be in idle mode while maintaining the previously set configuration (V_{DD} and assist techniques) and implements low-power mode to reduce active power. The ASC keeps updating the configurations until Built-In Self Test (BIST) ensures fault-free operation.

4.4.3   Results

To demonstrate the proposed Canary-based failure detection and assist controller based adaptive tuning, we fabricated a testchip in a commercial 130nm technology. Figure 4.13 shows the die photo of the testchip with various blocks highlighted with the main features of the design. Figure 4.14 shows the measured cumulative distribution functions for the SRAM with three peripheral assists: (1) V_{DD} boosting (VDDB) for low-voltage readability and half-select [52, 55] read-stability; (2) wordline (WL) boosting (WLB); and (3) negative bitline (NBL) for write-ability. Using all the three assists, the proposed architecture improves V_{MIN} by 240mV (at 90th percentile). However, using fewer assists can further save the power overhead when the target V_{DD} is higher for a given frequency requirement. To provide a quantitative measure of a wide operating range, Figure 4.15 shows the measured Shmoo plot with the CPA — extended range highlighted for the 256kb SRAM.
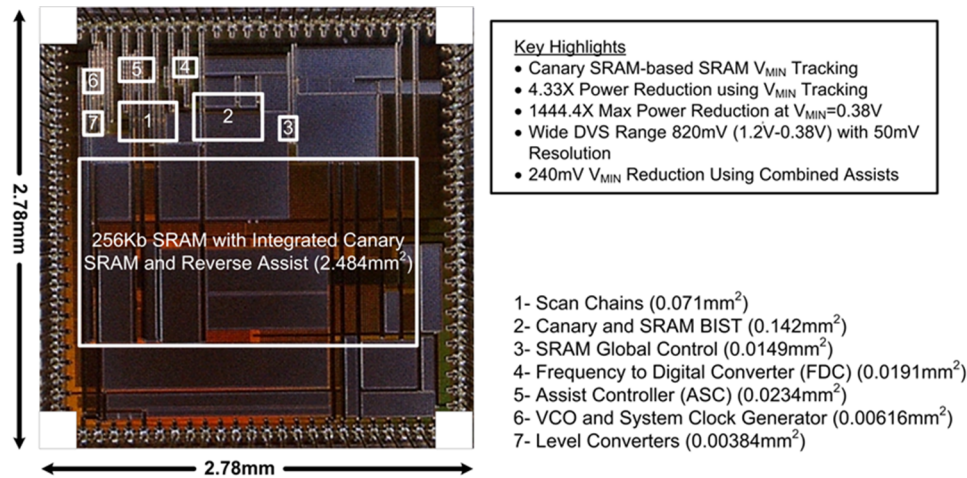
Figure 4.13: Die photo of the fabricated 256kb SRAM sub-system with sub-blocks and various features of the architecture.
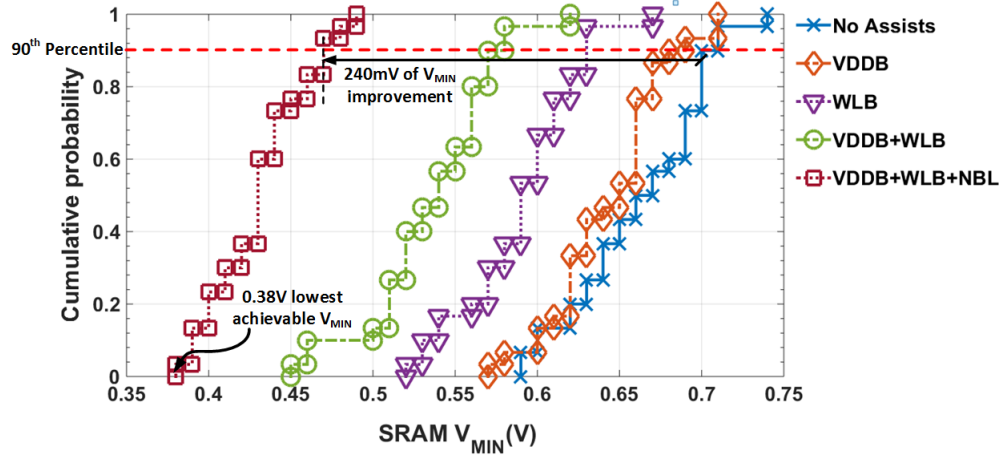
Figure 4.14: Measured CDF of 256kb SRAM V_{MIN} showing 90th percentile V_{MIN} improvement of 240mV using combined assists of VDDB, WLB, and NBL
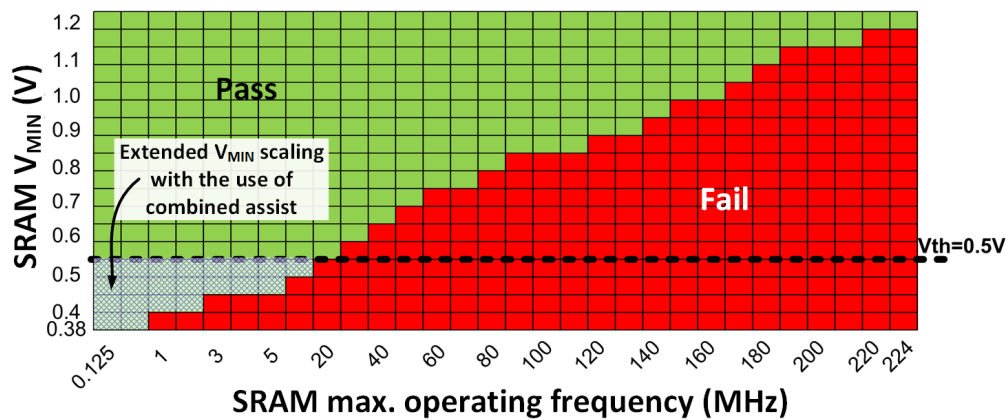
Figure 4.15: Measured Shmoo plot highlighting a wide range of operation over frequency and V_{DD}

4.5   Conclusions

In this chapter, we proposed a self-tuning SRAM architecture to achieve an energy efficient SRAM over a wide range of operation to enable IoT platform. To achieve the low-power and high-performance, we explore various write assist techniques.

First, we evaluated different write assist techniques for minimizing SRAM array write energy to enable energy-constrained applications such as BSN. Based on this study, we conclude:

1. The reliability (WM) trend of write-assist techniques in subthreshold are different from super-threshold. Thus the design decision for assist implementation will be different for given system level constraints in subthreshold.

2. The half-select failures, while not an issue at super-threshold, limit the array V_{MIN} as well as the degree of assist that can be applied to scale the V_{DD}.

3. Lowering V_{DD} provides maximum supply scaling down to 0.5V with the least energy, but negatively impacts write delay.

4. NegBL provides the optimal solution for systems that require higher performance (speed) with only a marginally higher energy at 0.5V compared to lowering V_{DD}, and lower robustness (WM).

5. WL boosting shows the energy per operation optimal solution for a system that has a higher system level V_{DD}. In this work, we provided an in depth system-level evaluation of assist techniques studying the different trade-offs between the array V_{MIN}, array V_{MIN}, write reliability, and performance for reliable and energy-constrained applications.

Later, we consider the results from the assist evaluation study and demonstrate a closed loop self-tuning 256kb 6T SRAM with 0.38V–1.2V extended operating range using combined read and write assists using Canary-based V_{MIN} tracking. The proposed architecture allows

337X power reductions and 3.4X V_{MIN} tracking using multiple assists. A testchip fabricated in 130nm extends the 6T SRAM operating range by over 67% using three combined read/write assists and Canary-based V_{MIN} tracking. The SRAM self-tunes to the V_{MIN} across process and temperature for a target frequency. An adaptive digital assist controller is designed to select an optimal assist based on the frequency of operation required. This adaptive solution enables a range of IoE applications and achieves up to 1444X active power reduction. Table 4.5 compares our proposed architecture with the available state-of-the-art.

Table 4.5: Comparison of proposed Canary-based close loop SRAM sub-system with present state-of-the-art work.

		This Work	ISSCC '10	ISSCC '12	VLSI '14	ISSCC '15
Memory Features	Technology	130nm	45nm	22nm	180nm	28nm
	Cell Type	6T	8T	6T	8T	6T
	Capacity	256Kb	512Kb	576KB	16KB	256Kb
DVS Features	DVS range	1.2-0.38V (850mV)	1.2V-0.57V (630mV)	1V-0.625V (375mV)	1.8V-0.6V (1200mV)	0.9V-0.58V (320mV)
	SRAM VMIN tracking	Y	N	N	N	N
Supply and Power	VMIN	0.38V	0.57V	0.7V	0.6V	0.58V
	Sub-VT Operation	Y	-	-	-	-
	Active Power	18mW @1.2Vand 12.6uW @ 0.38V	169mW @1.2V	-	250mW @ 1.8V, 15mW @0.82V	-
	Power Reduction using regular DVS	234.22X	-	-	16.4X	
	Power Reduction using VMIN Tracking	4.33X	N	N	N	N
	Max Power Reduction	1444.4X	-	-	16.4X	-
	Min Energy/op	81.43pJ/op @1.2V and 25.33pJ/op @0.38V	-	-	3850pJ/op @ 1.8V, 690pJ/op @ 0.82V	-
	Standby Leakage Power/bit	9.5pW/bit	644nW/bit	-	-	-
	Standby Leakage Power Reduction to VMIN	12.39X	9.4X	-	-	-
SRAM Features	Sensor	2Kb Canary SRAM with reverse assist (0.77% area penalty)	DRV Sensor	-	Energy monitor	-
	Combined Assist	Any combination of NBL, WLB, VDDB	N	TVC-WA, WLUD	N	WLUD, NBL
	Adaptive Assist Selection	Y	N	N	N	N

Chapter 5

Reliability at Ultra-Low Voltage Operation

5.1   Introduction

Aggressive technology scaling over the past decade and more to double the performance demand has been followed Moore's law successfully. However, the trend of linear technology scaling significantly reduced the reliability of the digital system. Few of the significant sources affecting the reliability of the todays digital designs are: variations in process, voltage, and temperature (PVT) increased the sensitivity of high-energy particle strikes, electrostatic discharge, and many others. Additionally, the growing demand for ubiquitous sensing of various health and environmental parameters require an ultra-low power (ULP) platform with an energy-efficient operation as one of the few essential requirements. The supply voltage (V_{DD}) is scaled below the threshold voltage (V_T) to optimize energy per operation ($CV_{DD}^2 f$) [62, 63]. However, an exponential dependence on device V_T results in functional failures [24, 30] with a minute variation in PVT. Figure 5.1 shows the increase in failure probability in write operation between the typical corner and the worst case process corner. Therefore, the subthreshold operations ($V_{DD} < V_T$) present an entirely new set of design

reliability challenge in the presence of PVT variations and another reliability aspect such as Soft-errors. An over-margined design approach may address many reliability concerns such as PVT variation and reduced noise margins even at the cost of area and energy. However, the transient change in the node value due to high-energy particle strike has a complex relationship with an over-designed approach and decrease in reliability.
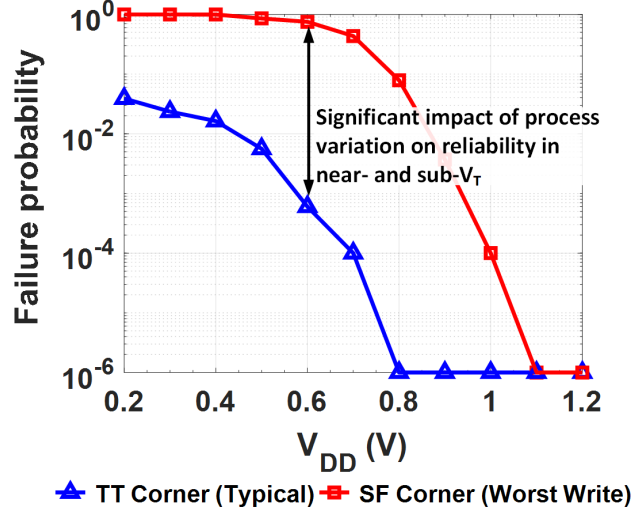
Figure 5.1: Impact of process variation on reliability with V_{DD} scaling.

The circuit operating at low-voltages also experiences reliability challenges due to the particle strike from the atmosphere. The authors in [64] highlighted the severity of the device scaling below 1μm length by showing the short circuit between drain and source due to a single particle strike. With excessive V_{DD} scaling to the subthreshold voltage range, the circuit node charge responsible for holding the state reduces resulting in a smaller particle strike causing the flip of the logic state. Likewise, the soft-errors also increases significantly with altitude from the sea level. Figure 5.2 shows the increasing impact of radiation. The figure exaggerates the scope and the potential risk of device failure modes due to soft-errors based on the location and usage. Authors in [5, 65] presented a comprehensive analysis of terrestrial cosmic radiation as a function of the altitude and places. With growing demand of bio-sensing and other Body Sensor Network (BSN) applications; those require an extremely low power, the reliability issue due to soft-error in subthreshold become a very critical problem
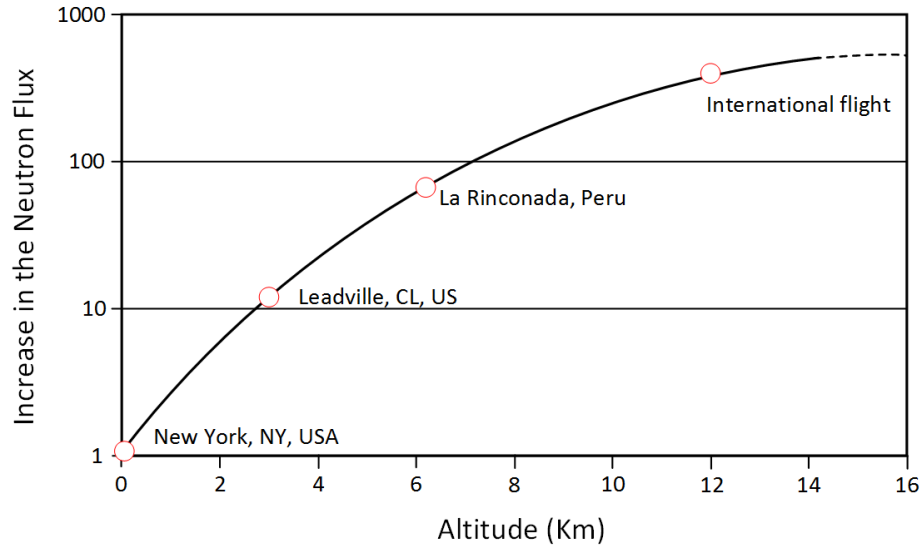
Figure 5.2: Increase in Neutron Flux increase with altitude. Highest altitude place in the earth (La Rinconada, Peru with ~6Km) experiences ~100X higher flux while the international flight achieving 39000ft (~12Km) has ~500X greater risk of particle strike then New York City, NY, USA. [5]

to be addressed. The magnitude of the impact on circuits due to radiation varies from a temporal change in the storage to a complete application failure [6]. The soft-error occurs when a particle with enough charge reverse the state. Based on the nature of the error being temporal, not a permanent failure, circuit functional disruption, it is called as *Soft-Error*.

In this chapter, we address both, the challenge of PVT variability and also study the impact of particle strike, on the SRAM in ULP operations. First, we address PVT variation challenges using various controls using a ring oscillator (RO)-based process corner and V_{DD} droop detection: 1) body-biasing, 2) adaptive peripheral assist technique selection for an SRAM, and 3) droop detection and mitigation. We fabricated a testchip with a PVT controller demonstrating detection of the process corner and different control signal generation in a 130nm technology. The proposed controller occupied only $5400\mu m^2$ and dissipated no noticeable power compared to the traditional SRAM power requirements. Later, we study the impact of the radiation-induced soft errors in ULP applications operating at subthreshold voltages and compared the same trend under the technology scaling.

5.2   Research Approach

To address the reliability of subthreshold SRAM, we divide our approach into two categories.

1. Process-Voltage-Temperature compensated SRAM using a smart digital controller:

   To mitigate or adapt to PVT variation, it is important to detect the specific process corner, change in operating temperature, and droop in the supply voltage rail. We will explore the various effect of process and temperature on different metrics of the SRAM such as stability, power, and performance. We will use the results of this exploration to design a real-time smart controller taking inputs from process and temperature sensors. The controller then will provide various control signals to SRAM to adjust according to the changes in external parameters.

   Similar to the process and temperature effect on stability aspect of design, we analyze the leakage power and performance(delay) metrics for the robust design. Based on the available results from static, power, and performance metrics, we propose a controller that can provide corrective action for the change in external factors (process and temperature). This controller is aimed to add negligible area overhead while providing a significant contribution to the system level robustness. In summary, controlling the process and temperature using a smart controller result in both, reliability and energy, efficient SRAM design.

2. The effect of soft-errors in subthreshold SRAM design:

   Since the first report of cosmic ray caused failure in 1975 in space applications, much research has been done, from device structure to different protection schemes [5, 65, 66]. In this work, we would like to explore and address different challenge for the radiation-induced soft error in ULP applications targeting subthreshold voltage operation. The impact of technology scaling has been well-studied, the exploration of soft-error on low-voltage operation is still an explored area. Therefore, we will perform a study of

the effect of the particle strike on a low-power SRAM operating in the subthreshold region. We will also compare the impact of technology scaling versus supply scaling on particle strike to understand the behavior under the technology and supply scaling.

5.3 Adapting PVT Variation using Digital Controller

[1] In [67, 68], the authors explored the impact of variability on performance in a subthreshold design. Body-biasing and device sizing have been proposed to mitigate variation in subthreshold. Previously, variations have been addressed by increasing device sizes and logic depth to mitigate timing failures [68, 69]. However, a significant increase in device sizes results in an energy and area overhead. Hence, sizing is unfavorable in energy-constrained ULP applications. The impact of PVT variations on various design components is different. For example, a digital design in a synchronous system faces challenges meeting timing requirements (i.e., timing yield). Timing variation in digital designs can potentially cause ∼20% yield loss [70]. SRAM leakage power varies by ∼25% (power yield loss) under the parametric variation across process corners [71]. Device aging-related reliability issues due to variations such as negative bias temperature instability (NBTI), hot carrier injection (HCI), and time-dependent dielectric breakdown (TDDB) are reduced by lowering gate-to-source voltage (V_{GS}) in subthreshold operation. However, the reduced I_{ON}-to-I_{OFF} of devices makes ratioed circuits such as SRAM functionally unreliable. An over-margined design approach to address the effects of variation results in ∼20% area penalty [72].

Many compensation techniques are proposed to address the PVT variation. The authors in [73] proposed a data randomization scheme for a subthreshold SRAM to reduce bitline swing degradation against PVT variations. In [74], the authors present a digital controller with static body-bias control generation schemes for an SRAM array and peripherals to

---

[1]This section is based on the publication [HNP11] and [HNP13]

optimize the leakage and improve the performance. This work addresses PVT variation challenges by dynamically controlling three design parameters using a ring oscillator (RO) based process corner and V_{DD} droop detection: 1) body-biasing, 2) adaptive peripheral assist technique selection for an SRAM, and 3) droop detection and mitigation. We fabricated a testchip with a PVT controller demonstrating detection of the process corner and different control signal generation in a 130nm technology. The proposed controller occupied only $5400\mu m^2$ and dissipated no noticeable power compared to the traditional SRAM power consumption.

Achieving variation tolerance can improve system power/performance by reducing the guard band. While process and temperature (PT) vary largely for a system, having a smart controller that can provide information of such external changes and ensure the SRAM functionality can significantly reduce the guard-band. Identifying the major contributor in process variation (intra-die versus inter-die) can further optimize the design by targeting specific sources of the variation. Validating similar effects for different technologies can allow a generalized solution across technologies. Controlling the effect of external parameters such as process and temperature variation can also save time-to-market by producing an SRAM operating over a wide range of external conditions.

5.3.1   Variation: Impact and Mitigation

Variation in device parameters (e.g., in V_T, L, W) in the subthreshold significantly affects the functionality of the ratioed designs, such as an SRAM. Figure 5.3 shows the V_T variation of devices across different process corners and a wide range of temperatures. We used a conventional RO design to detect various process corners and temperatures. The RO aims to have a unique frequency mapping for each corresponding process corner and temperature (Figure 5.4). We also place a group of 16 ROs at each corner of the fabricated chips to observe the effect of spatial variation. Figure 5.5 shows the measured RO frequency of 16 ROs placed 1.4mm apart at each corner. The RO offers minuscule variation within the blocks

(Left Top, Left Bottom, Right Top, and Right Bottom). However, the frequency of RO between the blocks differs by up to 200 kHz. The difference in frequencies of ROs is due to the spatial process variation that resulted because of variation in the strength of the devices based on their location. A 256kb of SRAM bitcell array in 130nm node occupies an area of 1mm x 2.1mm [75] where spatial variation in the bitcell array results in a higher functional V_{DD} (i.e., V_{MIN}). In state-of-the-art System-on-Chips (SoCs), an SRAM may require the substantially increased area to address increasing complexity. Therefore, the spatial variation has a significant impact on subthreshold SRAMs while achieving the smallest operating V_{DD}.

(a) V_T distribution across process corners (T=25°C).

(b) V_T distribution across temperatures (TT Corner).
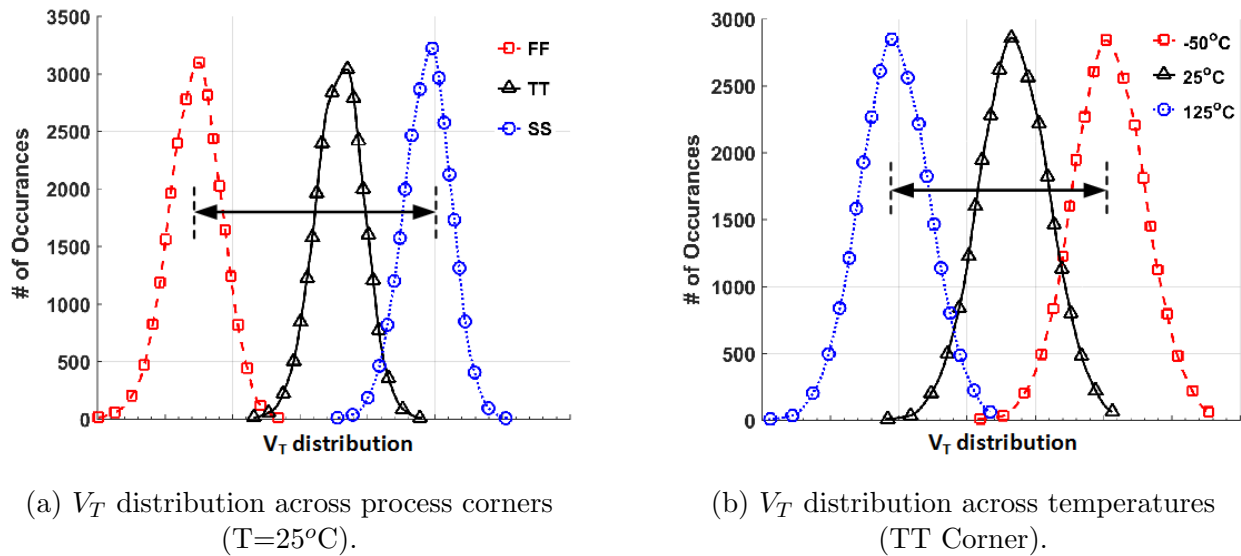
Figure 5.3: Process and temperature variation in device V_T. The plots show distribution of 10000 point Monte Carlo simulations at given process corner and temperature.

(a) Frequency mapping of process corners.

(b) Frequency mapping of temperature.
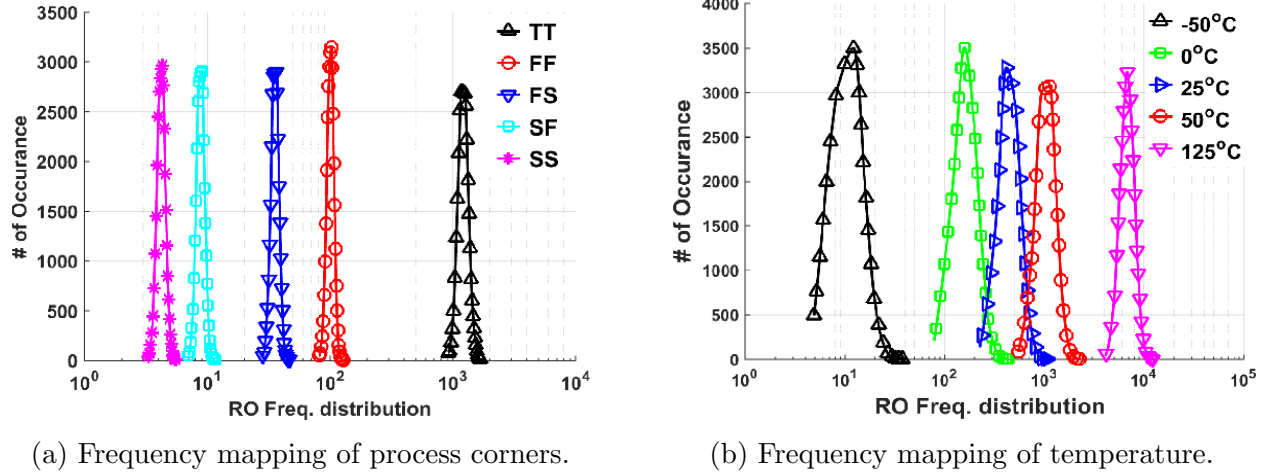
Figure 5.4: Process and temperature mapping by detection of the different frequencies.
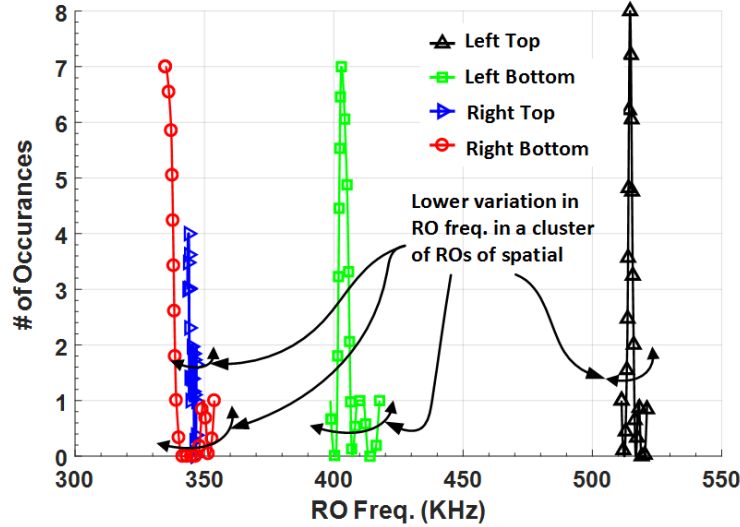
Figure 5.5: Impact of Intra-die variation: Due to spatial variation (blocks placed at each corner of the chip), the measured frequency of 16 ROs varies from 300KHz to 500KHz at $V_{DD}=0.4$V

Figure 5.6a shows the block diagram of the proposed PVT controller. The process and temperature information determined using RO frequencies are provided as inputs to the PVT controller. The Frequency-to-Digital Converter (FDC) is used to map the RO frequencies into a 16-bit digitized output. The droop in the voltage rail is sensed by a voltage droop detection circuit (discussed in sub-section later), and digitized data is provided to the PVT controller. Based on the control knobs, the controller generates different signals to give resiliency against PVT variations: Body Biasing (BB), assist selection, or droop mitigation.

The controller also employs temperature variation mitigation where the increase in leakage power is controlled by applying reverse BB. The controller also generates signals to indicate the independent strengths (slow/fast) of NMOSs and PMOSs. The information about the strength of individual devices can further be used in any analog design to control DC biasing generation based on NMOS/PMOS strength. Figure 5.6b maps multiple control generations for selected control given the corresponding process corner detection.
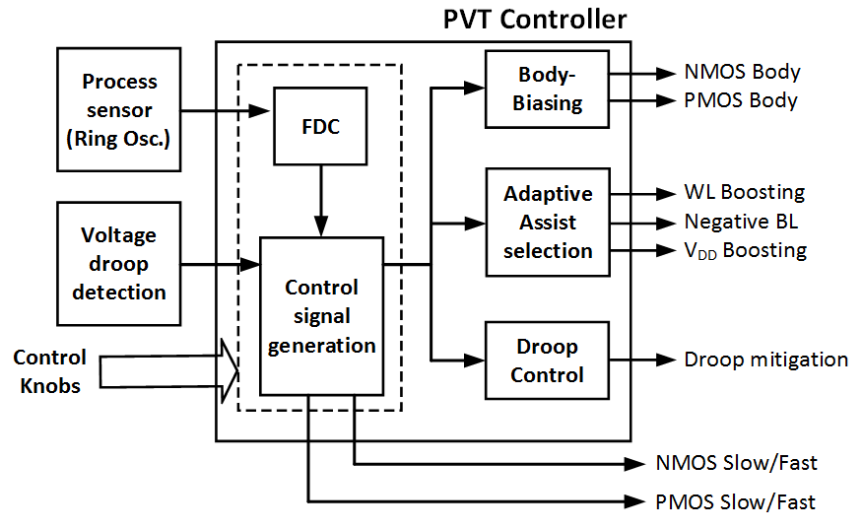
(a) Block diagram of the proposed PVT controller with process and voltage droop detection sensors and controlled signal.
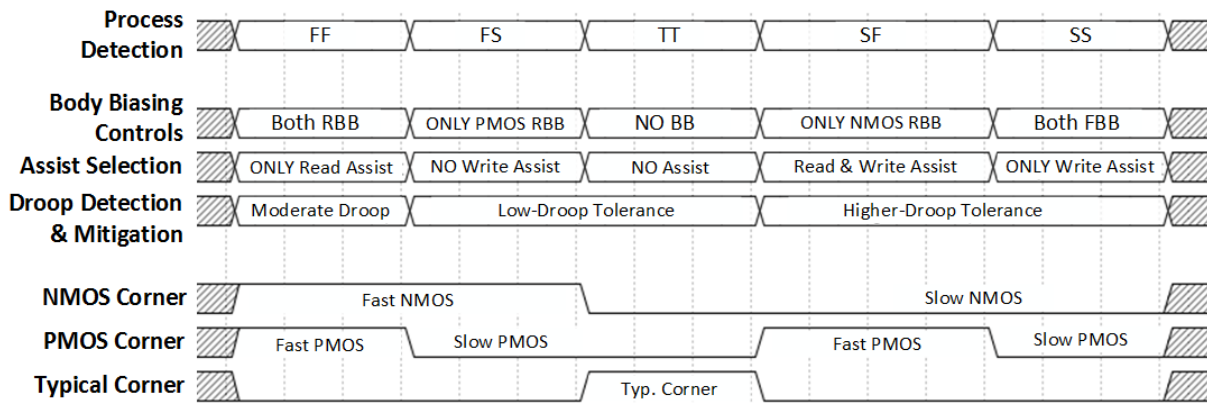
(b) Control signal generation based on process corner detection using RO; selection of a control knob allows specific mitigation control to improve the yield of subthreshold design.

Figure 5.6: Evaluating optimal assist technique for the performance enhancement at super-threshold

5.3.2   Results

In this section, we discuss the impact of process variation and its mitigation using BB and peripheral assist techniques by showing the measurement results of the fabricated chip. Finally, we will consider the implications of the variation in voltage rail. Figure 5.7 shows the fabricated chip in 130nm with ROs and voltage monitor blocks.
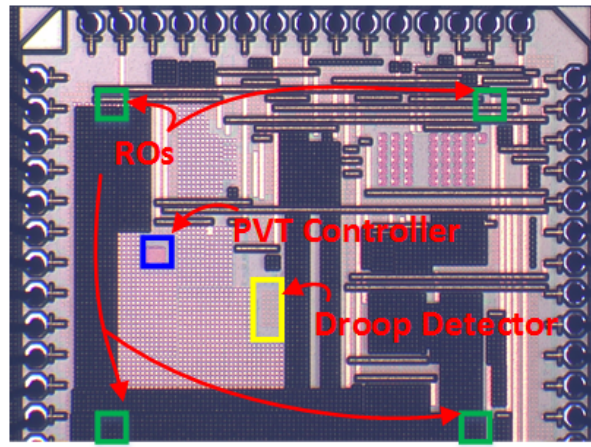
Figure 5.7: Die photo of the fabricated chip with PVT controller, ROs as the process sensor, and droop detection sensors

5.3.2.1   Body-Biasing: Reliability, Leakage, and Energy Optimization

Existing research has highlighted the benefits of BB to optimize leakage power and performance [30, 76]. The selective BB (i.e., forward BB (FBB) or reverse BB (RBB)) for NMOS or PMOS improves different metrics of an SRAM. For example, RBB reduces leakage current by increasing the device V_T while FBB improves performance with V_T reduction. Fig. 4 a) shows how selective BB (FBB for NMOS and RBB for PMOS) can improve the worst case write margin (reliability) of an SRAM bitcell. Similarly, RBB allows up to a 100X leakage current reduction in an SRAM array (Fig. 4b). However, a static selection of BB applied

across different process corners results in the area, power, and energy loss. For example, the FS corner (Fast NMOS; Slow PMOS) inherently assists the write operation for an SRAM and hence does not require any additional NMOS-bias based BB technique. Therefore, a dynamic selection of BB based on the process corner enables optimization of power consumption and hence ensures reliable operation.

Figure 5.6b shows the decision taken from the PVT controller for the required BB based on the detected process corner. To demonstrate the benefits of selective BB, we applied controller-generated signals to a 1kb SRAM fabricated separately in multiple process corners. We paired our process monitoring chips for a given corner with an SRAM chip fabricated in the same corner and provided the same temperature for both chips. Figure 5.8 highlights the benefits of BB by showing improvement in reliability, power, and energy. In Figure 5.8a, applying selective BB improves the write failure in fabricated SRAM. While NMOS and PMOS RBBs are selected to reduce high leakage in the FF corner to achieve target leakage power same as that if TT corner (Figure 5.8c). Finally, we measure the energy versus delay plot for a 1kb SRAM while applying controller generated BB signals. Here, the energy per access of an SRAM reduced by around 10X with BB = 0.3V, as shown in Figure 5.8c. Careful control of BB from the PVT controller for a specific process corner and temperature further optimizes energy consumption by only enabling the BB when required. To allow finer control over BB, the controller provides a separate BB for SRAM bitcell array and peripherals, individually for NMOS and PMOS devices. Managing BB of peripherals can improve the timing yield in SRAM.
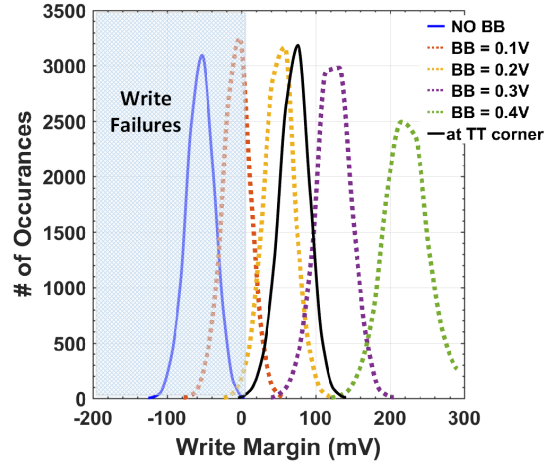
(a) Write Stability improvement: Worst case write failures improves by selective BB based on the process corner detection
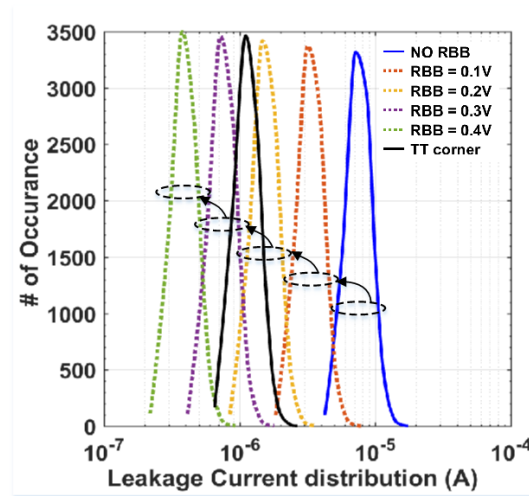
(b) Leakage power reduction: Higher leakage at FF corner is optimized by applying RBB resulting in 100X leakage current reduction
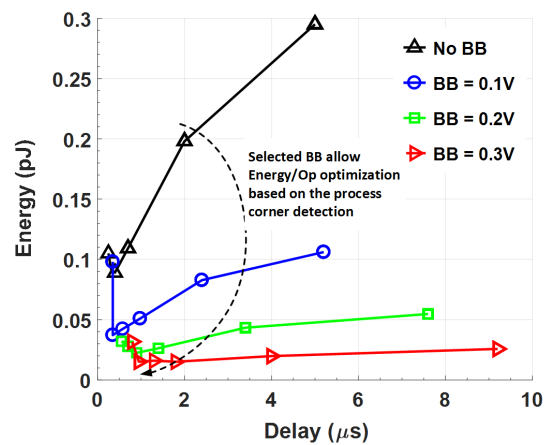
(c) Energy per access optimization: measured result from 1kb SRAM with adaptive BB lowers energy/access with selective BB based on the process corner detection

Figure 5.8: Effectiveness of PVT based body-bias control on a) Reliability (Write margin), b) Leakage (Power), and c) Energy (Energy per operation)

5.3.2.2 Adaptive Assist Technique Selection

As shown in Figure 5.6b, the PVT controller selects the required write and read assist techniques in optimizing energy and yield. For example, at FF (Fast NMOS; Fast PMOS) corner, only a read assist technique is needed to avoid half-select failures. While at SF (Slow NMOS; Fast PMOS) corner, a write assist is enabled to help the write operation. Additionally, in Chapter 4 we showed that the optimal assist technique is a strong function of operating supply voltage. Therefore, the proposed controller can be re-programmed for an optimal assist selection across a wide range of operating voltages. The assist enabling controls are provided to a 256kb SRAM to optimize the SRAM V_{MIN}. Based on available assist controls, we consider wordline (WL) boosting, and negative Bitline (NegBL) as write assist techniques and V_{DD} boosting as a read assist technique. Enabling the only required assist based on the process corner will save energy from the peripherals and the degree of assist for the functional guarantee.
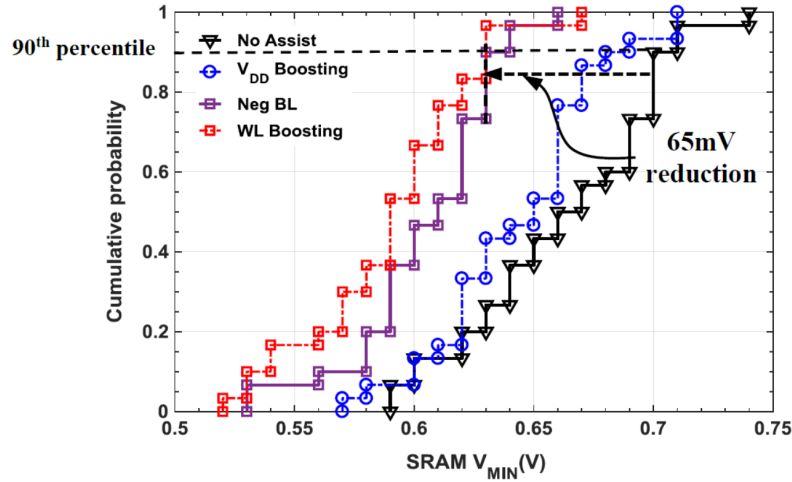
Figure 5.9: Measured SRAM V_{MIN} improvement using selective peripheral assist

Detection of the process corner can further reduce SRAM minimum operating voltage (V_{MIN}). In [77], the authors indicated that the knowledge of the process corner could help in reducing V_{MIN} and the degree of assist needed to achieve the targeted V_{MIN}. The authors in [4] showed that the optimal assist technique changes with V_{DD}, the extent of applied assist, and offers different trade-offs between reliability (margins), performance (delay), and energy

(sustainability). Additionally, the worst case corner for the write operation for an SRAM (i.e., SF) is different than the worst case read stability (i.e., half-select) corner (FS). Therefore, it is crucial to select an assist technique based on the process corner and temperature to achieve true V_{MIN} across different process corners. Figure 5.9 shows the CDF of the measured SRAM V_{MIN} across chips by enabling a single assist technique. However, it is important to note that applying a combination of multiple assists may further allow scaling the SRAM V_{MIN} at the cost of higher energy. As a result of selective assist, 90th percentile of SRAM V_{MIN} improves by 65mV. Allowing lower V_{MIN} will reduce the leakage at guaranteed functionality.

5.3.2.3 Voltage Droop Detection and Mitigation Strategy

The voltage rail experiences variation (i.e., droop) due to the load change and inherent losses in the regulator. The V_{DD} in subthreshold is scaled significantly to optimize the energy in ULP application. However, the droop in V_{DD} rail results in functional failures. Figure 5.10 highlights the worst case Data Retention Voltage (DRV) of an SRAM across different process corners and temperatures. Here, it is important to observe that at each corner a wide range of
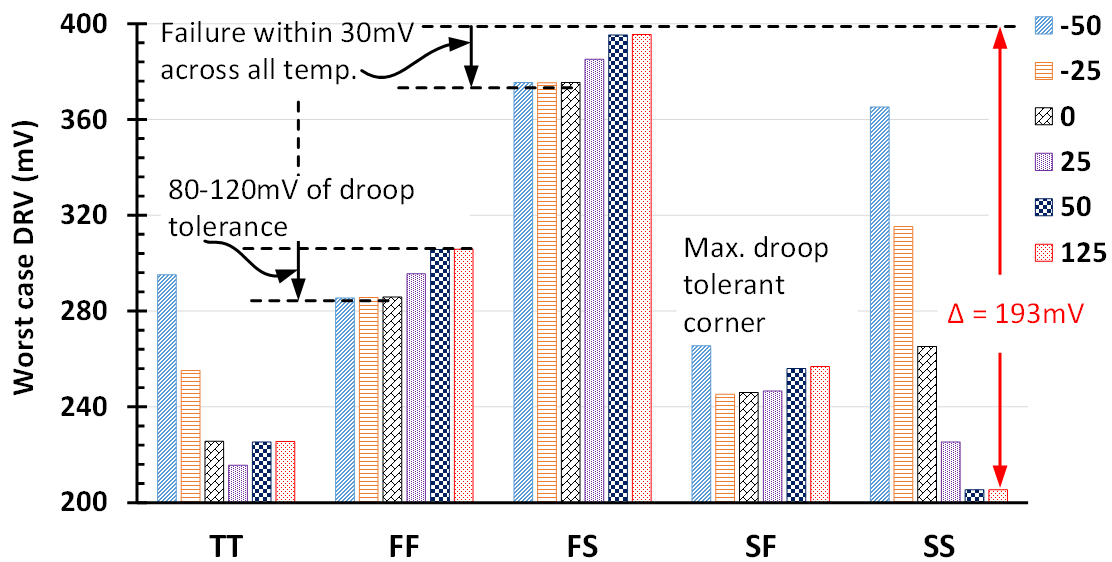
Figure 5.10: Data Retention Voltage (DRV) across different process corners and temperatures (in ^{o}C).
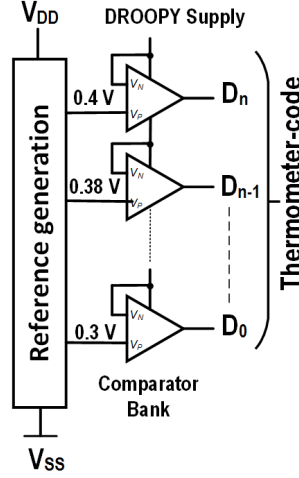
Figure 5.11: Implemented voltage droop monitor with 20mV of droop granularity.

temperature samples are considered. The worst case DRV at a given corner and temperature is calculated from 10000 point Monte Carlo simulations to account for the mismatch due to the local variations. The figure indicates that the different process corners pose a different degree of variation tolerance in voltage rail (droop). For example, an SRAM in FS corner requires very stable V_{DD} to operate at 0.4V. A 30mV of droop in V_{DD} at SF corner will result in retention failures. In contrast, the SF corner allows maximum droop. The TT and SS corners experience a wide droop tolerance across temperatures. Therefore, to mitigate the voltage droop, we implement a voltage monitor (Figure 5.11) to identify the droop in the voltage rail. Based on the process Corner detection, the PVT controller generates signals to control the switching activity of a switch-capacitor (SC) based DC-DC regulator to ensure a reliable V_{DD}.

5.4   Soft Errors: Reliability Challenges at Ultra-Low Voltage

[2] The high energy alpha particles generated from radioactive decay of uranium and thorium impurities located in the packaging become a major reliability challenge [78]. The area-constrained subthreshold SRAM operating at scaled supply voltages can be another critical concern for such particles. Because the soft-error is due to "extra" charge generation from alpha particles, the effect of node capacitance and supply voltage plays a very critical role in physical error generation. In [79–81], the authors demonstrated the impact of radiation on advanced technology nodes and high-performance systems with possible compensation using different design parameters and techniques (Sizing, ECC, BIST, etc.), while the effect of technology scaling, Random Dopant Fluctuation (RDF), and process variations on soft-errors are addressed in [82–84]. However, the effect of soft-errors in subthreshold is still examined. The change in MOS capacitance from super-threshold to subthreshold follows a nonlinear trend that might cause a higher probability of failure. For BSNs-like applications, where higher frequency may not be the significant source of the soft-error, scaled supply (V_{DD}) might significantly increase soft-error rates. Soft errors can cause a significantly higher failure rate than all the other reliability mechanisms combined. The rate at which soft errors occur is called the soft-error rate (SER). The measurement of the SER in any design is being calculated by failure in time (FIT). Therefore, one FIT represents failures in device per billion hours. For example, a typical failure rate for a "hard" reliability failures such as gate oxide breakdown, electromigration, etc., is about 1-50 FIT. There are half-a-dozen other critical reliability mechanisms which cumulatively cause 50-200 FIT. However, for an unprotected design, the SER can exceed 50000FIT per chip [65].

   In this work, we study the impact of the radiation-induced soft errors in ULP applications

---

[2]This section is based on the publication [HNP2]

operating at subthreshold voltages and possible solutions to improve soft-error resiliency for subthreshold operation. We discuss the mechanism behind the soft errors followed by the experimental results. First, we provide the basic theory and physical mechanism behind the SEU generation. The experimental test setup to emulate the impact of particle strike for the storage node is demonstrated using an SRAM bitcell and discussed later in this chapter. We used this setup to measure the Q_{Crit} of the node across different parameters.

5.4.1 Theory and Mechanism Behind the Single Event Upset (SEU)

There can be many sources of radiation-induced high energy particle generation. However, there are two main causes: 1) alpha particles from packaging impurities, and 2) neutrons from the atmosphere. An alpha particle, consisting of two protons and two neutrons, emitted by radioactive nuclei such as uranium and thorium impurities in the packaging materials. The charged alpha particle interacts directly with electrons, while neutrons impact the silicon by inelastic or elastic collisions [85].
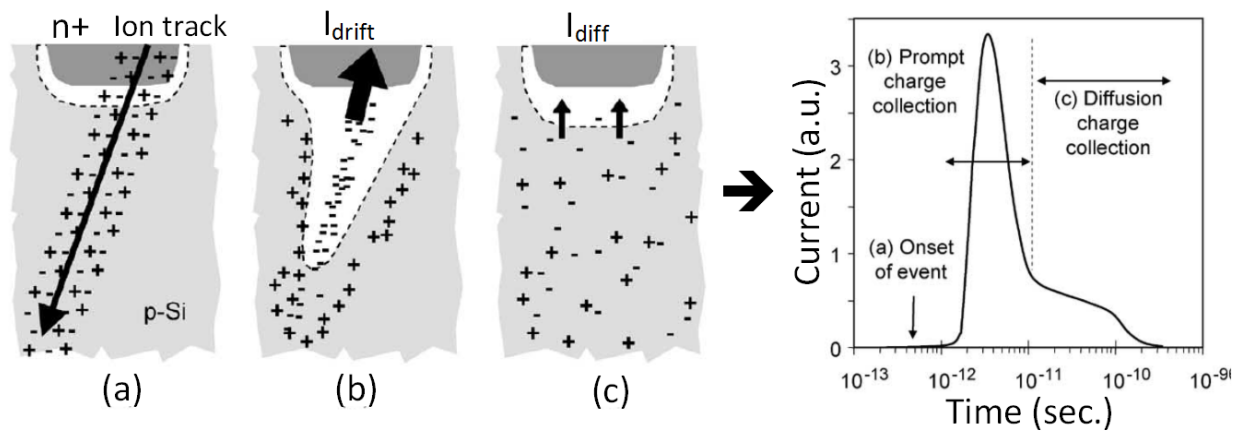
Figure 5.12: Charge generation and collection mechanism in a reverse-biased junction and the resultant current pulse because of the high-energy particle strike [6].

To illustrate the mechanism of the particle strike and hence the resulting soft-error, Figure 5.12 considers a reverse-biased junction as it is the most sensitive part of the circuit. Figure 5.12 a) shows an example event of the impact of an ion in a cylindrical track of electron-hole pairs with very high concentration. At the onset of the track coming closer to the depletion region, electrons and holes are collected by the corresponding electric field potentials (b). For Silicon substrate devices, the range of a 10MeV alpha particle is <100m and therefore alpha particles from outside the package cannot produce soft-errors, while particles emitted by device packaging materials become a concern [6]. Similar to the alpha particle, another source of soft-errors is high-energy cosmic rays. Neutrons have higher flux, which varies with altitude [5]. In addition to Q_{Coll}, device sensitivity to a particle strike (i.e., a critical charge (Q_{Crit}) also depends on the node capacitance, operating voltage (V_{DD}), and the strength of the feedback path (e.g., FF and SRAM). With the technology and voltage scaling, effective node capacitance reduces and hence reduces Q_{Crit}. Therefore, a smaller Q_{Coll} that is greater than Q_{Crit} can result into a soft-error.

Present Integrated Circuits (ICs) consist of many individual blocks such as analog circuits to interface with analog signals, digital processors for computation, and embedded memories (mainly SRAM) for the storage. The impact of particle strikes depends on the circuit design
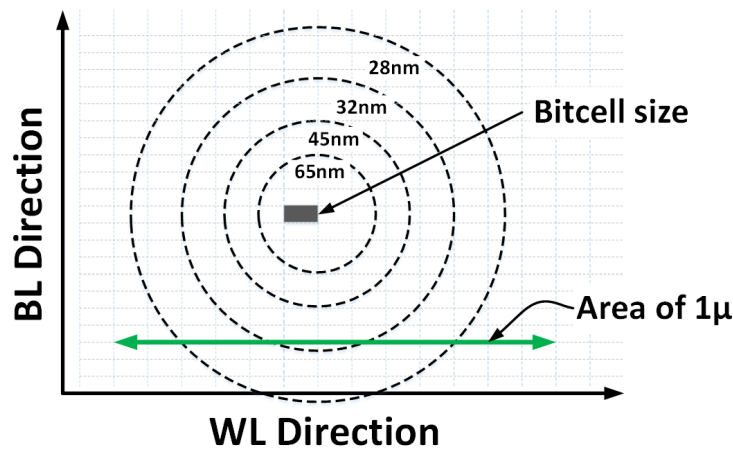
Figure 5.13: Soft-error concern with technology scaling: Area of an SRAM bitcell array covered by 1μm space indicating an increasing number of bitcells under the particle strike hit.

and therefore the concern of soft-error. Although the first soft-error been discovered in DRAM [64], it is currently one of the most robust electronic devices due to the 3D structure of the capacitive storage of the charge. The SRAMs, storing the state in two bi-stable elements, is considered to be robust because of the inherent feedback nature of the storage. However, with supply and technology scaling, denser SRAM with lower power requirement reduces the resiliency of the SRAM over technologies. Therefore, in our experiments, we consider a 6T SRAM bit cell design in 130nm technology as the similar design has been used in present BSN that represents the state-of-the-art [2]. Figure 5.13 represents an estimated SRAM bitcell being covered by a radiation-induced particle with technology scaling. Additionally, an exponential growth in the SRAM capacity in microprocessor and other applications has increased SER significantly. Therefore, SRAM with technology and supply scaling is a grave concern to achieve a sustainable functional-yield.

5.4.2 Simulation Setup

As discussed in the previous section, the particle strike on a reversed-bias pn junction results in a current pulse with a particular rise time (t_R) and fall time (t_F). As mentioned in the previous section, the current pulse depends on many factors, including the size of the device, biasing supply voltage, substrate doping, particle strike distance from the junction, the type of the ion particle, strength of the feedback path in the case of SRAM or FF, and well doping. For our experiment, we consider double-exponential pulse (Equation (5.1)) from Hazucha and Svensson [66] that is well-defined and validated to understand the current pulse behavior generated by a particle strike. Figure 5.14 shows the simulation setup with current injection at one of the storage nodes in a 6T SRAM bitcell. During the particle strike (I(t)), the bitcell is considered to be in the 'Hold' state (WL=0, BL=BLB). An iterative process performs a binary search for an applied Q_{Coll} for which the node flip value is checked. Figure 5.15 shows different current pulses based on the different t_F values. Here, the purpose of exploring various current pulse with different t_F is to understand the impact of the current pulse on

the Q_{Crit}. It is also important to note that different current pulses can be represented as different technology nodes and substrate doping [86].

$$I(t) = \frac{Q_{Coll}}{t_F - t_R}\left\{ e^{\frac{t}{t_F}} - e^{\frac{t}{t_R}} \right\} \tag{5.1}$$

Where,

Q_{Coll} = Charge collected by the reversed bias junction

t_F = Fall time of the current pulse

t_R = Rise time of the current pulse
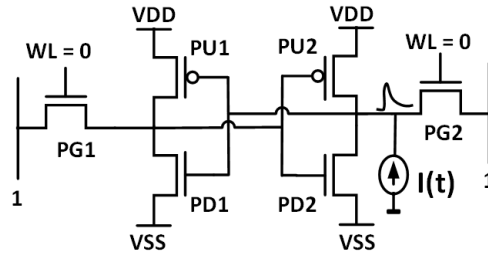
Figure 5.14: Experimental setup for calculating Q_{Crit} for a 6T SRAM bitcell.
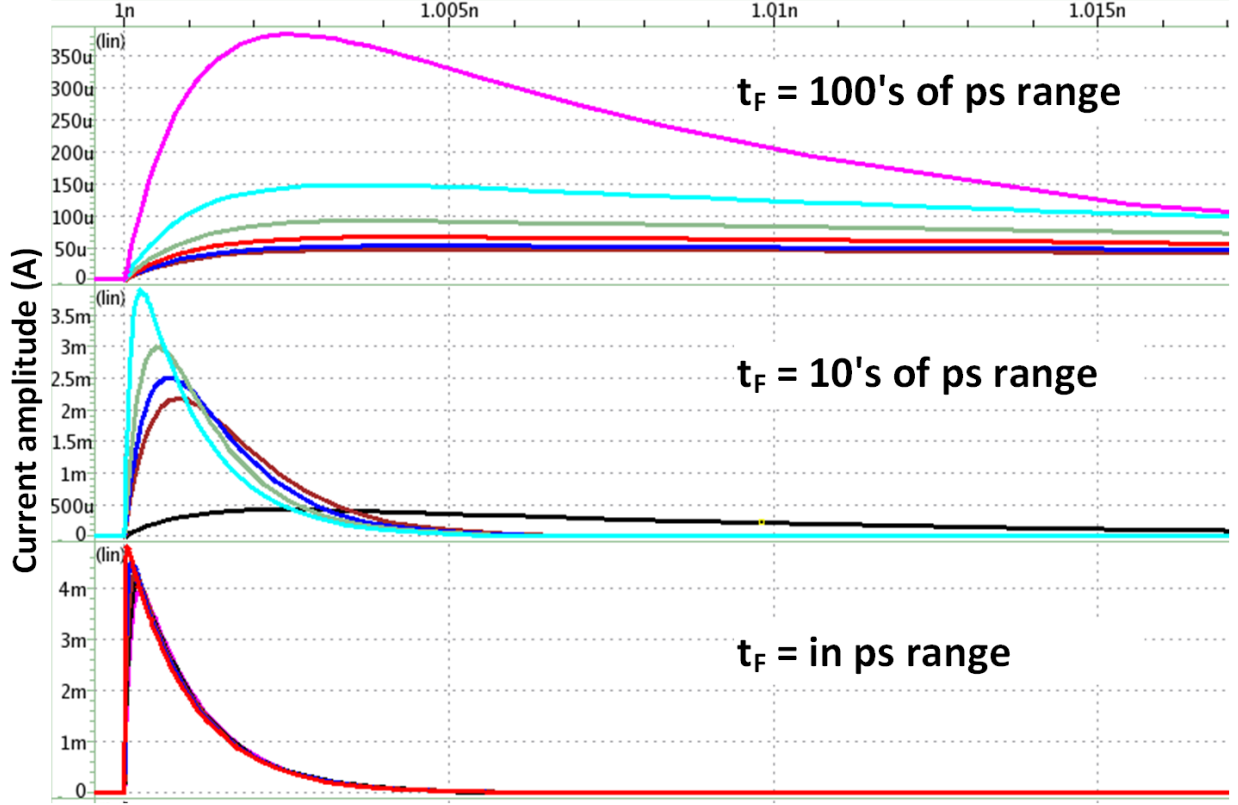
Figure 5.15: Change in the pulse width of the current pulse based on charge collection property of the device material and structure (e.g., bulk, FDSOI, FinFET, etc.) for a constant V_{DD} and Q_{Coll}. The charge collection property is emulated by varying t_F of the pulse.

5.4.3 Results

We used an iterative loop of varying different values of particle strike and observe the bit flip in the 6T bitcell as per the experimental setup discussed in the previous section. Figure 5.16 shows Q_{Crit} values for different particle strike having different t_F. In the super-threshold range of supply voltages, the faster diffusion (shorter t_F) results in 100X Q_{Crit} reduction for the same amount of charge collected. The trend is predominant at super- and near-threshold range. While in the subthreshold, even larger current pulse causes the flip due to the smaller node capacitance. The authors in [87, 88] showed the difference in the charge collection property using a 3D device model for bulk/planner and FinFET. The result shows how different devices with different physical structure (e.g., bulk, FinFET, FDSOI) with various
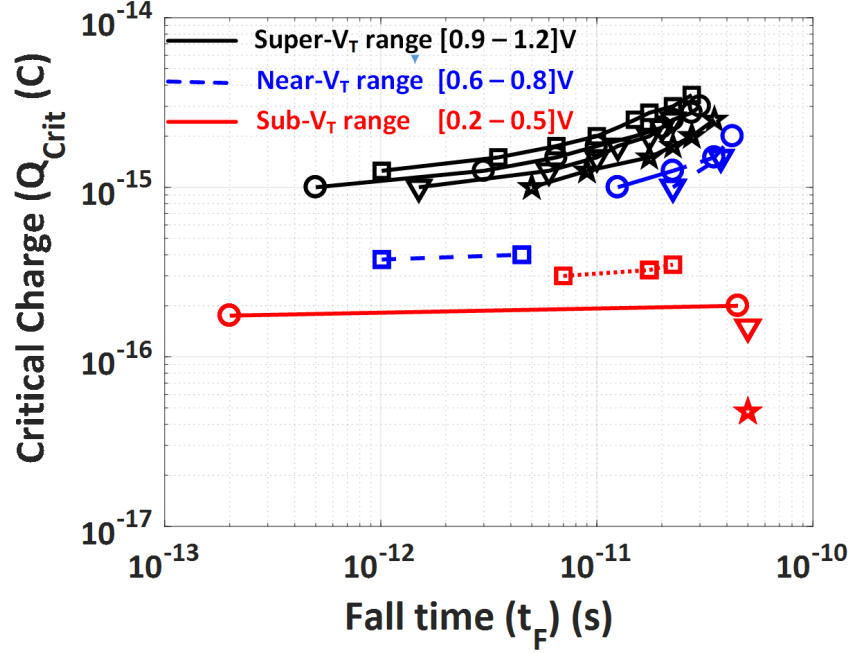
Figure 5.16: Effect of t_F (i.e., charge collection property) on Q_{Crit} for the same value of Q_{Coll} at given V_{DD}.

charge collection properties impact the Q_{Crit} for similar devices.

Figure 5.17 shows the effect of supply scaling on Q_{Crit}. In the super-threshold, scaling the supply voltage only reduces Q_{Crit} by less than 0.5fC/V and hence does not impact reliability until the supply scaling is limited to this range. However, in the near- and subthreshold voltages, Q_{Crit} reduces exponentially. In this range of V_{DD}'s, the Q_{Crit} becomes so small that even a particle strike at nominal altitude with a lower flux can flip the node value. This is a critical challenge for ULP biomedical applications where a person is required to travel at different altitudes and the neutron flux increasing with altitude [5]. Because the ON current (I_{ON}) has an exponential dependency on the V_T in subthreshold [24], process variation impacts Q_{Crit}. A reduced I_{ON} results in a weaker feedback path for the storage node in a crossed couple bitcell. Figure 5.18 highlights the impact of process variation on Q_{Crit}. To summarize the impact of supply scaling on soft errors, we compare the impact of technology scaling on the Q_{Crit} with the supply voltage scaling. We consider low-power (LP) device models from the PTM [89]. Figure 5.19 shows linear scaling of the Q_{Crit} value

with the technology scaling while an exponential decay with supply voltage scaling towards the subthreshold range of V_{DD}s and hence suffices the need of better design consideration at scaled supply compared to a scaled technology node.
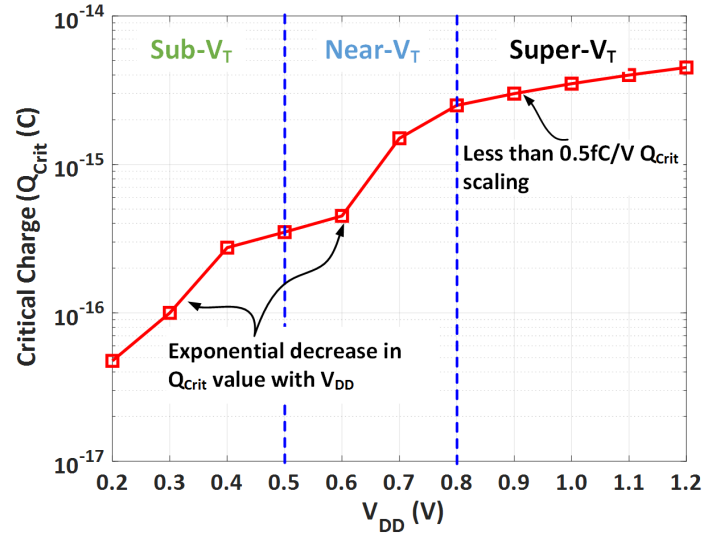
Figure 5.17: Effect of Soft-Error on V_{DD} scaling.
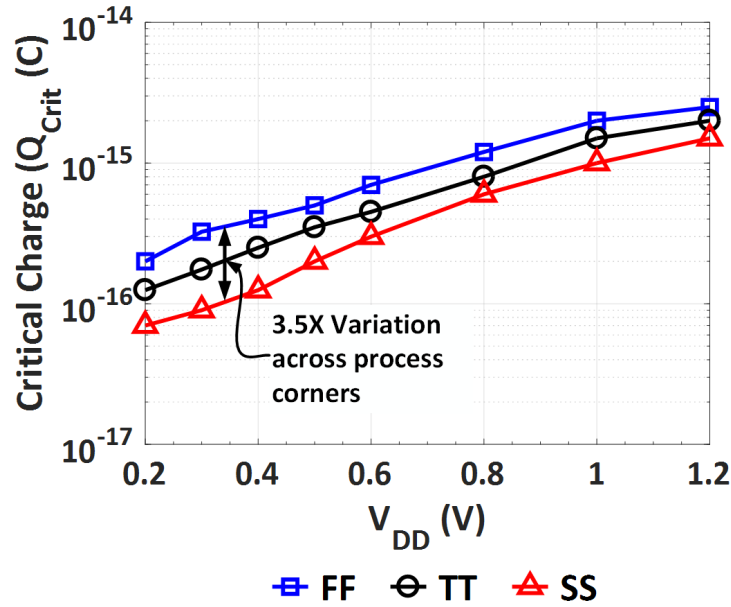
Figure 5.18: Q_{Crit} has 3.5X variation across process variation for the 130nm technology node for the subthreshold range of V_{DD}s.
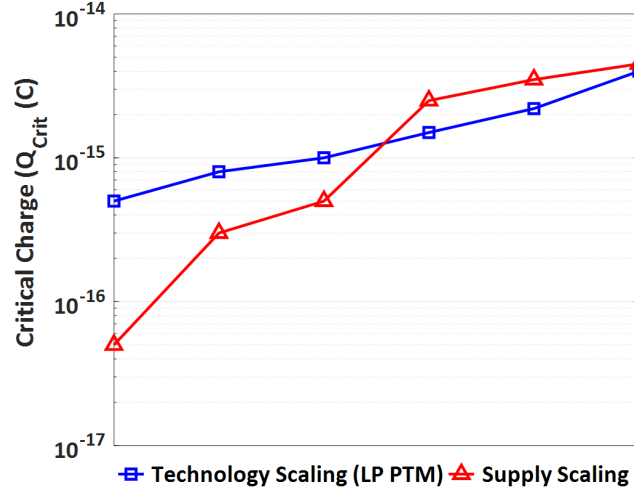
Figure 5.19: Comparison: Impact of technology scaling vs. supply scaling on the Q_{Crit}.

Since the first report of cosmic rays causing failures in 1975 space applications, a significant amount of work has been done, from device structure [86–88, 90] to different protection schemes [91–93]. The device like FDSOI makes device robust against the particle strike by structure while Deeply Depleted Channel (DDC) [30] reduces doping concentration in the channel helps SEU reduction. In addition to the physical structures, new FF structures [92] prevents the propagation of the soft-error generated. Different circuit techniques used the low-power such as Body Biasing (BB), also minimize soft-error based on the degree of biasing [94]. At the architecture level, various techniques are used to minimize functional disruption due to the particle strike [95].

5.5 Conclusion

The resiliency of the SRAM can improve system power/performance by reducing the guard band. Therefore, we demonstrated a digital controller to mitigate the PVT variations dynamically. The fabricated controller used RO-based process detection and voltage droop monitor. During the BB-based PVT variation, selectively enabling BB improves reliability and improves energy by 10X. The adaptive peripheral assist selection for the SRAM helps to lower V_{MIN} by 65mV. The use of selective BB, peripheral assist technique, and droop

control significantly improve the reliability, leakage power, and energy of the subthreshold ULP applications.

While process and temperature (PT) vary largely for a system, having a smart controller that can provide information of such external changes and ensure the SRAM functionality can significantly reduce the guard-band. Identifying the major contributor in process variation (intra-die versus inter-die) can further optimize the design by targeting specific sources of the variation. Also, evaluating similar effects for different technologies can provide a generalized solution across technologies. Controlling the effect of external parameters such as process and temperature variation can also save time-to-market by producing an SRAM operating over a wide range of external conditions.

In addition to PVT variations, soft-errors limit the scope of subthreshold SRAM applications. To address the reliability, later in this chapter, we explored the reliability challenges for ULP applications induced by soft-errors. The supply scaling has been proven as an efficient metric to optimize power and energy for battery-less BSN application. However, reliability at scaled supply voltage operation is a big challenge. We also studied the impact of multiple current pulses on a critical charge (Q_{Crit}). The results help to understand Q_{Crit} behavior for different device structures based on charge diffusion property. Finally, the supply scaling effect on Q_{Crit} has been studied. The impact of process variation in subthreshold resulted in 3.5X higher Q_{Crit} shows significant reliability issue due to subthreshold operation. Further, exploring the effect of device and circuit level techniques such as sizing/layouts on the soft-error will help to address this challenge in subthreshold designs.

Chapter 6

Conclusions

In addition to the technology scaling, supply voltage scaling to reduce power has created many design challenges for both, low-power requirement and reliability. subthreshold characteristics such as increased in variability and reduced ON current demands new circuit techniques. With IoT, battery-less ultra low-power healthcare applications such as Body Sensor Network demands power optimization — from the circuit design to the architectural level design decisions. This dissertation contributes to low-power SRAM design space mainly in the following ways: evaluating reliability-power-performance design trade–offs for evaluating optimal assist techniques to reduce SRAM V_{MIN}, evaluating the impact of various bitcell based on the selection of different threshold devices, emphasizing technology-circuit co-design for energy-efficient design ULP SRAM, proposal of a wide range operating SRAM using dynamic selection of assist techniques. Certainly, the variation in device parameter due to the variation in the process, voltage, and temperature (PVT) impact the reliability of the SRAM at lower supply voltage operation. To address the reliability challenge, we designed a digital controller to adapt PVT variation for achieving reliability and energy efficiency and evaluating the impact of particle strike on ULP operation.

6.1   Summary of Contributions

Write-Assist Techniques for an Array Energy Minimization

- Evaluated different write assist techniques for minimizing SRAM array write energy to enable energy-constrained BSNs applications.

- Investigated energy minimization trend for different assist techniques with supply voltage and degree of applied assist as design knobs.

- Proposed Pareto curves for various metrics with trade-offs between reliability, energy, and system V_{MIN}.

- System-level evaluation of assist techniques studying the trade-offs between the array V_{MIN}, array E_{MIN}, write reliability, and performance for reliable and energy-constrained applications.

- Concluded the following:

  - The reliability trend of write-assist techniques in subthreshold are different from super-threshold thus the design decision for assist implementation will be different for given system level constraints in subthreshold.

  - The half-select failures, not an issue at super-threshold, limit the array V_{MIN} as well as the degree of assist that can be applied to scale the V_{DD}.

Ultra-Low Power Subthreshold SRAM Design Optimizing Leakage[1]

- Evaluated optimal choice of the bitcell based on device type for given constraints and suitable read/write assist techniques for High-V_T and Multi-V_T arrays.

- Fabricated 1KB SRAM chip in 130nm CMOS that operates between 350mV and 700mV for ULP subthreshold operation.

---

[1]Part of this work has been done in collaboration with Farah B. Yahya

- A read burst mode is implemented to reduce read energy when consecutive addresses are accessed and saves 22% active read energy.

- Aggressive power gating reduces the power consumption down to 12.29nW with retention and 1.09nW when data is not needed.

- Proposed design achieves the lowest leakage power per bit at 1.5pW/bit that is 4X lower leakage for an 8T bit-cell array.

- Proposed low-power SRAM with a tight couple with the low-power controller allowed the controller to take full advantage of a various low-power feature from SRAM that resulted in up to 66% measured power savings at the system level and helps to achieve sub-μW power budget.

Technology-Circuit Co-Design with Body Biasing Control[2]

- Circuit techniques such as subthreshold operation and reverse body biasing (RBB) are co-designed with the technology to maximize the energy/power saving.

- A test chip implemented with a 1Kb 6T SRAM, an FIR filter, and a 51-stage RO demonstrating the benefits of technology knobs to minimize energy.

- Demonstrated circuit-technology co-optimization with body biasing to minimize 98% leakage power and 7X energy reduction for an SRAM design using an ultra-low leakage (ULL) device.

- Proposed design achieved the lowest leakage power per bit at 1.5pW/bit for a 6T bit-cell array.

- Achieved one of the lowest possible V_{MIN} SRAMs ($V_{MIN} = 0.2$V) to enable battery-less low-power platform.

---

[2]This work has been done in collaboration with Abishek Roy, Farah Yahya, and Ningxi Liu

Self-tuned SRAM System for a Wide Range Operation[3]

- Co-designed an adaptive, closed loop memory system that leverages combinations of bias-based peripheral assists to expand the operating range of a 256kb 6T SRAM by over 67%.

- Analyzed an optimal assist technique selection based on performance requirement for a broad range of V_{DD}.

- Implemented an Assist Controller in RTL for a dynamic range of assist selection to operate from 0.38V to 1.2V.

- Demonstrated the impact of proposed techniques reducing the SRAM power from 14.4mW to 11.4W (1263X power reduction).

- Combined assist selection extends the operating range down to 0.38V, and gives a 12.4X lower leakage power (9.5pW/bit).

- Integrated complete design with multi-voltage domain crossing achieving 240mV V_{MIN} reduction on silicon.

PVT Variation Mitigation for Reliable and Energy Optimization

- Designed RO-based process and temperature detection.

- Demonstrate a digital controller to mitigate the PVT variations dynamically.

- Fabricated a testchip in 130nm with PVT detection mechanism and controller.

- Silicon validation of the proposed controller architecture with body-biasing, assist technique selection and power gating control signals to 100X leakage power, 10X energy, and 65mV SRAM V_{MIN} reduction.

---

[3]This work has been done in collaboration with Arijit Banerjee and Ningxi Liu

Impact of Particle-Strike Induced Soft-Error on ULP Design

- Studied the impact of particle strike on critical charge (Q_{Crit}) in the subthreshold operation

- Compared the impact of Soft-Error Rate (SER) on storage node for the technology scaling versus supply scaling.

- Studied the impact of the process variation in sub-V_T resulting in 3X Q_{Crit} variation.

- Demonstrated a trend of an exponential reduction in the Q_{Crit} of a storage node with supply in near- and sub-V_T design.

6.2 Conclusions and Open Problems

In this work, we addressed many low-power challenges in SRAM to enable a battery-less platform. However, there are still many open questions in order to optimize the energy with a block and system level optimization. Also, many efforts have been made to lower the SRAM V_{MIN}.

In Chapter 3 we studied various design knobs such as technology, device selection, and circuit and architecture-level techniques to allow SRAM V_{MIN} scaling in subthreshold range. An advanced technology, such as Deeply Depleted Channel (DDC), provided an Ultra-Low Leakage (ULL) device as a technology knob while the Reverse Body Biasing (RBB) is implemented as circuit design knob to reduce the leakage of the 6T SRAM by 98% and energy/cycle of SRAM by 83%. In an effort to reduce the leakage of the SRAM array, we achieved the one of the lowest operating V_{MIN} ($V_{MIN} = 0.2$V). The use of circuit and various architecture level optimization showed 460% leakage power reduction compared to the available similar technology node design [18]. With high-threshold devices, the SRAM array leakage has been reduced substantially. While many efforts made to optimize the leakage of the SRAM array using various techniques, the peripheral circuitry starts dominating the total

power. Therefore, there are many areas of further optimization in the peripheral circuitry at both, device selection and circuit techniques, levels.

A complex SoC implements two types of power domain implementation for different blocks (Figure 4.4): a shared supply voltage between all blocks or split supply voltages between SRAMs and other blocks. For a shared supply configuration, minimizing the SRAM array energy may or may not minimize the full system energy, since other components in the SoC beside the memory may dominate the system power (e.g., Radios). In this case, minimizing the SRAM V_{MIN} extends the operating range of SRAM and allows more flexibility for co-optimizing SRAM with a non-SRAM block for the system energy minimization. Therefore, to advance the goal of energy per operation for a battery-less self-sustainable IoT system, it is essential to model system-level requirements with energy optimal tracking with a system perspective.

In Chapter 4, we evaluated the effectiveness of various assist techniques to achieve energy optima. We demonstrated a closed loop control system where an adaptive selection of the assist techniques is made to optimize the energy of an SRAM for the required performance. The proposed closed loop self-tuning 256kb 6T SRAM sub-system allowed a wide range operation with 0.38V–1.2V extended operating range using combined read and write assists using canary-based V_{MIN} tracking.

The proposed architecture allows 337X power reductions and 3.4X V_{MIN} tracking using multiple assists. Although, the proposed system selects the assist techniques based on the performance requirement (and hence operating V_{DD}), it is still implemented based on a look-up table that requires configuration. The system can further be optimized with a co-design with another block(s) in the SoC that may also control overall power efficiency of the system. For example, the power management unit (PMU) required higher power to provide a stable power rail. While achieving this, PMU power efficiency reduces and hence delivers lower power to the system. An SRAM-PMU co-design may allow droop in the rail until the failure in the SRAM been detected. Such a system level co-design using the power

modeling will enable various system level knobs for the energy optimization. The proposed circuit techniques with detailed trade-off analysis applied for the used process technology node can also be applied. As a result, the impact of various leakage optimization techniques can be validated irrespective of the process technology node selection.

Finally, we addressed the reliability concerns at the ultra low voltages in Chapter 5. The supply scaling has been proven as an effective metric to optimize power and energy for battery-less BSN application. However reliability at scaled supply voltage operation is a big challenge. We demonstrated that improving the resiliency of the SRAM can improve system power/performance by lowering the guard band. The proposed digital controller to mitigate the PVT is fabricated using 130nm technology to highlight the benefits. An RO-based process detection and voltage droop monitor enabled selective body biasing to improve reliability and energy by 10X. While the similar control signals helped to lower SRAM V_{MIN} by 65mV. However, the proposed controller legs the closed loop control that , in future, may allow an autonomous control for the energy optimization. We also studied the reliability challenges for ULP applications induced by soft-errors. Future IoT applications operating at scaled supply voltages to achieve lower power need to be evaluated for the soft-error resiliency. Various radiation hardening techniques proposed before [91–93] required to be re-evaluated for the subthreshold operation under lower performance requirement.

Appendices

Appendix A

Publications

Completed Publications

[HNP1] **Patel, H. N.**; Yahya F.; Calhoun B. H., "Subthreshold SRAM: Challenges, Design Decisions, and Solutions", 2017 IEEE International Midwest Symposium on Circuits and Systems is the oldest Circuits and Systems Symposium (MWCAS), Boston, MA, USA — August 6th-9th, 2017

[HNP2] **Patel, H. N.**; Mann, R. W.; B. Calhoun, "Soft Errors: Reliability Challenges in Energy-Constrained ULP Body Sensor Networks Applications", 23rd IEEE International Symposium on On-Line Testing and Robust System Design (IOLTS), Thessaloniki, Greece, July 3-5, 2017.

[HNP3] Yahya, F.B.; Lukas C.J.; Breiholz J.; Roy A.; **Patel H. N.**; Liu N.; Chen X.; Kosari A.; Li S.; Akella D.; Ayorinde O.; Wentzloff D.; Calhoun B.H., "A Battery-less 507nW SoC with Integrated Platform Power Manager and SiP Interfaces", Symposia on VLSI Technology and Circuits (VLSIC), 2017.

[HNP4] Banerjee, A.; Liu, N.; **Patel, H.N.**; Poulton J.; Gray T.; Calhoun, B.H., "A 256kb 6T Self-Tuning SRAM with Extended 0.38V-1.2V Operating Range using Multiple

Read/Write Assists and VMIN Tracking Canary Sensors, IEEE Custom Integrated Circuits Conference (CICC), 2017.

[HNP5]   **Patel H. N.**, et al., "Ultra Low Leakage 55nm Deeply Depleted Channel Technology with Reverse Body Biasing and Subthreshold Operation to Minimize 6T SRAM and Logic Energy", 2016 46th European Solid-State Device Research Conference (ESSDERC), Lausanne, 2016, pp. 37-40.

[HNP6]   **Patel, H. N.**; Yahya F.; Calhoun B., "Optimizing SRAM Bitcell Reliability and Energy for IoT Applications", International Symposium on Quality Electronic Design (ISQED), Santa Clara, CA, 2016.

[HNP7]   Yahya, F.B.;**Patel, H.N.**; Boley, J.; Banerjee, A.; Calhoun, B.H., "A Sub-threshold 8T SRAM Macro with 12.29nW Standby Power and 6.24pJ/access for Battery-less SoCs , Journal of Low Power Electronics and Applications (JLPEA), 2016 6,8.

[HNP8]   **Patel, H.N.**, Yahya F.; Calhoun B. H., "Improving Reliability and Energy Requirements of Memory in Body Sensor Networks.", International Conference on VLSI Design (VLSID), Kolkata, India, 2016.

[HNP9]   Yahya, F.B.; **Patel,H. N.**; Chandra, V.; Calhoun, B. H., "Combined SRAM Read/Write Assist Techniques for Near/Sub-Threshold Voltage Operation, Quality Electronic Design (asQED), 2015 9th Asian Symposium on , 4-5 August 2015

[HNP10] Yahya, F.B.; **Patel,H. N.**; Chandra, V.; Calhoun, B. H., "Combined SRAM Read/Write Assist Techniques for Near/Sub-Threshold Voltage Operation, Design Automation Conference (DAC) , 7-11 June 2015 (accepted as work-in-progress)

[HNP11] Akella, D. K.; **Patel,H. N.**; Roy, Abhishek; Calhoun, B. H., "A 28 nW CMOS Supply Voltage Monitor for Adaptive Ultra-Low Power IoT Chips, IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), 2017.

Anticipated

[HNP12] **Patel H. N.**, et al., "A 55nm Ultra Low Leakage Deeply Depleted Channel Technology Optimized for Low-Power Internet of Everything".

[HNP13] **Patel, H.N.**; Akella D.; Calhoun B. H., "Improving PVT Variation Immunity using Digital Controller for Reliability and Energy Optimization for Subthreshold".

Appendix B

Acronyms

ASIC   Application Specific Integrated Circuits

BIST   Built-In Self Test

BL   Bitline

BSN   Body Sensor Node

CDF   Cumulative Distribution Function

CR   Cell Ratio

DDC   Deeply Depleted Channel

DRV   Data Retention Voltage

DMEM   Data Memory

FF   Fast NMOS; Fast PMOS process corner

FS   Fast NMOS; Fast PMOS process corner

HS   Half-Select bitcell

IMEM   Instruction Memory

IoT   Internet of Thing

MC   Monte Carlo

PD   Pull-down device

PG   Pass-gate device

PR   Pull-up Ratio

PU   Pull-up device

RBB   Reverse Body Biasing

RDF   Random Dopant Fluctuation

SA   Sense Amplifier

SER   Soft Error Rate

SF   Slow-NMOS, Fast-PMOS process corner

SNM   Static Noise Margin

SoC   System-on Chip

SOI   Silicon-on-Insulator

SRAM   Static Random Access Memory

SS   Slow-NMOS, Slow-PMOS process corner

Sub-V_T   sub-threshold

TEG   Thermoelectric generator

TT   Typical-NMOS, Typical-PMOS process corner

ULP   Ultra-Low Power

VDD   Supply Voltage

VMIN   Minimum operation voltage with reliable operations at a target yield

VT   Threshold Voltage of device

WL   Word-Line

WM   Write Margin

Bibliography

[1] H. Kim, S. Kim, N. Van Helleputte, A. Artes, M. Konijnenburg, J. Huisken, C. Van Hoof, and R. F. Yazicioglu. A configurable and low-power mixed signal soc for portable ecg monitoring applications. *IEEE Transactions on Biomedical Circuits and Systems*, 8(2):257–267, April 2014.

[2] A. Klinefelter, N.E. Roberts, Y. Shakhsheer, P. Gonzalez, A. Shrivastava, A. Roy, K. Craig, M. Faisal, J. Boley, Seunghyun Oh, Yanqing Zhang, D. Akella, D.D. Wentzloff, and B.H. Calhoun. 21.3 a 6.45 uw self-powered iot soc with integrated energy-harvesting power management and ulp asymmetric radios. In *Solid- State Circuits Conference - (ISSCC), 2015 IEEE International*, pages 1–3, Feb 2015.

[3] Naveen Verma. *Ultra-Low-Power SRAM Design In High Variability Advanced CMOS*. PhD thesis, Massachusetts Institute of Technology, 2009.

[4] Jacob Breiholz Abhishek Roy Harsh N. Patel NingXi Liu Xing Chen Avish Kosari Shuo Li Divya Akella Oluseyi Ayorinde David Wentzloff Benton H. Calhoun Farah Yahya, Christopher J. Lukas. A battery-less 507nw soc with integrated platform power manager and sip interfaces. In *2017 Symposium on VLSI Technology and Circuit*, July 2017.

[5] Shubu Mukherjee. *Architecture Design for Soft Errors*. 2008.

[6] R. C. Baumann. Radiation-induced soft errors in advanced semiconductor technologies. *IEEE Transactions on Device and Materials Reliability*, 5(3):305–316, Sept 2005.

[7] Cambridge. *Element Energy (EE): Pathways to high penetration of electric vehicles,*. PhD thesis, 2013.

[8] I. Korhonen, J. Parkka, and M. Van Gils. Health monitoring in the home of the future. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):66–73, May 2003.

[9] J. Kwong and A. P. Chandrakasan. An energy-efficient biomedical signal processing platform. *IEEE Journal of Solid-State Circuits*, 46(7):1742–1753, July 2011.

[10] E.J. Marinissen, B. Prince, D. Keltel-Schulz, and Y. Zorian. Challenges in embedded memory design and test. In *Design, Automation and Test in Europe, 2005. Proceedings*, pages 722–727 Vol. 2, March 2005.

[11] B.H. Calhoun, A. Wang, and A. Chandrakasan. Modeling and sizing for minimum energy operation in subthreshold circuits. *Solid-State Circuits, IEEE Journal of,* 40(9):1778–1786, Sept 2005.

[12] M. F. Chang, M. P. Chen, L. F. Chen, S. M. Yang, Y. J. Kuo, J. J. Wu, H. Y. Su, Y. H. Chu, W. C. Wu, T. Y. Yang, and H. Yamauchi. A sub-0.3 v area-efficient l-shaped 7t sram with read bitline swing expansion schemes based on boosted read-bitline, asymmetric-v_{rmTH} read-port, and offset cell vdd biasing techniques. *IEEE Journal of Solid-State Circuits*, 48(10):2558–2569, Oct 2013.

[13] S. Lutkemeier, T. Jungeblut, H. K. O. Berge, S. Aunet, M. Porrmann, and U. Ruckert. A 65 nm 32 b subthreshold processor with 9t multi-vt sram and adaptive supply voltage control. *IEEE Journal of Solid-State Circuits*, 48(1):8–19, Jan 2013.

[14] D. Kim, G. Chen, M. Fojtik, M. Seok, D. Blaauw, and D. Sylvester. A 1.85fw/bit ultra low leakage 10t sram with speed compensation scheme. In *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, pages 69–72, May 2011.

[15] P. Meinerzhagen, O. Andersson, B. Mohammadi, Y. Sherazi, A. Burg, and J. N. Rodrigues. A 500 fw/bit 14 fj/bit-access 4kb standard-cell based sub-vt memory in 65nm cmos. In *2012 Proceedings of the ESSCIRC (ESSCIRC)*, pages 321–324, Sept 2012.

[16] Y. Zhang, F. Zhang, Y. Shakhsheer, J. D. Silver, A. Klinefelter, M. Nagaraju, J. Boley, J. Pandey, A. Shrivastava, E. J. Carlson, A. Wood, B. H. Calhoun, and B. P. Otis. A batteryless 19 muw mics/ism-band energy harvesting body sensor node soc for exg applications. *IEEE Journal of Solid-State Circuits*, 48(1):199–213, Jan 2013.

[17] B. H. Calhoun, A. Wang, and A. Chandrakasan. Modeling and sizing for minimum energy operation in subthreshold circuits. *IEEE Journal of Solid-State Circuits*, 40(9):1778–1786, Sept 2005.

[18] N. Verma and A. P. Chandrakasan. A 256 kb 65 nm 8t subthreshold sram employing sense-amplifier redundancy. *IEEE Journal of Solid-State Circuits*, 43(1):141–149, Jan 2008.

[19] J. Kulkarni, M. Khellah, J. Tschanz, B. Geuskens, R. Jain, S. Kim, and V. De. Dual-vcc 8t-bitcell sram array in 22nm tri-gate cmos for energy-efficient operation across wide dynamic voltage range. In *2013 Symposium on VLSI Technology*, pages C126–C127, June 2013.

[20] A. Raychowdhury, B. Geuskens, J. Kulkarni, J. Tschanz, K. Bowman, T. Karnik, S. L. Lu, V. De, and M. M. Khellah. Pvt-and-aging adaptive wordline boosting for 8t sram power reduction. In *2010 IEEE International Solid-State Circuits Conference - (ISSCC)*, pages 352–353, Feb 2010.

[21] K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, and M. Bohr. A 3-ghz 70mb sram in 65nm cmos technology

with integrated column-based dynamic power supply. In *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*, pages 474–611 Vol. 1, Feb 2005.

[22] Randy W. Mann, Jiajing Wang, Satyanand Nalam, Sudhanshu Khanna, Geordie Braceras, Harold Pilo, and Benton H. Calhoun. Impact of circuit assist methods on margin and performance in 6t {SRAM}. *Solid-State Electronics*, 54(11):1398 – 1407, 2010.

[23] B. Zimmer, Seng Oon Toh, Huy Vo, Yunsup Lee, O. Thomas, K. Asanovic, and B. Nikolic. Sram assist techniques for operation in a wide voltage range in 28-nm cmos. *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 59(12):853–857, Dec 2012.

[24] H. N. Patel, F. B. Yahya, and B. H. Calhoun. Optimizing sram bitcell reliability and energy for iot applications. In *2016 17th International Symposium on Quality Electronic Design (ISQED)*, pages 12–17, March 2016.

[25] J. F. Ziegler. Terrestrial cosmic ray intensities. *IBM Journal of Research and Development*, 42(1):117–140, Jan 1998.

[26] E. Normand. Single-event effects in avionics. *IEEE Transactions on Nuclear Science*, 43(2):461–474, Apr 1996.

[27] E.J. Marinissen, B. Prince, D. Keltel-Schulz, and Y. Zorian. Challenges in embedded memory design and test. In *Design, Automation and Test in Europe, 2005. Proceedings*, pages 722–727 Vol. 2, March 2005.

[28] G. Z. Yang. *Body Sensor Networks*. Springer, 2006.

[29] American Heart Association. What is atrial fibrillation (afib or af)?

[30] H. N. Patel, A. Roy, F. B. Yahya, N. Liu, B. Calhoun, K. Kumeno, M. Yasuda, A. Harada, and T. Ema. A 55nm ultra low leakage deeply depleted channel technology optimized for energy minimization in subthreshold sram and logic. In *2016 46th European Solid-State Device Research Conference (ESSDERC)*, pages 37–40, Sept 2016.

[31] L. T. Clark, D. Zhao, T. Bakhishev, H. Ahn, E. Boling, M. Duane, K. Fujita, P. Gregory, T. Hoffmann, M. Hori, D. Kanai, D. Kidd, S. Lee, Y. Liu, J. Mitani, J. Nagayama, S. Pradhan, P. Ranade, R. Rogenmoser, L. Scudder, L. Shifren, Y. Torii, M. Wojko, Y. Asada, T. Ema, and S. Thompson. A highly integrated 65-nm soc process with enhanced power/performance of digital and analog circuits. In *Electron Devices Meeting (IEDM), 2012 IEEE International*, pages 14.4.1–14.4.4, Dec 2012.

[32] X. Chen, S. Samavedam, V. Narayanan, K. Stein, C. Hobbs, C. Baiocco, W. Li, D. Jaeger, M. Zaleski, H. S. Yang, N. Kim, Y. Lee, D. Zhang, L. Kang, J. Chen, H. Zhuang, A. Sheikh, J. Wallner, M. Aquilino, J. Han, Z. Jin, J. Li, G. Massey,

S. Kalpat, R. Jha, N. Moumen, R. Mo, S. Kirshnan, X. Wang, M. Chudzik, M. Chowdhury, D. Nair, C. Reddy, Y. W. Teh, C. Kothandaraman, D. Coolbaugh, S. Pandey, D. Tekleab, A. Thean, M. Sherony, C. Lage, J. Sudijono, R. Lindsay, J. H. Ku, M. Khare, and A. Steegen. A cost effective 32nm high-k/ metal gate cmos technology for low power applications with single-metal/gate-first process. In *2008 Symposium on VLSI Technology*, pages 88–89, June 2008.

[33] F. Hamzaoglu, K. Zhang, Y. Wang, H. J. Ahn, U. Bhattacharya, Z. Chen, Y. G. Ng, A. Pavlov, K. Smits, and M. Bohr. A 153mb-sram design with dynamic stability enhancement and leakage reduction in 45nm high-k metal-gate cmos technology. In *2008 IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, pages 376–621, Feb 2008.

[34] G. Tsutsui, K. Tsunoda, N. Kariya, Y. Akiyama, T. Abe, S. Maruyama, T. Fukase, M. Suzuki, Y. Yamagata, and K. Imai. Reduction of vth variation by work function optimization for 45-nm node sram cell. In *2008 Symposium on VLSI Technology*, pages 158–159, June 2008.

[35] K. Cheng, A. Khakifirooz, P. Kulkarni, S. Ponoth, B. Haran, A. Kumar, T. Adam, A. Reznicek, N. Loubet, H. He, J. Kuss, M. Wang, T. M. Levin, F. Monsieur, Q. Liu, R. Sreenivasan, J. Cai, A. Kimball, S. Mehta, S. Luning, Y. Zhu, Z. Zhu, T. Ya-mamoto, A. Bryant, C. H. Lin, S. Naczas, H. Jagannathan, L. F. Edge, S. Allegret-Maret, A. Dube, S. Kanakasabapathy, S. Schmitz, A. Inada, S. Seo, M. Raymond, Z. Zhang, A. Yagishita, J. Demarest, J. Li, M. Hopstaken, N. Berliner, A. Upham, R. Johnson, S. Holmes, T. Standaert, M. Smalley, N. Zamdmer, Z. Ren, T. Wu, H. Bu, V. Paruchuri, D. Sadana, V. Narayanan, W. Haensch, J. O'Neill, T. Hook, M. Khare, and B. Doris. Etsoi cmos for system-on-chip applications featuring 22nm gate length, sub-100nm gate pitch, and 0.08um2 sram cell. In *2011 Symposium on VLSI Circuits - Digest of Technical Papers*, pages 128–129, June 2011.

[36] K. J. Kuhn. Cmos scaling for the 22nm node and beyond: Device physics and technology. In *Proceedings of 2011 International Symposium on VLSI Technology, Systems and Applications*, pages 1–2, April 2011.

[37] L. T. Clark, D. Zhao, T. Bakhishev, H. Ahn, E. Boling, M. Duane, K. Fujita, P. Gregory, T. Hoffmann, M. Hori, D. Kanai, D. Kidd, S. Lee, Y. Liu, J. Mitani, J. Nagayama, S. Pradhan, P. Ranade, R. Rogenmoser, L. Scudder, L. Shifren, Y. Torii, M. Wojko, Y. Asada, T. Ema, and S. Thompson. A highly integrated 65-nm soc process with enhanced power-performance of digital and analog circuits. In *2012 International Electron Devices Meeting*, pages 14.4.1–14.4.4, Dec 2012.

[38] K. Fujita, Y. Torii, M. Hori, J. Oh, L. Shifren, P. Ranade, M. Nakagawa, K. Okabe, T. Miyake, K. Ohkoshi, M. Kuramae, T. Mori, T. Tsuruta, S. Thompson, and T. Ema. Advanced channel engineering achieving aggressive reduction of vt variation for ultra-low-power applications. In *2011 International Electron Devices Meeting*, pages 32.3.1–32.3.4, Dec 2011.

[39] N. Kimizuka, Y. Yasuda, T. Iwamoto, I. Yamamoto, K. Takano, Y. Akiyama, and K. Imai. Ultra-low standby power (u-lstp) 65-nm node cmos technology utilizing hfsion dielectric and body-biasing scheme. In *Digest of Technical Papers. 2005 Symposium on VLSI Technology, 2005.*, pages 218–219, June 2005.

[40] M. Yamaoka, N. Maeda, Y. Shimazaki, and K. Osada. 65nm low-power high-density sram operable at 1.0v under 3sigma systematic variation using separate vth monitoring and body bias for nmos and pmos. In *2008 IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, pages 384–622, Feb 2008.

[41] W. H. Ma, J. C. Kao, V. S. Sathe, and M. Papaefthymiou. A 187mhz subthreshold-supply robust fir filter with charge-recovery logic. In *2009 Symposium on VLSI Circuits*, pages 202–203, June 2009.

[42] H. N. Patel, F. B. Yahya, and B. H. Calhoun. Improving reliability and energy requirements of memory in body sensor networks. In *2016 29th International Conference on VLSI Design and 2016 15th International Conference on Embedded Systems (VLSID)*, pages 561–562, Jan 2016.

[43] Hong Zhu and V. Kursun. A comprehensive comparison of data stability enhancement techniques with novel nanoscale sram cells under parameter fluctuations. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 61(5):1473–1484, May 2014.

[44] V. Sharma, S. Cosemans, M. Ashouie, J. Huisken, F. Catthoor, and W. Dehaene. Ultra low-energy sram design for smart ubiquitous sensors. *Micro, IEEE*, 32(5):10–24, Sept 2012.

[45] E. Seevinck, F.J. List, and J. Lohstroh. Static-noise margin analysis of mos sram cells. *Solid-State Circuits, IEEE Journal of*, 22(5):748–754, Oct 1987.

[46] K. Leochico and E. John. Data retention voltage analysis of various low-power sram topologies. In *Circuits and Systems (MWSCAS), 2014 IEEE 57th International Midwest Symposium on*, pages 913–916, Aug 2014.

[47] J. Boley, P. Beshay, and B. Calhoun. Virtual prototyper (vipro): An sram design tool for yield constrained optimization. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 23(12):3109–3113, Dec 2015.

[48] J. Kwong and A. P. Chandrakasan. An energy-efficient biomedical signal processing platform. *IEEE Journal of Solid-State Circuits*, 46(7):1742–1753, July 2011.

[49] A. Roy, P. J. Grossmann, S. A. Vitale, and B. H. Calhoun. A 1.3 uw, 5pj/cycle subthreshold msp430 processor in 90nm xlp fdsoi for energy-efficient iot applications. In *2016 17th International Symposium on Quality Electronic Design (ISQED)*, pages 158–162, March 2016.

[50] B. H. Calhoun and A. P. Chandrakasan. Standby power reduction using dynamic voltage scaling and canary flip-flop structures. *IEEE Journal of Solid-State Circuits*, 39(9):1504–1511, Sept 2004.

[51] J. Wang, A. Hoefler, and B. H. Calhoun. An enhanced canary-based system with bist for sram standby power reduction. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 19(5):909–914, May 2011.

[52] M. Qazi, K. Stawiasz, L. Chang, and A. Chandrakasan. A 512kb 8t sram macro operating down to 0.57v with an ac-coupled sense amplifier and embedded data-retention-voltage sensor in 45nm soi cmos. In *2010 IEEE International Solid-State Circuits Conference - (ISSCC)*, pages 350–351, Feb 2010.

[53] E. Karl, Y. Wang, Y. G. Ng, Z. Guo, F. Hamzaoglu, U. Bhattacharya, K. Zhang, K. Mistry, and M. Bohr. A 4.6ghz 162mb sram design in 22nm tri-gate cmos technology with integrated active vmin-enhancing assist circuitry. In *2012 IEEE International Solid-State Circuits Conference*, pages 230–232, Feb 2012.

[54] Y. Sinangil, S. M. Neuman, M. E. Sinangil, N. Ickes, G. Bezerra, E. Lau, J. E. Miller, H. C. Hoffmann, S. Devadas, and A. P. Chandraksan. A self-aware processor soc using energy monitors integrated into power converters for self-adaptation. In *2014 Symposium on VLSI Circuits Digest of Technical Papers*, pages 1–2, June 2014.

[55] M. F. Chang, C. F. Chen, T. H. Chang, C. C. Shuai, Y. Y. Wang, and H. Yamauchi. 17.3 a 28nm 256kb 6t-sram with 280mv improvement in vmin using a dual-split-control assist scheme. In *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pages 1–3, Feb 2015.

[56] A. Feki, B. Allard, D. Turgis, J. Lafont, and L. Ciampolini. Proposal of a new ultra low leakage 10t sub threshold sram bitcell. In *SoC Design Conference (ISOCC), 2012 International*, pages 470–474, Nov 2012.

[57] G. Harutunvan, V.A. Vardanian, and Y. Zorian. Minimal march tests for unlinked static faults in random access memories. In *VLSI Test Symposium, 2005. Proceedings. 23rd IEEE*, pages 53–59, May 2005.

[58] G. Chen, D. Sylvester, D. Blaauw, and T. Mudge. Yield-driven near-threshold sram design. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 18(11):1590–1598, Nov 2010.

[59] James Boley, Vikas Chandra, Robert Aitken, and Benton Calhoun. Leveraging sensitivity analysis for fast, accurate estimation of sram dynamic write vmin. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2013*, pages 1819–1824, March 2013.

[60] S. Nalam, M. Bhargava, K. Ringgenberg, Ken Mai, and B.H. Calhoun. A technology-agnostic simulation environment (tase) for iterative custom ic design across processes. In *Computer Design, 2009. ICCD 2009. IEEE International Conference on*, pages 523–528, Oct 2009.

[61] A. Banerjee, J. Breiholz, and B. H. Calhoun. A 130nm canary sram for sram dynamic write vmin tracking across voltage, frequency, and temperature variations. In *2015 IEEE Custom Integrated Circuits Conference (CICC)*, pages 1–4, Sept 2015.

[62] B. H. Calhoun and A. Chandrakasan. Characterizing and modeling minimum energy operation for subthreshold circuits. In *Proceedings of the 2004 International Symposium on Low Power Electronics and Design (IEEE Cat. No.04TH8758)*, pages 90–95, Aug 2004.

[63] Bo Zhai, D. Blaauw, D. Sylvester, and K. Flautner. Theoretical and practical limits of dynamic voltage scaling. In *Proceedings. 41st Design Automation Conference, 2004.*, pages 868–873, July 2004.

[64] P. E. Dodd and L. W. Massengill. Basic mechanisms and modeling of single-event upset in digital microelectronics. *IEEE Transactions on Nuclear Science*, 50(3):583–602, June 2003.

[65] T. Karnik and P. Hazucha. Characterization of soft errors caused by single event upsets in cmos processes. *IEEE Transactions on Dependable and Secure Computing*, 1(2):128–143, April 2004.

[66] P. Hazucha, T. Karnik, J. Maiz, S. Walstra, B. Bloechel, J. Tschanz, G. Dermer, S. Hareland, P. Armstrong, and S. Borkar. Neutron soft error rate measurements in a 90-nm cmos process and scaling trends in sram from 0.25-/spl mu/m to 90-nm generation. In *IEEE International Electron Devices Meeting 2003*, pages 21.5.1–21.5.4, Dec 2003.

[67] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester, and D. Blaauw. Exploring variability and performance in a sub-200-mv processor. *IEEE Journal of Solid-State Circuits*, 43(4):881–891, April 2008.

[68] Bo Zhai, S. Hanson, D. Blaauw, and D. Sylvester. Analysis and mitigation of variability in subthreshold design. In *ISLPED '05. Proceedings of the 2005 International Symposium on Low Power Electronics and Design, 2005.*, pages 20–25, Aug 2005.

[69] S. Mukhopadhyay, H. Mahmoodi, and K. Roy. Reduction of parametric failures in sub-100-nm sram array using body bias. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 27(1):174–183, Jan 2008.

[70] M. Orshansky, L. Milor, Pinhong Chen, K. Keutzer, and Chenming Hu. Impact of spatial intrachip gate length variability on the performance of high-speed digital circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 21(5):544–553, May 2002.

[71] Rajeev R. Rao, A. Devgan, D. Blaauw, and D. Sylvester. Parametric yield estimation considering leakage variability. In *Proceedings. 41st Design Automation Conference, 2004.*, pages 442–447, July 2004.

[72] Seung Hoon Choi, B. C. Paul, and K. Roy. Novel sizing algorithm for yield improvement under process variation in nanometer technology. In *Proceedings. 41st Design Automation Conference, 2004.*, pages 454–459, July 2004.

[73] A. T. Do, Z. C. Lee, B. Wang, I. J. Chang, X. Liu, and T. T. H. Kim. 0.2 v 8t sram with pvt-aware bitline sensing and column-based data randomization. *IEEE Journal of Solid-State Circuits*, 51(6):1487–1498, June 2016.

[74] S. T. Eid, M. Whately, and S. Krishnegowda. A microcontroller-based pvt control system for a 65nm 72mb synchronous sram. In *2010 IEEE International Solid-State Circuits Conference - (ISSCC)*, pages 184–185, Feb 2010.

[75] Harsh N. Patel John Poulton C. Thomas Gray Arijit Benerjee, NingXi Liu and Benton H. Calhoun. A 256kb 6t self-tuning sram with extended 0.38v-1.2v operating range using multiple read/write assists and vmin tracking canary sensors. In *2017 IEEE Custom Integrated Circuits Conference*, July 2017.

[76] S. Ghosh and K. Roy. Parameter variation tolerance and error resiliency: New design paradigm for the nanoscale era. *Proceedings of the IEEE*, 98(10):1718–1751, Oct 2010.

[77] F. B. Yahya, H. N. Patel, V. Chandra, and B. H. Calhoun. Combined sram read-/write assist techniques for near/sub-threshold voltage operation. In *2015 6th Asia Symposium on Quality Electronic Design (ASQED)*, pages 1–6, Aug 2015.

[78] S. Sayil and J. Wang. Single-event soft errors in cmos logic. *IEEE Potentials*, 31(2):15–22, March 2012.

[79] P. Royer, F. Garca-Redondo, and M. Lpez-Vallejo. Evolution of radiation-induced soft errors in finfet srams under process variations beyond 22nm. In *Nanoscale Architectures (NANOARCH), 2015 IEEE/ACM International Symposium on*, pages 112–117, July 2015.

[80] M. C. Casey, B. L. Bhuva, S. A. Nation, O. A. Amusan, T. D. Loveless, L. W. Massengill, D. McMorrow, and J. S. Melinger. Single-event effects on ultra-low power cmos circuits. In *Reliability Physics Symposium, 2009 IEEE International*, pages 194–198, April 2009.

[81] R. Garg and S. P. Khatri. 3d simulation and analysis of the radiation tolerance of voltage scaled digital circuit. In *Computer Design, 2009. ICCD 2009. IEEE International Conference on*, pages 498–504, Oct 2009.

[82] A. Balasubramanian, P. R. Fleming, B. L. Bhuva, O. A. Amusan, and L. W. Massengill. Effects of random dopant fluctuations (rdf) on the single event vulnerability of 90 and 65 nm cmos technologies. *IEEE Transactions on Nuclear Science*, 54(6):2400–2406, Dec 2007.

[83] T. Heijmen, D. Giot, and P. Roche. Factors that impact the critical charge of memory elements. In *On-Line Testing Symposium, 2006. IOLTS 2006. 12th IEEE International*, pages 6 pp.–, 2006.

[84] V. Chandra and R. Aitken. Impact of voltage scaling on nanoscale sram reliability. In *Design, Automation Test in Europe Conference Exhibition, 2009. DATE '09.*, pages 387–392, April 2009.

[85] J. T. Wallmark and S. M. Marcus. Minimum size and maximum packing density of nonredundant semiconductor devices. *Proceedings of the IRE*, 50(3):286–298, March 1962.

[86] R. Garg and S. P. Khatri. 3d simulation and analysis of the radiation tolerance of voltage scaled digital circuit. In *Computer Design, 2009. ICCD 2009. IEEE International Conference on*, pages 498–504, Oct 2009.

[87] P. Royer, F. Garca-Redondo, and M. Lpez-Vallejo. Evolution of radiation-induced soft errors in finfet srams under process variations beyond 22nm. In *Nanoscale Architectures (NANOARCH), 2015 IEEE/ACM International Symposium on*, pages 112–117, July 2015.

[88] A. Balasubramanian, P. R. Fleming, B. L. Bhuva, O. A. Amusan, and L. W. Massengill. Effects of random dopant fluctuations (rdf) on the single event vulnerability of 90 and 65 nm cmos technologies. *IEEE Transactions on Nuclear Science*, 54(6):2400–2406, Dec 2007.

[89] Arizona State University PTM Libraries. Ptm libraries, arizona state university.

[90] M. C. Casey, B. L. Bhuva, S. A. Nation, O. A. Amusan, T. D. Loveless, L. W. Massengill, D. McMorrow, and J. S. Melinger. Single-event effects on ultra-low power cmos circuits. In *Reliability Physics Symposium, 2009 IEEE International*, pages 194–198, April 2009.

[91] J. F. Ziegler. Terrestrial cosmic ray intensities. *IBM Journal of Research and Development*, 42(1):117–140, Jan 1998.

[92] S. Mitra, M. Zhang, N. Seifert, T. Mak, and K. S. Kim. Soft error resilient system design through error correction. In *2006 IFIP International Conference on Very Large Scale Integration*, pages 332–337, Oct 2006.

[93] S. Sriram, H. Nan, and K. Choi. Low power latch design in near sub-threshold region to improve reliability for soft error. In *2011 12th International Symposium on Quality Electronic Design*, pages 1–4, March 2011.

[94] T. Heijmen, D. Giot, and P. Roche. Factors that impact the critical charge of memory elements. In *On-Line Testing Symposium, 2006. IOLTS 2006. 12th IEEE International*, pages 6 pp.–, 2006.

[95] S. Raasch, A. Biswas, J. Stephan, P. Racunas, and J. Emer. A fast and accurate analytical technique to compute the avf of sequential bits in a processor. In *2015 48th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, pages 738–749, Dec 2015.

[96] Y. C. Lai, S. Y. Huang, and H. J. Hsu. Resilient self-v$_{rmDD}$-tuning scheme with speed-margining for low-power sram. *IEEE Journal of Solid-State Circuits*, 44(10):2817–2823, Oct 2009.

[97] N. Verma and A. P. Chandrakasan. A 256kb 65 nm 8t subthreshold sram employing sense-amplifier redundancy. *IEEE Journal of Solid-State Circuits*, 43(1):141–149, Jan 2008.

[98] S. Lutkemeier, T. Jungeblut, H. K. O. Berge, S. Aunet, M. Porrmann, and U. Ruckert. A 65 nm 32 b subthreshold processor with 9t multi-vt sram and adaptive supply voltage control. *IEEE Journal of Solid-State Circuits*, 48(1):8–19, Jan 2013.

[99] P. Meinerzhagen, O. Andersson, B. Mohammadi, Y. Sherazi, A. Burg, and J. N. Rodrigues. A 500 fw/bit 14 fj/bit-access 4kb standard-cell based sub-vt memory in 65nm cmos. In *2012 Proceedings of the ESSCIRC (ESSCIRC)*, pages 321–324, Sept 2012.

[100] M. E. Sinangil, N. Verma, and A. P. Chandrakasan. A reconfigurable 65nm sram achieving voltage scalability from 0.25 to 1.2v and performance scalability from 20khz 200mhz. In *ESSCIRC 2008 - 34th European Solid-State Circuits Conference*, pages 282–285, Sept 2008.

[101] Dave Evans. The internet of things how the next evolution of the internet is changing everything.

[102] S. Lutkemeier, T. Jungeblut, H. K. O. Berge, S. Aunet, M. Porrmann, and U. Ruckert. A 65 nm 32 b subthreshold processor with 9t multi-vt sram and adaptive supply voltage control. *IEEE Journal of Solid-State Circuits*, 48(1):8–19, Jan 2013.

[103] B.H. Calhoun, A. Wang, and A. Chandrakasan. Modeling and sizing for minimum energy operation in subthreshold circuits. *Solid-State Circuits, IEEE Journal of*, 40(9):1778–1786, Sept 2005.

[104] M.E. Sinangil, H. Mair, and A.P. Chandrakasan. A 28nm high-density 6t sram with optimized peripheral-assist circuits for operation down to 0.6v. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pages 260–262, Feb 2011.

[105] N. Verma and A.P. Chandrakasan. A 256 kb 65 nm 8t subthreshold sram employing sense-amplifier redundancy. *Solid-State Circuits, IEEE Journal of*, 43(1):141–149, Jan 2008.

[106] A. Wang and A. Chandrakasan. A 180mv fft processor using subthreshold circuit techniques. In *Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC. 2004 IEEE International*, pages 292–529 Vol.1, Feb 2004.

[107] S. Lutkemeier, T. Jungeblut, H.K.O. Berge, S. Aunet, M. Porrmann, and U. Ruckert. A 65 nm 32 b subthreshold processor with 9t multi-vt sram and adaptive supply voltage control. *Solid-State Circuits, IEEE Journal of*, 48(1):8–19, Jan 2013.

[108] Jun Zhou, S. Jayapal, B. Busze, Li Huang, and J. Stuyt. A 40 nm inverse-narrow-width-effect-aware sub-threshold standard cell library. In *Design Automation Conference (DAC), 2011 48th ACM/EDAC/IEEE*, pages 441–446, June 2011.

[109] N. Verma and A.P. Chandrakasan. A 256 kb 65 nm 8t subthreshold sram employing sense-amplifier redundancy. *Solid-State Circuits, IEEE Journal of*, 43(1):141–149, Jan 2008.

[110] K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, and M. Bohr. A 3-ghz 70mb sram in 65nm cmos technology with integrated column-based dynamic power supply. In *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*, pages 474–611 Vol. 1, Feb 2005.

[111] A. Raychowdhury, B. Geuskens, J. Kulkarni, J. Tschanz, K. Bowman, T. Karnik, Shih-Lien Lu, V. De, and M.M. Khellah. Pvt-and-aging adaptive wordline boosting for 8t sram power reduction. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, pages 352–353, Feb 2010.

[112] James Boley, Vikas Chandra, Robert Aitken, and Benton Calhoun. Leveraging sensitivity analysis for fast, accurate estimation of sram dynamic write vmin. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2013*, pages 1819–1824, March 2013.

[113] I. J. Chang, J. J. Kim, S. P. Park, and K. Roy. A 32 kb 10t sub-threshold sram array with bit-interleaving and differential read scheme in 90 nm cmos. *IEEE Journal of Solid-State Circuits*, 44(2):650–658, Feb 2009.

[114] T. H. Kim, J. Liu, J. Keane, and C. H. Kim. A high-density subthreshold sram with data-independent bitline leakage and virtual ground replica scheme. In *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, pages 330–606, Feb 2007.

[115] K. Fujita, Y. Torii, M. Hori, J. Oh, L. Shifren, P. Ranade, M. Nakagawa, K. Okabe, T. Miyake, K. Ohkoshi, M. Kuramae, T. Mori, T. Tsuruta, S. Thompson, and T. Ema. Advanced channel engineering achieving aggressive reduction of vt variation for ultra-low-power applications. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 32.3.1–32.3.4, Dec 2011.

[116] N. Kimizuka, Y. Yasuda, T. Iwamoto, I. Yamamoto, K. Takano, Y. Akiyama, and K. Imai. Ultra-low standby power (u-lstp) 65-nm node cmos technology utilizing hfsion dielectric and body-biasing scheme. In *VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on*, pages 218–219, June 2005.

[117] V. Chandra and R. Aitken. Impact of voltage scaling on nanoscale sram reliability. In *Design, Automation Test in Europe Conference Exhibition, 2009. DATE '09.*, pages 387–392, April 2009.

[118] S. Sayil and J. Wang. Single-event soft errors in cmos logic. *IEEE Potentials*, 31(2):15–22, March 2012.

[119] E. Normand. Single-event effects in avionics. *IEEE Transactions on Nuclear Science*, 43(2):461–474, Apr 1996.

[120] S. Shambhulingaiah, C. Lieb, and L. T. Clark. Circuit simulation based validation of flip-flop robustness to multiple node charge collection. *IEEE Transactions on Nuclear Science*, 62(4):1577–1588, Aug 2015.

[121] R. C. Baumann. Radiation-induced soft errors in advanced semiconductor technologies. *IEEE Transactions on Device and Materials Reliability*, 5(3):305–316, Sept 2005.

[122] T. Karnik and P. Hazucha. Characterization of soft errors caused by single event upsets in cmos processes. *IEEE Transactions on Dependable and Secure Computing*, 1(2):128–143, April 2004.

[123] J. T. Wallmark and S. M. Marcus. Minimum size and maximum packing density of nonredundant semiconductor devices. *Proceedings of the IRE*, 50(3):286–298, March 1962.

[124] P. Hazucha, T. Karnik, J. Maiz, S. Walstra, B. Bloechel, J. Tschanz, G. Dermer, S. Hareland, P. Armstrong, and S. Borkar. Neutron soft error rate measurements in a 90-nm cmos process and scaling trends in sram from 0.25-/spl mu/m to 90-nm generation. In *IEEE International Electron Devices Meeting 2003*, pages 21.5.1–21.5.4, Dec 2003.

[125] P. E. Dodd and L. W. Massengill. Basic mechanisms and modeling of single-event upset in digital microelectronics. *IEEE Transactions on Nuclear Science*, 50(3):583–602, June 2003.