Subthreshold SRAM: Challenges, Design Decisions, and Solutions

Harsh N. Patel, Farah B. Yahya, Benton H. Calhoun

Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, USA
E-mail: {hnpatel, fby5bb, bcalhoun}@virginia.edu

Abstract— **This paper presents an overview of various challenges, optimization strategies, and design requirements for subthreshold SRAM arrays targeting Ultra-Low Power (ULP) applications in the Internet of Things (IoTs). We study the impact of threshold voltage (V_T) change due to process and temperature variations on various SRAM design decisions for ULP operation. We explore different solutions to enable reliable subthreshold operation ranging from technology to cell to architecture and assist. We also highlight the impact of process variations on optimal peripheral assist selection, and degree of assist requirements. We present trade-offs between reliability, energy, and performance to an application-specific SRAM design. Six different types of SRAM bitcells are compared for various subthreshold metrics to provide an optimal bitcell selection for the targeted application.**

Keywords— *body-biasing, margin, SRAM, Subthreshold, variation;*

I. INTRODUCTION

The number of ubiquitous sensors has increased to almost thrice the human population on the planet and is expected to continue growing in the near future [1]. The pervasive use of personal healthcare, bio-sensing, and remote sensing such as Body Sensor Networks (BSNs) and other Internet of Things (IoTs) platforms presents a bigger challenge for sustainability under available power source. With the predicted growth of 50 billion devices to be connected to the Internet by 2020 [1], recharging or replacing batteries at a regular interval will result in enormous time and cost overhead. Therefore, modern System-on-Chips (SoCs) typically have a stringent constraint on power – they either have a limited battery source or harvest energy from the ambient environment – and thus operating supply voltage (V_{DD}) is scaled down to subthreshold voltages to permit quadratic savings in active digital energy (CV_{DD}^2) [2]. Additionally, the major criteria for such applications change from the traditional performance-driven to reliability, low-energy, and longer battery life. To address these challenges, it has become critical to re-evaluate the different design decisions made under high-performance requirements and to re-validate any different trade-offs among the metrics under consideration.

The authors in [3] highlighted the Static Random Access Memory (SRAM) as one of the major contributors to the power dissipation of the digital design in Ultra-Low Power (ULP) SoCs. The demand for extended capabilities requiring larger embedded memories (mainly SRAM) is increasing in highly integrated SoCs (70-80% of the total chip area) [3], thus further tightening the design constraints on power, performance, and energy.

This paper presents an overview of the design challenges and various solutions for an ULP SRAM design targeting low-power IoT platforms. We explore various design considerations to address the subthreshold SRAM challenges such as fabrication technology, choice of an SRAM bitcell (6T vs. 8T), and peripheral assist techniques required to optimize the subthreshold SRAM design. Various subthreshold design challenges are discussed in Section II. The changes in the design metric consideration as a result of V_{DD} scaling – including optimal assist – are explored in Section III. Section V concludes the paper.

II. SUBTHRESHOLD DESIGN CHALLENGES

Subthreshold SRAM faces additional challenges compared to super-threshold SRAM. In [5], 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. Fig. 1 demonstrates the impact of these challenges on the minimum operating voltage (V_{MIN}) of the SRAM. In this plot, we show 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), even though a relaxed clock frequency (~20KHz) is provided, write delay (WD) still limits V_{MIN} (= 0.65V). On the other hand, half-select (HS) failures – measured by the RSNM or dynamic HS – limit V_{MIN} for FS and FF corners. While HS is seldom a concern for nominal operation, it becomes a critical concern for subthreshold designs operating closer to V_{MIN}.
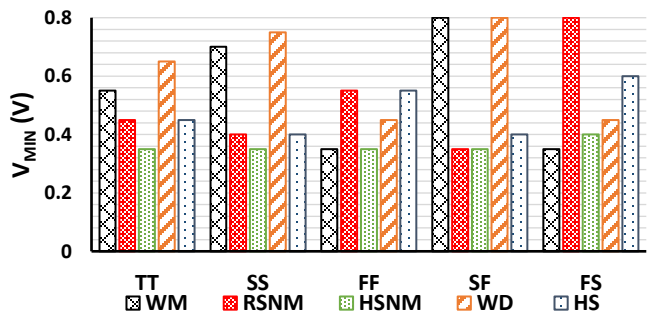

Fig. 1. Limitation on the SRAM V_{MIN} imposed by different static and dynamic metrics.

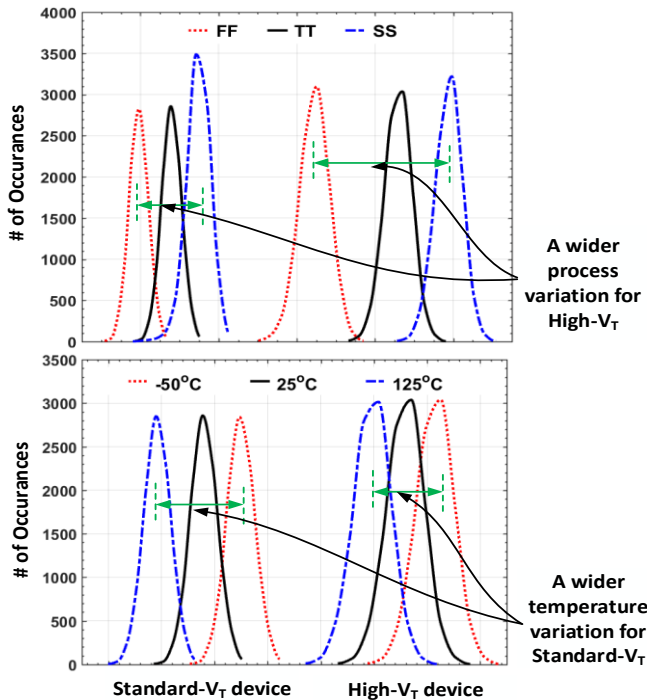Fig. 2. Distribution of V_T variation across process and temperature: The high-V_T devices experience higher process variation (top); while the standard-V_T devices have a larger temperature variation (bottom).

In [6], the authors proposed a subthreshold SRAM with reduced leakage using high-V_T devices for the array and standard-V_T devices for the peripheral circuits to ensure timely signal generation. While using high-V_T devices reduces leakage these devices suffer from increased variations. Fig. 2. shows the V_T distributions (i.e., local variation) of standard-V_T and high-V_T devices across three corners (i.e., global variation) and over an extended range of temperatures based on 10000 Monte-Carlo simulations. As shown in Fig. 2, high-V_T devices experience wider variation due to process. For ratio-ed design such as an SRAM, the functionality of the circuit is largely based on the relative strength of the devices. Therefore, the leakage reduction in the SRAM array using high-V_T devices demands peripheral assist techniques to ensure the functionality across process corners. Similarly, the standard-V_T devices are susceptible to huge variations due to temperature (Fig. 2 bottom) that induce timing failures.

At nominal V_{DD}, the device sizing is controlled to ensure the functionality. However, at lower V_{DDS}, I_{ON} does not increase linearly with the width of the transistors due to Inverse Narrow Width Effect (INWE) [8] as shown in Fig. 3. Therefore, at subthreshold voltages, sizing is a weak knob to control I_{ON}. Fig. 3 indicates that at $V_{DD} = 1.2V$ I_{ON} can be increased linearly with sizing (increasing W) for the selected technology. In contrary, a similar increase in I_{ON} at $V_{DD} = 0.2V$ 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
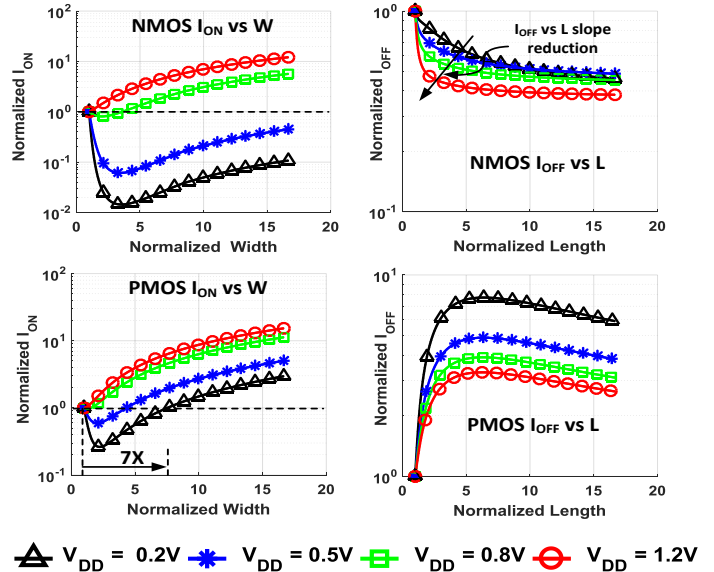

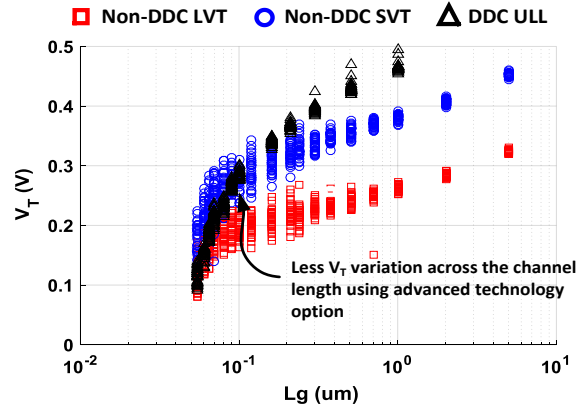Fig. 3. Sizing in subthreshold is very weak knob to control the I_{ON} and I_{OFF}

Fig. 4. Technology such as DDC is optimized for subthreshold operation where the V_T roll-of over a wide channel length is reduced significantly[6].

fixed I_{OFF} reduction. Therefore, the subthreshold SRAM design requires new design knob to reduce the leakage. In addition to the impact of variation and sizing being a weak knob in the subthreshold, a major concern for the reliability in subthreshold is particle strike-induced soft errors. The magnitude of the impact on circuits due to radiation varies from a temporary change in the storage to a complete application failure [12].

III. DESIGN CONSIDERATION AND SOLUTIONS

To address the subthreshold challenges discussed in the previous section, we consider technology, bitcell design, assist selection, and device type as design parameters for a reliable and energy optimal SRAM design.

A. Technology

Advanced technologies such as Ultra Low Leakage (ULL) Deeply Depleted Channel (DDC) [7], optimizes V_T variations. Therefore, such technologies can allow a reliable subthreshold SRAM for ULP applications. Fig. 4 compares V_T roll-off of
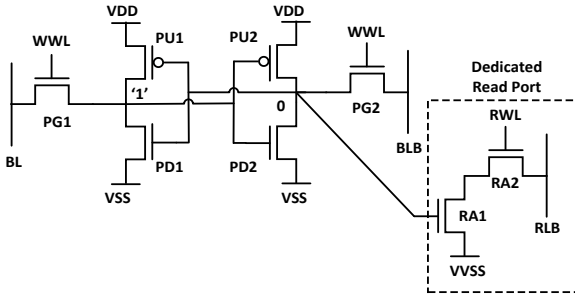
Fig. 5. Conventional 6T with dedicated read port in an 8T bitcell.

DDC ULL devices with conventional standard (SVT) and Low V_T (LVT) devices. ULL DDC shows a strong control over V_T across a wide range of channel lengths. Reduced V_T variation ensures reliable operations for the subthreshold SRAM.

B. Bit-cell

After considering an advanced fabrication technology for the subthreshold SRAM design, we explore the type of cell a design knob. From Fig. 1, we show that HS failures in subthreshold play crucial role in SRAM V_{MIN} reduction. Since HS failures occur during read and write operations in a 6T cell, an 8T design is proposed to decouple read and write operations. While an 8T bitcell decouples read from write and addresses the HS problem during read operations, it still faces write HS problems. However, several solutions were proposed to address HS in 8T cells, such as employing a read-before-write approach [6], applying read assist techniques [8], or other non-6T bitcell designs [4][9][10]. Due to an area optimized design, an 8T bitcell is considered as an optimal choice for subthreshold SRAM. Fig. 5 shows the 8T SRAM bitcell with a pull-up (PUs), pull-down (PDs), and pass-gate (PGs) devices.

C. Peripheral Assist Technique

Peripheral assist techniques are used to ensure a reliable operation. The authors in [11][12] evaluated various SRAM peripheral assist techniques for a broad range of supply voltages and different metric considerations. Fig. 6a) shows the sensitivity of the WM to V_T of each of the six transistors. At super-threshold voltage ($V_{DD} = 0.8V$), the WM is very sensitive to changes in the V_T of device PGR (Fig. 5). Devices PUR, PDL, and PGL also impact the WM, while PUL and PDR have a negligible impact on WM. Therefore, at super-threshold voltages, peripheral assist targeting PGR (WL Boosting and NegBL) improves the write margin (Fig. 6b). NegBL enhances the strength of PGR only and thus gives a lower WM than WL Boosting. In contrary, at a subthreshold voltage ($V_{DD} = 0.4V$), the sensitivity of WM to PUR and PDL increases and hence V_{DD} lowering (Fig. 6b) proves to be more efficient for the write margin improvement here.

Interestingly, the increasing degree of assist (10-40% of V_{DD}) in subthreshold also changes the optimal choice for the assist technique as shown in (Fig. 6b) where the V_{DD} lowering assist improves WM more than WL boosting with an increase in applied assist due to the same reason discussed before. While Fig. 6b shows the impact of assist on WM for the worst case process corner, Fig. 7 shows the required percentage of assist
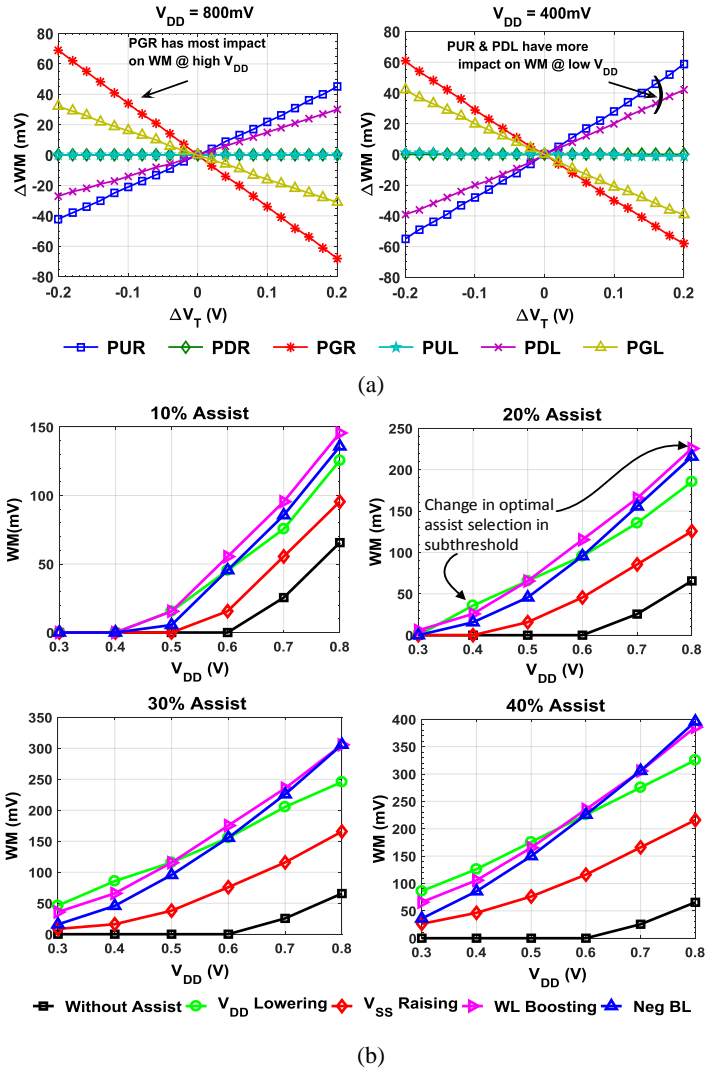

(a)

(b)

Fig. 6. Assist selection in subthreshold operation changes from the super-threshold operation: a) WM sensitivity to V_T variation changes in subthreshold operation [3] b) The impact of different assist technique [12].
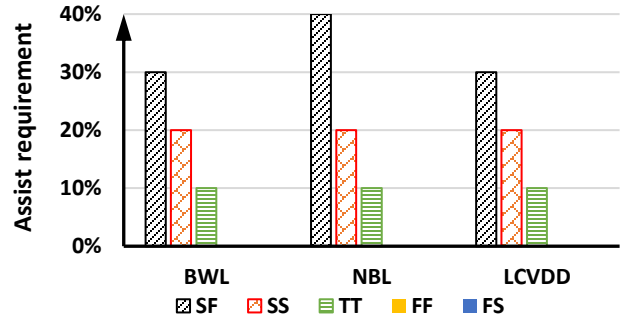

Fig. 7. Percentage (of V_{DD}) assist required to achieve a V_{MIN} of 0.5V. Fast corners (FS, FF) do not need a write assist technique, however slow corners could benefit from applying varying degrees of assist.

for each process corner to achieve an array V_{MIN} of 0.5V. Fast process corners (FF, FS) can easily scale their supply voltage without the need for the assist. However, slow corners (SS, SF) require assist in ensuring functionality. Thus, the degree of applied assist can be controlled through a process monitor to

reduce the margining required, the half-select failures, and the power consumed.

Fig. 8 provides an overview of various trade-offs between three essential metrics - reliability (write margin), sustainability (energy), and performance (delay). The selection of an assist depends on such trade-off against the metric of interest. For example, V_{DD} lowering assist can be best suited for a reliable and energy optimized system that operates at a very slow frequency. WL boosting can be used with better overall metrics but demands higher system V_{MIN} (= 0.6V). Because of these design considerations, the subthreshold SRAM design optimization demands an application specific customization for the targeted metric.

D. Device Type

Similar to the optimal assist technique selection, the choice of the device types within a bitcell also impacts the metric of interest and requirements. In [5][9], the authors considered a different type of 8T bitcells using various types of devices as a tuning knob to optimize a particular metric for the optimization. Table I defines the six different bitcells based on the type of device used for each of its transistors. Fig. 9 maps the best-to-worst choice of a bitcell using four contours. Here, the outermost contour represents the best option for the selection while the inner contour shows the worst choice for the given metric. For example, the MVT3 and HVT bitcells are the best choices for the DRV (i.e. lowest DRV possible), and SVT represents the worst. Similarly, the SVT bitcell is write-performance optimal but proves worst for the leakage. Therefore, an optimal selection of the device in bitcell depends on the metric of the consideration.

TABLE I DIFFERENT BITCELLS WITH DEVICE TYPE MAPPING [12]

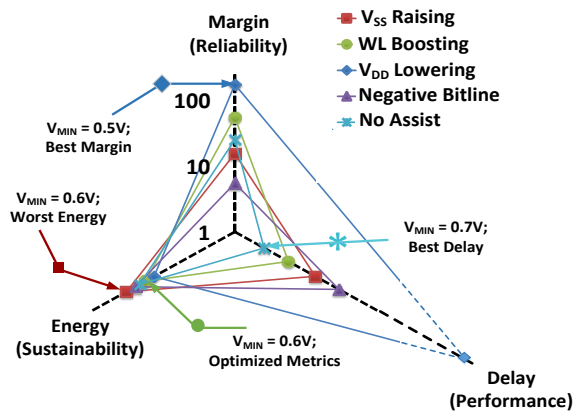| Bitcell	Device Usage			
	PU	**PD**	**PG**	**RAs**
HVT	high-V_T	high-V_T	high-V_T	high-V_T
SVT	standard-V_T	standard-V_T	standard-V_T	standard-V_T
MVT1	high-V_T	high-V_T	standard-V_T	standard-V_T
MVT2	standard-V_T	standard-V_T	high-V_T	high-V_T
MVT3	high-V_T	high-V_T	high-V_T	standard-V_T
MVT4	high-V_T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

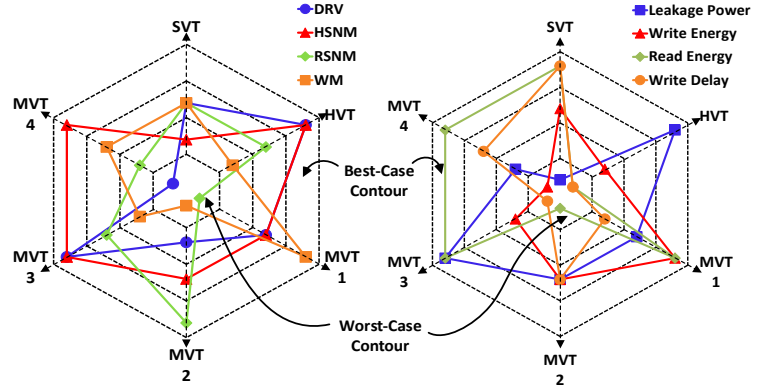Fig. 8. Optimal assist technique selection in subthreshold depends on the metric of interest.

Fig. 9. Optimal Bitcell selection based on static and dynamic metrics for the consideration.

IV. CONCLUSION

In this paper, we explored different aspects of subthreshold SRAM design including challenges and required changes in design consideration. We highlighted the significance of process and temperature variation on subthreshold SRAM design. The half-select issue in subthreshold is found limiting SRAM V_{MIN} scaling. We explored technology, different bitcell design, and various assist technique to address subthreshold challenges. Selection of an optimal assist technique changes with operating supply voltage, the degree of assist, and metric under consideration.

REFERENCES

[1] https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

[2] B. Calhoun, A. Wang, A. Chandrakasan, "Modeling and Sizing for Minimum Energy Operation in Subthreshold Circuit," JSSC, 2005.

[3] F. B. Yahya et al., "Combined SRAM read/write assist techniques for near/sub-threshold voltage operation," *ASQED* 2015.

[4] S. Lutkemeier et al., "A 65 nm 32 b Subthreshold Processor With 9T Multi-Vt SRAM and Adaptive Supply Voltage Control," JSSC 2013.

[5] H. N. Patel, F. B. Yahya and B. H. Calhoun, "Optimizing SRAM bitcell reliability and energy for IoT applications," *ISQED* 2016.

[6] F. Yahya et al., "A Sub-Threshold 8T SRAM Macro with 12.29 nW/KB Standby Power and 6.24 pJ/access for Battery-Less IoT SoCs," *JLPEA*, May 2016.

[7] Harsh N. Patel et al., "A 55nm Ultra Low Leakage Deeply Depleted Channel technology optimized for energy minimization in subthreshold SRAM and logic," ESSCIRC 2016.

[8] Marinissen, E.J.; Prince, B.; Keltel-Schulz, D.; Zorian, Y., "Challenges in embedded memory design and test," DATE, 2005

[9] Meinerzhagen, P. et al., "A 500 fW/bit 14 fJ/bit-access 4kb standard-cell based sub-VT memory in 65nm CMOS", ESSCIRC, 2012.

[10] Kim, D. et al., "A 1.85fW/bit ultra low leakage 10T SRAM with speed compensation scheme", ISCAS, 2011.

[11] V. Chandra, C. Pietrzyk and R. Aitken, "On the efficacy of write-assist techniques in low voltage nanoscale SRAMs," *DATE 2010*.

[12] H. N. Patel et al., "Improving Reliability and Energy Requirements of Memory in Body Sensor Networks," *VLSID* 2016.

[13] P. E. Dodd and L. W. Massengill, "Basic mechanisms and modeling of a single-event upset in digital microelectronics,"TNS, 2003.

[14] N. Verma and A. P. Chandrakasan, "A 65nm 8T Sub-Vt SRAM Employing Sense-Amplifier Redundancy," ISSCC *2007*.