Optimizing SRAM Bitcell Reliability and Energy for IoT Applications

Harsh N. Patel[1], Farah B. Yahya, Benton H. Calhoun
University of Virginia, VA USA
[1]Email: hnpatel@virginia.edu

***Abstract* — This paper compares six different 8T SRAM bitcells targeting different design space requirements - such as reliability and low power/energy - for Internet of Things (IoT) applications. Different bitcells leverage the varying characteristics of high-threshold (high-V_T) and standard-threshold (standard-V_T) devices to affect SRAM metrics like write margin (WM), Data Retention Voltage (DRV), Hold Static Noise Margin (HSNM), Read Static Noise Margin (RSNM), write and read energy, standby leakage power, and variability. The reliability for each bitcell over process (intra- and inter-die variation) and temperature variation is also evaluated. Measured results for a commercial 130nm test chip compare the most promising two 8T bitcell structures targeting low leakage and low energy.**

Keywords—V_{MIN}, high-V_T, standard-V_T, multi-V_T, Variation Index, Write Margin, robustness.

I. Introduction

With the continuously changing electronic market, specifications and design requirements for different applications vary widely. For example, for the Internet of Things (IoTs) and Body Sensor Nodes (BSNs), robustness and energy efficiency are the most important design constraints; while for general purpose and graphics (GPU) processors, servers, and other high-end applications, energy is traded off for higher performance. Generally, IoT devices need to 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. The authors in [1] show the Static Random Access Memory (SRAM) as the major contributor to the power dissipation of the digital design in Ultra-Low Power (ULP) Systems-on-Chip (SoCs). In addition, the need for larger embedded memories (mainly SRAM) is increasing in highly integrated SoCs to support a wide range of capabilities, thus further tightening the design constraints on power, performance, and energy. V_{DD} scaling to the sub-threshold region minimizes the total energy consumption [2]. With V_{DD} scaling, the contribution of the active energy and the leakage energy to the total energy changes. The active energy (CV_{DD}^2) dominates at higher voltages, while leakage energy ($V_{DD}*I_{Leak}$) dominates at sub-threshold voltages. Thus, an optimal V_{DD} exists that minimizes the total energy of a design, and usually that V_{DD} lies in the sub-threshold region [4]. Since this optimum point changes for different performance needs and designs, it is important to explore the design space while varying different knobs to determine the optimal voltage that minimizes the energy consumption of the design for each application. In this paper, we explore the design space of SRAM considering the threshold voltage of its transistors as one of the important design knobs needed to achieve a robust and energy efficient memory for IoT applications.

Sub-threshold SRAM faces additional challenges compared to super-threshold SRAM. For ratio-ed design such as an SRAM, the functionality of the circuit depends largely on relative strength of the devices. Fig. 1 shows the normalized I_{ON}/I_{OFF} characteristics of NMOS and PMOS devices across V_{DD}s at different process corners for the 130nm CMOS technology. At lower V_{DD}s, the reduced I_{ON}-to-I_{OFF} ratio and significantly higher variation across process corners lead to stability and performance degradation of the SRAM. Fig. 1 shows ~1000X I_{ON}/I_{OFF} reduction at V_{DD}=0.5V compared to the nominal V_{DD} (1.2V) with ~20X variation in I_{ON}/I_{OFF} ratio across process corners. Also at lower V_{DD}s, I_{ON} does not increase linearly with the width of the transistors due to Inverse Narrow Width Effect (INWE) [2]. Therefore, at sub-threshold voltages, sizing is a weak knob to control I_{ON}.

This paper compares stability metrics and energy consumption of different 8T bitcells for a given design requirement. The optimal choice of transistor threshold voltage within a bitcell varies significantly with the application requirements. The paper also compares different variation sources (intra-die, inter-die, and temperature) in sub-threshold SRAM and their impact on various metrics. The remainder of the paper is divided as follows: Section II describes the present solutions and justifies the choice of different device selections within the bitcell for the comparison. Section III compares different metrics for all of the considered bitcells. Section IV presents the results from the chip fabricated with two of the bitcells chosen for low power IoT application. Section V concludes with a comparison of bitcells highlighting the best choice given a target application.

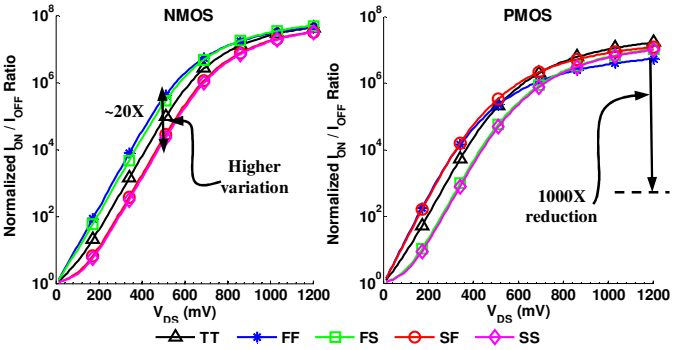


Fig. 1. Sub-threshold challenges: Reduces I_{ON}/I_{OFF} ratio and higher process variation at lower V_{DD}s

II. State of the Art and Choice of Bitcells

In [5], the authors performed a similar analysis at nominal voltage with a focus on performance. In that study, the authors evaluated performance of different bitcells at nominal voltages. Thus, sub-threshold effects such as higher impact of variation on SRAM performance were not examined. In addition to that, previous bitcell exploration targeting nominal voltage did not evaluate energy, an important metric for battery-less and other energy constrained IoT applications. Therefore, in this paper we explore the impact of variation on the stability and energy consumption of different bitcells at sub-threshold voltages. For low power applications, the conventional 6T bitcell does not allow V_{DD} scaling to the sub-threshold region because of write and read-disturb failures [8]. Instead, the 8T bitcell (Fig. 2) is widely used in sub-threshold SRAMs to enable independent design of the read and write ports. Thus, we will focus on the 8T bitcell in our analysis.

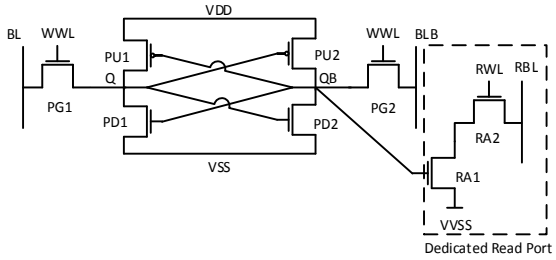


Fig. 2. An 8T SRAM bitcell [6]

In our evaluation, we chose V_T of the devices within the bitcells based on the targeted metric improvement. For example, weakening the pull-up devices (PU1/PU2) and/or strengthening pass transistors (PG1/PG2) improves the write functionality. Therefore, a bitcell with high-V_T pull-up devices and/or 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 I shows the different bitcells studied in this paper. Different combinations of high-V_T and standard-V_T devices are used within these bitcells to improve various metrics like WM, write delay, DRV, etc as discussed above. 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

TABLE I: 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

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.

III. Comparison of Evaluation Metrics

In this section, we compare six proposed bitcells in Table I using different evaluation metrics. The main evaluation metrics for the bitcells vary based on the targeted application. For example, an application operating at sub-threshold voltages can trade off performance to guarantee functionality (read/write) and low energy. In other applications, the system-level power consumption might limit the V_{DD} at which the SRAM must operate to a sub-optimal voltage. In such application, guaranteeing functionality is a main concern. For this reason, we divide evaluation metrics into two categories: reliability and dynamic metrics. The reliability metrics include the static functionality metrics (e.g. DRV, WM, HSNM, and RSNM) while the dynamic category includes metrics such as leakage, read/write energy, and operating speed. Table II shows the evaluation metrics under each category.

TABLE II 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 sub-section, we study the reliability of an SRAM bitcell against intra-die variation by running 1000-point Monte Carlo(MC) simulation for each metrics, while the robustness against inter-die (across process corners) 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 Data Retention Voltage (DRV). Therefore, to operate an SRAM at lower V_{DD}s, the DRV of the bitcell is required to be as low as possible. 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 [12] are used to measure the HSNM and RSNM.

Fig. 3 and Fig. 4 show the distribution of DRV, HSNM, and RSNM evaluated at T=25℃ with the worst case (min/max) and ±3σ variation results for each metric. The plots highlight the best and worst choice of bitcell. Since we target IoT applications operating in near or sub-threshold 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 the different devices on DRV by varying their V_T and measuring the sensitivity of the DRV to this change [10]. Fig. 3 b) 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 per Fig. 3 b), when the I_{ON} is increased by stronger PU1 and PD2 while 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 Fig. 4 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. 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, IoT applications not employing these solutions for low V_{DD}s operation require an SRAM with smaller DRV (HVT/MVT3), higher HSNM (HVT/MVT3), and higher RSNM (MVT2).

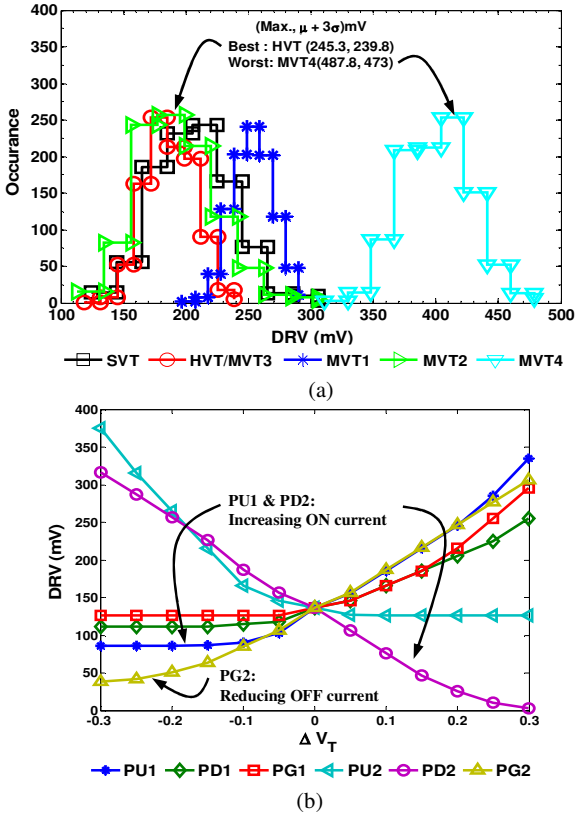


Fig. 3. a) DRV distribution of different bitcells b) Impact of device choice on DRV mapped as change in V_T (bitcell with Q node holding '1')

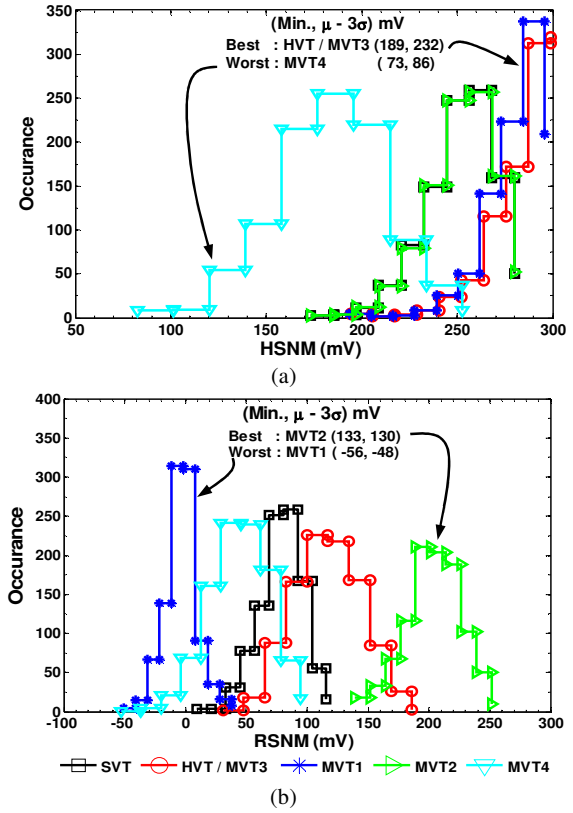


Fig. 4. Noise margin comparison: a) HSNM, b) RSNM - Distribution under local variation and optimal choice.

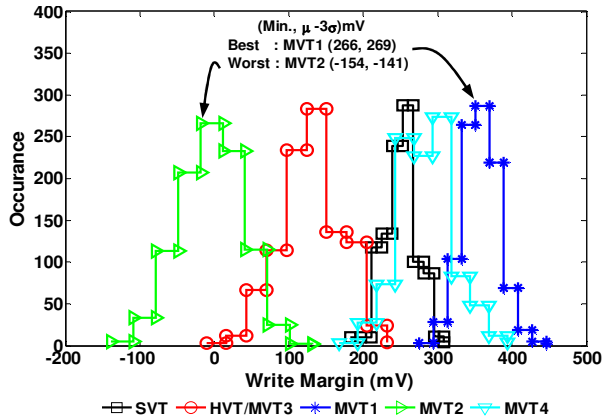


Fig. 5. WM distribution of different bitcells for V_{DD}=0.5V

Fig. 5 evaluates the WM of the different bitcells for V_{DD} =0.5V as the sub-threshold 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 (Fig. 5).

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 (Fig. 1), 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 linear dependence on temperature [2]. Due to these facts, 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] °C.

Fig. 6 a) 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. 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 a chip, fabricated at any corner, can experience due to temperature variation. For example, SVT bitcell experiences maximum VI (worst-case variation impact) of 130mV for the selected range of temperatures across the different corners. Fig. 6 b) 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 Fig. 6 b), MVT4 shows an optimal choice for a temperature variation resilient (less variation) design while HVT/MVT3 cells experience different trade-offs for the different metrics compared to SVT bitcell.

The threshold voltage choice 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 mainly due to the characteristics of individual device at a given corner and temperature. To better understand these results, Fig. 7 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 a good insight on the impact of threshold voltage choice on the stability metrics. As shown in Fig. 7, 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. As HSNM and DRV metric depends on I_{ON}/I_{OFF} characteristics of the devices used, these trends help explain the results in Fig. 6 b). The bitcell with more standard-V_T devices (e.g. SVT) experience higher variations in the 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 (Fig. 6 b) that requires device physics knowledge. Therefore, we propose VI to be considered as one of the most influencing metrics for a robust SRAM design for IoT applications operating in sub-threshold region - where variation is a major concern - and under a wide range of environmental conditions such as temperature.

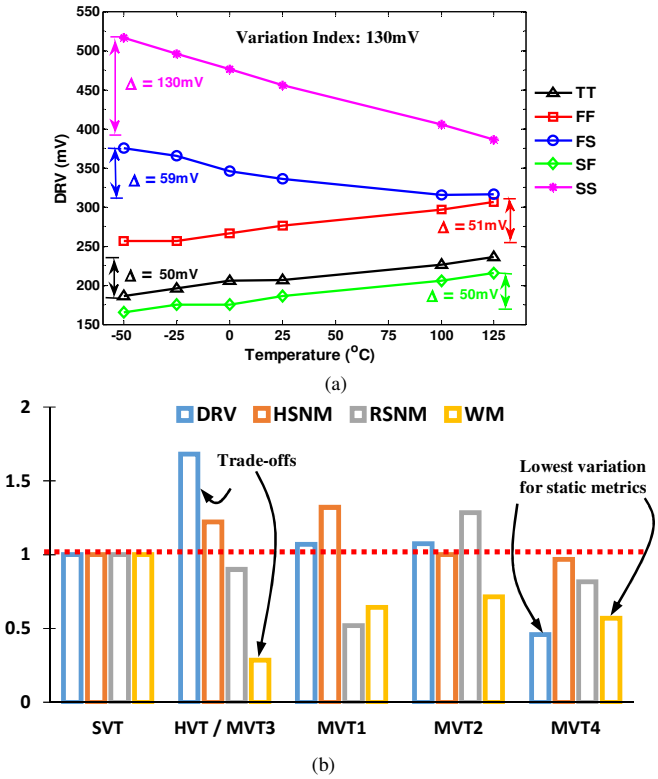


Fig. 6. a) Impact of variation on DRV across corners and temperatures. b) Comparison of VI for different metrics across bitcells

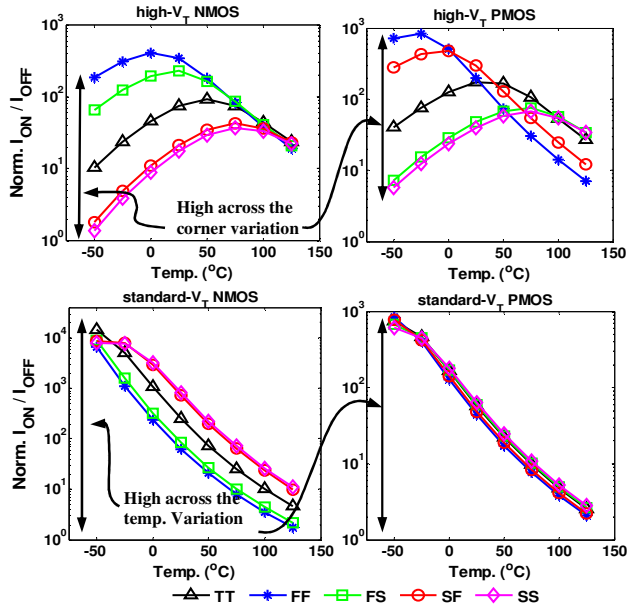


Fig. 7. Normalized I_{ON}/I_{OFF} characteristics for different devices used across process corners and temperatures

Table III summarizes the optimal SRAM bitcells for each static metric at low V_{DD}s. Table III provides the trade-offs between different static metrics and orders the bitcells according to their effectiveness in improving the corresponding metric. Each color corresponds to one bitcell.

TABLE III. OPTIMAL BITCELL CHOICE FOR STATIC METRICS

Metrics	Best-to-Worst Choice (left to right)				
DRV	HVT/MVT3	MVT1	SVT	MVT2	MVT4
HSNM	HVT/MVT3	MVT1	MVT2	SVT	MVT4
RSNM	MVT2	HVT/MVT3	SVT	MVT4	MVT1
WM	MVT1	SVT	MVT4	HVT/MVT3	MVT2

B. Dynamic Metrics

After evaluating static metrics for robustness, we explore the dynamic metrics in this sub-section. We consider an array of 1KB to calculate leakage power, write/read energy, and operating frequency at low V_{DD}s.

Fig. 8 shows the leakage power of different bitcells while taking into account process and temperature variations. Here, 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 nominal temperature (25°C). The percentage reduction in leakage power varies across temperature from ~755% (-50°C) to ~142% (125°C). Many IoT applications with less activity factor and higher standby time will significantly benefit from the lower leakage of the HVT/MVT3 bitcells.

Fig. 9 shows the comparison of write/read energy across V_{DD}s at their respective worst-case corner. We calculate the energy as the 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 sub-threshold 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 Fig. 9 shows how the power-delay product (energy) of different bitcells impacts the optimal choice of bitcell. Also, the selection of operating voltage influences the optimal bitcell. Table IV shows the consolidated results of the dynamic metrics for the bitcells.

IV. CHIP RESULTS

Fig. 8 and Fig. 9 show that the applications being in 'Standby' mode for longer time require low-leakage bitcells, while 'Always ON' applications require lower active energy bitcells. Since many IoT 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. Fig. 10 shows a die micrograph of the fabricated chip in a commercial 130nm technology. Fig. 11 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 reliable

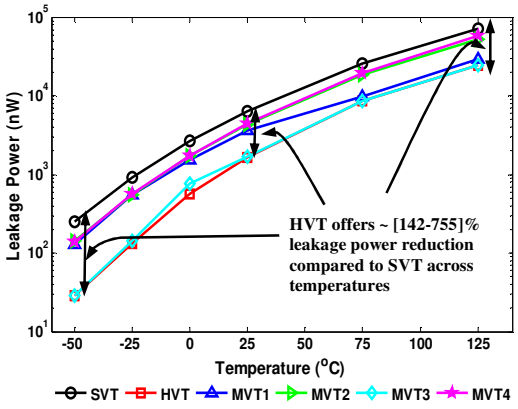


Fig. 8. Leakage power of 1KB array across temperature for different bitcells

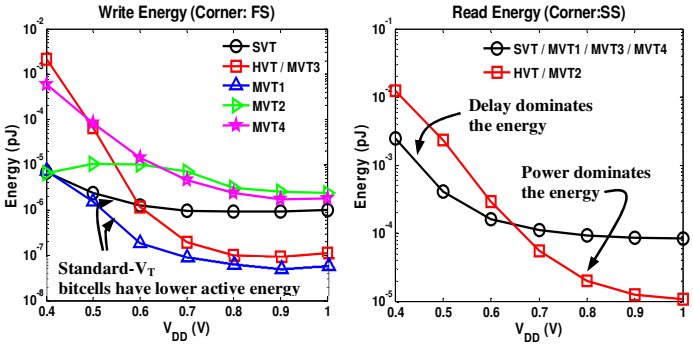
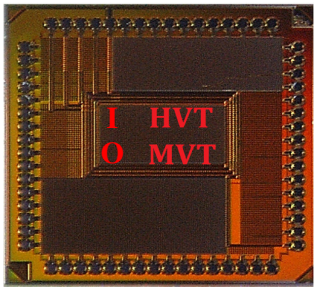


Fig. 9. Write and Read energy comparison of different bitcells across the V_{DD}s

TABLE IV. OPTIMAL BITCELL CHOICE FOR PERFORMANCE METRICS

Metrics	Best-to-Worst Choice (left to right)				
Leakage Power	HVT/MVT3	MVT1	MVT2	MVT4	SVT
Write Energy	MVT1	SVT	MVT2	HVT/MVT3	MVT4
Read Energy	SVT / MVT1 / MVT3 / MVT4			HVT/ MVT2	

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 ~2X reduction in the measured energy numbers as compared to the HVT bank. The



Fabricated Chip:

Size : 1KB Bank of each bitcells
Process : 130nm
Bitcell : HVT and MVT3

Fig. 10. Die micrograph of chip showing HVT and MVT banks

reduction 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 (Fig. 9) due to the additional peripheral circuits implemented to make the array functional. The results highlight that MVT3 bitcell reduces active energy compared to HVT bitcell while providing lowered leakage compared to other bitcells.

V. Conclusion

In this paper, we address 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 was studied. The paper also presents an in-depth study of the effect of variations on the different static metrics required for low power applications. Table V provides a summary of the metrics studied, possible IoT applications were these metrics are most significant, traded-off metrics and the optimal bitcell choice.

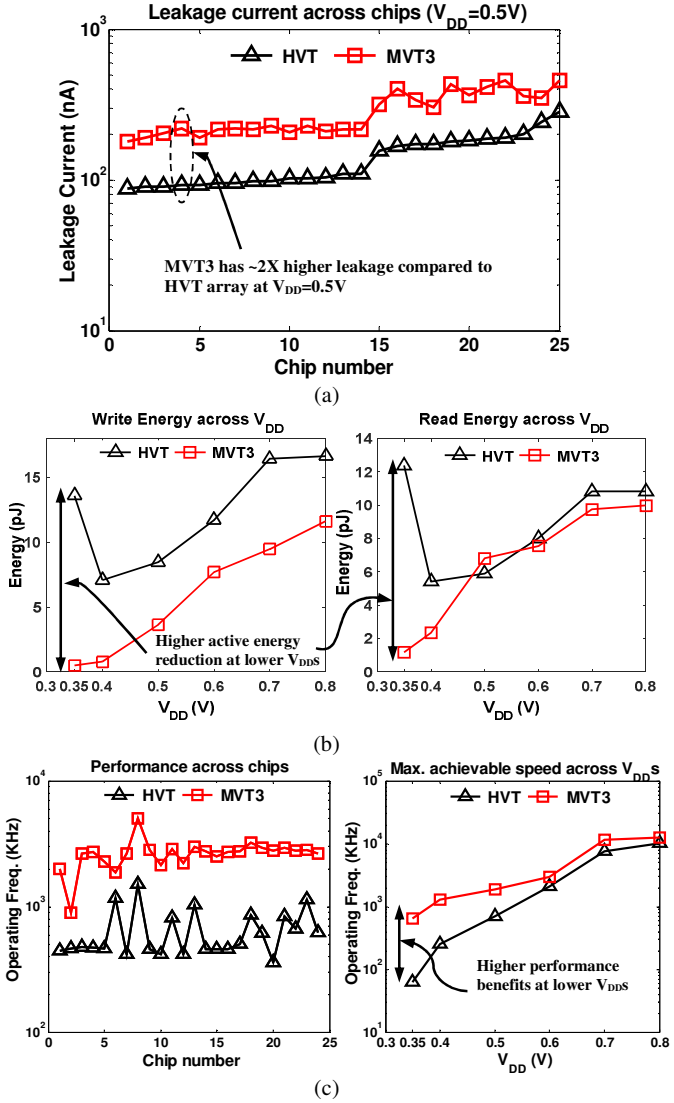


Fig. 11. Measurement result comparison between HVT and MVT3 bitcells from 24 chips: a) Leakage measurement (V_{DD}=0.5V) b) Write and Read energy across V_{DD}s c) Sub-threshold performance with variation and across the V_{DD}s

TABLE V. Optimal Bitcell Selection Summary with Possible Trade-offs for Targeted Application

Design Metric	Best bitcell choice	Trade-off metric(s)	Possible IoT Application
DRV	MVT2	WM + Leakage power	Applications w/ longer stand by time (e.g. remote sensing)
HSNM	HVT / MVT3	WM + Write Speed	Robust design to operate in noisy environments (difficult terrain sensing)
RSNM	MVT2	WM + Leakage power	Low V_{DD} write operation (low-power application)
WM	MVT1	RSNM + DRV	
Leakage Power	HVT	Write/Read Delay + WM	'Mostly Stand-by' application (e.g. body sensing)
Active Energy	MVT1 / MVT3		'Mostly Active' application (e.g. DSP, speech processing)

In summary, the paper contributes an exploration of a low power/energy optimal SRAM design with an optimal bitcell choice for a targeted IoT application.

VI. Acknowledgement

This work was funded in part by the SRC, ARM, and the NSF NERC ASSIST Center (EEC-1160483). This work was funded in part by DARPA through a subcontract from NVIDIA. This research was, in part, funded by the U.S. Government. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

References

[1] Lutkemeier, S.; et al., "A 65 nm 32 b Subthreshold Processor With 9T Multi-Vt SRAM and Adaptive Supply Voltage Control," JSSCC 2013
[2] Marinissen, E.J.; Prince, B.; Keltel-Schulz, D.; Zorian, Y., "Challenges in embedded memory design and test," DATE, 2005
[3] J. Zhou; et al."A 40 nm inverse-narrow-width-effect-aware sub-threshold standard cell library", DAC, 2011.
[4] B. Calhoun, et al., "Modeling And Sizing For Minimum Energy Operation In Subthreshold Circuit". JSSC, 2005.
[5] H. Zhu, V. Kursun, "A Comprehensive Comparison of Data Stability Enhancement TechniquesWith Novel Nanoscale SRAM Cells Under Paramet er Fluctuations", IEEE TCAS I, VOL. 61, NO. 5, MAY 2014
[6] N. Verma, A. P. Chandrakasan, "A 256 Kb 65 Nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy", JSSC 2007.
[7] K. Zhang, et al. "A 3-Ghz 70MB SRAM In 65nm CMOS Technology With Integrated Column-Based Dynamic Power Supply", ISSCC 2005.
[8] Sharma,S. et al. "Ultra-Low Energy SRAM Design for Smart Ubiquitous Sensors",IEEE Micro.
[9] A. Raychowdhury, et al., "PVT-and-Aging Adaptive Wordline Boosting for 8T SRAM Power Reduction", ISSCC 2010
[10] K. Leochico, E. John, "Data retention voltage analysis of various low-power SRAM topologies," MWSCAS 2014.
[11] Boley, J.; Chandra, V.; Aitken, R.; Calhoun, B., "Leveraging sensitivity analysis for fast, accurate estimation of SRAM dynamic write V_{MIN}," DATE 2013.
[12] Seevinck, E.; List, F.J.; Lohstroh, J., "Static-noise margin analysis of MOS SRAM cells," JSSC, 1987.
[13] L. Dolecek, et al. "Breaking the simulation barrier: SRAM evaluation through norm minimization,", ICCAD, 2008