Low Power SRAM Final Report

Braden Desman and Silas Schroer

Abstract—**In this paper, we describe the design, simulation, and performance of a 1Mb SRAM implemented in the FreePDK 45nm technology for PICo's low power application. By using a Dynamic Leakage Suppression (DLS) bitcell and structuring large interconnects hierarchically, we are able to realize substantial reductions in the active energy per access and idle power at the cost of increased delay and area. Comparisons are made between the conventional 6T bitcell and the DLS bitcell throughout the paper. We achieved 718fJ per access with a delay of 20ns and an idle power of 123.2µW at 27ºC and the TT process corner. In terms of PICo's low power design metric, the array reaches 2.05 x 10^{-33} J^2*sec*mm*W while satisfying the company's specifications in terms of capacity, functionality, and robustness against PVT variation.**

I. INTRODUCTION

PICo has commissioned team SSD to design an SRAM chip for its low power microsensor nodes. In order to maximize the lifetime of the node, PICo has requested that the SRAM design be optimized in terms of energy consumption. To that end, the design metric specified by PICo is as follows:

$$(ActiveEnergy/Access)^2(IdlePower)(Delay)(Area) \quad (1)$$

In order to design a memory that best suits PICo's needs, we have kept this metric in mind at all stages of the design process and have used it to set our priorities as follows:

1) **Lower V$_{DD}$ as much as possible.** Given that V$_{DD}$ is squared in the active energy equation, which itself is then squared in PICo's design metric, V$_{DD}$ is the strongest knob available to us to design a suitable low power memory. We have chosen to operate at V$_{DD}$=0.4V.
2) **Break up large capacitances along WL, BL, etc.** Using a hierarchical structure for the large interconnects inherent to SRAM arrays, we can lower the amount of capacitance that must be charged up during each memory access and thus reduce the active energy per access.
3) **Select a low-leakage bitcell.** By using a dynamic leakage suppression (DLS) based bitcell, we can achieve a reduction in the array's leakage power by several orders of magnitude while lowering the energy per access compared to the 6T bitcell.

The remainder of this paper is structured as follows. In Section II, we will provide an overview of the architecture of our memory array and emphasize how certain architectural choices lowered our design's energy consumption. In Section III, we will provide details about the implementation of our DLS bitcell and peripheral circuits. We will also highlight some of the tradeoffs, especially in terms of delay and area, that were made in order to achieve substantially lower energy per access compared to our initial, unoptimized memory design.
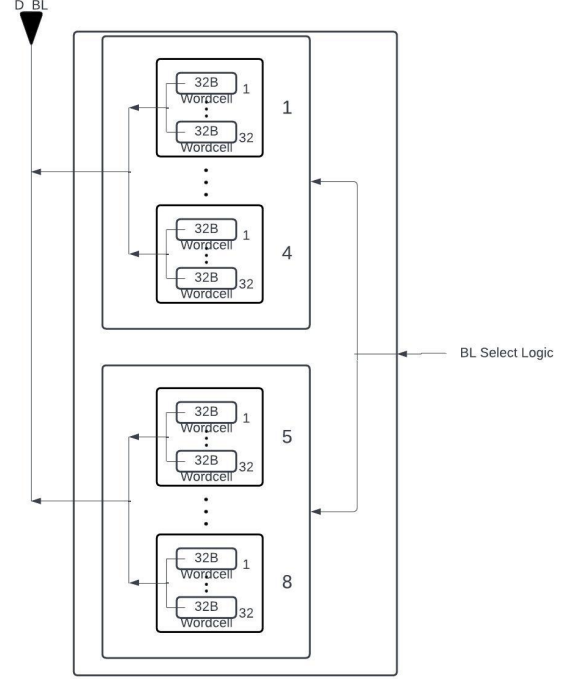

Fig. 1: Bit Line Heirarchy

II. ARCHITECTURE

To meet PICo's capacity specifications and optimize the power consumption of the memory, we have introduced several levels of hierarchy in our design. At the top level, the memory is composed of 16 banks, with each bank consisting of 2,048 32b words. The bank block diagram is provided in Appendix A. Each bank consists of 256 rows with 8 words per row, with an additional "dummy" column used to help with timing the read operation. The operation of this dummy column will be explained in more detail in Section III-B.

In order to achieve substantial reductions in energy per access, we introduced hierarchical read and write bit lines which break up the large capacitance along the interconnects. Figure 1 shows the grouping of cells used for each level of hierarchy, the levels are separated by PMOS transistors with a width of 540 nm enabled by certain address bit combinations. This limits the capacitance on the worst level of hierarchy to that of 128 PMOS drains, which is 4 times less than 256 DLS bitcells.

A similar hierarchy was also implemented in the architecture of the write word lines (WWL) and read word lines (RWL) to reduce capacitance. This structure can be seen in Figure 2.
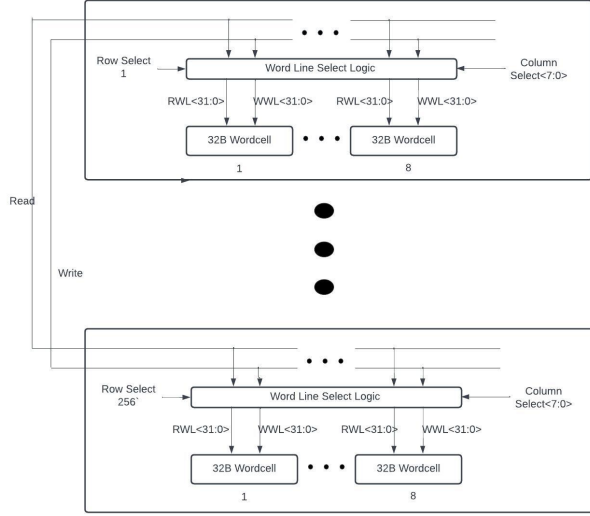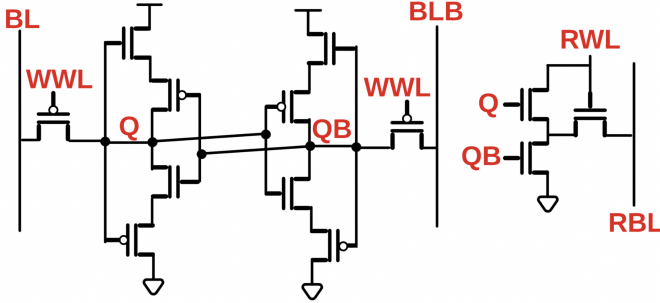
Fig. 2: Word Line Heirarchy

Fig. 3: The DLS Bitcell. All transistors are minimum width aside from the header and footer devices, which are twice minimum width. The access transistors are VTG, the internal transistors VTH, and all remaining devices are VTL. All NMOS body connections are grounded and all PMOS body terminals are tied to V_{DD}.

III. MEMORY COMPONENTS

A. The DLS Bitcell

We evaluated several bitcell designs including the conventional 6T design, a low-power 9T design, and our own novel 7T design before ultimately proceeding with the DLS design proposed in [1]. Although the 9T design proposed in [2] and our 7T design showed initial promise, a loss of functionality at longer clock periods rendered them unusuable for our array. Due to functionality issues, we regard comparisons to the 9T and 7T designs as unfair and thus omit them from the following figures. In this section, we will describe the operation of the DLS bitcell and compare its energy consumption and idle power to that of the 6T bitcell. The schematic of the DLS bitcell is provided in Fig. 3

1) Hold Operation: In Fig. 3, we can see that the DLS bitcell simply consists of two cross-coupled DLS inverters, a pair of access transistors, and a separate read port. Thus, due to the regenerative property of the cross-coupled inverters, the hold operation is conceptually identical to that of the 6T
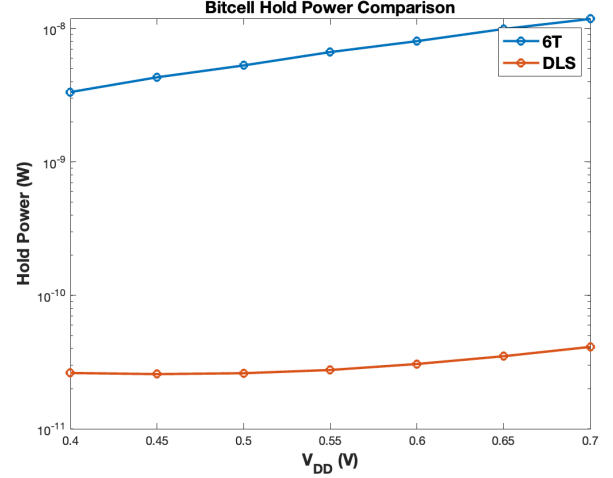

Fig. 4: Hold power of the DLS bitcell vs. the 6T bitecll. Recall that our array operates at $V_{DD} = 0.400$ mV. At this supply level, the DLS bitcell consumes 127 times less hold power than the 6T.

bitcell. What distinguishes the DLS bitcell however is its vast reduction in leakage power. At every supply level tested, the hold power of the DLS bitcell was less than the hold power of the 6T bitcell by two orders of magnitude as shown in Fig. 4.

This improvement in leakage power is a result of the use of DLS inverters instead of static CMOS inverters. DLS inverters operate though leakeage currents and feature a feedback mechanism to regulate the current through "OFF" devices. By placing "OFF" transistors in a "supercutoff" state, the leakage current through the gate is orders of magnitude lower than a static CMOS implementation [3].

Because the inverters operate through leakage currents which are a strong function of the device's threshold voltage, and because NMOS and PMOS devices have asymmetrical V_t, the output transitions of the inverters are asymmetrical. For the bitcell, this means that there are separate noise margins for holding a logic 0 versus a logic 1. The hold static noise margin for holding a 0 is shown in panel A of Fig. 5. Note that the stable logic 0 state and the stable logic 1 state do not quite reach ground or V_{DD}, respectively, due to the fact there are NMOS devices in the pull-up network and PMOS devices in the pull-down network. To reduce this disparity, header and the footer devices are sized to be twice minimum width, sacrificing small increases in area and leakage for improved margins [4].

2) Read Operation: The read operation of the DLS bitcell is relatively straightforward thanks to the separate read port. When performing a read operation, the read bitline (RBL) is precharged and the read wordline (RWL) is driven high. If Q is a 1, RBL is held high. If Q is a 0, QB is a 1 and RBL is pulled towards ground. A comparison of the energy per read access between the 6T bitcell and DLS bitcell is shown in Fig. 6

Unlike the 6T bitcell, the DLS bitcell features strong noise margins during the read operation. See panel B of Fig. 5 for the noise margin of a read 0 operation.

(a) Butterfly plot for hold 0. The hold 1 noise margin is provided in Appendix A and shows a similarly large SNM.

(b) Butterfly plot for read 0. The read 1 noise margin is provided in Appendix A and shows a similarly large SNM.

(c) Butterfly plot for write 0. The write 1 noise margin consists of a vertical line at V_1=0.4V and a horizontal line at V_2=0.4V.
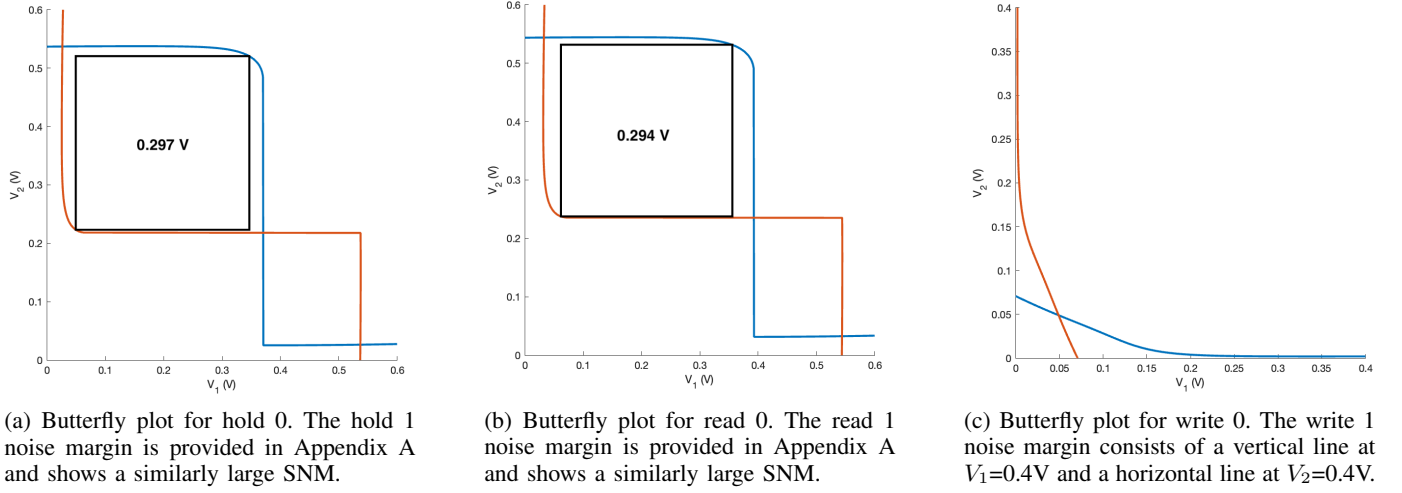
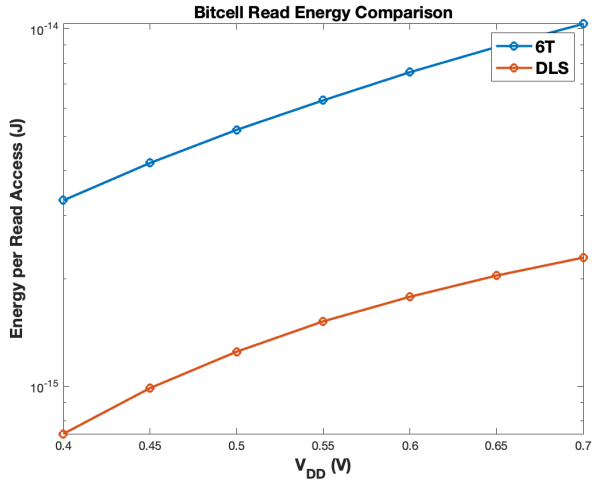Fig. 5: Hold, Read, and Write Noise Margins

Fig. 6: Energy per 1 bit read of DLS bitcell vs. 6T bitcell. Note that the energy per read access for the DLS bitcell is about 5x lower than for the 6T bitcell at V_{DD} = 400 mV.
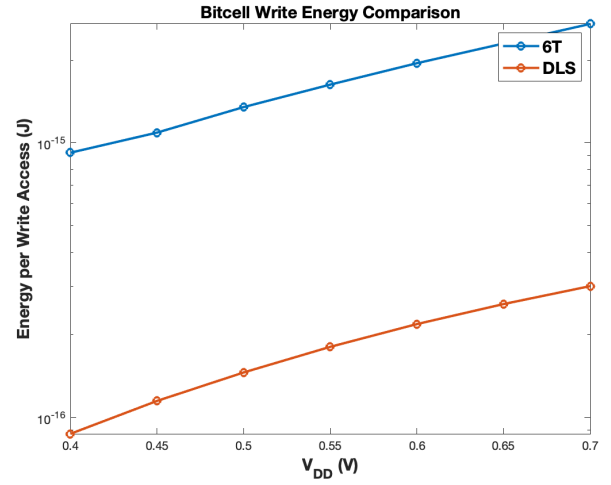

Fig. 7: Energy per 1 bit write of DLS bitcell vs 6T bitcell. The DLS bitcell outperforms the 6T bitcell in terms of energy per write access by a factor of at least 10x at every supply level tested.

3) Write Operation: The write operation is also relatively simple due to the fact that the inverters are operating through leakage currents, whereas the access transistors are strongly "ON" while writing to the cell. This means that the access devices are able to write to the cell with relative ease, and the cell ratio is not an important factor when sizing the devices in the cell. Visually, we can see that the DLS bitcell writes appropriately by looking at the write SNM provided in panel C of Fig. 5. Once again, we find that the DLS bitcell outperforms the 6T design in terms of energy per write access at all supply levels tested as shown in Fig. 7.

4) DLS versus 6T Bitcell Tradeoffs: Although the DLS bitcell clearly outperforms the 6T design in terms of energy per operation and idle power, there are some tradeoffs of note. First, there is a 218% increase in the area of the cell as opposed to the 6T bitcell. Since area is not emphasized as heavily in PICo's design metric, this does not result in a massive penalty. However, it does reduce the adaptability of the array

to applications where both energy and area are limiting factors.

In terms of delay, the DLS bitcell also finds itself outperformed by the 6T bitcell. For a single bit read or write, the 6T bitcell can operate 5 times faster. Again, this tradeoff was made to achieve orders of magnitude reduction in energy and power, which dominate PICo's metric for success.

B. Read Timing ("Dummy" Column)

In order to properly time the single-ended read, a "Dummy" read column is used to obtain the worst-case delay for reading a 0 at runtime. This column consists of a bitcell holding a 0 in a column with every other cell holding a 1. Because of leakage from the cells holding a 1, this is the worst case delay for RBL to be pulled to ground. Once the inverter at the bottom of RBL can sense the 0, this signal is driven to the bank register's enable input. While this solution adds a non-negligible amount of area to the array, it also ensures that the

timing remains robust at all process corners and temperatures tested.

IV. RESULTS AND SIMULATIONS

In order to confirm the functionality of the array, transient simulations were performed in Cadence Virtuoso. In Fig. 8, we simulate writing 0xFFFFFFFF to address 0x0000 and writing 0x00000000 to address 0x7FFF. The write operations are performed in the first two clock cycles, followed by a read at each location. The array then holds its state for several clock cycles. Finally, two more reads are performed during the last two cycles shown.
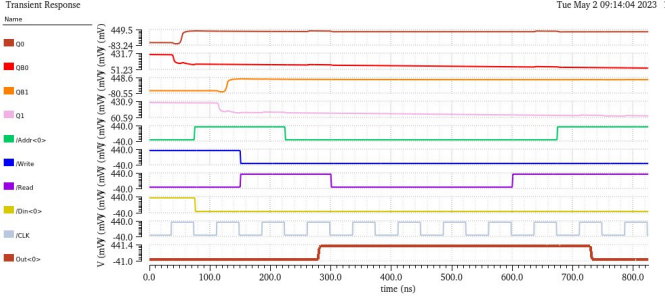

Fig. 8: Transient Simulation of Read and Write operations in nominal conditions (TT, 27 ºC)

We also used Cadence simulations to obtain the delay, energy per access, and power metrics for our design. The energy per read access was obtained by simulating several reads and taking the average energy of the operation. The same procedure was followed to obtain the write energy. To measure the idle power, the array was simulated in the hold state (i.e., with no accesses occurring). The results of these array measurements are provided in Tab. I. The read delay was calculated by measuring the delay from when the rising clock edge during a read phase to the changing edge of the output data. Write delay was calculated by measuring the delay from the rising edge of the clock during a write phase to the moment when data within the cell being written to is stable.

TABLE I: Final Array Metrics

Metric	Quantity
(Active Energy/Access)^2(Delay) (Area)(Idle Power)	2.05×10^{-33}
Total Area (mm)	1612.02
32b Read Energy (J)	687×10^{-15}
32b Write Energy (J)	875×10^{-15}
32b Average Energy/Access (J)	718×10^{-15}
Read Delay (ns)	20
Write Delay (ns)	20
Total Delay (ns)	20
Idle Power (W)	123.2×10^{-6}

These simulations were performed for several different iterations of the design, with the main independent variable being the implementation of bit line heirarchy. The two designs explored most heavily involved using transmission gates versus PMOS gates to separate each level of the BL hierarchy. The metrics shown above correspond to the implementation of PMOS gates, while the implementation of transmission gates yielded write energy metrics of 407.7×10^{-15} J and read energy of 810×10^{-15} J. Although write energy was lower for the transmission gate implementation, the final design choice was to implement heirarchy on the bit lines with PMOS transistors because read energy was reduced by 123 fJ and the area was greatly reduced, which had a heavier combined weight on the final score metric.

V. CONCLUSION

In this paper, we explained our design of a 1 Mb SRAM memory for PICo's low power application using the DLS bitcell and hierarchy within large interconnects to reduce energy consumption. We succeeded at creating a memory that is functional at all PVT corners while meeting PICo's specifications for capacity and realizing substantial reductions in energy per access and idle power.

References

[1] S. Gupta, D. Truesdell, and B. Calhoun, "A 65nm 16kb sram with 131.5pw leakage at 0.9v for wireless iot sensor nodes," *2020 IEEE Symposium on VLSI Circuits*, 2020. DOI: 10.1109/VLSICircuits18222.2020.9162772. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9162772.

[2] S. Pal and A. Islam, "Variation tolerant differential 8t sram cell for ultralow power applications," *IEEE Transactions On Computer-Aided Design Of Integrated Circuits And Systems*, 2016. DOI: 10.1109/TCAD.2015.2474408.. [Online]. Available: https://ieeexplore.ieee.org/document/7229279.

[3] D. Bol, R. Ambroise, D. Flandre, and J. Legat, "Building ultra-low-power low-frequency digital circuits with high-speed devices," *2007 14th IEEE International Conference on Electronics, Circuits and Systems*, 2007. DOI: 10.1109/ICECS.2007.4511262. [Online]. Available: https://ieeexplore.ieee.org/document/4511262.

[4] Y. Wang, G. Chen, X. Yu, X. Chen, and K. Niitsu, "A 22nm cmos 0.2v 13.3nw 16t sram using dynamic leakage suppression and half-selected free technique," *2021 IEEE Asia Pacific Conference on Circuits and Systems*, 2021. DOI: 10.1109/APCCAS51387.2021.9687693. [Online]. Available: https://ieeexplore.ieee.org/document/9687693.
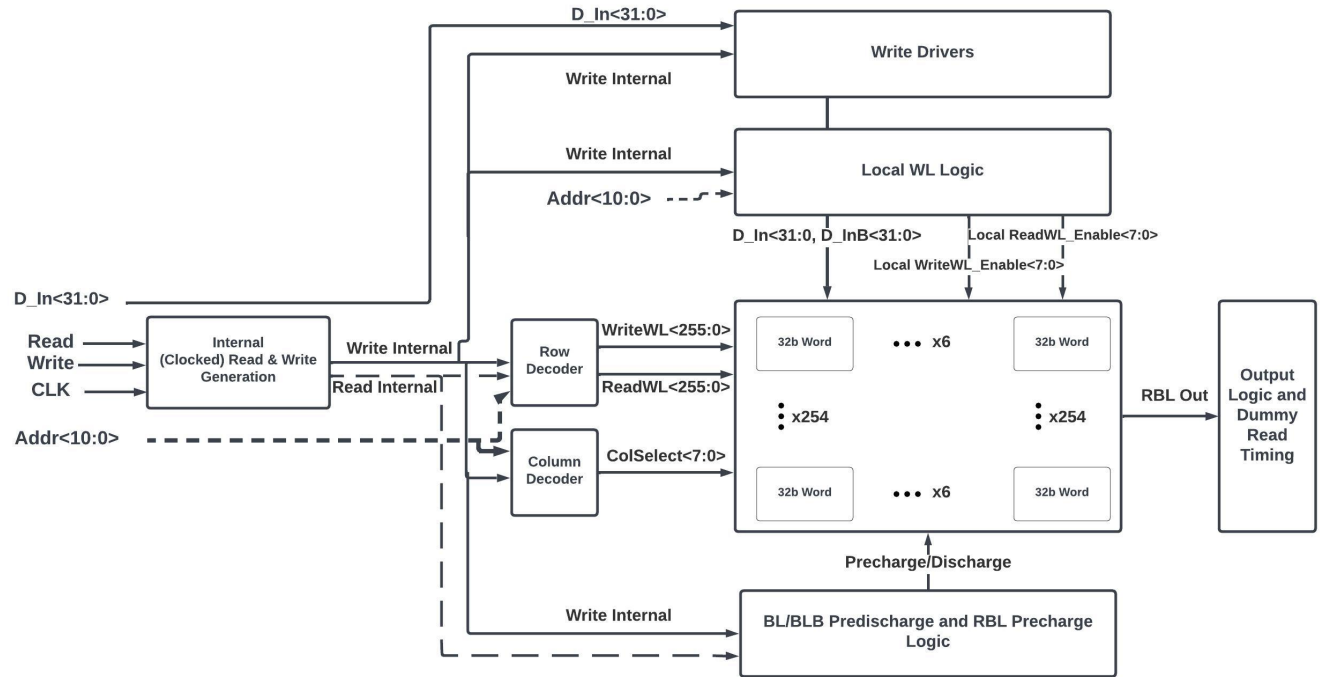
APPENDIX A
OVERFLOW FIGURES

Fig. 9: Block diagram of one bank of memory.

TABLE II: Read Static Noise Margins of DLS Bitcell

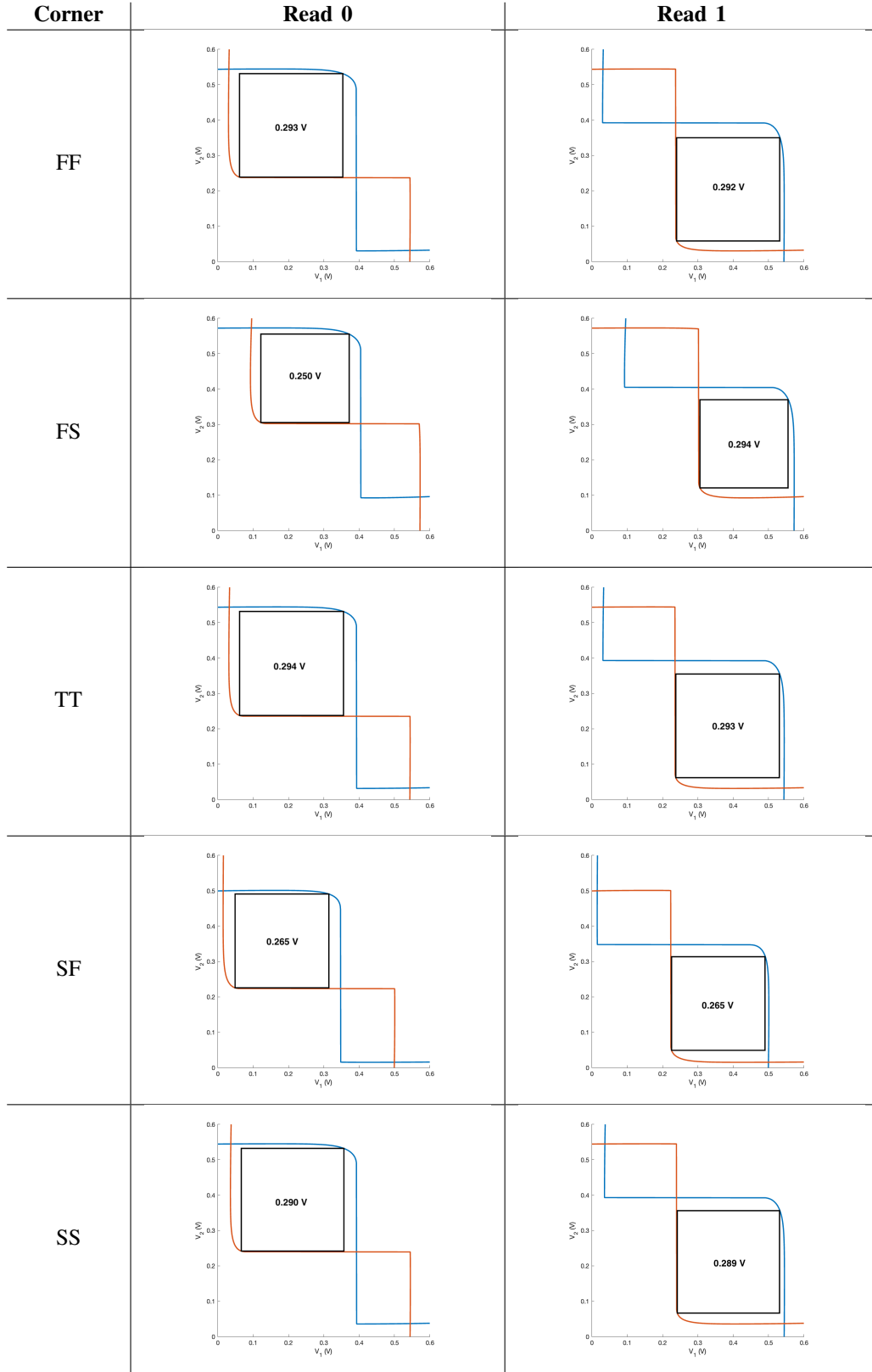| Corner	Read 0	Read 1
FF	0.293 V	0.292 V
FS	0.250 V	0.294 V
TT	0.294 V	0.293 V
SF	0.265 V	0.265 V
SS	0.290 V	0.289 V

TABLE III: Hold Static Noise Margins of DLS Bitcell

Corner	Hold 0	Hold 1
FF	0.297 V	0.296 V
FS	0.257 V	0.257 V
TT	0.297 V	0.297 V
SF	0.265 V	0.264 V
SS	0.294 V	0.292 V