Variation-Tolerant Sub-threshold SRAM

Ksenia Suglobova, Gidon Gautel

1. Introduction

Many modern industries (e.g., medical, power, shipping) are becoming increasingly reliant on low-power, local computing resources. Subthreshold logic can allow for lower power use where the associated drop in speed is acceptable. Specifically, given that dynamic energy is proportional to total system capacitance and voltage:

$$E \propto C \cdot VDD^2$$

Reducing the VDD of both logic and memory can reduce dynamic energy usage with the square of voltage reduction. However, such an approach is not typically feasible with standard 6T SRAM designs, due to primarily a combination of the reduction in noise margins and array-wide process variation. This paper examines the theory of subthreshold SRAM, implements an approach proposed by Zhai et al. (2008), and compares this against a standard 6T SRAM cell.

2. Underlying Theory

2.1. Subthreshold SRAM Challenges and Zhai et al. (2008) Solution

2.1.1. Noise Sensitivity

In traditional designs, voltage changes on the bit lines are amplified by a sense amplifier that produces a full logic level. However, at lower voltages, this becomes unreliable as the signal to noise ratio drops. In the Zhai et al. design, a single bitline can be driven rail to rail to produce a plain logic signal, eliminating the need for a sense amplifier and the uncertainty that comes from reading small voltage differences. Noise is also isolated to the single bit line in single-ended designs, increasing reliability.

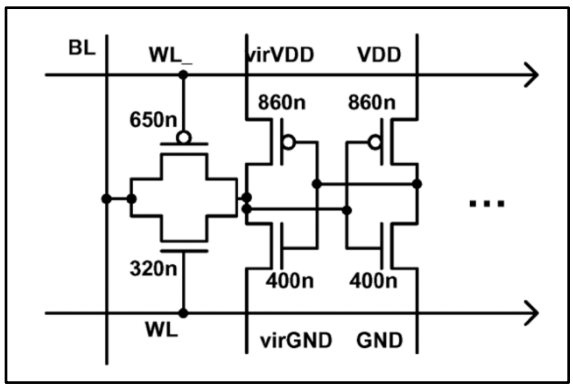

Figure 1. Core

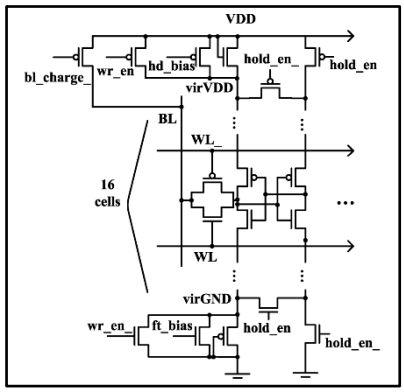

Figure 2. Cell array structure

2.1.2. RDF-Induced Vth Mismatch

In subthreshold designs, RDFs (random dopant fluctuations) become a significant risk. In transistor manufacturing, dopants added to the silicon channel to set the threshold voltage experience random fluctuation, resulting in small differences in Vth values between transistors. At standard voltages, these small differences do not impact the system. However, in the sub-threshold regime, the impact is severe, since the current becomes exponentially dependent on Vth. Transistors with small differences in Vth will carry vastly different currents, unbalancing the pullup and pulldown networks. In the Zhai et al. design this is in part mitigated by increasing gate size (minimum of 320 nm) since larger transistors require more dopant to achieve the same Vth. Increasing total dopant volume reduces the impact of small random variation across transistors.

2.1.3. Tradeoffs Of 1 Bitline

The Zhai et al. design is not without tradeoffs. A single bitline makes it harder to overpower the feedback inverter when writing to the cell. Zhai et al. solve this by introducing virtual GND and VDD rails controlled by a write enable input. When wr_en is on, VDD is drooped and GND raised, weakening the inverter and allowing the pass transistors to write more reliably. During hold or read modes, wr_en is disabled, and both virGND and virVDD

remain at the default GND and VDD levels. Droop in virGND and virVDD can be tuned via the hd_bias signal and an on-chip bias generator. Finally, subthreshold leakage in standby mode is reduced by connecting both virGND and virVDD to the feedback and forward inverters, allowing the middle node to float up and reducing the Vgs of the bottom transistor, greatly reducing current.

3. Subthreshold 6T Cell Implementation

The schematic for the subthreshold 6T SRAM cell was implemented in Cadence (Fig 1), and tested to ensure it functioned as expected in the subthreshold regime.

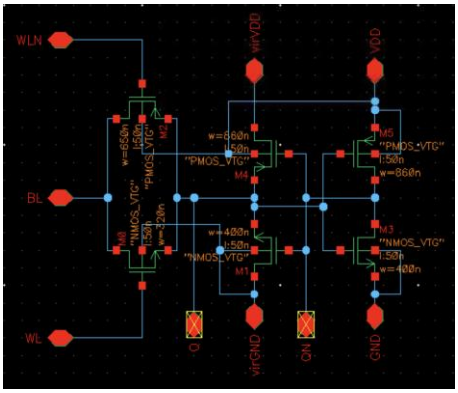

Figure 3. Cell implementation

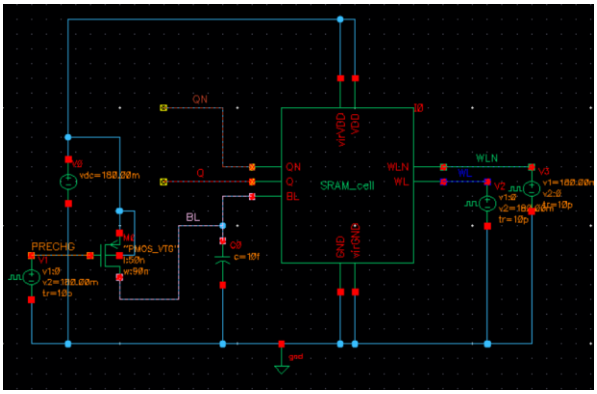

Figure 4. Read testbench

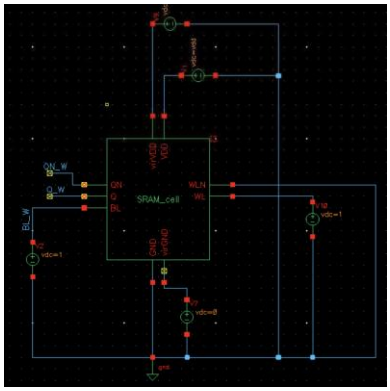

Figure 5. Write testbench

3.1. Writing Implementation

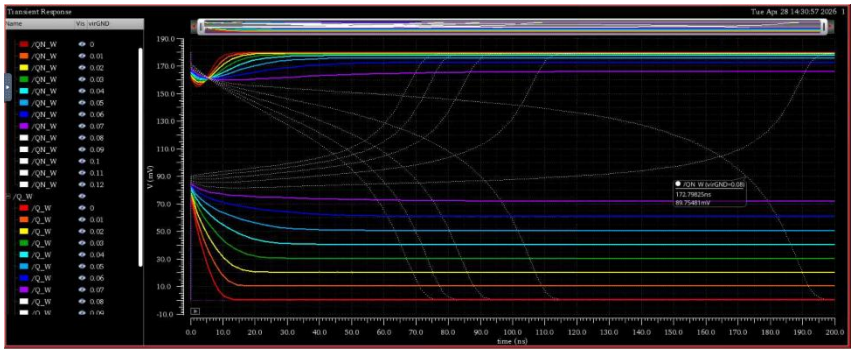

Figure 6. virGND sweep

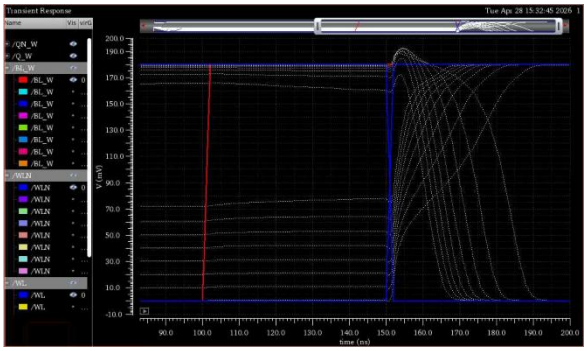

Figure 7. Write with 50ns BL delay

Writing was tested at 180mV VDD. For writing a 0 to the cell when the initial value is 1, the following parameters were set: virVDD = VDD = 180mV, GND = 0V, $Q_{initial}$ = 1V, $QN_{initial}$ = 0V. The values for virGND were then swept from 0V to 12mV to find the optimal voltage for performing a write. When trying to write a logical 1, the access transistors are fighting the pull-down NMOS, and raising virGND makes it weaker, easing the write.

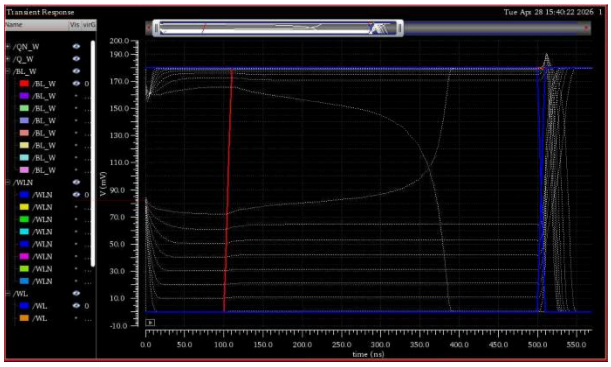

Figure 8. Write with 200ns BL Delay

Sweeping virGND, the cell destabilizes and switches values when it exceeds 70mV (Fig. 4). The writing was then tested at virGND values up to 70mV. While writing succeeds for all virGND values when the delay between switching the BL and WL/WLN voltages is at 50ns (Fig. 5), when this delay exceeds 200ns, the 70mV virGND is pulled up by the voltage of the bitline, resulting in a failed write (Fig. 6). This matches the findings in Section II-C of Zhai et al. (2008), where bitline leakage also creates cell instability, making the timing of the write just as important as

the drooping voltage values. For our cell, 70mV is the optimal virGND if write timing is reliable. Otherwise, 60mV is the safer choice.

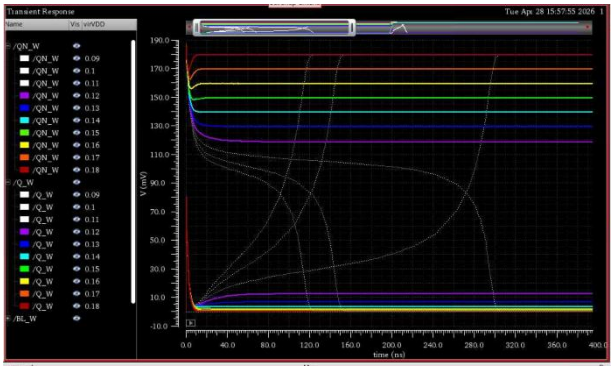

Figure 9. virVDD sweep

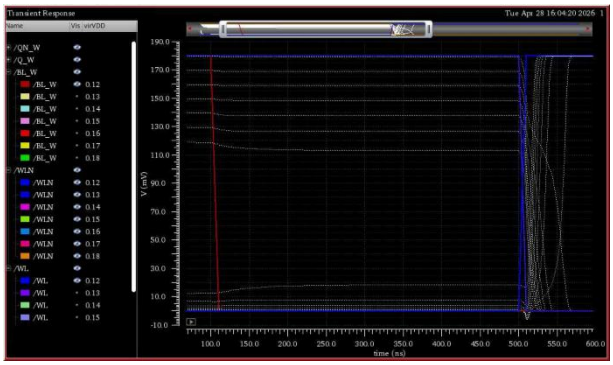

Figure 10. Write with 200ns BL delay

The same analysis was conducted for writing a 1 to the cell when the initial value is 0, with the following parameters: VDD = 180mV, virGND = GND = 0V, $Q_{initial}$ = 0V, $QN_{initial}$ = 1V. The voltage of virVDD was then swept from 180mV - 90mV. Here, the cell became unstable for values of virVDD below 120mV. Writing was then tested within this range. Unlike the previous results, writing worked for all tested voltages even with a delay of 200ns between the assertion of the bitline and WL/WLN. This implies the NMOS in this system is stronger, so writing a 0 to the cell is easier, making it the less critical case. This follows the findings of Zhai et al. (2008), where the NMOS/PMOS current ratio rises significantly in subthreshold, meaning that the NMOS becomes

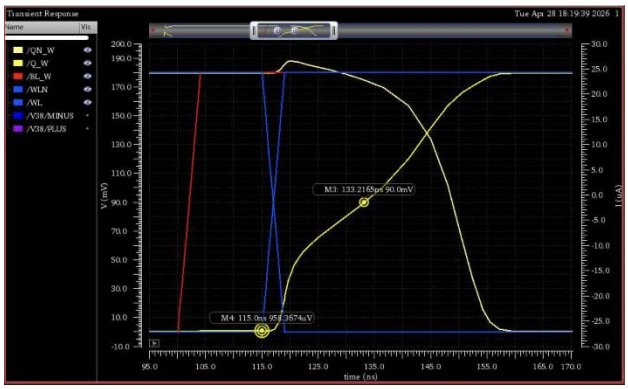

Figure 11. No droop write

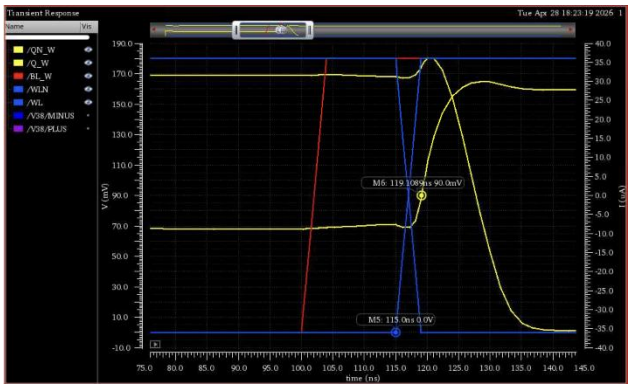

Figure 12. Optimized droop write

dominant. Therefore, for our cell, drooping virVDD to 90mV results in the most efficient write, with no major inaccuracy risks.

Finally, we tested both droops in combination on the critical case, which is writing a logical 1 to the cell when the initial value is 0. Setting virVDD to 120mV and virGND to 70mV resulted in a ~4ns write as opposed to an ~18ns write when virVDD = VDD = 180mV and virGND = GND = 0V. This ~4.5x speedup confirms the virtual power rail approach of Zhai et al. (2008) as an effective method for write speed-up. This said, weakening virVDD and virGND is not inconsequential. By drooping virVDD, the stored value is degraded, as it only reaches the full VDD voltage after virVDD is restored, reducing noise margin and increasing leakage. Therefore, our droop values are optimized for speed-up, but are more susceptible to noise and leakage-induced failures.

3.2. Reading Implementation

The implemented schematic was also tested for read functionality at subthreshold by precharging the bitline (whose capacitance was set to 10fF) to 180mV, then pulsing WL to 180mV and WLN to 0V. As can be seen the bitline discharges to GND, indicating a successful read. A read disturb can be seen on the Q line. This is because, when WL goes high during a read, BL connects to Q, resulting in a momentary spike before the pull-down system 'wins'.

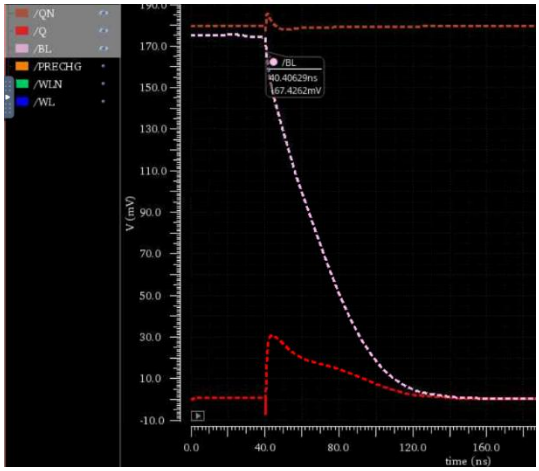

Figure 13. Read at 180mV

4. Comparative Analysis

To make clear the benefit of the optimizations implemented by Zhai et al., a number of comparative analyses were conducted. These involved producing identical metrics for a standard 6T SRAM cell, and the subthreshold optimized cell. Identical transistor sizing was used for the inverters and access transistors, in order to produce a fair comparison.

4.1. Static Noise Margins (SNMs)

Firstly, butterfly curves were produced for both cells at 1V and 0.2V, in both hold and read mode, in order to gauge static noise margins (SNMs). A butterfly curve is made by plotting the voltage transfer characteristics of the two cross-coupled inverters in an SRAM cell against each other, producing two "lobes" that show how strongly the cell holds either stored logic state. The static noise margin, or SNM, is the side length of the largest square that can fit inside one of those lobes, and it represents the maximum DC noise voltage the SRAM cell can tolerate before the stored bit becomes unstable or flips. To produce butterfly curves for SNM estimation, one link of the inverter cross-coupling was removed, and one of the internal nodes swept.

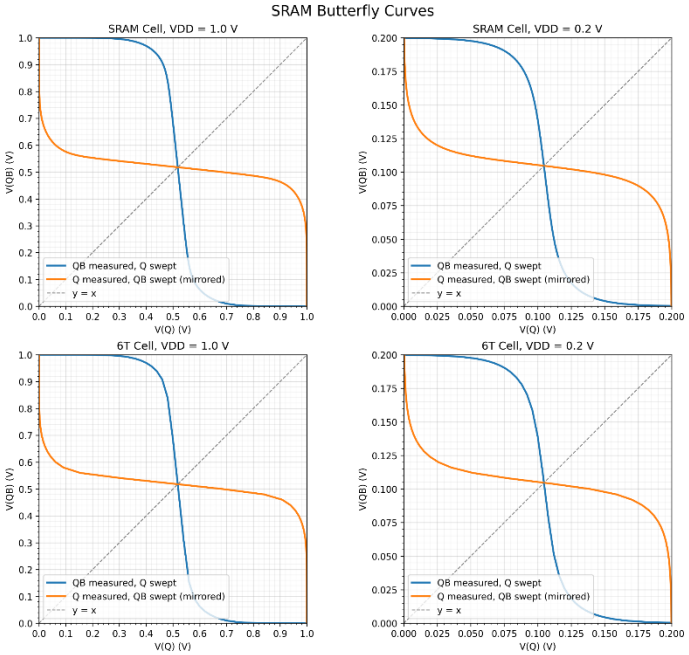

Figure 14. Hold butterfly curves for both cells (subthreshold cell top, standard 6T bottom)

Both hold and read butterfly curves were produced at 1V and 0.2V VDD. Unsurprisingly, the hold SNMs were near-identical for both cells, as no external disturbance affected what is in this case simply a cross-coupled inverter. This said, the SNM is seen to reduce at 0.2V, as is expected.

For the read SNM, the advantage of the subthreshold cell becomes clear. While SNM does reduce from 1V to 0.2V, this contrasts starkly with the standard 6T cell, where read SNM completely collapses at 0.2V.

The superior read performance of the subthreshold cell can be explained by its single bitline and full transmission gate. The standard 6T cell loses read SNM at subthreshold voltages as the internal nodes connect to two precharged bitlines. The 0 node is pulled upward through the access transistor, creating a read disturb. At low VDD, the pull-down/access ratio is unreliable as device currents are exponentially sensitive to threshold mismatch. The single-ended structure and transmission-gate structure of the subthreshold cell means noise is isolated to one bitline, the bitline can swing rail-to-rail (and the design avoids a low-voltage differential sense amplifier). Further, since writeability is handled separately using gated-feedback write-assist, the read path does not need to be compromised as much for write strength.

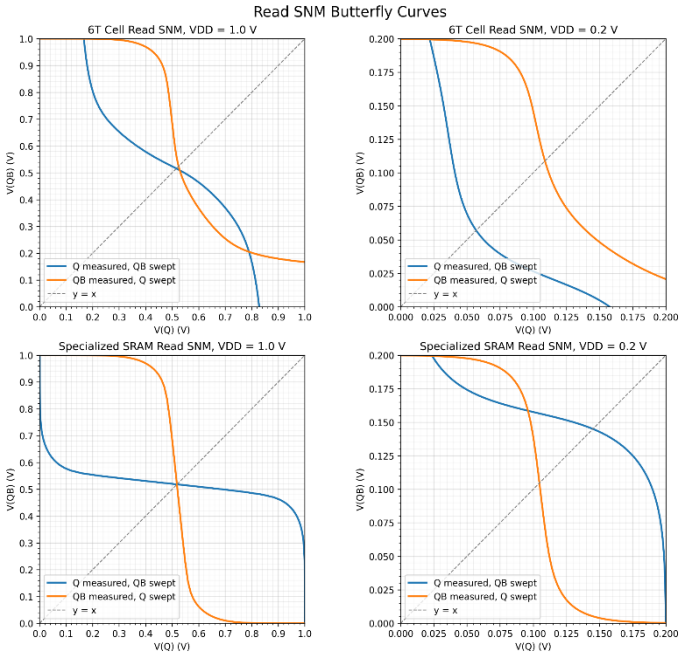

Figure 14. Read butterfly curves for both cells (standard 6T top, subthreshold cell bottom)

4.2. Write Performance

To compare write performance between the two cells, a write transient was run and the write delay measured for the standard 6T cell and the subthreshold cell at subthreshold voltage. This was done using the optimized values of virVDD and virGND for the subthreshold cell (120mV and 70mV respectively), and thus, a nominal VDD of 180mV was used for both cells. The results were fairly stark, with the write delay for the subthreshold cell being approximately 2ns, whereas for the standard 6T cell, the delay was ~16ns. Notwithstanding the fact that the standard 6T cell fails reading completely at the subthreshold, this result further indicates that writing without topological optimizations results in a significantly reduced potential operating frequency.

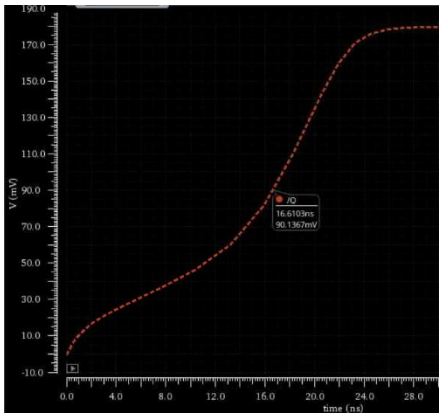

Figure 15. Slow standard 6T write

5. Conclusion and Relevance to Digital VLSI Circuit Design

Through reproducing the paper's design and testing this standalone and comparatively, it is clear that simply scaling a conventional 6T SRAM cell into the subthreshold regime is insufficient for reliable operation. At low VDD, reduced drive current, increased sensitivity to variation, degraded SNMs, and read/write competition are all critical constraints. The proposed subthreshold cell addresses these issues using design techniques including single-ended read operation, gated-feedback write assist, and transistor sizing. Drawing conclusions for the broader area of subthreshold digital VLSI design, lowering VDD requires optimizing the circuit topology, device sizing, timing assumptions, and robustness margins in unison. Further, from the paper, in addition to unit-level verification, it is necessary to characterize and validate at the system level, for example through statistical techniques like Monte-Carlo analysis, and implementation and testing in hardware.

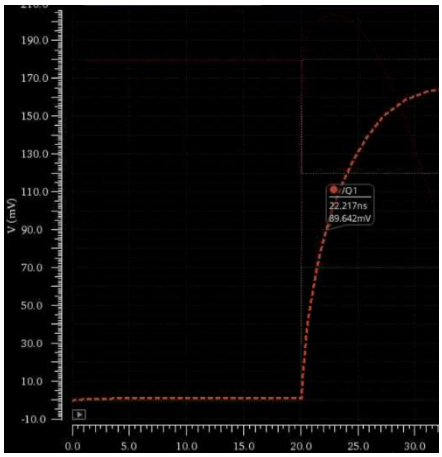

Figure 16. Fast subthreshold-optimized cell write