

Abstract

SRAM is a class of physical memory storage elements that utilize direct connections to rail voltages to establish and maintain voltage levels. A library of associated circuit elements is presented here, developed in an iterative and scaffolded fashion to facilitate future extensions or forks. Analysis of these elements explore the processes, considerations, and challenges involved when designing and creating physically realizable SRAM memory elements. Additional design restrictions are imposed in terms of size and speed, such that the research thrust of this report remains focused: analytical emphasis is placed on comparing the corresponding design requirements and device performance indices. To wit: late-stage candidate devices produced in this process feature partially-implemented address control circuitry while maintaining a reasonably small silicon footprint, minimized/standardized transistor gate dimensions, and write speeds in the MHz – GHz range. Final results demonstrate a 2x2 SRAM memory array designed to operate at mid-level electronics voltages under the performance conditions as defined.

All circuit iterations, design work, simulations, and analyses were implemented and/or executed via schematic- and layout-level representations, using appropriate development environments within the CMC Cadence CMOSIS5 design kit (c/o T. Obuchowicz and Concordia University, Montreal, QC, Canada).

Table of Contents

Project Overview <i>Introduction; Design strategy; Analytical strategy.</i>	1
Design Phase 1 – Single 6T-SRAM Cell	2
Design Phase 2 – 2x2 SRAM Array with Write-Only Bit Lines	7
Design Phase 3 –SRAM with Word Line Control and Generalized Bit Lines <i>Design methodologies; Analytical results; Challenges encountered.</i>	12
Additional Discussion	19
Conclusions	20
Reference	20

Table of Figures

Figure 1: Six-transistor SRAM memory cell, schematic representation . . 2
Figure 2: Six-transistor SRAM memory cell, custom min-size layout representation 3
Figure 3: Performance of 6T-SRAM memory cell, schematic representation . 4
Figure 4: Performance of 6T-SRAM memory cell, layout representation . . 5
Figure 5: 2x2 SRAM memory array, schematic representation . . 7
Figure 6: 2x2 SRAM memory array, custom min-size layout representation . 8
Figure 7: Performance of 2x2 SRAM memory array, schematic representation . 9
Figure 8: Performance of 2x2 SRAM memory array, layout representation . 10
Figure 9: 2x2 SRAM with word line control, schematic representation . . 13
Figure 10: 2x2 SRAM with word line control, custom min-size layout representation 15
Figure 11: Performance of 2x2 SRAM with word line control, schematic representation 16
Figure 12: Performance of 2x2 SRAM with word line control, layout representation 17

Project Overview

Static RAM (SRAM) is an important class of data storage devices that finds utility in several applications: caching, real-time systems, and FPGAs, to name a few. The theory behind its functionality seems straightforward and rather convenient: by applying static gate logic to latches, one can take advantage of their regenerative feedback mechanism and store user data using access voltages with relatively short pulse widths. This simplicity, however, belies some important design considerations [1] that must be respected when realizing such a circuit with physical materials. This project attempts to demonstrate the feasibility of scaling up SRAM-based memory towards an end-user accessible product in mid-voltage electronics systems.

Design Strategy: For this project, the selected primary restriction is to keep all transistor gates as small as possible; this is self-imposed to provide direction and focus. From this point the design proceeds in phases: Phase 1 is a single SRAM cell constructed to provide baseline performance parameters, and to serve as a building block for later phases. Practical gate widths for physical layout are selected here to ensure good low-level performance before scaling up. In Phase 2, this cell is duplicated to create a 2x2 SRAM array as a “proof of concept”, as well as to observe the change in performance as the number of cells in an array begins to increase. Finally, Phase 3 endeavours to make the memory array more field-deployable by abstracting the functionality, e.g., by adding control devices (i.e., additional transistors) and reducing the number of required inputs. These modifications are also expected to affect the device’s performance.

For cohesion of analysis, consistency of scope, and out of respect for the length requirements of this report, only the SRAM’s “write” mode is considered in this discussion.

Analytical Strategy: The performance of each design iteration is compared in terms of size (footprint, number of transistor devices), speed (delay, and corresponding peak operating speed), and power consumption. For each individual phase, observations and deductions are made regarding the shift in performance between the schematic “ideal” circuit and that of the physical layout (including parasitic capacitances). The iterations are also compared against one another to quantify the effects of expanding and abstracting the memory storage array under analysis.

All design work, simulations, and analyses are performed using the Cadence CMOSIS5 software suite.

Design Phase 1 – Single 6T-SRAM Cell



Design: Initial experimentation resulted in a simple latch-based 6T-SRAM cell realized in schematic representation (Figure 1). After simulating functionality with generous input signal pulse widths (e.g., 10 ns), the circuit was quickly optimized for space and speed. This was achieved by reducing all transistor (W/L) ratios to the smallest size achievable in CMOSIS5 and shortening the input pulses. While creating the layout representation of this circuit, the smallest feasible transistor ratio was determined to be 1200/600 nm, which allows good performance while also facilitating layout design (see below). The schematic blueprint was recreated to reflect this choice.

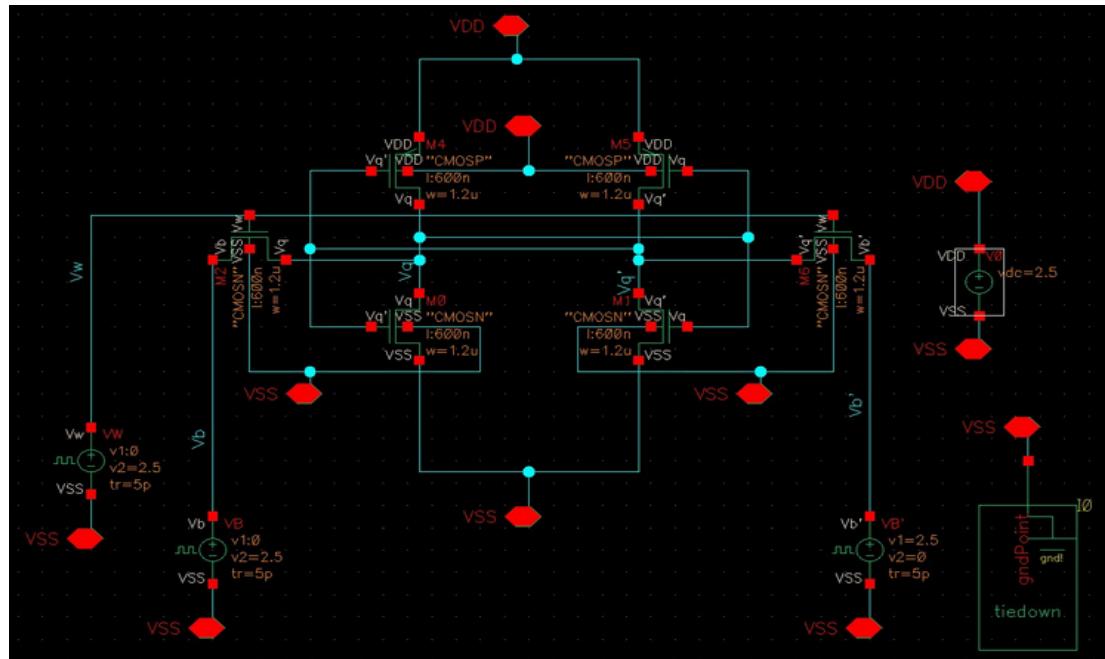


Figure 1: Six-transistor SRAM memory cell, schematic representation

The circuit design was then translated into its corresponding physical layout representation (Figure 2). As mentioned, minimum transistor gate dimensions in layout were defined as 1200/600 nm; this design change was reflected back to the schematic representation before moving to Phase 2. These dimensions were chosen to balance small size with clean geometries, and allow key parts of the circuit to be modularized and repeated in later iterations.

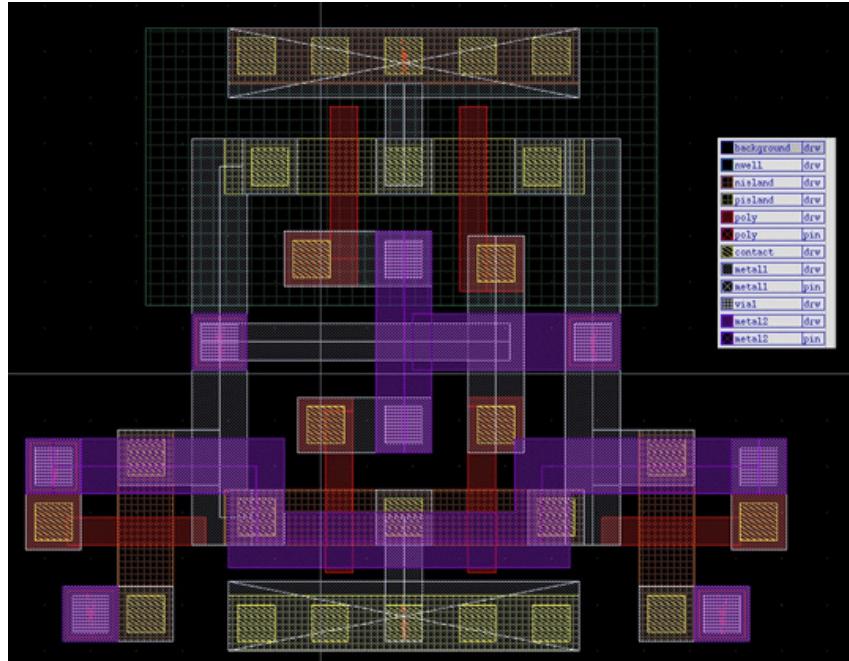


Figure 2: Six-transistor SRAM memory cell, custom min-size layout representation

Analysis: The layout representation circuit was extracted (with parasitic capacitances) before creation of an abstracted symbolic representation. LVS verification was carried out to ensure matching functionality between the layout and schematic circuits. The Phase 1 schematic circuit was then simulated using transient analysis of the relevant input and output signals; representative waveforms were captured to help determine the performance (Figure 3). The corresponding layout (symbolic) circuit was then simulated with identical testing conditions to produce a comparative set of circuit activity data (Figure 4).

Finally, quantitative comparison of key performance indices was carried out between the two circuit representations (Table 1). Data sources include the Cadence data calculator applet, manual analysis of the transient waveforms, and direct inspection or observation of circuit states and simulation layouts.

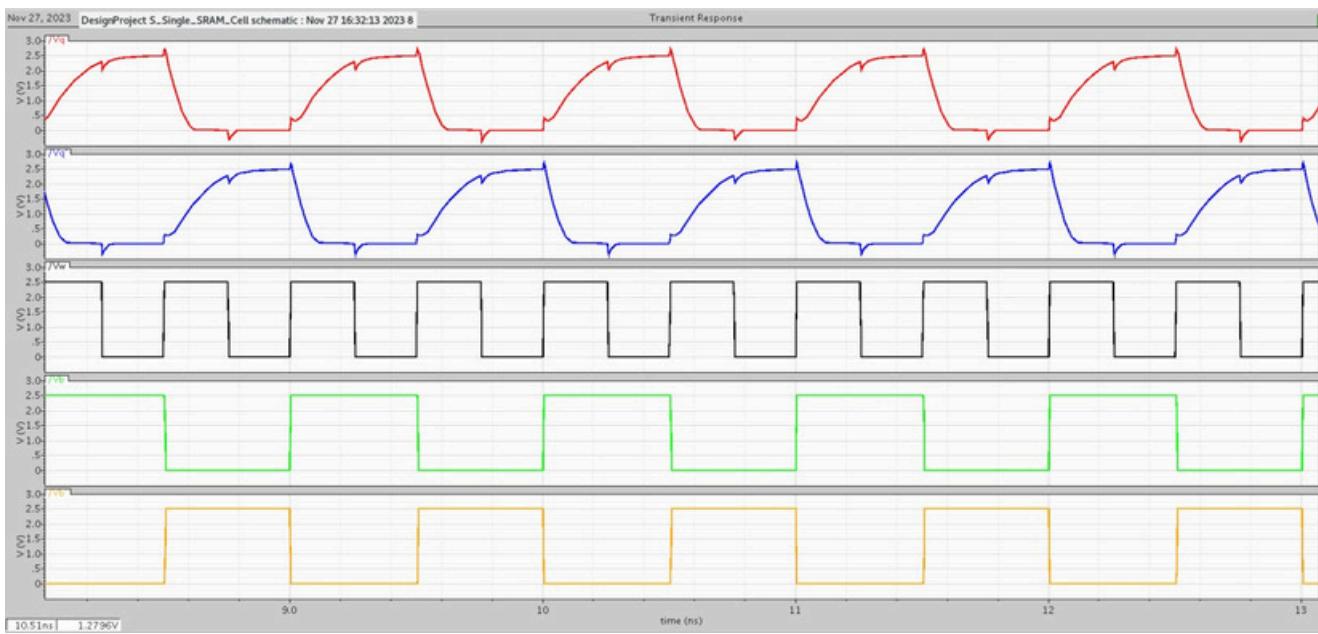


Figure 3a: Performance of 6T-SRAM memory cell (operating limit), schematic representation: I/O

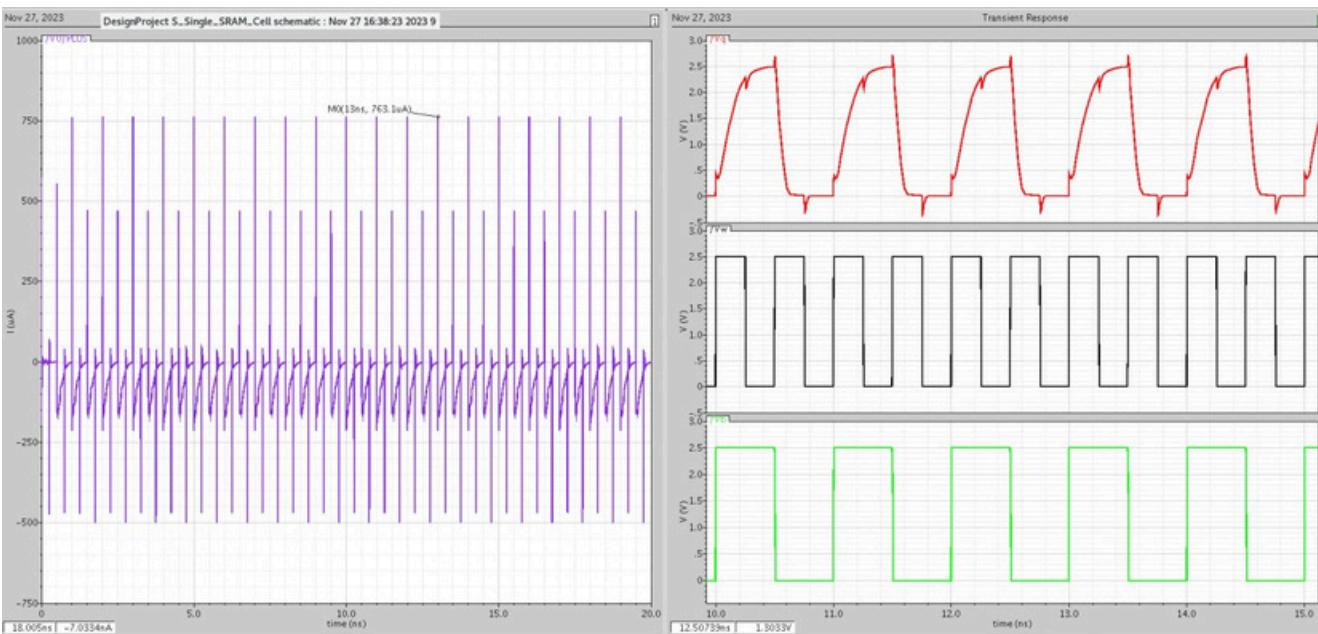


Figure 3b: Performance of 6T-SRAM memory cell (operating limit), schematic representation: current vs. voltages

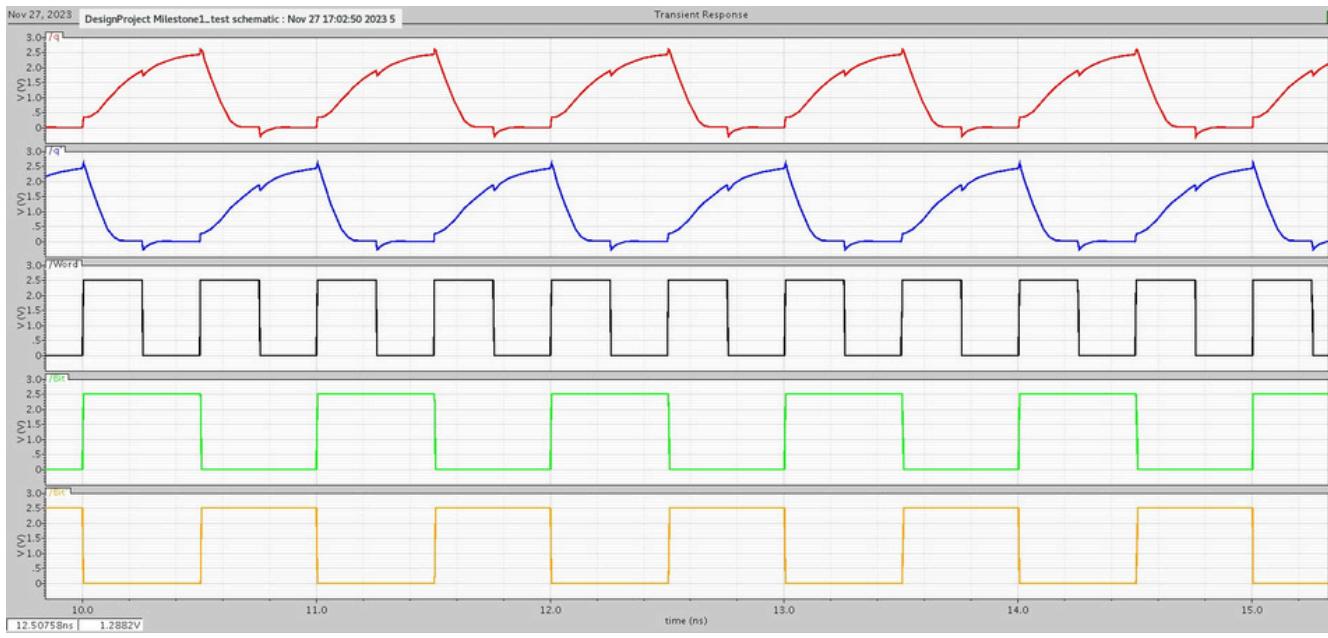


Figure 4a: Performance of 6T-SRAM memory cell (operating limit), layout representation: I/O

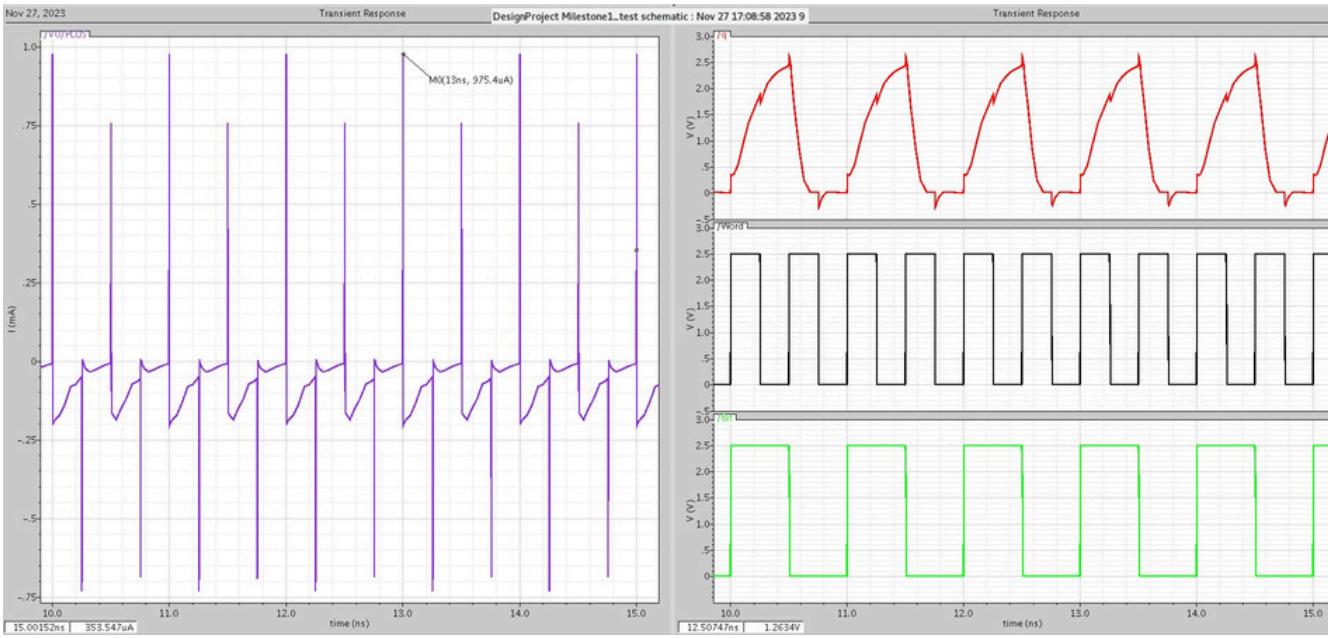


Figure 4b: Performance of 6T-SRAM memory cell (operating limit), layout representation: current vs. voltages

Table 1: Comparative analysis of Phase 1 design and performance indices

	Index	Schematic Representation	Physical Layout Representation	
	Supply Voltage	2.5 V		
Size	Number of Transistors <ul style="list-style-type: none">• PMOS• NMOS	2, all min size 1 4, all min size		
	Maximum Footprint on Silicon	N/A	16.50 μm width x 13.50 μm height	
Speed	Test parameters at operating limit	Tword = 0.50 ns Tbit = 1.00 ns Wword = 0.25 ns ^{Wbit} _{≤ 0.50} ns	Tword = 0.50 ns Tbit = 1.00 ns Wword = 0.25 ns Wbit = 0.50 ns	
	Max writes per second 2	2 x 10 ⁹	2 x 10 ⁹	
Power	Propagation Delay 3 <ul style="list-style-type: none">• Rise time• Fall time• Transition Hi-to-Lo• Transition Lo-to-Hi	tr = 0.25 ns tf = 0.07 ns tphl = 0.10 ns tplh = 0.04 ns	tr = 0.35 ns tf = 0.11 ns tphl = 0.09 ns tplh = 0.15 ns	
	Peak Device Operating Current	763.1 μA	975.4 μA	
	Peak Operational Static Power	1.908 mW	2.439 mW	
	Average Device Operating Current	47.38 μA	68.38 μA	
	Average Static Power Consumption	118.45 μW	170.95 μW	

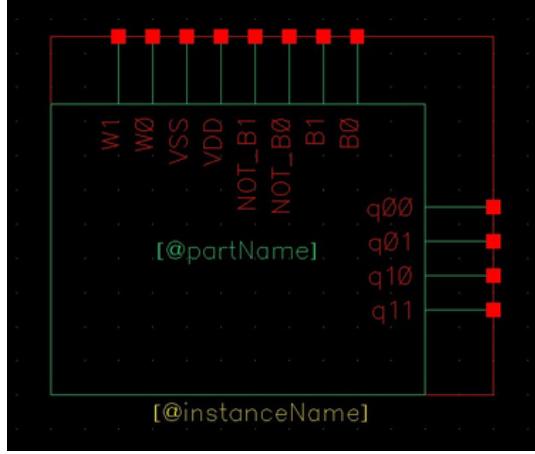
Notes

¹ Min size transistors defined above as $(W/L) = 1200/600$ nm for practicality.

² With respect to Tword. Higher rates possible under specific conditions; listed rate holds in general.

³ tr, tf defined in range $V_q = [0.1VDD, 0.9VDD]$. tphl, tplh defined between $V_w = 0.5VDD$ and $V_q = 0.5VDD$.

Design Phase 2 – 2x2 SRAM Array with Write-Only Bit Lines



Design: At the schematic level, the modularized SRAM cell was duplicated three times and the copies arranged in a 2x2 array (Figure 5). Each cell was given an arbitrary address (B,W) for use throughout the remainder of the design process. While chiefly aesthetic, this is an important high-level design decision to help chart signal flow and facilitate any debugging that might be needed later. Finally, bit lines and word lines were constructed and connected to the cells.

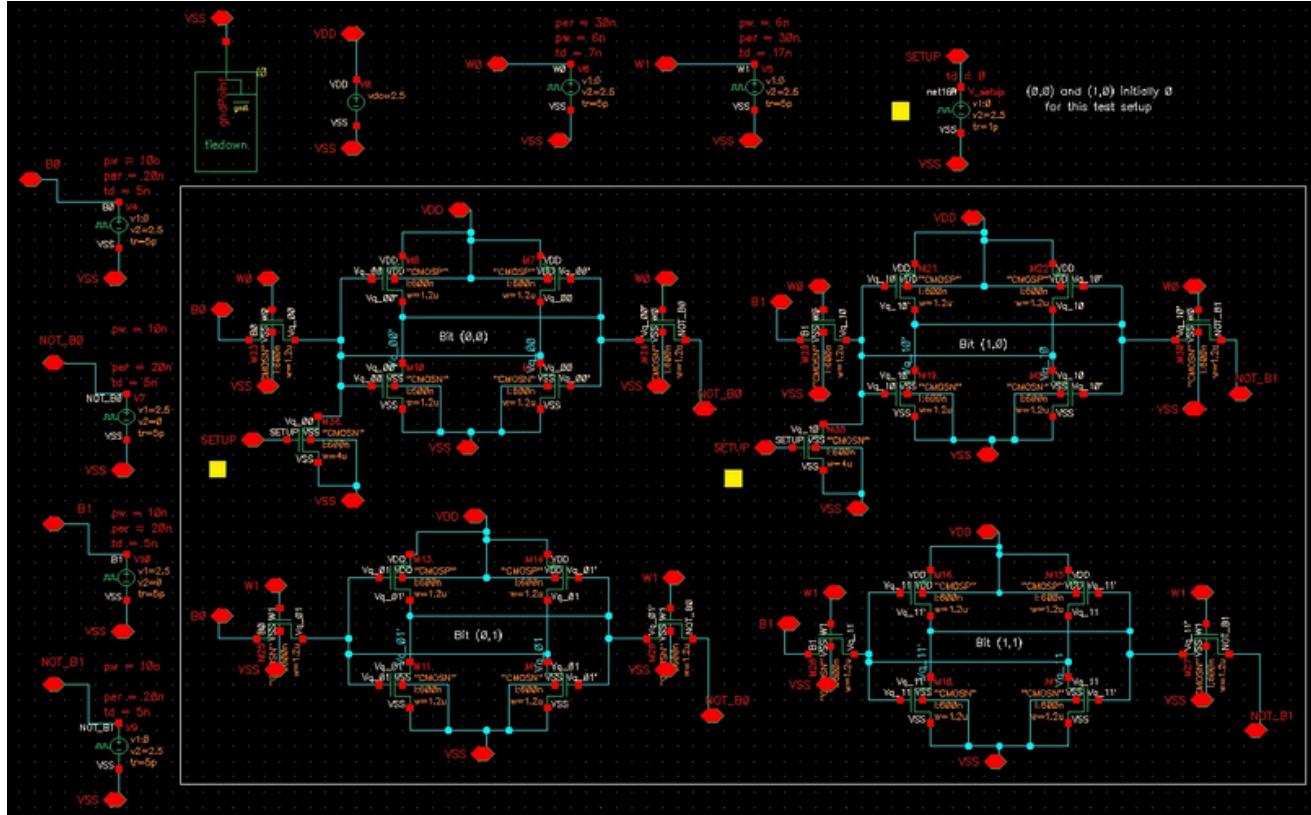


Figure 5: 2x2 SRAM memory array, schematic representation

The stated objective of Phase 2 is to observe the effect of increasing the number of cells in an array; to this end, control signals were held as nearly-ideal pulse voltages in this phase.

Further, the complement $B_{n'}$ of each bit line B_n is explicitly defined as $B_{\bar{n}} = \overline{B_n}$ as a way of

“hardcoding” the write functionality and obviating any further parametrical interference. In addition, “setup” transistors were added during simulation (see Figure 5) to assert logic “0” in arbitrary cells during certain debugging tests. These transistors are not present in the final design.

The circuit design was then translated into its corresponding physical layout representation (Figure 6). With careful attention to topology, it was possible to clone the cell linearly in space and achieve the desired logic without a third metal layer, while also maintaining a minimum of unused space in the horizontal dimension. By inspection, however, the likely tradeoff comes in the form of increased signal delay due to steadily increasing material overlap.

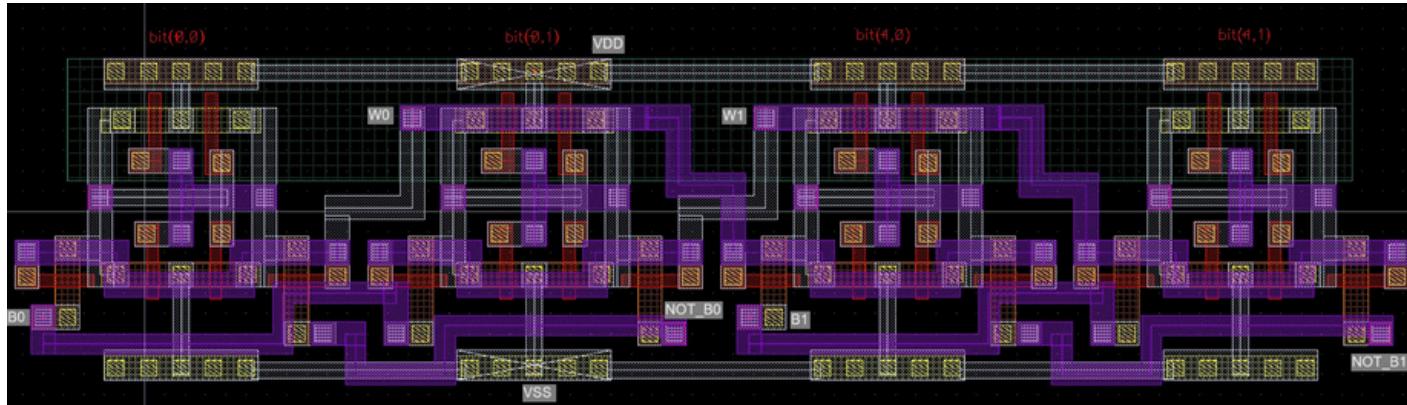


Figure 6: 2x2 SRAM memory array, custom min-size layout representation

Analysis: In a similar fashion as Phase 1, the layout representation circuit was extracted with parasitic capacitances before abstracting it into a symbolic representation. LVS verification in this phase was successfully executed before proceeding. The Phase 2 schematic circuit was then simulated using transient analysis of the relevant input and output signals; the representative waveforms are summarized in Figure 7. The corresponding layout (symbolic) circuit was also simulated, again with identical testing conditions, to produce the respective dataset (Figure 8). Finally, quantitative comparison of key performance indices was carried out between the two circuit representations, the results of which are presented in Table 2 below.

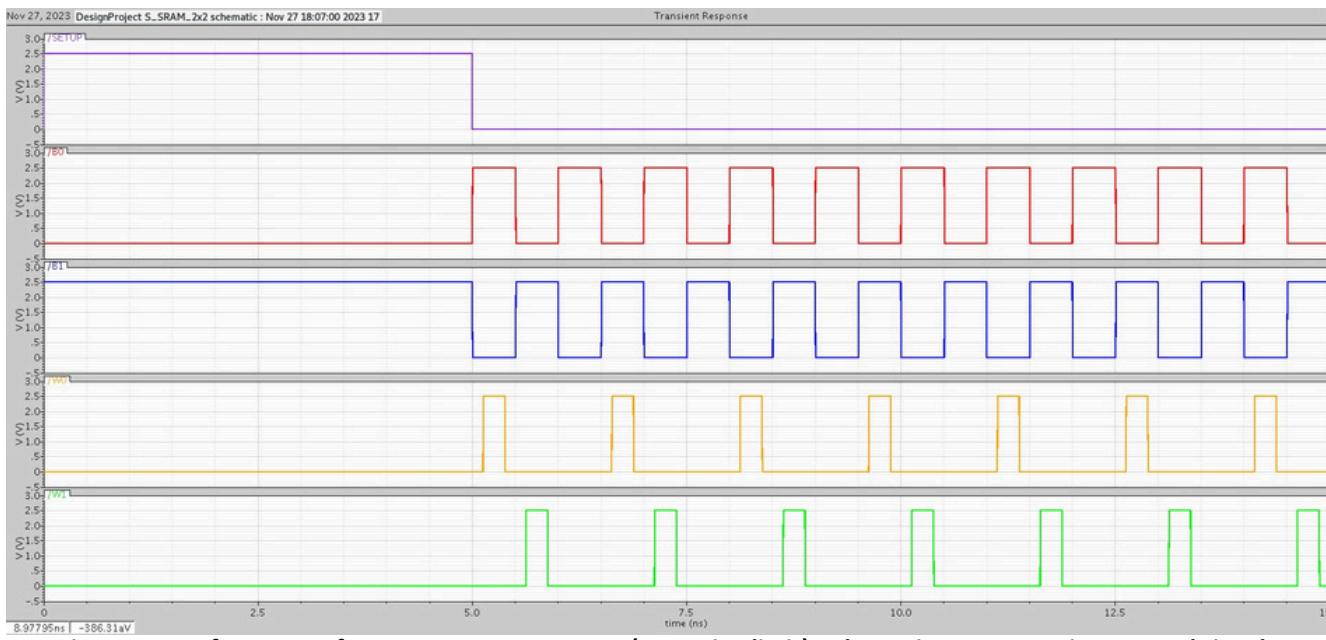


Figure 7a: Performance of 2x2 SRAM memory array (operating limit), schematic representation: control signals

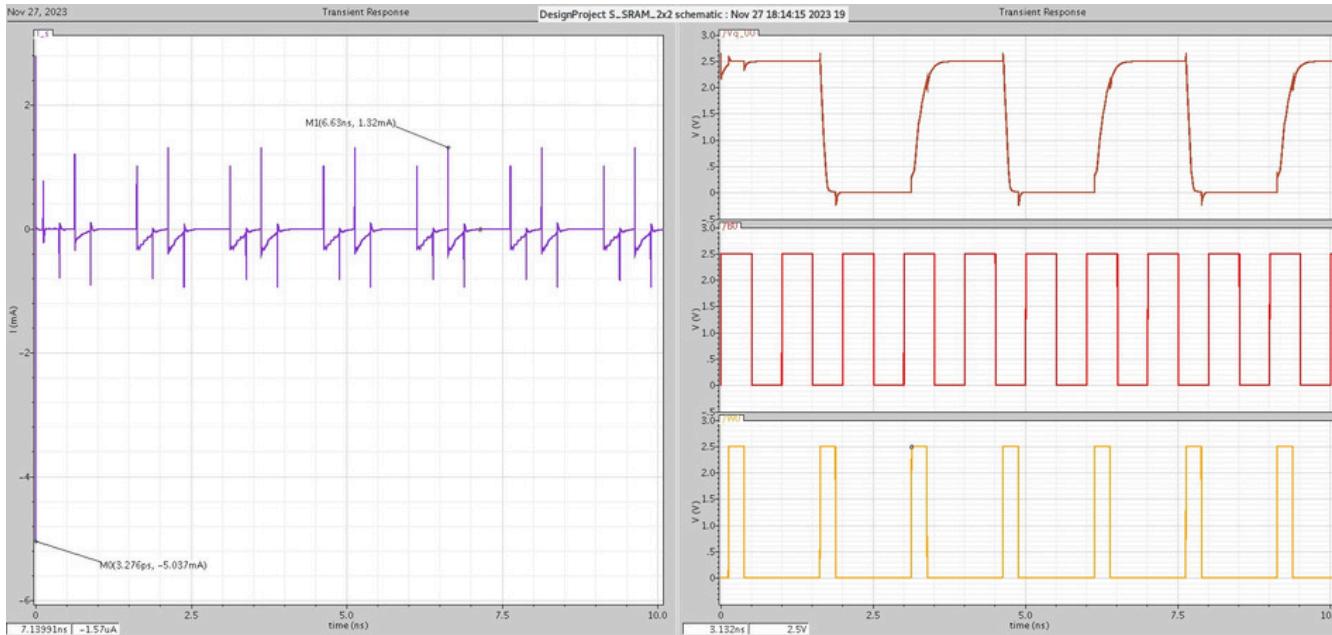


Figure 7b: Performance of 2x2 SRAM memory array (operating limit), schematic representation: current vs. voltages

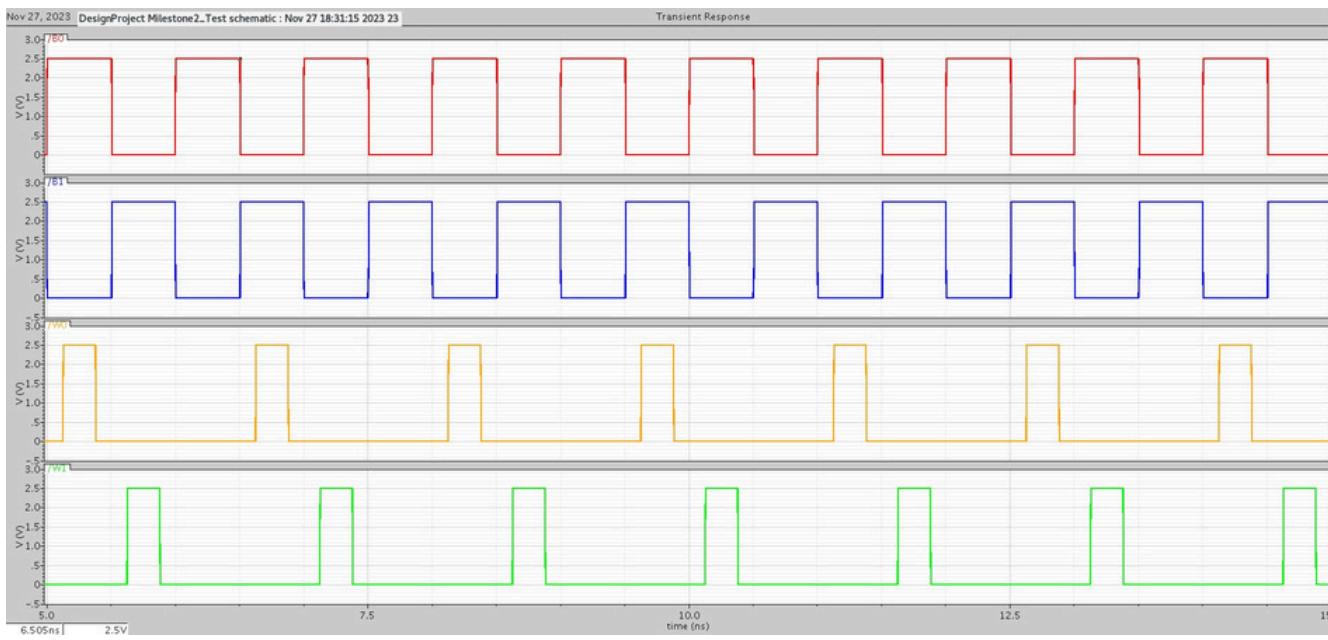


Table 2: Comparative analysis of Phase 2 design and performance indices

	Index	Schematic Representation	Physical Layout Representation	
	Supply Voltage	2.5 V		
Size	Number of Transistors <ul style="list-style-type: none">• PMOS• NMOS	8, all min size 1 16, all min size		
	Maximum Footprint on Silicon	N/A	68.70 μm width x 16.00 μm height	
Speed	Test parameters at operating limit 2	Tword = 1.50 ns Tbit = 1.00 ns Wword = 0.25 ns ^{Wbit} _{≤ 0.50} ns	Tword = 1.50 ns Tbit = 1.00 ns Wword = 0.25 ns Wbit = 0.50 ns	
	Max writes per second 3	2 x 10 ⁹	2 x 10 ⁹	
Power 5	Propagation Delay 4 <ul style="list-style-type: none">• Rise time• Fall time• Transition Hi-to-Lo• Transition Lo-to-Hi	tr = 0.29 ns tf = 0.09 ns tphl = 0.06 ns tplh = 0.11 ns	tr = 0.40 ns tf = 0.10 ns tphl = 0.10 ns tplh = 0.15 ns	
	Peak Device Operating Current	1.32 mA	2.00 mA	
	Peak Operational Static Power	3.30 mW	5.00 mW	
	Average Device Operating Current	63.06 μA	96.54 μA	
	Average Static Power Consumption	157.65 μW	214.35 μW	

Notes

¹ Min size transistors defined above; see Phase 1. “Setup” transistors not included.

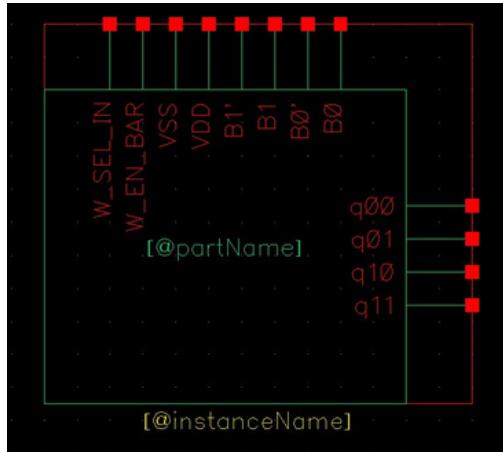
² Test conditions include VW0 = VW1 = 0 once per period to confirm consistent “no write” functionality.

³ With two writes per Tword. Higher rates theoretically possible under specific conditions; see Note 2.

⁴ tr, tf defined in range Vq_xy = [0.1VDD, 0.9VDD]. tphl, tplh defined between VWy = 0.5VDD and Vq_xy = 0.5VDD.

⁵ Schematic representation values calculated without “Setup” transistors present (see e.g., Fig. 7b)

Design Phase 3 – SRAM with Word Line Control and Generalized Bit Lines



Design: After confirming the functionality and performance of the original 2x2 SRAM array, steps were taken to introduce control mechanisms and improve device portability. The most suitable entry point was deemed to be the word lines, being easier to route than the bit lines and requiring only one driver signal per row. The control logic in this case is also easier to implement.

Circuit extensions continued in this vein until the end of the experiment's design iteration period. The resultant circuit features "enable"-triggered word line control and generalized bit lines, both intended to facilitate future expansion efforts and further iterations on the system.

This final device was constructed in schematic representation (Figure 9) to confirm correct operation and allow for low-level parametric testing.

The word line control circuitry includes a buffered input port for processing the raw selection signal. This pre-device is in place to encourage a more predictable device delay when inputs are not ideal, while allowing easy modularization of the design if additional rows are ever desired. To reduce transistor count (and thus silicon footprint), the word line drivers use negative-level (NOR gate) logic rather than positive-level (AND gate) logic. Terminal NMOS devices are added to ensure timely word line discharge once any enable signals are de-asserted.

Finally, bit line voltage ports have been generalized – i.e., the restriction imposed during Phase 2 that $B'n = \bar{B}n$ is lifted in this case. This can allow more flexibility when fine-tuning a memory system, as well as eventually allowing read functionality to be realized. While efforts were undertaken to implement full bit line control, no suitable prototype could be constructed within the allotted timeframe that fully satisfied the desired design constraints (see Discussion).

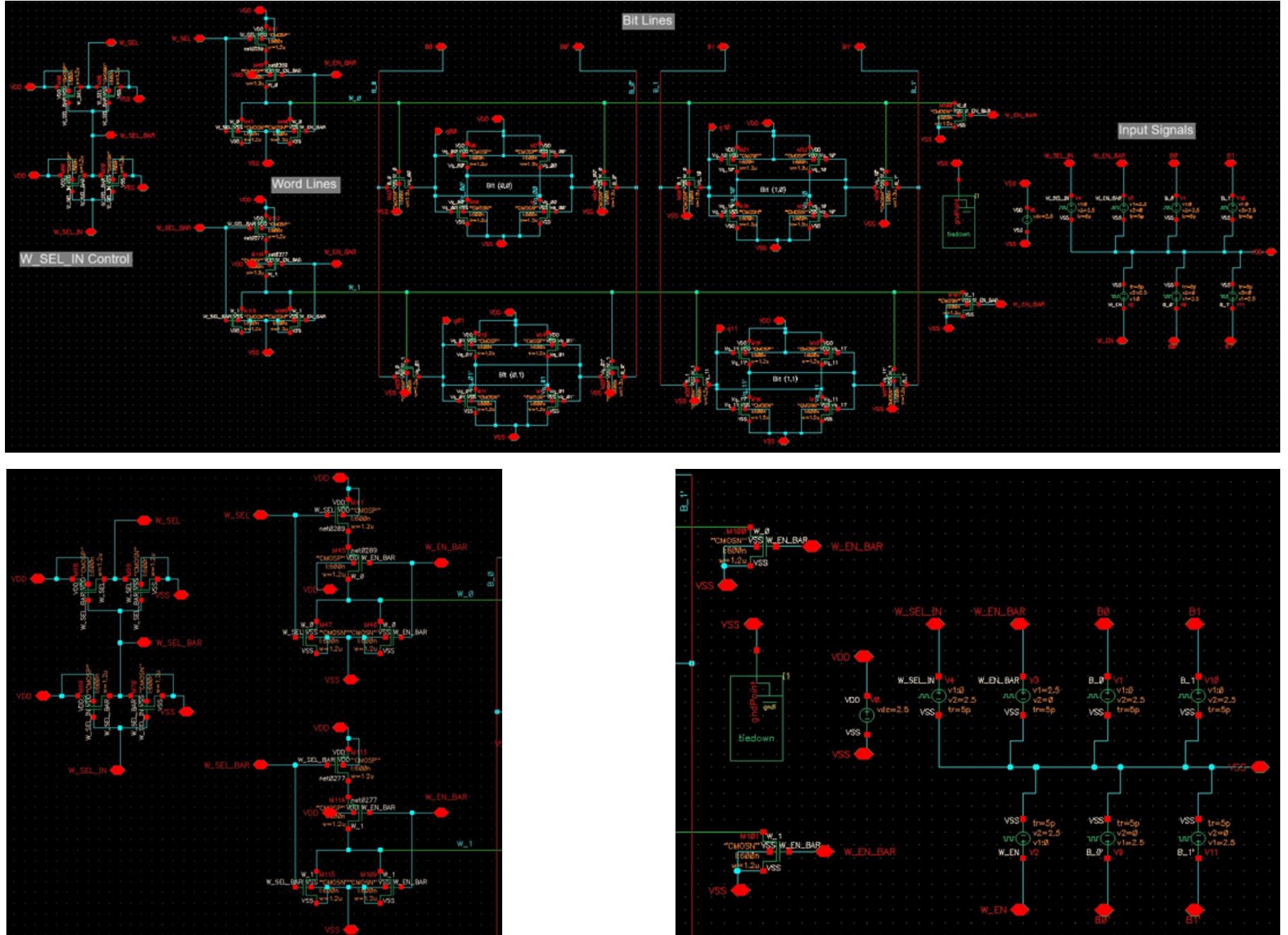


Figure 9: 2x2 SRAM with word line control, schematic representation. Top: Overview. Left: Word line detail. Right: Input signal detail.

Extending the circuit into physical space between Phases 2 and 3 is clearly more involved than between Phases 1 and 2 (see Figure 10). A fundamental obstacle is the essential difference in functionality: Phase 2 is easy to implement by modularizing Phase 1, copying the cells, and paying reasonable attention to the high-level logic. Realizing Phase 3, on the other hand, involves new circuitry and many more interconnects due to the word line drivers and terminal NMOSs. Consequently, in order to respect the original set of design decisions (minimum gate dimensions; avoid metal-3; etc.), additional precision was required in the project scope at this point.

Therefore, the decision was made to minimize silicon footprint at the expense of other parameters. Thus, all transistor sizes were kept uniform regardless of their role in the circuit. The expected tradeoff, of course, is a mitigation of the desired high processing rate: increased wire lengths affect maximum swing at target nodes due to parasitic resistances, while the accumulating material overlaps will extend signal propagation delays due to the parasitic capacitors. As these parameters grow, the parasitic elements become increasingly non-negligible.

Analysis: Following the protocol established heretofore, an extracted view of the Phase 3 circuit was created from its layout representation, and LVS verification confirmed the device logic.

The Phase 3 schematic circuit was then simulated using transient analysis of the relevant input and output signals; the representative waveforms are presented in Figure 11. While the corresponding layout (symbolic) circuit was also simulated, attempts that use identical testing conditions ultimately fail at this level, as the voltage swing ΔVq for a given memory cell qXY always falls short of $V_{th,inv}$. These conditions cannot initiate the crucial regenerative processes in the latches upon which SRAM relies. As predicted, this barrier approximates the device performance limit when using the selected design methodology, i.e., emphasizing small gates and silicon footprint over other performance indices. This presented an opportunity to explore this extreme.

To this end, it was determined after further testing (see Discussion) that the circuit was able to demonstrate the desired functionality by permitting an increase of 50% in the “enable” signal pulse width. Simulation was carried out under these circumstances to produce the layout circuit’s I/O dataset (Figure 12). Finally, a quantitative comparison of key performance indices was carried out between the two representations used in this final phase of design iteration (Table 3).

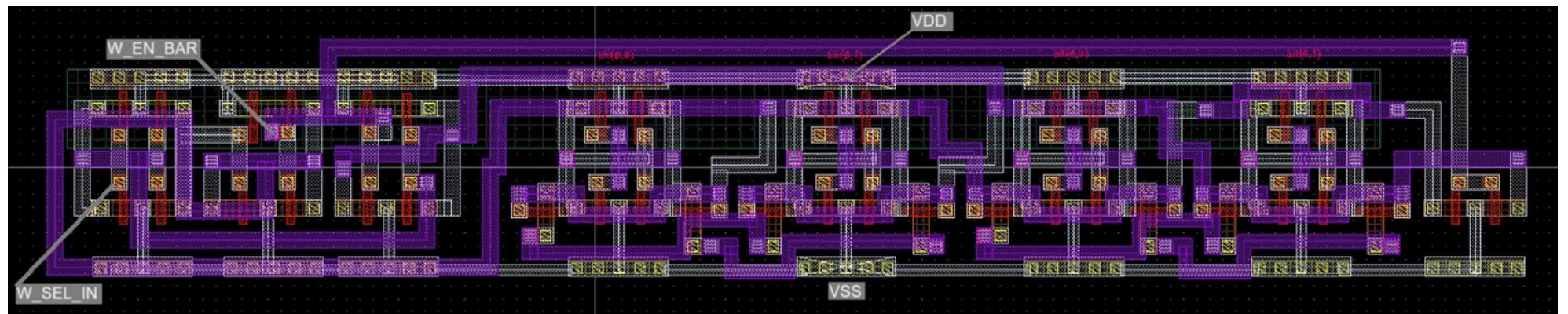


Figure 10: 2x2 SRAM with word line control, custom min-size layout representation

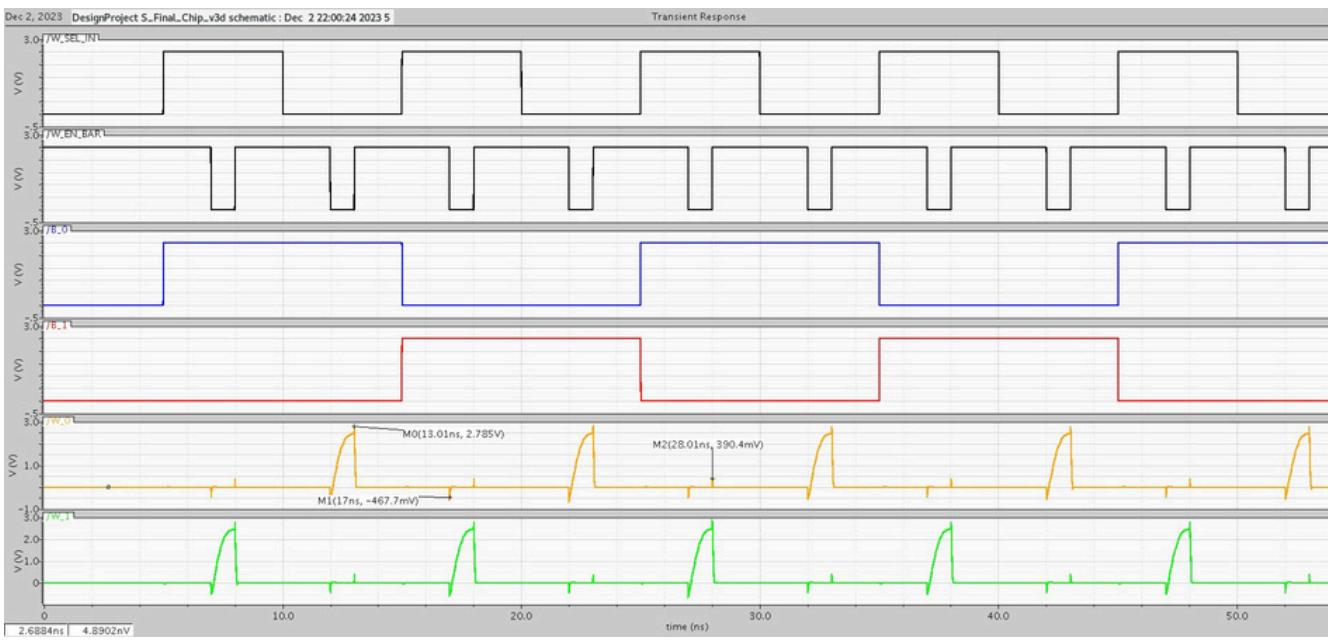


Figure 11a: Performance of 2x2 SRAM with word line control (guaranteed stable operation), schematic representation: control signals

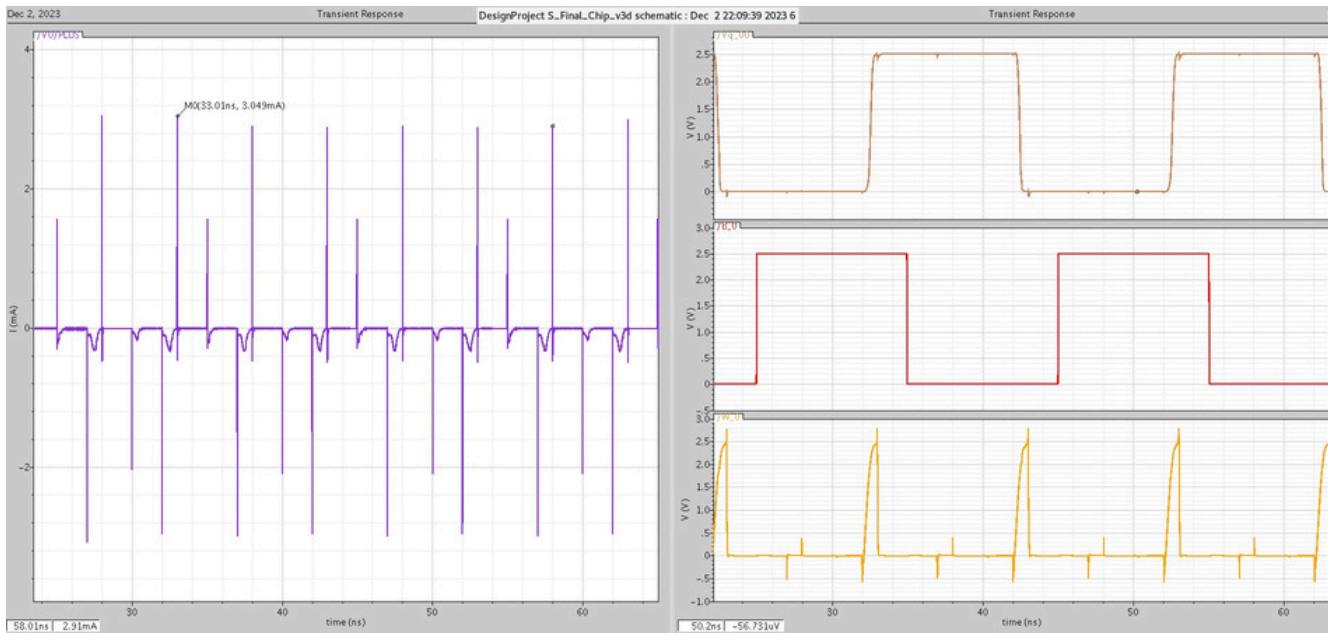


Figure 11b: Performance of 2x2 SRAM with word line control (guaranteed stable operation), schematic representation: current vs. voltages

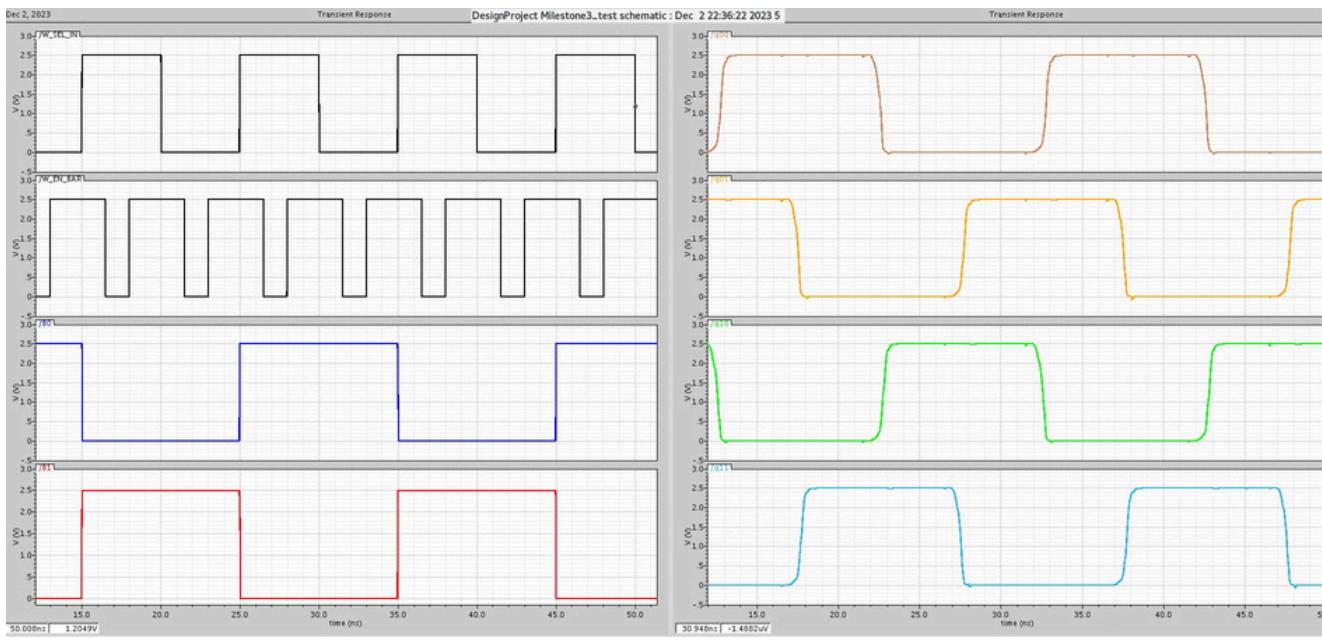


Figure 12a: Performance of 2x2 SRAM with word line control (guaranteed stable operation), layout representation: I/O

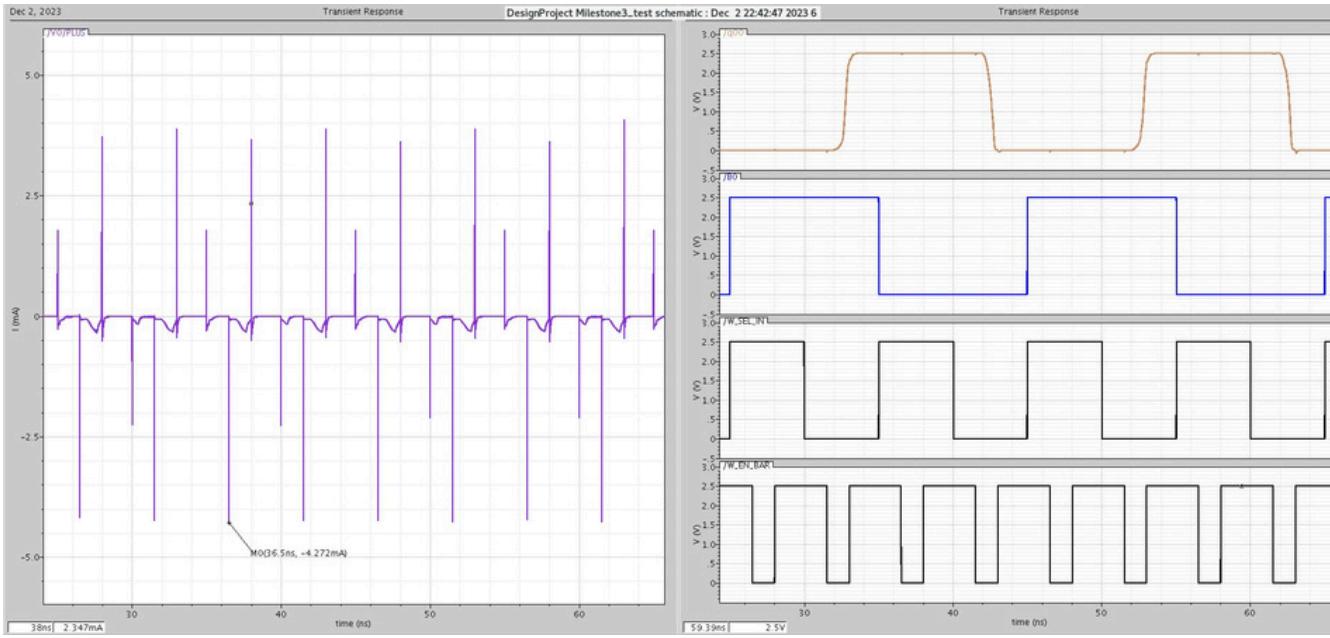


Figure 12b: Performance of 2x2 SRAM with word line control (guaranteed stable operation), layout representation: current vs. voltages

Table 3: Comparative analysis of Phase 3 design and performance indices

	Index	Schematic Representation	Physical Layout Representation	
	Supply Voltage	2.5 V		
Size	Number of Transistors <ul style="list-style-type: none">• PMOS• NMOS	14, all min size 1 24, all min size		
	Maximum Footprint on Silicon	N/A	113.20 μm width x 18.30 μm height	
Speed	Test parameters for stable operation 2	WW_EN_BAR = 1.0 ns WW_SEL_IN = 5.0 ns	WW_EN_BAR = 1.5 ns WW_SEL_IN = 5.0 ns	
	Max writes per second 3	2 x 108 - 1 x 109	2 x 108 - 6.67 x 108	
	Propagation Delay 4			
	<ul style="list-style-type: none">• Rise time• Fall time• Transition Hi-to-Lo• Transition Lo-to-Hi	tr = 0.28 ns tf = 0.23 ns tphl = 0.12 ns tplh = 0.24 ns	tr = 0.46 ns tf = 0.50 ns tphl = 1.07 ns tplh = 1.27 ns	
Power⁵	Peak Device Operating Current	3.094 mA	4.29 mA	
	Peak Operational Static Power	7.735 mW	10.725 mW	
	Average Device Operating Current	35.75 μA	61.91 μA	
	Average Static Power Consumption	89.375 μW	154.775 μW	

Notes

¹ Min size transistors defined above as $(W/L) = 1200/600$ nm for practicality. See Phase 1.

² That is, fully enable-triggered writes with wide enough pulse widths to accommodate non-ideal B_n , B_n' setup times; this tolerance is held over from the (failed) experiments in controlling the bit lines.

³ Listed range is for writes initiated with / without bit line tolerances; see Note 2.

⁴ tr, tf defined in range $V_{q_xy} = [0.1VDD, 0.9VDD]$. tphl, tplh defined between $VW_EN_BAR = 0.5VDD$ and $V_{q_xy} = 0.5VDD$.

⁵ Due to differences in test parameters, avoid direct comparisons against this layout representation.

Additional Discussion

A summary of the parametrical analysis amongst all design phases presented in this paper can be found in Table 4 below, demonstrating the worst-case results in physically realized devices operating at maximum performance. As expected, adding functionality to the circuit necessarily increases the number of transistors required, and therefore the area it must occupy. This also drives up delays and power consumption: the magnitude of supply current increases alongside effective gate dimensions, as do the parasitic elements with increasing wires and semiconductors.

Table 4: Comparative summary of representative performance indices in physically realized memory

Iteration	Area (μm^2)	Delay (ps)	Writes per second	Pstatic,pk (mW)	Pstatic,av (μA)
Phase 1: Single Cell	222.75	350	2×10^9	2.439	170.95
Phase 2: 2x2 Array	1099.20	400		5.00	214.35
Phase 3: Array / Controls	2071.56	1270		10.725	154.775

A point about dynamic power analysis should be mentioned here. While the concept of switching frequency does not strictly apply to this circuit (as it does not require a clock to function), dynamic power consumption may still be of interest when using this design at e.g., a very high number of writes per second. Ultimately, this analysis was omitted due to the difficulty in quantifying the parasitic capacitors in Phase 3 without additional tools, e.g., Cadence's Assura.

Regrettably, every creative project has a deadline, and compromises must be made to ensure that a complete product can be delivered once it expires. In this case, for example, the bit line circuitry proved far too sensitive to control reasonably within the project timeframe, especially given the primary design constraints. While promising results were under development by ignoring these constraints (i.e., varying device transistor gate dimensions), analytical efforts were ultimately insufficient to allow a reasonable, functional product to be delivered on time.

Additionally, comprehensive library analysis was stymied by the Phase 3 circuit and its inability to function at the established write speed. As mentioned above, plans for improving the circuit effectively consisted of two divergent methods: either enlarge some transistors to drive more current (thereby reducing delay), or change the “enable” pulse width and allow the cells more time to process the data signal. The former scheme increases silicon footprint while

potentially maintaining high write speed, while the latter sacrifices write speed to retain its small size. Ultimately, the latter was chosen to respect the primary design goal and to avoid lengthy circuit redesign. While “real world” circumstances would dictate the “correct” choice in this case, one can still note that making these (important!) decisions inevitably complicates the analysis. For example, the true meaning behind the average static power consumption in Phase 3 (see e.g., Table 4) is obscured, as it was unable to operate under the same conditions as its predecessors.

Conclusions

As predicted, extending the basic SRAM memory cell by enlarging the array and/or adding circuitry necessarily decreases performance. The two primary reasons for this degradation include weakening currents due to an increase in effective transistor gate length (a factor observable in high-level schematic representation), and poorer charging or discharging of node voltages due in part to parasitic elements (a factor that has meaning only in physical layout). Between the two circuit representations per phase, and across the three phases that demonstrate various approaches to extending circuit functionality, this experiment provided a reasonable cross-section of useful data. These results can be used to estimate the performance of devices that exhibit similar construction or function, even if they have not yet been built by the designer.

Such an experiment also presents the opportunity for one to self-critique their use of a new skill set. Given the chance to revisit the project, the author would approach the analysis much differently: for example, Cadence tools make design iteration quite effortless in that components or whole cellviews can be easily copied and modified. A refocused analysis would take advantage of this fact to create many proposed solutions for a given problem and perform cross-strategy parametric studies at every phase – a task that seemed daunting at the outset of the experiment, and now seems reasonable to achieve (and rather, the most complete approach!).

With this increase in analytical power, interesting extensions could be more easily achieved: enlarging the array dimensions further, exploring non-square arrays, optimizing the control circuitry layout, solving the bit line problem, and implementation of the SRAM “read” function.

Reference

- [1] C. Wang, “Memory Circuits: Random Access Memory”. Course lecture (2023): COEN 314, Concordia University, Montreal, QC, Canada.