

Asta Skirbekk

Design of SRAM for Sub-100mV Operation Using 22 nm FD-SOI

Master's thesis in Electronic Systems Design and Innovation

Supervisor: Snorre Aunet

Co-supervisor: Trond Ytterdal

June 2023

Asta Skirbekk

Design of SRAM for Sub-100mV Operation Using 22 nm FD-SOI

Master's thesis in Electronic Systems Design and Innovation
Supervisor: Snorre Aunet
Co-supervisor: Trond Ytterdal
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems



Preface

This master thesis has been performed at the Department of Electronic Systems at the Norwegian University of Science and Technology (NTNU) in Trondheim in the spring of 2023.

The objective of this project has been to create a custom Static Random Access Memory circuit (SRAM) using 22 nm FD-SOI transistors, and to study how the minimum supply voltage needed for proper operation can be minimised.

I would like to thank my supervisor, professor Snorre Aunet at NTNU, for his useful feedback and support throughout my work, and my co-supervisor, professor Trond Ytterdal at NTNU, for his support throughout my work.

Trondheim, June 26th 2023



Asta Skirbekk

Sammendrag

Energihøsting er en lovende løsning for tingenes internett (IoT), siden dette fjerner behovet for hyppig bytting av batterier. Mange energihøstingsmetoder strever med å lage høye forsyningsspenninger, og dette er et problem for minnekretsene på chipen siden disse ofte er flyktige og derfor er avhengig av et stabilt høyt spenningsnivå til enhver tid for å klare å holde på informasjonen som er lagret. For å kunne benytte energihøsting er man derfor avhengig av å designe minnekretser som fungerer på ultralave spenningsnivåer.

Målet med denne masteroppgaven har vært å bruke en 22 nm FD-SOI (Fully Depleted Silicon On Insulator) transistorteknologi til å lage en statisk RAM (SRAM, Static Random Access Memory) for bruk ved spenninger under 100 mV, og å studere hvordan prosessvariasjoner og lokale transistorvariasjoner påvirker hva den minste mulige spenningsforsyningen for minnekretsen er. For å få til dette må man designe og studere alle delkretser i minnet nøye, og det har derfor utgjort en stor del av dette prosjektet. Siden målet har vært å minimere forsyningsspenningen har dette blitt gjort selv når det medfører økt chip-areal og/eller en økning i effektforbruk sammenlignet med andre SRAMer. SRAMen ble designet for å kunne fungere ved temperaturer fra 0°C til 50°C, siden dette gjør at den kan brukes i de fleste innendørs applikasjoner så vel som i medisinske applikasjoner.

Fysiske utlegg har blitt laget for en 4Byte SRAM, en 16Byte SRAM, og en 64Byte SRAM, og for alle SRAMens delkretser, for å kunne generere mer pålitelige og nøyaktige simuleringresultater. De tre SRAM-utleggene krevde alle en forsyningsspenning på minimum 85 mV for å fungere for alle prosess- og temperaturvariasjonene som ble testet. Alle simuleringer som ble gjort etter å ha lagt utlegg viste at kretsene fungerte dårligst i SF-hjørnet, og det ble konkludert med at en bedre balansering av styrken til PMOS og NMOS transistorene vil føre til en betydelig forbedring i dette hjørnet. Dette kan gjøres ved å endre litt på strategien for å bestemme transistorstørrelser og ved å bytte fra å bruke flettede (merged) transistorer til å bruke uflettede (non-merged) transistorer i utlegget.

Monte Carlo simuleringer av lokale transistorvariasjoner ble kjørt for både 4B SRAMen og 16B SRAMen, og det ble observert at disse hadde gode yield etter utlegg for en forsyningsspenning på 80 mV med $yield_{4BSRAM,80mV} = 97\%$ og $yield_{16BSRAM,80mV} = 97.6\%$. Den gode ytelsen observert for SRAMens delkretser tyder på at yelden vil forbli høy også for større SRAMer.

Abstract

Energy harvesting is a promising solution for Internet of Things (IoT) devices, as this removes the need for frequent changes of batteries. Many energy harvesting solutions struggle to supply a high voltage, and this provides a problem for on-chip memory which is often volatile and therefore requires a reliable power supply at all times. On-chip memory must therefore be designed to work at ultra low supply voltages.

The objective of this master's thesis has been to use a 22 nm FD-SOI (Fully Depleted Silicon On Insulator) transistor technology to create a custom Static Random Access Memory (SRAM) for sub-100mV operation and to study how the minimum supply voltage is affected by process variation and transistor mismatch. To achieve this one must also carefully design and study the SRAM's subcircuits, and this has therefore been a major part of this project. As minimising the supply voltage has been the aim, this has been done even when it is at the cost of higher power consumption and/or an increased chip area compared to SRAM circuits operating at higher supply voltages. The SRAM was designed to operate at temperatures in the range 0°C to 50°C, as this would allow it to be used in most indoor applications as well as in medical applications.

Physical layouts were created for a 4B SRAM, 16B SRAM, and a 64B SRAM, as well as for all the SRAM's subcircuits, to get more reliable and accurate simulation results. The three SRAM layouts were found to operate at a minimum supply voltage of 85 mV when process and temperature variations were considered. The SF corner had the worst post layout performance for all circuits, and it was concluded that better balancing of the PMOS and NMOS transistors would improve the performance in this corner considerably. This improvement can be done by changing the transistor sizing strategy slightly as well as switching from merged to non-merged transistors in the layout.

Monte Carlo simulations of transistor mismatch were run on the 4B SRAM and the 16B SRAM, and good post layout yields were achieved for a supply voltage of 80 mV with $yield_{4BSRAM,80mV} = 97\%$ and $yield_{16BSRAM,80mV} = 97.6\%$. The performance of the SRAM's subcircuits indicate that the yield will remain high for larger SRAM circuits as well.

Contents

List of Figures	1
List of Tables	7
List of Abbreviations	8
1 Introduction	9
1.1 Thesis Outline	10
2 Theory	11
2.1 FD-SOI Transistor Technology	11
2.1.1 Single Well	12
2.2 Ultra Low Voltage SRAM	12
2.3 MOSFET Operation in Subthreshold Region	13
2.3.1 Operation of Simple NOT gate in the Subthreshold Region	15
2.4 Schmitt Trigger Logic Gates	16
2.4.1 Principles of a ST Logic Gate	16
2.4.2 Transistor Sizes in ST NOT Circuit	18
2.4.3 Creating Other ST Logic Gates	20
2.5 Transistor Scaling	21
2.5.1 Transistor Scaling and Mismatch Variation	21
2.5.2 Transistor Scaling and Layout Considerations	22
3 SRAM Architecture	23
3.1 SRAM	23
3.2 Bitcell - D Latch	25
3.3 Decoder	26

3.4	Output Selection Module	28
3.5	NOT and NAND Topology	29
3.6	Logic Levels	29
4	Method	31
4.1	Determining the Type and Size of Transistors	31
4.1.1	FD-SOI Transistor Type	31
4.1.2	Transistor Scaling Method	31
4.1.3	Transistor Widths in the NOT gate	31
4.1.4	Transistor Widths in the NAND gate	33
4.2	Physical Layout	35
4.2.1	Single P-Well and Substrate Contact	35
4.2.2	Transistor Folding and Euler's Path	35
4.2.3	Block Regularity	39
4.2.4	Layout of 8-bit Row	41
4.2.5	SRAM layout	43
4.3	Layout Extraction	44
4.4	Monte Carlo Mismatch Simulations	44
4.5	Simulating Process Variations	44
4.6	NOT Gate Testbench	45
4.7	NAND Gate Testbench	46
4.8	D Latch Testbench	47
4.9	2x4 Decoder Testbench	48
4.10	Output Selection Module Testbench	51
4.11	4B SRAM Testbench	52
4.12	16B SRAM Testbench	54

4.13	64B SRAM Testbench	56
5	Results	59
5.1	NOT Simulation Results	59
5.1.1	Monte Carlo Mismatch Simulation Results for Different Supply Voltages .	59
5.1.2	Process Corner Simulation Results	60
5.2	NAND Simulation Results	61
5.2.1	Monte Carlo Mismatch Simulation Results	61
5.2.2	Process Variation Simulation Results	63
5.3	D Latch Results	65
5.3.1	Monte Carlo Mismatch Simulation Results	65
5.3.2	Process Variation Simulation Results	69
5.4	Decoder Results	70
5.4.1	Monte Carlo Mismatch Simulation Results	71
5.4.2	Process Variation Simulation Results	73
5.5	Output Selection Module Results	76
5.5.1	Monte Carlo Mismatch Simulation Results	77
5.5.2	Process Variation Simulation Results	78
5.6	4B SRAM Simulation Results	80
5.6.1	Monte Carlo Mismatch Simulation Results	80
5.6.2	Process Corner Simulation Results	87
5.7	16B SRAM Simulation Results	91
5.7.1	Monte Carlo Mismatch Simulation Results	91
5.7.2	Process Variation Simulation Results	96
5.8	64B SRAM Simulation Results	100
5.8.1	Results from Simulations of Process Variation	100

6 Discussion	104
6.1 Performance under Process Variation	104
6.2 Effect of Transistor Mismatch on the Minimum Supply Voltage	105
6.3 Pre Layout and Post Layout Differences	106
6.4 Comparison with State of the Art	107
7 Conclusion	108
8 Suggestions for Further Work	109
References	110
Appendices	112
A NAND Results	113
B Results from Simulations on the D Latch	114
C Results from Simulations on the 2to4 Decoder	115
D Output Selection Module Results	123
E Results from Simulations on the 4B SRAM	126
F Results from Simulations on the 16B SRAM	136
G Results from Simulations on the 64B SRAM	142
H Layout of the 2to4 Decoder	146
I Layout of the Output Selection Module	147
J Layout of the 4B SRAM	148
K Layout of the 16B SRAM	149
L Layout of the 64B SRAM	150

List of Figures

1	The figure shows the general structure of a traditional bulk CMOS (left) and FD-SOI (right). The figure is based on a figure from chapter 2 (page 11) in [8].	11
2	Illustrations of a NMOS transistor (left) and a PMOS transistor (right), and their respective terminals.	13
3	Schematic of a simple NOT gate. Bulk node ignored for simplicity.	15
4	Schematic of ST inverter.	16
5	Schematic of ST NAND.	21
6	8-bit row of bitcells used as bulding block for SRAM. The input signals <i>RW</i> and <i>Select</i> are used to create the internal signals <i>Read</i> and <i>Write</i> . Only the first (Bit0) and last (Bit7) bitcell are included in the figure.	24
7	4B SRAM.	25
8	A D latch implemented using NOT and NAND gates.	26
9	A 2-to-4 decoder with enable.	26
10	4-to-16 decoder implemented using 2-to-4 decoders. The decoder can be enabled/disabled by using the <i>EN</i> -input.	27
11	Output selection module used to set the correct output from the SRAM based on the outputs from the 8-bit rows.	28
12	Illustration of the legal ranges for logic high and logic low values, and the illegal middle range where the logic value is undefined.	30
13	Voltage transfer curves (VTCs) for different configurations of the NOT gate, both pre layout (blue lines) and post layout (red lines).	32
14	Voltage transfer curves (VTCs) for different configurations of the NAND gate, both pre layout (blue lines) and post layout (red lines).	34
15	Left: Schematic illustration of ST NOT where all transistors have been replaced with unit transistors ($W = W_{base}$). The equivalent widths are the same as for <i>Layout (final)</i> in Table 2. Right: Illustration of the two Euler paths.	36

16	Layout of the ST NOT. The pull-up network is the row on top, and the pull-down network the row below. The three first gate polys in each network belong to the outer transistor ($N0 / P0$), the next two belong to the inner transistors ($N1 / P1$) and the final gate polys belong to the feedback transistors ($N1 / P2$). $Area = 3.0251\mu m^2$ (with $h = 1.79 \mu m$ and $w = 1.69 \mu m$). The dummy polys and substrate contact are not included here, as this layout is used as a building block for larger layouts. To use this NOT gate on its own, a substrate contact must be connected to the VSS -rail and dummy polys added at the ends.	37
17	Left: Schematic illustration of ST NAND where all transistors have been replaced with unit transistors ($W = W_{base}$). The equivalent widths are the same as for <i>Layout (final)</i> in Table 3. Right: Illustration of the two Euler paths.	38
18	Layout of the ST NAND gate. The pull-up network (PMOS) is the row on top, and the pull-down network (NMOS) the row below. $Area = 6.981\mu m^2$ ($h = 1.79 \mu m$ and $w = 3.9 \mu m$). The dummy polys and substrate contact are not included here, as this layout is used as a building block for larger layouts. To use this NOT gate on its own, a substrate contact must be connected to the VSS -rail and dummy polys added at the ends.	39
19	Layout of the D-latch. The NOT gates are outlined in yellow, and the NAND gates are surrounded by a green outline. $Area = 40.94766\mu m^2$ (with $h = 3.51 \mu m$ and $w = 11.666 \mu m$). The dummy polys and substrate contact are not included here, as this layout is used as a building block for larger layouts. To use this NOT gate on its own, a substrate contact must be connected to the VSS -rail and dummy polys added at the ends.	41
20	Layout of the 8-bit row. The D latches that store bit 0 and bit 1 are outlined in yellow. The output NAND gate is outlined in grey for each of these two D latches. The row on top (NOT, NAND, NOT, NAND, NOT) are the logic used to create <i>Read</i> and <i>Write</i> . $Area = 444.06978\mu m^2$ (with $h = 29.085 \mu m$ and $w = 15.268 \mu m$). The dummy polys and substrate contact are not included here, as this layout is used as a building block for larger layouts. To use this NOT gate on its own, a substrate contact must be connected to the VSS -rail and dummy polys added at the ends.	42
21	Layout of the 4B SRAM. The four 8-bit rows (or rather: columns) are outlined with yellow. A Decoder is placed in the top left corner. Eight Output Selection Modules are placed in a column to the right, and the first two are outlined in green. $Area = 1961.33244\mu m^2$ (with $h = 29.085 \mu m$ and $w = 76.616 \mu m$).	43
22	NOT gate testbench schematic. <i>DUT</i> , the third NOT gate from the left, is the device under test.	45
23	NAND gate testbench schematic. Two NOT gates are used as load for the <i>DUT</i> . Two NOT gates in series are used to drive each of the inputs. See Table 6 for combinations of input stimuli A and B and expected output Y	46

24	Timing diagram for the NAND. The transient analysis lasts 200 μ s.	47
25	D Latch testbench. The DUT drives a similar load as what it will do when part of the SRAM. The timing diagram shows inputs, expected intermediate values and expected output for the transient analysis.	48
26	Testbench for the 2x4 decoder. The DUT is driving the same load as it would be if connected to an 8 bit row of bit-cells in the SRAM. Another decoder, <i>drive</i> , is used to create the input to the DUT's <i>EN</i> -pin.	49
27	Timing diagram showing how the input signals <i>EN</i> , <i>I0</i> , <i>I1</i> , <i>I2</i> , and <i>I3</i> are varied during a transient analysis on the decoder testbench in Figure 26. Expected values for the intermediate signal <i>Sel</i> is shown, as well as expected values for the output signals <i>Out0</i> , <i>Out1</i> , <i>Out2</i> , and <i>Out3</i> . <i>RW</i> = <i>VDD</i> . The analysis lasts 8 periods, where each period is marked by a dotted vertical line.	50
28	Block diagram showing the testbench for the output selection module.	51
29	Timing diagram showing the stimuli applied to the testbench in Figure 28 during the transient analysis. Expected values of the intermediate signals <i>A1</i> , <i>B1</i> , <i>C1</i> , and <i>D1</i> are also shown, as well as expected values for the DUT's output signal <i>Y1</i> and the load's output signal <i>Y2</i> . <i>A</i> , <i>B</i> , <i>C</i> , <i>D</i> , <i>AA</i> , <i>CC</i> , and <i>DD</i> are kept at the constant values given in the list to the right. The transient analysis lasts for ten 120 μ s periods (the end of each period is marked by a vertical dotted line).	52
30	Block diagram of the 4B SRAM testbench, which consists only of the 4B SRAM. Input signals are applied as shown in Figure 31, and the values of the eight output signals are sampled.	53
31	Timing diagram showing how the input signals <i>Sel</i> , <i>RW</i> , <i>Addr</i> [1 : 0], and <i>Data</i> [7 : 0] vary during a transient analysis of the 4B SRAM. The expected output values, <i>Out</i> [7 : 0], are also shown. The transient analysis lasts 1 ms. Each period of 100 μ s is marked by a dotted vertical line.	53
32	Block diagram of the 16B SRAM testbench, which consists only of the 16B SRAM. Input signals are applied as shown in Figure 33, and the values of the eight output signals are sampled.	55
33	Timing diagram showing how the input signals <i>Sel</i> , <i>RW</i> , <i>Addr</i> [3 : 0], and <i>Data</i> [7 : 0] vary during a transient analysis of the 16B SRAM. The expected output values, <i>Out</i> [7 : 0], are also shown. The transient analysis is divided into ten periods, where each period is marked by a dotted vertical line.	55
34	Block diagram of the 64B SRAM testbench, which consists only of the 64B SRAM cell. Input signals are applied as shown in Figure 35, and the values of the eight output signals are sampled.	57

35	Timing diagram showing how the input signals Sel , RW , $Addr[5 : 0]$, and $Data[7 : 0]$ vary during a transient analysis of the 64B SRAM. The expected output values, $Out[7 : 0]$, are also shown. The transient analysis is divided into ten periods, where each period is marked by a dotted vertical line.	57
36	The NAND's output Y plotted for all pre layout corners.	64
37	The NAND's output Y plotted for all post layout corners.	65
38	The D latch's output signal Out plotted for the 1000 Monte Carlo points simulated with $VDD = 70$ mV.	66
39	The D latch's output signal Out plotted for the 1000 Monte Carlo points simulated with $VDD = 65$ mV.	66
40	Plot of the DUT 's output Out for all Monte Carlo points that failed the pre layout simulation with $VDD = 65$ mV.	67
41	The D latch's output signal Out plotted for the 1000 Monte Carlo points simulated with $VDD = 80$ mV.	68
42	Plot of the DUT 's output Out for all Monte Carlo points that failed the post layout simulation with $VDD = 80$ mV.	68
43	The pre layout D latch process variation results for $VDD = 70$ mV. FS50 has the lowest logic high value, and SF50 the highest logic low value.	69
44	The post layout D latch process variation results for $VDD = 75$ mV. SF50 and SF27 fail.	70
45	The post layout D latch's output signal $Out0$ plotted for all process corners at $VDD = 80$ mV. All corners pass.	70
46	Each of the DUT 's output signals plotted for the failing post layout simulation points with $VDD = 80$ mV.	72
47	A plot of $Out3$ for the only failing point in the post layout Monte Carlo mismatch simulations on the decoder when $VDD = 85$ mV. The logic high value is marginally below $V_{H,min} = 63.75$ mV.	73
48	The DUT 's output signal $Out0$ plotted for pre layout process corners at $VDD = 70$ mV.	74
49	The DUT 's output signals plotted for all pre layout corners at $VDD = 75$ mV.	75
50	The DUT 's output signals plotted for all post layout corners at $VDD = 85$ mV.	76

51	Output Selection Module. $Y1$ and $Y2$ are plotted for the failing pre layout Monte Carlo simulation points at $VDD = 80$ mV. Both signals fail for simulation points 362 and 587.	77
52	Output Selection Module. $Y1$ and $Y2$ are plotted for the failing post layout Monte Carlo simulation points at $VDD = 80$ mV. All failing points are unique. . .	78
53	Plots of $Y1$ and $Y2$ for all pre layout process corners simulated with $VDD = 75$ mV. The logic high values produced for FS50 are below $V_{H,min} = 56.25$ mV. . .	79
54	Plots of $Y1$ and $Y2$ for all pre layout process corners simulated with $VDD = 80$ mV. All corners pass.	79
55	Plots of the DUT 's output $Y1$ for all post layout process corners simulated with $VDD = 80$ mV and $VDD = 85$ mV.	80
56	The 4B SRAM's output signals plotted for the pre layout simulation points failing when $VDD = 70$ mV. $Out0$, $Out1$, and $Out2$ should have a square pulse in the seventh period, and be logic low otherwise. $Out5$, $Out6$, and $Out7$ should be logic low until they go high in the last period.	82
57	The 4B SRAM's output signals $Out0-Out3$ plotted for the failing post layout simulation points with $VDD = 80$ mV. The signals should have a square pulse in the seventh period, between $t = 600$ μ s and $t = 700$ μ s.	84
58	The 4B SRAM's output signals $Out4-Out7$ plotted for the failing post layout simulation points with $VDD = 80$ mV. The signals should have a square pulse in the last period, starting at $t = 900$ μ s.	86
59	The 4B SRAM's output signals plotted for all pre layout process corners simulated with $VDD = 75$ mV. All corners pass.	88
60	The 4B SRAM's output signals plotted for all post layout process corners simulated with $VDD = 85$ mV. All corners pass.	90
61	The 16B SRAM's output signals plotted for the 1000 simulated points in the pre layout Monte Carlo mismatch simulation with $VDD = 75$ mV. All simulated points pass.	92
62	The 16B SRAM's output signal $Out0-Out2$ plotted for the failing post layout simulation points with $VDD = 80$ mV. The signals should have a square pulse in the seventh period, between $t = 900$ μ s and $t = 1050$ μ s.	93
63	The 16B SRAM's $Out4$ plotted for the failing post layout simulation points with $VDD = 80$ mV.	95
64	The 16B SRAM's output signals plotted for all pre layout process corners simulated with $VDD = 80$ mV. All corners pass.	97

65	The 16B SRAM's output signals plotted for all post layout process corners simulated with $VDD = 85$ mV. All corners pass.	99
66	The 64B SRAM's output signals plotted for all pre layout process corners simulated with $VDD = 80$ mV. All corners pass.	101
67	The 64B SRAM's output signals plotted for all post layout process corners simulated with $VDD = 85$ mV. All corners pass.	103

List of Tables

1	Truth table for the 2-to-4 decoder.	27
2	Transistor widths for the size configurations tested in the NOT gate, both pre and post layout.	33
3	The widths of the transistors in the different NAND configurations tested (both pre and post layout). All widths are a multiple of the base width $W_{base} = 320$ nm. $W_{N0A,B} = W_{N0A} = W_{N0B}$, $W_{N1A,B} = W_{N1A} = W_{N1B}$, $W_{P0A,B} = W_{P0A} = W_{P0B}$, and $W_{P1A,B} = W_{P1A} = W_{P1B}$	35
4	The combination of process corner and temperature tested when simulating process variation.	45
5	The upper limit for logic low ($V_{L,max}$) and lower limit for logic high ($V_{H,min}$) for different supply voltages (VDD).	46
6	Truth table for NAND gate, showing expected output value Y for all combinations of inputs A and B	47
7	Post and pre layout yield for the NOT gate at different supply voltages, based on logic low values observed at the input of the <i>DUT</i> (signal <i>In</i>).	60
8	Post and pre layout yield for the NOT gate at different supply voltages, based on logic high values observed at the output of the <i>DUT</i> (signal <i>Out</i>).	60
9	The operating point value for the <i>DUT</i> 's input signal <i>In</i> and output signal <i>Out</i> for all pre layout process corners simulated at $VDD = 70$ mV. The NOT gate passes for all corners.	61
10	The operating point value for the <i>DUT</i> 's input signal <i>In</i> and output signal <i>Out</i> for all post layout process corners simulated at $VDD = 70$ mV. The NOT gate passes for all corners.	62
11	Results from the pre layout Monte Carlo mismatch simulations on the NAND for both $VDD = 70$ mV and $VDD = 75$ mV. The yield, maximum value of Y , and minimum value of Y are given for each input combination.	62
12	Results from the post layout Monte Carlo mismatch simulations on the NAND for both $VDD = 70$ mV and $VDD = 75$ mV. The yield, maximum value of Y , and minimum value of Y are given for each input combination.	63

List of Abbreviations

B	Byte. 1B = 8 bits.
DIBL	Drain-Induced Barrier-Lowering
DUT	Device Under Test
FD-SOI	Fully Depleted Silicon on Insulator
hvt	High Threshold Voltage (FD-SOI transistor type)
LOD	Length Of Diffusion
lvt	Low Threshold Voltage (FD-SOI transistor type)
RDF	Random Dopant Fluctuations
rvt	Regular Threshold Voltage (FD-SOI transistor type)
slvt	Super Low Threshold Voltage (FD-SOI transistor type)
SRAM	Static Random Access Memory
ST	Schmitt Trigger
uhvt	Ultra High Threshold Voltage (FD-SOI transistor type)
ULV	Ultra-Low Voltage
VDD	Supply voltage
V_{H,min}	Lower Limit for a Logic High Value
V_{L,max}	Upper Limit for a Logic Low Value
VSS	Ground
V_T	Thermal voltage, $V_T = \frac{kT}{q}$
V_{th}	Threshold voltage
WPE	Well Proximity Effect

1 Introduction

The Internet of Things has been growing rapidly, and by the end of 2021 there were 12.2 billion connected IoT devices. This number is expected to grow, and it is estimated that the number of connected IoT devices will reach 27 billion by 2025 [1]. A major factor in limiting this growth is that most IoT devices are powered by batteries. Even a ten year battery lifetime would mean that batteries would have to be changed in millions of IoT devices every day [2]. This would obviously be a problem. In addition to being time consuming, changing the batteries might also be difficult to do depending on where the IoT device has been placed.

In order to try to solve this battery problem, there has been an increased focus on energy harvesting in IoT devices. This way IoT devices can be self-powered, omitting the need for people to change or recharge their batteries. It is, however, important that the IoT device still works as expected. IoT devices require an on-chip memory, such as a Static Random Access Memory (SRAM), to enable key functionalities such as storing instructions and sensor/measurement data [3]. For IoT devices using energy harvesting, supplying a large enough voltage for the memory to operate and retain data might be difficult. The benefit of minimising the supply voltage needed for a memory will in these cases be greater than the cost of an increase in chip area, leakage, and/or energy per operation [4].

Additionally, the SRAM is a volatile memory, which means that it will only be able to retain data for as long as the power supply is on. As a result, on-chip SRAMs require a large proportion of the devices' energy budget, especially when the rest of the device is in sleep mode [5] [6]. In some situations having a sleep mode at all might be difficult when the SRAM must remain on. Minimising the SRAM's minimum supply voltage would then allow the other parts of the chip to operate at a lower supply voltage as well, and thus reduce the power consumption in the rest of the chip.

The main objective of this master's thesis has been to create a custom SRAM for sub-100mV operation using 22 nm FD-SOI transistor technology and to study how the minimum supply voltage is affected by process variation and transistor mismatch. The SRAM should be able to operate at temperatures in the range 0°C to 50°C, as this would allow it to be used in most indoor and medical applications. Designing a robust SRAM circuit so that the supply voltage can be decreased as much as possible has been the main focus, even when this is at the cost of increased area and/or a higher power consumption compared to SRAM circuits operating at higher supply voltages. In doing so one also has to carefully design and study the SRAM's subcircuits, and this has therefore been a major part of this thesis.

To verify the functionality of the SRAM if it were to actually be manufactured, a physical layout has been constructed for the SRAM and all its subcircuits. These layouts have then been extracted with parasitics. Transistor mismatch and process variation was simulated on these extracted layouts. Simulations were also run on the pre layout design to check that the layout methodology did not cause any large or unexpected changes in performance.

1.1 Thesis Outline

Relevant background theory is introduced in Section 2. The SRAM circuit architecture is presented in Section 3, together with the architecture of all its subcircuits. The sizing strategy for the transistors in the NOT and NAND gate, the layout methodology, and the test environments and simulations run to verify the designs are described in Section 4. Results from these simulations are given in Section 5, and discussed in Section 6. Concluding remarks are made in Section 7, before suggestions for further work are given in Section 8.

2 Theory

This section introduces relevant background theory. The Fully Depleted Silicon on Insulator (FD-SOI) transistor technology is presented in Subsection 2.1. An introduction to the Static Random Access Memory (SRAM) and current state of the art in ultra low voltage SRAMs is given in Subsection 2.2. Subsection 2.3 describes the operation of MOSFET transistors and a simple NOT gate in the subthreshold region. The ST logic gate is introduced and described in Subsection 2.4, and Subsection 2.5 presents aspects that affect the choice of absolute value for the widths and lengths of transistors.

2.1 FD-SOI Transistor Technology

The following subsection is copied, with minor modifications and corrections, from [7], which was written by the author at the end of the project work preceding this thesis.

Fully depleted silicon on insulator (FD-SOI) is a planar CMOS technology process. An ultra-thin buried oxide separates the channel from the substrate, allowing the channel to become fully depleted [8].

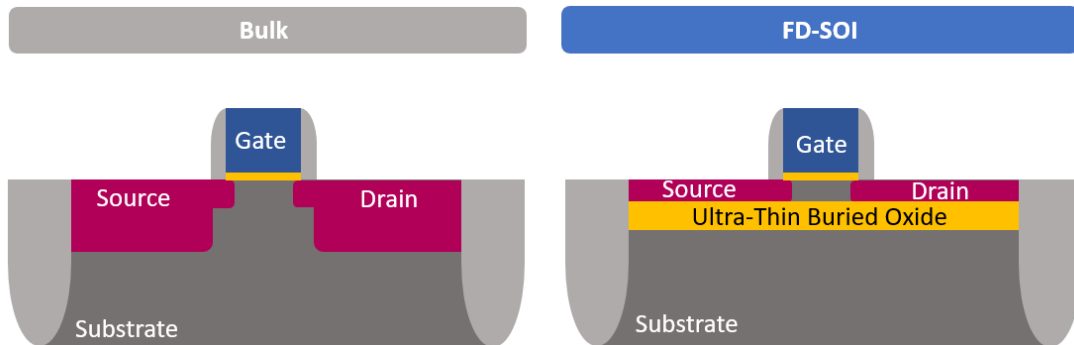


Figure 1: The figure shows the general structure of a traditional bulk CMOS (left) and FD-SOI (right). The figure is based on a figure from chapter 2 (page 11) in [8].

Figure 1 illustrates the difference between traditional bulk CMOS and FD-SOI. An important difference is that FD-SOI does not require channel doping to adjust the threshold voltage. Since random dopant fluctuations in the channel is a common cause for transistor mismatch, using FD-SOI transistors without channel doping greatly reduces the transistor mismatch [9]. Note that the FD-SOI not requiring channel doping is a truth with moderation, as some FD-SOI transistor variants might still require channel doping to achieve the desired threshold voltage. Compared to bulk CMOS, FD-SOI also offers a reduction in drain induced barrier lowering (DIBL) and other short channel effects (SCE) [9].

Body-biasing can be used to control the threshold voltage of an FD-SOI transistor to a large extent, even dynamically to adjust for process variation, which can be a huge advantage [8]. Unfortunately, this requires a supply voltage range larger than what is used in this thesis.

2.1.1 Single Well

The FD-SOI transistor comes in different V_{th} -variants: super-low V_{th} (slvt), low V_{th} (lvt), regular V_{th} (rvt), high V_{th} (hvt), and ultra-high V_{th} (uhvt). The type of well a NMOS or PMOS transistor is placed in varies between the different types. For example is the slvt NMOS placed in a N-well, while the rvt NMOS is placed in a P-well. Similarly, the slvt PMOS is placed in a P-well while the rvt PMOS is placed in a N-well. This opens up the possibility of using a single well by mixing types. One could combine rvt NMOS and slvt PMOS, using a single P-well, or combine slvt NMOS and rvt PMOS and use a single N-well. One should consider the relative driving strengths of the NMOS and PMOS when deciding on which types to combine, as the driving strength is typically increased for the transistors with lower V_{th} .

A benefit of using a single P-well is that the PMOS transistors can be designed with a lower threshold voltage, increasing cell stability and reducing certain ageing effects, while the NMOS transistors can have a higher threshold voltage, thus optimising leakage and static noise margins at high temperatures [8]. Another benefit of having only one well is that the impact of well proximity effects (WPE) on the circuit is greatly reduced, as the distance from a device to the nearest well edge is much larger when there is only one well.

2.2 Ultra Low Voltage SRAM

SRAM (Static Random Access Memory) is a term used to describe a type of memory circuit which is volatile, i.e. retains the data as long as the power is on. It does not need to be refreshed as long as the power is kept on, in contrast to a dynamic memory. The memory is a "random access" memory, which in contrast to a ROM (Read Only Memory) can be written to as well as read from. Both RAMs and ROMs are randomly accessible, i.e. the addresses in the memory can be accessed in any order, so the term "random access memory" is slightly misleading [10].

A typical SRAM uses a latch structure created by a few transistors to store a single bit. The most common type uses 6 transistors (6T bit-cell), but variations with a couple of transistors more or fewer also exist. Creating a SRAM for ultra low supply voltages is difficult, as small variations caused by mismatch and process variation are more likely to affect the behaviour of the circuit as the supply voltage is lowered. Using 22 nm FD-SOI technology, a SRAM based on a 7T bit-cell was in [11] documented to be able to operate at a supply voltage of 300 mV and to retain data at a supply voltage of 240 mV. In [12] a SRAM design based on a 10T bit-cell, which used the more robust 0.13 μm CMOS technology, was found to work for a supply voltage of only 160 mV.

Since the bit-cell is the main building block of a SRAM, creating a more robust bit-cell is the key to improving the minimum supply voltage. One way of doing this is by using logic gates instead

of transistors in the bit-cell, as logic gates can be made very robust so that their functionality are less likely to be affected by process variations and transistor mismatch. This will come at the cost of a notable increase in area, as each bit-cell will need to use a couple of logic gates which each consist of several transistors. When ultra low voltage operation is required, especially when the supply voltage is in the sub-100mV region, the increased robustness is likely to be worth this extra cost in area.

2.3 MOSFET Operation in Subthreshold Region

A MOSFET transistor, as illustrated in Figure 2, is said to operate in the subthreshold region (or weak inversion) when $V_{eff} < 0$ [13], where

$$V_{eff,NMOS} = V_{GS} - V_{thn} \quad (1)$$

and

$$V_{eff,PMOS} = V_{SG} - |V_{thp}|. \quad (2)$$

Here, V_{th} is the transistor's threshold voltage.

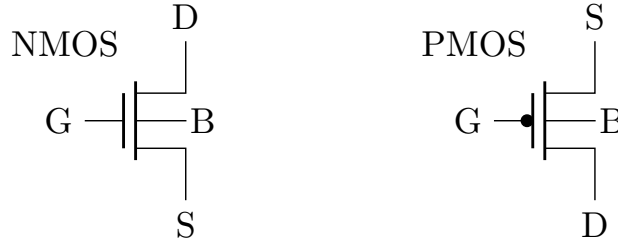


Figure 2: Illustrations of a NMOS transistor (left) and a PMOS transistor (right), and their respective terminals.

The drain current of a MOSFET operating in the sub-threshold region is given by Equation 3,

$$I_{D,sub} = I_0 e^{\frac{V_{GS} - V_{th} - \eta V_{DS}}{nV_T}} (1 - e^{-\frac{V_{DS}}{V_T}}), \quad (3)$$

where V_{GS} is the gate-source voltage, V_{th} is the threshold voltage, η is the drain-induced barrier-lowering (DIBL) coefficient, V_{DS} is the drain-source voltage, n is the subthreshold ideality factor, and $V_T = \frac{kT}{q}$ is the thermal voltage [14]. T is the absolute temperature (measured in Kelvin), k is Boltzmann's constant and q is the elementary charge. The constant I_0 , which can be expressed as

$$I_0 = \frac{W}{L} (n-1) \mu C_{ox} \left(\frac{kT}{q}\right)^2, \quad (4)$$

represents factors that determine the drive strength of the transistor [13]. The values of C_{ox} , which is the gate capacitance per unit area, and the carrier mobility μ are largely dictated by the chosen transistor technology. Changing the factor $\frac{W}{L}$ is therefore the designers best tool to manipulate the transistor's strength. The subthreshold ideality factor, or subthreshold swing coefficient, n is given by Equation 5 [13],

$$n = \frac{C_{ox} + C_{j0}}{C_{ox}} \approx 1.5. \quad (5)$$

The leakage current $I_{D,leak}$ through a transistor that is supposed to be off can be found by setting $V_{GS} = 0$ in Equation 3. This results in:

$$I_{D,leak} = I_0 e^{\frac{-V_{th} - \eta V_{DS}}{nV_T}} (1 - e^{-\frac{V_{DS}}{V_T}}). \quad (6)$$

when operating in the subthreshold region.

For a digital circuit to operate properly it has to be able to differentiate between a logic high and a logic low value. This requires a distinct difference between the on current I_D and the off (leakage) current $I_{D,leak}$. The ratio between on and off current is found by dividing Equation 3 by Equation 6:

$$RATIO_{on/off} = \frac{I_0 e^{\frac{V_{GS} - V_{th} - \eta V_{DS}}{nV_T}} (1 - e^{-\frac{V_{DS}}{V_T}})}{I_0 e^{\frac{-V_{th} - \eta V_{DS}}{nV_T}} (1 - e^{-\frac{V_{DS}}{V_T}})} = e^{\frac{V_{GS}}{nV_T}}. \quad (7)$$

This can be further simplified to

$$RATIO_{on/off} = e^{\frac{V_{DD}}{nV_T}}, \quad (8)$$

assuming that the voltage applied to the gate of an NMOS is $V_G = V_{DD}$ and that the source node is connected to ground.

From Equation 8 it is evident that a higher supply voltage and/or a lower thermal voltage (i.e. lower temperature) will increase the on/off current ratio. For a logic gate operating in the subthreshold region, increasing the temperature will reduce the robustness of the circuit and possibly lead to the output voltage deviating from the expected logic value. Similarly, decreasing the supply voltage, which is the main focus of this thesis, will result in a smaller difference between on/off currents.

By applying a biasing voltage to the bulk node, marked with a B in Figure 2, the threshold voltage can be modified (this can even be done dynamically, by adjusting the bias in response to other V_{th} -changing effects such as process variation)[8]. As stated in Subsection 2.1, a notable change in V_{th} will require a higher voltage potential than the sub-100mV supply voltage the design in this thesis is limited to. Since the effect of a sub-100mV V_B on the threshold voltage is very limited, other considerations can be prioritised in stead when choosing which voltage potential the bulk should be connected to. If for instance a single well is used, all bulk nodes can be connected to the same voltage potential even if this results in different V_{BS} -values for all transistors in the design.

2.3.1 Operation of Simple NOT gate in the Subthreshold Region

A simple NOT gate, see Figure 3, can be constructed from only two transistors. Ideally, the NMOS should function as a short circuit and the PMOS as an open circuit when the input is high, as this would connect the output node directly to ground. When the input is low, the PMOS should ideally function as a short circuit to connect the output node to VDD while the NMOS should be an open circuit. As explained above, even off transistors will in the subthreshold region conduct a leakage current $I_{D,leak}$. This will lead to a slight deviation of the output value, and limit the minimum supply voltage. The theoretical minimum supply voltage for a chain of NOT gates, like the one in Figure 3, was derived by Swanson and Meindl in 1972 to be

$$VDD_{min} = 2 \ln(2) V_T, \quad (9)$$

assuming ideal process parameters and a maximum steepness of more than -1 for the voltage transfer curve (VTC) [15]. A less steep transfer curve will result in a degradation of the output from each stage in the chain of gates, until the logic value of the output can no longer be determined with certainty. At room temperature (300K, or 27°C), this equates to a minimum supply voltage $VDD_{min} = 35.87$ mV.

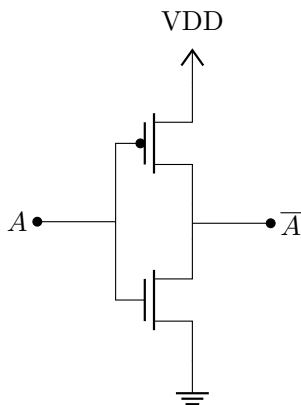


Figure 3: Schematic of a simple NOT gate. Bulk node ignored for simplicity.

In spite of the low theoretical value, this circuit topology has many weaknesses when it comes to ultra low voltage operation. When implementing this circuit on silicon, process variations and transistor mismatch will skew the balance between the PMOS and the NMOS. This will result in the output signal deviating from the logic values. When several NOT gates are connected in series, a small deviation can continue to grow as the next gate receives a deviated input value and therefore produce an even worse output. This can eventually lead to a signal value that is misinterpreted as the opposite logic value. Extra measures must therefore be taken to reduce the output voltage deviation when the supply voltage is decreased far into the subthreshold region. One such strategy that limits the output voltage deviation in spite of the low on/off current ratio is the use of a Schmitt Trigger topology, and this will be described in Subsection 2.4.

2.4 Schmitt Trigger Logic Gates

Lotze and Manoli used a Schmitt Trigger (ST) structure to increase the on/off current ratio and reduce the output level deviation of a logic gate, reporting a minimum supply voltage $VDD_{min} = 62$ mV for an 8x8 multiplier made from ST logic gates [14].

2.4.1 Principles of a ST Logic Gate

To create a Schmitt Trigger NOT gate, each of the transistors in the basic NOT gate in Figure 3 are replaced by two transistors in series. This is illustrated in Figure 4, where the NMOS transistors $N0$ and $N1$ and the PMOS transistors $P0$ and $P1$ are connected in series. The reason for replacing one transistor in the basic NOT gate with two transistors in the ST NOT gate is to create a middle node in both the pull-up and the pull-down networks. These are the nodes marked X and Y in Figure 4.

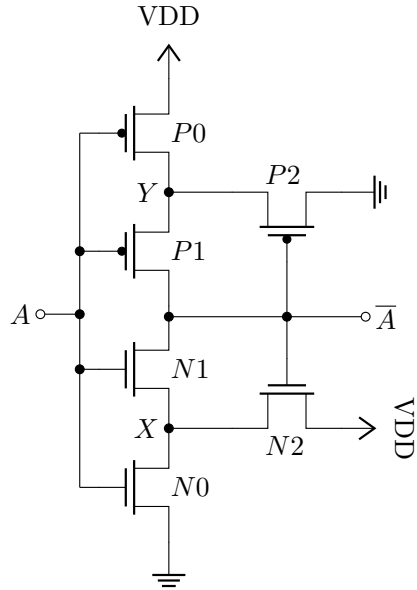


Figure 4: Schematic of ST inverter.

To each of these middle nodes the source node of another transistor is connected, which is used to create a feedback. The feedback transistor $N2$ in the pull-down network has $V_D = VDD$, and the feedback transistor $P2$ in the pull-up network has $V_D = VSS$. The gate nodes of both feedback transistors are connected to the output of the NOT gate (\bar{A} in Figure 4), which means that the voltage potential observed at their gate nodes is the inverse of the voltage potential observed at the gate nodes of the driving transistors ($N0, N1, P0, P1$). As a result, the feedback transistor $N2$ is active when the NMOS pull-down network is off and the feedback transistor

$P2$ is active when the PMOS pull-up network is off [14].

To examine how the feedback transistors work, we can first assume that the value applied to the input A is a logic zero. $P0$ and $P1$ are then on, while $N0$ and $N1$ are off. A logic one value can be observed at the output \bar{A} , but it has slightly deviated from the ideal value due to the leakage current conducted by the off transistors $N0$ and $N1$. This is where the feedback transistor comes in. The logic high value at \bar{A} means that the feedback transistor $N2$ is on, and conducting a current that flows to the middle node X . This current must flow through $N0$, which is off and leaking, in order to reach ground. The current from $N1$ is also flowing through $N0$. This causes a larger voltage drop over $N0$, forcing the voltage potential at X up. An increased V_X means smaller V_{DS} for both $N1$ and $N2$, while V_{DS} for $N0$ is increased to allow for more current flowing through $N0$.

The voltage potential at X will increase until it reaches a value where the leakage current through $N0$ is equal to the current through $N2$ and the leakage current through $N1$. As a result, the drain-source voltage $V_{DS,N1}$ over node $N1$ will be close to zero and the gate-source voltage $V_{GS,N1}$ will be negative (as $V_G = 0$ and $V_X \rightarrow VDD$). As seen from Equation 6, the negative $V_{GS,N1}$ will effectively quench the leakage current through $N1$. Since the leakage current through $N1$ is the leakage current observed at the output of the NOT gate, quenching this current will drastically reduce the output level degradation [14]. It is important to note that this technique does not reduce the overall leakage current in the NOT gate, it only changes the path of the leakage in such a way that it has a smaller impact on the output voltage of the gate [14].

A similar behaviour will occur for the pull-up network when the input A is logic one, which will cause $P0$ and $P1$ to be off and leaking and $P2$ to be on and conducting. The leakage current flowing through $P0$ will essentially meet with a current divider consisting of $P1$ and $P2$. As $P2$ is on and $P1$ is off, nearly all of the current from $P0$ will flow through $P2$. The voltage potential at Y will be forced down so that KCL is fulfilled:

$$I_{P0(off)} = I_{P1(off)} + I_{P2(on)}. \quad (10)$$

This can be used to calculate an approximate value for the voltage potential V_Y at node Y . For simplicity, assume that the three transistors have the same driving strength (I_0 is the same for all three) and that they have the same threshold voltage V_{th} . Using the formula for leakage current (Equation 6) and the formula for subthreshold drain current (Equation 3), assuming $\bar{A} = 0$, and ignoring the DIBL effect (i.e. $\eta = 0$), Equation 10 can be rewritten as

$$\left(1 - e^{-\frac{(VDD - V_Y)}{V_T}}\right) = \left(1 - e^{-\frac{V_Y}{V_T}}\right) + e^{\frac{V_Y - V_{th}}{nV_T}} \left(1 - e^{-\frac{V_Y}{V_T}}\right) e^{\frac{V_{th}}{nV_T}}. \quad (11)$$

Solving Equation 11 for V_Y , we find that

$$e^{-\frac{2V_Y}{V_T}} - e^{\frac{V_Y}{nV_T}} e^{-\frac{V_Y}{V_T}} + e^{\frac{V_Y}{nV_T}} e^{-\frac{2V_Y}{V_T}} = e^{-\frac{VDD}{V_T}}. \quad (12)$$

By substituting $x = e^{-\frac{V_Y}{V_T}}$, we get the following expression:

$$x^2 - x^{-\frac{1}{n}}x + x^{-\frac{1}{n}}x^2 = e^{-\frac{V_{DD}}{V_T}} \quad (13)$$

Further assuming that $n \approx 1.5$, see Equation 5, this can be rewritten as

$$x^2 - x^{\frac{1}{3}} + x^{\frac{4}{3}} = e^{-\frac{V_{DD}}{V_T}}. \quad (14)$$

For this example, the supply voltage is set to $V_{DD} = 100$ mV. Solving Equation 14 for x gives $x = 0.599318$. Substituting $x = e^{-\frac{V_Y}{V_T}}$ and taking the natural logarithm on both sides results in

$$V_Y = -V_T \ln(0.599318) \approx 13.3\text{mV}, \quad (15)$$

using that $V_T \approx 26$ mV at room temperature.

The leakage current flowing through $P1$, which is the leakage current observed at the output node of the NOT gate, is then

$$I_{D,leak,P1} = I_{0,P1} e^{-\frac{V_{th}}{nV_T}} (1 - e^{-\frac{V_Y}{V_T}}) \approx I_{0,P1} e^{-\frac{V_{th}}{nV_T}} \cdot 0.4 \quad (16)$$

assuming $V_T = 26$ mV.

An equivalent NOT gate without feedback transistors would have an equal voltage divide over $P0$ and $P1$ leading to $V_Y = 0.5V_{DD}$ when the pull-up network is off, given the above assumption that all the PMOS transistors are of equal driving strength and have the same V_{th} . The leakage current through $P1$ would then be

$$I_{D,leak,P1} = I_{0,P1} e^{-\frac{V_{th}}{nV_T}} (1 - e^{-\frac{V_Y}{V_T}}) \approx I_{0,P1} e^{-\frac{V_{th}}{nV_T}} \cdot 0.85, \quad (17)$$

for $V_{DD} = 100$ mV and $V_T = 26$ mV. This is approximately 2.1 times the leakage current calculated for the NOT gate with a feedback transistor $P2$, which illustrates how effective the leakage quenching in the ST NOT is.

Note that the actual value of n will depend among other things on the transistor technology used, so $n \approx 1.5$, which was used in Equation 14, might not be a good estimate. The general tendencies observed from the calculations above, i.e. that the ST NOT gate quenches leakage much more effectively, should still hold for other values of n .

2.4.2 Transistor Sizes in ST NOT Circuit

To get a balanced voltage transfer curve for the NOT gate, the strength of the pull-up network should be equal to the strength of the pull-down network. If one of the networks is weaker than the other, this can be compensated for by increasing $\frac{W}{L}$ of the driver transistors ($P0$ and $P1$ for the pull-up network, and $N0$ and $N1$ for the pull-down network) or by decreasing $\frac{W}{L}$ for the drivers in the opposite network. Another way to compensate for a weaker pull-up or

pull-down network is by increasing $\frac{W}{L}$ for the feedback transistor in the opposite network, as this will decrease the current flow through the inner driver transistor ($P1$ or $N1$) in the stronger network and therefore have the same effect as if the driver transistors in the stronger network were made weaker [14].

How effective the leakage quenching in the ST NOT is depends on the relative driving strengths of the different transistors within the same network. The optimal transistor sizing for minimum supply voltage has been discussed thoroughly by Lotze and Manoli in [4]. They use a standard $0.13\ \mu\text{m}$ CMOS technology, but the arguments made should hold for FD-SOI technology as well. In the following explanation, only the size of the NMOS transistors in the pull-down network are considered to make the explanation clearer. The same line of argumentation can be applied to the sizing of the PMOS transistors in the pull-up network.

Firstly, one must consider the relative strength of the feedback transistor $N2$ compared to $N0$. When $N0$ and $N1$ are off, the feedback transistor must be able to pull node X as high as possible so that the leakage through $N1$ is effectively quenched and the output level deviation is limited. The stronger $N2$ is compared to $N0$, the better will it be able to pull the middle node X high.

When $N0$ and $N1$ are on, there should ideally be no voltage drop over them. $N2$ will in this situation be off, but leaking. This leakage current will have to flow through $N0$ and cause a higher voltage drop over this transistor. To limit this additional voltage drop as much as possible, a weak feedback transistor (weak compared to $N0$) is needed. The sizing of the feedback transistor is in other words a trade-off between on-behaviour and off-behaviour of the pull-down network, i.e. between deviation of the logic low level and deviation of the logic high level.

When $N0$ and $N1$ are on, the pull-up network is off and leaking. The voltage drop over $N0$ will therefore be caused by a combination of the leakage current from the pull-up network and leakage current from the feedback transistor $N2$. Assuming that the pull-up and pull-down networks are balanced and the feedback transistors $P2$ and $N2$ have the same drive strength, a stronger feedback transistor would mean less leakage from the pull-up network and more from $N2$ while a weaker feedback transistor would have the opposite effect. Lotze and Manoli used this to find the optimum drive strength ratio for which the voltage drop over the on transistors is minimised. For the nominal case, this ratio was found to be $\frac{N2}{N0} \approx \frac{2}{3}$ [4].

In reality, global variations will affect the threshold voltages of all NMOS and PMOS transistors. This can result in the pull-up and pull-down network having different driving strengths, which means that the weaker network when on will experience a larger leakage current caused by the stronger network compared to the nominal case. To combat this, the feedback transistors must be made stronger so that there is less leakage from the opposite network [4].

Secondly, the driving strength of the outer transistor compared to the inner transistor, i.e. $\frac{N0}{N1}$, must be considered. When the pull-down network is on, leakage current flowing through $P1$ will cause a voltage drop over $N1$ and $N0$ and degrade the output level. Increasing the driving strength of the outer transistors compared to the inner transistors will reduce this voltage drop and minimise the output level degradation caused by this leakage from the opposite network. It was concluded in [4] that the outer transistor should be made as strong as possible compared

to the inner transistor. Based on the results in [4], going from $\frac{N0}{N1} = 1$ to $\frac{N0}{N1} = 2$ gives the same reduction in VDD_{min} as going from $\frac{N0}{N1} = 2$ to $\frac{N0}{N1} = 6$. Although the exact numbers are expected to be different for other transistor technologies, this shows that the continued increase of $\frac{N0}{N1}$ has diminishing returns, and that the most important point is to ensure $\frac{N0}{N1} > 1$.

2.4.3 Creating Other ST Logic Gates

By replacing $N0$, $N1$, $P0$, and $P1$ in the NOT gate in Figure 4 with other transistor configurations, it is possible to design a variety of different ST logic gates. A ST NAND gate with two inputs A and B is made by replacing each NMOS with two transistors in series and each PMOS with two transistors in parallel, as shown in Figure 5.

A good starting point for determining the sizes of a ST NAND is to assume that the two inputs A and B are tied together. This transforms the NAND into a gate that functions like a NOT, since $\overline{1 \cdot 1} = \overline{1} = 0$ and $\overline{0 \cdot 0} = \overline{0} = 1$, and reduces the problem to finding appropriate sizes for transistors in a ST NOT which has been described in Subsection 2.4.2. As each of the drive transistors $P0$ and $P1$ in the NOT are replaced with two transistors in parallel in the NAND, $\frac{W}{L}$ for each transistor should be halved compared to the NOT to maintain the same total driving strength, so that

$$\left(\frac{W}{L}\right)_{P0} = \left(\frac{W}{L}\right)_{P0A} + \left(\frac{W}{L}\right)_{P0B} \quad (18)$$

and

$$\left(\frac{W}{L}\right)_{P1} = \left(\frac{W}{L}\right)_{P1A} + \left(\frac{W}{L}\right)_{P1B}. \quad (19)$$

The value of the feedback transistors can be kept the same. $N0$ and $N1$ in the NOT are each replaced with two transistors in series in the NAND. This is equal to doubling the length L of $N0$ and $N1$. To maintain the same drive strength as in the NOT, $\frac{W}{L}$ must therefore be doubled for each of the NMOS drive transistors $N0A$, $N0B$, $N1A$, and $N1B$.

After creating this starting point, the drive strengths of the transistors in the ST NAND must be adjusted to compensate for the case where $A \neq B$. In this case, half of the pull-up network will be on and the other part off. An increased output level deviation should be expected for these cases, but the circuit must still manage to produce a logic high output, $\overline{0 \cdot 1} = 1$, just as if both halves of the pull-up network was on. To reduce the voltage drop over the pull-up network, $\frac{W}{L}$ can be increased slightly for $P0$ and/or $P1$ (increasing the strength of $P0$ is preferable, to maintain a decent ratio between $P0$ and $P1$). In addition, the pull-down network may be made weaker. As a large $\frac{N0}{N1}$ -ratio is desirable, reducing $\frac{W}{L}$ for $N1$ is the preferred way of weakening the pull-down network. In addition, the strength of the feedback transistor $N2$ can be increased to increase the voltage drop over $N0$ and reduce the leakage through $N1$.

To check that the values chosen give a balanced NAND gate, a voltage transfer curve (VTC) can be created for the transition from input $AB = 01$ to $AB = 11$. The switching point should occur when $A = \frac{VDD}{2}$.

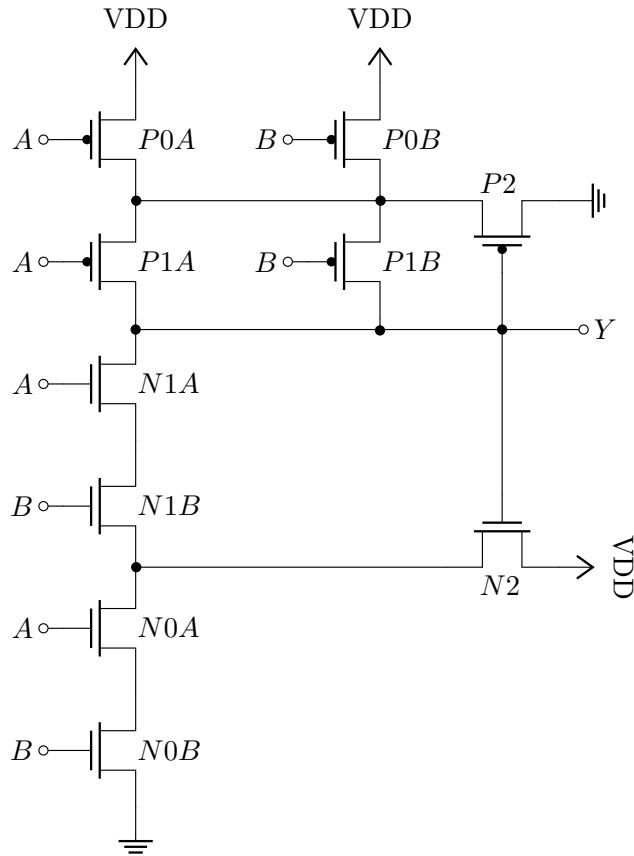


Figure 5: Schematic of ST NAND.

2.5 Transistor Scaling

The previous subsection discussed the relative sizing of the transistors in relation to the other transistors in a NOT and NAND gate. This subsection will present different aspects to consider when choosing the absolute value of the transistors' widths and lengths. First, transistor scaling in relation to mismatch will be explained in Subsection 2.5.1. The impact of transistor dimensions on layout will then be presented briefly in Subsection 2.5.2.

2.5.1 Transistor Scaling and Mismatch Variation

According to [14], random dopant fluctuations (RDF), which impact the threshold voltage of a transistor, are the dominant cause of variations caused by transistor mismatch in ST logic gates. The threshold voltage variance caused by transistor mismatch is inversely proportional to the

area of the transistor, a relation often referred to as Pelgrom's law [16]. As such, increasing the transistor area increases the robustness of the transistors and the circuit it is a part of.

Transistor area is often scaled either by increasing both the gate width and the gate length or by only increasing the gate width. Lotze and Manoli found in [14] that the choice of scaling technique only had a minor impact on the VDD_{min} achieved for ST NAND and ST NOT gates, with the technique scaling both width and length being marginally better. Other aspects can therefore be considered when deciding on how to increase the transistor area.

As seen from Equation 4, the factor $\frac{W}{L}$ is an important part of the constant I_0 , and the factor that the circuit designer has most control over. Increasing the length L will, for a given width W , reduce the leakage current $I_{D,leak}$ (see Equation 6). Increasing L will also reduce the transistor's on current $I_{DS,sub}$, see Equation 3, which will lead to an increased delay through the logic gate [17] [14]. There is, in other words, a trade-off between having a fast logic gate (small L) and reduced leakage (large L). Since the Schmitt Trigger topology effectively reduces the leakage current at the logic gate's output node, which is the leakage causing output level deviation, further reduction of the leakage current is not likely to make a drastic difference. It might therefore be wise to prioritise speed, both for the functionality of the design (at least if it consists of a long chain of gates, as the delay through the circuit might become unreasonable) and as a way to reduce the power consumption which is at its highest during a transition when both the pull-up and pull-down networks are partly on.

2.5.2 Transistor Scaling and Layout Considerations

To make it easier on the lithography and process control it is desirable to have a high degree of transistor regularity, e.g. by using the same length for all transistors and limiting the number of different widths [18]. Increased regularity also makes it easier to stack subcircuits together when creating a modular layout design. This can save time for the person designing the layout, as well as enabling the design of a more compact layout.

3 SRAM Architecture

The memory created in this thesis uses logic gates to create a D latch that stores a single bit, which is different from the memory architectures typically encompassed by the term "SRAM". The underlying principles, such as the static nature of the memory, are the same and the memory circuit designed in this thesis is meant to fulfil the same purpose as a typical SRAM. The definition of the term "SRAM" has therefore been stretched to include the memory architecture designed for this thesis.

The circuit architecture was decided upon in the project preceding this thesis [7]. For the benefit of any reader unfamiliar with the contents of that work, the following subsections will describe the architecture of the asynchronous SRAM and all the subcircuits it consists of. Note that this section is copied in its entirety from [7], with only minor changes and corrections.

An overview of the SRAM's composition is given in Subsection 3.1. The following three subsections describe components that were introduced in Subsection 3.1: Subsection 3.2 describes the implementation of the bitcell, Subsection 3.3 presents the address decoder, and Subsection 3.4 describes the module used to select the correct driver for the output of the SRAM. NOT and NAND gates are the logic gates used to implement all other modules. The NOT and NAND topologies used are presented in Subsection 3.5. Subsection 3.6 defines the upper limit of logic low values and the lower limit for logic high values used in this design.

3.1 SRAM

The SRAM designed in this thesis consists of several 8-bit rows of bitcells, where each bitcell can hold one bit of memory. Each 8-bit row has a *Select* signal as input, which enables reading/writing to this row if it is set high. Each row also takes a signal *RW* as input, which decides whether to read ($RW = 1$) or write ($RW = 0$) when *Select* is high. Each row has 8 data inputs, which are used when writing to the row, and 8 data outputs, which are used when reading from the row. Figure 6 shows the schematic of this SRAM-row.

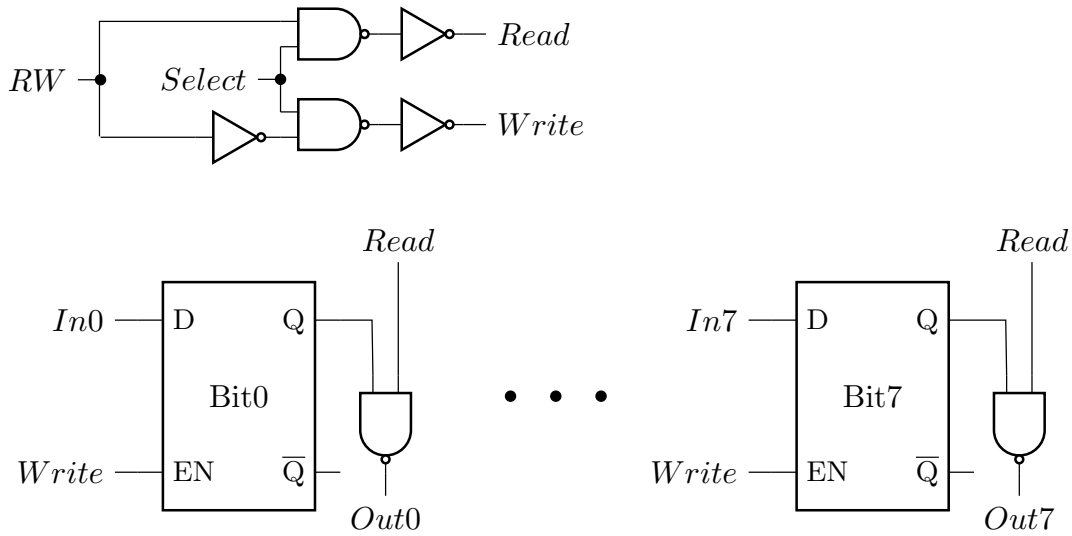


Figure 6: 8-bit row of bitcells used as building block for SRAM. The input signals RW and $Select$ are used to create the internal signals $Read$ and $Write$. Only the first ($Bit0$) and last ($Bit7$) bitcell are included in the figure.

An enable signal, EN , is used to enable reading or writing from the memory. If $EN = 0$, no 8-bit row is selected, which means that nothing is being read or written. If $EN = 1$, a binary address is used to select which 8-bit row to read/write from. This is done using decoders to create the correct configuration of $Select$ -signals, so that only the chosen 8-bit row's $Select$ -signal is set high. A signal RW is then used to choose whether to read or write, where $RW = 1$ equals read and $RW = 0$ means write. Note that the output value will be the inverse of the actual value read from the storage element. This knowledge is used later on when designing the output selection module, so that the output of the actual SRAM is correct (and not the inverse).

A 4 Byte (4B) SRAM configuration is shown in Figure 7. A 16B SRAM is made by adding a new decoder, combining four 4B SRAMs, and adding a new output selection module. The new output selection module must have a NOT gate at each input so that it is given the inverse of what it should output. A 64B SRAM is made from four 16B SRAMs, a 256B SRAM is made from four 64B SRAMs, and a 1024B SRAM is made from four 256B SRAMs.

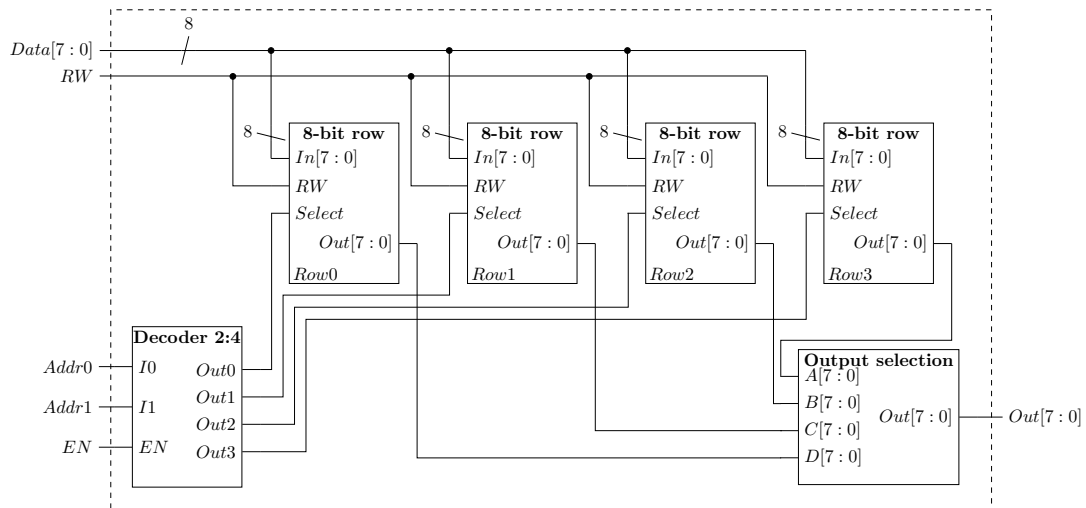


Figure 7: 4B SRAM.

3.2 Bitcell - D Latch

A D latch, see Figure 8, is used as a bitcell for this design. The value on input D will be written to the latch if $EN = 1$. If $EN = 0$, the latch will store the value it is currently holding. This stored value can be read from output Q . The D latch is implemented using NAND and NOT gates.

The two NOT gates connected between the upper right NAND and the output Q in Figure 8 function as buffers for the value outputted by the NAND. The same goes for the two NOT gates between the lower right NAND and the output \bar{Q} . As Q and \bar{Q} are fed back to the NAND gates on the right, a slightly bad logic value on Q or \bar{Q} will create an even worse logic value which again will be fed back and create a worse logic value etc. Buffering these values stops this negative cycle from occurring, thus contributing to the robustness of the D latch at lower supply voltages.

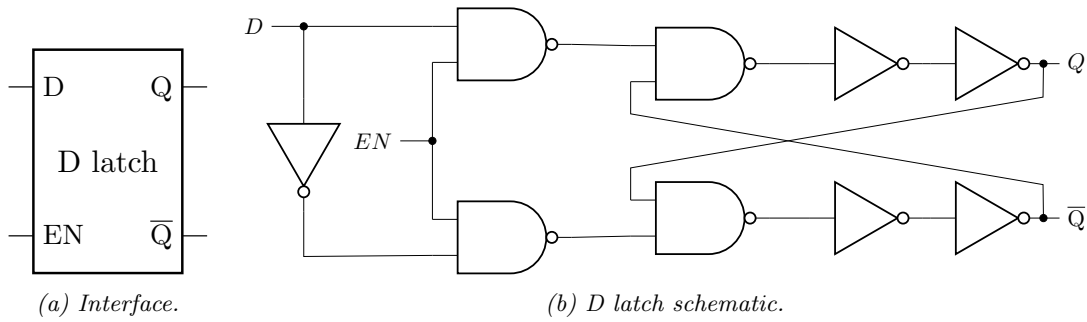


Figure 8: A D latch implemented using NOT and NAND gates.

3.3 Decoder

Decoders are used to select the correct 8-bit row based on the binary input address that the SRAM receives. The decoder implemented for this thesis is a 2-to-4 bit decoder with an enable bit. The interface is shown in Figure 9a. Figure 9b shows how the 2-to-4 decoder was implemented using NAND and NOT gates. The outputs created for different combinations of inputs are summarised in Table 1.

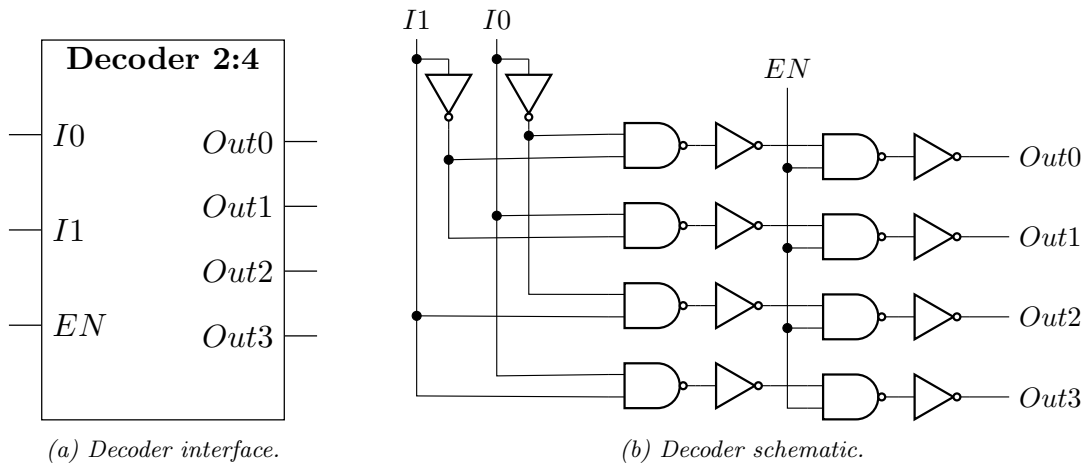


Figure 9: A 2-to-4 decoder with enable.

By connecting multiple 2-to-4 decoders together it is easy to create a decoder of the desired size. This is useful as it enables the creation of SRAMs of many different sizes. An example of how a 4-to-16 decoder can be made from five 2-to-4 decoders is shown in Figure 10.

Table 1: Truth table for the 2-to-4 decoder.

EN	$I1$	$I0$	$Out3$	$Out2$	$Out1$	$Out0$
1	0	0	0	0	0	1
1	0	1	0	0	1	0
1	1	0	0	1	0	0
1	1	1	1	0	0	0
0	X	X	0	0	0	0

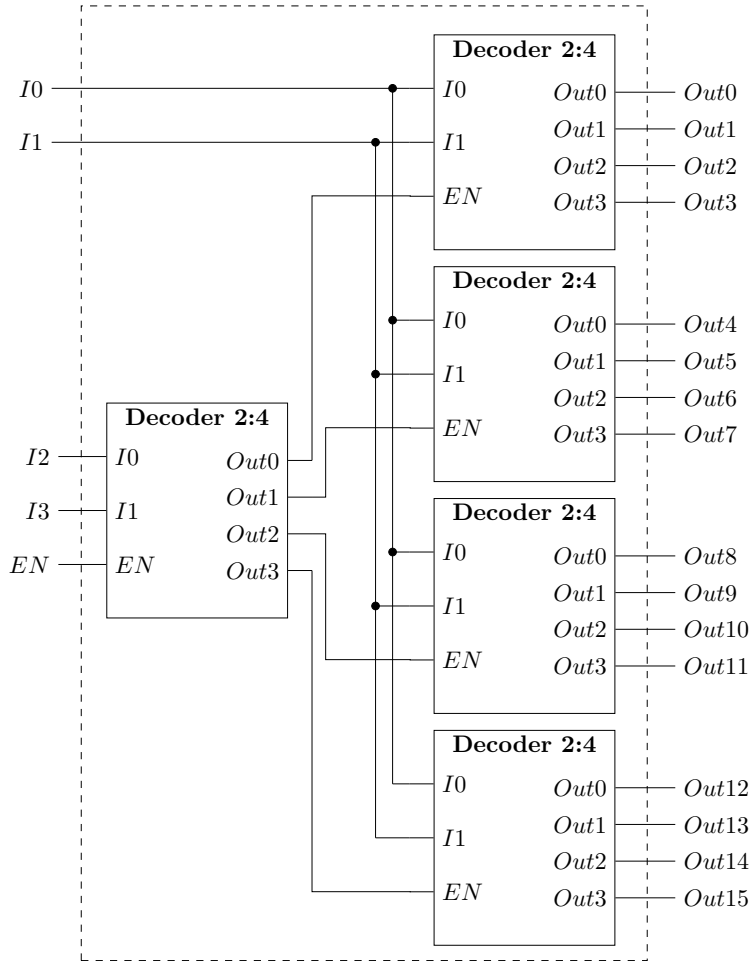


Figure 10: 4-to-16 decoder implemented using 2-to-4 decoders. The decoder can be enabled/disabled by using the EN -input.

3.4 Output Selection Module

Only one bitcell is supposed to drive each of the SRAM's output signals at any given time, but the inactive bitcells might contribute with some noise. This is most likely not a problem for a normal supply voltage, as the noise contribution is small compared to the signal value. For operation in the sub-100 mV this might constitute a problem, as noisy bitcells are more likely to occur and the small difference between logic low and logic high makes even small noise signals problematic. A module has therefore been created which takes the outputs of four bitcells driving the same line, and uses combinatorial logic to determine which logic value to use as output. This way each of the SRAM's output signals only has a single driver. The module has been given the name *Output selection module*.

If no row in the SRAM has been selected for a read, the SRAM outputs a logic low value. The user of the SRAM is expected to know that this is an invalid output (and not a value read from memory), based on the values of the signals going in to the SRAM which the user is in control of. When a row has been selected for read, the values from the bitcells in this row should be visible at the SRAM's 8 output lines. In order to avoid multiple drivers and unnecessary noise on the output lines while still outputting the correct value, a simple output selection module is made from NAND and NOT gates.

Only one row is ever selected at a time, meaning that only the outputs from a selected row have the potential to be logic low (which happens if the row is enabled and the stored bit is logic high). Figure 11 shows the output selection module, which will give a logic high at the output if one of the inputs is logic low and a logic low if all inputs are logic high. The output selection module will, in other words, use the fact that the output from the bitcell's output NAND is the inverse of what should be outputted to the user. Had the output from the bitcell's NAND not been inverse, an additional NOT would be needed at all the output selection module's inputs. This has to be done when connecting several output selection modules in series, as the output from each module (which will be the input of the next) is not the inverse.

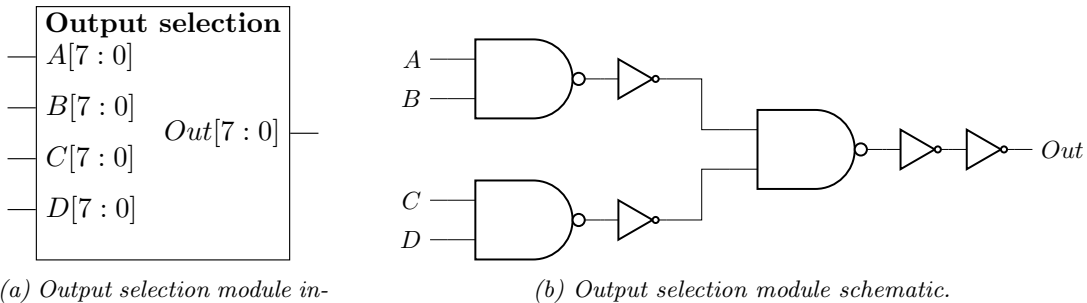


Figure 11: Output selection module used to set the correct output from the SRAM based on the outputs from the 8-bit rows.

3.5 NOT and NAND Topology

The ST NOT gate topology presented in Figure 4 in Subsection 2.4 is used for this design, as it reduces the output level deviation of the inverter compared to the traditional inverter-design. The NAND gate is implemented using the ST topology shown in Figure 5 in Subsection 2.4.3.

3.6 Logic Levels

A struggle when designing logic gates for ultra low supply voltage is ensuring that the output values are unambiguous, i.e. that they are either a clear logic low or a clear logic high. An ambiguous value might be misinterpreted by the next logic gate, leading to the design having unreliable functionality. It is therefore common to choose a range of voltage values that will be accepted as logic low, and a range that will be accepted as logic high. Any value that falls between these two ranges is considered ambiguous, and might therefore be interpreted either way. This is illustrated in Figure 12.

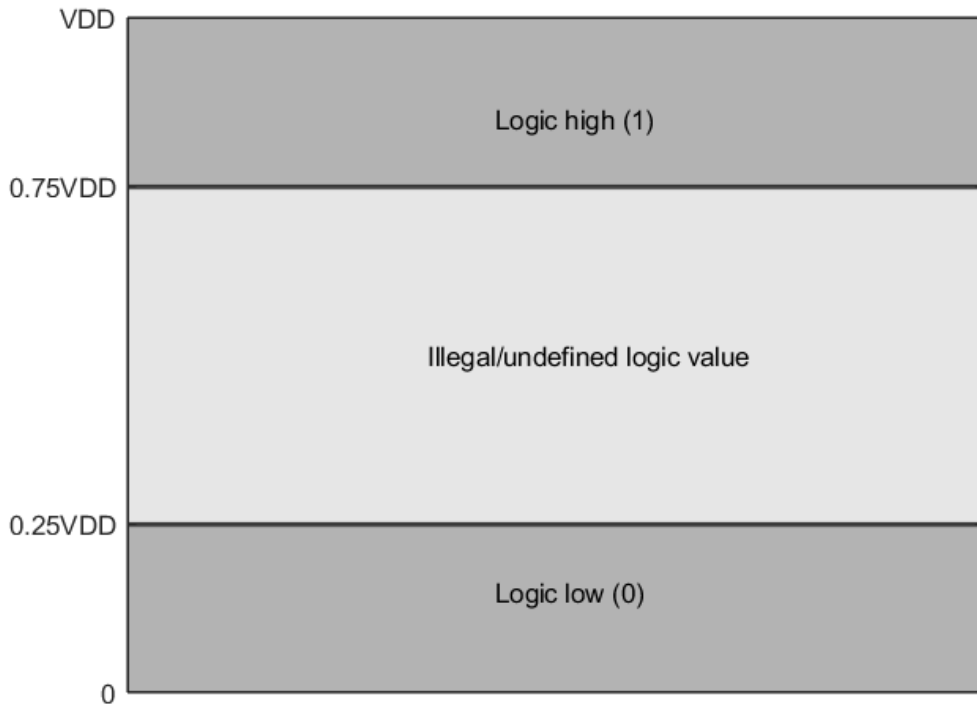


Figure 12: Illustration of the legal ranges for logic high and logic low values, and the illegal middle range where the logic value is undefined.

There is no definitive rule when choosing the lower and upper bounds of these ranges, it will depend on the specification of the design. Normal ranges vary from 30%/70% to 10%/90%. One can argue that the limits should be stricter the lower the supply voltage is, since the difference between logic low and logic high will be smaller in absolute value and therefore more susceptible to being misinterpreted due to process variations etc. The logic limits have for this thesis been chosen as 25%/75%.

4 Method

4.1 Determining the Type and Size of Transistors

In this subsection, the process of finding appropriate types and sizes for the transistors in both the NOT gate and the NAND gate are presented together with the values chosen for the final design.

4.1.1 FD-SOI Transistor Type

As described in Subsection 2.1, the transistors in the 22nm FDSOI technology used here come in five different threshold voltage variations: super low V_{th} (slvt), low V_{th} (lvt), regular V_{th} (rvt), high V_{th} (hvt), and ultra high V_{th} (uhvt). In [7], rvt and slvt transistors were found to perform best in the simulations of transistor mismatch, which was expected as these variations do not have any channel doping while some of the others do. It was therefore decided to use a combination of rvt and/or slvt transistors.

A large difference in driving strength would be reflected in the relative sizes of the NMOS transistors compared to the PMOS transistors needed to achieve equally strong pull-up and pull-down networks, and a larger chip area would be needed. To facilitate for a symmetrical logic gate layout, slvt PMOS and rvt NMOS were chosen as these were found to have comparable driving strengths (the slvt PMOS was found to be slightly stronger than the rvt NMOS) [7]. An additional benefit is that both rvt NMOS and slvt PMOS use a P-well, which means a single P-well could be used for all transistors as explained in Subsection 2.1.1.

4.1.2 Transistor Scaling Method

All transistors are assigned the same length L to increase layout regularity (see Subsection 2.5.2). Methods for scaling the transistor area, which reduces the effect of transistor mismatch in accordance with Pelgrom's law, were presented in Subsection 2.5.1. To keep a decent circuit speed, the transistor length L is chosen to be only slightly longer than the minimum length allowed for the transistor technology used. $L = 32$ nm for all transistors, as this significantly decreased the number of layout design rules compared to an even shorter length. As such, increased area is mainly achieved by increasing the transistor width.

4.1.3 Transistor Widths in the NOT gate

To facilitate layout regularity (see Subsection 2.5.2) the widths were all chosen to be a multiple of a base width $W_{base} = 320$ nm. The widths were then chosen based on the principles described in Subsection 2.4, with $W_{N0} = 3W_{base}$, $W_{N1} = 2W_{base}$, and $W_{N2} = 2W_{base}$. This gives $\frac{W_{N0}}{W_{N2}} = \frac{2}{3}$,

which was found in [4] to be a good ratio. $\frac{W_{N0}}{W_{N1}} = 1.5$, thus fulfilling the main requirement of $W_{N0} > W_{N1}$ (see Subsection 2.4.2).

As a starting point the PMOS and the NMOS were assumed to have the same driving strength, and each PMOS was therefore given the same width as the corresponding NMOS (i.e. $W_{P0} = W_{N0}$, $W_{P1} = W_{N1}$, and $W_{P2} = W_{N2}$). The voltage transfer curve (VTC) was then simulated for the nominal corner (TT) at room temperature (27 degrees), with $VDD = 70$ mV

The VTC should be symmetrical, with $v_{out} = \frac{VDD}{2}$ when $v_{in} = \frac{VDD}{2}$, so that the NOT gate is not skewed in favour of one of the logic values. The lack of symmetry seen for VTC_{start} in Figure 13 is caused by the pull-up network being stronger than the pull-down network. Decreasing the size of $P0$ and/or $P1$ or the size of $N2$ would weaken the pull-up network, but this option was not chosen as the performance of smaller transistors are more affected by transistor mismatch (as explained in Subsection 2.5.1). Instead, the width of feedback transistor $P2$ was increased to $W_{P2} = 3W_{base}$, which resulted in the symmetrical VTC_{final} shown in Figure 13.

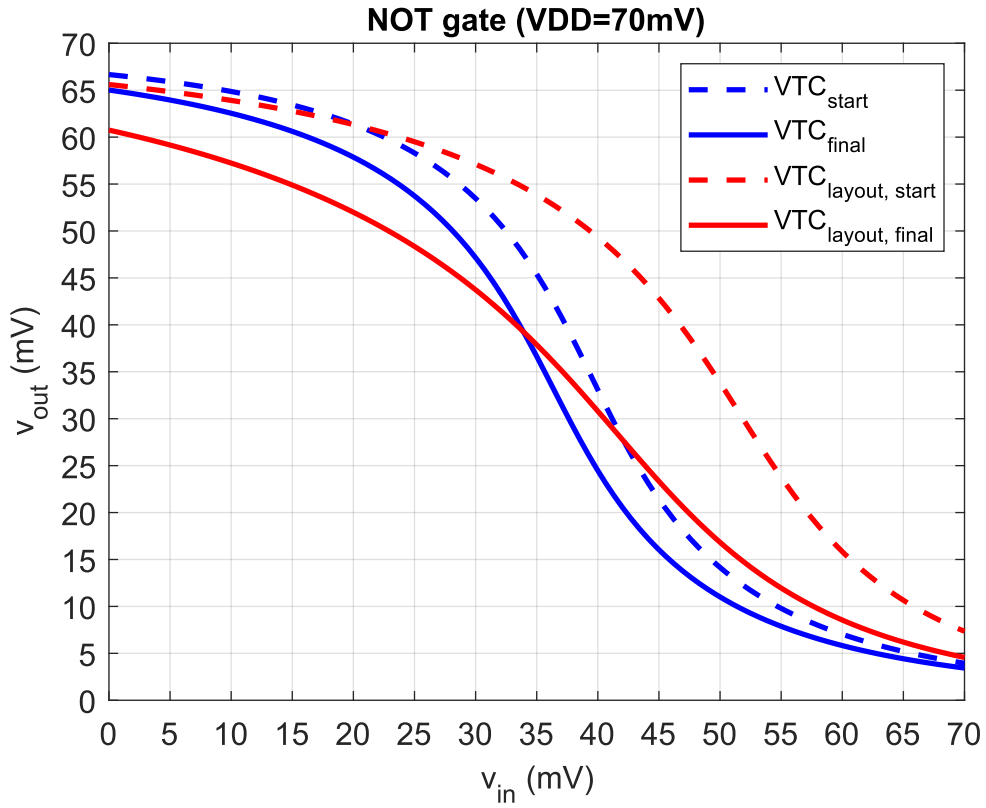


Figure 13: Voltage transfer curves (VTCs) for different configurations of the NOT gate, both pre layout (blue lines) and post layout (red lines).

Layout for the NOT gate was created with the same transistor sizes as in the schematic. As is evident when comparing the post layout transfer curve $VTC_{layout,start}$ to the pre layout transfer curve VTC_{final} in Figure 13, there is a shift of the VTC to the right post layout which means that the pull-up network is stronger (compared to the pull-down network) in the circuit post layout. This shift is caused by the layout methodology used, which will be explained further in Subsection 4.2. To compensate for this shift, the width of the feedback transistor $P2$ is increased to $W_{P2} = 5W_{base}$ in order to achieve the symmetrical transfer curve $VTC_{layout,final}$ in Figure 13. Note that the VTC for the final layout configuration is a bit less steep than the other transfer curves. The switching point comes slightly too late for $VTC_{layout,final}$ ($v_{out} = \frac{V_{DD}}{2} = 35$ mV when $v_{in} = 37$ mV), but this was deemed to be acceptable. The widths of all the transistors in each of the four configurations described are given in Table 2.

Table 2: Transistor widths for the size configurations tested in the NOT gate, both pre and post layout.

	W_{N0}	W_{N1}	W_{N2}	W_{P0}	W_{P1}	W_{P2}
Schematic (start)	960 nm	640 nm	640 nm	960 nm	640 nm	640 nm
Schematic (final)	960 nm	640 nm	640 nm	960 nm	640 nm	960 nm
Layout (start)	960 nm	640 nm	640 nm	960 nm	640 nm	960 nm
Layout (final)	960 nm	640 nm	640 nm	960 nm	640 nm	1600 nm

4.1.4 Transistor Widths in the NAND gate

For the schematic NAND design, the widths found for the NOT schematic (see Table 2), were used as a starting point. The widths of the transistors in parallel were then halved and the widths of the transistors in series were doubled, as described in Subsection 2.4.3. To strengthen the pull-up network compared to the pull-down network, so that the NAND can handle the case where $A \neq B$ better (see Subsection 2.4.3), the widths of $P0A$ and $P0B$ were slightly increased. This is also convenient as the chosen width for $P0$ in the NOT was $3W_{base}$, and halving this would result in a width that is not equal to some integer multiplied by W_{base} . $W_{P0A} = W_{P0B} = 2W_{base}$ is used instead. All widths used for this first schematic NAND configuration are given in Table 3.

The VTC was then simulated for the nominal corner (TT) at room temperature (27°C), with $V_{DD} = 70$ mV. Input B was kept at a constant logic high value ($B = V_{DD}$), and input A was swept from 0 to V_{DD} . From Figure 14, it is clear that VTC_{start} must be shifted to the right to obtain symmetry. This is done by weakening the pull-down network. Since a larger $\frac{W_{N0}}{W_{N1}}$ has been found in [4] to be good in a ST logic gate (see Subsection 2.4.2), only W_{N1} is decreased. W_{P2} is also decreased, as this has the same effect as strengthening the pull-up network, so that $W_{P2} = 2W_{base}$. W_{N2} is slightly increased, which effectively weakens the pull down network. Updated transistor widths can be found in Table 3, in the line for *Schematic (final)*, and the simulated voltage transfer curve is included in Figure 14 as the graph line denoted VTC_{final} . As this is almost symmetrical around $\frac{V_{DD}}{2}$ (the switching point occurs for $v_{in} = 33$ mV), this configuration of widths was selected for the pre layout NAND.

A layout was created using the widths from the final pre layout configuration, see Table 3 for the value of all transistor widths. This resulted in the transfer curve $VTC_{layout,start}$ in Figure 14. The VTC has been shifted drastically to the right compared to the pre layout VTC for the same configuration of widths. This shift is caused by the layout methodology, just like for the NOT, and will be explained in Subsection 4.2. To increase the strength of the pull-down network (which will shift the VTC left), all the driving NMOS transistors are increased. W_{N2} is also slightly increased in order to maintain the same relation between the NMOS transistors as in the schematic, even though this will have the opposite effect on the VTC. It is still worthwhile, as this helps limit the output level deviation. The width of $P2$ is increased a lot, until W_{P2} is the same as in the post layout NOT. The VTC for this final post layout configuration, $VTC_{layout,final}$ in Figure 14, has a much better symmetry. The switching point is the same as for the final pre layout curve VTC_{final} . Note that $VTC_{layout,final}$ is less steep than the pre layout VTC.

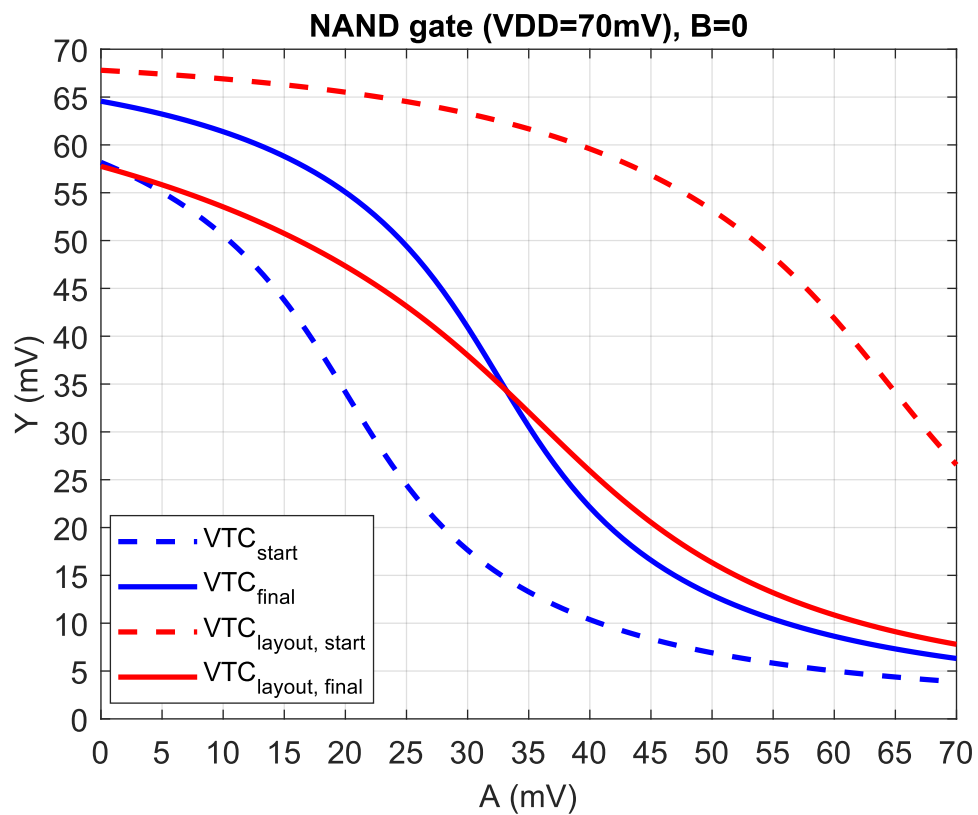


Figure 14: Voltage transfer curves (VTCs) for different configurations of the NAND gate, both pre layout (blue lines) and post layout (red lines).

Table 3: The widths of the transistors in the different NAND configurations tested (both pre and post layout). All widths are a multiple of the base width $W_{base} = 320$ nm. $W_{N0A,B} = W_{N0A} = W_{N0B}$, $W_{N1A,B} = W_{N1A} = W_{N1B}$, $W_{P0A,B} = W_{P0A} = W_{P0B}$, and $W_{P1A,B} = W_{P1A} = W_{P1B}$.

	W_{N0}	W_{N1}	W_{N2}	W_{P0}	W_{P1}	W_{P2}
Schematic (start)	1920 nm	1280 nm	640 nm	640 nm	320 nm	960 nm
Schematic (final)	1920 nm	960 nm	960 nm	640 nm	320 nm	640 nm
Layout (start)	1920 nm	960 nm	960 nm	640 nm	320 nm	640 nm
Layout (final)	2560 nm	1280 nm	1280 nm	640 nm	320 nm	1600 nm

4.2 Physical Layout

In this section, the creation of the physical layout for the circuits is described.

4.2.1 Single P-Well and Substrate Contact

Since slvt PMOS transistors and rvt NMOS transistors were used, see Subsection 4.1.1, all transistors could be placed in a single P-well. When there is only one well, the distance to the nearest well edge is the same as the distance to the end of the chip area. Well proximity effects (WPE) can therefore be assumed to be negligible or non-existing.

As mentioned in previous sections, the impact of the bulk node potential on the transistor's behaviour is assumed to be small since the voltage potentials in the circuit are so low. For simplicity, a substrate contact was therefore included to tie all the bulk nodes to the same voltage potential (ground).

4.2.2 Transistor Folding and Euler's Path

All transistors have a width which is a multiple of a base width W_{base} , and the transistors are therefore treated as transistors with x fingers of width W_{base} . This, combined with the common transistor length L , means the transistor layouts have a high degree of regularity.

Figure 15a shows a simplified schematic illustration of a NOT where each transistor is replaced by a number of unit transistors in parallel. The total width of each transistor group is equal to the ones used for the NOT layout, see *Layout (final)* in Table 2. An Euler path, i.e. a path that never crosses the same transistor twice, can then be drawn for both the pull-up and pull-down network, as illustrated in Figure 15b. Two transistors that are next to each other on the Euler path share a common node (either drain or source), and the entirety of the pull-up/pull-down network can therefore be implemented in layout as if it consisted of one very wide folded transistor. The exception is of course that not all the transistors should be connected to the same gate voltage, and that there are four different source/drain node voltages used in contrast to the normal two voltages that would be present if this was actually just one big transistor.

Creating an Euler path and using transistor folding reduces the S/D junction area, thus reducing the overall area of the design. Connecting all the transistors in one network together also means that the array created has different voltage potentials connected to the gate polys (the feedback transistors have gates connected to the output, while the other transistors' gates are connected to input). In addition, the different shared diffusion areas are connected to different voltage potentials. This, together with the different gate potentials, means that each finger experience a different amount of stress, which makes it hard to predict their actual behaviour [19]. Stress has been found to either enhance or inhibit the diffusion (depending on the stress), and this can cause changes in the threshold voltage [19]. This is called the Length of Diffusion (LOD) effect [14]. The adjustments of the transistor widths from the schematic design to the physical layout, see Subsection 4.1, were done to compensate for this shift.

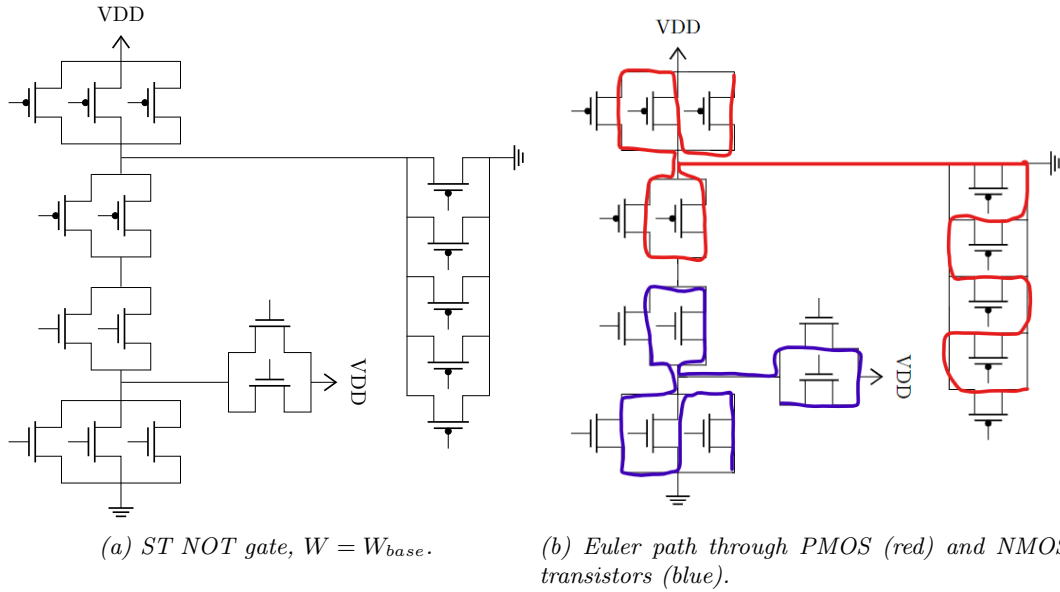


Figure 15: Left: Schematic illustration of ST NOT where all transistors have been replaced with unit transistors ($W = W_{base}$). The equivalent widths are the same as for Layout (final) in Table 2. Right: Illustration of the two Euler paths.

A picture of the NOT gate layout is included in Figure 16. The PMOS transistors are placed in the top row, where the first three gate polys belong to $P0$ and the next two belong to $P1$. All of these are connected to the input pin. Since the input pin is in the $M1$ metal layer, a via is created for each poly to create a connection between the poly and $M1$ layers. The remaining five gate polys in the top row belong to $P2$, and are all connected to the output pin. The NMOS transistors are placed in the bottom row, in the same order as the PMOS transistors (i.e. output transistor first, then inner transistor, then feedback transistor) so that the gate polys match up. This way, a poly strip can be drawn from a PMOS poly to the corresponding NMOS poly, which makes a cleaner layout. As $P2$ is wider than $N2$, three dummy polys have been added after the $N2$ polys to maintain symmetry in the design.

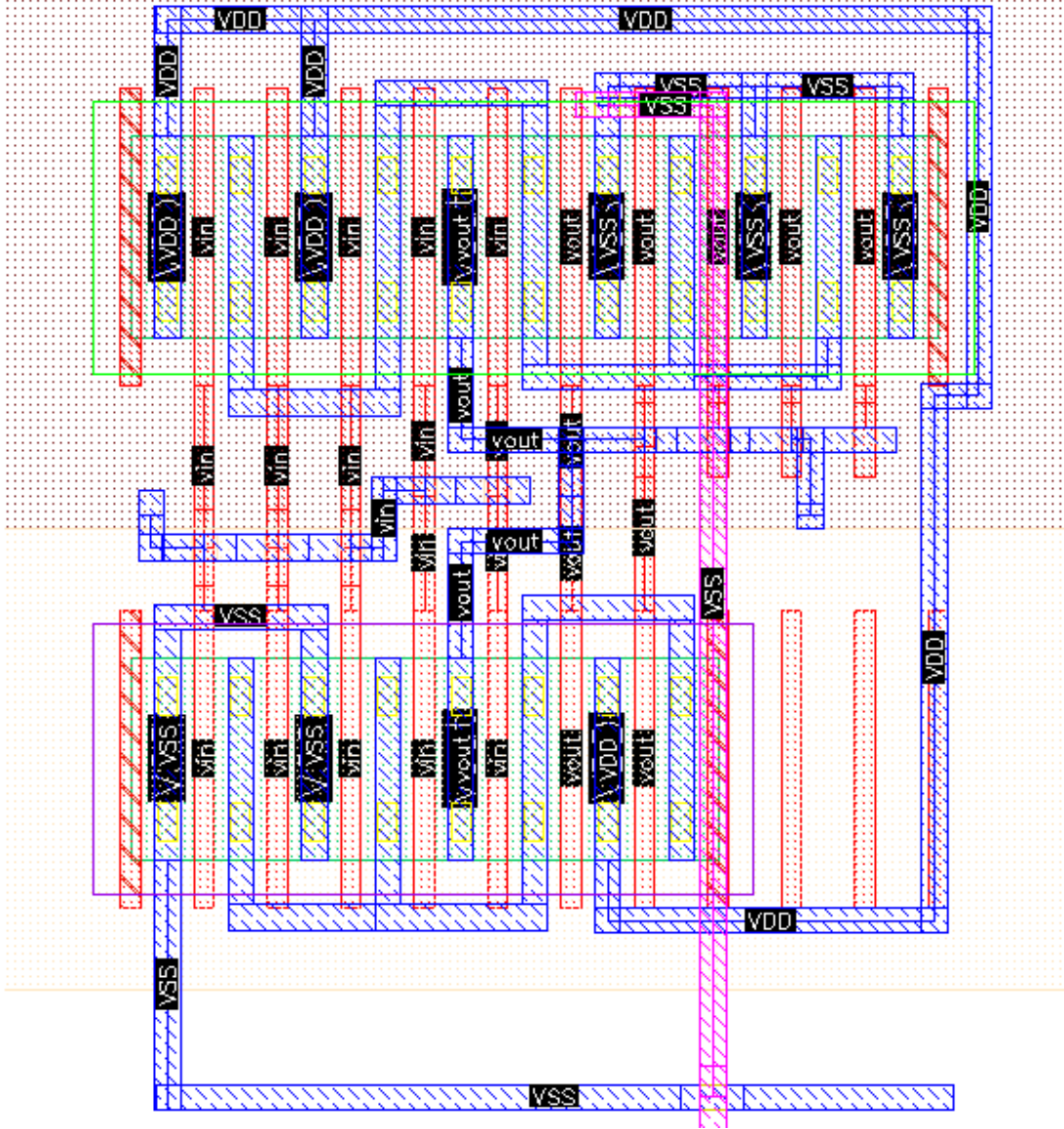


Figure 16: Layout of the ST NOT. The pull-up network is the row on top, and the pull-down network the row below. The three first gate polys in each network belong to the outer transistor (N0 / P0), the next two belong to the inner transistors (N1 / P1) and the final gate polys belong to the feedback transistors (N1 / P2). Area = $3.0251\mu\text{m}^2$ (with $h = 1.79\ \mu\text{m}$ and $w = 1.69\ \mu\text{m}$). The dummy polys and substrate contact are not included here, as this layout is used as a building block for larger layouts. To use this NOT gate on its own, a substrate contact must be connected to the VSS-rail and dummy polys added at the ends.

In the same way as for the NOT, the NAND layout is constructed by first creating an equivalent schematic (see Figure 17a) where all transistors are replaced by a number of unit transistors in parallel. The total width of each parallel group is equal to the width of the equivalent transistor used for the NAND layout, which are listed for *Layout(final)* in Table 3. An Euler path is then created for the pull-up and pull-down networks in the NAND, as illustrated in Figure 17b. There are several possible Euler paths for both the pull-up and pull-down network, and a choice was therefore made to create paths that allowed the pull-up network's gates to mirror the pull-down network's gates as best as possible such that two gate polys that are opposite each other are connected to the same gate voltage potential. When creating the layout, the PMOS gate polys can then simply be extended vertically so that they connect to the NMOS gate polys, as seen in Figure 18.

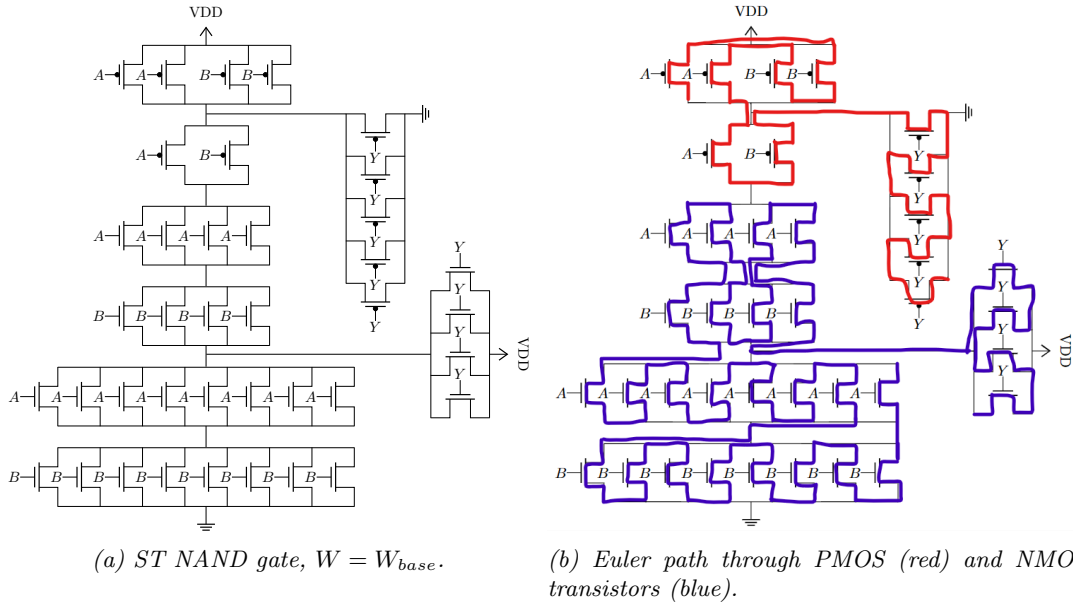


Figure 17: Left: Schematic illustration of ST NAND where all transistors have been replaced with unit transistors ($W = W_{base}$). The equivalent widths are the same as for *Layout (final)* in Table 3. Right: Illustration of the two Euler paths.

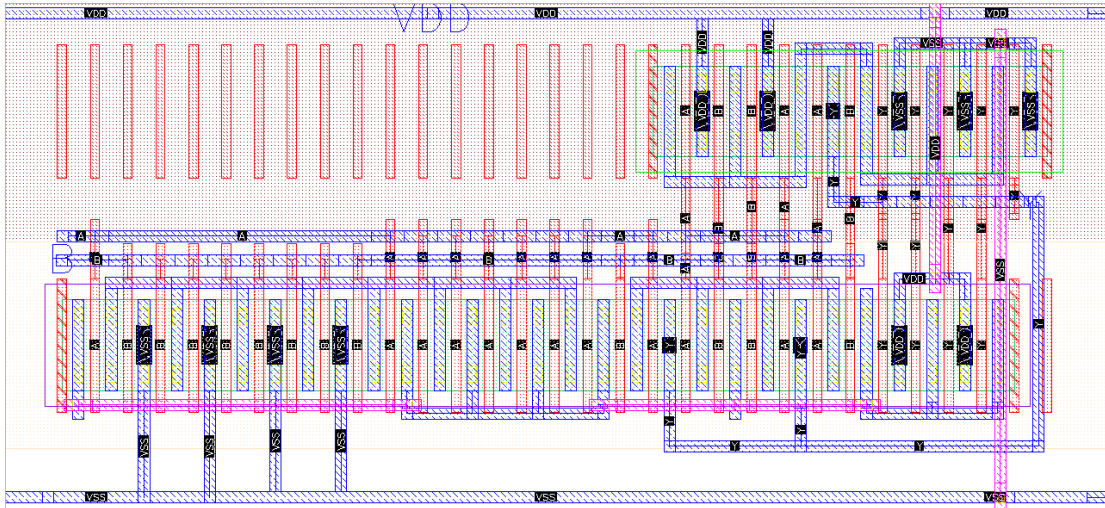


Figure 18: Layout of the ST NAND gate. The pull-up network (PMOS) is the row on top, and the pull-down network (NMOS) the row below. Area = $6.981\mu\text{m}^2$ ($h = 1.79\ \mu\text{m}$ and $w = 3.9\ \mu\text{m}$). The dummy polys and substrate contact are not included here, as this layout is used as a building block for larger layouts. To use this NOT gate on its own, a substrate contact must be connected to the VSS-rail and dummy polys added at the ends.

The NOT layout in Figure 16 and NAND layout in Figure 18 are building blocks used to create larger circuits. Since the requirement is that any part of the design must within a $40\ \mu\text{m}$ distance from a substrate contact, the substrate contacts are not included in each NOT and NAND building blocks but rather applied where necessary in the larger designs. If the NOT or NAND are being used by themselves, a dummy poly must be added to each end of the rows of gate polys.

Both dummy polys on the side and a substrate contact are therefore included in the NOT layout design and NAND layout design that were extracted for post layout simulation.

4.2.3 Block Regularity

NOT and NAND are the two building blocks which all the other circuits are based on. To enable stacking of a NAND and NOT next to each other, the layouts of the two logic gates are coordinated so that the distances between the PMOS and NMOS polys, between the PMOS polys and the power rail (VDD), and between the ground rail (VSS) and the NMOS polys are the same for both. The height of the NOT and NAND layout are therefore the same, $h = 1.79\ \mu\text{m}$.

Figure 19 shows how NAND and NOT gates are stacked together to create the D-latch (with topology as described in Subsection 3.2). A green outline has been drawn around the NANDs

and a yellow outline around the NOTs to make the illustration more readable. The symmetry with regards to poly placement and heights in the two logic gates made this layout easy to design, as all gates could simply be stacked together. The placement of the logic gates in the D-latch layout corresponds to the placement of the logic gates in the schematic illustration in Figure 8.

There are two rows of logic gates in the D-latch layout in Figure 19. The first gate on the top row is a NAND gate. An input pin for the D-latch enable signal EN is connected to the B -input of the NAND. A M2 wire (pink) connects the other NAND input to input pin D which is placed at the input of the NOT gate to the left of the bottom row of logic gates. The second NAND gate in the top row is placed so that the distance between the final polys in the first NAND and the first polys in the second NAND are the same as within the logic gates themselves. This stacking means that no dummy polys are needed between the two NANDs. A M2 wire connects the output of the first NAND to input A of the second NAND. Similarly, the output of the second NAND is connected to the first NOT in the upper row, and the output of the first NOT is connected to the input of the second. All the gates in the top row are stacked together as described for the NAND gates, thus omitting the need for dummy polys between the logic gates.

All the logic gates in the bottom row have been flipped vertically, so that the VSS rail is on top and the VDD rail on the bottom. The VSS rail in the top row can then overlap with the VSS rail in the bottom row, which saves chip area. The D-latch's input pin D is connected to the input of the first NOT in the bottom row. All the gates in the bottom row are stacked together as described for the top row, and dummy polys are therefore unnecessary between the logic gates.

The D latch is used as a building block to create the 8 bit memory row, see Subsection 3.1. It might therefore be stacked together with other circuits on the left and/or right, in which case dummy polys are not needed at the ends of the D-latch. No dummy polys are therefore included in the layout in Figure 19, but rather added on later depending on how this building block is used. Note that to run post layout simulations on the D-latch by itself, a dummy poly at the end of each row of polys is necessary to pass the DRC and thus ensure proper behaviour.

As mentioned in the case of the NOT and the NAND, the distance from any part of the layout to the substrate contact must be less than 40 μm to ensure proper body connection for all transistors. The D-latch is 3.51 μm high and 11.666 μm wide, so it is not necessary to include a substrate contact for every D-latch when it is used as a building block in the larger memory circuits. This has therefore been excluded from the layout of the D-latch building block in Figure 19. Should one wish to use the D-latch by itself, i.e. to run post layout simulations, a substrate contact must be included before the layout is extracted. A space is left for the substrate contact between the second NAND in the top row and the second NAND in the bottom row in Figure 19, so that it can be connected to the VSS -rail.

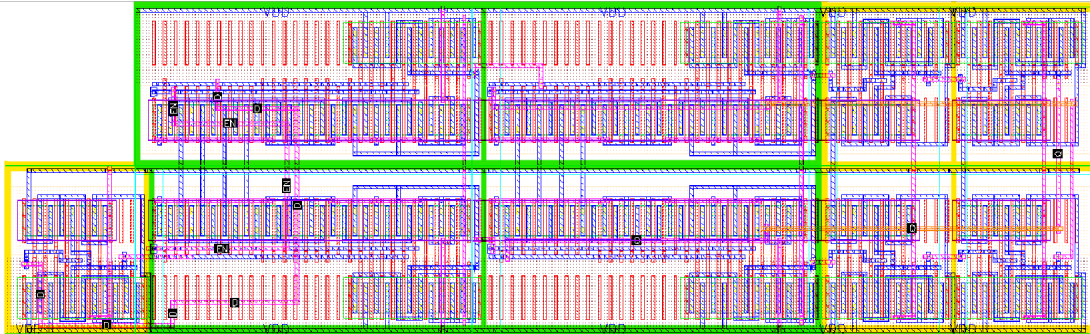


Figure 19: Layout of the D-latch. The NOT gates are outlined in yellow, and the NAND gates are surrounded by a green outline. Area = $40.94766\mu\text{m}^2$ (with $h = 3.51\ \mu\text{m}$ and $w = 11.666\ \mu\text{m}$). The dummy polys and substrate contact are not included here, as this layout is used as a building block for larger layouts. To use this NOT gate on its own, a substrate contact must be connected to the VSS-rail and dummy polys added at the ends.

The Decoder and the Output Selection Module are made up from NOTs and NANDs, just like the D latch. Since the layout principles used are the same as for the D latch, these layouts will not be described in detail here. The Decoder layout is included in Subsection H, and the Output Selection Module layout is included in Subsection I.

4.2.4 Layout of 8-bit Row

Eight D latches were stacked on top of each other, with the VDD -rails overlapping, to create the main body of the 8-bit row. Every other D latch was flipped vertically, to create an overall shape that lends itself to further stacking. The layout is shown in Figure 20. The two first D latches, which store bit 0 and bit 1, are highlighted by a yellow outline in Figure 20. A NAND gate is placed at the output of each D latch, as described in Subsection 3.1. The output NANDs for D latch 0 and D latch 1 are outlined in grey in Figure 20.

Three NOTs and two NANDs are used to create *Read* and *Write* signals from *RW* and *Select*, see Subsection 3.1. These are all stacked one after each other in a row, and then flipped vertically before the row is placed on top of the stack of D latches in such a way that the VDD -rails overlap. The 8-bit row is only a step towards building a SRAM, so dummy polys at the row ends and substrate contacts are not included at this stage.

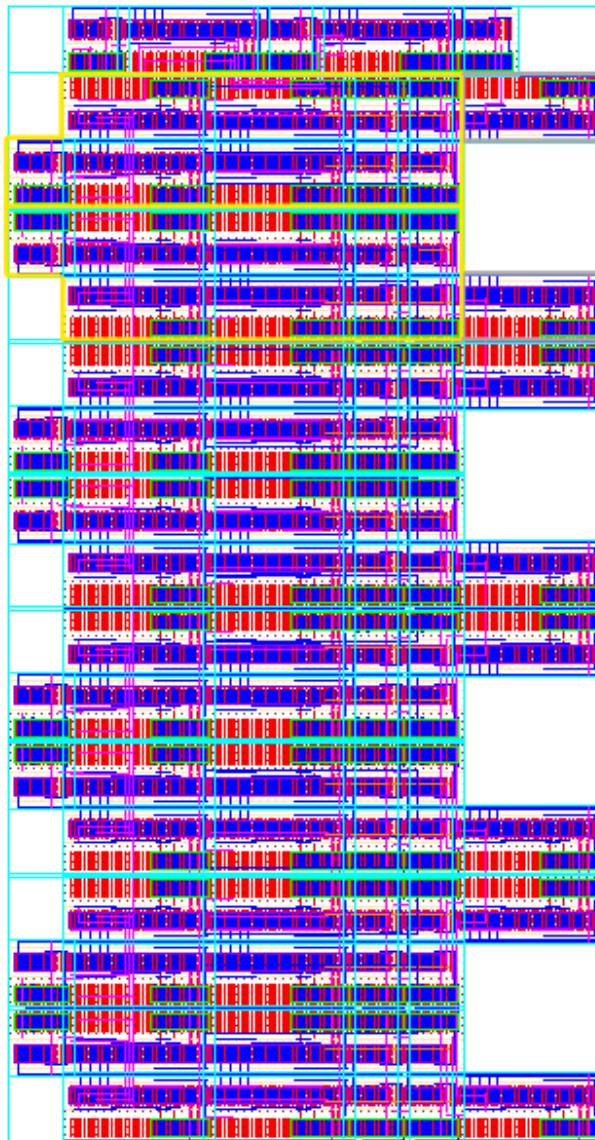


Figure 20: Layout of the 8-bit row. The D latches that store bit 0 and bit 1 are outlined in yellow. The output NAND gate is outlined in grey for each of these two D latches. The row on top (NOT, NAND, NOT, NAND, NOT) are the logic used to create Read and Write. Area = $444.06978\mu\text{m}^2$ (with $h = 29.085\ \mu\text{m}$ and $w = 15.268\ \mu\text{m}$). The dummy polys and substrate contact are not included here, as this layout is used as a building block for larger layouts. To use this NOT gate on its own, a substrate contact must be connected to the VSS-rail and dummy polys added at the ends.

4.2.5 SRAM layout

A layout for the 4B SRAM, see Figure 21, is created by placing four 8-bit rows next to each other in such a way that the output NANDs of one 8-bit row line up with the D latch without needing any dummy polys. Each of the 8-bit rows are outlined in yellow to make it easier to see how they are fitted together. A small gap remains between the NOT at the end of a D latch and the NOT at the start of the equivalent D latch in the next 8-bit row, and dummy polys were therefore added here.

A Decoder is placed in the top left corner, and each of its outputs is connected to one of the 8-bit rows *Select*-input. The column to the right in Figure 21, where the two modules at the top are outlined in green, is a stack of eight Output Selection Modules with overlapping *VDD*-rails.

A larger version of Figure 21 without any annotations is included in Subsection J.

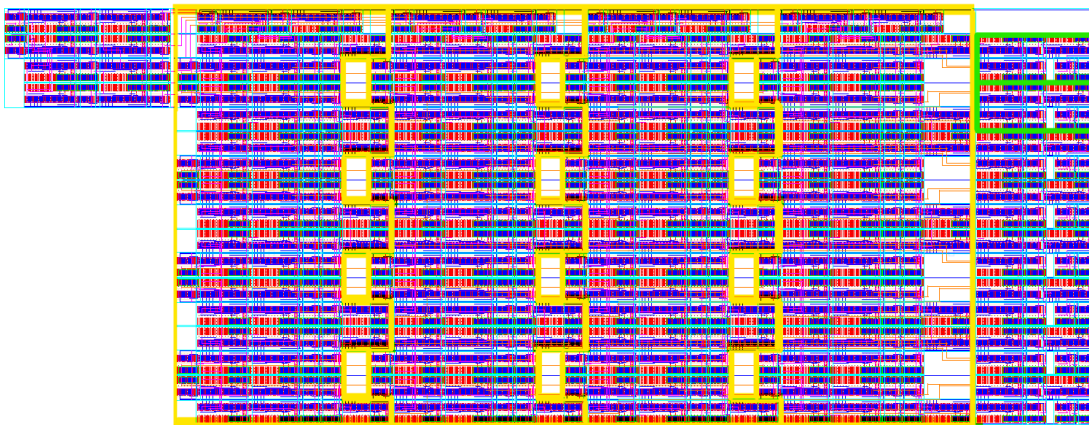


Figure 21: Layout of the 4B SRAM. The four 8-bit rows (or rather: columns) are outlined with yellow. A Decoder is placed in the top left corner. Eight Output Selection Modules are placed in a column to the right, and the first two are outlined in green. Area = $1961.33244\mu\text{m}^2$ (with $h = 29.085\ \mu\text{m}$ and $w = 76.616\ \mu\text{m}$).

To create the 16B SRAM, four 4B SRAMs must be stacked together either vertically or horizontally. Horizontal stacking was chosen, for no other reason than that this was used for 4B SRAM and meant that the work needed to create the 16B SRAM largely resembled the work done for 4B (thus being easier to create from a layout designer perspective). Dummy polys are added at all poly row ends that are not immediately followed by another row of polys. The additional Decoder is placed in the top left corner and output logic is placed in a column to the right, just like for the 4B SRAM layout. A picture of the 16B layout is included in Subsection K.

The 64B SRAM was made using vertical stacking of 16B SRAMs, in order to restore some degree of proportion between height and width. The power/ground-rails at the end of a 16B SRAM overlaps with the power/ground-rails of the next 16B SRAM. A picture of the 64B SRAM layout

is included in Subsection L. Every other 16B SRAM is flipped vertically to achieve correct rail overlaps. The decoder, which creates the EN -signal for all the 16B SRAM's decoders, is placed at the input of the second 16B SRAM. The output logic (eight Output Selection Modules with a NOT at each input) is placed to the right of the layout, at the output of the second 16B SRAM from the top.

4.3 Layout Extraction

All the physical layouts are tested to see that they pass the Design Rules Check (DRC) and the LVS (Layout Versus Schematic) check. The Calibre-tool in Cadence Virtuoso's Layout Suite XL environment is used to run both checks. When the layout passes both LVS and the DRC, xACT (another part of the Calibre tool) is used to extract the layout effects and parasitics for all corners at $27^{\circ}C$. As the voltages and currents in this design are very low, the parasitic resistances are negligible. Only the parasitic capacitances and coupling capacitances ($C+CC$ in xACT) are therefore extracted, as excluding the resistance will reduce the simulation time which is very desirable for larger designs.

Some errors remain in the DRC for the 64B layout. All these errors report that the M2 density is too low in certain parts of the layout. Attempts were made to solve the errors, but since the M2 layer is used a lot for wires it was difficult and time consuming to find places where M2-fill could be placed. These errors do not affect the functionality of the 64B SRAM, and not solving them are assumed to only have minor effects on the circuits performance.

4.4 Monte Carlo Mismatch Simulations

The effect of transistor mismatch on the circuits' yield and general performance is simulated by running Monte Carlo simulations.

If nothing else is explicitly stated, all references to a Monte Carlo simulation or a mismatch simulation means that a Monte Carlo mismatch simulation has been run with 1000 points using the Latin Hypercube method for sampling. 12345 was used as the seed. The temperature is $27^{\circ}C$, and the simulations are done on the typical (TT) corner.

By using a relatively high number of simulation points, the circuit yield found can be assumed to be a good estimate.

4.5 Simulating Process Variations

All references to simulations of process variations / process corners refer to simulations of the process corner and temperature combinations listed in Table 4. Each temperature and corner combination has been assigned a name, e.g. SS20, so that they can be referred to more easily.

Table 4: The combination of process corner and temperature tested when simulating process variation.

Temperature \ Corner	TT	SS	FF	FS	SF
0°C	TT0	SS0	FF0	FS0	SF0
20°C	TT20	SS20	FF20	FS20	SF20
27°C	TT27	SS27	FF27	FS27	SF27
50°C	TT50	SS50	FF50	FS50	SF50

4.6 NOT Gate Testbench

The NOT gate is tested, both pre and post layout, using the testbench shown in Figure 22. Two NOT gates, *DRIVE0* and *DRIVE1*, are used to drive the DUT, and another NOT gate, *LOAD*, is used as the output load. A signal V_{in} is applied to the input of the first driving inverter. *DRIVE0* inverts the input V_{in} and this signal is again inverted by *DRIVE1*, resulting in a signal In at the input of the DUT which should hold the same value as the input signal V_{in} .

The dimensions of the transistors in the pre layout NOT and post layout NOT are as found in Subsection 4.1.3.

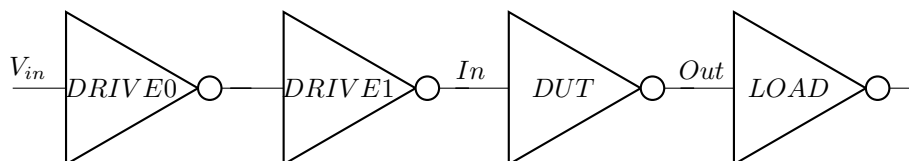


Figure 22: NOT gate testbench schematic. DUT, the third NOT gate from the left, is the device under test.

The circuit yield of the NOT gate is simulated, both pre and post layout, using a Monte Carlo simulation as described in Subsection 4.4 with a DC operating point analysis as the basis. This simulation is repeated for several different supply voltages in order to observe how the NOT gate's yield is affected when lowering the supply voltage. $V_{in} = 0$ mV for all tests, so In should be logic low and Out should be logic high. The upper limit for logic low is $0.25VDD$ and the lower limit for logic high is $0.75VDD$, as described in Subsection 3.6. The numerical value of these limits are listed in Table 5 for each of the supply voltages tested.

The same DC operating point analysis, with $V_{in} = 0$ mV, is used as the basis when simulating process variation. Tests are run for the corner and temperature combinations described in Subsection 4.5, with $VDD = 70$ mV.

Table 5: The upper limit for logic low ($V_{L,max}$) and lower limit for logic high ($V_{H,min}$) for different supply voltages (VDD).

VDD	$V_{L,max}(0.25VDD)$	$V_{H,min}(0.75VDD)$
75 mV	18.75 mV	56.25 mV
70 mV	17.5 mV	52.5 mV
65 mV	16.25 mV	48.75 mV
60 mV	15 mV	45 mV
55 mV	13.75 mV	41.25 mV
50 mV	12.5 mV	37.5 mV

4.7 NAND Gate Testbench

In the larger circuits, such as the D-latch, decoder, and output selection module, a NAND gate is most commonly paired with a NOT load and driven by a NOT gate. The testbench in Figure 23, which uses NOT gates as drivers of each NAND input and as load at the NAND output, is therefore used to simulate the behaviour of the NAND gate both pre and post layout. The dimensions used for the transistors in the pre layout NAND and post layout NAND are described in Subsection 4.1.4.

Monte carlo simulations are run as described in Subsection 4.4 and with a DC Operating Point analysis as the base, in order to find how transistor mismatch affects the NAND gate yield pre and post layout. Simulations are run for all four input combinations in Table 6, for both $VDD = 70$ mV and $VDD = 75$ mV (acceptable range for logic low and logic high values are as listed in Table 5).

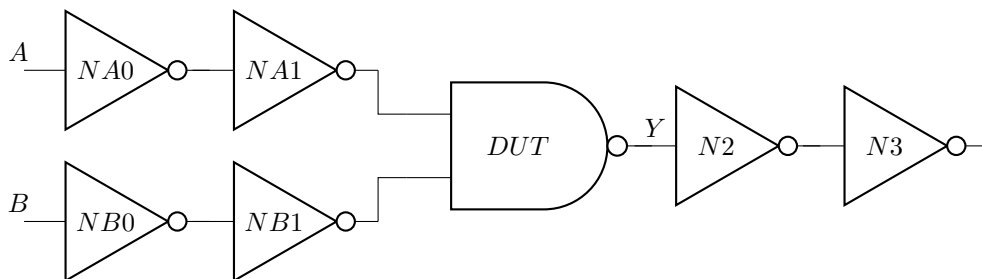


Figure 23: NAND gate testbench schematic. Two NOT gates are used as load for the DUT. Two NOT gates in series are used to drive each of the inputs. See Table 6 for combinations of input stimuli A and B and expected output Y .

A transient analysis lasting 200 μ s, illustrated by the timing diagram in Table 24, is used as the base for the simulations of process variation. The pre layout and post layout process corners are simulated as described in Subsection 4.5, for supply voltages $VDD = 80$ mV and $VDD = 85$ mV. Post layout simulations are run for $VDD = 87$ mV as well.

Table 6: Truth table for NAND gate, showing expected output value Y for all combinations of inputs A and B .

A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

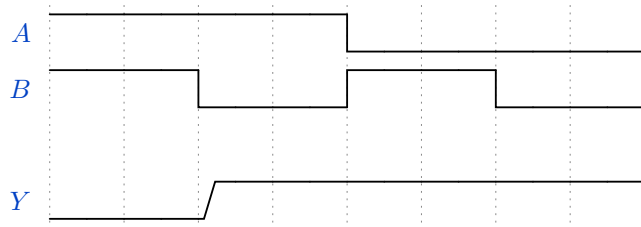
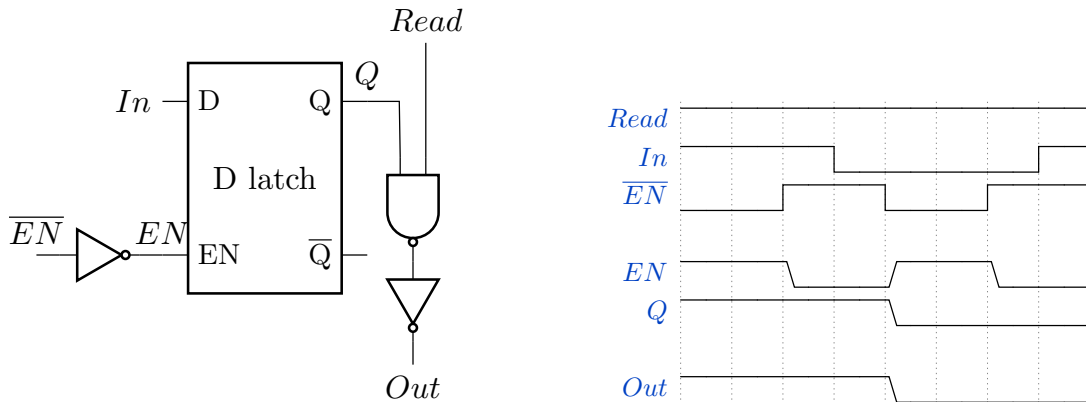


Figure 24: Timing diagram for the NAND. The transient analysis lasts $200 \mu\text{s}$.

4.8 D Latch Testbench

When part of the SRAM, the D latch will receive an enable signal EN which has been generated by the input logic (decoder and read/write-logic). This signal is assumed to be within the legal range for the logic value it represents, but is not expected to be a perfect logic high or logic low value. This expectation is reflected in the D latch testbench in Figure 25a by using a NOT gate to create the enable signal, as this will create a natural logic level degradation of EN . A NAND gate followed by a NOT gate is used as the LOAD for the D latch's Q output, as this will be the load it sees when part of the SRAM (see Subsection 3.1).

A transient analysis is used as the basis for all pre and post layout simulations on the D latch testbench. The testbench inputs In and \overline{EN} are varied throughout the analysis as illustrated in the timing diagram in Figure 25b, while $Read$ is kept at a constant logic high value equal to VDD throughout the analysis. Expected values of the intermediate signals EN and Q are also shown in Figure 25b, as well as the expected output Out .



(a) Block diagram of the testbench.

(b) Timing diagram for transient analysis lasting 8 periods (end of each period is marked with a vertical dotted line).

Figure 25: D Latch testbench. The DUT drives a similar load as what it will do when part of the SRAM. The timing diagram shows inputs, expected intermediate values and expected output for the transient analysis.

The transient analysis lasts 8 periods, see Figure 25b, and a period $p = 12.5 \mu\text{s}$ is used when running Monte Carlo simulations on the testbench as described in Subsection 4.4 (i.e. 100 μs total duration). The pre layout mismatch simulation is run for both $VDD = 65 \text{ mV}$ and $VDD = 70 \text{ mV}$, while the post layout simulation is run for $VDD = 80 \text{ mV}$.

The effect of process variation on the D latch is simulated by running the transient analysis for all the corner and temperature combinations described in Subsection 4.5. Pre layout simulations are run for both $VDD = 70 \text{ mV}$ and $VDD = 75 \text{ mV}$ with a transient analysis period $p = 25 \mu\text{s}$ (200 μs total duration). The post layout process variation is simulated for $VDD = 75 \text{ mV}$ using a period $p = 50 \mu\text{s}$ (total duration 400 μs) and for $VDD = 80 \text{ mV}$ using a period $p = 25 \mu\text{s}$ (total duration 200 μs).

4.9 2x4 Decoder Testbench

The testbench used to verify the behaviour of the 2x4 decoder is illustrated by the block diagram in Figure 26, where the decoder marked *DUT* is the device under test. In the finished memory, the output of a decoder is either driving another decoder or driving the input to an 8-bit row of bitcells (see Subsection 3.1). This latter has been replicated in the testbench in Figure 26, where each of the DUT's outputs is driving a load equivalent to the input of an 8-bit row of bitcells.

A decoder's *EN*-input will in some cases come from outside the SRAM and be created by circuitry that is unknown to us when designing the SRAM. In all other cases *EN* will be created

by another decoder inside the SRAM, and it is this that the decoder marked *drive* in Figure 26 replicates. The input signals I_0 , I_1 , I_2 , and I_3 , which combined creates an address = $I[3 : 0]$, are always applied from the outside of the SRAM, and are for the purpose of these simulations assumed to be perfect logic values (either V_{DD} or 0 mV).

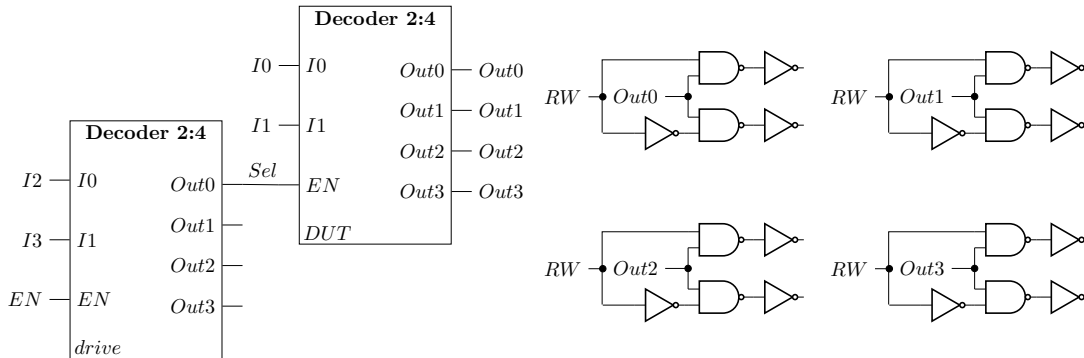


Figure 26: Testbench for the 2×4 decoder. The DUT is driving the same load as it would be if connected to an 8 bit row of bit-cells in the SRAM. Another decoder, *drive*, is used to create the input to the DUT's *EN*-pin.

A transient analysis, illustrated by the timing diagram in Figure 27, is used as the basis for all simulations on the testbench in Figure 26. This way the combinatorial behaviour of the decoder and its performance with respect to speed are tested simultaneously. The transient analysis lasts for 8 periods, marked by the vertical dotted lines in the timing diagram in Figure 27.

The timing diagram shows how the input signals EN , I_0 , I_1 , I_2 , and I_3 in Figure 26 are varied during the analysis. The expected value of the intermediate signal *Sel* is also shown, as well as the expected values for the output signals Out_0 , Out_1 , Out_2 , and Out_3 . RW , which is an additional input signal to the DUT's load circuits in Figure 26, is kept at a constant logic high value throughout the analysis and is therefore not shown in the timing diagram.

For the first four periods in Figure 27, I_2 and I_3 are kept low and EN high so that the signal *Sel* from the driving decoder's *Out0*-pin is kept high and the DUT is enabled. All four possible combinations of I_0 and I_1 are tested during this, to check that the decoder produces correct output values. I_3 goes high at the start of the fifth period, which should force *Sel* to go low as another of the driving decoders output pins is selected. As the DUT is no longer selected, all outputs are expected to be low. I_0 and I_1 go low at the start of the sixth period, but as I_3 remains high the DUT is still not selected and no changes should occur at the outputs. At the start of the seventh period I_3 goes low so that *Sel* is again high and the DUT is enabled. As $I_0 = I_1 = 0$, Out_0 is expected to go high as soon as the DUT is enabled. For the eighth and final period, the EN signal is set low so that the driving decoder is no longer selected. This should force *Sel* low, and all the DUT's outputs should be low as the DUT is no longer enabled.

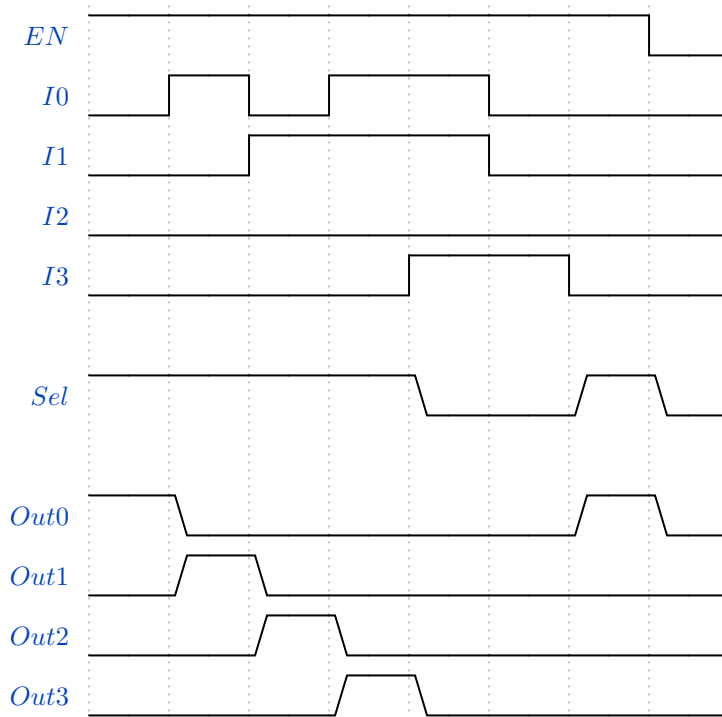


Figure 27: Timing diagram showing how the input signals EN , $I0$, $I1$, $I2$, and $I3$ are varied during a transient analysis on the decoder testbench in Figure 26. Expected values for the intermediate signal Sel is shown, as well as expected values for the output signals $Out0$, $Out1$, $Out2$, and $Out3$. $RW = VDD$. The analysis lasts 8 periods, where each period is marked by a dotted vertical line.

The effect of transistor mismatch on the decoder is simulated by running Monte Carlo simulations as described in Subsection 4.4. The transient analysis described above is used as the base for the simulation, with a period $p = 50 \mu\text{s}$ and a total time of $400 \mu\text{s}$. Pre layout simulations are run for both $VDD = 70 \text{ mV}$ and $VDD = 75 \text{ mV}$, while the post layout design is simulated for the supply voltages $VDD = 80 \text{ mV}$ and $VDD = 85 \text{ mV}$.

Process variation is simulated by testing all corner and temperature combinations described in Subsection 4.5. For the pre layout simulations of process variation, a transient period $p = 50 \mu\text{s}$ is used and the simulations are run for $VDD = 70 \text{ mV}$ and $VDD = 75 \text{ mV}$. To compensate for an increase in the circuit delay, the transient period was increased to $p = 100 \mu\text{s}$ for the post layout process variation simulation (total transient run time of $800 \mu\text{s}$). The post layout process variation was simulated for $VDD = 75 \text{ mV}$, $VDD = 80 \text{ mV}$, and $VDD = 85 \text{ mV}$.

4.10 Output Selection Module Testbench

Figure 28 is a block diagram of the testbench used when simulating the behaviour of the output selection module. NAND gates are used to create the four input signals to the DUT, as it is the NAND at the outputs of each bitcell which will drive these signals when the module is used in a SRAM (see Subsection 3.1). A NOT gate followed by another output selection module is used as load for the DUT. When integrated into the SRAM, an output selection module is expected to have either this load or an unknown load (whatever is attached to the output of the SRAM).

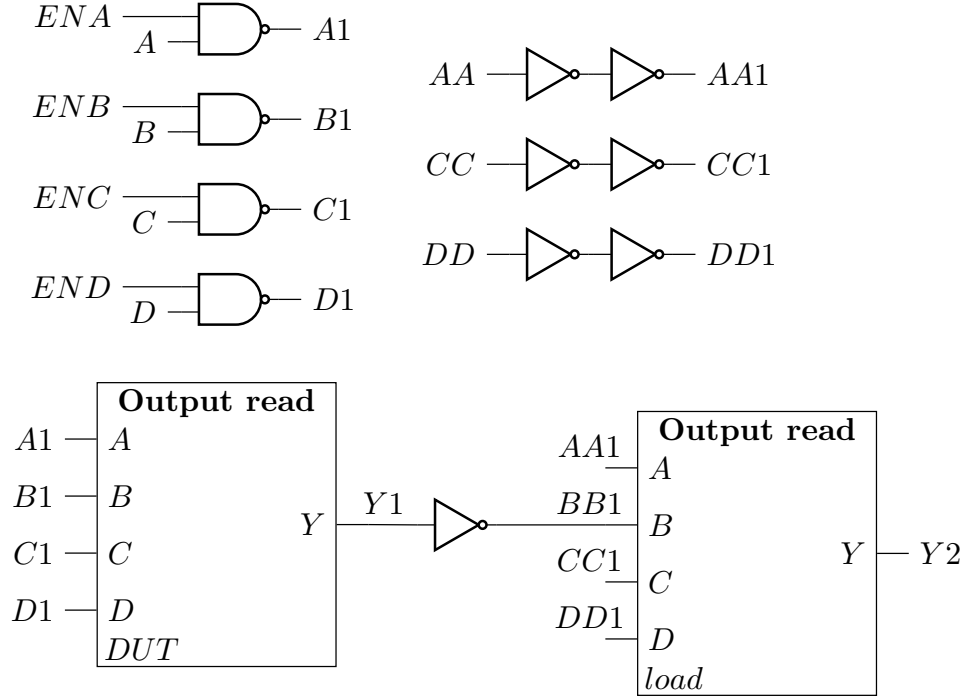


Figure 28: Block diagram showing the testbench for the output selection module.

A transient analysis lasting 1200 μs is used as the basis for all simulations in order to both verify that the combinatorial logic works correctly and to observe the delay through the module. The applied signals A , C , D , AA , CC , and DD are kept at a constant logic high value, while B is kept at a constant logic low value. The other input signals, ENA , ENB , ENC , and END , are varied as shown in the timing diagram in Figure 29. Expected values for the intermediate signals $A1$, $B1$, $C1$, and $D1$, which are applied to the input pins of DUT in Figure 28, are also illustrated in the timing diagram. The transient simulation is split into ten 120 μs periods, where each period is marked by a vertical dotted line in Figure 29.

Monte Carlo simulations are run as described in Subsection 4.4 to see how transistor mismatch affects the performance of the output selection module. Both the pre layout and post layout

simulations are run for the supply voltages $VDD = 80$ mV and $VDD = 85$ mV.

Process variation is simulated using the process corner and temperature combinations described in Subsection 4.5. The pre layout simulations of process variation are run for the supply voltages $VDD = 75$ mV and $VDD = 80$ mV, while the post layout process simulations are run for $VDD = 80$ mV and $VDD = 85$ mV.

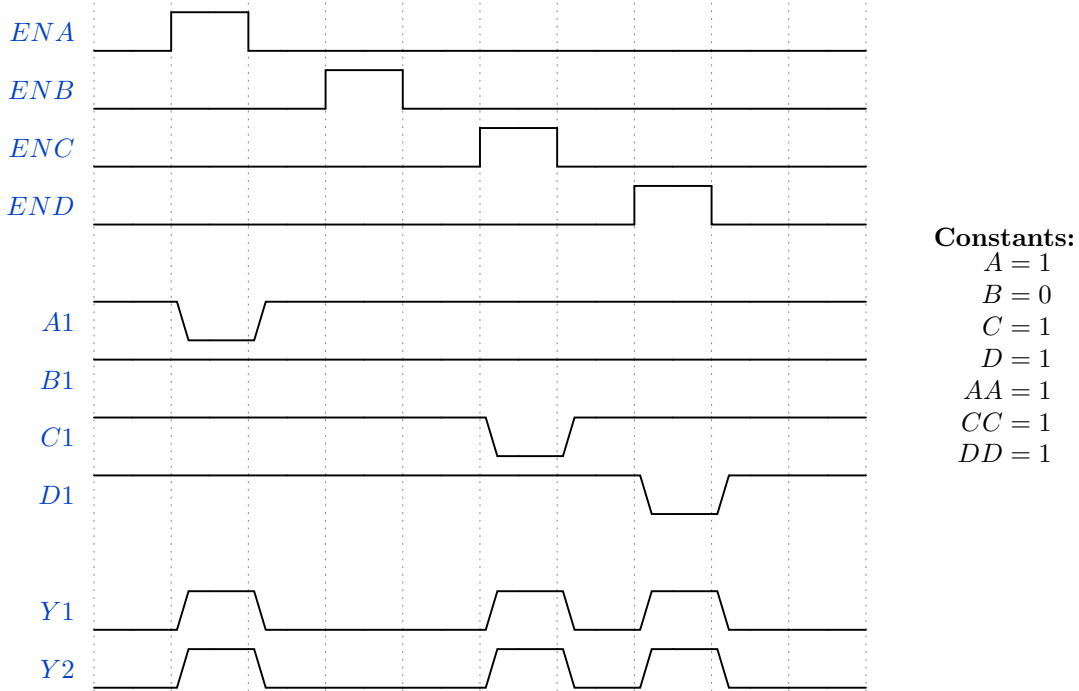


Figure 29: Timing diagram showing the stimuli applied to the testbench in Figure 28 during the transient analysis. Expected values of the intermediate signals $A1$, $B1$, $C1$, and $D1$ are also shown, as well as expected values for the DUT’s output signal $Y1$ and the load’s output signal $Y2$. A , B , C , D , AA , CC , and DD are kept at the constant values given in the list to the right. The transient analysis lasts for ten $120 \mu\text{s}$ periods (the end of each period is marked by a vertical dotted line).

4.11 4B SRAM Testbench

The 4B SRAM is the smallest of the memories designed in this thesis. It is a self contained memory circuit able to store four bytes (1 byte = 8 bit). Figure 30 shows the 4B SRAM and the names of all inputs and outputs. Stimuli is applied to the inputs in the form of square pulses, and the outputs are sampled to check that the circuit behaves correctly. No additional circuitry is added to drive the 4B SRAM or to be used as load in the testbench.

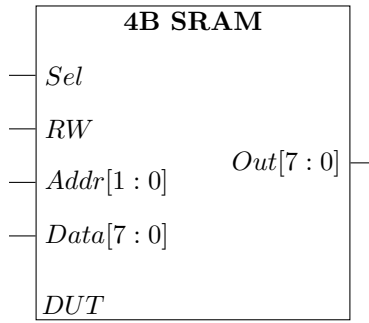


Figure 30: Block diagram of the 4B SRAM testbench, which consists only of the 4B SRAM. Input signals are applied as shown in Figure 31, and the values of the eight output signals are sampled.

A transient analysis lasting 1 ms is used as the base for all simulations on the 4B SRAM. It is divided into ten 100 μ s periods, as indicated by the dotted vertical lines in the timing diagram in Figure 31. The timing diagram shows the applied inputs and expected outputs for each stage of the transient analysis.

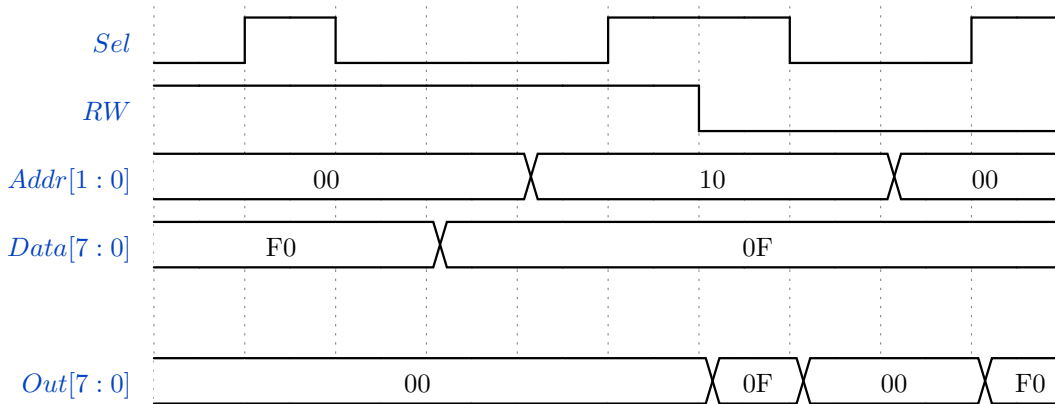


Figure 31: Timing diagram showing how the input signals Sel , RW , $Addr[1:0]$, and $Data[7:0]$ vary during a transient analysis of the 4B SRAM. The expected output values, $Out[7:0]$, are also shown. The transient analysis lasts 1 ms. Each period of 100 μ s is marked by a dotted vertical line.

Sel is kept low for the first period, which means that it is not possible to write or read from the SRAM and all outputs should therefore be low. This is done to allow the applied input values to propagate through the SRAM so that the 4B SRAM enters a legal state (e.g. no undefined values except for what is stored in the D latches). At the beginning of the second period Sel goes high, and since $RW = 1$ this means that the values of $Data[7:0]$ are written to the address given by $Addr[1:0]$. For the stimuli applied in this testbench, see Figure 31, this means that

F0 (or 11110000 in binary) is written to address 00 (i.e. word 0).

Sel goes low again at the start of the third period, but all other signals are kept constant to give the *Sel*-signal time to propagate through the circuit. At the beginning of the fourth period *Sel* is assumed to have propagated through the circuit so that it is safe to change *Data*[7 : 0] without accidentally overwriting the value stored in word 0. *Addr*[1 : 0] is changed at the beginning of the fifth period, and *Sel* is kept low to allow this change time to propagate through the 2x4 Decoder at the SRAM's input. *Sel* goes high at the start of the sixth period, which should result in the value of *Data*[7 : 0] to be written to the address given by *Addr*[1 : 0]. In other words should 0F (or 00001111 in binary) be written to address 10 (i.e. word 2).

RW = 1 for the first six periods of the transient analysis, so reading is disabled even when *Sel* is high. All the outputs are therefore expected to be 0 during this time. At the beginning of the seventh period *RW* goes low, signalling a read operation. As *Sel* is high, the values stored in address 10 (the address given by *Addr*[1 : 0]) should become visible at the outputs. This is used to verify that the read operation performed during the previous period succeeded. The outputs should then go back to 0 as *Sel* goes low at the start of the eighth period, as this means that both read and write is disabled.

The final two periods of the transient analysis are used to verify the success of the first write operation, which was performed during the second period. The address is therefore changed back to *Addr*[1 : 0] = 00 at the start of the ninth period, and *Sel* is kept low while this change is given time to propagate throughout the circuit. At the start of the tenth and final period, *Sel* goes high and a read operation is enabled as *RW* = 0. The value F0 (or 11110000 in binary), which was stored in word 0 in the read operation during the second period, should then become visible at the outputs.

Monte Carlo simulations are run as described in Subsection 4.4, testing with *VDD* = 70 mV, *VDD* = 75 mV, and *VDD* = 80 mV for the pre layout design, and *VDD* = 80 mV and *VDD* = 85 mV for the post layout simulations.

Process variation simulations are run with *VDD* = 70 mV, *VDD* = 75 mV, and *VDD* = 80 mV for the pre layout design, testing for all process corner and temperature combinations described in Subsection 4.5. Post layout simulations of process variation are run for the same corners, testing with *VDD* = 75 mV, *VDD* = 80 mV, and *VDD* = 85 mV.

4.12 16B SRAM Testbench

Stimuli in the form of square pulses is applied directly to the 16B SRAM's input during testing, and the outputs are sampled to verify the circuit's behaviour. An illustration of the 16B SRAM and the names of all inputs and outputs can be found in Figure 32.

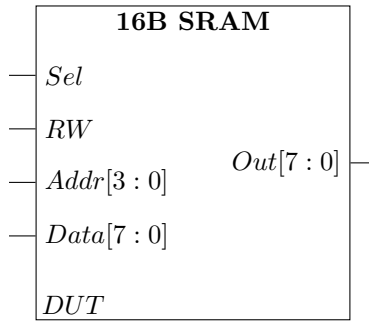


Figure 32: Block diagram of the 16B SRAM testbench, which consists only of the 16B SRAM. Input signals are applied as shown in Figure 33, and the values of the eight output signals are sampled.

A transient analysis is used as the base for all simulations on the 16B SRAM. It is divided into ten periods, as indicated by the dotted vertical lines in the timing diagram in Figure 33. The timing diagram shows the applied inputs and expected outputs for each stage of the transient analysis. The transient analysis follows the same pattern as the transient analysis used for the 4B SRAM in Subsection 4.11, only differing in the choice of addresses to read and write from. The explanation of the transient analysis given for the 4B SRAM in Subsection 4.11 is valid for 16B as well, and the reader is therefore encouraged to consult this explanation if the behaviour illustrated in the timing diagram in Figure 33 is unclear.

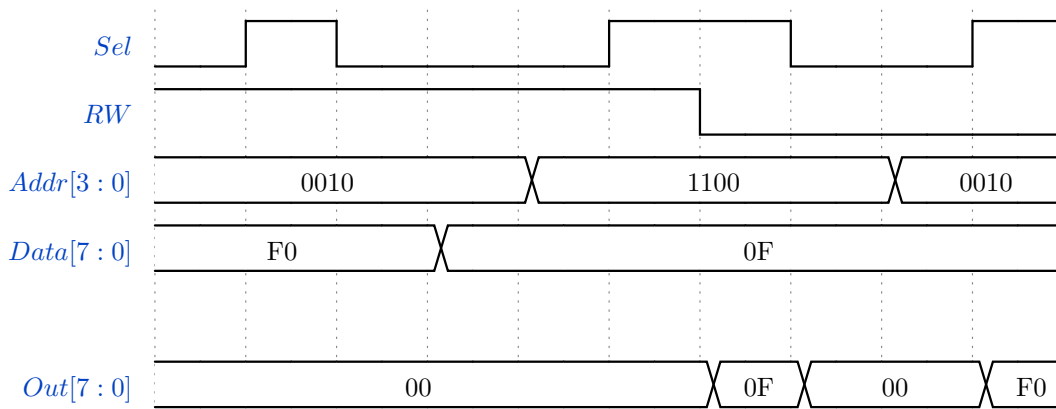


Figure 33: Timing diagram showing how the input signals Sel , RW , $Addr[3:0]$, and $Data[7:0]$ vary during a transient analysis of the 16B SRAM. The expected output values, $Out[7:0]$, are also shown. The transient analysis is divided into ten periods, where each period is marked by a dotted vertical line.

The two addresses used, $Addr[3:0] = 0010$ and $Addr[3:0] = 1100$, are chosen with some degree of care to ensure that different parts of the decoder circuitry is tested. The first two bits

($Addr[3]$ and $Addr[2]$) are inputs to the first 2x4 Decoder in the 16B SRAM. The first address ($Addr[3 : 0] = 0010$) sets $Out0$ high, while the second address ($Addr[3 : 0] = 1100$) sets $Out3$ high. Each of the outputs of this first decoder is connected to the EN -pin of another decoder, and the two addresses chosen will therefore test two of these four decoders. The two last bits of the address ($Addr[1]$ and $Addr[0]$) are inputs to the second level of decoders, and are different for each of the addresses so that more input combinations are tested during the analysis.

Monte Carlo simulations are run as described in Subsection 4.4, but the number of simulation points is reduced to 500 for the post layout simulations due to a long simulation time. The pre layout simulation, with 1000 Monte Carlo points, is run with $VDD = 75$ mV. The post layout simulations are run for $VDD = 80$ mV and $VDD = 85$ mV. The transient analysis is set to last 1500 μ s, with ten 150 μ s periods, for both the post layout simulation with $VDD = 80$ mV and the pre layout simulation with $VDD = 75$ mV. The post layout simulation with $VDD = 85$ mV is run for 750 μ s, with ten 75 μ s periods, in the hope that this will decrease the simulation time somewhat.

Process corner simulations are run with $VDD = 75$ mV, $VDD = 80$ mV, and $VDD = 85$ mV for the pre layout design, testing for all process corner and temperature combinations described in Subsection 4.5. The transient analysis is set to last 1500 μ s for these pre layout simulations, which means it has a period $p = 150$ μ s. Post layout simulations of process variation are run for the same corners, using a transient analysis lasting 3 ms (each period lasts 300 μ s) and testing with $VDD = 80$ mV and $VDD = 85$ mV.

4.13 64B SRAM Testbench

No additional circuitry is added to drive the 64B SRAM or to be used as load in the testbench, and stimuli is applied directly to the 64B SRAM shown in Figure 34. A transient analysis is used as the base for all simulations on the 64B SRAM. The stimuli is applied in the form of square pulses, and the 64B SRAM's outputs are then sampled to check that the circuit behaves correctly. The timing diagram in Figure 35 shows the applied inputs and the expected outputs for each stage of the analysis, which has been divided into ten periods as indicated by the dotted vertical lines in the timing diagram.

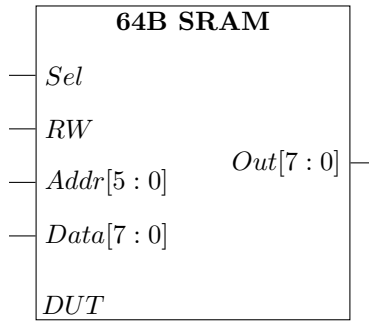


Figure 34: Block diagram of the 64B SRAM testbench, which consists only of the 64B SRAM cell. Input signals are applied as shown in Figure 35, and the values of the eight output signals are sampled.

The transient analysis follows the same pattern as the transient analysis used for the 4B SRAM in Subsection 4.11, only differing in the choice of addresses to read and write from. The explanation of the transient analysis given for the 4B SRAM in Subsection 4.11 holds for 64B as well, and the reader is therefore encouraged to consult this explanation if the behaviour illustrated in the timing diagram in Figure 35 is unclear.

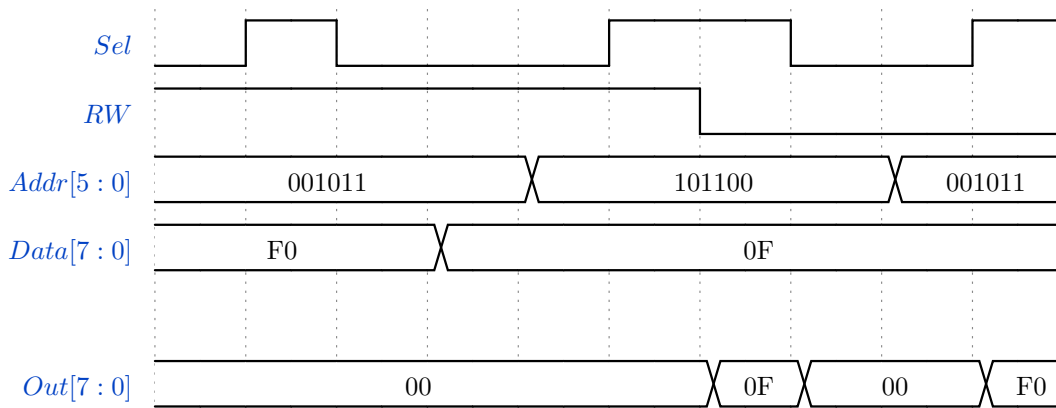


Figure 35: Timing diagram showing how the input signals Sel , RW , $Addr[5:0]$, and $Data[7:0]$ vary during a transient analysis of the 64B SRAM. The expected output values, $Out[7:0]$, are also shown. The transient analysis is divided into ten periods, where each period is marked by a dotted vertical line.

The two addresses used, $Addr[5:0] = 001011$ and $Addr[5:0] = 101100$, are chosen so that different parts of the decoder circuitry is tested during the transient analysis. The first two bits ($Addr[5]$ and $Addr[4]$) are inputs to the first 2x4 Decoder in the 64B SRAM. The first address ($Addr[5:0] = 001011$) sets $Out0$ high, while the second address ($Addr[5:0] = 101100$) sets $Out2$ high. Each of the outputs of this first decoder is connected to the EN -pin of another

decoder, each one placed at the input of a 16B cell, and the two addresses chosen will therefore test two of these four decoders. The two middle bits of the address ($Addr[3]$ and $Addr[2]$) are inputs to the second level of decoders, and are different for each of the addresses so that more input combinations are tested during the analysis. Each output of these second level decoders is connected to the input of another decoder, which is placed at the input of a 4B cell. These third level decoders have the two last address bits ($Addr[1]$ and $Addr[0]$) as inputs, which have also been chosen so that a different combination is tested for each of the addresses.

Process corner simulations are run with $VDD = 75$ mV, $VDD = 80$ mV, and $VDD = 85$ mV for the pre layout design, testing for all process corner and temperature combinations described in Subsection 4.5. The transient analysis is set to last 1500 μ s for the simulation at $VDD = 85$ mV, 2000 μ s for the simulation at $VDD = 80$ mV, and 2500 μ s for the simulation at $VDD = 75$ mV.

Post layout simulations of process variation are also run for the process corner and temperature combinations described in Subsection 4.5. The transient analysis is set to last 4 ms (each period lasts 400 μ s), and the simulation is run for both $VDD = 80$ mV and $VDD = 85$ mV.

When simulating on the other circuits, it was found that process variation restricted the minimum supply voltage more than the transistor mismatch. As Monte Carlo mismatch simulations are very time consuming for large circuits, as well as taking up a lot of disk space on the server, no such simulations are run for the 64B SRAM.

5 Results

Simulation results for both pre and post layout simulations are presented in this section. Results are presented for the NOT gate in Subsection 5.1 and the NAND gate in Subsection 5.2. D latch results are presented in Subsection 5.3, 2to4 Decoder results are presented in Subsection 5.4, and results from the simulations on the Output Selection Module are presented in Subsection 5.5. The final subsections present simulation results for the SRAM circuits: 4B SRAM in Subsection 5.6, 16B SRAM in Subsection 5.7, and 64B SRAM in Subsection 5.8.

5.1 NOT Simulation Results

Simulations were run on the NOT gate testbench as described in Subsection 4.6. The results of the Monte Carlo mismatch simulations are presented in Subsection 5.1.1, and process corner simulation results are presented in Subsection 5.1.2.

A DC operating point analysis was used as the base for all simulations, and the *DUT*'s input signal *In* and output signal *Out* were sampled. *In* is expected to be logic low, and *Out* is expected to be logic high.

5.1.1 Monte Carlo Mismatch Simulation Results for Different Supply Voltages

Both pre and post layout mismatch simulations were run on six different supply voltages: $VDD = 50$ mV, $VDD = 55$ mV, $VDD = 60$ mV, $VDD = 65$ mV, $VDD = 70$ mV, and $VDD = 75$ mV. The results for *In* are listed in Table 7, and the results for *Out* in Table 8.

The pre layout NOT has a decent yield for a supply voltage of 60 mV and above, with the lowest being 97.7% for *Out* with $VDD = 60$ mV. The lowest logic high voltage observed for *Out* at $VDD = 60$ mV is still relatively close to the lowest legal logic high value ($V_{H,min}$). The pre layout yield is degraded slightly for $VDD = 55$ mV, with a yield of 83.8% for *Out* at this supply voltage. Note that $V_{L,max,pre55} = 21.28$ mV is quite close to $V_{H,min,pre55} = 27.31$ mV at this supply voltage, which is undesirable.

A large increase in the pre layout yield degradation occurs when going from $VDD = 55$ mV to $VDD = 50$ mV, where the yield for *Out* is only 45.4 %. At $VDD = 50$ mV, $V_{H,min,pre50} = 18.66$ mV which is lower than $V_{L,max,pre50} = 23.76$ mV, which means there is an overlap between the logic high and logic low regions. This is detrimental as it is no longer possible to separate the logic low values from the logic high values.

The performance of the post layout NOT is notably worse, only presenting a good yield for $VDD = 75$ mV. The yield at $VDD = 70$ mV is above 90% for both *In* and *Out*, and can be classified as an acceptable yield. It is still possible to distinguish between logic low and logic high values at $VDD = 65$ mV, as $V_{L,max,post65} = 26.35$ mV and $V_{H,min,post65} = 32.79$ mV, but the yield has been notably degraded compared to $VDD = 70$ mV.

Table 7: Post and pre layout yield for the NOT gate at different supply voltages, based on logic low values observed at the input of the DUT (signal *In*).

VDD	$V_{L,max}$	Yield (pre)	$V_{L,max}$ (pre)	Yield (post)	$V_{L,max}$ (post)
75 mV	18.75 mV	100 %	7.767 mV	100 %	16.94 mV
70 mV	17.5 mV	100 %	10.2 mV	98.7 %	22.46 mV
65 mV	16.25 mV	100 %	13.45 mV	92.1 %	26.39 mV
60 mV	15 mV	99.9 %	17.41 mV	74.9 %	28.96 mV
55 mV	13.75 mV	95.8 %	21.28 mV	48.6 %	29.7 mV
50 mV	12.5 mV	80.1 %	23.76 mV	25.2 %	28.89 mV

Table 8: Post and pre layout yield for the NOT gate at different supply voltages, based on logic high values observed at the output of the DUT (signal *Out*).

VDD	$V_{H,min}$	Yield (pre)	$V_{H,min}$ (pre)	Yield (post)	$V_{H,min}$ (post)
75 mV	56.25 mV	100 %	64.11 mV	98.5 %	51.66 mV
70 mV	52.5 mV	100 %	56.73 mV	91.0 %	42.2 mV
65 mV	48.75 mV	99.9 %	48.62 mV	60.3 %	32.79 mV
60 mV	45 mV	97.7 %	39.73 mV	27.3 %	24.63 mV
55 mV	41.25 mV	83.8 %	27.31 mV	7.8 %	18.77 mV
50 mV	37.5 mV	45.4 %	18.66 mV	1.6 %	14.85 mV

For supply voltages of 60 mV and below, the post layout yield is significantly degraded and $V_{L,max,post} > V_{H,min,post}$ which makes it impossible to separate logic high and logic low values. The yield found by observing *Out* is the worst, with a yield of only 27.3% for $VDD = 60$ mV and a yield of less than 10% for $VDD = 55$ mV and $VDD = 50$ mV.

5.1.2 Process Corner Simulation Results

Process variation was simulated with $VDD = 70$ mV for both the pre and post layout simulations. The upper limit of a logic low value is therefore $V_{L,max} = 0.25VDD = 17.5$ mV and $V_{H,min} = 0.75VDD = 52.5$ mV.

The logic low value observed for *In* and the logic high value observed for *Out* is listed for each of the pre layout corners in Table 9, together with FAIL/PASS information. All corners are found to pass, with SF50 producing the highest logic low value ($V_{L,SF50} = 10.17$ mV) and FS50 producing lowest logic high value ($V_{H,FS50} = 59.35$ mV). The lowest logic low value is found for FS0, $V_{L,FS0} = 2.582$ mV, and SS0 has the highest logic high value with $V_{H,SS0} = 65.82$ mV.

The logic low value observed for *In* and the logic high value observed for *Out* for each post layout corner is listed in Table 9, together with FAIL/PASS information. All corners are found to pass, with SF50 producing both the highest logic low value ($V_{L,SF50} = 17.39$ mV) and lowest logic high value ($V_{H,SF50} = 52.95$ mV). The lowest logic low value is found for FS0, $V_{L,FS0} = 3.373$ mV, and FF0 has the highest logic high value with $V_{H,FF0} = 61.78$ mV.

Table 9: The operating point value for the DUT’s input signal In and output signal Out for all pre layout process corners simulated at $VDD = 70$ mV. The NOT gate passes for all corners.

Process corner (pre)	In	Out	PASS/FAIL
SS0	2.954 mV	65.82 mV	PASS
SS20	3.805 mV	64.64 mV	PASS
SS27	4.184 mV	64.17 mV	PASS
SS50	5.421 mV	62.42 mV	PASS
TT0	3.151 mV	65.78 mV	PASS
TT20	4.06 mV	64.57 mV	PASS
TT27	4.421 mV	64.09 mV	PASS
TT50	5.754 mV	62.32 mV	PASS
FF0	3.296 mV	65.7 mV	PASS
FF20	4.257 mV	64.44 mV	PASS
FF27	4.633 mV	63.94 mV	PASS
FF50	6.007 mV	62.11 mV	PASS
FS0	2.582 mV	64.4 mV	PASS
FS20	3.336 mV	62.63 mV	PASS
FS27	3.635 mV	61.93 mV	PASS
FS50	4.738 mV	59.35 mV	PASS
SF0	6.114 mV	65.78 mV	PASS
SF20	7.584 mV	64.6 mV	PASS
SF27	8.153 mV	64.14 mV	PASS
SF50	10.17 mV	62.43 mV	PASS

By comparing the post layout results in Table 10 to the pre layout results in Table 9, it becomes clear that there is a larger logic level degradation in the post layout design.

5.2 NAND Simulation Results

The results from the simulations run on the NAND gate testbench described in Subsection 4.7 are presented here. Monte Carlo mismatch simulation results are presented in Subsection 5.2.1, and process corner simulation results in Subsection 5.2.2.

5.2.1 Monte Carlo Mismatch Simulation Results

A DC operating point analysis was used as the base for the mismatch simulations, and the DUT’s output signal Y was sampled. The DC operating analysis was repeated for all four possible combinations of input signals A and B , as listed in Table 6.

Both the pre and post layout Monte Carlo mismatch simulations were run for $VDD = 70$ mV and $VDD = 75$ mV.

Table 10: The operating point value for the DUT's input signal In and output signal Out for all post layout process corners simulated at $VDD = 70$ mV. The NOT gate passes for all corners.

Process corner (post)	In	Out	PASS/FAIL
SS0	5.883 mV	60.58 mV	PASS
SS20	7.336 mV	58.23 mV	PASS
SS27	7.913 mV	57.33 mV	PASS
SS50	9.949 mV	54.14 mV	PASS
TT0	6.017 mV	61.16 mV	PASS
TT20	7.547 mV	58.87 mV	PASS
TT27	8.14 mV	57.99 mV	PASS
TT50	10.22 mV	54.88 mV	PASS
FF0	5.93 mV	61.78 mV	PASS
FF20	7.511 mV	59.56 mV	PASS
FF27	8.111 mV	58.71 mV	PASS
FF50	10.21 mV	55.68 mV	PASS
FS0	3.373 mV	60.97 mV	PASS
FS20	4.424 mV	58.65 mV	PASS
FS27	4.848 mV	57.76 mV	PASS
FS50	6.419 mV	54.55 mV	PASS
SF0	13.34 mV	59.51 mV	PASS
SF20	14.96 mV	56.99 mV	PASS
SF27	15.54 mV	56.07 mV	PASS
SF50	17.39 mV	52.95 mV	PASS

The results from the pre layout simulations are listed in Table 11. Two of the simulated points fail for the input combination $AB = 11$ at $VDD = 70$ mV, as the value produced for Y is higher than $V_{L,max} = 0.25VDD = 17.5$ mV. $Y_{max,AB=11} = 19.51$ mV at this supply voltage, so the two failing points are relatively close to the legal range. No failing points are found for the other input combinations at $VDD = 70$ mV, and no failing points are found for any input combination at $VDD = 75$ mV.

Table 11: Results from the pre layout Monte Carlo mismatch simulations on the NAND for both $VDD = 70$ mV and $VDD = 75$ mV. The yield, maximum value of Y , and minimum value of Y are given for each input combination.

VDD	A	B	Yield	Simulated Y_{min}	Simulated Y_{max}
70 mV	0 mV	0 mV	100 %	63.60 mV	68.81 mV
70 mV	0 mV	70 mV	100 %	54.23 mV	67.48 mV
70 mV	70 mV	0 mV	100 %	54.47 mV	66.89 mV
70 mV	70 mV	70 mV	99.8 %	4.617 mV	19.51 mV
75 mV	0 mV	0 mV	100 %	69.71 mV	74.02 mV
75 mV	0 mV	75 mV	100%	61.94 mV	72.90 mV
75 mV	75 mV	0 mV	100 %	62.37 mV	72.39 mV
75 mV	75 mV	75 mV	100 %	3.677 mV	14.61 mV

The results from the post layout simulations are listed in Table 12. There are several failing simulation points for all input combinations at $VDD = 70$ mV. The input combination $A = B = 0$ produces the best results, while the other three generate many logic values in the middle of the illegal range. The maximum logic low value, $Y_{max,AB=11} = 39.26$ mV, is higher than the minimum values for the input combinations $AB = 10$ and $AB = 01$, and it is therefore impossible to distinguish between logic low and logic high values. As such, the post layout NAND can not be said to work for $VDD = 70$ mV.

Table 12: Results from the post layout Monte Carlo mismatch simulations on the NAND for both $VDD = 70$ mV and $VDD = 75$ mV. The yield, maximum value of Y , and minimum value of Y are given for each input combination.

VDD	A	B	Yield	Simulated Y_{min}	Simulated Y_{max}
70 mV	0 mV	0 mV	99.8 %	51.02 mV	66.17 mV
70 mV	0 mV	70 mV	73.6 %	35.20 mV	63.34 mV
70 mV	70 mV	0 mV	86.6 %	38.75 mV	64.67 mV
70 mV	70 mV	70 mV	58.5 %	7.552 mV	39.26 mV
75 mV	0 mV	0 mV	100 %	58.92 mV	71.79 mV
75 mV	0 mV	75 mV	91.6 %	43.19 mV	69.19 mV
75 mV	75 mV	0 mV	97.6 %	47.71 mV	70.41 mV
75 mV	75 mV	75 mV	91.9 %	6.015 mV	32.42 mV

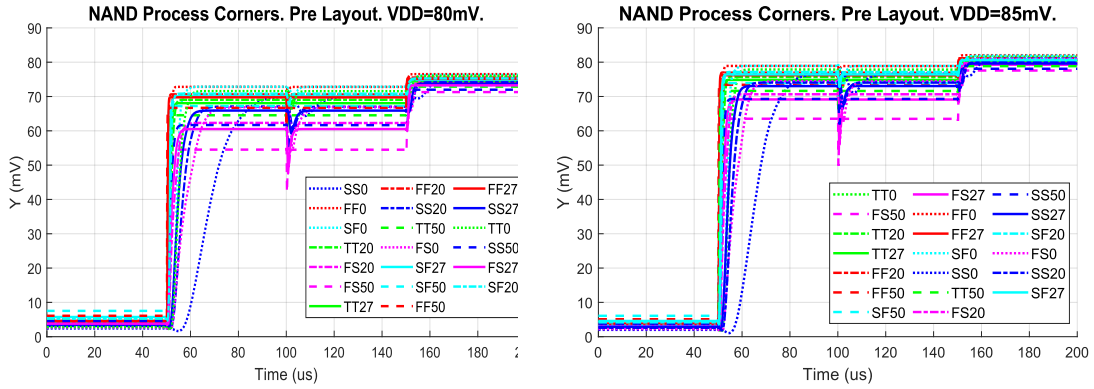
The yield is greatly improved for the post layout simulations with $VDD = 75$ mV, with yield above 90% for all input combinations. No failing points are found for the input combination $AB = 00$. Even though there are failing points for the three other input combinations, they are not critical as the maximum logic low value ($Y_{max,AB=11} = 32.42$ mV) is lower than the minimum logic high value ($V_{min,AB=01} = 43.19$ mV). As for the simulations at $VDD = 70$ mV, the input combination $AB = 11$ presents the worst yield. This was also the case for the pre layout simulations, as seen in Table 11.

5.2.2 Process Variation Simulation Results

A transient analysis was used as the base for the process corner simulations, and the DUT 's output signal Y was sampled. All four NAND input combinations are tested during the analysis in the following order: $AB = 11$, $AB = 10$, $AB = 01$, $AB = 00$.

Pre Layout Simulations

The NAND's output Y is plotted for all pre layout process corners simulated with $VDD = 80$ mV in Figure 36a and with 85 mV in Figure 36b. The logic high level is lower when $AB = 10$ or $AB = 01$ than when it is $AB = 00$, which is expected. The spike at 100 μ s is caused by the switching from $AB = 10$ to $AB = 01$. FS50 fails for $VDD = 80$ mV since the logic high value produced for $AB = 10$ and $AB = 01$ is lower than $V_{H,min} = 0.75VDD = 60$ mV. All corners pass for $VDD = 85$ mV.



(a) FS50 fails for $VDD = 80 \text{ mV}$.

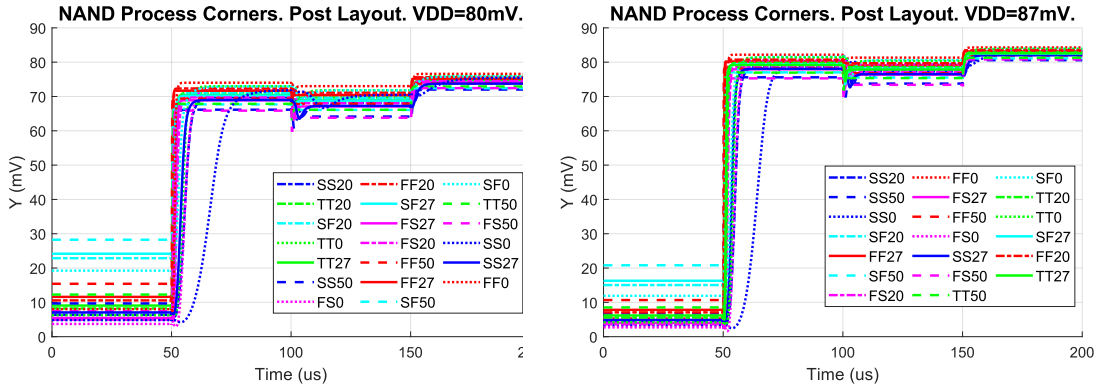
(b) All corners pass for $VDD = 85 \text{ mV}$.

Figure 36: The NAND's output Y plotted for all pre layout corners.

Post Layout Simulations

Y is plotted for all post layout process corners simulated with $VDD = 80 \text{ mV}$ in Figure 37a and with 87 mV in Figure 37b. As expected, the logic high level is lower when $AB = 10$ or $AB = 01$ than when it is $AB = 00$. The logic high values produced by both $AB = 10$ and $AB = 01$ have improved compared to the pre layout design. For the pre layout corners $AB = 10$ and $AV = 01$ produced equally good logic high values, but as can be seen from Figure 37b this is not the case post layout. $AB = 10$ performs slightly better than $AB = 01$. The logic high values produced by both $AB = 10$ and $AB = 01$ has improved compared to the pre layout design.

A plot of Y for the post layout corners at $VDD = 85 \text{ mV}$ is included in Subsection A. For the post layout NAND it is the SF50 corner that has the worst performance, causing fails for both $VDD = 80 \text{ mV}$ and $VDD = 85 \text{ mV}$, and all logic low values are in general worse than for the pre layout design. All corners pass for $VDD = 87 \text{ mV}$.



(a) $SF50$, $SF27$, and $SF20$ fail for $V_{DD} = 80$ mV. (b) All corners pass for $V_{DD} = 87$ mV.

Figure 37: The NAND's output Y plotted for all post layout corners.

5.3 D Latch Results

The results from simulations run on the D latch testbench, as described in Subsection 4.8, will be presented in this subsection. Expected output of the transient simulation is illustrated in the timing diagram in Figure 25b. The output is expected to go high at the start of the simulation, and remain at a logic high value until $t = 0.5t_{total}$, where t_{total} is the total duration of the transient analysis. Out should then go low at $t = 0.5t_{total}$ and remain at a logic low value until the analysis ends at $t = t_{total}$.

5.3.1 Monte Carlo Mismatch Simulation Results

For all Monte Carlo mismatch simulations, the transient analysis was set to have a period $p = 12.5 \mu\text{s}$ and a total runtime $t_{total} = 100 \mu\text{s}$.

Pre Layout Simulation Results

Pre layout simulations of mismatch were first run for $V_{DD} = 70$ mV, and plots of the D latch's output signal Out for all 1000 Monte Carlo points is included in Figure 38. All points pass, giving a tentative yield of 100%.

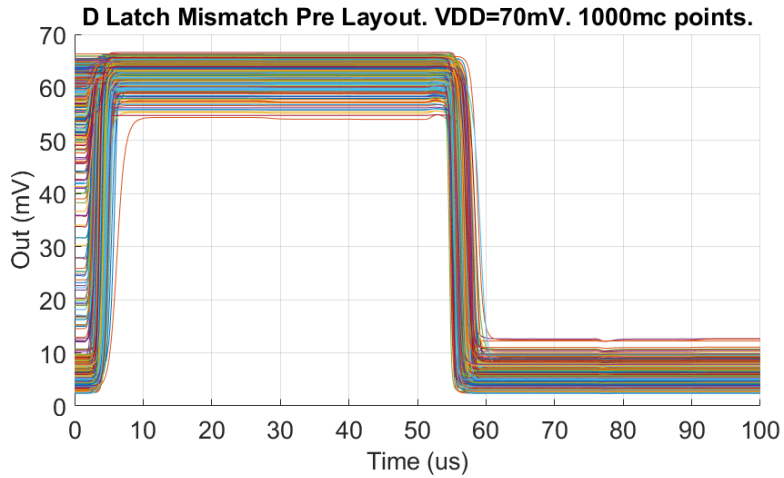


Figure 38: The D latch's output signal *Out* plotted for the 1000 Monte Carlo points simulated with $VDD = 70 \text{ mV}$.

The pre layout Monte Carlo simulation was then repeated for $VDD = 65 \text{ mV}$. The *DUT*'s output signal *Out* is plotted in Figure 39 for all 1000 Monte Carlo simulation points, and it is clear that the D latch fails for several of these simulation points.

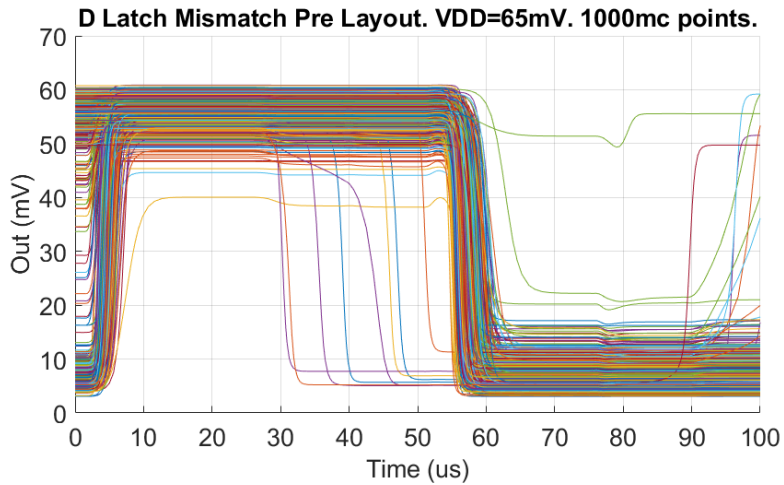


Figure 39: The D latch's output signal *Out* plotted for the 1000 Monte Carlo points simulated with $VDD = 65 \text{ mV}$.

A closer look at the 20 simulation points failing in the first half of the transient analysis is given in Figure 40a. *Out* goes low between $t = 25 \mu\text{s}$ and $t = 50 \mu\text{s}$ for seven of the simulation points. This is caused by *EN* going low at $t \approx 25 \mu\text{s}$, which should disable writing to the D-latch,

followed by In going low at $t = 37.5 \mu\text{s}$. Of the seven, the first three go low while In is still high. For all seven, this shows a lack of ability to retain the value stored. The remaining 13 failing points all display the correct behaviour, by holding the value until $EN = 1$ at $t = 50 \mu\text{s}$ and the new value (logic 0) is stored, but fail because the logic high value they hold is below $V_{H,min} = 0.75V_{DD} = 48.75 \text{ mV}$.

18 simulation points fail in the second half of the transient analysis, and a closer look at these is given in Figure 40b. For one of the failing points Out remains high the entire time, which is an obvious fail as this means a logic low value could not be written to or stored in the latch. The remaining 17 failing points all go low after writing is enabled by EN going high at $t = 50 \mu\text{s}$. As was the case for the failing points in the first half, some of the points in the second half fail simply by settling at a logic value outside the legal limits (i.e. a logic low value larger than $V_{L,max} = 0.25V_{DD} = 16.25 \text{ mV}$) while others fail by going high again when In goes high at $t = 87.5 \mu\text{s}$. While all these are classified as fails, the point that fails to go low and the points that go high again are the most critical errors as this means the D latch no longer functions as a D latch.

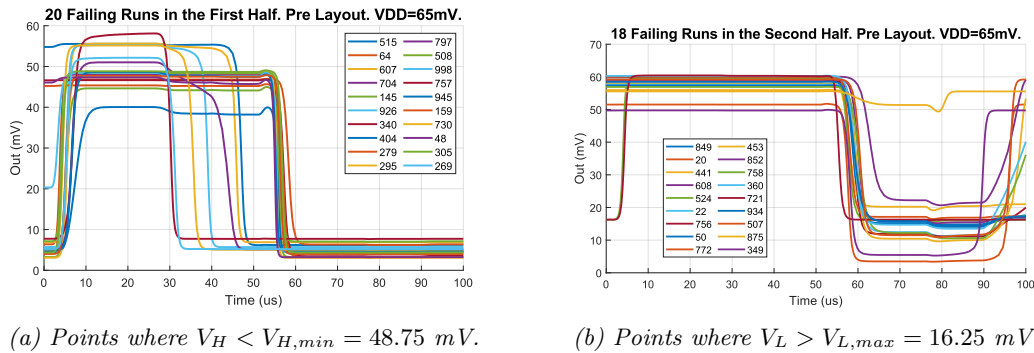


Figure 40: Plot of the DUT's output Out for all Monte Carlo points that failed the pre layout simulation with $V_{DD} = 65 \text{ mV}$.

Post Layout Simulation Results

A post layout Monte Carlo simulation was run for $V_{DD} = 80 \text{ mV}$, and Out is plotted in Figure 41 for all 1000 simulation points. Out shows the correct overall behaviour for all points by going high and holding a high value for the first half of the analysis, and then going low and remaining low for the second half.

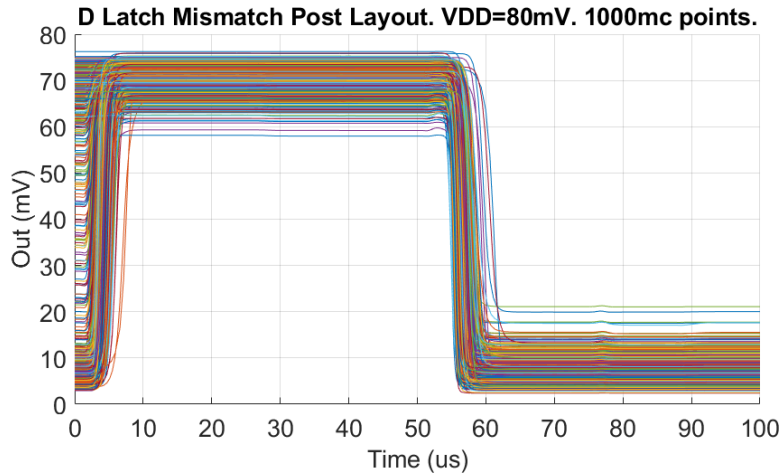
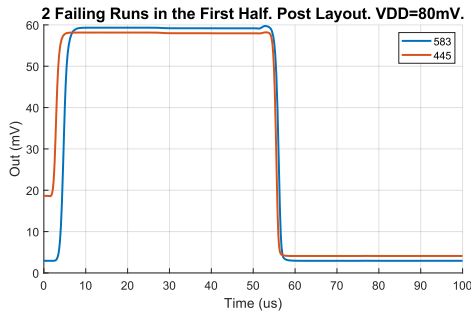
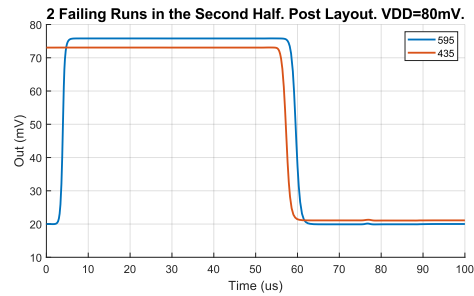


Figure 41: The D latch's output signal Out plotted for the 1000 Monte Carlo points simulated with $VDD = 80$ mV.

Two of the simulation points fail in the first half of the analysis, and these are plotted again in Figure 42a. Simulation point 445 settles at a logic high value $V_{H,445} \approx 58$ mV, and simulation point 583 settles at $V_{H,583} \approx 59$ mV. The yield for the D latch holding a logic high value at $VDD = 80$ mV is 99.8%.



(a) Points where $V_H < V_{H,min} = 60$ mV.



(b) Points where $V_L > V_{L,max} = 20$ mV. NB! The y-axis starts at 10 mV.

Figure 42: Plot of the DUT's output Out for all Monte Carlo points that failed the post layout simulation with $VDD = 80$ mV.

There are two failing simulation points in the second half of the analysis as well, and these are plotted in Figure 42b. The yield for the D latch holding a logic low value at $VDD = 80$ mV is therefore 99.8%. Simulation point 435 settles at a logic low value $V_{L,435} \approx 21$ mV, and simulation point 595 settles at $V_{L,595} \approx 20$ mV. It is the small bump at $t \approx 77$ μ s that causes 595 to exceed $V_{L,max} = 0.25VDD = 20$ mV. Overall yield for the D latch is estimated to 99.6%,

since there were 4 unique failing points.

5.3.2 Process Variation Simulation Results

Pre Layout Simulation Results

Pre layout simulations were run for both $VDD = 70$ mV and $VDD = 75$ mV. Since all corners were found to pass for both supply voltages, only the results from the simulation with the lowest supply voltage ($VDD = 70$ mV) are presented here. A figure showing the results for the pre layout process variation simulation at $VDD = 75$ mV is included in Subsection B.

The results from the pre layout simulation with $VDD = 70$ mV are shown in Figure 43. The transient analysis lasted $t_{total} = 200$ μ s, and Out should therefore go low at $t = 100$ μ s. As the supply voltage is 70 mV, $V_{L,max} = 17.5$ mV and $V_{H,min} = 52.5$ mV (see Table 5). From Figure 43 it is clear that all process corners pass this requirement, with $V_{H,FS50} \approx 57.5$ mV as the lowest logic high value and $V_{L,SF50} \approx 13.5$ mV as the highest logic low value.

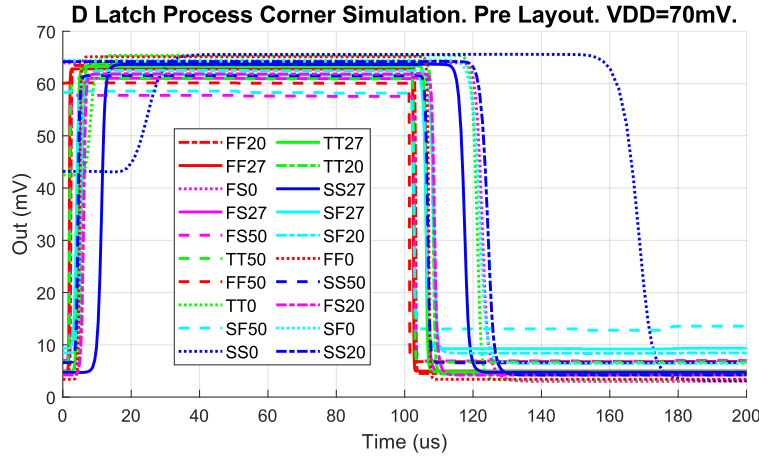


Figure 43: The pre layout D latch process variation results for $VDD = 70$ mV. FS50 has the lowest logic high value, and SF50 the highest logic low value.

Post Layout Simulation Results

Figure 44 shows the results for the post layout simulation at $VDD = 75$ mV, where $t_{total} = 400$ μ s. The SF corner struggles to reach a legal logic low value, but manages to reach a value less than $0.25VDD$ at both 0°C and 20°C . The logic low value for SF27 remains at approximately 20 mV, which is higher than the upper limit for logic low (18.75 mV for $VDD = 75$ mV). SF27 is therefore classified as a fail. SF50 fails more critically, by first settling at a logic low value of approximately 26.5 mV and then going high at $t = 350$ μ s. This is triggered by In going high at $t = 350$ μ s, but should not have affected the D-latch since EN is kept low.

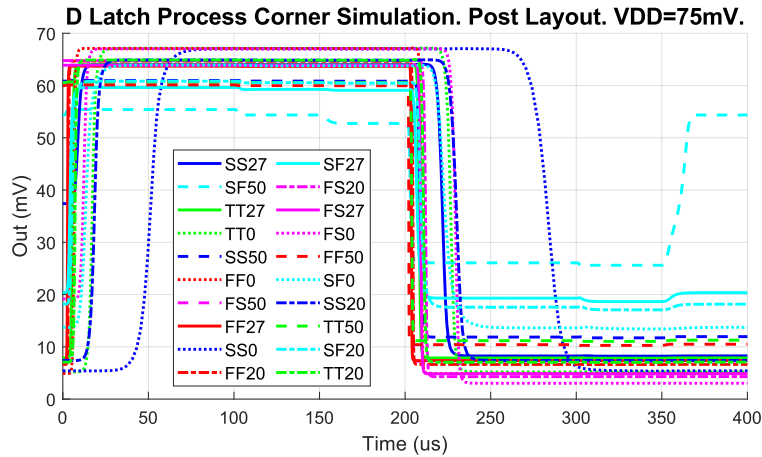


Figure 44: The post layout D latch process variation results for $V_{DD} = 75$ mV. SF50 and SF27 fail.

For the post layout simulation at $V_{DD} = 80$ mV, a shorter transient analysis was run ($t_{total} = 200$ μ s). To pass, all logic high values must be above 60 mV and all logic low values below 20 mV (see Table 5). As seen from Figure 45, $V_{H,SF50} \approx 64.5$ mV is the lowest logic high value and $V_{L,SF50} \approx 19$ mV is the highest logic low value, meaning that all corners pass.

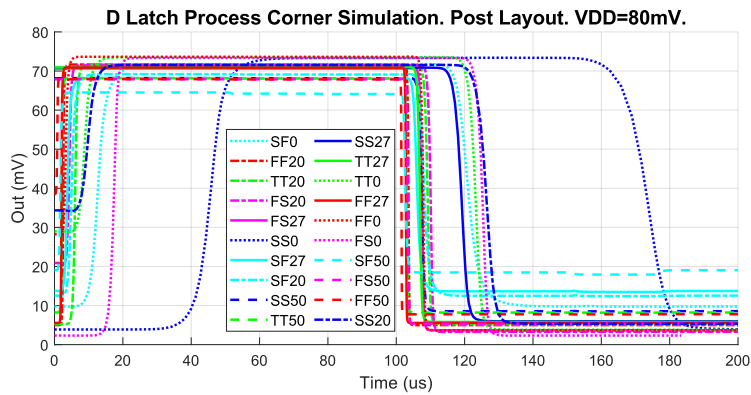


Figure 45: The post layout D latch's output signal Out0 plotted for all process corners at $V_{DD} = 80$ mV. All corners pass.

5.4 Decoder Results

The Decoder testbench is described in Subsection 4.9, together with an explanation of the simulations that were run. The results of the Monte Carlo mismatch simulations will be presented

in Subsection 5.4.1, and results from the process corner simulation are presented in Subsection 5.4.2.

A transient analysis was used as the basis for all these simulations, and a timing diagram showing the applied stimuli and expected outputs can be found in Figure 27.

5.4.1 Monte Carlo Mismatch Simulation Results

The transient simulation was set to last $t_{total} = 400 \mu\text{s}$ for all Monte Carlo mismatch simulations.

Pre Layout Simulation Results

Pre layout mismatch simulations were first run for $VDD = 70 \text{ mV}$. While *Out0*, *Out1*, *Out2*, and *Out3* all maintained a correct general behaviour, going low and high as intended, some of the simulation points only managed to produce logic high values in the range of 45 mV to 52.5 mV. These points were classified as failures, since the logic high values were $V_{H,min} = 0.75VDD = 52.5 \text{ mV}$. There were in total 17 failing points: 6 failing points for *Out0*, 7 failing points for *Out1*, 3 failing points for *Out2*, and 1 failing point for *Out3*. All 17 points were unique (no simulation point caused a failure for more than one output signal), which gives a yield of 98.3% for the pre layout Decoder at $VDD = 70 \text{ mV}$.

Plots of *Out0*, *Out1*, *Out2*, and *Out3* for all the 1000 Monte Carlo points simulated are included in Subsection C, as well as plots of the failing simulation points for each output signal.

The pre layout mismatch simulation was then repeated for $VDD = 75 \text{ mV}$, which resulted in no failed simulation points and an estimated yield of 100%. Plots of *Out0*, *Out1*, *Out2*, and *Out3* for all 1000 pre layout Monte Carlo points simulated at $VDD = 75 \text{ mV}$ are included in Subsection C.

Post Layout Simulation Results

A post layout mismatch simulation was first run for $VDD = 80 \text{ mV}$. All logic low values were below $V_{L,max} = 0.25VDD = 20 \text{ mV}$ for all output signals, but some of the logic high values were not high enough. This resulted in 10 failing points for *Out0*, plotted in Figure 46a, 14 failing points for *Out1*, see Figure 46b, 6 failing points for *Out2*, plotted in Figure 46c, and 8 failing points for *Out3*, see Figure 46d. Except for simulation point 755, which is shared between *Out0* and *Out2*, all the failing points are unique. This means that there is a total of 37 failing points during the post layout simulations at $VDD = 80 \text{ mV}$, and the yield can be estimated to 96.3%.

Plots of *Out0*, *Out1*, *Out2*, and *Out3* for all 1000 Monte Carlo simulation points are included in Subsection C.

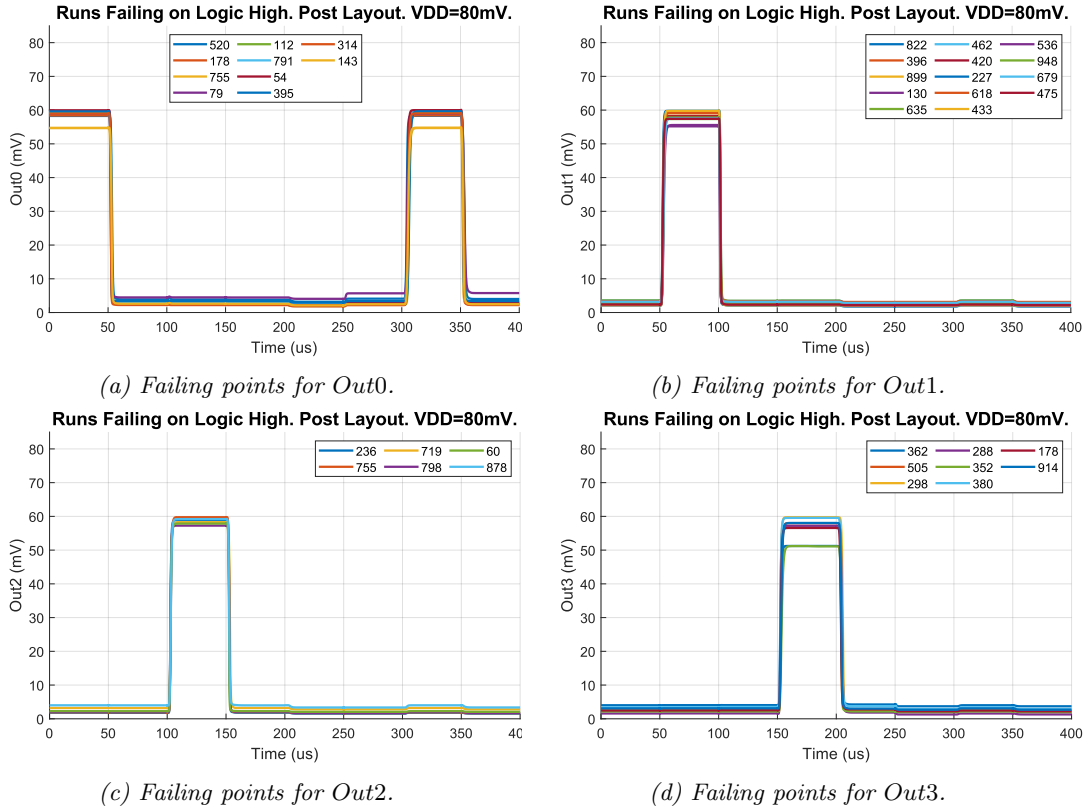


Figure 46: Each of the DUT's output signals plotted for the failing post layout simulation points with $VDD = 80\text{ mV}$.

The post layout Monte Carlo mismatch simulation was then repeated using $VDD = 85\text{ mV}$. Simulation point 362 failed to produce a logic high value above $V_{H,min} = 0.75VDD = 63.75\text{ mV}$ for *Out3*, but this was also the only failing simulation point. The estimated yield is therefore 99.9%. *Out3* is plotted for the failing point in Figure 47. Plots of the DUT's output signals for all the 1000 simulated points are included in Subsection C.1.

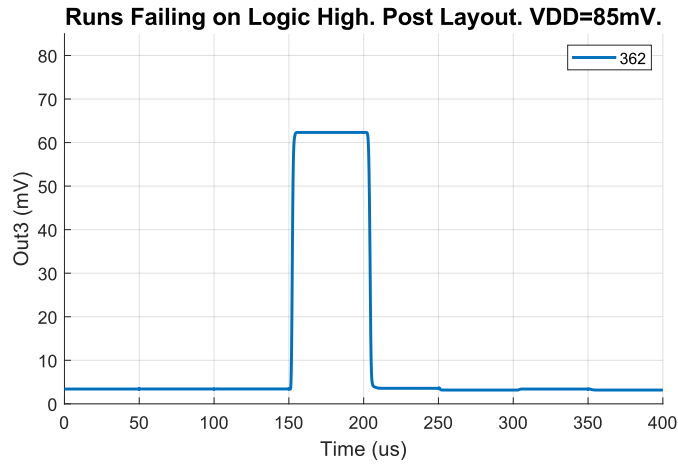


Figure 47: A plot of *Out3* for the only failing point in the post layout Monte Carlo mismatch simulations on the decoder when $VDD = 85$ mV. The logic high value is marginally below $V_{H,min} = 63.75$ mV.

5.4.2 Process Variation Simulation Results

Pre Layout Simulation Results

The transient analysis used for these pre layout process corner simulations has a period $p = 50$ μ s and a total duration of $t_{total} = 400$ μ s.

Pre layout simulations were first run with $VDD = 70$ mV, and the resulting *Out0*-plots are shown in Figure 48. The SF50 corner fails to produce a high enough logic high value ($V_{H,min} = 52.5$ mV for $VDD = 70$ mV). The same failure occurs for *Out1*, *Out2*, and *Out3*, and plots of these are Subsection C.2.

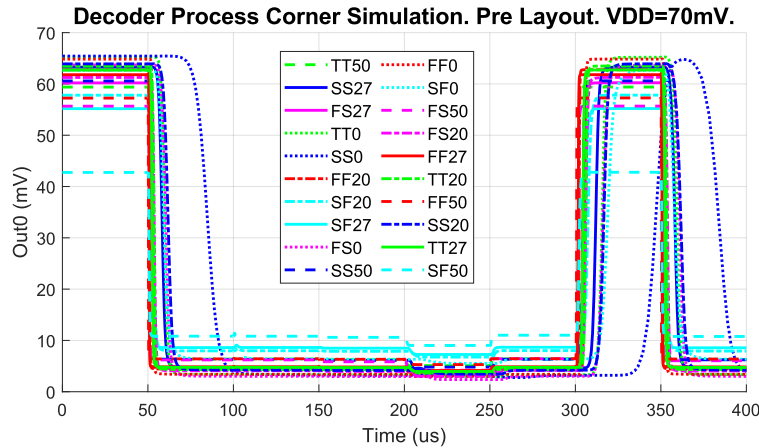


Figure 48: The DUT's output signal *Out0* plotted for pre layout process corners at $VDD = 70$ mV.

The pre layout simulation was repeated for $VDD = 75$ mV, which resulted in all corners passing. *Out0* is plotted in Figure 49a, *Out1* is plotted in Figure 49b, *Out2* is plotted in Figure 49c, and *Out3* is plotted in Figure 49d.

The transition from high to low takes longer time to start for *Out3*, see Figure 49d, than for the other signals. This transition is caused by *In3*, the input to the driving Decoder, going high so that *Sel* goes low and the DUT is disabled. This change in *In3* must propagate through the driving Decoder (see Figure 26) before the change is visible at the input of the DUT, so it makes sense that this takes longer time than when one of the inputs applied directly to the DUT (i.e. *I0* or *I1*) change. This added delay is also present at the rising edge of the second pulse of *Out0* in Figure 49a.

The falling edge of the second pulse in *Out0* is triggered by the fall of *EN*, which is an input to the driving Decoder. *EN* falling will lead to *Sel* falling, thus disabling the DUT. The enable input is connected directly to the second column of NANDs in the Decoder, see Figure 9, and the change in *EN* must therefore only propagate through the second halves of each Decoder before *Out0* falls. This explains why the delay before the second pulse of *Out0* goes low in Figure 49a is not as long as the delay seen at the falling edge of *Out3* in Figure 49d.

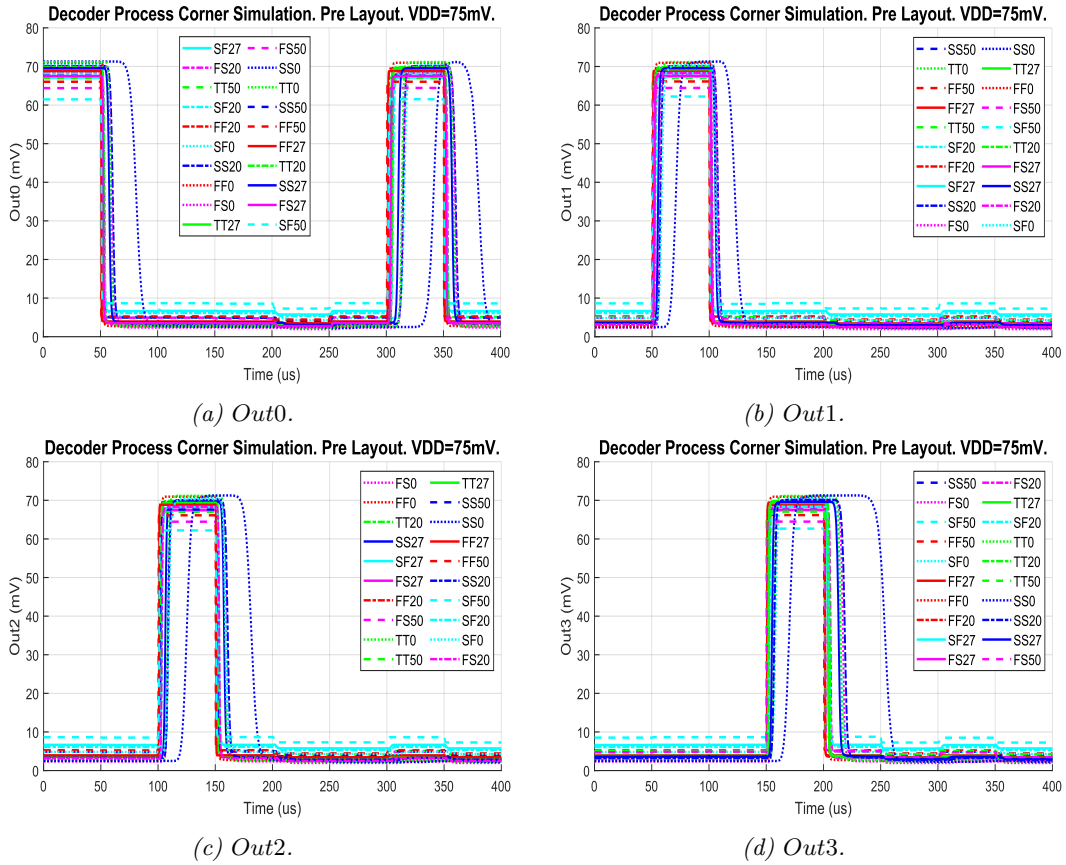


Figure 49: The DUT’s output signals plotted for all pre layout corners at $VDD = 75 \text{ mV}$.

Post Layout Simulation Results

The transient analysis used for these pre layout process corner simulations has a period $p = 100 \text{ }\mu\text{s}$ and a total duration of $t_{total} = 800 \text{ }\mu\text{s}$.

Post layout process corner simulations were first run on $VDD = 75 \text{ mV}$ and $VDD = 80 \text{ mV}$. SF50 fails to produce a logic high value above $V_{H,min}$ for both supply voltages, and the same goes for several other corners at $VDD = 75 \text{ mV}$. Plots of Out_0 , Out_1 , Out_2 , and Out_3 for these two failing process corner simulations can be found in Subsection C.

All corners passed when running the simulation with $VDD = 85 \text{ mV}$. Out_0 is plotted in Figure 50a, Out_1 is plotted in Figure 50b, Out_2 is plotted in Figure 50c, and Out_3 is plotted in Figure 50d.

The same increase in delay observed at the falling edge of the Out_3 -pulse in Figure 49d pre layout is observed for this post layout simulation. This can be seen when comparing the delay

before *Out3* goes low in Figure 50d to e.g. the delay before *Out2* goes low in Figure 50c. As explained for the pre layout simulation above, this is because the critical path for the input signal causing the falling edge of *Out3* is longer.

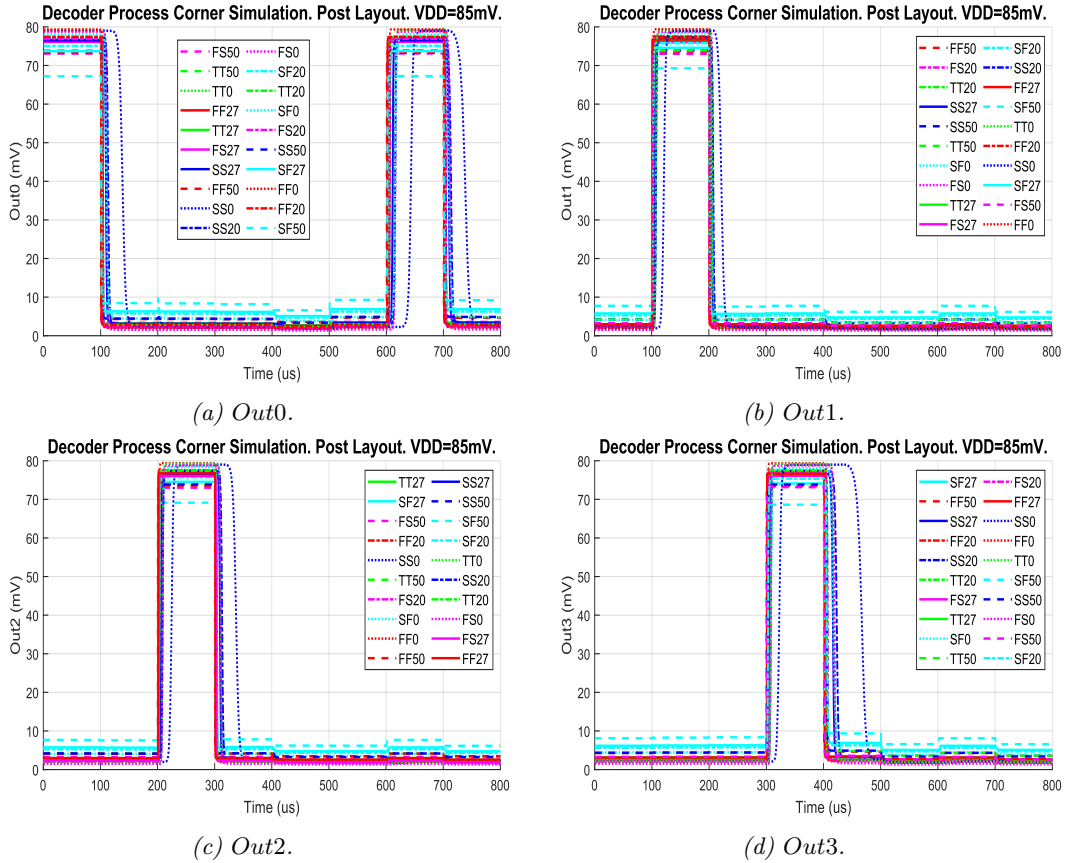


Figure 50: The DUT's output signals plotted for all post layout corners at $VDD = 85 \text{ mV}$.

5.5 Output Selection Module Results

The Output Selection Module was tested by running process variation simulations and Monte Carlo mismatch simulation as described in Subsection 4.10. The results of these simulations, both pre layout and post layout, will be presented here.

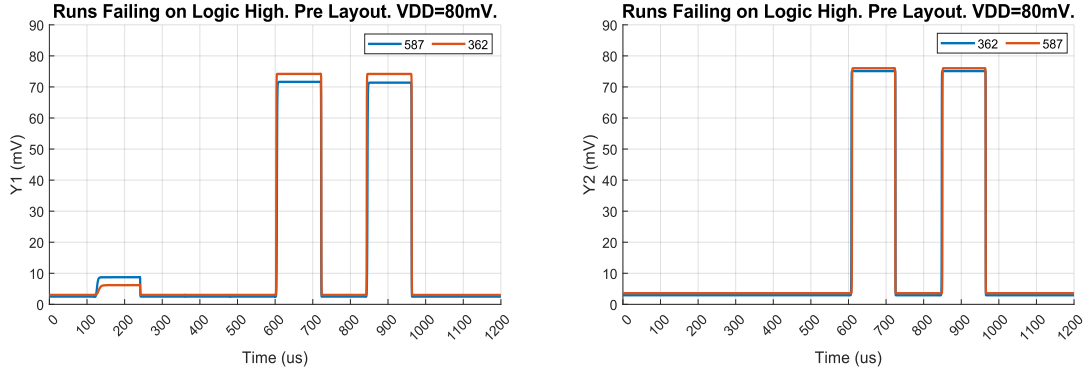
The expected values for the DUT's output signal $Y1$ and the LOAD's output signal $Y2$ can be found in the transient analysis timing diagram in Figure 29. The transient analysis lasts 1200 μs and is split into 10 periods, where the output signals $Y1$ and $Y2$ are expected to be logic high during the second, sixth, and eighth period, and logic low the rest of the time.

5.5.1 Monte Carlo Mismatch Simulation Results

Pre Layout Simulation Results

Pre layout Monte Carlo mismatch simulations were run for two different supply voltages: $VDD = 80$ mV and $VDD = 85$ mV. All simulated points passed for $VDD = 85$ mV, and plots of $Y1$ and $Y2$ for the 1000 simulated points at this supply voltage are included in Subsection D.

Plots of $Y1$ and $Y2$ for all 1000 pre layout points simulated with $VDD = 80$ mV are included in Subsection D.1. A couple of the simulation points failed, and plots of $Y1$ and $Y2$ for these failing points are shown in Figure 51. $Y1$ and $Y2$ fail for the same two points, 362 and 587, which is expected since $Y1$ is used as input to the Output Selection Module that creates $Y2$. The signals should have gone high during the second period (from $t = 120$ μ s to $t = 240$ μ s), but $Y1$ only has a very slight increase in value and $Y2$ remains constant. The estimated pre layout yield at $VDD = 80$ mV is therefore 99.8%.



(a) The DUT's output $Y1$. The pulse in the second period is very small.

(b) $Y2$ (LOAD) doesn't go high at $t = 120$ μ s.

Figure 51: Output Selection Module. $Y1$ and $Y2$ are plotted for the failing pre layout Monte Carlo simulation points at $VDD = 80$ mV. Both signals fail for simulation points 362 and 587.

Post Layout Simulation Results

Post layout mismatch simulations were run for both $VDD = 80$ mV and $VDD = 85$ mV. All simulated points passed for $VDD = 85$ mV, and plots of $Y1$ and $Y2$ for the 1000 simulated points at this supply voltage are included in Subsection D.1.

Some simulation points failed for $VDD = 80$ mV, and these are plotted in Figure 52. In contrast to the pre layout results, the failing points are unique for $Y1$ and $Y2$. This makes sense when comparing the plots of the failing signals, as the failing points found for $Y1$ post layout all have the possibility of being interpreted correctly. The most critical failing points post layout are point 924 (see Figure 52a) and 360 (see Figure 52c), as the logic low values for these signals

are in the middle of the illegal range, and might therefore easily be interpreted one way or the other. All other failing values are quite close to the legal range.

As there are five unique failing points for both the *DUT* (*Y1*) and the *LOAD* (*Y2*), the post layout yield of the Output Selection Module at $VDD = 80$ mV is estimated to 99.5%. Plots of *Y1* and *Y2* for all 1000 post layout simulation points with $VDD = 80$ mV are included in Subsection D.1.

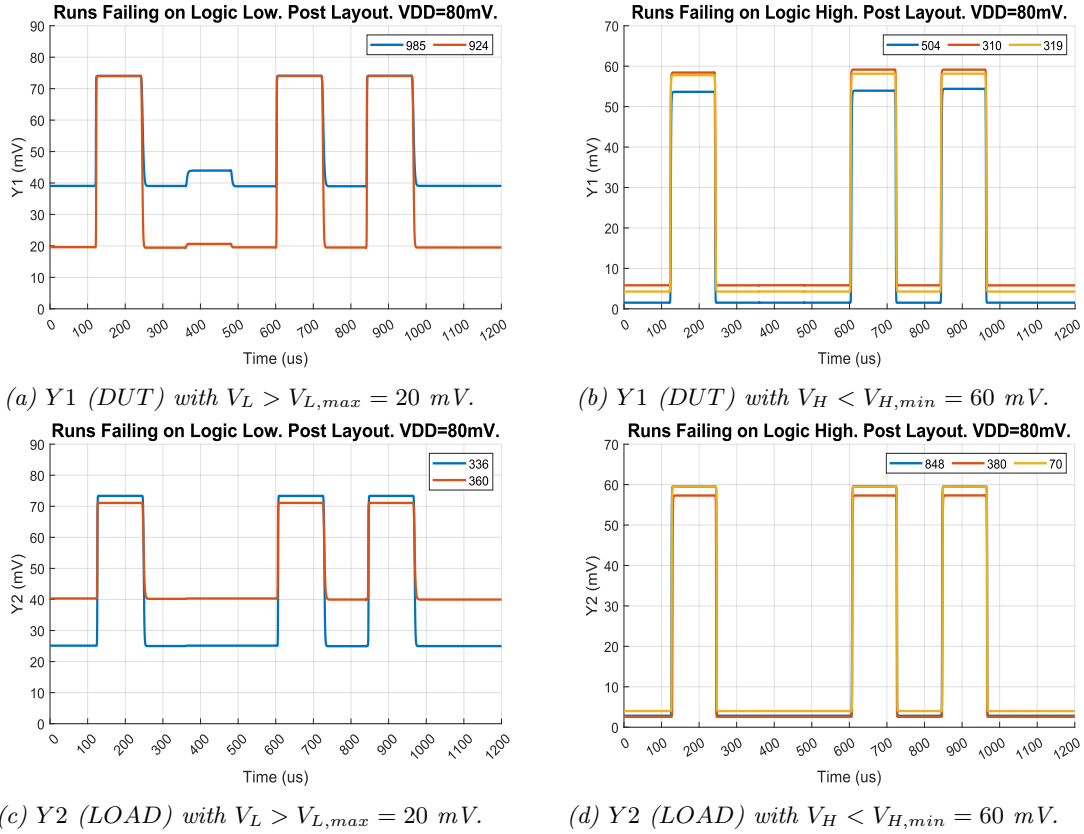


Figure 52: Output Selection Module. *Y1* and *Y2* are plotted for the failing post layout Monte Carlo simulation points at $VDD = 80$ mV. All failing points are unique.

5.5.2 Process Variation Simulation Results

Pre Layout Simulation Results

From the plot in Figure 53, which shows the results for the pre layout process corner simulation at $VDD = 75$ mV, we see that the logic high value produced for the FS50 corner is below $V_{H,min} = 56.25$ mV. All other corners pass. The *LOAD* receives *Y1* as input, which means that

a bad logic value is received for SF50. The already bad logic value is then further degraded when passing through circuit in the SF50 corner, which results in the increased deviation observed for Y2 (Figure 53b) compared to Y1 (Figure 53a) in this corner.

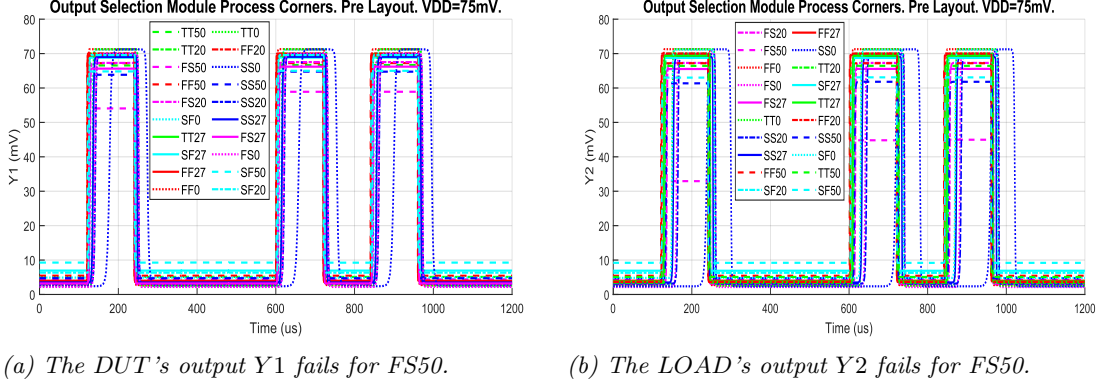


Figure 53: Plots of Y1 and Y2 for all pre layout process corners simulated with $VDD = 75 \text{ mV}$. The logic high values produced for FS50 are below $V_{H,min} = 56.25 \text{ mV}$.

The pre layout process corners were then simulated for $VDD = 80 \text{ mV}$, and all corners were found to pass. From the plots of Y1 and Y2 in Figure 54 it is clear that all logic values are well within the limits dictated by $V_{L,max} = 0.25VDD = 20 \text{ mV}$ and $V_{H,min} = 0.75VDD = 60 \text{ mV}$.

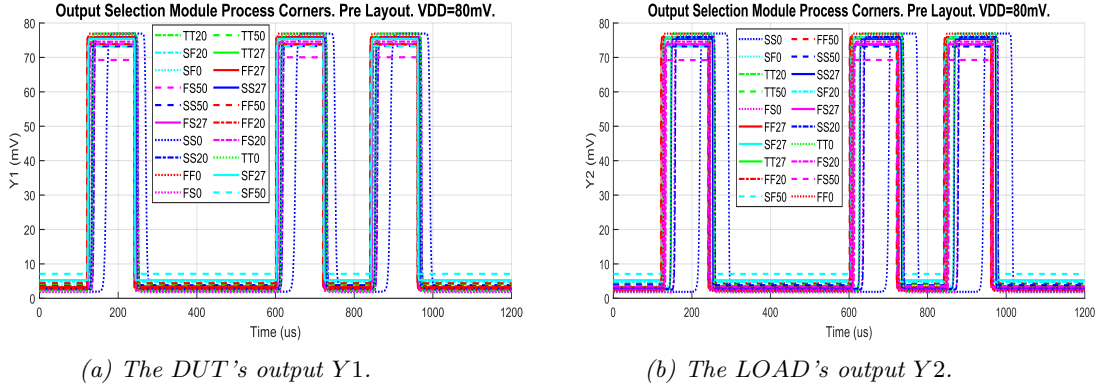


Figure 54: Plots of Y1 and Y2 for all pre layout process corners simulated with $VDD = 80 \text{ mV}$. All corners pass.

Post Layout Simulation Results

Post layout process corner simulations were first run with a supply voltage $VDD = 80 \text{ mV}$, and it is clear from the plot of Y1 in Figure 55a that the SF50 corner fails since $V_{L,SF50} > V_{L,max} = 20$

mV. $Y2$'s logic value is degraded more than $Y1$'s for SF50, as was also the case for the pre layout simulation. A plot of $Y2$ for all post layout corners at $VDD = 80$ mV is included in Subsection D.2.

The post layout simulation was then repeated for $VDD = 85$ mV, and it can be seen from Figure 55b that all process corners pass at this supply voltage. A plot of $Y2$ for all post layout corners at $VDD = 85$ mV is included in Subsection D.2.

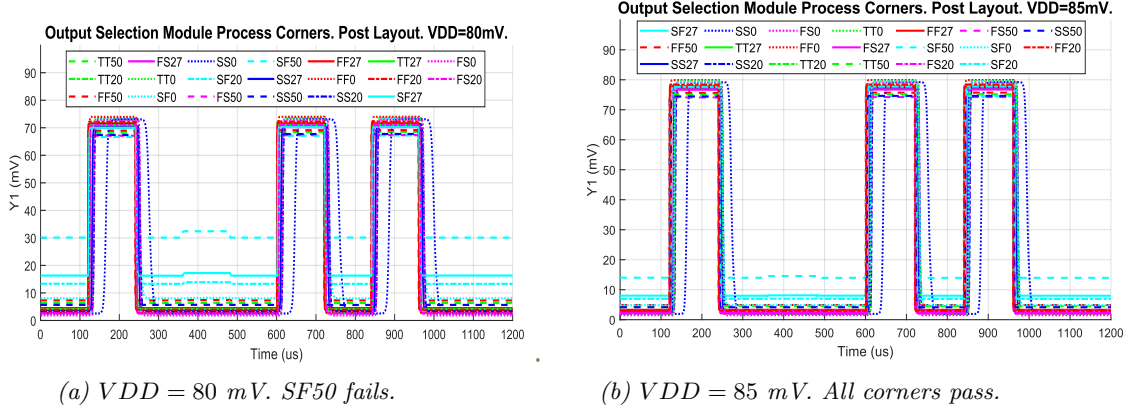


Figure 55: Plots of the DUT's output $Y1$ for all post layout process corners simulated with $VDD = 80$ mV and $VDD = 85$ mV.

5.6 4B SRAM Simulation Results

A description of the simulations run on the 4B SRAM testbench can be found in Subsection 4.11. The results from the Monte Carlo mismatch simulations are presented in Subsection 5.6.1, and the process corner simulation results are presented in Subsection 5.6.2.

A transient analysis was used as the basis for all simulations, and expected values for the eight output signals are illustrated in the timing diagram in Figure 31. The transient analysis lasts $t_{total} = 1$ ms, and is divided into 10 periods of $100 \mu\text{s}$. The output signals $Out0$, $Out1$, $Out2$, and $Out3$ are expected to go high at $t = 600 \mu\text{s}$ and remain high for one period. The output signals $Out4$, $Out5$, $Out6$, and $Out7$ are expected to go high at $t = 900 \mu\text{s}$ and remain high for one period.

5.6.1 Monte Carlo Mismatch Simulation Results

Pre Layout Simulations

Pre layout mismatch simulations were run for $VDD = 70$ mV, $VDD = 75$ mV, and $VDD = 80$ mV. All simulated points were found to pass for $VDD = 75$ mV and $VDD = 80$ mV, and

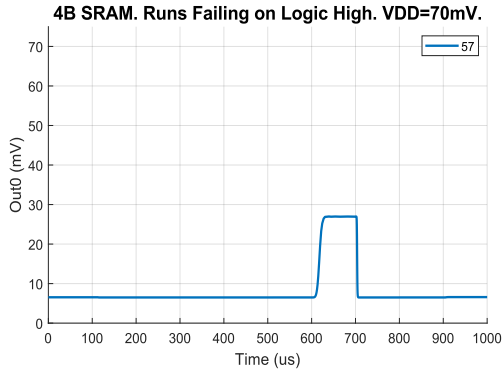
plots of *Out0*, *Out1*, *Out2*, *Out3*, *Out4*, *Out5*, *Out6*, and *Out7* for all simulated points at these supply voltages are included in Subsection E.

The pre layout Monte Carlo simulation at $V_{DD} = 70$ mV had some failing points. *Out0* failed to produce a high enough logic high value for simulation point 57, see Figure 56a. The logic high value it produces is below 30 mV, and is closer to the legal range for logic low values than to the legal range for logic high values. *Out1* is plotted for the failing simulation point 893 in Figure 56b, where it settled at a too high logic low value. From the plot of *Out2* for the failing point 248 in Figure 56c it becomes clear that the failure is caused by the logic low value increasing to a value above $V_{L,max} = 17.25$ mV in the last period. This error is so small that it would likely be interpreted correctly if it was used as an input to either a NAND or a NOT. *Out3* and *Out4* were correct for all simulated points.

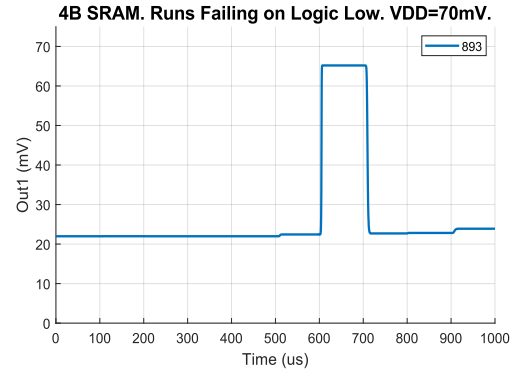
The logic low value produced for *Out5* is slightly too high for simulation point 169, see Figure 56d, but the deviation from legal logic low values is so small that it can be considered to nearly pass. *Out6* fails for two simulation points, plotted in Figure 56e, where point 580 fails because the logic high value produced is marginally lower than $V_{H,min} = 52.5$ mV and point 88 fails because *Out6* remains a constant logic low. *Out7* is plotted for the failing point 290 in Figure 83h, which fails because the logic low value settles at a value higher than $V_{L,max} = 17.5$ mV.

The failure seen in *Out6* for simulation point 88 is the most serious, as the 4B SRAM seems to either have stored an incorrect value or to not respond to the read operation. The failing point for *Out0* is also critical, as it is very unlikely that the logic high value produced will be interpreted as logic high.

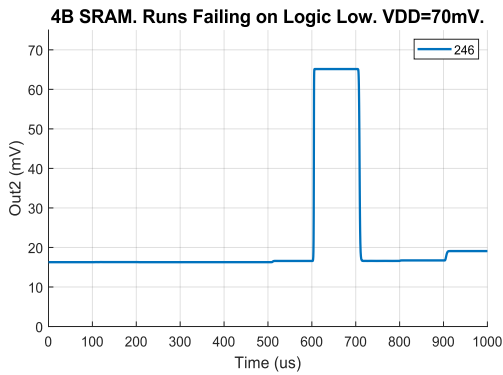
None of the output signals fail for the same simulation points, so there are 7 failing points in total. The yield is estimated to be 99.3% for the 4B SRAM at $V_{DD} = 70$ mV. Plots of *Out0*, *Out1*, *Out2*, *Out3*, *Out4*, *Out5*, *Out6*, and *Out7* for all simulated pre layout mismatch points with $V_{DD} = 70$ mV are included in Subsection E.



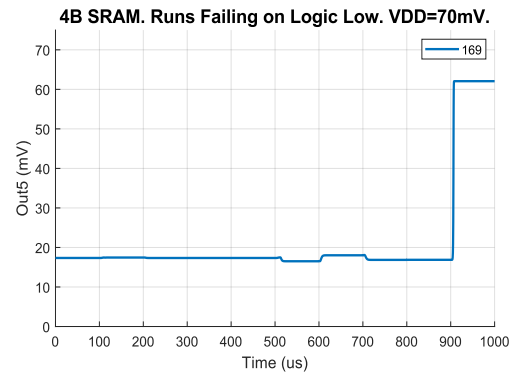
(a) *Out0* fails because $V_H < V_{H,min} = 52.5$ mV.



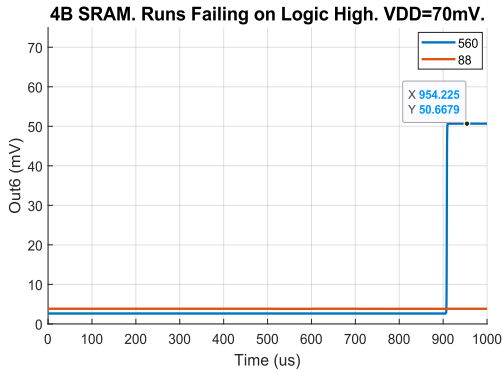
(b) *Out1* fails because $V_L > V_{L,max} = 17.5$ mV.



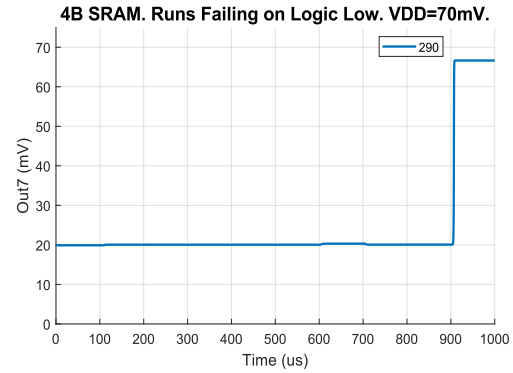
(c) *Out2* fails because $V_L > V_{L,max} = 17.5$ mV.



(d) *Out5* fails because $V_L > V_{L,max} = 17.5$ mV.



(e) *Out6* fails because $V_H < V_{H,min} = 52.5$ mV.



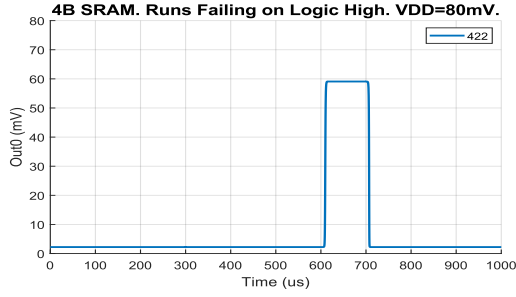
(f) *Out7* fails because $V_L > V_{L,max} = 17.5$ mV.

Figure 56: The 4B SRAM's output signals plotted for the pre layout simulation points failing when $VDD = 70$ mV. *Out0*, *Out1*, and *Out2* should have a square pulse in the seventh period, and be logic low otherwise. *Out5*, *Out6*, and *Out7* should be logic low until they go high in the last period.

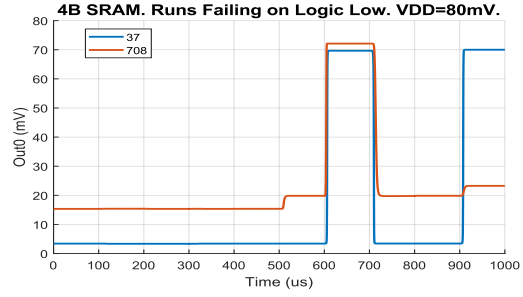
Post Layout Simulations

The post layout mismatch simulations were run with $VDD = 80$ mV and with $VDD = 85$ mV. All simulated points were found to pass for $VDD = 85$ mV, and plots of the 4B SRAM's output signals for all simulated points are included in Subsection E.

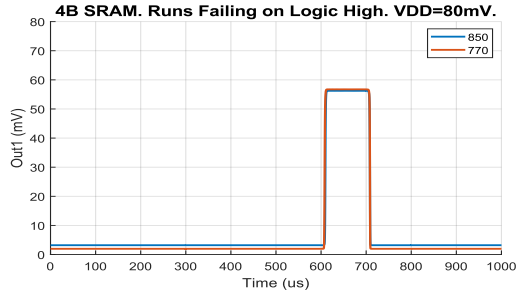
A total of 3 unique simulation points fail for *Out0*. The logic high value that *Out0* settles on is slightly below $V_{H,min} = 60$ mV for point 422, see Figure 57a, and this is therefore classified as a fail. Figure 57b contains plots of *Out0* for the two points that fail to maintain a correct logic low value. Point 708 fails because the logic low value is slightly above $V_{L,max} = 20$ mV in the last period. Point 37 is a more critical error, as *Out0* goes high in the last period. This reflects an error in the functionality of the SRAM for this simulation point, as either a wrong value has been stored (should have been logic 0) or the wrong address is being read from.



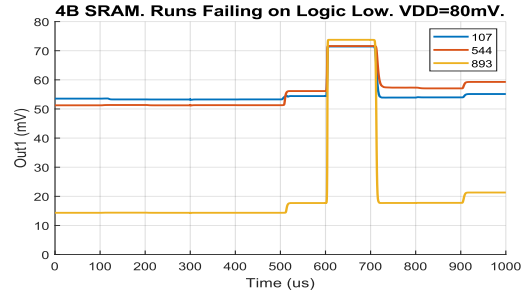
(a) *Out0 fails because $V_H < V_{H,min} = 60$ mV.*



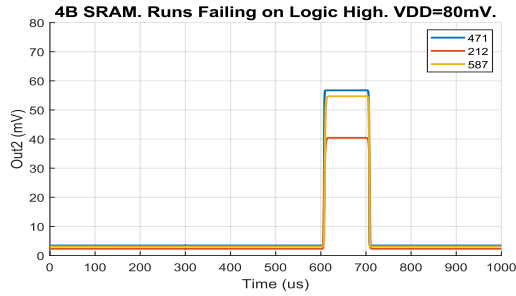
(b) *Out0 fails because $V_L > V_{L,max} = 20$ mV.*



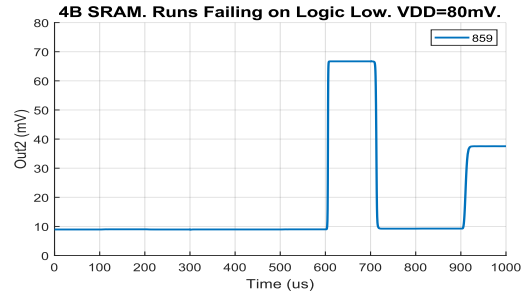
(c) *Out1 fails because $V_H < V_{H,min} = 60$ mV.*



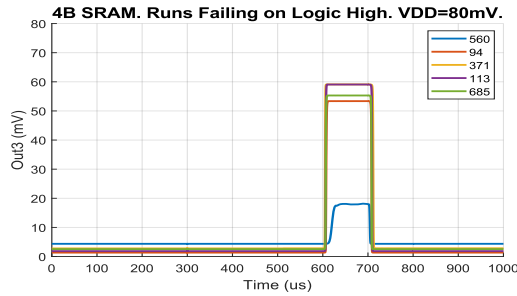
(d) *Out1 fails because $V_L > V_{L,max} = 20$ mV.*



(e) *Out2 fails because $V_H < V_{H,min} = 60$ mV.*



(f) *Out2 fails because $V_L > V_{L,max} = 20$ mV.*



(g) *Out3 fails because $V_H < V_{H,min} = 60$ mV.*

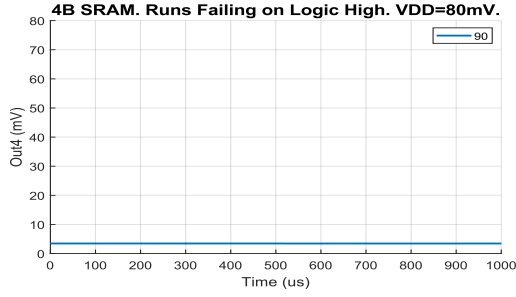
Figure 57: The 4B SRAM's output signals Out0-Out3 plotted for the failing post layout simulation points with $V_{DD} = 80$ mV. The signals should have a square pulse in the seventh period, between $t = 600$ μ s and $t = 700$ μ s.

Five simulation points fail for *Out1*. The logic high value that *Out1* settles on is below $V_{H,min} = 60$ mV for points 850 and 770, see Figure 57c. Figure 57d contains plots of *Out1* for the three points that fail to maintain a correct logic low value. Point 893 fails because the logic low value is slightly above $V_{L,max} = 20$ mV in the last period. Points 107 and 544 settle at a logic level that is closer to the logic high range than to $V_{L,max}$, so these will likely be misinterpreted as logic high.

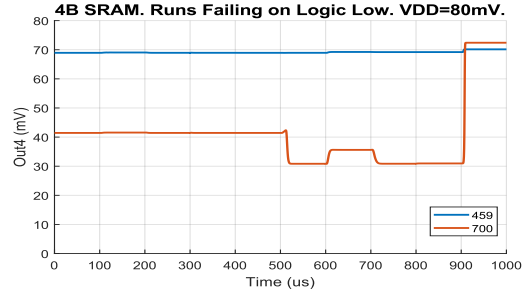
As can be seen from Figure 57e, points 471, 212, and 587 fail because the value of *Out2* is below $V_{H,min}$ in the seventh period. For point 859, *Out2* settles at an illegal value of approximately $0.5V_{DD}$ in the last period when it was supposed to remain low (see Figure 57f).

Correct logic low values are produced for *Out3* for all simulated points, but 5 points fail to produce a legal logic high value in the seventh period. *Out3* is plotted for these 5 failing points in Figure 57g. 560 is the most critical point, since the top of the pulse in the seventh period is still within the legal range of logic low values. It should therefore be interpreted as a logic low, which is incorrect. The four other points, 94, 371, 113, and 685, all settle at logic high values that are relatively close to the legal range for logic high.

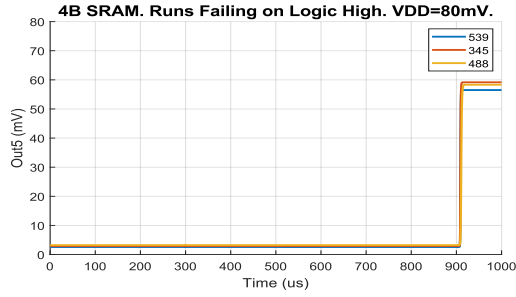
Three simulation points cause *Out4* to have incorrect values. *Out4* remains at a constant logic low value the whole time for simulation point 90, see plot in Figure 58a, even though it should have gone high in the last period. Similarly, *Out4* remains at a constant logic high value for point 459, as shown in Figure 58b. Point 700 produce a more correct plot for *Out4*, but fails because the logic low value remains in the middle of the illegal range.



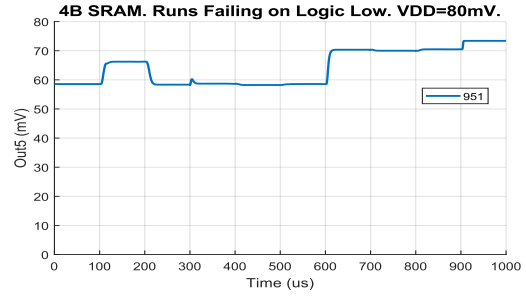
(a) *Out4 fails because $V_H < V_{H,min} = 60$ mV.*



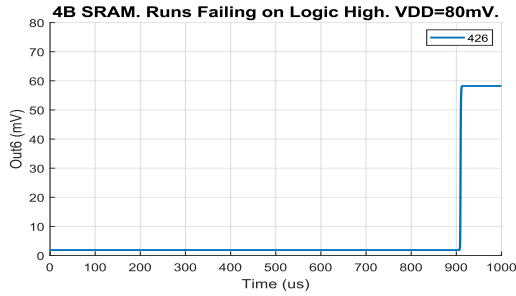
(b) *Out4 fails because $V_L > V_{L,max} = 20$ mV.*



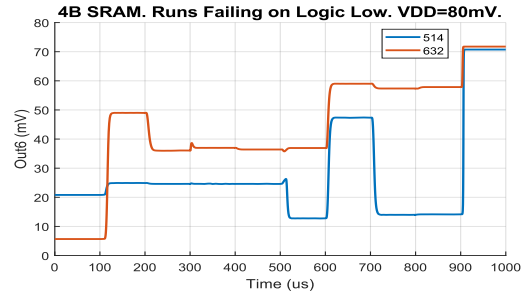
(c) *Out5 fails because $V_H < V_{H,min} = 60$ mV.*



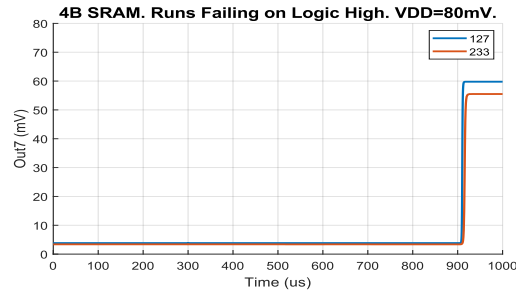
(d) *Out5 fails because $V_L > V_{L,max} = 20$ mV.*



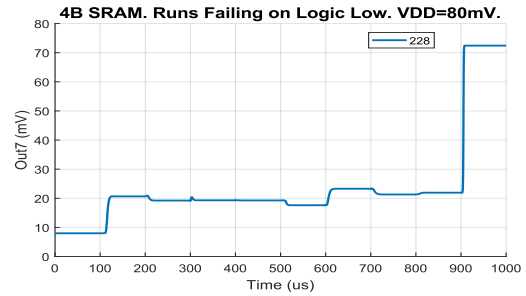
(e) *Out6 fails because $V_H < V_{H,min} = 60$ mV.*



(f) *Out6 fails because $V_L > V_{L,max} = 20$ mV.*



(g) *Out7 fails because $V_H < V_{H,min} = 60$ mV.*



(h) *Out7 fails because $V_L > V_{L,max} = 20$ mV.*

Figure 58: The 4B SRAM's output signals Out4-Out7 plotted for the failing post layout simulation points with $V_{DD} = 80$ mV. The signals should have a square pulse in the last period, starting at $t = 900$ μ s.

Simulation points 539, 345, and 488, which are plotted in Figure 58c, cause *Out5* to fail, but these are not critical since they settle at a logic high value only slightly below $V_{H,min} = 60$ mV. The fail for point 951, plotted in Figure 58d, is a critical error, as *Out5* remains at a logic high value the whole time.

The logic high value created for *Out6* in the last period is slightly too low for point 426, see Figure 58e. The value of *Out6* goes up and down for points 514 and 632 in the first nine periods, see Figure 58e, instead of remaining at a constant logic low value.

Simulation points 127 and 233 create logic high values for *Out7* that are slightly below $V_{H,min} = 60$ mV in the last period, see Figure 58g, and *Out7* has a logic low value that oscillates around $V_{L,max} = 20$ mV for point 228, see Figure 58h.

Since none of the simulated points cause a failure for more than one of the output signals, there are a total of 30 unique failing points. The estimated yield for the post layout 4B SRAM at $VDD = 80$ mV is therefore 97.0%. Plots of *Out0*, *Out1*, *Out2*, *Out3*, *Out4*, *Out5*, *Out6*, and *Out7* for all 1000 simulated post layout Monte Carlo points at $VDD = 80$ mV are included in Subsection E.

5.6.2 Process Corner Simulation Results

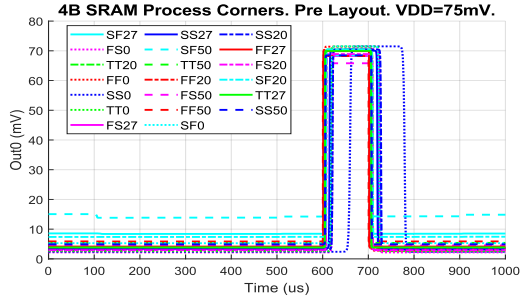
Pre Layout Simulation

Pre layout simulations of process variation were run for $VDD = 70$ mV, $VDD = 75$ mV, and $VDD = 80$ mV.

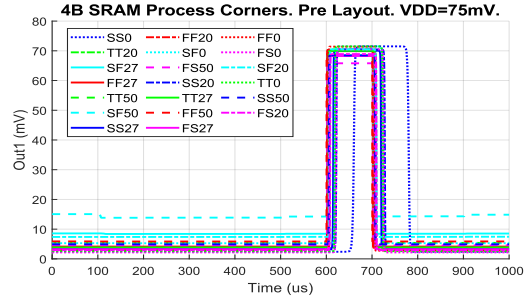
All the 4B SRAM's output signals fail to produce logic low values within the legal range for SF20, SF27, and SF50 when $VDD = 70$ mV. SF20 and SF27 produce logic low values close to $V_{L,max}$, but SF50 settles at values in the middle of the illegal range. Plots of the output signals for all corners are included in Subsection E.2.

All pre layout corners pass for both $VDD = 75$ mV and $VDD = 80$ mV. Plots of the output signals for all corners at $VDD = 80$ mV are included in Subsection E.2. The output signals for $VDD = 75$ mV are plotted for all corners in Figure 59.

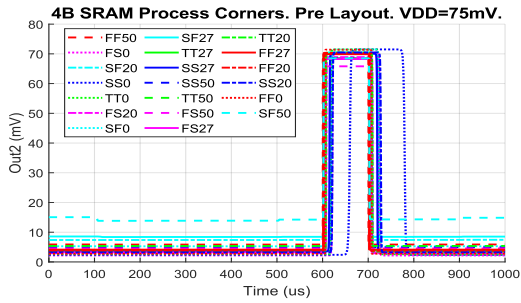
All the output signals are affected in more or less the same way by the different process corners. Looking at the plots for *Out0* in Figure 59a, it is clear that SS0 causes a much longer delay than any of the other corners. The highest logic low value is produced for SF50, and the lowest logic high value for FS50.



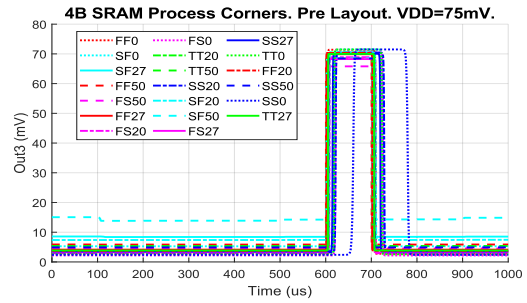
(a) Plot of Out0.



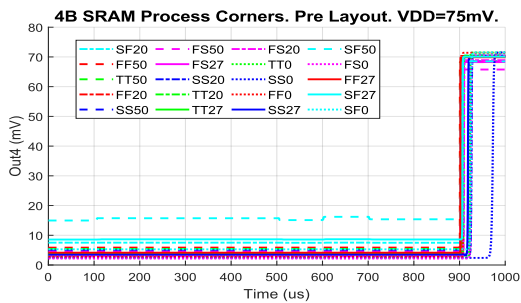
(b) Plot of Out1.



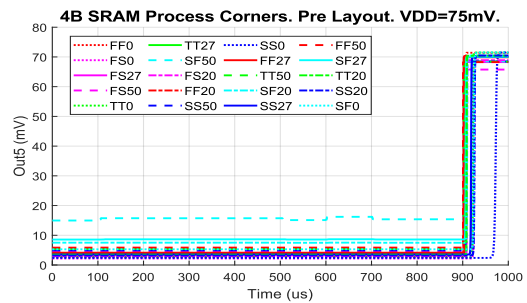
(c) Plot of Out2.



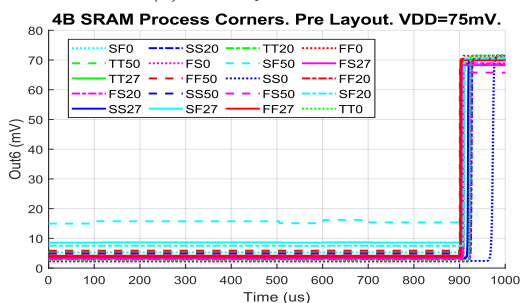
(d) Plot of Out3.



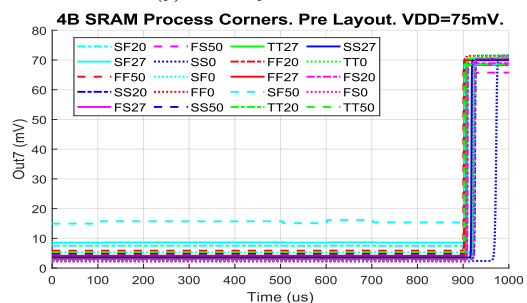
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

Figure 59: The 4B SRAM's output signals plotted for all pre layout process corners simulated with $VDD = 75\text{ mV}$. All corners pass.

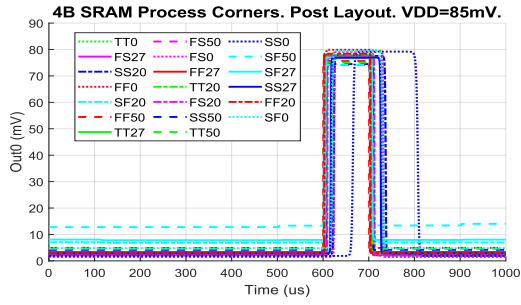
Post Layout Simulation

Post layout process corner simulations were run for $VDD = 75$ mV, $VDD = 80$ mV, and $VDD = 85$ mV.

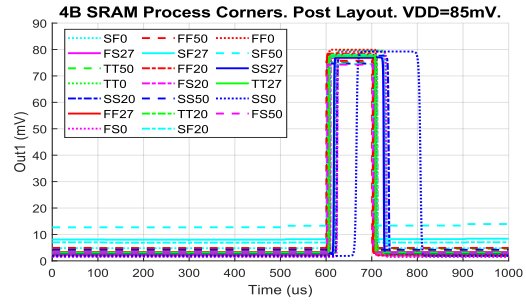
When $VDD = 75$ mV, all the 4B SRAM's output signals fail to produce logic low values within the legal range for SF0, SF20, SF27, and SF50. *Out4-Out7* also fail to produce legal logic low values for FF50. SF0 and FF50 produces logic low values close to $V_{L,max}$, but the others settle at values in the middle of the illegal range. SF50 fails to create a logic high value for *Out4*, *Out5*, *Out6*, and *Out7*. Plots of the output signals for all corners are included in Subsection E.2.

For $VDD = 80$ mV, all output signals fail to produce logic low values within the legal range for SF50. Plots of the output signals for all corners are included in Subsection E.2.

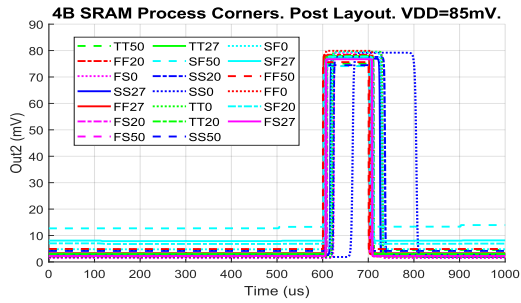
As can be seen from the plots in Figure 60, all corners pass when simulating with $VDD = 85$ mV. SS0 is the slowest corner, and from Figure 60e it is clear that *Out4* goes high close to the end of the period (the same can be observed for *Out5*, *Out6*, and *Out7*). The delay before the output starts to go low in the SS0 corner is slightly shorter for *Out0-Out3*. SF50 produces the highest logic low values for all output signals, which is expected since this was the corner that failed in the simulation at $VDD = 80$ mV.



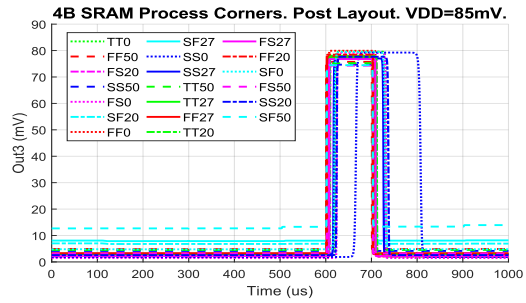
(a) Plot of Out0.



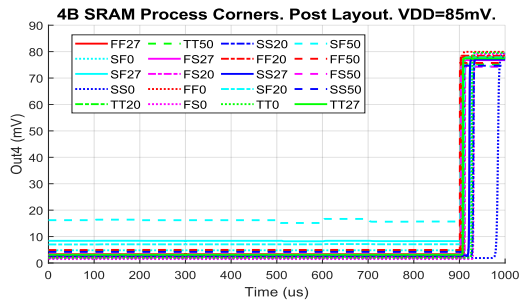
(b) Plot of Out1.



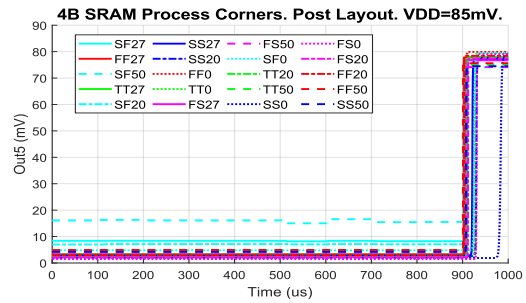
(c) Plot of Out2.



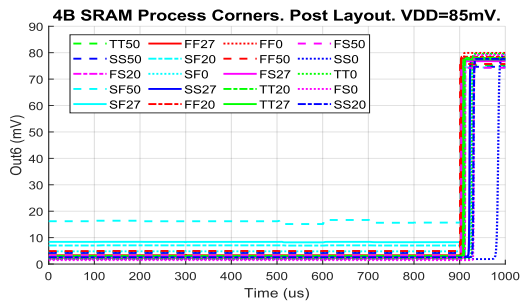
(d) Plot of Out3.



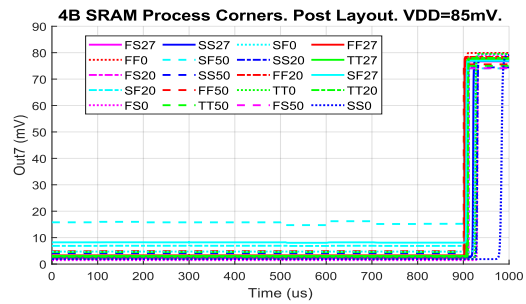
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

Figure 60: The 4B SRAM's output signals plotted for all post layout process corners simulated with $VDD = 85\text{ mV}$. All corners pass.

5.7 16B SRAM Simulation Results

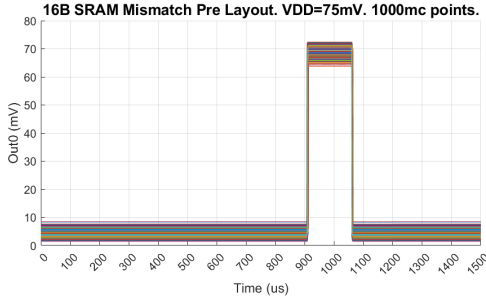
A description of the simulations run on the 16B SRAM testbench can be found in Subsection 4.12. Results from the Monte Carlo mismatch simulations are presented in Subsection 5.7.1, and the process corner simulation results are presented in Subsection 5.7.2.

A transient analysis was used as the basis for all simulations, and expected values for the eight output signals are illustrated in the timing diagram in Figure 33. The transient analysis is divided into 10 periods. The output signals *Out0*, *Out1*, *Out2*, and *Out3* are expected to go high in the seventh period, while *Out4*, *Out5*, *Out6*, and *Out7* are expected to go high in the tenth period.

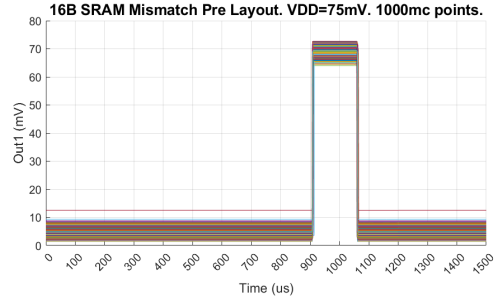
5.7.1 Monte Carlo Mismatch Simulation Results

Pre Layout Simulation

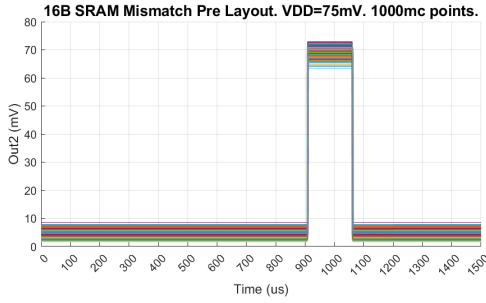
The transient analysis used for the pre layout mismatch simulation with $VDD = 75$ mV lasted 1500 μs , with each period lasting 150 μs . *Out0-Out3* are therefore expected to go high at $t = 900$ μs and then go low again at $t = 1050$ μs . *Out4-Out7* should go high at $t = 1350$ μs and remain high for the rest of the analysis. *Out0-Out7* are plotted in Figure 61 for the 1000 simulated points, and it is clear that all the simulated points pass.



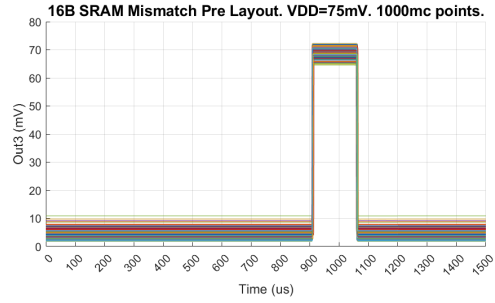
(a) Plot of Out0.



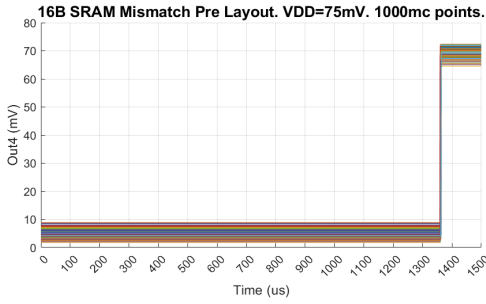
(b) Plot of Out1.



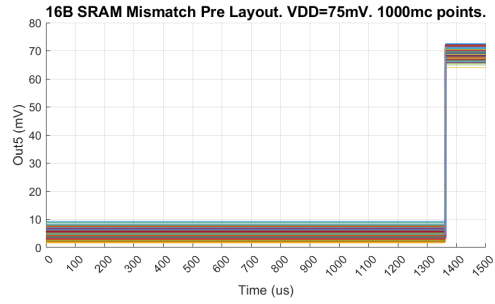
(c) Plot of Out2.



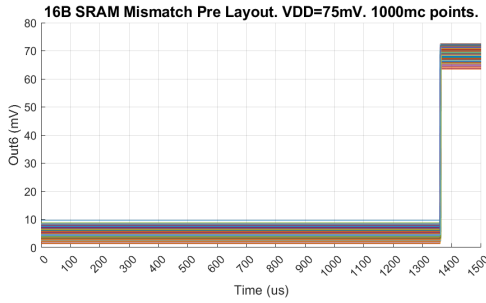
(d) Plot of Out3.



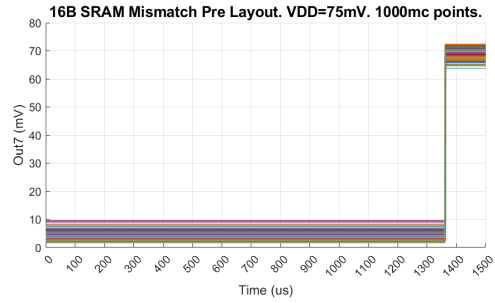
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

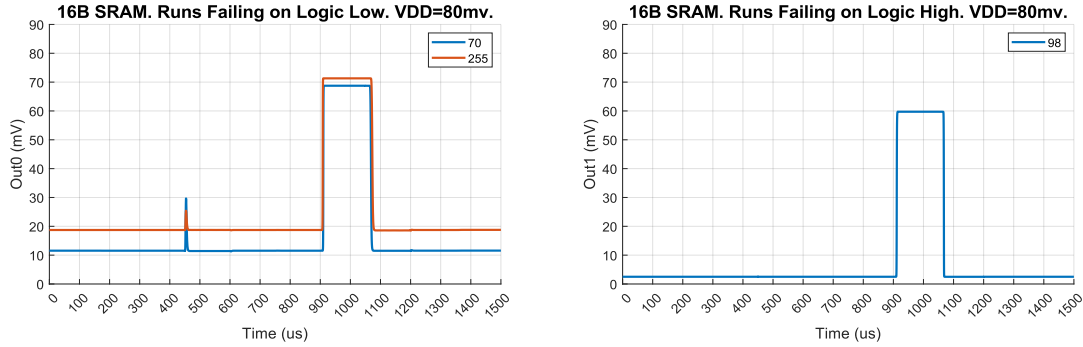
Figure 61: The 16B SRAM's output signals plotted for the 1000 simulated points in the pre layout Monte Carlo mismatch simulation with $VDD = 75 \text{ mV}$. All simulated points pass.

Post Layout Simulations

The transient analysis used for the post layout mismatch simulation with $VDD = 85$ mV lasted $750 \mu\text{s}$, with each period lasting $75 \mu\text{s}$. All the simulated points passed, and plots of $Out0$ - $Out7$ for all the 500 points simulated are included in Subsection F.1.

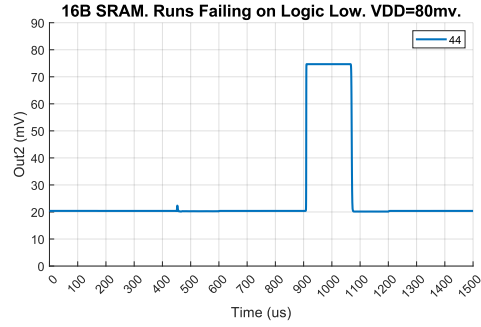
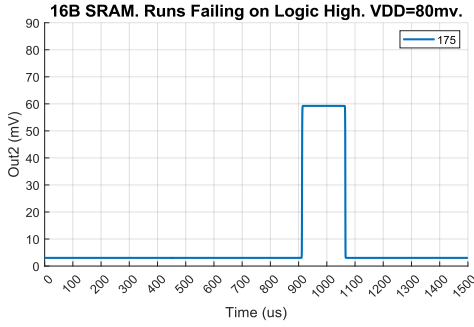
For the post layout mismatch simulation with $VDD = 80$ mV, the total runtime of the transient simulation was $1500 \mu\text{s}$. $Out0$ - $Out3$ are therefore expected to go high at $t = 900 \mu\text{s}$ and then go low again at $t = 1050 \mu\text{s}$. $Out4$ - $Out7$ should go high at $t = 1350 \mu\text{s}$ and remain high for the rest of the analysis.

A total of two simulation points fail for $Out0$, and these are plotted in Figure 62a. For both points, the fail occurs because the logic low value has a spike at $t = 450 \mu\text{s}$. The spikes are very narrow and the peaks are below 30 mV, so these failing points are likely to not be critical.



(a) $Out0$ fails because $V_L > V_{L,max} = 20$ mV.

(b) $Out1$ fails because $V_H < V_{H,min} = 60$ mV.



(c) $Out2$ fails because $V_H < V_{H,min} = 60$ mV.

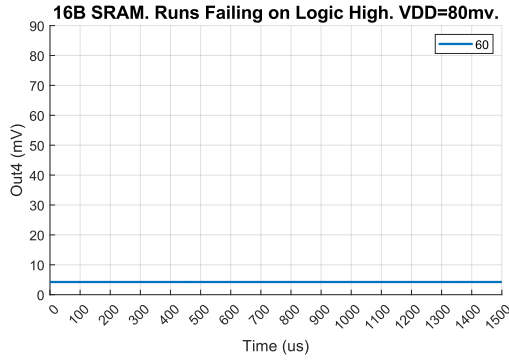
(d) $Out2$ fails because $V_L > V_{L,max} = 20$ mV.

Figure 62: The 16B SRAM's output signal $Out0$ - $Out2$ plotted for the failing post layout simulation points with $VDD = 80$ mV. The signals should have a square pulse in the seventh period, between $t = 900 \mu\text{s}$ and $t = 1050 \mu\text{s}$.

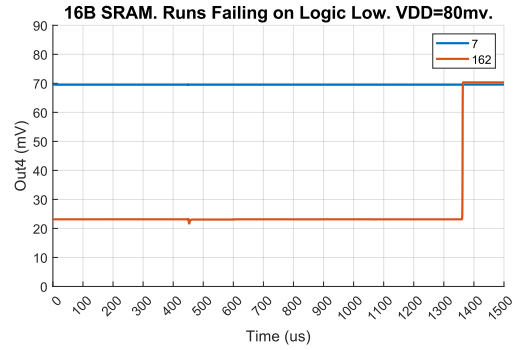
$Out1$ is plotted in Figure 62b for the failing point 98. This is a fail because the logic high value is marginally below $V_{H,min} = 60$ mV, but it is so close that it can be considered to nearly pass.

For simulation point 175, see Figure 63a, *Out2* settles at a logic high value marginally below $V_{H,min} = 60$ mV, and this can be classified as a near pass. Point 44, which is plotted in Figure 62d, is similarly a near pass, as *Out2* settles at a logic low value which is slightly higher than $V_{L,max}$.

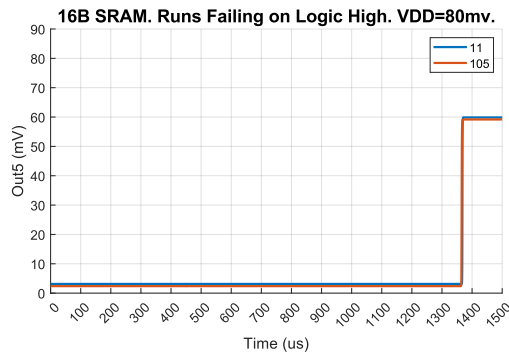
Out4 fails for three simulation points, and these are plotted in Figure 63a and Figure 63b. *Out4* remains at a constant logic low value for point 60, which is a critical error. Point 7 is also a critical error, as *Out4* remains at a constant high value for this simulation point. Simulation point 162 fails because the logic low value is slightly above $V_{L,max} = 20$ mV, but *Out4* could still be interpreted correctly and this is therefore a non-critical error.



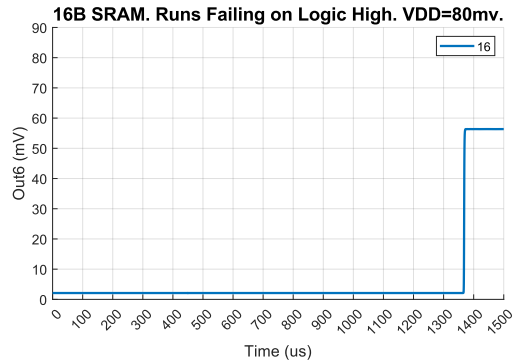
(a) *Out4 fails because $V_H < V_{H,min} = 60$ mV.*



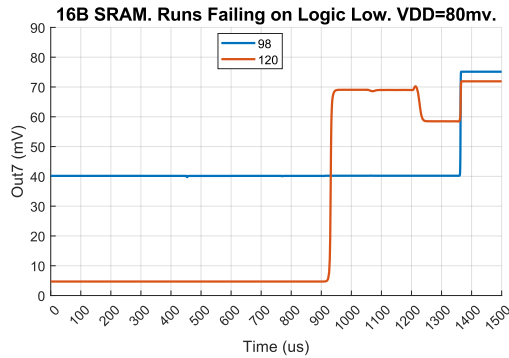
(b) *Out4 fails because $V_L > V_{L,max} = 20$ mV.*



(c) *Out5 fails because $V_H < V_{H,min} = 60$ mV.*



(d) *Out6 fails because $V_H < V_{H,min} = 60$ mV.*



(e) *Out7 fails because $V_L > V_{L,max} = 20$ mV.*

Figure 63: The 16B SRAM's Out4 plotted for the failing post layout simulation points with $V_{DD} = 80$ mV.

Two failing simulation points are caused by *Out5* settling at a logic high value that is very slightly below $V_{H,min} = 60$ mV, and these are plotted in Figure 63c.

Figure 63d contains a plot of *Out6* for point 16, which is the only simulation point which causes a fail for this output signal. The logic high value produced is below $V_{H,min} = 60$ mV.

Two failing simulation points are caused by *Out7* not being logic low when it should be, see Figure 63e. *Out7* is 40 mV the entire time it should be logic low, and as this is in the middle of the illegal value range it is a clear fail. Point 120 produces correct values until $t = 900$ μ s, at which point *Out7* goes to logic high when it should remain low. This error might be caused by the wrong value being stored in the bitcell or by reading from the wrong bitcell, and is therefore a critical error.

As point 98 causes a fail in both *Out1* and *Out7*, the total number of unique failing points for the post layout simulation with $VDD = 80$ mV is 12. The circuit yield for the 16B SRAM post layout can be estimated to 97.6% for this supply voltage. Plots of the eight output signals for all 500 simulated points are included in Subsection F.1.

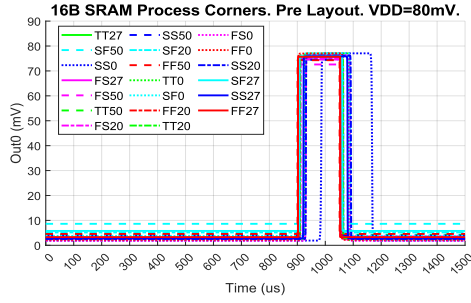
5.7.2 Process Variation Simulation Results

Pre Layout Simulations

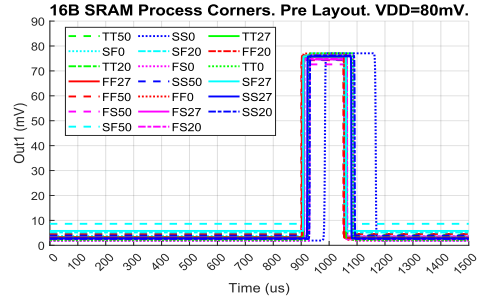
The transient analysis was set to last 1500 μ s for the pre layout process corner simulations. *Out0-Out3* are expected to go high at $t = 900$ μ s and then go low again at $t = 1050$ μ s. *Out4-Out7* should go high at $t = 1350$ μ s and remain high for the rest of the analysis.

Pre layout process variation was simulated for $VDD = 75$ mV, $VDD = 80$ mV, and $VDD = 85$ mV. For $VDD = 75$ mV, the SF50 corner fails because none of the output signals produce a low enough logic low value. Plots of *Out0-Out7* for all simulated pre layout corners at $VDD = 75$ mV are included in Subsection F.2. All corners pass for both $VDD = 80$ mV and $VDD = 85$ mV. Plots of the output signals for all corners at $VDD = 85$ mV are included in Subsection F.2. Figure 64 contains the plots of *Out0-Out7* for all corners at $VDD = 80$ mV.

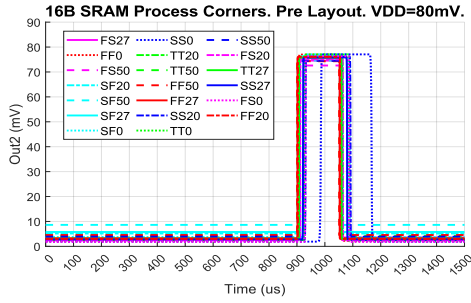
The process corners affect the different output signals in more or less the same way. Looking at the plots for *Out0* in Figure 64a, it is clear that SS0 causes a much longer delay than any of the other corners. The highest logic low value is produced for SF50, and the lowest logic high value for FS50.



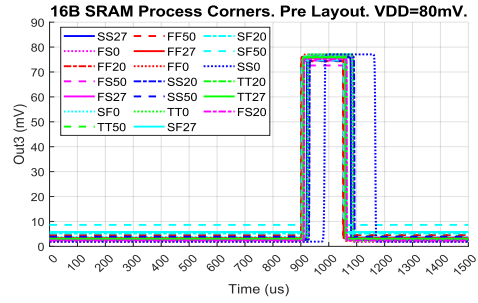
(a) Plot of Out0.



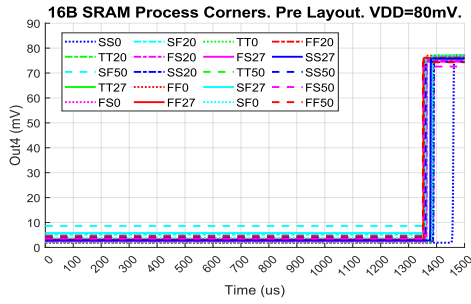
(b) Plot of Out1.



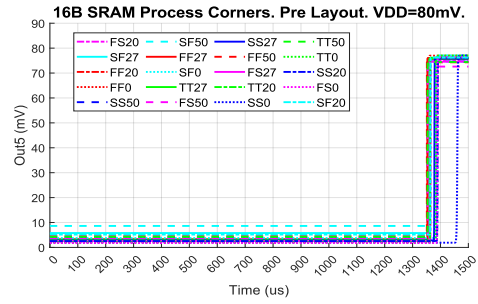
(c) Plot of Out2.



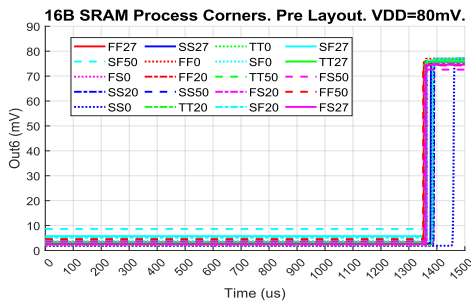
(d) Plot of Out3.



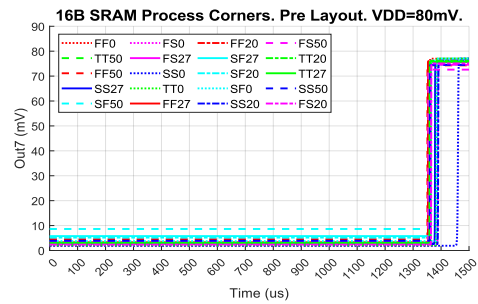
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

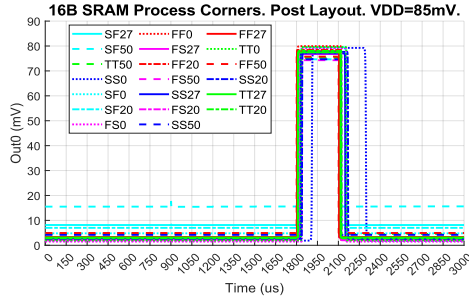
Figure 64: The 16B SRAM's output signals plotted for all pre layout process corners simulated with $VDD = 80\text{ mV}$. All corners pass.

Post Layout Simulations

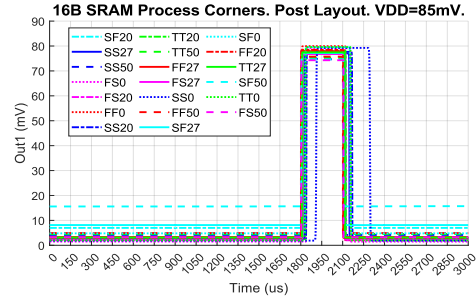
The transient analysis was set to last 3000 μs for the post layout process corner simulations. *Out0-Out3* are expected to go high at $t = 1800 \mu\text{s}$ and then go low again at $t = 2100 \mu\text{s}$. *Out4-Out7* should go high at $t = 2700 \mu\text{s}$ and remain high for the rest of the analysis.

Post layout process corner simulation was first run for $VDD = 80 \text{ mV}$. All the 16B SRAM's output signals fail to produce a legal logic low value for SF50 and SF27. SF27 is a near pass, but SF50 produces logic low values in the logic high range which is unacceptable. Plots of *Out0-Out7* for all the simulated process corners are included in Subsection F.2.

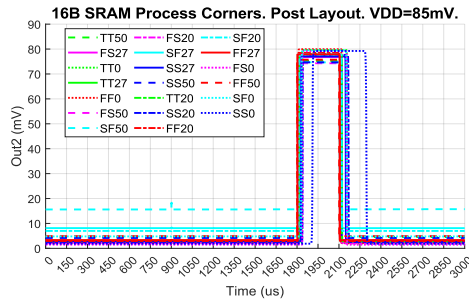
The post layout process corner simulation was then repeated for $VDD = 85 \text{ mV}$, and all points were found to pass. *Out0-Out7* are plotted in Figure 65 for all simulated post layout process corners. SS0 stands out as the slowest corner. SF50 produces the highest logic low values for all output signals, but it is still clearly below $V_{L,max} = 21.25 \text{ mV}$.



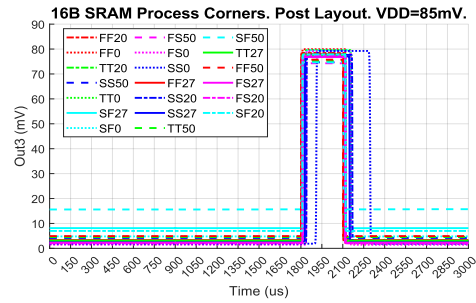
(a) Plot of Out0.



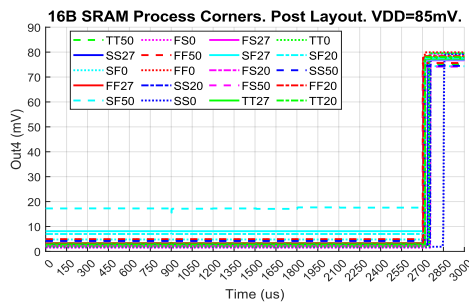
(b) Plot of Out1.



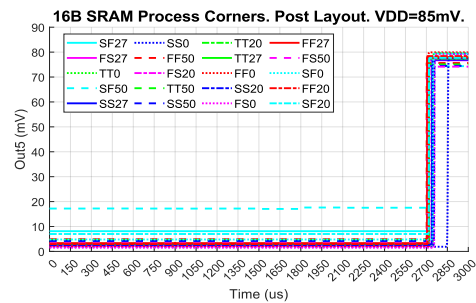
(c) Plot of Out2.



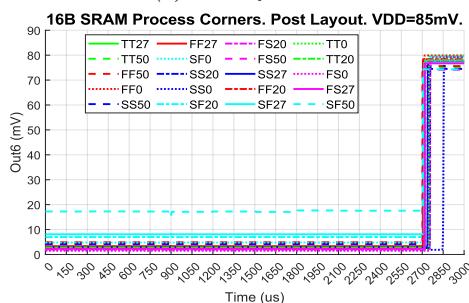
(d) Plot of Out3.



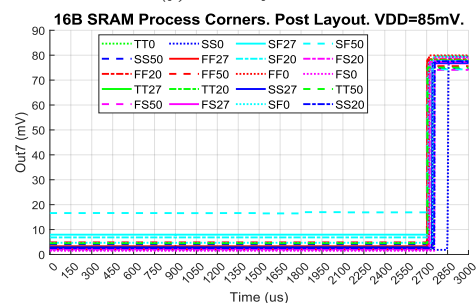
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

Figure 65: The 16B SRAM's output signals plotted for all post layout process corners simulated with $V_{DD} = 85 \text{ mV}$. All corners pass.

5.8 64B SRAM Simulation Results

A description of the simulations run on the 64B SRAM testbench can be found in Subsection 4.13. Results from the process corner simulations are presented in Subsection 5.8.1.

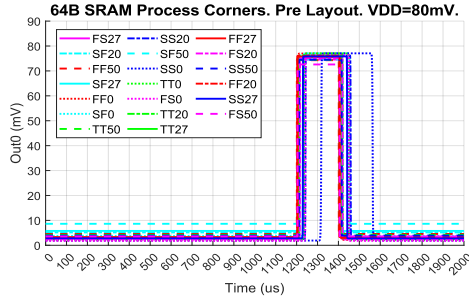
A transient analysis was used as the basis for all simulations, and expected values for the eight output signals are illustrated in the timing diagram in Figure 35. The transient analysis is divided into 10 periods. The output signals *Out0*, *Out1*, *Out2*, and *Out3* are expected to go high in the seventh period, while *Out4*, *Out5*, *Out6*, and *Out7* are expected to go high in the tenth period.

5.8.1 Results from Simulations of Process Variation

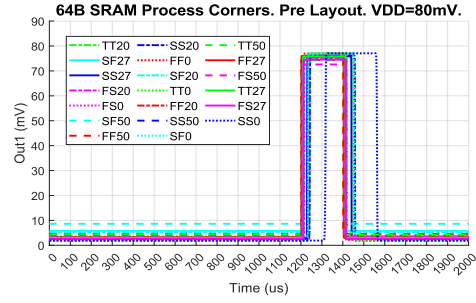
Pre Layout Simulations

Pre layout process corner simulations were first run for $VDD = 75$ mV, using a transient analysis lasting 2500 μ s. All the 64B SRAM's output signals fail to produce a legal logic low value for the SF50 corner, but all other corners pass. Plots of *Out0-Out7* for all simulated pre layout corners at $VDD = 75$ mV are included in Subsection G.1.

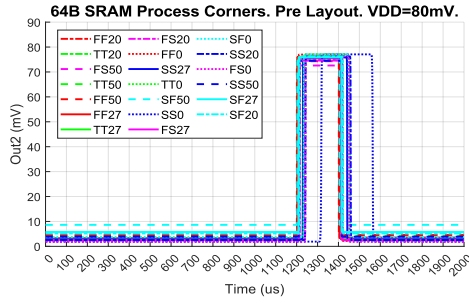
All pre layout process corners pass for both $VDD = 80$ mV and $VDD = 85$ mV. Plots of *Out0-Out7* for all corners at $VDD = 85$ mV are included in Subsection G.1. *Out0-Out7* are plotted in Figure 66 for all corners simulated with $VDD = 80$ mV. The delay for SS0 is considerably longer than for the other corners. SF50 produces the highest logic low value, which is expected as this was the failing point for $VDD = 75$ mV, but it is < 10 mV and thus well below $V_{L,max} = 20$ mV.



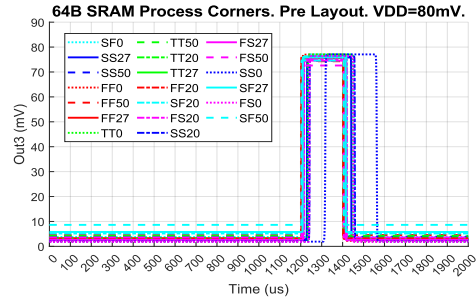
(a) Plot of Out0.



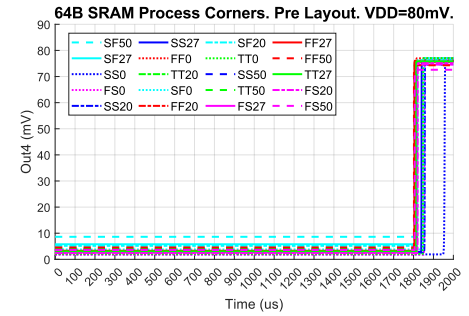
(b) Plot of Out1.



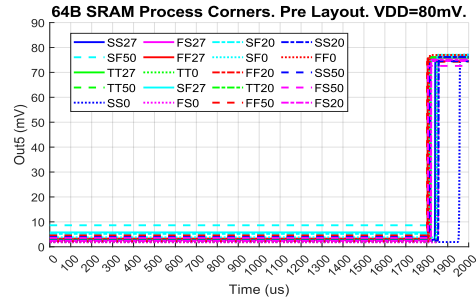
(c) Plot of Out2.



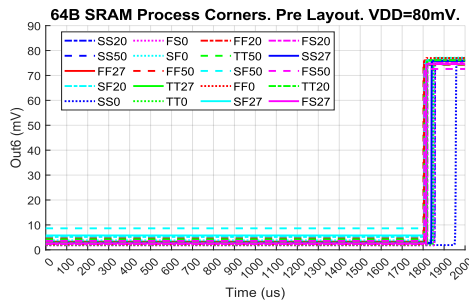
(d) Plot of Out3.



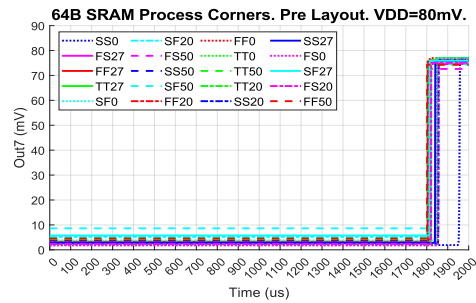
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

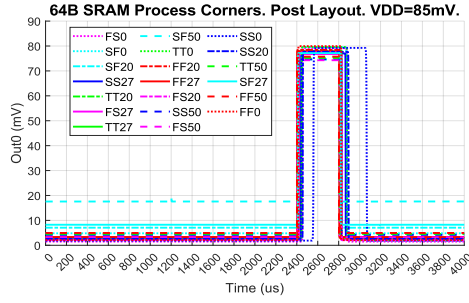
Figure 66: The 64B SRAM's output signals plotted for all pre layout process corners simulated with $V_{DD} = 80\text{ mV}$. All corners pass.

Post Layout Simulations

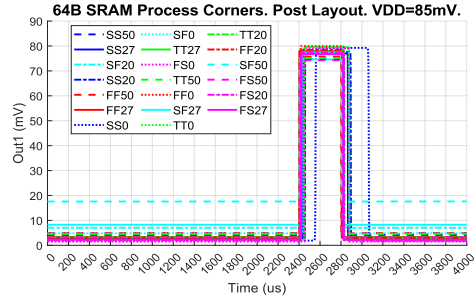
The transient analysis was set to last 4000 μs for the post layout process corner simulations. *Out0-Out3* are expected to go high at $t = 2400 \mu\text{s}$ and then go low again at $t = 2800 \mu\text{s}$. *Out4-Out7* should go high at $t = 3600 \mu\text{s}$ and remain high for the rest of the analysis.

Post layout process corners were first simulated for $VDD = 80 \text{ mV}$. All the 64B SRAM's output signals fail to produce a legal logic low value for SF50 and SF27. SF27 produces logic low values in the middle of the illegal range, while all output signals maintain a constant logic high value for SF50. Plots of *Out0-Out7* for all the simulated process corners are included in Subsection G.1.

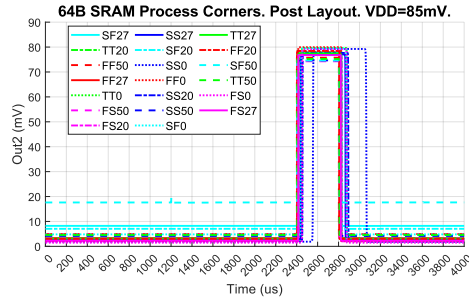
The post layout process corner simulation was then repeated for $VDD = 85 \text{ mV}$, and all points were found to pass. *Out0-Out7* are plotted in Figure 67 for all simulated post layout process corners. The delay for SS0 is much longer than for any of the other corners. SF50 produces the highest logic low values for all output signals, but it is still clearly below $V_{L,max} = 21.25 \text{ mV}$.



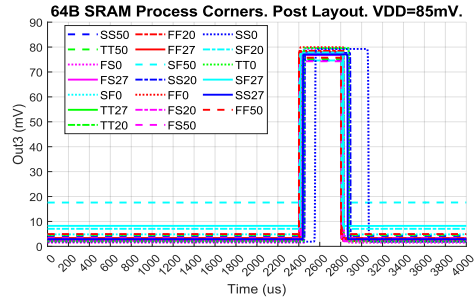
(a) Plot of Out0.



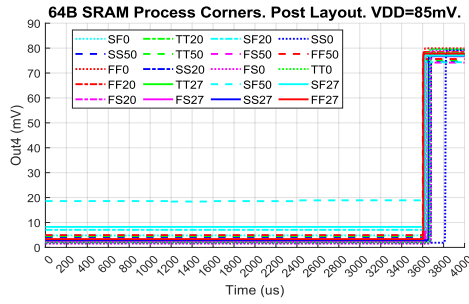
(b) Plot of Out1.



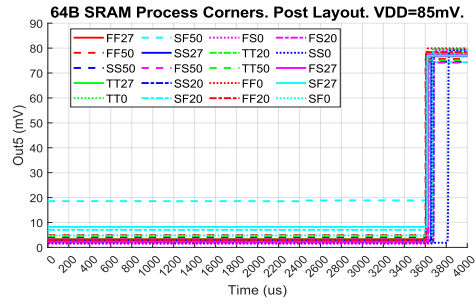
(c) Plot of Out2.



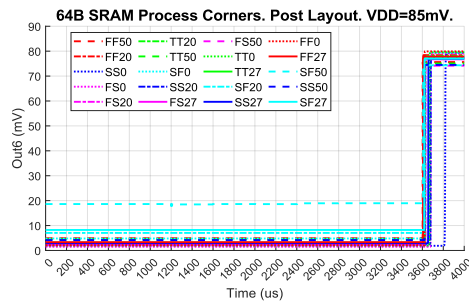
(d) Plot of Out3.



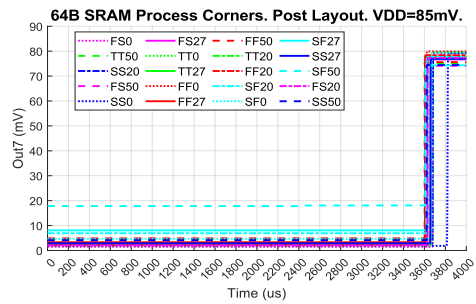
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

Figure 67: The 64B SRAM's output signals plotted for all post layout process corners simulated with $V_{DD} = 85 \text{ mV}$. All corners pass.

6 Discussion

6.1 Performance under Process Variation

For the post layout simulations, the 4B SRAM, 16B SRAM, and 64B SRAM all had some failing process corners with $VDD = 80$ mV and no failing process corners with $VDD = 85$ mV. The SF50 corner is the closest to failing for all, and the logic low value produced for SF50 increases slightly with each increase in SRAM size (from $V_{L,4B(SF50)} \approx 16.0$ mV, to $V_{L,16B(SF50)} \approx 16.6$ mV, to $V_{L,64B(SF50)} \approx 17.8$ mV for *Out7*). That the change is so small from one size to the next is very promising, as this means one could expect a larger size (e.g. 256B, which is the next logical jump) to also pass all process corner simulations at this supply voltage.

The SF corner was found to be the worst post layout process corner for all circuits. Though the FS corner (FS50 in particular) was found to produce the lowest logic high values in many of the post layout process corner simulations, it always performed much better than the SF corner. This considerable difference in performance between the SF and FS corners indicates that there must be a strength difference between the networks in the nominal corner, where the pull-down network is weaker than the pull-up network. This is very undesirable, as it has likely caused a large degradation in the performance of all circuits. An effort was made to make the NOT's VTC and the NAND's VTC symmetrical, see Subsection 4.1, precisely to avoid such a difference. The question is therefore why the imbalance between the pull-up and pull-down networks is still present, and what can be done to improve it.

The sizing strategy chosen was to use the same base width for all transistors, both PMOS and NMOS. As the slvt PMOS and rvt NMOS were found to have similar threshold voltages, it was assumed that they could be treated as having the same driving strength. The relative strengths within the pull-up and pull-down networks were then tweaked to get a symmetrical VTC, mainly by increasing the size of the feedback transistor *P2* as a way of weakening the pull-up network. Though it improved the performance in the nominal corner, it did not actually change the driving strength of the PMOS transistors in the pull-up network. This is likely to be the explanation as to *why* the SF corner had a much worse performance than the FS corner in the post layout simulations. This is also supported by Melek, who in [20] stresses the importance of having the same driving strength for corresponding NMOS and PMOS transistors in order to achieve optimal behaviour.

Solving this problem must therefore mean changing the sizing strategy slightly, so that the PMOS and NMOS transistors actually have the same driving strength in the nominal corner. One way of doing this could be to use a different base width for PMOS and NMOS, instead of having one base width for all transistors. The base widths could then be chosen so that the strength of a PMOS with base width $W_{base,PMOS}$ is equal to the strength of a NMOS with base width $W_{base,NMOS}$, and it would not be necessary to alter the optimal strength relations within the pull-up and pull-down network to get a symmetrical VTC. High layout regularity was the main argument for using a common base width for all transistors, but this would still be achieved to a large degree with two different base widths.

Even after improving this, which hopefully would make SF and FS equally good/bad so that

the voltage supply can be decreased further, the performance in the skewed corners (FS and SF) are likely to be a limiting factor when minimising VDD . To reduce the output level deviation in these corners, the feedback transistors can be made stronger so that the leakage from the stronger network is decreased [4].

6.2 Effect of Transistor Mismatch on the Minimum Supply Voltage

The minimum supply voltage is also affected by transistor mismatch. For $VDD = 80$ mV, an estimated yield of 97% was found for the post layout 4B SRAM, while the yield was estimated to 97.6% for the post 16B SRAM. Fewer points were simulated for the 16B SRAM, so this yield estimate is somewhat less reliable. Regardless of this, it is clear that the two circuits have a very similar performance. The overall circuit yield is expected to decrease for larger circuits, as more transistors are used and the likelihood of one or more of them causing a critical error increases. Additionally, a longer critical path in the larger SRAMs creates more opportunities for the signal level to be degraded. There are several possible reasons as to why this was not observed for the 4B and 16B SRAMs. The first is that the number of Monte Carlo points simulated might be insufficient, which would cause the yield estimate to be unreliable. This hypothesis can be tested by running more post layout Monte Carlo simulations on the 16B SRAM, which will be a bit time consuming but fully possible.

Secondly, only reading/writing from/to two different memory addresses was tested. This means that half the 4B SRAM was tested, but only an eighth of the 16B SRAM. Errors in any of the Output Selection Modules would still be noticed, as all these are connected to the critical path, as well as a large proportion of errors caused by the Decoder. But any errors in the bitcells that are not written to/read from will go unnoticed. The transient analysis needs to be much more extensive if it is going to test all bitcells, and this will cause the Monte Carlo simulations to be even more time-consuming.

A third possibility is that the yield actually remains relatively constant when increasing the memory size. An argument in the favour of this possibility is that an increase in memory mainly means a large increase in the number of D latches used. The D latch was found to have a post layout yield of 99.6% for $VDD = 80$ mV, where all of the failing points were very close to passing. So close in fact, that they are highly unlikely to be misinterpreted. An increase in the number of D latches should therefore not be detrimental to the SRAM's yield. The same goes for the Decoder, which at $VDD = 80$ mV was estimated to have a yield of 96.3%. This is the worst yield of all the subcircuits, but the failing points are quite close to passing just as they are for the D latch at this supply voltage. It is therefore very likely that the yield estimates found for the 4B SRAM and 16B SRAM are quite accurate, even if only some of the addresses were tested.

6.3 Pre Layout and Post Layout Differences

Post layout simulations are more accurate, as these include the effects of parasitics in the circuit. It was therefore expected that the post layout simulations might require a higher supply voltage to pass all process corners and to achieve a decent circuit yield with regards to transistor mismatch.

It was not expected that the process corners failing pre layout would be different from the process corners failing post layout, but this was observed for both the NAND and the Output Selection Module. For both, FS50 was worst pre layout and SF50 worst post layout. All other circuits were worst for SF50 both pre and post layout. The NAND is an important component in the Output Selection Module, as well as being used to create the stimuli for the Output Selection Module *DUT*. As the SF and FS corner had quite similar performances in the pre layout NOT, and the SF corner is worst in the post layout NOT, it is most likely the NAND that causes this difference between pre and post layout for the Output Selection Module.

When the VTC curve was simulated for the pre layout NAND, see Subsection 4.1.4, it was at first shifted to the left which confirms that the pull-down network was stronger initially. Widths were then tweaked to help move the VTC to the right, but it still ended up slightly unsymmetrical. In the FS corner, the NMOS transistors become even stronger and the PMOS transistors weaker which increases the pre-existing difference in strength between the two networks. For the initial layout configuration of the NAND, the VTC was shifted quite a lot to the right which means that the pull-up network was stronger. This was mainly improved by increasing $\frac{P2}{P0}$. As discussed in Subsection 6.1, a wiser strategy would have been to instead use different base widths for PMOS and NMOS. Increasing *P2* would then improve the performance in the SF corner more effectively. This difference in the relative strengths of the pull-up network and pull-down network between the pre layout NAND and post layout NAND explains why the FS corner was worst pre layout and the SF corner worst post layout.

Most post layout designs required the supply voltage to be increased by approximately 5 mV compared to the pre layout designs. This is quite small, and might just be due to the inclusion of the effect of parasitic devices in the post layout simulations. But it might also be a result of the layout strategy chosen, or more specifically the Length of Diffusion effect which cause a shift in V_{th} . Even though attempts were made to compensate for this in the layout so that the VTC was balanced, it might not have been enough to counter the effects of LOD on the circuits. This suspicion is further supported by comparing the logic high levels achieved for $AB = 10$ and $AB = 01$ in the post layout NAND, which are found to be slightly different (see Subsection 5.2.2. These were equal in the pre layout simulations, which indicates that something has happened during layout to change this.

A better layout methodology might have been to create non-merged layout, i.e. to only allow transistors that are really fingers of the same larger transistor device to share diffusion regions with each other. This would create a symmetry within each transistor, and the transistor's fingers would all experience a similar stress [19]. This would increase the area somewhat, as extra dummies and space would have to be added. On the other hand, if this made the transistors in the layout behave more like they did in the schematic it might not be necessary to increase

the widths as much to compensate for change in V_{th} and this would reduce the overall area. The increase in area might therefore not be very significant.

6.4 Comparison with State of the Art

The D latch was found to be operational for a minimum supply voltage of 80 mV (post layout). In comparison, a D flip-flop (DFF) created with ST logic gates was found in [14] to work with a supply voltage as low as 61 mV. There is in other words room for improvement in the design presented here. It is, however, important to note that a 130 μm CMOS technology was used in [14], while a 22 nm FD-SOI technology has been used here. As larger transistor technologies are more robust, it might not be possible to obtain as good a result with 22 nm FD-SOI. But there is clearly room for improvement regardless. Improving the robustness of the NAND gate must be the main focus, as this had the worst post layout performance in the process variation simulation of any of the circuits tested. The change in sizing strategy, as discussed in Subsection 6.1, together with an increased strength for the $P2$ transistors (improving performance in the SF corner) should improve the performance of both the NAND and the NOT, and therefore also of all the other circuits.

Robustness of a circuit is affected by the transistor technology used. It is therefore most fair to compare the performance of the SRAMs designed here to the performance of other SRAMs designed using 22nm FD-SOI transistors. A 7T SRAM was found in [11] to be able to operate at $VDD = 300$ mV and retain data at $VDD = 240$ mV. Compared to this, the 85 mV supply voltage achieved for the SRAMs here is quite good. This large improvement in V_{min} has cost a lot of area as each bitcell (D latch) uses 4 ST NANDs and 5 ST NOTs, which adds up to a total of 70 transistors. This does not include the logic needed for reading and writing. Compared to the 7 transistors used in the bitcell in [11] this is a lot. The DFF that achieved $VDD_{min} = 61$ mV in [14] used primarily ST NANDs and ended up with 90-100 transistors, which is in the same size range as the D latch design presented here.

7 Conclusion

A Static Random Access Memory (SRAM) circuit operating at a minimum supply voltage of 85 mV has been designed using a 22 nm FD-SOI (Fully Depleted Silicon On Insulator) transistor technology, with a D latch created from NOTs and NANDs as bitcell. The SRAM was designed to be quite modular, so that it is easy to create different sized memories.

Monte Carlo mismatch simulations were run on the 4B SRAM and the 16B SRAM, and good post layout yields were achieved for supply voltages as low as 80 mV with $yield_{4BSRAM,80mV} = 97\%$ and $yield_{16BSRAM,80mV} = 97.6\%$. The performance of the SRAM's subcircuits indicate that the yield will remain high for larger SRAMs as well.

Process variations were simulated by testing for all corners (TT, SS, FF, SF, and FS) at temperatures as low as 0°C and as high as 50°C. For the post layout simulations, the SF corner at high temperatures required the largest supply voltage to pass for all circuits. Improving the transistor sizing strategy is likely to improve the performance in this corner. The 4B SRAM, 16B SRAM, and 64B SRAM all required $VDD_{min} = 85$ mV to pass all corners. The increase in the logic low value in the SF50 corner was very slight from one size to the next, and larger sized SRAMs are therefore also expected to perform well at $VDD = 85$ mV.

The results are likely to be improved by balancing the driving strengths of NMOS and PMOS better, and to change the layout strategy to using non-merged devices instead of merged devices to reduce the effect of stress on the physical layout.

8 Suggestions for Further Work

- Change to using a different base width for NMOS and PMOS, as discussed in Subsection 6.1, so that the pull-up and pull-down networks can be balanced without changing $\frac{N_0}{N_1}$ and $\frac{N_2}{N_0}$ (or $\frac{P_0}{P_1}$ and $\frac{P_2}{P_0}$). This is likely to improve the robustness of the NOT and NAND, and in extension improve the robustness of all the other circuits as well.
- Change from using merged transistors to non-merged transistors, as this is likely to impact the threshold voltage less. Test to see if this improves the performance of the NOT and NAND gates compared to what was achieved by using merged transistors.
- Increase $\frac{P_2}{P_0}$ to improve the performance in the SF corner.
- To get a more reliable yield estimate and a better understanding of how local and global variations affect the SRAM's performance, transistor mismatch should be estimated in combination with process variation as these might cancel each other out or make each other worse.
- Perform more post layout Monte Carlo mismatch simulations on the 16B SRAM, to increase the confidence of the yield estimate.
- Perform Monte Carlo simulations on the 4B and 16B SRAM using a more extensive transient analysis, so that all parts of the circuit is tested. Use this to estimate how the circuit's yield scales with increased size.
- Create a script that automatically generates a SRAM with the desired size. This should not be too difficult for sizes $4^x \cdot 4B$ since the SRAM created is very modular.

References

- [1] Mohammad Hasan. *State of IoT 2022: Number of connected IoT devices growing 18% to 14.4 billion globally*. en-US. May 2022. URL: <https://iot-analytics.com/number-connected-iot-devices/> (visited on 05/12/2022).
- [2] Rice U ECE. *Prof. Benton Calhoun - Self Powered System Design for Next Generation Wireless Sensors*. Sept. 2020. URL: <https://www.youtube.com/watch?v=qhI8Ddn3I10> (visited on 05/12/2022).
- [3] Sayeed Ahmad, Naushad Alam and Mohd Hasan. “Robust TFET SRAM cell for ultra-low power IoT application”. In: *2017 International Conference on Electron Devices and Solid-State Circuits (EDSSC)*. Oct. 2017, pp. 1–2. DOI: 10.1109/EDSSC.2017.8333263.
- [4] Niklas Lotze and Yiannos Manoli. “Ultra-Sub-Threshold Operation of Always-On Digital Circuits for IoT Applications by Use of Schmitt Trigger Gates”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 64.11 (Nov. 2017). Conference Name: IEEE Transactions on Circuits and Systems I: Regular Papers, pp. 2920–2933. ISSN: 1558-0806. DOI: 10.1109/TCSI.2017.2705053.
- [5] Xin Fan et al. “Synthesizable Memory Arrays Based on Logic Gates for Subthreshold Operation in IoT”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 66.3 (Mar. 2019). Conference Name: IEEE Transactions on Circuits and Systems I: Regular Papers, pp. 941–954. ISSN: 1558-0806. DOI: 10.1109/TCSI.2018.2873026.
- [6] Koji Nii. “Ultra-Low Standby Power Embedded SRAM Design Techniques for Smart IoT Applications”. In: *2019 IEEE 11th International Memory Workshop (IMW)*. ISSN: 2573-7503. May 2019, pp. 1–1. DOI: 10.1109/IMW.2019.8739660.
- [7] A. Skirbekk. *Design of Ultra-Low Voltage SRAM in 22nm FD-SOI*. Project report in TFE4580. Department of Electronic Systems, NTNU – Norwegian University of Science and Technology, Dec. 2022.
- [8] Sylvain Clerc, Thierry Di Gilio and Andreia Cathelin. *The Fourth Terminal*. Springer, 2020. ISBN: 978-3-030-39495-0.
- [9] Thomas Skotnicki et al. “Innovative Materials, Devices, and CMOS Technologies for Low-Power Mobile Multimedia”. In: *IEEE Transactions on Electron Devices* 55.1 (Jan. 2008). Conference Name: IEEE Transactions on Electron Devices, pp. 96–130. ISSN: 1557-9646. DOI: 10.1109/TED.2007.911338.
- [10] Andrew S. Tanenbaum and Todd Austin. *Structured Computer Organization*. English. 6th (international edition). Pearson, 2013. ISBN: 978-0-273-76924-8.
- [11] Somayeh Hossein Zadeh, Trond Ytterdal and Snorre Aunet. “Comparative Study of Single, Regular and Flip Well Subthreshold SRAMs in 22 nm FDSOI Technology”. In: *2020 IEEE Nordic Circuits and Systems Conference (NorCAS)*. Oct. 2020, pp. 1–6. DOI: 10.1109/NorCAS51424.2020.9265001.
- [12] Jaydeep P. Kulkarni, Keejong Kim and Kaushik Roy. “A 160 mV, fully differential, robust schmitt trigger based sub-threshold SRAM”. In: *Proceedings of the 2007 international symposium on Low power electronics and design (ISLPED '07)*. Aug. 2007, pp. 171–176. DOI: 10.1145/1283780.1283818.

-
- [13] Tony Chan Carusone, David Johns and Kenneth Martin. *Analog Integrated Circuit Design*. English. 2nd. WILEY, 2013.
- [14] Niklas Lotze and Yiannos Manoli. “A 62 mV 0.13 μ m CMOS Standard-Cell-Based Design Technique Using Schmitt-Trigger Logic”. In: *IEEE Journal of Solid-State Circuits* 47.1 (Jan. 2012). Conference Name: IEEE Journal of Solid-State Circuits, pp. 47–60. ISSN: 1558-173X. DOI: 10.1109/JSSC.2011.2167777.
- [15] R.M. Swanson and J.D. Meindl. “Ion-implanted complementary MOS transistors in low-voltage circuits”. In: *IEEE Journal of Solid-State Circuits* 7.2 (Apr. 1972). Conference Name: IEEE Journal of Solid-State Circuits, pp. 146–153. ISSN: 1558-173X. DOI: 10.1109/JSSC.1972.1050260.
- [16] Marcel J.M. Pelgrom and Aad C.J. Duinmaijer. “Matching properties of MOS transistors”. In: *ESSCIRC '88: Fourteenth European Solid-State Circuits Conference*. Sept. 1988, pp. 327–330. DOI: 10.1109/ESSCIRC.1988.5468276.
- [17] Kai Chen et al. “The impact of device scaling and power supply change on CMOS gate performance”. In: *IEEE Electron Device Letters* 17.5 (May 1996). Conference Name: IEEE Electron Device Letters, pp. 202–204. ISSN: 1558-0563. DOI: 10.1109/55.491829.
- [18] Ban P. Wong et al. *Nano-CMOS Circuit and Physical Design: Wong/Nano-CMOS*. en. Hoboken, NJ, USA: John Wiley & Sons, Inc., Nov. 2004. ISBN: 978-0-471-65382-0 978-0-471-46610-9. DOI: 10.1002/0471653829. URL: <http://doi.wiley.com/10.1002/0471653829> (visited on 16/12/2022).
- [19] P. G Drennan, M. L. Kniffin and D. R. Locascio. “Implications of Proximity Effects for Analog Design”. In: *IEEE Custom Integrated Circuits Conference 2006* (2006), pp. 169–176.
- [20] Luiz A.P. Melek, Márcio C. Schneider and Carlos Galup-Montoro. “Ultra-low voltage CMOS logic circuits”. In: *2014 Argentine Conference on Micro-Nanoelectronics, Technology and Applications (EAMTA)*. July 2014, pp. 1–7. DOI: 10.1109/EAMTA.2014.6906070.

Appendices

A NAND Results

The NAND failed on corner SF50 in the post layout process corner simulation at $VDD = 85$ mV. The NAND's output signal Y is plotted for all corners in Figure 68.

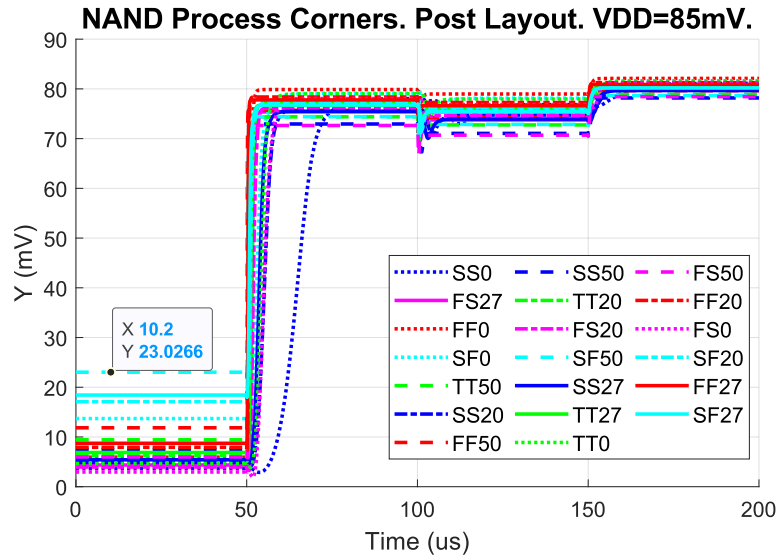


Figure 68: The NAND's output Y plotted for all corners. SF50 fails because the logic low value is slightly above $V_{L,max} = 0.25VDD = 21.25$ mV.

B Results from Simulations on the D Latch

Results from the pre layout process corner simulations at $V_{DD} = 75$ mV are presented in Figure 69.

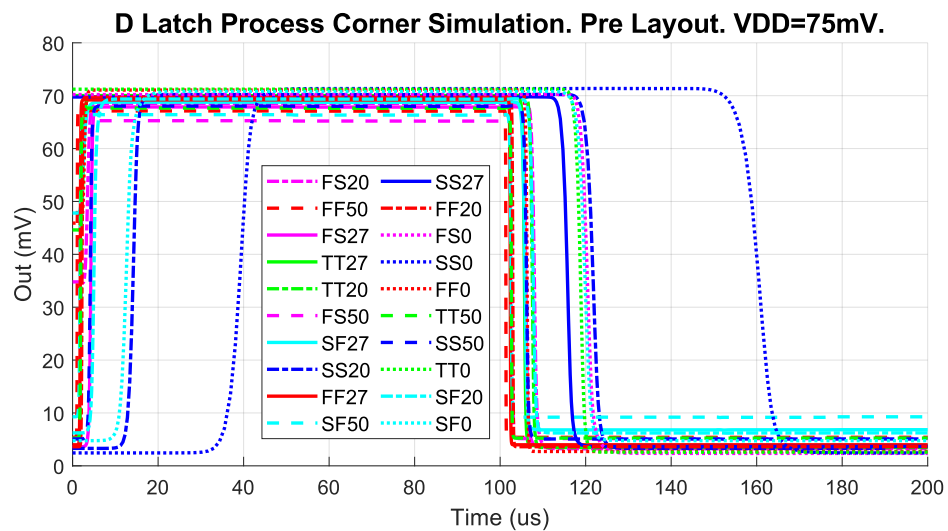


Figure 69: The pre layout D latch process variation results for $V_{DD} = 75$ mV. Note that SS0 has the longest delay. FS50 has the lowest logic high value, and SF50 the highest logic low value.

C Results from Simulations on the 2to4 Decoder

C.1 Monte Carlo Mismatch Simulation Results

Out0 is plotted in Figure 70a for all simulation points in the pre layout Monte Carlo Simulation at $VDD = 70$ mV, and the six failing points are plotted again in Figure 71a. *Out1* is plotted in Figure 70b for all simulation points in the pre layout Monte Carlo Simulation at $VDD = 70$ mV, and the seven failing points are plotted again in Figure 71b. *Out2* is plotted in Figure 70c for all simulation points in the pre layout Monte Carlo Simulation at $VDD = 70$ mV, and the three failing points are plotted again in Figure 71c. *Out3* is plotted in Figure 70d for all simulation points in the pre layout Monte Carlo Simulation at $VDD = 70$ mV, and the one failing point is plotted again in Figure 71d.

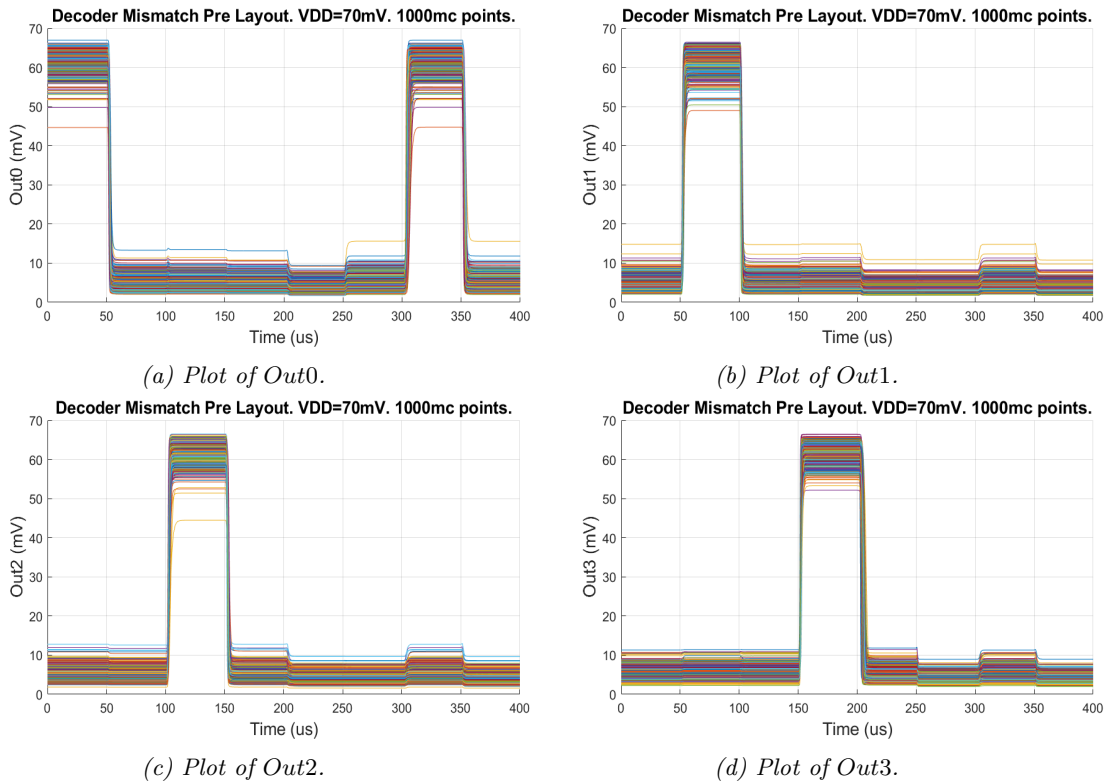


Figure 70: The DUT's output signals plotted for the 1000 pre layout Monte Carlo points simulated with $VDD = 70$ mV.

Pre layout Monte Carlo mismatch simulations were run for $VDD = 75$ mV, where all points tested were found to pass. Plots of *Out0*, *Out1*, *Out2*, and *Out3* for all 1000 simulation points are shown in Figure 72a, Figure 72b, Figure 72c, and Figure 72d.

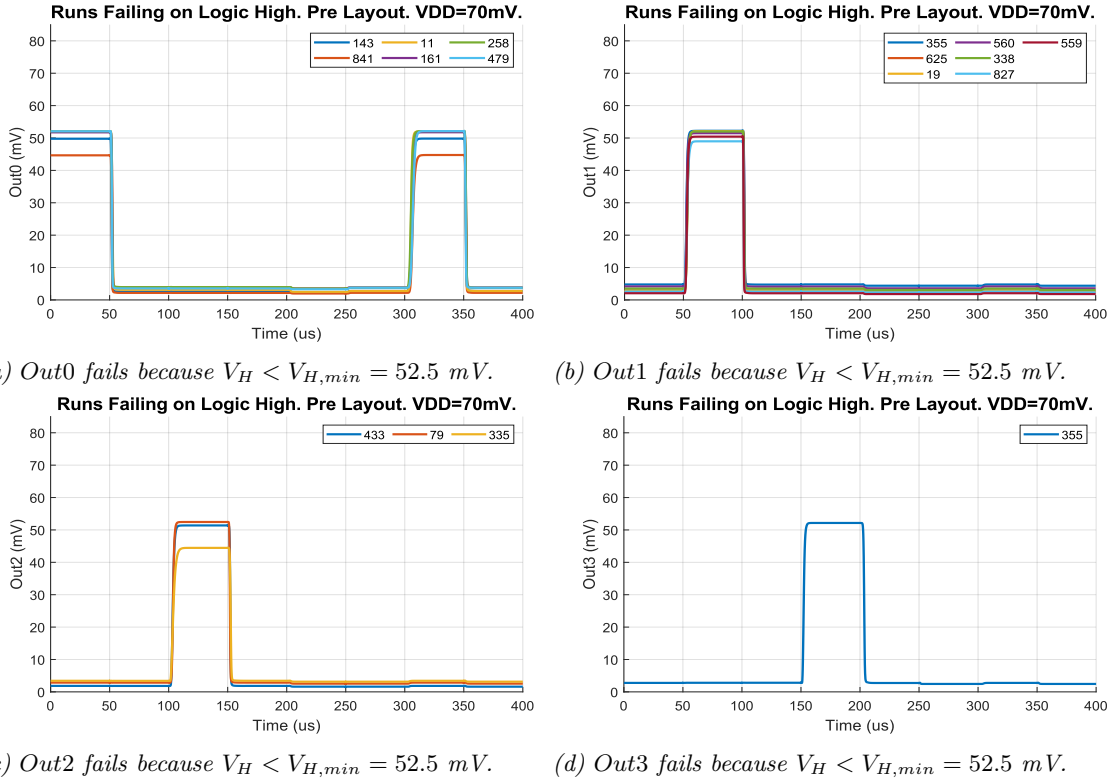


Figure 71: The DUT's output signals plotted for the failing points in the pre layout Monte Carlo simulation with $V_{DD} = 70$ mV.

Post layout Monte Carlo mismatch simulations were run for $V_{DD} = 80$ mV. The DUT's output signals were plotted for all 1000 simulated points: *Out0* in Figure 73a, *Out1* in Figure 73b, *Out2* in Figure 73c, and *Out3* in Figure 73d.

Post layout Monte Carlo mismatch simulations were run for $V_{DD} = 85$ mV. The DUT's output signals were plotted for all 1000 simulated points: *Out0* in Figure 74a, *Out1* in Figure 74b, *Out2* in Figure 74c, and *Out3* in Figure 74d.

C.2 Process Corner Simulation Results

Results from the pre layout process corner simulations at $V_{DD} = 70$ mV are presented for *Out0* in Figure 75a, *Out1* in Figure 75b, *Out2* in Figure 75c, and *Out3* in Figure 75d. The logic high value produced for SF50 is too low for all the DUT's output signals.

Results from the failing post layout process corner simulations at $V_{DD} = 75$ mV are presented for *Out0* in Figure 76a, *Out1* in Figure 76b, *Out2* in Figure 76c, and *Out3* in Figure 76d. The

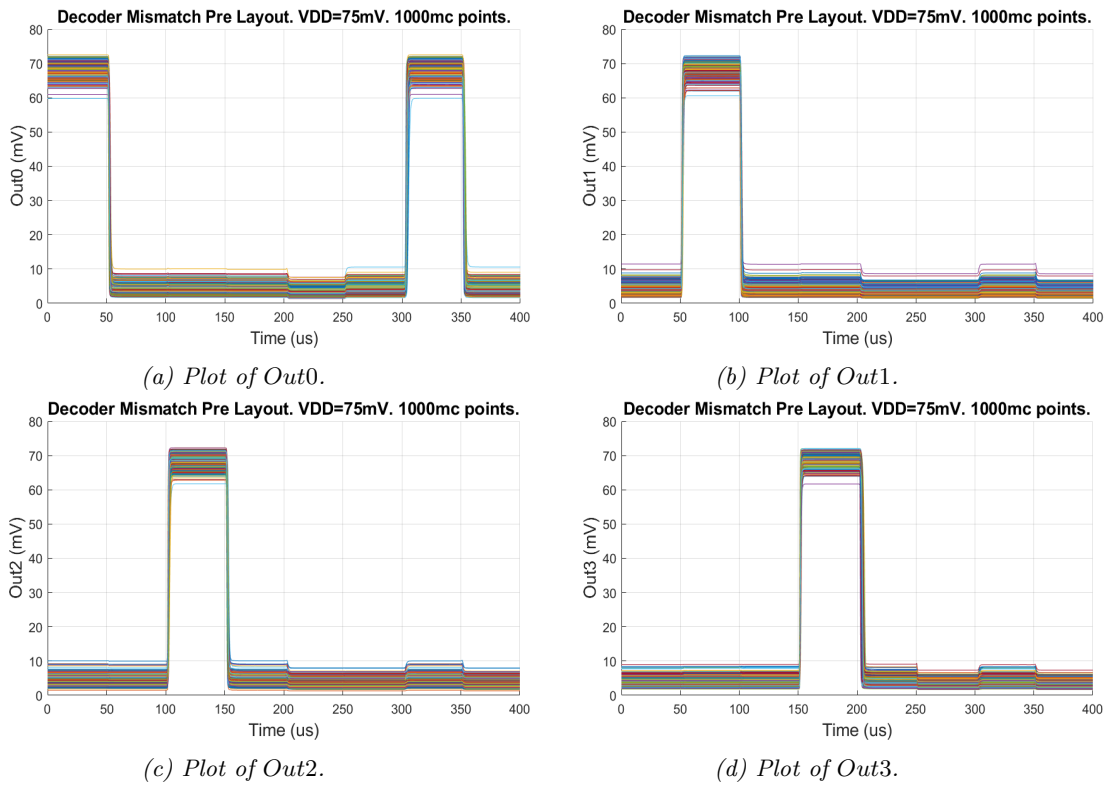
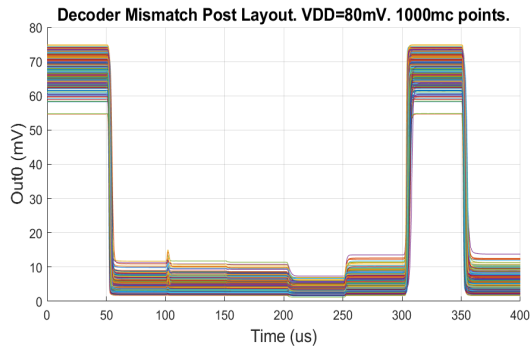


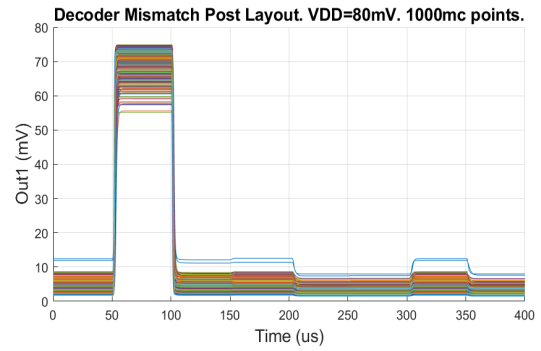
Figure 72: The DUT's output signals plotted for the 1000 pre layout Monte Carlo points simulated with $VDD = 75 \text{ mV}$.

logic high value produced for several of the corners is too low for all the DUT's output signals.

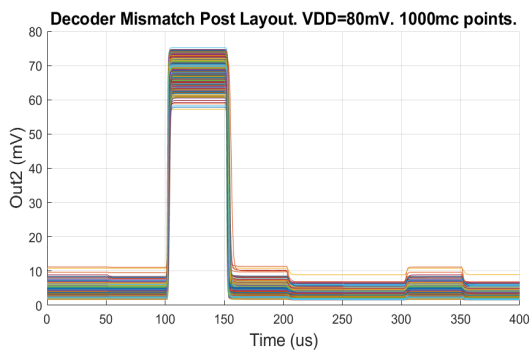
Results from the failing post layout process corner simulations at $VDD = 80 \text{ mV}$ are presented for *Out0* in Figure 77a, *Out1* in Figure 77b, *Out2* in Figure 77c, and *Out3* in Figure 77d. The logic high value produced for SF50 is too low for all the DUT's output signals.



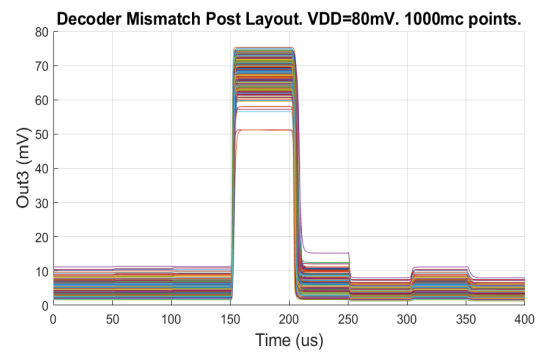
(a) Plot of Out0.



(b) Plot of Out1.

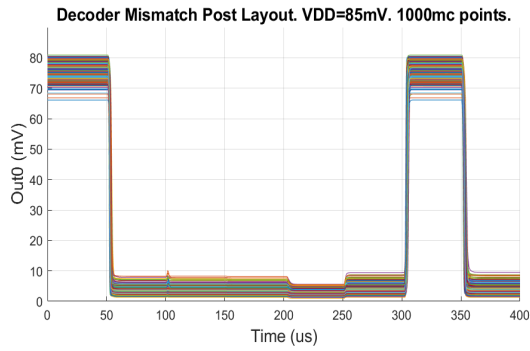


(c) Plot of Out2.

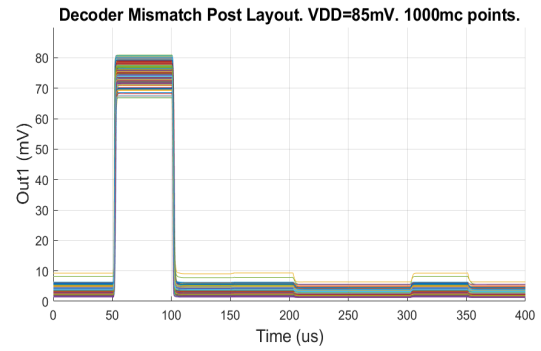


(d) Plot of Out3.

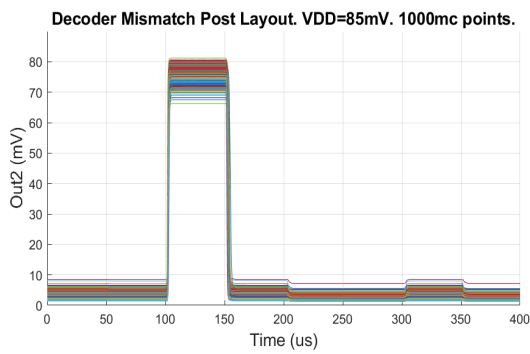
Figure 73: The DUT's output signals plotted for the 1000 post layout Monte Carlo points simulated with $V_{DD} = 80 \text{ mV}$.



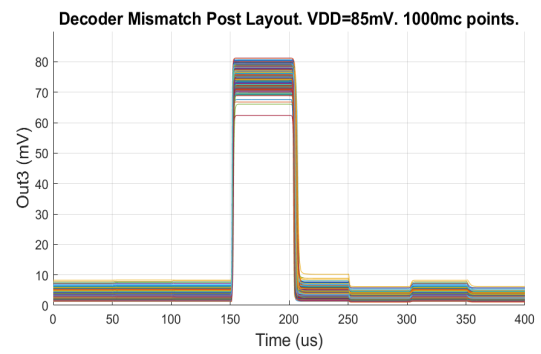
(a) Plot of Out0.



(b) Plot of Out1.

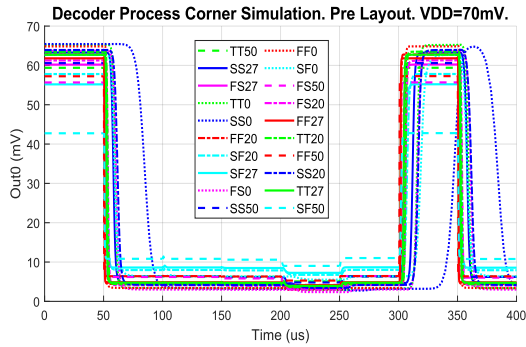


(c) Plot of Out2.

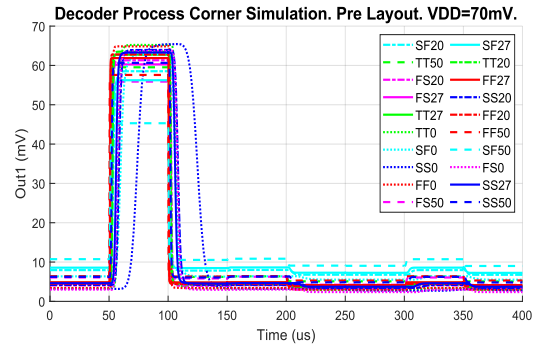


(d) Plot of Out3.

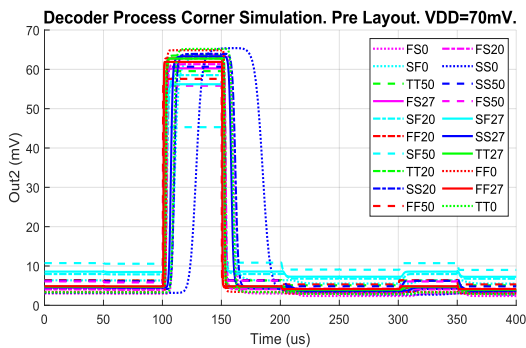
Figure 74: The DUT's output signals plotted for the 1000 post layout Monte Carlo points simulated with $VDD = 85\text{ mV}$.



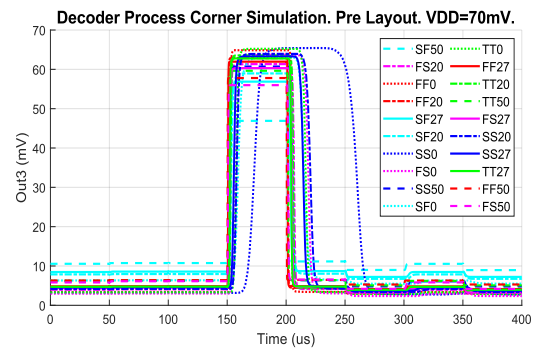
(a) Plot of Out0.



(b) Plot of Out1.

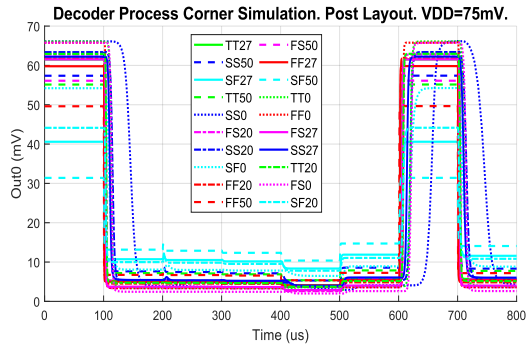


(c) Plot of Out2.

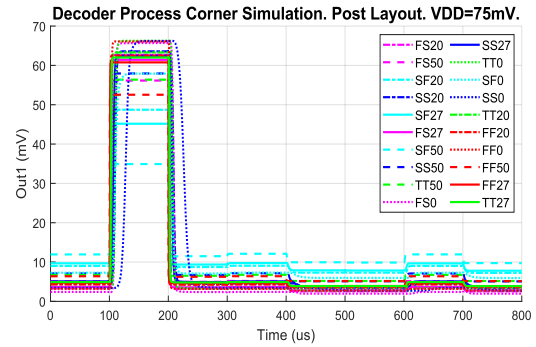


(d) Plot of Out3.

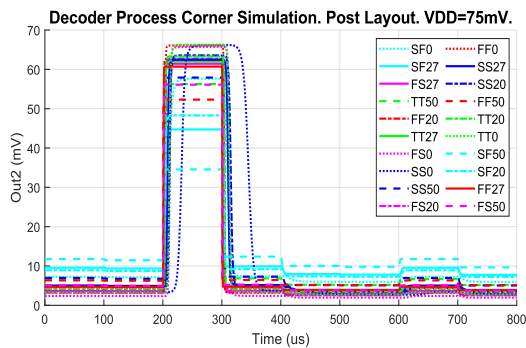
Figure 75: The DUT's output signals plotted for all post layout process corners at $V_{DD} = 70$ mV. Several corners fail to produce a logic high value larger than $V_{H,min} = 52.5$ mV.



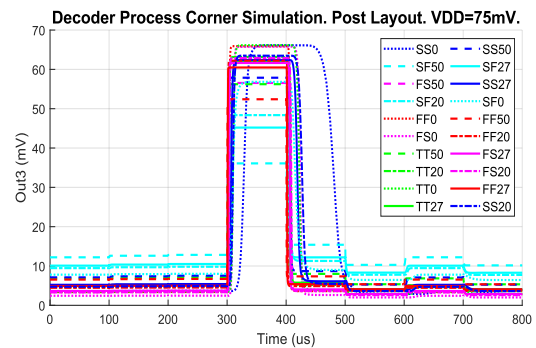
(a) Plot of Out0.



(b) Plot of Out1.



(c) Plot of Out2.



(d) Plot of Out3.

Figure 76: The DUT's output signals plotted for all post layout process corners at $V_{DD} = 75$ mV. Several corners fail to produce a logic high value larger than $V_{H,min} = 56.25$ mV.

D Output Selection Module Results

D.1 Monte Carlo Mismatch Simulation Results

All simulated points passed when running the pre layout mismatch simulation with a supply voltage $VDD = 85$ mV. $Y1$ is plotted for all simulated points in Figure 78a, and $Y2$ is plotted for all simulated points in Figure 78b.

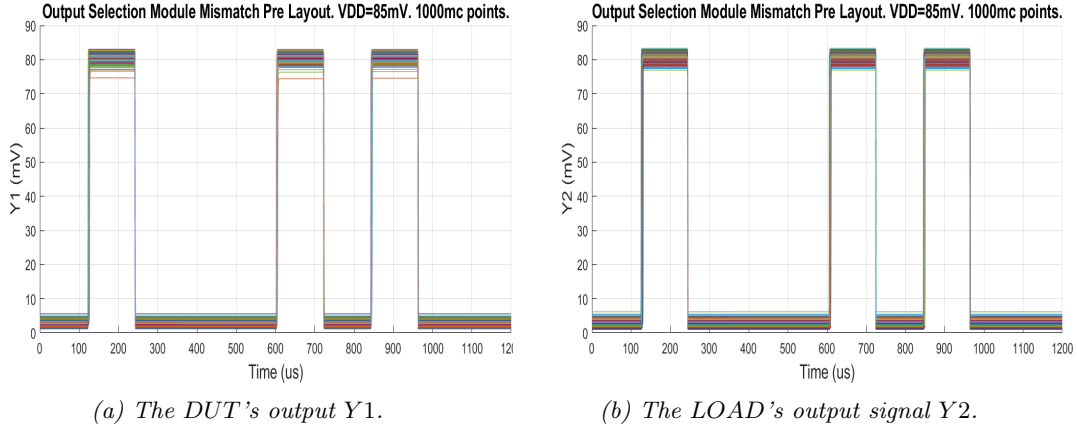


Figure 78: $Y1$ and $Y2$ are plotted for all pre layout Monte Carlo points simulated with $VDD = 85$ mV. All points pass.

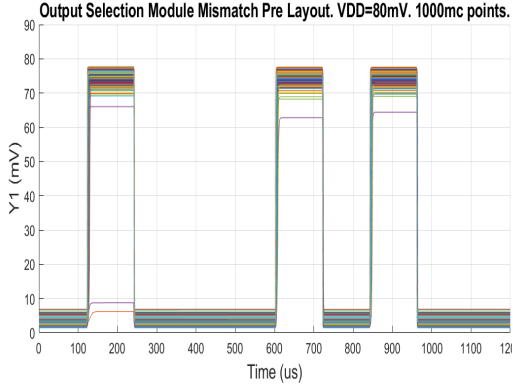
The pre layout mismatch simulation with $VDD = 80$ mV failed for simulation points 362 and 587. $Y1$ is plotted for all simulated points in Figure 79a and $Y2$ is plotted for all simulated points in Figure 79b.

All simulated points passed when running the post layout mismatch simulation with a supply voltage $VDD = 85$ mV. $Y1$ is plotted for all simulated points in Figure 80a, and $Y2$ is plotted for all simulated points in Figure 80b.

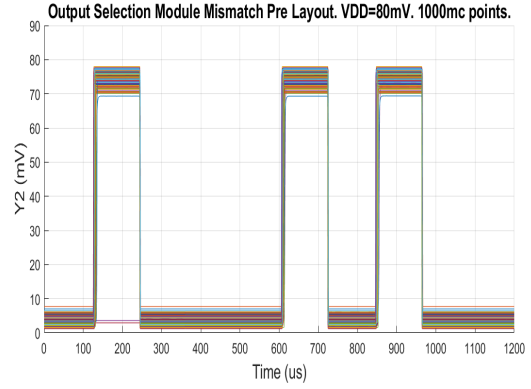
$Y1$ and $Y2$ are plotted in Figure 81 for all 1000 simulated points in the post layout mismatch simulation with $VDD = 80$ mV. 5 simulation points fail for both the DUT ($Y1$) and the LOAD ($Y2$)

D.2 Process Variation Simulation Results

$Y2$ is plotted in Figure 82a for all post layout process corners simulated with $VDD = 80$ mV, and in Figure 82b for all post layout process corners simulated with $VDD = 85$ mV

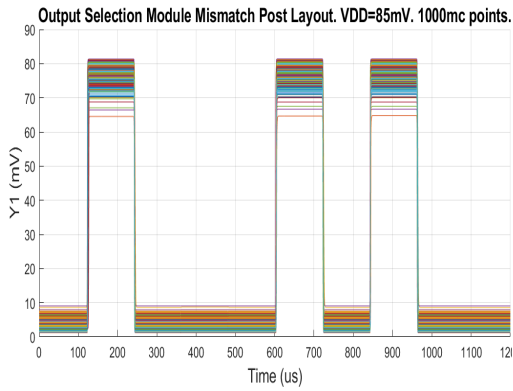


(a) The DUT's output Y1.

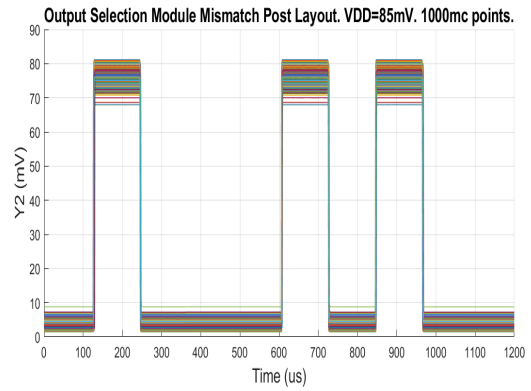


(b) The LOAD's output signal Y2.

Figure 79: Y1 and Y2 are plotted for all pre layout Monte Carlo points simulated with $V_{DD} = 80\text{ mV}$. Points 362 and 587 fail.

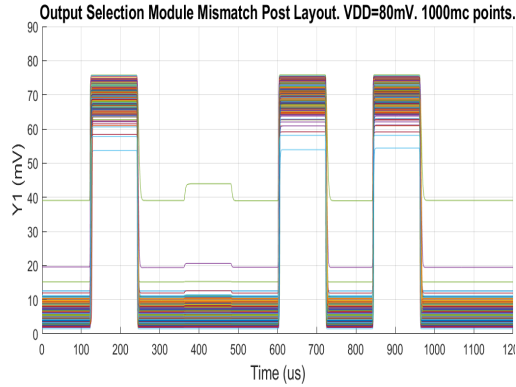


(a) The DUT's output Y1.

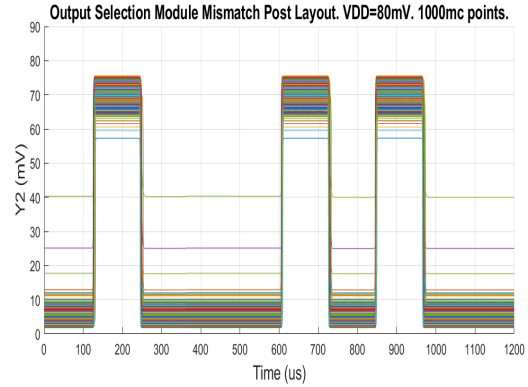


(b) The LOAD's output signal Y2.

Figure 80: Y1 and Y2 are plotted for all post layout Monte Carlo points simulated with $V_{DD} = 85\text{ mV}$. All points pass.

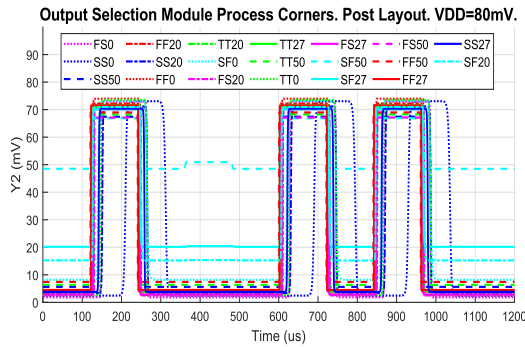


(a) The DUT's output Y1.

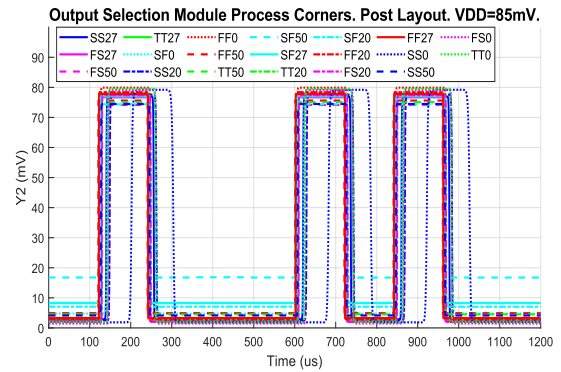


(b) The LOAD's output signal Y2.

Figure 81: Y1 and Y2 are plotted for all post layout Monte Carlo points simulated with $VDD = 80\text{ mV}$. All points pass.



(a) $VDD = 80\text{ mV}$. SF50 fails.



(b) $VDD = 85\text{ mV}$. All corners pass.

Figure 82: Plots of the LOAD's output Y2 for all post layout process corners simulated with $VDD = 80\text{ mV}$ and $VDD = 85\text{ mV}$.

E Results from Simulations on the 4B SRAM

E.1 Monte Carlo Mismatch Simulation Results

All simulated points passed for the pre layout mismatch simulation at $VDD = 75$ mV. *Out0* is plotted in Figure 83a for all 1000 simulated points, *Out1* is plotted in Figure 83b, *Out2* is plotted in Figure 83c, *Out3* is plotted in Figure 83d, *Out4* is plotted in Figure 83e, *Out5* is plotted in Figure 83f, *Out6* is plotted in Figure 83g, and *Out7* is plotted in Figure 83h.

All simulated points passed for the pre layout mismatch simulation of the 4B SRAM at $VDD = 80$ mV. *Out0* is plotted in Figure 84a for all 1000 simulated points, *Out1* is plotted in Figure 84b, *Out2* is plotted in Figure 84c, *Out3* is plotted in Figure 84d, *Out4* is plotted in Figure 84e, *Out5* is plotted in Figure 84f, *Out6* is plotted in Figure 84g, and *Out7* is plotted in Figure 84h.

The pre layout mismatch simulation of the 4B SRAM at $VDD = 70$ mV resulted in a couple of failed points, but most of them passed. *Out0* is plotted in Figure 85a for all 1000 simulated points, *Out1* is plotted in Figure 85b, *Out2* is plotted in Figure 85c, *Out3* is plotted in Figure 85d, *Out4* is plotted in Figure 85e, *Out5* is plotted in Figure 85f, *Out6* is plotted in Figure 85g, and *Out7* is plotted in Figure 85h.

All simulated points passed for the post layout mismatch simulation at $VDD = 85$ mV. *Out0* is plotted in Figure 86a for all 1000 simulated points, *Out1* is plotted in Figure 86b, *Out2* is plotted in Figure 86c, *Out3* is plotted in Figure 86d, *Out4* is plotted in Figure 86e, *Out5* is plotted in Figure 86f, *Out6* is plotted in Figure 86g, and *Out7* is plotted in Figure 86h.

Post layout mismatch simulations were run at $VDD = 85$ mV, and the 4B SRAM's output signals are plotted for all 1000 simulation points in Figure 86.

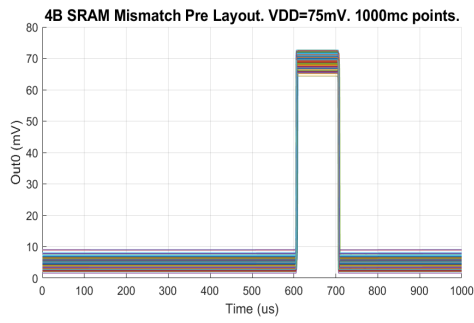
E.2 Process Corner Simulation Results

The 4B SRAM's output signals are plotted in Figure 88 for all the pre layout process corners simulated with $VDD = 70$ mV. SF20, SF27, and SF50 fail to produce a legal logic low value for all outputs.

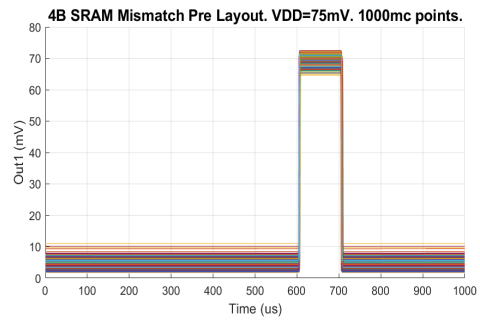
The 4B SRAM's output signals are plotted in Figure 89 for all the pre layout process corners simulated with $VDD = 80$ mV. All corners pass.

The 4B SRAM's output signals are plotted in Figure 90 for all the post layout process corners simulated with $VDD = 75$ mV. SF0, SF20, SF27, and SF50 fail to produce legal logic low values for all output signals. SF50 fails to produce a legal logic low for *Out4-Out7*. SF50 fails to produce a legal logic high value for *Out4-Out7*.

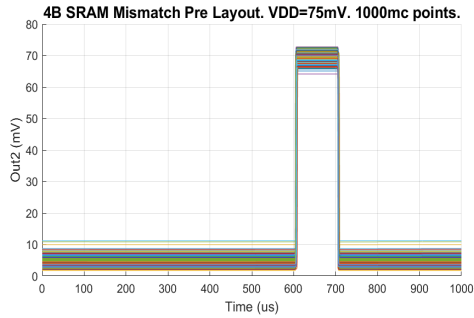
The 4B SRAM's output signals are plotted in Figure 91 for all the post layout process corners simulated with $VDD = 80$ mV. SF50 fails to produce legal logic low values for all output signals.



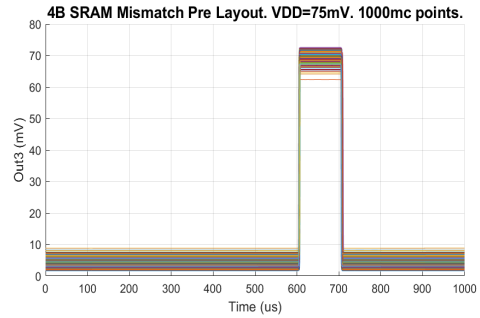
(a) Plot of Out0.



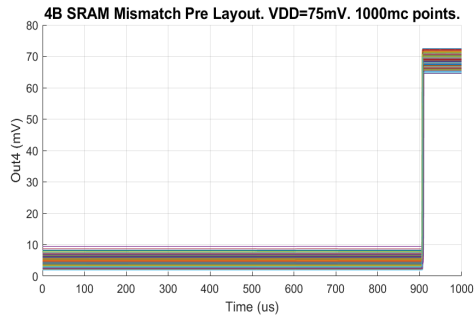
(b) Plot of Out1.



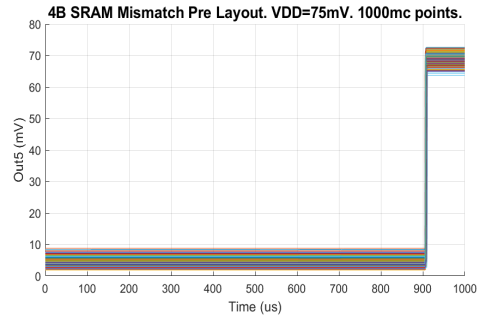
(c) Plot of Out2.



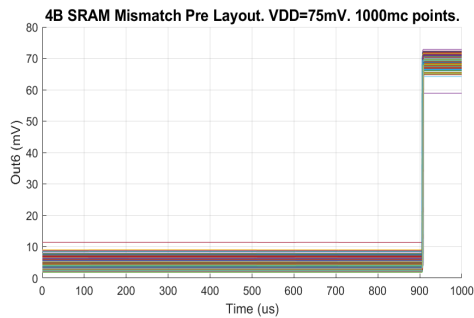
(d) Plot of Out3.



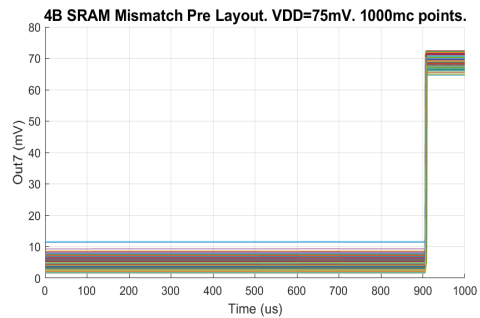
(e) Plot of Out4.



(f) Plot of Out5.

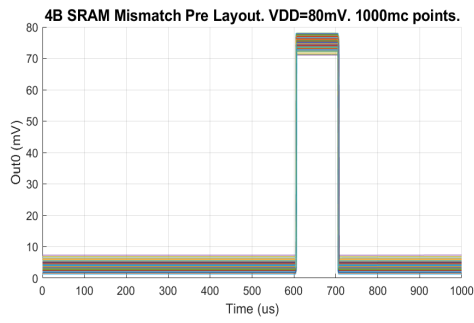


(g) Plot of Out6.

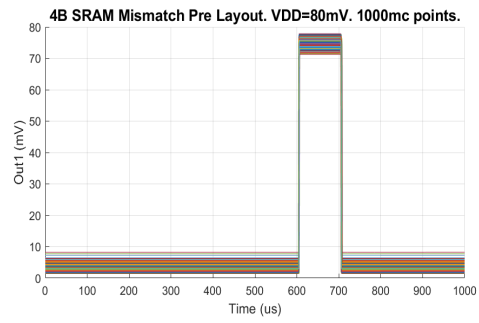


(h) Plot of Out7.

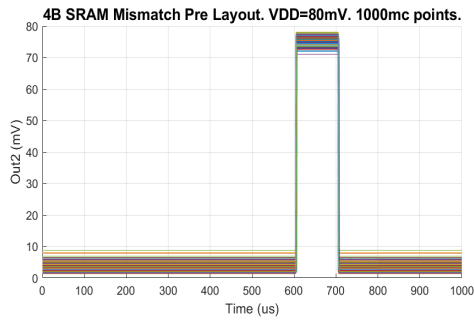
Figure 83: The 4B SRAM's output signals plotted for the 1000 pre layout Monte Carlo points simulated with $V_{DD} = 75 \text{ mV}$.



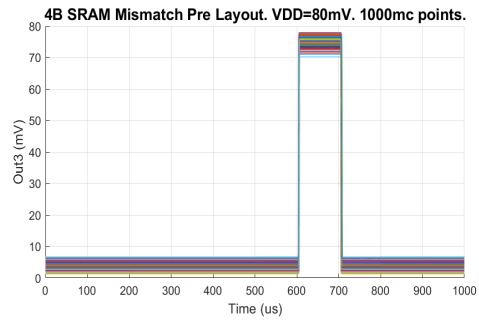
(a) Plot of Out0.



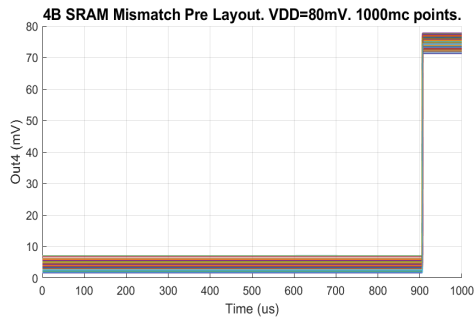
(b) Plot of Out1.



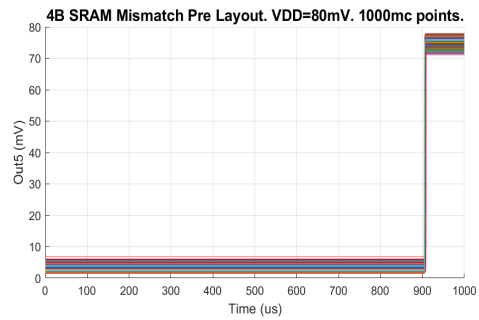
(c) Plot of Out2.



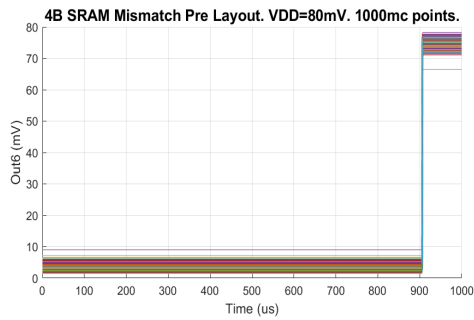
(d) Plot of Out3.



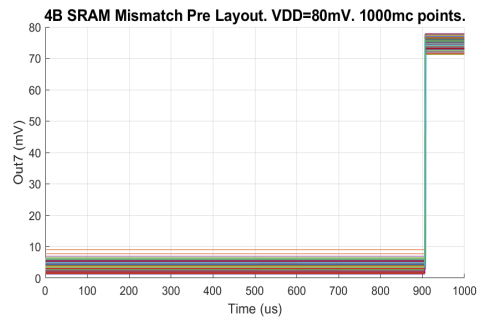
(e) Plot of Out4.



(f) Plot of Out5.

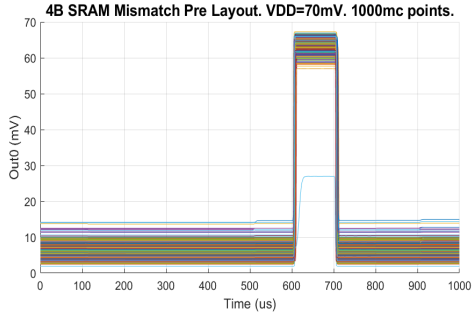


(g) Plot of Out6.

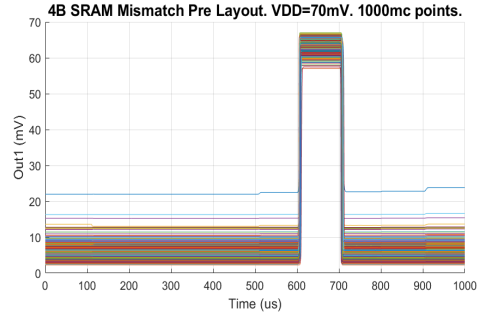


(h) Plot of Out7.

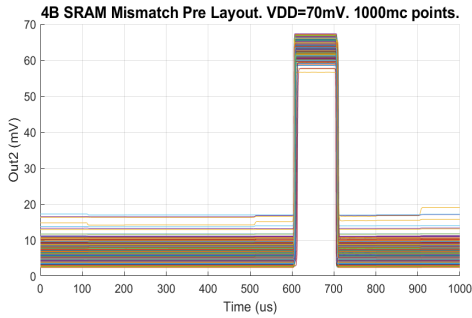
Figure 84: The 4B SRAM's output signals plotted for the 1000 pre layout Monte Carlo points simulated with $V_{DD} = 80\text{ mV}$.



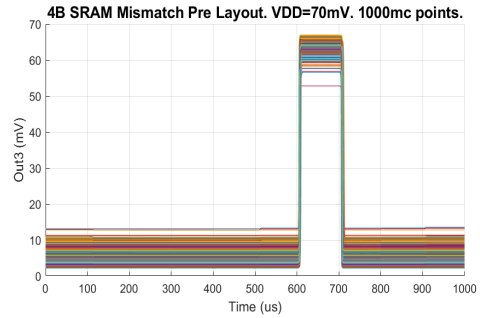
(a) Plot of Out0.



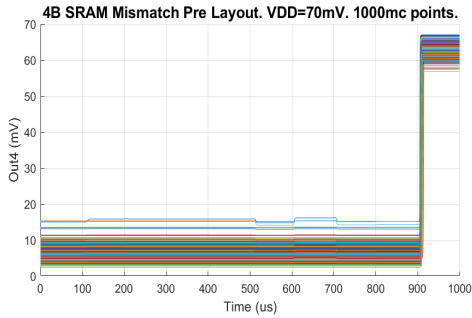
(b) Plot of Out1.



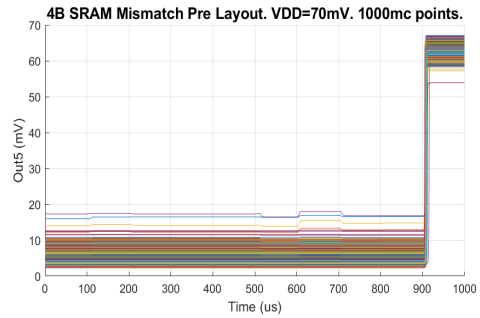
(c) Plot of Out2.



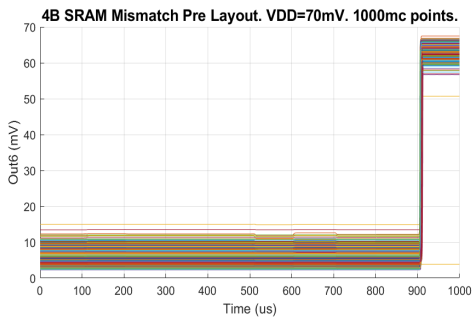
(d) Plot of Out3.



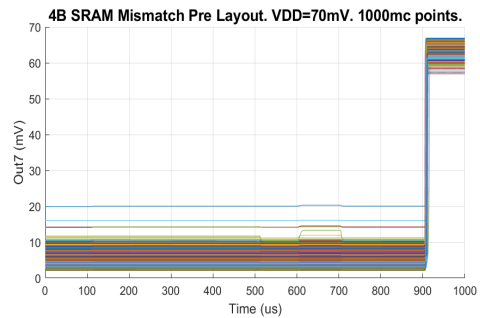
(e) Plot of Out4.



(f) Plot of Out5.

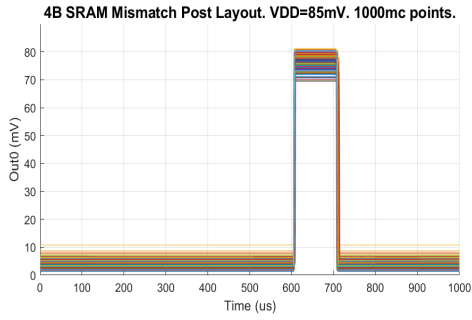


(g) Plot of Out6.

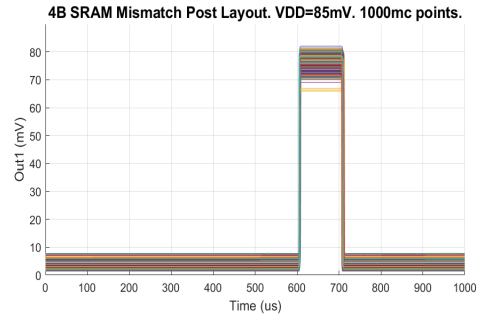


(h) Plot of Out7.

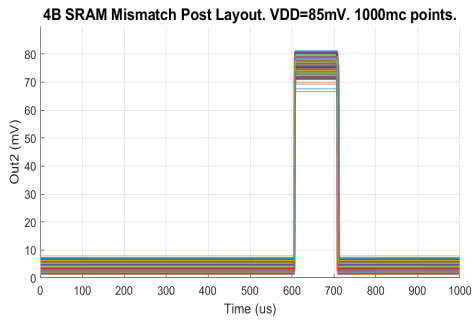
Figure 85: The 4B SRAM's output signals plotted for the 1000 pre layout Monte Carlo points simulated with $V_{DD} = 70 \text{ mV}$.



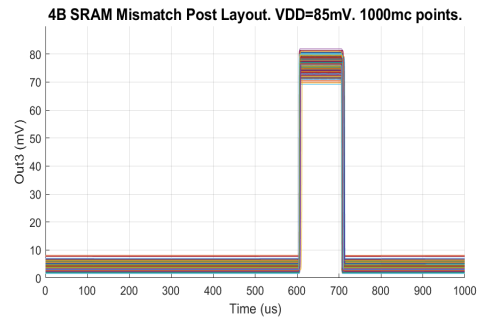
(a) Plot of Out0.



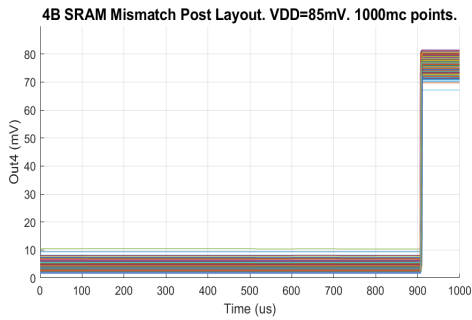
(b) Plot of Out1.



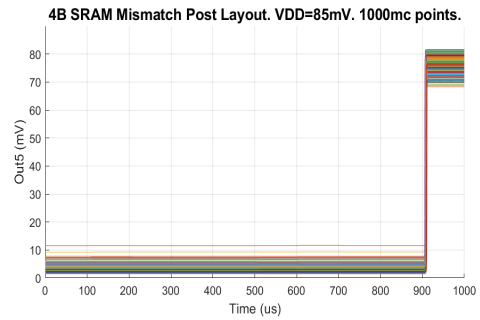
(c) Plot of Out2.



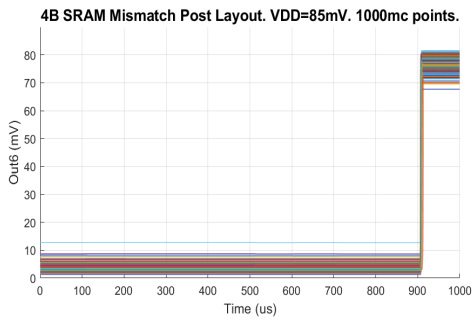
(d) Plot of Out3.



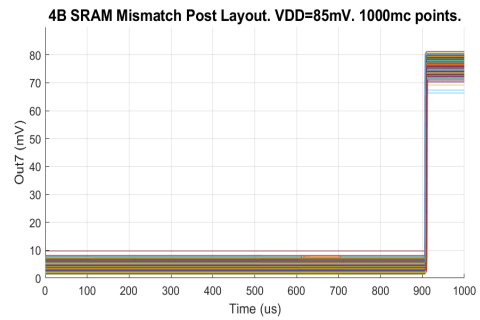
(e) Plot of Out4.



(f) Plot of Out5.

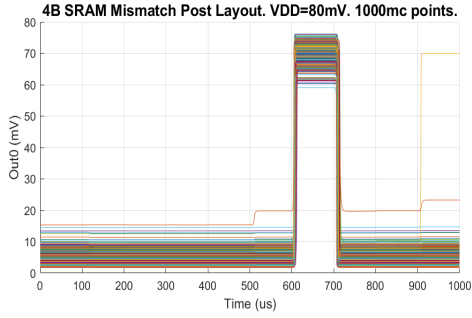


(g) Plot of Out6.

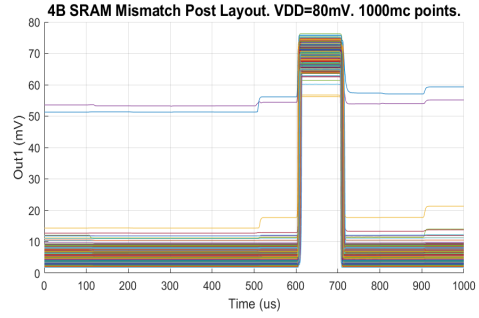


(h) Plot of Out7.

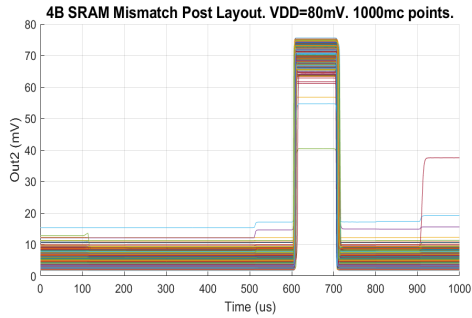
Figure 86: The 4B SRAM's output signals plotted for the 1000 post layout Monte Carlo points simulated with $V_{DD} = 85 \text{ mV}$.



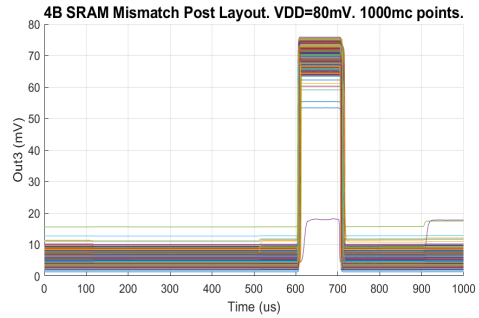
(a) Plot of Out0.



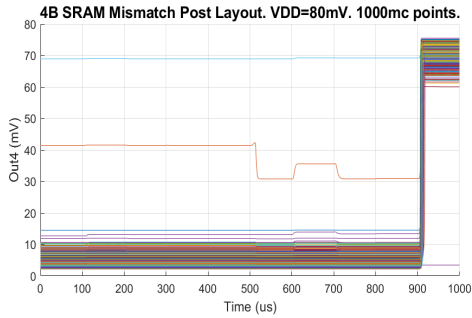
(b) Plot of Out1.



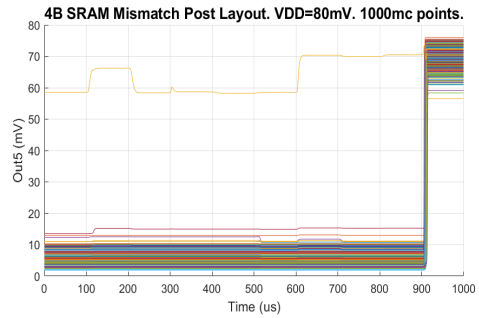
(c) Plot of Out2.



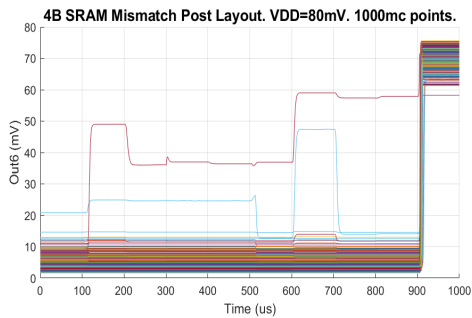
(d) Plot of Out3.



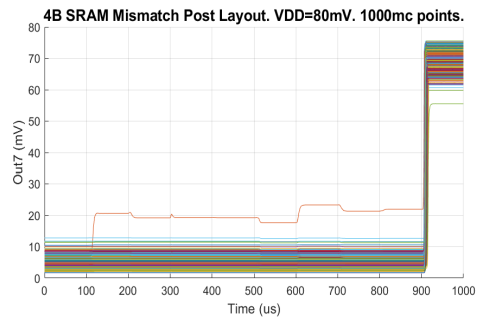
(e) Plot of Out4.



(f) Plot of Out5.

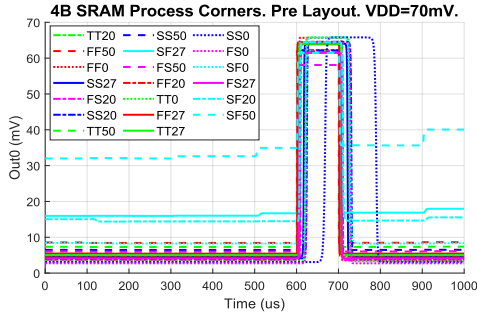


(g) Plot of Out6.

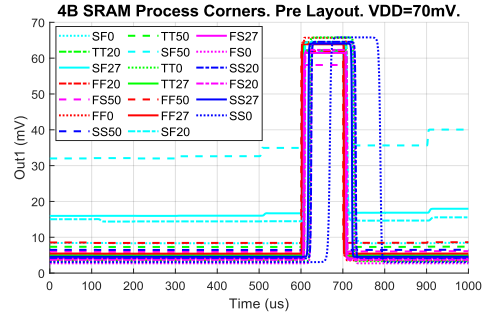


(h) Plot of Out7.

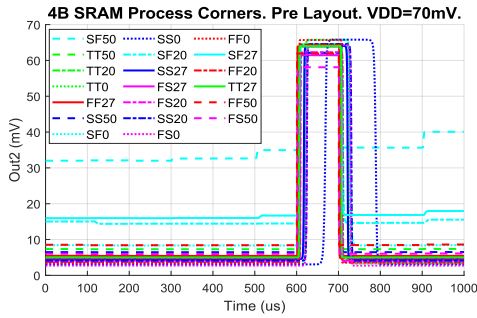
Figure 87: The 4B SRAM's output signals plotted for the 1000 post layout Monte Carlo points simulated with $V_{DD} = 80\text{ mV}$.



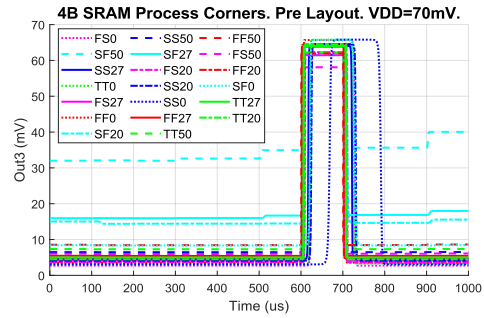
(a) Plot of Out0.



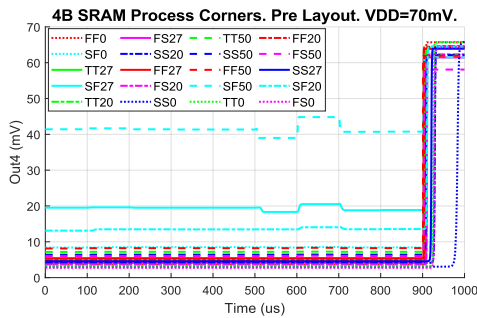
(b) Plot of Out1.



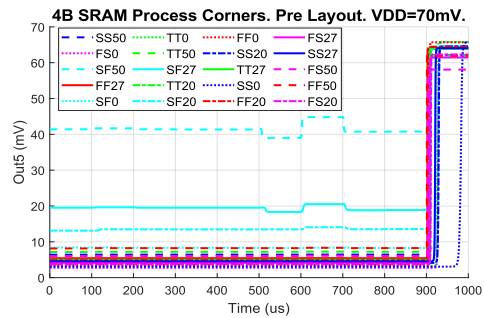
(c) Plot of Out2.



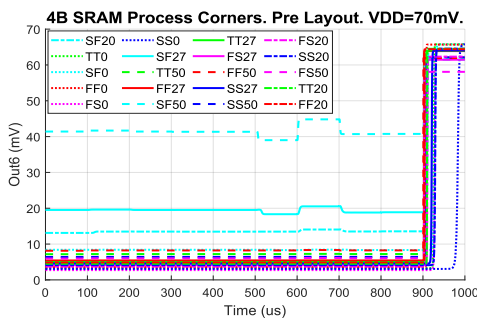
(d) Plot of Out3.



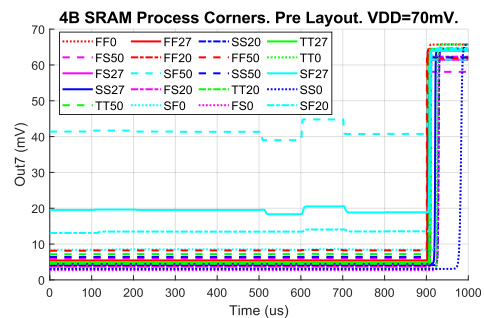
(e) Plot of Out4.



(f) Plot of Out5.

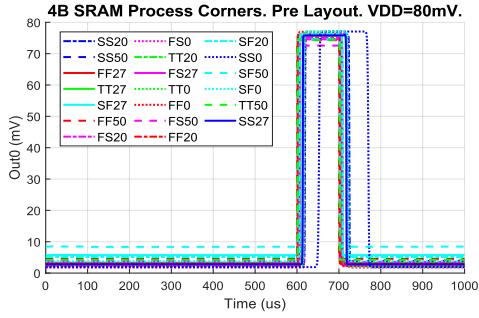


(g) Plot of Out6.

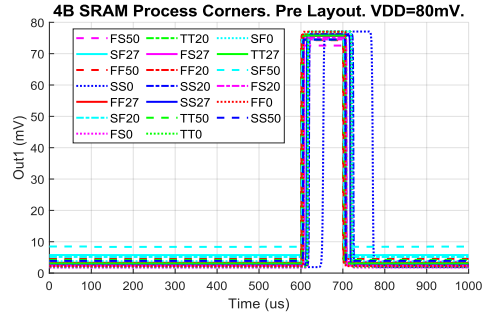


(h) Plot of Out7.

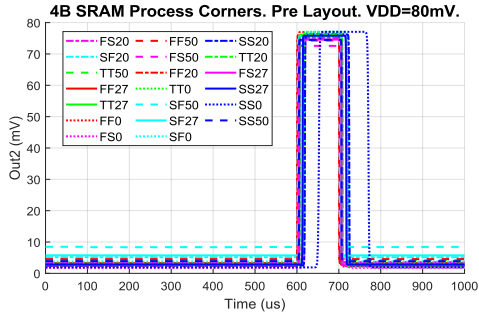
Figure 88: The 4B SRAM's output signals plotted for all pre layout process corners simulated with $V_{DD} = 70\text{ mV}$. *SF20, SF27, and SF50 fail to produce a legal logic low value for all outputs.*



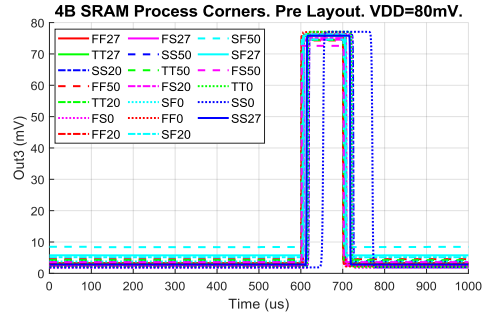
(a) Plot of Out0.



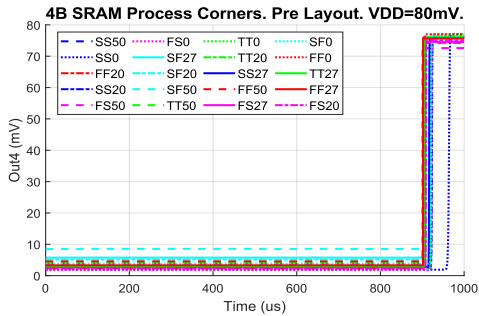
(b) Plot of Out1.



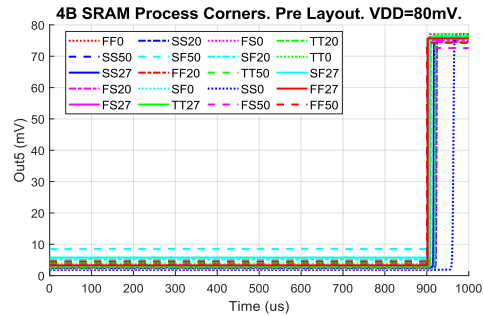
(c) Plot of Out2.



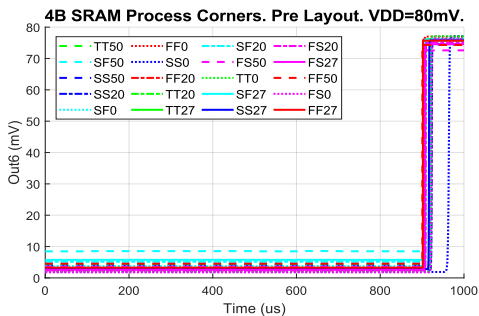
(d) Plot of Out3.



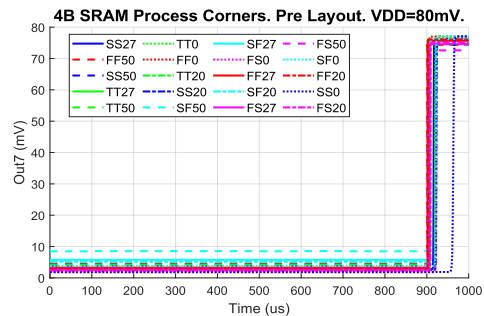
(e) Plot of Out4.



(f) Plot of Out5.

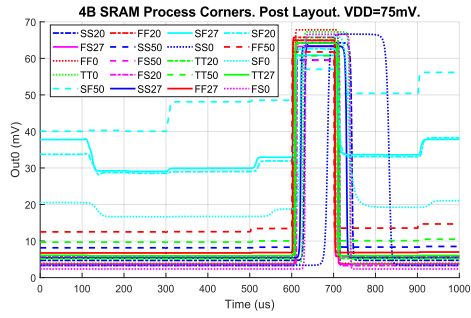


(g) Plot of Out6.

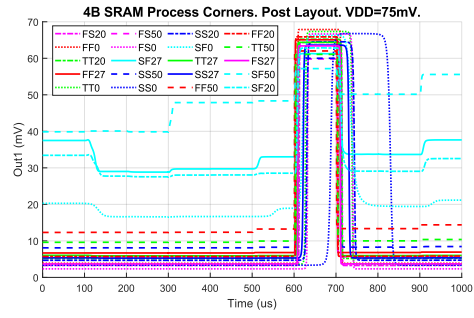


(h) Plot of Out7.

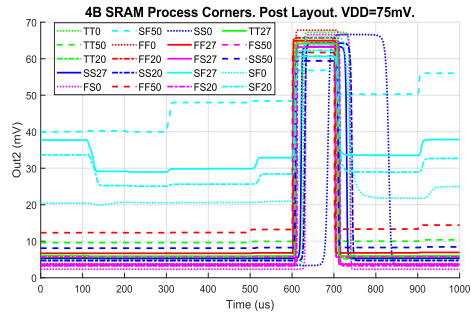
Figure 89: The 4B SRAM's output signals plotted for all pre layout process corners simulated with $V_{DD} = 80\text{ mV}$. All corners pass.



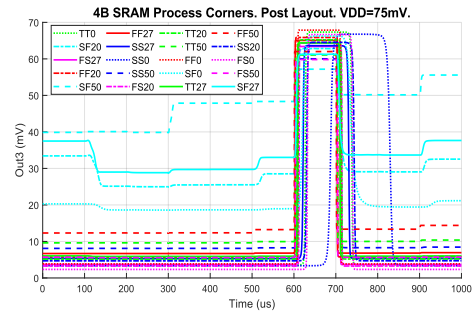
(a) Plot of Out0.



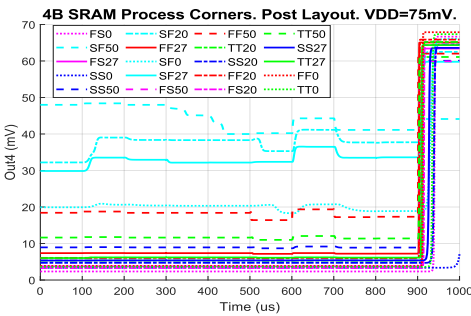
(b) Plot of Out1.



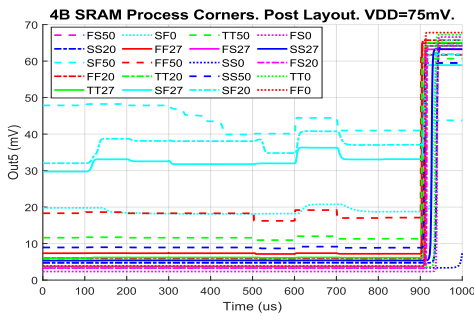
(c) Plot of Out2.



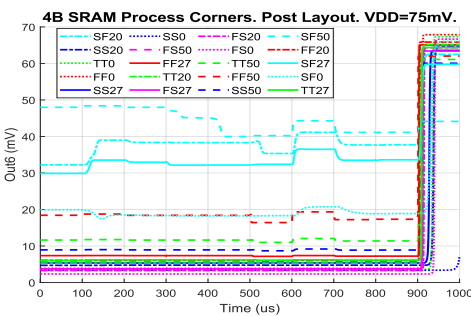
(d) Plot of Out3.



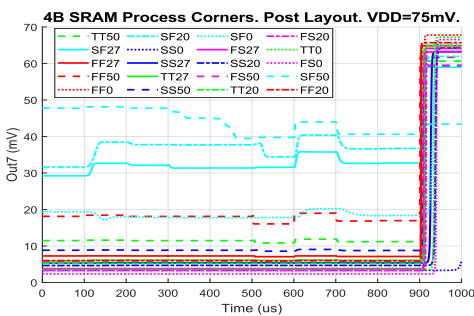
(e) Plot of Out4.



(f) Plot of Out5.

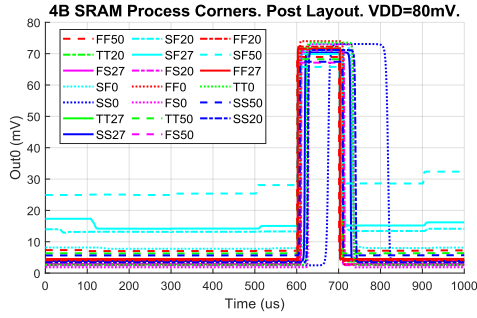


(g) Plot of Out6.

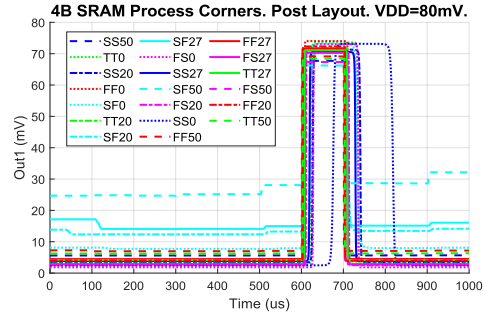


(h) Plot of Out7.

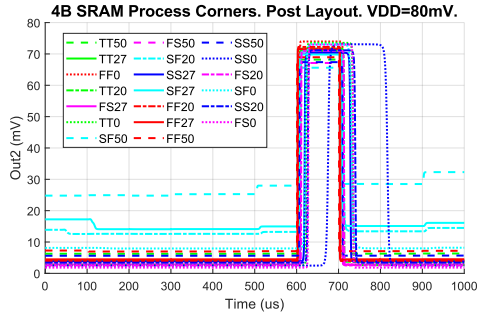
Figure 90: The 4B SRAM's output signals plotted for all post layout process corners simulated with $V_{DD} = 75 \text{ mV}$. The SF corner fails for all temperatures. FF50 fails for Out4-Out7.



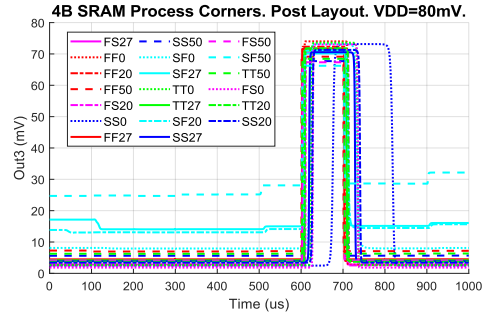
(a) Plot of Out0.



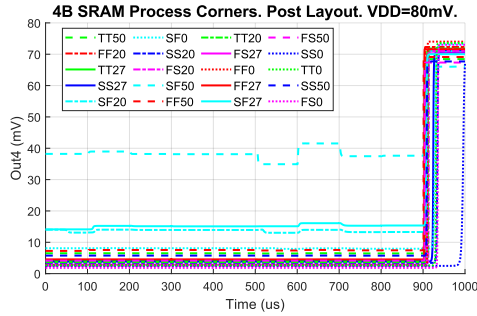
(b) Plot of Out1.



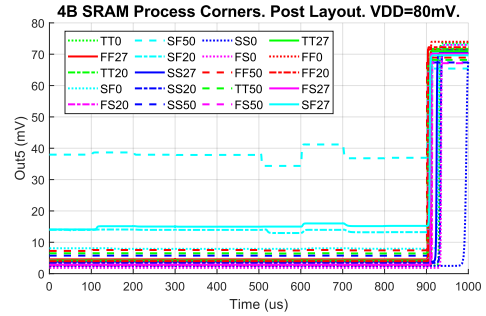
(c) Plot of Out2.



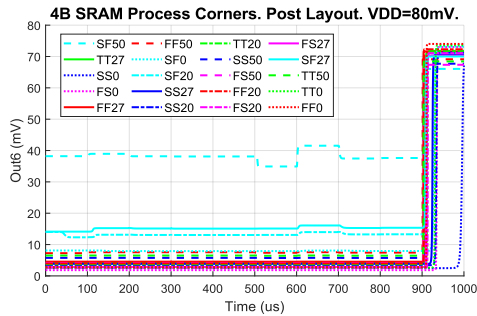
(d) Plot of Out3.



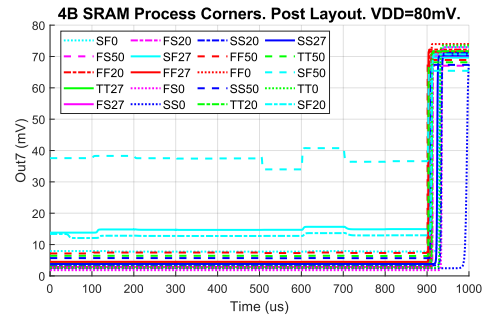
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

Figure 91: The 4B SRAM's output signals plotted for all post layout process corners simulated with $V_{DD} = 80\text{ mV}$. The SF50 corner fails.

F Results from Simulations on the 16B SRAM

F.1 Monte Carlo Mismatch Simulation Results

All 500 simulated points passed for the post layout mismatch simulation at $VDD = 85$ mV. *Out0* is plotted in Figure 92a for all 500 simulated points, *Out1* is plotted in Figure 92b, *Out2* is plotted in Figure 92c, *Out3* is plotted in Figure 92d, *Out4* is plotted in Figure 92e, *Out5* is plotted in Figure 92f, *Out6* is plotted in Figure 92g, and *Out7* is plotted in Figure 92h.

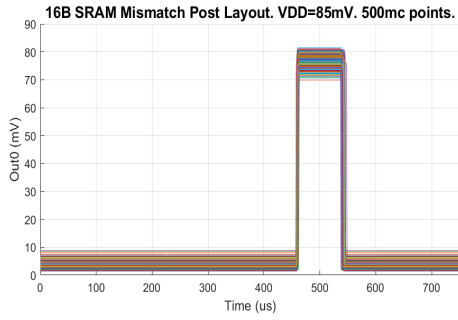
Post layout mismatch simulations were run at $VDD = 80$ mV, and the 16B SRAM's output signals are plotted for all 500 simulation points in Figure 93.

F.2 Process Corner Simulation Results

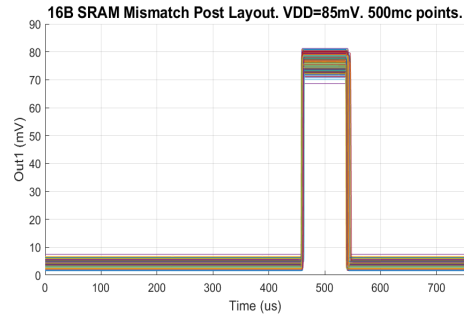
The 16B SRAM's output signals are plotted in Figure 94 for all the pre layout process corners simulated with $VDD = 75$ mV. SF50 fails to produce a legal logic low value for all outputs.

All corners pass for the pre layout simulation of process variation with $VDD = 85$ mV. Figure 95 contains plots of the 16B SRAM's output signals for all simulated process corners.

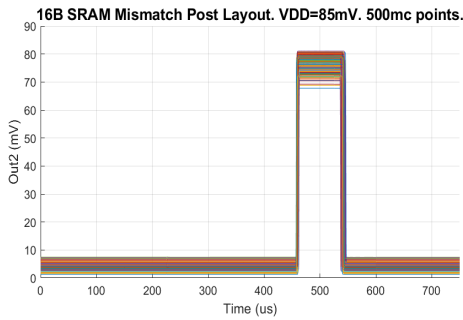
The 16B SRAM's output signals are plotted in Figure 96 for all the post layout process corners simulated with $VDD = 80$ mV. SF27 and SF50 fail to produce legal logic low values for all output signals.



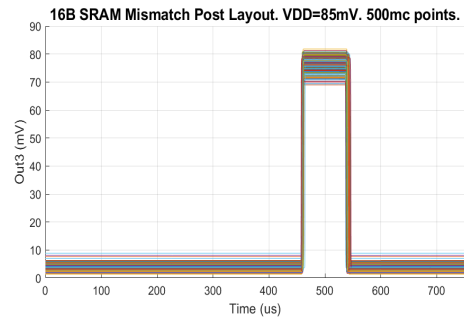
(a) Plot of Out0.



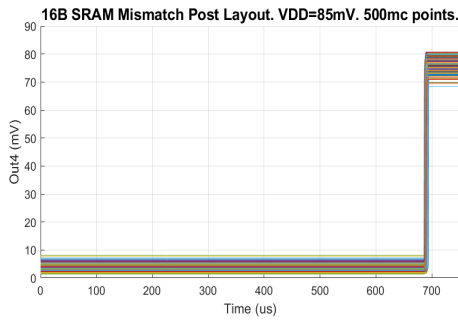
(b) Plot of Out1.



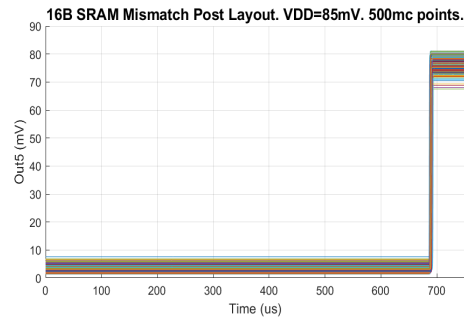
(c) Plot of Out2.



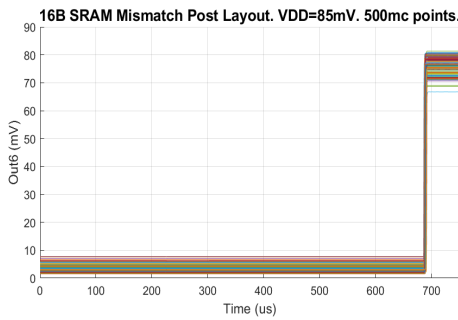
(d) Plot of Out3.



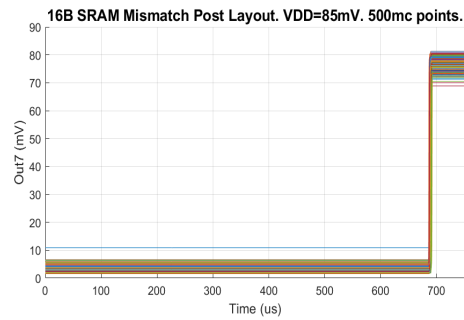
(e) Plot of Out4.



(f) Plot of Out5.

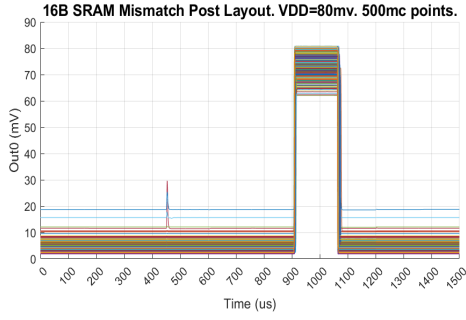


(g) Plot of Out6.

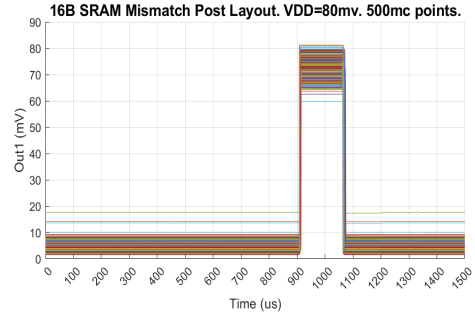


(h) Plot of Out7.

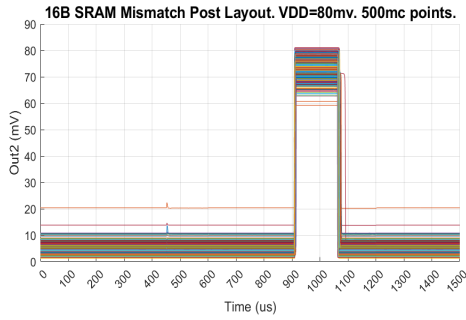
Figure 92: The 16B SRAM's output signals plotted for the 500 post layout Monte Carlo points simulated with $V_{DD} = 85 \text{ mV}$.



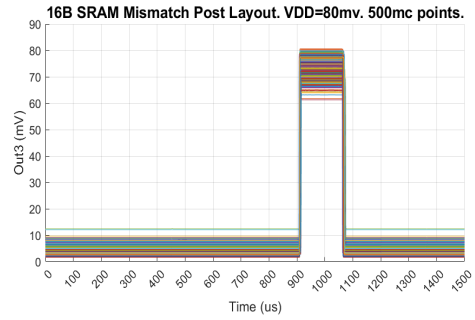
(a) Plot of Out0.



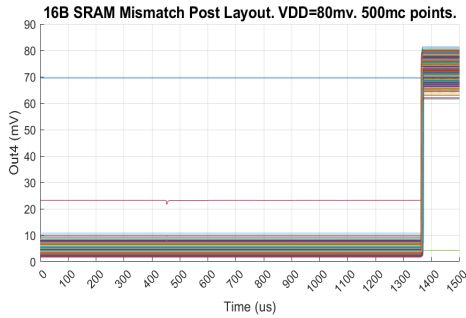
(b) Plot of Out1.



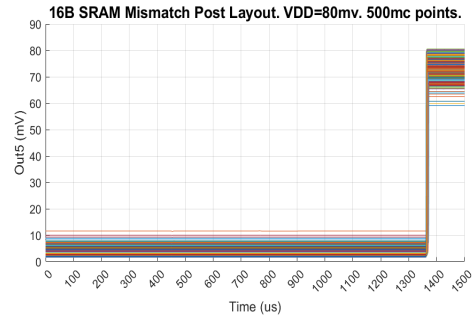
(c) Plot of Out2.



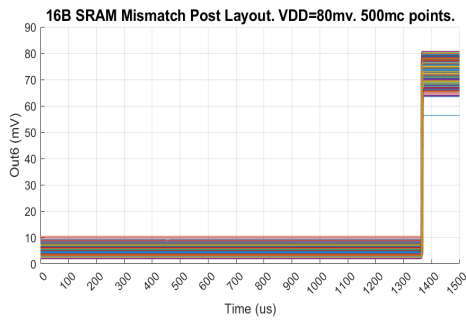
(d) Plot of Out3.



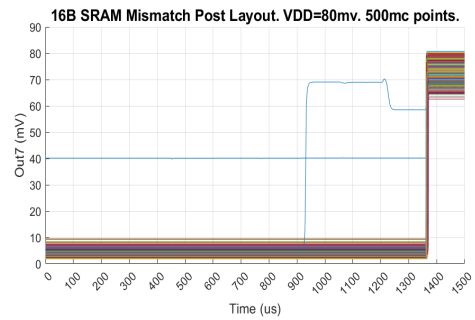
(e) Plot of Out4.



(f) Plot of Out5.

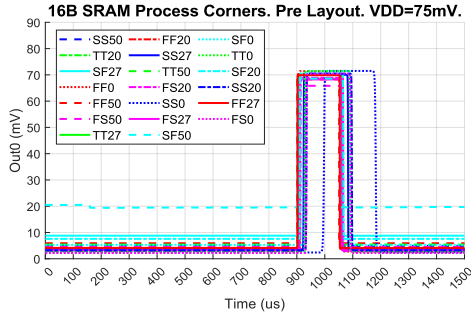


(g) Plot of Out6.

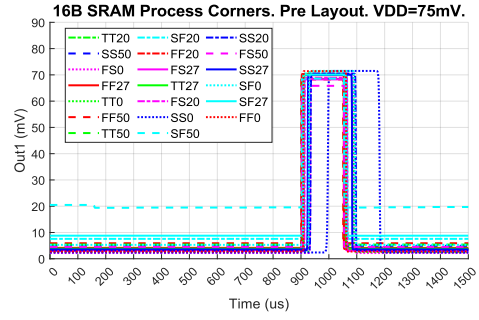


(h) Plot of Out7.

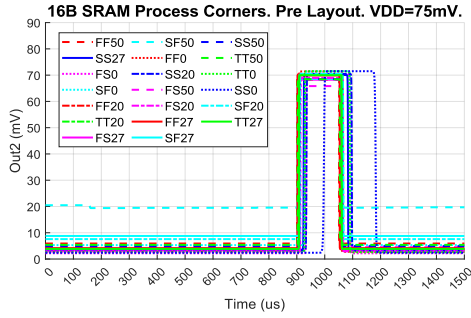
Figure 93: The 16B SRAM's output signals plotted for the 500 post layout Monte Carlo points simulated with $V_{DD} = 80\text{ mV}$.



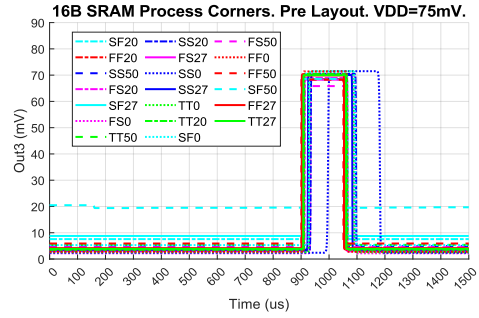
(a) Plot of Out0.



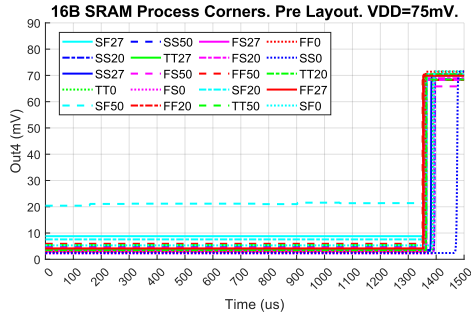
(b) Plot of Out1.



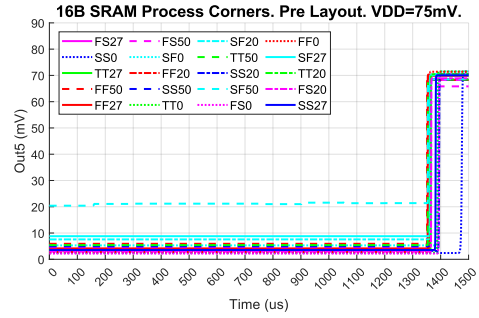
(c) Plot of Out2.



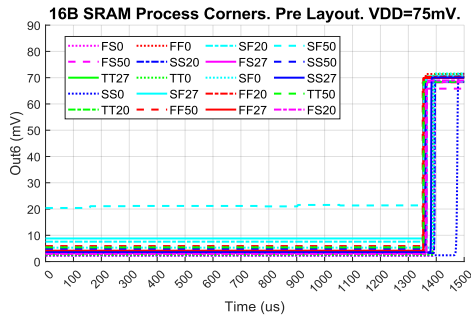
(d) Plot of Out3.



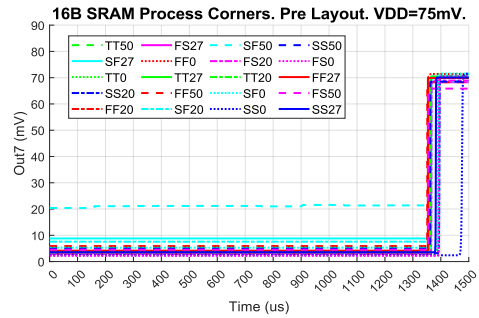
(e) Plot of Out4.



(f) Plot of Out5.

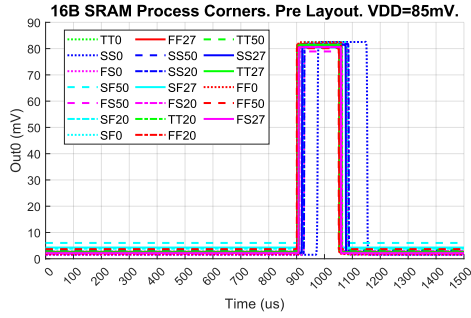


(g) Plot of Out6.

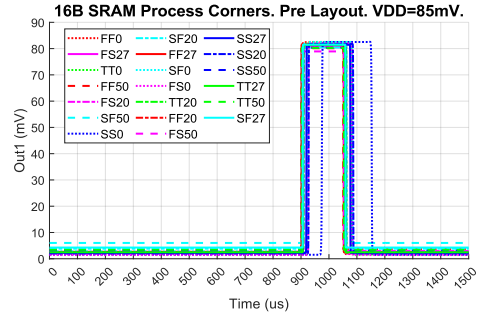


(h) Plot of Out7.

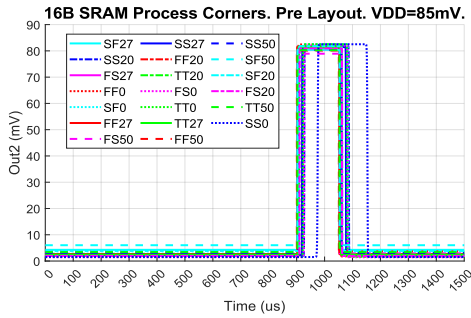
Figure 94: The 16B SRAM's output signals plotted for all pre layout process corners simulated with $VDD = 75\text{ mV}$. The SF50 corner fails.



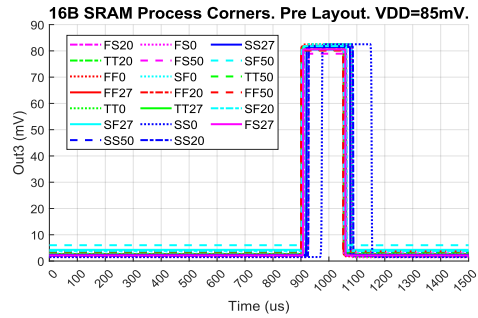
(a) Plot of Out0.



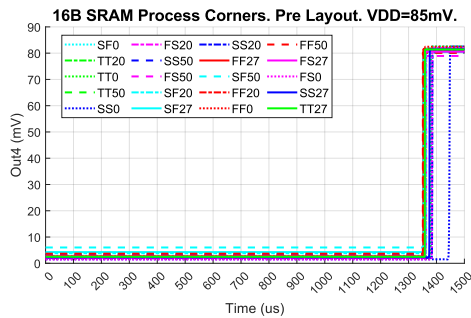
(b) Plot of Out1.



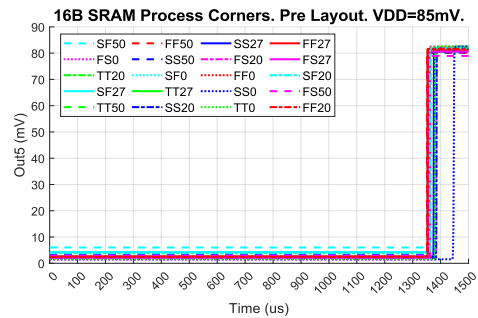
(c) Plot of Out2.



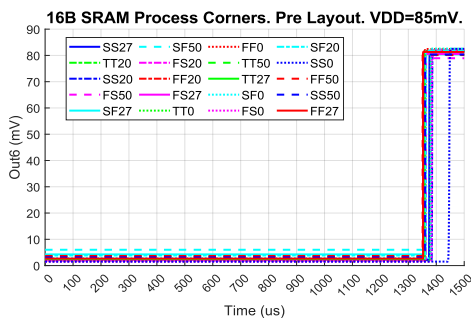
(d) Plot of Out3.



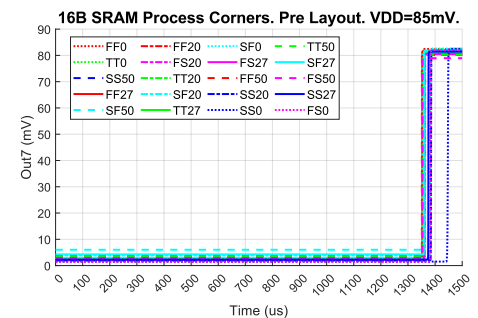
(e) Plot of Out4.



(f) Plot of Out5.

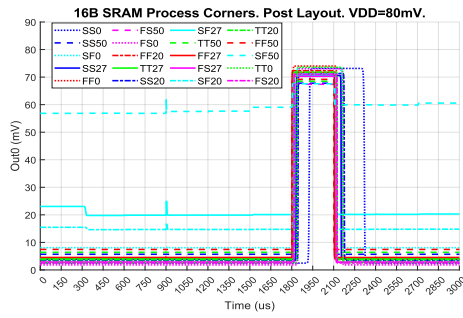


(g) Plot of Out6.

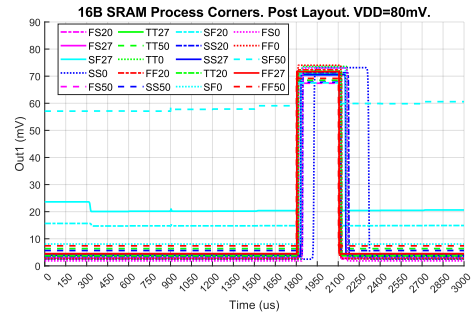


(h) Plot of Out7.

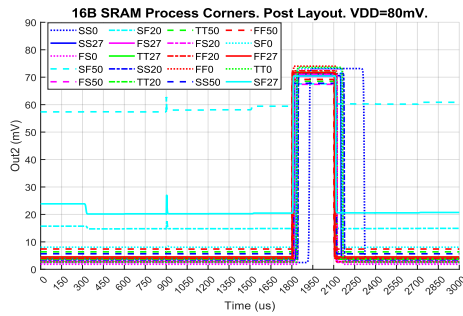
Figure 95: The 16B SRAM's output signals plotted for all pre layout process corners simulated with $VDD = 85\text{ mV}$. All corners pass.



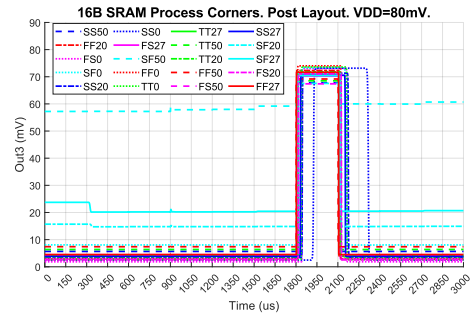
(a) Plot of Out0.



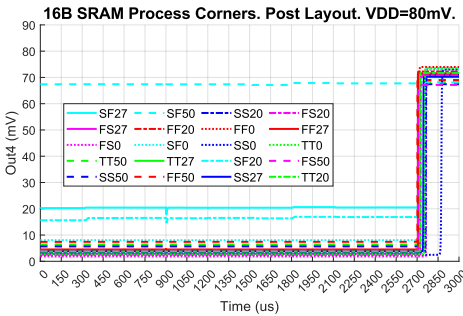
(b) Plot of Out1.



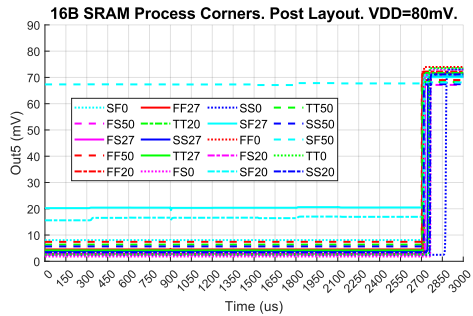
(c) Plot of Out2.



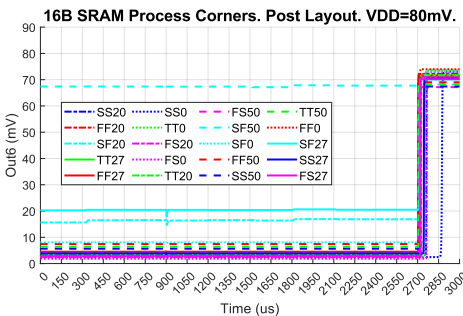
(d) Plot of Out3.



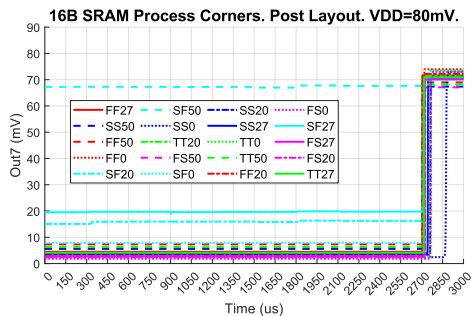
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

Figure 96: The 16B SRAM's output signals plotted for all post layout process corners simulated with $V_{DD} = 80 \text{ mV}$. SF27 and SF50 fail.

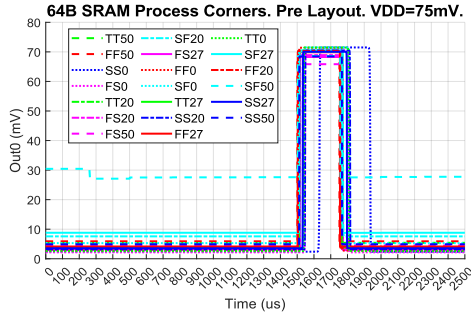
G Results from Simulations on the 64B SRAM

G.1 Process Corner Simulation Results

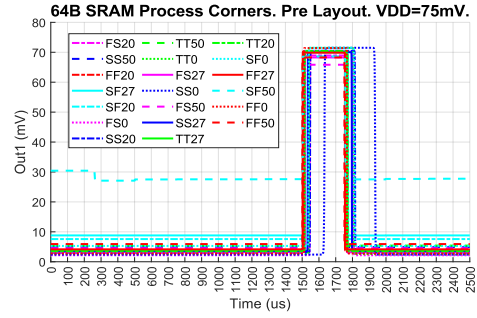
The 64B SRAM's output signals are plotted in Figure 97 for all the pre layout process corners simulated with $VDD = 75$ mV. SF50 fails to produce a legal logic low value for all outputs.

All corners pass for the pre layout simulation of process variation with $VDD = 85$ mV. Figure 98 contains plots of the 64B SRAM's output signals for all simulated process corners.

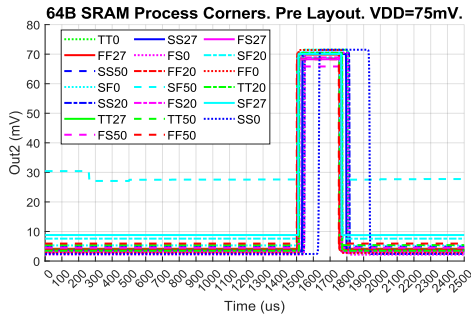
The 64B SRAM's output signals are plotted in Figure 99 for all the post layout process corners simulated with $VDD = 80$ mV. SF27 and SF50 fail to produce legal logic low values for all output signals.



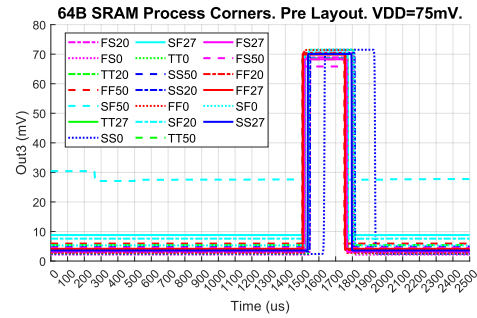
(a) Plot of Out0.



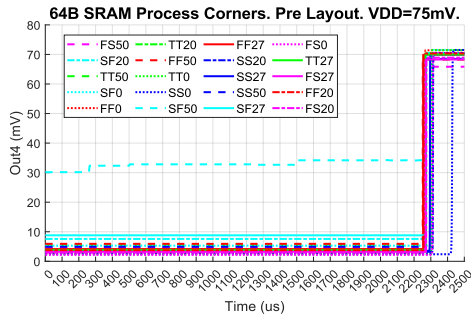
(b) Plot of Out1.



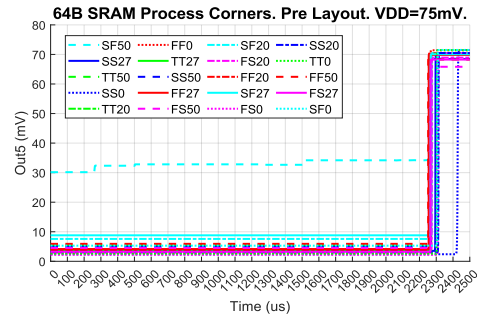
(c) Plot of Out2.



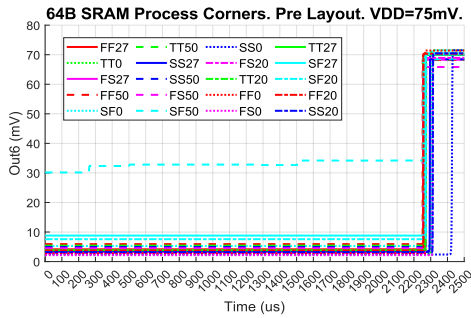
(d) Plot of Out3.



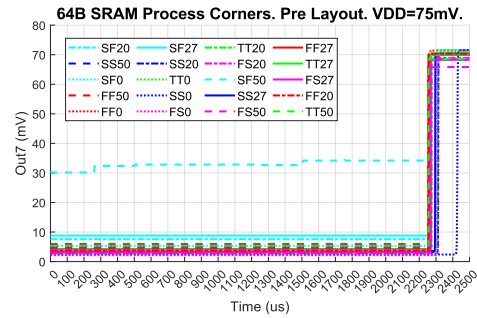
(e) Plot of Out4.



(f) Plot of Out5.

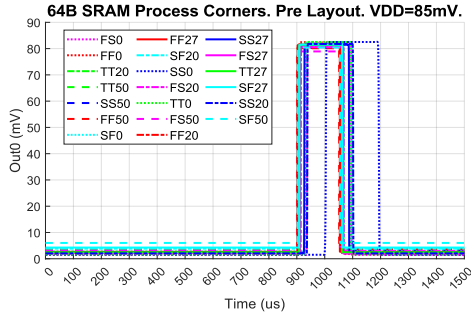


(g) Plot of Out6.

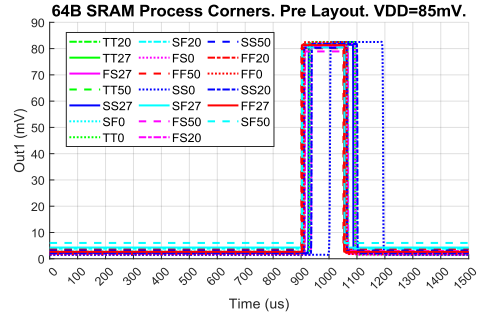


(h) Plot of Out7.

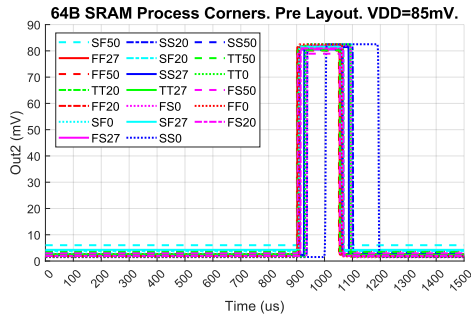
Figure 97: The 64B SRAM's output signals plotted for all pre layout process corners simulated with $VDD = 75\text{ mV}$. The SF50 corner fails.



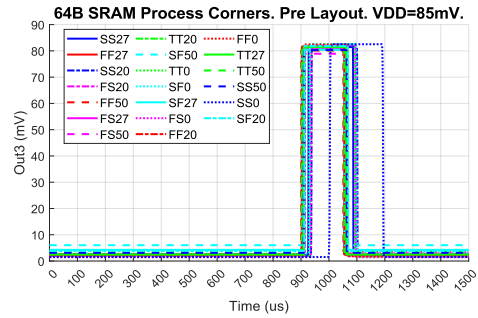
(a) Plot of Out0.



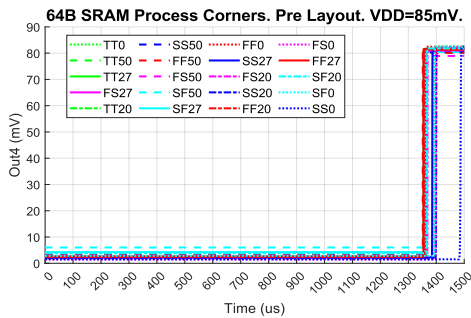
(b) Plot of Out1.



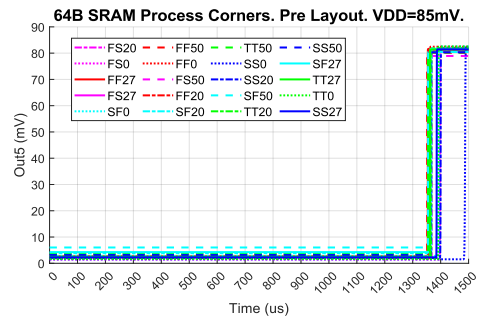
(c) Plot of Out2.



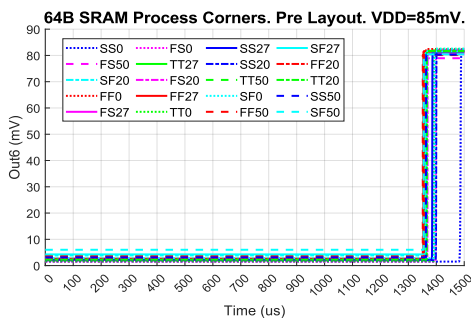
(d) Plot of Out3.



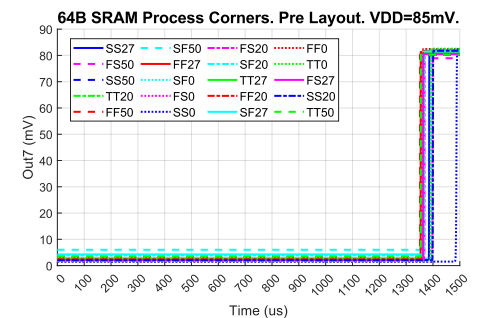
(e) Plot of Out4.



(f) Plot of Out5.

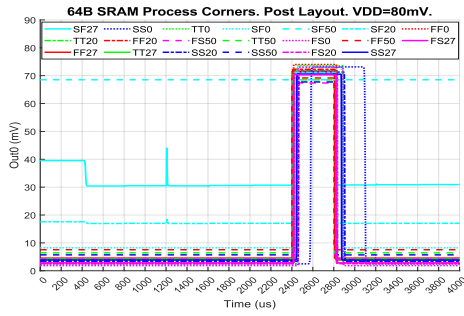


(g) Plot of Out6.

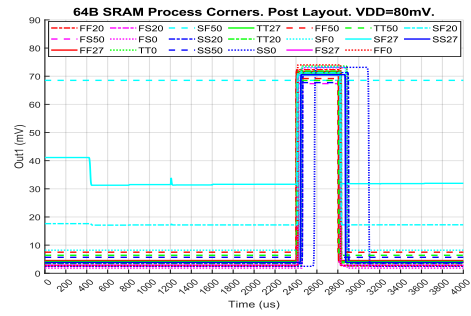


(h) Plot of Out7.

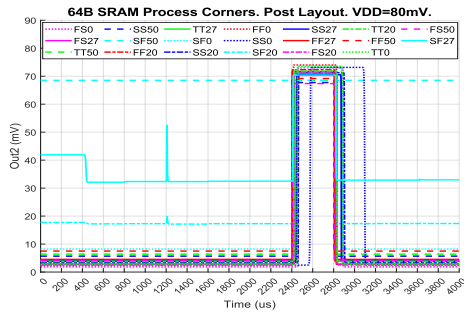
Figure 98: The 64B SRAM's output signals plotted for all pre layout process corners simulated with $VDD = 85\text{ mV}$. All corners pass.



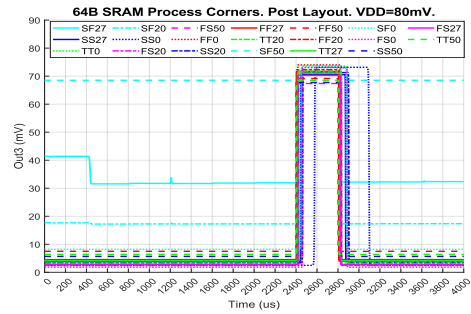
(a) Plot of Out0.



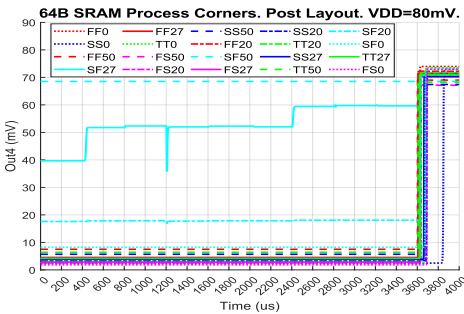
(b) Plot of Out1.



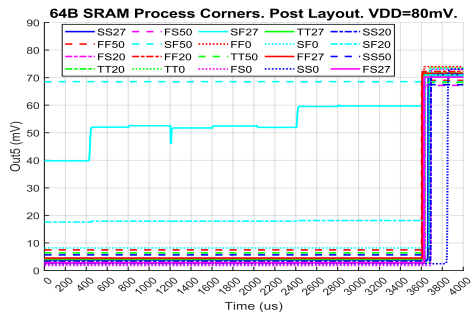
(c) Plot of Out2.



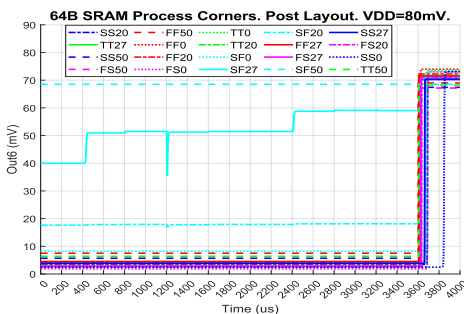
(d) Plot of Out3.



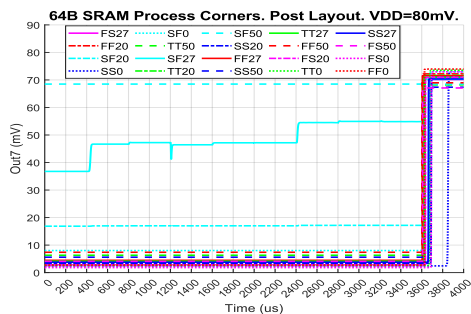
(e) Plot of Out4.



(f) Plot of Out5.



(g) Plot of Out6.



(h) Plot of Out7.

Figure 99: The 64B SRAM's output signals plotted for all post layout process corners simulated with $VDD = 80\text{ mV}$. SF27 and SF50 fail.

H Layout of the 2to4 Decoder

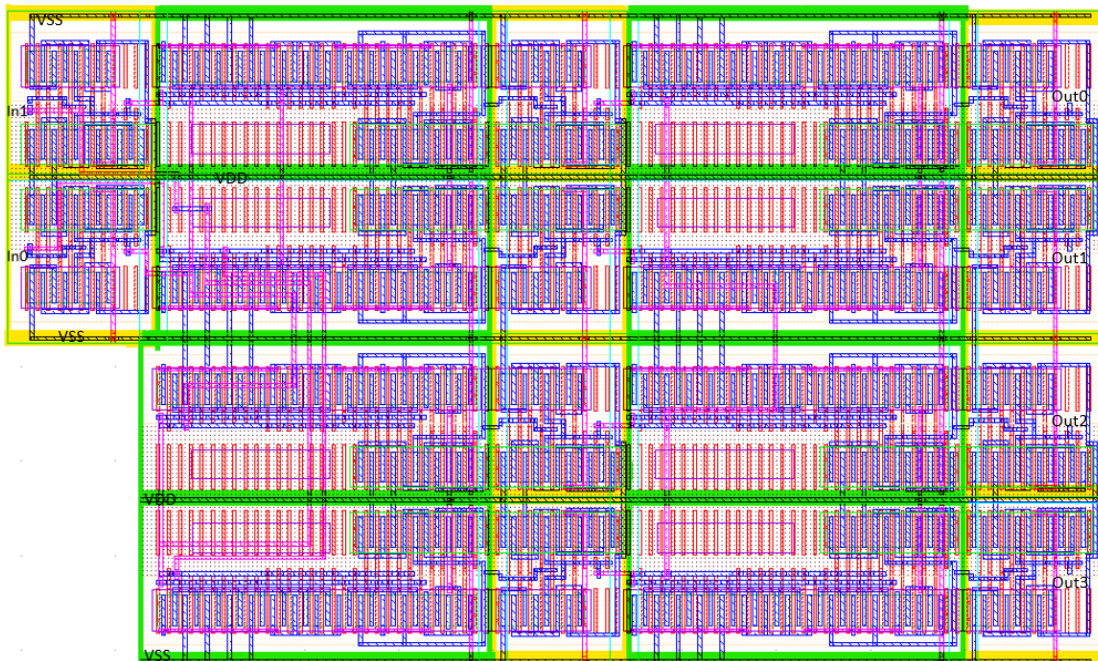


Figure 100: Layout of the Decoder. The NOT gates are outlined in yellow, and the NAND gates are surrounded by a green outline. Area = $75.982\mu\text{m}^2$ (with $h = 6.92\mu\text{m}$ and $w = 11.666\mu\text{m}$). Since the Decoder is used as a building block in the SRAM, dummy polys at the end of the rows and a substrate contact are added later. These must also be added to pass DRC and LVS, and run post layout simulations on the module.

I Layout of the Output Selection Module

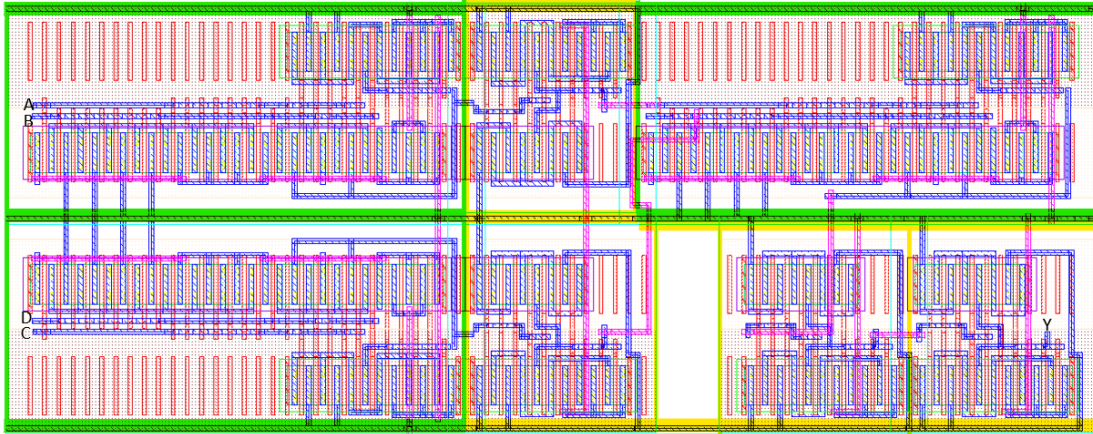


Figure 101: Layout of the Output Selection Module. The NOT gates are outlined in yellow, and the NAND gates are surrounded by a green outline. Area = $31.19688\mu\text{m}^2$ (with $h = 3.51\ \mu\text{m}$ and $w = 8.96\ \mu\text{m}$). Since it is used as a building block in the SRAM, dummy polys at the end of the rows and a substrate contact are added later. These must also be added to pass DRC and LVS, and run post layout simulations on the module. The rail at the top and at the bottom are VDD, and the middle rail is VSS. The NAND and NOTs in the bottom row have been flipped vertically, so that the pull down network is on top.

J Layout of the 4B SRAM

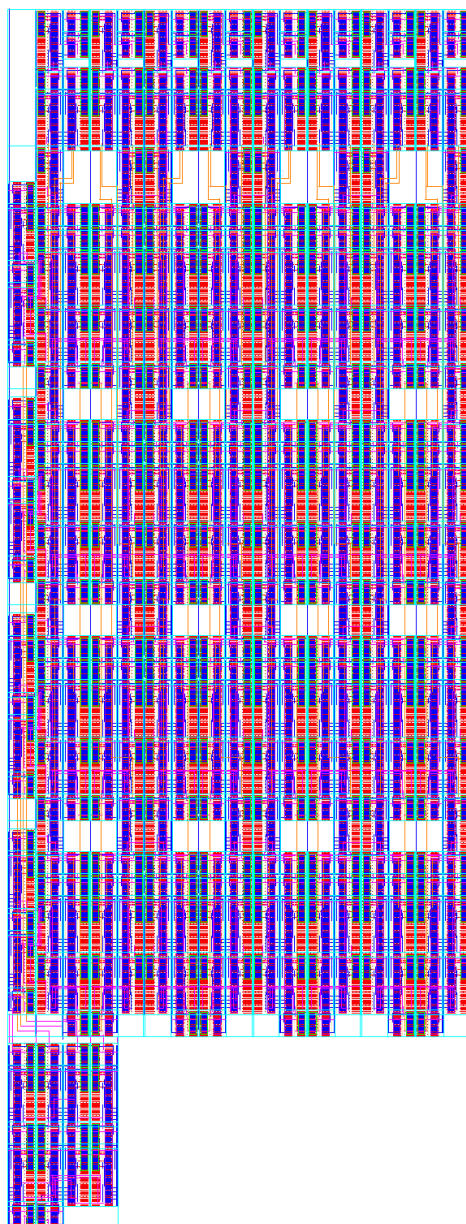


Figure 102: Layout of the 4B SRAM without annotations. The figure has been rotated 90 degrees counter clockwise to better fit the page. Area = $1961.33244\mu\text{m}^2$ (with $h = 29.085\ \mu\text{m}$ and $w = 76.616\ \mu\text{m}$). Metal fill is needed for the higher metal layers to pass the DRC, but are not included in the picture to increase readability.

K Layout of the 16B SRAM

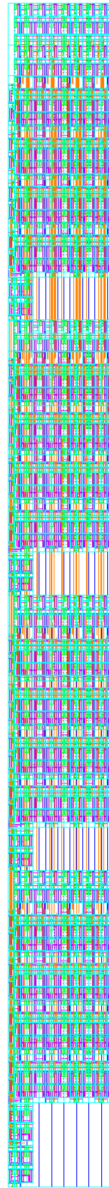


Figure 103: Layout of the 16B SRAM without annotations. The figure has been rotated 90 degrees counter clockwise to better fit the page. Four 4B SRAMs are stacked horizontally. A column of output logic (Output Selection Modules with NOT gates at each input) is placed to the right. A decoder is placed in the top left corner, and creates EN-signals for the 4B SRAM's Decoders. Area = $9579.08658\mu\text{m}^2$ (with $h = 29.085\ \mu\text{m}$ and $w = 329.348\ \mu\text{m}$).

L Layout of the 64B SRAM

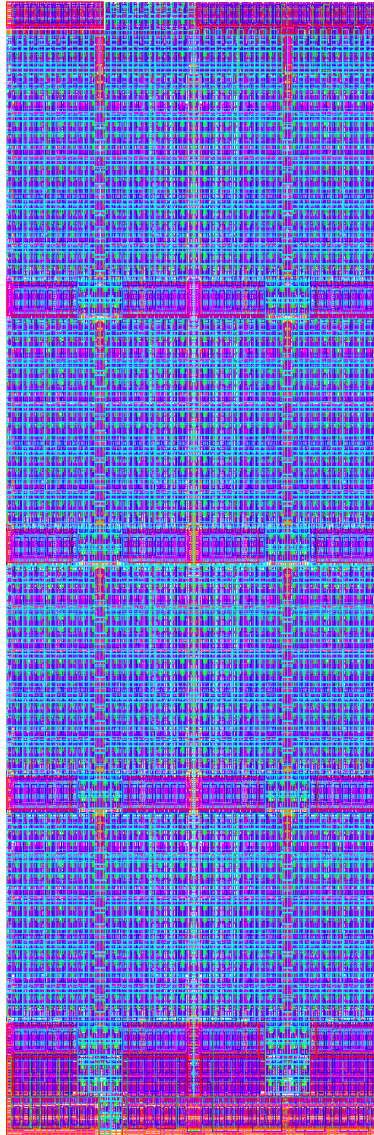


Figure 104: Layout of the 64B SRAM. The figure has been rotated 90 degrees counter clockwise to better fit the page. Four 16B SRAMs are stacked vertically (every other 16B SRAM is flipped vertically, so that the rails at each end could overlap with the neighbour's end rails). A column of output logic (Output Selection Modules with NOT gates at each input) is placed to the right, at the output of the second 16B SRAM from the top. The decoder that creates EN-signals for the 16B SRAM's Decoders is placed to the left, at the input of the second 16B SRAM from the top. Area = $40561.401 \mu\text{m}^2$ (with $h = 116.04 \mu\text{m}$ and $w = 351.832 \mu\text{m}$).



 **NTNU**

Norwegian University of
Science and Technology