Circuit Design of a TCAM using an unmodified eDRAM bit-cell
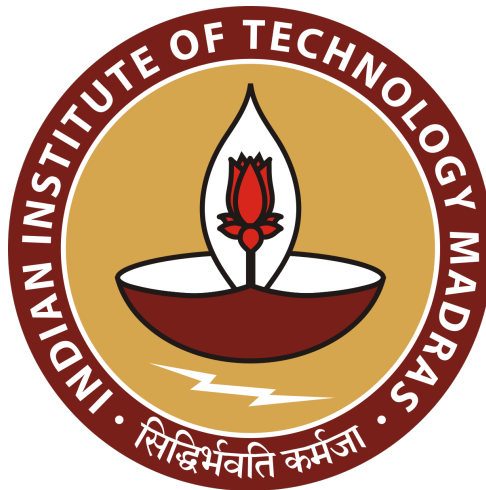
A project thesis

submitted by

Dasari Shirisha

in partial fulfillment of the requirements
for the award of the degree of

Master of Technology

Dept. of Electrical Engineering
IIT Madras
Chennai 600 036

Thesis Certificate

This is to certify that the thesis titled Circuit Design of a TCAM using an unmodified eDRAM bit-cell, submitted by Dasari Shirisha, to the Indian Institute of Technology, Madras, for the award of the degree of Master of Technology, is a bona fide record of the research work done by Dasari Shirisha under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Janakiraman Viraraghavan
Project Guide,
Assistant Professor,
Dept. of Electrical Engineering,
IIT Madras, 600 036

Place : Chennai
Date : June 2021

Acknowledgements

I am greatly indebted to Prof. Janakiraman Viraraghavan for guiding me through the entire course of my M.Tech. project. He always took the time and effort to discuss the problem and to suggest different methods to experiment. His valuable remarks always gave new directions to my project.

A special thanks to Shruthi Parvathi who was my fellow associate in doing this project. This project would not have been possible without her contributions and insightful observations. A special word of thanks to my family, and my friends Alen C Philip, Balaji Vijayakumar for their constant support throughout the course of the project.

Abstract

This project proposes a new eDRAM architecture for Ternary Content Addressable Memory (TCAM). The main motive was to perform a fast search operation without increasing the area of the memory array.TCAM in SRAM is popular as the addition of extra logic transistors is not expensive. The available TCAM in DRAM architectures do not conserve much area due to the additional area expense of adding extra transistors to a thick oxide access transistor.

We have proposed a new DRAM architecture that is area effective and takes lesser number of cycles for a search operation and also does not require extra peripheral circuitry to do a match operation. It can also perform fundamental DRAM operations such as Read, Write and Refresh. 4 Transistor Micro Sense Amplifier was the peripheral circuitry that was used for both reading and search operations. The results obtained from simulations met the goal of the project.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to TCAMs in SRAM

1.1 Content Addressable Memory

A Content Addressable Memory (CAM) compares the input search data against the table of stored data and returns the address of matched data [1]. Highly parallel multi-data search makes a CAM an indispensable component for high-associative caches, computer networking devices to provide faster lookup in routing tables and register renaming. CAMs are much faster than RAMs in data search applications. In a RAM, when a user supplies the memory address, it returns the word stored in the address. Whereas CAM access the words based on the content itself. A lookup of a word in a CAM can be performed in a single clock cycle, but a RAM module requires multiple clock cycles to make a single memory fetch. The speed of CAM comes at the cost of increased area and power consumption due to highly parallel comparisons.
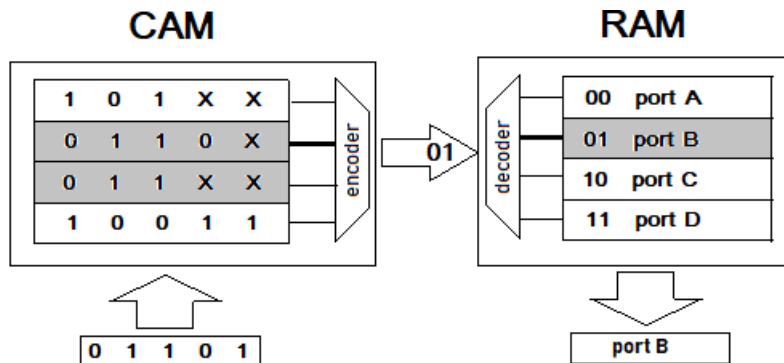
Figure 1.1: CAM-based implementation of the routing table

The parallel comparisons of CAM consume huge power and amount to increase in Silicon area.

1.2 CAM Architecture

There are two types of CAM- Binary CAM (BCAM) and Ternary CAM (TCAM). A Binary CAM performs binary look-up and returns either 0 or 1. A single bit is enough to store the data. A TCAM stores 0,1 as well as a don't care state. The "Don't Care" is stored at an additional cost over binary CAM since the internal memory cell must now encode three possible states. This is usually implemented by adding a mask bit ("care" or "don't care" bit) to every memory cell, and hence 2 bits are required to store a single data. The input to the CAM system is a search word, which is broadcast on the search lines of the stored data.
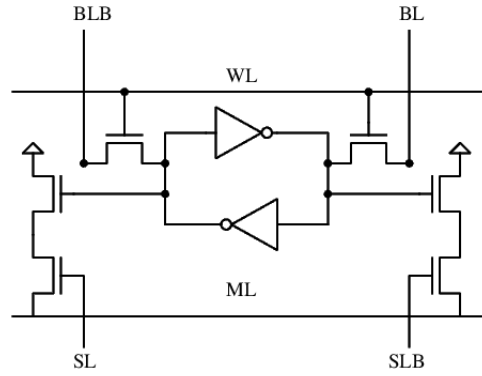
Figure 1.2: Binary CAM 10T bit cell

Figure 1.3: Tenary CAM 16T bit cell

1.2.1 NOR based SRAM TCAM

The NOR cell implements [1] the comparison between the bit, D and $\overline{D}$ and data on search lines, SL and $\overline{SL}$, by implementing dynamic XOR logic gate with inputs SL and D. Each pair of transistors, (M_1, M_3) and (M_2, M_4), forms a pull-down path from the match line, ML. A mismatch of SL and D

activates at least one of the two pull-down paths, connecting ML to ground. A match of SL and D disables both pull-down paths, disconnecting ML from the ground.

Don't care state is stored in cell by setting both D and $\overline{D}$ equal to logic "1", which disables both pulldown paths and forces the cell to match regardless of the inputs in the search line. The state where D and $\overline{D}$ are both '0' is not allowed.

A mask condition is implemented by sending '0' to both the Search Lines SL and $\overline{SL}$, which disables both pulldown paths and forces the cell to match regardless of the stored bits.
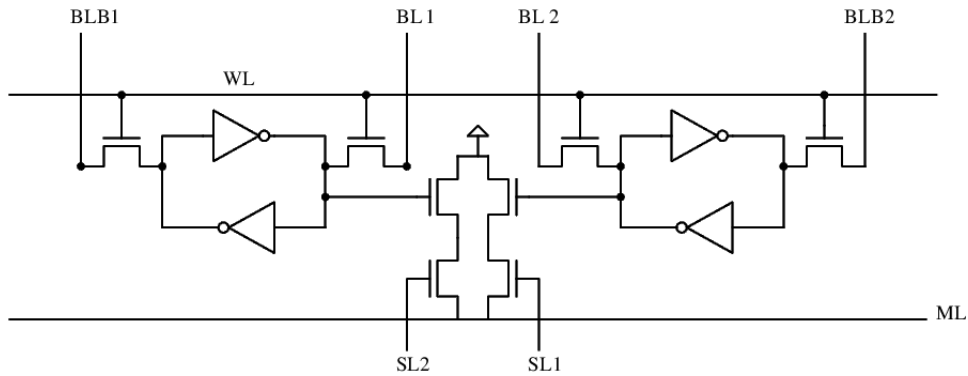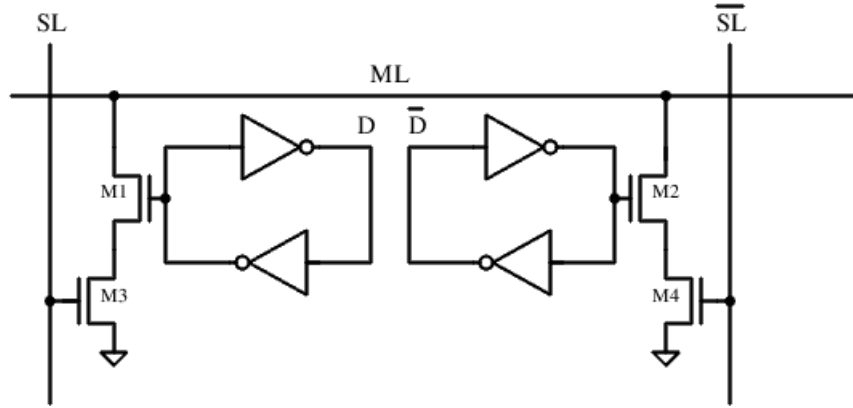

Figure 1.4: NOR based SRAM TCAM

1.2.2   NAND based SRAM TCAM

The NAND cell implements the comparison between the stored bit, D, and $\overline{D}$ and the corresponding search data on the search lines, SL, and $\overline{SL}$, using the three comparison transistors M_1, M_D, and $M_{\overline{D}}$. In case of a match when $SL = 1$ and $D = 1$ or when $SL = 0$ and $D = 0$, pass transistor M_D is ON and passes the logic '1' on the SL to node B. Node B is the bit-match node which is at logic 1 if there is a match in the cell[1]. In this case, the transistor $M_{\overline{D}}$ passes a logic HIGH to raise node B. In cases where $SL \neq D$, that result in a mismatch,, the node B is at logic '0' and the transistor M_1 is OFF.

To store a Don't care state, the mask bit is set to "1". This forces the transistor M_{mask} to turn ON, regardless of the value of D, ensuring that the cell always matches.

To implement a mask condition, both the search lines SL, and $\overline{SL}$ are set to

logic '1' which enables at least one of the two transistors M_D or $M_{\overline{D}}$ to pass the logic '1' to node B.
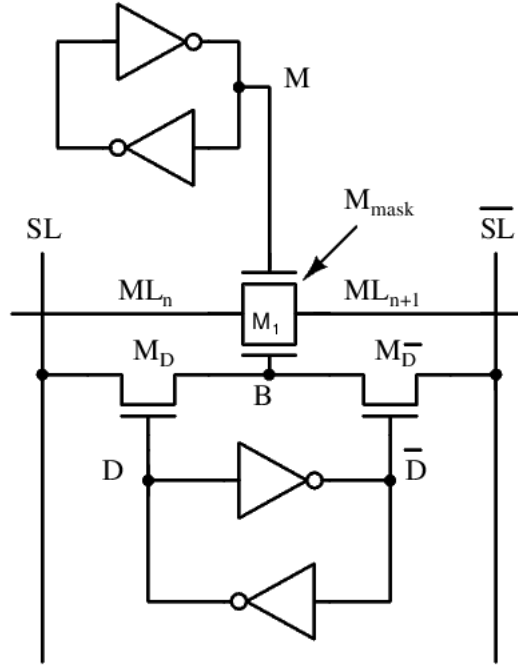
Figure 1.5: NAND based SRAM TCAM

An important property of the NOR based SRAM TCAM cell is that it provides a full rail voltage V_{DD} at the gates of all comparison transistors. On the other hand, a deficiency of the NAND based SRAM TCAM cell is that it provides only a reduced logic "1" voltage at node B, which can reach only $V_{DD} - V_{tn}$ when the search lines are driven to V_{DD} (where V_{DD} is the supply voltage and V_{tn} is the nMOS threshold voltage).

Chapter 2

Introduction to TCAMs in DRAM

2.1 Dynamic TCAM implementations

Dynamic CAM has an inherent advantage over static CAM since it can store three states of CAM in a small area.

2.1.1 4T Dynamic CAM cell

The four transistor (4T) Dynamic CAM cell [2] shown in Figure 2.1 consists of two transistors T_{C0} and T_{C1} to perform XOR operation of the data presented at Bit and NBIT and the data stored in the cell. The gates of these transistors serve as dynamic storage elements, and are labelled as S_{b1} and S_{b0}.
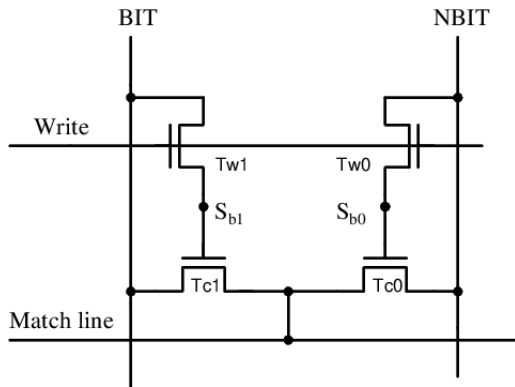
Figure 2.1: 4T DTCAM

S_{b1}	S_{b0}	State
0	0	Don't care
0	1	0
1	0	1
1	1	Not allowed

Figure 2.2: DRAM cell stored states

Since the output of XOR operation is connected to the match line, in case of a match, the Matchline is not discharged. The table 2.1 gives information on results on various cases of search operation.

Stored data	Sb1	Sb0	bit	NBit	Match condition
0	0	1	0	1	Match
0	0	1	1	0	Mismatch
1	1	0	1	0	Match
1	1	0	0	1	Mismatch
X	0	0	0	0	Match

Table 2.1: Table for match operation in 4T DCAM

2.1.2   6T Dynamic CAM cell

Dynamic TCAM [3] has 6T structure storing a trit i.e., logic 0 for 01(data=0, mask=0), logic 1 for 10(data=1, mask=0) and don't care for 00(data=x,mask=1)
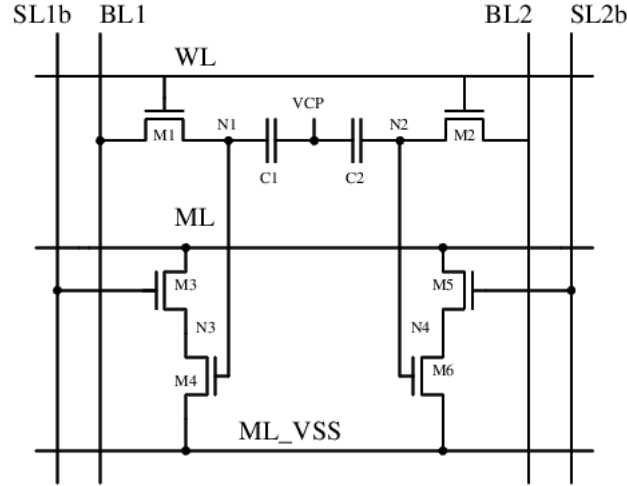
Figure 2.3: 6T Dynamic TCAM

Search operation in the above architecture is performed by pre-charging the Matchline to a value below Vdd and the search lines are held at ground. Then the precharge of Match line is released and the inverted input bits are placed on search line. If the search data matches the cell data or masked, the match line would not discharge. Match line is discharged, in a case of mismatch.

The transistors M3, M4 and M5, M6 implement XOR operation of search input and the contents of the cell. For example, if '1' is stored in cell (10), placing '0' on search line (SL1b=0, SL2b=1), the XOR stack is turned OFF and the Match line would not discharge. If '1' is placed on search lines (SL1b=1, SL2b=1), there is a discharge path from ML to ML_VSS from

M3,M4 thereby implying a mismatch.

2.1.3   Problems with implementing TCAM in DRAM over SRAM

In SRAM based implementation, the provision for comparison is done by adding four extra transistors per SRAM cell. The gates of these transistors are connected to the contents of the cell and the input search data, like in Fig 1.2. The comparison is performed without turning on the Wordline.

DRAMs store their contents on a capacitor rather than in a feedback loop. Deep Trench Capacitors (DTC) form the main storage element in eDRAM and have an advantage that they offer higher capacitance per unit area. But for cell architecture in Fig 2.2, if the gate of the mosfet M4 has to be connected to the capacitor, a contact has to be made for metal connection, affecting the density of eDRAM.

The access transistor in DRAM is a thick oxide device to reduce the subthreshold leakage and higher voltage swing. Adding two thin oxide transistors for comparison would affect the density as a lot of isolation is required between thin and thick oxide transistors for transistor stability. If chosen to add two extra thick oxide transistors for comparison would reduce the density by 3x, losing the density advantage over SRAMs.

Due to the foresaid problems, there is a need to develop a technique to implement TCAM in DRAM without affecting the density of DRAMs.

Chapter 3

Proposal of TCAM in eDRAM

3.1 Implementing TCAM in DRAM

The goal of the project is to implement TCAM in DRAM without affecting the density of eDRAM.

3.1.1 Pass transistor Logic in DRAM

In case of a SRAM based TCAM, the search operation was implemented in a stack based XNOR CMOS logic. Such a stack is not desired in DRAM. Hence the XNOR logic was implemented using pass transistor logic.

Figure 3.1: XNOR logic using pass transistor logic
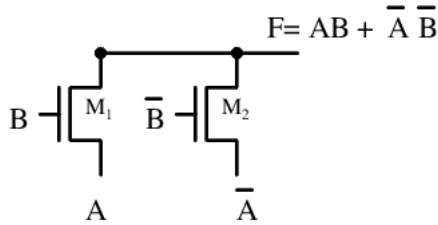
A	B	F
0	0	1
0	1	0
1	0	0
1	1	1

Figure 3.2: Truth table of XNOR gate

The eDRAM cell consists a thick oxide nmos access transistor and the bit is stored as voltage across deep trench capacitor. A single DRAM cell is used to represent Binary CAM, as it stores only two possible logic states, 0 or 1. For TCAM implementation, each data bit is represented by two bits, to store additional don't care state. Table 3.1 gives information on cell state representation for TCAM.

The conventional TCAM has search lines to place the search input data, which runs in parallel to the bitlines.The match lines runs across the Word-Lines to indicate status of search.

In the proposed eDRAM based TCAM, we intend to use Wordlines as the

Cell Value	Logic
01	0
10	1
00	Don't care
11	Forbidden

Table 3.1: Cell states for TCAM

search lines, and implement pass transistor based XNOR logic as shown in Figure 3.1. The bitlines are used as Matchlines. This results in area reduction and discards the need for extra external circuitry to detect a match. The word has to be stored in the same axis as that of the matchline. In the proposed case it has to be stored along the column.

3.2 Search operation in DRAM

The search operation is performed by placing the complementary bits or mask bits on the Word-Lines i.e., if we intend to search for logic 0(01) we place 1 (10) on the Wordlines and detecting change in Bit-Line voltage through the already available read circuitry.
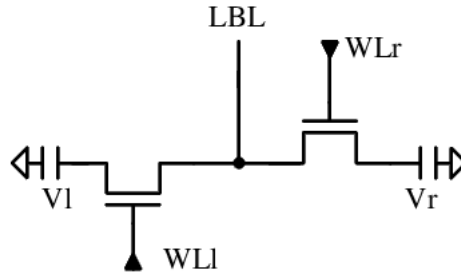
The access transistor in DRAM is an NMOS, which cannot write a 1 fully. The Wordline in eDRAM cell swings between 1.6V (V_{pp}) and -0.3 (V_{ss}) to write a strong 1 into the cell and to reduce the leakage during the off state respectively.

The Bit-Lines (BLs) are pre-discharged to the ground. In case of a match and mask, the Bit-Line is floating at the same pre-discharged state (gnd), and in case of a mismatch, the Bit-Line charge shares with the cell, and it is pulled to a weak '1'. During a search operation, it was assumed that only the Word-Lines of one cell (2 bits) is activated, and the remaining Word-Lines in the Bit-Line are turned off.

3.2.1 Detecting a match

The search bits are complementary to the contents of the eDRAM cell. This
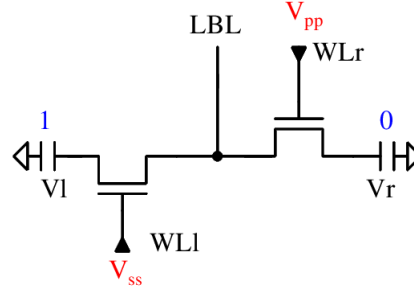
leaves the Bit-Line in the original pre-discharged state. The read circuitry detects the voltage on the Bit-Line and resolves it as a match.
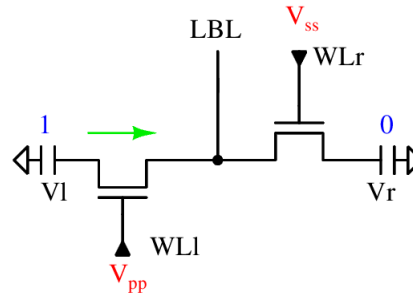
3.2.2 Detecting a mismatch

In case of a mismatch, the Word-Lines and the bits in the cell have the same value. The set where the Wordline is turned on and the cell storing a 1,

charge shares with the bitline. The bitline develops a voltage high enough that the read circuitry resolves it as a mismatch. The state 11 in a cell is forbidden.

3.2.3 Don't care and mask condition

The cell can be in the don't care state when both the bits are storing a 0. In this case turning on the Word-Line of either bit will not charge the Bit-Line.
  The proposed eDRAM TCAM allows masking any number of bits in the word. In case of a mask, both Word-Lines are turned off. As a result, the Bit-Line won't charge irrespective of contents of the cell.
In the above two conditions as the Bit-Line remains in its pre-discharged state the read circuitry resolves it as a match.
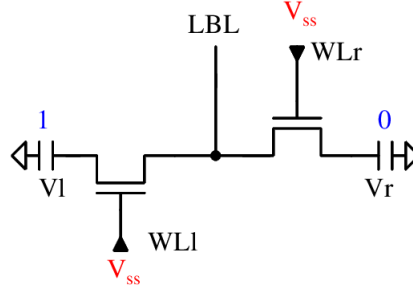
The table 3.2 gives the summary of the implementation of search in eDRAM by XNOR operation.

Vl	Vr	WLl	Wlr	Match Condition
0	1	1	0	Match
1	0	0	1	Match
0	1	0	1	Mismatch
1	0	1	0	Mismatch
0	0	X	X	Match
X	X	0	0	Mask

Table 3.2: XNOR operation in DRAM

In the proposed TCAM, the word is stored along the bitlines. To increase the efficiency of the search, multiple Wordlines should be turned on at once. Turning on multiple Wordlines of the same Local bitline result in charge
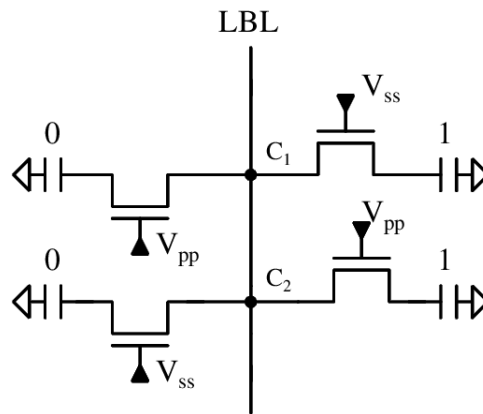

Figure 3.3: Conflict while turning on multiple WLs

conflict. The first data bit corresponding to node C1 indicate a match. The second data bit indicate mismatch and charge share with the bitline to increase bitline voltage. But as the wordline in the first data bit is turned on, there's a path from the local bitline to ground and discharges the local

bitline. Due to this, we would never resolve a mismatch, in the case where match and mismatch in data bits occur together. We proceed by doing a serial search i.e., searching for a single data bit in each local bitline at once. To improve the efficiency of the search, it is advised to have multiple local bitlines under each Sense Amplifier.

3.3 3T Micro Sense Amplifier

The proposed eDRAM architecture has multiple local bitlines, each with its own Sense Amplifier (SA). 3T Micro Sense Amplifiers are chosen to optimise the area and are connected to a Global Sense Amplifier (GSA) by Read BitLine(RBL) and Write Bitline (WBL) signals. The Global Sense Amplifiers(GSA) are then connected to a Data Sense Amplifier(DSA) by Local Data Line True(LDLC) and Local Dataline Complement (LDLC) signals.
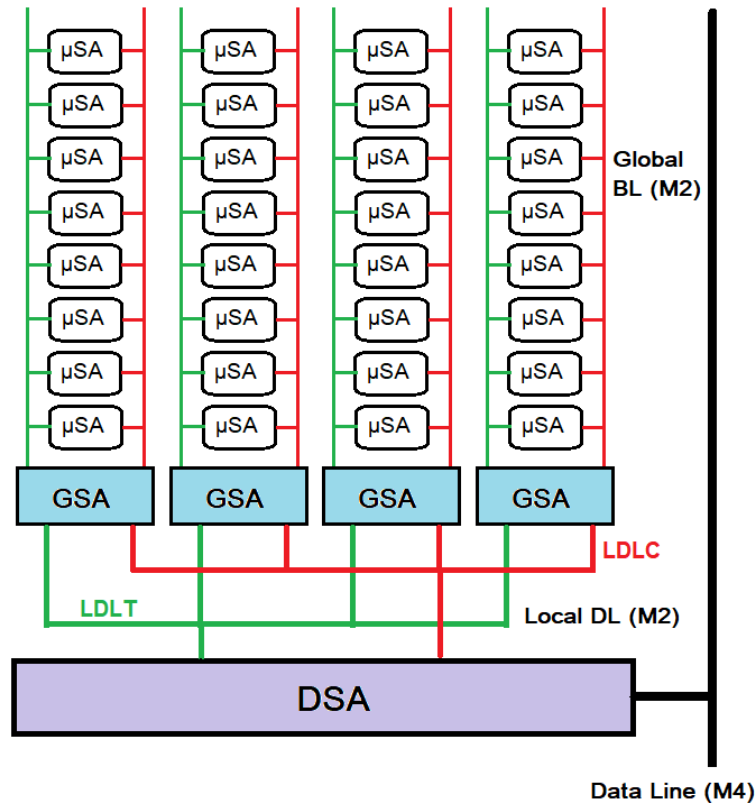

Figure 3.4: Sense Amplifier Architecture

Each DSA is connected to four such GSAs which are selected by Column Select.

3.3.1 3T μSA

As the name suggests, 3T μSA consists of three transistors, PCW0, FB and RH connected to the Local Bitline (LBL). The PCW0 (Precharge-Write 0) transistor is responsible for precharging the LBL to ground and also writing 0 into the cell. The FeedBack(FB) transistor is responsible for writing 1 into the cell and for autowrite back of 1 after a read operation. The Read Header (RH) is connected to the Read BitLine(RBL) and Write Bitline (WBL),which are controlled by the GSA.
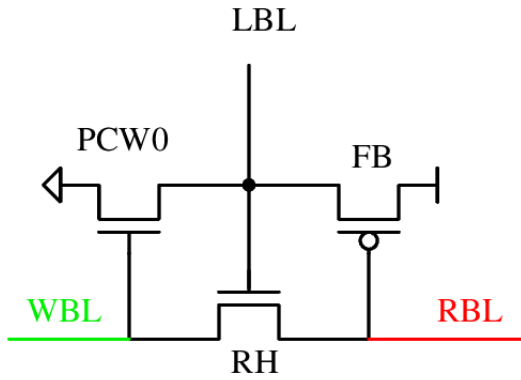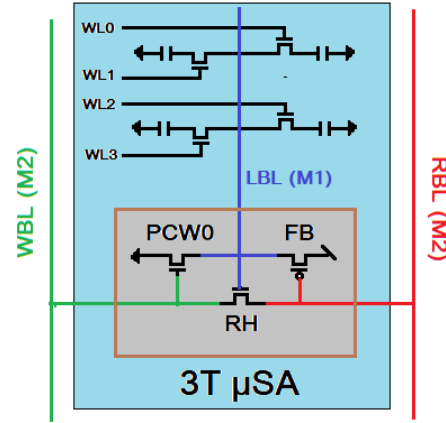

Figure 3.5: 3T μSA

Figure 3.6: 3T μSA Architecture

Before any operation, the local Bit-Line LBL is pre-discharged to ground by pre-charging WBL to HIGH.RBL is also pre-charged to HIGH. After discharging LBL, WBL is driven LOW by GSA and RBL is floating HIGH. The Word-Lines are activated, and the cell charge shares with the Bitline depending on the contents of the cell.

If the bit stored is a '0', then LBL is floating at ground and RBL is floating HIGH. For read '1', the LBL voltage rises after the charge sharing, sufficiently enough to turn on the RH. The RH transistor, pulls RBL LOW, thereby turning ON FB transistor for auto writeback of 1 into the cell.

RBL LOW is detected as Read 1, and RBL HIGH is detected as Read 0 by the GSA.

3.3.2 GSA

The Global Sense Amplifier (GSA) controls the two Global Bit-Lines WBL and RBL and is controlled by signals SEQn, SETp, BEQn, CSL and from the Local Data-Lines LDLT and LDLC from the Data Sense Amplifier.
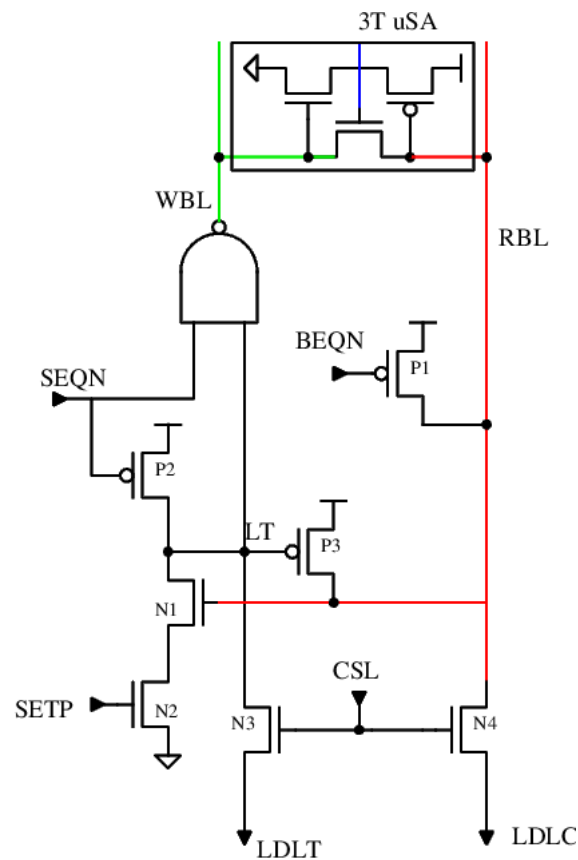
During Pre-Charge before a read or a search operation, RBL should be pre-charged to HIGH and WBL should be turned High and then driven to LOW. To achieve this, BEQn signal is driven LOW to precharge RBL to V_{DD}. The SEQN signal is LOW initially, to discharge the LBLs by making WBL HIGH. Turning SEQN LOW also drives LT node HIGH. SETp signal is kept LOW to turn OFF nmos N2, thus preventing the discharge of LT node. After the pre-charge state BEQn and SEQn signals are turned HIGH, thereby floating RBL and WBL at V_{DD}, ground respectively. CSL is the column select which connects a particular GSA to the DSA by LDLC and LDLT.

Figure 3.7: GSA

If the bit read is '0', then RBL remains in its pre-charged state. LT node is floating HIGH. After some time delay SETp signal is taken HIGH and N1 is also turned on due to RBL signal and hence LT node is pulled LOW. This drives the output of NAND HIGH and turns on PCW0 and enables a write back of a strong '0'. The nmos M3 connects LT and LDLT nodes. This reduction in LDLT node is detected by the Data Sense Amplifier as a '0'.
If the bit read is '1', then the RBL node goes LOW. The nmos N4 connects the RBL node to LDLC node. SETp signal is turned on after a time delay in this case too, but as RBL node goes low there is no pull down path for

the LT node.The Data Sense Amplifier resolves this reduction in voltage of LDLC node as a '1'.

3.3.3 Data Sense Amplifier

The bidirectional data sense amplifier (DSA) is responsible for transferring data between Global data lines in Metal layer 4 to the selected GSA. One of the four GSAs are selected by column select(CSL). When the DSA is not en-
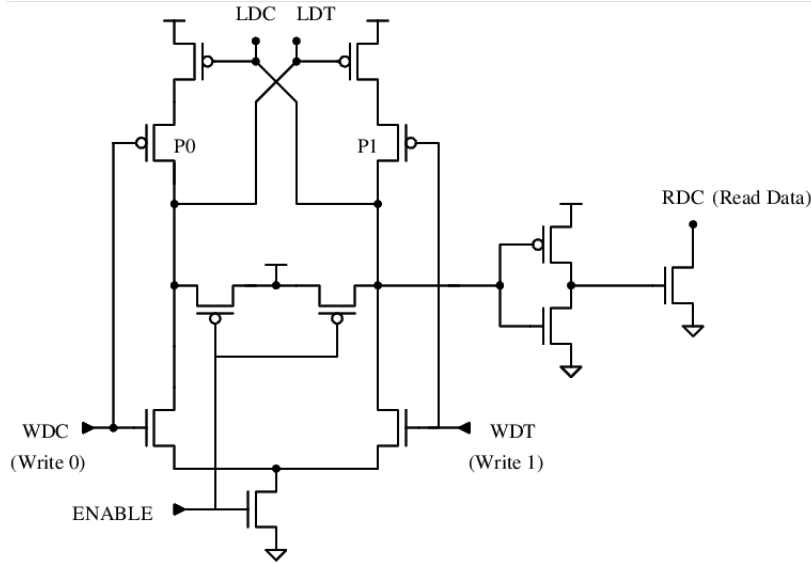

Figure 3.8: Data Sense Amplifier

abled, the two equalisation transistors, hold the LDLT and LDLC signals are held at the supply voltage. For a write operation, the global write data lines, WDT and WDC are driven high in a lower voltage domain. Using a pair of cross coupled PMOS transistors, a Cascade Switch Voltage Logic(CVSL) has been implemented for translating the signal to higher voltage domain. The transistors P0 and P1 provide improved voltage shifting by strongly turning off one of the PMOS stacks.

During a read operation,LDLC is either at logic LOW or logic HIGH depending on whether it is a read '1' or read '0' respectively. By default, LDLC is held at V_{dd} during the precharge. The inverter takes care of inverting LDLC signal to connect it to Global read data line (RDC). Hence, the global read dataline discharges for a read 0, and is floating at vdd for a read 1.

3.3.4 Integrating DSAs

Each Data Sense Amplifier would cater 256 wordlines, if we have 32 wordlines per LBL. To have 1024 wordlines, we integrate output of four such DSAs by

a dynamic NOR gate. Depending on the wordline accessed, the output of the corresponding DSA is reflected on the Global Dataline.
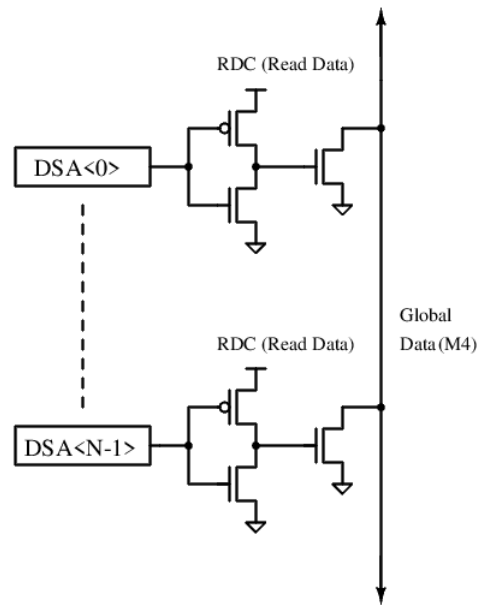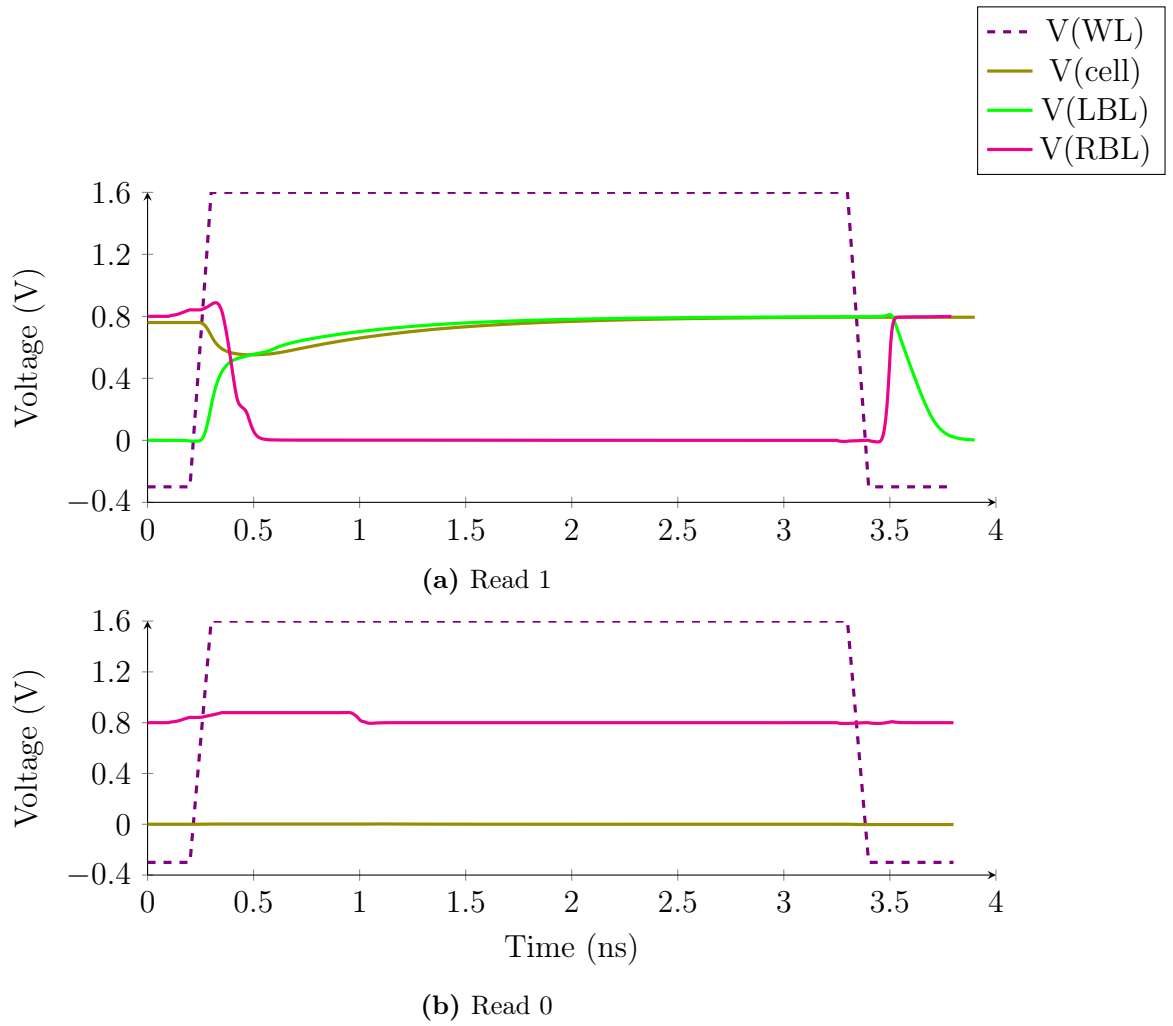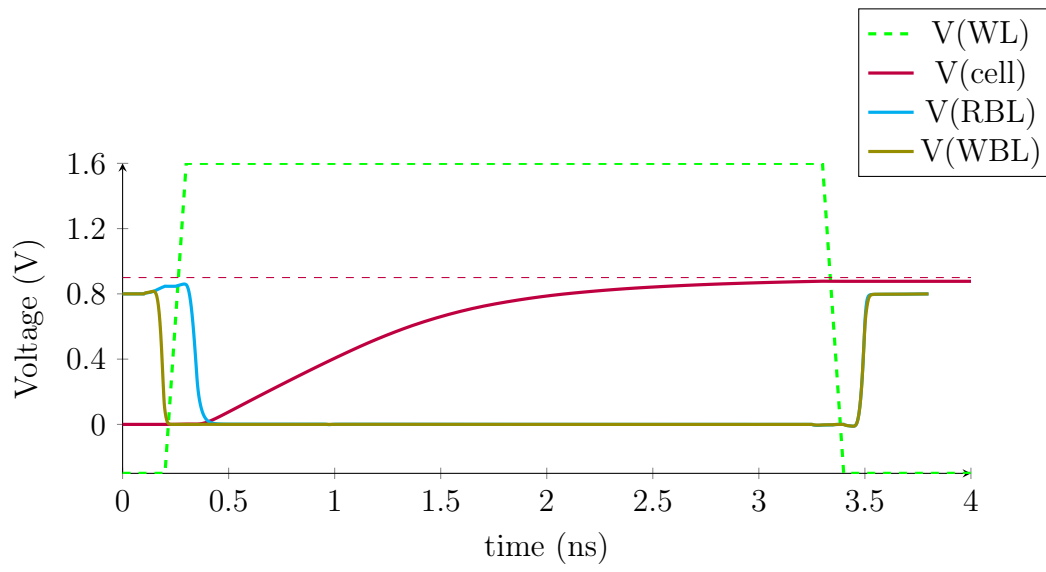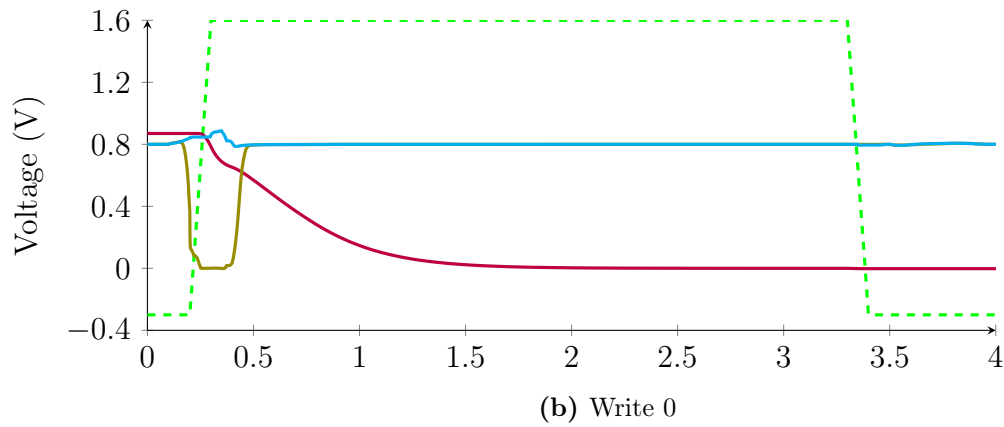

Figure 3.9: Dynamic NOR configuration to integrate DSA

3.3.5   Read Waveforms

(a) Read 1

(b) Read 0

3.3.6 Write waveforms

(a) Write 1

(b) Write 0

Chapter 4

Implementation of TCAM in eDRAM

4.1 Search operation using 3T uSA

Read operation is performed by letting the cell charge-share with the bitline. In our project, the word is stored across the bitline unlike the conventional DRAM, where data is stored across the wordline. Turning on multiple wordlines under the same Sense amplifiers would cause false reads and the incorrect data written into the cell. We intend to proceed further by performing a hierarchical search, by operating with single data bit, i.e., two DRAM bits in each uSA in parallel.

To improve the performance, where area overhead is not a concern, reducing bitline length is an accepted method to improve DRAM performance.Each time the bitline length is reduced,the sense circuits and data buffers must be doubled. [4]

4.1.1 Procedure for evaluating a match

Consider a 1Mb memory with 1024 wordlines, 1024 bitlines and 8 columns, i.e., 128 Datalines. As data is stored across the datalines, with 1024 wordlines, we store 512-bit word over 4 DSAs connected by a dynamic NOR gate.

1. The serial search operation begins by placing the search bits on wordlines of two bits in each microsense amplifier array. The subsequent wordline sets are turned on hierarchically.

2. Local bitlines, which are initially predischarged to ground, charge share with the cells. Local bitlines are at ground voltage in case of a match and are charged up, in case of a mismatch.

In Figure 4.1, it can be noted that the mismatch in the first bit (WL0, WL1) causes LBL00 to charge up, whereas LBL01 is floating at ground voltage(due to WL32, WL33). Similarly, other LBLs under each μSA are evaluated based on the search data.
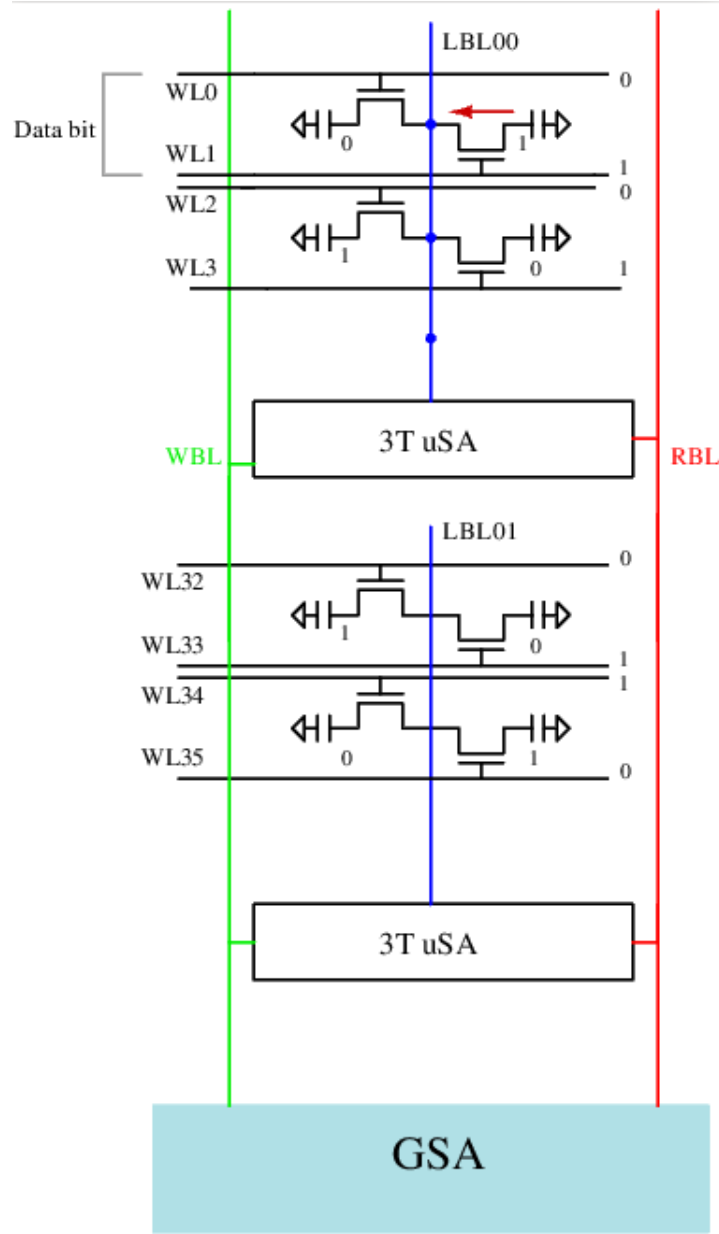
Figure 4.1: Search using 3T μSA

3. The evaluation of Match is similar to that of read 0 and mismatch, we read a '1'. Once there's a mismatch, the Bitline voltage rises and the global data line is pulled down to ground throughout the search operation indicating a mismatch in the word. For match and mask conditions, the global data line is at vdd.

With 32 WLs per LBL, in each cycle, we are able to evaluate 32 bits of the 512 bit word across 128 such datalines. To improve the performance of search,we reduce the number of WLs per LBL to 16 and doubling the μSAs under each GSA. In each cycle, we evaluate 64 bits.The status of the global dataline gives the result of this serial search.

4.2 Implementation of 4T μSA to prevent auto write-back

Read Bit line (RBL) is shared among 8 μSAs. In regular operation of eDRAM, depending on the WL accessed, one μSA is activated. But in search operation, as all μSAs are activated in parallel, conflicts occur when match and mismatch occur together. Due to the auto writeback nature through FB transistor of 3T μSA, incorrect value is written into the cell(The cell value becomes 11 which is invalid).
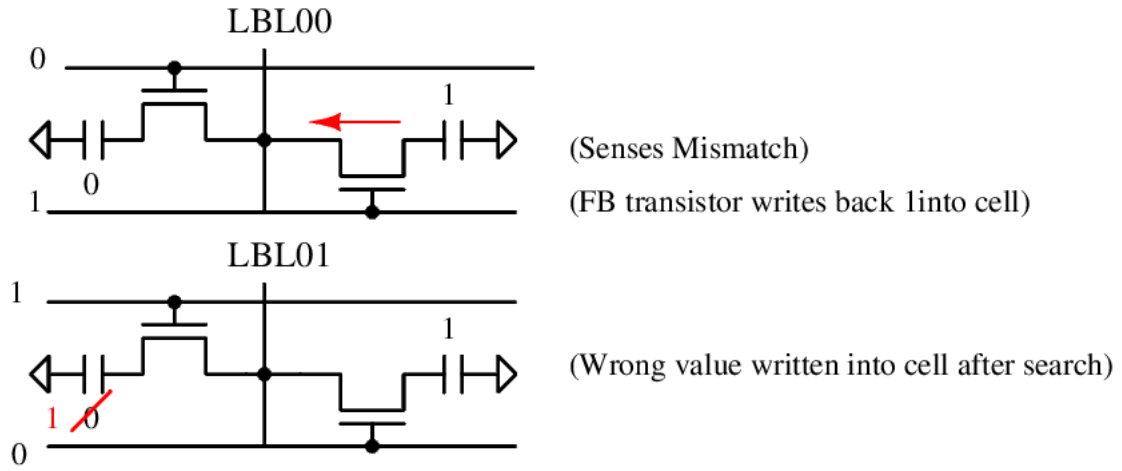

Figure 4.2: Wrong write-back

An extra PMOS is added in series with the FB transistor as in Fig4.3, to disable the writeback during the search. Note that the Search Enable is common to all the μSAs and can be routed easily. To restore the contents of the cell, a a refresh is performed after search operation.
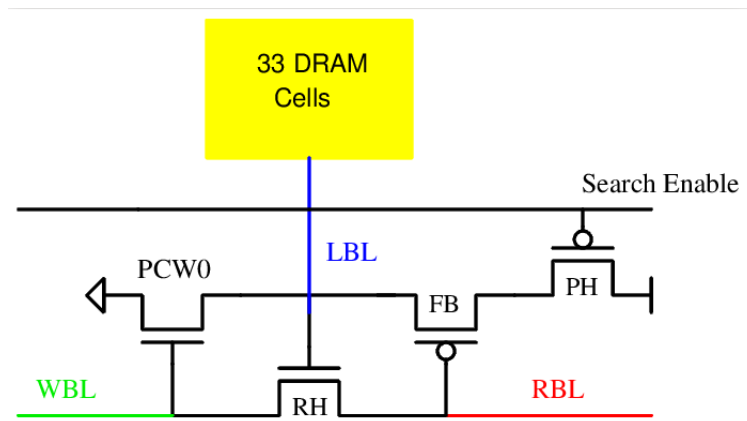

Figure 4.3: 4T μSA to prevent auto writeback

Chapter 5

Simulation Results

5.1 Control Signals

The following are the control signals for the GSA and DSA peripherals and they are plotted with respect to the word line (WLa).
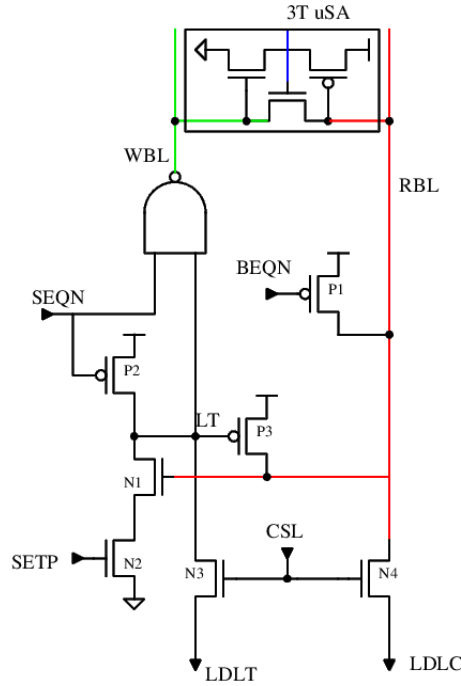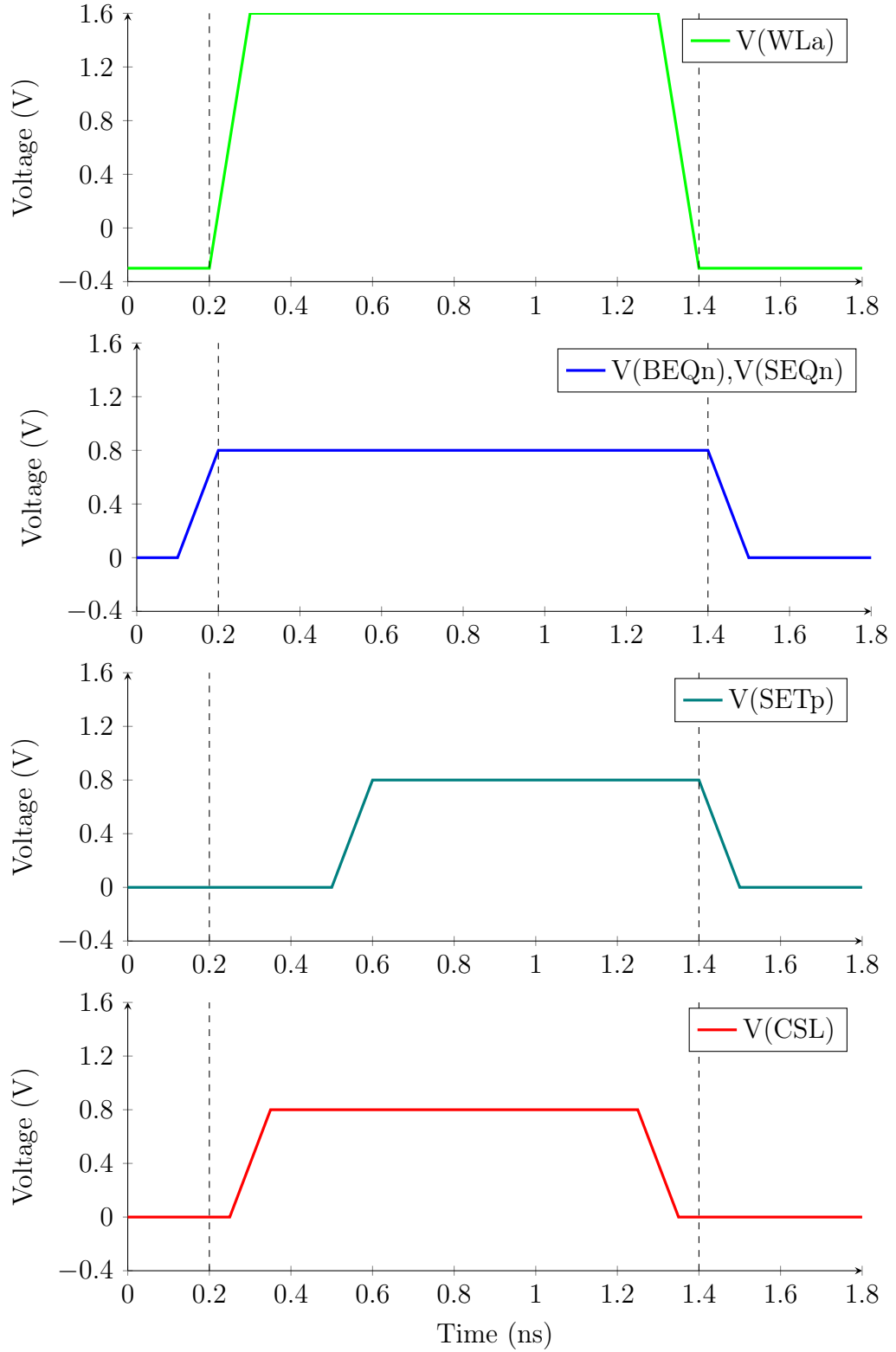
5.1.1 Control Signals for GSA

Figure 5.1: Schematic of GSA

Global Bit Lines RBL and WBL are connected to 4T μSA, and the Local Data Lines LDLC and LDLT are connected to the DSA. The control signals BEQn and SEQn are identical.

Figure 5.2: GSA control signals

The two vertical dashed Lines represent the start of the rising edge, and the end of the falling edge of a Word Line respectively when it is turned on.
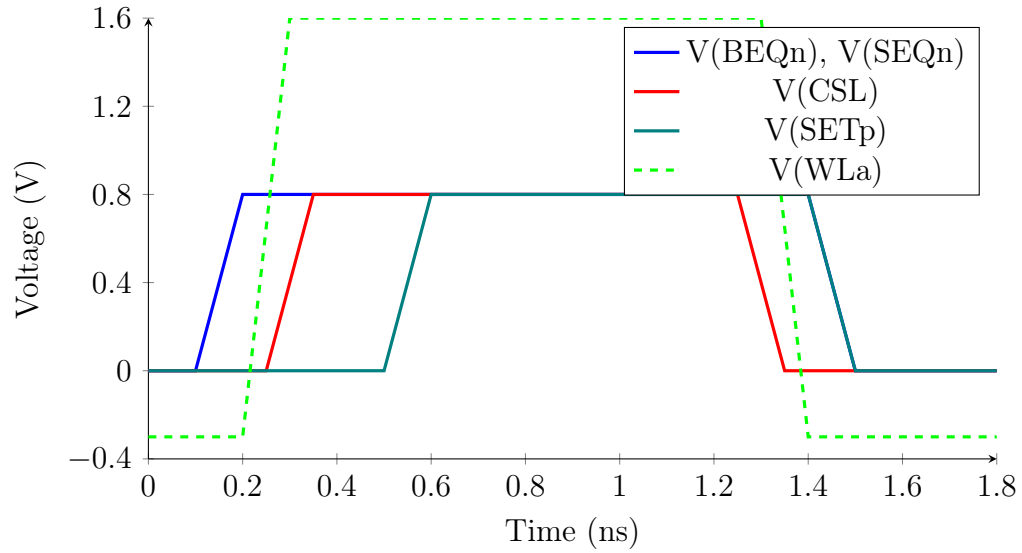
Figure 5.3: Control signals for GSA combined

5.1.2 Control Signals for DSA

RDC is the Global Data Line which is common to all DSAs in the y-direction. It is pre-charged to V_{dd}. In the event of a mismatch it is pulled low. In the remaining cases it remains floating at V_{dd}.
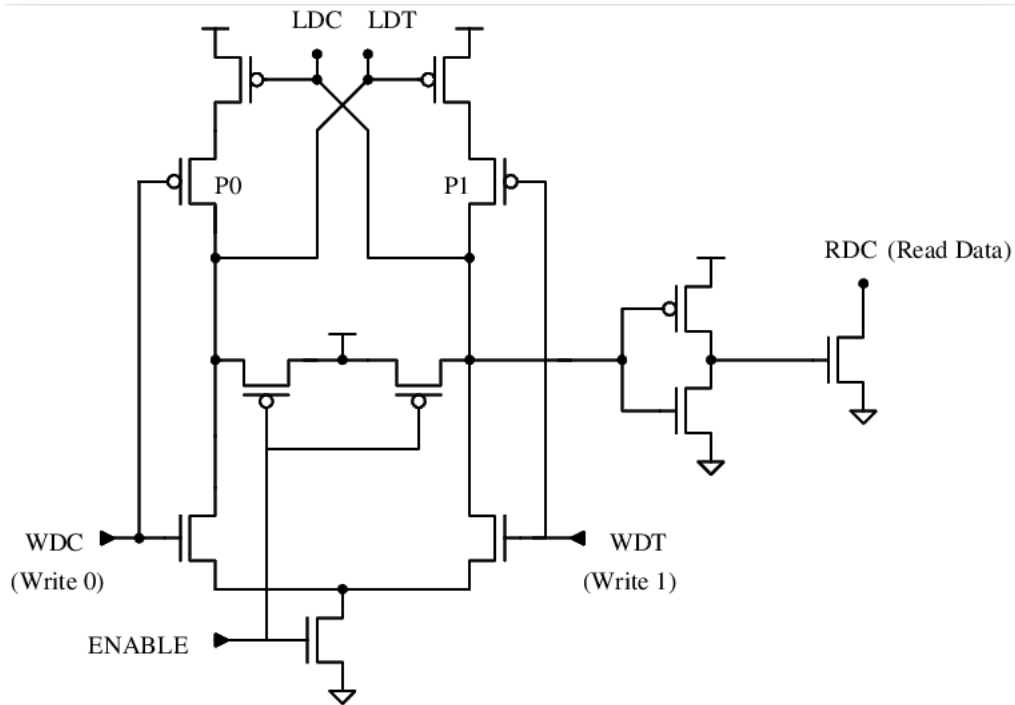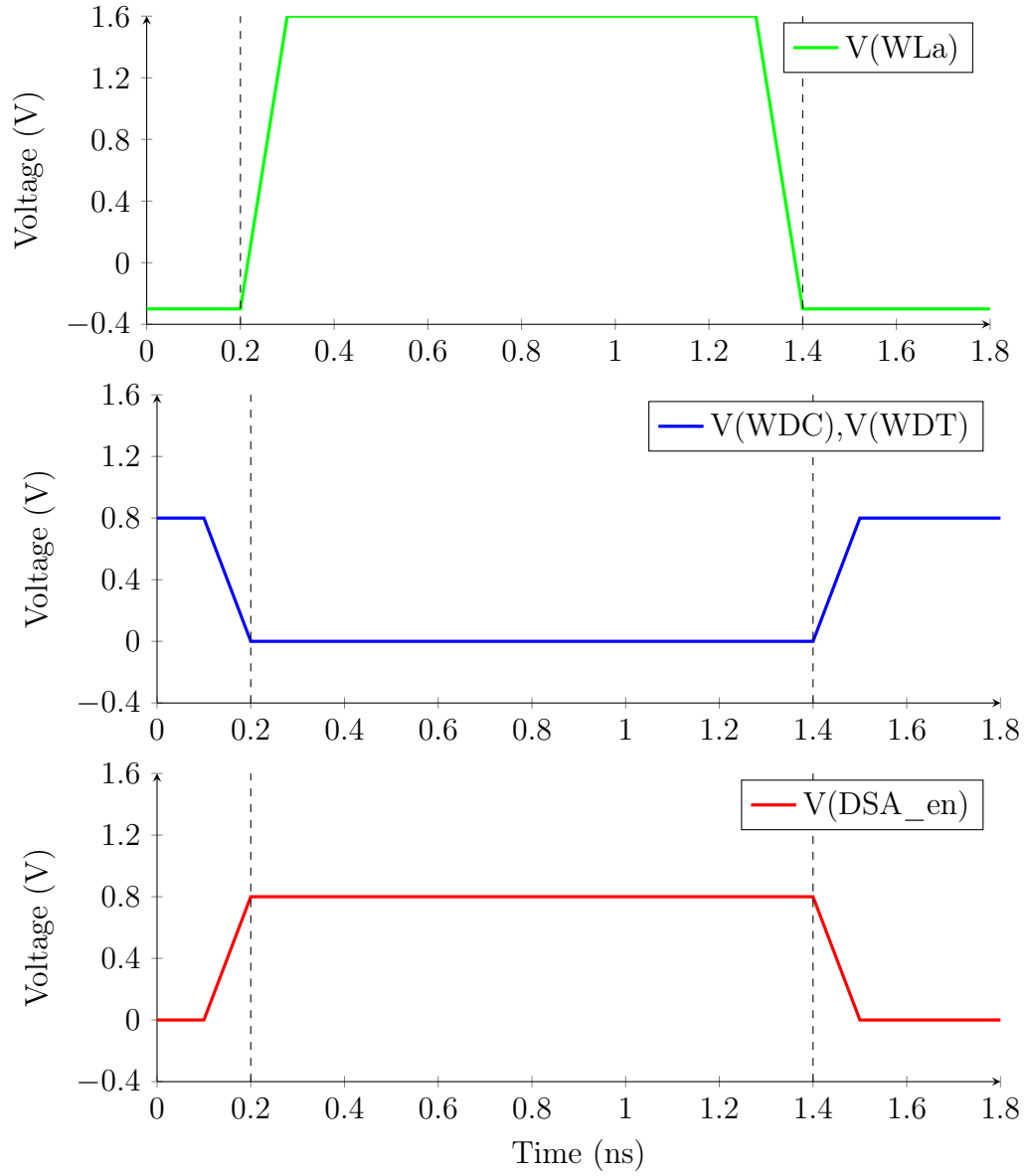

Figure 5.4: Schematic of DSA
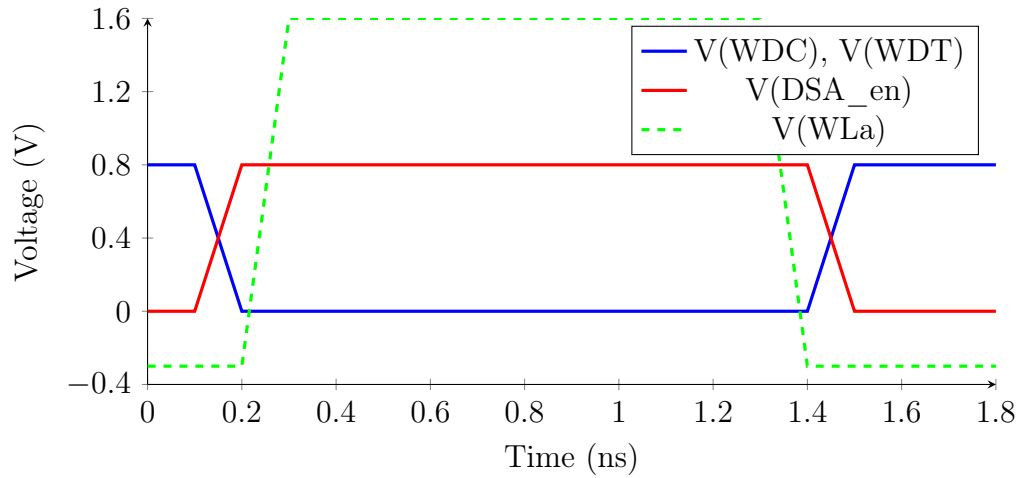
Figure 5.5: DSA control signals

Figure 5.6: Control signals for DSA combined

5.2 Search Operation Results

The following wave forms are the simulation results for a search operation of a single data bit match, mismatch, don't care and mask cases. The remaining data bits in the Local Bit Line are assumed to be in a match condition. VL and VR are the cell voltages of the left and right bit of a data bit. WLa and WLb are the word lines that control the left and right bit respectively.

5.2.1 Match

When there is a match, the cell voltages remain unchanged.
The Local Bitline (LBL) is in its pre-discharged state. In this case RBL node stays HIGH. SETp signal enable the write back of a strong '0' after some time delay, pulling down the LT node which then drives WBL HIGH.
When the column select is active, LDLC is connected to RBL and LDLT is connected to LT. In the case of a match the RDC node (Global Data Line) remains in its pre-charged state.
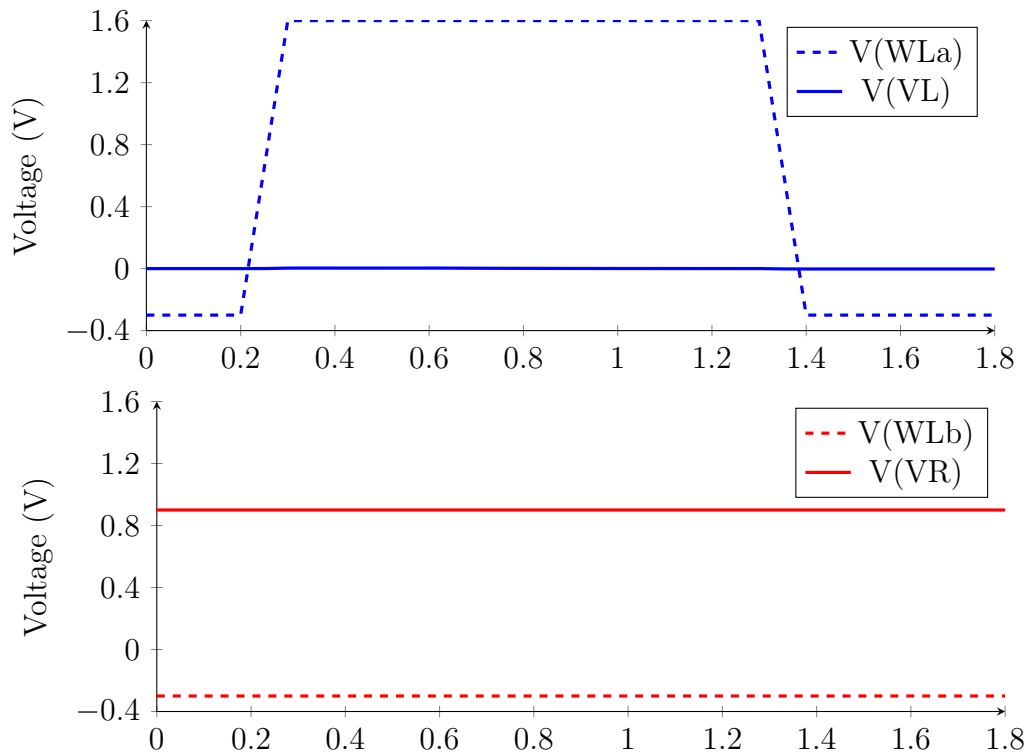

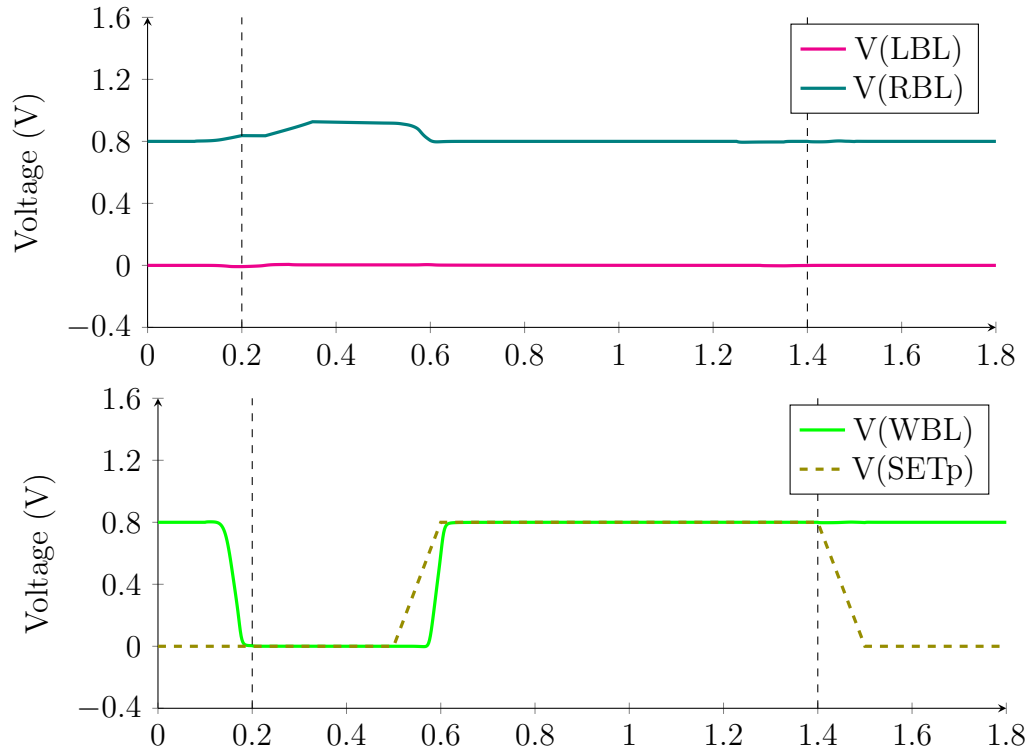**Figure 5.7:** Cell voltages $V_l = 0$ and $V_r = 1$ during a match

Figure 5.8: Bit Line Voltages LBL, WBL, and RBL during a match
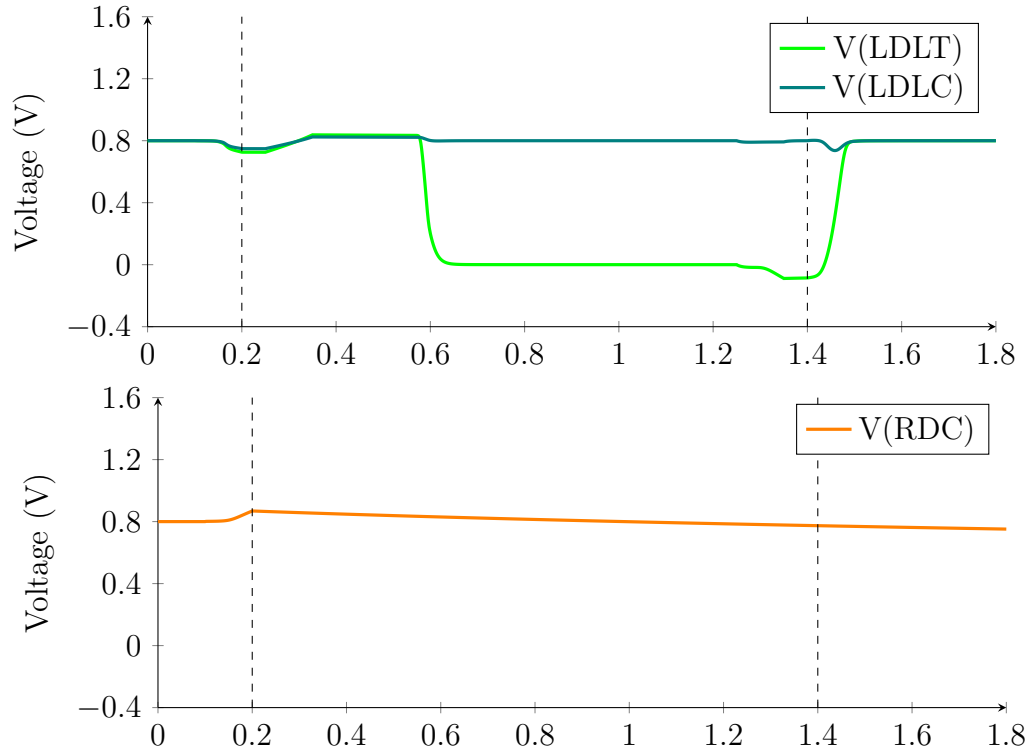

Figure 5.9: Data-Lines during a match

5.2.2 Mismatch

When there is a mismatch, the the bit that stored a strong '1' charge shares with the Local Bit Line (LBL), and they both becomes a weak '1'.

In this case RBL node is pulled down by 3T uSA. The write-back of a '1' is disabled through the pmos in 4T uSA to prevent wrong write-back.

When the column select is active, LDLC is connected to RBL and LDLT is connected to LT. The RDC node (Global Data Line) is pulled down during a mismatch.
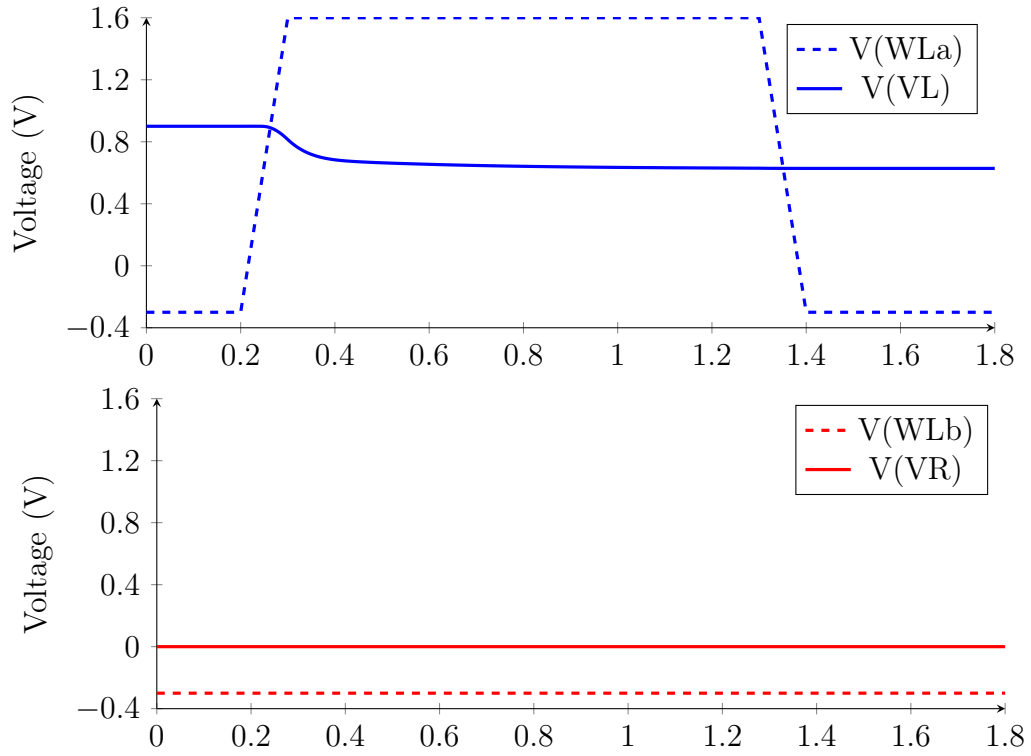

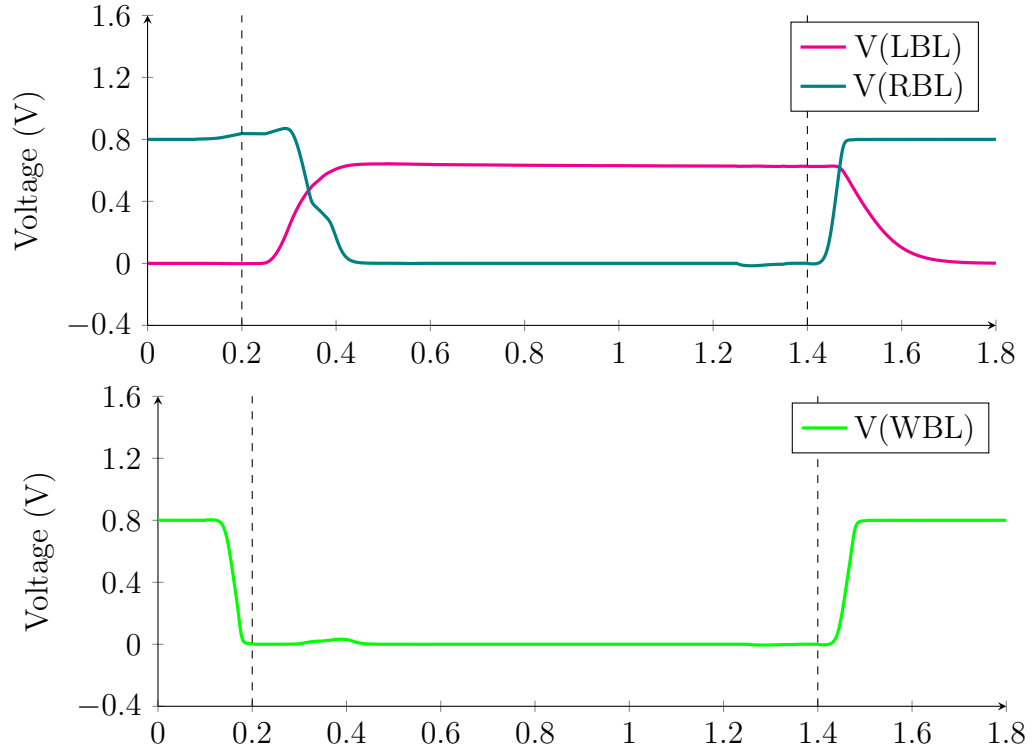**Figure 5.10:** Cell voltages $V_l = 1$ and $V_r = 0$ during a mismatch

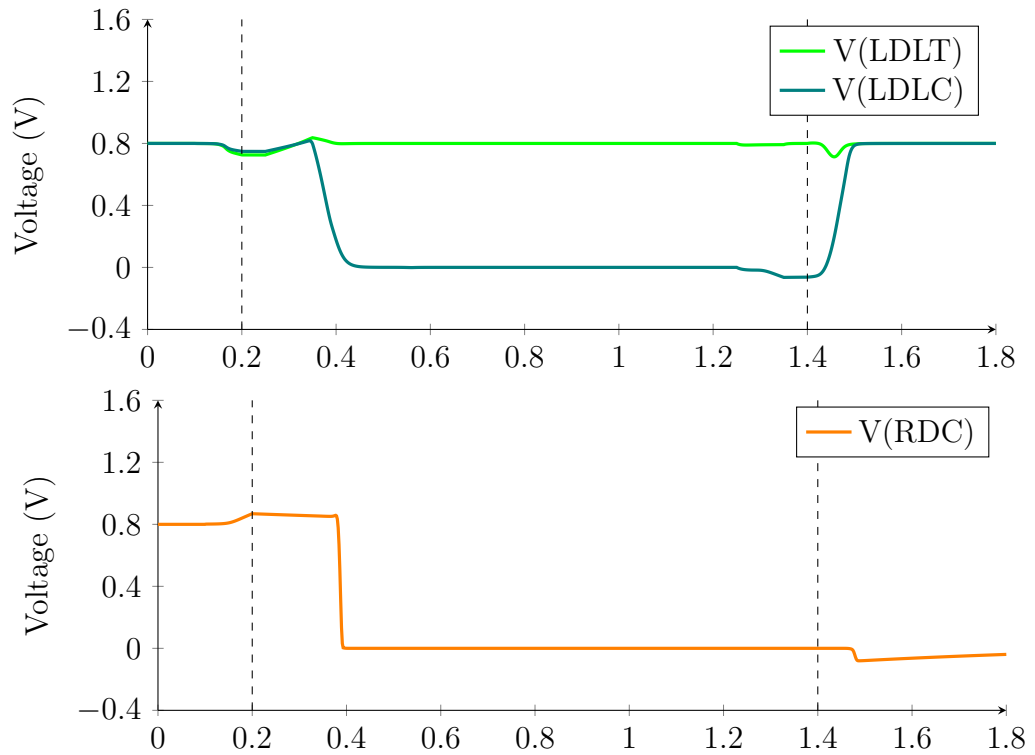**Figure 5.11:** Bit Line Voltages LBL, WBL, and RBL during a mismatch

Figure 5.12: Data-Lines during a mismatch

5.2.3   Mask and Don't Care condition

In a mask condition, both the Word Lines are turned OFF. In this case, the Local Bit Line LBL doesn't charge irrespective of the bits stored in a cell.

When the cell is in the Don't Care State, both the cell voltages are storing a '0'. In this case, the Local Bit Line LBL doesn't charge irrespective of turning on either one of the two Word Lines.

In both the above cases, RBL remains in its pre-discharged state. The Bit-Lines and Datalines are the in the same state similar to that of a match condition. RDC stays HIGH indicating a match.
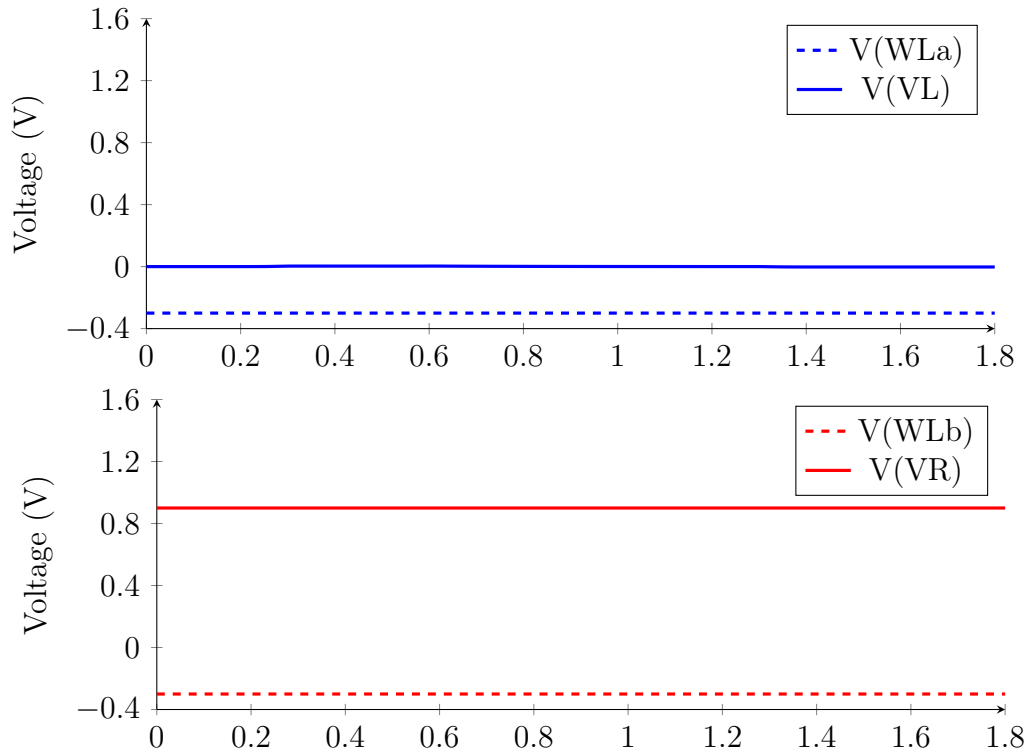

Figure 5.13: Cell voltages $V_l = 0$ and $V_r = 1$ during a mask
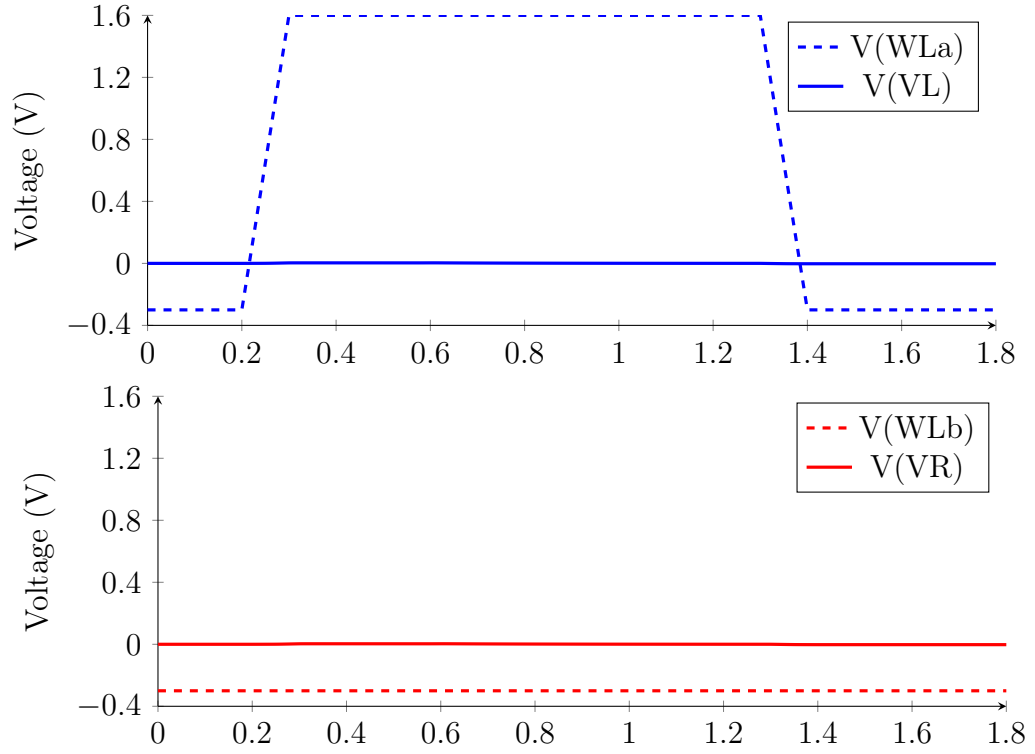
Figure 5.14: Don't Care state $V_l = 0$ and $V_r = 0$
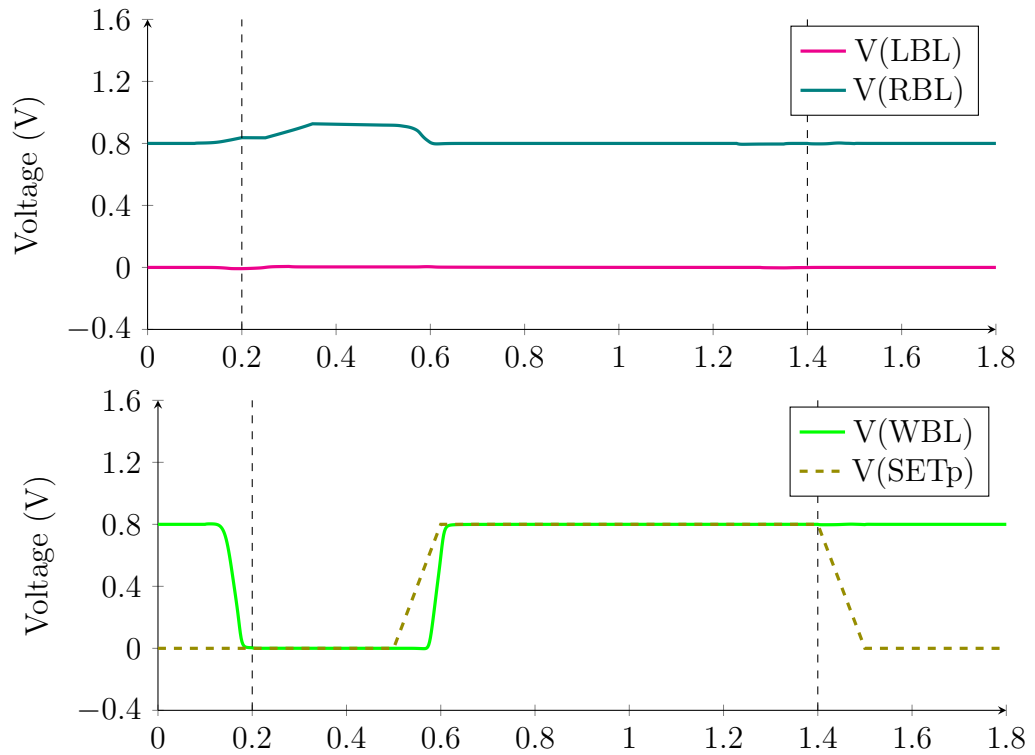

Figure 5.15: Bit Line Voltages LBL, WBL, and RBL during Don't care and mask condition
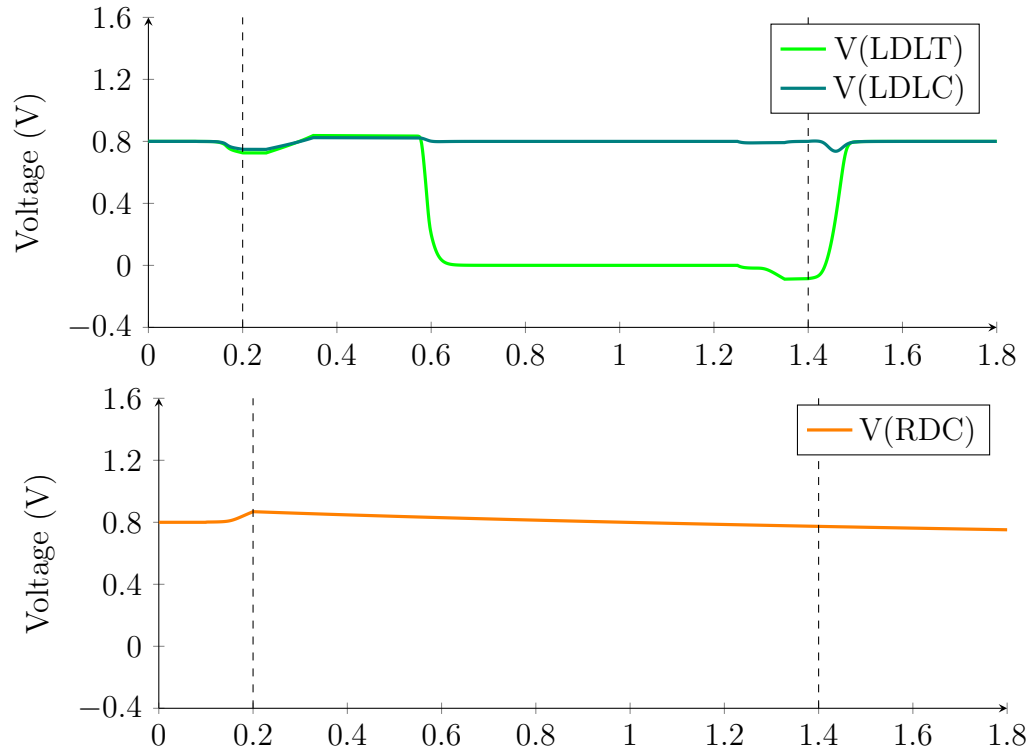
Figure 5.16: Data-Lines during don't care and mask condition

Chapter 6

Conclusion

The proposed DRAM architecture for TCAM was implemented on LT Spice with 22nm technology. The provision for comparison is made by implementing the XNOR using pass transistor logic. The status of the search is detected through the already available circuitry 3TμSA. 4T Micro Sense Amplifier was implemented to prevent the auto write-back during search mode. A refresh operation is performed after the search to restore the contents of the cell.We were able to perform the serial search without any extra peripheral circuitry for evaluating a match.

Appendix A

Leakage reduction in FinFET using circuit techniques

A.1 Self Controllable Voltage Level Circuit

A self controllable voltage level circuit[5] has been implemented to decrease the leakage power while maintaining high speed performance. The circuit supplies a higher dc voltage to the load circuit in active mode and decreases the dc voltage given to the load circuit in standby mode. The total power dissipation can be reduced by applying the Upper SVL circuit that results in decreased supply potential and Lower SVL circuit that results in raised ground potential[5].

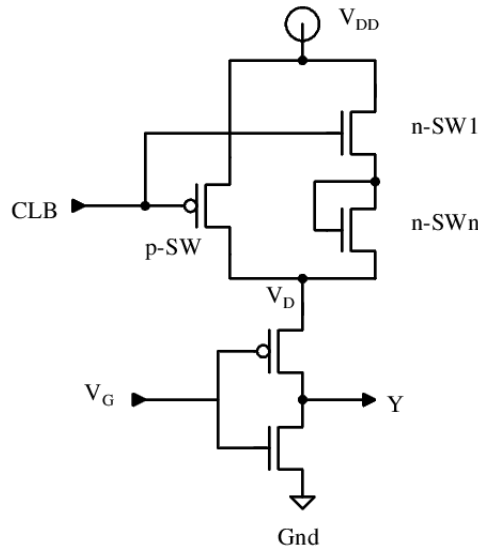
Figure A.1: Upper SVL circuit

The Upper SVL circuit has single PMOS switch and n NMOS switches.In active mode, the p-SW is turned on and it passes the the full supply voltage to the load circuit. In standby mode, when CLB becomes 1, the nmos

switches are weakly turned on, and the voltage V_D is expressed as

$$V_D = V_{DD} - n v_n$$

where v_n is the voltage across NMOS stack.

V_D can be decreased by increasing v_n which increases the barrier height by reducing DIBL. This results in increase of the threshold voltage of nmos and decreases sub-threshold leakage.
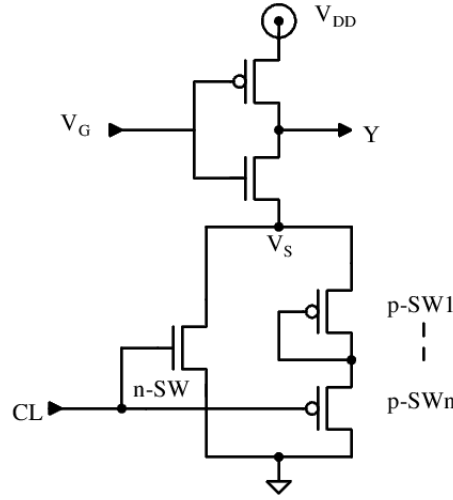
Figure A.2: Lower SVL circuit

In Lower SVL circuit operating in standby mode, pmos stack is turned on and nmos is turned off. The voltage V_S is expressed as

$$V_S = v_p$$

where v_p is the voltage across PMOS stack. The voltage across the load circuit inverter (V_{DS}) is V_{DD}-V_S. The reduction DIBL due to reduced V_{DS} results in lower sub-threshold leakage current.

DIBL effect is further reduced by combining upper and lower SVL circuits together.

$$V_{DS} = V_{DD} - v_n - v_p$$
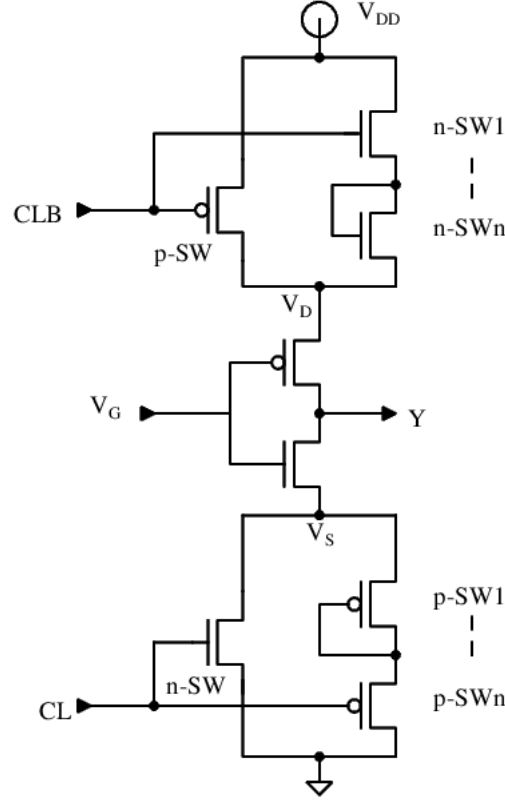
Circuit	Vgs=0	Vgs=0.8V
Inverter	14.567nA	11.063nA
UpperSVL	11.58nA	3.72nA
Upper+Lower SVL	3.433nA	3.50nA

Table A.1: Leakage currents of inverter load circuit with SVL

By employing SVL circuit, leakage current of PMOS has reduced by 68.36% and leakage current of NMOS has reduced by 76.43% compared to that of conventional inverter.

The header and footer devices in SVL circuits are upsized to accommodate load circuits in active mode, to reduce the wake up time. But when the devices are upscaled, the observed reduction in supply voltage of load circuit in sleep mode is less. This is due to the lower resistance offered by the devices, which increases the leakage. It is observed that if the circuit is left in sleep mode for a long time, sub-threshold leakage drifts the node voltages V_D and V_S to V_{DD} and ground respectively. Leakage reduction can also be improved by increasing the number of switches in the stack. Due to the area penalty and other disadvantages stated, this is not a scalable approach.

A.2 Study of effect of fingering and impact of process variations

Lower transistor of two stack nmos is split into 10 fingers and the impact of process variations was studied. The threshold voltage of the fingers was varied as a Gaussian distribution with a variance of $+/-$ 5%. From the simulations it was noted that even if there was a one finger with considerably low V_T compared to nominal device, the observed reduction in leakage is less. To overcome this problem, we need to systematically increase the V_T of the fingers. It was intended to use bad layout techniques to reduce leakage by introducing process variations. But this requires fabrication of devices to verify the claims. This study did not yield a satisfactory result, hence we proceeded by the study of TCAMs and Implementation of TCAM in eDRAM.

Bibliography

[1] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (cam) circuits and architectures: a tutorial and survey," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 3, pp. 712–727, 2006.

[2] J. Delgado-Frias, A. Yu, and J. Nyathi, "A dynamic content addressable memory using a 4-transistor cell," in *Proceedings of the Third International Workshop on Design of Mixed-Mode Integrated Circuits and Applications (Cat. No.99EX303)*, 1999, pp. 110–113.

[3] V. Lines, A. Ahmed, P. Ma, S. Ma, R. McKenzie, H.-S. Kim, and C. Mar, "66 mhz 2.3 m ternary dynamic content addressable memory," in *Records of the IEEE International Workshop on Memory Technology, Design and Testing*, 2000, pp. 101–105.

[4] J. Barth, W. R. Reohr, P. Parries, G. Fredeman, J. Golz, S. E. Schuster, R. E. Matick, H. Hunter, C. C. Tanner, J. Harig, H. Kim, B. A. Khan, J. Griesemer, R. P. Havreluk, K. Yanagisawa, T. Kirihata, and S. S. Iyer, "A 500 mhz random cycle, 1.5 ns latency, soi embedded dram macro featuring a three-transistor micro sense amplifier," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 1, pp. 86–95, 2008.

[5] Y. O. Tadayoshi Enomotoe and H. Shikano, "A self- controllable-voltage-level (svl) circuit for low-power, high-speed cmos circuits," pp. 411–413, 2002.