**CHAPTER 1**

**INTRODUCTION**

**1.1 MOTIVATION**

Power reduction in SoC based embedded systems, is one of the important design specifications of VLSI design. This project deals with the reduction in delay and power consumed by memory in SoC with acceptable trade off of stability and performance.

SoCs integrate multiple functions on a single silicon die. As process geometries have scaled, designs which use more and more of the additional silicon real estate available on chips to integrate embedded memories evolved. These embedded memories allow for significantly better system performance and lower power compared to a solution where off-chip memories are used. Most current designs have over 50% of their area used by embedded memories and these memories account for 50-70% of the total SoC power dissipation. Clearly, any attempt to reduce SoC power is incomplete if it does not attempt to reduce the power consumed by the embedded memories in the design.

Embedded cache memories are implemented using SRAM cells.SRAM is a type of [semiconductor](http://en.wikipedia.org/wiki/Semiconductor) memory that uses [bi-stable](http://en.wikipedia.org/wiki/Multivibrator)[latching circuitry](http://en.wikipedia.org/wiki/Flip-flop_(electronics)) to store each bit. A conventional SRAM has a 6T design. It is volatile in the conventional sense that data is eventually lost when the memory is not powered. So designing the SRAMs which consume less power becomes crucial. The cache is a smaller, faster memory which stores copies of the data from the most frequently used [main memory](http://en.wikipedia.org/wiki/Main_memory) locations. As long as most memory accesses are cached memory locations, the accesses time of memory will be closer to the accesses time of cache than to the access time of main memory. Hence improving the speed of cache becomes important.

**1.2 OBJECTIVE**

The objective of this project is to implement a cache memory using 4T SRAM and compare it with conventional 6T SRAM cache for parameters like power, area and speed of operation, then implement buffered bit-line technique on 4T SRAM cache to get an optimized design with low power and lesser delay.

**CHAPTER 2**

**LITERATURE SURVEY**

**2.1 INTRODUCTION**

This chapter includes all the reference papers. Objective, concepts and experimental results of each paper are presented. Advantages and drawbacks of each paper are discussed.

**2.2 PAPER 1**

**Title:**ArashAziziMazreah, Mohammad Noorollahi Romani, Mohammad TaghiManzuri, Ali Mehrparvar, *“A low power and high density cache memory on novel SRAM cell”*, IEICE Electronics Express, Vol.6, August 2009

**2.2.1 OBJECTIVE**

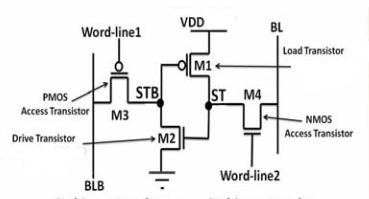
The objective is to develop an SRAM cell with four transistors to reduce the cell area size and power consumption with no performance degradation.

**2.2.2 INTRODUCTION**

Conventional SRAMs that use 6T SRAM cells have difficulty in meeting the growing demand for a larger memory capacity in mobile applications. Furthermore, in a conventional SRAM cell one of the two bit-lines must be discharged regardless of the written value. Therefore the power consumption in writing both “0” and “1” are the same. Also during the read operation one of the two bit-lines must be discharged irrespective of the value stored in the cell. Therefore there are always transitions on bit-lines in both writing “0” and “1” and reading “0” and “1”. Furthermore 6T SRAM cell uses full swing on wordlines and these cause high dynamic power consumption during read and write operations. In response to this requirement an SRAM with 4 transistors is developed. The power consumption of writing and reading zeros in novel cell is much smaller than ones, thus the average power consumption is reduced in caches based on this novel cell.

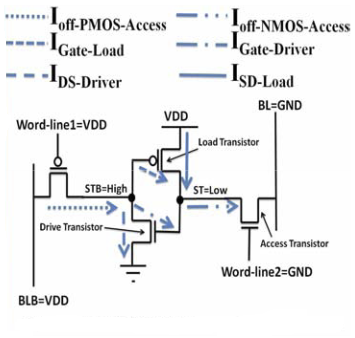
**2.2.3 CELL DESIGN CONCEPT**

Fig (2.1) shows the circuit equivalent to the developed 4T SRAM cell.

****

**Fig (2.1) 4T SRAM Cell**

When “1” is stored in the cell, load and drive transistors are ON and there is a positive feedback between ST and STB node. Therefore STB node is pulled to GND by drive transistor and ST node is pulled to VDD by load transistor.

****

**Fig (2.2) 4T SRAM Cell in idle mode when “0” is stored**

When “0” is stored in the cell both load and drive transistors are OFF and for data retention without refresh cycle, the following conditions must be satisfied.

IOff-NMOS-access > ISD-Load + IGate-Driver + IGate-Load **Eq (2.1)**

IOff-PMOS-access > IDS-Load + IGate-Driver + IGate-Load **Eq (2.2)**

Fig (2.2) shows leakage current of cell during idle mode for data retention when “0” is stored in the cell. For satisfying the above condition when “0” is stored in the cell we use leakage current of access transistors, especially sub-threshold current of access transistors (IOff-NMOS-access and IOff-PMOS-access ).

To achieve this we can use high threshold voltage for load and drive transistors to reduce sub-threshold currents of these transistors. Low threshold voltage can be used for access transistors; hence leakage current of access transistors will be greater than leakage currents of load and drive transistors. With this threshold voltage assignments above conditions can be satisfied and “0” can be stored in a cell successfully. Leakage currents also depend on the width of the transistors. As width of a transistor increases, leakage currents also increase. As the leakage currents of access transistors should be greater than leakage currents of load and drive transistors, the width of access transistors should be greater than the width of load and drive transistors.

The specifications for 4T SRAM cell are mentioned as follows:

Technology used: 65nm

VDD = 1.2V

(W/L)M1=130nm/65nm­ (W/L)M2 =130nm/65nm­

(W/L)M3=260nm/65nm­ (W/L)M4 =260nm/65nm­

Normal threshold voltage: VTN = 0.32V VTP = 0.33V

High threshold voltage: VTN = 0.52V VTP = 0.53V

**2.2.4 WRITE AND READ OPERATION**

* **Write operation:** When a write operation is issued the memory cell will go through the following steps.

**1) Bit-line driving:** For a write, data is placed on BL and its compliment is placed on BLB, then WL1 and WL2 are asserted to GND and VDD respectively.

**2) Cell flipping:** This step includes two states as follows.

**(a) When data is logic ‘0’:** In this state, ST node is pulled down to GND by NMOS access transistor STB node will be pulled up to VDD by PMOS access transistor.

**b) When data is logic ‘1’:** In this state, ST nodepulled up to VDD-VTN by NMOS access transistor, and therefore the drive transistor will be ON and STB node will be pulled up to VTP by PMOS access transistor. Load and drive transistors will be ON and positive feedback is created by load and drive transistors between ST and STB nodes.

**3) Idle mode:** At the end of write operation, cell will go to idle mode and WL1 and WL2 are asserted to VDD and GND respectively and BL and BLB return to GND and VDD respectively.

* **Read operation:**

When a read operation is issued the memory cell will go through the following steps.

**1) Bit-line Pre-charging:** For a read, BLB is pre-charged to VDDand then floated. Since, in idle mode BLB maintained at VDD, this step doesn’t include any dynamic energy consumption.

**2) Word-line activation:** In this step WL1is asserted to GND and two states can be considered.

**(a) Stored data is “1”:** When voltage of STB node is low, the voltage of BLB pulled up to low voltage by NMOS access transistor. We refer to this voltage of BLB as VBLB-LOW.

**(b) Stored data is “0”:** When voltage of STB node is high, the voltage of BLB and STB node are equalized. Since in this state, there is very small different between BLB and STB node, power consumption is very small.

**3) Sensing:**After WL1is deactivated, the sense amplifier is enabled to read data on BLB.

**4) Idle mode:** At the end of read operation, cell will go to idle mode and BLBis asserted to VDD.

**2.2.5 LEAKAGE CURRENTS**

4T SRAM cell has to retain its value using leakage currents in one state (when “0” is stored) whereas in another state it retains the data using positive feedback (when “1” is stored). In idle mode when “1” is stored, load and drive transistors are ON thereby creating a positive feedback leakage and access transistors have sub-threshold current. But since load and drive transistors have high threshold voltage, therefore there are paths from supply to ground. Thus leakage current in “1’’ state is greater than that in “0’’ state.Leakage current of 6T SRAM cell when ‘‘0’’ or ‘‘1’’ stored is approximately equal to leakage current of 4T SRAM cell when ‘‘1’’ is stored.Leakage current of 4T SRAM cell when ‘‘0’’ is stored is much smaller than leakage current of 6T SRAM cell when ‘‘0’’ or “1” is stored. Most bits of caches are zeros for both data and instruction streams, so average leakage current is smaller in case of 4T SRAM cache when compared to 6T SRAM cache.

**2.2.6 EXPERIMENTAL RESULTS**

**Dynamic power consumption:**

In idle mode BL and BLB are maintained at GND and VDD, hence there is no transition on bit-lines while writing ‘0’.Hence in 4T SRAM power consumed for writing ‘‘0’’ is smaller than that needed for writing ‘‘1’’, whereas in 6T SRAM both cases consume same power. Further in 4T SRAM cell, when “0” is read from cell, there is no transition on BLB and hence power consumption of reading ‘0’ is smaller than reading “1”. In 6T power consumed for reading and writing “1” is same.

From H-SPICE simulations it is obtained that:

* Power consumption of writing “1” in 4T cell is approximately equal with power consumption of writing “1” or “0” in 6T cell.
* Power consumption of reading “1” or “0” in 4T cell are smaller than power consumption of reading “1” or “0” in 6T cell.
* Average dynamic power consumption of new cell is 40% smaller than that of 6T cell.
* Static power consumption of new cell is 20% less than that of 6T cell.

**Cell area:**The new cell size is 20% smaller than conventional 6T cell size.

**2.2.7 CONCLUSION**

* **Advantages:**Cell area, static power consumption and dynamic power consumption have decreased.
* **Disadvantages:** Delay has increased and SNM has decreased.

**2.3 PAPER 2**

**Title:**

Stefan Cosemans, WimDehaene and FranckyCatthoor, *“A Low-Power Embedded SRAM for Wireless Applications”*, IEEE Journal of Solid- State circuits, Vol. 42, No. 7, July2007

**2.3.1 OBJECTIVE**

The objective of this paper is to introduce a novel ultra-low-power SRAM in which large power reduction is obtained by the use of four new techniques that allow for a wider and better trade-off between area, delay and active and passive energy consumption.

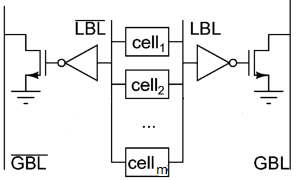
**2.3.2 INTRODUCTION**

Embedded memories play a crucial role in contemporaryelectronic systems. They are used in different sizes ranging from a few kilobytes for local scratchpads to a few megabytes for on-chip caches. This paper focuses on embedded SRAMs smaller than 1 Mb for use in the lowest levels of the memory hierarchy. Memories at this level of the hierarchy are used very intensively, so their energy consumption has a signiﬁcant impact on the energy consumption of the entire system. In this paper, a novel SRAM design is introduced. The implemented design techniques consist of a more efﬁcient memory databus, the exploitation of the dynamic read stability of SRAM cells, a new low-swing write technique and a distributed decoder.

**2.3.3 SHORT BUFFERED BITLINE TECHNIQUE**

In traditional low-power memory designs, the cell read current Iread,cell must create a large enough voltage difference onthe bit-lines. Since the bit-line capacitance is very large, this stepmakes up a large part of the memory access delay. Therefore Iread,cell must be made as large as possible, which results in high cell leakage currents because a large requires large transistor widths, low threshold voltages or a high supply voltage for the cell. Nominal value of Iread,cell is important. Because the cells need to be very small, the intra-die variation of Iread,cell will be large. This requires a large safety margin on the memory delay. Since all cells will remain activated until the slowest cell is ready, this also causes an important increase in energy consumption.

These problems are remedied when the amount of charge that the cell must draw from the bit-line is reduced. Therefore, the bit-line is divided into shorter local bit-lines. Buffer connects local and global bit-lines as shown in fig (2.3).Inverter acts as a sensing element for the buffer.LBL uses a large voltage swing but since capacitance is less, it has limited impact on energy consumption. Buffer consists of a scaled up NMOS transistor. Iread,buffer can be much larger than Iread,cell and suffers less from intra-die variation. Low voltage power supply can be used as precharge voltage to the GBL which reduces power consumption.



**Fig (2.3) Buffered Bit-lines**

**Impact on Memory Databus:**

In traditional designs, amplification at the column level is required to limit Iread,cell impact on memory speed. In buffered bit-line approach, buffer can easily deliver more current so global bit-lines can be directly extended to memory output. Only one set of sense amplifiers is required for entire memory so area overhead is reduced.

|  |  |
| --- | --- |
| **Traditional Solution** | **Proposed Solution** |
| C:\Users\nikhila\Desktop\Capture2.PNG | C:\Users\nikhila\Desktop\Capture3.PNG |

**Fig (2.4) Memory databus**

**2.3.4 CONCLUSION**

* **Advantages:** Since short bit-lines are used, capacitance offered has decreased. Hence power and delay have reduced.
* **Disadvantages:**Instead of using a single bit-line, number of LBLs and a GBL are being used, due to whicharea overhead has increased.
  1. **PAPER 3**

**Title:**

Karandikar and K. K. Parhi, *“Low power SRAM design using hierarchical divided bit-line approach,”* in Proc. Int. Conf. Computer Design: VLSI in Computers and Processors, 1998, pp. 82–88.

**2.4.1 OBJECTIVE**

This objective of this paper is to present a novel hierarchical divided bit-line approach for reducing active power in SRAMs byreducing bit-line capacitance.

**2.4.2 INTRODUCTION**

Designing a low power system not only reduces weight and size ofbatteries for portable systems butalso helps in reducing the ever important packagingcosts of integrated circuits. To this end, the design of low power digital systems is becoming increasingly important. This paper describes a novel divided bit-line approach for reducing the active power by reducing the bit-line capacitance and then extends it to a hierarchical divided bit-line approach. It is shown that by reducing this capacitance, not only power reduction is achieved but also access time is reduced.Two or more 6T SRAM cells are combined together to divide the bit-line in to several sub bit- lines. These sub bit-lines are again combined to form two or more levels of hierarchy. Optimum values for numberof levels of hierarchy and number of blocks combinedat each level have been derived.

**2.4.3 CONCEPT**

**2.4.4 EXPERIMENTAL RESULTS**

A 2K x 8 bits SRAM chip using MAGIC layout tool is designed; 6 and 8 cells are combined at the sub bit-line level.

* Active power consumption has reduced approximately by 50-60%.
* Access time is reduced by 30%.

**2.4.5 CONCLUSION**

* **Advantages:**By dividing bit-line into sub bit-lines, SRAM cells become more stable, as they areguarded from the noise on bit-lines through pass transistors. Active power consumption and access time have decreased.
* **Disadvantages:** Due to hierarchical bit-line setup area overhead increases.

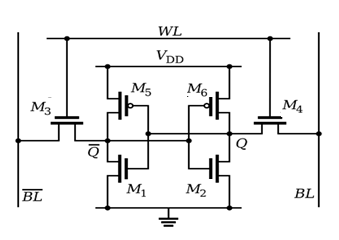
**CHAPTER 3**

**6T SRAM CONCEPTS**

In this chapter the concepts of 6T SRAM cells are discussed. Different operations (read, write and hold) taking place, design considerations and the working in detail is explained for 6T SRAM cell.

**3.1 6T SRAM CELL**

6T SRAM cell shown in Fig (3.1) uses a simple bistable latch circuit to hold a data bit. A pair of cross coupled inverters provides the storage while two access transistors (NFETs) provide read and write operations. The access transistors are controlled by the word line signal WL that defines the operational modes.



**Fig (3.1) 6T SRAM Cell**

**3.2 DESIGN CONSIDERATIONS OF 6T SRAM CELL**

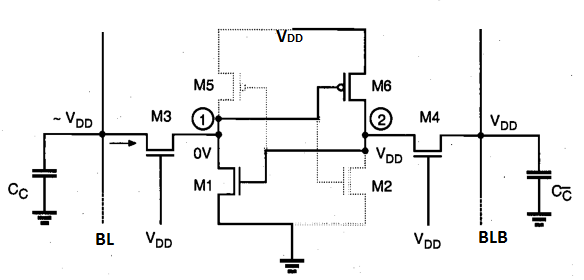
The two basic requirements which dictate the (W/L) ratios in a typical CMOS SRAM cell are:

(a) The data-read operation should not destroy the stored information in the SRAM cell.

(b) The cell should allow modification of the stored information during the data-write phase.

**Read “0" operation:**

Consider the data read operation first, assuming that logic "0" is stored in the cell. The voltage levels in the CMOS SRAM cell at the beginning of the "read" operation are depicted in Fig (3.1). Here, the transistors M2 and M5 are turned OFF, while the transistors M1 and M6 operate in the linear the linear mode. Thus, the internal node voltages are V1=0 and V2=VDD before the cell access (or pass) transistors M3 and M4 are turned on. The active transistors at the beginning of the data-read operation are highlighted in Fig (3.2).



**Fig (3.1) Voltage levels in the SRAM cell at the beginning of the read “0” operation**

After the pass transistors M3 and M4 are turned on by the row selection circuitry, the voltage level of column C will not show any significant variation since no current will flow through M4. On the other half of the cell, however, M3and Ml will conduct a nonzero current and the voltage level of column C will begin to drop slightly. Note that the column capacitance Cc is typically very large; therefore, the amount of decrease in the column voltage is limited to a few hundred millivolts during the read phase. The data- read circuitry is responsible for detecting this small voltage drop and amplifying it as a stored "0" While MI and M3 are slowly discharging the column capacitance, the node voltage V1, will increase from its initial value of 0 V. Especially if the (W/L) ratioof the access transistor M3 is large compared to the (W/L)ratio of M1, the node voltage V1may exceed the threshold voltage of M2 during this process, forcing an unintended change of the stored state.

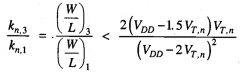
The key design issue for the data-read operation is then to guarantee that the voltage V1, does not exceed the threshold voltage of M2, so that the transistor M2 remains turned off during the read phase, i.e.

C:\Users\nikhila\Desktop\6tr1.PNG**Eq (3.1)**

We can assume that after the access transistors are turned ON the column voltage Vc remains approximately equal to VDD. Hence, M3 operates in saturation while M1 operates in the linearregion.

C:\Users\nikhila\Desktop\eqre2.PNG**Eq (3.2)**

Combining this equation with Eq(4.1) results in:

**Eq (3.3)**

The upper limit of the aspect ratio found above is actually more conservative, since a portion of the drain current of M3 will also be used to charge-up the parasitic node capacitance of node1.

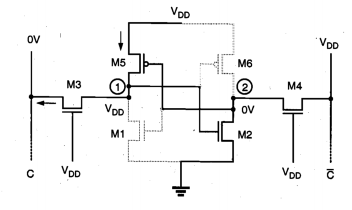
To summarize, the transistor M2 will remain in cut-off during the read "0" operation if Eq (3.3) is satisfied. A symmetrical condition also dictates the aspect ratios of M2 and M4.

Now substituting VDD = 0.5V and VT,n = 0.2V ; standard values in 50nm technology in “Microwind”, we get Eq (3.3) as W3 ≤ 8(W1) and similarly we obtain W4 ≤ 8( W2). From the layout W1 = W2 = 9 lambda, hence we obtain

W3 ≤ 72 lambda and W4 ≤ 72 lambda.

**Write “0” when “1” is stored:**

Now consider the write "0” operation, assuming that a logic "1" is stored in the SRAM cell initially. Fig(3.3) shows the voltage levels in the CMOS SRAM cell at the beginning of the data-write operation. The transistors M1 and M6 are turned OFF, while the transistors M2 and M5 operate in the linear mode. Thus,the internal node voltages are V1 = VDD and V2 = 0 V before the cell access (or pass) transistors M3 and M4 are turned ON.

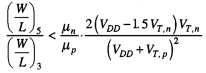


**Fig (3.3) Voltage levels in the SRAM cell at the beginning of the "write"operation.**

The column voltage VC is forced to logic "0" level by the data-write circuitry; thus,we may assume that Vc is approximately equal to 0 V. Once the pass transistors M3 and M4 are turned ON by the row selection circuitry, we expect that the node voltage V2remains below the threshold voltage of Ml, since M2 and M4 are designed according toEq (3.3). Consequently, the voltage level at node (2) would not be sufficient to turn ON Ml. To change the stored information, i.e., to force V1 to 0 V and V2to VDD, thenode voltage V1 must be reduced below the threshold voltage of M2, so that M2 turns OFF first. When V1 = VTn the transistor M3 operates in the linear region while M5 operatesin saturation.

C:\Users\nikhila\Desktop\eq wr1.PNG**Eq (3.4)**

C:\Users\nikhila\Desktop\eqwr2.PNG**Eq (3.5)**

**Eq (3.6)**

To summarize, the transistor M2 will be forced into cut-off mode during the write"0" operation if Eq (3.6) is satisfied. This will guarantee that Ml subsequently turns ON, changing the stored information. Note that a symmetrical condition also dictatesthe aspect ratios of M6 and M4.

Now substituting VDD = 0.5V and VT,n = 0.2V ; standard values in 50nm technology in “Microwind” and taking µn/µp = 2.5 we get Eq (3.6) as

W3 > 2.45 (W5) and similarly we obtain W4>2.45 (W6).

From the layout W5 = W6 = 5 lambda, hence we obtain W3 > 12.25 lambda and W4> 12.25 lambda.

**3.3WORKING**

The cell works in three operating modes.

**Hold state:** When WL=0, both access FETs are OFF and the cell is isolated. This defines the “hold’’ condition and the cell retains the value stored. Both bit and bit-bar lines are precharged to VDD during this state.

**Write operation:**

To perform ‘‘write’’ operation, the word line is brought up to a value of WL=1. This turns ON the access transistors and bit, bit-bar lines get connected to the write circuitry. Value to be written is placed on bit line and its complimentary value is placed on bit-bar line. Precharge remains off during write operation.

When logic ‘1’ is written into the cell, bit line is made high and bitbar line is made low. Referring to the Fig (3.1), access transistors M5 and M6 are turned ON. Node Q is charged to VDD which turns M1 ON and M2 OFF. M1 pulls Q̅node to ground. M3 turns OFF and M4 is turned ON which pulls node Q to VDD.

When logic ‘0’ is written into the cell, bit line is made low and bitbar line is made high. Referring to the Fig (3.1), access transistors M5 and M6 are turned ON. Node Q is pulled down to logic ‘0’ value which turns M2 ON and M1 OFF. M2 pulls Q̅node to VDD. M4 turns OFF and M3 is turned ON which pulls node Q to ground.

**Read operation:**During the read operation word-line is made high, WL=1. This turns on the access transistors and bit, bit-bar lines get connected to read circuitry. Precharge remains OFF during read operation. Bit and bitbar lines act as outputs and are fed into a sense amplifier that determines the stored state.

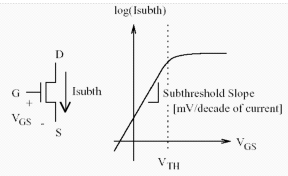
**CHAPTER 4**

**LEAKAGE CURRENTS**

High leakage current in deep-sub micrometer regimes is becominga significant contributor to power dissipation of CMOScircuits as threshold voltage, channel length, and gate oxide thicknessare reduced. Consequently, the identification and modeling of different leakage components is very important for estimationand reduction of leakage power, especially for low-power applications. In this chapter various transistor intrinsic leakagemechanisms are explained.

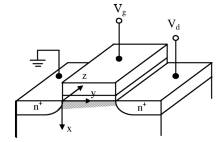
**4.1 SUBTHRESHOLD LEAKAGE**

Sub-threshold or weak inversion conduction current between source and drain in NMOS transistor occurs whengate voltage is below VT. The weak inversion region isseen in Fig (4.1) as the linear region of the curve (semilog plot).



**Fig (4.1) Sub-threshold leakage in NMOS transistor**

In the weak inversion, the minority carrier concentration is small, but not zero. Fig (4.2) shows the variation of minority carrier concentration along the length of the channel for n-channel MOSFET biased in the weak inversion region. Letus consider that the source of the n-channel MOSFET is grounded. Gate voltage, VG<VT and the drain to source voltageVDS > 0.1V. For such weak inversion condition, VDSdrops almostentirely across the reverse-biased substrate-drain pn junction. As a result, the variation of the electrostatic potentialØS atthe semiconductor surface along the channel (the y axis) issmall. Thecomponent of the electric field vectorE(Ey) being equal to, is also small. With both the number ofmobile carriers and the longitudinal electric field small, thedrift component of the sub-threshold drain-to-source currentis negligible. Therefore, unlike the strong inversion regionin which the drift current dominates, the sub-threshold conductionis dominated by the diffusion current. The carriersmove by diffusion along the surface similar to charge transport across the base of bipolar transistors.



**Fig (4.2) variation of minority carrier concentration in the channel of a MOSFET biased in weak inversion region**

The exponentialrelation between driving voltage on the gate and the drain current is a straight line in a semilog plot of IDversus VG. Weak inversion typically dominates modern deviceoff-state leakage due to the low VT. The weak inversioncurrent can be expressed as:

C:\Users\nikhila\Desktop\subthresheq1.PNGC:\Users\nikhila\Desktop\subthresheq1cont.PNG**Eq (4.1)**

**C:\Users\nikhila\Desktop\subthresheq2.PNGEq (4.2)**

WhereVthis the threshold voltage, and*V*T = kT/q is thethermal voltage. Cox is the gate oxide capacitance; µois thezero bias mobility; and m is the sub-threshold swing coefficient(also called body effect coefficient).Wdmis the maximumdepletion layer width, andtoxis the gate oxide thickness.Cdm isthe capacitance of the depletion layer.

**4.2 GATE OXIDE TUNELLING**

Reduction of gate oxide thickness results in an increase inthe field across the oxide. The high electric field coupled withlow oxide thickness results in tunneling of electrons fromsubstrate to gate and also from gate to substrate through the gate oxide, resulting in the gate oxide tunneling current. When a positive bias is applied to the gate, due to the small oxide thickness, which results in a small widthof the potential barrier, the electrons at the strongly invertedsurface can tunnel into or through the SiO2layer and hencegive rise to the gate current. On the other hand, if a negativegate bias is applied, electrons from the poly-silicon cantunnel into or through the oxide layer and give rise to the gate current.

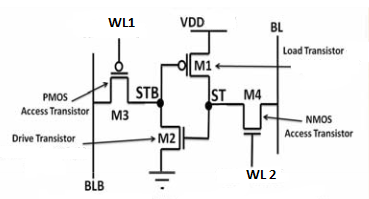
**CHAPTER 5**

**4T SRAM CONCEPTS**

In this chapter the concepts of 6T SRAM cells are discussed. Different operations (read, write and hold) taking place, design considerations and the working in detail is explained for 6T SRAM cell.

**5.1 4T SRAM CELL**

4T SRAM cell consists of load and driver transistors and two access transistors. Unlike in 6T cell where both the access transistors are NFETs, in 4T cell one of the access transistors is a PFET. Two different word lines WL1 and WL2 are used for controlling PFET and NFET access transistors respectively. Circuit diagram of 4T SRAM cell is shown in the Fig (5.1).



**Fig (5.1) 4T SRAM Cell**

**5.2 DESIGN CONSIDERATIONS**

**Conventional model 1:** In this model, read and write operations are implemented as given in the reference paper1. Read operation is done using BLB and write operation is done using BL.

**Problems encountered:**

* In this case, precharge is always kept ON. Hence even during read operation BLB is kept at VDD. Considering the case where “1” is stored in the cell, STB node is kept at GND. While reading this value BLB line which is at VDD, is now pulled to GND. This results in high current flow which in turn results in high power consumption.
* Consider two cells in the same column. Assuming “1” is stored in the first cell and “0” is stored in the second cell. During the read operation, while reading “0” from BLB line, from the first cell, the BLB line which is kept at VDD is pulled down to GND, which alters the value “1” at STB node of second cell. This happens because, “1” at STB node is retained due to leakage currents of PMOS access transistor connecting BLB (at VDD) and STB node.

**Values used:**VDD= 0.5V VSS = 0V VNWELL = 0.5V

VTN = 0.2V VTP = 0.2V

VWLN = 0.5 and 0V (ON and OFF)

VWLP = 0 and 0.5V (ON and OFF)

**Model 2:** To overcome the above problems, both read and write operations are implemented using BL. Precharge on BLB is always kept ON, hence BLB line is always at VDD. During read and write operation, pre-charge on BL is turned OFF.

As per the reference paper 1, leakage currents of access transistors have to be more than that of load and drive transistors. To achieve this VT of access transistors is kept greater than that of load and drive transistors. Since threshold voltages cannot be changed for each transistor in Microwind tool, an alternative approach is used. Referring to Eq(4.1) VSG of PMOS access transistor is increased in order to increase the leakage current.

**Problem encountered:** When “1” is stored in the cell, STB node is at “0”. Since load transistor is ON it pulls STB node to GND. PMOS access transistor is OFF but as we have increased its VSG value the leakage currents have increased which try to pull STB node to “1”. This results in increased power consumption.

**Values used:**VDD= 0.5V VSS = 0V VNWELL = 0.5V

VTN = 0.2V VTP = 0.2V

VWLN = 0.5 and 0V (ON and OFF)

VWLP = 0 and 0.3V (ON and OFF)

**Improved model 3:** In this model, to decrease the power consumption problem we have reduced VDD to 0.45V and increased VSS to 0.05V. Due to this bit-line swing reduces resulting in less power consumption. Write and read operations are performed using BL, as mentioned in model2.

Sub threshold leakage currents come into picture when VGS < VT. When “0” is stored in the cell, load and drive transistors are OFF. In model2, when load transistor is in OFF state, its VGS value is equal to 0V which results in flow of IOFF current. In this model since we have increased VSS value to 0.05V, when load transistor is in OFF state, its VGS value becomes -0.05V which reduces the sub threshold leakage current contribution of load transistor. Here, VDD value is decreased to 0.45V so when drive transistor in in OFF state its VSG value decreases, which reduces the sub threshold leakage current contribution of drive transistor. As VGS and VSG values of load and drive transistors have reduced, gate tunneling effect reduces. All these conditions are desirable as per Eq(2.1) and Eq(2.2).

**Values used:**VDD= 0.45V VSS = 0.05V

VTN = 0.2V VTP = 0.2V

VNWELL, ACCESS = 0.6V

VNWELL, DRIVE = 0.9V

VWLN = 0.5 and 0V (ON and OFF)

VWLP = 0 and 0.3V (ON and OFF)

From the layout,

WNMOS,ACCESS= 4 lamda WLOAD = 4 lamda

WPMOS,ACCESS= 4 lamda WDRIVE = 4 lamda

**5.3 WORKING**

The cell works in three operating modes.

* **Hold state:** When WL1= 0V and WL2= 0.3V, both access transistors are OFF and the cell is isolated. This defines the “hold’’ condition and the cell retains the value stored. Different precharge voltages are used for bit and bitbar lines. Bit line is precharged to VDD and bitbar line is precharged to ground during this state.
* **Write operation:**

When a write operation is issued the memory cell will go through the following steps.

**1) Bit-line driving:** For a write, data is placed on BL, and then WL2 asserted to VDD, but voltages on word-line1 and BLB maintained at idle mode ( WL1= 0.3V and VBLB=VDD). Only precharge on bit-line is turned OFF.

**2) Cell flipping:** This step includes two states as follows.

**(a) When data is logic ‘0’:** In this state, ST node is pulled down to GND by NMOS access transistor, and therefore the drive transistor will be OFF. STB node will be floated and then pulled up to voltage of BLB (VDD) by leakage current of PMOS access transistor and thus load transistor will be OFF.

**b) When data is logic ‘1’:** In this state, ST nodepulled up to VDD-Vtnby NMOS access transistor, and therefore the drive transistor will be ON , and STB node will be pulled down to GND, thus load transistor will be ON which further pulls up ST node to VDD.

**3) Idle mode:** At the end of write operation, cell will go to idle mode and WL2 and BL are asserted to GND.

* **Read operation:**

When a read operation is issued the memory cell will go through the following steps.

**1) Bit-line Pre-charging:** For a read, BL pre-charged to GNDand then floated. Since, in idle mode BL maintained at GND, this step doesn’t include any dynamic energy consumption.

**2) Word-line activation:** In this step WL2 asserted to VDD and two states can be considered.

**(a) Voltage of ST node is high:** When voltage of ST node is high, the voltage of BL pulled up to high voltage by NMOS access transistor. We refer to this voltage of BL as VBL-HIGH.

**(b) Voltage of ST node is low:** When voltage of ST node is low, the voltage of BL and ST node equalized. Since in this state, there is very small different between BL and ST node, power consumption is very small.

**3) Sensing:**After WL2 is deactivated, the sense amplifier is enabled to read data on BL.

**4) Idle mode:** At the end of read operation, cell will go to idle mode and WL2 and BL are asserted to GNDrespectively.

**CHAPTER 6**

**CACHE MEMORY**

Cache memory is random access memory that a computer microprocessor can access more quickly than it can access regular RAM. It is a high speed memory in the CPU that is used for faster access to data. It provides the processor with the most frequently requested data and increases the performance. In this chapter basic components of a cache memory array and concepts of 6T and 4T cache memories are discussed.

**6.1 COMPONENTS OF SRAM ARRAY**

SRAM arrays are created by replicating the basic storage cell and adding the necessary peripheral circuitry. The basic architecture of SRAM memory array comprises of the following components.

**Dual core:**It consists of two core regions of storage cells that share central word line circuits. Memory cells are tiled to produce the left and right core regions. Width of core region is multiple of word size.

**Row decoder:**The output of a centrally located decoder provides wordline signals to the storage cells. The address word specifies a particular row, which is then driven high. The access transistors of the selected row cells are turned on, permitting the read or write operations to take place. The location of the circuitry allows a single decoder to be used for both left and right memory cores.

**Column decoder:** To choose a particular word in a row, a group of column decoder circuits is added, to select a particular set of columns in the matrix. Mux sections are driven by column decoder to steer the selected bit, bit-bar groups.

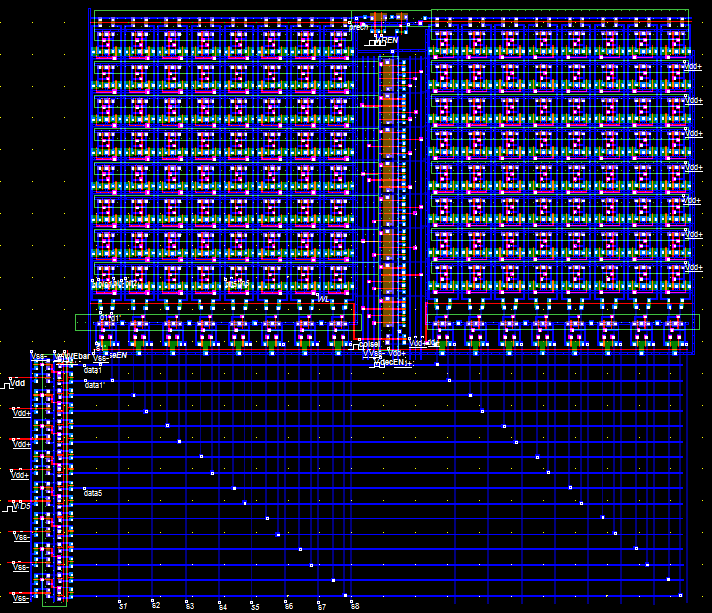
**Precharge circuit:** This circuit at the top is controlled by a clock signal that is used to synchronize the operation and data flow. During hold state these circuits are active and maintain constant voltages on bit and bit-bar lines. During read and write operations precharge circuitry is deactivated.

**Read/ Write circuitry:** It performs the following functions

* Directing the data flow into the array during a write operation or out of array during a read operation.
* Connecting the read and write circuits to the bit and bit-bar lines of every column.
* Providing amplifiers to detect and amplify the outputs during a read operation.

**6.2 6T SRAM CACHE**

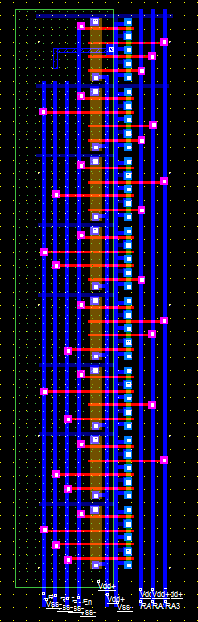
Conventionally, cache memory is implemented using 6T SRAM cells. Here, layout of cache memory of size 8x16 is implemented using Microwind tool. Storage cell is made up of 6 transistors. A bi-stable latch circuit is used. Dual core is made by replicating 6T SRAM cells. Layout of 6T SRAM cache is shown in Fig (6.1).



**Fig (6.1) Layout of 6T SRAM cache**

**Pre-charge circuitry:**A PMOS transistor is used which is activated during hold state which pulls bit and bit-bar lines to VDD. It is deactivated during read and write operation.

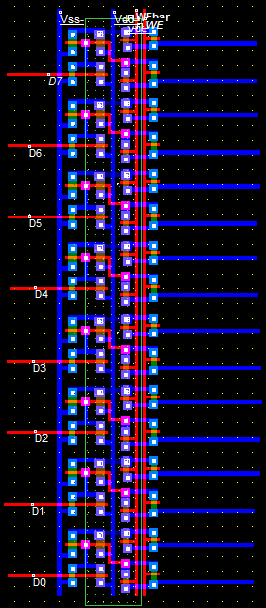
**Row decoder:** Since we have 8 rows, to activate them we require 3 address bits. These three address bits along with enable signal are ANDed to activate the respective wordline that activates the corresponding row. Layout of row decoder circuitry is shown in Fig (6.2)



**Fig (6.2) Layout of Row decoder circuit**

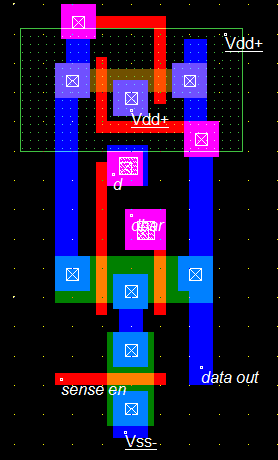
**Column decoder:** Since we have only two sets of words, we require one address bit to activate them. This address bit acts as column select signal. Transmission gates are used for this purpose as they provide strong “1” and “0” values.

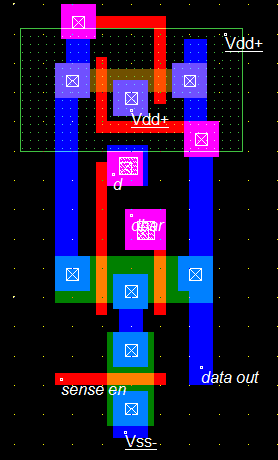
**Write circuitry:**The input bits are inverted and buffered to provide complimentary pairs (d and dbar). When write enable control bit has a value WE=1, transmission gates act as closed switches connecting the data pairs to bit and bit-bar lines. Every bit pair is fed to the appropriate locations. Layout of write circuitry is shown in Fig (6.3).



**Fig (6.3) Layout of write circuit**

**Read circuitry:**It consists of a sense amplifier which detects the values on bit and bit-bar lines, amplifies the value and transfers it onto the output line. When read-enable and sense-enable signals are high sense amplifier read is activated and the value is read. Layout of sense amplifier circuitry is shown in Fig (6.3).

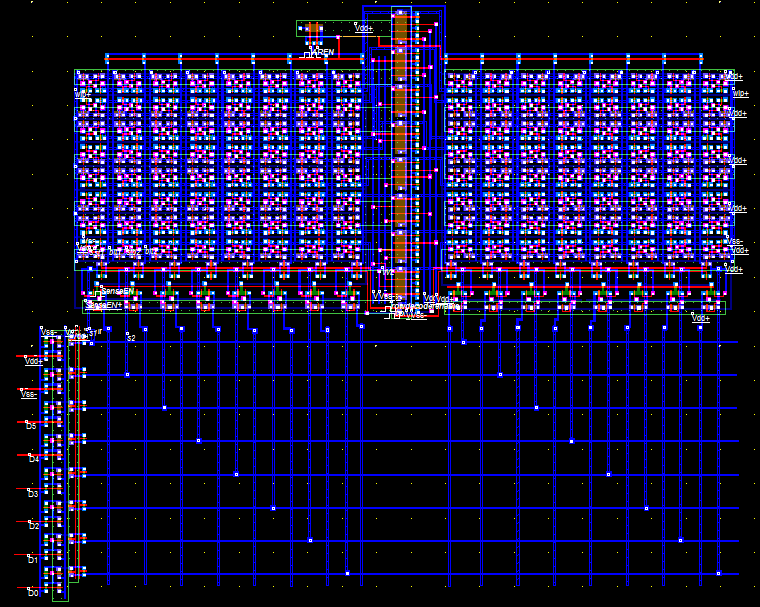




**Fig (6.4) Layout of sense amplifier**

**6.3 4T SRAM CACHE**

Cache memory is implemented using 4T SRAM cells. Here, layout of cache memory of size 8x16 is implemented using Microwind tool. Storage cell is made up of 4 transistors. Dual core is made by replicating 4T SRAM cells.Layout of 4T SRAM cache is shown in Fig (6.2).



**Fig (6.2) Layout of 4T SRAM cache**

**Pre-charge circuitry:**A NMOS transistor and PMOS transistor are used which are activated during hold state to pull bit and bit-bar lines to GND and VDD respectively. They are deactivated during read and write operation.

**Row decoder:** Since we have 8 rows, to activate them we require 3 address bits. These three address bits along with enable signal are ANDed to activate the respective word-line that activates the corresponding row.

**Column decoder:** Since we have only two sets of words, we require one address bit to activate them. This address bit acts as column select signal. Transmission gates are used for this purpose as they provide strong “1” and “0” values.

**Write circuitry:**The input bit (d) is buffered onto the bit-line. When write enable control bit has a value WE=1, transmission gates act as closed switches connecting the data to bit-line. Every bit is fed to the appropriate location.

**Read circuitry:** It consists of a sense amplifierwhich detects the values on bit-line, amplifies the value and transfers it onto the output line. When read enable and sense enable signals are high sense amplifier is activated and the value is read.A current mirror along with differential amplifier circuit is used. A reference voltage VREF is used. Any voltage above this value is detected as logic “1” and any voltage value below this value is detected as logic “0”.

**CHAPTER 7**

**DIVIDED BITLINE**

Divided bit-line approach is used for reducing access time and active power consumption by reducing bit-line capacitance. Two or more SRAM cells can be combined together to divide the bit line into several bit lines. These sub bit lines are again connected to a global bit line through a TG or a NFET.

There will be a increase in area about 5-10% ,because of precharge and pass transistor for every row and AND gate used for each group wordling consisting of M input(M=no. of cells in a group) and gate and a not gate for a group precharge signal.

Let us take NFET drain capacitence = Cd, metal capacitance =Cm, metal resistance=Rm,drain resistance=Rd

Toatal no of cells=N

From the tool we can observe that Cd=Cm=.08fF , Rd=310 ohm and Rm=3ohm

From the above observations we can take Cd=Cm=C and Rd=100Rm or Rm=0.01Rd

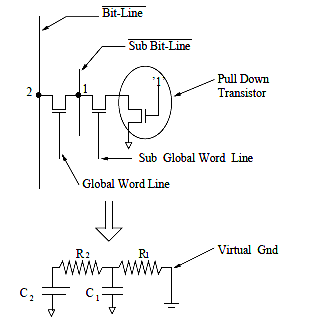
In a single bitline approach total delay is due to all drain and metal connect.this can be written as

Td=NCmRm + NCdRd  **Eq (7.1)**

For N=128

Td=(128C x 0.01R) + (128xCxR) **Eq (7.2)**

Now consider a divided bitline with N rows grouped with 8 cells each. Here delay is due to global bit line and one sub bitline.



**Fig (7.1) RC model**

From the above figure total global bit-line delay can be written as sum of delays due to N metal pieces and N/m drain capacitance of pass transistors(N/m because N rows of m bits each will have N/m groups and n/m pass transistors).

T2=NCmRm + (N/m)CdRd **Eq (7.3)**

= NCx0.01R + (N/m)CR **Eq (7.4)**

Sub bit line delay can be written as sum of delays of (m+2)drains (m drains in a m cell group plus one pass transistor drain plus one precharge drain ) and m metal lines which ca be written as

T1= (MxCmxRm) + (M+2)xRdxCd **Eq (7.5)**

= MxCx0.01R + (M+2)RC **`Eq (7.6)**

Now total delay can be considered as t1+t2

Td = MxCx0.01R + (m+2)RC + NCx0.01R + (N/m)RC **Eq(7.7)**

Now optimize the value of m for which the delay is minimum. To find the m for minimum value of delay we differentiate td with respect to m and equate it to zero.

Differentiating Eq (7.7) w.r.t to m

=

= + + +

=0.01RC + RC - (1/m2)NRC **Eq (7.8)**

Now equating Eq (7.8) to zero, we get

0.01RC + RC - (1/m2)NRC = 0

1.01RC = (1/m2)NRC

1.01= (N/m2)

m=

For N=128

m= = 11.2575 ~ 12

Substituting m=12 and N=128 in Eq (7.7)

Td = 12Cx0.01R + (12+2)RC + 128 Cx0.01R + (128/12)CR

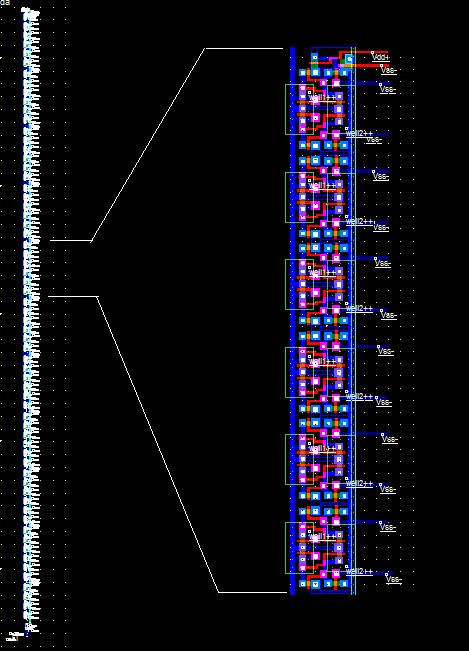
= 0.12CR +14CR + 1.28CR + 10.66667CR **Eq (7.9)**

Now dividing Eq (7.9) with Eq(7.1), we get

= 0.2

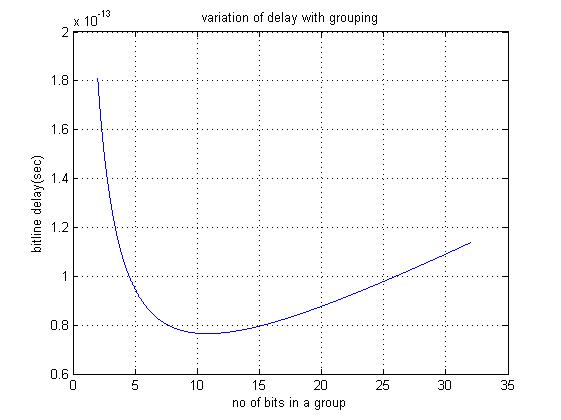
Delay due to use of divided bit line gets reduced by 80% theoretically.

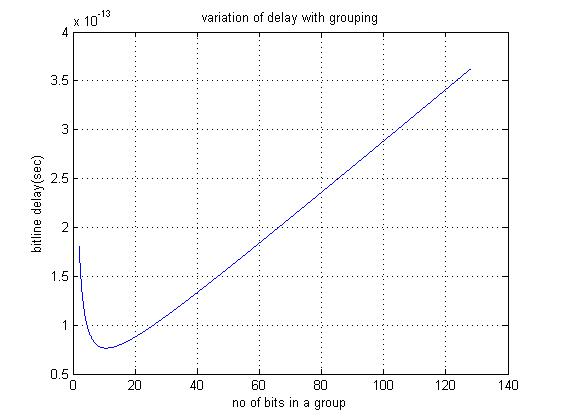
Layout of 128 cells in a column on which divided bit-line technique is applied grouping 8 cells, is shown in Fig (7.2).



**Fig (7.2) Layout of 128 cell column with a subgroup of M=8**

Variation of delay with change of grouping



****

**CHAPTER 8**

**SIMULATIONS AND RESULTS**

**CHAPTER 9**

**CONCLUSION AND FUTURE SCOPE**