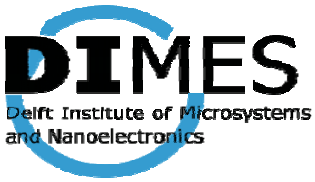


SRAM in Three Dimensional Integrated Circuits

Negin Golshani

The Laboratory of **Electronic Components, Technology and Materials** (**ECTM**)

June 2009

Table of Contents

Chapter 1

Introduction

1-1) Introduction to 3DIC and motivation of this thesis

In most of the electronics and communication devices such as mobile, video phone and handheld video games low power and high density SRAM (Static Random Access Memory) is a favor. On the other hand, integration of many functions such as digital, memory, RF and analog circuits is necessary in near future. Scaling is one of the solutions to increase density of memories and functionality of integrated circuits. However, the increase of leakage current, process complexity and process variation of parameters limit scaling.

This limitations force us to think about new dimension in integrated circuits. Three dimensional integrated circuits can solve some of the problems. They can give us low power, high density memories and high functionality circuits in same area of planar ICs. Different technologies can be merged in different layers of 3DIC to finally make a high performance system.

In this thesis we realize SRAM cells in 3DIC to increase the capacity and performance of them. In chapter 2 we will introduce different memories and particular case SRAM. The principle of operation and design metrics are discussed in this chapter. Then in chapter 3 we talk about 3DIC and its advantages and disadvantages. Heat generation is main issue in 3DIC. New 3DIC fabrication technology called µ-Czochralski Process is introduced. Next in chapter 4 we design SRAM cells using analytic approach and we confirm the design by circuit simulation tools. Different SRAM cells and sense amplifier and output buffers are designed in this chapter. To fabricate design circuits we need layout for SRAM cells. In chapter 5 we extensively look to the design rules for SRAM circuits in one layer and two layers of silicon. Using double gate and H-Gate transistors to increase the performance of SRAM cells are discussed in this chapter. Then in chapter 6 we show fabrication process flow of one layer and two layers single grain silicon devices. Finally fabricated circuits are characterized electrically and results are reported in chapter 7.

Chapter 2

CMOS SRAM Circuits

Content:

2-1) Introduction to data storage devices:

In general, data storage devices most frequently are classified by fabrication technology of them. Fig. 2.1 shows four examples of storage technologies. Semiconductor, Magnetic, Optical and Biological technologies are some types of memories that are being used [1].

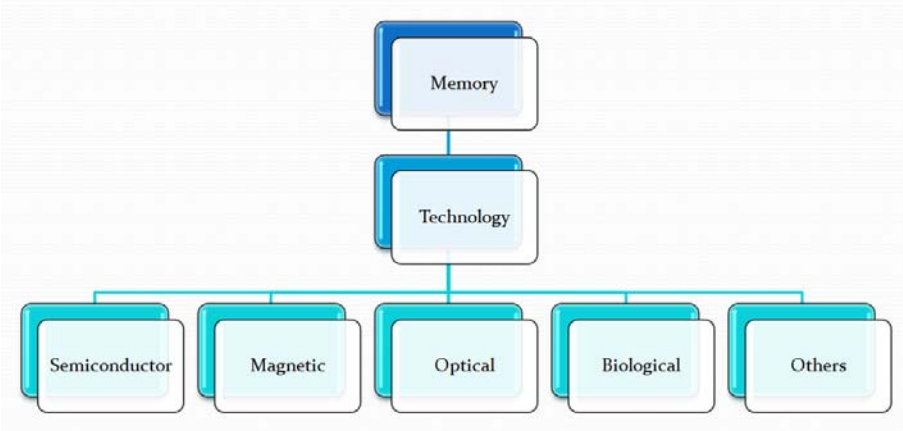


Fig.2.1: Data storage technologies

From the variety of technologies which can applied to create data storage devices, the semiconductor integrated circuit technology, and within that, the CMOS technology (Fig.2.2) has become as the dominant technology in fabrication of system-internal memories (mainframe, cache, buffer, scratch-pad, etc.). In CMOS technology it is possible to realize memories in bulk and SOI (silicon on insulator) substrates [1].

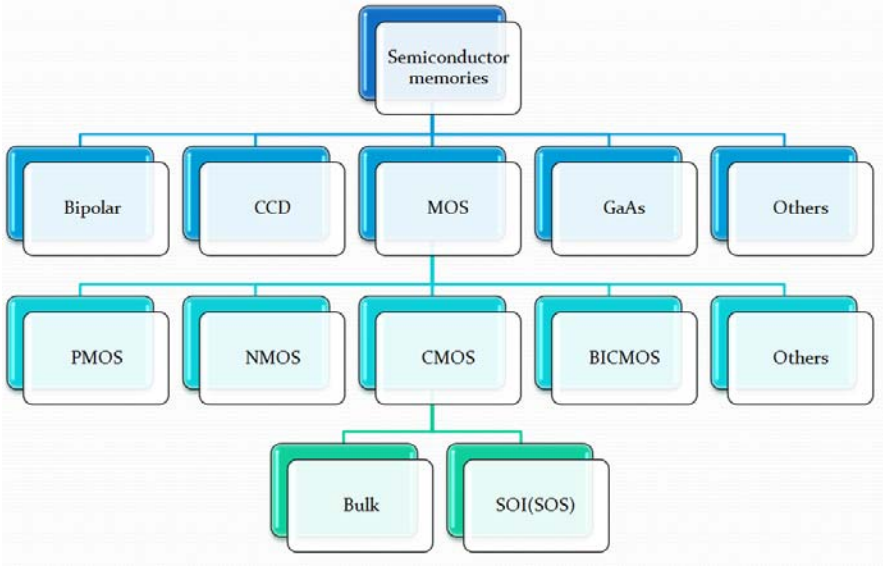


Fig.2.2: Semiconductor Memories

Operational speed versus memory-capacity at a certain state of the industrial development, (Figure 2.3) is of primary importance in choosing memories to a specific system application. It can be seen that CMOS static memories are faster than dynamic memories. Magnetic discs that have highest capacities are widely used in hard discs of computers. Application areas of the

diverse CMOS dynamic, CMOS static, magnetic disk, magnetic tape, bipolar, gallium-arsenide, and other memory devices alter rapidly [1].

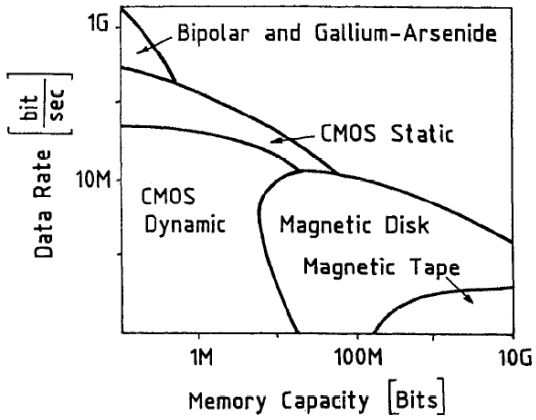


Fig.2.3: Data-rate versus memory-capacity diagram indicating application areas [1].

Memories based on CMOS are used in a much greater quantity than all the other types of semiconductor integrated circuits, and appear in a wide variety of circuit organizations.
As Fig.2.4 shows, most frequently, CMOS memories are categorized by the operation of the storage cells into four categories [1, 2]:

- Dynamic RAMs (DRAMs)
- Static RAMs (SRAMs)
- Fixed program or mask-programmed read-only memories (ROMs)
- User-programmable read-only memories (PROMS).

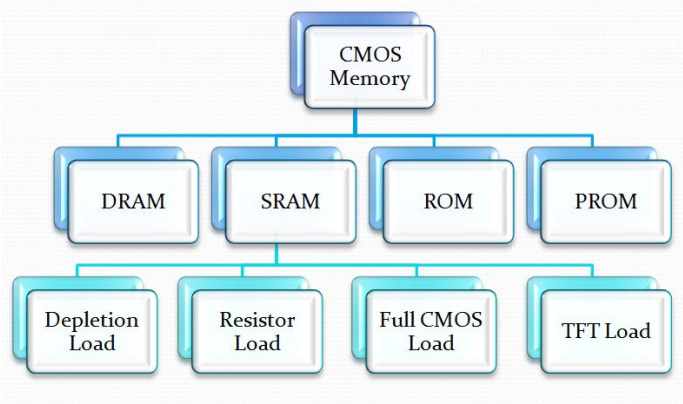


Fig.2.4: Types of CMOS memories

2-2) Static Random Access Memories (SRAMs):

Random access memories which retain their data content as long as electric power is supplied to the memory device, and do not need any rewrite or refresh operation, are called static random access memories or SRAMs.
As shown in Fig. 2.4 depending on load type of SRAM cell, there are four representative MOS SRAM cells [2]:

- Depletion Load

- Resistor Load
- TFT (Thin-Film Transistor) Load
- Full CMOS Load

The table 2.1 summarizes the advantages and disadvantages of these cells. Depletion load is the oldest structure and uses nMOS depletion mode transistor as a load. Resistor load uses high resistance polysilicon resistor as a load. Since this resistor is made over the nMOS drivers in cell, this type has advantage of very small cell size hence high density memory. However, this cell has more standby current. To overcome this problem, TFT load is used. In this case both advantages of high density and less standby current can be achieved. But TFTs are fabricated using polysilicon and can create stability problems especially at low voltages [2]. Using single grain TFTs we will have high density cells, low standby current and high cell stability.
The full CMOS load uses a bulk pMOS transistor as a load and it is compatible with standard CMOS process. However, area is more than resistor load. Standby current is lowest in CMOS cells [2].

Table 2.1: Comparing different loads for SRAM cells [2]

	Depletion Load	Resistor Load	Full CMOS	TFT Load
Density	Medium	High	Low	High
Standby current	High	Medium	Low	Low
Cell stability	High	Low	High	Medium

Fig. 2.5 shows circuit schematics of resistor load 4T structure [3]. The main advantage of static 4T cells with polysilicon resistor load (R_L) is the approximately 30% smaller area as compared to 6T CMOS SRAM cells. However, it is sensitive to noise and soft errors because of high resistivity of resistors (in the range of Giga ohm) and also it is not fast [4].

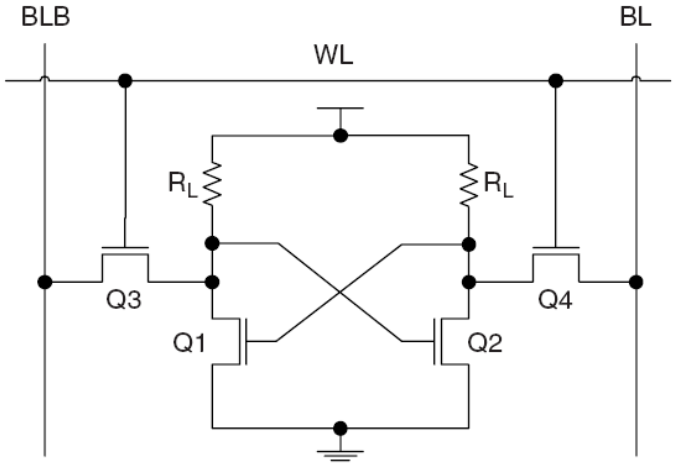


Fig.2.5: Four transistor SRAM cell with polysilicon resistor load [3]

Poly two is used to fabricate resistor load on top of driver nMOS transistors as shown in Fig.2.6. The P^+ regions in the poly 2 are for isolation of resistors. As the technology scaled into sub-micron regime (beyond 0.8µm technology generation), the scalability of a resistor SRAM cell

became an issue [3]. The polysilicon resistor in the RL cell could not be scaled as rapidly as the cell's transistors.

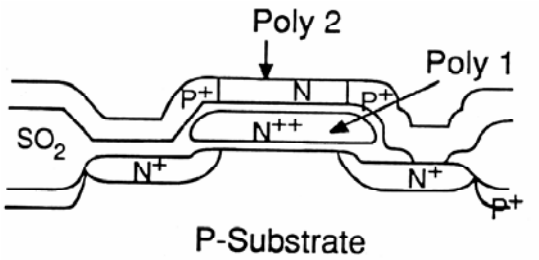


Fig.2.6: Using Poly two to make resistor on top of drive nMOS transistor in 4T SRAM cell

TFT load could solve the mentioned problems. Fig.2.7 shows TFT SRAM cell. In the conventional TFT load cell, that has pMOS TFT, the channel of transistor is poly and has high S factor, resulting high threshold voltage. Thus this cell can not operate below 3V. Also the ratio of on-current to off-current is low in poly pMOS transistors [5]. Process and cell size are closer to a 4T cell technology than a 6T cell technology because the area of the TFT transistors is above the NMOS transistors [9].

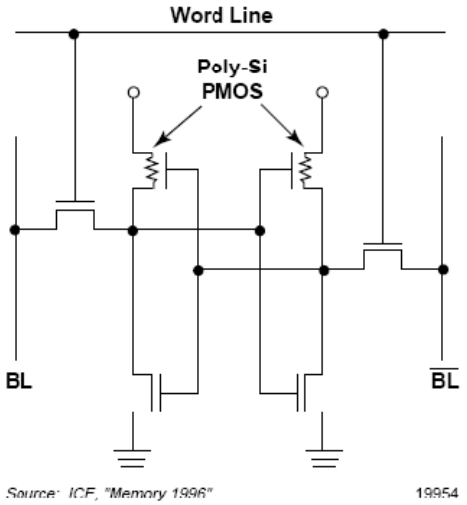


Fig.2.7: SRAM cell using TFT loads [5,9]

On the other hand, stacked-TFT SRAMs cells [5,6,7,8] are promising for some applications where we need tiny cells such as mobile applications. Using an almost single crystal TFT is one solution [6]. However, in these process the size of grains are small and the drain current is still two thirds that of the bulk, and the S-factor of pMOS TFT is as large as 140mV/decade [6]. Thus, it seems that this type of TFT is not suitable to make a low-voltage, high-speed embedded SRAM. Nevertheless, it has the great advantage of having high density for stand-alone SRAMs. Fig.2.8 shows SRAM cell that uses stacked TFT transistors.

A conventional 6T SRAM cell uses $84\times F^2$ (F: feature size) area while using one stacked TFT layer for loads it reduced to $45\times F^2$ [7]. In this case n-well has not been used. It was further reduced to $25\times F^2$ by using two stacking layer of load pMOS TFTs and nMOS access transistors over bulk driver nMOS transistors [8] as shown in Fig. 2.8. In this paper the cell size was comparable to DRAM cells.

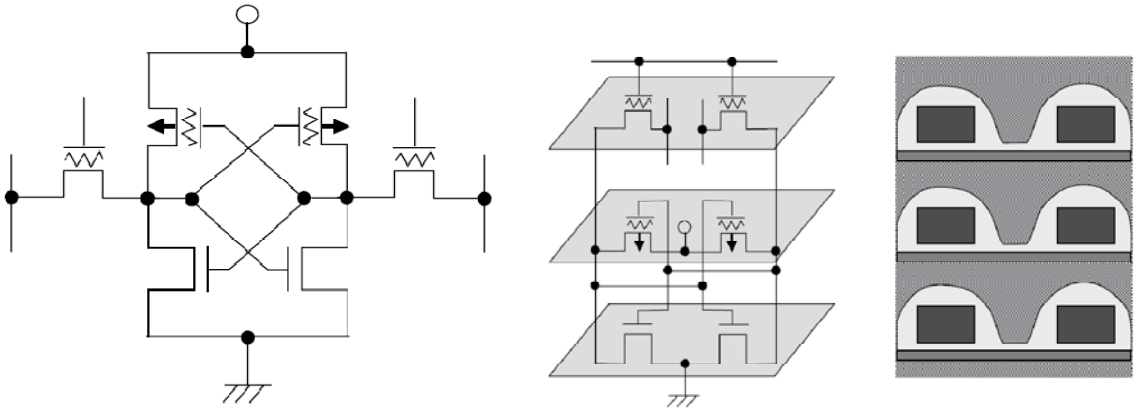


Fig.2.8: SRAM cell using stacked TFTs for high density applications [5,6,7,8]

Using 80nm SRAM processes and with 1.3V supply voltage, die area of 28.5mm^2, 49.2ns access time, 64Mb chip capacity and 0.288μm^2 or 45×F^2 cell size parameters are achieved with one layer TFT stacking [7]. With two layer TFT stacking and 1.8V, die area of 61.1mm^2, 144MHz, 256Mb chip capacity and 0.16μm^2 or 25×F^2 has been reported [8]. For both SRAM cells, the voltages for word-line and cell power supply were selectively increased to compensate for the limited current driving capability of the TFTs.

Figure 2.9 shows a cross section drawing of the TFT cell. The TFT technology requires the deposition of two more films and at least three more photolithography steps that makes it complex for high capacity SRAMs [9].

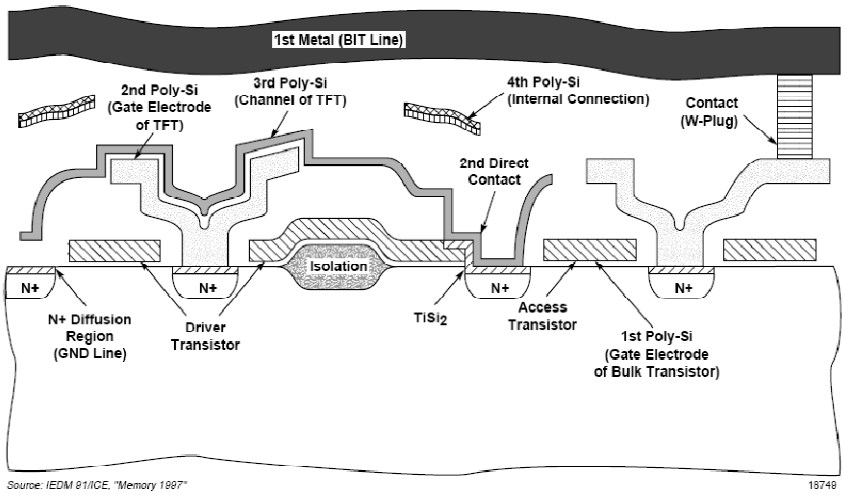


Fig.2.9: Cross section of a TFT SRAM cell

The other type of SRAM is CMOS SRAM that widely is used in commercial products. CMOS SRAMs feature very fast write and read operations and can be designed to have extremely low standby power consumption and can operate in radiation hardened and other severe environments. We will discuss about it with more details.

2-3) SRAM Block Structure

Figure 2.10 shows an example of the basic SRAM block structure. A row decoder is used, by appropriate timing block signals, to decodes X row address bits and selects one of the word lines WL_0 to WL_(N-1) which N is the 2^X. The SRAM core consists of a number of arrays of N×M, where N is the number of rows and M is the number of bits. If an SRAM core is organized as a number of arrays in a page manner, an additional Z-decoder is needed to select the accessed page. Figure 2.5 shows an example of an SRAM with four pages of N×M arrays with the corresponding I/O blocks [3].

SRAMs are called based on bit or word. In a bit-oriented SRAM, each address accesses a single bit, whereas in a word-oriented memory, each address addresses a word of *n* bits (where the popular values of *n* include 8, 16, 32 or 64). Column decoders or column multiplexers (MUXs) (YMUXs) addressed by Y address bits allow sharing of a single sense amplifier among 2, 4 or more columns. The majority of modern SRAMs are self-timed, i.e. all the internal timing is generated by the timing block within an SRAM instance. An additional Chip Select (CS) signal, introducing an extra decoding hierarchy level, is often provided in multi-SRAM chip architectures [3].

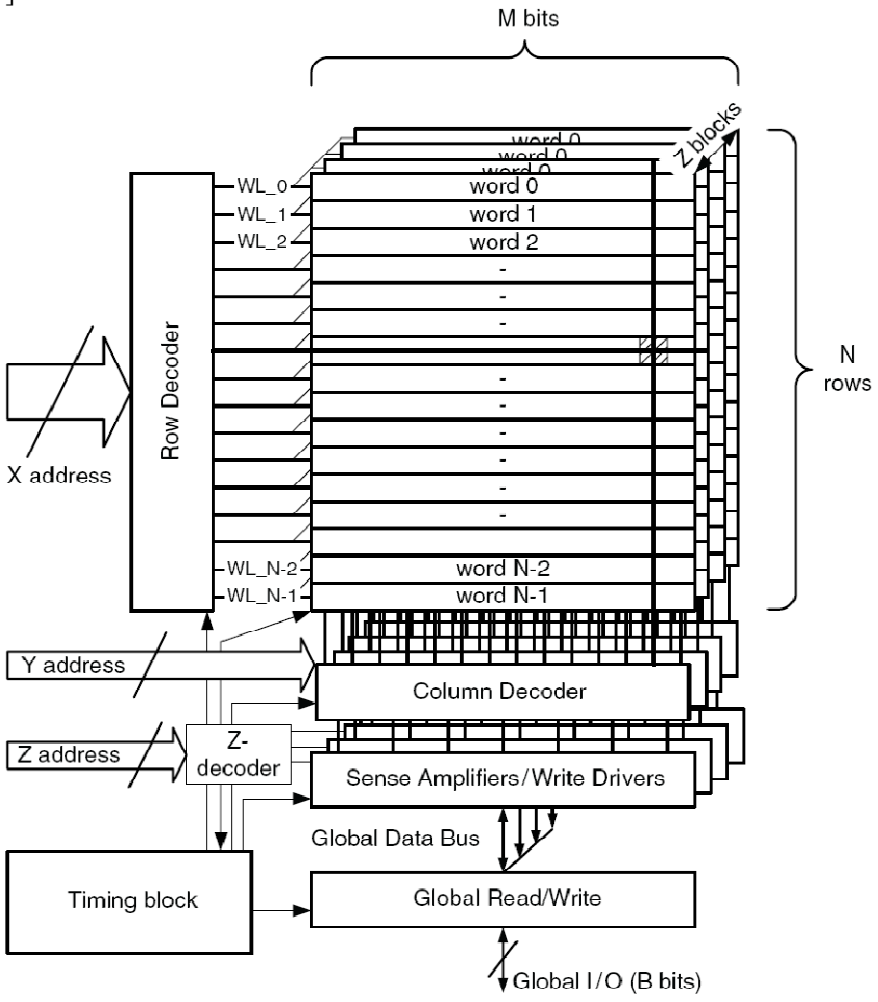


Fig.2.10: SRAM block diagram [3]

In SRAM block diagram, cells are the key component storage for binary information [3].

2-4) Six transistor (6T) CMOS SRAM cell

A 6T CMOS SRAM cell is the most popular SRAM cell due to its superior robustness, low power and low voltage operation, short access time, high frequency data rate, radiation hardness, operation in space, high temperature and noisy environments [1].

The structure of six-transistor (6T) CMOS SRAM cell is shown in Figure 2.11. It consists of six transistors. Four transistors ($Q1-Q4$) comprise cross-coupled CMOS inverters that use positive feedback to store a value and two nMOS transistors $Q5$ and $Q6$ provide read and write access to the cell and provide cell isolation during the not-accessed state. Q1 and Q2 are called drivers (or pull down) and Q3 and Q4 are pull up transistors. The function of the pull-up transistors is only to maintain the high level on the "1" storage node and prevent its discharge by the off-state leakage current of the driver transistor during data retention (hold) and to provide the low-to-high transition during overwriting. Q5 and Q6 are access transistors [3].

The positive feedback between two complementary inverters provides a stable data storage, and makes high speed write and read operations. Upon the activation of the word line, the access transistors connect the two internal nodes of the cell to the bitline (BL) and the complementary bit line (BLB).

An SRAM cell must be designed such that it provides a non-destructive read operation and a reliable write operation. These two requirements impose contradicting requirements on SRAM cell transistor sizing. SRAM cell transistors ratios must be observed for successful read and write operations [10].

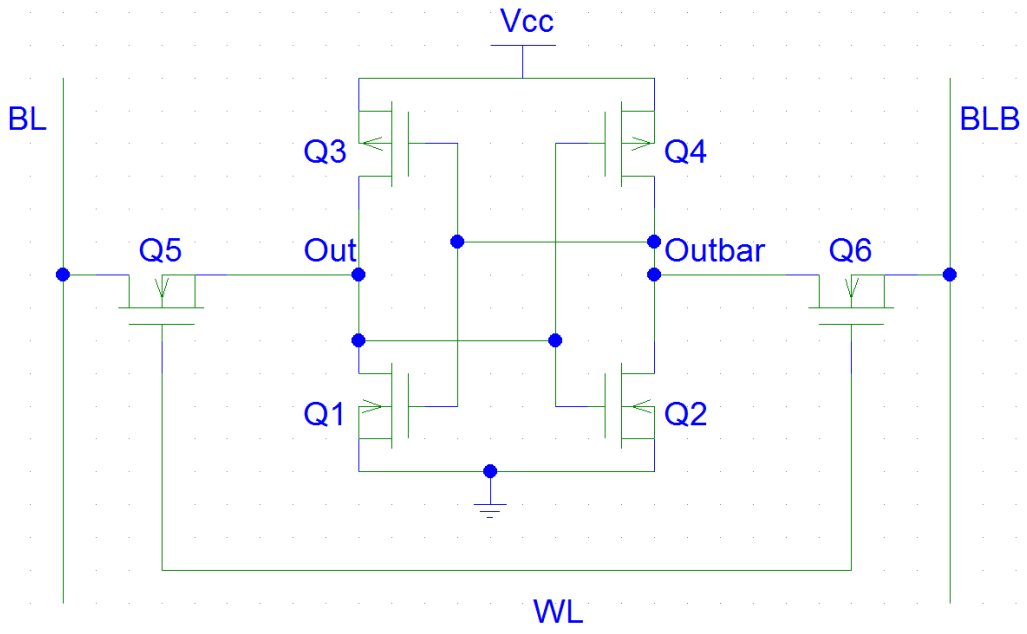


Fig.2.11: Six transistor (6T) CMOS SRAM cell

In general, the cell design must have a balance between cell area, robustness, speed, leakage and yield. Cell size minimization is one of the most important design objectives. A smaller cell allows the number of bits per unit area to be increased and thus, decreases cost per bit. Reduced cell area can indirectly improve the speed and power consumption due to the reduction of the associated cell capacitances. Smaller cells result in a smaller array area and hence smaller bit line and word line capacitances, which in turn helps to improve the access speed performance.

Reducing the transistor dimensions is one of the effective ways to achieve a smaller cell area. However, the transistor dimensions cannot be reduced without compromising the other parameters. For instance, smaller transistors can influence on the cell stability.
To achieve minimum area without losing stability, performance and radiation hardening, three dimensional integrated circuits can be used. In the next chapter we will discuss about this technology with more details.

2-5) Operation of 6T CMOS SRAM cell

A memory cell must be designed very carefully for reliable operation and minimum area. Three working modes can be distinguished among which two are active, the read and write operations, and one is passive, the retention mode. The key parameters for performance characterization in the three modes are summarized below [1,2,3,5,10,12,27]:

- Retention (Hold)
  - Adequate stability in retention to preserve the stored data
  - Minimum static current ($I_{leakage}$) for low standby power applications
- Read
  - Adequate stability in read mode for reliable operation
  - Maximum read cell current to achieve short read access time
- Write
  - Adequate cell write margin for reliable operation
  - Adequate cell write disturb for reliable operation
  - Maximum write cell current to achieve short write access time
- Area
  - Minimum cell size for highest density
- Minimum Bitline (C_{BL}) and World line (C_{WL}) load capacitors to achieve high speed and low operating power

2-5-1) Read Operation

The read operation is used to transfer the contents of the storage cell to the BL and BLB. Before starting to Read operation, both BL and BLB are precharged to Vcc. The read operation is started by enabling the word line (WL) and connecting the precharged bit lines, BL and BLB, to the internal nodes of the cell. A high value on the BL does not change the value in the cell, so the cell will pulls one of the lines low. The important point is that data read operation should not destroy the stored information. The capacitance of bitlines is large enough to act like a voltage source.
Assume that a logic 0 is stored in the cell (Out=0 and Outbar=1) (see Fig.2.12). Upon read access, the bit line voltage V_{BLB} remains at the precharge level. The bit line voltage V_{BL} is discharged through transistors Q1 and Q5 connected in series. Effectively, transistors Q1 and Q5 form a voltage divider and their output is connected to the input of inverter Q2−Q4 (Figure 2.12). Sizing of Q1 and Q5 should ensure that inverter Q2−Q4 does not switch causing a destructive

read. In other words, the voltage of "out" point (0+ΔV) should be less than the switching threshold of inverter Q2−Q4 plus some safety margin or Noise Margin.
In this case Q1 and Q4 are in linear region and Q2, Q3 are off. When WL (world line) is low Q5 and Q6 are off. Upon activating WL, Q5 goes to saturation and Q6 goes to linear modes. Then BL voltage decreases and BLB voltage increases slowly. During data reading, Q2 must remain turn off. It means the increased voltage of "Out" point (ΔV) should be less than V_{th2}. Also currents of Q1 and Q5 are same. Then the condition of non-destructive reading can be obtained:

$$\Delta V_{max} \leq V_{th2} \; and \; I_{Q5}(sat.) = I_{Q1}(lin.)$$

Now we define cell ratio (CR or β) as a ratio of driver $(\frac{W}{L})$ to access $(\frac{W}{L})$:

$$\beta \text{ or } CR = \frac{(\frac{W}{L})_1}{(\frac{W}{L})_5} = \frac{(\frac{W}{L})_2}{(\frac{W}{L})_6} = \frac{\text{driver transistor}}{\text{access transistor}}$$

Using the above equations it is possible to draw ΔV_{max} versus CR. Fig.2.13 shows a simulation result in 0.13µm technology [3].

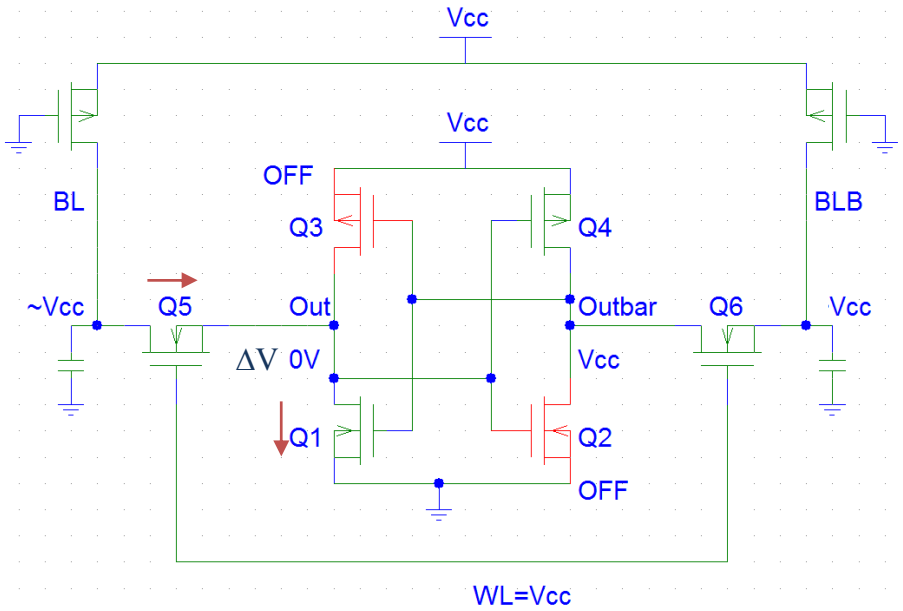


Fig.2.12: Read operation

Typically, in order to ensure a non-destructive read and an adequate noise margin, *CR* must be greater than from approximately 1 to 2.5. Larger *CRs* provide higher read current (and hence the speed) and improved stability at the expense of larger cell area. Smaller *CRs* ensure a more compact cell with moderate speed and stability. Leakage through the access transistors should be minimized to ensure robust read operation and to reduce the leakage power [3].

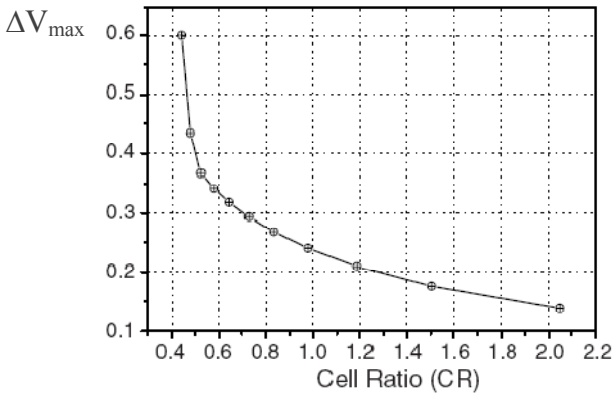


Fig.2.13: The rise ΔV of the "0" node (Out) in 6T SRAM cell [3]

Access time of a cell can be defined during read operation. The time difference from 50% of Vcc in world line to BL is access time that has been illustrated in Fig.2.14. As it can be seen in Fig.2.14 the access time is approximately 100pS [15]. In this picture we see that during read, stored data on points 1 and 2 are increased slightly. In node 1 value of 0 and in point 2 value of 1 (V_{cc}) were stored. When world line is high, point 1 goes a little bit higher than 0 that with appropriate CR it will not be more than V_{th} of N3 in Fig.2.14. In this case precharge voltage is V_{cc}.

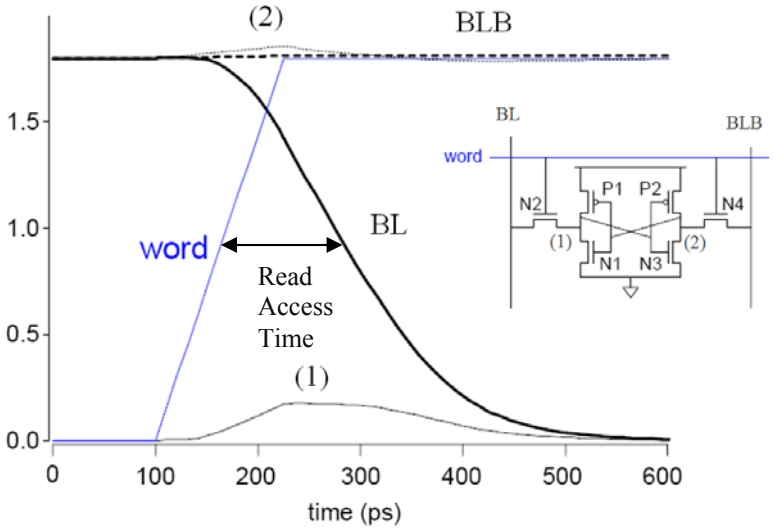


Fig.2.14: Definition of access time. Precharge voltage is V_{dd}.

Fig.2.15 shows two examples of "Pass" and "Fail" for reading. Precharge voltage is V_{cc} /2 and we can see that during reading the voltage of V_L drops a little bit. When the sizing has not been selected correctly, V_R increases more than V_{thn} of drivers and V_L decreases more than V_{cc}-$|V_{thp}|$ and inverters flip the data from 0 to 1 and we lose the correct data [12].
A preferred sizing solution is using a minimum-width access transistors with a slightly larger than the minimal length channel and a larger than minimal width with a minimal length driver transistors [3].

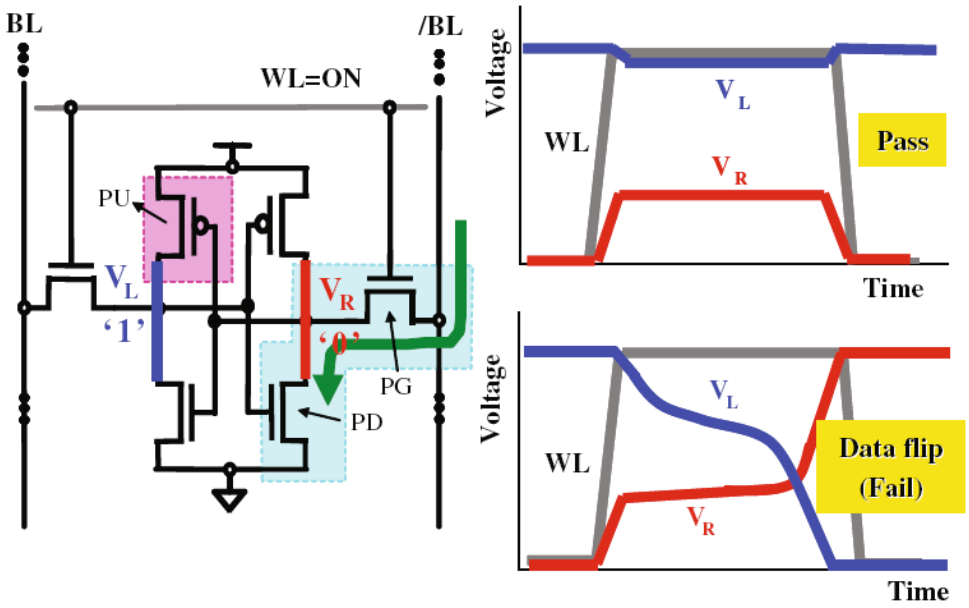


Fig.2.15: SRAM read operation and its waveforms for two cases of "Pass" and "Fail". Precharge voltage is $V_{dd}/2$ [12].

2-5-2) SRAM Data Retention or Hold operation

When WL is not high, SRAM cell is in data retention or hold mode. Data should remain correctly in the cell until a read or write action happens. Usually to reduce the leakage current during hold mode, supply voltage is reduced to a retention or hold voltage. A sufficient level (V_{DDH}) is required to turn-on the inverters as shown in Fig. 2.16.

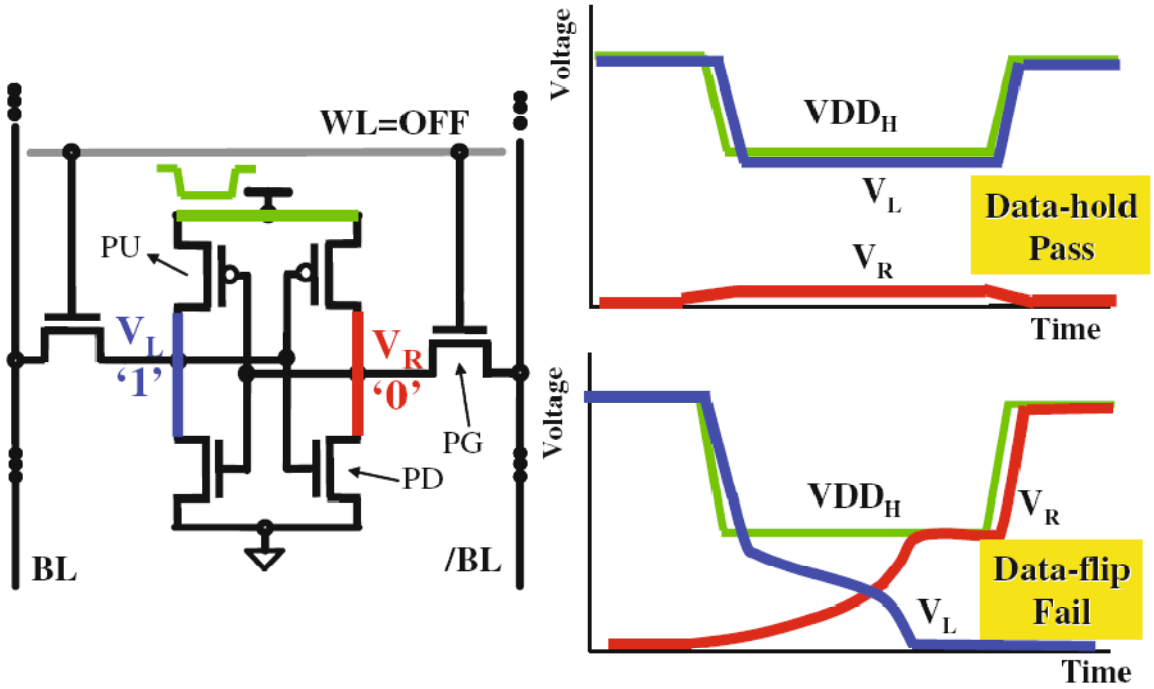


Fig.2.16: SRAM data retention. Hold supply voltage should be enough to keep the data correctly [12].

The cross-coupled inverters will reinforce each other without any disturbance from BL or BLB through pass transistors (PG). As a result, SRAM data can hold the full potential difference of (V_{DDH}–V_{SS}). However, when V_{DDH} gets lower than a certain point, which is referred to as SRAM data retention voltage (V_{HOLD}), no longer the inverters are able to hold the state, the storage nodes can latch into the wrong state as shown in Fig. 2.16 [12]. In nanometer-scale, SRAM cells need to lower the V_{DDH} as low as possible to reduce the leakage while maintaining V_{HOLD}. However, the trend of V_{HOLD} gets higher as SRAM size is scaling.

2-5-3) Write Operation

During write operation one of the bit lines, BL in Figure 2.17, is driven from precharged value (Vcc) to the ground potential. If transistors Q3 and Q5 are properly sized, then the cell is flipped and its data is effectively overwritten. Consider "Out" has "1" and "Outbar" has "0" stored information. As it can be seen Q1 and Q4 are off and Q2 and Q3 are in linear region. We want to write "0" on node "Out" and consequently "1" on node "Outbar". When WL is activated Q5 and Q6 go to saturation region. In order to change the stored data, Q1 should be on and Q2 should be off. But from read operation design we have $V_{outbar}<V_{th1}$ and Q1 can not be turned on. Hence Q2 must be switched off. It means the voltage of "Out" node must be lower than V_{th2}. Here are the equations:

$V_{out} \leq V_{th2}$ and $I_{Q3}(lin) = I_{Q5}(sat)$

Now we define Pull up ratio as a ratio of load (W/L) to access (W/L):

$$PR = \frac{(\frac{W}{L})_3}{(\frac{W}{L})_5} = \frac{(\frac{W}{L})_4}{(\frac{W}{L})_6} = \frac{\text{pull up transistor}}{\text{access transistor}}$$

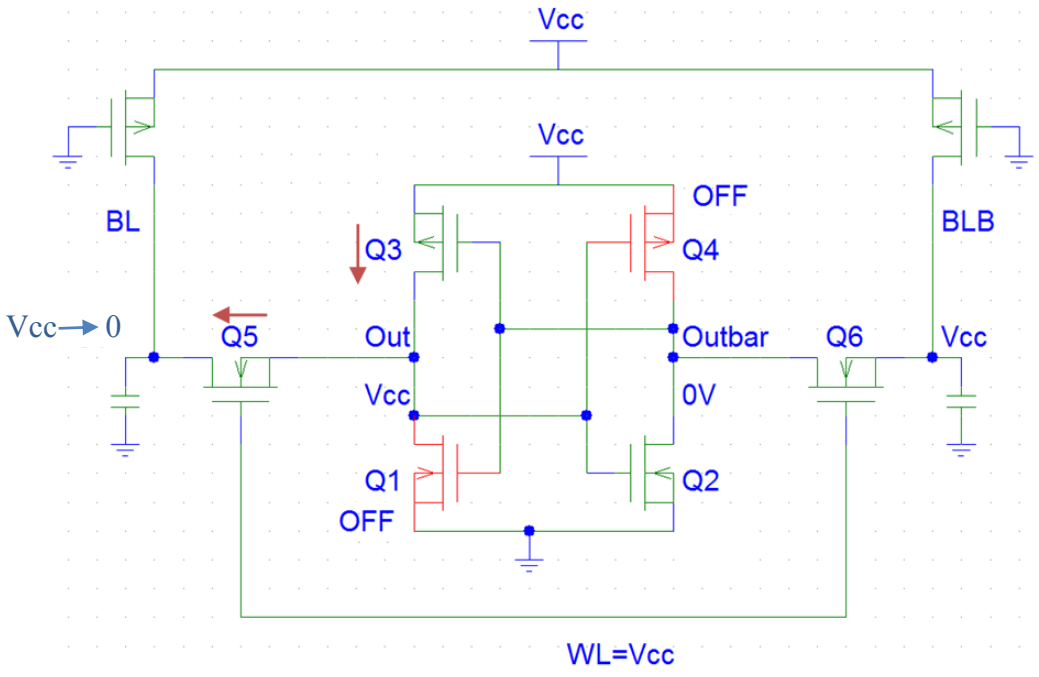


Fig.2.17: Write operation.

In order to write the cell, the access transistor Q5 must be more conductive than the Q3 to allow node "Out" to be pulled to a value low enough for the inverter pair (Q2/Q4) to begin amplifying

the new data. Fig.2.18 shows simulated result of the drop voltage of node V_{out} versus PR in 0.13µm CMOS technology at 1.2V supply voltage. The exact maximum allowed *PR* is defined by the V_{thn} process option and by the switching threshold of inverter Q2−Q4 in Figure 2.17.
Normally, to minimize the cell area and hence, increase the packing density, the sizes of the pull-up and access transistors are chosen to be minimal and approximately the same (PR=1). However, stronger access transistors and/or weaker pull-up transistors may be needed to ensure a robust write operation under the worst process conditions e.g., in the fast PMOS and slow NMOS process corner. On the other hand, a relatively strong pull-up PMOS also benefits the read stability due to the increased P/N ratio of the back-to-back inverters (Q1-Q3) and *(Q2-Q4)* in Figure 2.17. The read stability of an SRAM cell on one hand and the write ability of the cell on the other hand are conflicting design requirements. It is getting increasingly more difficult to balance these requirements by conventional transistor sizing and V_{TH} optimization as the design window becomes increasingly narrower [3] with the technology scaling.

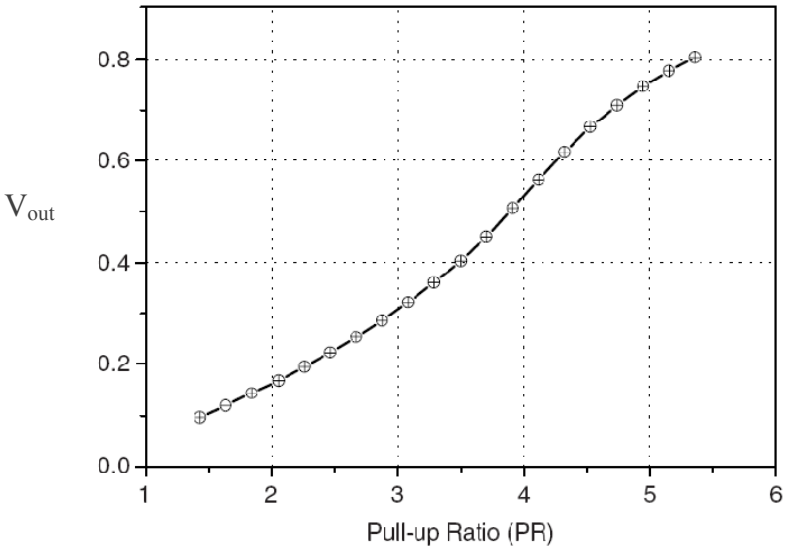


Fig.2.18: The voltage drop at node V1 during write access as a function of the pull up ratio (PR) at 0.13µm CMOS technology and 1.2V supply voltage [3]

Write access time refers to the memory speed in write mode with respect to the activation of the word line as shown in Fig.2.19. The write delay is measured between the time when WL reaches to 50% of Vcc and the time when node “Out” reaches 50% of Vcc [15]. In this case it is approximately 100pS.

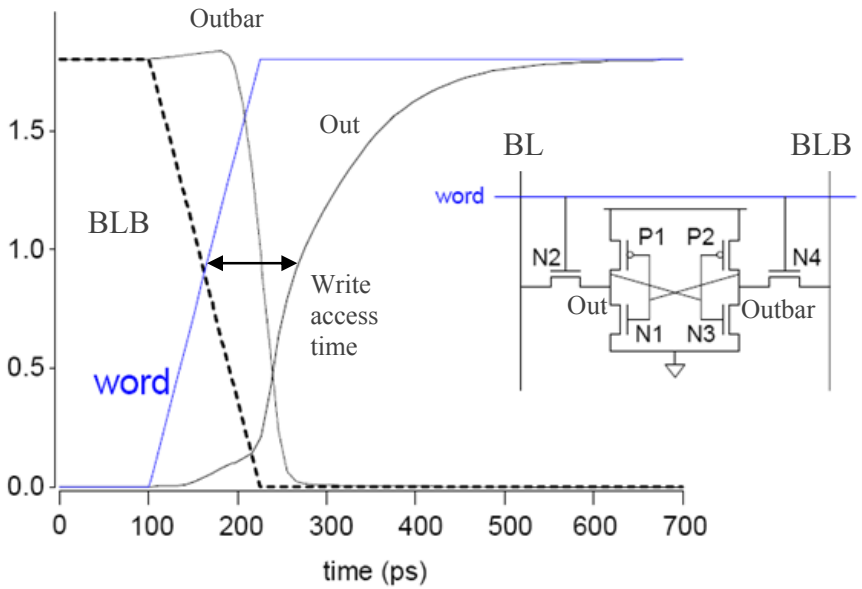


Fig.2.19: Write access time measurement

A statistical measure of SRAM cell write ability is defined as *write margin*. Write margin is defined as the minimum bit line voltage required to flip the state of an SRAM cell [3]. The write margin value and variation is a function of the cell design, SRAM array size and process variation. A cell is considered not writeable if the worst-case write margin becomes lower than the ground potential [3]. Fig 2.20 shows "pass" and "fail" cases in SRAM cell writing [12].

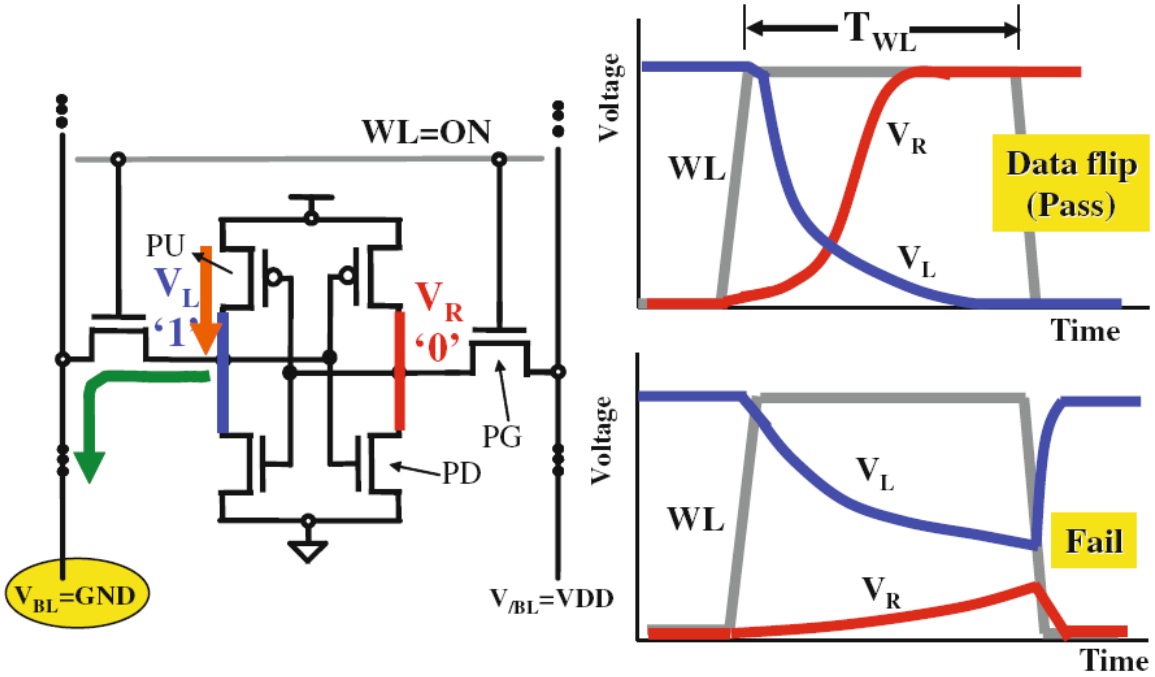


Fig.2.20: SRAM cell write operation and its waveforms [12]

In order to write a SRAM cell, write driver is used. The function of the SRAM write driver is to quickly discharge one of the bit lines from the precharge level to below the write margin of the

SRAM cell. Normally, the write driver is enabled by the Write Enable (WE) signal and drives the bit line using full-swing discharge from the precharge level to ground. Fig.2.21 shows a sample of write driver. As it can be seen three state buffers are used to feed the data to BL and BLB when Wr is enabled.

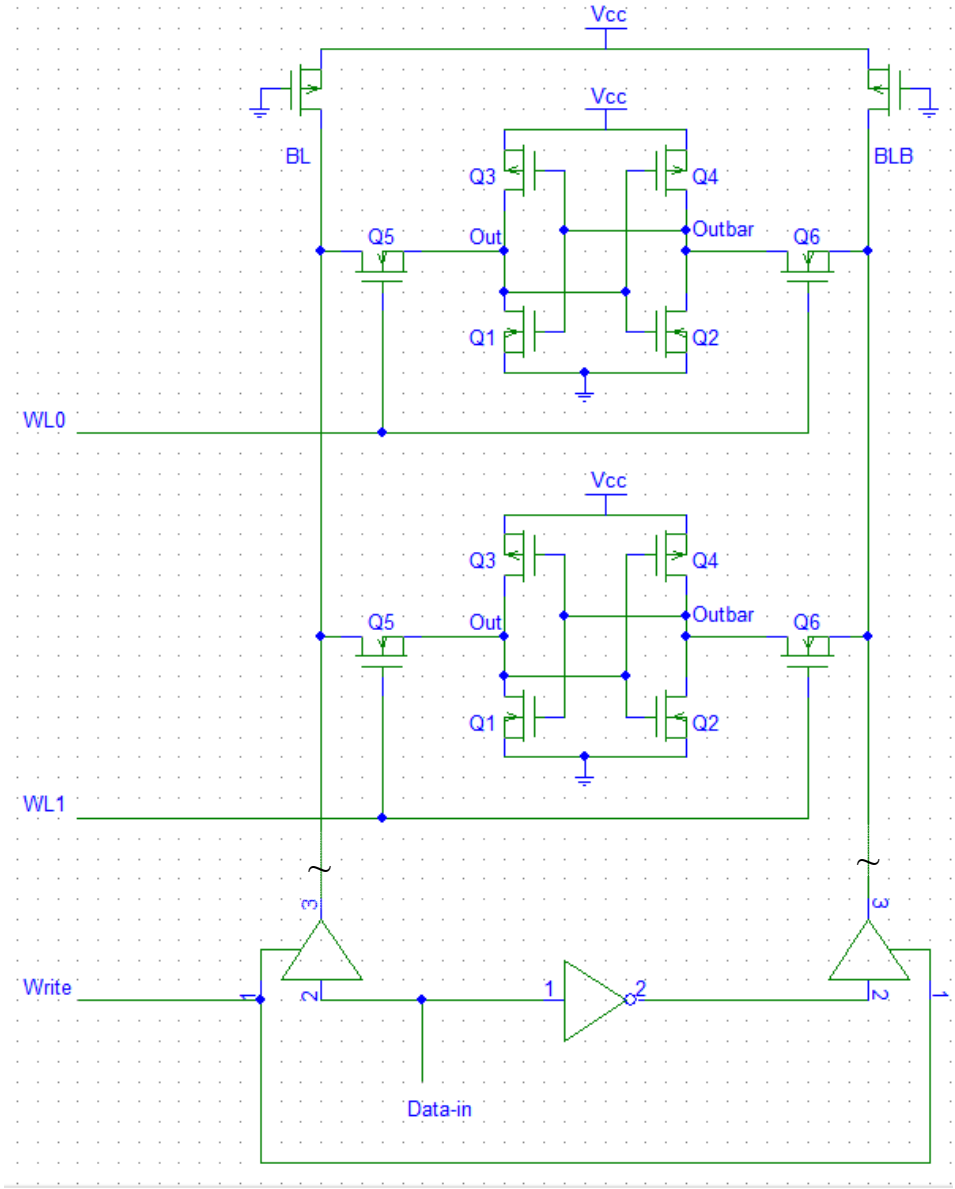


Fig.2.21: Write drivers circuit to write SRAM cell

2-5-4) SRAM Cell Precharging

Precharge aims to initialize the bitlines to the same voltage, in order to make sensing of the state of a cell possible, avoid errors, increase the performance and reduce the overall power consumption. Usually, the precharge equalizes the bitlines to V_{dd}, but in some memories, the bitlines are precharged at 0V or intermediate values such as $V_{dd}/2$. Precharging to Vdd makes it hard to write. Using $V_{dd}/2$ is more popular and it decreases power consumption. Here is the reason of reducing power consumption.

$\Delta t = c\frac{\Delta V}{I}$ and Energy=C×V_{dd}×ΔV

Where C is total capacitance in the Bitline and ΔV is voltage swing and V_{dd} is precharge voltage. Therefore making ΔV small and precharge voltage to $V_{dd}/2$, reduces the delay and energy. However, lower swing should be converted to high swing and this needs energy and delay and needs to be optimized. Possibly read instability can be happened with using $V_{dd}/2$. To solve this problem it is possible to increase V_{cc} (Inverters supply voltage) above V_{dd} during read and decrease it to V_{dd} during write mode dynamically [11].

2-6) SRAM Stability

The stability of SRAM cells determines its soft-error rate (SER) and its sensitivity to process tolerances and operating conditions [18].

2-6-1) Soft error:

Soft errors are random, nonrecurring errors that are not due to the physical defects in a device (hard error). They are "temporary" in a sense that when new data are written, the memory operates normally, whereas a hard error causes the memory location to fail permanently. Soft errors can be caused by various mechanisms such as system noise, voltage marginality, pattern sensitivity, alpha particles from chip packaging materials decay, thermal neutrons and cosmic rays that create energetic neutrons and protons [2, 3, 17].

Radiation can create localized ionization problems in the semiconductor devices especially memories, either directly or as secondary reaction products. Many of these radiation-induced events create enough electron hole pairs to degrade the storage nodes of SRAM cells. While such degradation can cause a data error, the device structures are not permanently damaged. If the voltage disturbance on a storage node of an SRAM cell is smaller than the noise margin of that node, the cell will continue to operate properly to keep the data. However, if the noise margin of a cell is not sufficient to withstand the disturbance caused by ionizing radiation, a "soft" error will result [3].

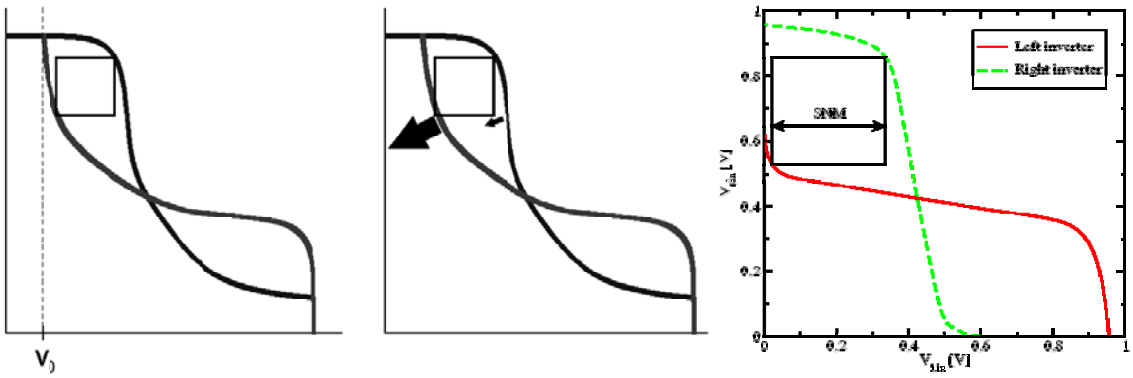
2-6-2) Static noise margin (SNM) or read margin

In many cases, the stability of the cell is a critical factor to obtain a desired yield and to lower the cost of the chip. Many different tests and methods exist that try to capture different aspects of the cell's stability. From many of metrics, Static Noise Margin (SNM) is the most important parameter [13].

Static noises are DC disturbances such as external noise or offsets and mismatches due to process variations and changes in operating conditions [14, 18]. The SNM of an SRAM cell is the maximum value of DC disturbances at the internal nodes that can be tolerated before the cell's storage value is flipped.

The SNM of SRAM cells can be determined graphically by drawing and mirroring the two voltage transfer curves (VTC) of the involved CMOS inverters and finding the maximum possible square between them as shown in Fig. 2.22(c)[14, 18]. The SNM is measured during the read, since the cell is most vulnerable when the pass-transistors are conducting.

When an external DC noise is larger than the SNM, the state of the SRAM cell can change and data is lost. Improved cell stability requires high values of SNM. SNM can be improved by upsizing the driver or increasing the gate length of the access device, but at the cost of the cell area and write performance. As it can be seen in Fig.2.22, studying the butterfly curve indicates that to enlarge the size of the SNM square, designers must lower the value of V_0 in Fig.2.22(a). Since V_0 is determined by the inverse of $driver/access$ transistors, this ratio must be high.



(a) (b) (c)

Fig.2.22: Butterfly curve to determine read SNM [14]

The large arrow in Figure 2.22(b) illustrates the effect of increasing the $driver/access$ ratio. However, increasing driver width has the side effect of decreasing the $pull\ up/driver$ ratio (highlighted by the small arrow), which would slightly decrease SNM. Following this we reach to this conclusion that increasing pull up to achieve higher $pull\ up/driver$ ratio could also improve a cell's SNM. On the other hand, to achieve good write margin, a "0" on the bitline must be able to overcome pull up holding the storage node at "1" through access transistor. Therefore, decreasing access width and increasing pull up transistors width to improve SNM would negatively influence on the write margin. Especially in nanometer technology and with voltage scaling and increased device variation, it is becoming difficult to satisfy both read and write margins. [14] Fig. 2.23 shows the simulation and measurement setup for SNM. As it can be seen both bitlines and WL are connected to Vcc and we look to the inverters voltage transfer characteristics. This simulation shows SNM shrinks with each generation [19].

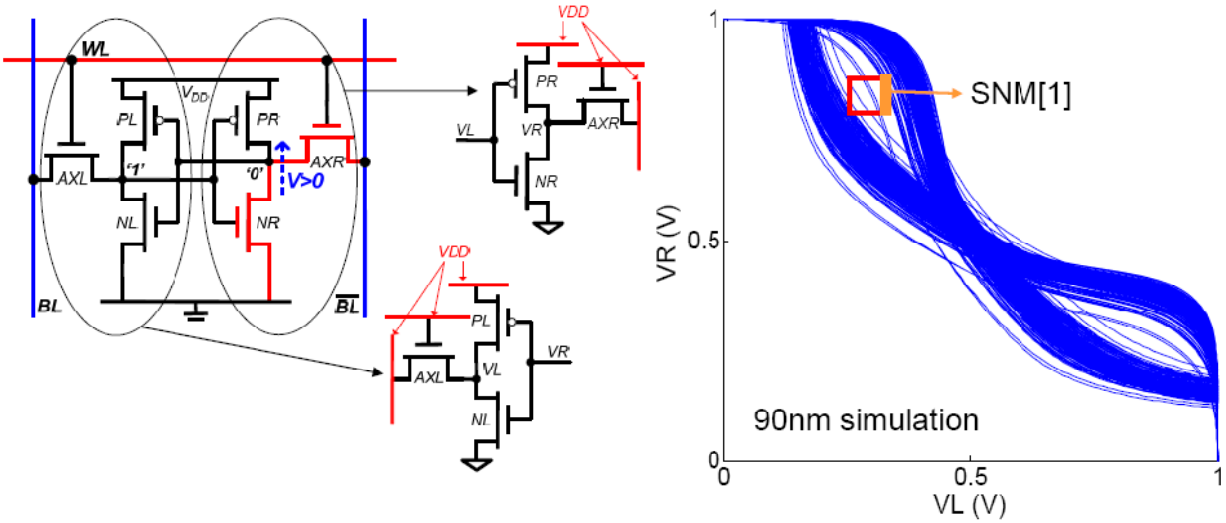


Fig.2.23: Simulation and measurement setup for read SNM and shrinking SNM with scaling [19]

Normally, the real VTC of an SRAM cell is asymmetrical due to mismatches of devices or defects. The SNM of the cell is proportional to the amount of asymmetry of the VTCs and the SNM value corresponds to the side of the smallest square. Variation in process parameters such as V_{TH} (spreads of over 10% of the typical values), W, L and presence of defects and poorly

formed contacts and Vias and Poor transistor matching can weaken the driving strength of one of the cell inverters. This can shifts the metastability point of the cell that results asymmetrical butterfly curve. This will decrease SNM and any noise disturbance more than the metastable (switching) point of the cell will cause such a cell to flip states. As shown in Fig.2.24, VM_{good} and VM_{weak} represent the voltages corresponding to the metastable points Z_{good} and Z_{weak} of the good and weak cell, respectively [3].

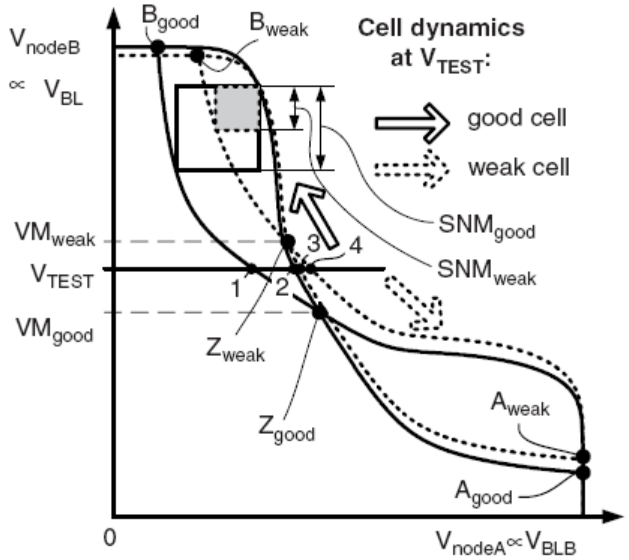


Fig.2.24: VTC curve of typical and weak SRAM cell [3].

2-6-3) Retention noise margin (RNM) or hold margin

Retention or hold noise margin of cell is the SRAM static noise margin (SNM) in the standby mode. RNM represents the maximum amount of voltage noise that can be introduced at the outputs of the two inverters such that the cell retains its data. 6-T cells present good retention as long as the supply voltage is high enough (>data retention voltage, DRV). It is determined as the side of the largest square that can fit within the butterfly curve for standby operation (access transistors are not active [13]. As it can be seen in Fig.2.25 noise margin in read operation mode is much lower than standby due to loading effect of nMOS access transistors during read operation mode [25].

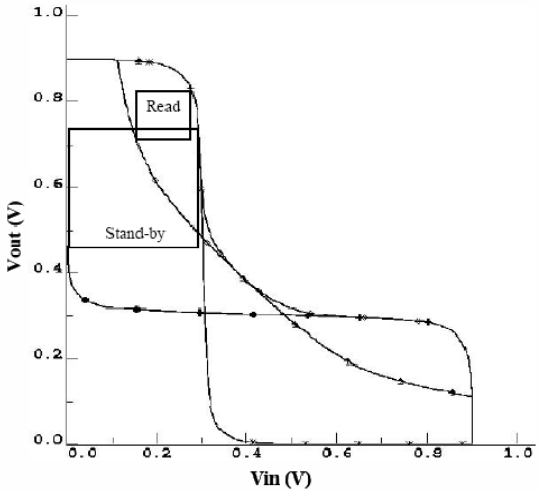


Fig.2.25: Comparing read and standby SNM, Vdd=0.9V [25]

2-6-4) Write ability current (I_{wr})

Write ability current is the minimum current flowing through the storage node during write operation. It can be determined by the current margin available for stable writing through bitlines. To write easily, higher I_{wr} is preferred. The write ability of a cell can be improved by increasing the access or decreasing the load device, but that would degrade read SNM and result in area penalty. SRAM cell sizing is a tradeoff between stability and write ability [13]. One of the solutions is to use double gate transistors that we will discuss about it in this chapter.

2-6-5) Write SNM or Write noise margin (WNM)

Write margin is the maximum voltage of bitlines nodes to be able to flip the cell state during write operation. Write margin can be measured from write ability current (I_{wr}, as defined in section 2-6-4) or the write noise margin derived from the butterfly curves. Fig.2.26 shows measuring write ability current and definition of WNM [25].

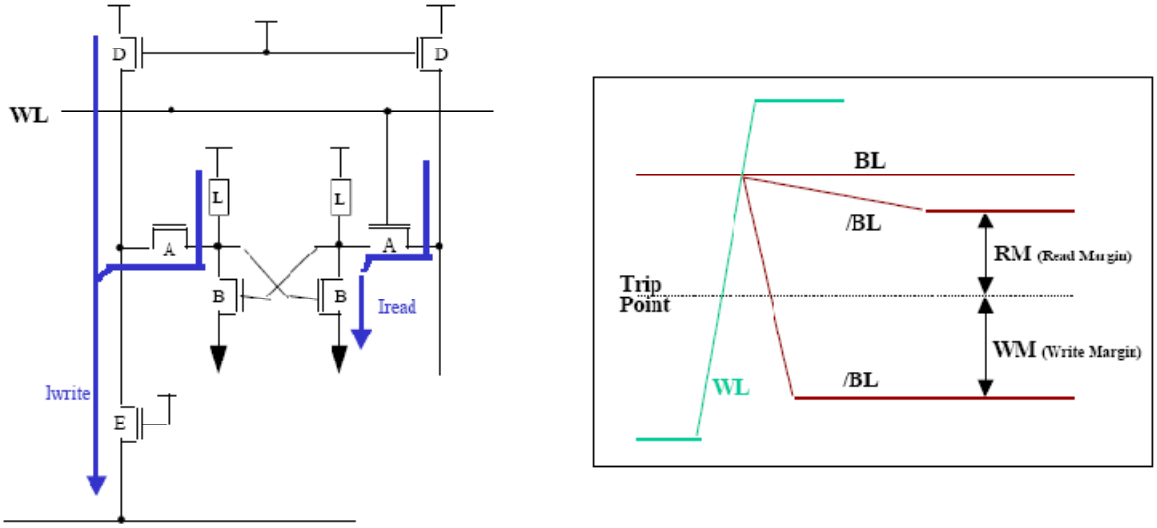


Fig.2.26: Measuring write noise margin [20]

Fig.2.27 shows measuring WNM from butterfly curve. To measure WNM one of the bitlines is connected to 0 and the other one is connected to Vdd and WL also is connected to Vdd [19]. Write stability is becoming more tighter with scalling.

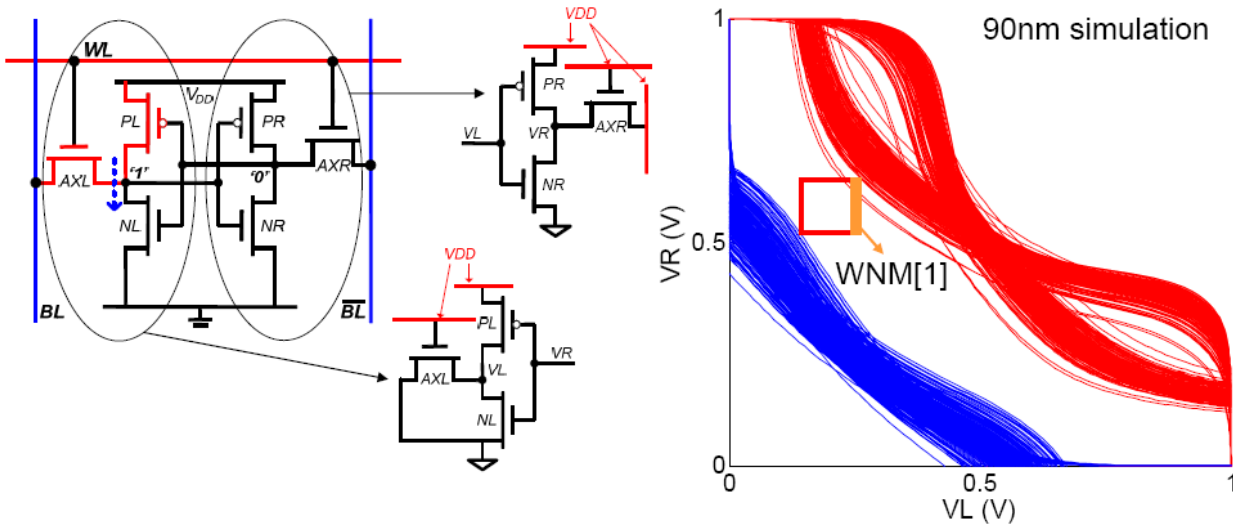


Fig.2.27: Measuring WNM from butterfly curve [19]

2-6-6) Static Power Consumption

Static power consumption is a critical parameter in some applications such as portable devices where memories should have a quite low activity and very low leakage current. Subthreshold current, gate tunneling current and GIDL (gate induced drain leakage) current are the dominant leakage mechanisms in sub nanometer technologies as shown in Fig.2.28. In CMOS 6T SRAM cell, leakage from V_{DD} to ground through the cell inverters, the leakage from the bitlines to ground and the leakage from bitlines to the word line are main paths for leakage [10,12].

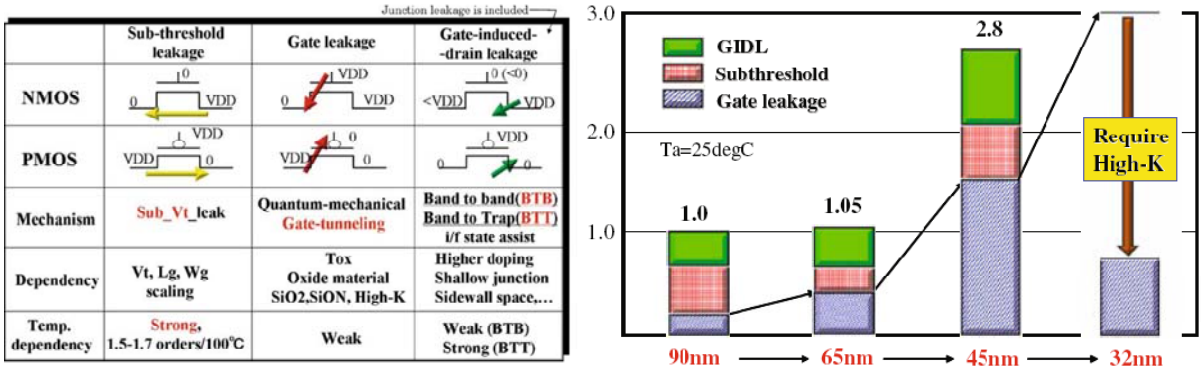


Fig.2.28: Leakage sources and trend of SRAM leakage with scaling [12].

2.7) Sense amplifier

Sense amplifiers are key elements in defining the performance and environmental tolerance of CMOS memories. It strongly influences the memory access time and power [1, 3, 21]. Due to large arrays of SRAM cells, the resulting signal, during read operation, has a much lower voltage swing as shown in Fig.2.29a.

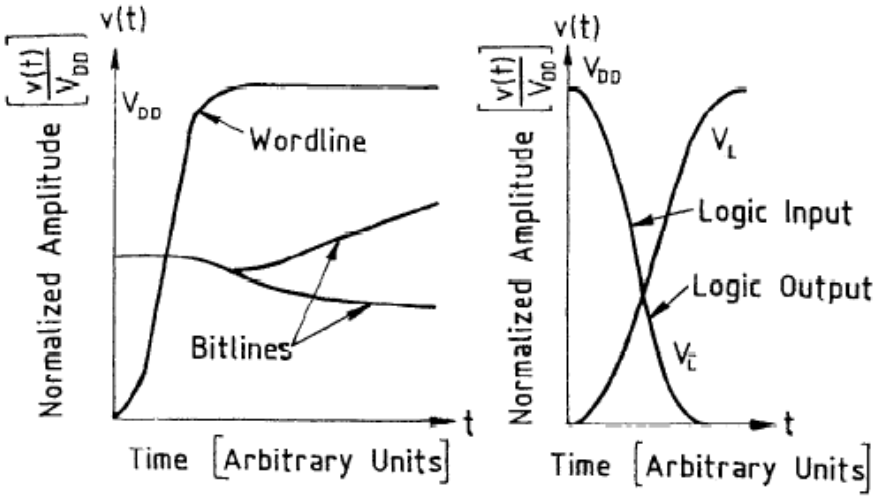


Fig.2.29: Unamplified data signals on the bitlines before sense amplifier and standard logic levels after sense amplifier [1].

To compensate output voltage swing a sense amplifier is used to amplify voltage coming from bitlines. The output voltage of the sense amplifier has fully voltage swing. Sense amplifier also helps reducing the delay times and power dissipation in the overall SRAM chip. Furthermore it reduces the time required for a read operation. Fig.2.29 shows the level of voltages before and after sense amplifier.

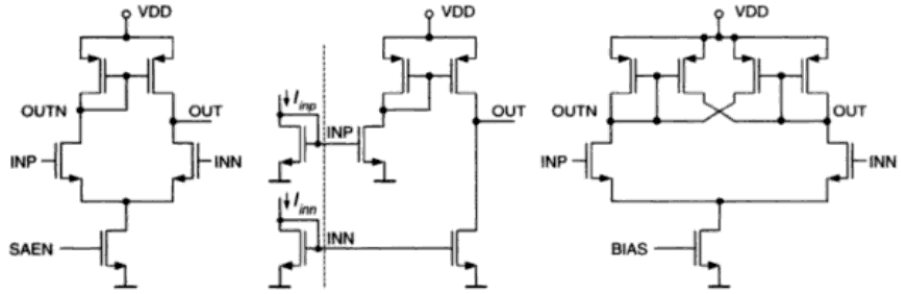
The delay of sense amplifier can be modeled with $t_p = (C \times \Delta V) / I_{av}$ where C is the capacitance of bitlines and ΔV is the differential voltage of bitlines and I_{av} is the sense amplifier's current. In large SRAMs capacitance is large and I_{av} is small. Then to decrease delay, we should make ΔV as small as possible. ΔV is defined by precharge voltage and typically it is in the range of 50mV to 200mV [3, 21].

In designing sense amplifiers many of parameters should be considered. Normally it is difficult to meet all requirements in one design. Depends on the application some of these conditions can be provided. Design constraints for sense amplifier are defined by the following parameters [1, 3]:

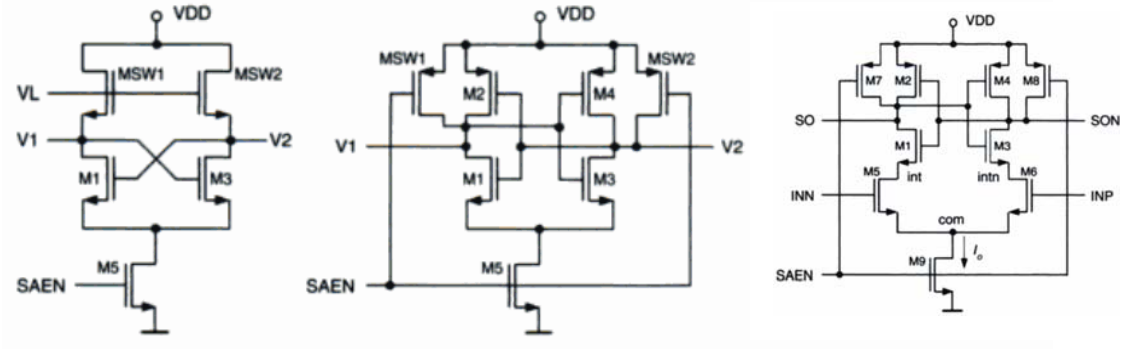
- The range of minimum differential input signal amplitude or resolution of sense amplifier
- Should have minimum sense delay
- Should have minimum power consumption
- What is the minimum gain (worse case gain)
- Should fit to the restricted layout area (area penalty)
- Should have less tolerance to the environmental conditions and mismatches (such as particles impact, temperature and hot carrier).

Sense amplifiers are classified by the operation modes (voltage, current, charge) and circuit types (differential and non-differential). It is possible to realize them in circuit level by static or dynamic structures depend on the application and power consumption constraints [1, 3, 21]. Fig.2.30 shows typical static and dynamic sense amplifiers. As it can be seen in static sense amplifiers inputs are separated from outputs. While in latch or dynamic sense amplifiers inputs and outputs can be same. In this type upon activation of sense amplifier the level of voltage boosted to Vdd on one side (BL) and GND on the other line (BLB) by positive feedback. It

means there should be a pass gate or multiplexer between sense amplifier and SRAM cells. Otherwise BL and BLB voltages will not be in precharge voltages (Q6 and Q7 in Fig.2.32).



(a)



(b)

Fig.2.30: Typical circuits for static (a) and dynamic (b) voltage sense amplifiers [21].

Table 2.2 compares the static and dynamic sense amplifiers [21].
In most of the SRAM applications the differential sensing is used since it allows rejection of the common mode noise that may present on both the bit lines. Noise sources, such as power spikes, capacitive coupling between the bit lines and between the word line and the bit lines, can inject common-mode noise to both sense amplifier inputs.

Table 2.2: Comparing two types of sense amplifiers

	Static	**Dynamic**
Complexity, Implementation	Good +	Good +
Speed	Medium -	High +
Power Consumption	High -	Low +
Can recover during sensing	Yes +	No -
Control effort	Good +	Difficult -

A typical current mirror differential sense amplifier with active load and the voltages of nodes is shown in Figure 2.31. Once the differential voltage exceeds the sensitivity of the sense amplifier a “Sense Amplifier Enable” (SAE) signal is applied and the sense amplifier converts the differential voltage on the bit lines to the full swing output level (out) [3].

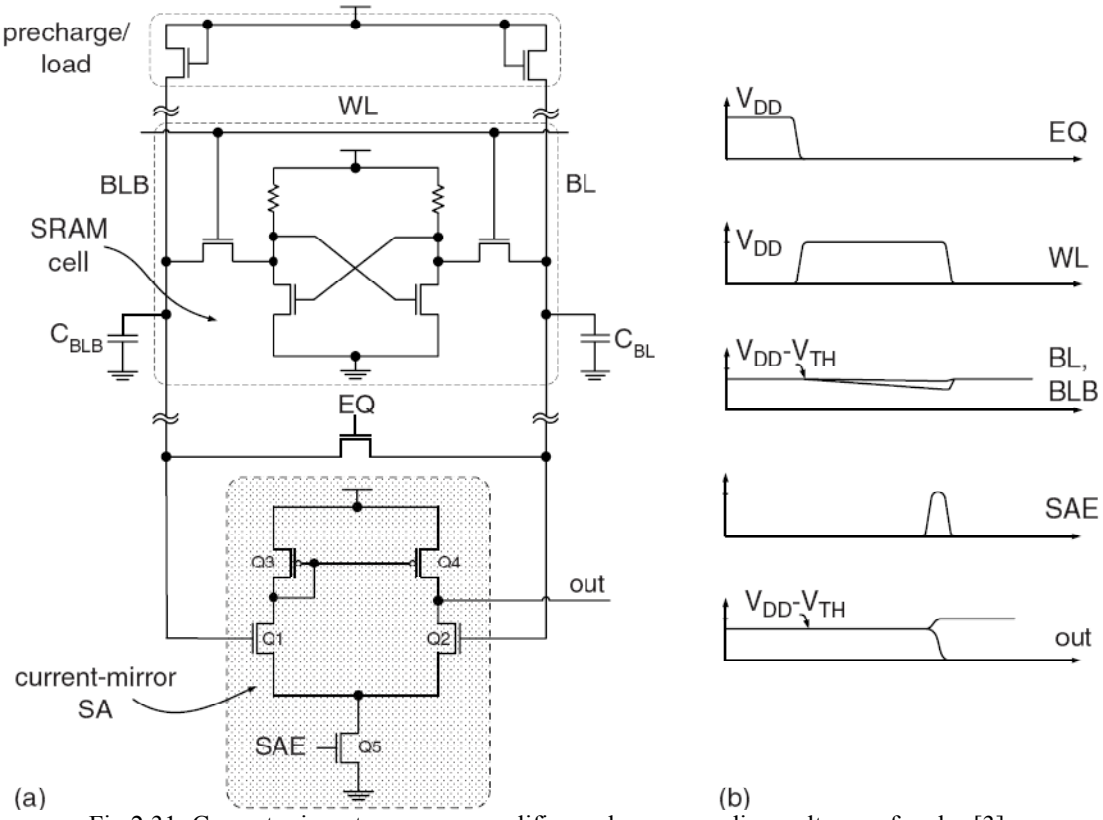


Fig.2.31: Current mirror type sense amplifier and corresponding voltages of nodes [3].

A latch type sense amplifier together with signals of nodes is shown in Figure 2.32. This type of a sense amplifier is formed by a pair of cross coupled inverters as a positive feedback, same as 6T SRAM cell. Positive feedback quickly drives the low capacitance outputs “out” and “outbar” to the full swing complementary voltages.

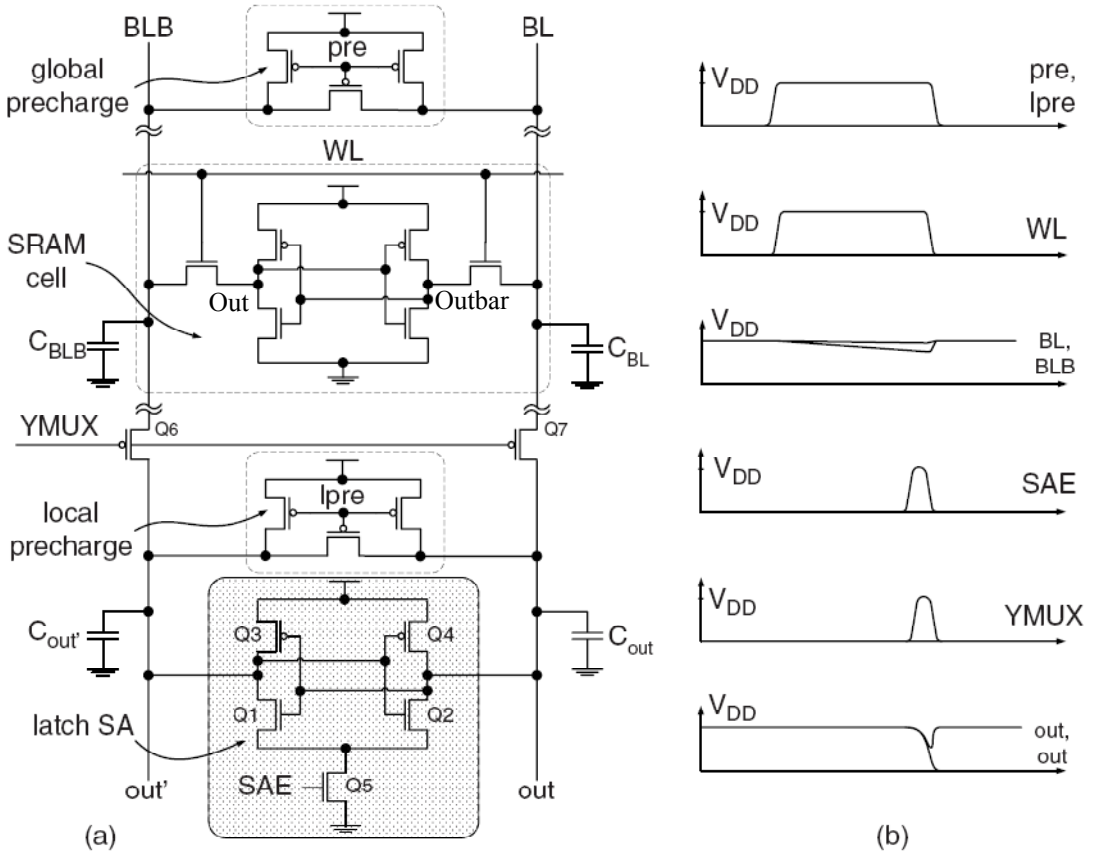


Fig.2.32: Lath type sense amplifier with a 6T SRAM cell and signals during read operation [3]

2-8) Principles of layout design in SRAM Cell

Compare to digital circuit design, specific and very aggressive layout design rules are used for the memory cell, in order to obtain the minimum area without impacting the reliability of the memory. To achieve high density, low power and high performance at the same time, the design must be carefully optimized to obtain the best trade-off. More particular in nanometer technologies layout should be designed in a way to improve critical dimension (CD) [22].

Designing SRAM layout influences strongly on the parameter variation. For example threshold voltage can be changed by layout style. Pattern deformation after processing (lithography and etching), mask misalignment and size fluctuation are major sources of V_{th} variation. This Variation is strong in the conventional cell layout shown in Fig.2.33a. This layout is called "tall" SRAM layout and has poor pattern reproducibility especially on the corners. Unfortunately, with device scaling, process variation become worse and worse that cause a larger variation of V_{th}. The solution is to use lithographically symmetric layout cell (LS cell) that shown in Fig.2.33b. As it can be observed in this layout it uses an advanced lithography, called optical proximity correction (OPC[1]), to be applied to every layer layout to reduce pattern deformation [3,5].

[1]OPC: Corrects the irregularities of shape and size and applies corrective compensation to the photo mask images, which then produce a light beam that more closely approximates the intended shapes. This is done by dividing polygon edges into small segments and moving the segments around, and by adding additional small polygons to strategic locations in the layout [23].

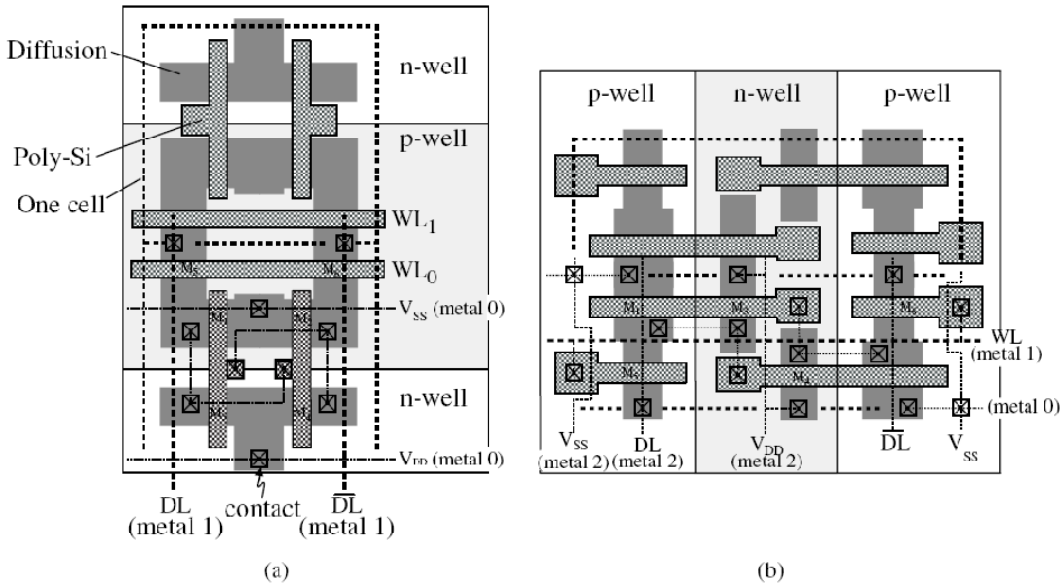


Fig.2.33: Comparing conventional (called "tall") and lithographical symmetrical (LS) (called "wide") layouts [3,5]

This type of layout is not sensitive to mask misalignments and offers excellent electrical stability with scaling and low voltage operations. Moreover, having same orientation of all the transistors in the cell provide better pattern reproducibility (W and L matching) and matching of threshold voltage of transistors. Since the word lines in the LS or so called "wide" layout are implemented in metal, it can offer reduced word line resistance and shorter propagation time of the WL signal to the parts of SRAM array which are far from the word line drivers. Another benefit of using this layout is the shorter bitline length per cell. Therefore, in same area it increases the number of cells and also reduces the capacitance of bitlines and capacitance coupling noise from next bitlines. This is because of the short and wide bitlines (DL in Fig.2.33) pitch and shielding effects of Vdd and Vss that surrounds bitlines in same metal layer. Table 2.3 summarizes the specifications of conventional and LS layouts [3,5].

Table 2.3: Comparison between conventional and LS SRAM layouts [5]

	Conventional cells	LS cells
Cell aspect ratio (height/width)	>1	<1
Well layer	Parallel to WL	Parallel to DL
Diffusion layer	Bended	Straight and parallel to DL
Poly–silicon layer	Two directions	Straight and parallel to WL
Application of advanced lithography (phase shift or OPC*)	Difficult	Easy
Pattern fluctuation	Large	Small
Immunity to mask misalignment	Poor	Good
Electrical balance	Poor	Good
Scalability	Poor	Good
Low voltage operation	Difficult	Possible

*OPC: Optical Proximity Correction

The critical masking layers in SRAM cells are diffusion, poly, contact and metal1. To increase the density of cells, the tightened design rules applied to the SRAM array in these layers that increase the risk of bridging/shorting. Having correct size of the patterns in these layers is critical to cell stability, drive and leakage currents. To improve the patterns in deep submicron processes, patterns are corrected using optical proximity correction (OPC). OPC of SRAM cells is especially challenging considering the tighter design rules in layout. To do SRAM OPC a collection of lithographic conditions and properties of the photoresist for each layer are needed. There are automated OPC algorithms in SRAM design layouts. Table 2.4 summarizes the issues associated with the critical layers in SRAM cell layout. Layout optimization and modification for high speed, low leakage and high density applications are different and needs different algorithms and design rules [3,5].

Table 2.4: Issues associated with the critical layers in SRAM cell layout [3].

Layer	Issues addressed
Diffusion	(a) Critical Dimensions (CDs) (for I_{read}) (b) Bridging (c) Adequate area for contact (for low $R_{contact}$)
Poly	(a) Critical Dimensions (CDs) (for I_{read}) (b) Symmetry (for SNM) (c) Bridging (d) Adequate end-cap
Contact	(a) Adequate coupled contact coverage over diffusion and poly (b) Bridging with adjacent contacts (c) Adequate clearance from potential ILD voids
Metal 1	(a) Bridging (b) Adequate overlap of contact and via (for low $R_{contact}$)

Fig.2.34 shows SEM image of fabricated SRAM cell in 65nm technology. In this layout unidirectional poly gates improve CD (critical dimension) control; shorter bitlines enable better cycle time and having metal wordlines result lower RC delay [intel].

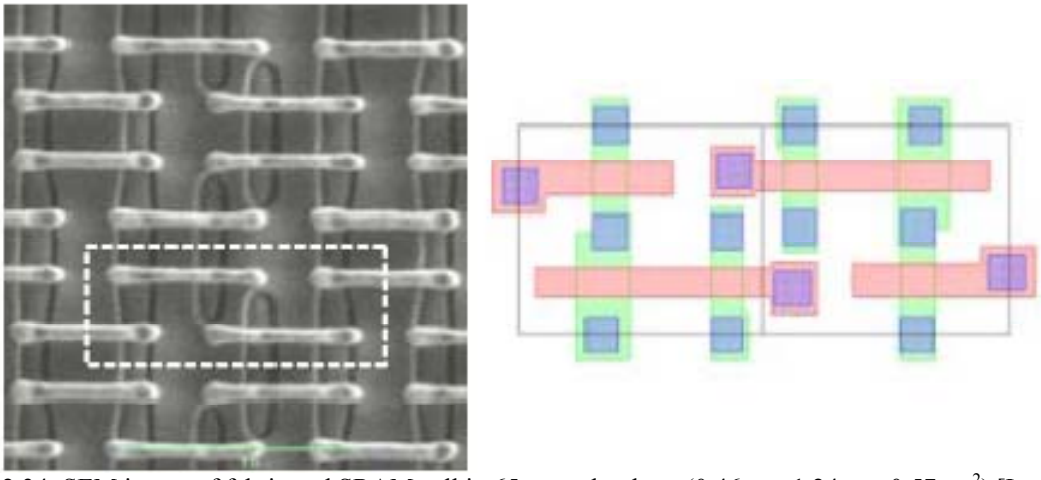


Fig.2.34: SEM image of fabricated SRAM cell in 65nm technology (0.46µm×1.24µm=0.57µm^2) [Intel].

2-9) Double-Gate and Multi-Gate SRAM Memories

2-9-1) SRAM Double Gate Circuits

As we discussed in previous sections during read and write operations, different requirements of transistors sizing are needed. There is a trade off in choosing transistor's sizing to meet read, write and area limitations. However, with scaling and in some applications it is very difficult to reach a good write noise margin (WNM) together with a good read noise margin (SNM) due to process variations in large memory capacity. To improve the SRAM cell electrical and stability characteristics, many innovative 6T SRAM cells have been proposed in multiple gate technology, such as DGMOS, FINFET and Ultra-thin-BOX FDSOI. All of them use the advantage of the independent biasing of the second gate to increase the cell ratio (CR) and decrease the pull up ratio (PR). To do this, two main approaches have been extensively investigated. In static approach second gate is biased to the different voltages permanently to improve CR and PR. In dynamic approach the voltage of second gate dynamically changes using feedback and signaling to again improve the CR and PR ratios.

Hitachi proposed a static double gate SRAM cell to improve both CR and PR at the same time. Fig.2.35 shows this circuit. As it can be seen second gate of pMOS load transistors are connected to Vcc and second gate of access transistors are connected to GND to make them weak. This improves write margin and read margin dramatically. However, weakness of access transistors causes lower read current [15].

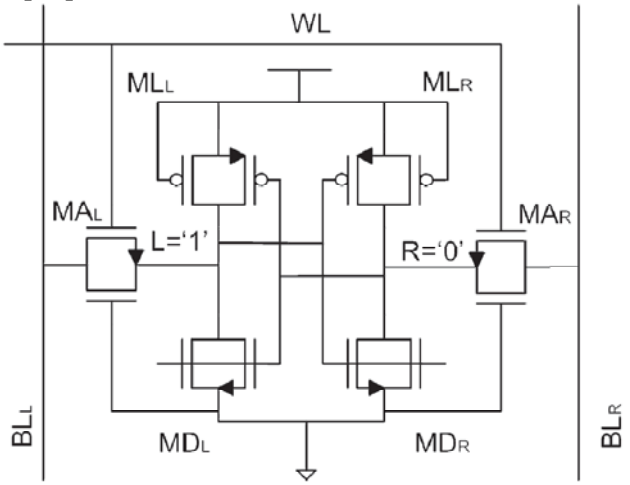


Fig.2.35: Hitachi 6T SRAM Cell [15]

In dynamic approach as it can be seen in Fig.2.36 different voltages are applied to the WL during read and write. As illustrated in this figure one of the gates (RWWL) is connected to Vcc during read and write. The second gate of access transistors (WWL) is connected to Vcc during write and to GND during read mode to make access transistors weak in read mode. Although this structure improves CR and PR, again read current is low. To improve read current it is possible to give a voltage in the range of 200mV to the WWL during read operation mode [15].

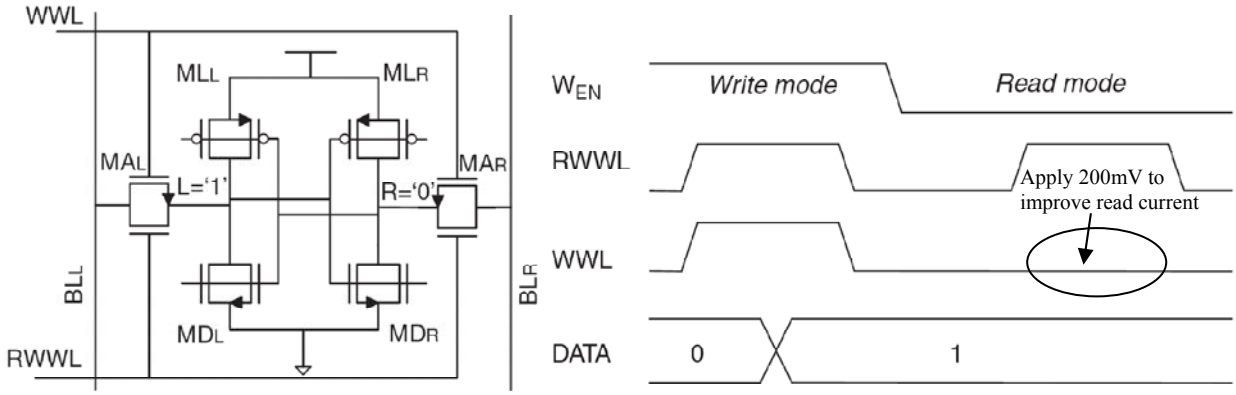


Fig.2.36: 6T-2WL RAM Cell [15]

An innovative structure of double gate SRAM uses feedback to improve read SNM and leakage current drastically by dynamically controlling access transistors without area penalty that is called Berkeley Pass-Gate Feed-Back (PGFB) cell. As illustrated in Fig.2.37 second gates of access transistors are connected to the storage nodes of "out" and "outbar". For instance when "outbar" point has "0" that connected to the second gate of MA_R, this transistor will be weak and CR improves during read. However, weakness of access transistor will degrade write margin.

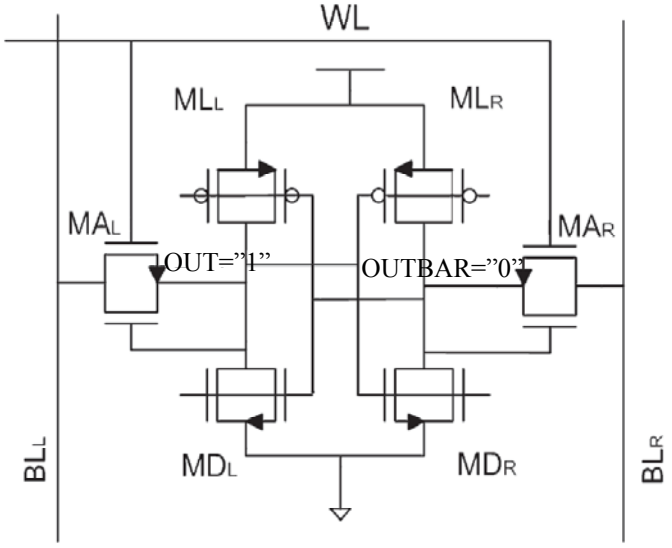


Fig.2.37: PGFB 6T SRAM Cell [15]

With adding write control signal to Pass-Gate Feed-Back (PGFB) cell new Pull-Up Write Gating (PGFB+PUWG) cell can be developed that has compensated degradation of write margin. As it can be seen in Fig.2.38 second gates of pMOS transistors (load transistors) are connected to new write word line (WWL). During write operation mode this line is connected to Vcc to limit the drain current of load pMOS transistors. This improves write margin a lot [15]. Nevertheless giving 0V to WWL during read degrades read SNM. But with applying an optimum voltage (in the range of 700mV) to WWL during read, it is possible to improve both write and read margins.

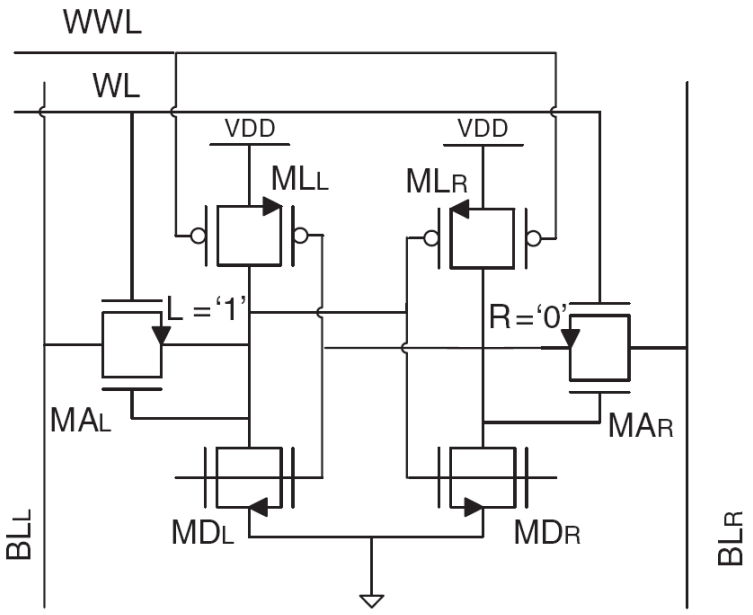


Fig.2.38: PGFB+PUWG 6T SRAM Cell [15]

2-9-2) SRAM Double Gate and Multi Gate technology

Double-Gate and Multi-Gate MOSFET improve subthreshold slope, leakage current and short channel effects and also are more scalable. Fig.2.39 shows a typical view of DGMOSFET. In planar double gate MOSFET threshold voltage of top gate is modulated with bottom gate. The relation between threshold voltages is given by the following equation [29]:

$V_{G1}-V_{TH1}=T_{OX}/T_{BOX}\ (V_{G2}-V_{TH2})$

Where V_{TH1}, V_{TH2}, T_{OX}, T_{BOX} are front and back threshold voltages and gate oxides, respectively. As shown in this figure Id-Vg curve changes with second gate biasing. In this picture "SYM" is symmetric case where both oxide gates are same. In this case the effect of both gate voltages are same. "PA" is partially symmetric where front gate oxide is T_{OXFG} =1.2nm and back gate oxide thickness is T_{OXBG} = 3.5nm. In this case the influence of gates voltages are not same and bigger oxide gate will influence less on the current. In case of fully asymmetrical (FA) gate oxides are same as "PA" and metal gate of back gate is different (different work-function) so threshold voltage is shifted from 325 to 800mV. As we can see the threshold voltage and leakage current are changed with applying voltage to the bottom gate [10,24].

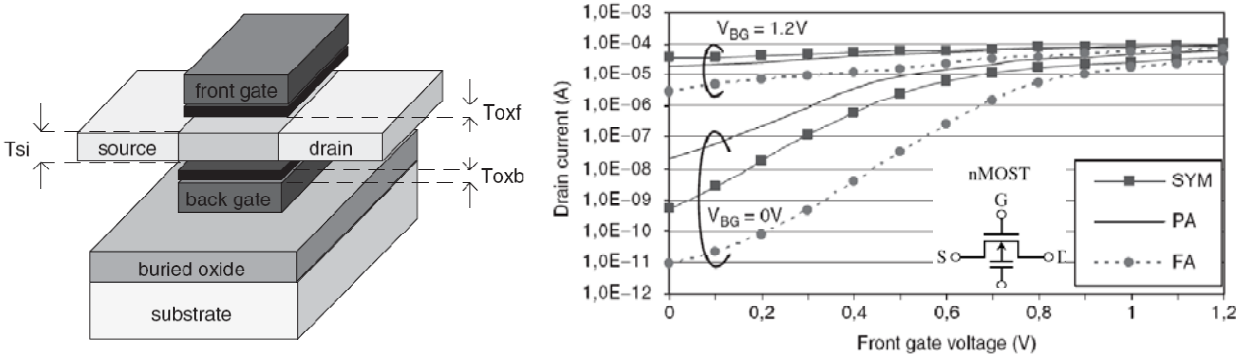


Fig.2.39: Cross section view of double gate MOSFET and corresponding Id-Vg with Toxf=Toxb, [10]

Depend on the application bottom gate can be connected to the top gate to act as a conventional single gate transistor. Bottom gate can be independent gate that results a device with two gates. In SRAM applications second gate is dynamically controlled to change V_{th}.
One of the main advantages of DG MOS is their lower leakage current. Different gate materials (such as metal, p+ poly, n+ poly), oxide thickness and different oxide materials and channel doping (can be intrinsic that results no random dopant fluctuation and high mobility) are main parameters that influence on the leakage current [24,30].
In double gate transistors top and bottom gates have overlap on the extension part of gates. This will create a parallel capacitor between two gates. This capacitor will influence the electrical characteristics of Id-Vg curve. The schematic of double gate transistor shown in Fig. 2.39 is in SOI (silicon on insulator) technology. In fully depleted (FD) SOI technology soft errors and leakage current of p-n junctions are reduces and drive current is increased. To increase the mobility, channel can be un-doped and also source/drain doping can be alternatively replaced by silicide [15].
Fig.2.40 shows another dynamic SRAM cell fabricated in fully depleted SOI technology. During read mode the bottom gates of pMOS and nMOS transistors are biased at "0" volt. This results high current for pMOS transistors during read. Consider node "N_1" has "0" and node N_2 has "1". During read N_1 should not have voltage more than V_{thn}. Because pMOS current is higher in read mode, point N_1 will not reach to V_{thn} and it means read SNM is increased. During write mode bottom gates are biased at V_{dd} and the current of nMOS transistors are increased. Therefore during writing changing voltage of nodes will be easy because access transistors can discharge nodes through driver transistors with higher current. pMOS transistors are weak in this mode. During hold or standby (retention) mode bottom gate of pMOS transistors are biased at higher voltage (2.5V) to increase threshold voltage to obtain lower subthreshold current for them. Vss is also biased at o.3V in this mode to decrease the subthreshold current of nMOS transistors [5,31].

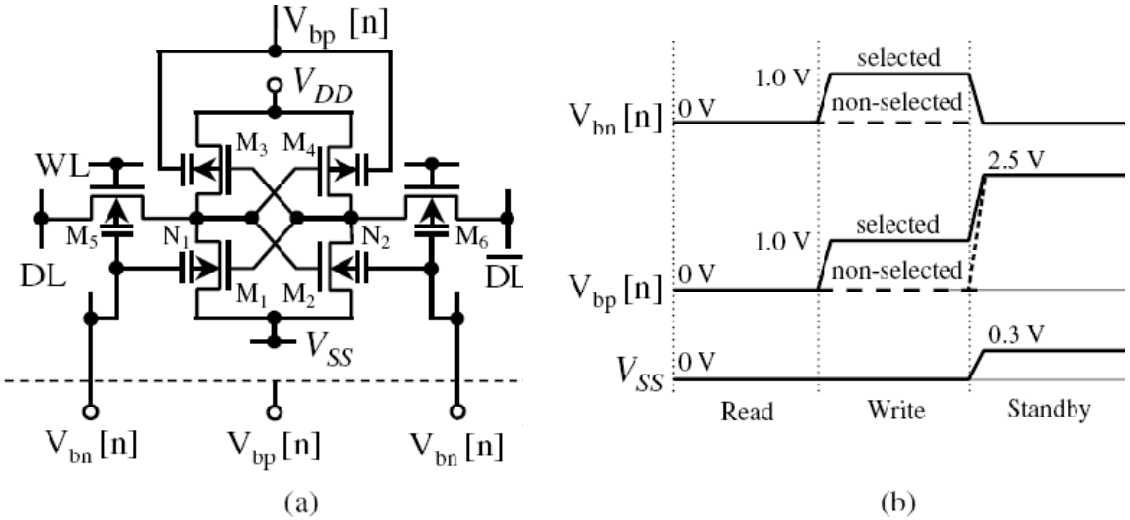


Fig.2.40: A dynamic SRAM cell with double gate transistors fabricated in FDSOI technology (a) and pulse timing (b) [5].

2-9-3) SRAM Cell using FINFET Transistors

The other structure of transistor used in SRAM cells is the FINFET transistors. FINFET technology is promising technology for future with transistor scaling and improved control of short channel effects. In this transistors the width of transistors are constant and defined by the

height of FIN as described in Fig.2.41. Therefore the remaining degree of freedom to design with this type of transistor is gate length. It is possible to change the gate of FINFET to the two independent gates using CMP (chemical mechanical polishing) or back etching using photoresist [26,28,32]. The second gate can control the Id-Vg curve as shown in this figure [28].

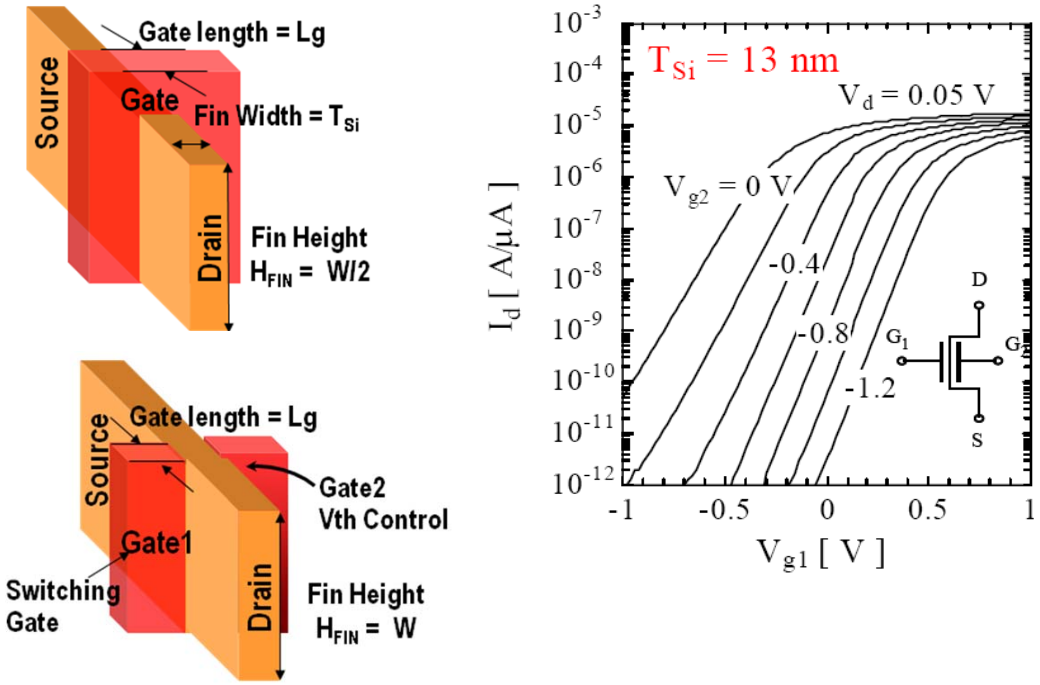


Fig.2.41: FINFET transistor [16,28]

A sample of fabricated SRAM cell using FINFET transistors is shown in Fig.2.42. The measured static noise margin was 185mV for VDD=1V. To improve the cell stability it is possible to use the independent FINFET structure [16,32].

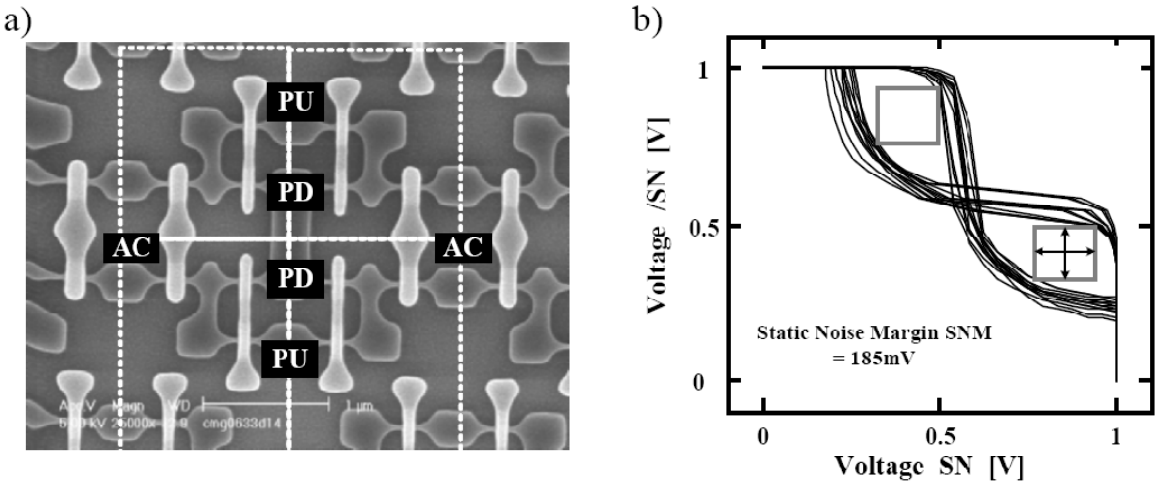


Fig.2.42: Fabricated SRAM cell using FINFET technology and SRAM butterfly curves with 1V supply voltage [16].

2-9-4) SRAM Cell using H-Gate Transistors

Another way of changing the threshold voltage of FDSOI transistors to improve cell stability in SRAM cells is to use H and T gate transistors. In this structure special shape of gate is employed to have access to the channel of the transistors and finally change the V_{th}. Fig. 2.43 shows H gate transistor and a cross section view of this structure. Parasitic body resistor from terminal BODY1 to terminal BODY2 ($P^+P^-P^+$) generally functions like a PMOS controlled by Gate. One part of the resistor area is doped more lightly than the other part, thus, when a high voltage is applied to gate, the p^- area quickly depletes, thereby greatly enhancing the resistance and shut down current between BODY1 and BODY2. With different biasing of points A and A' it is possible to change the threshold voltage of transistor. As it can be seen in this figure with changing V_A from 0V to 0.5V threshold voltage has been changed [33,34].

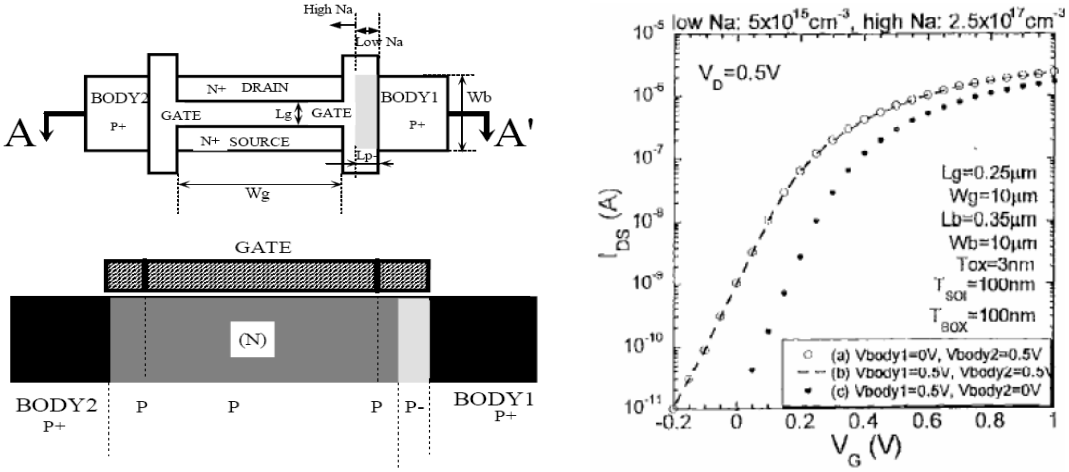
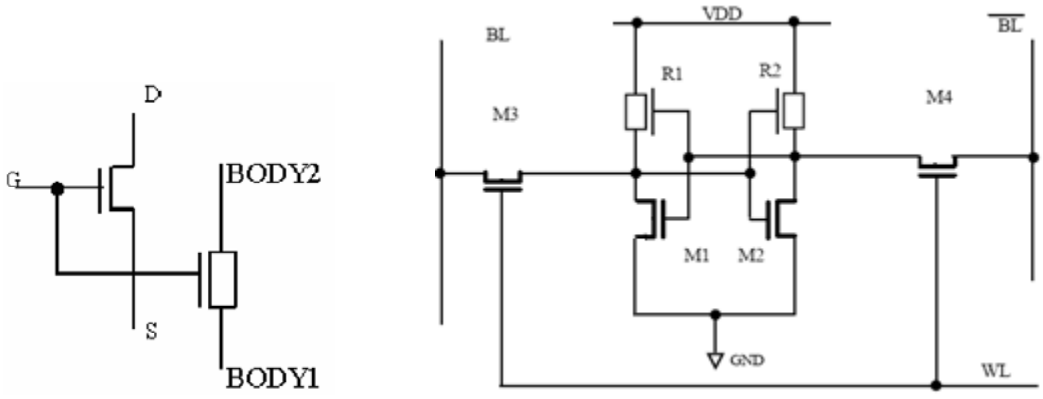


Fig. 2.43: H gate transistor and I_{DS}-V_{GS} curve. With biasing the body the V_{th} of transistor is changed [33].

In Fig.2.44(a) the equivalent circuit of H-gate transistor is demonstrated. It is consist of nMOS transistor with a resistor that acts like a pMOS transistor. The schematic of SRAM cell using H-gate transistors is show in Fig.2.44(b). As it can be seen this cell can work same as 6T SRAM Cell [33,34].



(a) (b)

Fig. 2.44: Model of H-gate transistor (a) and 4T SRAM Cell (b) [33]

The designed layout for this cell is demonstrated in Fig. 2.45. The black H shape is gate and can be in T or H form [33,34].

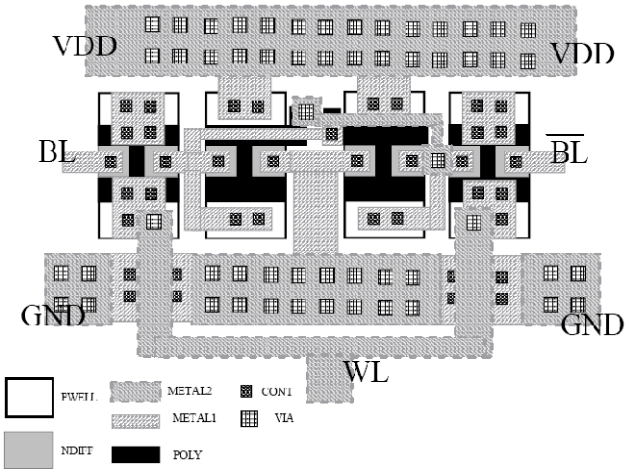


Fig. 26 Layout of the proposed 4T SRAM cell

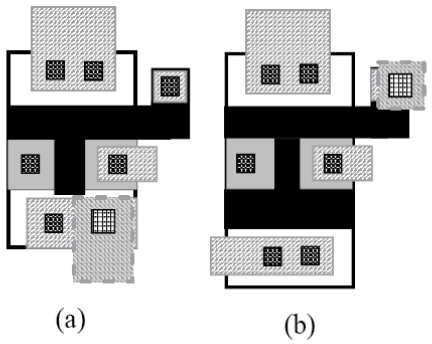


Fig. 27 T-gate structure (a) has less area than H-gate structure (b)

Fig. 2.45: Designed layout of 4T SRAM Cell using H gate transistor (a) layout of T and H gate transistors (b) [33]

2-10) Conclusions:

In this chapter the most frequently used structures for SRAM were introduced. The basic operation principles of 6T SRAM cell extensively was discussed in all operation modes including read, write, precharge and hold states and then the design constraints for SRAM cells were defined. There is a trade off in designing SRAM in Read and Write operation modes and also area penalty. Double gate transistors could improve both read and write margins and with different circuits configurations. Furthermore we saw layout designing is critical in SRAM as we want to achieve high density memory without sensitivity to process variation. OPC algorithms can improve SRAM stability. The different sense amplifier circuits to use in read operation modes were discussed in details.

Chapter 3

3DIC and μ-Czochralski Process

Content:

3-1) Three dimensional integrated circuits (3DIC)

3-1-1) Interconnection challenges in submicron technology

Request for minimum cost and power dissipation in VLSI circuits is increasing dramatically in computer industry and information technology. To increase functionality and performance of circuits and devices, scaling is the way to decrease gate delay in MOS devices. However, this scaling will increase the interconnect delay rapidly. In high performance chips, due to long wiring for clock distribution and loading effects of interconnects, significant fraction of the power consumption is interconnects. Moreover, designing these interconnects using computer tools needs much more cost and time.

As interconnect scaling continues, RC delay is increasingly becoming the dominant factor determining the performance of advanced ICs. Fig.3.1 illustrates this problem, where the gate delay and the interconnect delay are shown as functions of various technology nodes [38].

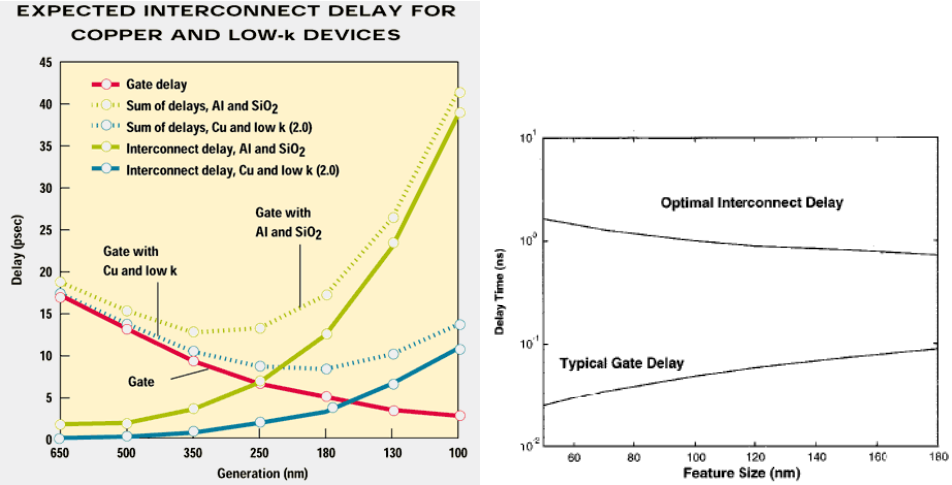


Fig.3.1: Comparing gate delay and interconnect delay with scaling. Decrease in interconnect delay and improved performance are achieved using copper and low-k dielectrics [38,39].

As it can be seen in this figure at 0.25µm technology Cu and low dielectric materials are used to decrease the interconnect delay. However, below 0.13µm this solution will not be able to decrease the interconnect delay.

One of the best example of interconnect problem is system on chip devices where approximately all aspects of a system design are on a single chip and long global wires are needed. In system on chip devices, CPU, SRAM, DRAM, MEMS, FPGA, RF and analog systems are integrated in same chip. Increasing the functionality, minimum area, cost and power consumption are the most challenges in system on chip. Furthermore, noise and interference between different circuit blocks are other challenges in the system on chip devices [38].

3-1-2) Three dimensional approach

Three dimensional integrated circuits (3DIC) offer significant performance advantages over two dimensional integrated circuits based on the electrical and mechanical properties that come from the new placement and arrangement of active and passive devices. Access to the third dimension significantly simplify communications between devices and the transfer of information and also provide rapid access to the memory. For instance, 3DIC technology would enable extremely dense memory cells with greater numbers of interconnects between active layers that will reduce

access times. This enables 3DICs to operate at higher clock rates and can consume less power over their 2D implementations due to minimizing the length of circuit interconnects [35].
In 3DIC each 2D block is stacked on top of each other and connected vertically (called vertical interlayer interconnects (VILICs)) and common global interconnects. It offers much more flexibility in system design, placement and routing. Fig.3.2 shows a typical 3DIC structure that consists of fabrication of active transistors in top silicon layers. 3DIC has this potential to integrate different technologies such as logic, memory, analog, RF and optical I/O circuits on different active layers [38,40,41].

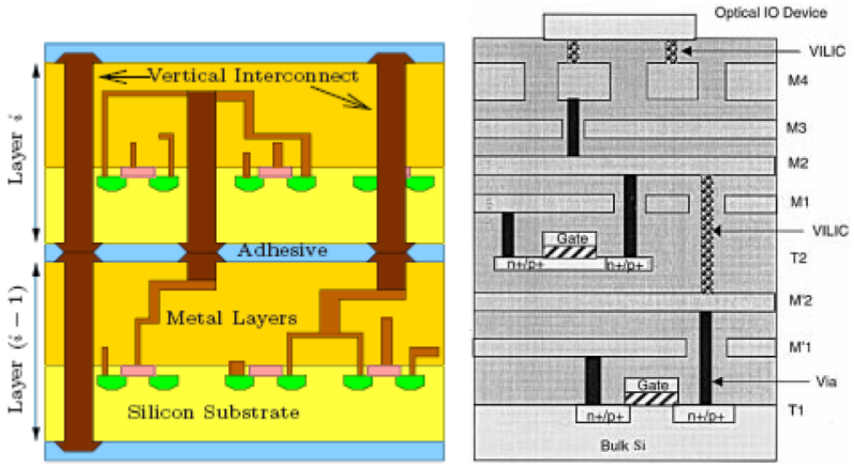


Fig.3.2: 3D-IC process [38,40,41]

With increasing the number of active layers, interconnects delay decreases as demonstrated in Fig. 3.3. However, stacking more than 3 silicon layers will not decrease interconnect delay significantly when we compare it with complexity of interconnection's design in 3DIC [38].

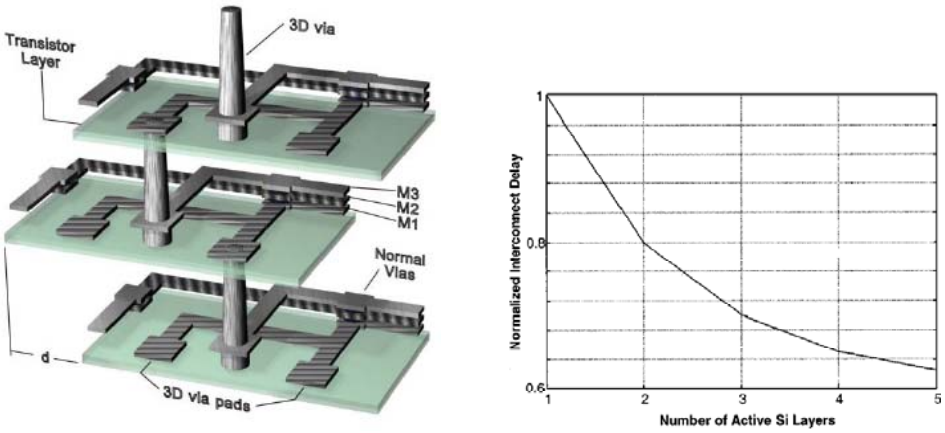
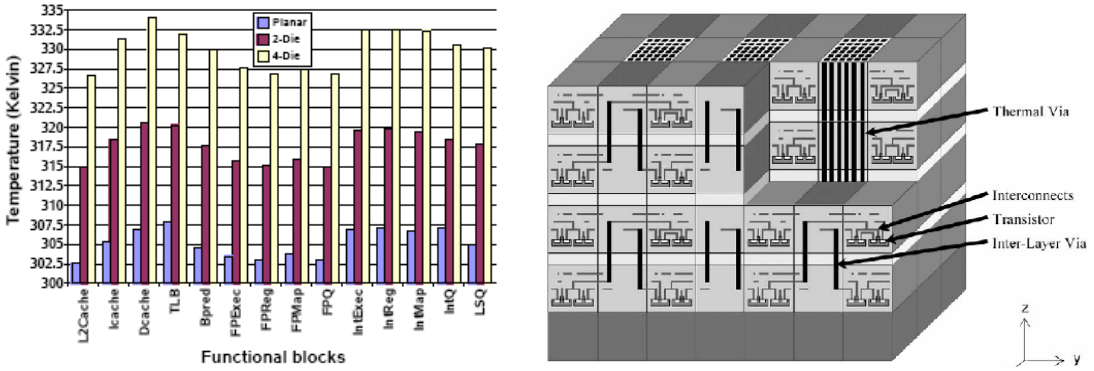


Fig.3.3: interconnect delay versus the number of active silicon layers in 3DIC [38,51]

3-1-3) Challenges in 3DIC

As mentioned before, stacking more silicon layers has a significant effect on reduction of wire length and power consumption. However, stacking active layers increase the temperature of silicon layers during their operation. The topmost active devices will experience high

temperature than bottom layers since they are far from substrate. This will influence on reliability and performance of devices. Fig.3.4(a) compares temperature difference between some circuits that have been made using planar, two stack layers and 4 stack layers fabricated by die stacking method. This figure shows in 4 stacked layers the temperature is approximately 25 degree higher than planar technology [42].



(a) (b)

Fig.3.4: Temperature differences between planar, two stacked layers and 4 stacked layers for some circuits [41,42,51].

Vertical interconnections can remove the generated heat, in addition to their function as electrical connections; they can act as heat pipes. In fact, some dummy vertical interconnections are used as generated heat removers as shown in Fig.3.4(b). These are connected to the thermal sinks in the package [41,51].

The other challenge associated with 3DIC is electromagnetic coupling between active layers. In 3DIC extra capacitance is exist between the top layer and the metal of bottom layer. This capacitor will couple two layers and will cause EMI (electromagnetic interference) problems. Moreover, because of high clock speed, wiring paths will show an inductive load. It will effect on waveforms in high frequencies. In 3DICs, the reduction of wire lengths will surely help to reduce inductance [38].

3-1-4) 3DIC Fabrication Methods

3D-IC has many advantages but realizing high quality devices using deposited silicon layers is a challenging subject. Most of the methods have complex fabrication process or they give poor quality and reliability. Beam Re-crystallization, Epitaxial Growth, Metal Induced Lateral Crystallization (MILC), Germanium as a seed for crystallization, chip level and wafer level bonding are the most common fabrication methods for 3DIC [38,41,43,44]. Among these methods wafer and die level bonding methods are the most frequently used in research and industrial products. In this technique wafers are processed in parallel including interconnects and then they are bonded together. This bonding can be done on wafer or chip level as shown in Fig.3.5.

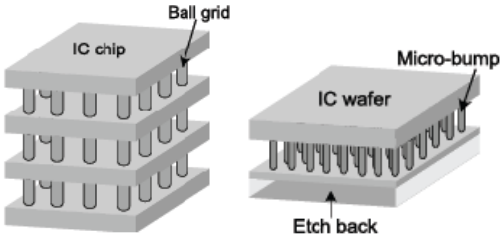


Fig.3.5: Chip and wafer level bonding to make 3DIC [41]

This technique basically uses bonding the wafers, which are processed at optimized process conditions, and then etching-back of one of the wafers [41].
In other approach before bonding, a glass substrate is bonded to the top part of devices and then silicon substrate is removed using wet etching or grinding and polishing. This removing of wafer is one of the issues in this technique. Then inter-chip vias are etched to electrically connect both wafers after metallization and before to the bonding process at 400^{o}C. This technique is very suitable for further processing or the bonding of more pairs in this vertical fashion. Other advantages of this technology is in the similar electrical properties of devices on all active levels and the independence of processing temperature since all chips can be fabricated separately and later bonded. Second limitation of this technique is its lack of precision alignment [36,43].
To achieve final 3DIC structure this method always involves the integration of four key technology areas: thinning of the wafers, inter device layer alignment, bonding, and interlayer contact patterning. An additional challenge in achieving high density I/O signal through the stack layers comes from thermal mismatch between the bonded layers that cause alignment tolerance. All of these 3DIC integration challenges require new material and process innovations. Fig 3.6 shows schematic diagrams of IBM assembly process, which uses layer transfer methodology to fabricate 3DICs [36,43].

- Attach circuit to glass substrate
- Remove original substrate
- Align and bond top circuit to bottom circuit
- Remove glass substrate and adhesives
- Form vertical interconnects

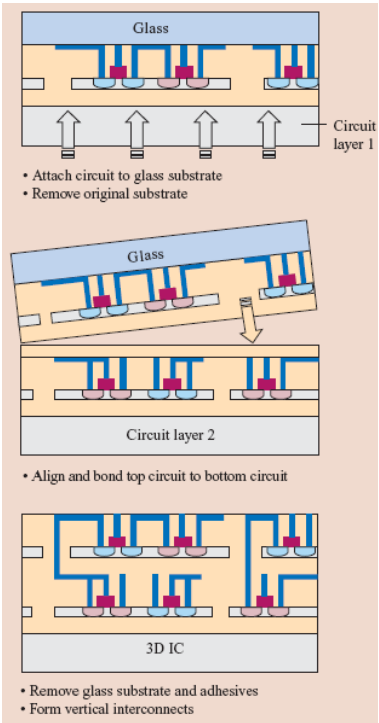


Fig.3.6: Schematic diagrams of IBM assembly process, which uses layer transfer methodology to fabricate 3DICs [43].

3-2) µ-Czochralski or Grain-Filter Process

One of the promising technologies to stack silicon layers to realize 3DIC at low temperature is grain filter or µ-Czochralski technology developed at TU-Delft. In this process a-Si is deposited on oxide layer which has some small holes. Then using Excimer laser amorphous silicon is melted while the bottom of this hole is solid. Crystallization starts from this solid seeds and distributed to whole surface. Single grain silicon can be obtained using this method. Fig.3.7 shows a schematic of this process [45]. First 700nm thermal oxide is grown on <100> p-type silicon wafer. After making 1µm holes on oxide, 750nm second oxide is deposited by PECVD with TEOS source at 350^{o}C to make holes as narrow as possible. In average they are in the range of 0.1µm and called grain filter. Then 250nm amorphous silicon is deposited by LPCVD at 550^{o}C. Silicon layer penetrates to the bottom of grain filter. During Excimer laser crystallization at 400^{o}C and 1500mJ/cm^{2} laser energy, silicon layer is melted in the surface except in the deepest part of grain filter. Solid silicon in this region acts as a seed for crystallization. The crystallized areas are single grain and thin film transistors (TFTs) are designed inside this grains [47,48,49,50].

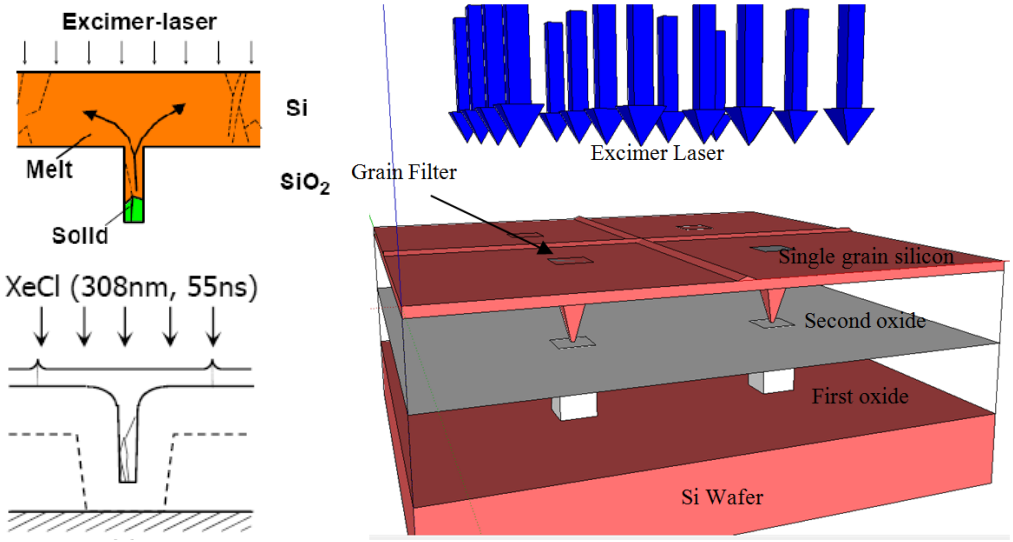


Fig.3.7: Schematic of silicon crystallization with µ-Czochralski process

In this process large single grain silicon such as 6µm×6µm can be obtained [45,46]. Fig.3.8 shows an SEM picture of crystallized region and AFM image of grains. As we can see the sizes of grains are in the order of 7µm. The brighter points are locations where four grains meet each other.

The maximum temperature of this process is 550^{o}C for LPCVD deposition of silicon. It is possible to use sputtering silicon in this process deposited at very low temperature (100^{o}C) enables this technology for flexible electronics [47].

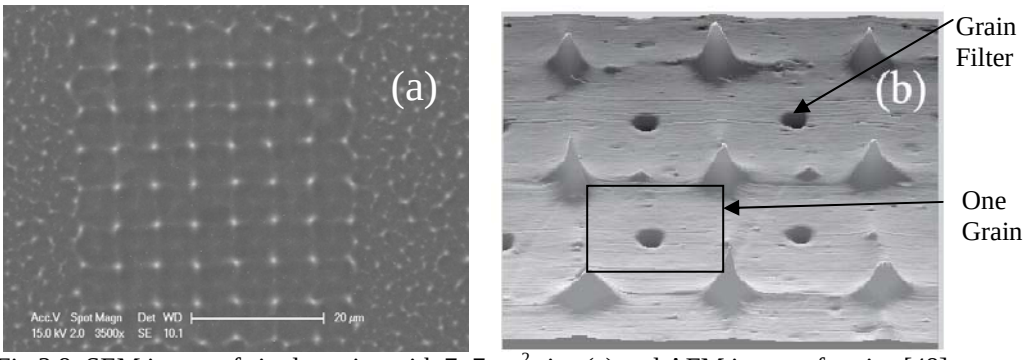


Fig.3.8: SEM image of single grains with 7×7µm^2 size (a) and AFM image of grains [48]

Because of large silicon grains and controlled location of grain filter by lithography, several thin film transistors can be fabricated inside of this grain. The quality of these devices is quite excellent and similar to SOI wafers. It means they can operate at lower supply voltages and lower power consumption, there is no latch up problem and also short channel effects are reduced. The mobility of nMOS and pMOS transistors is 600cm^2/VS and 150cm^2/VS, respectively. Fig.3.9 shows a schematic of TFT process [48].

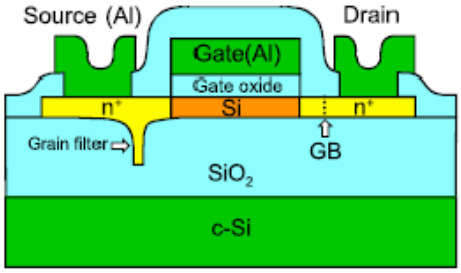


Fig.3.9: Using µ-Czochralski process to fabricate high performance TFTs with similar characteristics of SOI devices [48]

Table 3.1 compares the electrical characteristics of amorphous, poly and single grain TFTs [52].

Table 3.1: Comparison of various types of TFTs

	α-Si TFT	poly-Si TFT	SG-TFT
Field-effect mobility [cm^2/V·sec]	0.1–1	50–300	500–600
Off-current [pA]	< 1	~ 0.1	~ 0.1
Type of TFT	NMOS	P-/N-MOS	P-/N-MOS
Uniformity	Good	Poor	Good

The electrical properties of fabricated single grain thin film transistors are strongly dependent to the location of grain filter with respect to channel. Fig.3.10 shows four cases where crystallization direction can have 0^o (X position), 45^o (XY position) and 90^o (Y position) angles with respect to the current direction between source and drain. The fourth case is when channel is situated on top of grain filter (C position) [53].

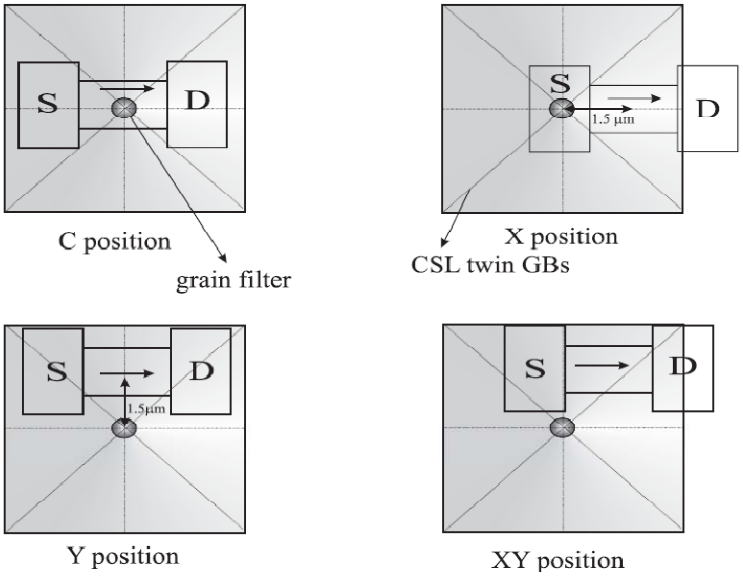


Fig.3.10: Schematic diagram of TFT channels at various positions inside a location-controlled single grain [53].

Table 3.2 illustrates the electrical properties of nMOS and pMOS devices fabricated on the mentioned locations. As it can be seen X position has highest mobility, lowest subthreshold slope (S), lowest leakage current (I_{OFF}) and C position is highest case for uniformity of mobility. The highest mobility for SGTFTs (μ_{nMOS}=597cm^2/VS and μ_{pMOS}=273cm^2/VS) at the X position is due to the fact that the carriers do not feel the coincidence site lattice (CSL) twin boundaries, which are parallel to the direction of current flow [53].

Table 3.2: Electrical Characteristics of nMOS and pMOS single grain thin film transistors (SGTFTs) in 4 positions

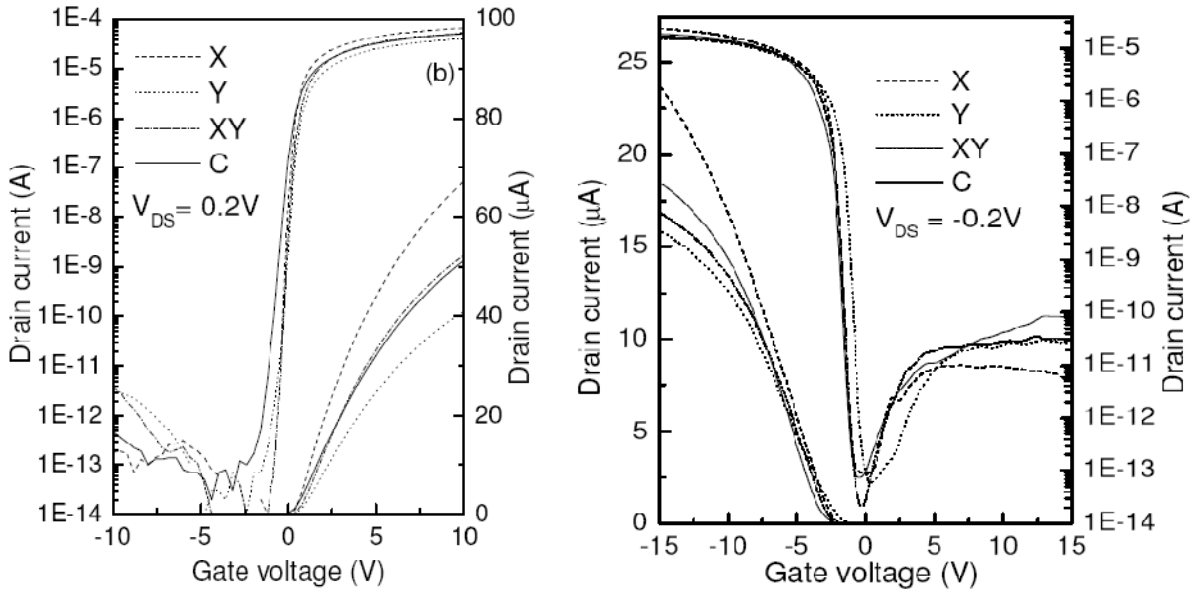
nMOS transistors [53]

TFTs	*Position*	$\mu_{FE,e}[cm^2/Vs](\sigma\%)$	$S[V/dec.](\sigma\%)$	$V_{TH}[V](\sigma\%)$
SG	*X*	597±101(17%)	0.21±0.03(13%)	1.7±0.2(11%)
	Y	528±57(10%)	0.25±0.04(14%)	1.8±0.3(15%)
	XY	505±55(7%)	0.22±0.01(6%)	1.9±0.1(6%)
	C	471±32(7%)	1.1±0.13(12%)	0.86±0.3(32%)
SOI		727±18(2.4%)	0.18±0.006(3.6%)	1.1±0.09(8%)

pMOS transistors[53]

Position	$\mu_{FE,h}[cm^2/Vs](\sigma\%)$	$S[V/dec.](\sigma\%)$	$V_{TH}[V](\sigma\%)$
X	273±27(11%)	0.14±0.04(24%)	$-2.2\pm0.38(17\%)$
Y	202±26(13%)	0.15±0.02(21%)	$-2.3\pm0.52(24\%)$
XY	219±44(20%)	0.18±0.02(8%)	$-3.36\pm0.58(23\%)$
C	228±22(10%)	0.16±0.03(16%)	$-3.3\pm0.64(25\%)$

Figure 3.11 shows I_{DS}-V_{GS} curves for nMOS and pMOS transistors in different positions in logarithmic and linear scales. The TFT made with SOI wafer is also plotted as a reference [53].



(a) nMOS (b) pMOS

Figure 3.11 shows I_{DS}-V_{GS} curves for nMOS and pMOS transistors in different positions in logarithmic and linear scales [53].

The high performance of SG Si TFTs makes them a suitable candidate for several applications [48,49,54]. Since this process is similar to SOI, it has all SOI advantages and enables it to use at high frequency and low power applications such as LNA [54]. Fig.3.12 shows an inverter characteristics fabricated with this process [49].
However, the reliability of SG Si TFTs would be an important issue for circuit applications [49].

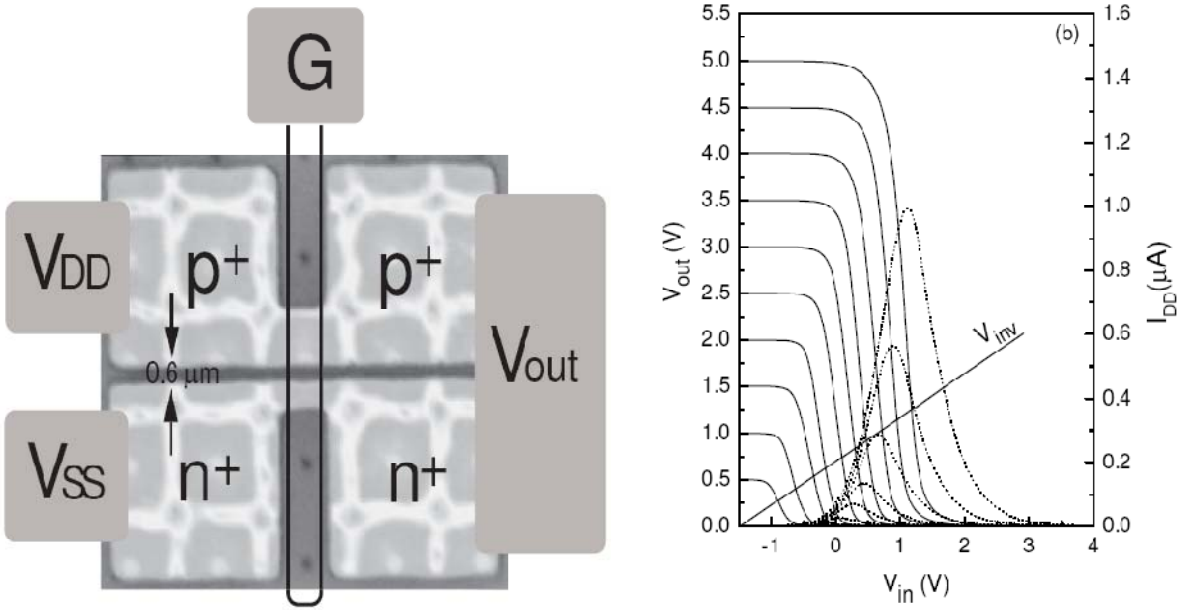


Fig.3.12: Fabricated inverter circuit using SGTFT [49]

3.3) Conclusions:

In this chapter we introduced 3DIC and its benefits in decreasing delay and area and solving interconnects problem that exist in planar IC technology. Heat management is the main issue involved in this process especially top most active devices have higher temperature. To solve this problem interconnect materials can be changed to act as interconnect and thermal path to the heat sink. Dummy heat paths can be designed to transfer heat. Several methods are being used to stack silicon layers to realize 3DIC. Wafer level and die level bonding are the most used methods to fabricate 3DIC. However, wafer thinning, alignment accuracy and fabrication of interconnections are main issues in these process.

µ-Czochralski process can offer high performance devices (similar to SOI) and circuits and monolithic stacking of silicon layers with high density of interconnects and without the mentioned problems that exist in wafer and die bonding methods. It is a low temperature process and can be used in low temperature applications.

Chapter 4

Designing 6T SRAM Cell and Sense Amplifier

Content:

4-1) SRAM Cells design

In this chapter we will discuss about designing 6T SRAM cells using analytical method. Then we will evaluate the design with circuit simulator. In this project Advanced Design System (ADS) software with modified BSIM-SOI model extracted from measurements of experimental single grain TFTs has been used.

4-1-1) Analytical Approach

As we discussed in chapter 2 designing SRAM cell needs a trade off and balance in transistor sizing to meet read, write and area constrains. Based on these three parameters (successful read and write and minimum area) we design SRAM cells and then we confirm and modify the design by DC and AC simulation.

In read operation the stored date in internal nodes is transferred to the BL and BLB which were precharged to Vcc. By enabling the word line (WL) cell will pull down one of the bitlines. As shown in Fig.4.1 "out" node has 0 data and will pull down BL to 0. The voltage of "out" node should not be more than Vth of Q2 or switching voltage of inverter with Q2 and Q4 transistors, plus some safety room for noise margin. Sizing of Q1 and Q5 should ensure non-destructive read.

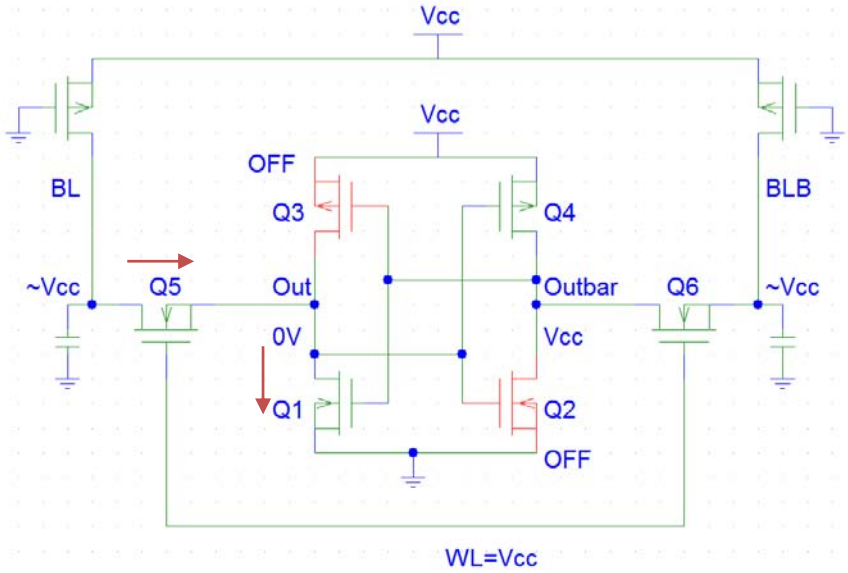


Fig.4.1: Read operation

In this circuit Q1 is in linear and Q5 is in saturation regions. Therefore we can calculate the increased voltage of "out" node (V_{out}) from current equivalent equation:

Q1 in linear region: $I_{Q1}(lin.) = \mu_n C_{ox} \frac{W_1}{L_1}\left(V_{cc} - V_{tn1} - \frac{V_{out}}{2}\right) V_{out}$

Q5 in saturation region: $I_{Q5}(Sat.) = \frac{1}{2}\mu_n C_{ox} \frac{W_5}{L_5}(V_{cc} - V_{out} - V_{tn5})^2$

$$\Big\} \quad I_{Q1} = I_{Q5}, V_{tn1} = V_{tn5}$$

Where μ_n , V_{tn} , W and L are mobility, threshold voltage, channel width and channel length of nMOS transistors. Vcc is supply voltage. In this design we used the values of our process

parameters where threshold voltage of pMOS (Vtp) was -1V, for nMOS (Vtn) was 0.5V, mobility of pMOS (μ_p) was 100cm^2/VS and for nMOS (μ_n) was 300cm^2/VS.

And we define cell ratio (CR or β): $\beta = CR = \frac{W1/L1}{W5/L5}$

Then we can calculate: $\Delta V = V_{out} = (V_{cc} - V_{tn}) \frac{1+CR \pm \sqrt{(CR(1+CR))}}{(1+CR)}$

Fig.4.2 shows the voltage of node "out" that should be less than Vth of nMOS drivers.

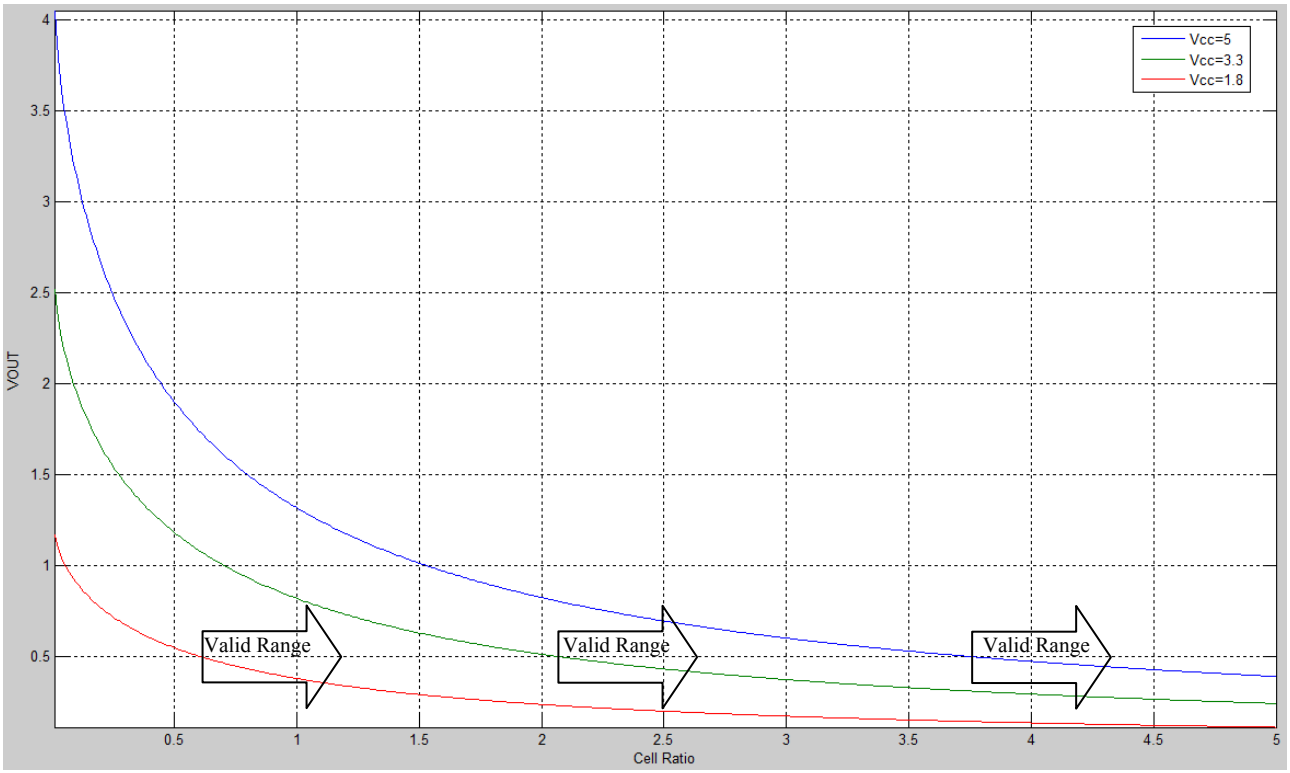


Fig.4.2: Vout versus CR for Vcc=1.8, 3.3, 5V

This curve describes the Vout versus CR for 1.8V, 3.3V and 5V supply voltage. It can be seen that for instance at Vcc=3.3V, CR should be more than 2.07 to have readable cell. It means the following relation should be held between access and driver transistors:

$$\mathrm{CR} = \frac{\mathrm{Wd}}{\mathrm{Wa}} = \frac{\mathrm{W1}}{\mathrm{W5}} > 2.07 \;\; at\; Vcc = 3.3V$$

On the other hand we should be able to write the cell. Suppose node out has 1 and node outbar has 0 stored data. In this case pull up transistor Q3 is on and driver transistor Q1 is off. We want to write 0 on node out. During writing, BL in Figure 4.3, is driven from precharged value (Vcc) to the ground potential (0 state) and then by enabling WL, through access transistor Q5 we change the value of node V_{out}. If transistors Q3 and Q5 are properly sized, then the cell is flipped and its data is effectively overwritten. It means V_{out} should be less than V_{th2} to turn off Q2.

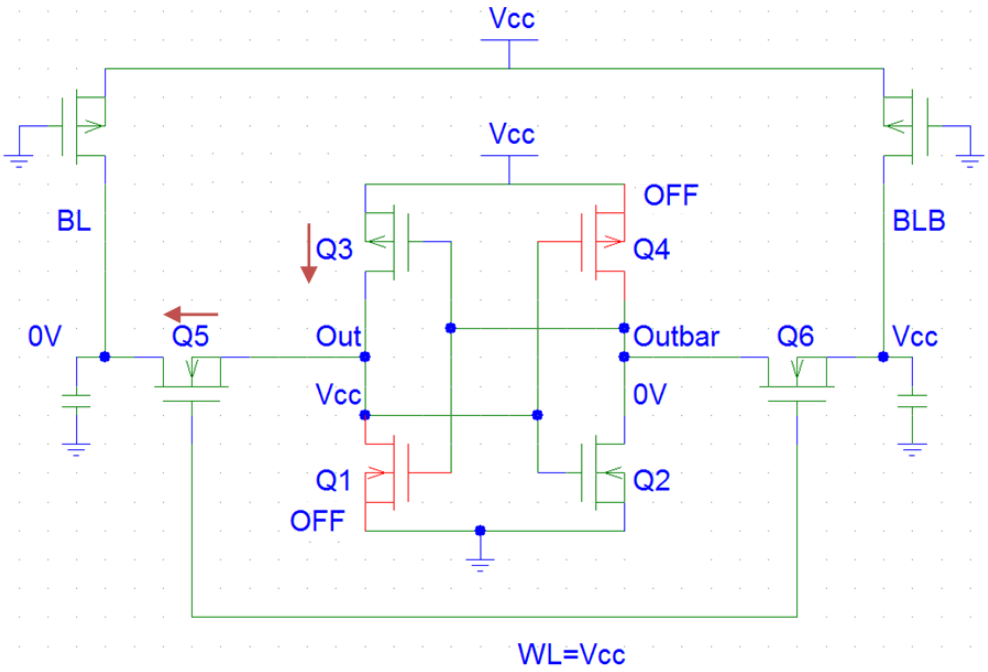


Fig.4.3: Write operation

As it can be seen in Fig. 4.3 Q5 is in saturation and Q3 is in linear region. The current of Q3 and Q5 are same and we can calculate the Vout voltage from equal current equation.

$$V_{out} = (V_{cc} - V_{tn}) - \sqrt{((V_{cc} - V_{tn})^2 - (\frac{\mu_p}{\mu_n})(\mathrm{PR})((V_{cc} - V_{tn} - |V_{tp}|)^2)}$$

Where μ_p and V_{tp} are mobility and threshold voltage of pMOS transistor.

And we define Pull up ratio (PR): $PR = \frac{W3/L3}{W5/L5}$

Fig.4.4 shows the Vout versus PR for Vcc=1.8, 3.3 and 5V. As it can be seen PR should be less than 2.36 in case of Vcc=3.3V. It means we should have the following relation between transistors size:

$$\mathrm{PR} = \frac{\mathrm{Wp}}{\mathrm{Wa}} = \frac{\mathrm{W3}}{\mathrm{W5}} < 2.36 \ \ at\ Vcc = 3.3V$$

From this value for PR and CR value we can see that there is a trade off in transistor sizing. Beside these values for CR and PR the area of cell should be kept as minimum as possible. Furthermore, if we design the cells for Vcc=5V those cells can successfully work at lower voltages too.

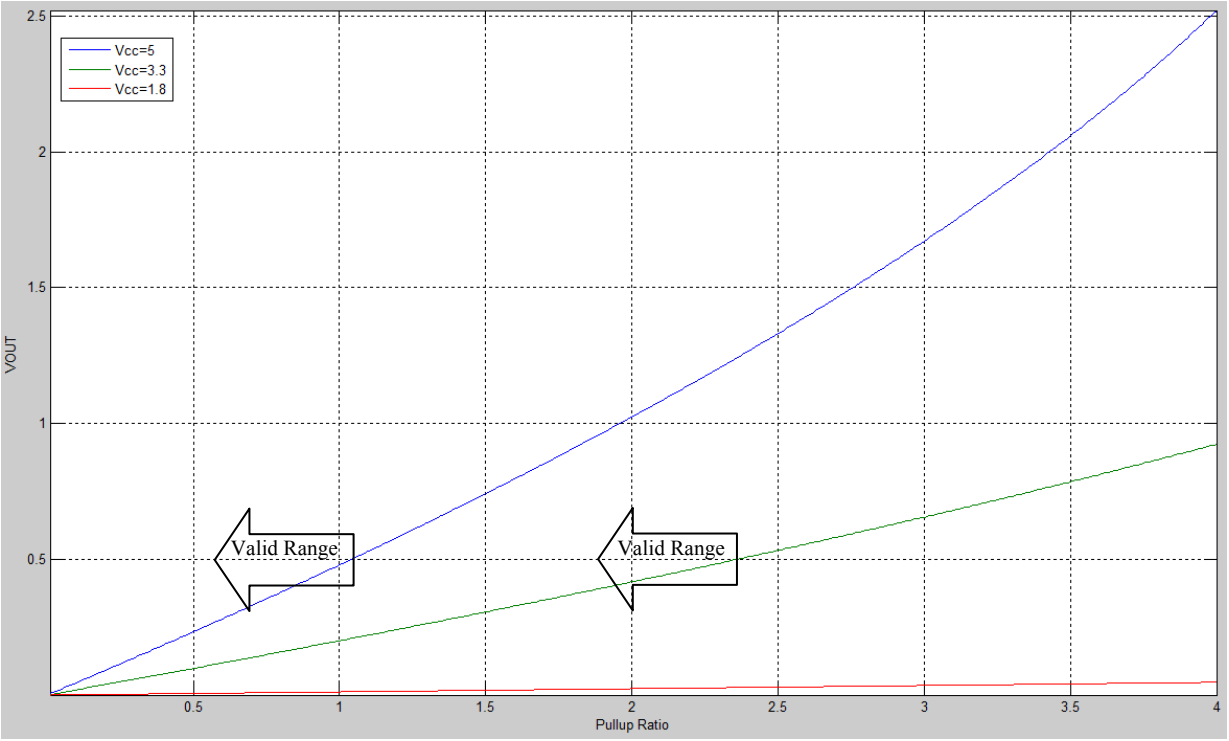


Fig.4.4: Vout versus PR for Vcc=1.8, 3.3, 5V

The calculated values for CR and PR at different Vcc are shown in table 4.1.

Table 4.1: The values of CR and PR at different Vcc

Vcc	5	3.3	1.8
CR	3.76	2.07	0.61
PR	1.04	2.36	35
Vswitching	1.3	1.18	0.744

In both calculations for CR and PR we compared V_{out} with threshold voltage of nMOS driver transistors (V_{th} of Q1 and Q2). However, sometimes the switching voltage of an inverter can be less than V_{thn}. Therefore we should confirm that at specific transistor sizing, switching voltage is large enough to take V_{thn} as a lowest value to compare with V_{out}.
The switching voltage of an inverter is given by this equation:

$$Vswitching = \frac{\sqrt{\frac{\mu_p W_p}{\mu_n W_n}}\left(V_{cc} - |V_{tp}|\right) + V_{tn}}{1 + \sqrt{\frac{\mu_p W_p}{\mu_n W_n}}}$$

Where μ_p, μ_n and V_{tp}, V_{tn} and W_p, W_n are mobility, threshold voltage and channel width of pMOS and nMOS transistors, respectively. In this equation the length of the transistors are assumed to be same. The calculated values for switching voltage are shown in table 4.1. As we can see the switching voltages are larger than V_{thn}.

We discussed that there is a trade off in transistors sizing in read and write operation modes. On the one hand, if we make access transistors strong (large W_{access}) we will have good write margin and bad read margin but read current is higher that results higher speed for cell. If we make it weak (small W_{access}), read stability improves but write margin and read current will be poor. On the other hand for higher CR and lower PR we have to increase the size of nMOS drivers and decrease the size of load pMOS transistors. We can select the pMOS transistors as low as possible (equal to gate length L) to improve PR. But we can not increase the size of nMOS drivers as we want to keep cell area as minimum as possible.
We have selected various widths of transistors to meet the CR and PR requirements at different supply voltages. We will investigate these selections using DC and AC simulation in the coming section.
We saw in previous chapter one solution is using double gate transistors for access. With biasing back gate with different voltages during read and write we can change threshold voltage of access transistors to improve both read and write margins. If we do not take threshold voltage of nMOS drivers equal to threshold voltage of nMOS access transistors then we will have following relation for Vout during read operation:

$$(1+CR){V_{out}}^2 - 2V_{out}[V_{cc} - V_{tna} + CR(V_{cc} - V_{tn1})] + (V_{cc} - V_{tna})^2 = 0$$

$$a = a + CR \ , \quad b = V_{cc} - V_{tna} + CR(V_{cc} - V_{tn1}) \quad , c = (V_{cc} - V_{tna})^2$$

$$V_{out} = \frac{b - \sqrt{b^2 - ac}}{a}$$

And the following equation for write operation:

$$V_{out} = (V_{cc} - V_{tna}) - \sqrt{((V_{cc} - V_{tna})^2 - (\frac{\mu_p}{\mu_n})(\mathrm{PR})((V_{cc} - V_{tna} - |V_{tp}|)^2)}$$

Where V_{tna} is the threshold voltage of access transistors. Fig.4.5 and 4.6 show the calculated Vout for both read and write operations. As we can see with increasing Vth of access transistors CR improves that means we can use small size of transistor for driver or means we have good read margin.

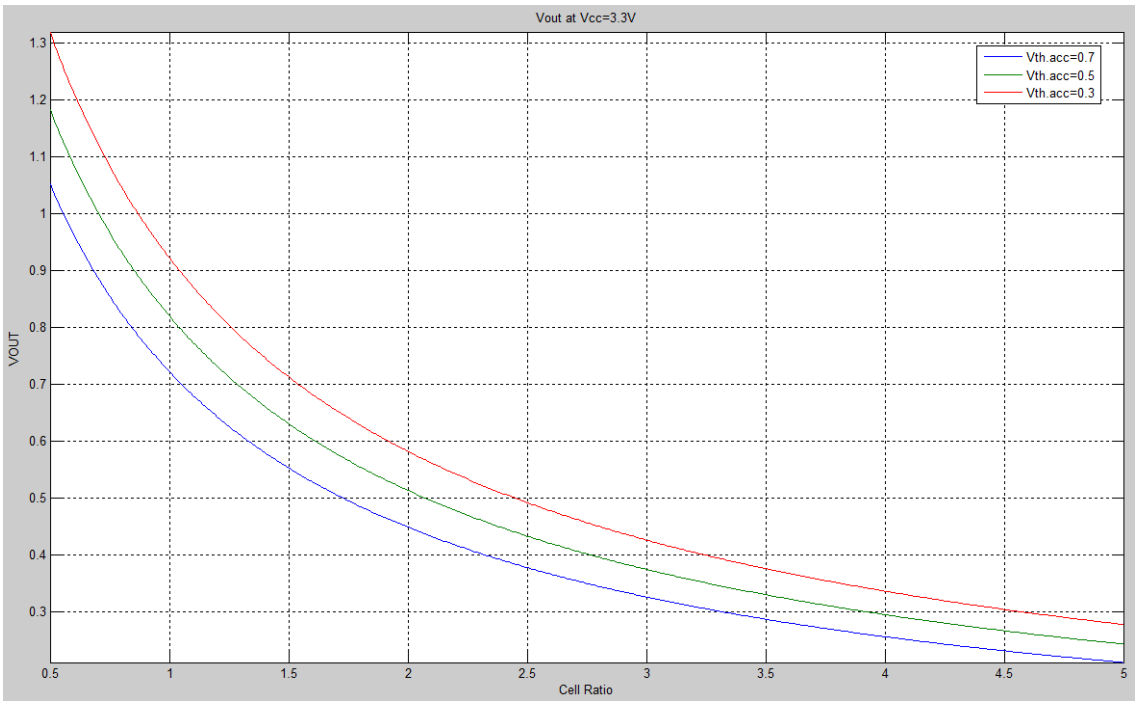


Fig.4.5: Vout during read for Vth of access transistors equal to 0.3, 0.5 and 0.7V. Higher Vth improves CR.

However, as it can be seen from Fig. 4.6 having low value for threshold voltage of access transistors is better for PR. It improves PR and write margin.

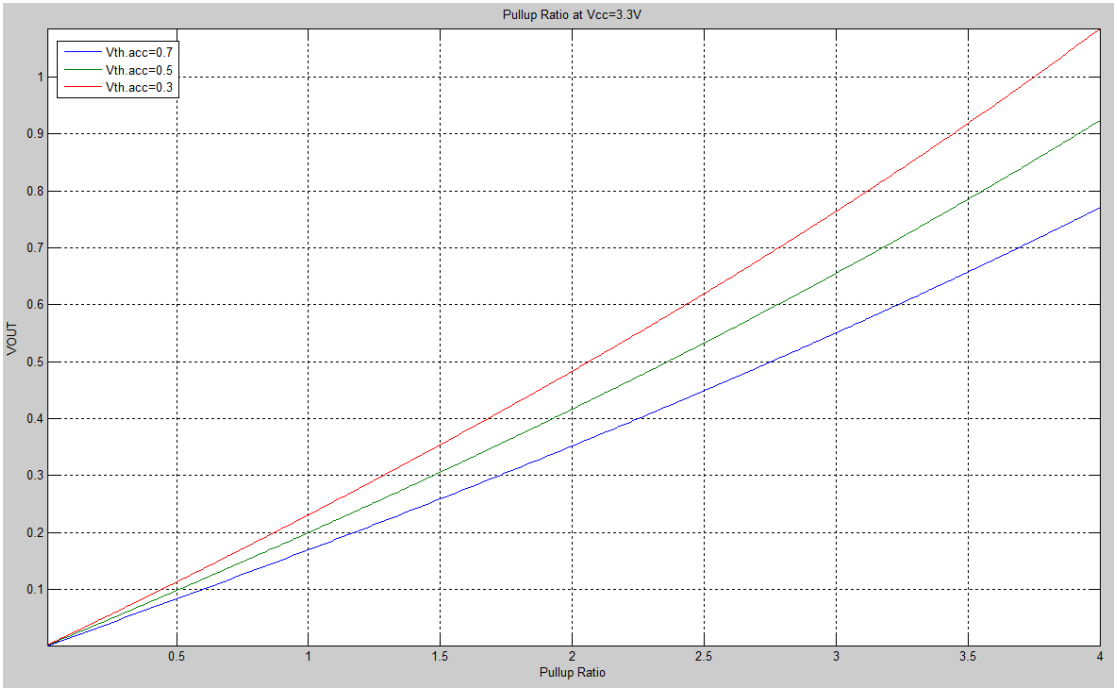


Fig.4.6: Vout during write for Vth of access transistors equal to 0.3, 0.5 and 0.7V. Lower Vth improves PR.

Table 4.2 shows the calculated values for CR and PR for Vth of access transistors equal to 0.3, 0.5 and 0.7V. This table shows with increasing 0.2V threshold voltage of access transistors CR

improves 16.42% and with decreasing 0.2V threshold voltage of access transistors PR improves 12.71%. These improvements are significant values for read and write noise margins without area penalty. Hence using double gate or H-gate transistors will improve the both noise margins certainly.

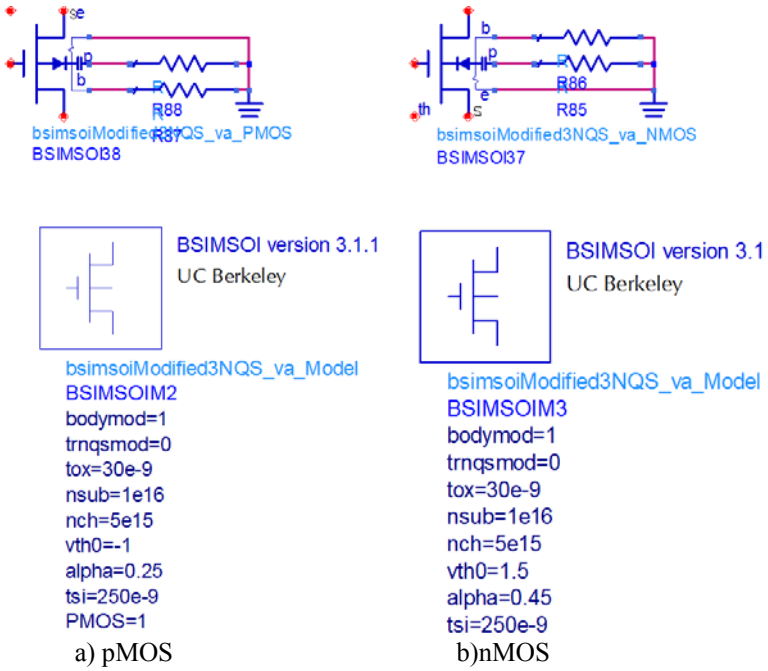
Table 4.2: Calculated CR and PR at different threshold voltage of access transistors

Vth.acc	0.3	0.5	0.7
CR	2.45	2.07	1.73
PR	2.06	2.36	2.75
Vswitching	1.12	1.18	1.25

4-1-2) DC Simulation

After selecting the values for transistors sizing from calculation we measure the hold, read and write SNM values from DC simulation. Always hold margin is bigger than read and write margins. Advanced design system (ADS) software is used to simulate the cells. We have used a modified BSIM-SOI model extracted from measurements of experimental single grain TFTs. The parameters of this technology have been fitted to the experimental thin film transistors. In this design we assumed that threshold voltage of pMOS (V_{tp}) is -1V, for nMOS (V_{tn}) is 0.5V, mobility of pMOS (μ_p) was 100cm^2/VS and for nMOS (μ_n) was 300cm^2/VS.

Fig.4.7 shows the model of pMOS and nMOS transistors and their related parameters. The resistors in the model are for modeling the effect of body in fully depleted SOI transistors. The assigned value is 50MΩ.



a) pMOS b)nMOS

Fig.4.7: The model of pMOS and nMOS transistors and their parameters.

If we look at the parameters in the model there are several parameters to show the physical process values. The thickness of gate oxide (tox) is 30nm, substrate and channel doping (nsub and nch) are 1×10^{16} and 5×10^{15}/cm^3 and silicon thickness is 250nm. To model the effect of grain boundaries “alpha” parameter is used. When alpha is near 1 it means we have poly and when it is

close to 0 we have single grain. The mentioned values for alpha will result the mobility of $100cm^2/VS$ for pMOS and $300cm^2/VS$ for nMOS transistors. Another parameter in the model is "vth0" that is related to threshold voltage of transistors. For nMOS it is 1.5 and equals to 0.5V threshold voltage and for pMOS it is -1 that equals with -1V threshold voltage. Finally "trnqsmod" is a parameter to determine transient or DC/AC simulation. When it is 1 the modified parameters are for transient simulation. When it is 0 the simulation is valid only in DC or AC simulations.

Fig.4.8 shows the schematic of circuit to measure the read SNM. As it can be seen in order to measure it WL and both BL and BLB are connected to Vcc and a variable voltage source is applied to "Out" node and then the voltage of "Outbar" node is monitored.

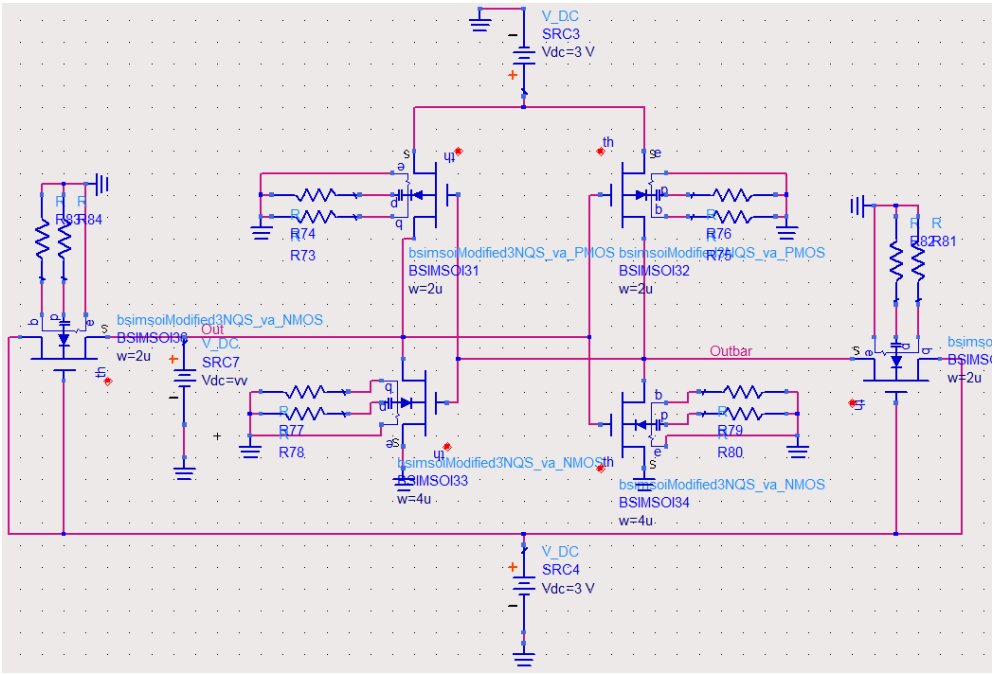


Fig.4.8: Schematic of circuit to measure read SNM.

In this circuit the width of access transistors are 2µm and nMOS drivers are 4µm and load (pull up) pMOS are 2µm. Therefore we have PR=1 and CR=2. The result of simulation at 3V supply voltage is shown in Fig.4.9. In this simulation we sweep the voltage source at "Out" node and we look to the "Outbar" voltage. Then "Outbar" is mirrored and is drawn in same curve to obtain butterfly curve. The maximum square that we can fit in this curve is 0.5V for read and 1V for hold SNM. Hence the read SNM is 0.5V and hold SNM is 1V. In hold mode WL and both BL and BLB are connected to GND. Always hold SNM is more than read SNM.

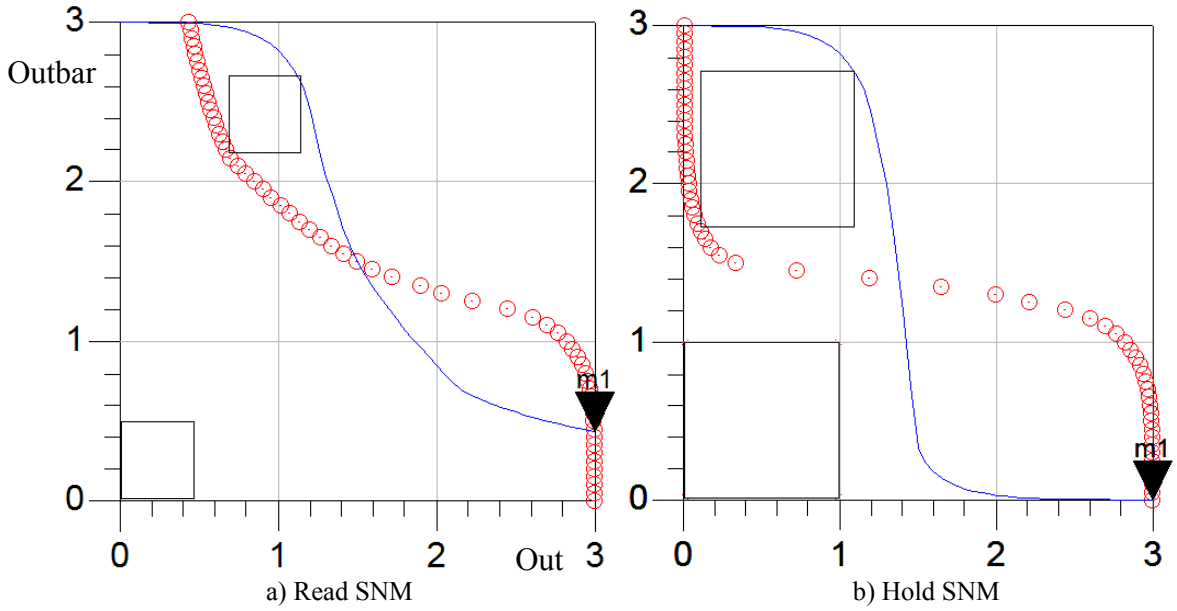


Fig.4.9: The simulated SNM for Wa=2µm, Wd=4µm and Wp=2µm. Read SNM is 0.5V and hold SNM is 1V at Vcc=3V.

To simulate the write SNM one of the BL or BLB should be connected to GND and the other one to Vcc. WL is enabled. As shown in Fig. 4.10, suppose BL is connected to Vcc and BLB is connected to GND. Then we put a sweep voltage source on the "Out" node and we measure the voltage of "Outbar" point. This write curve ("Outbar" versus "Out") is added to the read SNM butterfly curve.

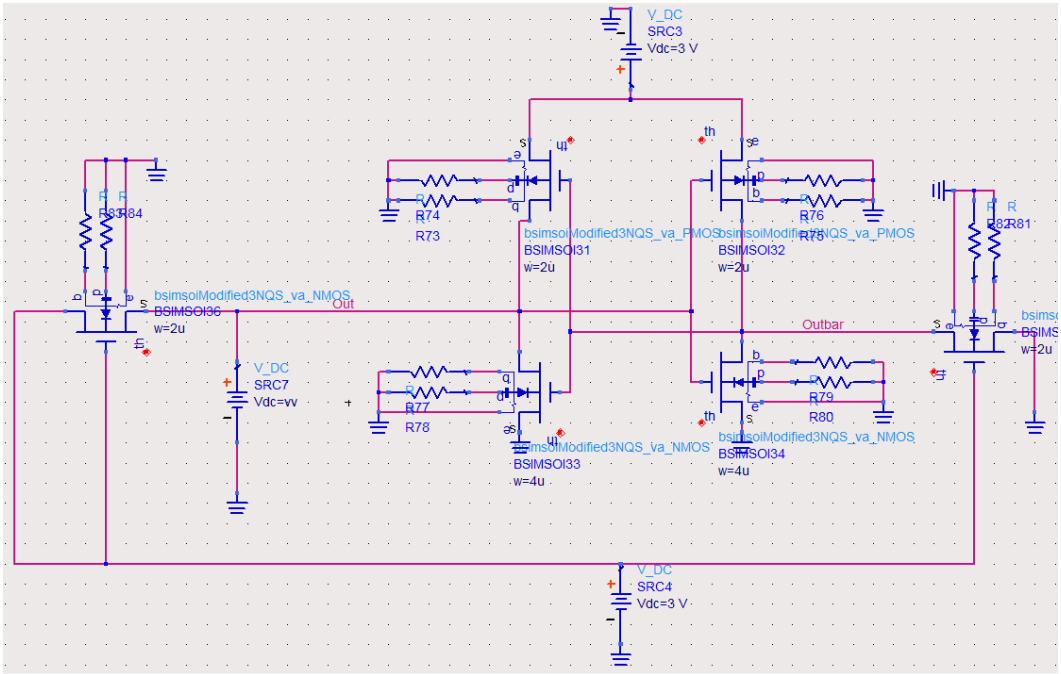


Fig.4.10: Circuit schematics to simulate write SNM. One of the BL or BLB is connected to GND.

The result of final curve will be same as one as shown in Fig. 4.11. The butterfly curve is for read and the other one comes from write operation mode. With fitting a minimum square between read and write curves, write SNM can be determined. For the mentioned transistors sizing write SNM is 1.1V as shown in Fig.4.11.

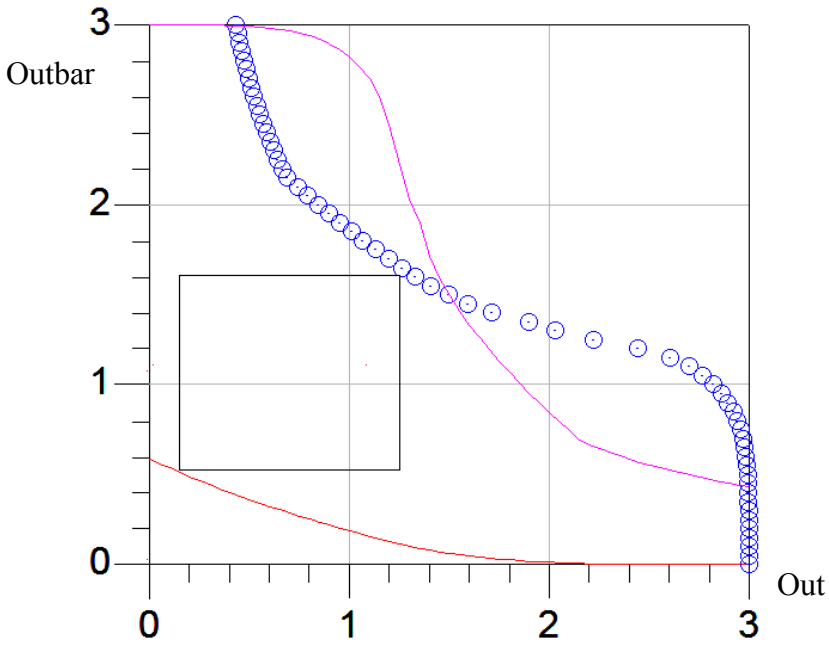


Fig.4.11: The simulated write SNM for Wa=2µm, Wd=4µm and Wp=2µm. Write SNM is 1.1V at Vcc=3V.

Fig. 4.12 shows two cases of simulation that in one case cell is not writable and in the other case cell has low value for write SNM. In the left cell the access transistors have 4µm and drivers have 24µm and pull up transistors have 16µm width. At 3.3V supply voltage it is not possible to write it. Read SNM is 0.7 in this case. In the right cell the access transistors have 4µm and drivers have 9µm and pull up transistors have 16µm width. At 3.3V supply voltage write SNM is 0.2 and read SNM is 0.6 in this cell.

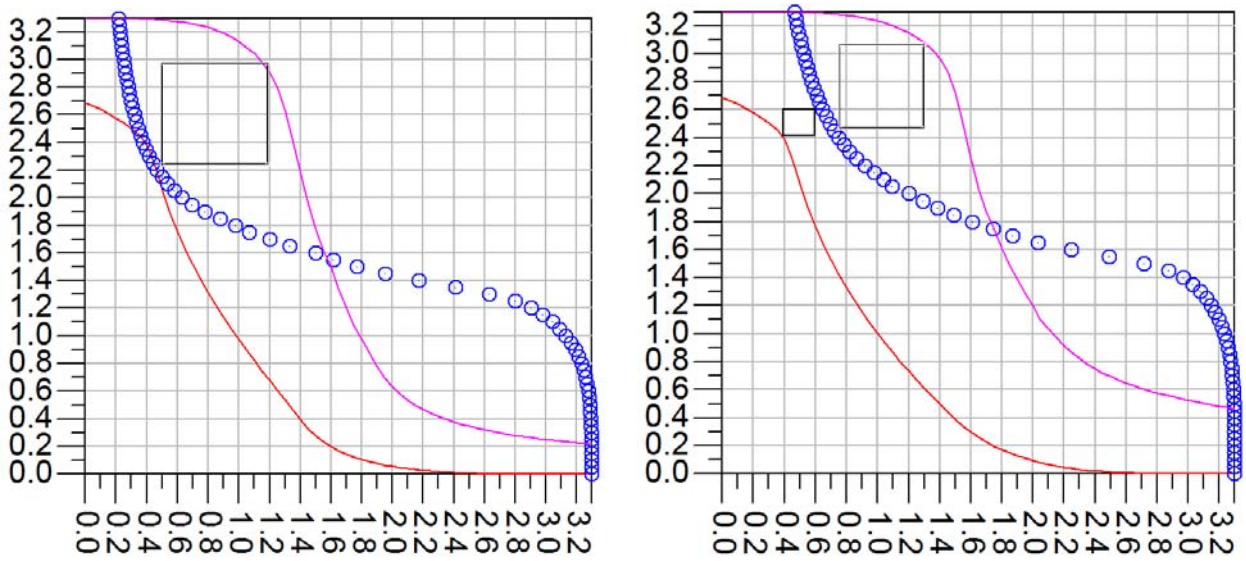


Fig.4.12: Two cases that cell is not writable and small write margin.

The result of simulations for selected designs is shown in table 4.3. The supply voltage is 3.3V. These values for width of transistors are used in designing layout for fabrication. We put non-writable cells to see the characteristics of cell in practice for comparing purposes. The length of transistors is 2µm. Furthermore, these designed values and resulted static noise margins are resistance with process variation. As it can be seen in this table the simulation results show higher values for SNM that is an evident for free process variation design. Later in characterization chapter we will see the practical vales.

Table 4.3: Designed width of transistors and simulation result for read and write SNM. Vcc=3.3V

W-access	W-nMOS (Drivers)	W-pMOS (Pull Up)	CR	PR	Read SNM	Write SNM
2	2	2	1	1	0.3	1.3
2	4	2	2	1	0.4	1.1
2	4	3	2	1.5	0.45	0.9
2	2	10	1	5	0.45	0.3
2.6	5.4	8	2	3	0.5	0.4
2.6	8	5.4	3	2	0.5	0.6
2.6	6.6	5.4	2.5	2	0.5	0.65
4.4	5.4	5.4	1.2	1.22	0.35	1.1
4.4	5.4	16	1.2	3.6	0.5	0.4
2.6	10.8	9.6	4.15	3.69	0.6	0.1
2.6	10.8	6.6	4.15	2.54	0.6	0.4
2.6	13.4	9.4	5.15	3.6	0.6	0.1
2.6	10.6	4	4	1.5	0.5	0.8
2.6	10.6	5.4	4	2	0.55	0.7
4	10.6	9.4	2.65	2.35	0.6	0.6
4	13.4	9.4	3.35	2.35	0.6	0.5
2.6	24	8	9.23	3	0.7	0.1
2.6	24	16	9.23	6.15	0.8	Not writable
4	24	16	6	4	0.7	Not writable
4	9	16	2.25	4	0.6	0.2

4-1-3) Transient Simulation

After finding suitable transistors sizing we need to confirm it with transient simulation. Unfortunately the transient simulation in our model is not completed yet. It means the model is not valid for transient simulation. However, just to show the approach we did a simulation. Fig.4.13 shows the schematics of the circuit. It consists of write drivers and cell. In this circuit there is a transistor as a driver to let the signal to the BL during write. Precharge transistors can

be nMOS or pMOS connected to Vcc/2. In nMOS case gate is connected to drain that has Vcc/2 voltage and always is in saturation region. To keep it on the voltage of nodes BL and BLB should be less than $V_{cc}/2-V_{thn}$. In pMOS case source is connected to GND therefore it is always on in linear region and BL and BLB voltages can go to $V_{cc}/2$.

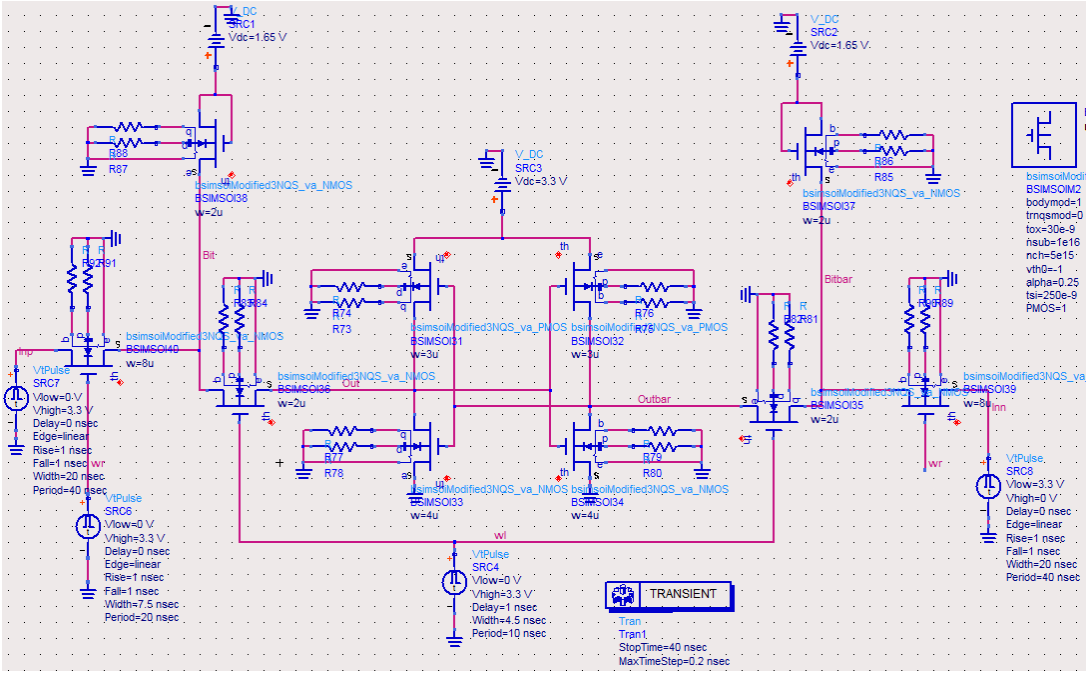


Fig.4.13: Schematics of circuit for transient simulation

The result of simulation is shown in Fig.4.14. The right signals in this figure are WL, Wr (write), input signal (Inp or Inn), BL and Out. As it can be seen data can be written and read successfully. In the left hand of figure measured values for write and read access time are shown. It can be seen that write and read access times are approximately 7ns and 5ns. It is good to mention that read access time is the time between WL and BL and write access time is time between WL and "Out" node.

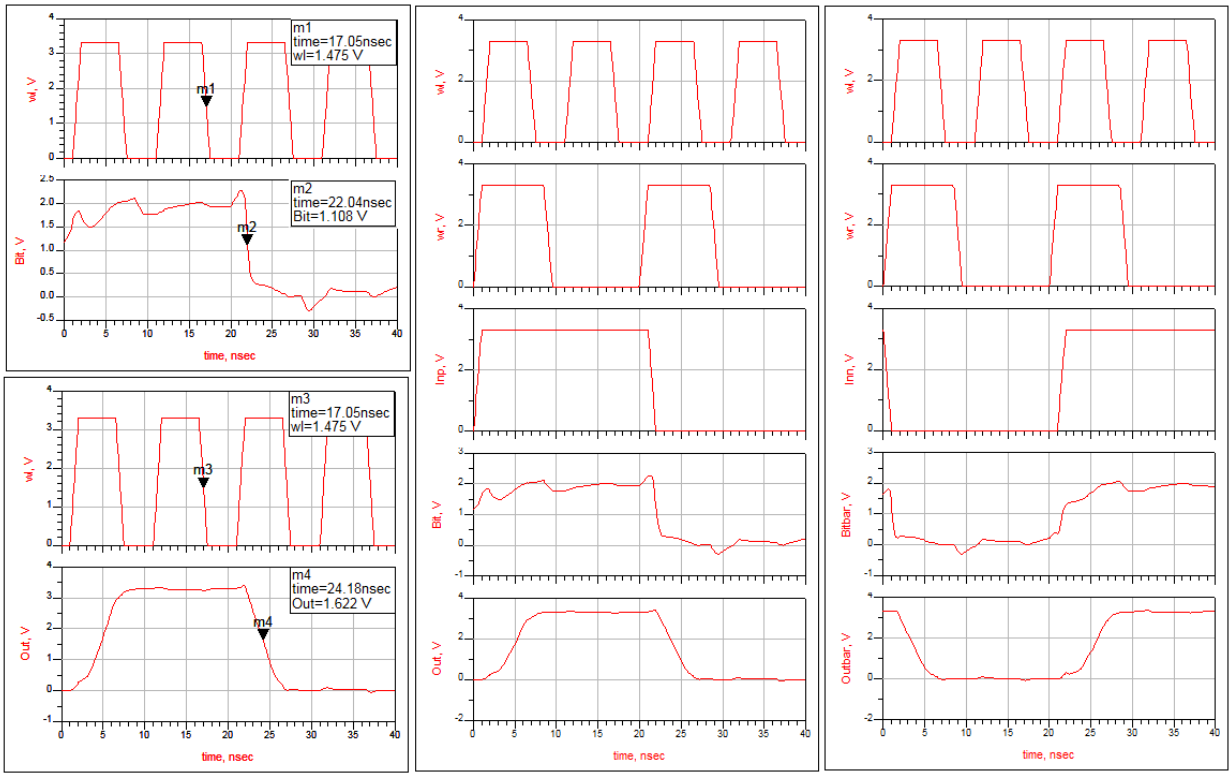


Fig.4.14: The result of transient simulation. Read access time is 5ns and write access time is 7ns.

4-2) Using Double-Gate and H-Gate Transistors to improve SNM and WNM

As we mentioned there is a trade off in choosing the size of transistors to have large SNM for write and read. If we want to have small area for SRAM cells, it is not possible to have large margins for both read and write. From equations for pull up ratio (PR) and cell ratio (CR) we know that the width of access transistors is in both parameters. If we make access transistors large, we will have good PR (or good write SNM) but poor CR (or poor read SNM). One solution is using double gate or H-gate transistors to dynamically change the threshold voltage of access transistors during read and write. Fig.4.15 shows a simulation result to see the effect of threshold voltage of access transistors on SNM value. As it can be seen, with increasing threshold voltage read SNM improves but write SNM degraded. At V_{th}=1.5V cell is almost not writable. At lower threshold voltage write SNM improves. This simulation suggests dynamically changing V_{th} during read and write. For write, lower V_{th} and for read, higher V_{th} are better. This can be done by double gate transistors. In these transistors the biasing of bottom gate will modulate the threshold voltage of top gate. Moreover, it will improve subthreshold slope (S) and leakage current of transistor. In this SRAM cells area is same as normal layout. The only drawback of this transistor is extra fabrication process steps and also need for second WL for bottom gate.

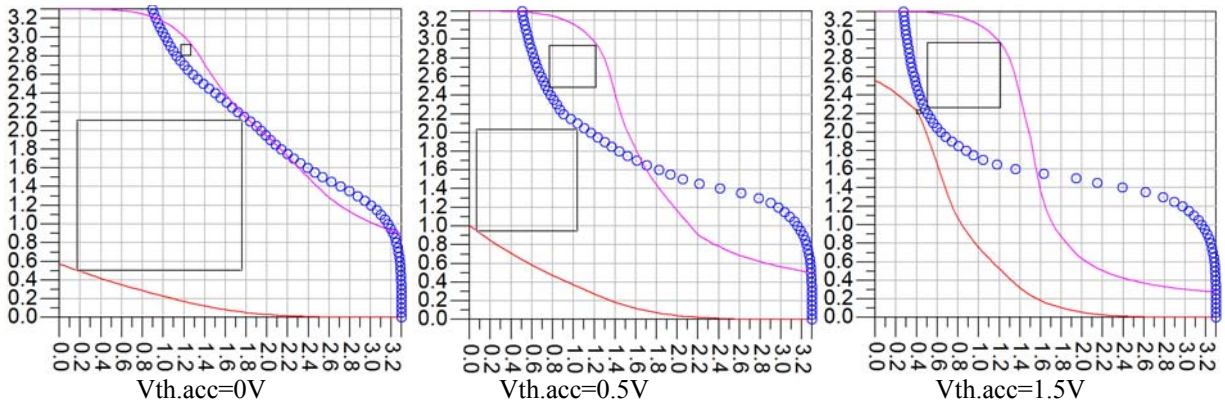


Fig.4.15: Influence of threshold voltage of access transistors on write and read SNM

4-3) Designing Sense Amplifier

Sense amplifiers, in association with memory cells, are key elements in defining the performance and environmental tolerance of CMOS memories. Because of their great importance in memory designs, sense amplifiers became a very large circuit-class. Fig. 4.16 shows the schematics of designed voltage sense amplifier. This sense amplifier is a voltage differential amplifier that amplifies the signals and rejects the common mode signals. To increase the gain of this amplifier one more stage is added. The sizes of transistors are shown in the schematics.

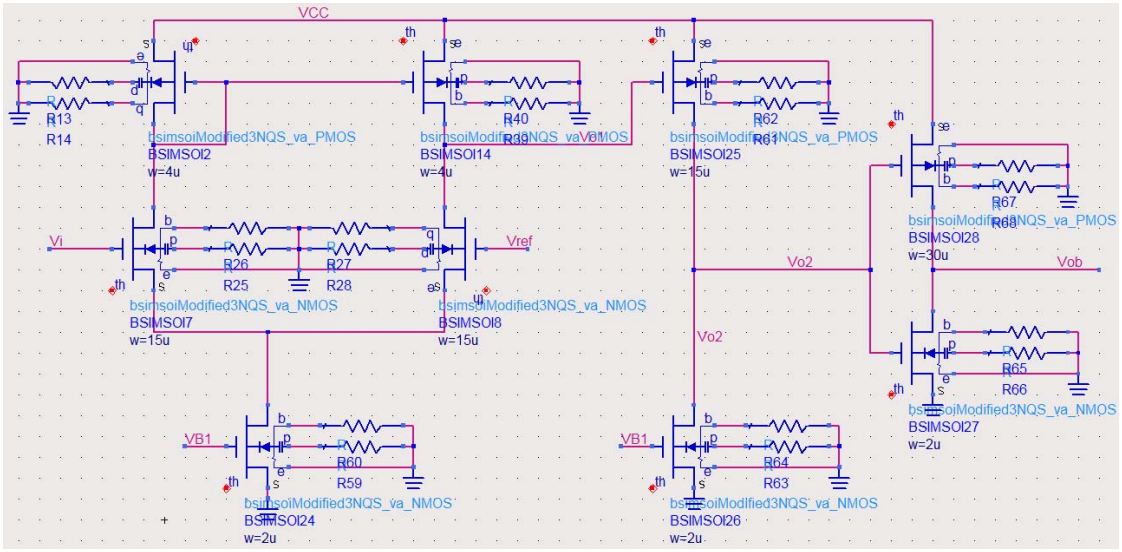


Fig.4.16: The designed sense amplifier for SRAM cells

Fig.4.17 shows ac response of this amplifier. This amplifier has 65dB DC gain and 160MHz unity gain band width. The delay of this sense amplifier is approximately 1.76µs. In order to drive the load of measurement system, output buffers are used.

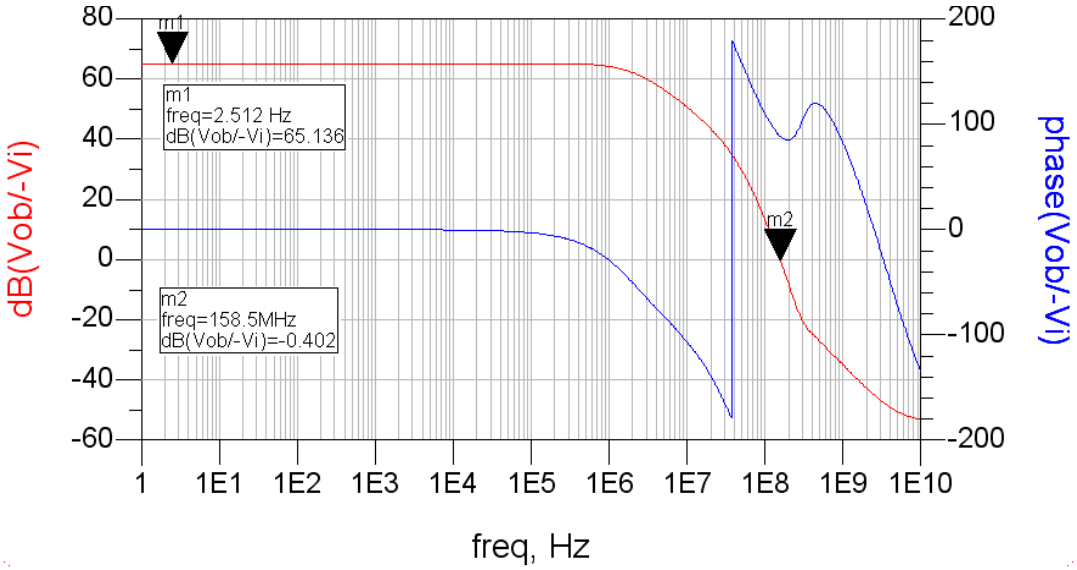


Fig.4.17: Ac response of sense amplifier

4-4) Designing Output Buffers

To measure the transient response of SRAM cells we need output buffers to drive the load of measurement system. The idea is shown in Fig.4.18. The load is consist of a 1MΩ resistor and 25pF capacitor. We use same rules of VLSI for output buffers to have minimum delay. The input capacitance of our inverters is typically in the order of 0.1pF. Putting one inverter with large size can drive the load but it will have long delay. Inverter chain is used to make minimum delay. We should calculate the number and size of inverters.

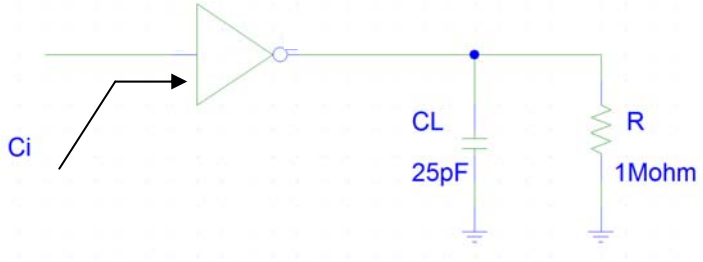


Fig.4.18: Measurement system load

The delay of a buffer chain is given by this equation:

$$tp = N \times tp0\left(1 + \frac{\sqrt[N]{F}}{\gamma}\right)$$

Where in this formula t_p is delay, $\gamma \approx 1$, $F = \frac{CL}{Ci}$, N is the number of inverters and t_{p0} is the delay of standard inverter. The size of inverters is increased with multiply factor of $f = \sqrt[N]{F}$. In our case we assume that C_{BL}=0.1pF and C_L=25pF. Then F=250 and with taking N=2 minimum delay can be obtained. Therefore the multiply factor will be f=16. Fig.4.19 shows the final inverter sizing. With Wn=2µm and Wp=10µm inverter is symmetric.

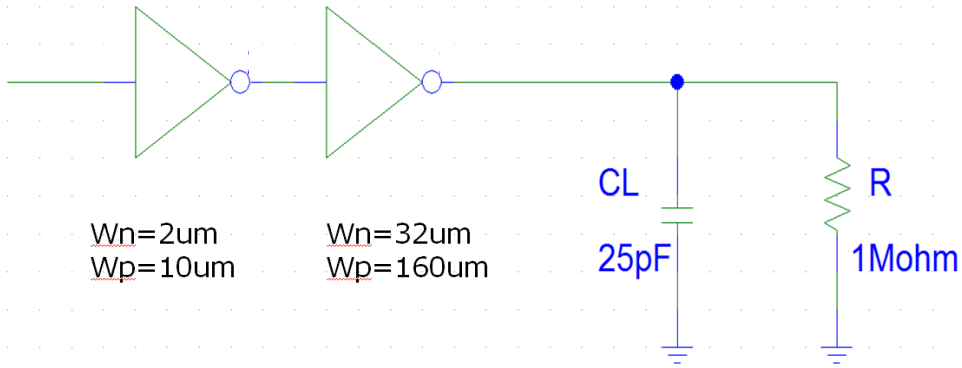


Fig.4.19: The size of output buffer

4-5) Conclusions

To sum up in this chapter we designed the SRAM cells by analytical and simulation approaches to have good read and write SNM cells. Various transistor sizing are considered to compare their characteristics. Double gate transistors can improve both read and write margins. Sense amplifier is designed to have minimum delay and high common mode rejection ratio. Finally output buffers are designed to drive 25pF load of measurement system.

Chapter 5

6T SRAM Layout Design in One Layer and Two Layers of Single Grain Silicon

Content:

5-1) Layout design rules

In designing layout using single grain technology (same as FDSOI technology) there are several design rules. Design rules and layers are exactly same SOI rules. The only difference is that we have an extra mask layer called grain filter (GF). We discussed about grain filter and their locations relative to gate and island masks in previous chapters. Here are the rules for etch layer:

5-1-1) Grain Filter:

Rule Index	Description	Value (µm)
2.1	Grain Pitch	6×6
2.2	Grain Size	1×1

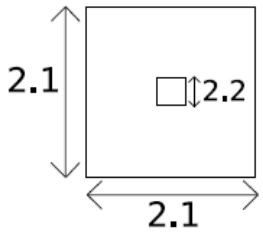


Fig.5.1: DRC for Grain Filter

5-1-2) Active layer:

Rule Index	Description	Value (µm)
3.1	Minimum width of active area	1.5
3.2	Spacing between the active areas	0.6
3.3	Minimum length of source/drain	3.5
3.4	Spacing between different implantations	0.6

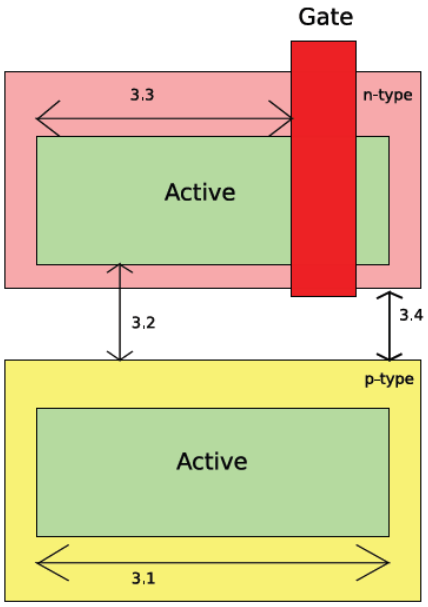


Fig.5.2: DRC for Active layer

5-1-3) Gate:

Rule Index	Description	Value (µm)
4.1	Minimum width of gate	1.5
4.2	Minimum spacing over active region	2
4.3	Minimum gate extension over active area	1
4.4	Minimum active extension of gate	0.6
4.5	Minimum spacing between gate and active area	0.6

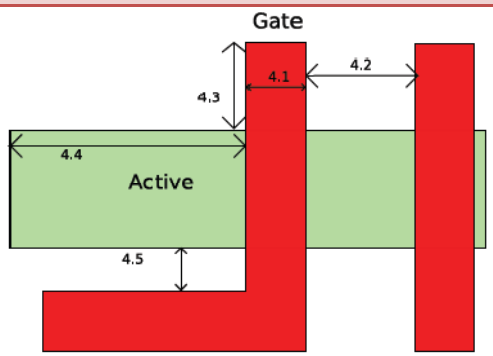


Fig.5.3: DRC for Gate layer

5-1-4) Contact:

Rule Index	Description	Value (µm)
5.1	Minimum contact hole dimension	2.1×2.1
5.2	Minimum VIA	1.5×1.5
5.3	Minimum contact and /or VIA spacing	1
5.4	Minimum spacing to gate	0.9
5.5	Minimum overlap of the active area	0

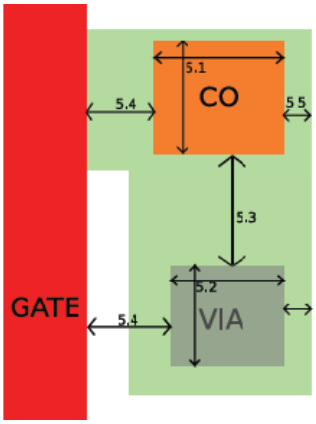


Fig.5.4: DRC for Contact layer

5-1-5) Metal:

Rule Index	Description	Value (µm)
6.1	Minimum metal separation	1
6.2	Minimum metal VIA/contact overlap	0.3
6.3	Minimum metal width	1

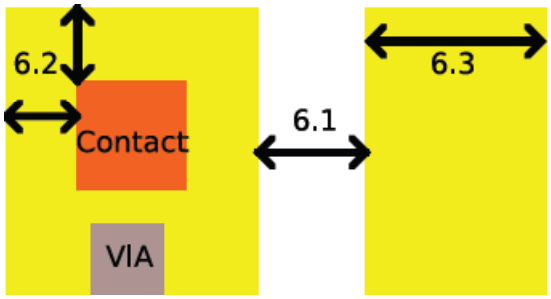
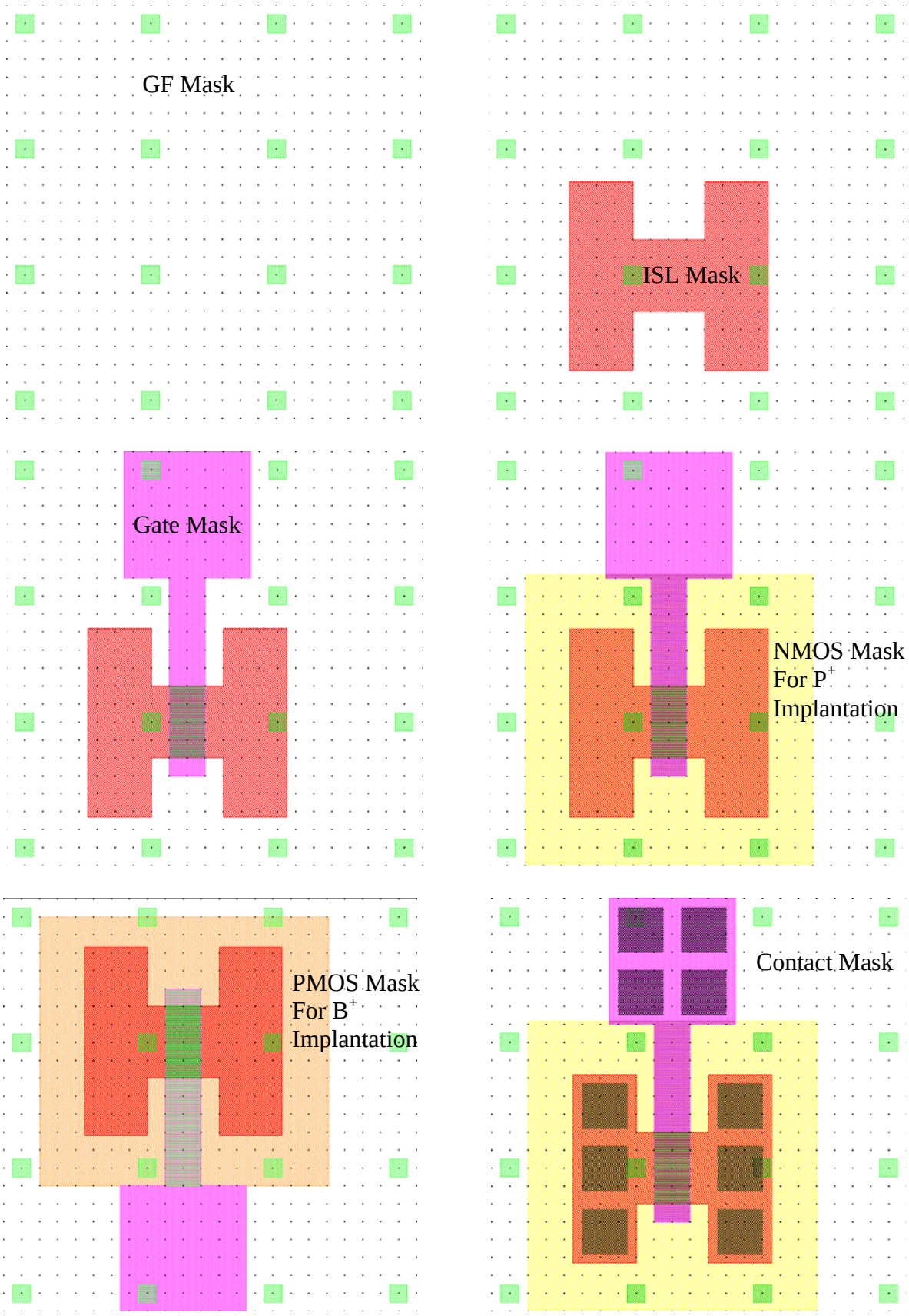


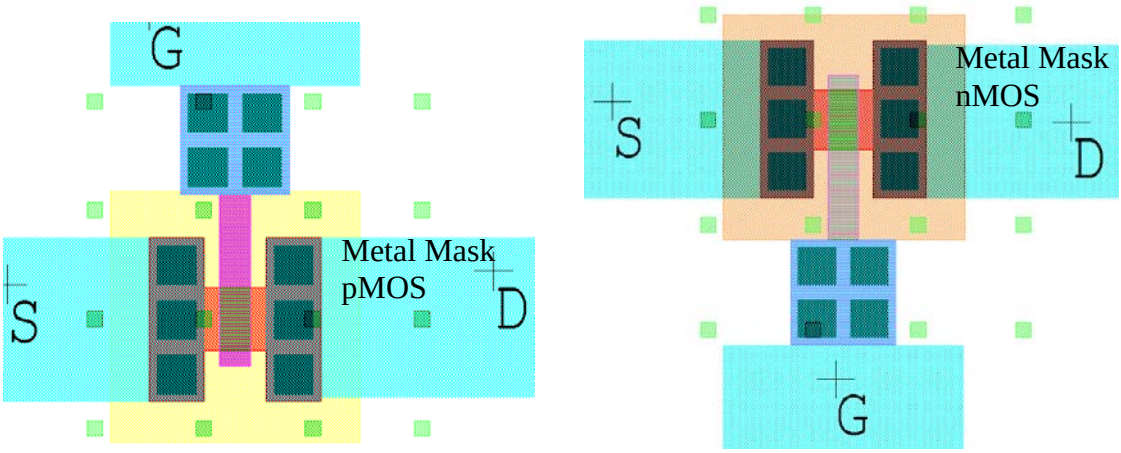
Fig.5.5: DRC for Metal layer

5-2) Designing layout inside of single grain silicon

All the designed devices should be inside of single grain silicon to have high quality characteristics. In designing layout to have high mobility devices, gate is place 0.5µm away and only in one position (X) from grain filter in all cases. This will help us to have stable devices that can result working circuit.

Fig. 5.6 shows the process flow of designed pMOS and nMOS transistors. The first layer is grain filter mask to define the single grain silicon area. Next step is island or active mask to define the silicon layer. Then gate mask is applied to define the gate area. Channel of transistors is under gate and it must be inside single grain area and not in the grain boundaries. In order to make nMOS and pMOS transistors we need mask to implant. To make nMOS transistors p^+ is used as a dopant and we cover the pMOS areas by nMOS mask. For pMOS transistors we use pMOS mask to cover nMOS transistors and dope B^+ in pMOS area. Finally we open the oxide to reach S/D and gate by Contact mask and metallization by M1 mask. The dimensions of transistors are 2µm gate length and 4µm width of transistors.


GF Mask
ISL Mask
Gate Mask
NMOS Mask
For P+
Implantation
PMOS Mask
For B+
Implantation
Contact Mask



5.6: The process flow of designed pMOS and nMOS transistors

In some of the circuits we need large width for transistors. In this case parallel transistors are used where all channels are inside of single grain silicon area. Fig.5.7 shows a transistor with W=24µm that is combination of 6 parallel transistors with W=4µm. The advantage of this type of layout is cancelling the process variation of devices by taking average of parameters.

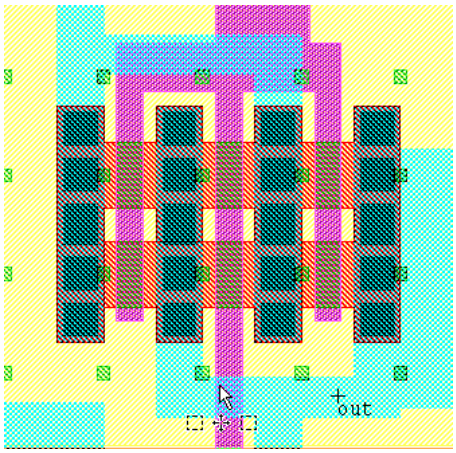


Fig.5.7: Parallel transistors to make one big transistor

Another point is the size of laser shots for crystallization and annealing. Fig.5.8 shows designed frame for layout. This frame consist of 24 areas with 1500µm×2330µm size and 163µm overlap between those rectangles. During laser crystallization two shots of layer make overlap in this area and we should avoid putting active layers there.

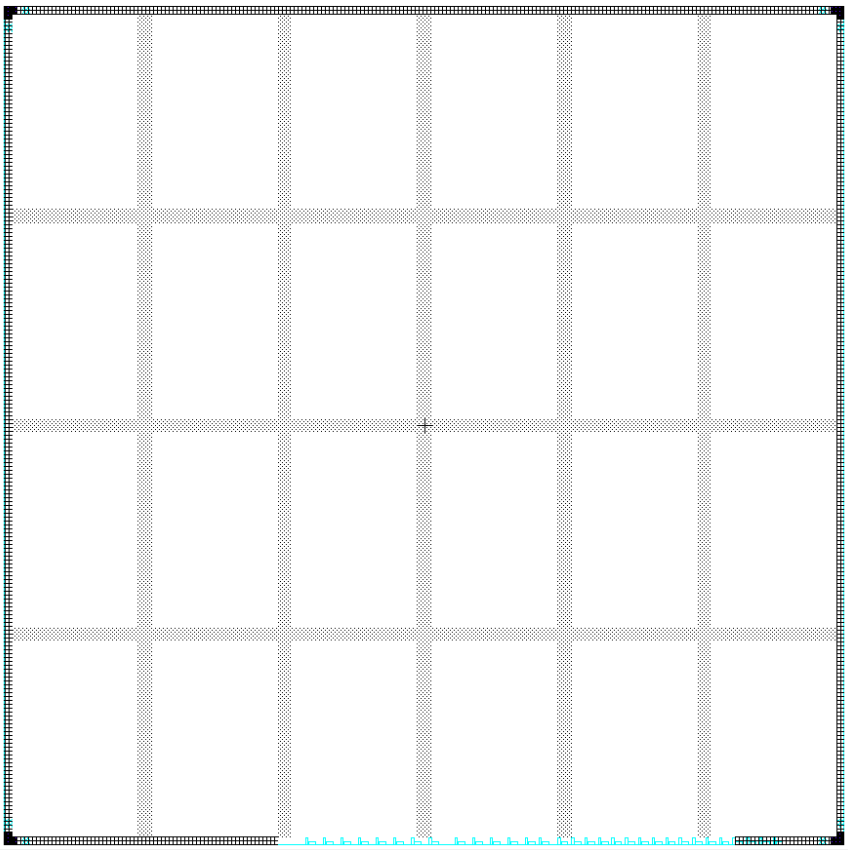


Fig5.8: Designed frame for layout. Whole area is one die. Laser shots during crystallization make overlap in gray areas and we should avoid putting active devices there.

During laser crystallization we need to check the grain sizes to find an optimum value for laser energy. Fig. 5.9 shows a designed structure with different grain sizes (grain pitch). When the laser energy is enough and optimized the grains get square. For example when we need 7µm grains the area related to 7 should show squared grains. Then all 4, 5, 6 areas will be squared too. But 8 will be rounded grains. If we increase the laser energy then ablation starts to happening and silicon layer will be evaporated. Therefore there is a trade-off between grain sizes and ablation.

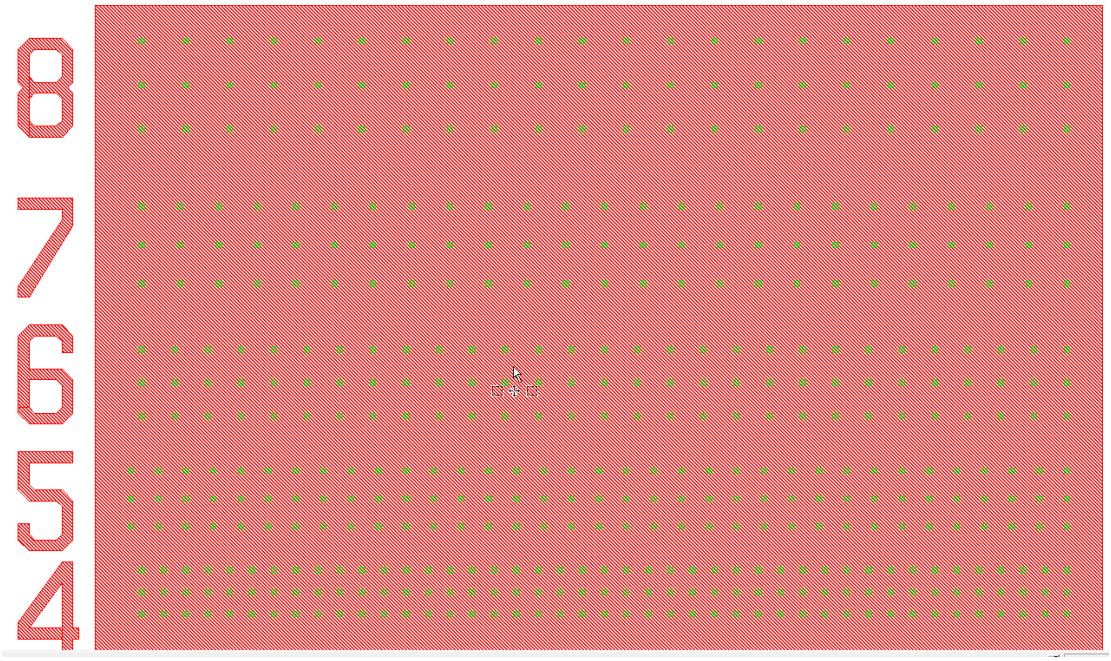


Fig.5.9: Different grain sizes to check and optimize laser energy.

5-3) One Layer SRAM

In designing layout for SRAM cells area should be minimized. We designed several test cells to check their characteristics and compare them. In general for small size cells we put one transistor inside one grain.

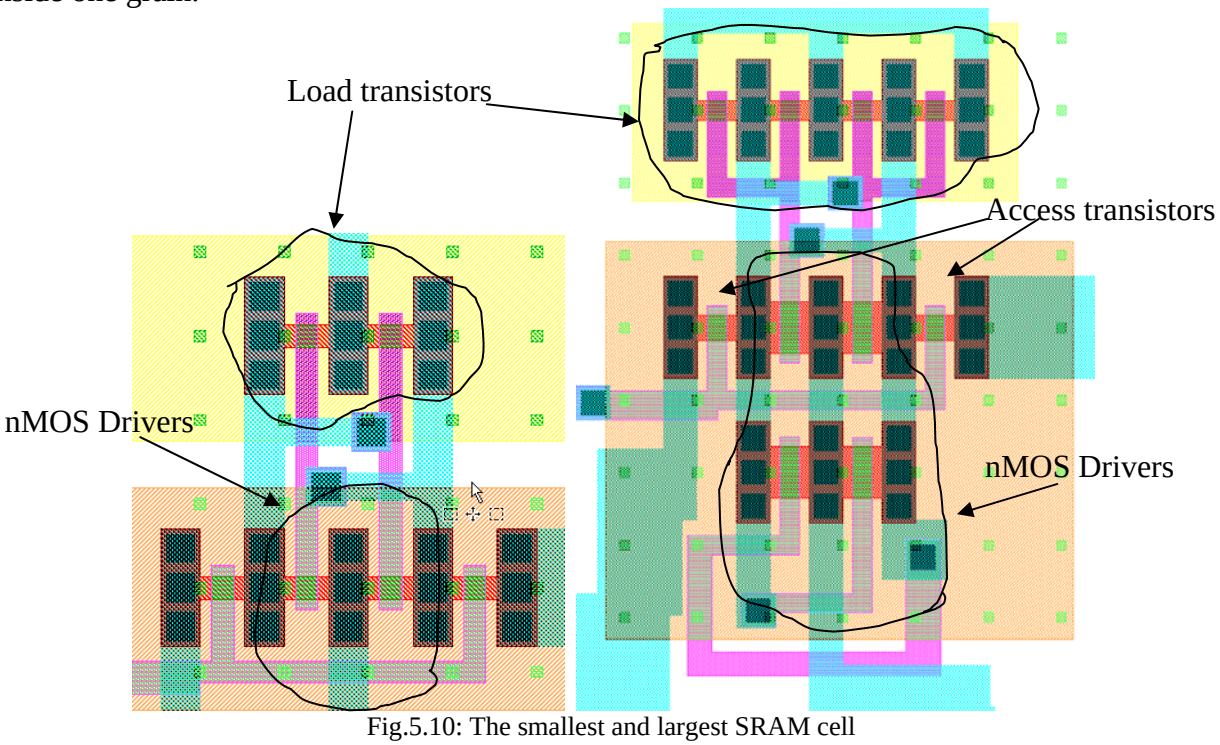


Fig.5.10: The smallest and largest SRAM cell

For larger width of transistors we used parallel transistors. Fig. 5.10 shows small and large cells. Furthermore to compare with 6T SRAM cells we have designed resistor load and poly TFT load

SRAM cells. Fig.5.11 shows both resistor and poly TFT SRAM cells. In resistor type still we use gate mask to prevent heavy implantation of resistor and finally to make a high value resistor for loads. In poly TFT pull up transistors we removed grain filter from design and during laser crystallization it will be a poly TFT.

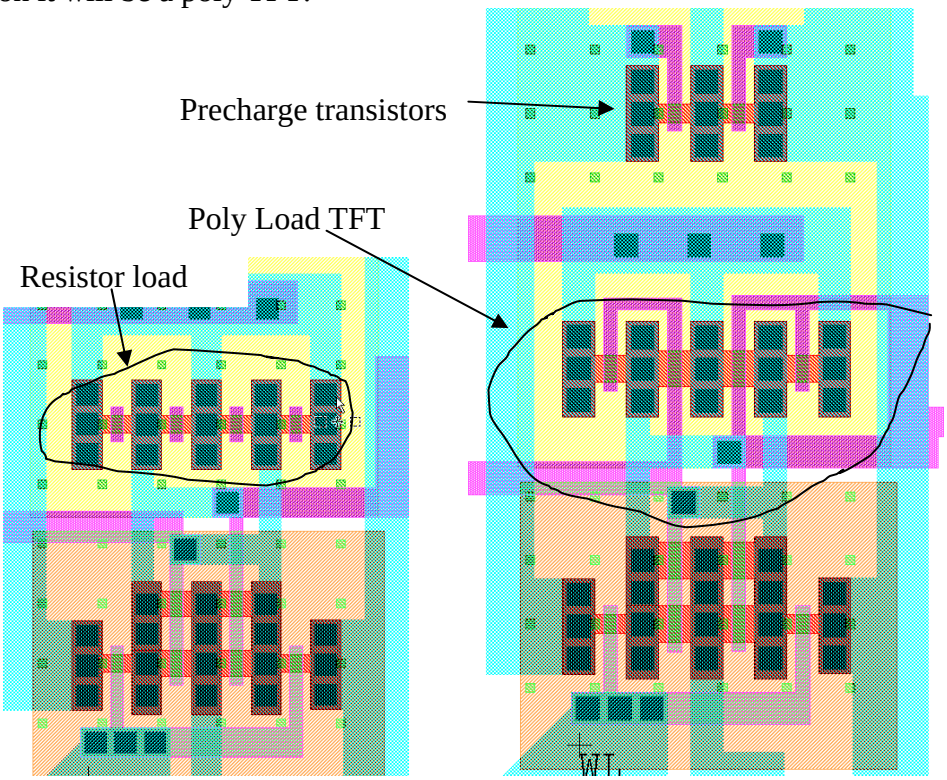


Fig.5.11: Poly load pMOS transistors Resistor load cell

The complete SRAM cell with precharge transistors, sense amplifier and output buffers is shown in Fig.5.12. Later we will talk about sense amplifier and buffers.

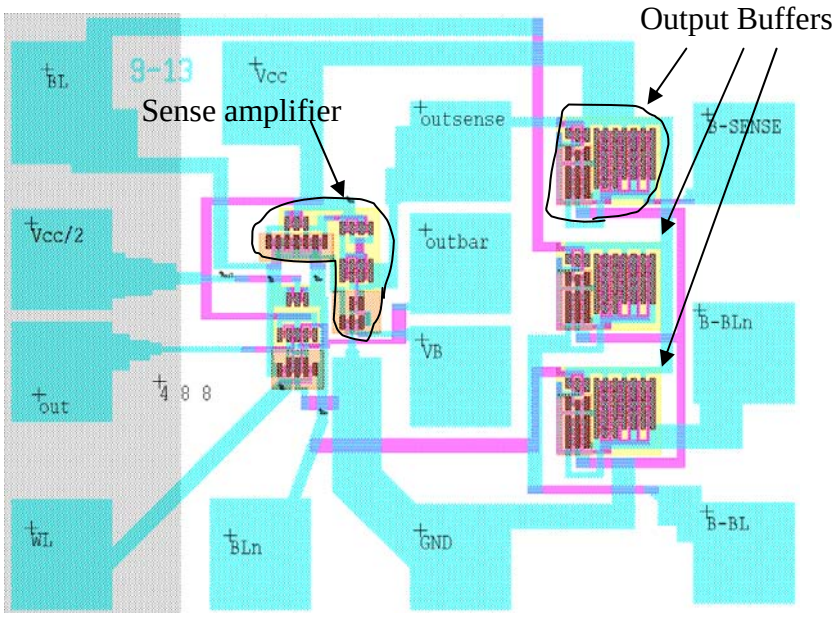


Fig.5.12: Complete SRAM cell layout

An SRAM with 4 cells and sense amplifier is shown in Fig.5.13. As it can be seen it has 4 WL to have access to the cells and precharge transistors.

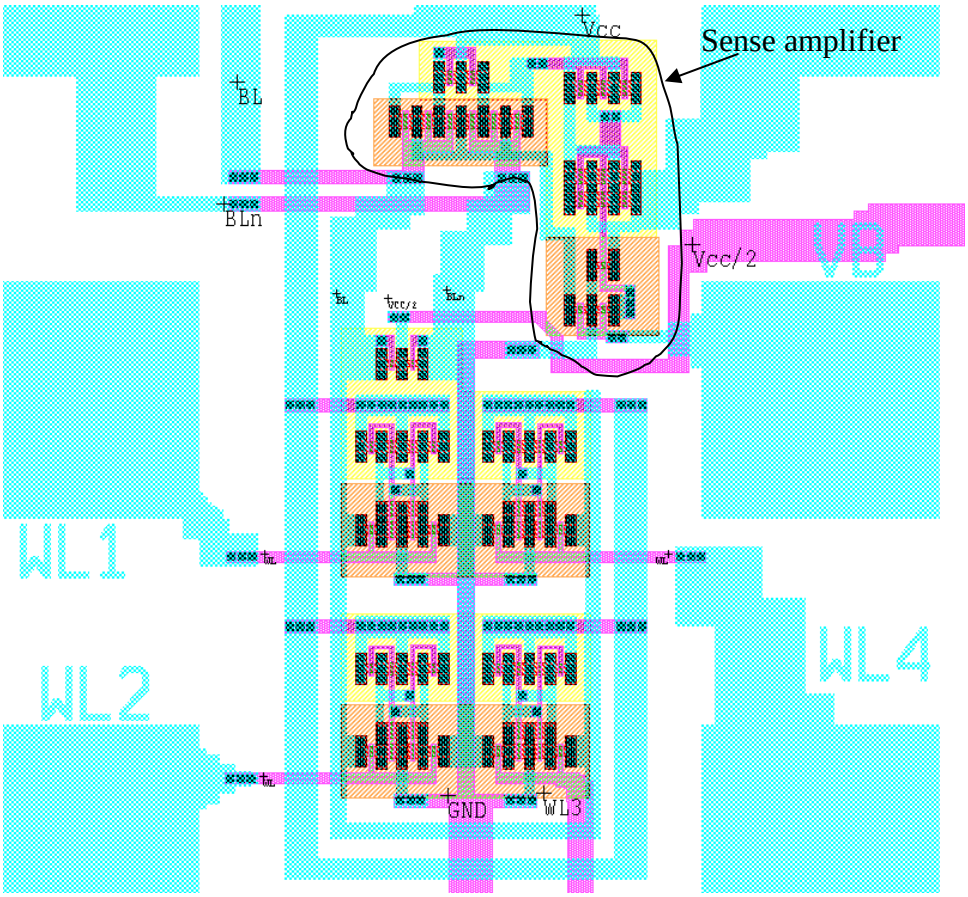


Fig.5.13: Four SRAM cells and sense amplifier

5-4) Two Layers SRAM

Designing SRAM cell layout for two layers can be done in two methods. In first case the top active silicon layer can be exactly on top of first layer and VIA interconnection will connect the top layer S/D and gate to the extended S/D and gate of bottom layer. It can be seen in the three dimensional view of Fig.5.14. In this case bottom active layers are not symmetric. Fig. 5.14 illustrates the active layers and interconnection between top and bottom layers. All pMOS transistors are on top layer and nMOS transistors are on bottom layer. The advantage of this layout is having minimum area. For example for SRAM cell with $W_a=W_d=W_p=2\mu m$ the area is $46\times40\mu m^2$.

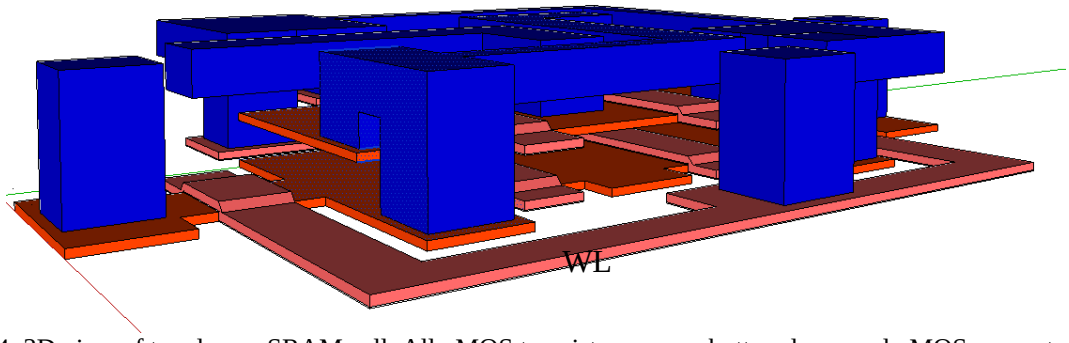


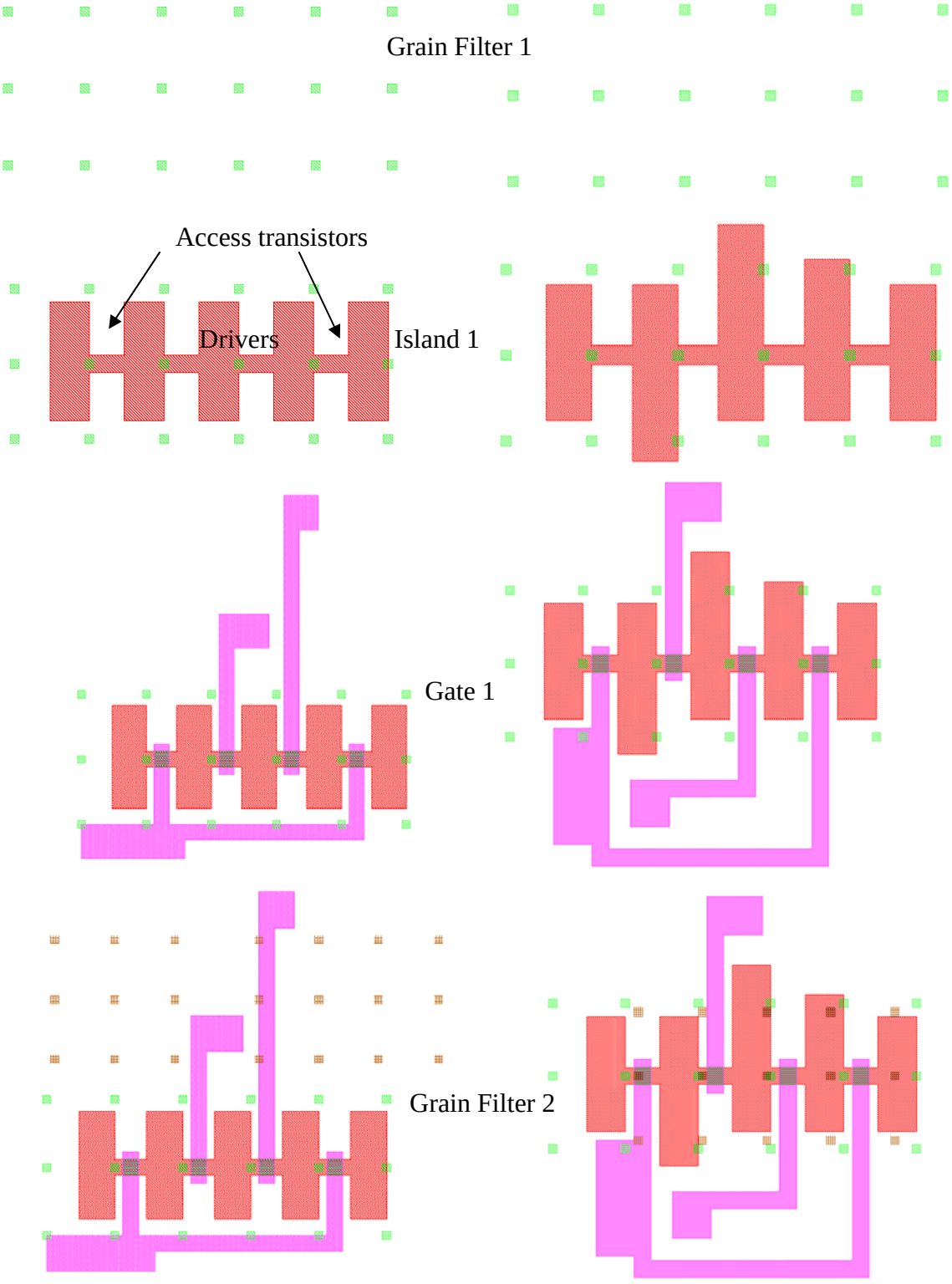
Fig.5.14: 3D view of two layers SRAM cell. All nMOS transistors are on bottom layer and pMOS are on top layer.

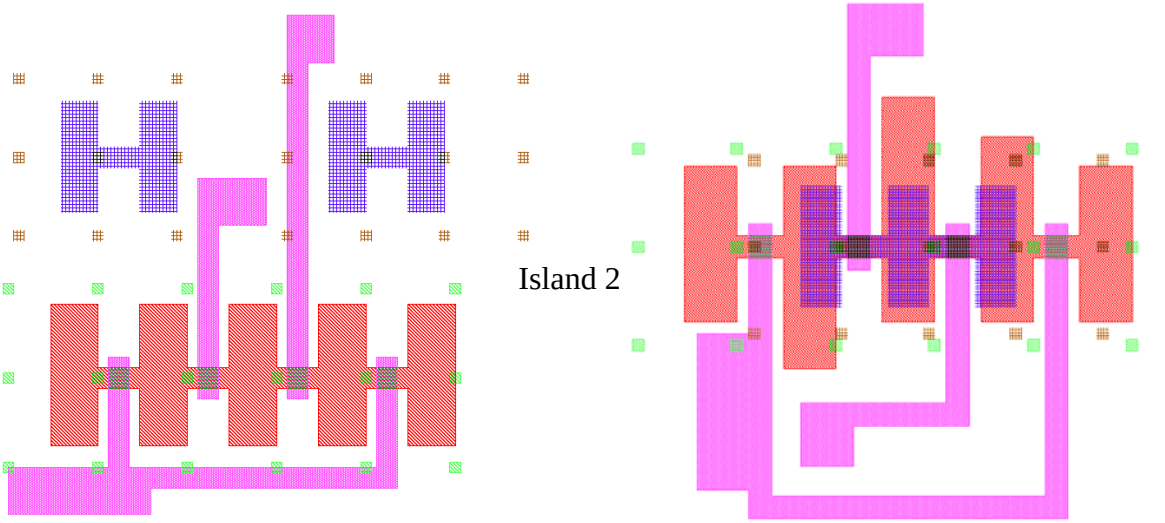
In the second option top silicon layer is shifted so that interconnection can be done between top and bottom layers without any extension of S/D and gate of bottom devices. In this case the area for SRAM cell with $W_a=W_d=W_p=2\mu m$ is $46\times47\mu m^2$.
In 3D layout pMOS transistors are on one layer and nMOS transistors are on the other layer. In general soft errors can only effect on top layer. Therefore we can put less sensitive transistors on top layer such as nMOS drivers. Soft error is effective in write mode.
In placement of transistors, different combinations of access, drivers and pull up transistors have been designed. Combinations which result less interconnection number and Small area are selected. Table 5.1 shows the specifications of designed different layouts. In this table area and number of VIA between two layers have been compared for different placements. In all cases except last column active layers are on top of each other. This table shows when nMOS transistors are on top layer and pMOS transistors are on bottom layer the number of VIA interconnects are less and we do not need mask for S/D implantation.

Table 5.1: Comparing different placements of transistors on two layers of silicon

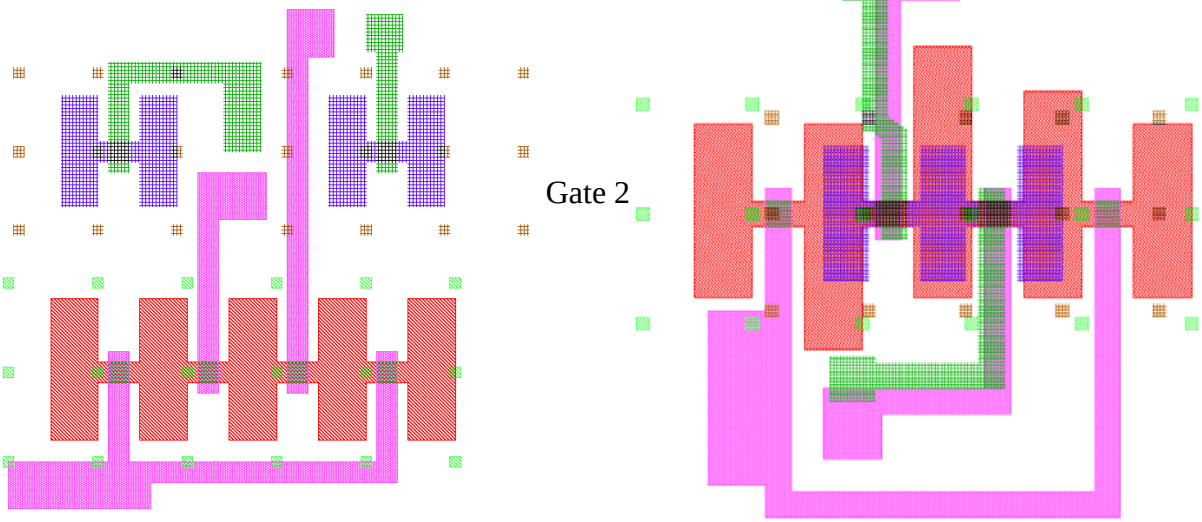
Top layer	nMOS Access and drivers	pMOS Pull up	nMOS Drivers	nMOS Access and pMOS pull up	pMOS Pull up Shifted
Bottom layer	pMOS Pull up	nMOS Access and drivers	nMOS Access and pMOS pull up	nMOS Drivers	nMOS Access and drivers
Number of VIA interconnects	5	8	10	5	8
Area	2080	1380	2130	1860	1570
Other Advantages	No mask for S/D doping	No mask for S/D doping	Less sensitive to soft errors		No mask for S/D doping Symmetric

Fig.5.15 shows process flow of layout design for symmetrical and non-symmetrical silicon islands on two layers. This process is shown for nMOS transistor on bottom layer and pMOS transistors on top layer. NMOS and PMOS masks are not shown here since we do not need them in this case. For simplicity of the process only one layer of metal is used.

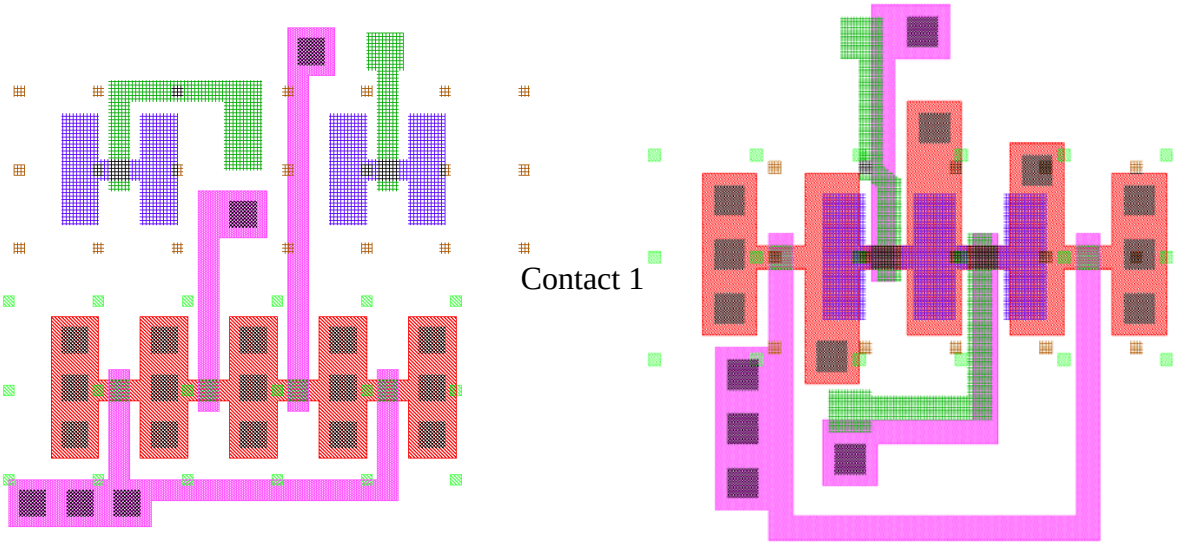

Grain Filter 1
Access transistors
Drivers
Island 1
Gate 1
Grain Filter 2



Island 2



Gate 2



Contact 1

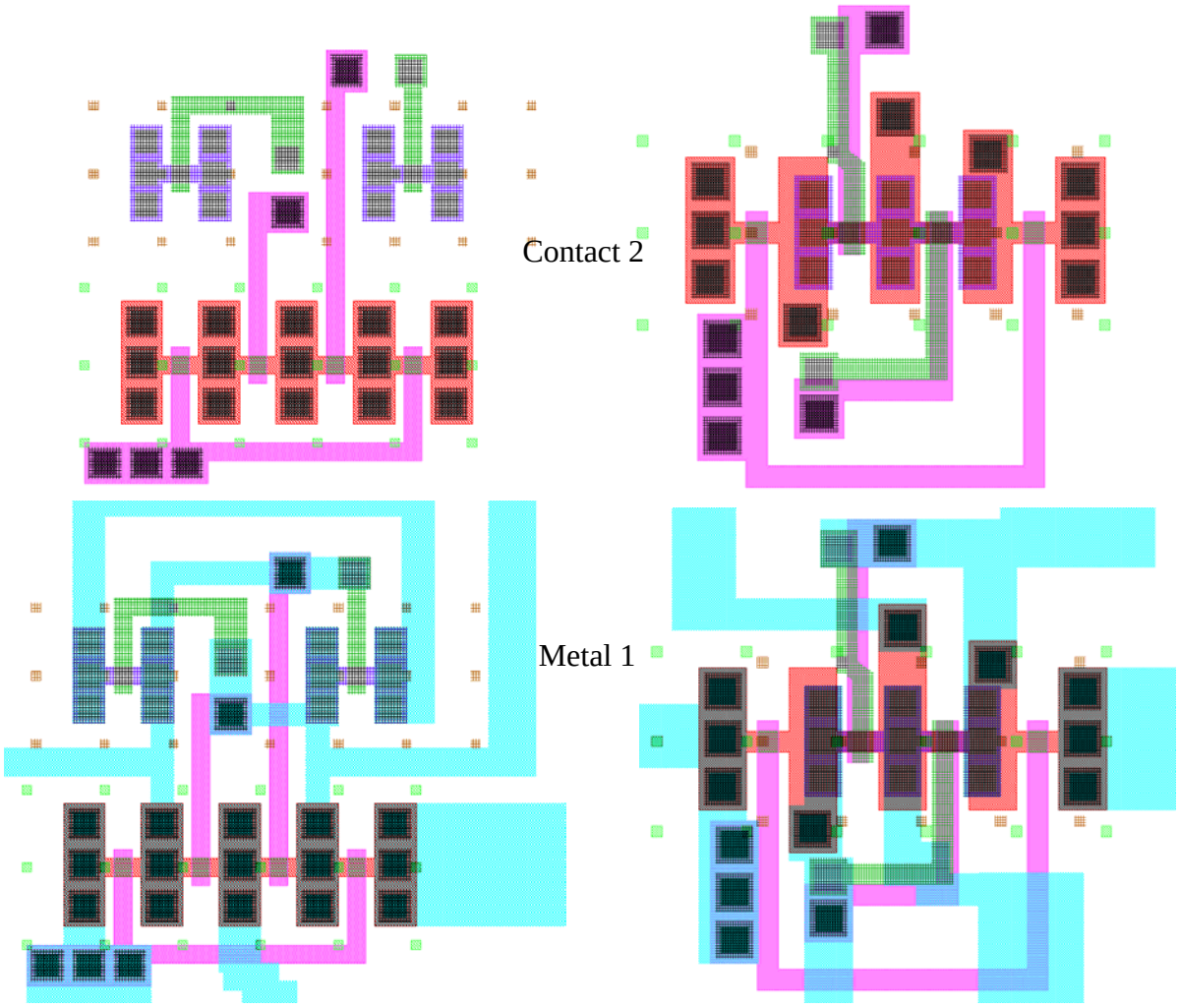
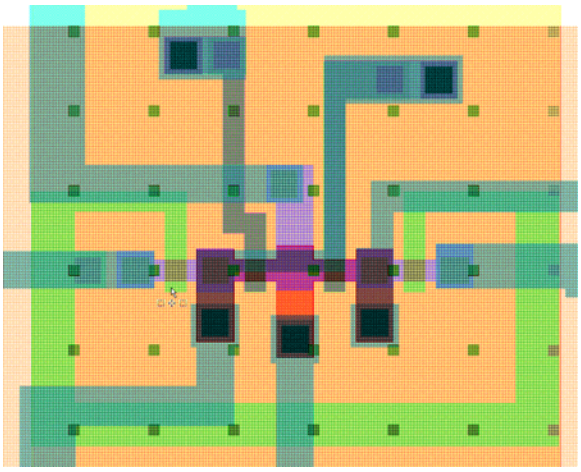
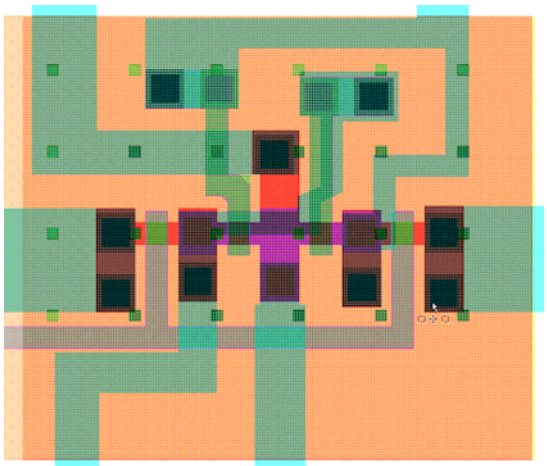


Fig.5.15: Process flow of layout designing for symmetrical and non-symmetrical islands in 3DIC

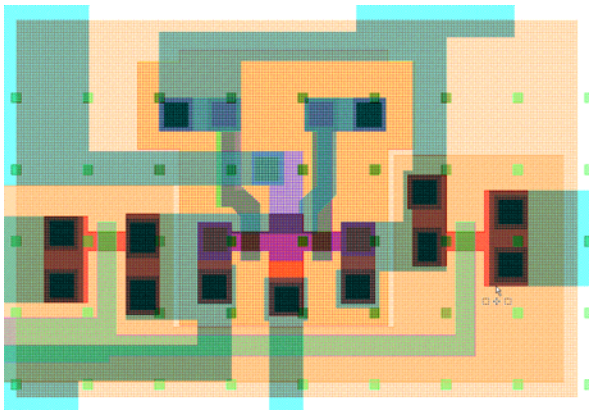
The complete designed layouts for all cases that mentioned in table 5.1 are shown in Fig.5.16.



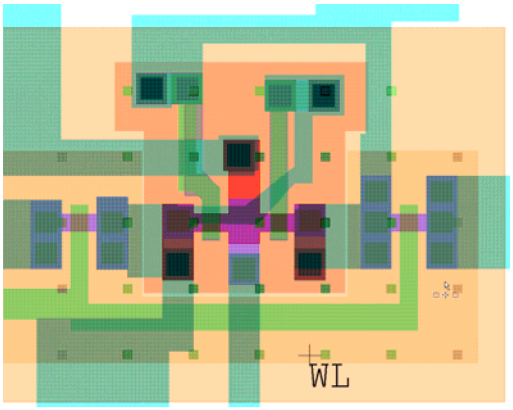
a) Top: nMOS, Bottom: pMOS



b) Top: pMOS, Bottom: nMOS



c) Top: nMOS drivers, Bottom: nMOS access, pMOS



d) Top: nMOS access, pMOS, Bottom: nMOS drivers

Fig.5.16: Designed layouts for different placement of transistors on two layers of silicon

5-5) Double-Gate and H-Gate SRAM

Double gate transistors are used in the access transistors to change the threshold voltage of transistors to improve pull up ratio and cell ratio during write and read operation modes. Fig.5.17 shows 3D view of double gate transistor. It can be seen that there is overlap area between top and bottom gates.

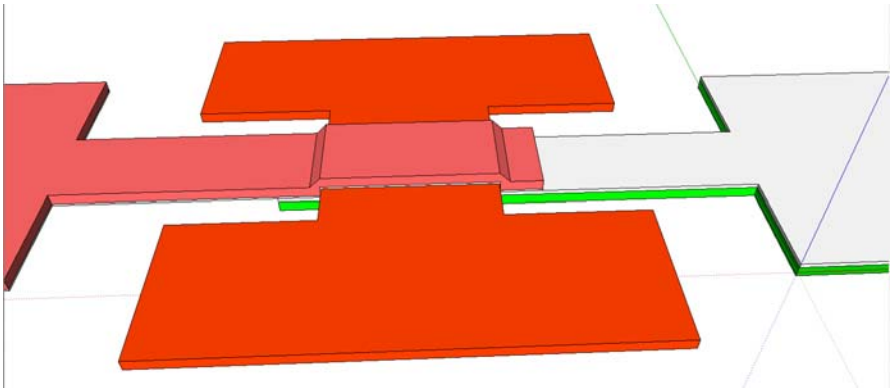


Fig.5.17: 3D view of double gate transistor

In designing layout for double gate transistors we designed different overlap area between top and bottom gates. Fig.5.18 shows layout of double gate transistor and using that in SRAM cell.

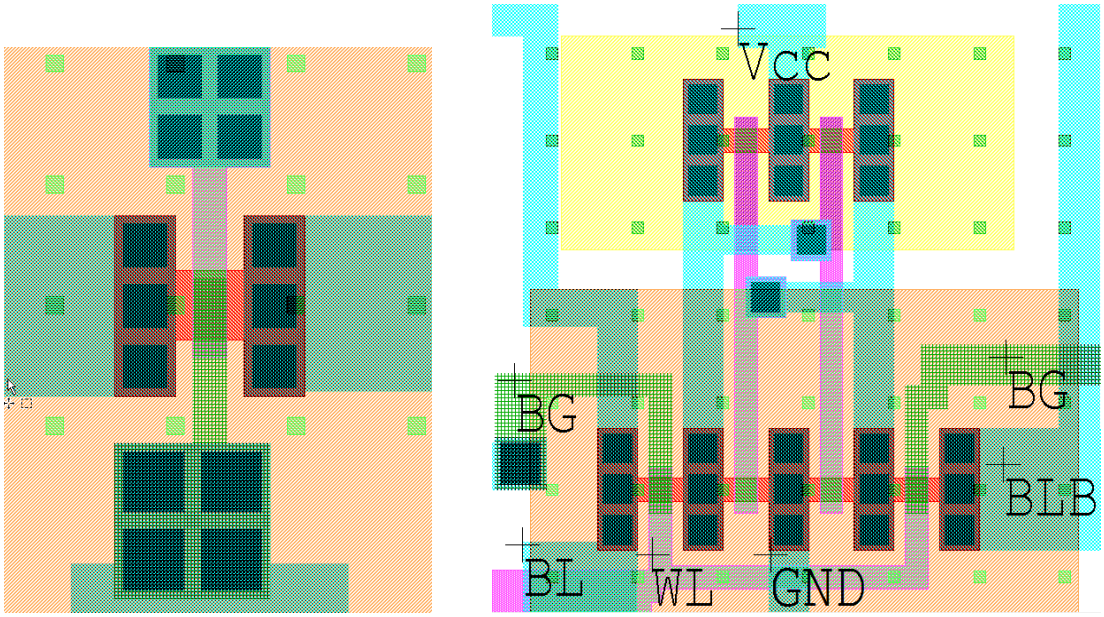


Fig.5.18: Designed layout for double gate transistors and using them in SRAM cell

Another way of changing threshold voltage is using H-gate transistor that has been shown in the Fig. 5.19.

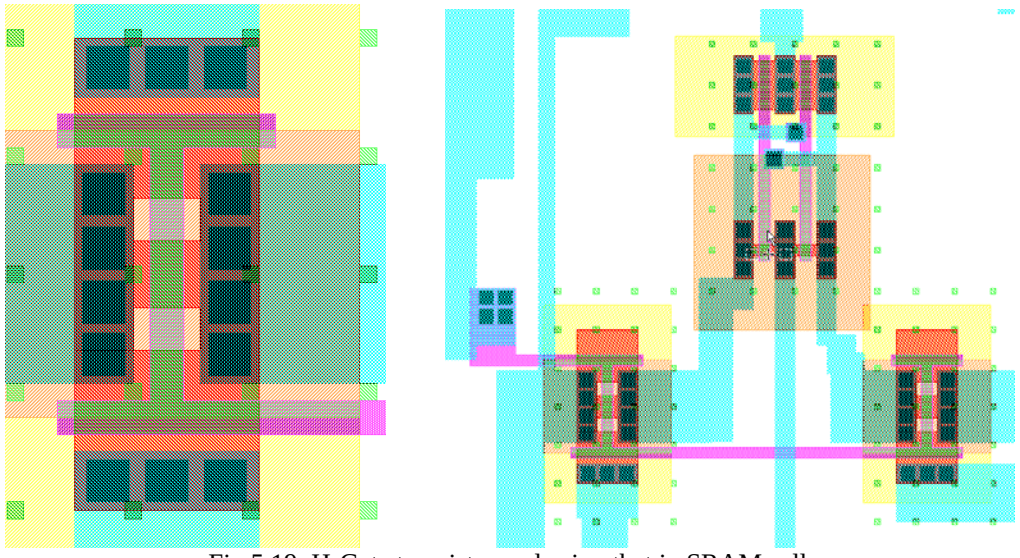


Fig.5.19: H-Gate transistor and using that in SRAM cell

5-6) Sense amplifier

Designing layout of sense amplifier is critical because of matching between transistors in differential amplifier. To decrease the influence of parameters process variation we have used parallel transistors to make an average of parameters. Since the input transistors have large width (W=15µm) it is not possible to put them in one grain. Fig.5.20 shows designed sense amplifier.

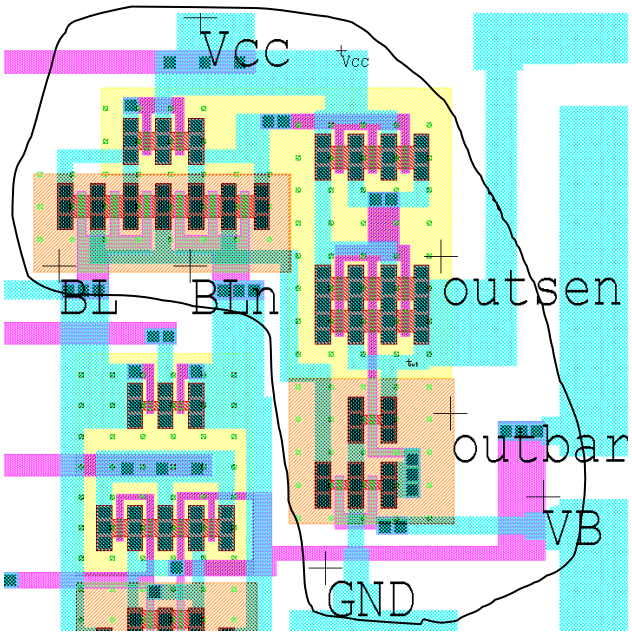


Fig.5.20: Designed layout for sense amplifier

5-7) Output Buffers

Output buffer is designed to drive load of measurement system. Designed layout is demonstrated in Fig. 5.21. The small and large nMOS and pMOS transistors are for first and second inverters, respectively.

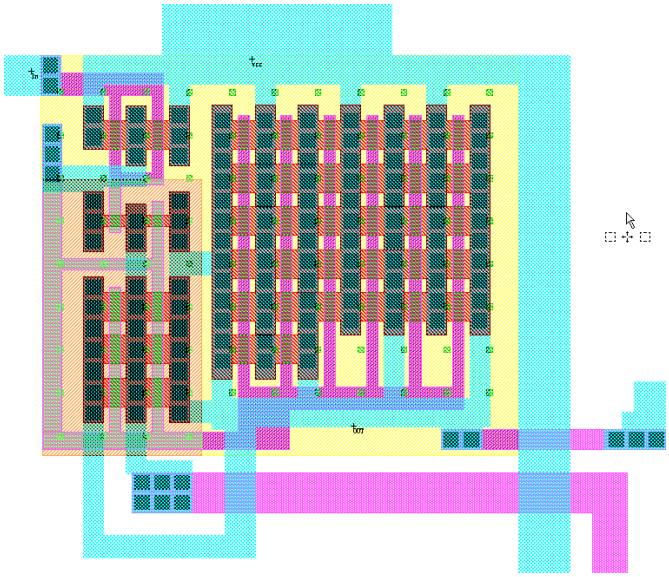


Fig.5.21: Designed layout for output buffer

5-8) Conclusions

In this chapter the principle of designing SRAM cells have been investigated. The layout of one layer and two layers SRAM cells, sense amplifier and output buffers have been designed and compared in term of area and number of VIA interconnections. Having nMOS transistors on top

layer and pMOS transistors on bottom layer can save the number of masks and have less interconnection VIA.

Chapter 6

Fabrication Process of SRAM in Planar and 3DIC Technologies

Content:

6-1) Single Grain TFTs fabrication process

In order to make thin film transistors using µ-Czochralski process wafers with thermal oxide are etched to open 1µm holes with grain filter mask. Then second oxide is deposited on wafers to make those holes narrow in the order of 0.1µm. Next amorphous silicon is deposited using LPCVD system at 550^{o}C. To control the threshold voltage all wafers are implanted with 10KeV energy and 2.5×10^{11} ions/cm^2 dose of boron. Then wafers are irradiated using Excimer laser with approximately 1500mJ/cm^2 energy at 450^{o}C to make single grain silicon layer. During laser crystallization silicon layer is molten except bottom of cavity. When energy is enough grains are squared as shown in Fig.6.1. In these images 4, 5, 6 and 7µm grains are squared but 8µm grains are rounded. Both optical and SEM images show squared grains. Our design is 7µm grain size therefore this energy is enough for our purposes.

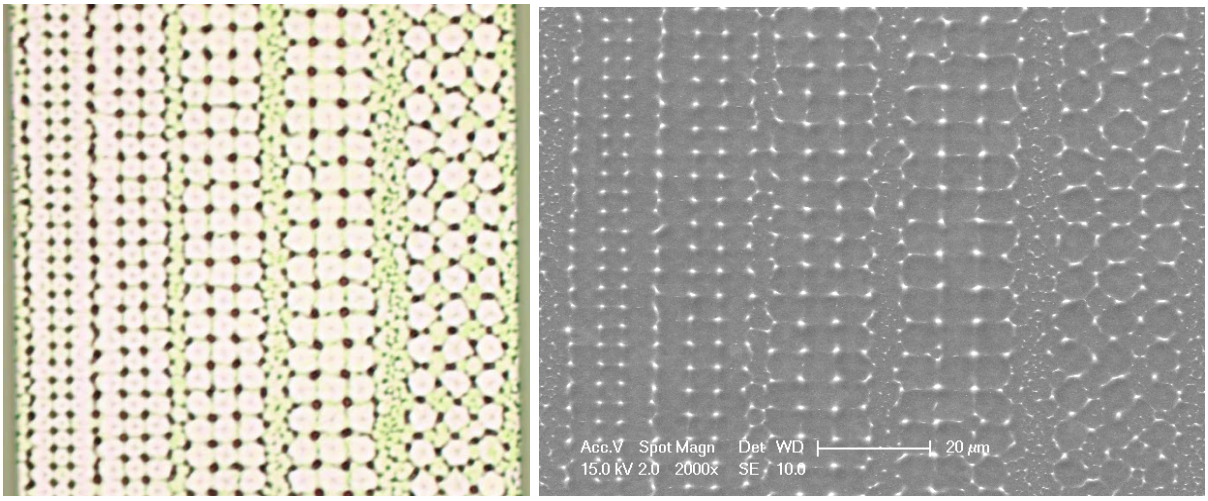


Fig.6.1: Optical and SEM images of grains.

After silicon crystallization we use island mask to pattern the silicon layer. Fig.6.2 shows SEM picture of island related to an SRAM cell. The names of areas have been written on image.

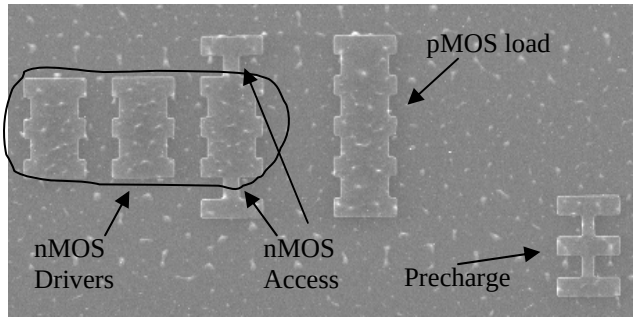


Fig.6.2: SEM image of islands

Fig.6.3 show optical and SEM picture of island area related to output buffers.

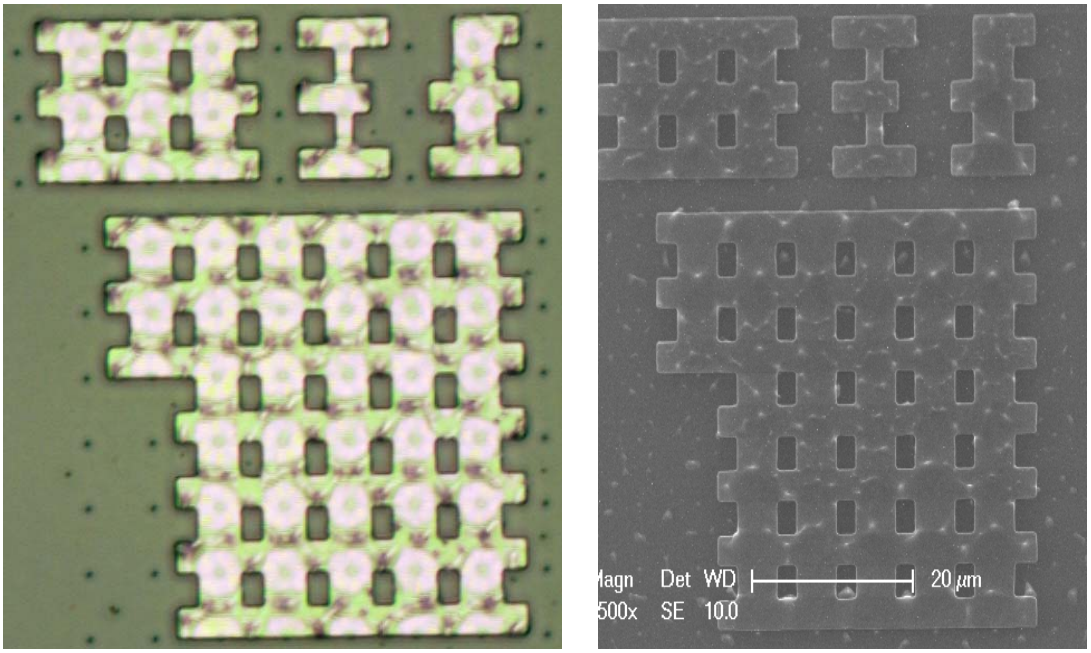


Fig.6.3: Optical and SEM image of islands related to output buffers

After cleaning, 30nm gate oxide is deposited and followed by 675nm Al. Gate mask is used to etch Al and oxide and then using NMOS mask we implant the source and drain of transistors with P^+ with 30KeV energy and 5×10^{15} ions/cm^2 dose and PMOS transistors with B^+ with 20KeV energy and 5×10^{15} ions/cm^2 dose. Next using Excimer laser with 300mJ/cm^2 dopants are activated. Then 800nm TEOS oxide is deposited on devices and is opened by contact mask to reach to source, drain and gate. Finally 1.4µm Al is deposited and patterned by metal one mask. In the last step alloying is done at 400^{o}C for 20min.

Fig.6.4 shows SEM image of completed SRAM cell. Bright lines are metal one AL layer. Silicon islands and gate are visible in this image. As it can be seen interconnection to the islands are done by three 2.5µm×2.5µm opened holes in oxide using Al. This will increase the reliability of devices.

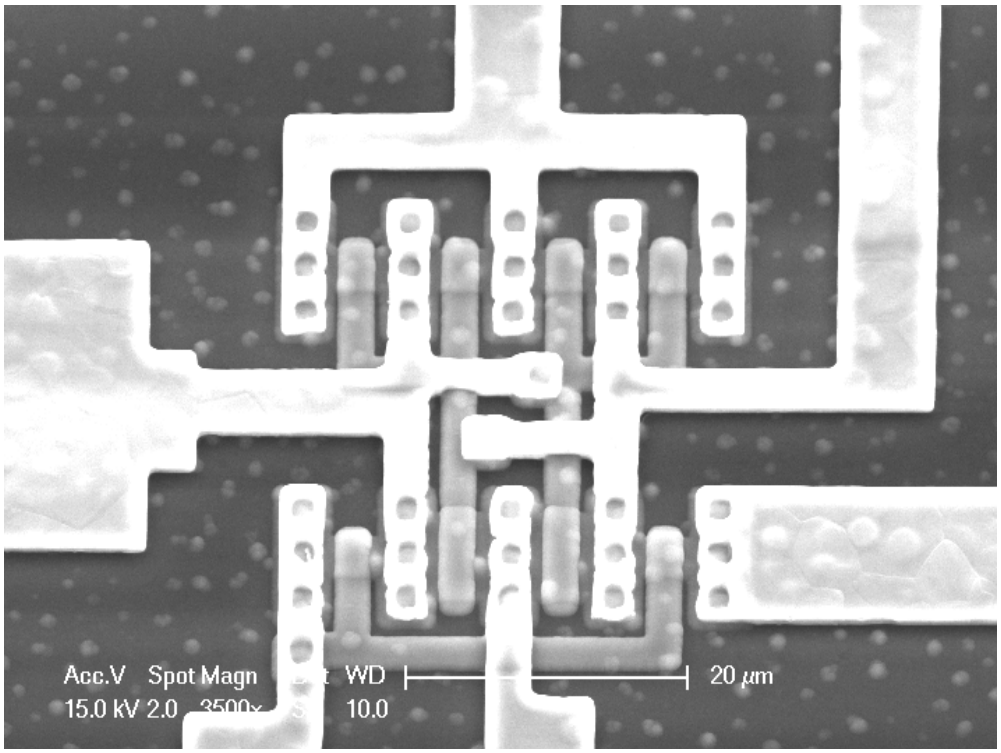


Fig.6.4: SEM image of completed SRAM cell.

Fig. 6.5 shows complete SRAM cell with sense amplifier and output buffers. Same picture was shown in previous chapter in layout.

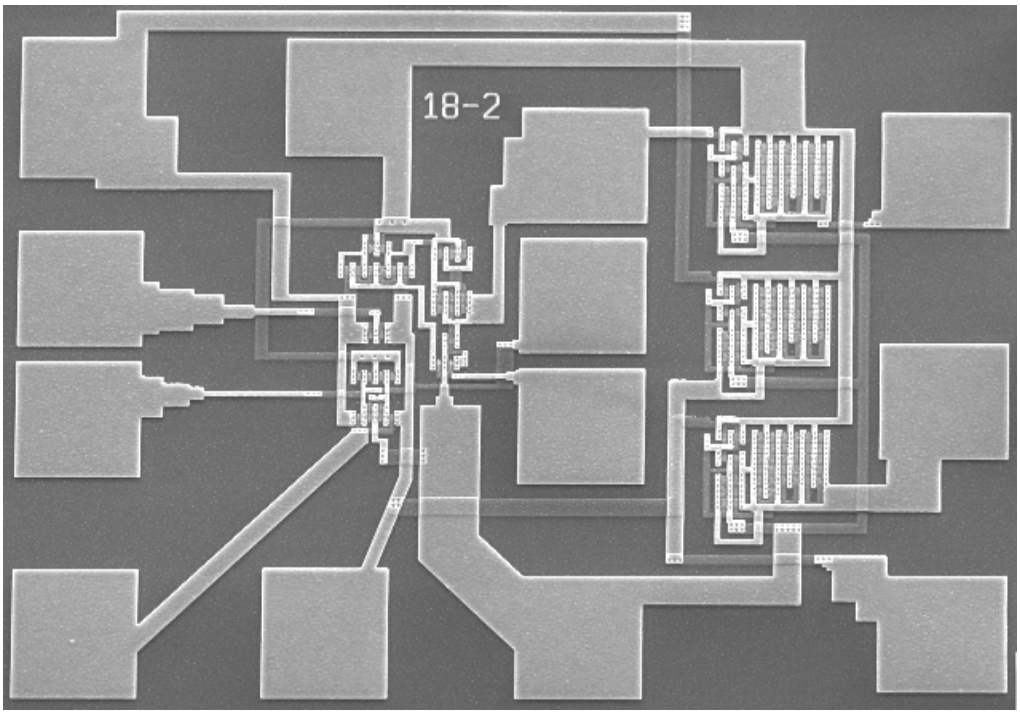
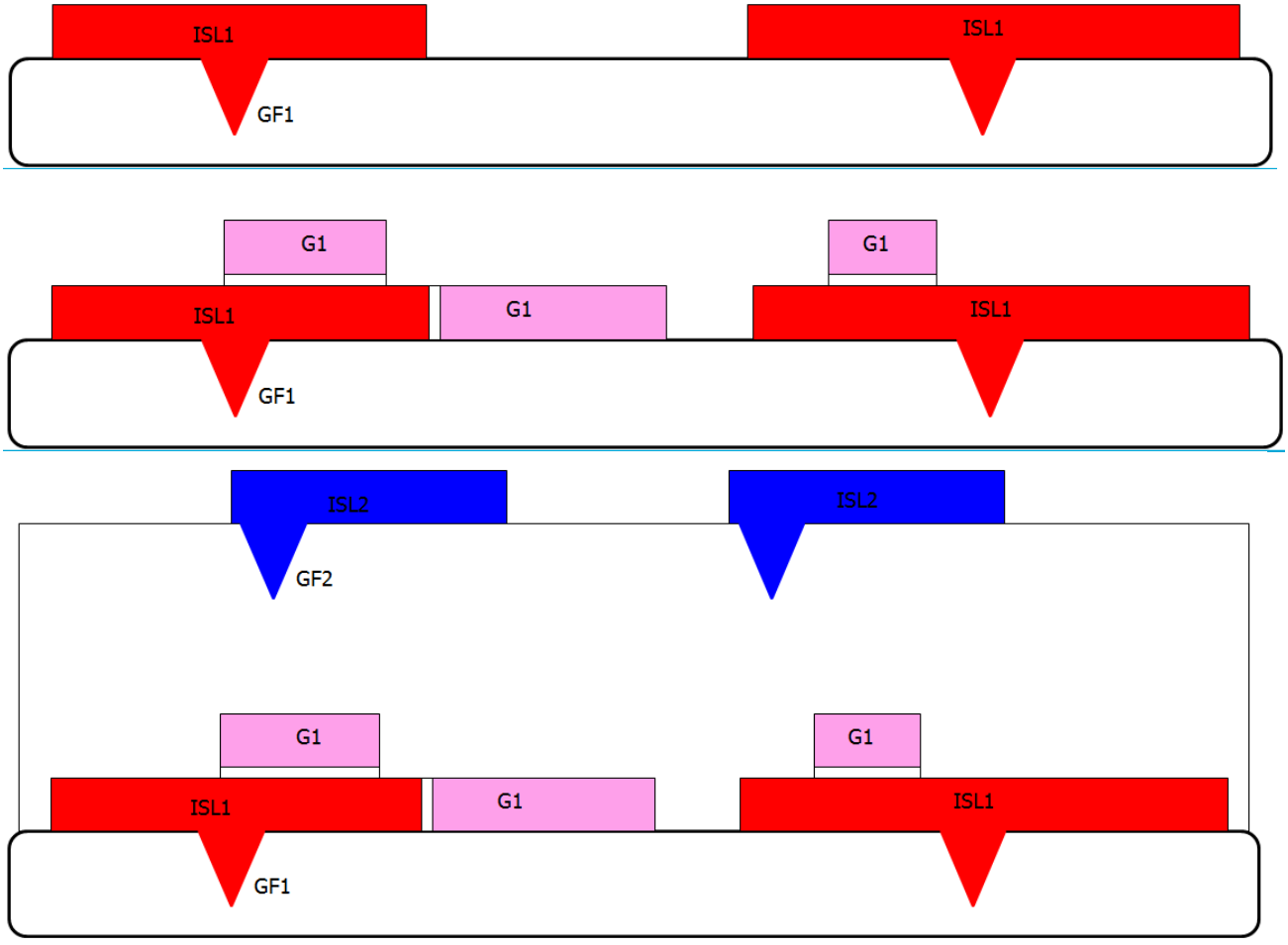


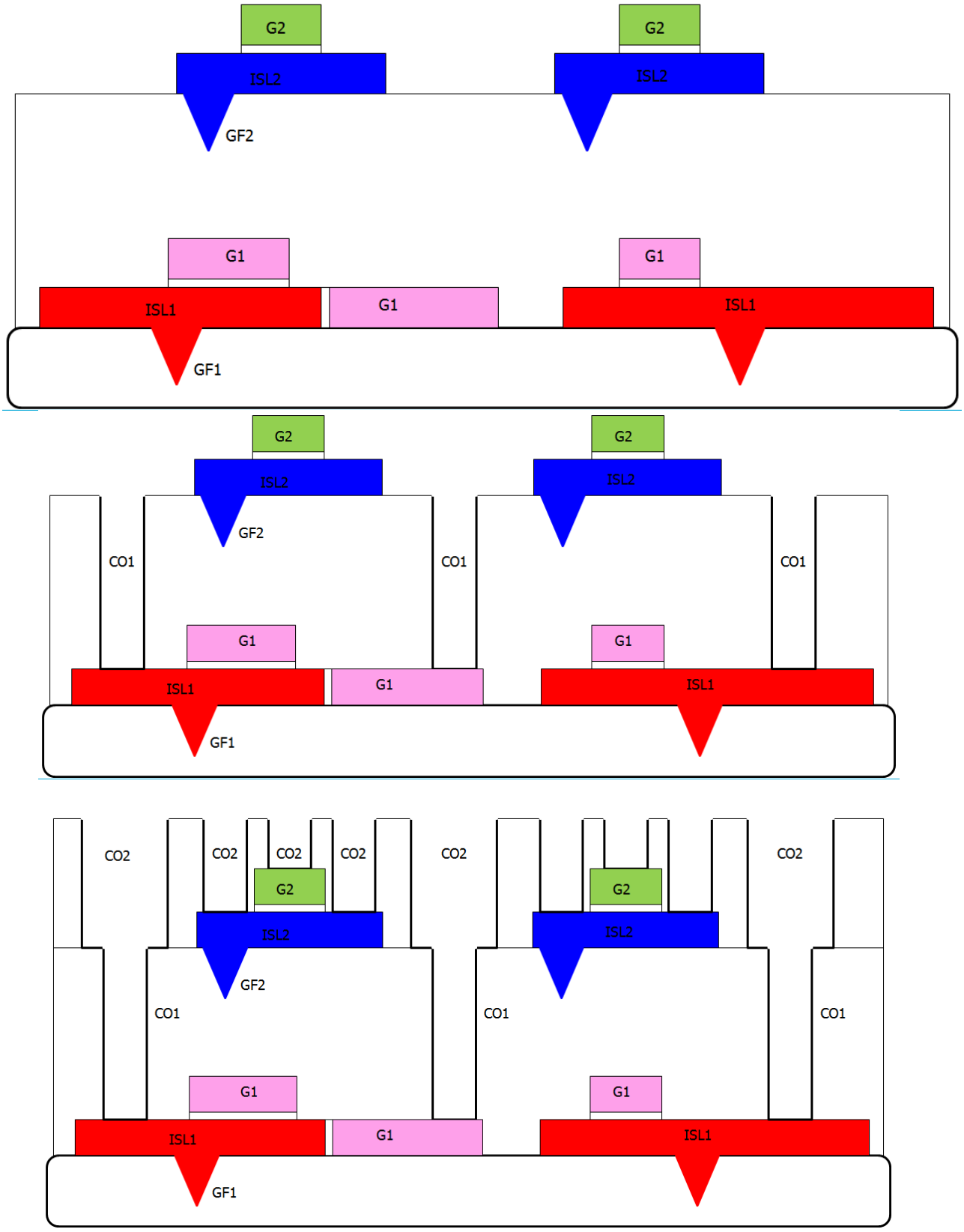
Fig.6.5: SEM image of complete SRAM cell with sense amplifier and output buffers

6-2) Fabrication Process of Two Active Layers to make SRAM Cells

Fabrication process of two layers is same as one layer. For two layers we repeat the fabrication process of first layer. Since after finishing first layer we need to deposit silicon layer by LPCVD we can not use AL as a gate in bottom layer. Instead we use poly silicon with 250nm thickness and during doping and activation of bottom layer source and drain; poly gate is doped and activated. Fig.6.6 shows the steps of fabrication process in detail. After making top gate to have access to bottom layer source, drain and gate we use contact one (CO1) mask to reach to bottom devices. Then 1.2µm TEOS oxide is deposited on wafers and then we use contact two (CO2) mask to reach to bottom and top devices. The openings of CO1 mask are 2.5µm×2.5µm and CO2 mask are 3.5µm×3.5µm. Then we will cancel any chance of misalignment between two CO masks. After that 3µm Al is deposited at 350^{o}C to make good contact between top and bottom devices.

Another point is influence of top silicon crystallization on bottom devices. During crystallization of top layer heat is generated by laser irradiation and can redistribute the dopants of bottom layer. However, if the thickness of oxide layer is more than 1µm we will not have any problem for bottom devices.




G2
ISL2
GF2
G2
ISL2
G1
ISL1
G1
GF1
G1
ISL1
G2
ISL2
GF2
CO1
CO1
G2
ISL2
CO1
G1
ISL1
G1
GF1
G1
ISL1
CO2
CO2
CO2
CO2
CO2
CO2
G2
ISL2
GF2
G2
ISL2
CO1
CO1
CO1
G1
ISL1
G1
GF1
G1
ISL1

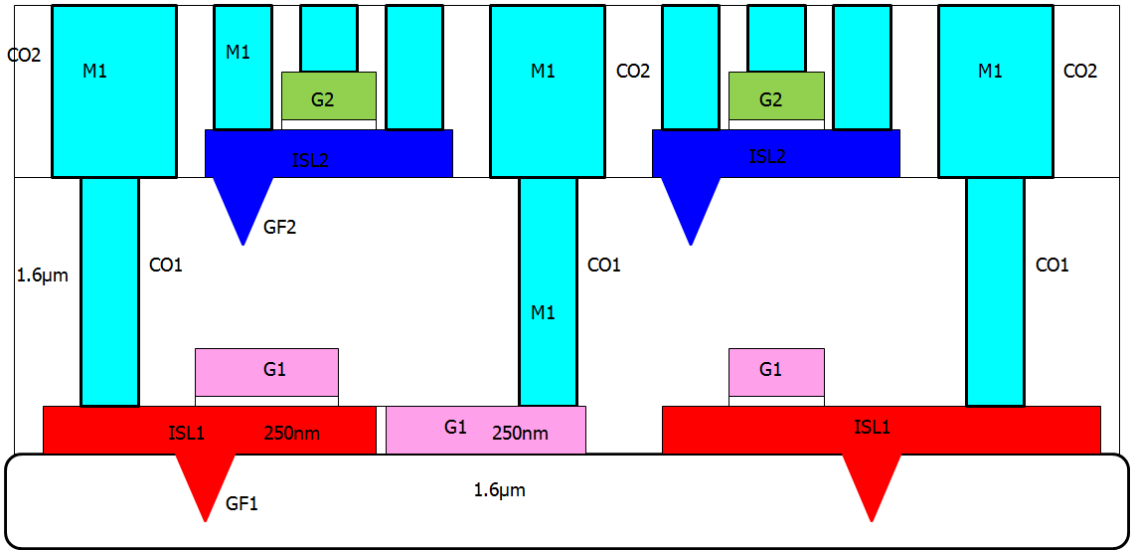


Fig.6.6: Fabrication flow of two layers of devices.

6-3) Double gate TFTs for SRAM

The fabrication process of double gate transistors starts after deposition of oxide on grain filters. Using inverse of bottom gate mask 350nm oxide is etched and then 250nm amorphous silicon is deposited and etched by bottom gate mask. Then this silicon layer is doped and activated to create p^+ and n^+ bottom gates. Next is depositing 100nm gate oxide for bottom gate and then depositing main silicon layer. The rest of process is same as TFT process.
Fig.6.7 shows SEM image related to etched area of bottom gate.

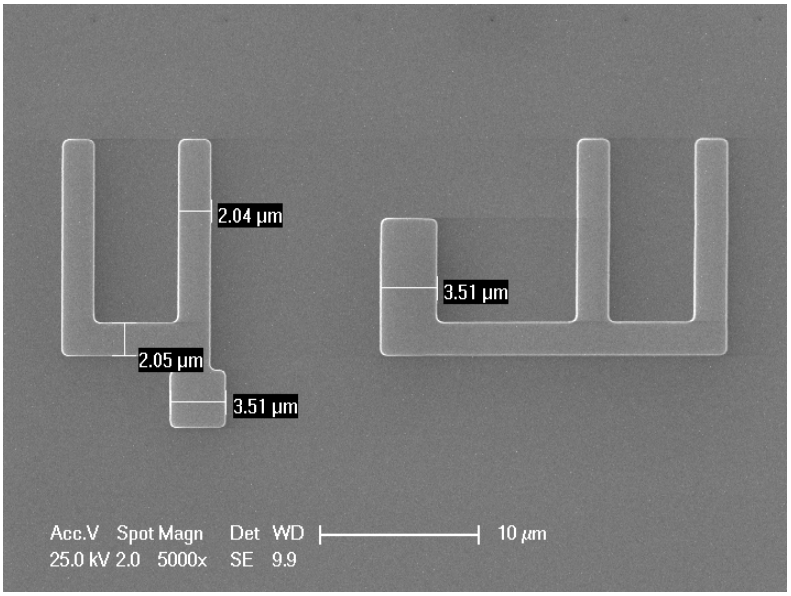


Fig.6.7: Etched area one oxide to make embedded gate.

After etching oxide and depositing poly gate and main silicon island we crystallize the silicon. Fig. 6.8 shows an SEM image of Silicon Island and embedded poly gate.

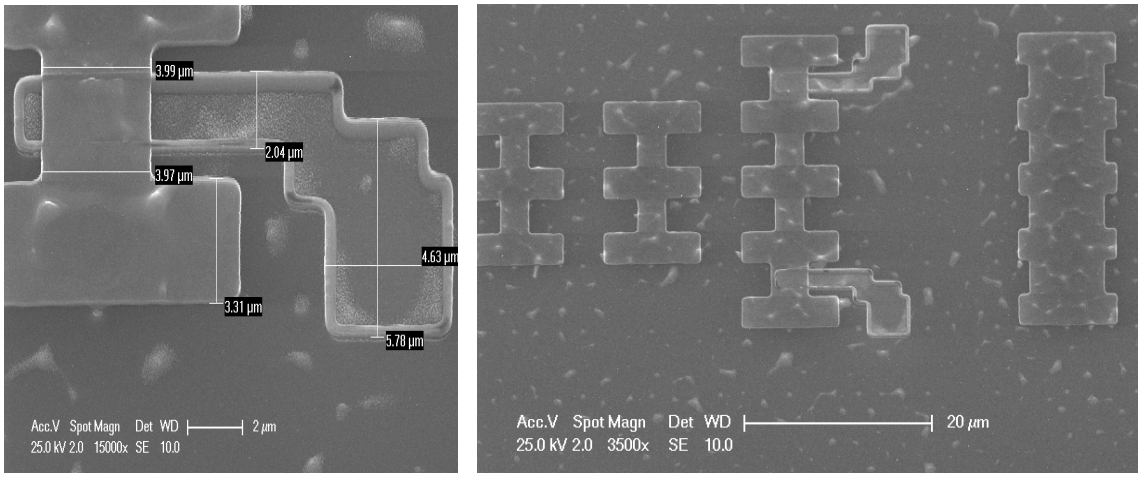


Fig.6.8: SEM image of silicon island and bottom gate

SEM image of completed transistor is shown in Fig.6.9. Both bottom and top gates can be seen in this picture.

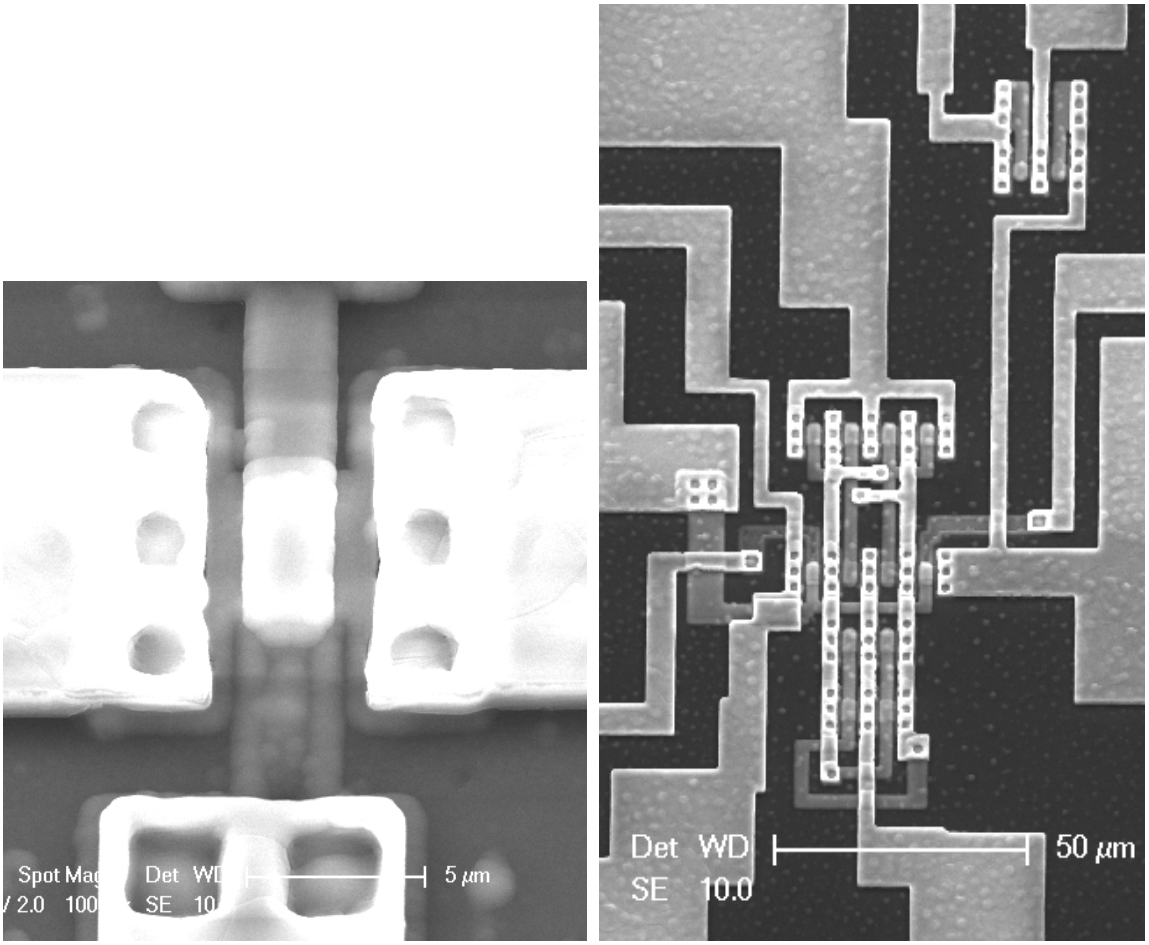


Fig.6.9: Completed SEM image of double gate transistor.

The fabrication process of H-Gate transistors is same as normal single grain TFT process. Fig.6.10 shows SEM image of fabricated transistors and SRAM cells.

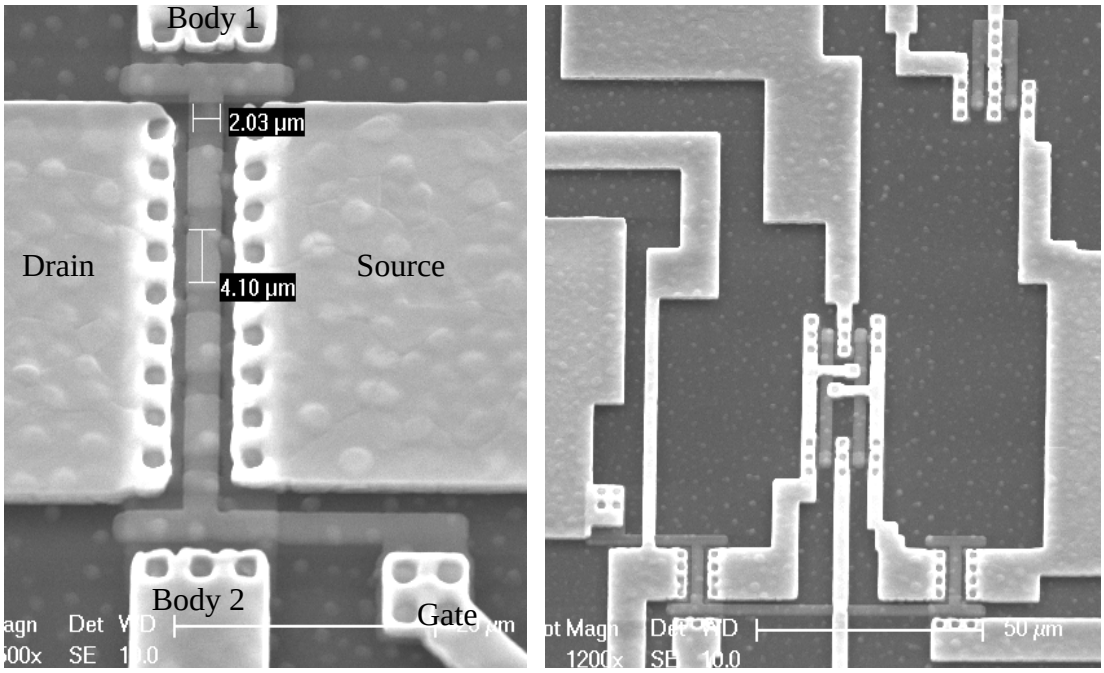


Fig.6.10: SEM image of fabricated H-Gate transistor and SRAM cell

6-4) Conclusions

In this chapter fabrication process of single grain TFT was discussed. The SEM image of fabricated one layer SRAM cell, Double gate and H-Gate SRAM cells have been shown.

Chapter 7

Electrical Characterization of SRAM Cells

Content:

7-1) Thin Film Transistors

The electrical characteristics of thin film transistors are shown in Fig. 7.1 and Fig.7.2. The Id-Vg curve of nMOS transistors shows high quality devices. It can be seen from linear curve that threshold voltage of nMOS devices is 0.4V and from logarithmic curve the on current is around 100µA and off current is 5×10^{-12}A. Therefore the ratio of I_{on} to I_{off} is 2×10^{7}. For this transistor the subthrehsold slope is 150mV/dec.

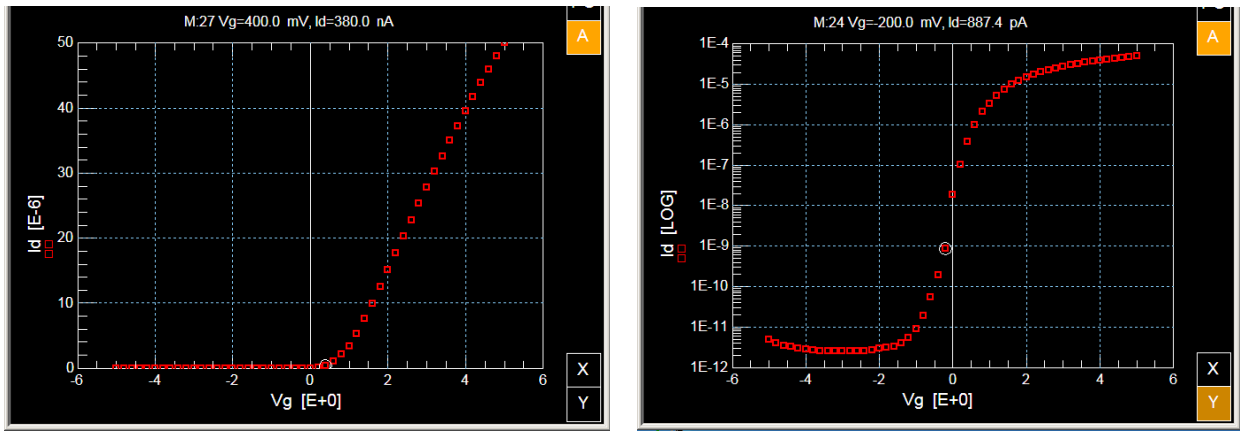


Fig.7.1: Linear and logarithmic Id-Vg curve of fabricated nMOS transistors

For pMOS transistors as illustrated in Fig.7.2 threshold voltage is -1.2V and on current is 10µA and off current is 5×10^{-14}A. Therefore the ratio of on to off is 2×10^{8}. The subthreshold slope is 400mV/dec.

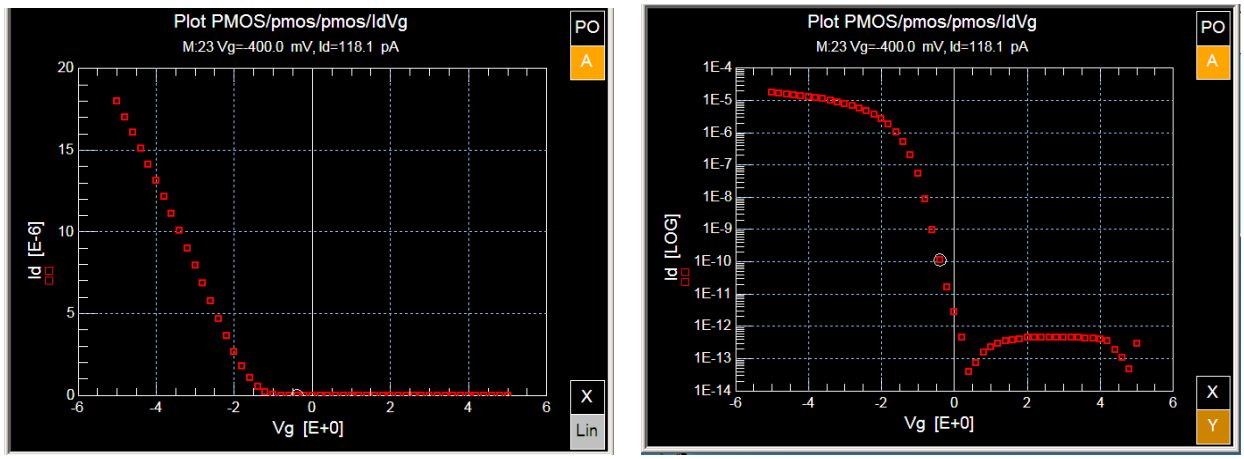


Fig.7.2: Linear and logarithmic Id-Vg curve of fabricated pMOS transistors.

Fabricated double gate transistors are working and bottom gate can control the threshold voltage of top gate. Fig.7.3 shows Id-Vg curve of nMOS double gate transistors with different voltages applied to bottom gate. As it can be seen in this curve, transistor parameters such as S, I_{on}, I_{off} and thereshold voltage are changing with biasing bottom gate.

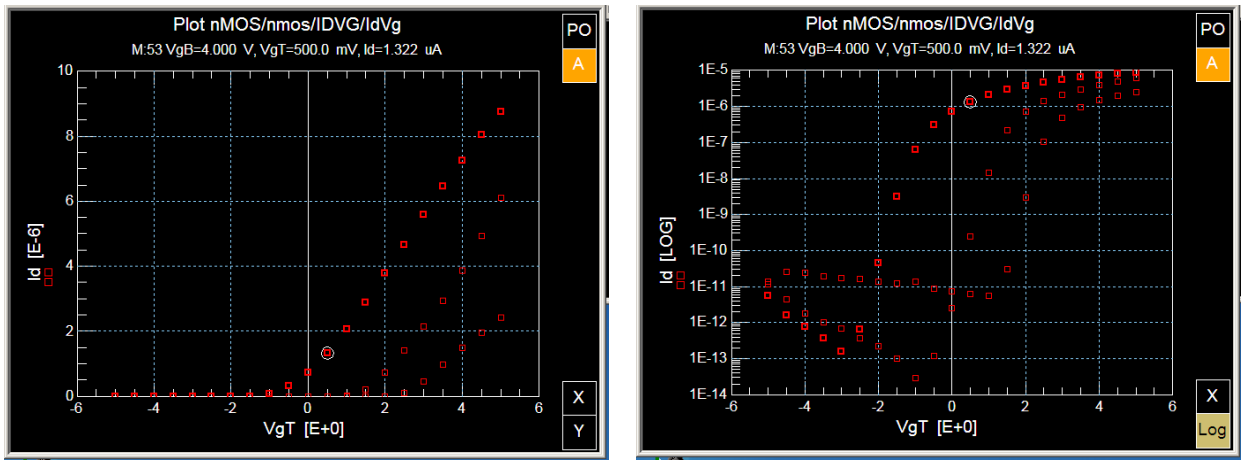


Fig.7.3 Id-Vg curve of nMOS double gate transistors

Table 7.1 shows the value of parameters with biasing bottom gate in nMOS transistors. It can be seen that with positive biasing bottom gate threshold voltage and I_{off} or leakage current decreses dramatily. On current is increasisng with positive biasing.

Table 7.1: The measured parameters of double gate nMOS transistors with bottom gate biasing

Bottom Gate (V)	Vth (V)	I_{off} (A)	I_{on} (A)
+4	-0.5	10^{-14}	10^{-5}
0	1.5	10^{-14}	5×10^{-6}
-4	2.5	10^{-11}	10^{-6}

Fig7.4 shows Id-Vg curve of double gate pMOS transistors. As it can be seen in linear curve threshold voltage is changing with biasing bottom gate. Same as nMOS with positive biasing bottom gate, the threshold voltage gets more negative. Moreover it improves both off current and subthereshold slope (S).

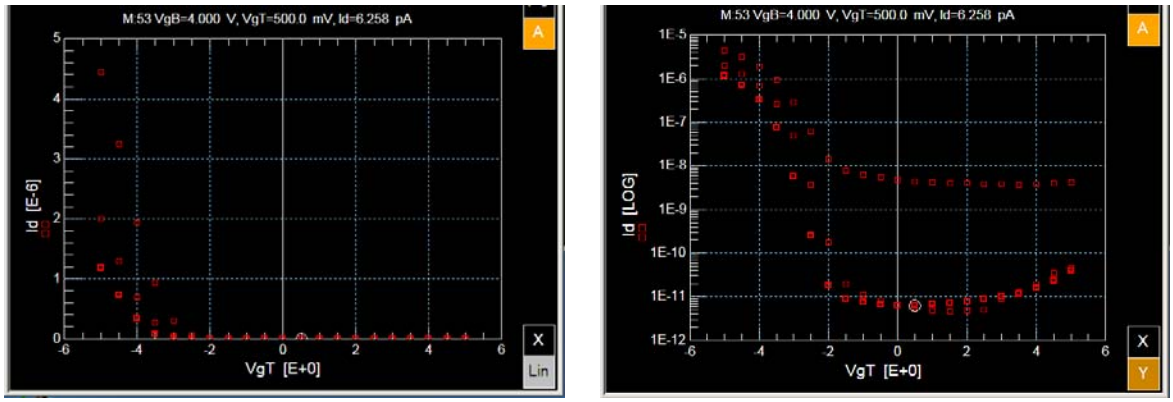


Fig.7.4: Id-Vg of pMOS double gate transistor.

Table 7.2 shows the measured data of double gate pMOS transistors.

Table 7.1: The measured parameters of double gate pMOS transistors with bottom gate biasing

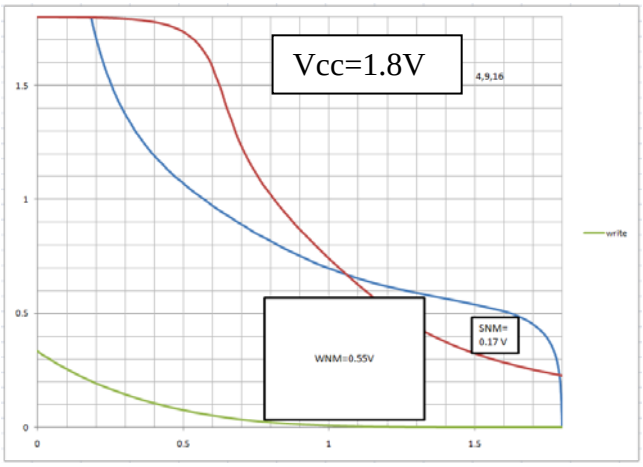
Bottom Gate (V)	Vth (V)	I_{off} (A)	I_{on} (A)
+4	-3.5	10^{-11}	10^{-6}
0	-3	10^{-11}	5×10^{-6}
-4	-2.5	5×10^{-9}	10^{-5}

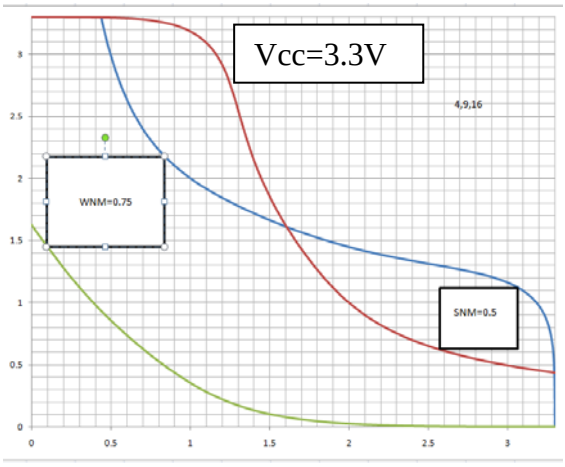
7-2) One Layer SRAM Cells

We performed DC and ac measurements to characterize SRAM cells. In DC measurement we look to the inverter characteristics to find SNM. To find read SNM we connect WL, BL and BLB to Vcc and we apply DC voltage between 0 to Vcc to “out” and we measure the “outbar” voltage. We plot outbar versus out. Then again we connect WL, BL and BLB to Vcc and “outbar” to DC voltage between 0 to Vcc and we measure “out” voltage. (“out” and “outbar” are outputs of two inverters). This time again we plot outbar versus out in same curve. This will give us butterfly curves. To find write SNM we connect BL and WL to Vcc and BLB to GND. Then we sweep the voltage source in “out” node between 0 to Vcc and the voltage of “outbar” node is measured. We draw this voltage in the read SNM curve.

The SNM of a CMOS SRAM cell is defined the minimum DC noise voltage necessary at both of the two cell storage nodes, during a read access, to flip the state of a cell. The smaller side of butterfly curve where a maximum square nested between the static characteristics of the two cell inverters are equal to cell read SNM. The minimum square nested between the write curve and one of the read static characteristics is write SNM.

Fig.7.5 shows the result of measurement for one of the designed cells at different supply voltages. This SRAM cell has W_{access}=4µm, W_{driver}=9µm and W_{pullup}=16µm. At Vcc=1.8V we have 0.55V write SNM (WNM) and 0.17V read SNM. With incresing supply voltage to 3.3V both WNM and SNM increse to 0.75V and 0.5V, respectively . Further increasing supply voltage to 5V gives WNM=0.65V and SNM=0.7V.





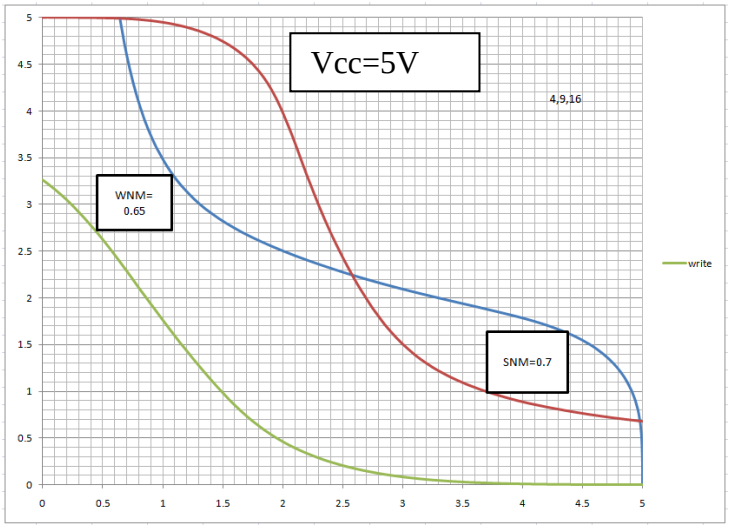


Fig.7.5: Measured WNM and SNM of an SRAM Cell at Vcc=1.8, 3.3 and 5V. W_{access}=4µm, W_{driver}=9µm and W_{pullup}=16µm

Table 7.2 shows the summary of measured parameters for this cell.

Table 7.2: Measured WNM and SNM for a cell with W_{access}=4µm, W_{driver}=9µm and W_{pullup}=16µm

V_{CC}	1.8 V	3.3 V	5 V
SNM	0.17 V	0.5 V	0.7 V
WNM	0.55 V	0.75 V	0.65 V

Fig.7.6 shows an SRAM cell that is not writable. In simulation this cell was not writable too. The size of cell is W_{access}=4µm, W_{driver}=24µm and W_{pullup}=16µm. In this case write SNM is -0.45V. It means if we want to write 0 in this cell we should give negative voltage to BL.

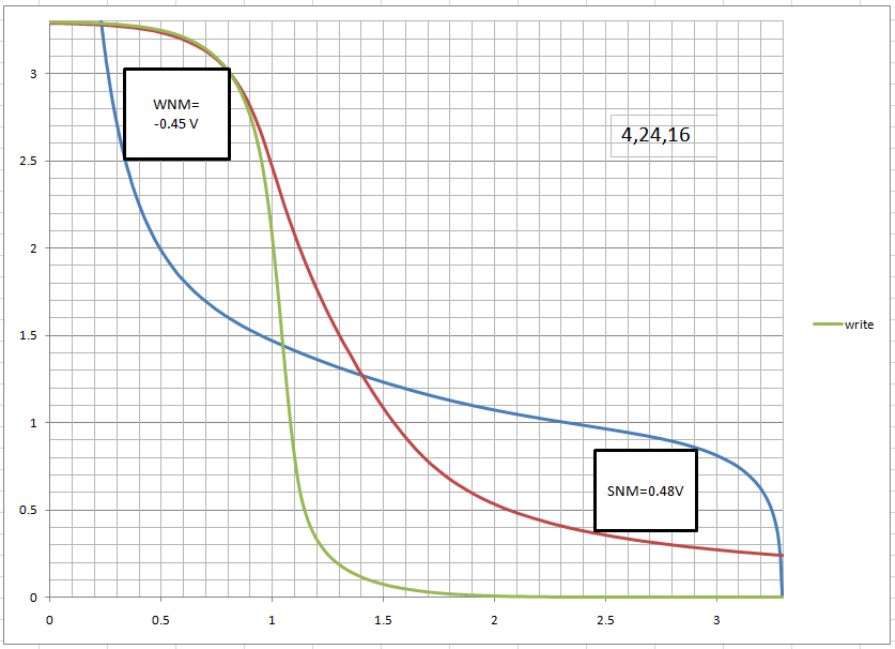


Fig.7.6: An SRAM cell that is not writable. Vcc=3.3V. W_{access}=4µm, W_{driver}=24µm and W_{pullup}=16µm.

Another SRAM cell that has W_{access}=2.6µm, W_{driver}=5.4µm and W_{pullup}=8µm is shown in Fig.7.7. Both write SNM (WNM) and read SNM (SNM) increase with increasing Vcc.

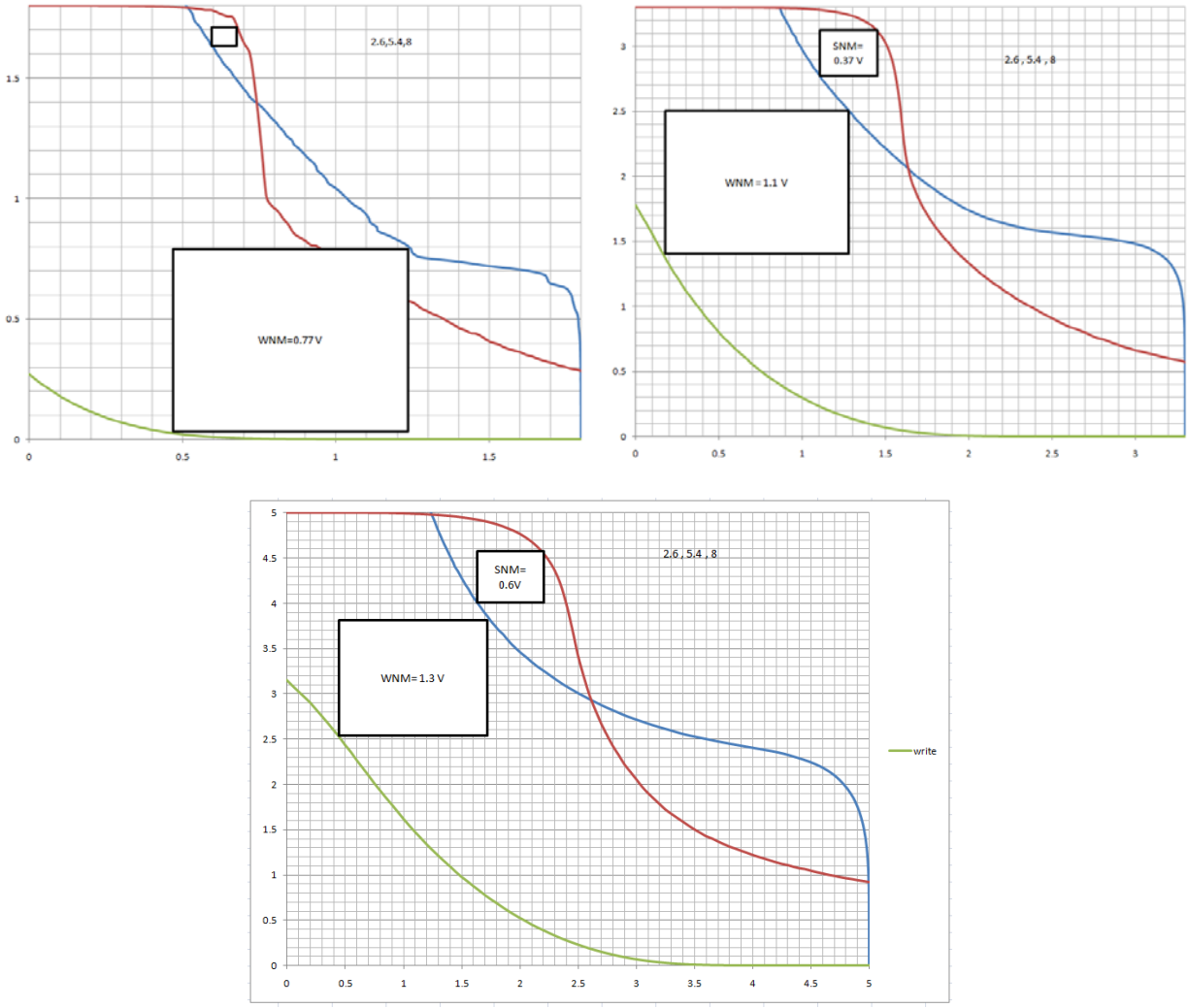


Fig.7.7: Measured WNM and SNM of an SRAM Cell at Vcc=1.8, 3.3 and 5V. W_{access}=2.6µm, W_{driver}=5.4µm and W_{pullup}=8µm

The measured SRAM cells have been compared with simulation results at table 7.3. As it can be seen there is a good consistency between simulation and measurement results. The rest of the SRAM cells are being measured.

Table7.3: Comparing measurement and simulation results for different cells.

W-access	W-nMOS (Drivers)	W-pMOS (Pull Up)	CR	PR	Read SNM: Simulation	Write SNM: Simulation	Read SNM: Measurement	Write SNM: Measurement
2	2	2	1	1	0.3	1.3	0.1	1.2
2	4	2	2	1	0.4	1.1	0.2	1.3
2	4	3	2	1.5	0.45	0.9	0.2	Big (not measured)
2	2	10	1	5	0.45	0.3	0.12	1.2

2.6	5.4	8	2	3	0.5	0.4	0.37	1.1
2.6	8	5.4	3	2	0.5	0.6	0.12	1.1
2.6	6.6	5.4	2.5	2	0.5	0.65	0.12	1.1
4.4	5.4	5.4	1.2	1.22	0.35	1.1	----------	----------
4.4	5.4	16	1.2	3.6	0.5	0.4	0.05	1.1
2.6	10.8	9.6	4.15	3.69	0.6	0.1		
2.6	10.8	6.6	4.15	2.54	0.6	0.4		
2.6	13.4	9.4	5.15	3.6	0.6	0.1		
2.6	10.6	4	4	1.5	0.5	0.8		
2.6	10.6	5.4	4	2	0.55	0.7		
4	10.6	9.4	2.65	2.35	0.6	0.6		
4	13.4	9.4	3.35	2.35	0.6	0.5		
2.6	24	8	9.23	3	0.7	0.1		
2.6	24	16	9.23	6.15	0.8	Not writable		
4	24	16	6	4	0.7	Not writable	0.5	Not writable(-0.5)
4	9	16	2.25	4	0.6	0.2	0.5	0.75

As we mentioned in layout chapter we have designed poly TFT and resistor load SRAM cells. Fig.7.8 shows one of the measured poly TFT load SRAM cell. The measured SNM is very low in comparision with single grain TFT. This cell has W_{access}=3.3µm, W_{driver}=4µm and W_{pullup}=4µm.

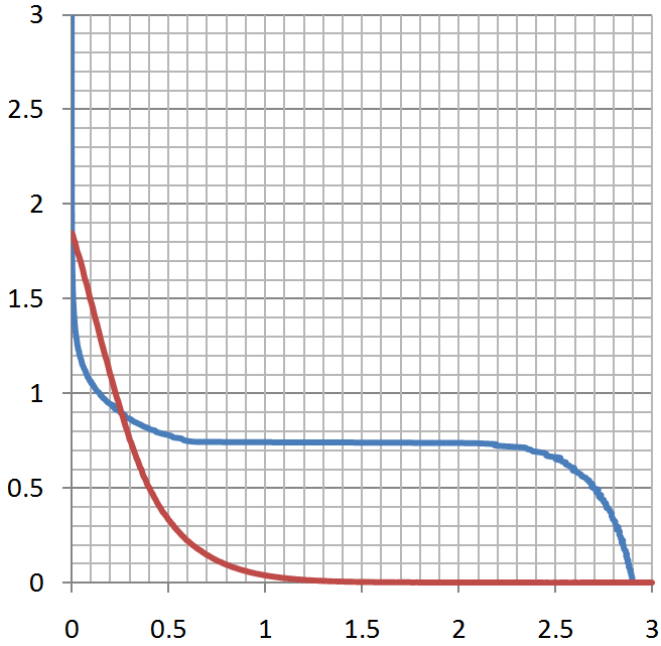


Fig.7.8: Measured SNM for poly TFT load SRAM cell. This cell has W_{access}=3.3µm, W_{driver}=4µm and W_{pullup}=4µm.

The rest of designed SRAM cells are being measured.
Figure 7.9 shows waveforms of transient measurement of an SRAM cell. First waveform is given to WL. Second waveform is write signal (Wr) that enables writing. We give Wr signal to the external three state buffer to enable the bit and bitbar generated inputs to write in the cell. During read operation Wr signal is 0. Therefore, when Wr is high we write “Bit” and “Bitbar” waveforms (that are inverterd etch other) to the SRAM and when Wr is low we read the stored data in SRAM. Third waveform is generated Bit signal from pulse generator that we want to write. Finally last signal is Bit that shows both write and read states. As it can be seen from this picture when WL is high and Wr is low, stored data could recover successfully for both 0 and 3V stored voltages. We measured the circuits until 1MHz without changing the output signals. This shows SG-TFTs can operate at higher frequencies. Measurements with high frequency pulse generators are in investigation.

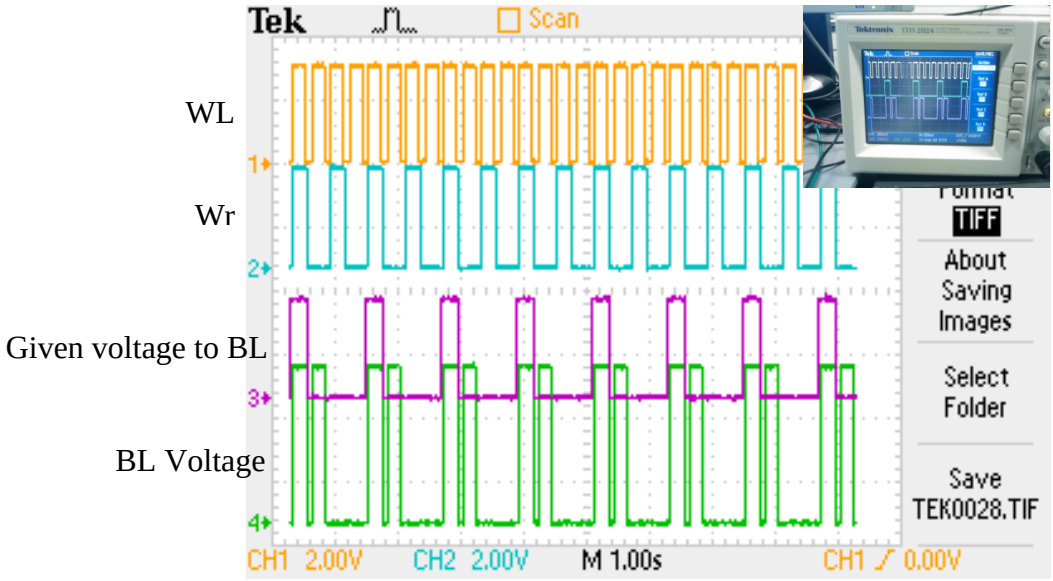


Fig.15. Output signals of SRAM cells

7-3) Conclusions

In this chapter we showed the measurement results of designed SRAM cells. The result of DC measurements for write and read SNM are consistent with simulation results. Transient measurement shows working SRAM in all three modes of operation (hold, read and write). Measurering the rest of the SRAM cells and double gate and H-Gate SRAM cells and their transient behaivieral are being pursued.

Conclusions:

In this thesis we have designed and fabricated SRAM cells on one layer and two layers of single grain silicon. Firs we most popular SRAM cells have been investigated. 6TSRAM cell has many advantages such as low power operation, better read and write stability, suitable for scaling and better performance over the other styles. The most important design metrics of SRAM cell such as read SNM, write SNM, read access time and write access time, hold or retention voltage and leakage sources have be defined. The design procedure of this type SRAM cell was discussed. We saw that there is a trade off in transistors sizing considering read and write operation modes and area penalty. Double gate transistors could improve transistors parameters such as subthreshold slope and leakage current. Using double gate and H-Gate transistors static noise margin of SRAM cell can be improved by different biasing of bottom gate. Furthermore we saw layout designing is critical in SRAM as we want to have high density memory without sensitivity to process variation. OPC algorithms can improve SRAM stability. Then we talked about different sense amplifier circuits to use in read operation modes. Integrating SRAM cells in three dimensional can improve the performance, density in lower area. It can solve global interconnection problems. Main issues in 3DIC are heat management and fabrication process and maybe design tools. µ-Czochralski process offers high performance devices and circuits that easily can be used in 3DIC. Other process like wafer and chip level bonding can offer same quality devices but wafer thinning, alignment accuracy and fabrication of interconnections are main issues in these process. Next we have designed the SRAM cells by analytical and simulation approaches to have good read and write SNM cells. Various transistor sizing are considered to compare their characteristics. Double gate transistors can improve both read and write margins. The size of double gate transistors was designed to improve read and write margins. Sense amplifier is designed to have minimum delay and high common mode rejection ratio. Finally output buffers are designed to drive 25pF load of measurement system. Layout of one layer and two layers of single grain silicon were discussed and compared in term of area and number of VIA interconnections. Having nMOS transistors on top layer and pMOS transistors on bottom layer can save the number of masks and have less interconnection VIA. Then we explained the fabrication process of planar and 3DIC using single grain TFTs shortly. We showed the optical and SEM images of fabrication process steps. Electrical characterization showed working devices with high SNM for read and write. The result of DC measurements for write and read SNM are consistent with simulation results. Transient measurement shows working SRAM in all three modes of operation (hold, read and write). Measurering the rest of the SRAM cells and double gate and H-Gate SRAM cells and their transient behaivieral are being pursued.

References……

[1] Book: “CMOS MEMORY CIRCUITS”, Tegze P. Haraszti

[2] Book: “ULSI devices”, C.Y. Chang and S.M.Sze

[3] Book: “CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies” Andrei Pavlov. Manoj Sachdev

[4] http://www.eeherald.com/section/design-guide/esmod15.html

[5] Book: “Ultra-Low Voltage Nano-Scale Memories”, Itoh, Kiyoo; Horiguchi, Masashi; Tanaka, Hitoshi (Eds.)

[6] S-M Jung, J. Jang, W. Cho, J. Moon, K. Kwak, B. Choi, B. Hwang, H. Lim, J. Jeong, J. Kim, and K. Kim, “The revolutionary and truly 3-dimensional 25F^2 SRAM technology with the smallest S^3 (Stacked Single-crystal Si) cell, 0.16μm^2,and SSTFT (Stacked Single-crystal Thin Film Transistor) for ultra high density SRAM,” Symp. VLSI Tech. Dig. Tech. Papers, pp. 228–229, June 2004.

[7] H-J An, H-Y Nam, H-S Mo, J-P Son, B-T Lim, S-B Kang, G-H Han, J-M Park, K-H Kim, S-Y Kim, C-K Kwak, and H-G Byun, “64Mb mobile stacked single crystal Si SRAM(S^3RAM) with selective dual pumping scheme (SDPS) and multi cell burn-in scheme (MCBS) for high density and low power SRAM,” Symp. VLSI Circuits Dig. Tech. Papers, pp. 282–283, June 2004.

[8] Y. H. Suh, H. Y. Nam, S. B. Kang, B. G. Choi, H. S. Mo, G. H. Han, H. K. Shin, W. R. Jung, H. Lim, C. K. Kwak, and H. G. Byun, “A 256Mb synchronous-burst DDR SRAM with hierarchical bit-line architecture for mobile applications,” ISSCC Dig. Tech. Papers, pp. 476–477, Feb. 2005.

[9] http://smithsonianchips.si.edu/ice/cd/MEM96/SEC08.pdf

[10] Book: “SRAM Circuit Design”, Bastien Giraud, Olivier Thomas, Amara Amara, Andrei Vladimirescu, and Marc Belleville

[11] Course lecture, “Computer Systems Laboratory, Stanford University”, Mark Horowitz

[12] Book: “Embedded SRAM Design in Nanometer-Scale Technologies”, Hiroyuki Yamauchi

[13] “6-T SRAM cell design with nanoscale double-gate SOI MOSFETs: impact of source/drain engineering and circuit topology” Rashmi, Abhinav Kranti and G Alastair Armstrong

[14] Book: “Dynamic and Adaptive Techniques in SRAM design”, John J. Wuu

[15] Book: “Planar Double-Gate Transistor from Technology to Circuit”, Amara Amara, Olivier Rozeau

[16] Book: “FinFETs and Other Multi-Gate Transistors”, J.-P. Colinge

[17] http://en.wikipedia.org/wiki/Soft_error

[18] E. Seevinck Sr., F. J. List, and J. Lohstroh. “Static-noise margin analysis of MOS SRAM cells”, *IEEE Journal of Solid-State Circuits*, 22:748–754, October 1987.

[19] Course Presentation: “Robust SRAM Design Under Process Variations” Zheng Guo, Prof. BorivojeNikolić January 8, 2007.

[20] Presentation:”Static random access memories”, 변 현 근 삼성전자 **(주)** SAMSUNG

[21] Book: “Current sense amplifiers”, Bernhard Wicht

[22] Book: “Advanced CMOS cell design”, Etienne Sicard, Sonia Ben Dhia, Sonia Delmas Bendhia

[23] http://www.layouteditor.net/

[24] Kaushik Roy, Hamid Mahmoodi, Saibal Mukhopadhyay, Hari Ananthan, Aditya Bansal, and Tamer Cakici, ”Double-Gate SOI Devices for Low-Power and High-Performance Applications” , Computer-Aided Design, 2005. ICCAD-2005. IEEE/ACM International conference , Page(s): 217 - 224

[25] “Noise Margin in Low Power SRAM Cells”, S. Cserveny, J. -M. Masgonty, C. Piguet, CSEM SA, Neuchâtel, CH

[26] Endo, K. O'uchi, S.-i. Ishikawa, Y. Yongxum Liu Matsukawa, T. Masahara, M. Sakamoto, K. Tsukada, J. Ishii, K. Yamauchi, H. Suzuki, E. “Enhancing noise margins of FinFET SRAM by integrating Vth-controllable flexible-pass-gates”, 38th European Solid-State Device Research Conference, 2008. ESSDERC 2008.

[27] Oliver Thomas, et al.” A power-efficient improved-stability 6T SRAM Cell in 45nm multi-channel FET technology”, 38th European Solid-State Device Research Conference, 2008. ESSDERC 2008.

[28] Y. X. Liu, M. Masahara, K. Ishii, and E. Suzuki, “Double-Gate FinFET Innovation: From 3-Terminal to Flexible Threshold Voltage 4-Terminal“,Abs. 858, 206th Meeting, © 2004 The Electrochemical Society, Inc.

[29] S. Eminente, S. Cristoloveanu, et al, “Ultra-thin fully-depleted SOI MOSFETs: Special charge properties and coupling effects”, Solid-State Electronics, Volume 51, Issue 2, February 2007, Pages 239-244

[30] Huaxin Lu,Wei-Yuan Lu and Yuan Taur, “Effect of body doping on double-gate MOSFET characteristics”, Semicond. Sci. Technol. 23 (2008) 015006 (6pp)

[31] Keunwoo Kim Kuang, J.B. Gebara, F. Ngo, H.C. Ching-Te Chuang Nowka, K.J., **“**Stable high-density FD/SOI SRAM with selective back-gate bias using dual buried oxide”, IEEE International SOI Conference, 2008.

[32] Robert Chau, Brian Doyle, Jack Kavalieros, Doug Barlage, Anand Murthy, Mark Doczy, Reza Arghavani and Suman Datta,” Advanced Depleted-Substrate Transistors: Single-gate, Double-gate and Tri-gate” Intel company

[33] www.paper.edu.cn/download_feature_paper.php?serial_number=AgilentC-02

[34] Ying Li; Guofu Niu; Cressler, J.D.; Patel, J.; Marshall, C.J.; Marshall, P.W.; Kim, H.S.; Reed, R.A.; Palmer, M.J**.,** **”**Anomalous radiation effects in fully depleted SOI MOSFETs fabricated on SIMOX”, IEEE Transactions on Nuclear Science, Volume 48, Issue 6, Dec 2001 Page(s):2146 – 2151

[35] http://www.darpa.mil/mto/programs/index.html

[36] www.leti.fr

[37] www.yole.fr

[38] K.Banerjee et al, " 3-D ICs: A Novel Chip Design for Improving Deep-Submicrometer Interconnect Performance and Systems-on-Chip Integration" Proceedings of the IEEE, Vol. 89, NO. 5, MAY 2001

[39] http://www.semiconductor.net/article/CA164243.html

[40] Roshan Weerasekera, Li-Rong Zheng,” Extending Systems-on-Chip to the Third Dimension: Performance, Cost and Technological Tradeoffs”

[41] M. Koyanagi et al, "Future system on chip LSI chips", IEEE micro, July–August 1998

[42] Kiran Puttaswamy, Gabriel H. Loh," Thermal Analysis of a 3D Die Stacked High Performance
Microprocessor", glsvlsi2006.

[43] A. W. Topol et al," Three-dimensional integrated circuits", IBM J. RES. & DEV. VOL. 50 NO. 4/5 JULY/SEPTEMBER 2006

[44] Victor W**.** C. Chan, "Three Dimensional CMOS Integrated Circuits on Large Grain Polysilicon Films", IEDM 00- 161

[45] Paul Ch. van der Wilt, B. D. van Dijk, G. J. Bertens, R. Ishihara, and C. I. M. Beenakker, Formation of location controlled crystalline islands using substrate embedded seeds in excimer laser crystallization of silicon films, Appl. Phys. Lett., Vol 79, Pages 1819, 2001

[46] R. Ishihara et al," Microstructure characterization of location-controlled Si-islands crystallized by excimer laser in the μ-Czochralski (grain filter) process **",** Journal of Crystal Growth**,** Volume 299, Issue 2, 15 February 2007, Pages 316-321

[47] M. He et al," Preparation of large poly silicon grains by excimer laser crystallization of sputtered a-Si film with a process temperature at 100^{o}C", Jpn. J. Appl. Phys. 2007, p:1245-1249

[48] Rana, V. et al," Single-Grain Si TFTs and Circuits Inside Location-Controlled Grains Fabricated Using a Capping Layer of silicon dioxide" IEEE Transactions on Electron Devices, Volume 54, Issue 1, Jan. 2007 Page(s):124–130

[49] R. Ishihara, Vikas Rana, Ming He, Y. Hiroshima, S. Inoue, Wim Metselaar and Kees Beenakker**,** "Single-Grain Si TFTs and Circuits Fabricated Through Advanced Excimer-Laser Crystallization", Solid State Electronics, 2008, 52, 353-358

[50] Ishihara, R.; Hiroshima, Y.; Abe, D.; van Dijk, B.D.; van der Wilt, P.C.; Higashi, S.; Inoue, S.; Shimoda, T.; Metselaar, J.W.; Beenakker, C.I.M," Single-grain Si TFTs with ECR-PECVD gate SiO_2" , IEEE Transactions on Electron Devices, Volume 51, Issue 3, March 2004 Page(s): 500 - 502

[51] Goplen.B, "Placement of thermal vias in 3DICs using various thermal objectives", IEEECAD April 2006

[52]V. Rana, R. Ishihara, Y. Hiroshima, D. Abe, S. Inoue, T. Shimoda, J.W. Metselaar, and C.I.M. Beenakker, "High Performance Single Grain Si TFTs Inside a Location-Controlled Grain by mu-Czochralski Process with Capping Layer", Electron Devices Meeting (IEDM) Technical Digest 5 (2005), no. 5, 919–922.

[53]V. Rana, R. Ishihara, Y. Hiroshima, D. Abe, S. Inoue, T. Shimoda, J.W. Metselaar, and C.I.M. Beenakker, "Dependence of Single-Crystalline Si Thin-Film Transistor Characteristics on the Channel Position inside a Location-Controlled Grain", IEEE Trans. Electron Devices 52 (2005), no. 12, 2622–2628.

[54] N. Saputra, M. Danesh, A. Baiano, R. Ishihara, J. R. Long, N. Karaki and S. Inoue*,"* An Assessment of μ-Czochralski, Single-Grain Silicon Thin-Film Transistor Technology for Large-Area, Sensor and 3D Electronic Integration" **,** IEEE Journal of Solid State Circuits, 2008, 43, 7, 1563-1576

Fabrication of 6T SRAM cell using single grain TFTs obtained by μ-Czochralski process

Negin Golshani, R. Ishihara, J. Derakhshandeh, and C.I.M Beenakker
Delft University of Technology, Delft Institute of Microsystems and Nanoelectronics (DIMES), ECTM,
Feldmannweg 17, P. O. Box 5053, 2600 GB Delft, the Netherlands
Phone: +31-15-2786104 Fax: +31-15-2622163 E-mail: n.golshani@student.tudelft.nl

In this paper we will report successfully fabricated 6T SRAM cells using high quality single-grain Thin-Film Transistors (TFTs). TFTs are fabricated by μ-Czochralski process including making grain filter and excimer laser crystallization. Either the single grain or poly silicon thin film transistors has been used for pMOS pull up and compared in terms of output parameters of SRAM cells. Fabricated SRAM cells based on single grain TFTs, show good static noise margin equal to 0.6V at 3V power supply and also an excellent read and write speeds in 1.5μm gate length of TFTs.

1. INTRODUCTION

SRAM cells are used in several applications where high speed and low power consumption are needed such as cache memory of CPU, FPGA, specific ICs, home appliances and many of other usages.

For large area and low temperature applications Epson Company has developed a flexible and low temperature SRAM with poly silicon TFT technology. [1)] However, the speed of fabricated SRAM in poly silicon technology is low (around 2MHz), due to the low mobility of the poly-Si.

In this study we have used single grain thin film transistors which have been fabricated in low temperature process called μ-Czochralski process. These TFTs have high mobility and high frequency operation speeds. The mobility of nMOS TFTs are around 600cm^2/VS and for pMOS is 300cm^2/VS.[2,3)] These types of TFTs have the same advantages as fully depleted SOI devices. Having simple layout designing, avoiding Latch up problem, no diode junctions for source and drain (S/D) areas, fast operation frequency are advantages of this technology.[4,5)]

We have developed this technology for analog and RF applications. Several OPAMP structures, voltage reference and cascode RF amplifies have been fabricated successfully using this technology. [6)] They all were analog circuits.

In this paper we report fabrication of SRAM circuits using single grain TFTs with 1.5μm gate length. Fabricated SRAMs have 0.6V SNM at 3V supply voltage and high speed for both read and write operations. These cells can work in 1.8V supply voltage as well.

2. μ-Czochralski process

Fig.1 shows 3D schematics of μ-Czochralski process. First 700nm thermal oxide is grown on <100> p-type silicon wafer. After making 1μm holes on oxide, 780nm second oxide is deposited by PECVD with TEOS source at 350^{o}C to make holes as narrow as possible. In average they are in range of 0.1μm and called grain filter. Then 250nm amorphous silicon is deposited by LPCVD at 550^{o}C. Silicon layer penetrates to the bottom of grain filter. During excimer laser crystallization at 400^{o}C and 1500mJ/cm^2 laser energy, silicon layer is melted in the surface except in the deepest part of grain filter. Solid silicon in this region acts as a seed for crystallization. The crystallized areas are single grain and we design TFTs inside this grains.[2,3,7)]

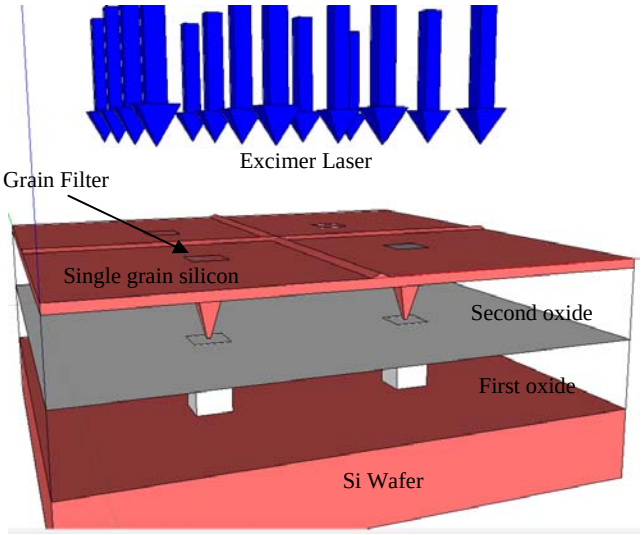


Fig.1. μ-Czochralski process

Fig.2 shows a sample SEM picture of crystallized region. As we can see the sizes of grains are in the order of 7μm. Brighter points are locations where four grains meet each other. Thin film transistors are designed inside single grain area to obtain single crystalline equivalent characteristics for devices. In designing circuits we placed channel of transistors in 0.5μm distance from away the grain filter to obtain a high mobility and also good stability for threshold voltage. For larger width of transistors that are more than 5 μm, we put many of them in parallel.

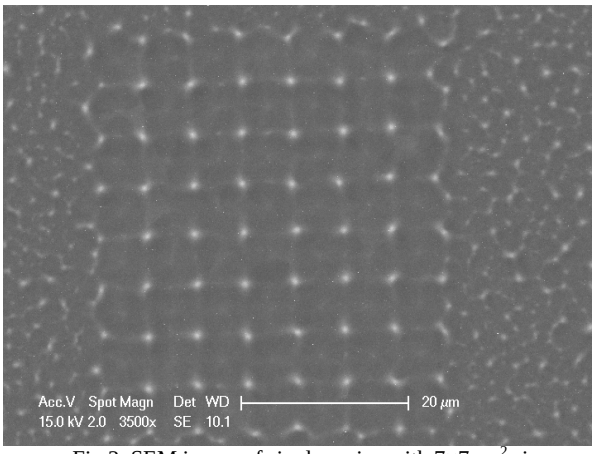


Fig.2. SEM image of single grains with 7×7μm^2 size

3. Designing 6T SRAM cells

We designed conventional 6T SRAM structure for our process. In 6T structure that is shown in Fig. 3, we have two cross coupled inverters as a positive feedback to store a value. To write and read the cell, two access transistors M5 and M6 are used. With connecting word line (WL) to supply voltage (Vcc), inverters are connected to the bit line (BL) and bit line bar (BLB) columns.

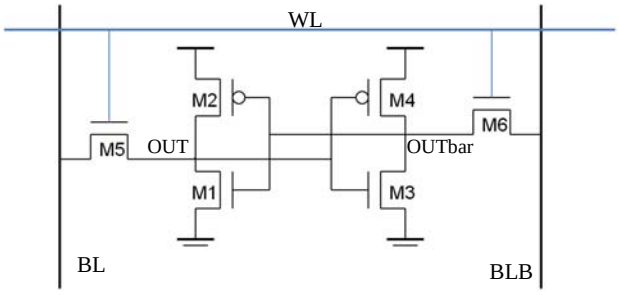


Fig. 3. 6T SRAM cell

For high speed and low power consumption SRAM cells, lower capacitance of source and drain is needed because of connection between bit lines of cells in column. Lower capacitance increases memory performance.[4,5] Single grain TFTs have this requirement that makes it suitable for this application. Moreover, to reduce power consumption it is recommended to connect precharge transistors to half Vcc.[8]
In designing SRAM cells there is a trade-off between requirements of read and write states. In this design we used the values of our process parameters where threshold voltage of pMOS (Vtp) was -1 V, for nMOS (Vtn) was 0.5 V, mobility of pMOS (μ_p) was 100 cm^2/VS and for nMOS (μ_n) was 300 cm^2/VS. During read operation mode we should be able to read the stored value without disturbing the data. Cell Ratio (CR) is defined as a driving current of M1 divided by driving current of M5.

$$CR=(\mu_{n1}/\mu_{n5})[(Vcc-V_{tn1})^2/(Vcc-V_{tn5})^2](W_{M1}/L_{M1})/(W_{M5}/L_{M5})$$

In this formula W and L are the width and length of transistors, respectively.
In table 1 different CR are shown in various supply voltages. To avoid read-disturb, the voltage on node OUT should remain below the trip point of the inverter pair (Vswitching) for all process, noise, and operating conditions. The voltage of OUT point in Fig. 3 is computed for different Vcc. In order to have the right data for zero state, CR should be more than the values that are mentioned in table 1. Then obtained values for width of nMOS transistors are written in the last raw. This means, for instance at Vcc=3V the width of nMOS in inverters should be more than twice width of access transistors. Also it means in lower supply voltage we need less area.

Table 1. Read mode: CR values computed for different Vcc

Vcc	5 V	3 V	1.8 V
V_{out}	0.475 V	O.458 V	0.5
Vswitching	1	0.98	0.6
CR	4 {(w1/w5)>4}	2 {(w1/w5)>2}	0.7 {(w1/w5)>0.7}

In write operation mode, in order to write the cell, the access transistor M6 must be more conductive than the M4 to allow node OUTbar to be pulled to a value low enough for the inverter pair (M2/M1) to begin amplifying the new data. Here is the equation for pull up ratio (PR) which is driving current of M4 divided by driving current of M6.

$$PR=(\mu_{p4}/\mu_{n6})[(Vcc-V_{tp4})^2/(Vcc-V_{tn6})^2](W_{M4}/L_{M4})/(W_{M6}/L_{M6})$$

Table 2 shows designed values for PR. The maximum ratio of the pull up size is required to guarantee that the cell is writable where M6 is in linear and M4 in saturation.

Table 2. Wire mode: Designed PR for different Vcc

Vcc	5 V	3 V	1.8 V
V_{outbar}	0.5	0.5	0.5
Vswitching	1	0.98	0.6
PR	1.04 {(w4/w6)<3.95}	3 {(w4/w6)<14}	23 {(w4/w6)<182}

Therefore from both tables the width of nMOS and pMOS in inverters and access transistors should meet W1>2×W5 and W2<14×W5 for 3V supply voltages. As we can see one needs less W5 and one needs more W5. We have trade-off between these two conditions. From area considerations we take minimum value for nMOS access transistors and we compute the rest.
Fig. 4 illustrates the simulated SNM values for two different sizes of SRAM. Simulation was done in ADS software using a modified BSIM-SOI model extracted from measurements of experimental single grain TFTs.[6] The left picture in figure 4 is for 3.3μm, 4μm and 4μm width of access, nMOS and pMOS transistors, respectively. The right one is for 1.5μm, 4μm and 48μm width of access, nMOS and pMOS

transistors, respectively. As it can be seen larger sizes of transistors give better SNM compared with small sizes.

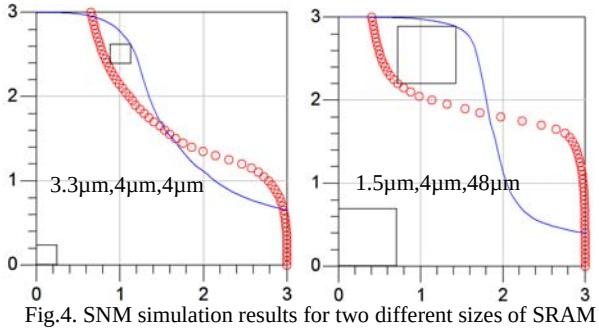


Fig.4. SNM simulation results for two different sizes of SRAM

Designed layout in LEDIT software is shown in Fig. 5. The small green squares are locations of grain filters and pMOS transistors are placed in orange area. The precharge transistors are pMOS.

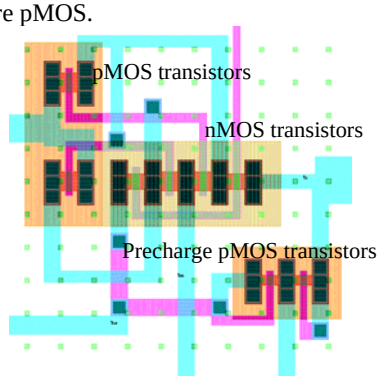


Fig.5. Designed layout of SRAM

Moreover, we designed same cells with poly pMOS transistors to compare the results. In this case we removed grain filter from pMOS area to obtain poly silicon instead of location controlled grains.
The schematic of designed sense amplifier is shown in Fig.6. Designed sense amplifier has 60dB gain and 160MHz unity gain bandwidth.

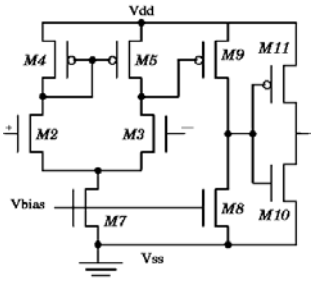


Fig.6. Schematic of used sense amplifier

Table 3 shows the designed width of transistors. The gate length is 1.5µm.

Table 3. Designed width of transistors for sense amplifier

	M2,3	M4,5	M7,10	M8	M9	M11
Width	96µm	20µm	2µm	4µm	56µm	24µm

4. Fabrication process

As we discussed in second part, µ-Czochralski process is used to fabricate SRAM cells. After crystallization of silicon and forming single grains, Island mask is applied to define the silicon islands using dry etching process. Cleaning before depositing gate oxide is very important. Maragoni cleaning is used and immediately after cleaning we use TEOS PECVD oxide to deposit 30nm gate oxide followed by 675nm Al gate. Using gate mask we etched the Al and oxide by Dry etching to reach to S/D. Next step is implantations for nMOS and pMOS transistors, respectively and then activation of dopants using excimer laser at room temperature and 300mJ/cm^2 laser energy is followed. Then 800nm TEOS oxide is deposited on wafers and contact hols are opened to have access to gate, source and drain. Finally 1.5µm Al is deposited and patterned using metal 1 mask.
Figure 7, 8 show two SEM images of SRAM cells for different sizes corresponding to layout shown in Fig.5.

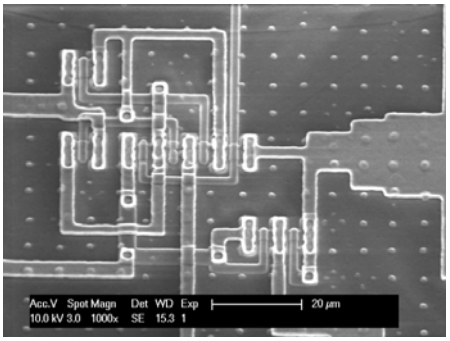


Fig.7. SEM image of small size cell

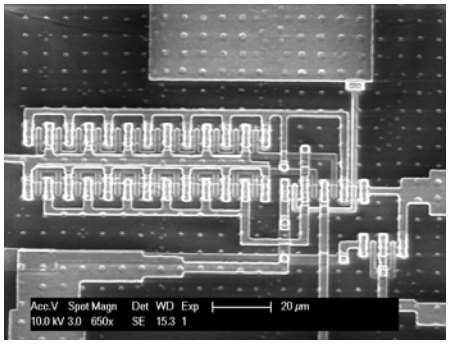


Fig.8. SEM image of big size cell

The gate length of all transistors is 1.5µm. Minimum size of contact openings are 2.5µm and minimum width of metal layer is 2µm. The area of small size cell is 40µm×40µm and for bigger one is 120µm×40µm.

5. Measurement results

We performed DC and ac measurements to characterize SRAM cells. In DC measurement we look to the inverter characteristics to find SNM. To find read SNM we connect WL, BL and BLB to Vcc and we apply DC voltage between 0 to Vcc to out and we measure the outbar voltage. We plot

outbar versus out. Then again we connect WL, BL and BLB to Vcc and outbar to DC voltage between 0 to Vcc and we measure out voltage. (out and outbar are outputs of two inverters). This time again we plot outbar versus out in same curve. This will give us butterfly curves.
The SNM of a CMOS SRAM cell is defined as the minimum DC noise voltage necessary at both of the two cell storage nodes, during a read access, to flip the state of a cell. The smaller side of butterfly curve where two maximum squares nested between the static characteristics of the two cell inverters are equal to cell SNM.[9)]
Figures from 9 to 12 show obtained butterfly curves for four different sizes of transistors and their SNM values. The best SNM is obtained with width of transistors mentioned in Fig. 12 that is consistent with simulation results (Fig.4).

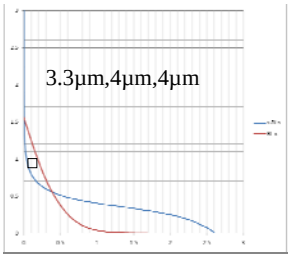


Fig.9. SNM for Wa=3.3µm, Wn,p=4µm is equal to 0.2V

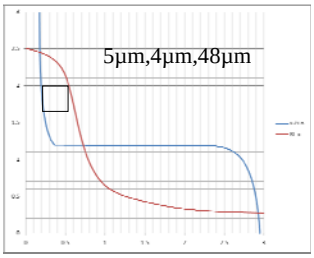


Fig.10. SNM for Wa=5µm, Wn=4µm and Wp=48µm is 0.4V

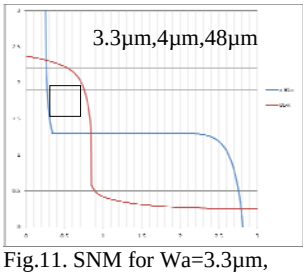


Fig.11. SNM for Wa=3.3µm, Wn=4µm and Wp=48µm is 0.5V

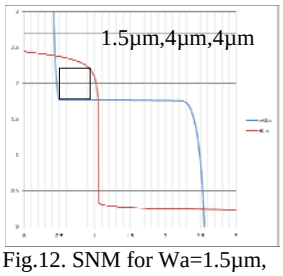


Fig.12. SNM for Wa=1.5µm, Wn=4 and Wp=48µm is 0.5V

In case of poly pMOS transistors we obtained lower SNM comapred with single garin silicon pMOS. Figure 13 and 14 show butterfly curves for poly pMOS SRAMs.

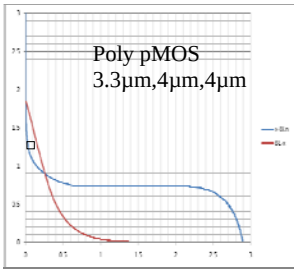


Fig.13. SNM for Wa=3.3µm, Wn,p=4µm is equal to 0.1V

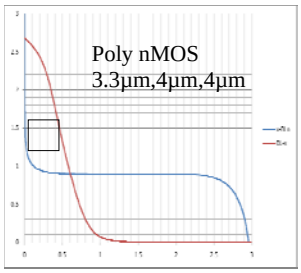


Fig.14. SNM for Wa=3.3µm, Wn=4µm and Wp=48µm is 0.4V

Figure 15 shows waveforms of tested SRAM. First waveform is given to WL. Second waveform is write signal (Wr) that we give it to the external three state buffer to enable the bit and bitbar bar generated inputs during writing and disable them during read operation. Therefore, when Wr is high we write Bit and Bit bar waveforms (that are inverterd etch other) to the SRAM and when Wr is low we read the stored data in SRAM. Third waveform is generated Bit signal from pulse generator that we want to write. Finally last signal is Bit that shows both write and read states. As it can be seen from this picture when WL is high and Wr is low, stored data could recover successfully for both 0 and 3V stored voltages. We measured the circuits until 1MHz, which was the upper limit of pulse generator, without changing the output signals. This shows SG-TFTs can operate in higher frequencies. Measurements with high frequency pulse generators are in investigation.

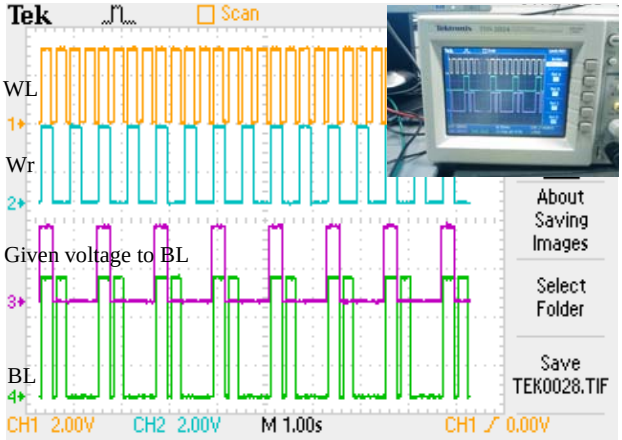


Fig.15. Output signals of SRAM cells

6. Conclusions

We reported design and fabrication of 6T SRAM cells with µ-Czochralski process. Working SRAMs show this technology is suitable for memory applications, particularly for low temperature applications. Good stability for both read and write operation modes are achieved.

Acknowledgments
We would like to express sincere thanks to all DIMES clean room staff and Prof. Nick van der Meijs from TU Delft and special thanks to Dr. Navid Azizi from University of Toronto for their helps during design, fabrication and test.

References
1) H.Ebihara, N.Karaki, et al. from Epson Company.
2) R.Ishihara, et al, Solid state electronics52 (2008) 353-358.
3)R.Ishihara, et al, IEEE Transaction on electron device, VOL. 51,NO. 3, March 2004
4) S. Park et al., IEEE J. Solid-State Circuits, vol. 34, no. 11, Nov. 1999, pp. 1436-1445.
5) J. B. Kuang et al., IEEE J. of Solid-State Circuits, vol.32, no. 6, Jun. 1997, pp. 837-844.
6) Nitz Saputra, et al, JSSC. VOL. 43,NO.7,July 2008.
7) Ming He, Ryoichi Ishihara et al, “Japanese Journal of Applied Physics” Vol. 45 No. 1A, Part 1, 2006 pp. 1-6
8) Kaushik Roy and Sharat Prasad, “ Low-power CMOS VLSI circuit design “, New York : Wiley, 2000
9) Andrei Pavlov, Manoj Sachdev, “CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled … “ , 2008,chapter 2, p. 17

High speed 6T SRAM cells using single grain TFTs fabricated by µ-Czochralski process at low temperature

Negin Golshani, Jaber Derakhshandeh, R. Ishihara and C.I.M Beenakker
Delft University of Technology, Delft Institute of Microsystems and Nanoelectronics (DIMES), ECTM, Feldmannweg 17, P. O. Box 5053, 2600 GB Delft, the Netherlands,
Phone: +31-15-2786104 Fax: +31-15-2622163 E-mail: n.golshani@student.tudelft.nl

Keywords: 6T SRAM Cell, SNM, high speed SRAM, Laser crystallization

In this paper we report designing and characterization of 6T SRAM cells using single grain thin film transistors (TFT) fabricated by low temperature µ-Czochralski process. In this process Excimer laser is used to melt amorphous silicon deposited on oxide with some holes. Crystallization starts from bottom part of the holes where silicon is still solid, and distributes over the surface to make a single square grain [1]. Single grains with size of bigger than 7um were achieved using 1500mJ/cm^2. Fig.1 shows three dimensional view of this process. TFTs are designed inside these grains to have high quality in electrical characteristics. In general to make big size transistors we use parallel transistors.

6T SRAM cell structure has low power consumption and short access time. This structure has cross-coupled CMOS inverters that make a positive feedback to store a value. Fig.2 shows 6T SRAM cell circuit. The size of transistors in SRAM cells are designed by means of analytic and simulation methods to have good read and write static noise margin (SNM). In designing a SRAM cell there is a trade-off between read and write performances. Fig. 3 shows a simulation result of a cell with access transistors of 4um, nMOS drivers of 9um and pMOS load transistors of 16um. This cell has 0.6V read SNM and 0.3V write SNM [2].

Fig.4 shows one of the fabricated complete SRAM cells with sense amplifier and output buffers. The gate length of all transistors is 2um. Fig.5 shows the result of DC measurement of SRAM cell. It can be seen that this cell has 0.5V read SNM and 0.75V write SNM for same sizes of simulated cell. The transient measurement shows excellent dynamic characteristics for SRAM cell. The measured read SNM is 12nS at 87MHz world line frequency, which has been shown in Fig.6.

References:

[1] R.Ishihara, et al, IEEE Transaction on electron device, VOL. 51,NO. 3, March 2004

[2]) J. B. Kuang et al., IEEE J. of Solid-State Circuits, vol.32, no. 6, Jun. 1997, pp. 837-844.

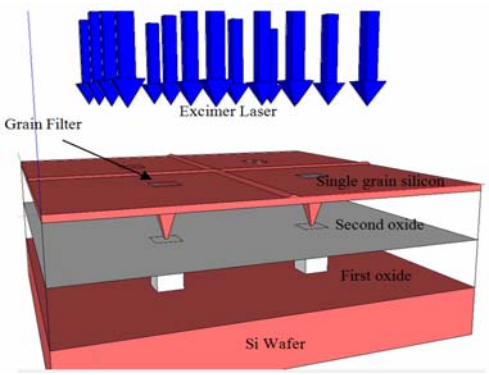


Fig. 1: Schematics of µ-Czochralski Process

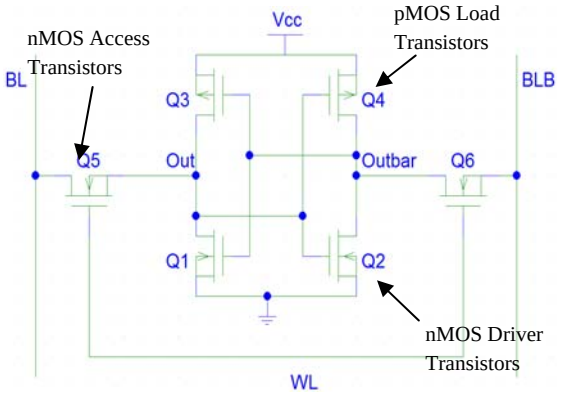


Fig.2: Schematics of 6T SRAM Cell

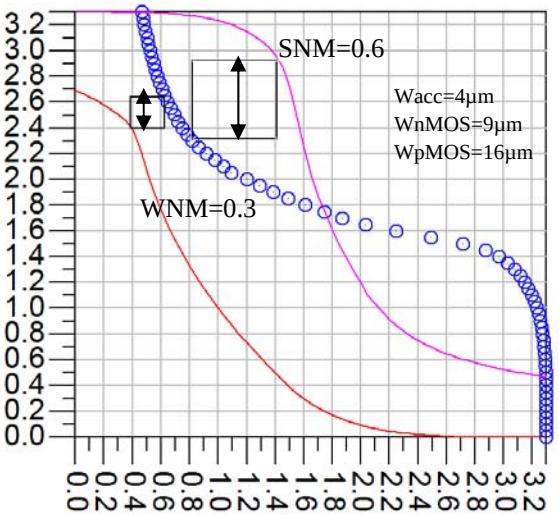


Fig.3: Simulated read and write SNM

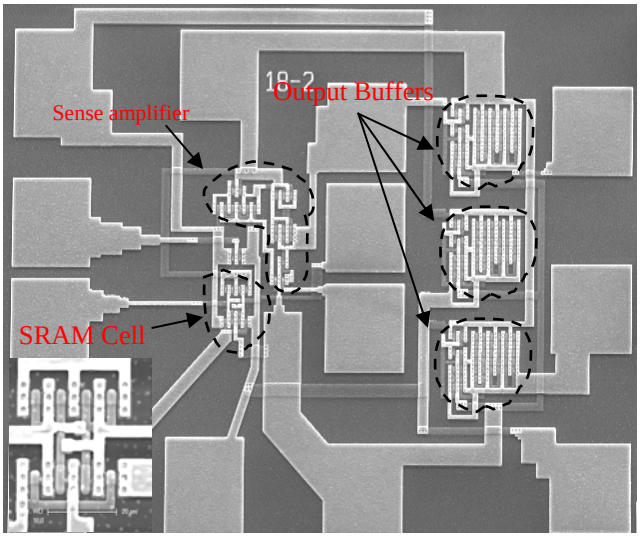


Fig.4: SEM image of fabricated SRAM cell

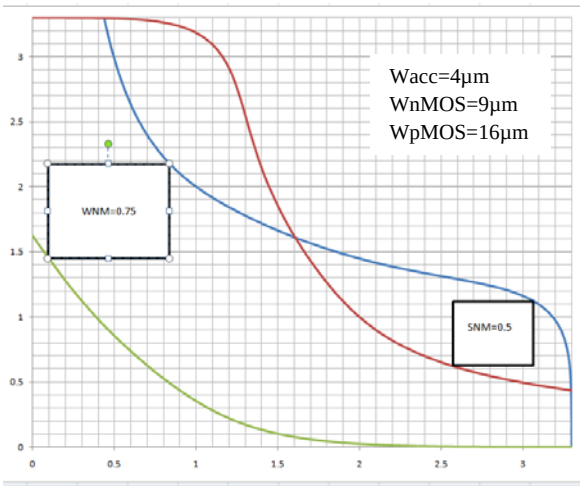


Fig. 5: Measured read and write SNM

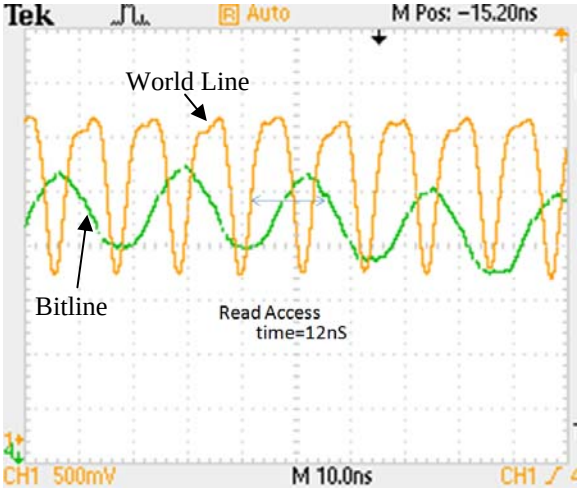


Fig.6: Measured read access time

High speed 6T SRAM cells using single grain TFTs fabricated by µ-Czochralski process at low temperature

Negin Golshani, Jaber Derakhshandeh, Ryoichi Ishihara and C.I.M Beenakker

Delft University of Technology, Delft Institute of Microsystems and Nanoelectronics (DIMES), ECTM, Feldmannweg 17, P. O. Box 5053, 2600 GB Delft, the Netherlands, Phone: +31-15-2786104 Fax: +31-15-2622163 Email:negingolshani@yahoo.com

In this paper we will report successfully fabricated 6T SRAM cells using single-grain Thin-Film Transistors (TFTs). SRAM cells have been designed by analytical calculations and verified by DC and transient simulations. TFTs are fabricated by µ-Czochralski process with 2µm gate length including making grain filter and Excimer laser crystallization at temperatures below 550°C. Fabricated SRAM cells based on single grain TFTs, show good read and write static noise margin equal to 0.55V and 0.75V at 3.3V power supply, respectively. Finally, excellent read and write access times equal to 12ns and 8ns in 87MHz worldline frequency were obtained.

Keywords: 6T SRAM Cell, SNM, high speed SRAM, read access time, write access time, Laser crystallization

1. INTRODUCTION

SRAM cells are used in several applications where high speed and low power consumption are needed such as cache memory of CPU, FPGA, specific ICs, home appliances and many of other usages.

For large area and low temperature applications Epson Company has developed a flexible and low temperature SRAM with poly silicon TFT technology. [1)] However, the speed of fabricated SRAM in poly silicon technology is low (around 2MHz), due to the low mobility of the poly-Si.

In this study we have used single grain thin film transistors which have been fabricated in low temperature process called µ-Czochralski process. These TFTs have high mobility and high frequency operation speeds. The mobility of nMOS TFTs are around 600cm^2/VS and for pMOS is 300cm^2/VS.[2,3)] These types of TFTs have the same advantages as fully depleted SOI devices. Having simple layout designing, avoiding Latch up problem, no diode junctions for source and drain (S/D) areas, fast operation frequency are advantages of this technology.[4,5)]

We have developed this technology for analog and RF applications. Several OPAMP structures, voltage reference, cascode RF amplifies and lateral photodiodes have been fabricated successfully using this technology. [6,10)] They all were analog circuits.

In this paper we report fabrication of SRAM circuits using single grain TFTs with 2µm gate length. Fabricated SRAMs have 0.55V read SNM and 0.75V write SNM at 3.3V supply voltage and read access time of 12ns and write access time of 8ns. These cells can work in 1.8V supply voltage as well.

2. µ-Czochralski process

Fig.1 shows three dimensional schematics of µ-Czochralski process. This process starts with 750nm thermal oxidation on <100> p-type silicon wafer. Next 1µm holes were formed in oxide using dry etching and then 780nm TEOS based second oxide is deposited by PECVD at 350°C to make holes as narrow as possible. In average they are in the range of 0.1µm and called grain filter. Then 250nm amorphous silicon is deposited by LPCVD at 550°C. During Excimer laser crystallization at 450°C and 1500mJ/cm^2 laser pulse energy, silicon layer is melted in the surface except in the deepest part of grain filter. Solid silicon in this region acts as a seed for crystallization. The crystallized areas are single grain and we design TFTs inside this grains.[2,3,7)]

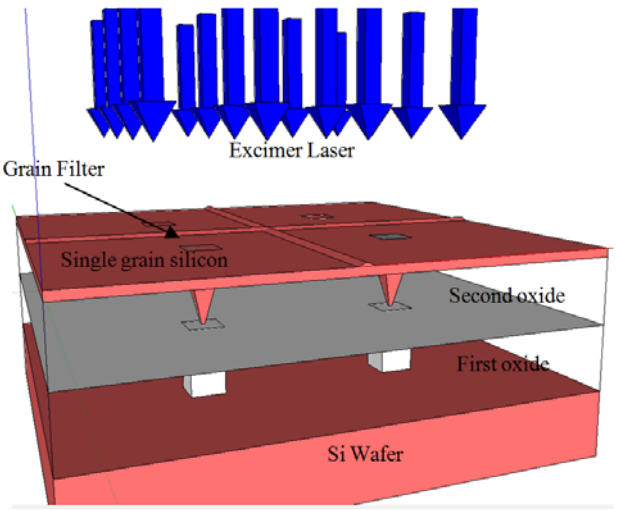


Fig.1. μ-Czochralski process

Fig.2 shows a sample SEM picture of crystallized region. As it can be seen the sizes of grains are in the order of 7μm. Brighter points are locations where four grains meet each other. Thin film transistors are designed inside single grain area to obtain single crystalline equivalent characteristics for devices. In designing circuits we placed channel of transistors in 0.5μm distance from away the grain filter to obtain a high mobility and also good stability for threshold voltage. For larger width of transistors that are more than 5μm, we used parallel transistors.

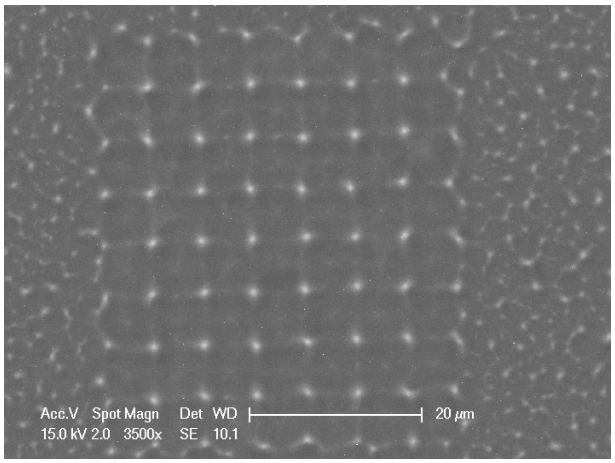


Fig.2. SEM image of single grains with 7×7μm^2 size

3. Designing 6T SRAM cells

3-1) Analytical approach

We designed conventional 6T SRAM structure for our process. In 6T structure that is shown in Fig. 3, we have two cross coupled inverters as a positive feedback to store a value. To write and read the cell, two access transistors M5 and M6 are used. With connecting word line (WL) to supply voltage (Vcc), inverters are connected to the bit line (BL) and bit line bar (BLB) columns.

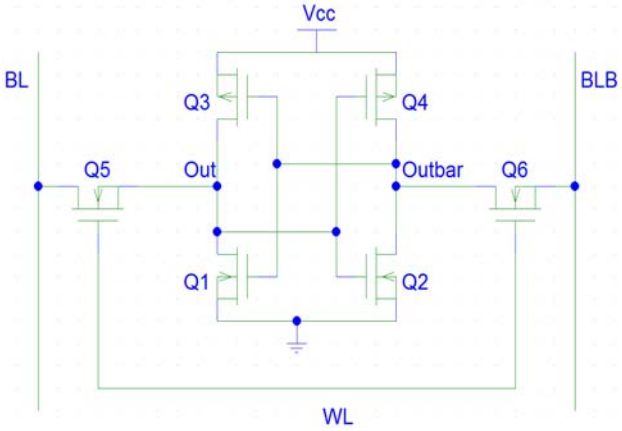


Fig. 3. 6T SRAM cell

For high speed and low power consumption SRAM cells, lower capacitance of source and drain is needed because of connection between bit lines of cells in column. Lower capacitance increases memory performance.[4,5] Single grain TFTs have this requirement that makes it suitable for this application. Moreover, to reduce power consumption it is recommended to connect precharge transistors to half Vcc.[8]

In designing SRAM cells there is a trade-off between requirements of read and write states. In this design we used the values of our process parameters where threshold voltage of pMOS (Vtp) was -1V, for nMOS (Vtn) was 0.5 V, mobility of pMOS (μ_p) was 100 cm^2/VS and for nMOS (μ_n) was 300 cm^2/VS. During read operation mode we should be able to read the stored value without disturbing the data.

In read operation the stored date in internal nodes is transferred to the BL and BLB which were precharged to Vcc. By enabling the word line (WL), one of the bitlines will pull down to GND. As shown in Fig.4 “out” node has 0 data and will pull down BL to 0. The voltage of “out” node

should not be more than Vth of Q2 or switching voltage of inverter with Q2 and Q4 transistors, plus some safety room for noise margin. Sizing of Q1 and Q5 should ensure non-destructive read.

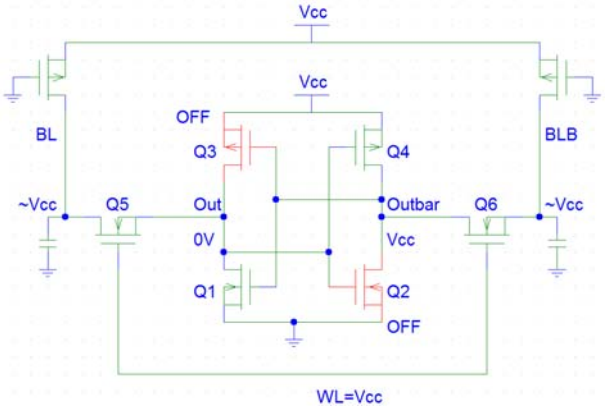


Fig.4: Read operation

In this circuit Q1 is in linear and Q5 is in saturation regions. Therefore we can calculate the increased voltage of "out" node (V_{out}) from current equivalent equation.
Cell Ratio is defined by ratio of driver transistor to the access transistor dimension:

$$\beta = CR = \frac{W1/L1}{W5/L5}$$

Then we can calculate:

$$V_{out} = (V_{cc} - V_{tn}) \frac{1+CR \pm \sqrt{(CR(1+CR))}}{(1+CR)}$$

Fig.5 shows the voltage of node "out" that should be less than Vth of nMOS drivers.

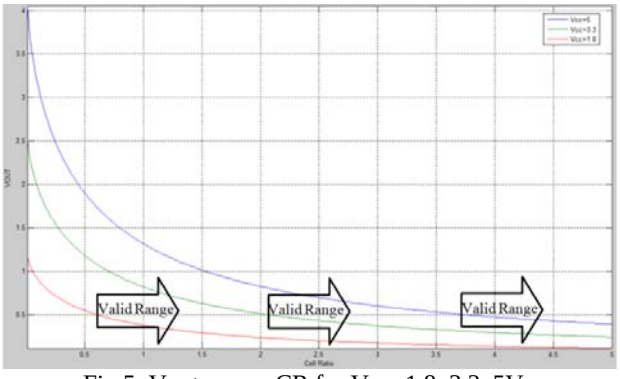


Fig.5: Vout versus CR for Vcc=1.8, 3.3, 5V

This curve describes the Vout versus CR for 1.8V, 3.3V and 5V supply voltage. It can be seen that for instance at Vcc=3.3V, CR should be more than 2.07 to have readable cell. It means the following relation should be held between access and driver transistors:

$$CR = \frac{Wd}{Wa} = \frac{W1}{W5} > 2.07 \ \ at\ Vcc = 3.3V$$

On the other hand we should be able to write the cell. Suppose node out has 1 and node outbar has 0 stored data. In this case pull up transistor Q3 is on and driver transistor Q1 is off. We want to write 0 on node out. During writing, BL in Figure 6, is driven from precharged value (Vcc) to the ground potential (0 state) and then by enabling WL, through access transistor Q5 we change the value of node V_{out}. If transistors Q3 and Q5 are properly sized, then the cell is flipped and its data is effectively overwritten. It means V_{out} should be less than V_{th2} to turn off Q2.

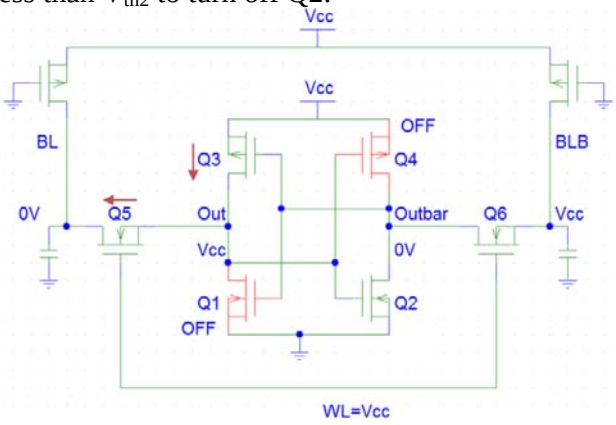


Fig.6: Write operation

As it can be seen in Fig. 6 Q5 is in saturation and Q3 is in linear region. The current of Q3 and Q5 are same and we can calculate the Vout voltage from equal current equation.
Pull up ratio (PR) is defined by ratio of pull up transistor to the access transistor dimension:

$$PR = \frac{W3/L3}{W5/L5}$$

Fig.7 shows the Vout versus PR for Vcc=1.8, 3.3 and 5V. As it can be seen PR should be less than 2.36 in case of Vcc=3.3V. It means we should

have the following relation between transistors size:

$$\mathrm{PR} = \frac{\mathrm{Wp}}{\mathrm{Wa}} = \frac{\mathrm{W3}}{\mathrm{W5}} < 2.36 \;\; at\; Vcc = 3.3V$$

From these values for PR and CR, we can see that there is a trade off in transistor sizing. Beside these values for CR and PR the area of cell should be kept as minimum as possible. Furthermore, if we design the cells for Vcc=5V those cells can successfully work at lower voltages too.

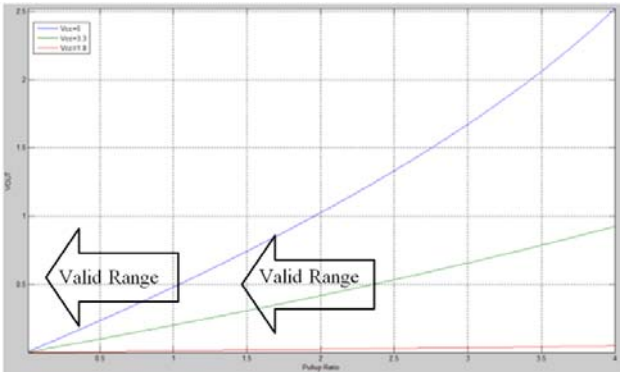


Fig.7: V_{out} versus PR for Vcc=1.8, 3.3, 5V

3-2) DC Simulations

After selecting the values for transistors sizing from calculations we measure the read and write SNM values from DC simulation. Advanced design system (ADS) software is used to simulate the cells. We have used a modified BSIM-SOI model extracted from measurements of experimental single grain TFTs. The parameters of this technology have been fitted to the experimental thin film transistors. [6)]

In order to simulate read SNM, WL and both BL and BLB are connected to Vcc and a variable voltage source is applied to "Out" node and then the voltage of "Outbar" node is monitored. Then "Outbar" is mirrored and is drawn in same curve to obtain butterfly curve. The maximum square that we can fit in this curve is read SNM.

To simulate the write SNM one of the BL or BLB should be connected to GND and the other one to Vcc. WL is enabled. Suppose BL is connected to Vcc and BLB is connected to GND. Then we put a sweep voltage source on the "Out" node and we measure the voltage of "Outbar" point. This write curve ("Outbar" versus "Out") is added to the read SNM butterfly curve.

The result of final curve is shown in Fig. 8. The butterfly curve is for read and the other one is for write operation mode. With fitting a minimum square between read and write curves, write SNM can be determined. For the SRAM cells with two different sizes the read and write margins are shown in this figure. For cell with Wa=2µm, Wd=2µm and Wp=2µm sizing, read SNM is 0.3V and write SNM is 1.4V and for SRAM cell with Wa=4µm, Wd=9µm and Wp=16µm sizing, read SNM is 0.6V and write SNM is 0.3V at Vcc=3.3V.

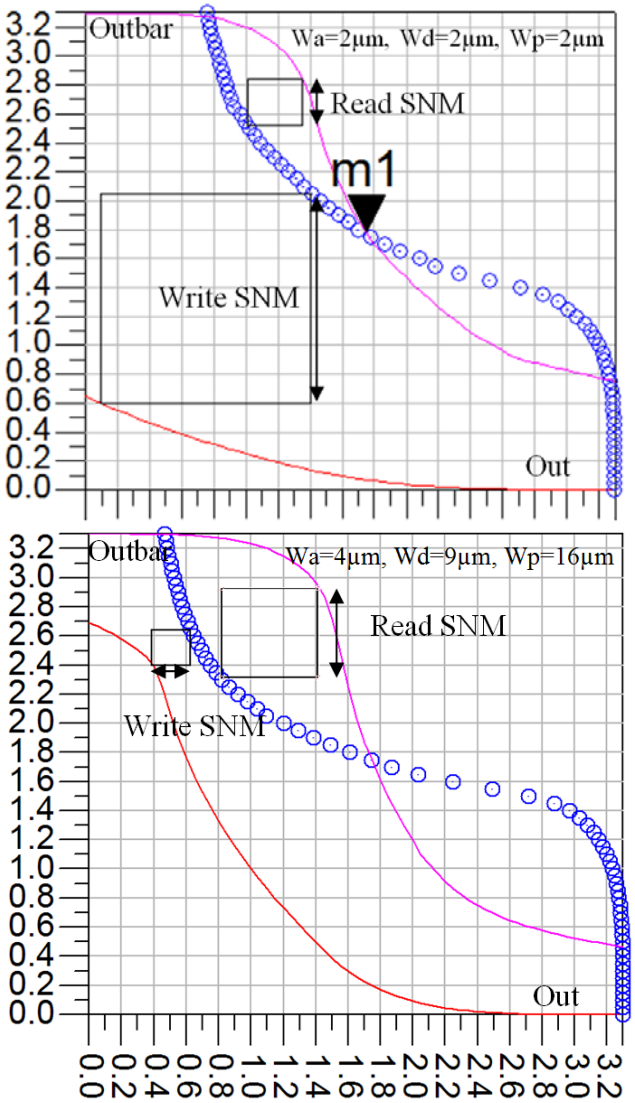


Fig.8: The simulated write SNM for Wa=2µm, Wd=2µm and Wp=2µm. Write SNM is 1.1V at Vcc=3V.

Designed layout in LEDIT software for one SRAM cell and sense amplifier and output buffers is shown in Fig. 9.

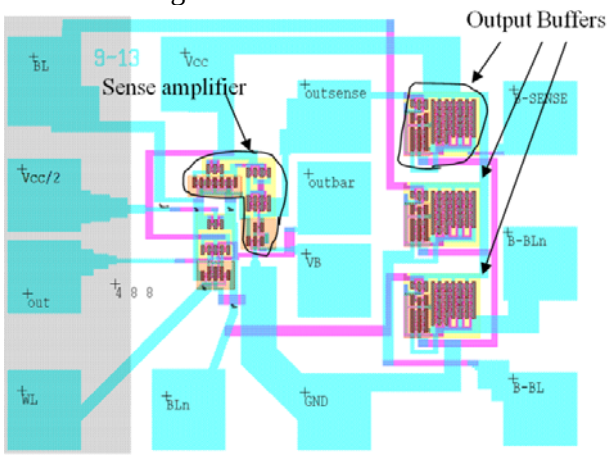


Fig.9. Designed layout of SRAM

4. Fabrication process

As we discussed before, µ-Czochralski process is used to fabricate SRAM cells. After crystallization of silicon and forming single grains, Island mask is applied to define the silicon islands using dry etching process. Cleaning before depositing gate oxide is very important. Maragoni cleaning is used and immediately after cleaning we use TEOS PECVD oxide to deposit 30nm gate oxide followed by 675nm Al gate. Using gate mask we etched the Al and oxide by Dry etching to reach to S/D. Next step is implantations for nMOS and pMOS transistors, respectively and then activation of dopants using excimer laser at room temperature and 300mJ/cm^2 laser energy is followed. Then 800nm TEOS oxide is deposited on wafers and contact hols are opened to have access to gate, source and drain. Finally 1.5µm Al is deposited and patterned using metal 1 mask.
Figure 10 shows SEM images of complete SRAM cell corresponding to layout shown in Fig.9.

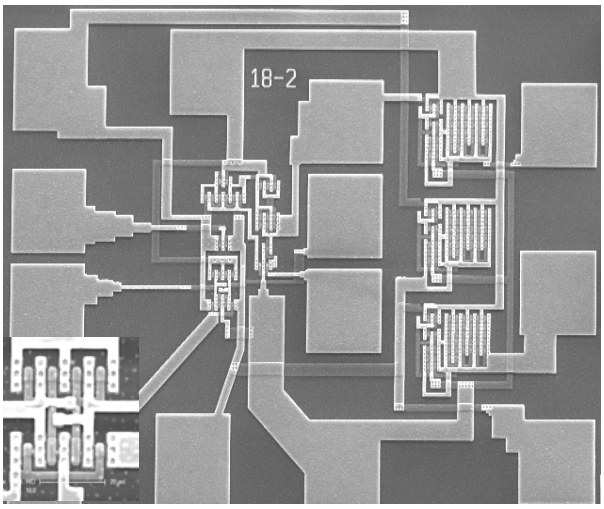


Fig.10. SEM image of complete SRAM cell

The gate length of all transistors is 2µm. Minimum size of contact openings are 2.5µm and minimum width of metal layer is 2µm.

5. Measurement results

We performed DC and transient measurements to characterize SRAM cells. In DC measurement we look to the inverter characteristics to find SNM. To find read SNM we connect WL, BL and BLB to Vcc and we apply DC voltage between 0 to Vcc to “out” node and we measure the “outbar” node voltage. We plot outbar versus out. Then again we connect WL, BL and BLB to Vcc and outbar to DC voltage between 0 to Vcc and we measure out voltage. (out and outbar are outputs of two inverters). This time again we plot outbar versus out in same curve. This will give us butterfly curves.
The SNM of a CMOS SRAM cell is defined as the minimum DC noise voltage necessary at both of the two cell storage nodes, during a read access, to flip the state of a cell. The smaller side of butterfly curve where two maximum squares nested between the static characteristics of the two cell inverters are equal to cell SNM.[9)]
To measure write SNM we connect BL to Vcc and BLB to the GND. Then we put a sweep voltage source on the “Out” node and we measure the voltage of “Outbar” point. This write curve (“Outbar” versus “Out”) is added to the read SNM butterfly curve.

Figure 11 shows measured DC curve of two SRAM cells with different sizes at 3.3V supply voltages.

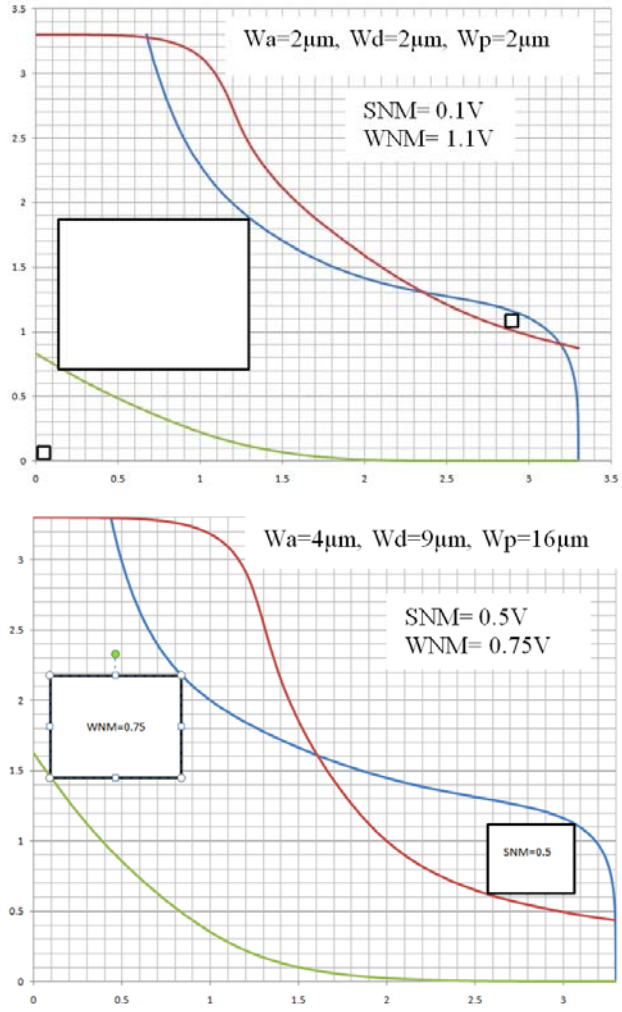


Fig.11: Measured read and write SNM for three SRAM cells

As it can be seen in small cell we have good write SNM but read SNM is low. The metastable point is not in middle and this shows process variation can affect on cell performance. With incresing the size of transistors to Wa=4µm, Wd=9µm and Wp=16µm we obtain good SNM for both read and write modes. In this cell the metastable point is in middle and shows using parallel transistors to make a big transistor in our process can cancell the process variations.

We measured the transient responce of our SRAM cells using pulse generators connected to BL, Wr and WL. Pulse generators can work till 100MHz. Fig. 12 shows the result of transient measurement. First waveform is given to WL. Second waveform is write signal (Wr) that we give it to the external three state buffer to enable the bit and bitbar bar generated inputs during writing and disable them during read operation. Therefore, when Wr is high we write Bit and Bitbar waveforms (that are inverterd etch other) to the SRAM and when Wr is low we read the stored data in SRAM. Third waveform is generated Bit signal from pulse generator that we want to write. Finally last signal is Bit signal that shows both write and read states. As it can be seen from this picture when WL is high and Wr is low, stored data could recover successfully for both 0 and 3V stored voltages at 1MHz WL frequency.

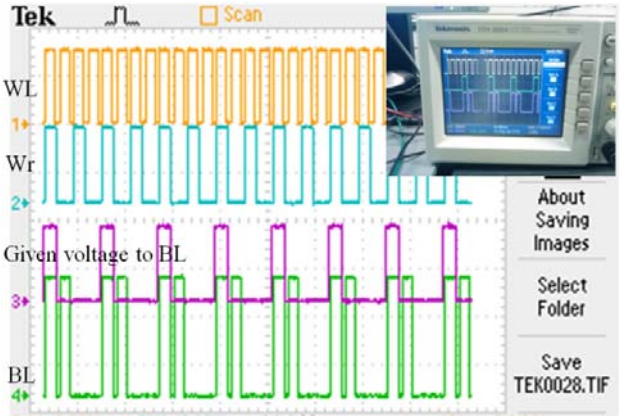


Fig.12. Output signals of SRAM cells

Furthermore, we increased the frequency till 100MHz to measure read and write access times. The SRAM cells can work till 87MHz WL frequency and after that read signal is noisy and we need sense amplifier to amplify the signals. Our measurements were based on definitions of Ref. 11 that shown in Fig.13. The time difference from 50% of Vcc in world line to BL is read access time. The write access time refers to the memory speed in write mode with respect to the activation of the word line as shown in Fig.13. The write delay is measured between the time when WL reaches to 50% of Vcc and the time when node “Out” reaches 50% of Vcc [11].

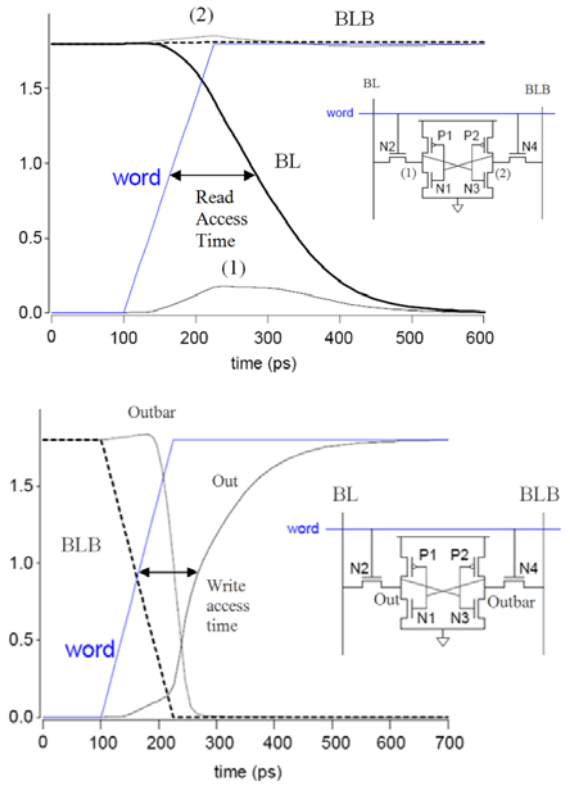


Fig.13: Definitions of read and write access times [11].

The measured read and write access times are shown in Fig.14.

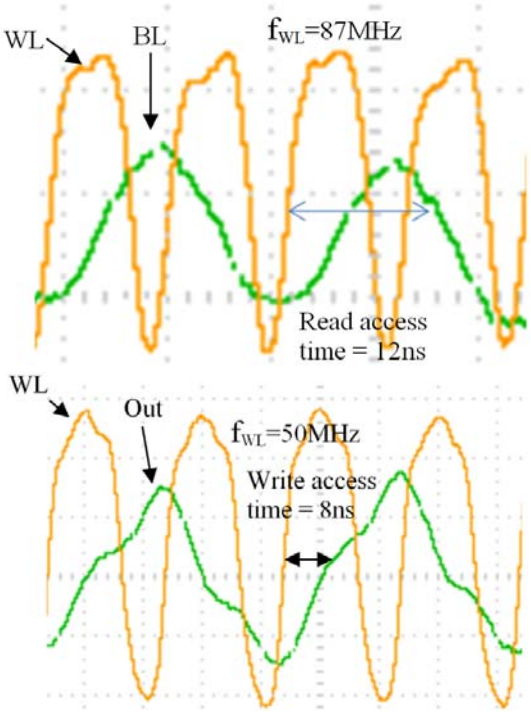


Fig.14: Measured read and write access times

As it can be seen at 87MHz worldline frequency we can measure read access time equal to 12ns. The write access time was measured at 50MHz and equal to 8ns. This shows SG-TFTs can operate in high frequencies successfully.

6. Conclusions

We reported design and fabrication of 6T SRAM cells with µ-Czochralski process. Working SRAMs show this technology is suitable for memory applications, particularly for low temperature applications. We obtained high read and write SNM equal to 0.5V and 0.75V at 3.3V supply voltage. The measured read and write access times are 12ns and 8ns, respectively.

Acknowledgments

We would like to express sincere thanks to all DIMES clean room staff and Prof. Nick van der Meijs from TU Delft and special thanks to Dr. Navid Azizi from University of Toronto for their helps during design, fabrication and test.

References

1) H.Ebihara, N.Karaki, et al. from Epson Company.

2) R.Ishihara, et al, Solid state electronics52 (2008) 353-358.

3)R.Ishihara, et al, IEEE Transaction on electron device, VOL. 51,NO. 3, March 2004

4) S. Park et al., IEEE J. Solid-State Circuits, vol. 34, no. 11, Nov. 1999, pp. 1436-1445.

5) J. B. Kuang et al., IEEE J. of Solid-State Circuits, vol.32, no. 6, Jun. 1997, pp. 837-844.

6) Nitz Saputra, et al, JSSC. VOL. 43,NO.7,July 2008.

7) Ming He, Ryoichi Ishihara et al, "Japanese Journal of Applied Physics" Vol. 45 No. 1A, Part 1, 2006 pp. 1-6

8) Kaushik Roy and Sharat Prasad, " Low-power CMOS VLSI circuit design ", New York : Wiley, 2000

9) Andrei Pavlov, Manoj Sachdev, "CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies: Process-Aware SRAM Design and Test ", 2008,chapter 2, p. 17

10) Jaber Derakhshandeh, M. R. Tajari Mofrad, R. Ishihara and C.I.M Beenakker, "Analog and digital output lateral photodiodes fabricated by µ-Czochralski process at low temperature", Proceeding of Device Research Conference 2009, State college, USA

[11] "Planar Double-Gate Transistor From technology to circuit" Amara, Amara; Rozeau, Olivier (Eds.) 2009**,** ISBN: 978-1-4020-9327-2.