International Journal on Emerging Technologies (Special Issue on ICRIET-2016) **7**(2): 347-353(2016)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Estimation of Power Dissipation of CMOS and finFET based 6T SRAM Memory

M. R. Govind * and Chetan Alatagi**

**Asst Prof, Department of Electronics and Communication Engineering, VSMIT, Nipani, Karanataka, India*

***Asst Prof, Department of Electronics and Communication Engineering, VSMIT, Nipani, Karanataka, India*

(Corresponding author: M. R. Govind)

(Received 28 September, 2016 Accepted 29 October, 2016)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: This paper provides the estimation of power dissipation of CMOS and finFET based 6T SRAM Memory. CMOS expertise feature size and threshold voltage have been scaling down for decades for achieving high density and high performance. The continuing reduce in the feature size and the corresponding increases in chip density and operating frequency have made power consumption a major concern in VLSI design. Extreme power dissipation in integrated circuits discourages their use in moveable systems. Low threshold voltage also results in enlarged sub-threshold leakage current because transistors cannot be turned off completely. For these reasons, leakage power dissipation, has become a major part of total power consumption for current and future silicon technologies. FinFET evolving to be a promising technology in this regard .In this the designing, modeling and optimizing the 6-T SRAM cell device is done.

Keywords: CMOS, FinFET, Static RAM, Read/Write, Sense Amplifier

I. INTRODUCTION

It is found that FinFET-based 6T SRAM cells designed with built in feedback realize significant improvements in the cell static noise margin without area penalty, read/write in time analysis. Improvement in SNM (signal to noise margin) can be achieved in 6-T FinFET-based SRAM cells. Improvements in SNM as the 6T cell, making them attractive for low-power, low-voltage applications. The long-channel-device-based SRAM cell is slightly robust than optimized SRAM; however, increased gate-edge tunneling and leakage parasitic capacitances degrade the power consumption and access time.

A. Overview of FinFET

A multigate device refers to a MOSFET which incorporates more than one gate into a single device. FinFETs are better substitutes for bulk CMOS at the nanoscale. Mostly FinFETs having double-gate control devices are chosen. In some experiment the two gates of a FinFET can either be shorted for higher performance, higher gain or independently controlled for lower leakage or reduced transistor count. These modifications give rise to a rich design space. Since nanometer process technologies have advanced, Chip density and operating frequency have increased, that makes power burning up in battery operated portable devices a major concern. FinFETs have been adopted for the high-volume production of CMOS integrated circuits beginning at the 22-nm technology generation [1], due to the superior electrostatic integrity of these multigate transistor structures [4].

Fig.1.1 shows the top view of double gate FinFET devices. Double gate FinFET consists of two SOI gates connected together. The thickness of a single fin equals to silicon channel thickness. The current flows from the source to drain along the wafer plan. For the FinFET devices, widths are quantized into units of the fins. Large width of device is obtained by using multiple fins.

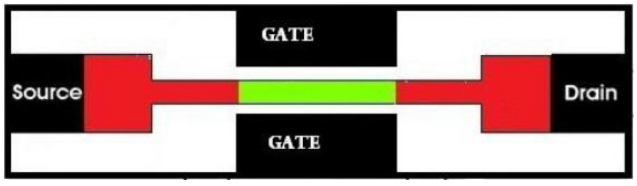

Fig. 1. Top View of FinFET.

B. Problem over CMOS based SRAM

1) While it is possible to scale the classical bulk-Si MOSFET structure down into the sub-20nm rule, short channel effect control requires heavy channel doping (>1018 cm^3) and heavy super-halo implants to control sub-surface leakage currents. As a result, carrier mobility's are solemnly degraded due to impurity dispersion and a high transverse electric field in the ON state.

2) Also, the increased depletion charge density results in a larger depletion capacitance hence a larger sub threshold slope. And thus, for a given off-state leakage current specification, on-state drive current is degraded. Off-state leakage current is boosted due to band-to-band tunneling between the body and drain.

3) Vt variability caused by random dopant fluctuations is another concern for nanoscale bulk Si MOSFETs. Designing large arrays must have design for 5 or more standard deviations. With increasing variations, it becomes difficult to guarantee near minimum sized cell stability for large arrays in embedded, low-power applications. Increasing transistor sizes, on the other hand, is counter to the fundamental reason for scaling in the first place – to increase density.

4) Access time is reliant on wire delays and column height. So to speed up arrays, segmentation method is commonly employed. With additional reductions in bit line height, the overhead area of sense amplifiers becomes substantial.

C. Challenges Overcome by FinFET

The FinFET has initially developed to manufacture of self-aligned double-gate MOSFETs. Also to address the need for improved gate control to reduce Drain Induced Barrier Lowering, Sub threshold current and process-induced variability for Lg < 25nm.Tri-Gate and Bulk variations of the FinFET have been developed to improve manufacturability and cost. Multi-gate MOSFETs provide a pathway to achieving lower power and/or improved performance[1].The double-gate FinFET offers distinct advantages for simultaneously suppressing the sub-threshold current and gate dielectric leakage current as compared to the traditional single-gate MOSFETs. The haphazard dopant fluctuation is a form of process variation resulting from variation in the implanted impurity concentration. In MOSFET transistors, RDF in the channel region can alter the transistor's properties, especially threshold voltage. Thus Finfet technology is being carried out in order to reduce the dopant fluctuation which leads to the variation of threshold voltage. [7].

II. SRAM CELL

A high level interface to the SRAM is very similar to that for the Read Only Memory. Read-write random-access memories (RAM) may store information in flip flop like circuits or simply as charge on capacitors. Because read-write memories store data in active circuits, they are volatile; that is, stored information is lost if the power supply is sporadic. If the terms were consistent, both read-only and read-write memories would be called RAMs. The widely used common types of RAMs are the static RAM (SRAM) and the dynamic RAM (DRAM). Static RAMs hold the stored value in flip-flop circuits as long as the power is on. SRAMs tend to be high-speed memories with clock cycles in the range of 5 to 50 ns. Dynamic RAMs store values on capacitors. They are inclined to noise and leakage problems, and are slower than SRAMs, clocking at 50 ns to 200 ns. However, DRAMs are much denser than SRAMs up to four times denser in a given generation of technology.

Read-only memories (ROMs) store information according to the presence or absence of transistors joining rows to columns. All ROMs are nonvolatile, but they vary in the method used to enter (write) stored data. The simplest form of ROM is programmed when it is manufactured by formation of physical patterns on the chip; subsequent changes of stored data are impossible. These are termed mask-programmed ROMs (PROM).

A. Features of SRAM

1. Data is stored as long as supply is applied
2. Fast - so used where speed is important (e.g.,caches)
3. Differential outputs
4. Low power consumption
5. Compatible with CMOS technology

B. Why SRAM Cell

SRAM cells are typically used to implement memories that involve short access times, low power dissipation, and easiness to environmental circumstances. There are many reasons to use an SRAM in a system design. Design tradeoffs include density, speed, volatility, cost, and features. All of these factors should be considered before you select a RAM for your system design.

1. Speed -The key advantage of an SRAM over a DRAM is its speed. The fastest DRAMs on the market still require five to ten processor clock cycles to access the first bit of data. Fast, synchronous SRAMs can operate at processor speeds of 250MHz and beyond, with access and cycle times equal to the clock cycle used by the microprocessor. With a well-designed cache using ultra-fast SRAMs, conditions in which the processor has to wait for a DRAM access become rare.

2. Density – The DRAM and SRAM memory cells are designed, without much difficulty available DRAMs have significantly more densities than the largest SRAMs. Thus, when 64 Mb DRAMs are rolling off the production lines, the largest SRAMs are expected to be only 16 Mb.

3. Volatility - Whereas SRAM memory cells need more space on the silicon chip, they have other advantages that translate directly into improved performance. Unlike DRAMs, SRAM cells do not need to be refreshed operaton. This means they are available for reading and writing data 100 percent of the time.

4. Cost - If cost is the primary factor in a memory design, then DRAMs win hands down and on the other hand, performance is a critical factor, then a well-designed SRAM is an effective cost performance solution.

5. Custom features - Most DRAMs come in only one or two flavors. This keeps the cost down, but doesn't help when you need a particular kind of addressing sequence, or some other custom feature. SRAMs are customized, via metal and substrate, for the processor or application that will be using them. Features are connected or disconnected according to the requirements of the user. Likewise, interface levels are selected to match the processor levels.

III. POWER CONSUMPTION

The power dissipation in CMOS digital circuits is classified into two types:

-Peak power and

-Time-averaged power consumption

1. Peak power is a dependability issue that determines both the chip lifetime and performance. The voltage drop effects, caused by the excessive instantaneous current flowing through the resistive power network, affect the performance of a design due to the increased gate and interconnect delay. This large power consumption causes the device to overheat which reduces the reliability and lifetime of the circuit. Also noise margins are declined, increasing the chance of chip failure due to crosstalk.

2. The time-averaged power consumption in conventional CMOS digital circuits occurs in two forms: *dynamic* and *static*. Dynamic power dissipation occurs in the logic gates that are in the process of switching from one state to another.

During this process, any internal and external capacitance associated with the gate's transistors has to be charged, thereby overwhelming power. Static power dissipation is associated with inactive logic gates (i.e., not currently switching from one state to another). Dynamic power is important during normal operation, especially at high operating frequencies, whereas static power is more important during standby, especially for battery-powered devices.

The power consumption in digital CMOS circuits can be deliberated by:

$Pt=P_{dyn} + P_{sh} + P_{st} + P_{lea}$ ---------(3.1)

Where P_t is the total power consumption, P_{dyn} is the dynamic power consumption due to switching of transistors, P_{Sh} is the short-circuit current dissipation when there is direct path from the power source down to ground. P_{dyn} and P_{sh} together can be called as P_{dyn}. P_{sh} is the static power consumption, and P_{lea} is the power consumption due to leakage currents.

A. Dynamic Power Dissipation

Dynamic power is primarily caused by the current flow from the charging and discharging of parasitic capacitances. Whenever the logic level changes at different points in the circuit because of the change in the input signals the dynamic power dissipation occurs. It consists of three components: switching power, short-circuit power, and glitching power. Switching power dissipation is caused by the charging and discharging of the node capacitance.

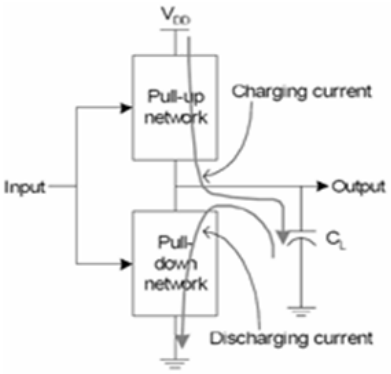

Fig. 2. Switching power dissipation.

Fig. 2 shows the Short circuit power dissipation caused by concurrent conduction of n and p blocks. Switching power dissipation mostly depends on supply voltage, physical capacitance and switching activity. Short circuit power dissipation mostly depends on the input, load capacitance, and the transistor size of the gate, supply voltage, frequency and threshold voltage. Dynamic power is calculated by the equation 3.2.

$Dyanmic = C_{load} \times V_{dd} \times N \times f$ ----------(3.2)

Where, C_{load} is parasitic and interconnect capacitance, is V_{dd} supply voltage, N is switching activity factor and f is frequency of signal. The figure 3. shows the short circuit current diagram

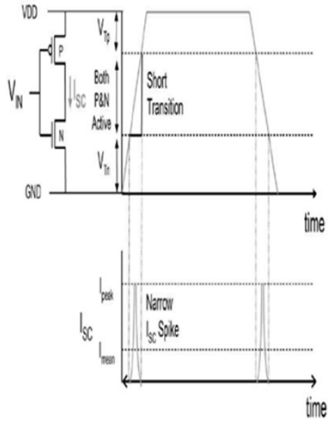

Fig. 3. Short circuit current.

B.Static Power Dissipation

Static power is caused by leakage currents while the gates are switch off condition; that is, no output transitions. Tentatively, CMOS gates should not be consuming any power in this mode. This is due to the fact that either pull-down or pull-up networks are turned off, thus preventing static power dissipation. In reality, however, there is always some leakage current passing through the transistors, indicating that the CMOS gates does consume a certain amount of power. Even though the static power consumption, associated with an individual logic gate is extremely small, the total effect becomes significant when tens of millions of gates are utilized in today's integrated circuits (ICs). Additionally, as transistors shrinks in size, the level of doping has to be increased, thereby causing leakage currents to become larger.

C. Leakage Currents

Leakage currents arise from a variety of sources within the transistor devices. For long-channel transistors, the leakage current is conquered by the reverse diode leakage and the sub-threshold leakage. Other leakage mechanisms are abnormal to the small-device geometries. There are six different leakage currents in short channel transistor illustrated in Fig. 4.

Where, I_1 is the reverse-bias pn junction leakage. A reverse bias pn junction leakage I_1 has two main components: one is minority carrier diffusion/drift near the edge of the depletion region; the other is due to electron-hole pair production in the depletion region of the reverse-biased junction. I_2 is the sub-threshold leakage; which occurs due to carrier diffusion between the source and drain when the gate-source voltage, V_{GS}, has increased the weak inversion point, but is still below the threshold V_T, where carrier drifts is dominant. In this rule, the MOSFET behaves similarly to a bipolar transistor, and the sub-threshold current is exponentially dependent on the gate-source voltage.

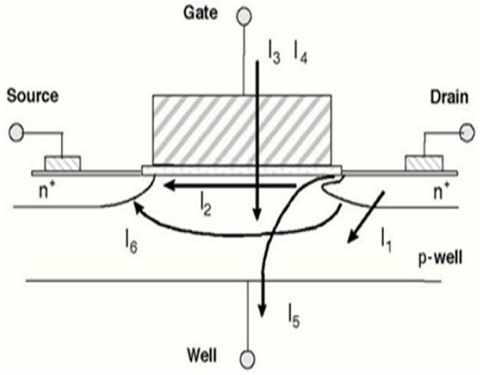

Fig. 4. Summary of leakage current mechanism.

I_3 is the oxide tunneling current; decrease of gate oxide thickness arises in an increase in the field across the oxide. The high electric field coupled with low oxide thickness results in tunneling of electrons from substrate to gate and also from gate to substrate through the gate oxide, resulting in the gate oxide tunneling current.

I4 is the gate current due to hot-carrier injection; in a short-channel transistor, due to high electric field near the Si–SiO_2 crossing point, electrons or holes can gain enough energy from the electric field to cross the interface potential barrier and enter into the oxide layer. This effect is known as hot-carrier injection. The insertion from Si to SiO_2 is more likely for electrons than holes, as electrons have a lower effective mass than that of holes, and the barrier height for holes (4.5 eV) is more than that for electrons (3.1 eV).

I_5 is the GIDL (Gate Induce Drain Leakage) where GIDL is due to high field effect in the drain junction of an MOS transistor. When the gate is biased to form an accumulation layer at the silicon surface, the silicon surface under the gate has almost same potential as the p-type substrate. Owing to presence of accumulated holes at the surface, the surface behaves like a p area more heavily doped than the substrate. This causes the depletion layer at the surface to be much narrower than elsewhere. The narrowing of the depletion layer at or near the surface causes field crowding or an increase in the local electric field, thereby enhancing the high field effects near that region. When the negative gate bias is large (i.e., gate at zero or negative and drain at V_{DD}), the n+ drain area under the gate can be depleted and even inverted.

This causes more fields crowding and peak field increase, resulting in a dramatic increase of high field effects such as avalanche multiplication and BTBT. The possibility of tunneling via near-surface traps also increases. As a result of all these effects, minority carriers are emitted in the drain region underneath the gate. Since the substrate is at a lower potential for minority carriers, the minority carriers that have been accumulated or formed at the drain depletion region underneath the gate are swept laterally to the substrate, completing a path for the GIDL. Thinner oxide thickness and higher (higher potential between gate and drain) enhance the electric field and therefore increase GIDL.

I_6 is the channel punch through current. In short-channel devices, due to the closeness of the drain and the source, the depletion regions at the drain-substrate and source-substrate junctions widen into the channel. As the channel length is reduced, if the doping is kept constant, the separation between the depletion region boundaries decreases. An increase in the reverse bias across the junctions (with increase in V_{DS}) also pushes the junctions nearer to each other.

D. SRAM cell leakage paths

In this section, we explain the major sub threshold leakage components in a 6T Static RAM cell. The sub threshold leakage current in an SRAM cell is typically divided into two kinds as shown in Fig. 5:

(i) cell leakage current that flows from *Vdd* to *Gnd* internal to the cell.

(ii) *bitline* leakage current that flows from *bitline* (or *bitline'*) to *Gnd*. Although an SRAM cell has two *bitline* leakage paths, the *bitline* leakage current and *bitline'* leakage current differs according to the value stored in the SRAM bit. If an SRAM cell holds „1 as shown in Fig. 4, the *bitline* leakage current passing through *N3* and *N2* is effectively suppressed due to two reasons. First, after precharging *bitline* and *bitline'* both to „1, the source voltage and the drain voltage of *N3* are the same and thus potentially no current flows through *N3*.

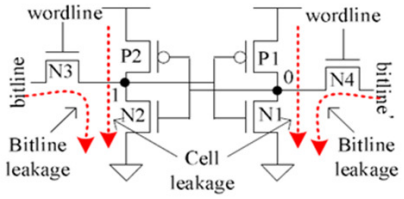

Fig. 5. SRAM cell leakage paths.

Second, two stacked and turned off transistors (*N2* and *N3*) produce the stack effect. Meanwhile, for this case where the SRAM bit holds value „1, a large *bitline'* leakage current flows passing through *N4* and *N1*. If, on the other hand, the SRAM cell holds „0, a large *bitline* leakage current flows while *bitline'* leakage current is suppressed.

IV. CMOS SRAM CIRCUIT

A.Schematic

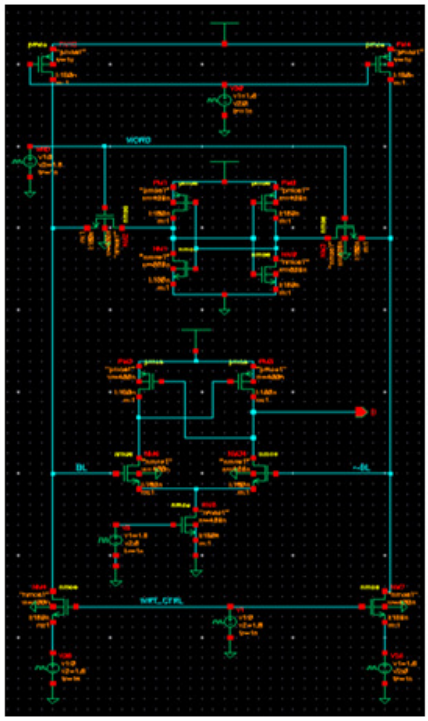

Fig. 6. 6T SRAM schematic.

Single bit SRAM memory cell is shown in Fig. 6 using CMOS. Static latches are used in the SRAM cell. SRAM cell is made up of flip flop comprising of two cross coupled inverters. Two access transistors are used to access the stored data in the cell. These transistors are turned ON/OFF by the control line called word line (WL). Generally this word line is connected to the output of row decoder circuits. When WL=V_{dd} the SRAM cell is connected to bit line (BL) and complement of bit line (~BL) allowing both read and write operations. Read-write operation is carried out by the help of access transistors.

B. Read Operation

Consider node Y as reference node of the SRAM cell. Cell is said to be storing 1 if node Y is high at V_{dd} and node Y bar is at 0V. For the reverse voltage conditions cell is said to be storing zero.

Let us assume that cell is storing 1.Before the read operation starts BL and ~BL lines are precharged to $V_{dd}/2$. When write line WL is activated the current enters through M5 and M6. Now current from V_{dd} will flow through M1 and M5 charging the bit line capacitance, say CBL. The existing capacitance on the line ~BL, say CBLbar discharges through the transistors M6 and M4.This process develops a voltage difference between node Y and node Y bar which is sensed by the sense amplifier to detect it as 1. Like that a 0 in the cell is also detected by the sense amplifier.

C. Write Operation

Let us consider the write operation of zero to the cell which is storing a value of 1. For this, sense amplifiers and precharge circuits are disabled. The cell is selected by activating the equivalent WL signal. To write zero to the cell, BL line held low and ~BL line is raised to V_{dd} by the write circuit. Thus the node Ybar is pulled up towards the $V_{dd}/2$ while node Y is pulled down to $V_{dd}/2$. When the voltage crosses this level on two nodes feedback action starts parasitic capacitances developed by M3, M5 and M4, M6 are charged and discharged respectively. Ultimately node Y stabilizes at the value 1. Since these parasitic capacitances offered by transistors are relatively much lesser than the bit line capacitances, write operation is faster than read operation.

D. Sense Amplifier

The figure 6 shows the Sense amplifier is an essential circuit in memory chips to speed up the *Read* function. Due to large arrays of SRAM cells, the resultant signal in the event of *Read* function has a much lower voltage swing. To recompense for that swing, a sense amplifier is used to amplify voltage coming off BL and ~BL. The voltage coming out of sense amplifier has a full swing voltage of (0 to 1.8 V). Sense Amplifier also helps reduce the delay times and power dissipation in the overall SRAM array.

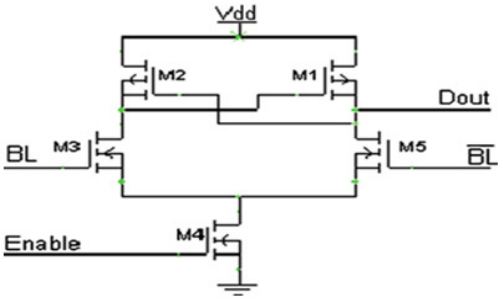

Fig. 6. Sense Amplifier.

To reimburse for that swing, a sense amplifier is used to amplify voltage coming off Bit Line (BL) and ~Bit Line(~BL). The voltage coming out of sense amplifier has a full swing voltage of (0–1.8 V). Sense Amplifier also helps reduce the delay times and power dissipation in the overall SRAM array.

V. FinFET SRAM CIRCUIT

A. Schematic Circuit Using FinFET

The design considerations for a robust 6T FinFET SRAM circuit are given in this section. [11]

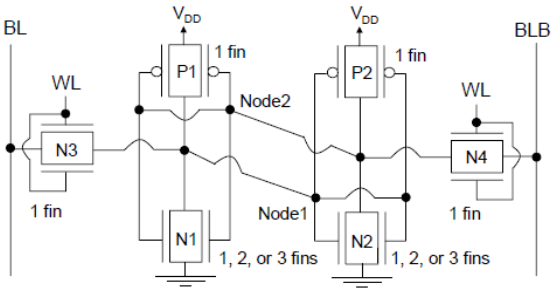

Fig. 7. A FinFET SRAM circuit.

VI. RESULTS

-SRAM= 1.0987E-4 W. (Using CMOS)
-SRAM= 3.06E-5 W. (Using FinFET) [3]

VII. CONCLUSION

From the above observation it is clear that the power dissipation is low for SRAM using FinFET compared to the SRAM using CMOS. With the continuous technology scaling devices, the outflow power is of great concern for designs in nanometer technologies and is becoming a major provider to the total power consumption; leakage power has become more governing as compared to Dynamic power. The gate leakage has become dominant sources of leakage and is predictable to increase with the technology scaling.

REFERENCES

[1]. Rashmi Verma, Prof. Rakesh Gajre, Prof. Ankit Adesara "Design SRAM Using FinFET -Review," International Journal of Innovative and merging Research in Engineering Volume 3, Issue 2, 2016,pp. 73-77.

[2]. H. Bu, "FinFET technology a substrate perspective," in Proc. IEEE Int. SOI Conf. (SOI), Oct. 2011, pp. 1–27.

[3]. Abhijith A Bharadwaj, H V Ravish Aradhya "Design and Performance Comparison of finFET, CNFET and GNRFET based 6T SRAM,", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438, pp. 399-403.

[4]. Xi Zhang, et al, " Analysis of 7/8-nm Bulk-Si FinFET Technologies for 6T-SRAM Scaling " , IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. **63**, NO. 4, APRIL 2016, pp. 1502 -1507.

[5]. C. Auth *et al.*, "A 22 nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *Proc. Symp. VLSI Technol. (VLSIT)*, Jun. 2012, pp. 131–132.

[6]. Farhana Afrin et al., "Statistical Analysis of Leakage Current of Trapezoidal FinFETs", IEEE International WIE Conference on Electrical and computer, 2015, pp. 407 to 410.

[7]. Lourts Deepak A, Likhitha Dhulipalla, Dr. Cyril Prasanna Ra P,"Performance comparison of CMOS and FINFET based SRAM for 22nm Technology", International Journal of Conceptions on Electronics and Communication Engineering Vol. **1**, Issue. 1, Dec' 2013; ISSN: 2357 – 2809.

[8]. Mugdha Sathe, Dr. Nisha Sarwade," Performance Comparison of CMOS and Finfet Based Circuits At 45nm Technology Using SPICE", Mugdha Sathe Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. **4**, Issue 7(Version 2), July 2014, pp.39-43.

[9]. Brian Swahn and Soha Hassoun," Gate Sizing: FinFET vs 32nm Bulk MOSFETs", Tufts University Medford, MA 0215.

[10]. Prateek Mishra, Anish Muttreja, and Niraj K. Jha." FinFET Circuit Design"

[11]. Sherif A. Tawfik et al., "Low-Power and Robust Six-FinFET Memory Cell Using Selective Gate-Drain/Source Overlap Engineering", ISIC 2009, pp.244-247.