Analysis of FinFET based Low Power SRAM Cell

Sudeep Bhattacharya[1] Devesh Tiwari[2] Mr. A.G Rao[3]
[1,2]M.Tech [3]Scientist B
[1,2,3]Department of Electronic Design and Technology
[1,2,3]NIELIT, Gorakhpur

Abstract— As CMOS electronic devices are continuously shrinking to nanometer regime, leads to increasing the consequences of short channel effects and variability due to the process parameters which lead to cause the reliability of the circuit as well as performance. To solve these issues of CMOS, FINFET is one of the promising and better technologies without sacrificing reliability and performance for its applications and the circuit design. Among the various embedded memory technologies, SRAM provides the highest performance along with low standby power consumption. In CMOS circuits, high leakage current in deep-submicron regimes is becoming a significant contributor to power dissipation due to reduction in threshold voltage, channel length, and gate oxide thickness. FinFET based SRAM design can be used as an alternative solution to the bulk devices. FinFET is suitable for Nano scale memory circuits design due to its reduced Short Channel Effects (SCE) and leakage current. As the impact of process variations become increasingly significant in ultra deep submicron technologies, FinFETs are becoming increasingly popular a contender for replacement of bulk FETs due to favorable device characteristics. The paper focuses on study of various design aspects of FinFET based SRAM.

Keywords: SRAM, FinFET, SCE, CMOS

I. INTRODUCTION

The past 3 decades CMOS IC technologies have been Scaled down continuously and entered into the nanometer region. In many designs the need of memory has increased vastly from consumer goods to industrial applications. It increases the necessity of improving memories in a single chip with the help of nanometer technologies. There are lots of applications and integrated memories are improved using nanotechnology especially SRAM cell.

The shrinking of the CMOS technology has been increased very aggressively with ultra-thin sizes. This shrinking of the design creates many significant challenges and reliability issues in design which causes augmented process variations, short channel effects, power densities and leakage currents etc. Ultra-thin sized CMOS technologies have been designed to use in many applications. Continuous shrinking of channel length is increases the high speed devices in very large scale circuits. This steady miniaturization of transistor with each new generation of bulk CMOS technology has yielded continual improvement in the performance of digital circuits. The scaling of bulk CMOS, however, faces significant challenges in the future due to the fundamental material and the process technology limits. The 22 nm FINFET based transistors are used as choice and solution for CMOS based technology with scaled device geometry. In these device structures, the effect of short-channel length can be controlled by limiting the off-state leakage. Moreover, FINFETs has advantages of suppressing short channel effects, gate-dielectric leakage currents etc.

The additional back gate of a FinFET gives circuit designers many options. The back gate can serve as a secondary gate that enhances the performances of the front (primary) gate. For example, if the front gate voltage is V_{DD} (transistor is ON) the back gate can be biased to V_{DD} to provide bigger current drive, which reduces transistor delay. If the front gate voltage is 0 (transistor is OFF), the back gate can be biased to 0, which raises the threshold voltage of the front gate and reduces the leakage current. The FinFET transistor structure has been introduced as an alternative to the bulk-Si MOSFET structure for improved scalability. The structure has two gates which can be electrically isolated and have two different voltages (back gate) for an improved operation. In the double-gate (DG) operating mode, the two gates have connected together to switch the FinFET on/off, whereas in the back-gate (BG) operating mode, they are biased independently – with one gate used to switch the FinFET on/off and the other gate used to determine the threshold voltage. The BG operation mode provides us with the ability to tune the dynamic and/or static performance characteristics.

II. DG FINFET DEVICE

The continuous down in scaling of bulk CMOS creates major issues due to its base material and process technology limitations. The main drawback of CMOS based design is the leakage in small channel size; due to this the leakage stems increased from the lower oxide thickness, more substrate doping's. The optimal performance of the device can be achieved by lowering the threshold voltage with low supply voltage worsen the leakage. The primary obstacles to the scaling of CMOS gate lengths to 22nm and beyond includes short channel effects, sub-threshold leakage, gate-dielectric leakage and device to device variation reduction which leads low yield. The International Technology Roadmap for Semiconductors (ITRS) predicts that double gate or multi-gate devices will be the perfect solution to obtain the device with reduced leakage problems and less channel length of the transistor. The FINFET based designs are known as double gate device which offers the better control over short channel effects, low leakage current and better yield in 22nm and beyond which helps to overcome the obstacles in scaling [1][2].

When threshold voltage V_t is less than a potential voltage, gates of the double gate or FINFET device activates the currently flow between drain to source with modulating the channel from both the sides instead of one side. The potential which is applied to two gates together influence potential of the channel which fighting against the drain impact and leads to solve and give the better shut off to the channel current and reduces Drain Induced Barrier Lowering (DIBL) with improved swing of the design.

This FINFET based transistors offers good tradeoff for Power as well offering interesting delay. The Figure 1 shows the structure of multi-FIN based field effect transistors.

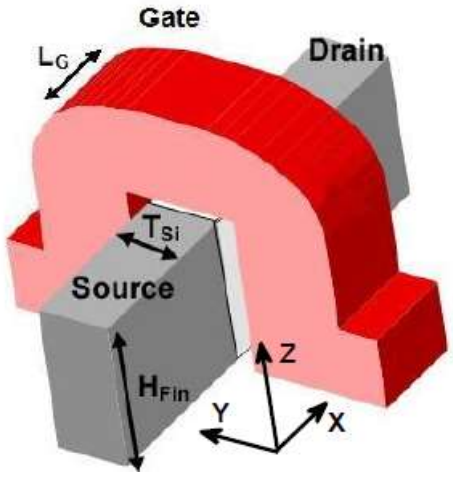

Fig. 1: FinFET Structure

A. Modes of FINFET

Double Gate (DG) devices have been used in a variety of Innovative ways in digital and analog circuit designs. A parallel transistor pair consists of two transistors with their source and drain terminals tied together. In Double-Gate (DG) FINFETS, the second gate is added opposite to the traditional gate, which has been recognized for their potential to better control short channel effects, as well as to control leakage current.

The modes of FINFETs are identified as short gate (SG)
mode with transistor gates tied together, the independent gate (IG) mode where independent digital signals are used to drive the two device gates, the low-power (LP) mode where the back-gate is tied to a reverse-bias voltage to reduce leakage power and the hybrid (IG/LP) mode, which employs a combination of low power and independent gate modes. Here independent control of front and back gate in DG FINFET can be effectively used to improve performance and reduce power consumption. Independent gate control is used to combine parallel transistors in non-critical paths.

B. FINFET Device Modelling Parameters

The FINFET is originally known as the folded channel MOSFET [3] which has narrow vertical fins from wafer. Usually width of the gate will be double of the fin height in FINFETs. FINFETs are nominated instead of CMOS in less than 22nm technology due to its cost effective manufacturing.

The geometric key parameters of FINFETs are

(1) L_g – Length of the gate,
(2) h – Height of the FIN,
(3) t_{ox} – Thickness of gate oxide,
(4) $t_{ox\text{-}top}$ - Oxide thickness of top gate and fin,
(5) T_{si}- Thickness of the fin
(6) Channel Doping.

III. SRAM DESIGN

SRAM is a volatile memory that retains data bits as long as power is being supplied. It provides fast access to data and is very reliable. SRAM arrays are widely used as cache memory in microprocessors and Application-Specific Integrated Circuits (ASICs) and occupy a large portion of the die area. Large arrays of fast SRAM help improve the performance of the system. Following are the requirements of SRAM cells for various applications:

Power dissipation: Embedded systems, particularly those targeted toward low duty cycles and portable applications (e.g. mobile phones), require extremely low energy dissipation as they are typically battery powered.

Performance: SNM can serve as a figure of merit instability evaluation of SRAM cells. The read SNM is defined as the minimum DC noise voltage which is required to flip the state of the SRAM cell during the read operation. It is measured as the length of the side of the largest square that fits inside the lobes of the butterfly curve of the SRAM.

Process variation: Millions of minimum-size SRAM cells are tightly packed making SRAM arrays the densest circuitry on a chip. Such areas on the chip can be especially susceptible and sensitive to manufacturing defects and process variations.

IV. SRAM BENEFITS

There are many reasons to use an SRAM as an embedded memory in a system design in place of the use of a DRAM. A couple of design tradeoffs include speed, density, volatility, cost, reliability, and features.

(1) Speed - The primary advantage of an SRAM over a DRAM is its speed. The fast, synchronous SRAMs can operate at processor speeds of 250 MHz 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 way DRAM and SRAM memory cells are designed, readily available DRAMs have significantly higher 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 - While SRAM memory cells require 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. This means they are available for reading and writing data.
(4) Cost - If cost is the primary factor in a memory design, then DRAMs win hands down. If, on the other hand, performance is a critical factor, then a well-designed SRAM is an effective cost performance solution.
(5) Features - Most DRAMs come in only one or two flavors. This keeps the cost down, but doesn't help when there is a need for a particular kind of addressing sequence, or some other custom feature. Features are connected or disconnected according to the requirements of the user. Likewise, interface levels are selected to match the processor levels [5].

V. FINFET BASED SRAM CELL

The six-transistor (6T) static memory cell, shown in Figure 3.2, is widely accepted as the standard memory cell. It is designed to achieve fast read times with the inclusion of sense amplifiers. The standard 6T cell requires that a logic value and its inverse be placed on the bit lines during a write operation. The word line (WL) is raised to logic 1 and the logic levels on the bit lines are passed into the cross-coupled inverter pair. Reading from the memory cell entails pre-charging the bit lines and then asserting logic 1 on the word line. The complexity of this cell is in arriving at the appropriate device sizes for proper functionality. Device sizing for a CMOS-based cell is driven primarily by area and functional operation constraints; sizing must be carefully performed to maintain a stored value, enable the cell to push a stored value to the bit lines with reasonable speed for a read operation, and to correctly transfer and overwrite new logic values into the cell for a write operation.

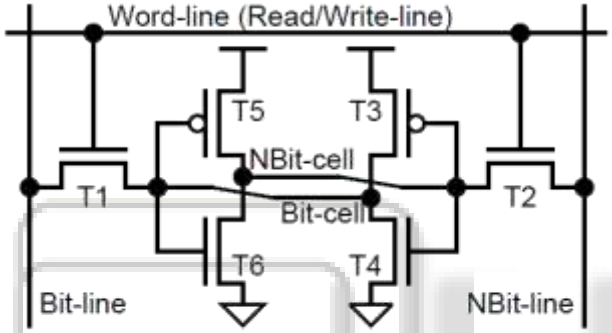

Fig. 2: 6T SRAM Cell

A. Read Operation

The row decoder and column decoder decode the address of the SRAM cell from where data has to read. Before the read operation the bit lines are pre charged to V . When the word line is enabled, one of the bit lines is discharged through an NMOS transistor connected to a"O" node of the cell. Because of very small current driving capability of the cell, voltage change on the bit line is very small [26] . The sense amplifier is used to sense this small voltage difference developed between the bit lines and finally we came to know whether " 1 " was stored or "0" was stored

B. Write Operation

The data to be written into a cell is provided by the external devices and row decoder and column decoder provide the address of a cell where data has to be written. When the word line is enabled the write circuitry will write the data into the cell depending upon the condition of the column gate [27] . Write operation will takes less time than the read operation because of the large current driving capability of the write buffer.

VI. FINFET SRAM PERFORMANCE MATRICES

(1) Static Noise Margin (SNM): Stability, the immunity of the cell to flipping during a read operation, is characterized by Static Noise Margin (SNM). SNM is calculated by the side of the largest square inside the SRAM cross-coupled inverter characteristic measured during the read condition (BL = BL' = V_{dd}, and WL = V_{dd}). Static Noise Margin is the standard metric to measure the stability in SRAM bit-cells. The SNM depends on the choice of the Vth for the transistors used in the SRAM cells. A high Vth means that drive current of these devices is small making the write operation more difficult, thus increasing the SNM. Thus, one approach to achieve a low power cell with high stability is to use high Vth devices at the cost of performance. FinFETs provide with a high drive current even with larger Vth thereby achieving high noise margins along with good write stability. The SNM is seen to be most sensitive to threshold voltage fluctuations in the access and pull down NMOSs and least sensitive to the fluctuations in the pull-up P-FinFET device. For FinFETs the effect of Lg variation on Vth is small, so the effect on the SNM is also small.

(2) Read Noise Margin (RNM): RNM is often used as the measure of the robustness of an SRAM cell against flipping during read operation. For read stability (High RNM) of FinFET based SRAM cell, pull down transistor is typically stronger than access transistor. The read margin can be increased by upsizing the pull-down transistor i.e nFinFET, which results in an area penalty and/or increasing the gate length of the access transistor, which increases the WL delay and hurts the write margin. A careful sizing of the FinFET device is required to avoid accidentally writing a 1 into the cell while trying to read a stored "0" thus resulting in a read upset.

(3) Write Noise Margin (WNM): Write Noise Margin (WNM) is the maximum bit line (BL) voltage that is able to flip the state of the FinFET basedSRAM cell while bit line bar (BL") voltage is kept high. Higher the WNM, greater is the stability. Use of a weaker pull up (pFinFET) and a stronger access transistor helps the node storing "1" to discharge faster, thus facilitating a quicker write of "0". The write margin can be measured as the maximum BL" voltage that is able to flip the cell state while BL is kept high. Hence, the write margin improves with a strong access and a weak pull up transistor at the cost of cell area and the cell read margin.

(4) Power and Delay: Power dissipation of the FinFET SRAM cell asses the utility of the cell in portable devices. The fundamental advantage of the FinFET based SRAM is in its low access time and power dissipation due to low SCE"s and leakage current in FinFET device. While a strong driving current reduces the access time it also increases the power dissipation in the SRAM cell. In SRAM, the propagation delay depends on the column height and wire delays. Thus segmentation is employed to reduce the delay. Since the power-delay-product is constant for a device increasing one decreases the other and vice-versa. Upsizing the FinFET device in SRAM cell decreases the delay at the cost of slightly increases power dissipation. However to reduce power dissipation and leakage currents need to be minimized which warrant an increase in the channel length or higher transistor threshold

voltages. Larger channel length results in higher delay and there exists a trade-off between these two performance indices [5-9].

VII. PROCESS VARIATION

Process variations can be environmental or lithographic in nature. Environmental factors arise during the operation of a circuit and include variations in supply voltage, switching activity and temperature across the chip. Lithographic variations are permanent in nature. These variations arise due to processing and masking limitations, and result in random or spatially varying deviations from the nominal value. A large amount of work has been done in analyzing conventional transistors under process variations. Various delay/leakage models have been proposed, which capture the deviations of process parameters. Further, these models have been used to propose statistical static timing analysis (SSTA) and full-chip leakage analysis algorithms to calculate the yield of circuits under process variations.

Process variations comprise of FinFET parameters (Channel length (Lg), Threshold voltage (Vth) etc.) which are no longer deterministic and die-to-die and within-die variations which may be random or correlated. Die-to-die fluctuations (from lot to lot and wafer to wafer) result from factors such as processing temperature and equipment properties. Conversely, within-die variations result from factors such as nondeterministic placement of dopant atoms and channel length variation across a single die. The reason behind the observed random distribution is due to the limited resolution of the photolithographic process which causes W/L variations in FinFET device. The variations in W and L are not correlated because W is determined in the field oxide step while L is defined in the poly and source/drain diffusion steps. In case of random variations the design parameters are totally uncorrelated as for instance, variations in FinFET length are unrelated to Vth variations.

With the scaling of technology, process imperfection is becoming a major concern in maintaining the reliability of an SRAM cell. The major sources of parameter variations in FinFET are Tsi and Lg. In FinFET based SRAM, these parameters include Fin widths (W_{fin}), Fin thickness (T_{fin}) and threshold voltage (Vth). FinFET based SRAMs are built using minimum size FinFET device to minimize area making it highly vulnerable to process variations . Memory designs are optimized for 6σ variations. SRAM failure can occur due to an increase in access time, failure to write a bit into the cell, accidental writing into memory during read, or loss of stored bit in standby mode. In scaled technologies, an optimal design strategy of FinFET based SRAM should consider minimization of area and access times in conjunction with reducing the failure probabilities due to variations [5-9].

VIII. CONCLUSION

Although, FinFET is most promising till date to replace bulk CMOS, and much energy and time have been devoted to the development of its process, modeling, and circuit design. The FinFET devices must be explored for stability and other issues under various conditions. This paper discussed various design aspects of FinFET based SRAM design. In future, FinFET based SRAM can be implemented in deep submicron technology.

REFERENCES

[1] Brain Swahn and Soha Hassoun (July 2006), Gate Sizing: FinFETs vs 32nm Bulk MOSFETs, In proceedings of DAC, Vol 2. pp. 218-224, California, USA.

[2] Y.Omura, S.Cristovoveanu, F. Gamiz, B-Y. Nguyen (2009), Advanced FinFET Devices for sub-32nm Technology Nodes: Characteristics and Integration Challenges, In proceedings of SoI Technology and Devices 14, Issue 4, pp.45–54.

[3] D. Hisamoto, W.-C. Lee, J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, T.-J. King, J. Bokor, and C. Hu.(1998) A folded-channel MOSFET for deep-sub-tenth micron era, International Electron Devices Meeting.

[4] Nirmal, Vijayakumar and Sam Jabaraj (2010), Nand Gate using FINFET for Nano-Scale Technology, In International Journal of Engineering Science and Technology, Vol. 2(5), pp-1351-1358.

[5] Alexei Nazarow, J. P. Colinge et al (2011). Semiconductor-On- InsulatorMaterial for Nanoelectronics Applications, Springer Heidelberg Dordrecht, London.

[6] Lourts Deepak A. and Likhitha Dhulipalla (2011), Design and Implementation of 32nm FINFET based 4x4 SRAM cell array using 1-bit 6T SRAM, In Proceedings of the International conference on Nanoscience, Engineering and Technology, ISBN:978–1–4673–0072–8, pp. 218 – 221.

[7] Saraju P.Mohanty JawarSingh, EliasKougianos, DhirajK.Pradhan, "Statistical DOE-ILP based powerperformance- process (P3) optimization of Nano-CMOS SRAM", Journal of VLSI, Volume 45, Issue 1, Jan 2012, pp 33-45.

[8] Zheng Guo, , Andrew Carlson, Liang-Teck Pang, Kenneth T. Duong, Tsu-Jae King Liu, and Borivoje Nikolic´, "Large- Scale SRAM Variability Characterization in 45 nm CMOS", IEEE Journal of Solid-State Circuits, Vol. 44, No. 11, Nov. 2009.

[9] Benton H. Calhoun, Yu Cao, Xin Li , "Performance Evaluation of Emerging Devices Digital Circuit Design Challenges and Opportunities in the Era of Nanoscale CMOS" , Proceedings of the IEEE, Feb 2008, pp-342-365.

