

Design of Nanoscale 3-T DRAM using FinFET

R.Baskar¹, R.Jayaprakash², M.Balaji³, M.Kannan⁴, A.Divya⁵, G.Neelakrishnan⁶
^{1,2,3,4,5,6}, (Department of ECE, Muthayammal College of Engineering/ Anna University, India)

Abstract: In the world of Integrated Circuits, Complementary Metal–Oxide–Semiconductor (CMOS) has lost its efficiency during scaling beyond 32nm. Scaling causes severe Short Channel Effects (SCE) which are difficult to hold back. As a result of such SCE many alternate devices have been studied. Some of the major contestants include Multi Gate Field Effect Transistor (MuGFET) like FinFET, Nano tubes, Nano wires etc. In this work, the basic gates and memory circuits like DRAM are designed in HSPICE software using CMOS structure and FinFET structures are analyzed and their performances like average power, standby power dissipation and leakage power are compared. A 32nm technology FinFET model is used to design DRAM.

Keyword: CMOS, Dynamic RAM, FinFET, Memory Cell, and Power Dissipation.

I. INTRODUCTION

Virtually Integrated Circuits rule the world. Laptops, I-pads and other digital appliances are now indissoluble parts of the structure of modern life. CMOS is a technology for constructing integrated circuits. High noise immunity and low static power consumption are the two key factors of CMOS. Considerable power is only used during the switching process of the transistors in the CMOS devices. Also CMOS devices do not produce thermal noise as other forms of logic, for example Transistor-Transistor Logic (TTL). It allows a high density of logic functions on a chip. This made the CMOS an adorable technique for the use in IC Technique. The leakage current in MOSFETs depends on various process parameters, the transistor size and the quiescent state of the circuit. One method of reducing leakage currents have to stack transistors in series. The stacking of transistors can exponentially decrease sub-threshold leakage in two ways. First the source nodes of the stacked transistors are no longer at ground, therefore, the source-bulk voltage, VSB becomes larger, thus increasing the effective threshold voltage through the body effect and lowering the leakage current. Secondly, a slight reverse bias between the gate and source (VGS) of the transistors has induced when stacked transistors are turned off and this reduces the effective driving voltage on the gate, and again decreases the sub-threshold leakage current.

1.1 Circuit Level Techniques to Reduce Power in Caches

The following subsections describe the method for reducing the dynamic and static power consumption of other VLSI circuits as Multiplexers, SRAM, etc. based on circuit-level techniques.

1.2 Voltage Scaling

The varying of supply voltage and clock frequency on demand, dynamic voltage scaling tries to provide high performance when it has required and low energy consumption during periods of standby. Since dynamic energy is a function of the supply voltage squared, (i.e. $E_d \propto V^2DD$) lowering the supply voltage can lower the dynamic energy consumption dramatically. A complete microprocessor with cache is developed where the impact of a changing supply voltage has measured on SRAM cells and on sense amplifiers.

1.3 Replica Technique

Due to the variability in the threshold voltage of transistors due to process variations, large delay margins must be designed in arrays. These large margins become the source of power inefficiencies along the bit line and through the sense amplifier. This method tries to minimize the power consumed along the bit lines and the sense amplifiers by using a self-timed approach. Self-timing has used to pulse the word lines to limit their bit line swing to the minimum needed by the sense amplifiers and to clock the sense amplifiers and minimize the time that they are ON. To clock the sense amplifiers, dummy columns are used to match the clock path to the data path, thus allowing the sense amplifiers to turn on only when the data has arrived and thus limiting the sense amplifiers power consumption.

II. BASICS OF FINFET TECHNOLOGY

The term FinFET was coined by University of California, Berkeley researchers (Profs. Chenming Hu, Tsu-Jae King-Liu and Jeffrey Bokor) to describe a non planar, double-gate transistor built on an SOI substrate, based on the earlier DELTA (single-gate) transistor design. The distinguishing characteristic of the FinFET is that the conducting channel is wrapped by a thin silicon "fin", which forms the body of the

device. The thickness of the fin (measured in the direction from source to drain) determines the effective channel length of the device.

2.1 Gate work function (ϕ_G)

Devices for high-performance logic need dual gate work function (ϕ_G) technology, which requires two different metals, one with a low work function (~ 4.4 eV) and one with a high work function (~ 4.9 eV), to set the required V_{TH} for NMOS and PMOS devices, respectively. Various approaches have been proposed to achieve dual work functions. However, the higher- V_{TH} requirement for SRAMs means that the ideal ϕ_G values for the two metals approach mid-gap. From a layout density perspective, in order to achieve minimal spacing between the NMOS and PMOS devices, dual gate implants needed to adjust the work functions are infeasible due to geometric shadowing effects. The gate running over the sidewalls of the fins precludes the possibility of dual implants to set the right V_{TH} , and so a single mid gap metal gate is needed from an ease of integration point of view. In addition, the n+/p+ drains need to be strapped using silicide. A single metal gate with $\phi_G = 4.75$ eV provides with symmetric NMOS/PMOS performance, with $\phi_G = 4.6$ eV requires the use of accumulation mode (Acc) PMOS. The NMOS and PMOS device require separate titled gate implants to set the correct V_{TH} , which is not possible to achieve in a dense DRAM cell due to shadowing effects.

2.2 Channel Doping

Channel doping is a way to set the correct V_{TH} in FinFET devices. However, since the Si fin thickness is very small, the level of channel doping required to set the correct V_{TH} is very high. Higher channel doping results in mobility degradation from Coulombic scattering and increased transverse electric field and causes random dopant fluctuation effects resulting in the statistical variation of the V_{TH} . Therefore, the channel is best left undoped, thereby eliminating the impact of dopant fluctuations on V_{TH} . If $\phi_G = 4.6$ eV is chosen, the PMOS load device must be doped and is in an accumulation mode device (acc-PMOS). The acc-PMOS has lower performance and shows greater sensitivity to variations, and therefore this design was not pursued.

2.3 Body Thickness

In order to control short channel effects, the body thickness needs to be in the $LG/2$ to $0.7 LG$ range. Achieving this with good dimensional control can be challenging. There are novel technologies such as spacer lithography, and more conventional approaches controlled photoresist ashing and sacrificial oxidation of the single crystalline Si-fin. Variation in fin width is potentially a major source of DRAM variations, if not controlled adequately.

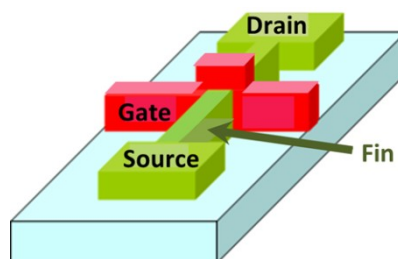


Fig.1. Structure of FinFET

2.4 Fin surface orientation

The FinFET sidewall surfaces fabricated on standard orientation (001) wafers lies along (110) planes, and along (100) planes if the layouts are rotated by 45° . For a (110) surface, hole mobility has enhanced while electron mobility has been degraded as compared to a (100) surface. Due to velocity saturation effects in nanoscale devices, only a small fraction of the mobility charge results in a change in $I_{D, SAT}$.

2.5 Sensitivity of FinFET performance to process-induced variations

Control of critical dimensions does not track their scaling, thus the ratio of the standard deviation (σ), over the average (μ) increases. Designing large arrays requires design for 5 or more standard deviations ($>5\sigma$). With increasing variations, it becomes difficult to guarantee near-minimum-sized cell stability for large arrays for 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 storage density.

The process-induced variation in FinFET performance arises from statistical variations in LG and T_{Si} . The devices with different T_{Si} are optimized individually, by changing the $LEFF$ by adjusting the gate sidewall spacer thickness and tuning ϕ_G to meet the IOFF target and $DIBL = 100$ mV/V. A thinner T_{Si} yields better leakage and control of short channel effects (SCE) so that a lower ϕ_G and $LEFF$ can be used, leading to larger ION. The simulations to study the impact of process variations assume that the same patterning technology has used to define the fins and the gates and therefore have the same absolute variability, with a $3\sigma = 10\%$ of LG .

This large variation in TSi results in a large spread in ION-IOFF. If spacer lithography has used to pattern the fins, the degree of variations in TSi can be reduced because the spacer thickness can be well controlled through the CVD deposition of the sidewall material [18]. The spread in ION-IOFF has comparable for the thinner and the thicker silicon body thickness, due to a tradeoff between better control of short channel effects using thinner TSi versus a larger relative variation in TSi for the thinner body case.

TABLE I
DEVICE PARAMETERS USED FOR SIMULATIONS

Parameters	FinFET	Bulk-Si
LG (nm)	32	32
LSD (nm)	0.7	24
Tox (Å)	14	11
Tsi (nm)	8.6	3.5
VDD (V)	0.3	1.0
Channel doping, NBODY (cm ⁻³)	2e16	4e18
HFIN (NM)	40	-
S/D doping gradient (nm/dec)	4	4

III. DRAM

As the trend for high-density RAM arrays forces the memory cell size to shrink, alternative data storage concepts must be considered to accommodate these demands. In a dynamic RAM cell, binary data is stored simply as charge in a capacitor, where the presence or absence of stored charge determines the value of the stored bit. Note that the data stored as charge in a capacitor cannot be retained indefinitely, because the leakage currents eventually remove or modify the stored charge. Thus, all dynamic memory cells require a periodic refreshing of the stored data, so that unwanted modifications due to leakage are prevented before they occur. The use of a capacitor as the primary storage device generally enables the DRAM cell to be realized on a much smaller silicon area compared to the typical SRAM cell. Notice that even as the binary data is stored as charge in a capacitor, the DRAM cell must have access devices, or switches, which can be activated externally for "read" and "write" operations. But this requirement does not significantly affect the area advantage over the SRAM cell, since the cell access circuitry is usually very simple. Also, no static power is dissipated for storing charge on the capacitance. Consequently, dynamic RAM arrays can achieve higher integration densities than SRAM arrays.

The circuit diagram of a typical three-transistor dynamic RAM cell is shown in Fig. 2 as well as the column pull-up (precharge) transistors and the column read/write circuitry. Here, the binary information is stored in the form of charge in the parasitic node capacitance *CI*. The storage transistor M2 is turned on or off depending on the charge stored in *CI*, and the pass transistors M1 and M3 act as access switches for data read and write operations. The cell has two separate bit lines for "data read" and "data write," and two separate word lines to control the access transistors.

The operation of the three-transistor DRAM cell and its peripheral circuitry is based on a two-phase non-overlapping clock scheme. The precharge events are driven by 01, whereas the "read" and "write" events are driven by 02. Every "data read" and "data write" operation is preceded by a precharge cycle, which is initiated with the precharge signal PC going high. During the precharge cycle, the column pull-up transistors are activated, and the corresponding column capacitances C2 and C3 are charged up to logic-high level. With typical enhancement type nMOS pull-up transistors (V.J 1.0 V) and a power supply voltage of 5 V, the voltage level of both columns after the precharge is approximately equal to 3.5 V.

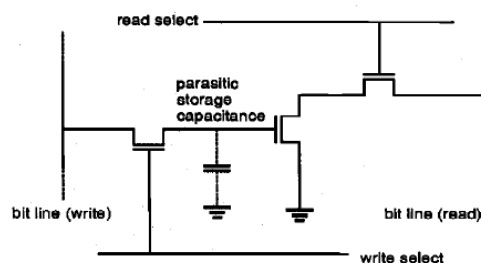


Fig.2. 3T DRAM Cell

IV. RESULTS AND DISCUSSIONS

4.1 Simulation Results

3-T DRAM memory cell is designed by using three transistors in CMOS transistors as well as FinFET devices. The simulation waveform for reading and writing the logic values '1' and '0' is obtained and verified

by using HSPICE tool. The simulation waveforms for CMOS based DRAM and FinFET based DRAM is shown below.

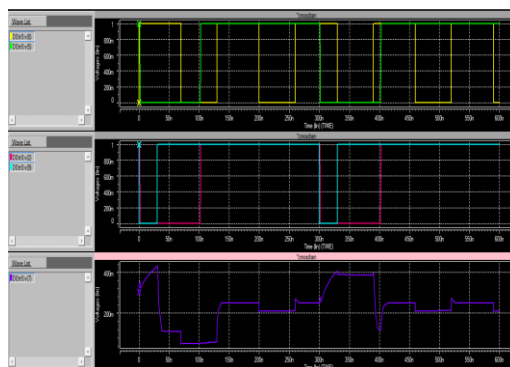


Fig.3. Simulation waveform for 32nm CMOS based 3T-DRAM

The above figure shows the simulation waveform for 32nm CMOS based 3T- DRAM memory cell. In the first panel, V(8) [yellow color waveform] represents write line (WR) and V(5) [green color waveform] represents read line (RL).

In the second panel, V(2) [pink color waveform] represents bit line (BL) and V(9) [blue color waveform] represents bit line bar (BLB).

Finally, V(7) [violet color waveform] shows the output waveform that indicates read and write operation based on charging and discharging of capacitor.

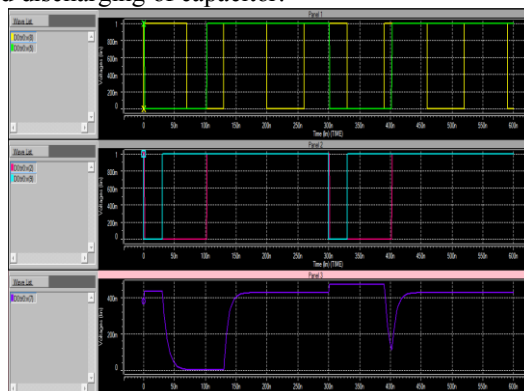


Fig.4. Simulation waveform for 32nm FinFET based 3T - DRAM

Figure 4 shows the simulation waveform for 32nm FinFET based 3T - DRAM memory cell. In the first panel, V(8) [yellow color waveform] represents write line (WR) and V(5) [green color waveform] represents read line (RL).

In the second panel, V(2) [pink color waveform] represents bit line (BL) and V(9) [blue color waveform] represents bit line bar (BLB).

Finally, V(7) [violet color waveform] shows the output waveform that indicates read and write operation based on charging and discharging of capacitor.

4.2 Performance Comparison

TABLE II
PERFORMANCE COMPARISON FOR CMOS AND FINFET BASED 3-T DRAM

DRAM	Average Power	Power Dissipation	Leakage Power
CMOS based Design	10.869uW	235.2313uW	1.8781uW
FinFET based Design	9.5136uW	144.7353uW	2.4303nW

From the above table, it clear that designing DRAM by using FinFET is better than DRAM design by using FinFET in terms of power dissipation, leakage and average power.

ACKNOWLEDGEMENT

Our sincere thanks to Ms.D.Sathya M.E., Assistant Professor, Department of Electronics and Communication Engineering, Maha Bharathi Engineering College, Villupuram(Dt).

REFERENCES

- [1] Kidong Kim, Ohseob Kwon, Jihyun Seo and Taeyoung Won ‘Nanoscale Device Modelling and Simulation: Fin Field Effect Transistor’, Japan Journal of Applied Physics. 43 (2004) pp. 3784-3789.
- [2] Brian Doyle, Reza Arghavani, Doug Barlage, Suman Datta, Mark Doczy, Jack Kavalieros, Anand Murthy and Robert Chau, “Transistor Elements for 30nm Physical Gate Lengths and Beyond”, Intel Technology Journal, volume 6, Issue 2, May 2002.
- [3] D.Hisamoto, W.C.Lee, J.Kedzierski, H.Takeuchi, C.Kuo, E.Anderson, T.J.King, J.Bokor and C.Hu “FinFET- A self Aligned Double Gate MOSFET scalable to 20 nm”, IEEE Trans. Electron Devices, Vol.47, No. 12, pp 2320-2325, 2000.
- [4] Kunihiro Suzuki, Tetsu Tanaka, Yoshiharu Tosaka, Hiroshi Horie and Yoshihiro Arimoto, “Scaling Theory for Double-Gate SOI MOSFET’s”, IEEE Trans. Electron Devices, Vol.40, No. 12, pp 2326-2329, 1993.
- [5] A. Carlson, Z. Guo, S. Balasubramanian, L.-T. Pang, T.-J. King Liu, and B. Nikolic “FinFET SRAM with Enhanced Read / Write Margins”, 2006 IEEE International SOI Conference Proceedings, PP 121-126.
- [6] Chevillon, N. Mingchun Tang, Pregaldiny, F. Lallement, C. and Madec, M. “FinFET compact modeling and parameter extraction”, Mixed Design of Integrated Circuits & Systems, 2009. MIXDES '09. MIXDES-16th International Conference, 2009.
- [7] C. Auth, P. Bai, et al (2005), ‘A 65nm logic technology featuring 35nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57 nm SRAM cell,’ presented at 2004 International Electron Devices Meeting. San Francisco, CA.