Analysis and Performance Comparison of CMOS and FinFET based DRAM Memory Cell

R.Baskar
ME-Applied electronics
Muthayammal Engineering College
Rasipuram- 637 408,Namakkal(Dt),Tamil Nadu

S.Jim Hawkinson.,M.E.,
Assistant Professor, Department of ECE
Muthayammal Engineering College
Rasipuram- 637 408,Namakkal(Dt),Tamil Nadu

ABSTRACT

In the world of Integrated Circuits, Complementary Metal–Oxide–Semiconductor (CMOS) has lost its credentiality during scaling beyond 32nm. Scaling causes severe Short Channel Effects (SCE) which are difficult to suppress. 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 modeled in HSPICE software using CMOS structure and FinFET structures are analyzed and their performances like standby power Consumption and static noise margin are compared. Also a low power and robust DRAM cells based on FinFET has been proposed for 32nm technology.

Keywords

CMOS, Dynamic RAM, FinFET, Memory Cell, Power Dissipation .

1.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.

A. 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.

B. 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. Ed α V2DD) 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.

C. 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.

2. RELATED WORK

Kidong Kim [1] reported the two-dimensional Quantum-Mechanical (QM) simulation of FinFET. Current-voltage (I-V) characteristics have compared with the experimental data. Device optimization has been performing in order to suppress the Short-Channel Effect (SCE) including the sub-threshold swing, threshold voltage roll-off, Drain Induced Barrier Lowering (DIBL).

Brain Doyle et al [2] have fabricated conventional planar transistors of various gate lengths down to as small as 10nm poly silicon gate lengths, in order to examine transistor scaling. One feasible method of significantly improving off-state leakage has through reducing the sub-threshold gradient. It has been shown that Depleted Substrate Transistors (DST), a broad category of devices that include single- and double gate transistors, whose active channel region stays fully depleted during operation, can achieve near-ideal sub threshold gradients

and a reduction in off-state leakage of at least two orders of magnitude over bulk transistors. In this paper, gate length scaling on bulk MOS Devices is examined.

A novel self-aligned double-gate MOSFET was proposed by Digh Hisamoto et al [3] in which FinFET suppresses the short channel effect, by using boron-doped Si04Ge06 as a gate material, the desired threshold voltage was achieved for the ultrathin body device. The quasi planar nature of this new variant of the vertical double-gate MOSFETs can be fabricated relatively easily using the conventional planar MOSFET process technologies. A novel self-aligned double-gate SOI structure (FinFET) is proposed as a nano CMOS device. In spite of its double-gate structure, FinFET is close to its root, the conventional MOSFET in terms of layout and fabrication. The result is a quasiplanar device.

A scaling theory for double-gate SO1 MOSFET's, which gives a guidance for the device design (silicon thickness tsi; gate oxide thickness tox) that maintaining a subthreshold factor for a given gate length LG. According to their theory, a device can be designed with a gate length of less than 0.1 μm while maintaining the ideal subthreshold factor, which has verified numerically with a two-dimensional device simulator. A natural length relevant to the scaling theory has been derived for double-gate SO1 MOSFET's, and described how to design tox and tsi for a given gate length maintaining a proper S-factor. According to the theory, almost the ideal S-factor value can be expected even with LG of less than 0.1 μm was proposed by Kunihiro Suzuki et al [4].

The impact of this pass-gate feedback (PGFB) technique on cell write ability and discussed gate work function (Dm) tuning for optimization of the trade-off with read margin was examined by A.Carlson et al[5]. The improved cell write-ability, the p-channel pull-up devices can also be operated in BG mode, with their back gates driven by a separate write word line. This has Pull-Up Write Gating (PUWG) technique is effective for maintaining larger than 6 standard deviations yield down to 0.4V VDD without area penalty, making FinFET-based DRAM compelling for high-density memory applications. FinFET-based DRAM designs with dynamic feedback and write word line gating have shown to have improved read and write performance and higher projected yield, without area penalty.

3. BASICS OF FINFET TECHNOLOGY

3.1 Gate work function (φG)
Devices for high-performance logic need dual gate wok function (φG) 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 VTH for NMOS and PMOS devices, respectively. Various approaches have been proposed to achieve dual work functions. However, the higher-VTH requirement for SRAMs means that the ideal φG 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 VTH, 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 φG = 4.75eV provides with symmetric NMOS/PMOS performance, with φG=4.6eV requires the use of accumulation mode (Acc) PMOS. The NMOS and PMOS device require separate titled gate implants to set the correct VTH, which is not possible to achieve in a dense DRAM cell due to shadowing effects.

3.2 Channel Doping
Channel doping is a way to set the correct VTH in FinFET devices. However, since the Si fin thickness is very small, the level of channel doping required to set the correct VTH is very high. Higher channel doping results in mobility degradation from Columbic scattering and increased transverse electric field and causes random dopant fluctuation effects resulting in the statistical variation of the VTH. Therefore, the channel is best left undoped, thereby eliminating the impact of dopant fluctuations on VTH. If φG =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.

3.3 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.

3.4 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 45o. 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 $I_{D, SAT}$.

3.5 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 (>5σ). 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 TSi. The devices with different TSi are optimized individually, by changing the LEFF by adjusting the gate sidewall spacer thickness and tuning φG to meet the IOFF target and DIBL = 100mV/V. A thinner Tsi yields better leakage and control of short channel effects (SCE) so that a lower φG 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 3s = 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 1 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

4. 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.
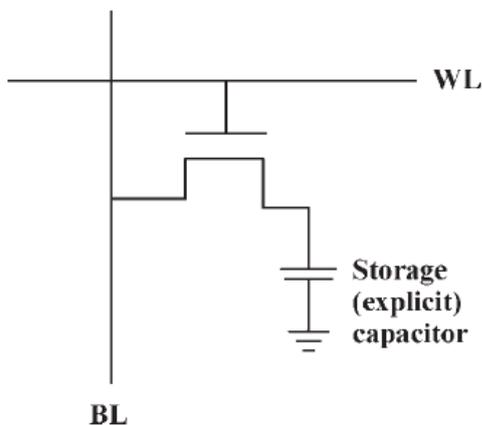

Fig 1. 1T DRAM Cell

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 *Cl.* The storage transistor M2 is turned on or off

depending on the charge stored in *C1,* and the pass transistors Ml 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.
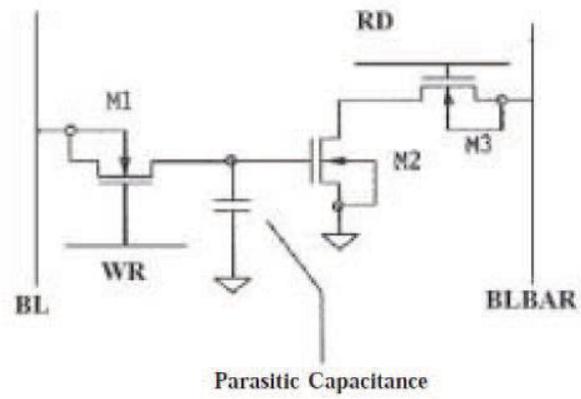

Fig 2. 3T DRAM Cell

5. RESULTS AND DISCUSSIONS

3-T DRAM memory cell is designed by using three transistors in MOS 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.

Table 1 Comparison of average power & power dissipation between CMOS as well as FinFET based DRAM

OUTPUT	CMOS	FINFET
Average Power	3.5013E10	4.4503E-09
PowerDissipation	15.2110nW	4.9156 nW

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

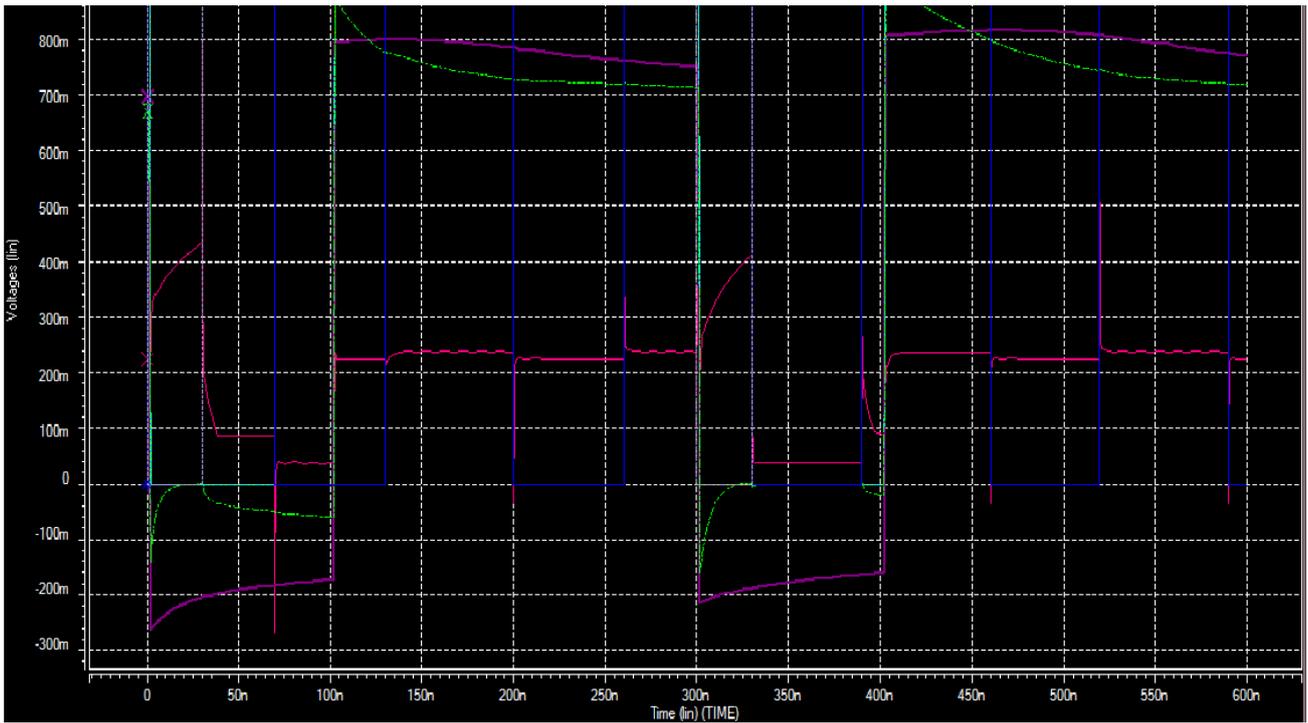The simulation waveforms for CMOS based DRAM and FinFET based DRAM is shown below.

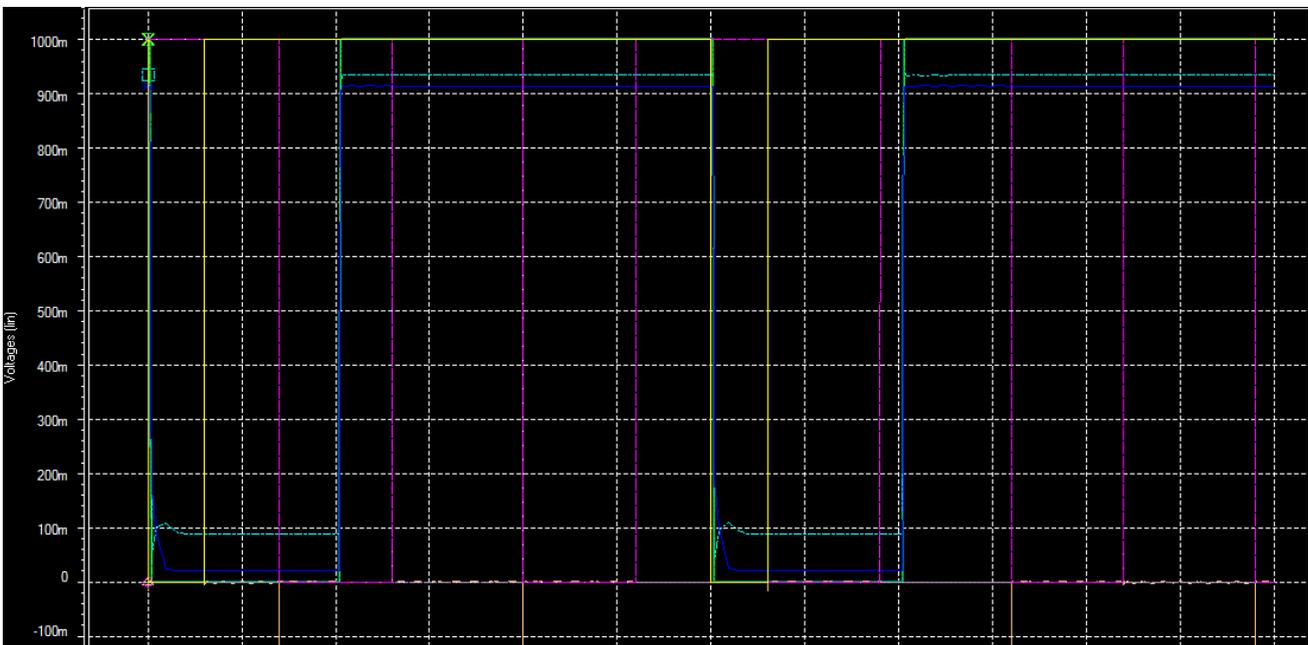**Fig 3. Simulation waveform for CMOS based DRAM memory cell**

Fig 4. Simulation waveform for FinFET based DRAM memory cell

6. ACKNOWLEDGEMENT

7. 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 Alligned Double Gate MOSFET scalable to 20 nm", IEEE Trans. Electron Devices, Vol.47, No. 12,pp 2320-2325, 2000.

[4]. Kunihoro Suzuki, Tetsu Tanaka ,Yoshiharu Tosaka , Hiroshi Horie and Yoshihiro Arimoto, "Scaling Theory

[5]. for Double-Gate SO1 MOSFET's", IEEE Trans. Electron Devices, Vol.40, No. 12,pp 2326-2329, 1993.

[6]. 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.

[7]. 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.

[8]. 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.

[9]. T. Su, J. B. Jacobs, J. E. Chung, and D. A. Antoniadis, 'Short-channel effects in deep-submicrometer SOI MOSFETS (1993),' presented at Proceedings of 1993 IEEE International SOI Conference. Palm Springs, CA