University of Oslo
Department of Informatics

Ultra Low Power Digital Circuit Design for Wireless Sensor Network Applications

PhD Thesis

Farshad Moradi

October 2011

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 1121

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Ultra Low Power Digital Circuit Design for Wireless Sensor Network Applications

Farshad Moradi

Institute of Informatics

UNIVERSITETET I OSLO

October 2011

To my family

ABSTRACT

Farshad Moradi, PhD, University of Oslo, July 2011, "Ultra Low Power Digital Circuit Design for Wireless Sensor Network Applications", Major advisor: Prof. Dag T. Wisland, Co-advisors: Prof. Snorre Aunet and Prof. Yngvar Berg

As the CMOS technology continues to scale down into the nano-scale regime, robustness of the circuit with respect to process variation and soft error are becoming major obstacles for circuit designers. Storage elements (SRAM, flip-flops) are particularly vulnerable to process variation and soft errors. Thus, in this work, we have focused on storage elements – to improve the yield loss in SRAM due to process variations and to design a soft error tolerant flip-flop. SRAMs are particularly vulnerable to failures due to process variation resulting in reduced yield. The main problem with SRAM is the conflicting requirements for read stability and writeability. In this work, we propose designs to overcome conflicting trade-off between read and write stability. Furthermore, new SRAM cells, namely 11T-SRAM, PMOS access transistor SRAM, are proposed with capability of working at near-threshold voltages, properly. The effect of body-biasing on SRAM cell is explored to show improvements from body-biasing in sub-threshold regions. Results show at least 30% improvement in read noise margin for proposed SRAM cells while write margin is improved. Furthermore, to overcome short channel effect, different candidate transistor structures have been investigated to replace the bulk MOSFETs. Among them, FinFET is considered to be a promising candidate for scaled CMOS devices in sub-22-nm technology nodes. In this work, by introducing a new device, the read and write stability for SRAM is improved by 20% and 9% respectively, while performance is improved by 56%compared to conventional designs. we study the double-gate FinFET SRAM technology-circuit design space to understand the interplay of device short-channel-effect (SCE), SRAM area, access time, soft error immunity, stability under process variations and leakage. Several Flip-Flop designs are designed to reduce power in DSP applications. Simulation results show 40% improvement in total power saving for some DSP applications such as FIR filters and other DSP applications. Design challenges in submicron CMOS technology are investigated in details for sub-threshold designs for wireless sensor network applications. In this work different CMOS model such as ST Microelectronic, TSMC and IBM models are used for different application.

Ultra Low power Digital Circuit Design for Wireless Sensor Network Applications

Farshad Moradi, PhD, University of Oslo, July 2011
Hovedveileder: Førsteamanuensis Dag T. Wisland
Medveiledere: Professor Snorre Aunet og Professor Yngvar Berg

Ny forskning innenfor feltet trådløse sensornettverk åpner for nye og innovative produkter og løsninger. Biomedisinske anvendelser er blant områdene med størst potensial og det investeres i dag betydelige beløp for å bruke denne teknologien for å gjøre medisinsk diagnostikk mer effektiv samtidig som man åpner for fjerndiagnostikk basert på trådløse sensornoder integrert i et "helsenett". Målet er å forbedre tjenestekvalitet og redusere kostnader samtidig som brukerne skal oppleve forbedret livskvalitet som følge av økt trygghet og mulighet for å tilbringe mest mulig tid i eget hjem og unngå unødvendige sykehusbesøk og innleggelser.

For å gjøre dette til en realitet er man avhengige av sensorelektronikk som bruker minst mulig energi slik at man oppnår tilstrekkelig batterilevetid selv med veldig små batterier. I sin avhandling " Ultra Low power Digital Circuit Design for Wireless Sensor Network Applications" har PhD-kandidat Farshad Moradi fokusert på nye løsninger innenfor konstruksjon av energigjerrig digital kretselektronikk. Avhandlingen presenterer nye løsninger både innenfor aritmetiske og kombinatoriske kretser, samtidig som den studerer nye statiske minneelementer (SRAM) og alternative minnearkitekturer. Den ser også på utfordringene som oppstår når silisiumteknologien nedskaleres i takt med mikroprosessorutviklingen og foreslår løsninger som bidrar til å gjøre kretsløsninger mer robuste og skalerbare i forhold til denne utviklingen. De viktigste konklusjonene av arbeidet er at man ved å introdusere nye konstruksjonsteknikker både er i stand til å redusere energiforbruket samtidig som robusthet og teknologiskalerbarhet øker. Forskningen har vært utført i samarbeid med Purdue University og vært finansiert av Norges Forskningsråd gjennom FRINAT-prosjektet "Micropower Sensor Interface in Nanometer CMOS Technology".

ACKNOWLEDGMENTS

I would like to take this opportunity to thank all the people without whom this dissertation could not have been possible.

I sincerely like to thank my thesis advisor **Prof. Dag Wisland**, for mentoring me in this research work. Observing his critical thinking, sound vision and great communication skills has really helped me to improve professionally as well as personally.

I would also thank my advisor, at Purdue University, **Prof. Kaushik Roy**, for giving me this opportunity to work in his Lab. His research visions and sharp insights were instrumental in the completion of this work. He taught me how to transform a new idea into a research accomplishment. My sincere thanks to him for giving advice and help.

I would like to thank my Co-advisors, **Prof**. **Snorre Aunet** and Prof. **Yngvar Berg** for their guidance and suggestions during the research work.

I would like to thank **Prof. Hamid Mahmoodi** for his contiguous help and advice during my research. My sincere thanks to him for providing me to work and learn from him. His continuous encouragement and motivation during this research has enabled me to study/explore various research areas. My sincere thanks to him for providing me the opportunity to learn and work with him.

I would like to thank Nanoelectronic group at University of Oslo.

I would like to thank Nanoelectronic Research Lab (NRL) members at Purdue University.

I would like to thank research council for their support during the project.

Specially, I would like to thank Charles Augustine(Intel), Georgios Karakonstantis(EPFL), Georgios Panagopoulos(Purdue), Sang Phill Park(Intel), Ashish Goel (Broadcom), Sumeet Kumar Gupta(Purdue), Niladri Mojumder(Global Foundries), Sabal Mukhapadhyay (Georgia Tech. University).

Last but not the least, my sincere thanks to my wife, Negin, my parents, and my family members whose support and love made this work possible.

Contents

List of Figures:

Chapter 1

Design Challenges in Nano-Scale Technology

Chapter 1

Design challenges in Nano-Scale Technology

Introduction

CMOS devices have been scaled down aggressively in the last few decades resulting in higher integration density and improved performance. However, due to short channel effects, threshold voltage (V_{th}) scaling, oxide thickness scaling and increased doping density, the "off" current in the devices has increased drastically with technology scaling. Hence, as we are approaching the end of the silicon roadmap, controlling leakage current is becoming a major problem. Moreover, statistical variations in process parameters, such as device structure (channel length, oxide thickness, width etc), location and number of dopants in channel (random dopant fluctuation), is increasing with technology scaling. The variation in process parameters results in large distribution in delay and leakage and significantly reduces robustness of a circuit. Hence, large leakage current and increasing process variations have emerged as two major obstacles for designing CMOS circuits (logic and memory) at the end of silicon roadmap.

Smaller transistors are inherently faster and consume less dynamic power. However, when millions of transistors are integrated together to create a complex VLSI system, we observe several new challenges threatening the reliability of computation. Some of them are as follows:

Leakage Power: With ever-increasing operating frequency and more transistors on a single die, switching power has increased significantly (Fig.1.1). By scaling the devices to sub-50nm regimes, controlling over the channel and second order effects such as short channel effects, DIBL, narrow width effect etc., has become less. Furthermore, by technology scaling, leakage components increase significantly. Fig.1.2 shows the leakage power components and the percentage of leakage power for different technology nodes. Both dynamic and leakage power consumption affect the reliability of the underlying devices and reduce the battery lifetime of handheld devices. In another word, I_{ON}/I_{OFF} ratio decreases with scaling technology. Especially, for ultra low voltage applications such as

wireless sensor nodes, using techniques to reduce the leakage current is crucial due to a very low I_{ON}/I_{OFF} ratio in near threshold or sub-threshold regions.

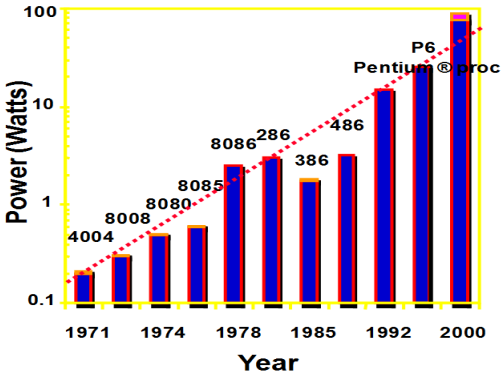


Fig.1.1. Increasing power over technology generations[Intel]

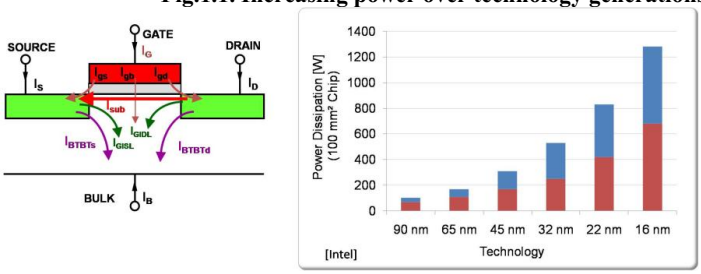


Fig.1.2 leakage power is becoming the dominant component of overall power consumption in scaled technologies

Process Variation: Sub-wavelength lithography has led to large variation in transistor geometries (*L;W; TOX*) and the flat-band voltage (*VFB*). Process variation effects are more stringent in small size devices. Variations in channel length, channel width, oxide thickness, threshold voltage, line-edge roughness, and random dopant fluctuations are the sources of the inter-die and the intra-die variations in process parameters [the random variations in the number and location of dopant atoms in the channel region of the device resulting in the random variations in transistor threshold voltage (RDF)]. One of this inter die effects is threshold voltage variations due to the changes of a single transistor (e.g. threshold voltage increases if temperature is reduced). However, intra-die variations may be different from one transistor to another (i.e. increase in V_{th} for one device on the other hand decrease in threshold voltage for another). An example of the systematic intra-die variation can be the change in the channel length of different transistors of a die that are spatially correlated. The RDF induced V_{th} variation is a classic example of the random intra-die variation. Such variations along with higher levels of integration can lead to large spread in circuit delay, power, and robustness across different dies (Fig.1.2). Therefore, a circuit designed to meet a target delay using nominal *VTH* transistors may not meet the delay target, leading to parametric delay failures. Parameter variations have severe impact (both in terms of errors as well as leakage) on minimum geometry circuits such as SRAM cells. Inter-die parameter variations, coupled with the intrinsic on-die *VTH* variation can result in stability failures in SRAM cells, degrading the memory yield. A cell failure can

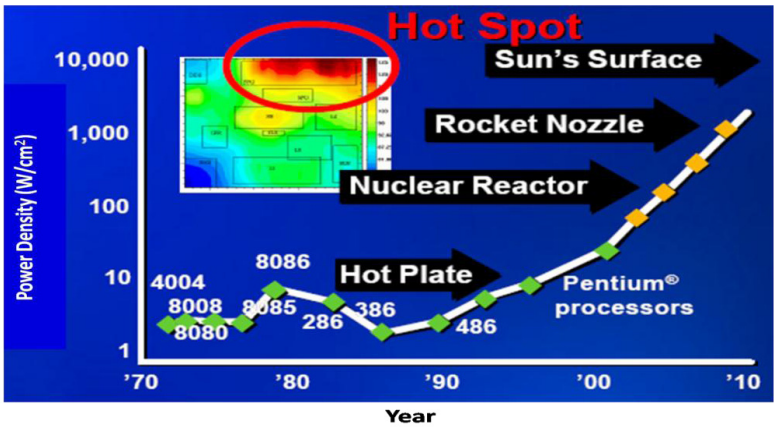


Fig.1.3. Increasing power density with technology scaling [Intel]

occur due to:(a) an increase in the cell access time (access failure), (b) unstable read/write operations (read/write failure),or (c) failure in the data holding capability of the cell at a lower supply voltage. Body-biasing has been used for mitigating the impact of inter-die process variation and reducing the parametric failures [1]. However, it has been observed that the effectiveness of body-biasing reduces with technology scaling [2]. Different challenges in ultra low supply voltage, especially for storage circuits, are investigated in this dissertation. Unfortunately, for lower supply voltages the effect of process variations increases significantly. Therefore looking at process variations in different designs, especially for very low voltages is interesting.

Power Density and Die-Temperature: Increasing power density due to faster clock, power consumption (due to dynamic power and leakage) and high device integration is becoming another issue. The increased power density translates into excessive heat while the cooling capability of the package remains limited. Hence, it gives rise to elevation of overall die temperature as well as localized heating at highly active regions of a chip (called "hot-spots", Fig.1.3). Traditionally, a circuit is designed to operate at worst-case temperature which is inefficient given the fact that the circuit may experience the worst-case condition only for short duration. Dynamic thermal management techniques like logic shutdown, clock gating, frequency scaling, voltage-frequency throttling etc. have been proposed in past [3] [4] [5]. Although these techniques are capable of bringing down power density and temperature, they also degrade the performance considerably. Dynamic voltage scaling can be an effective way to reduce temperature in the event of overheating because it reduces both the switching as well as leakage power. However, if the clock frequency is not scaled simultaneously, then increased path delays (due to lower voltage) may result into wrong computation. Combined V-f control technique reduces leads to cubic benefit in terms of power saving at the cost of complex control circuitry and pipeline stalls [5].

Nano-scale reliability: As dimensions of MOS devices have been scaled down, new reliability problems are coming into effect. One of these emerging reliability issues is aging effects that result in device performance degradation over time. NBTI (Negative biased temperature instability) is a well-known aging phenomenon, which is a limiting factor for future scaling of devices. NBTI results the generation of trapped charges, which cause threshold voltage degradation of PMOS. Another reliability issues are PBTI, Hot carrier injection (HCI), and TDDB that should be considered for ultra-scaled nano-scale devices.

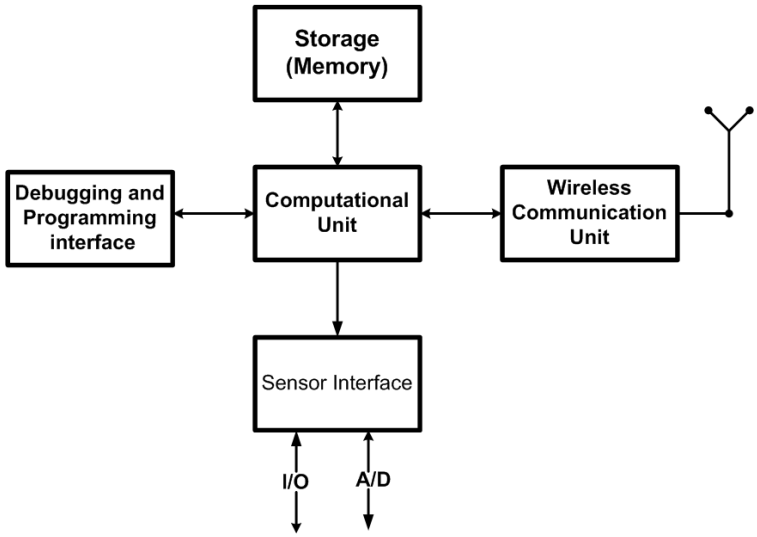


Fig.1.4. Basic sensor node block

Due to higher electric field in ultra-thin body (UTB) devices, HCI is increased significantly that should be considered for future technologies.

1.1. Thesis Organization:

In this thesis, we consider different challenges in ultra low power circuit design for wireless sensor network applications, portable devices etc. We present several leakage-tolerant, noise-immune, low power and high performance designs for different DSP application. The main contribution of this thesis is to design ultra low power digital circuit design for wireless sensor nodes at very low supply voltages (Near/Sub-threshold). However, in this thesis we tried to consider other potentials for future low power and reliable design by considering new devices such as double-gate MOSFETs. Therefore, the main contributions of this work are as follows:

- **High-Speed and leakage-tolerant design:** In this part of our work, we propose several low leakage logic circuits and basic blocks such as wide OR-gates, Full adder (FA) design, XOR, Multiplexer (MUX), comparators etc. Chapter 2, covers designing basic logic circuit blocks for high performance and low power targets.
- **Flip-Flop design:** Among the logic elements, latches and flip-flops (FF) are critical for the performance of a digital system. Flip-flops, the characteristic building block of any clocking or pipeline system, are the most important components in synchronous VLSI designs. Unless they are carefully designed and characterized, the performance of FFs has a critical effect on the system. Interestingly, the FFs often consumes the largest share of the total system power (20%-50%) [6]. We propose several new flip-flops to reduce the total power of DSP applications such as filters, and DCT. Results show at least 27% total saving compared to using conventional FFs.
- **SRAM Design:** SRAM bit-cells utilizing minimum sized transistors are susceptible to various random process variations. Hence reducing the memory operating supply voltage (V_{min}), while maintaining the yield is becoming extremely challenging in nano-scale technologies. Chapter 4 focuses on developing novel circuit techniques for robust memory operation with lower V_{min}. In this chapter we propose different SRAM design

at bit-cell level and SRAM architecture level. Simulation results show significant improvement in read and write robustness (2X) while lower leakage is achieved.

- **Challenges in Sub-Threshold Design**: challenges observed in 65nm technology for circuits utilizing sub-threshold region operation are presented. Different circuits are analyzed and simulated for ultra low supply voltages to find the best topology for sub-threshold operation. The effect of body-biasing technique is investigated in detail to find the optimal point to get the minimum power and maximum performance. SRAM and Flip-Flop designs are re-designed using body-biasing technique and stacking effect in chapter 5.
    - **Double-Gate Design:** To overcome short channel effects (SCE), different candidate transistor structures have been investigated to replace the bulk MOSFETs. Among them, FinFETs is considered to be a promising candidate for scaled CMOS devices in sub-22-nm technology nodes. This device shows increased immunity to SCE due to improved channel control by the gate voltage. Furthermore, threshold voltage (V_{th}) can be easily controlled by engineering the gate contact work function. Moreover, V_{th} variations due to random dopant in the channel region (RDF) are reduced due to almost intrinsic channel doping. In this chapter, new FinFET device is introduced to improve the device characteristics. Furthermore, a robust SRAM design is introduced to resolve the conflict between read and write margins. Results show 10 times less leakage while DIBL and subthreshold-slope are improved by three times and 20%, respectively. Furthermore, the conflict between read and write is resolved.

In general, the main goal of the thesis is to consider low power design for wireless sensor nodes. Fig.1.4 shows a typical wireless sensor node block that includes: Energy sources, wireless communication interface, sensor interface, including analog/digital converter and I/O parts, computational unit, that processes sensor reading and networking algorithms and is able to store small amount of intermediate data, Memory that stores program code, intermediate data, and queries and networking data, and finally debugging interface. Therefore, we focused on designing the storage blocks such as memory and flip-flops, and low power digital design for processing unit. At the same time, we explore the behavior of nano-scale transistors at ultra low supply voltages.

Chapter 2

Ultra Low Power Digital Circuit Blocks for Wireless Sensor Nodes

Chapter 2

Ultra Low Power Digital Circuit Blocks for Wireless Sensor Nodes

Introduction

With the rapid development of portable digital applications, the demand for increasing speed, compact implementation, and low power dissipation triggers numerous research efforts [7]-[10]. The role of power dissipation in VLSI systems is pervasive. For high performance design, power dissipation can be the limiting factor to clock speed and circuit density because of the inability to get power to circuits or to remove the heat that they generate. For portable information systems, power dissipation has a direct bearing on size, weight, cost, and battery life. Consequently, power dissipation is becoming widely recognized as a top-priority issue for VLSI circuit design. The challenge facing the VLSI designer is to find and effectively apply circuit techniques that can balance the needs for performance with those of power dissipation [11]. Therefore ultra low power circuits design becomes the major candidate for portable applications such as wireless sensor nodes. One common technique for reducing power is power supply scaling. For CMOS circuits the cost of lower supply voltage is lower performance. Scaling the threshold voltage can limit this performance loss somewhat but results in increased leakages [12]. Other techniques used in low power design include clock gating and dynamic voltage/frequency scaling [13], [14].

Subthreshold circuit design involves scaling the supply voltage below the threshold voltage, where load capacitances are charged/discharged by subthreshold leakage currents. Leakage currents are orders of magnitude lower than drain currents in the strong inversion regime, therefore there is a significant limit on the maximum performance of subthreshold circuits. Therefore, traditionally, subthreshold circuits have been used for applications which require ultra-low power dissipation, with low-to-moderate circuit performance [15].
In the first part of this chapter, different topologies for full-adders are presented along with some circuit modifications to reduce the power consumption. Due to the importance of leakage power consumption in digital CMOS circuits, new technique is proposed to reduce

the sub-threshold leakage current in high fan-in gates (e.g. OR gate) and basic blocks such as comparators and multiplexers for high-performance applications. Two circuit techniques are proposed to reduce the leakage current in nano-scale circuits as follows:

1. Leakage-tolerant circuit for domino logic
2. Adaptive supply voltage technique

Let us first look at different topologies for full-adders.

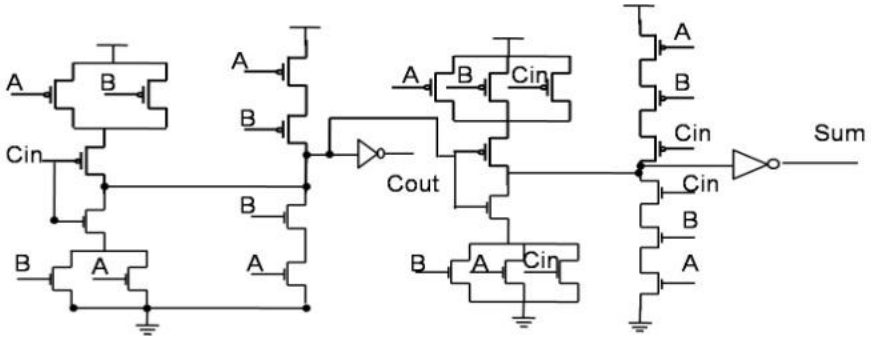


Fig.2.1. CMOS standard 28T full adder

Table.2.1 Truth table of SERF full adder design

A	B	Cin	Sum	Cout
0	0	0	0	0
0	0	1	Vdd-Vth	≈0
0	1	0	1	Vtp
0	1	1	0	1
1	0	0	1	>Vtp
1	0	1	0	1
1	1	0	Failure	Vdd -2Vth
1	1	1	>Vdd-2Vth	Vdd-Vth

2.1. Full Adder Topologies

One-Bit Full adder design is one of the most critical components of a processor that determines its throughput, as it is used in ALU, the floating point unit, and address generation in case of cache or memory accesses [15]. A variety of full adders have been reported in [16]-[24]. However, in this chapter, we focus on the full adder topologies specified for subthreshold design and wireless sensor networks. One of the most well known full adders is the standard CMOS full adder that uses 28 transistors as shown in Fig.2.1. In [16] the sense energy recovery full adder (SERF) is presented. The topology of this circuit is shown in Fig.2.2 which requires only 10 transistors to implement a full adder.

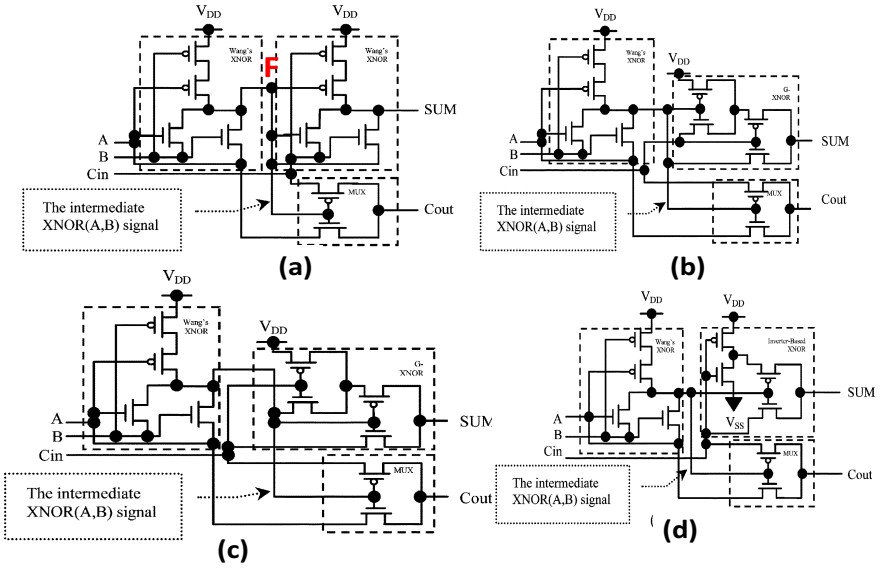


Fig.2.2. Different topologies for SERF full adder [17]

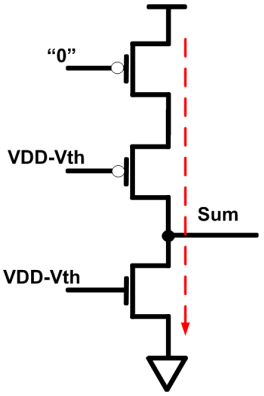


Fig.2.3. Equivalent circuit in input vector ABC_{in}="110"

Several full adder topologies have been presented in [18] with a low number of transistors (i.e. small area). However, in this chapter, we focus on two topologies including GDI and SERF full adders.

Let us, first, explore several full adder topologies based on the GDI technique. Then, different circuit topologies based on the SERF full adder are presented for ultra low supply voltage applications. The multi threshold technique is used to improve the operation of the SERF full adder design.

The SERF design uses only 10 transistors to implement a full adder. Although his circuit operates properly at higher supply voltages, if the supply voltage is scaled further to voltages lower than 0.3V, this circuit fails to work. Table.2.1 describes the behavior of this circuit for different inputs. As it can be seen, the SERF adder (Fig.2.2 (a)) is confronted with serious problems especially at lower supply voltages. Assume that one of the two

input vectors ABC_{in}="110" and "111" are applied. As seen from Fig.2.2 (a), when A=1 and B=1, the voltage at node F is at V_{DD}-V_{th}. Now if C_{in}=0 then C_{out} will be equal to V_{DD}-$2V_{th}$ and the Sum signal is discharged to zero driven by a MOS transistor with its gate connected to V_{DD}-V_{th}. When C_{in}=1, C_{out} is connected to V_{DD} (or lower) and the signal SUM raises to V_{DD}-V_{th}. Another problem with this design is the time in which the floating node is connected to 0 (A=0, B=1 or A=1, B=0).When C_{in} is "1", C_{out} is charged to V_{DD}, but when C_{in}=0, C_{out} must be discharged to ground using a PMOS pass transistor that cannot fully discharge the output. In this case, C_{out} is discharged to V_{tp} which is higher than V_{tn} (Where V_{tp} and V_{tn} are the threshold voltages of PMOS and NMOS transistors, respectively). This problem is intensified, when the circuit operates at subthreshold voltage. Let say A is at logic "1", current leaks to the C_{out} node which makes C_{out} to increase even more than V_{tp} in some cases depending on the sizing of the pass transistors. In this case the Sum voltage depends on the C_{in} state. For instance, if C_{in} is "1", the Sum output is charged to V_{DD}-V_{th} which is a problem in subthreshold region by assuming V_{DD}=0.3V while V_{th}=0.25V. In this situation, output is charged to 0.05V that is assumed as zero.

The most important problem with the SERF full adder is in the case when A=1, B=1 and C_{in}=0. For this input vector, as mentioned before the output signal reaches V_{DD}-V_{th}. Simulation results show that at V_{DD}=0.3V, the output signal reaches 0.1V which is not high enough to change the state of the next stage. To eliminate these problems a new topology is introduced. This limitation also causes a constraint for lowering the supply voltage. For instance, to have a correct output for SUM it seems that the supply voltage cannot be lowered less than V_{DD}/2+2Vtn indicating that the supply voltage must be higher than V_{DD}/2+0.28V in a 65nm CMOS technology. However this limit depends on the circuit design topology and also the sizing and the device types that are employed. To mitigate this problem, the gates of pass-transistor (PT) for C_{out} signal must be connected to V_{DD} during the challenging state (A=B=1, C_{in}=0). Then the supply voltage may be reduced to as low as V_{DD}/2+V_{th} which is estimated to be V_{DD}/2+0.14. For example when V_{DD}=0.3, in worst case C_{out} will be V_{DD}-V_{th}=0.16V, which can be used as a high logic. In addition the NMOS pass transistor may be upsized to further lower the supply voltage. It seems to be possible to lower the supply voltage to 0.25V.

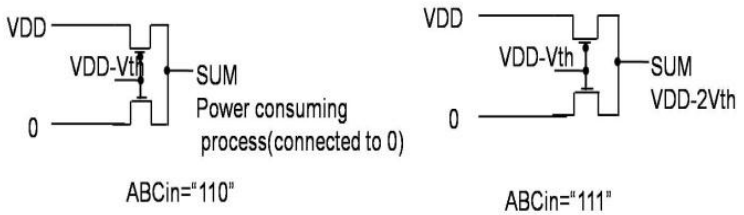


Fig.2.4. Equivalent circuit for related inputs (SUM)

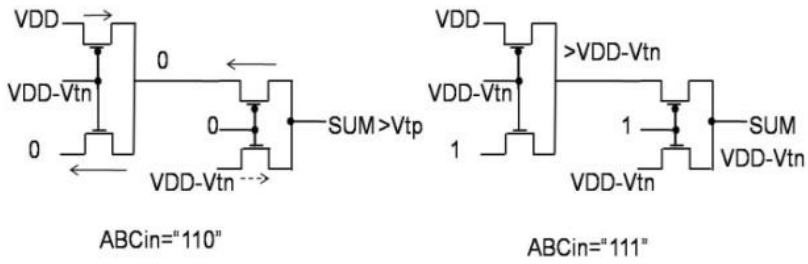


Fig.2.5. SERF behavior

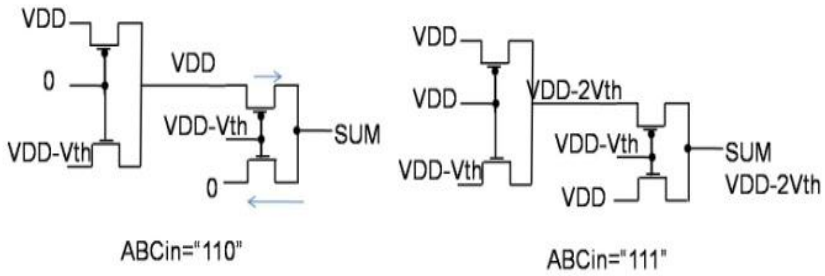


Fig.2.6. SERF behavior Analysis

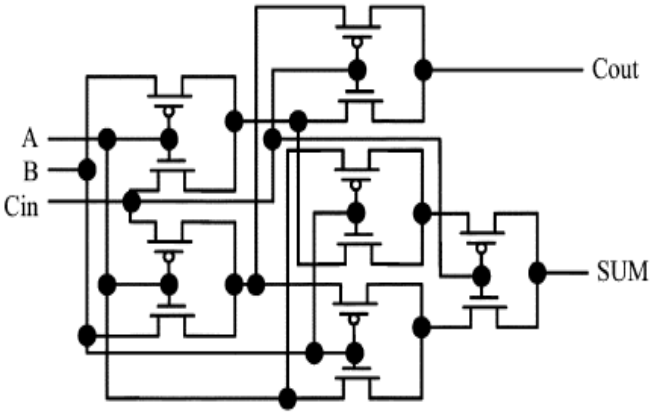


Fig.2.7. MUX-based SERF adder [17]

For input signals A=1, B=1, and C_{in}=0, the equivalent circuit for SUM signal is shown in Fig.2.3. As it can be seen the state of the output cannot be determined precisely, since two PMOS devices and also the NMOS transistor are ON, then the output state is roughly dependant on the transistor sizing. Consequently, the circuit fails to evaluate correctly in some cases. Let us consider other SERF full adders in Fig.2.4 for input vector ABC_{in}="11X". As it can be seen in Fig.2.4, this topology consumes much power with the input vector ABC_{in}="110". At ultra low supply voltage the probability of failure is higher than when operating at higher supply voltages. Fig.2.5 shows the behavior of the topology in Fig.2.2 (c) during these two input vectors (ABC_{in}="110" and ABC_{in}="111").

In this topology the behavior of the circuit is improved compared with Fig.2.2 (b), so we can reduce the supply voltage as much as V_{th} compared to the circuit shown in Fig.2.2 (b). In the topology of Fig.2.2 (d), there is the same problem for the SUM output signal. Fig.2.6 describes the behavior of this circuit. In Fig.2.6, if we add a circuitry, to connect the gates of the pass transistors to V_{DD} instead of V_{DD}-V_{th} leads to removing the power dissipating path due to the completely turned-off PMOS transistors, results the output connects to V_{DD}-V_{th} instead of V_{DD}-2V_{th}.

There are two possible solutions to improve the operation of the SERF circuit. The first one is connecting the output of the XNOR gate (first stage) to V_{DD} instead of V_{DD}-V_{th} (when the output of XNOR gate is high) and also modifying C_{out} to be connected to V_{DD} instead of V_{DD}-V_{th} or V_{DD}-2V_{th}. The multiplexer-based SERF adder circuit is presented in [17]. In this circuit which is shown in Fig.2.7, all parts are implemented using a multiplexer. Although this circuit consumes less power than previous SERF topologies, it

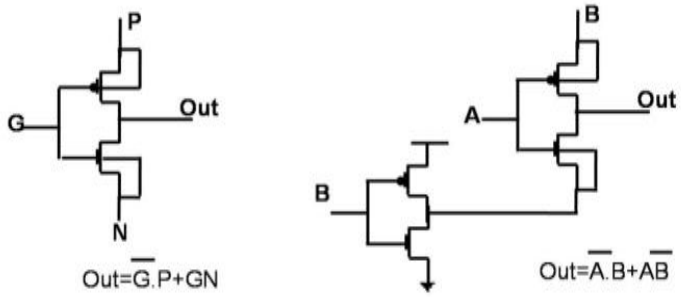


Fig.2.8 (a) GDI technique (b) XOR gate using GDI

Table.2.2. Truth table of MUX-based SERF

A	B	Cin	Sum	Cout
0	0	0	Vtp	0
0	0	1	Vdd-Vth	Vtp
0	1	0	Vdd-Vtn	Vtp
0	1	1	0	Vdd
1	0	0	Vdd	Vtp
1	0	1	Vtp	Vdd-2Vtn
1	1	0	Vtp	Vdd-Vth
1	1	1	Vdd-2Vth	Vdd-2Vth

has some serious problems at lower supply voltages. Table.2.2 shows the truth table of this circuit.

For the input vector "ABC_{in}"="111", the output signals are limited to V_{DD}-$2V_{tn}$ which limits the supply voltage scaling for this circuit. The supply voltage for this circuit must be higher than $2V_{tn}+V_{tp}$. Therefore, in 65nm technology, the supply voltage must be higher than 0.48 which is not sufficiently low for ultra low power applications. Therefore this circuit does not work properly in subthreshold region. Furthermore for input vector "101", C_{out} is limited to V_{DD}-$2V_{th}$.

In this section we analyze the Gate-Diffusion Input (GDI) full adder design that is proposed in [20]. The GDI technique is proposed by Morgenshtein et al. in [10]. This technique reduces the power dissipation and also makes the circuit smaller. The advantage of GDI technique two-transistor implementation of complex logic functions, and in-cell swing restoration under certain operating conditions, are unique within existing low-power design techniques [10]. Fig.2.8 shows the operation of GDI technique and also implementation of a XOR gate using this technique. As it can be seen most logic functions required can be implemented using a small number of devices. To implement a XNOR gate using GDI, the places for applying B and B_bar can be changed. The problem with GDI-based adder is the same as for the SERF adder. Fig.2.9 shows four topologies of GDI-based full adder. We analyze these full adders with different inputs. Table.2.3 shows the results for these full adders.

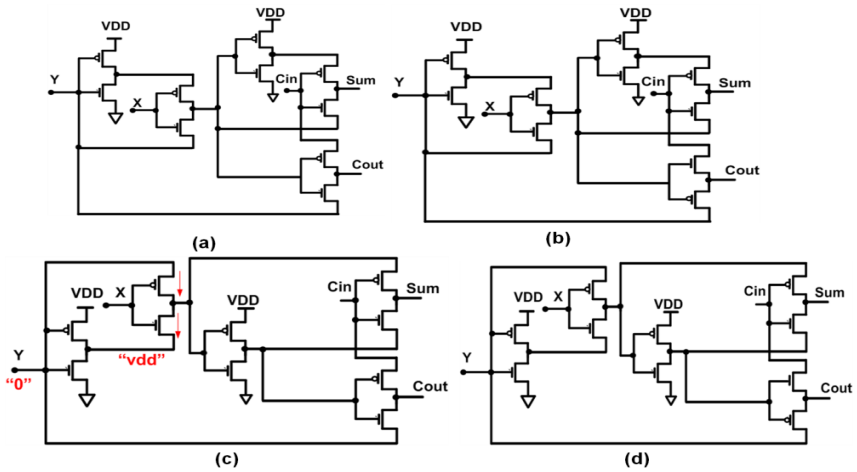


Fig.2.9. GDI XOR-based (a,b), and XNOR-based Full adder (c,d)

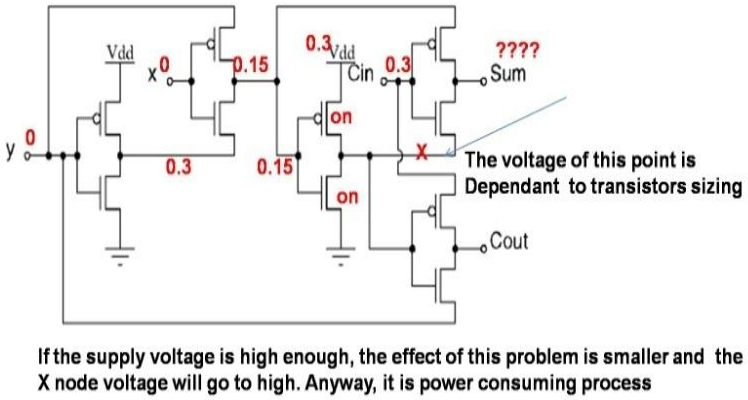


Fig.2.10. Failure in GDI full adder

These GDI-based full adders work better than SERF in most cases. However, there are some limitations. For instance, for the input vector ABC_{in}="001", suppose that the circuit is working with V_{DD}=0.3V and the threshold voltage for PMOS and NMOS circuit are -0.15 and 0.14 respectively. In this case, Fig.2.10 illustrates the problem which is even more degradation at lower supply voltages for the circuit from Fig.2.9(c). The SUM signal value is not stable, because of insistent power dissipation and contention between NMOS and PMOS to determine the X node state. Also in Fig.2.9(a), there are problems with discharging the SUM node to zero. The SUM output is discharged only to $2V_{tp}$ which is high enough to be assumed as "1". However, this circuit dissipates less power than SERF full adders. We can modify the SERF full adder to work at lower supply voltages and alleviate the problem with these special input vectors for the GDI-based adder design. For the GDI full adder, we can reduce the supply voltage lower than supply voltage for the SERF circuit.

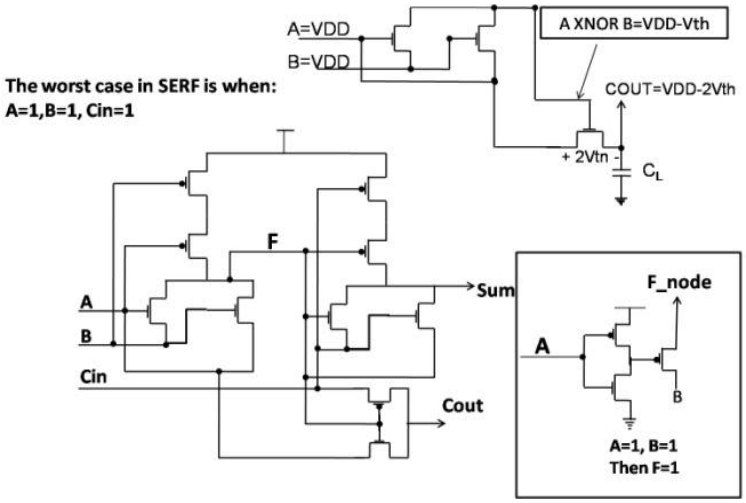


Fig.2.11. Proposed SERF full adder circuit

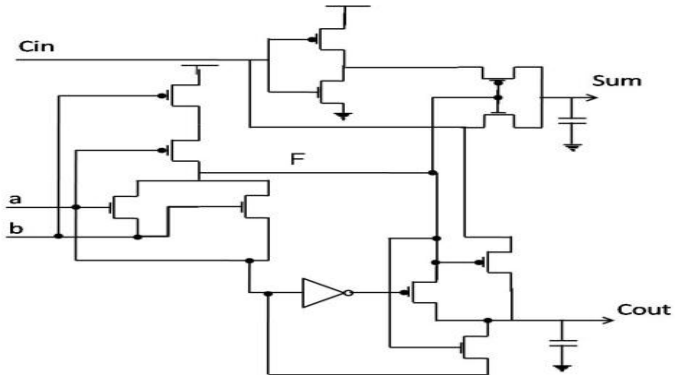


Fig.2.12. Modified SERF full adder (based on Fig.2.2 (d))

Table.2.3. GDI Full-adder truth table

A	B	Cin	Sum (Fig.2.9.a)	Sum (Fig.2.9.c)	Cout (Fig.2.9.a)	Cout (Fig.2.9.c)
0	0	0	Vtp	0	>2Vtp	2Vtp
0	0	1	Vdd-Vtn	Vdd-Vtn	0	0
0	1	0	Vdd	Vdd	>Vtp	Vtp
0	1	1	Vtp	0	Vdd-Vtn	Vdd
1	0	0	Vdd	Vdd-Vtn	Vtp	Vtp
1	0	1	0	0	Vdd	Vdd
1	1	0	Vtp	Vtp	>Vdd-Vn	Vdd-Vtn
1	1	1	Vdd-2Vtn	Vdd-2Vtn	Vdd-Vn	Vdd-Vtn

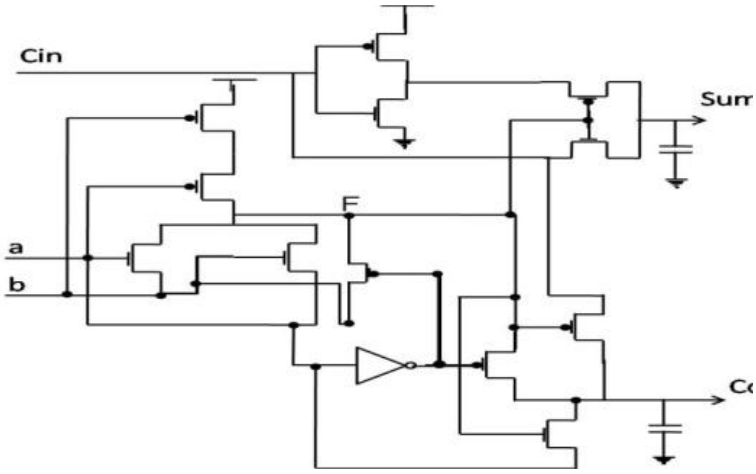


Fig.2.13. Proposed SERF full adder

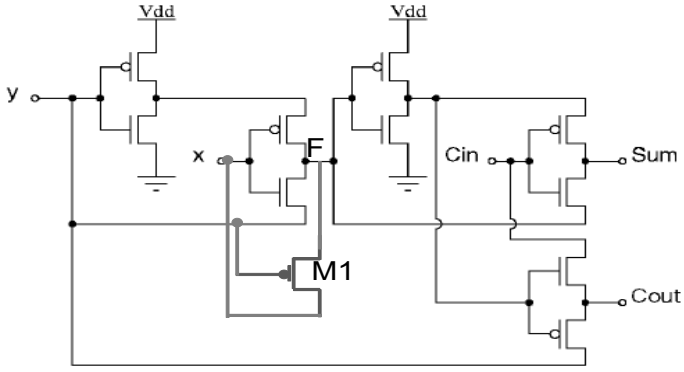


Fig.2.14. Modified GDI full adder (GDI#1)

Table.2.4. Truth table for full adder circuit Fig.2.11

A	B	Cin	Sum	Cout
0	0	0	0	0
0	0	1	Vdd-Vth	0
0	1	0	1	Vtp
0	1	1	0	1
1	0	0	1	<Vtp
1	0	1	Vtp	1
1	1	0	0 (Power consuming)	>Vdd -Vth
1	1	1	>Vdd-2Vth	Vdd-Vth

2.2. Proposed Full Adders (FA)

In this section some new full adders are proposed based on the SERF and GDI techniques. At first to modify the SERF full adder for different inputs, we add an extra circuit to the SERF adder as shown in Fig.2.11. By adding this circuit to the SERF, the F node voltage for input vectors ABC_{in}=(”110”, “111”), is connected to V_{DD}, which increases the output by V_{th}, so we can scale the supply voltage to $V_{tn}+V_{tp}$ that is estimated to be lower than 0.3V instead of 0.45V for the SERF full adder.

Another proposed idea is as follows: We use a different configuration for the C_{out} signal and a MUX to produce the SUM signal of the output of XNOR gate. This topology is shown in Fig.2.12. For this circuit Table.2.4 shows that that the logic levels of Cout are improved and the consuming path in SERF is mitigated. However this circuit uses one transistor more than the SERF full adder.

Another idea that significantly improves the operation of the SERF design is using a compound of the techniques proposed in Fig.2.11 and Fig.2.12. This proposed circuit is shown in Fig.2.13. This circuit enables even more scaling of the supply voltage lower than

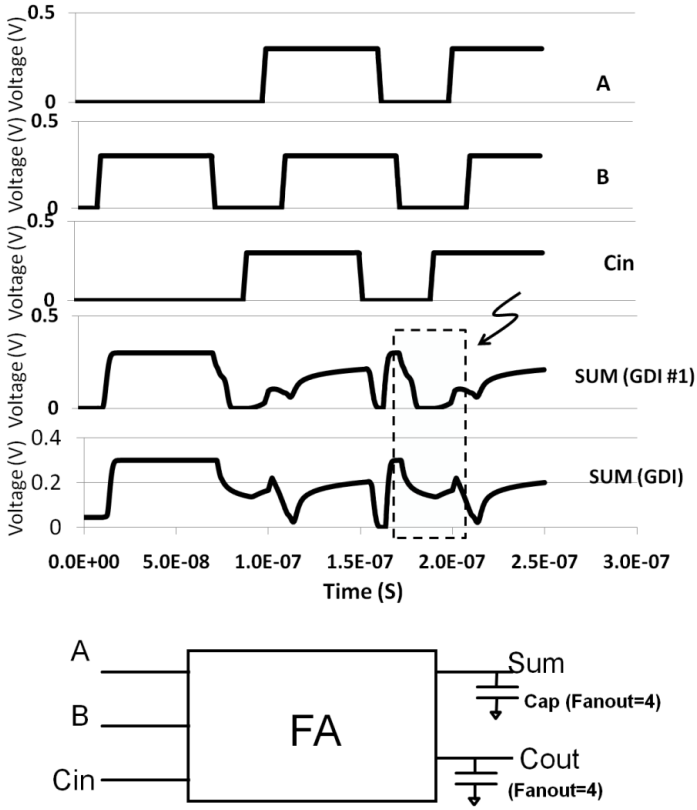


Fig.2.15. Output waveforms of GDI#1 and test bench

$V_{tn}+V_{tp}$ which is estimated to be 0.3V. Furthermore, we may use the precise sizing to enable the circuit to work at lower supply voltages. This circuit shows much better functionality compared with other SERF adder topologies. In this circuit, when A=1, B=1,

Table.2.5. Power consumption of full adders

Full Adder	Power (nW)	Total Width (L=Lmin)	VDD_{min}
Conv. SERF	61.12	6.1	0.48
SERF (Fig.2.11)	25.5	6.2	0.35
SERF (Fig.2.12)	41	6.1	0.4
SERF (Fig.2.13)	17.97	6.8	0.3
GDI #1	35.6	6.45	0.45
GDI#2	19.8	6.65	0.35
Conventional GDI	56.3	6.05	0.45

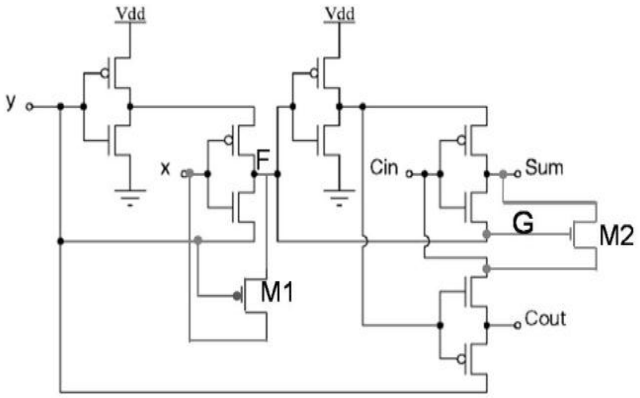


Fig.2.16. GDI #2 full adder circuit

C_{in}=0, there are no problems; because the F node is connected to "1", which eliminates the power consuming path in Fig.2.14. Because the gates of the pass transistors are connected to V_{DD}, the PMOS transistor is off. Moreover, in the state when ABC_{in}="111", the F node is connected to V_{DD}, and as a result, Sum is charged to V_{DD}-V_{th} and results C_{out} connecting to V_{DD}.

Another circuit is a new modified GDI based full adder design. The technique is used to improve the functionality of circuit shown in Fig.2.11 for lower supply voltages. The main drawback with the GDI full adder is when both x and y inputs are at logic low (AB="00"). As it can be seen the main problem is with the output of first stage XNOR gate in the GDI full adder. If we add a circuit to lower the output of first stage to logic zero instead of V_{tp}, the problem is resolved. Furthermore, the original circuit dissipates more power during this state which is now reduced using proposed circuit. The proposed GDI full adder is shown in Fig.2.14. The added circuit is shown in bold lines in Fig.2.14.

When y and x signals are "0", this turns on the added NMOS transistor and helps to discharge F node voltage to zero. Therefore this alleviates the problem with this circuit described in Fig2.10. Fig.2.15 shows the waveforms for proposed FA compared to conventional FA in Fig.2.10. We can add another transistor to the GDI full adder that is

Table.2.6 Simulation results (Delay and PDP)

Full Adder	Delay (ps)	PDP ($\times 10^{-18}$)
CMOS (28T)	900	55.8
Conv. SERF	150	38.94
SERF (Fig.2.11)	320	42.88
SERF (Fig.2.12)	450	78.12
SERF (Fig.2.13)	240	20.88
Hybrid SERF (Fig.2.13 and 2.14)	150	18.6
GDI #1	880	121
GDI#2	850	89.23
Conventional GDI	850	270

shown in Fig.2.16. In this topology, NMOS transistor M2 is added to connect the SUM output signal to ground when G node voltage is high. This added NMOS transistor M2 enables this circuit to work at lower supply voltages.

The proposed full-adders modify the operation of GDI and SERF full adders at very low supply voltages. Therefore, we consider new GDI full-adders as GDI#1 and GDI#2 and for SERF full adders by referring to the corresponding figures.

2.2.1. Simulations and results

We simulated FA circuits for different ranges of frequencies from 100 KHz to 10 MHz to find out which circuit performs well. We use the 65nm CMOS standard models from ST microelectronic. Also we simulated these circuits to find the lowest supply voltages that these circuits are able to work without failure. The results for all full adders are tabulated in Table.2.5 to show which full adder is best suited for operation in ultra low supply voltage applications.

For SERF full adder, the most important drawback is its limitation on supply voltage scaling that which cannot be reduced below $2V_{tn}+V_{tp}$. This problem is more challenging in corners especially in SS corner (Slow NMOS, Slow PMOS).

The supply voltage at which SERF outputs are satisfactory is higher than 0.5V with upsized transistors. The simulation results for different full adder circuits are shown in Table.2.5.

As it can be seen in Table.2.5, the power consumption for SERF (Fig.2.13) full adder is lower than other circuits. The supply voltage for this design is limited to voltages higher than 0.3V. For these designs we can use the lower supply voltages in some cases, but we have to use larger devices that increase the area significantly. To find the minimum supply voltages for full adder designs, different inputs were applied at different frequency. We considered the functionality of these circuits in different operating conditions. For instance, as it is shown in Fig.2.15, the dotted region shows the failure in SUM signal for GDI full adder at 0.3V supply voltage. In this case, for input vector ABC_{in}=”000”, SUM should be at zero logic. But it is discharged just to 0.16V, which can be assumed as high voltage causing a failure in circuit. As Table.2.5 shows, for proposed full adder topologies, power consumption is decreased significantly compared to conventional SERF and GDI full adder counterparts. The mail reason of lower power consumption is using the lower supply voltages that cause reduction in dynamic power (quadratic) and also subthreshold

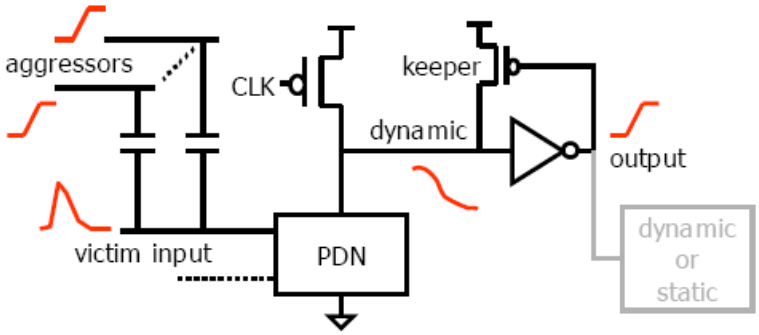


Fig.2.17. Te main sources of noise in domino logic circuit [26]

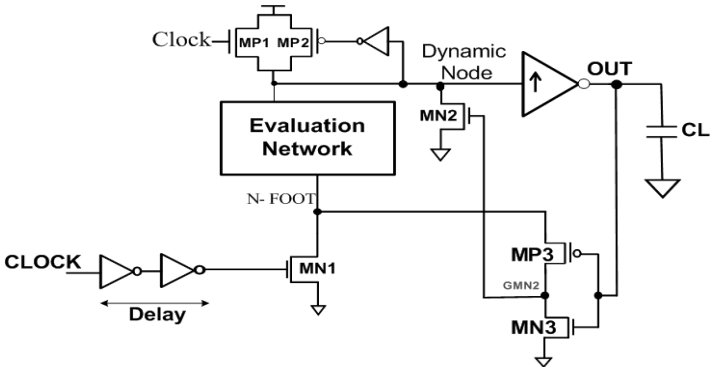


Fig.2.18. Proposed domino logic circuit-1 [30]

power consumption (exponential). Table.2.6 shows the results for delay comparison for proposed circuits compared to other topologies in literature. As it is shown, the proposed full adders improve the PDP in some cases by 2X times. For new GDI full adder designs, in proposed circuits, the PDP is improved compared to conventional GDI full adder circuit [10].

According to transistor level simulations, the power consumption is decreased with at least 62% for the SERF design and 86% for the GDI full adder design. The cost is a small area overhead; the proposed circuits have a small area overhead up to 11% compared with SERF and GDI full adders. Also the GDI technique showed that this logic can be suitable for ultra low power applications.

In next section, due to the importance of leakage current mechanisms in ultra deep sub-micron technologies, a new leakage tolerant logic circuit is proposed. Therefore, let us first look at different sources of noises causing lower robustness in logic circuit design. As we mentioned in chapter one, for ultra low voltage applications, I_n/I_{off} ratio decreases significantly. Therefore, by applying some techniques to reduce leakage current, let say by 10 times, I_{on}/I_{off} increases by 10 times, results in improved performance, power, and robust operation.

2.3. Leakage-Tolerant Logic Circuit Design

High fan-in compact dynamic gates are often used in high performance critical units of microprocessors. However, the use of wide dynamic gates is strongly affected by subthreshold leakage and noise sources [25]. This is mainly due to decreased threshold voltage that results in exponentially increased leakage currents in scaled technologies. To reduce power consumption, supply voltage scaling is used across technology scaling. However, threshold voltage needs to be scaled down as well to maintain transistor overdrive for large ON currents. Less threshold voltage means smaller gate switching trip point in domino circuits. Smaller trip points make the domino circuit more prone to input noise. Moreover, excessive leakage can discharge the precharge (dynamic) node of a domino circuit resulting in a logic failure (wrong evaluation). In addition to reduced trip pint and increased leakage, other noise sources such as supply noise and crosstalk noise also increase by technology scaling, further degrading the robustness of domino logic. Fig.2.17 shows different noise sources and how they impact the robustness of the domino logic [26]. Conventional approach for improving the robustness of domino circuits is keeper transistor upsizing. However, as the keeper transistor is upsized, the contention between keeper transistor and NMOS evaluation network increases in the evaluation phase. Such current contention increases evaluation delay of the circuit and increases power dissipation. Thus, keeper upsizing trades off delay and power to improve noise and leakage immunity. Such trade-off is not acceptable because it may make the circuit too slow or too power hungry.

There are techniques proposed in the literature to address this issue. High-speed domino logic [27] and conditional keeper [28] are among the most effective solutions for improving the robustness of domino logic. Fig.2.17. the main sources of noise in domino logic circuit [26]. In this section, we propose new domino circuits for high fan-in and high-speed applications in ultra deep submicron technologies. The proposed circuits employ a footer transistor that is initially OFF in the evaluation phase to reduce leakage and then turned ON to complete the evaluation. In order to avoid the delay penalty due to an initially OFF footer transistor, an extra path for evaluation is provided that is controlled by the output. According to simulations in a predictive 70nm process [29], the proposed circuits increase noise immunity by more than 26X for wide OR gates and shows performance improvement of up to 20% compared to conventional domino logic circuits. The proposed circuits reduce the contention between keeper transistor and NMOS evaluation transistors at the beginning of evaluation phase.

2.3.1. Proposed Domino Logic Circuit Design Using Feedback from Output

There are many proposed circuits that reduce leakage current and total power consumption. One of the existing leakage tolerant domino circuits is high-speed domino logic (HS). The description of this circuit has been explained in [27]. Another existing leakage tolerant domino circuit is conditional keeper domino logic (CKL) [28]. In our design, a conditional keeper that is turned on during the hold time is employed to dominate the leakage through parallel NMOS network. The schematic of our proposed circuit is shown in Fig.2.18 (referred as proposed circuit-1). The proposed circuit-1 employs stacking effect (by adding the footer transistor MN1) for noise immunity improvement and uses the steady state voltage of N_FOOT node at the beginning of evaluation phase to reduce leakage of the evaluation network. Below the operation of the circuit is analyzed for the different operational modes.

a) Precharge mode:

When clock is low, the circuit is in the precharge phase. MP1 is turned on and the dynamic node starts charging to V_{DD}. In addition, PMOS keeper transistor (MP2) is turned on helping the precharge. At the beginning of the precharge phase, MN1 is on. Thus, it pulls the N_FOOT node to ground. Meanwhile, node GMN2 is low and MN2 is in the off state. After the delay of the inverters (delay element), MN1 is turned off. In this case, the voltage of N_FOOT rises to an intermediate voltage level. The evaluation transistors are sized such that the DC voltage of GMN2 node does not exceed the threshold voltage of MN2 to avoid any possibility of short circuit current in the precharge phase. We have selected MN2 to be larger than other NMOS transistors.

b) All inputs at zero in evaluation

At the beginning of the evaluation phase, NMOS footer transistor MN1 is off. Thus, N-FOOT node is floating. Therefore, in this case, its voltage reaches a DC voltage. If this voltage exceeds $|V_{tn\text{-}MP3}|+V_{OUT}$, MP3 is turned on. In the other words:

$$if \quad V_{N-FOOT} \geq V_{tp-MP3} + V_{OUT} \quad MP3:ON \qquad (1)$$

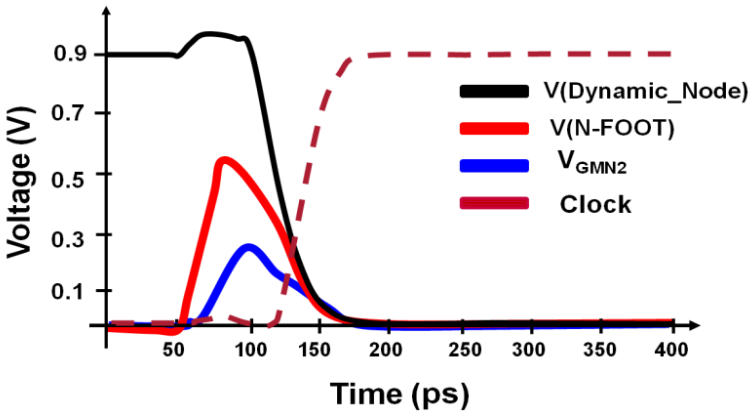


Fig.2.19. Waveforms for proposed Circuit-1

In that case, the GMN2 node is charged to $V_{N\text{-}FOOT}$, and therefore:

$$if \quad V_{GMN2} \geq V_{tn-MN2} \quad MN2:ON \qquad (2)$$

If condition (2) is satisfied, (MN2 is turned on), a wrong evolution occurs. However, in our design we have sized MN1, MN2, MP3, and MP4 considering the voltage of GMN2. The sidings are done in such a way that condition (1) and (2) do not happen. Therefore, the DC voltage of N_FOOT acts as a source biasing for the evaluation network without affecting the functionality of the circuit. This DC voltage reduces leakage of the evaluation network substantially, resulting in significant leakage tolerance. Our proposed circuit has significant immunity to input noise because due to the DC voltage of NMOS transistors source terminals in the evaluation network, their threshold voltage increases. Thus, their trip point increases and their subthreshold leakage current reduce significantly, due to

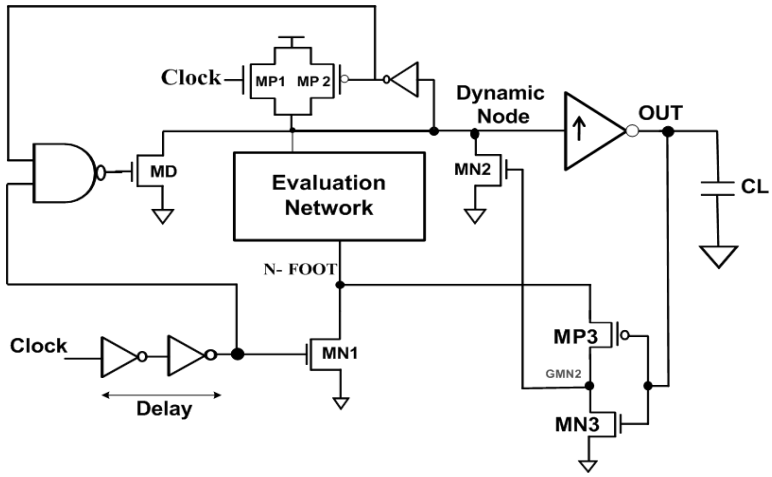


Fig.2.20. High speed proposed domino logic circuit-2

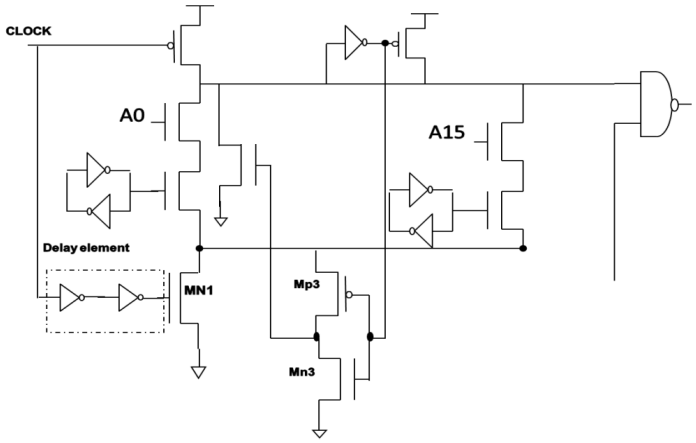


Fig.2.21. Proposed MUX-1

stacking effect. In our proposed circuit, performance improvement is achieved by upsizing MN2 transistor. This is further described in the following subsection.

c) An input switching high in evaluation phase:

The waveforms of the circuit in this mode are shown in Fig.2.19. As observed, the increased voltage of N_FOOT node at the beginning of the evaluation phase causes MP3 to be turned on. Therefore, the GMN2 node is charged to the voltage of N_FOOT node which rises above the threshold voltage of MN2. Therefore, the NMOS transistor MN2 is turned on at the onset of evaluation phase (when the footer transistor MN1 is off), connecting the dynamic node to ground. After delay of the delay element, N_FOOT node is strongly at zero voltage. Thus, the transistor MP3 switches to the off state. Since the output nod is at high now, it turns on the MN3, and connects GMN2 node to ground turning MN2 off.

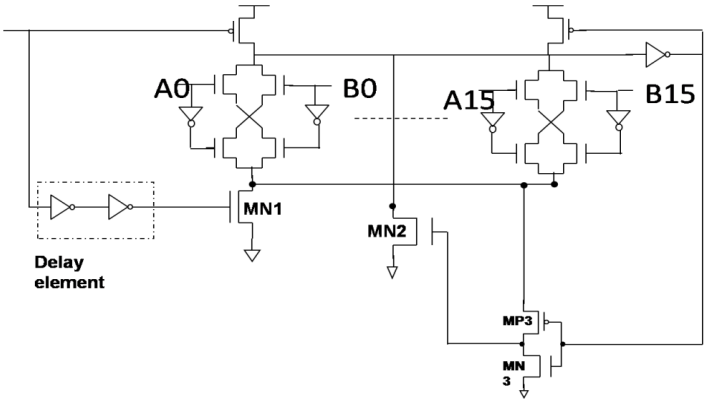


Fig.2.22. Proposed comparator-1

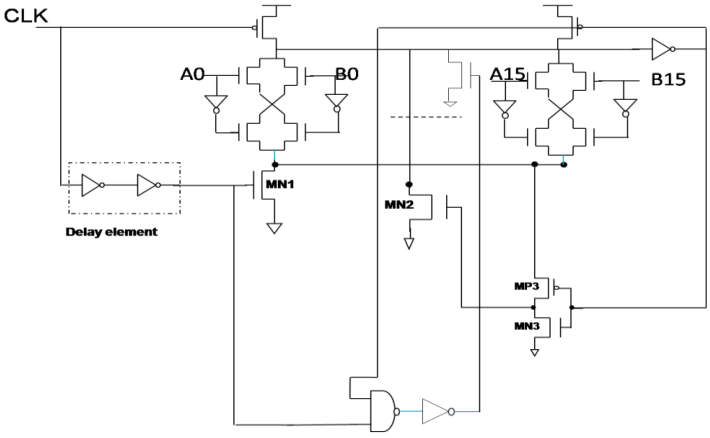


Fig.2.23. Proposed Comparator-2

However, the rest of evaluation phase (discharging of the dynamic node) completes through the evaluation network and the footer transistor that is fully on. Here we have more degree of freedom for increasing speed or enhancing noise immunity. For example, for improving speed, upsizing of MP3, MN2, MN1, evaluation transistors, and MN1 are all options.

To improve the operation of the proposed circuit-1 we use an extra circuit to improve the evaluation speed. This circuit is shown in Fig.2.20 (referred as proposed circuit-2). After the primary time of evaluation, the dynamic node starts to be discharged (when at least one of inputs is at high). The input of small keeper transistor MP2 starts to go high. So, both inputs of the NAND gate are at high logic, so the output of the NAND gate goes to low and the gate of NMOS MD transistor start to going high that helps to discharging dynamic node. Then the speed of our proposed circuit is improved significantly. In other times at least one of the inputs of the NAND gate is 0, so the voltage of the MD transistor is low. Therefore the MD transistor is switched off.

To show the efficacy of proposed domino logic circuit, a multiplexer and a comparator designs based on the proposed domino logic design are presented.

2.3.1.1. Proposed Multiplexer

High fan-in dynamic MUXs are commonly used in register files for implementation of bit-lines [31]. In register files, because of the fairly small size of the memory, the bit-lines are implemented using wide domino MUX gates [32]. Multiplexers with a high fan in are widely used in many applications, such as the column decoders of memories. In footless domino logic MUX, the excessive leakage of the evaluation network can cause logic failure during the read operation.

A method is proposed in [31] to improve the leakage-immunity of register file bit-lines is pseudo static bit-line. In this technique, the subthreshold leakage is reduced considerably. However, the technique either exhibits considerable increase in transistor count and delay penalty due to the use of many static OR gates. The proposed technique can be applied to the register file bit-line MUX as shown in the Fig.2.21. The worst case scenario for noise at the inputs is when all the inputs from the memory cells are high, and all the RS signals are low and receive same noise in the evaluation phase. In the FLDL MUX, the keeper transistor is upsized from a Keeper ratio of 1 to 2 in order to achieve different data points for delay and noise immunity. The FLDL MUX fails to operate for smaller keeper sizes because of high subthreshold leakage in the 70-nm technology. In the pseudo static MUX, the keeper transistor is upsized from a keeper ratio of 0.1 to 1 in order to achieve different data points for delay and noise immunity. In the proposed MUX, the keeper transistor is sized for a keeper ratio of 0.1 to 0.3 to achieve different data points for delay and noise immunity. Proposed MUX-2 is designed based on the configuration in Fig.2.18. To get the different data points, we increase the keeper ratio from 0.1 to 0.3. To improve the noise immunity of this circuit, we can use the minimum size devices for the NAND gate in proposed circui-2.

2.3.1.2. Proposed High Fan-in Comparators

The worst case for delay is when inputs A and B are different in only a single-bit position. In this case, only one of the evaluation paths conducts and discharges the precharge node. The worst case scenario for noise at the inputs is the case where all the inputs are low and receive the same noise in the evaluation phase. In the standard domino comparator, the keeper transistor is upsized from a keeper ratio of 1 to 2 in order to achieve different data points for delay and noise immunity. The standard domino comparator fails to operate for smaller keeper sizes because of high leakage in scaled technologies (70 nm). In the proposed circuit that is shown in Fig.2.22, the keeper size variations are very small. We changed the keeper size from minimum size to 0.15 to achieve data points. In fact, in proposed circuit the speed is almost independent of keeper size but it is due to Mn2 sizing. We can achieve even more speed by increasing the size of NMOS transistor Mn2. UNG is defined as the amplitude of input noise V_{noise} that causes an equal amplitude noise pulse at V_{out}.

Our simulation results are obtained in the worst case corner of the 70-nm predictive technology at 0.9 V and 110°C. Meanwhile the proposed circuit has a better performance and also UNG compared with Diode-Footed domino logic [32]. To configure the proposed comparator-2, a configuration based on Fig.2.20 is designed (referred as proposed comparator-2) that is shown in Fig.2.23.

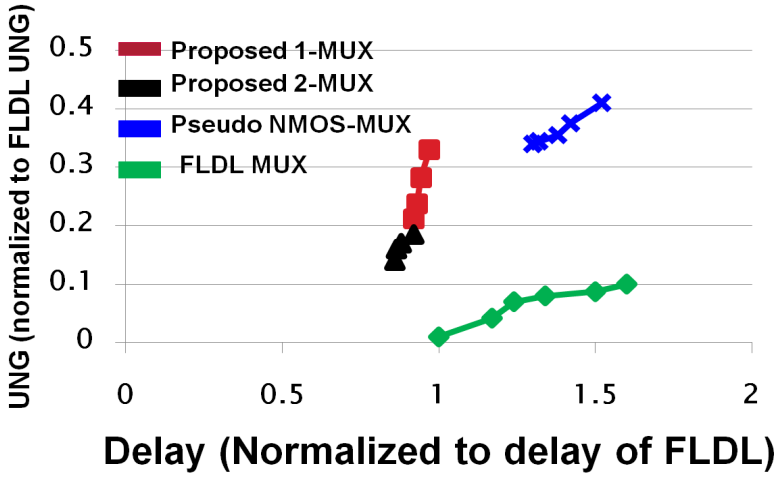


Fig.2.24. UNG vs. Delay for proposed MUX's

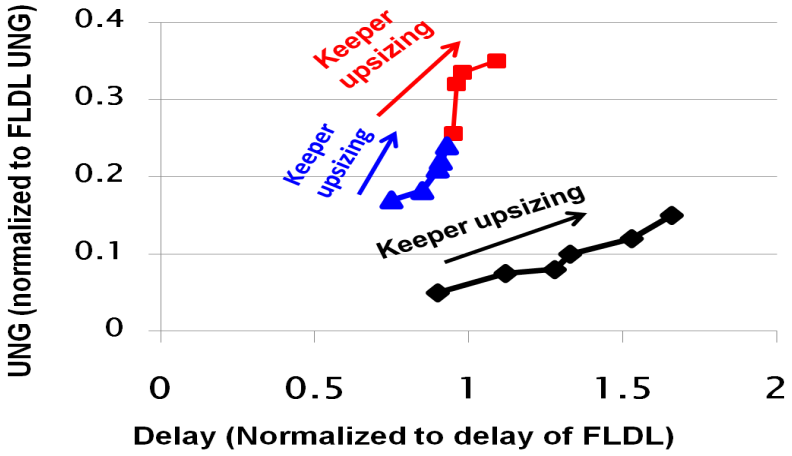
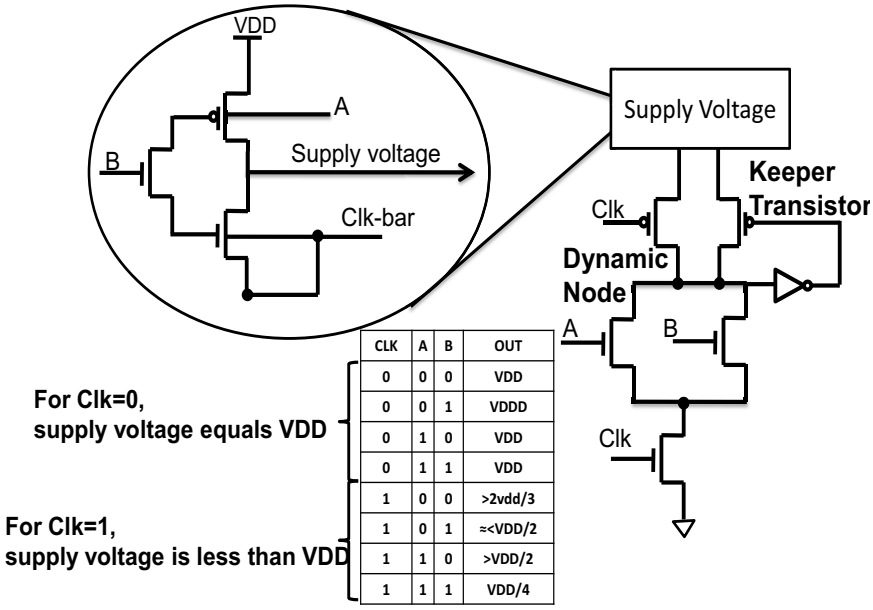


Fig.2.25. UNG versus Delay for proposed comparators

In this section, we study the behavior of our proposed circuits based on simulation results. The results are obtained using predictive technology model of 70nm technology at the temperature of 110C. The noise immunity metric used in our work is unity noise gain (UNG) [32]. UNG is the amount of DC noise at all inputs that result in the same amount of noise at the output noise [32-34]. Therefore, larger UNG indicate more noise (leakage) immunity. UNG of our proposed circuits is obtained by varying keeper transistor size from $0.3W_{EVAL}$ to $1W_{EVAL}$ (WEVAL being the width of evaluation transistors) for both circuits. The keeper ratio is defined as the size of keeper transistor to the size of evaluation transistors. The proposed circuit-1 shows very high UNG compared to other proposed circuits and higher speed than some existing designs [33-42]. Therefore, the proposed circuit-1 has a higher performance and very high UNG over conventional domino circuits. Results show that the improvement of UNG for our proposed circuits compared to

conventional circuits is as large as 26X. In addition, speed of our proposed circuit is acceptable and it shows 20% improvement for some cases. In proposed circuit-2 the performance is increased significantly due the added configuration that improves discharging the dynamic node during the evaluation phase (when at least one of the input signals is high). But the UNG of this circuit is lower than proposed circuit-1. In the evaluation mode with all inputs zero, we observed that the subthreshold leakage current has reduced significantly in our circuits. In our proposed circuits, by sizing transistors precisely, we can get less power dissipation compared to conventional circuits.



CLK	A	B	OUT
0	0	0	VDD
0	0	1	VDDD
0	1	0	VDD
0	1	1	VDD
1	0	0	>2vdd/3
1	0	1	≈<VDD/2
1	1	0	>VDD/2
1	1	1	VDD/4

Fig.2.26. Proposed domino logic circuit

Simulation results show that the power dissipation of our proposed circuits is lower compared to other domino logic circuits. The proposed circuits have employed small devices for the evaluation network, and therefore, the areas of our proposed circuits are less comparing with conventional circuits [30].

In summary, according to the simulation results, the propose circuits show 3.58X to 26X UNG improvement, 10% to 41% performance enhancement, and 6% to 22% power reduction compared to existing leakage tolerant domino techniques. New proposed domino circuit-2 has a lower UNG compared with proposed circuit in [30], but it improves speed of circuit with 10%-15%. We employ our two proposed circuits to design new comparators and MUXs.

Fig.2.24 shows the results of this experiment for 16-in MUX in the worst case I_{off} corner of the 70-nm predictive technology at 0.9 V and 110°C. As observed from Fig.2.24, the UNGs of the proposed circuits are larger than that of the FLDL design, and the proposed circuits design show the best delay among all the designs. Of course, the pseudo-static implementation has the highest UNG, but its delay is larger than our proposed circuit. Also, proposed circuit-2 shows a higher speed than FLDL MUX with a higher UNG. For the comparator circuit-1, as observed from Fig.2.25, the UNG of the proposed circuit is considerably larger than that of the Footless Domino Logic Comparator (FLDLC), and the

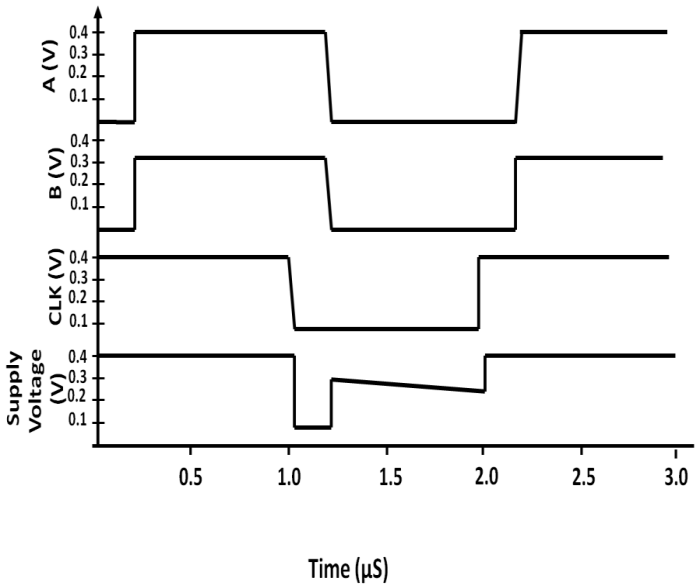


Fig.2.27. Waveforms of ASV circuit

delay of the proposed design is comparable to the best delay of the standard domino design, but for proposed comparator-2, the speed is improved compared to FLDL comparator in some cases. However the second domino design (e.g. circuit-2) has an area overhead more than 10% compared to other circuits.

2.3.2. New Domino Logic Style Using Adaptive Supply Voltage

The application of aggressive circuit design techniques which only focus on enhancing circuit speed without considering power is no longer an acceptable approach in most high complexity digital systems. The power consumed in high performance integrated circuits has increased to levels that impose a limiting factor on the system performance and functionality [43]-[46]. Compared to a typical static gate, a domino logic gate operates at a higher speed and occupies less area while implementing the same function [45]. However, deep sub micrometer (DSM) domino logic circuits utilizing low power supply and threshold voltages have decreased noise margins [43]. Fig.2.26 shows the proposed domino logic circuit using the adaptive supply voltage (ASV) circuitry.

This circuit was designed to work at very low supply voltages (lower than V_{DD}=0.4V), so the drive current of NMOS transistors decreases significantly which in turn degrades the performance of the circuit. For footed domino logic circuit (Fig.2.26. with constant supply voltage), NMOS devices are upsized to overcome the performance degradation that in turn increases the area overhead and also dissipates more power. To illustrate how this adaptive supply voltage circuit works, Fig.2.27 is constructive. As it can be seen when Clk is low (Clk_bar="1", Precharge phase), the supply voltage node is connected to V_{DD}. In evaluation mode (Clk_bar="0"), When all inputs are zero (idle mode), the supply voltage decreases to mitigate the power dissipation, but lowered supply voltage deteriorates the reliability and sensitivity to input noise because of reduced on-current through the PMOS keeper transistor to hold the dynamic node state. To overcome this problem in our design,

the NMOS devices in the evaluation network are downsized. In ultra low supply voltages the current through NMOS transistor is lower than PMOS transistor in the same conditions.

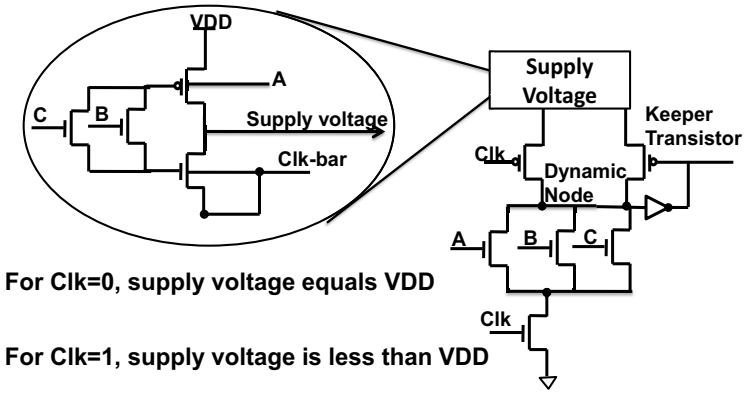


Fig.2.28. 3-inputs OR-Gate topology

Table.2.7. power consumption comparison (nW)

	FF			TT			SS			FS			SF		
	-40°C	27°C	110°C	-40°C	27°C	110°C	-40°C	27°C	110°C	-40°C	27°C	110°C	-40°C	27°C	110°C
FDL	4.2	30.5	211	1.96	7.54	57.9	2.57	2.57	19.5	1.27	3.02	30.2	21.7	105	364
New Logic	8	39.5	156	1.53	9	49.4	0.51	2.65	17.5	0.82	4.4	26.7	0.73	3.7	26.3

Furthermore, because of lowered supply voltage the DIBL is reduced significantly so the leakage current of NMOS transistor in idle mode is decreased significantly, but when at least one of the input signals is high, the supply voltage decreases to a level lower than $V_{DD}/2$ increasing the speed of evaluation. Furthermore, lowering the supply voltage in this phase mitigates the contention between PMOS keeper transistor and NMOS evaluation network to hold the state of the dynamic node. This circuit is appropriate for OR-gates with less than four inputs. To use it for higher fan-in OR gates, three signals of all inputs to the adaptive supply voltage circuit might be used.

Efficiency of this technique is strongly dependant of the input signals. When at least one of the selected input signals (applied signals to ASV) is at high level, the effect of ASV technique on reducing power consumption and improved speed is distinct. However, when all randomly selected signals applied to ASV circuit are zero, this circuit increases the area overhead and also the power consumption, but this degradation in power consumption is not significant. High fan-in gates are prone to failure due to very low I_{on}/I_{off} ratio. For two-input domino AND gate, due to three stacked NMOS transistors the probability of failure is very high. To avoid the failure NMOS transistors must be upsized significantly, that increases the area, but by applying this technique to the AND gate, when all inputs are high (A=1, B=1, Clk_bar=0), the supply voltage node is at the lowest rate which lets the evaluation network to discharge the dynamic node without any significant contention between the evaluation network and the PMOS keeper transistor. For higher fan-in (4-

inputs OR-gate), two input signals are applied to the ASV circuit randomly. The probability that A or B is high, is 0.5. Then due to the significant effect of the ASV circuit to mitigate the power dissipation and also increasing the speed, this circuit is useful. This probability is lowered for higher fan-in. For wireless network applications, using low fan-in gates (Fan-in=2) is usual. Therefore, in this case our proposed circuit has a higher speed and lower power dissipation.

Table.2.7. shows the results for power consumption of a 2-input OR gate for footed domino logic circuit employing the new logic circuit using adaptive supply voltage technique. As the results show, the power consumption decreases significantly while the new logic uses more transistors. Also, this technique enables domino logic circuit to work at ultra low supply voltages (subthreshold region), due to decreased contention between PMOS and NMOS transistors at the primary time of the evaluation phase. The delay of the proposed circuit does not increase in the same supply voltage (with the same size). In some cases the speed of domino logic using this technique increases. For higher fan-in gates, we can add NMOS transistor in parallel with input B in ASV circuit. The topology of 3-inputs domino OR-gate is illustrated in Fig.2.28.

2.4. Conclusions

In this chapter, several low power full-adder topologies were explained in details. We modified full-adder circuits to work in sub-threshold region properly. One of the most important issues in circuit design is leakage current through high fan-in gates for today's data path designs. In this chapter, some domino logic circuits were proposed for ultra low power data-path design along with utilizing domino style logic in some topologies such as high fan-in multiplexers and comparators. An adaptive supply voltage technique was proposed to reduce the failure in different logic design. In this chapter an application of proposed adaptive supply voltage technique was presented for domino logic style that showed significant improvement in power and performance of domino logic circuits.

Chapter 3

Flip-Flop Design for Low Power Applications

Chapter 3

Flip-Flop Design for Low Power Applications

Introduction

With the rapid growth of multimedia services in wireless applications, the demand for signal processing systems that deliver high-quality/high-performance levels under ever-diminishing power budgets has increased tremendously [46]. Therefore, power reduction techniques at various levels of design abstraction [48] have been proposed. Such techniques either try to reduce the complexity of computations or attempt to apply voltage and frequency scaling adaptively depending on the load of operations. Apart from such techniques, power can be reduced by reducing the power of basic elements of today's circuits. Interestingly, ubiquitous elements of today's complex systems are flip-flops (FF) which make up a major portion of the synchronous circuits. As a result, the structure of FF used in circuits has a large impact on the system power consumption. Moreover, the type of FF used determines the amount of clock load, which directly affects dynamic power consumption P_{DYN} of a circuit. Thus, it is prudent to come up with techniques to reduce the power consumption of FFs to reduce the overall system power.

Techniques to lower unnecessary internal node switching can be used to reduce the total power consumption, particularly for applications with lesser-input activities. It can be noted that the signal activity can be low in certain input bits (most significant bits, MSBs) of signal processing applications such as in image processing and video processing [49,50]. Unfortunately, conventional FFs do not take advantage of such low input activity and the power consumption due to such FFs can be large. Conditional-capturing flip-flop, presented in [51, 52], achieves improvement in power consumption for lower activities of inputs. However, the technique uses skewed inverters and extra circuitry causing significant area overhead. We propose a new Sense Amplifier FF (SAFF) in which the switching of the internal nodes is dependent on switching activity of the input data (Data-Dependent SAFF (DD-SAFF)). Hence, unnecessary transitions are eliminated during clock transitions when input data is unchanged. We show that the proposed FF has much lower power consumption (40%) compared to SAFF. However, the power savings with the proposed design comes at a cost of minor degradation (5%) in setup time and Clk-Q delay. Another technique that is used

to lower the power consumption along with the total delay is known as *pulsed flip-flop* technique. Such implementation results in reduction of negative setup time (T_{su}) and thus the total delay $T_{tot}=T_{cq}+T_{su}$ (where T_{cq} is the delay from the input clock to the output Q of the FF). Note that, T_{tot} is the Figure of Merit for a FF, since T_{cq}, T_{su} delays together with T_{pd} delay (propagation delay of a combinational logic), determine the operating frequency of a pipeline system [53]. Hence, the desire for reduced delay introduced by the FFs (that is T_{cq} and/or T_{su}) motivates for the development of a new and low-power pulse-based flip-flop. Recently the proposed pulsed FFs utilize a pulse generation technique to produce the desired clock. These FFs are often modifications of the master-slave latch-pair [54, 55] or pulse-triggered latch [56-58]. Note that, pulsed latches allow smaller data-to-output delay than the master-slave latches. This is attributed to the fact that their critical path consists of a single latch (assuming that the input data arrives when the latch is fully transparent). In addition, the pulse generating circuitry may be implemented (a) *internally* (i.e. as a part of a single flip-flop) or (b) *externally* (i.e. externally produced and then applied to adjacent flip-flops). The latter technique reduces the total power consumption significantly due to clock ring between neighboring FFs. We apply a clock-gating technique to produce a specific shaped pulse showing fewer penalties in area and delay compared to existing techniques. A narrow clock pulse is generated and applied to an array of FFs to reduce the total power of the clocking circuitry.

In this chapter we propose a design approach that utilizes the advantages of the new data dependent and pulsed flip-flops described above. Specifically, we propose a preferential usage of flip-flops in various architectures depending on the type of computation that they are sampling. Initially, we identify the FFs that can be replaced with the proposed data dependent flip-flops in various DSP applications. We show that the use of data dependent flip-flop for nodes that exhibit low switching activity in the various DSP applications leads to a significant power reduction. Interestingly, when the data dependent flip-flops are combined with pulsed approach, power is further reduced due to lower switching. To prove the efficacy of the proposed approach we apply it to various popular DSP applications such as discrete cosine transform (DCT) and FIR sub-systems. The contributions of our designs can be summarized as follows:

1) Proposes a data dependent flip-flop able to reduce power by up to 42% when used to sample bits with low switching activity.
2) Proposes a new low overhead pulse generation circuit that allows further power reduction and delay penalty, thereby increasing the efficiency of pulsed flip-flops.
3) Combines the data dependent flip-flops with the pulsed approach and uses them in locations of architectures that showcase low switching activity.
4) Apply the combined approach to various popular applications such as counter, DCT and FIR filtering. We first implement a counter where the most significant bits (MSBs) are implemented using DD-SAFF, since they have lower switching activity compared to the least significant bits (LSBs). In addition, we applied the proposed FF to widely used DSP applications, FIR filter and DCT in which FFs play a major role in total power consumption of the circuit. Results show large improvement (25.7%-42%) compared to using conventional sense-amplifier based flip-flops.
5) Compare the proposed flip-flops with various existing ones.

3.1. Flip-Flop Design

Several low-power flip-flop architectures have been proposed in literature to decrease the power consumption of a system [59]. The SAFF design can be applied to digital signal processing (DSP) applications that consist of systems such as Discrete Cosine Transform (DCT), FFT, Viterbi. By utilizing SAFF in these applications, lower power consumption is achieved [60]. Thus, it is obvious that the desire for delay improvement of the SAFF can be

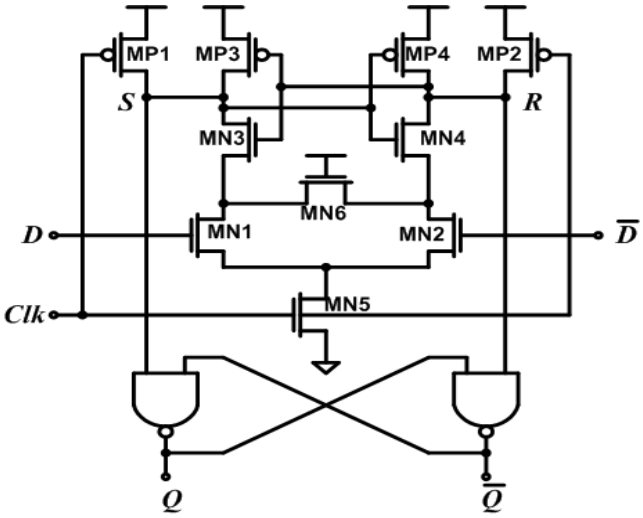


Fig.3.1. Sense-Amplifier Flip-Flop

achieved by decreasing the number of NMOS stack transistors. Note that the upsizing of the stacked NMOS transistors is another way to achieve higher performance. Furthermore, in case evaluation is done for a shorter time (after rising-edge of the clock), the flip-flop robustness is improved due to disconnected storage nodes (S and R) from ground.

In this section we explain different conventional flip-flops that show better performance in terms of speed and power compared to other counterparts. We focus on sense-amplifier flip-flop (SAFF) that has lower power consumption compared to other conventional flip-flops. Moreover, it is suitable for dual-rail data-paths [61]. Easier integration of logic circuitry with flip-flops and reduced swing of the clock pulses are other advantages of this circuit [62]. SAFF incorporates a precharged sense amplifier in the first stage to generate a negative pulse, and a Set-Reset (SR) latch in the second stage to capture the pulse and hold the results [63, 64]. Since the proposed design is based on SAFF, let us first consider the operation of sense-amplifier flip-flop in more detail.

3.1.1. Sense-Amplifier Flip-Flop

The schematic of sense-amplifier flip-flop is shown in Fig.3.1. The circuit works as follows: when the clock is in precharge mode, the master stage nodes (S and R) are both charged to V_{DD} using two PMOS transistors (MP1 and MP2). When the rising edge of the clock arrives, S or R is discharged to ground depending on the value of input (D). In the case when D is high, three stacked NMOS transistors discharge node S to ground that causes output (Q) to charge to "1". Due to the cross-coupled inverters, R remains high. Note that SAFF comes with a delay penalty. This is due to two reasons: the use of three stacked NMOS transistors and the low speed of the static output latch. In terms of power, SAFF outperforms its counterparts due to stacking effect and minimum sized design. However, it consumes unnecessary power even when the inputs do not change. The major internal nodes are precharged continuously even when the clock is low. For instance let us consider a scenario in which D remains constant (e.g. at "1") for long period of time. In this case, S and R hold "0" and "1", respectively. When clock is low, internal nodes are precharged to V_{DD}. Thereafter, on rising edge of clock, S is discharged to ground. Consequently, internal nodes of the SAFF charge/discharge, regardless of the input condition. Hence, the total power consumption of the SAFF can be significantly reduced by avoiding such redundant internal switching.

Other flip-flops that are explained are as follows: a) Hybrid latch flip-flop (HLFF). b) Modified transmission gate flip-flop (TGFF).

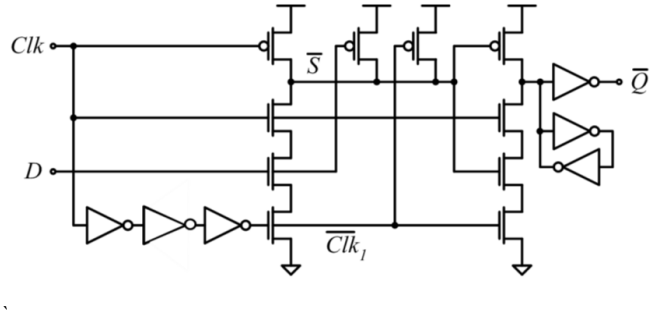


Fig.3.2. Hybrid-Latch Flip-Flop

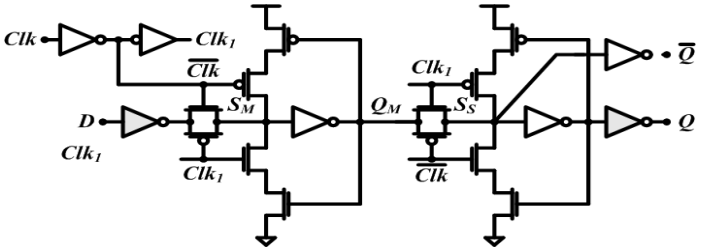


Fig.3.3. Modified Transmission-Gate Flip-Flop (MTGFF)

3.1.2. Hybrid Latch Flip-Flop (HLFF)

Fig.3.2 shows the *Hybrid-Latch Flip-Flop* (HLFF) circuitry [65]. Main advantages of this topology are the small delay and the low power consumption. To better explain its functionality let us consider two cases: i) at the rising edge of the clock, if input data is "1", the stacked NMOS network has $3\tau_{inv}$ time to discharge node $\overline{S}$ to '0', (τ_{inv} is the delay of a single inverter). When $\overline{S}$ is discharged to ground, the second stacked NMOS transistor network is OFF and the far right PMOS helps to charge Q to V_{DD}. ii) When input is low, the first stage stacked NMOS network stage is OFF and the second is ON. $\overline{S}$ is charged to V_{DD} and bottom NMOS transistor in second stage turns on for a delay of $3\tau_{inv}$ after the rising of the clock. At sufficiently high supply voltages, $3\tau_{inv}$ is enough time to evaluate the state of output. However, at low supply voltages (e.g. V_{DD}=0.8V), the delay increases significantly due to the decreased drive current of the stacked transistors. Consequently, the flip-flop fails to operate properly for such low voltages. Note that on the falling edge of the clock (precharge mode), $\overline{S}$ is "1" and both stacked NMOS branches are OFF (retention mode).

3.1.3. Transmission-Gate Flip-Flop

The *transmission-gate flip-flop* (TGFF) with input gate isolation is shown in Fig.3.3 and

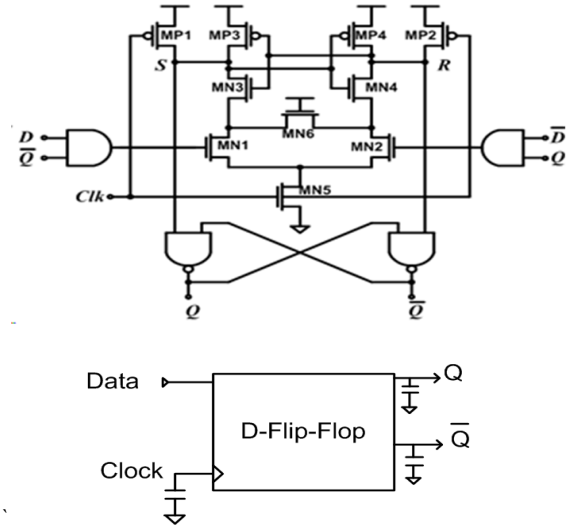


Fig.3.4. Data-Dependant FF (DD-SAFF) and Test-Bench

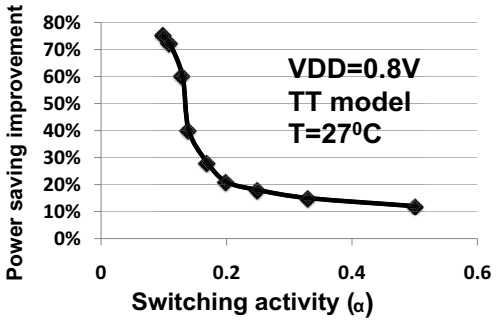


Fig.3.5. Power consumption improvement for DD-SAFF

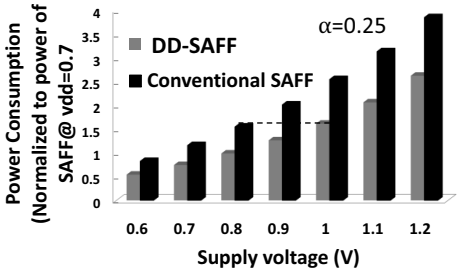


Fig.3.6. Power consumption for different supply voltages

was used first time in PowerPC603 latch-pair [54]. This FF works as a latch pair, where one is transparent high and the other is transparent low. Inverted data is applied to the input for better noise immunity. Furthermore, the output is inverted to provide non- inverting operation. This FF shows low power consumption along with high speed.

3.2. Data-Dependent Sense-Amplifier Flip-Flop

Fig.3.4 shows how the proposed circuit utilizes data-dependency to lower power consumption. We have added two AND gates (N1 and N2) whose inputs are D and Q_bar (D_bar and Q for N2). The purpose of the AND gate is to avoid MN1 and MN2 from discharging the nodes whenever the new data and previous data are the same. The flip-flop works as follows: When clock is low, the internal nodes (S and R) are precharged to V_{DD} through precharge transistors. During the evaluation phase (rising edge of the clock), MN5 is turned ON. Based on the values of Q and Q_bar, the status of MN1 and MN2 are determined. Let us consider a scenario by assuming that the previous value for Q is "0". As a result, output value of gate N2 will be "0". MN2 is turned OFF and MN1 status is dependent on D. If D is "1", MN1 is turned ON. Therefore, the evaluation is done through MN3, MN1, and MN5. Accordingly, S is discharged to ground and Q is changed to "1". Though, if D is "0" (i.e. same as previous state of Q), both MN1 and MN2 are turned OFF and the S and the R values are unchanged. Consideration of data-dependence significantly reduces the power consumption of the DD-SAFF.

Conversely, when the prior value for Q is "1", MN1 is OFF and MN2 status is determined by input D. In the case that D is "1" (i.e. same as the previous state of Q), both MN1 and MN2 are turned OFF and the internal nodes (S and R) of the FF do not discharge. In case D is low, discharging of node R to ground is through MN4, MN2, and MN5. Consequently, for lower activity inputs, assuming D remains at "1" for a long period of time, nodes S and R remain unchanged at "1". Avoiding continuous charge/discharge during the evaluation phase, leads to significant reduction in power dissipation for applications where input data remains unchanged for long periods of time

3.2.1. Simulation Results for DD-SAFF

To assess the performance of the proposed FF, we compared DD-SAFF with conventional SAFF in terms of power and delay. The power savings in DD-SAFF compared to conventional SAFF is strongly dependant on the input data activity. For different values of **α** (defined as $Tclk/T_{Data}$), it is observed that lower values of α leads to larger power savings (α=0.2 results in 2 times less power compared to SAFF). The worst case in terms of power consumption for DD-SAFF operation is when the clock and data frequency have the same activity (α=1). In this case, the power consumption for DD-SAFF deteriorates by 36% compared to a conventional SAFF (*f*=1GHZ, V_{DD}=0.8V). This is due to the extra AND gates used in the DD-SAFF. Thus, the DD-SAFF should be used judiciously in applications where the input data activity is low. Fig.3.5 illustrates the power savings (iso-delay) obtained by the proposed FF compared to conventional the SAFF for different switching activities at V_{DD}=0.8V. As it` can be observed for lower α, power savings are significant (e.g. for α=0.1, around 70% improvement in power consumption). Fig.3.6 illustrates the results for total power consumption of the data-dependant sense amplifier flip-flop (DD-SAFF) compared to conventional SAFF. To compare power consumption, circuits are simulated at different supply voltages. As shown, the power consumption of a DD-SAFF at V_{DD}=1V is approximately equal to that for a SAFF at V_{DD}=0.8V. The DD-SAFF is associated with some delay penalty as mentioned earlier. The DD-SAFF shows 5% delay degradation compared to a conventional SAFF. However, by upsizing the evaluation NMOS transistors in the DD-SAFF, delay degradation can be compensated, albeit at the cost of higher power consumption. Fig.3.7, illustrates the simulation results for the delay of DD-SAFF compared to a SAFF from V_{DD}=0.5V to V_{DD}=1.2V.

Table.3.1. Power and delay for different topologies (V_{DD}=0.8V, α=1/3, Fan_out=4)

Topology	$T_{C\text{-}Q}$(ps)	Power(μW)	PDP (fJ)
C²MOS	37	3.1	114.7
SAFF	53	1.93	102.29
TGFF	50	1.097	54.85
DD-SAFF	53	1.19	63.07

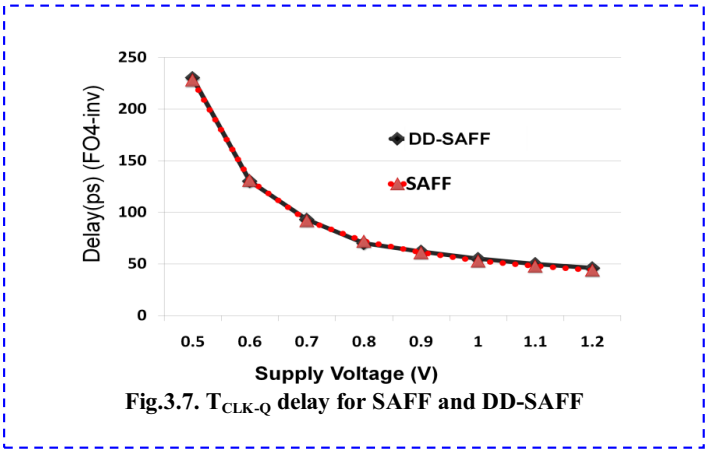


Fig.3.7. $T_{CLK\text{-}Q}$ delay for SAFF and DD-SAFF

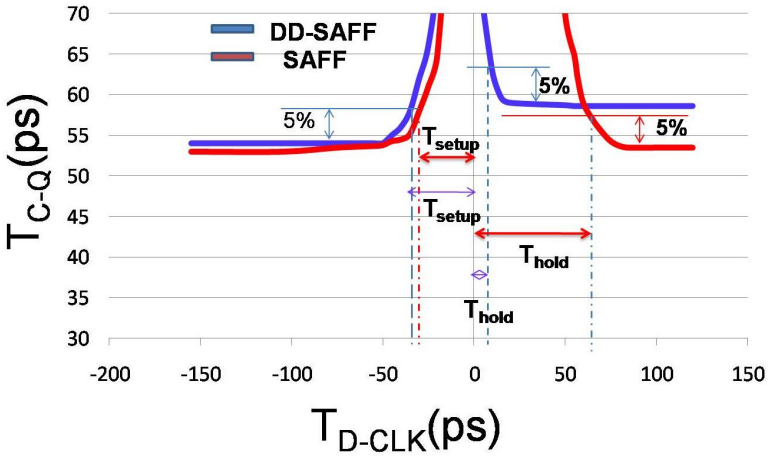


Fig.3.8. Setup and hold time for SAFF and DD-SAFF

We have also compared the performance of the proposed DD-SAFF with other energy

efficient FF architectures and have chosen the power-delay product as the comparison metric. For low power applications, the transmission-gate flip-flop (TGFF) has been proposed in [56], [66]. On the other hand, C^2MOS flip-flop is a good choice for high-speed applications. All the three architectures have been simulated under iso-voltage (V_{DD}=0.8V) and iso-frequency (f=1GHz) conditions. Table3.1 summarizes the performance results of our comparison study. As shown, the PDP of the DD-SAFF is 38% and 44.6% better compared to the conventional

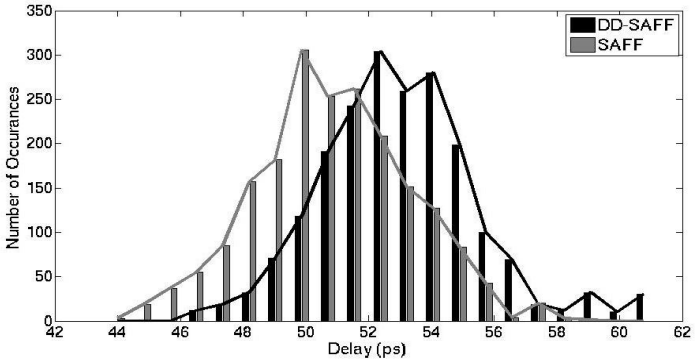
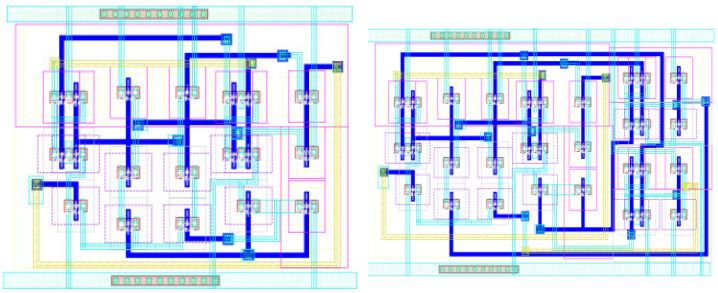


Fig.3.9. Monte-Carlo simulation for delay comparison



(a) (b)

Fig.3.10. Layout of (a) SAFF (b) DDSAFF

SAFF and C^2MOS FFs, respectively. Compared to the TGFF, the PDP results are comparable. However, the TGFF suffers from poor data-to-output latency due to positive setup time. Moreover, it also requires two clock phases.

Another metric that can be used to benchmark FF's is the setup/hold timing values. As shown in Fig.3.8, 5% increase in $T_{C\text{-}Q}$ delay is chosen as the reference point to measure the setup and hold time [69], [67] in our calculations. Although, setup time increases from 32ps to 35ps for the DD-SAFF, hold time improves considerably from 60ps to 10ps. In scaled CMOS technologies, the impact of process variations on the circuit performance can be substantial. Hence, we performed Monte-Carlo analysis (2000 trials) for both DD-SAFF and SAFF. It can be observed from Fig.3.9 that for the DD-SAFF, standard deviation and mean of delay variation are 6.3ps and 52.9ps, respectively while for the SAFF are 6ps and 50.9ps. Thus, DD-SAFF has similar σ/μ compared to SAFF. In terms of area, the proposed DD-SAFF design shows 25% area overhead compared to conventional SAFF. Fig.3.10 shows the layout of DD-SAFF and SAFF design.

Simulation results for the DD-SAFF compared to its counterparts in terms of delay, power, PDP, and setup/hold timing values have been produced and the design showed significant improvement in power (as high as 70% for α=0.125). Furthermore, it gives more than 38% and 45% improvement in PDP compared to the conventional SAFF and C^2MOS-FF, respectively.

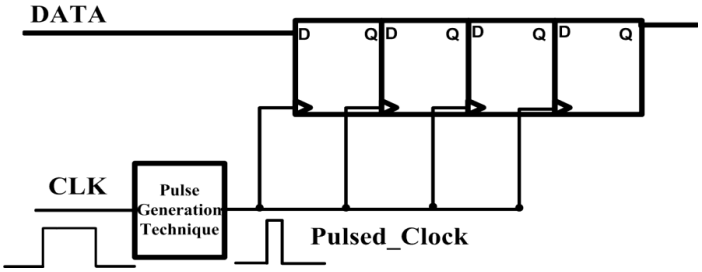


Fig.3.11. Pulse generation technique

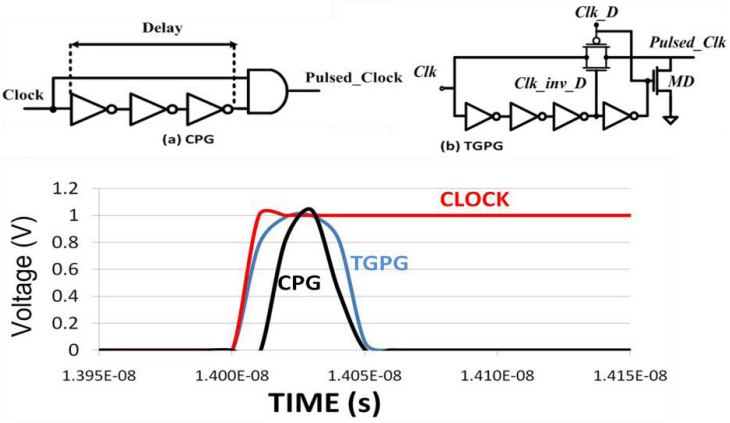


Fig.3.12. Output waveforms for CPG and TGPG designs

3.3. Pulsed-Flip-Flop Design

A pulse generator circuit and array of latches are shown in Fig.3.11. The shaped pulse is generated after the rising edge of the main clock. This enables data to arrive even later than the clock edge, and hence setup time becomes negative. This property is exploited to reduce the total delay given by *Tsu+Tcq*. The pulse generator circuit can be shared across several flip flops, as shown in Fig.3.11, mutually amortizing its area and power cost. Several design techniques have been proposed to generate the specified pulse [68-70]. The delayed clock ANDed with the clock is one of the most common methods. Note that, the corresponding output will be a pulse that is high for a time equal to $3\tau_{inv}$ starting from the rising edge of the clock (where τ_{inv} is the delay of a single inverter) and after that time it resets to '0'. Fig.3.12 shows the schematic of conventional pulse generation (CPG) incorporating the proposed transmission-gate pulse generator (TGPG) and the corresponding output waveforms. CPG poses penalty in area and delay due to an extra AND gate. To decrease the delay of the CPG, the TGPG technique is proposed to generate the desired pulse. The TGPG scheme is

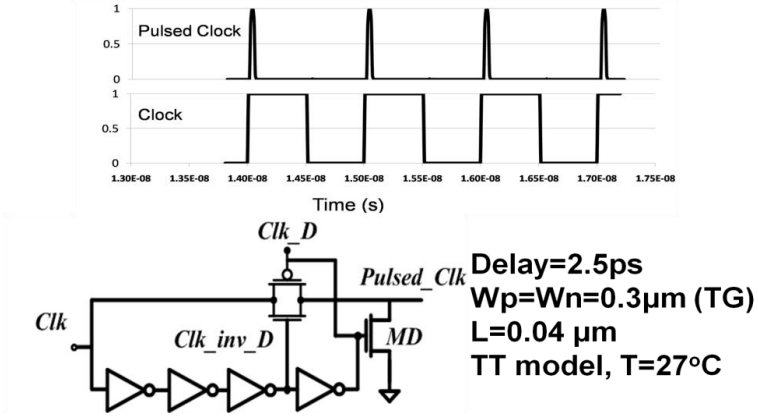


Fig.3.13. Transmission-Gate Pulse Generator (TGPG) output waves and scheme

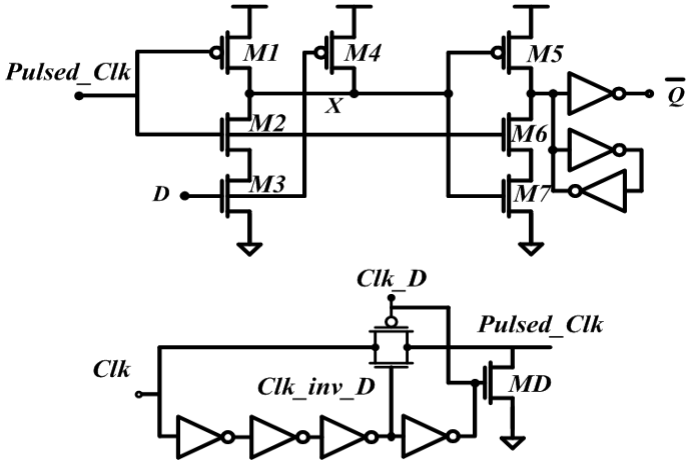


Fig.3.14. PHLFF design

illustrated in Fig.3.12 (b). At the rising clock edge, CLK_D is at "0", and CLK_inv_D is at "1". Therefore, the transmission gate is ON and the corresponding data is transferred to node Pulsed_Clock. Since PMOS devices are slower, the PMOS transistor has τ_{inv} more time to transfer the data to node Pulsed_Clock. After $4\tau_{inv}$ delay (delay between main CLOCK and CLK_D in Fig.3.13), the NMOS MD transistor turns on and discharges the node Pulsed_Clock to ground. At the falling clock edge, for the primary time ($4\tau_{inv}$) after the rising clock-edge, node Pulsed_Clock is connected to ground via NMOS MD. Afterwards, the transmission gate turns on and keeps the node Pulsed_Clock connected to "0".

At iso-size for the two pulse generator techniques, the TGPG (transmission gate pulse generator) shows better results in terms of speed and power. Fig.3.13 shows the output waveforms of node Pulsed_Clock of TGPG circuit. The delay of this circuit is 2.5ps compared to 15ps of the CPG topology. Furthermore, the proposed circuit uses one less transistor compared to the CPG topology (i.e. AND gate is implemented using a NAND gate and an inverter). The pulse-generator circuit is shared among neighboring FFs in a system.

Consequently, it does not impose significant penalty on both area and power consumption of a system.

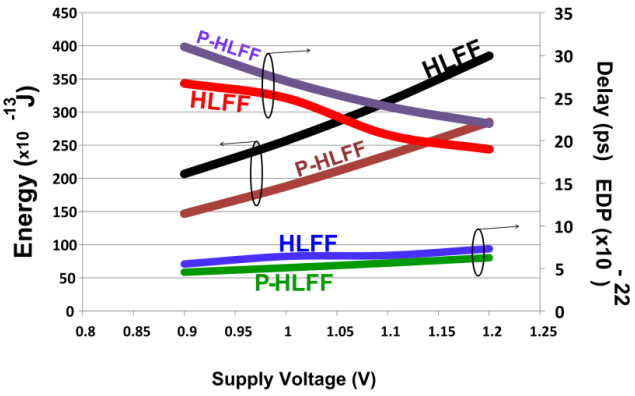


Fig.3.15. Delay and Energy comparison for HLFF and PHLFF

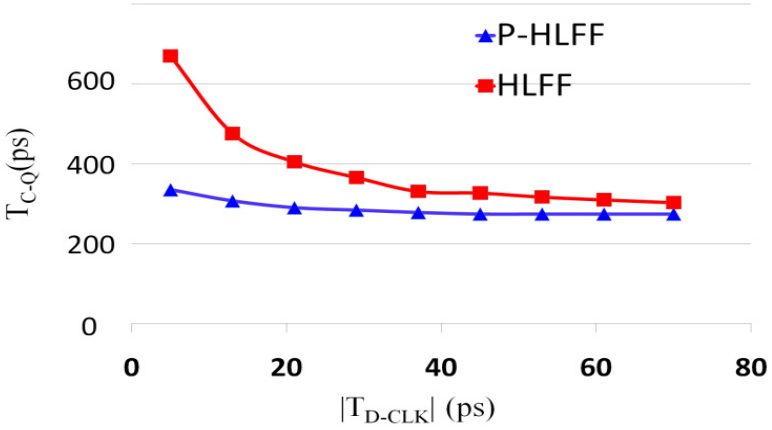


Fig.3.16. T_{C-Q} versus T_{D-CLK}

By applying the TGPG technique to conventional FFs, several pulsed-FFs are presented to show the efficacy of the proposed pulse-generation on power and delay. In this section, different FF topologies are explained along with the proposed topologies.

A. Proposed Pulsed-HLFF (PHLFF)

Fig.3.14 shows the circuit topology of the PHLFF flip-flop. It works as follows: at the rising clock edge, the transmission–gate transistor turns on and the corresponding input data is transferred to the drain of MD NMOS or the gate of M3. Then node X discharges to ground. In this case, M1 and M4 are turned off and the discharged node X causes M5 to turn on charging node Q to V_{DD}. When D is "0", M3 and M4 turn off and on, respectively connecting

the node X to V_{DD}. The transistor M6 is turned on due to Pulsed_Clock signal for $3\tau_{inv}$ after the rising edge of the main clock. This time is enough to discharge Q to ground due to the reduced number of stacked NMOS transistors. After time equal to $4\tau_{inv}$ of the rising clock edge, M6 turns off and disconnects the discharging path from ground. Node X goes high through M1 for the rest of the clock period. Therefore, the PHLFF works only during a short pulse for $4\tau_{inv}$ then it goes in retention mode resulting in saving power. Since Pulsed_Clock is off for the rest of the clock period after $4\tau_{inv}$ from the rising edge of the clock, changing data does not change the output because X is connected to "1" and M6 is OFF. The clocking topology is shared among neighboring FFs, therefore the area overhead is negligible. Compared to conventional HLFF, the proposed circuit has one extra inverter. Due to a smaller number of stacks in the PHLFF, delay is reduced by 16% at V_{DD}=1V compared to the conventional HLFF. Simulation results for energy and delay of HLFF and PHLFF are shown in Fig.3.15. It is obvious that the EDP of the PHLFF is improved significantly. Fig.3.16 shows the comparison for T_{C-Q} delay for HLFF and PHLFF. Note that, the setup time is very close to zero. Our results show that the negative setup time results in low $T_{su}+T_{C-Q}$.

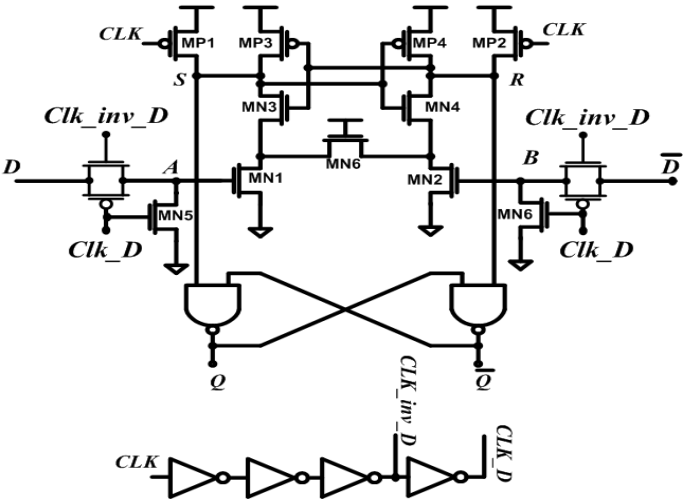


Fig.3.17. Pulsed-SAFF

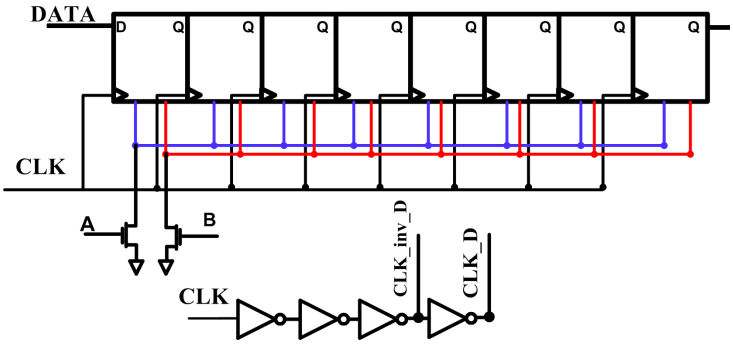


Fig.3.18. Shift register block

B. Proposed Pulsed-SAFF

To reduce the number of stacked NMOS devices, we employ the TGPG technique, while we use data as an input instead of the clock. Fig.3.17 shows the proposed circuit diagram using TGPG technique.

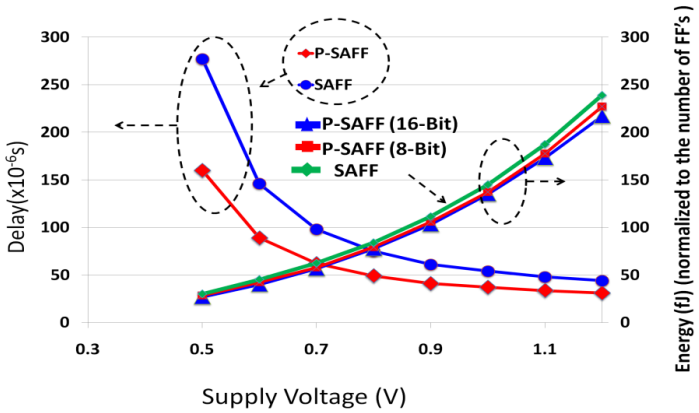


Fig.3.19. Energy and delay comparisons for SAFF and PSAFF

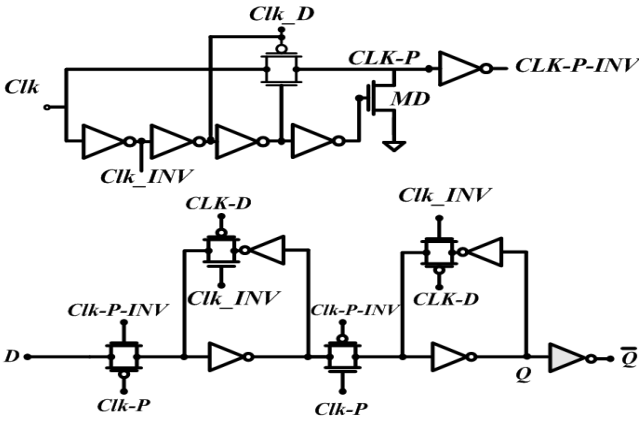


Fig.3.20. PTGFF topology

By assuming clock on rising edge, depending on the D value (e.g. "1"), after a delay of 3ps (delay of one PMOS transistor), the left side stacked NMOS network is turned on and discharges node S to ground forcing Q to go high and Qb to zero. The evaluation time should be less than 3τ since the TGPG disconnects the evaluation NMOS network from ground for the rest of the cycle. As we explained for the TGPG circuit, after $4\tau_{inv,}$ the CLK_D becomes "1", connecting the gate of the bottom NMOS in the stacked network to ground. The turned off bottom NMOS transistor cuts off the discharge path, resulting in no change in stored data on nodes S and R. When the clock is on the falling edge, nodes S and R are connected to "1" regardless to the input data, causing Q to remain unchanged. The pulsed-SAFF gives better results in terms of delay and power compared to the conventional SAFF due to lower internal switching and lower number of stacks, respectively.

For the proposed PSAFF, since two PMOS transistors are added to the main SAFF cell and the clocking circuitry is shared between adjacent FFs, the area overhead is negligible. Fig.3.18

shows how the proposed PSAFF is applied to an 8-bit register. As it can be seen, there is no area overhead due to the use of the pulsed SAFF technique. As it is illustrated in Fig.3.19, even though the power saving is improved at low supply voltages such as V_{DD}=0.5V, the corresponding delay is improved by 42% compared to the SAFF design.

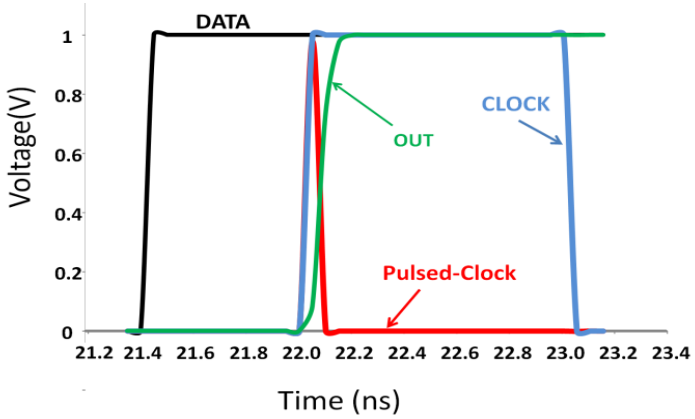


Fig.3.21. Output waveforms of PTGFF

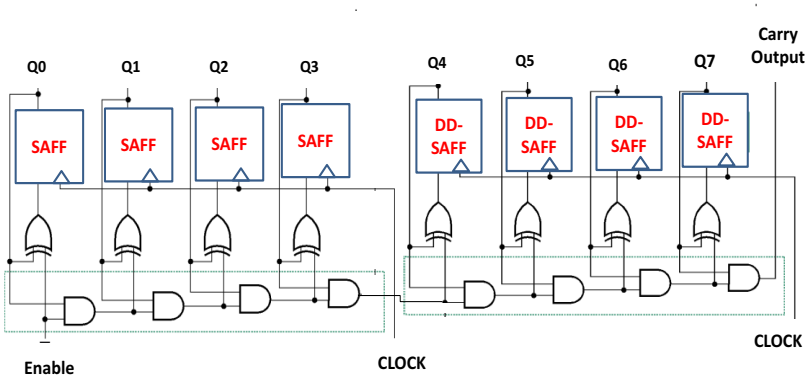


Fig.3.22. 8-bit counter

C. Proposed Pulsed-Transmission-Gate Flip-Flop

The *Pulsed-TGFF* topology is depicted in Fig.3.20. The PTGFF works as follows: on the falling clock edge, data is latched and transferred to the output when clock is on the rising clock edge. Transmission-gates turn on for a short time equal to the width of the shaped-pulse. The corresponding waveforms for PTGFF are demonstrated in Fig.3.21. When the clk-p signal goes high, the input data is transferred to the node Pulsed_Clock. Simulation results show 13% and 16% improvement in power dissipation for 8-bit and 16-bit shift-registers, respectively. Furthermore, delay comparison shows 20% improvement compared to MTGFF topology.

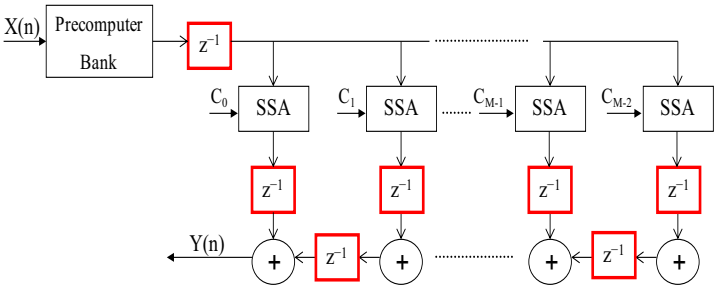


Fig.3.23. CSHM based Fir Filter of size M

Table.3.2. Energy saving percentage for data-dependant 8-bit Counter compared to using conventional SAFF (V_{DD}=0.8V)

Counter	8	16	32	64
power saving %	17.6%	28.4%	37.3%	47.1%

Table.3.3. Power comparison of FIR-Filter

Flip-Flop type	POWER
SAFF	1.34mW
DD-SAFF	771 µW
ST-standard FF	1.0385mW

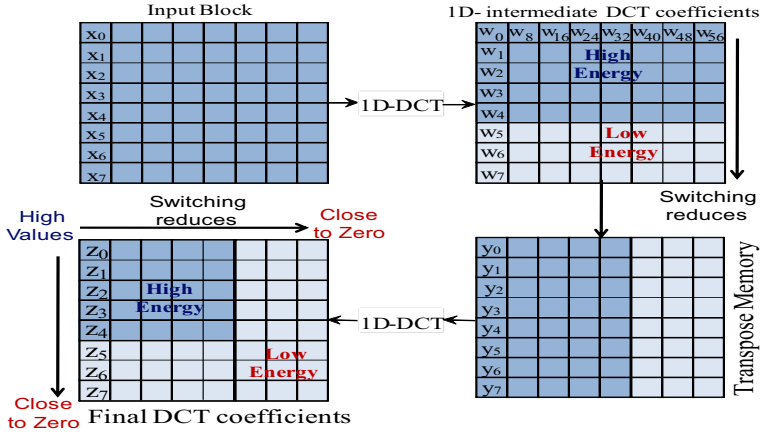


Fig.3.24. 2D-DCT operation, areas of high energy-high switching activity outputs and low energy outputs.

3.4. Flip-Flops in DSP Applications

As we discussed in chapter one, Flip-Flops are important to store the intermediate values in wireless sensor nodes. To investigate the efficacy of the proposed flip-flops, they are used in some DSP applications. In this section, we utilize the DD-SAFF and Pulsed-FFs in different applications such as counters, FIR filter and discrete cosine transform (DCT).

3.4.1. Data-Dependent Counter Design

To prove the efficacy of the DD-SAFF design, we employed DD-SAFF in the implementation of an 8-bit synchronous counter, the topology of which is shown in Fig.3.22. Counters are suitable choices for demonstrating the application of the DD-SAFF, due to lower input activity in MSBs of the counter [71]. The operation of the synchronous counter is as follows: new data values are evaluated every clock cycle and captured by related flip-flops. The switching activity of each bit in the counter is decreased by half compared to the preceding bit. Thus, in each clock cycle unnecessary transitions occur, especially for MSBs. Although for α=1, power consumption of DD-SAFF degrades by 36% (compared to conventional SAFF) but for lower input activity, power improves significantly. As a result, a conventional SAFF can be used for sections of logic with α close to one ($\alpha \cong 1$), while for α<0.5 the DD-SAFF can be used. Thus, for our implementation, the conventional SAFF is used for LSBs and we use the DD-SAFF for MSBs. As it is illustrated in Fig.3.22, for the 8-bit counter, four LSB bits are implemented using a conventional SAFF while four MSB bits are implemented using the DD-SAFF. The simulation results indicate that the improvement in power is close to 17%. For larger counters (e.g. 16 bits, 32bits, 64 bits), the power consumption is reduced even further since more bits have lower data activity. The performance results for different counters are presented in Table.3.2

3.4.2. FIR filter Design

Apart from counters, the proposed DD-SAFF has the potential to reduce power consumption of popular DSP applications, such as FIR filters. The basic operation of a FIR filter is the multiplication of a coefficient vector $C = [c_0, c_1, \ldots, c_{M-1}]$ with an input signal $x(n)$, referred to as vector scaling operation. Various low power techniques have been proposed for the design of FIR filters with most common among them being the removal of redundant computations. An example of low power FIR filter architecture, known as computation
sharing multiplier (CSHM) is shown in Fig.3.23. CSHM is based on the principle that in vector scaling operations, any scalar s_i can be decomposed into smaller bit sequences a_k (alphabets) such that s_i can be rebuilt from these sequences by few shifts and adds. Using these alphabets the coefficient vector (C) can be constructed spanning the entire set of filter coefficients [72]. This reduces the entire multiplication operation of the filter to shift/select and add operations using Select/Shift and Adder (SSA) units (Fig.3.23) instead of power intensive multipliers. Note that a pre-computer, computes the product of alphabets and input vector X in advance and stores them for re-use, thereby reducing the number of operations and maximizing the computation sharing with the help of SSAs.

However, irrespective of the techniques used to reduce power at the algorithm/architecture level, all resultant filter architectures have a
common component: - FFs. Interestingly, low power filter design techniques in their attempt to increase the sharing of computations and avoid the use of multipliers, sometimes increase the number of FFs. For instance, a (CSHM) [50] based FIR filter architecture requires more FFs in the precomputer output and in generating the delayed versions of input and output signals (Fig.3.23). In addition, in order to increase filter's performance (pipelining), it is customary to use flip flops between SSA units (Fig.3.23) and final adders. Intuitively, the additional number of flip flops reduces energy savings and results in area overhead. However, the proposed flip-flops presented in previous sections can potentially reduce power consumption taking advantage of low switching activity of such an application.

To prove the efficacy of the proposed flip flops, we replaced the conventional flip flops used in filter architecture by the proposed ones. Specifically, we obtained a synthesized Verilog netlist of a multiplier-less 8-bit FIR filter [69, 70] using Synopsys design compiler [71] and a 45nm STM library [72]. Then we imported the netlist in Cadence and replaced the conventional flip-flops with the proposed FFs. Using HSPICE we have simulated the FIR filter and results are shown in Table.3.3. It can be observed that using our proposed FF, power savings reach 42.4% compared to conventional SAFF based FIR. In addition, DD-SAFF shows 25.7% power improvement compared to low power ST FF.

3.4.3. DCT Design

One of the most popular blocks used in multimedia application is discrete cosine transform (DCT). DCT transforms an NxN image block from spatial domain to frequency domain and in matrix notation it can be expressed as: $Z=CXC^T$, where C is the coefficient matrix and X is the input NxN image block. Using row-column decomposition, DCT can be decomposed into two 1D-DCT units, which can be expressed as:

$$Z=C\,(X^T Z^T)^T \qquad (1)$$

In such an implementation, the 1D-DCT is applied to each row of input data and the result is transposed. A second 1D-DCT is applied to the rows of the transposed data to obtain DCT (eq. 1). The 1D-DCT output of a row in 8x8 block is expressed as:

$$w_k = \frac{c(k)}{2}\sum_{i=0}^{7} x_i \cos\frac{(2i+1)k\pi}{16},\ c(k)=\begin{cases}1/2 & k=0\\ 1 & k=1,\ldots 7\end{cases} \qquad (2)$$

Setting $c_k=cos(k\pi/16)$, $a=c_1$, $b=c_2$, $c=c_3$, $d=c_4$, $e=c_5$, $f=c_6$, $g=c_7$, the 1D-DCT can be written as [75]:

$$\begin{bmatrix} w_0 \\ w_2 \\ w_4 \\ w_6 \end{bmatrix} = \begin{bmatrix} d & d & d & d \\ b & f & -f & -b \\ d & -d & -d & d \\ f & -b & b & -f \end{bmatrix} \begin{bmatrix} x_0 + x_7 \\ x_1 + x_6 \\ x_2 + x_5 \\ x_3 + x_4 \end{bmatrix} \qquad (3)$$

$$\begin{bmatrix} w_1 \\ w_3 \\ w_5 \\ w_7 \end{bmatrix} = \begin{bmatrix} a & c & e & g \\ c & -g & -a & -e \\ e & -a & g & c \\ g & -e & c & -a \end{bmatrix} \begin{bmatrix} x_0 - x_7 \\ x_1 - x_6 \\ x_2 - x_5 \\ x_3 - x_4 \end{bmatrix} \qquad (4)$$

We can rearrange eqn. (3) in the following manner:

$$\begin{bmatrix} w_0 \\ w_2 \\ w_4 \\ w_6 \end{bmatrix} = (x_0+x_7)\begin{bmatrix} d \\ b \\ d \\ f \end{bmatrix} + (x_1+x_6)\begin{bmatrix} d \\ f \\ -d \\ -b \end{bmatrix} + (x_2+x_5)\begin{bmatrix} d \\ -f \\ -d \\ b \end{bmatrix} + (x_3+x_4)\begin{bmatrix} d \\ -b \\ d \\ -f \end{bmatrix} \qquad (5)$$

Eqn. (4) can be expressed in a similar format.

Interestingly, in such an application most of the input image energy (around 85% or more) is contained in the first 20 coefficients of the DCT matrix after the 2D-DCT operation [77]. This means that most of the coefficients within an 8x8 block will have small values (close to zero). The whole operation is depicted in Fig.3.24. Fig.3.24 highlights in the final DCT matrix the areas of high energy distribution and those of low energy in the final DCT matrix. High energy translates to high values and high switching activity as opposed to low energy and low switching outputs. It was noted for instance that in MPEG video compression application only 7 out of 64 coefficients [67] were not zero or close to zero for each 8x8 DCT output block. This actually, indicates that most of the

output DCT coefficients (at least the most significant bits) will have low switching activity, being most of the time close to zero. Such architecture makes for an excellent example in order to apply our data dependent flip-flops.

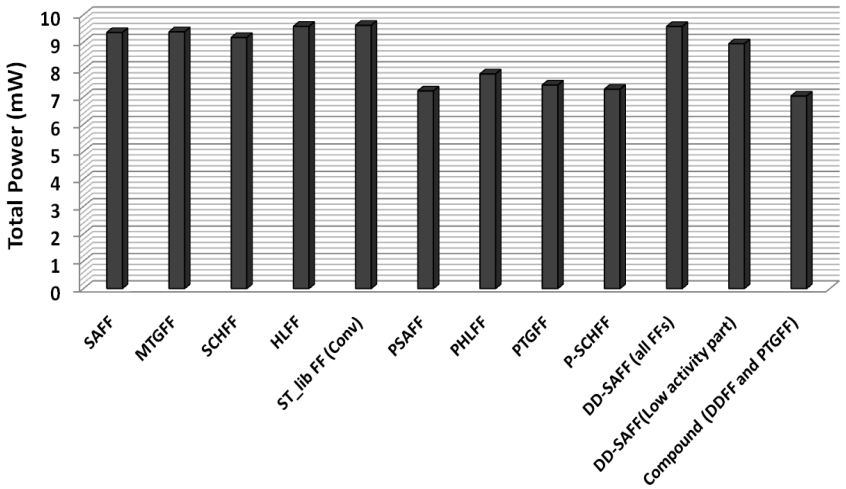


Fig.3.25. Power consumption result of DCT for different FFs

Specifically, we follow a preferential use of flip-flops. For outputs with low energy and thus low switching activity we use the data dependent flip-flops presented in section 3. On the other hand, for high energy outputs we use conventional flip-flops. More specifically, depending on the type of output bits we use different number of data dependent flip-flops. For instance, in case of z_5 to z_7 most of the bits/FFs are replaced with the proposed FFs. On the other hand, for all output bits of z_0 we use the conventional FFs and for z1 to z4 half of the MSB FFs are replaced with the proposed FFs. In addition, to further reduce the power we combine the data dependent flip-flops with pulse technique and the new pulse generation circuitry.

To prove the efficacy of the proposed approach we obtained a synthesized Verilog netlist of a DCT architecture using Synopsys design compiler [75] and 45nm STM libs. Then we imported the netlist in Cadence and replaced the conventional flip-flops with the proposed FFs in areas of low switching activity. Using HSPICE we have simulated the DCT and results are shown in Fig.3.25. It can be observed that using our proposed DDFF, the power saving is 7% compared to conventional SAFF based DCT. In addition, we can observe that by utilizing DD-SAFF in low switching parts of the DCT and the pulsed FF (e.g. PTGFF) for other parts, total power is reduced by 27%.

3.5. Conclusions

In this chapter several FFs that take advantage of data distribution in various applications have been presented. Specifically, we presented a data dependent FF that is able to limit significantly the charge/discharge of internal nodes of each FF and thus reduce power consumption. In addition we presented a low overhead pulse generation circuit that is able to reduce the clock power significantly. We followed a preferential usage of FFs in various applications depending on the type of computation/bit; high or low switching. By

utilizing the data-dependant flip-flop to a FIR filter design, total power consumption is reduced by 42% compared to using conventional flip-flop. Furthermore, by applying the proposed data dependent FF, combined with the pulsed approach resulted in more than 27% power reduction in DCT application. We believe that the proposed FFs can significantly reduce power in a wide range of DSP applications while effectively limiting any area and delay penalty. Furthermore, since flip-flops are one of the most important components in any design, such as portable designs and wireless sensor networks, lowering the power consumption of them helps to improve the total power of wireless sensor nodes significantly.

Chapter 4

Ultra Low Power SRAM Design

Chapter 4

Ultra Low Power SRAM Design

Introduction

Continued increase in the process variability is perceived to be a major challenge to future technology scaling. Process variability stems from systematic effects such as variations in critical dimensions and oxide thickness, as well truly random effects, like the dopant fluctuations. Current design methodology hardly distinguishes systematic variations from truly random ones. Commonly the entire process variability is lumped together, and included in process corners. Traditionally, integrated circuit designs capture the impact of variability by satisfying the design constraints at various process corners, where the process corners are the extreme deviations of the process parameters from their typical values. For digital logic design, the worst-case corners typically capture 3σ. To satisfy the worst-case performance requirements, often a large penalty is paid in power. Also, most of the today's designs are power limited. In such cases, satisfying the power budget often requires a back off in performance.

Memory design presents an extreme example of corner-based design. To satisfy the functionality of several tens of millions of SRAM cells, the designer has to capture 5 to 6 standard deviations of parameter variations. This is becoming increasingly challenging to satisfy, and may present a problem for continued scaling of memory density. Concurrently, high-end microprocessors have been increasing the amount of on-die cache to improve the performance.

Presently, to achieve highest possible densities with high parametric yields in bulk-CMOS and SOI technologies, designers use a combination of multi-layered ad-hoc techniques and heuristics, which include device sizing, supply and threshold voltage selection, SRAM column height and sense-amplifier optimization, and redundancy and error correction techniques.

Problems arising from technology scaling have recently made power reduction an important design issue for circuits and applications that traditionally are driven almost solely by performance constraints. The increased importance of power is even more notable for a new class of energy-constrained systems [76, 77]. Supply voltage scaling has been proposed for low-voltage/low-power digital designs. The leakage is reduced due to smaller voltage differences between the drain, source and body of a transistor [78, 79].
Scaling the supply voltage decreases the power consumption, but increases the delay of circuit. Therefore the threshold voltage are reduces to reduce the circuit delay. In order to satisfy the performance requirements demanded by applications, like wireless sensor networks (WSN), the threshold voltage (V_{th}) should also be lowered, to have both low power operation and satisfactory performance. However, there is a cost of higher static power dissipation due to large leakage currents [76].

Subthreshold circuit design involves scaling the supply voltage below the threshold voltage, where load capacitances are charged / discharged by subthreshold currents. In subthreshold region, the active current is defined by:

$$I_{sub} = I_0 e^{\frac{V_{GS}-V_T+\eta V_{DS}}{nV_{th}}}\left(1 - e^{\frac{-V_{DS}}{V_{th}}}\right) \quad (4.1)$$

$$I_0 = \mu_0 C_{ox} \frac{W}{L}(n-1)V_{th}^2 \quad (4.2)$$

Where n is the subthreshold slope factor ($1+C_d/C_{ox}$), and V_{th} is kT/q. So, the propagation delays equation with output capacitance C_g in subthreshold is [80]:

$$t_d = \frac{KC_g V_{DD}}{I_0 e^{(V_{GS}-V_T)/nV_{th}}} \quad (4.3)$$

where K is the delay fitting parameter. As shown by Eq.3, the speed of circuits operated in subthreshold region is proportional to the supply voltage. When $V_{GS}=V_{th}$, the propagation delay has an inverse dependence to I_0, so by changing temperature, device sizing, V_{GS}, and the physical parameters, the propagation delay changes. Eq.1 shows that lowering V_{DD} will produce a proportional reduction in subthreshold current degrading the performance. At very low V_{DS} values, the parenthetical term produce a more pronounced roll-off in subthreshold current [80].

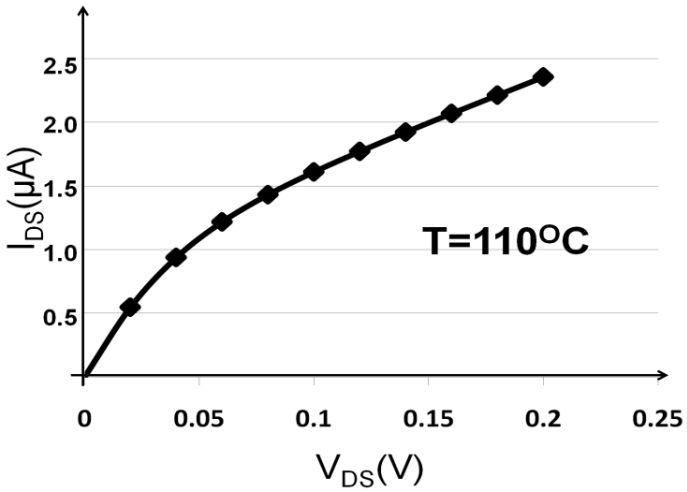


Fig.4.1. IDS versus VDS for single NMOS transistor in 65nm

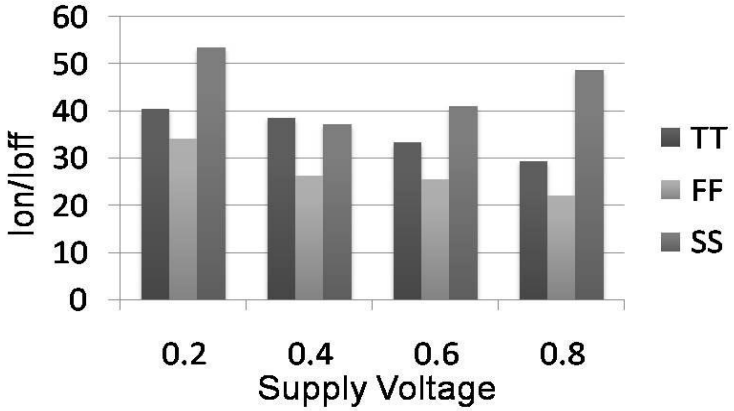


Fig.4.2. I_{on}/I_{off} ratio for different supply voltages

The most important concern in subthreshold region is the low ON current (referred as Ion). To increase the subthreshold current,

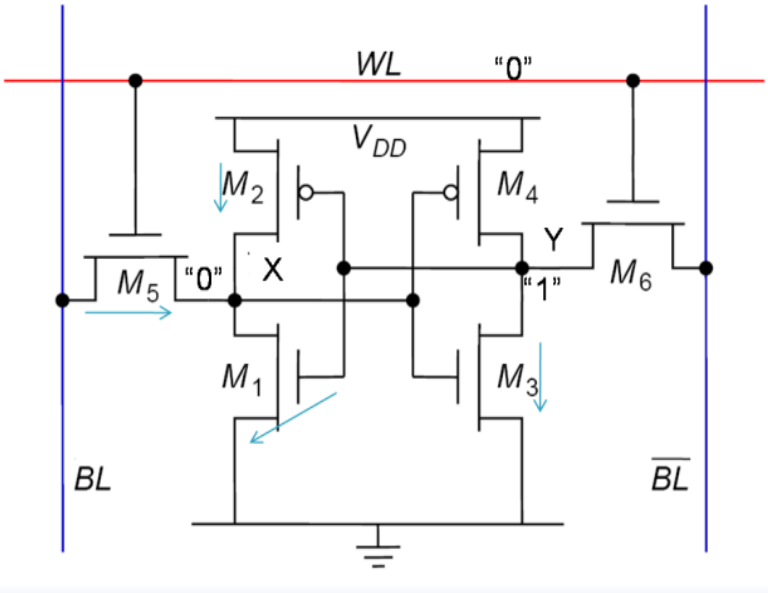


Fig.4.3. 6T SRAM cell

upsizing is one option. Upsizing the transistors will however increase the parasitic capacitances and in some cases (like for SRAM) cause failures [79]. Fig.4.1 shows Ion (V_{GS}=V_{DD}) and I_{off} (when VGS=0) for different processes corners. As illustrated in Fig.4.2, the ratio between I_{on} and I_{off} is small for such a low supply voltage. In the 65nm technology an unusual behavior is observed for specific supply voltage ranges. For minimum length devices, at low supply voltages, the PMOS current level is higher than NMOS current (Process dependant, STM 65nm). Due to the significant reduction in evaluation current in both NMOS and PMOS, subthreshold circuits are usually impractical in high speed applications. However for some applications such as wireless sensor networks (WSN), it is not necessary to have a high sampling frequency. In another words, the most important concern for WSN applications is to minimize the power consumption to increase the

battery lifetime. As a result, the frequency must be reduced to get correct evaluation. Lower ON current (I_{on}) causes some problems in SRAM design. SRAMs comprise a significant percentage of the total area for many digital systems as well as the total power consumption. Therefore SRAM leakage can dominate the total leakage of a chip. By reducing SRAM leakage, total leakage power of a system decreases. Pushing SRAM in subthreshold region reduces the leakage power significantly. However, it degrades the SNM of SRAM in read and write cycles due to lower read current and also the effect of BDGO (breakdown gate oxide) of NMOS transistors in SRAM circuit.

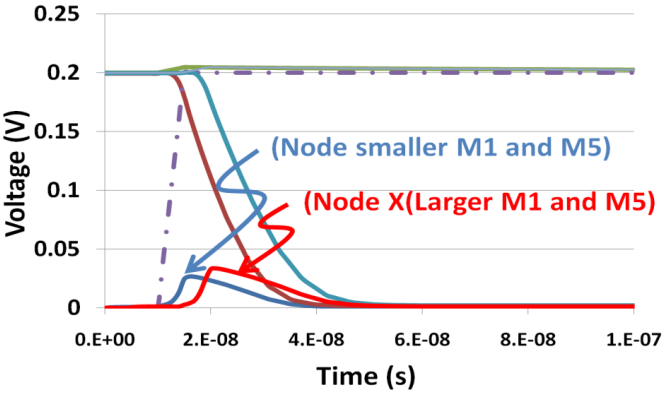


Fig.4.4. Output waveforms for 6T-SRAM cell

4.1. Conventional SRAM Design

Fig.4.3 shows the standard 6T-SRAM cell. In this circuit, storage nodes are specified by X, Y. Suppose that node X stores "0" and node Y stores "1". Due to low supply voltage, leakage sources are important. The leakage sources are shown in Fig.4.3. In this case, M1, M4 are turned on. Furthermore, M2 and M3 are turned off. During the hold time, when WL is not asserted (idle mode), M5 and M6 are off. In idle mode, M5 and M2 give a small rise on node X because of leakage currents, therefore node X has a low voltage in range of few mV (instead of zero) that causes an increase in leakage through M3 introducing failures [81]. When WL is selected (Read cycle) M6 is turned off and M5 is turned on. In this case, the read cycle is done through M5 and M1.However a rise at node X due to the stacking effect [82] causes more leakage through M3. This causes a drop in the voltage of node Y, therefore the read cycle speed is degraded results in flipping the data on the storage nodes. To improve the speed of the read cycle, the best way is upsizing M1 and M5. However there are some limitations on increasing sizes of transistors. Fig.4.4 shows the effects of upsizing on read noise margin in 6T-SRAM cell. As it can be seen from Fig.4.4, by increasing the ratio of WM1/WM5, discharging BL is faster, but due to the upsizing of these devices the voltage bump at node X increases which increases the leakage current through M3 and discharges node Y more that causes increasing leakage through PMOS that is higher than NMOS in 65nm CMOS technology (ST 65nm) for ultra low supply voltages.

During write cycle, when WL=1, storage nodes are discharged through the write path that include three NMOS stacked transistors including NMOS pass transistor in 6T-SRAM cell. Fig.4.5 shows the writing path of 6T-SRAM cell. As it is shown, the write cycle speed is dependent to these three stacked NMOS. When Data=0, Write=1, and WL=1, then node x is being discharged through M7, M9 and NMOS access transistor (M5). In this case, suppose that X holds "1", therefore, storage node X is discharged via stacked NMOS's.

this case there is a contention between NMOS transistors in stack and PMOS transistor of cross inverter (M2) in SRAM cell. PMOS transistor (Pull-up transistor) is trying to hold the stored data on node X. Therefore, sizing the NMOS transistors plays an important role in write speed. For high enough supply voltages, using small NMOS transistors satisfies write time. However, for lower supply voltage or sub-threshold design, M7-M10 should be upsized significantly.

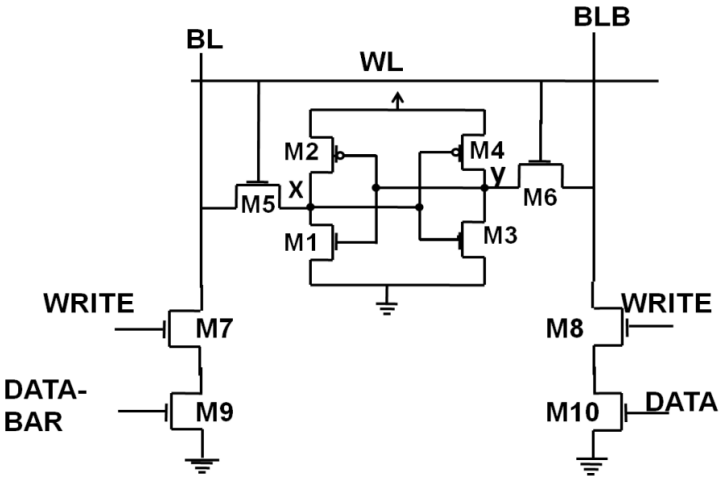


Fig.4.5. Write mode circuitry in 6T-SRAM cell

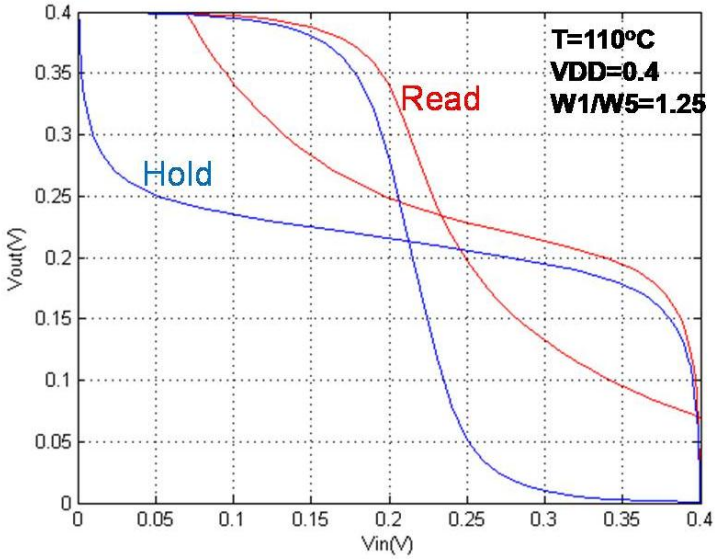


Fig.4.6. SNM for HOLD and READ in 6T-SRAM (V_{DD}=0.4)

Another concept that is used to evaluate the SRAM is read static noise margin (SNM) [83, 84]. SNM quantifies the amount of voltage noise required at the internal nodes of a bitcell to flip the contents of the cell [80, 85]. In 65nm, due to the different behavior in sub-Vt (PMOS is stronger than NMOS by an order of magnitude with minimum length) SNM is more affected. To balance the devices in an inverter, the NMOS must be upsized by an order of magnitude larger than PMOS. This problem is more challenging during the write cycle where stronger access NMOS transistors are needed.

At the onset of a read access, the wordline is "1" and the bitlines are precharged to "1". The internal node of the bitcell that represents a zero gets pulled upwards through the access transistor due to the voltage dividing effect across the access transistor and drive transistor, which degrades the read SNM. Fig.4.6 shows the butterfly curves during the hold and read cycles for V_{DD}=0.4V illustrating the degradation in SNM during read. As illustrated, the read noise margin is degraded at low supply voltages. Due to increased leakage through the transistors, especially PMOS, at higher temperatures,

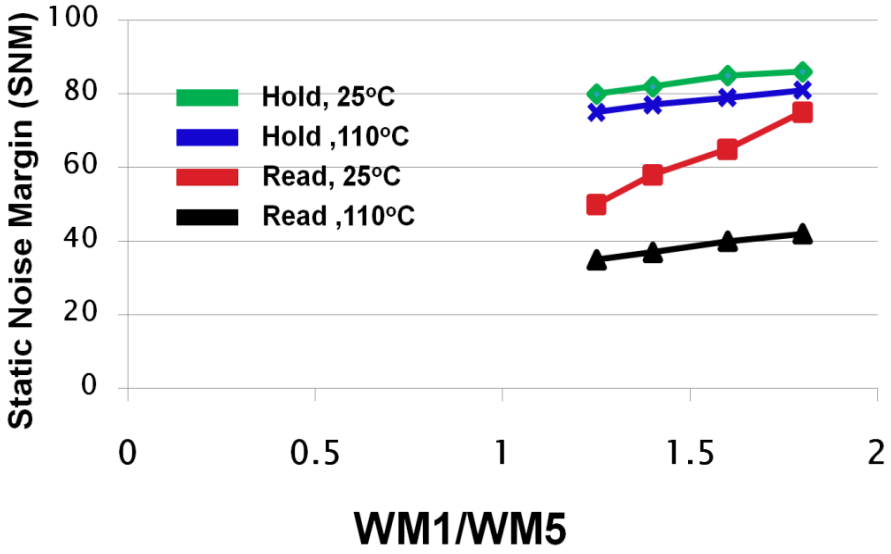


Fig.4.7. SNM vs. WM1/WM5 for different temperatures.

read noise margin is more decreased. Fig.4.7 shows SNM versus WM1/WM5 for different temperatures (27°C and 110°C) for TT CMOS models. To improve the SNM in subthreshold region, an 11T-SRAM design is proposed. In this circuit, the read bit-line is separated from the write wordline as proposed in [86].

4.2. Proposed Ultra Low Power SRAM Cell

In this section several topologies of SRAM topologies are proposed showing ultra low power consumption. Furthermore, the proposed SRAM cells are able to operate in ultra scaled supply voltage less than 0.3V.

4.2.1.Subthreshold 11T-SRAM Cell

Fig.4.8 shows the schematic of the proposed 11TSRAM bitcell. Transistors M2, M4, M5, and M6 are identical to 6T-SRAM, but two transistors M1 and M3 are downsized to the same size as the PMOS transistors. The read bitline and wordline are separated from the write wordline. In this case, memory can have distinct read and write ports. During the hold time, RDWL and WL are not selected. In the 6T-SRAM part, suppose that node Y stores "0" and node X stores "1" as was described for the 6T-SRAM part in the previous section. For the added circuitry the following behavior is observed when transistor M12 is turned off. Furthermore, M11 state is dependant to the voltage at node Y. If node Y stores "1", then M8 connects gate of M11 to ground then M11 turns off. However when node Y stores "0", M9 turns off and it starts to charge the gate of transistor M11. Therefore there is a leakage path through M11 that connects the node YN to zero. Minimum size transistors were used for the added 5T-circuitry, except the access transistor that has a larger size. The

most important part of the 11T-SRAM is a boosting capacitor (CB) that connects source of M9 to RDWL.

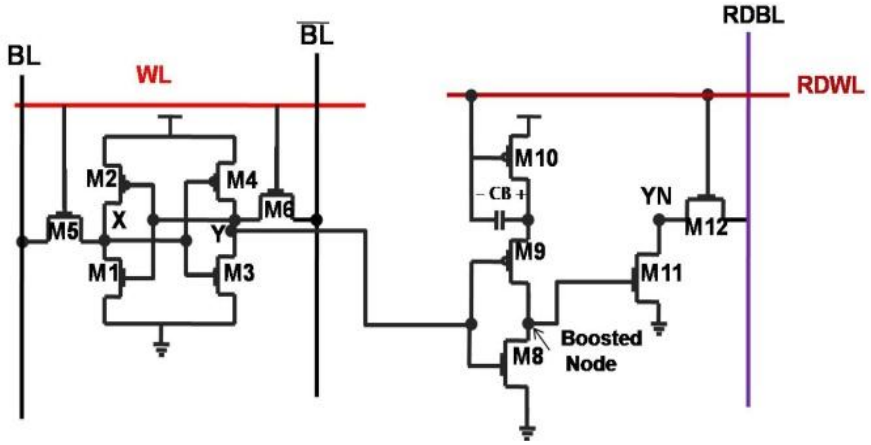


Fig.4.8. Proposed 11T-SRAM cell

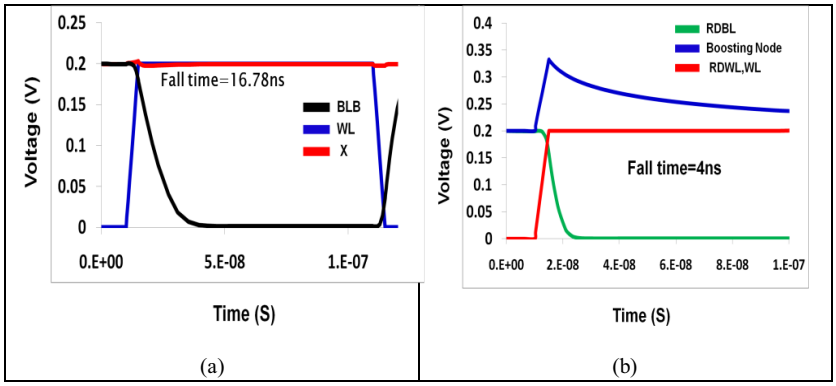


Fig.4.9. Output waveforms during read a. 6T-SRAM b. 11T-SRAM

Assuming node Y has stored "0". During the hold and write cycles, RDWL is connected to zero, then CB starts to be charged to V_{DD} but transistor M9 does not let CB being charged to V_{DD}. Simulations show that it is charged to $V_{DD}/2$. Thus the gate of M11 is connected to a voltage almost less than $V_{DD}/2$ discharging the node YN to ground slowly. When RDWL is selected, the source of M9 rises to $1.5V_{DD}$ connecting the gate of M11 to a voltage higher than V_{DD}. This increases the read current by an order of magnitude compared with the 6T-SRAM cell. Fig.4.9 shows the output waveforms of the 11TSRAM and the 6T-SRAM. As illustrated, the read time is decreased by more than 4X compared to 6T-SRAM cell.

In other hand, when node Y stores "1", M8 is turned on and connects the boosted node to ground then M11 turns off causes the read path disconnecting from ground. However leakage through M11 causes some reduction in RDBL voltage. Furthermore during this time, M9 is turned off, while its source is connected to high voltage, it will be leaky transistor. Since the drain of M9 is connected to ground via M8, this leakage has no effect on YN or RDBL,

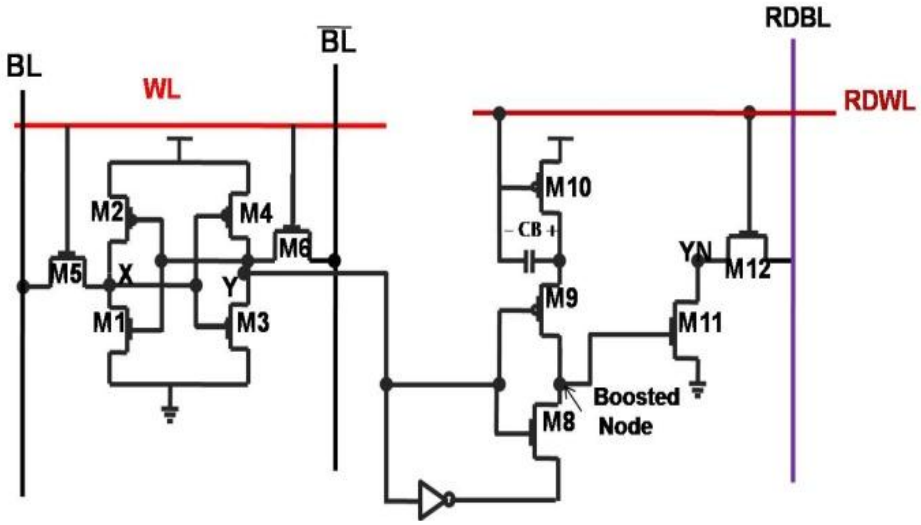


Fig.4.10. Modified 11T-SRAM (14T-SRAM)

However there is a leaky path during the read cycle, when Y has holds "0". In this case, the drain of M8 is at a high voltage more than $1.4V_{DD}$ causes more channel leakage through M8 and also gate leakage in M11. To suppress the channel leakage path via M8, a modified topology of SRAM cell is proposed that is shown in Fig.4.10. When Y is high the source of M8 is connected to ground through the added inverter, but when Y stores "0", the source of M8 is connected to V_{DD}, which suppresses the leakage path in M8 during the read cycle.
Fig.4.11 shows the simulation results for the proposed circuit for SNM in read and hold cycles. As it can be seen, SNM of the 11T circuit in hold and read mode is the same, which is due to higher read current of the proposed circuit. The circuit was also simulated for a temperature of 110^{o}C with a supply voltage V_{DD}=0.2. These results were compared to the results with a temperature of 27^{o}C in Fig.4.12 which shows that the SNM at higher temperature is lower. This is because of the many leakage sources and also dependence of current to temperature. Fig.4.13 shows read noise margin for 11T-SRAM in different process corners.

As mentioned before, the PMOS devices have a higher current than NMOS devices. This causes some problems in write mode. Therefore the circuit was redesigned for write mode using larger devices. The access NMOS transistors must be larger to dominate the current via PMOS transistors (M2, M4) to write a "0" in storage nodes. However there are some limitations on upsizing transistors. In this case (write mode) the sizes of M2 and M4 must be decreased while M5 and M6 must be increased to have a correct write with at high speed. Still the speed in write mode is the same as for 6T SRAM cell, so by working at a 0.25V supply voltage results in a very low speed during write mode. However, this problem is eliminated by increasing the supply voltage to a voltage more than 0.3V. In this case, both circuits are running at a satisfactory speed with a proper write noise margin.

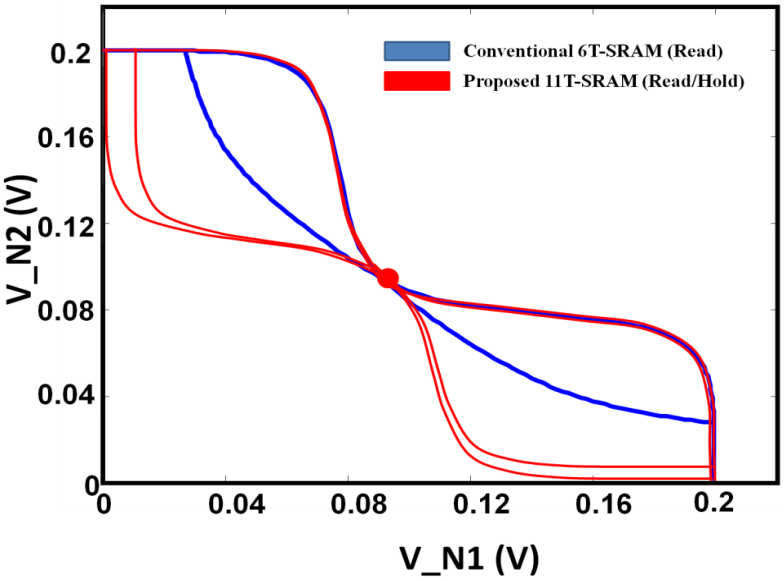


Fig.4.11. Butterfly plot for 11T-SRAM

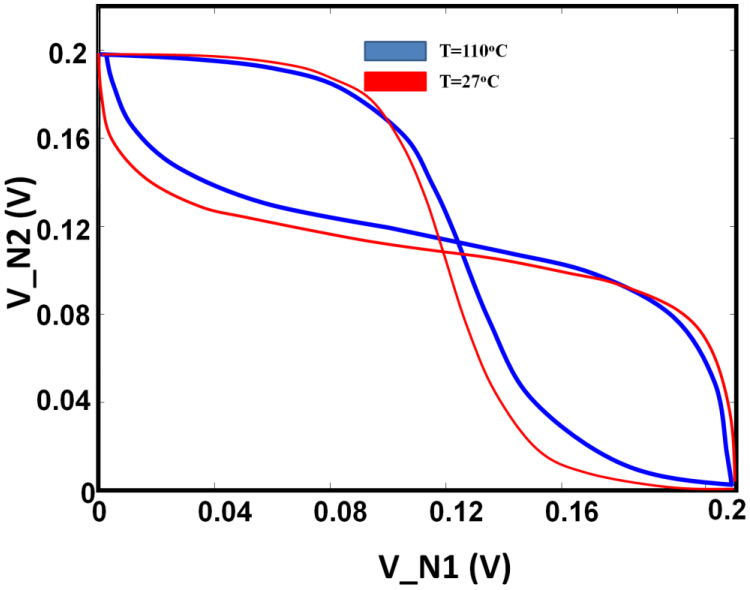


Fig.4.12. Butterfly plot for 11T-SRAM in different temperatures

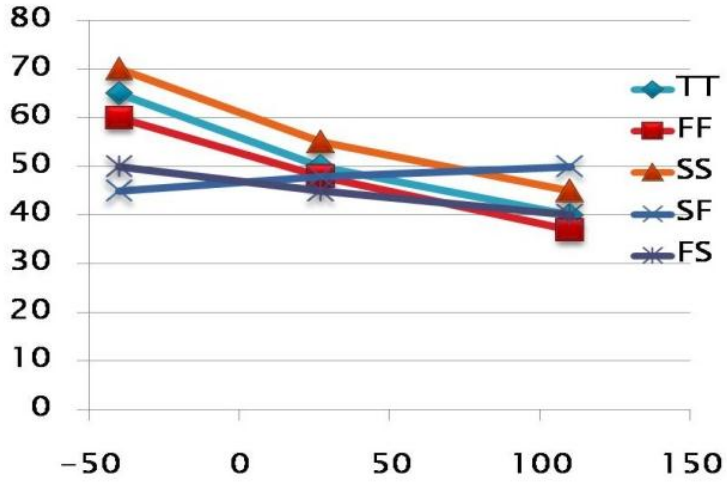


Fig.4.13. read noise margin for 11T-SRAM

The effective area overhead for the proposed circuit is typically between 22%-28%, but due to employing minimum size devices and lowering the sizes of PMOS devices in 6T-SRAM and also downsizing the NMSO transistors in 6T-SRAM cell, the area overhead may be reduced. The SNM is significantly increased (more than 6X in some cases compared with 6T-SRAM) and also simulations show that the speed of the proposed circuit is 4 times higher than the 6T-SRAM cell. Table 4.1 shows a comparison of the read noise margin between this work and the 10T-SRAM in [86]. As it can be seen the read noise margin of the 11T-SRAM is higher than SNM for SRAM in [86]. Since both topologies rely on the same 6T-circuit for the write cycle, the write noise margins are comparable.

An 11T-SRAM cell with higher SNM and higher speed operated at a supply voltage 0.2V was presented. In this circuit separate lines for read and write were used. Compared to the 6T-SRAM cell higher speed, higher SNM and lower power consumption were demonstrated through simulations at a cost of increased area. Furthermore, to further decrease the off current compared to 6T-SRAM, an additional inverter was added to 11TSRAM. The modified circuit shows a better SNM in read mode compared to the 6T-SRAM and the work reported in [86], maintaining the same SNM in write mode.

4.2.2. Subthreshold PMOS-Access Transistor SRAM (PAT-SRAM)

In [87] it is shown that for lower supply voltages, PMOS transistors are faster than NMOS devices. Therefore, in this case, it is shown that by using PMOS transistors to implement the write path, the write speed is improved. The proposed circuitry is shown in Fig.4.14. As it can be seen in Fig.4.14, in write cycle, when write=1, DATA=1, and WWL=0 (RDWL=0), node x is charged to V_{DD} through three stacked PMOS transistors. When WRITE=1, then the circuit is in the write mode and depending the DATA value, the storage nodes values are changed. If WRITE=0, then the circuit is in read mode and the BL and BLB are changed. Suppose that the circuit has been in hold mode, so the BL and BLB are at zero value. In Read cycle, when WWL=0, if node x stores “1”, then BL is charged to “1” through M5 and the value of BLB is not changed and is fixed at zero. Therefore, in this case, due to higher speed of evaluation for PMOS transistors in sub-threshold mode, read and write cycle speed are improved. Since upsizing the stacked PMOS transistors is possible, supply voltage can be lowered. In 6T-SRAM standard cell, this is not possible,

because the write cycle is not completed if we go in lower supply voltages. Therefore, we have to use very huge transistors to discharge storage nodes in write cycle. There are some ways to obviate this problem in write margin. One of these techniques is using sleep transistor connecting supply connecting sources of transistors M3 and M4 to V_{DD}. However, this technique comes with area penalty and degrades the reliability problems due to PMOS transistor between V_{DD} rail and M3 and M4.

Table4.1. Read noise margin comparison.

Process Corner	Temperature (°C)	Read Noise Margin (mV) [86]	Read Noise Margin (mV) 11T-SRAM
TT	-40	62	85
	27	57	76
	110	52	62
FF	-40	61	70
	27	54	61
	110	51	48
FS	-40	50	70
	27	45	63
	110	38	60
SF	-40	64	62
	27	59	67
	110	52	71
SS	-40	57	81
	27	55	72
	110	47	54

Another configuration to implement an SRAM cell is using separate read_WL and write_WL. Simulation results show that this configuration has a much higher speed in write cycle compared to conventional SRAM cell write circuitry.

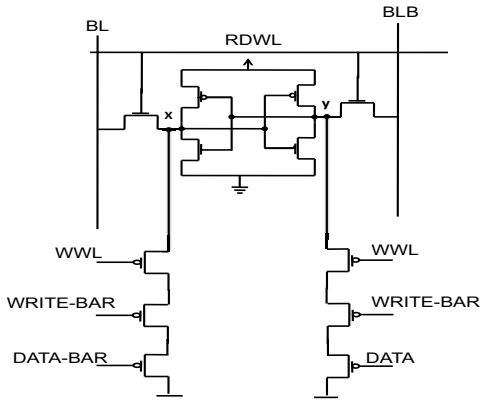


Fig.4.14. Configuration of modified write path in 6T-SRAM cell

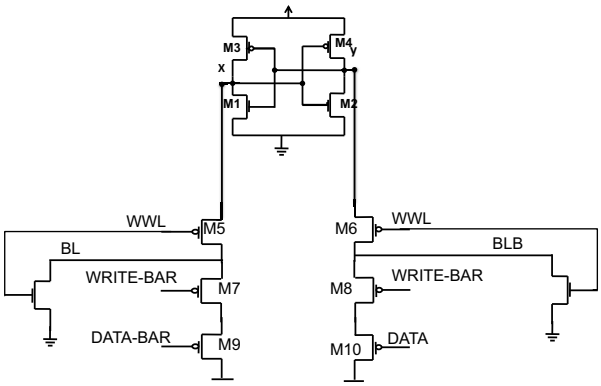


Fig.4.15. Proposed SRAM circuit

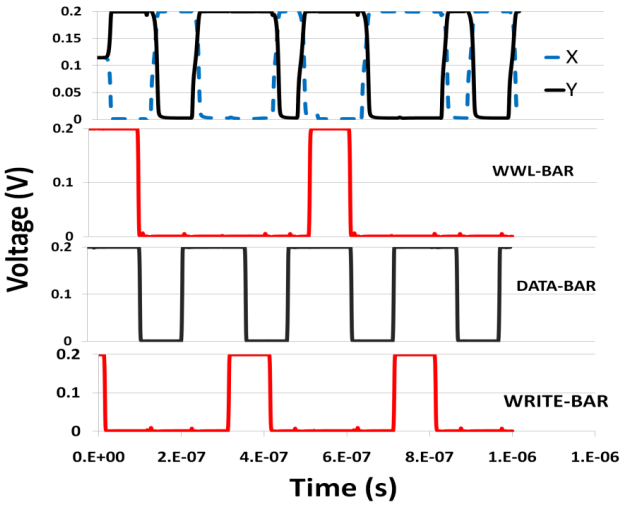


Fig.4.16. Curves of proposed SRAM cell

Fig.4.15 shows the configuration of this circuit. In this circuit NMOS M5 and M6 are replaced with PMOS access transistors. In this case, we use two NMOS transistors to discharge BL and BLB to ground when WWL=1. The Proposed SRAM works as follow: In hold mode, when WWL=1, M5 and M6 are turned off and separate the storage nodes from bit lines. During the hold cycle, BL and BLB are discharged to ground. In read cycle, when WWL=0 and WRITEBAR="1", depending the values on storage nodes BL and BLB are charged to X and Y values. suppose that X="1", Y="0", then BL is charged to "1", through M5 and this turns on M2 to discharge the Y node value. In this case, M6 turns off and separates the Y from the BLB value. Therefore, there is no way for BLB to be charged. There is only a leakage path via M6 helping node Y to be lowered during this mode. Then the BL and BLB lines are sensed through a sense amplifier to read the bit stored on bit-cell. For read mode, because PMOS devices are faster than NMOS devices in the sub-threshold design, read speed is improved. However it depends on the sizes for M5 and M6.

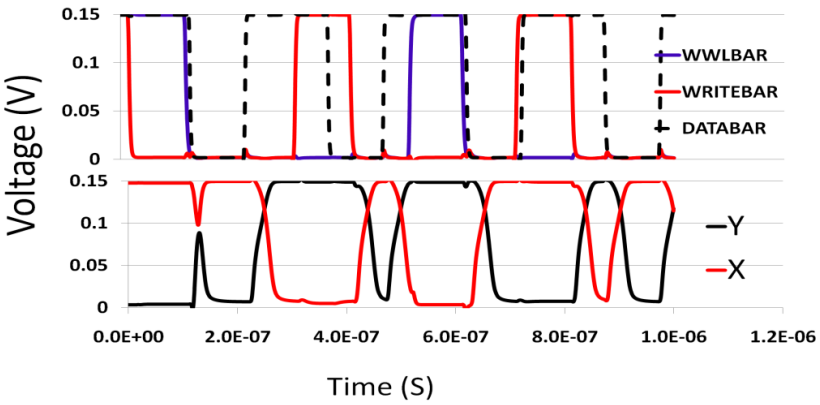


Fig.4.17. Waveforms of proposed circuit (Write cycle)

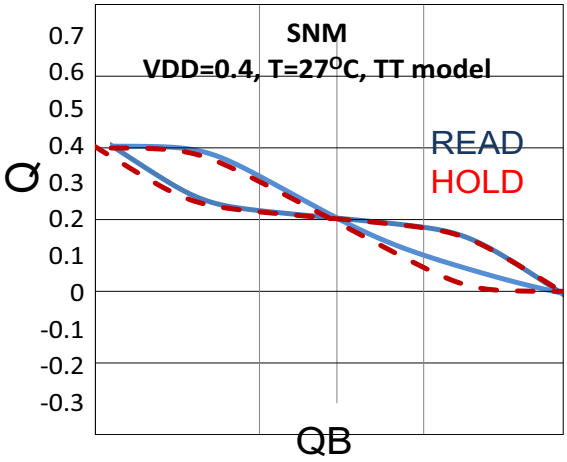


Fig.4.18. Butterfly curves for 6T-SRAM cell

During the write mode, as it can be seen in Fig.4.16, when WWL="0", WRITEBAR="0", the values of input data are written on storage nodes through PMOS transistors. In this case there are three stacked PMOS transistors that have a much higher speed than three stacked NMOS transistors. Therefore, there is no necessity not to use sleep transistor to separate M3 and M4 from supply voltage rail during the write cycle. As an example, for the proposed circuit, write circuit (two stacked PMOS's) are sized W_P=0.8um, but for NMOS stacked devices to get the same speed for write cycle, the NMOS transistors are sized W=2um. So it shows it is possible to get the lower area using PMOS transistors in sub threshold design. As you can see in Fig.4.17, the circuit operates well.

4.2.2.1. Simulation Results

Static noise margin are used as a metric to show the stability of an SRAM. To find the SNM in read and hold mode, butterfly approach is used [88]. Fig.4.18 shows the butterfly curve for a 6T-SRAM cell at V_{DD}=0.4V. By lowering the supply voltage in submicron technologies, SNM is degraded due to the process variations [89]. As it can be seen the

butterfly curve is degraded by lowering the supply voltage due to the lower drive current and process variations in submicron technology.

To find the write static noise margin some metrics are used [90-94]. In this chapter, the most common approach is used to calculate the SNM as a criterion [91] that is shown in Fig.4.19. To avoid the failure, the curves should have just one cross point that indicates the mono-stability of the cell. As an example the Write SNM for 6T-SRAM standard sell for TT model in T=27°C is 0.15V that shows there is no

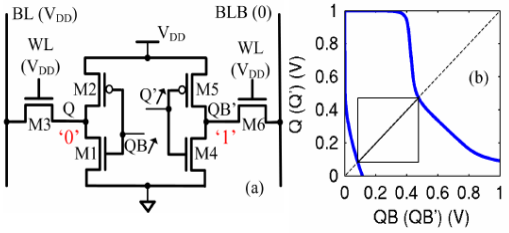


Fig.4.19. Write Margin calculation [90]

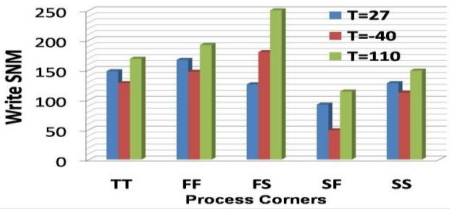


Fig.4.20. Write SNM for different temperature (at different corners)

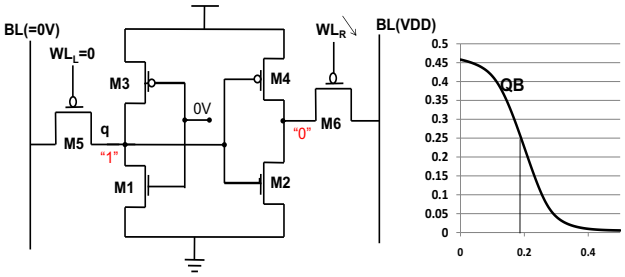


Fig.4.21. Circuit for sweeping WL_R to get WSNM

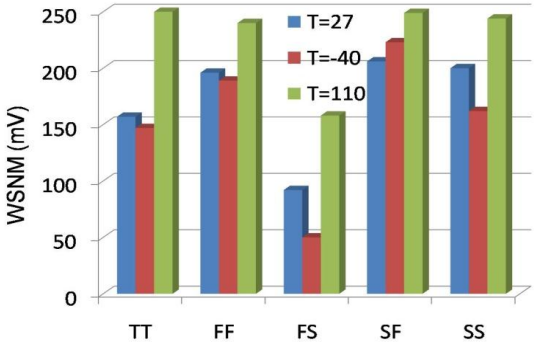


Fig.4.22. Write SNM for proposed circuit

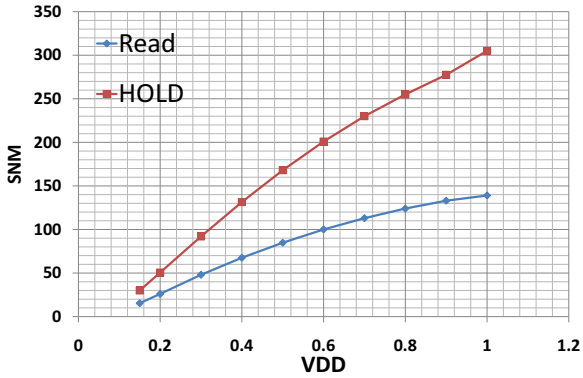


Fig.4.23. SNM versus V_{DD}

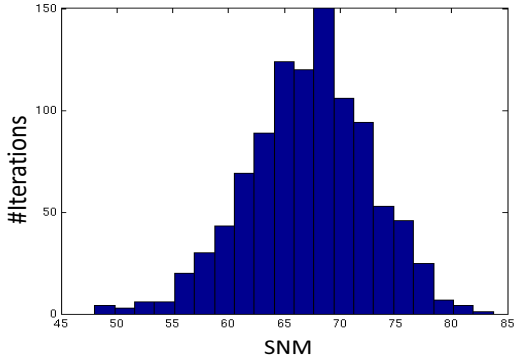


Fig.4.24. Monte Carlo simulation for read SNM for 6T-SRAM standard cell

failure. Fig.4.20, shows the results for WSNM for V_{DD}=0.5V for 6T-SRAM cell. In this case the method in [91] has been used to see the difference between results.

The topology of benchmark to find the write SNM is the topology used in [92] as illustrated in Fig.4.21. The results of write margin for proposed SRAM are illustrated in Fig.4.22. As it is shown, the write margin is improved by 47% compared to the conventional 6T-SRAM cell.

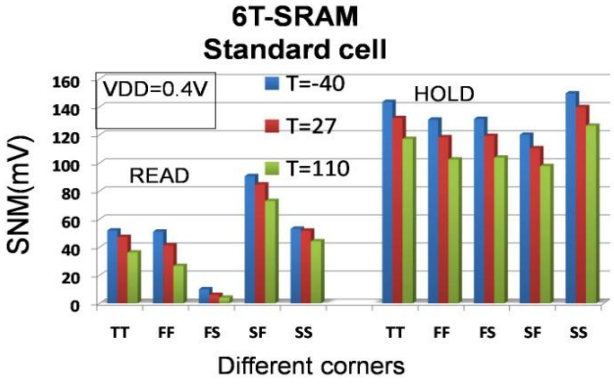


Fig.4.25. SNM in read and HOLD mode for proposed SRAM cell

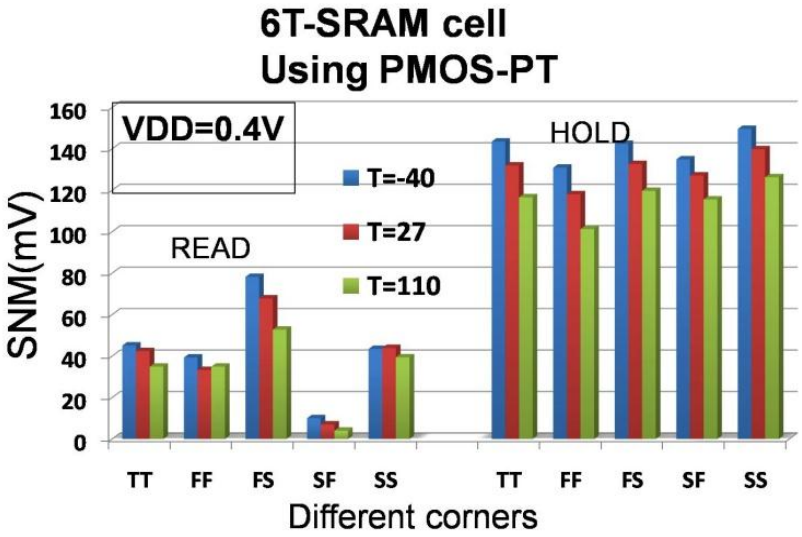


Fig.4.26. SNM in read and HOLD mode for proposed SRAM cell

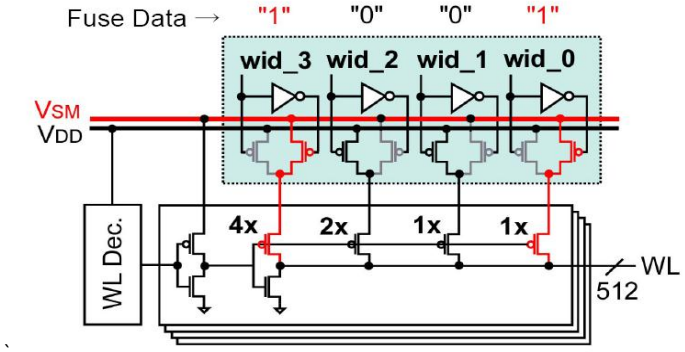


Fig.4.27. read assist circuitry [110]

Fig.4.23 shows the SNM for both read and hold mode for a 6T-SRAM cell. As it can be seen, at very low supply voltages the noise margin is low. To see the effect of process variations on SNM, Monte Carlo simulation is used. Fig.4.24 shows the Monte-Carlo simulation of SNM for a 6T-SRAM cell. As it can be seen, σ is around 5.6mV. The data have been acquired using Monte-Carlo simulation based on mismatch and process variations for a 6T-SRAM standard cell.

The effect of temperature on SNM in read and hold modes is shown in Fig.4.25. As it can be seen, the effect of temperature is not significant. The maximum change in SNM due to the temperature is 18mV in read cycle, and 27mV for Hold mode for a 6T-SRAM cell. As it can be seen in Fig.4.26, using PMOS transistor as a pass transistor (W/L=0.2/0.06 compared to W/L=0.4/0.06 for 6T-SRAM standard cell), shows good enough results compared to a standard SRAM cell. The hold SNM is improved due to the smaller PMOS pass-transistor and lower leakage current for proposed SRAM cell.

In this section, we showed that using PMOS transistors as a pass transistor and also in write path is much better compared to using NMOS transistors. Results showed significant

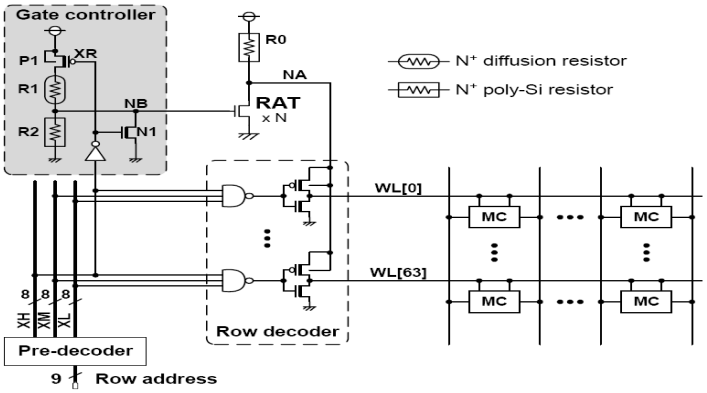


Fig.4:28. RAT scheme [96]

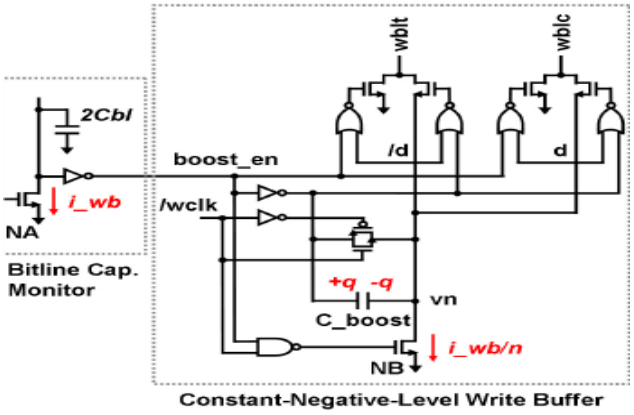


Fig.4.29. LPWD-SS [101]

improvement in write SNM with a lower area compared to standard 6T-SRAM cell. We showed that using proposed circuit the WRITE static noise margin is improved by around 50% for TT CMOS model. Simulation results showed the area overhead of proposed circuit is lower than 6T-SRAM standard cell due to using smaller pass transistors.

4.3. SRAM Design, Architecture Level

Several SRAM cell topologies were proposed that enables ultra low voltage scaling. However all the proposed SRAMs at cell level comes with a penalty in area (2X-3X larger area overhead). To cope with area, a new wordline driver is proposed that enables working in scaled supply voltage along with significant improvement in robustness.

In this section, we propose a technique with the following features: (a) Boosted voltage level to $V_{DD}+\alpha$ where α depends on the transistors sizing in wordline driver circuit, and boosting capacitance in wordline circuit. (b) Fixed voltage at V_{DD} for a short period of time from the start of the read access and then lowered voltage to around $V_{DD}/2$ for the rest of read mode. This improves the read noise margin. (c) Negative level of wordline voltage during standby mode to reduce bitline leakage.

4.3.1. Wordline Drives for SRAM Arrays

In literature, different topologies have been proposed to improve read and write static noise margins (SNM) [95-111]. In this work we consider the most recently proposed techniques that are relevant to wordline driving. In this section we briefly present some of them.

Fig.4.27 shows the schematic of Level-programmable Wordline Driver (LPWD) proposed by Hirabayashi et al. in [110]. They proposed a wordline compensation technique combined with a dual power supply scheme. As shown in Fig.4.27, there are two global power supplies named V_{SM} and V_{DD}. The authors reported that V_{SM} is 200mV higher than V_{DD} based on their measurements on a 512KB SRAM block. This technique uses an adaptive wordline (WL) level-control generated from dual power supplies in the WL

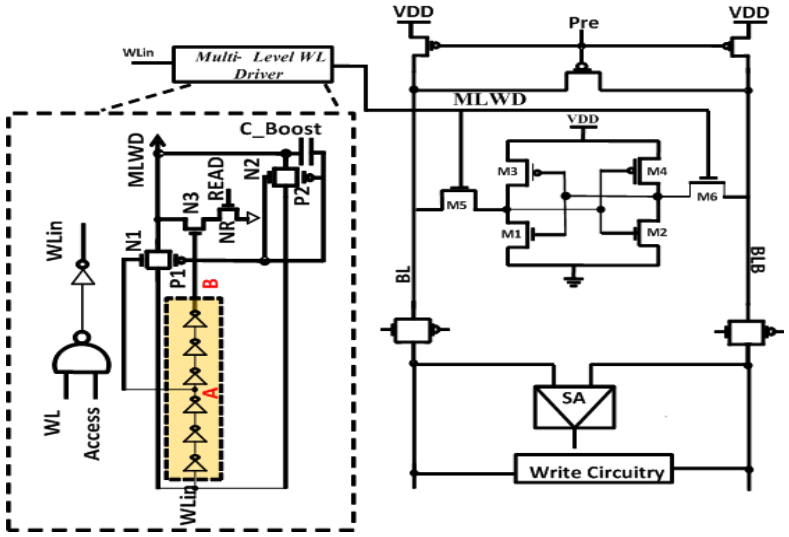


Fig.4.30. MLWD circuit topology

driver. In LPWD, the WL pull-up PMOS transistors are split in a binary manner. Even though the cell failure is reduced by 1000X, the usage of large PMOS transistors leads to large area penalty. Besides, this technique uses dual power supplies.

A replica access transistor (RAT) is proposed in [106] that self-calibrates the WL voltage suppression under dynamic control voltage and frequency scaling. The schematic of the design is shown in Fig.4.28. The technique reduces the cell current by 83% compared to conventional assist circuits [106]. Furthermore, the minimum operating voltage in the worst case was improved by 170 mV, confirming a high immunity against process and temperature variations with less than 10% area overhead. However, since the WL voltage is lowered, there is a concern about degradation of the operating speed. Furthermore, this technique uses resistances as voltage-divider (implemented by N^+ poly-Si) that imposes area penalty.

In [106, 107], the write capability is enhanced by negative write biasing without any reduction in the cell current Furthermore, read capability is enhanced by cell current boosting. This technique uses a negative voltage booster to boost the selected column's V_{SS} to a negative voltage to enhance the access current. However, in this technique the cells

are 8T-SRAM cells with separate read and write schemes. Therefore, these techniques suffer from large area and power penalties compared to 6T-SRAM array.

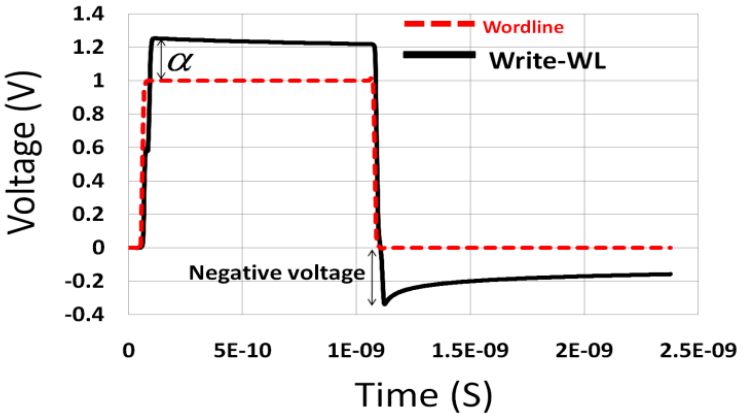


Fig.4.31. Output of MLVD during the write

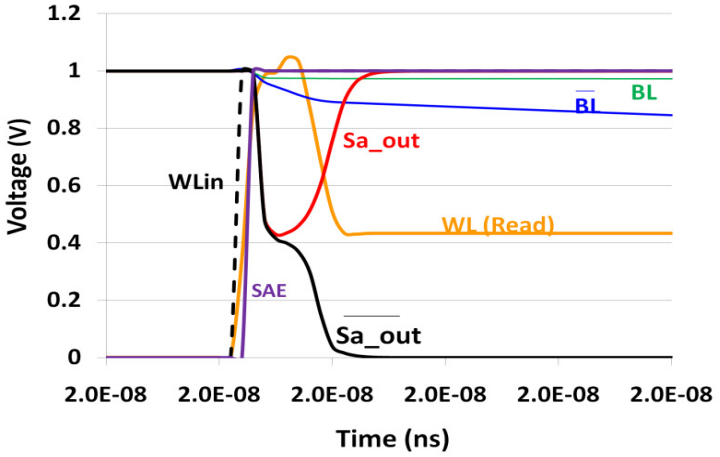


Fig.4.32. MLWL outputs during READ

In [108], capacitive coupling is used to generate a transient negative voltage at the low-going bit-line during Write operation without using any on-chip or off-chip negative voltage source. In this technique, bitline voltage is lowered to a negative level using a capacitance, and then it gets back to ground after a certain time depending on the design. This technique shows 1000X improvement in Write failures with no impact on the read stability.

In [112], the authors proposed a Level-programmable Wordline Driver for Single Supply (LPWD-SS) design, shown in Fig.4.29. This design uses a fixed negative voltage using boosting capacitance that gives better writeability to the SRAM cell. The design shows better rise time in WL compared to LPWD design. This design uses 32nm high-$_K$ metal-gate CMOS technology.

In following section, we propose a technique that improves read and write SNM, with comparable read access time compared to standard SRAM-cell, while showing lower power consumption compared to conventional wordline drivers. In addition, it gives lower leakage current due to the negative WL voltage during a percentage of standby mode.

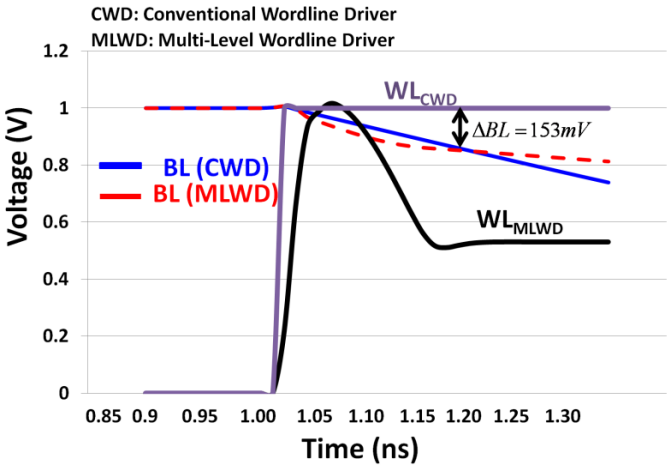


Fig.4.33. Simulation results for MLWD and CWD

Table4.2. MLWD output for different capacitances

Boosting Cap (fF)	MLWD output(V) (Write mode)
0.2	1.112
0.4	1.123
0.6	1.132
0.8	1.14
1	1.146

4.3.2. Proposed Wordline Driver

The schematic of the proposed wordline driver is shown in Fig.4.30. The Multi-Level wordline Driver circuit (MLWD) consists of three parts: (1) the delay element that determines the duration of read and the time during which WL has a negative level, (2) the pulse generator circuitry produces a special shaped pulse. (3) boosting capacitance. The MLWD scheme works as follows: let us consider the MLWD circuit operation in three different modes: (a) hold or standby mode (b) read mode and (c) write mode. During hold, the WL line is connected to a negative voltage for a short period of time ($6\tau_{inv}$, where τ_{inv} is the delay of a single inverter) and thereafter WL voltage level becomes zero where the circuit follows the conventional wordline driver operation. The negative voltage level of WL reduces the leakage current. The actual improvement in leakage power consumption due to the negative level of the wordline voltage depends on different factors such as the transistor sizing, boosting capacitance, and the inverter-chain delay. During write, the MLWD circuit, generates a $V_{DD}+\alpha$ voltage with $\alpha>0$. This WL voltage level (a wordline boosting technique) improves writeability. The value of α depends on the boosting capacitance while the rise time of WL depends on the transmission-gate sizing in the MLWD circuit. Larger transmission-gate transistors (TGT) gives lower rise time of the MLWD output during write or read modes. During write, due to the presence of higher current through pass
transistors in SRAM cell, writeability is improved significantly depending on the value of α. Hence, there is no need for dual supply voltages or negative ground (*Vss*). Finally,

during the read operation, for the primary time of read ($n \times \tau_{inv}$), where n is the number of

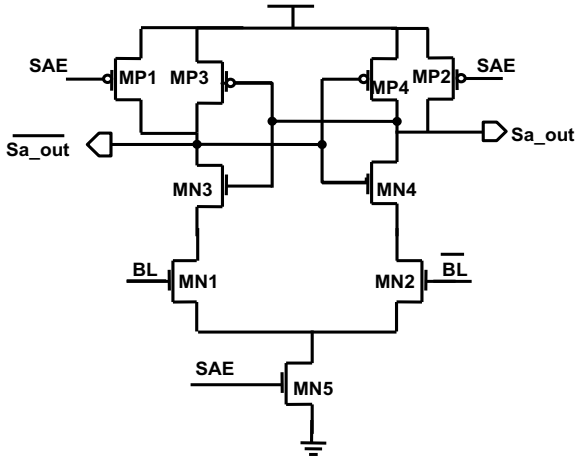


Fig.4.34. Voltage Sense-Amplifier used in SRAM array design

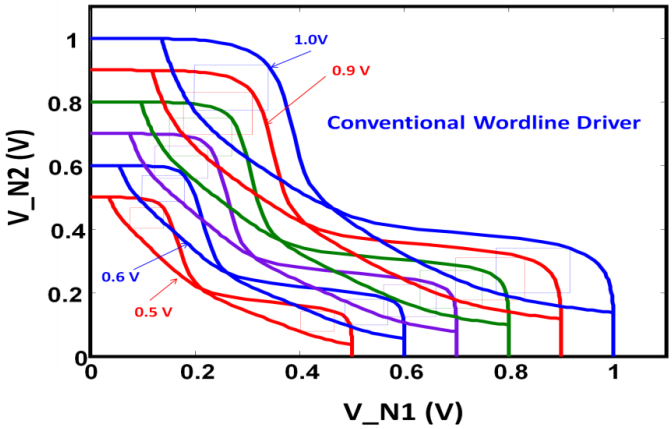


Fig.4.35. Read-SNM for 6T-SRAM at different operating voltages

inverter delays, WL is kept at a value around V_{DD}, and is then discharged to a level around $V_{DD}/2$. Due to the lowered level of WL voltage, the static noise margin of SRAM cell improves significantly compared to conventional WL driver with no degradation in read access time. The wordline voltage for a short time after starting the read operation is kept at V_{DD}. Consequently, the droop of the WL pulse width during read, does not affect the read access time.

To show the efficacy of the proposed MLWD circuit, simulations results are shown in the next section.

4.3.3. Simulation Results

As we have mentioned in the previous section, the proposed MLWD topology produces three voltage levels one for each operation region. Fig.4.31 shows the MLWD output during the write cycle. As illustrated, the wordline level is boosted to $V_{DD}+\alpha$ in order to

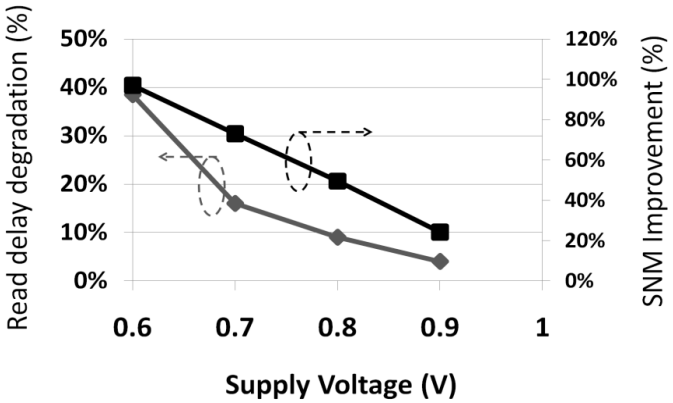


Fig.4.36. Lowering the WL voltage effect on SNM and Read access time

improve the Write time while it becomes negative when WL is at low. The value of α can be increased by making the boosting capacitance larger. The results are shown in Table4.2. During read for a short period of time from the beginning cell access ($\cong 6x\tau_{inv}$ seconds (70ps-100ps)), the MLWD output voltage is raised to $V_{DD}\pm\delta$ where δ is a very small value ($\cong$few mV), that depends on the transistor dimensions of the MLWD circuits, and the boosting capacitance. For the remaining WL pulse period, the MLWD output is set at a fixed voltage around $V_{DD}/2$. By applying such voltages, the read noise margin improves significantly while the read access time remains unchanged. Voltage waveforms showing the operation of a 6T-SRAM cell using MLWD are illustrated in Fig.4.32. The advantage of this technique is attributed to the non-fixed WL voltage level during read mode. Although some techniques lower the level of WL voltage during the read cycle [113] resulting in improved read noise margin, the penalty of degraded read access time is inevitably large. The proposed MLWD technique, in fact, slightly improves read access time. In Fig.4.33, we plot the results for BLs with conventional wordline driver (CWD) and Multi-Level Wordline Driver (MLWD). As it is shown, discharging the bitline for MLWD is faster than the CWD. Since there is enough time to read the data during the time after starting the read operation (70ps-100ps), the MLWD shows better results in terms of read access time compared to using conventional wordline driver topology due to the few mV boosted voltage during the read mode. For instance, read access time at $V_{DD}=1V$, $f=1GHz$, at TT (Typical NMOS, Typical PMOS) process corner, and at Temp=27^{O}C improves by 20% compared to the CWD. In this case, we use a sense-amplifier (SA) design in which ΔV_{BL} (Bitline differential) of 70mV can be sensed. For lower supply voltages, due to the delay of MLWD design (Transmission-gate transistors), read access time degrades by few percent compared to CWD. The sense amplifier (SA) used in our SRAM array design is shown in Fig.4.34 that can sense ΔV_{BL} less than 100mV.

Due to supply voltage scaling and process variations, the static noise margin in SRAM arrays degrades significantly, especially during read. To improve read SNM, different SRAM bit-cell topologies have been presented (8T, 10T, and Schmitt-Trigger SRAM) [114-116]. On other hand, different techniques have been presented to improve the SNM by lowering the level of WL voltage during read with a penalty on access time, weakening the pass transistors in SRAM cells using multi-threshold technology [96, 97], or applying a negative bitline voltage. We propose a technique to improve the read SNM by lowering the WL voltage level without any penalty on read access time. To show the improvement of

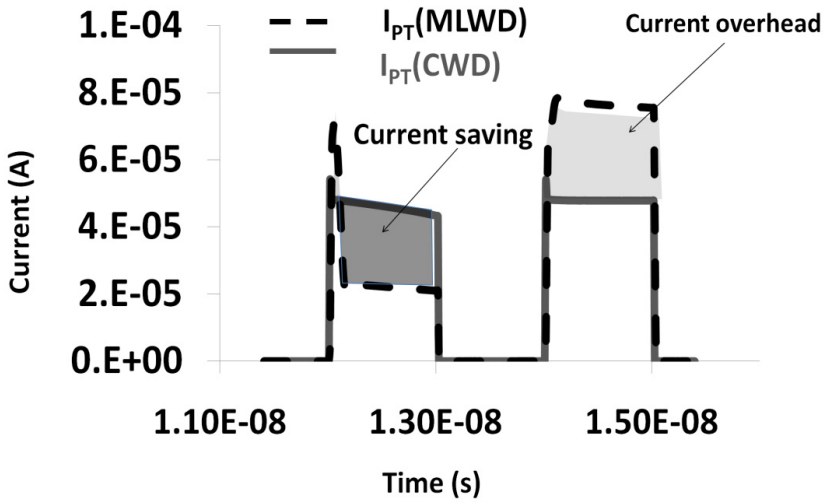


Fig.4.38. Current saving during Write and Read

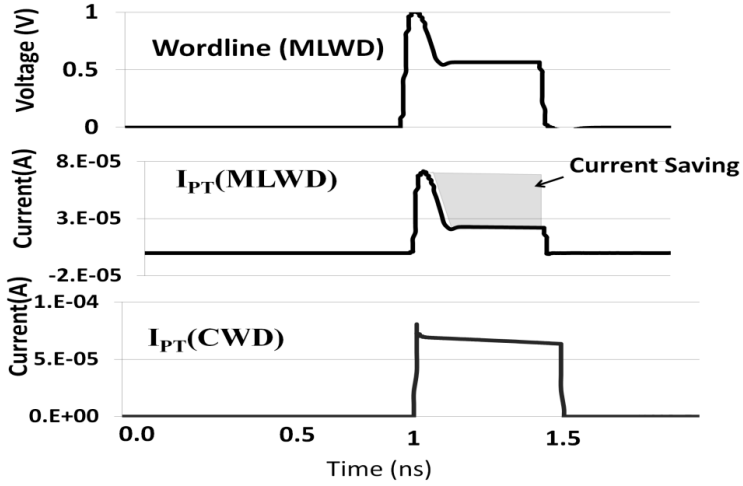


Fig.4.37. current saving of MLWD during Read.

MLWD technique, we perform HSPICE simulations on SRAM cell using conventional wordline driver and MLWD scheme. Fig.4.35 shows a family of butterfly curves and the corresponding read SNM of a conventional wordline driver for different supply voltages. It is known that lowering the supply voltage, SNM degrades significantly. To overcome this problem "weaker" access transistors are preferred. Lowering the WL voltage level is a way to make the access transistors "weaker" and thus to improve read SNM. Unfortunately, this technique degrades the read access time. Fig.4.36 explains how read SNM and access time are affected/ degraded by lowering the WL voltage. As it can be seen by lowering the WL voltage to 0.6V, read access time degrades by 40% while the read SNM improves by 2X.

By applying the MLWD technique, read access time does not get degrade while the read SNM improves by more than two times. Table 4.3, shows the results for read SNM using

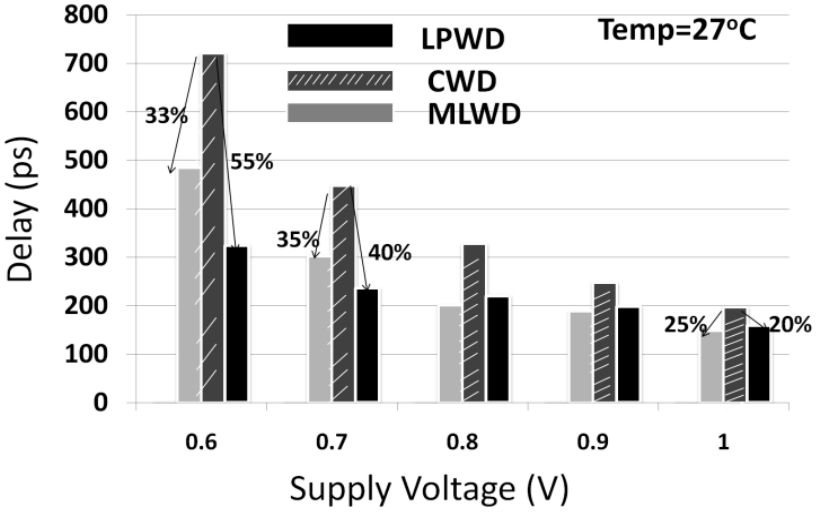


Fig.4.39. Write delay for MLWD compared to CWD and LPWD

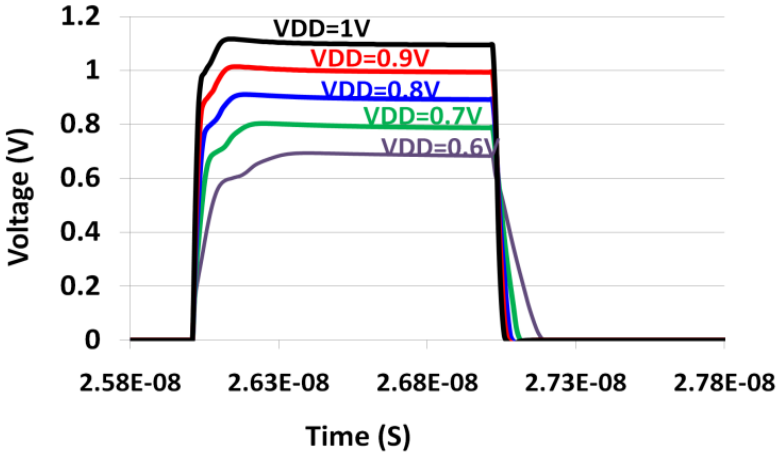


Fig.4.40. Level of MLWD at different operation voltage

Table 4.3. SNM results at different supply voltages

VDD (V)	READ SNM (mV)(CWD)	READ SNM (mV) (MLWD)	Improvement Percentage (%)
1	140.8	305.9	2.17X
0.9	132.6	284.2	2.14X
0.8	119.1	254.9	2.14X
0.7	101.0	219.4	2.17X
0.6	82	180.5	2.20
0.5	63.6	141.3	2.22

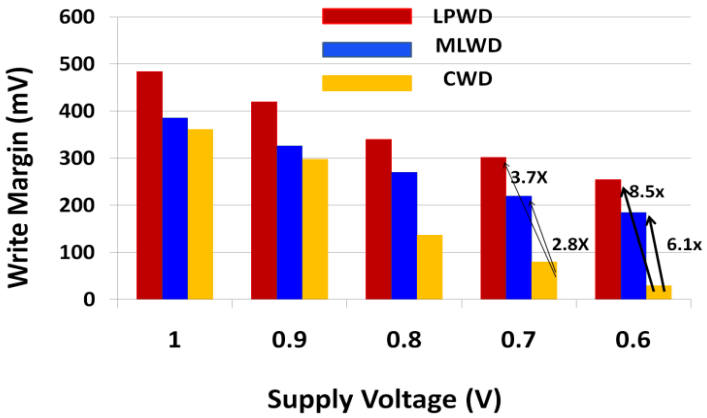


Fig.4.41. Write Margin comparison

MLWD and CWD topologies. MLWD design improves read SNM by at least 2.14Xcompared to the CWD scheme. This is attributed to the boosted WL (a few mV above V_{DD}). When the sense amplifier senses a voltage difference (ΔV_{BL}) between the bitlines, the WL voltage is lowered to suppress the access current for the remaining read cycle (90% of T_{WL}, where T_{WL} is the WL period). Fig.4.37 shows the access current flowing through the pass-transistors of the SRAM cell during read for MLWD and CWD . It is evident that our MLWD design provides more than 50% current savings during read compared to CWD technique. As a result, MLWD reduces the power consumed by the SRAM cell during read. This is attributed to decreased current flows through the access transistor (I_{PT}). In Fig.4.38, we have plotted I_{PT} during read and write modes for MLWD and CWD schemes. To calculate power saving, we perform HSPICE transient analysis for read, write, and hold modes. The results show that the MLWD design reduces the total power of an SRAM cell by 12% compared to the CWD design. Due to boosted wordline voltage during write, the MLWD scheme shows 37% increase in the I_{PT} current compared to CWD. Consequently, less improvement (12%) in total power saving is observed. It is known that the number of read accesses is three times larger than the number of write accesses based on test-benchmarks [118]. This suggests that the total savings will improve more than 12% for common applications, where read is more prevalent. Furthermore, writeability of the MLWD improves due to boosted WL voltage. In Fig.4.39, we compare the write delay for different wordline driver techniques. Since, the MLWD design shows less rise in voltage at lower supply voltages, write delay is less affected by using this technique. At lower supply voltages, the negative BL technique (e.g. LPWD and LPWD-SS) results in a smaller delay compared to MLWD technique. In addition, in Fig.4.40, we show the waveforms of the MLWD output during write for different supply voltages. The main reason of write speed degradation is due to the delay added by the MLWD circuitry. By upsizing the transistors in the MLWD (transmission-gate transistors), this problem can be mitigated for lower supply voltages. However, for higher supply voltages, the MLWD exhibits lower write delay compared to the negative BL methods (LPWD and LPWD-SS). The results of write margin for the negative bitline design compared to CWD and MLWD schemes are depicted in Fig.4.41. These simulation results show that the LPWD design provides better write margin compared to the MLWD technique. However, the MLWD achieves 2.75X improvement in write margin compared to the conventional design.

Furthermore, our design shows better write delay at higher supply voltages compared to the LPWD design.

To study the effect of process variations on the MLWD design, we perform Monte-Carlo simulations during read and write modes. The corresponding results are summarized in Fig.4.42 and Fig.4.43, respectively. For read mode, we observe the effect of process variations on the lowered WL voltage. As it can be seen, the μ and σ values are 0.53V and 17.4mV respectively. Furthermore, we observe that the mean value of α during write is 0.11V. Therefore, our circuit demonstrates robust behavior in presence of transistor process variations and mismatches.

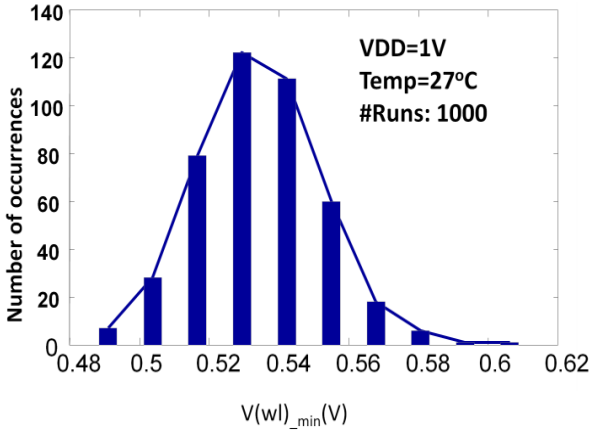


Fig.4.42. Process variations effect on WL voltage during the READ

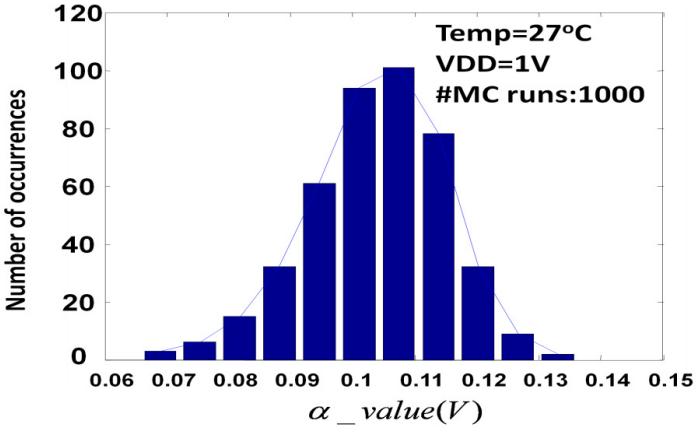


Fig.4.43. Effect of process variations on WL voltage level during write

4.3.4. Dynamic Noise Margin Comparison

Static noise margin (SNM) supposes that the word-line pulse width is infinite. Even though this assumption does not represent the reality, it provides an easy, fast and comprehensive way to measure the stability of an SRAM cell. Recently, increasing attention has been paid to dynamic stability analysis and dynamic noise margin (DNM)

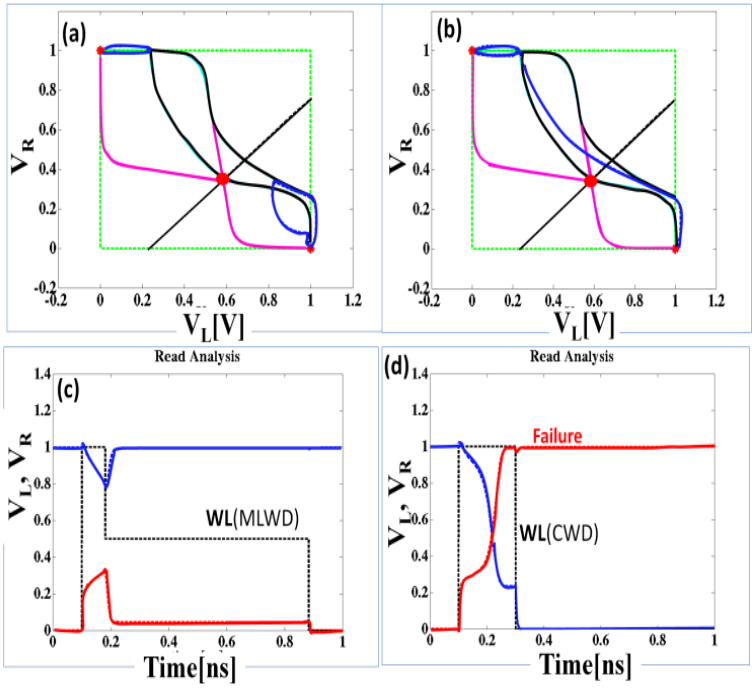


Fig.4.44. Simulation results for CWD and MLWD design

[119], [120]. The DNM incorporation for SRAM cell analysis is necessary in our case to grasp the improvements that are achieved by employing the proposed technique. The premise is capturing transient states that are not captured during the DC static analysis. The main reason for paying attention to dynamic noise margin here is that we used a shaped pulse during read cycle. Lowering the WL pulse width during read makes our design more stable compared to other counterparts. In our design, we lowered the level of voltage (V_F (WL)) after a period of time that improves the read stability.

Before we present our results, it will be more convenient to introduce some terms that help explain our results appropriately. By running transient analysis using a numerical simulator like SPICE we get the voltage waveforms (V_L, V_R and W_L) with respect to time as they are shown in Fig.4.44 (a). Their trace in the V_L-V_R plane is called a trajectory and this plane is called phase plane (or vector plane). Some points and curves on the phase plane have some special properties and play a significant role on SRAM analysis. If a point has the property that for a trajectory that starts from this point then it stays there forever then this point is called equilibrium point (EP). In our case these points can be determined graphically by the cross points of the butterfly curves. It is obvious that the points (V_{DD},0), (0,V_{DD}) and the metastable point are equilibrium point for the hold mode.

The first two are stable but the last one is unstable. Note that these EPs move during read and write operation. The curve that plays a significant role in our analysis is the separatrix. Separatrix is a curve that splits the V_L-V_R plane into two sub-planes. All the trajectories that start from one of these sub-planes converge to one of the two equilibrium points. In addition, separatrix passes through the metastable point. All these concepts are depicted in Fig.4.44.

DNM is important when we deal with cells that are statically unstable. Statically unstable cell is one where the read mode butterflies cross in one point. That case is shown

in Fig.4.45. For instance, for that cell, if a trajectory starts from EP (V_{DD},0) after some finite time it will cross the separatrix and then it will reach the unique read mode EP. When

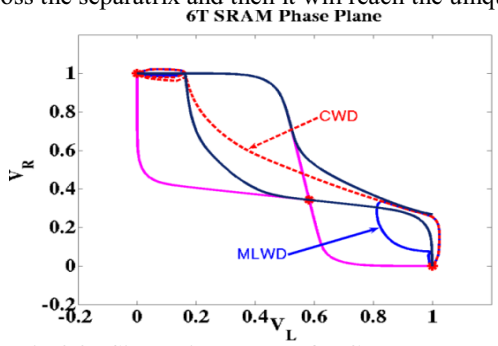


Fig.4.45. Simulations results for CWD and MLWD

the access transistor closes, it will converge to (0,V_{DD}) EP. It is obvious that this cell is dynamically unstable because the stored data is not the same before and after read operation. However if the time that the WL is high is short enough then the trajectory may not have enough time to cross the separatrix and reach the other EP, thereby, retaining the data. Note that the cell is dynamically stable in this case. These two scenarios are shown in Fig.4.46.

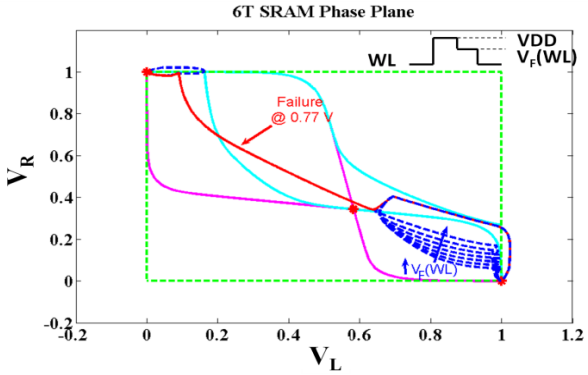


Fig.4.46. Simulation results for MLWD at different levels of V_F(WL)

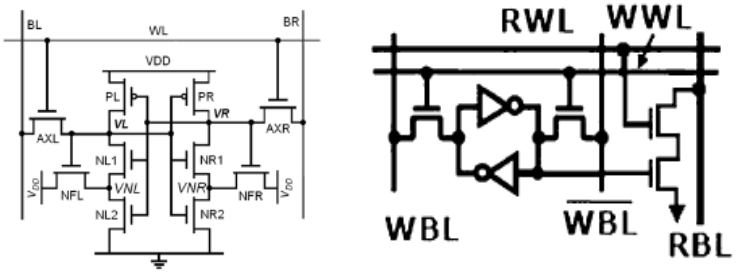


Fig.4.47. Schematics of 8T-SRAM [115] and ST-SRAM cells [117]

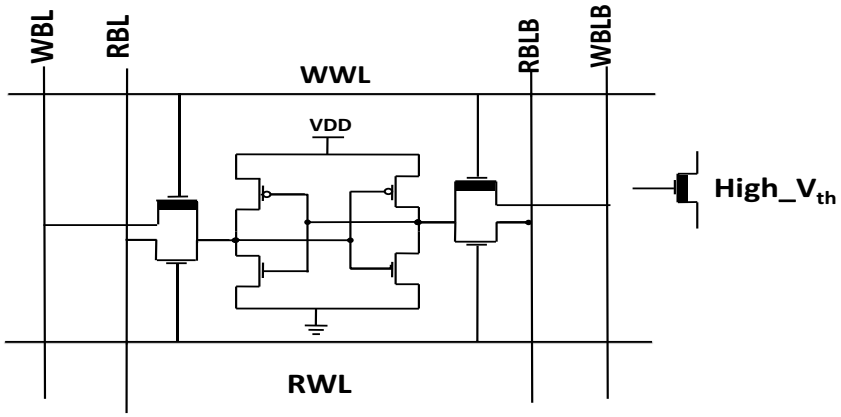


Fig.4.48. TGPT-SRAM cell

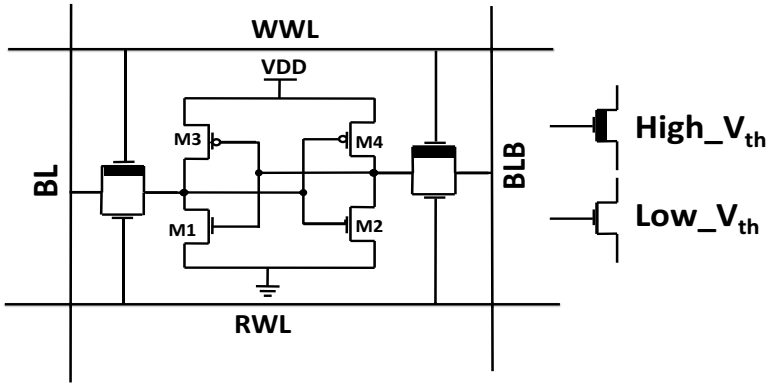


Fig.4.49. TGPT-SRAM cell with separate BLs

4.3.5. MLWD in Other SRAM Topologies

Fig.4.47 shows the schematics of the 8T-SRAM and ST-SRAM cells [115], [117]. By applying MLWD to 8T-SRAM cell, access time degrades due to the single-ended bitline for the 8T-SRAM cell. However, due to the boosted WL in write mode, the write margin improves similar to a 6T-SRAM cell. However, the 8T-SRAM cell does not suffering from the dummy-read when WL voltage is boosted to a value higher than V_{DD}.

Table4.4. Simulations results for CWD and MLWD

VDD(V)	READ SNM STSRAM (mV) (CWD)	READ SNM ST-SRAM (mV) (MLWD)	Improvement percentage (%)
1	206	399.8	1.94X
0.9	187.4	369.6	2.11X
0.8	166	338	2.03X
0.7	146.3	295.4	2.01X
0.6	122	255.6	2.09

The ST-SRAM cell [117] improves read SNM by 1.47X compared to the 6T-SRAM cell. Therefore, considering effect of MLWD on this design would be interesting. By MLWD to the ST-SRAM cell we can get at least 1.94X read SNM improvement compared to using the CWD scheme. Table 4.4 shows the results at different supply voltages for the MLWD

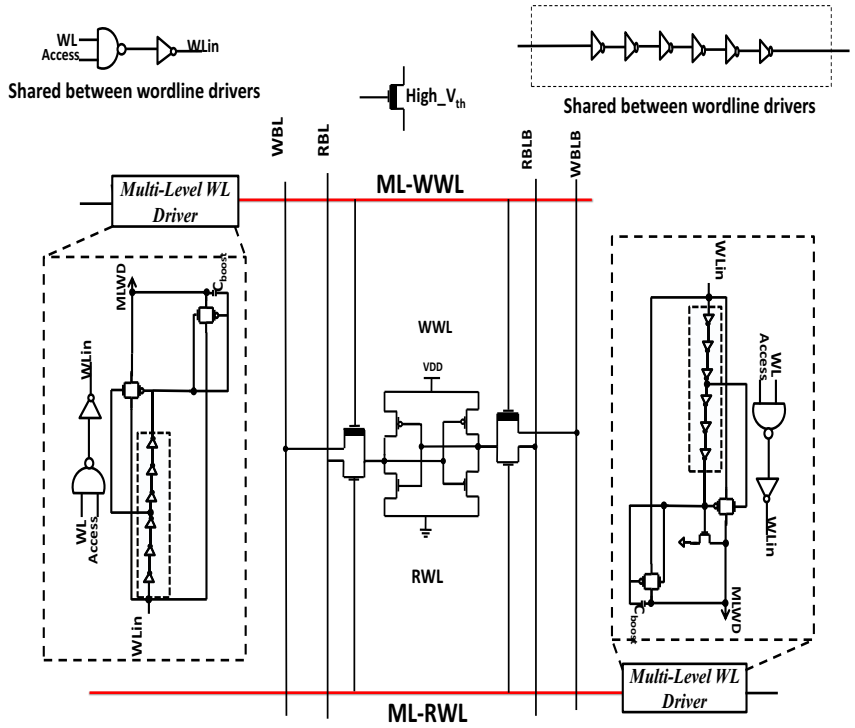


Fig.4.50. Separate bitlines TGPT-SRAM cell

compared to using a CWD design.

This technique can be applied to a transmission-gate pass-transistor SRAM (TGPT-SRAM) cell as shown in Fig.4.48. As it can be seen, by using high-V_{th} (HVT) devices for the write-assist access-transistors, dummy-read can be reduced due to the reduced effect of boosted voltage. Furthermore, by using Low-V_{th} device, the access time degradation at lower supply voltages can be compensated. However TGPT-SRAM cell leads to 25% area penalty compared to 6T-SRAM cell. Another configuration that can be used is separated RBL and WBL. The corresponding topology is shown in Fig.4.49. By using this SRAM topology, further improvements in read SNM and access time are observed. In this case, we use separate wordline drivers, read wordline (ML-RWL) and write wordline (ML-WWL). Furthermore, this topology is not suffering from the dummy-read due to the separated read and writes lines. The whole SRAM topology including WL drivers is shown in Fig.4.50. The number of transistors compared to using the MLWD in a 6T- SRAM cell is increased by two NMOS, one PMOS transistor, and one boosting capacitance that is less than 10% compared to using MLWD in single BLs for read and write.

In this section, we proposed a multi-level wordline driver to enable better write margin without any read access time degradation. Furthermore, by using the MLWD scheme, read noise margin is improved by at least 2X compared to the conventional wordline driver

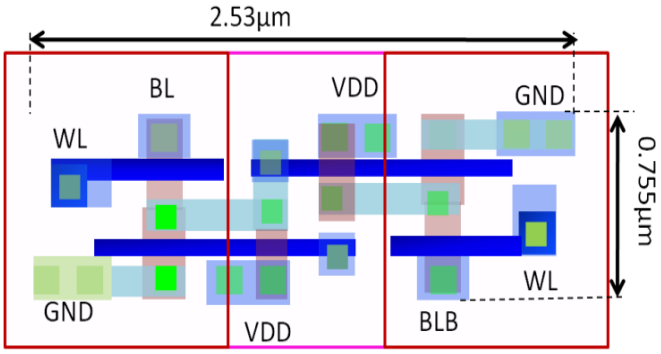


Fig.4.51. Thin-cell 6T-SRAM cell layout

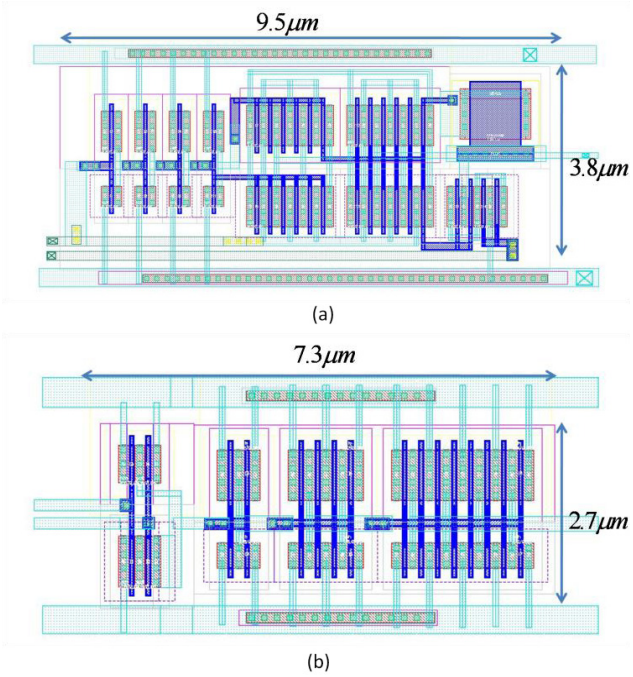


Fig.4.52. Layout of (a) MLWD (b) CWD designs

(CWD). When applying the proposed MLWD to an 8T-SRAM and a Schmitt-trigger SRAM cell, the read-SNM was improved by at least 1.96X compared to using a conventional wordline driver in the ST-SRAM. Total power is reduced by at least 12% compared to the conventional SRAM due to a lowered drive current through the pass-transistors in read mode. Furthermore, we proposed an 8T-SRAM cell (TGPT-SRAM) using MLWD scheme to separate read and write wordlines resulting in improved read SNM. Although the proposed MLWD scheme can be applied to any SRAM cell, it is more useful in SRAM cells with dual bitlines. However considering area overhead is an

important figure of merit for SRAM arrays. A 32KB SRAM array is was fabricated using TSMC 65nm technology using two different wordline driver topologies to show the

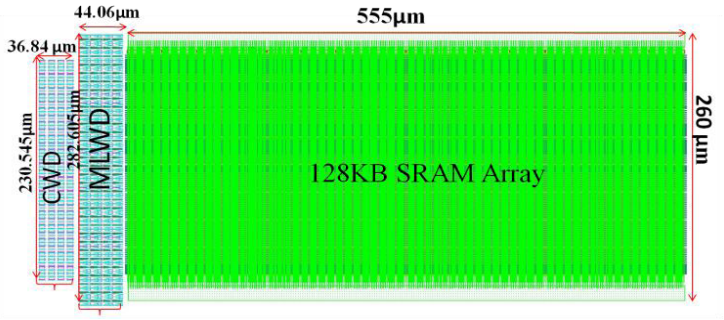


Fig.4.53. Layout of 128KB SRAM array including MLWD and CWD layouts

efficacy of proposed MLWD driver. Next sub-section includes area information for MLWD design.

4.3.6. Implementation Area

We have implemented the layout of a thin-SRAM cell as depicted in Fig.4.51. The layout of the MLWD and CWD designs are shown in Fig.4.52.Note that the capacitor C_{boost} is implemented using a MOSCAP device which shows different capacitance values at different biases, for instance C_{boost}=17fF, 10fF and 2fF for capacitance voltage equal to 1V, 0V, and -1V respectively. Even though the area overhead for a single bit wordline driver is 46% the total area overhead due to MLWD circuitry compared to CWD is less than 2.5% for a 128Kb memory array that is shown in Fig.4.53.

4.4. Conclusions

In this chapter, several topologies for SRAM design in circuit/architecture level were explained. Several SRAM cell topologies were proposed for ultra low power applications in sub/super threshold regions. 11T-SRAM and PAT-SRAM cells were proposed that showed significant improvement in robustness for read and write, respectively. However PAT-SRAM topology is process-dependent and 11T-SRAM comes with a penalty of a large area overhead. A new wordline driver was proposed for high-performance and low power applications with improved read and write margins. In multi-level wordline driver, different levels of signals were applied to wordline. Simulations results showed significant improvement in static and dynamic noise margin compared to conventional wordline drivers. Furthermore, to show the efficacy of proposed design, the MLWD technique was applied to 8T-SRAM and Schmitt-trigger SRAM cell.

Chapter 5

Nano-Scale CMOS Circuit Design Challenges for Ultra Low Power Applications

Chapter 5

Nano-Scale CMOS Circuit Design Challenges for Ultra Low Power Applications

Introduction

Making energy efficient electronic systems is currently becoming more and more important. The demands for combining high performance operation and long battery lifetime in portable and wireless applications have strong implications on how to design such systems. The need for cost optimized solutions targeting large volume applications results in combining all system components on one single chip usually referred to as a System on Chip (SoC). Such systems combine advanced RF and analog blocks in the front-end and analog/digital interfacing circuitry with complex digital parts performing signal processing, system control and storage. In order to make energy optimized solutions, all aspects must be taken into account. The introduction of intelligent power control and efficient modulation methods has reduced the energy consumption of such systems significantly. However some parts of the system may not be put in idle mode, especially those parts taking care of the power control and monitoring trigger events for leaving idle mode. The fact that these parts of the system are always on leaves a great potential of saving energy by reducing the supply voltage and introducing new and innovative techniques at device-level.

This chapter explains body-biasing techniques in order to find the optimum operating conditions for PMOS and NMOS devices in deep submicron technology. Body biasing techniques have been used to alter the MOS threshold voltage in order to either increase the speed or reduce leakage. According to theory and in more conventional devices, forward body biasing (FBB) reduces the threshold voltage while Reverse Body Biasing has the opposite effect. In deep submicron devices, the effect of these techniques is more complicated due to short channel effects. This work is focusing on analyzing the effect of these techniques using a standard 65nm CMOS process and to further determine how to utilize FBB and reverse body biasing (RBB) to establish optimal operating conditions for digital circuitry targeting high performance and low energy operation. In order to predict

the effect of FBB and RBB both as a function of physical parameters, geometry and temperature, an analytic approach combined with circuit simulations has been chosen to determine the physical origin of the observed behavior and the practical consequences for circuit design utilizing the subthreshold operating regime [121]-[124].

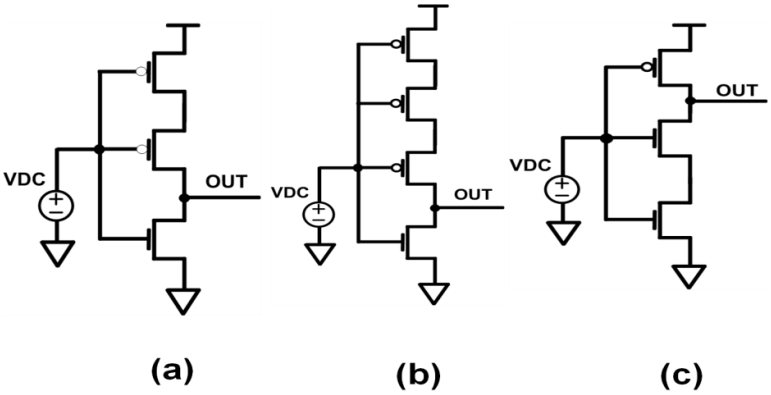


Fig.5.1. (a) 2PMOS (b) 3PMOS (c) 2NMOS circuits

Sub-threshold current rises due to lowering of threshold voltage which is scaled down to maintain transistor ON current in the face of falling power supply voltage. Voltage scaling for standby power reduction was suggested since both subthreshold current and gate current decrease dramatically (with V^4 for gate leakage) [124]. Lowering supply voltage thus saves standby power by decreasing both standby current and voltage [125]. The subthreshold region (weak inversion) is often utilized to implement power efficient circuits for ultra low power wireless applications, but due to the much lower current in subthreshold region compared with higher supply voltages, the evaluation speed of such circuits operating in weak inversion is decreased. Therefore, new techniques to improve circuit speed need to be developed.

Subthreshold design has emerged as a good potential for ultra low power applications such as wireless sensor networks, medical instruments, and portable devices.

We have observed some specific behaviors from devices operating in subthreshold region in the 65nm technology due to lack of well-engineered models for subthreshold region. Short channel devices have been optimized for regular strong inversion circuits to meet various objectives such as high mobility, reduced Drain-induced barrier lowering (DIBL), low leakage current, and minimal V_{th} roll-off. However, a transistor that is optimized for operating in superthreshold logics are not necessarily optimal in low voltage, low power dissipation applications designed for operation in the subthreshold region. Optimization problems include the transistor sizing, the drive current for PMOS and NMOS devices, the effects of some techniques such as FBB, RBB, and stacking effects on threshold voltage and drive current. Although, it would be ideal to have a dedicated process technology optimized for subthreshold circuits. In order to design optimal subthreshold circuits using CMOS devices that are targeted for superthreshold operation, it is crucial to develop design techniques that can utilize the side effects that appear in this new regime. However, in the absence of such a dedicated process the development of low voltage low power applications using the 65nm CMOS technology requires care and novelty in design.

5.1. DC Analysis for CMOS 65nm Technology

In this section three topologies for basic circuits are presented and simulated using DC analysis. Fig.5.1 illustrates the topologies that are simulated in 65nm technology with

supply voltage equal to V_{DD}=0.9V. In all topologies minimum sizes for the transistors are used. Fig.5.1 shows the three stacked devices (two PMOS and one NMOS referred to as 2PMOS). Fig.5.1 (b, c) are referred to as 3PMOS and 2NMOS respectively. Simulation results based on DC analysis for these three configurations are illustrated in Fig.5.2. As it can be observed, the short circuit current in 2NMOS is higher than for the other circuits, which implies that the delay for 2NMOS configuration is higher than the delay of other topologies causing increased short circuit current through this circuit.

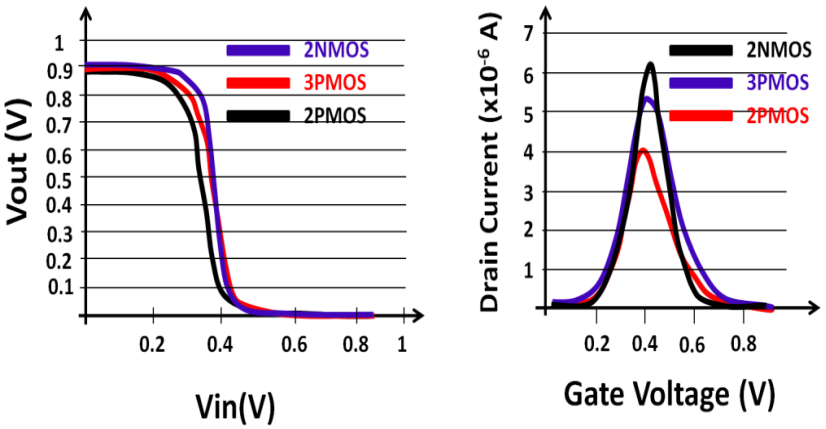


Fig.5.2. V_{out} vs. V_{in} (b) IDS current vs. gate voltage

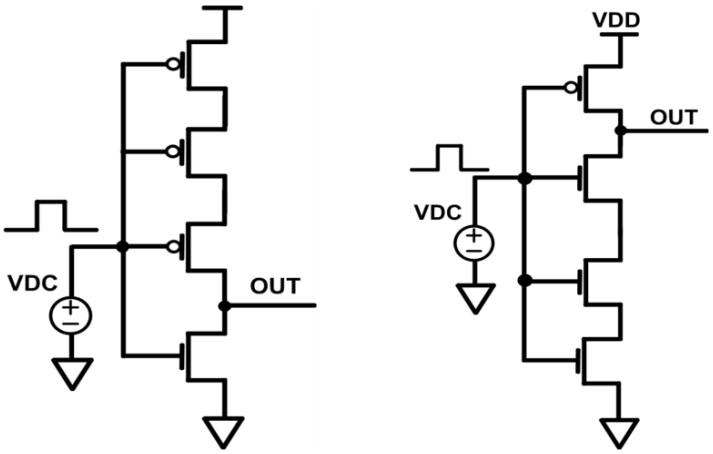


Fig.5.3. (a) circuit1 (b) circuit2 schematics

5.2. Stacking Effect in 65nm Technology

Stacking has been proposed as a technique to decrease the leakage current in subthreshold region [126]. This technique is based on increasing the threshold voltage of source to bulk as a result of which the threshold voltage will increase thereby reducing leakage current in idle mode. However, when the circuits are forced to work in ultra low supply voltage (subthreshold region), increased subthreshold current is desired to improve

the circuits speed. In order to find better circuit topologies which use the stacking technique, two topologies which utilize the stacking effect are simulated.

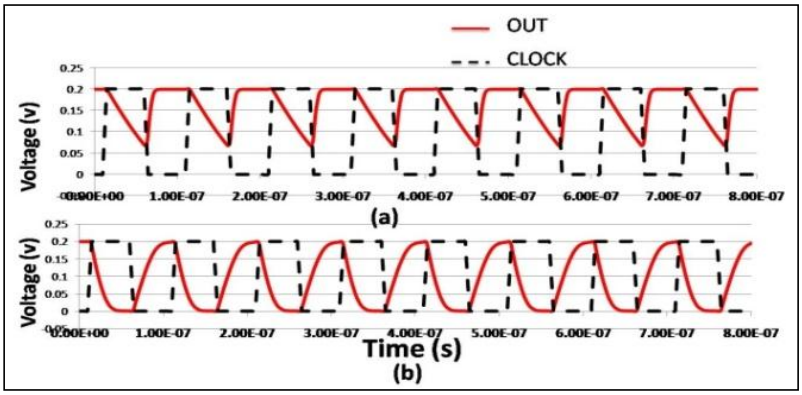


Fig.5.4. transient analysis for (a) circuit2 (b) circuit1 (with minimum size for all transistors in 27°C, TT model)

As known, in CMOS technologies, the speed of the NMOS is higher than PMOS because of higher mobility of electrons compared with holes, but simulations show that for 65nm technology in ultra low supply voltages (in subthreshold), this behavior is changed. The main reason of this phenomenon is that these models are engineered for superthreshold circuits. For superthreshold applications, the threshold voltage of PMOS devices are lowered to compensate the effect of lower mobility compared to NMOS devices. But in subthreshold region, the exponential dependence of subthreshold current to V_{th} causes some unexpected results [127]-[129]. Fig.5.3 shows two topologies used to illustrate the stacking effect on speed. These two circuits were simulated for ultra low supply voltages (V_{DD}=0.2V) to investigate which is the faster topology. Two similar pulses are applied to the inputs of the circuits, and the charging and discharging speeds of the output nodes are considered and compared for Circuit1 and Circuit2, respectively. Fig.5.4 shows the simulation results for these two configurations, which show that circuit1 has a higher speed compared with Circuit2 in ultra low supply voltages, so employing circuit1 configuration in circuits such as D-Flip-flops instead of circuit2 topology, gives much better results. To attain the same speed for circuit2, the NMOS transistors must be upsized 13 times, so the area overhead of circuit2 is higher than circuit1 in the same speed. Based on this concept The SAFF (Sense amplifier flip-flop) and the HLFF (Hybrid latch flip flop) are simulated using the complementary circuits of the NMOS stacked transistors configuration. Results of simulation utilizing circuit1 topology show that higher operating speeds are achievable at lower supply voltages as a result of which we attain significant reduction in power dissipation. Fig.5.5 shows the effect of body biasing technique on an inverter with different bulk voltages.

5.3. Body Biasing Technique

In this section the effect of forward body biasing (FBB) and reverse body biasing (RBB) techniques on the current through NMOS and PMOS devices are investigated. Body biasing is used to change the threshold voltage of transistors governed by the following equation:

$$V_T = V_{T0} + (\gamma\sqrt{2|\emptyset_F| + V_{SB}} - \sqrt{2|\emptyset_F|}) \qquad (5.1)$$

where γ is the body threshold parameter, V_{T0} is the threshold voltage in V_{SB}=0, and $\emptyset_F$ is the strong inversion surface potential which can be neglected compared with V_{SB} for lower supply voltages. The threshold voltage equation may then be simplified to this equation:

$$V_T = V_{T0} - \gamma V_{BS} \quad (5.2)$$

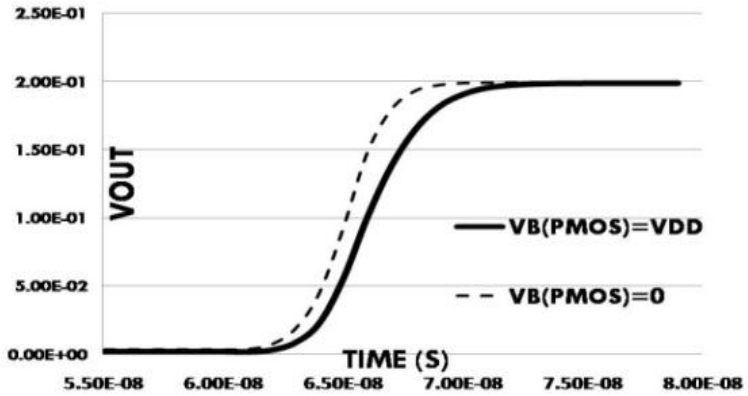


Fig.5.5. the effect of body biasing technique

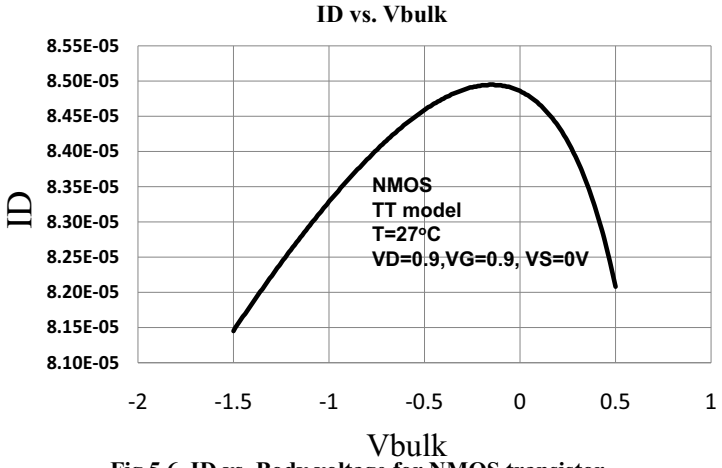


Fig.5.6. ID vs. Body voltage for NMOS transistor

As observed from this equation, by increasing the body voltage (NMOS) relative to the source voltage, the threshold voltage is decreased and then the current is increased, while in PMOS transistors by increasing the body voltage (Reverse body bias), the threshold voltage is increased and then the current is decreased [130]. By applying the RBB technique to NMOS devices, the source voltage is higher than body voltage, so by increasing the source voltage, the threshold voltage increases and then the current through the NMOS is decreased. By utilizing reverse body biasing technique in NMOS devices, the threshold voltage decreases and as a result, the drain current increases which in turn improves the speed of the circuit.

Therefore, the highest speed of the NMOS is expected to be when the body is connected to the maximum allowed forward body bias voltage. In the case of a PMOS, when the body voltage is connected to lowest possible voltage, the threshold voltage is expected to be at

its minimum, maximizing the ON current of the PMOS device. The limitation, however, is that the forward bias amount must be small enough to prevent the junction diodes from

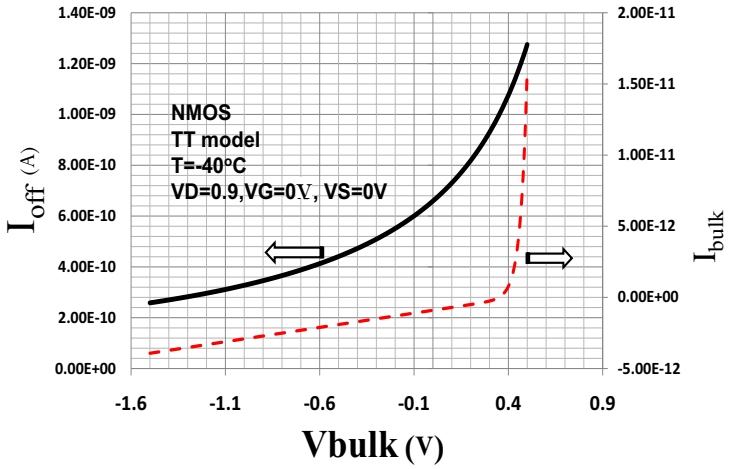


Fig.5.7. I_{off} and I_{bulk} vs. V_{bulk}

turning on.

This limits the forward bias range to be below the PN diode built-in potential, giving an acceptable forward bias range of about 500mV. Forward bias has been shown not only to improve performance, but to also reduce short channel effects [131]. Reverse body bias is while the body to source is reversely biased. In this case, the current is expected to decrease due to negative body to source voltage.

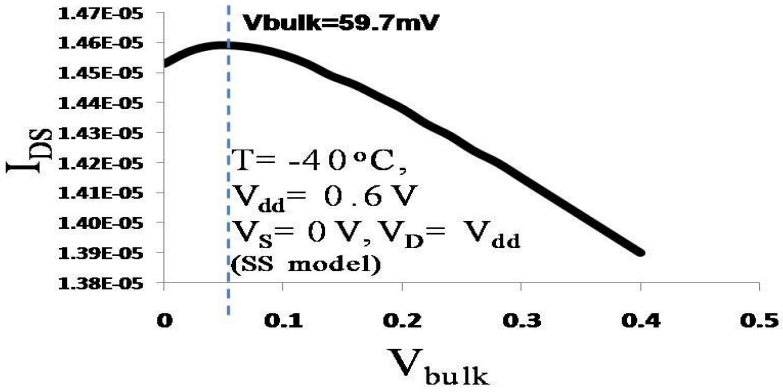


Fig.5.8. PMOS Drain current versus Body voltage

5.3.1. Body-Biasing in Scaled Technologies

RBB technique is typically applied to decrease the leakage currents in a system. In general, by applying the reverse body bias technique the threshold voltage is increased. By scaling the technology, the RBB effectiveness also diminishes with technology scaling primarily because of worsening SCE, especially when the target V_T value is low. In 65nm CMOS technology, body biasing technique has a different effect on NMOS and PMOS devices. Fig.5.6. shows the drain ON current at V_{DD}=0.9V versus the body (body) voltage

for a NMOS transistor. As illustrated, the maximum current is not while the body is connected to higher forward bias

voltage. In fact, the maximum ON current happens at a body in a negative voltage range (at Vbb=-0.23V). This is in contrast to the traditional effect of the BB technique on NMOS. Further, in this case, the maximum current occurs when the body voltage is around -230mV.

Fig.5.7. shows the drain off current (I_{off}) and body current with body voltage. As it can be seen, the minimum leakage is when the body terminal is connected to the lowest body voltage (maximum RBB). Therefore, to get the minimum leakage in 65nm technology, the body can be connected to the most negative bias. In Fig.5.7, it is shown the body current is negligible compared to the drain off current (1000X less). So the optimum point to get the minimum leakage is while the body is connected to the most possible negative voltage. Fig.5.8 shows the drain to source ON current versus body voltage for PMOS. it is observed that by lowering the body voltage until the body voltage is equal to V_{body}=50mV, the current increases and afterwards it decreases by further lowering the body voltage. As illustrated the maximum current is when the body voltage is V_{body}=59mV. This is also contrary to the traditional FBB effect which indicates that more FBB should result in more ON current. Table.5.1 shows the body voltage level at which the current of the device is at maximum.

In this section, we focus on simulation results based on short channel equations. We consider the channel depth, doping concentration and temperature effect on I_{on} ($V_{GS}>V_{th}$). First, let consider the short channel effects (SCE) in 65nm technology and beyond then the effects of body biasing techniques are shown for short channel devices. As technology is scaled, the oxide thickness is decreased and the threshold voltage is decreased. To have acceptable power consumption, the most efficient way is scaling the supply voltage, but by scaling the supply voltage the performance is degraded.

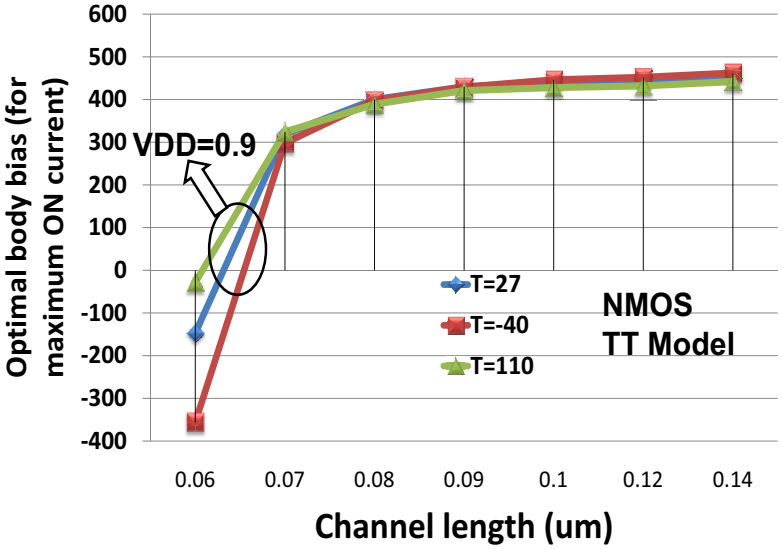


Fig.5.9. Optimal body bias vs. channel length for maximum ON current

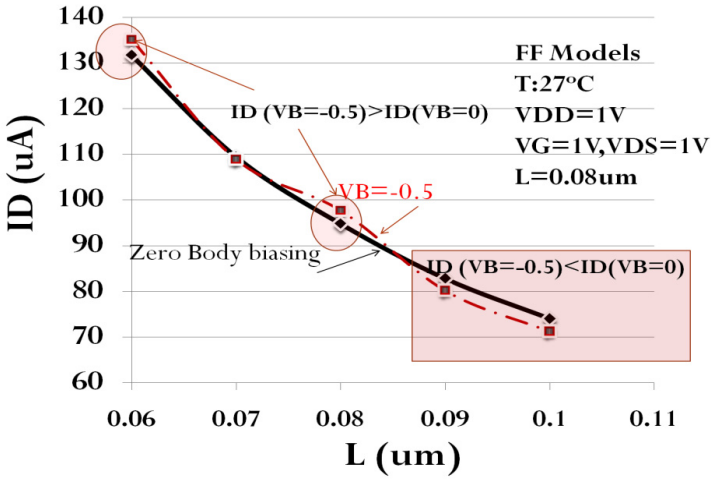


Fig.5.10. ID (ON current) vs. channel length

5.3.2. Body-Biasing Simulation Results

The simulation results shows that there is an optimal body bias point to get the highest ON current and lowest OFF currents for PMOS and NMOS devices in 65nm technology and these optimal bias points are different than traditionally expected. To understand the main reason behind these results, first let us consider the short channel effect. To have a higher speed, using minimum channel length is an alternative, but the SCE effect is worsened for short channel devices especially for 65nm technology and beyond. One of the processing options to mitigate the effect of SCE is halo doping to increase the threshold voltage along with lowering leakage current. This change in transistor characteristics, has some effect on body biasing techniques.

The following equations describe the DIBL and V_{th} roll-off for short channel devices:

$$V_t = V_{fb} + |2\phi_F| + \frac{\lambda_b}{C_{ox}}\sqrt{2qN\varepsilon_{Si}(|2\varphi_F| + V_{sb}} - \lambda_d V_{ds} \quad (5.3)$$

where V_{fb} is the flat-band potential. C_{ox} is the oxide capacitance, and N is substrate doping concentration. In this equation V_{sb} and V_{ds} are source to body and drain to source voltage respectively. λ_d and λ_b describes the DIBL and V_{th} roll-off effects in short channel devices, respectively. The following equations explain these two SCE effects:

$$\lambda_b = 1 - [(\sqrt{1 + \frac{2W}{X_j}} - 1]\frac{X_j}{L} \quad (5.4)$$

and

$$\lambda_d = [\frac{L}{2.2\mu m^{-2}(T_{ox}+0.012\mu m)(W_{sd}+0.15\mu m)(X_j+2.9\mu m)}]^{-2.7}$$

where $W = \sqrt{\frac{2\varepsilon_s}{qN_D}(V_{sb} + |2\phi_F|)}$, X_j is source or drain junction depth, and L is channel length. ε_s is the silicon permittivity. Regarding to these equations, using non-uniform doping in devices, causes an increase in doping concentration which results in increased threshold voltage due to the decreased channel depth. Therefore, by increasing the channel depth, the threshold voltage is decreased due to worsening the SCE effect.

5.3.3. Channel length effect on Body-Biasing technique

For short channel devices, there are two major effects: (a) DIBL and (b) V_{th} roll-off whose effects are shown in equations (5.3) and (5.4). By increasing the drain voltage, the depletion width is increased and due to reduced effective channel length, I_{on} increases. So by decreasing the channel length until a specific length, threshold voltage remains constant, but afterwards it starts decreasing. For halo doping devices, we have two effects. First, due to increased depletion width, the channel doping concentration raises causing an increase in the threshold voltage. The second effect is caused by DIBL and SCE that increase the current from Drain to Source. Hence, there is a contention between V_{th} roll-off and DIBL with doping concentration effect in channel. However at very low supply voltages DIBL effect is negligible due to very low V_{DS} voltage, and therefore, there is a contention between channel doping concentration and V_{th} roll-off effect. When the effective channel length is comparable with the depletion width, by increasing the body bias voltage, the depletion width is decreased. Lower depletion width decreases the SCE effect and DIBL causing a higher threshold voltage.

Simulations show that the effect of SCE is much higher than the effect of V_{bs} on threshold voltage increase. Therefore, by increasing the body voltage, the current decreases.

Fig.5.9 shows the NMOS Optimal body bias at different channel lengths. As it can be seen, more negative body voltage causes higher current for very short channel lengths, for longer channel lengths the behavior changes. Another point is that by just a 0.02um increase in length the current is decreased dramatically. This effect is due to mitigated short channel effects in longer channels. Fig.5.10 shows the drain current for different channel lengths. As illustrated, the dependence of drain current to body voltage is different

Table.5.1. Optimum body voltage to get the maximum I_{on} (PMOS)

VDD=0.9

Process	T=-40	T=27	T=110
TT	0.9	0.1500	0
FF	1	1	0.62
SS	0	0	0

VDD=0.8

Process	T=-40	T=27	T=110
TT	0	0	0
FF	1	0.65	0.305
SS	0	0	0

VDD=0.7

Process	T=-40	T=27	T=110
TT	0	0	0
FF	0.585	0.156	0
SS	0	0	0

VDD=0.6

Process	T=-40	T=27	T=110
TT	0	0	0
FF	0.585	0.156	0
SS	0	0	0

L=Lmin (0.06µm), W=0.12µm

at different channel lengths. Also the effect of body biasing technique in this figure is illustrated. Fig.5.11 shows the dependence of optimal body biasing on channel length. As illustrated in this figure, for higher supply voltages and minimum channel length we have DIBL and V_{th} roll-off effect and by increasing the body voltage, the current decreases. For longer channel lengths, we have the normal behavior because of lower influence of SCE

effect. Therefore, by increasing the body voltage the current increases due to a decrease in threshold voltage.

5.3.4. Efficacy of Body-Biasing Technique

In sub-threshold region, by increasing the body voltage, threshold voltage is decreased by the relation between V_{th} and V_{bs}, and I_{sub} is changed exponentially with the threshold voltage. In this case, the effect of threshold voltage on sub-threshold current is dominant factor. The body voltage increase has a normal effect on the sub-threshold current due to lower SCE effects. In Halo doping devices, by increasing the channel length, the overlap between Halo doping of the source and that of the drain is reduced. Therefore, first, the threshold voltage is decreased then it starts to increase again. This phenomenon is called reverse short channel effect. For sub-threshold region, there is an optimal channel length to maximize the I_{sub},

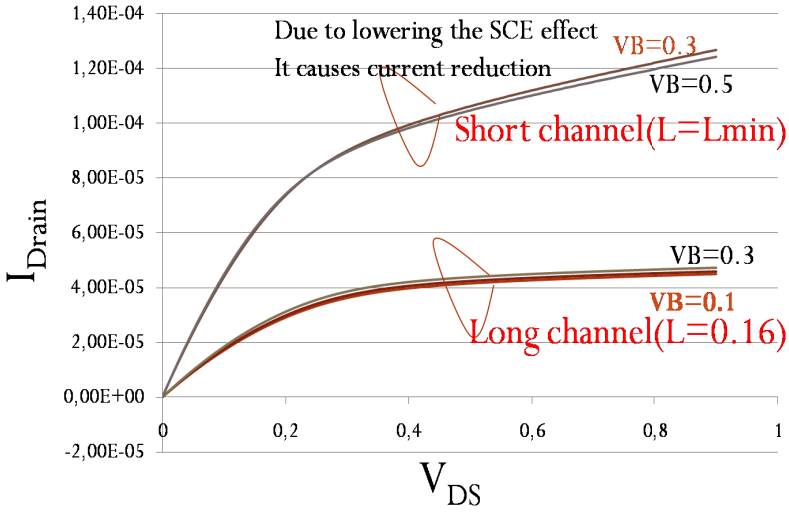


Fig.5.11. The effect of FBB on long and short channel devices

but the optimal point to get the highest I_{sub} is when bulk connected to a higher voltage. Table.5.1 shows that for higher supply voltages the effect of RBB technique is degraded, but for lower supply voltages due to improved DIBL and V_{th} roll-off, the regular behavior of RBB technique is observed. By lowering the temperature, the optimum body voltage to get the highest performance is decreased. As it can be seen in Table 1, the worst case in which the optimum body voltage is more negative is while FF models are used and the temperature is lowered. The reason is that, for FF models, the channel length is much smaller, so the effect of channel length increasing is more significant. For lower temperatures, the body voltage is lower, because the DIBL effect is more. As we discussed in this section, DIBL is a dominant factor to determine the optimum body voltage value. Short channel devices have been optimized for regular superthreshold circuits to meet various objectives such as high mobility, reduced DIBL, low leakage current, and minimal V_{th} roll-off. However, a transistor that is optimized for superthreshold logics might not be suitable to use in weak inversion (sub-threshold).

Although, it would be ideal to have a dedicated process technology optimized for subthreshold circuits this is currently not available. In order to design optimal subthreshold circuits using CMOS devices that are targeted for superthreshold operation, it is crucial to develop design techniques that can utilize the side effects that appear in this new regime.

SCE (or V_{th} roll-off) is an undesirable phenomenon in short channel devices where V_{th} decreases as the channel length is reduced [131]. Variation in device critical dimensions translates into a larger variation in the threshold voltage as SCE worsens with increasing DIBL.

Traditionally, non-uniform HALO doping is used to mitigate this problem by making the depletion widths narrow and hence reducing the DIBL effect. As a byproduct of HALO, a short channel device shows RSCE behavior where the V_{th} decreases as the channel length is increased. In subthreshold circuits, the SCE mechanism is not as strong as in superthreshold circuits. On the other hand, RSCE is still significant enough to affect the subthreshold performance, therefore in subthreshold design PMOS devices has the same speed as NMOS transistors or even higher. Due to changing characteristics of devices to reduce the SCE for superthreshold design, working in subthreshold region gives somehow different behaviors. So, variations in both PMOS and NMOS devices should be considered. By applying FBB technique, there is a contention between some effects to increase or decrease the threshold voltage. As known, the FBB technique is used to reduce the threshold voltage for high speed applications, but for short channel devices for sub 100nm design, this effect is alleviated. By applying FBB to a short channel device, depletion width is decreased. It means that threshold voltage is increased in one side and on the other hand threshold voltage is decreased. Because the effect of SCE is mitigated then it lowers the current through the channel. This technique also improves DIBL due to lower depletion width of drain and source, even though, this SCE effect improvement is dependent on the applied substrate voltage. As it was shown through simulations, minimum channel length devices at very low temperature and in FF corners are the worst case. In FF corner due to mobility degradation the effect of body biasing technique is more. Also for very low temperature, there is the maximum DIBL. Therefore, the maximum deviation from the normal behavior of FBB technique is in this corner. For lower temperature, mobility degradation is improved. It means that ON current has not been degraded. As a result, the effect of FBB to increase the current is the dominant factor. Moreover, the DIBL effect in lower supply voltages is less due to reduced V_{DS}. If we make the effective channel length larger, the effect of SCE is decreased and then FBB has no effect on improvement in SCE effect. As a result, in any design where minimum length devices are used, optimal points for body biases should be found. As it was shown, this effect is the same for PMOS devices, but these techniques (FBB and RBB) are useful to be used for lower supply voltages and in some wireless applications.

5.4. Design Examples

To observe the effect of body-biasing in scaled technologies, some applications such as flip-flops are examined using body-biasing technique in subthreshold region.

5.4.1. Flip-Flops in Subthreshold

A. Hybrid Latch Flip Flop

Hybrid-latch flip-flop (HLFF) presented in [132] is one of the fastest structures presented. It also has a very small PDP [133]. The major advantage of this structure is its soft-edge property, i.e., its robustness to clock skew. One of the major drawbacks of the hybrid design in general is the positive hold time. Due to the single-output design, the power-consumption range of the HLFF is comparable with that of the static circuits. However, depending on the data pattern, the precharged structures can dissipate more than static structures for data patterns with more “ones.”

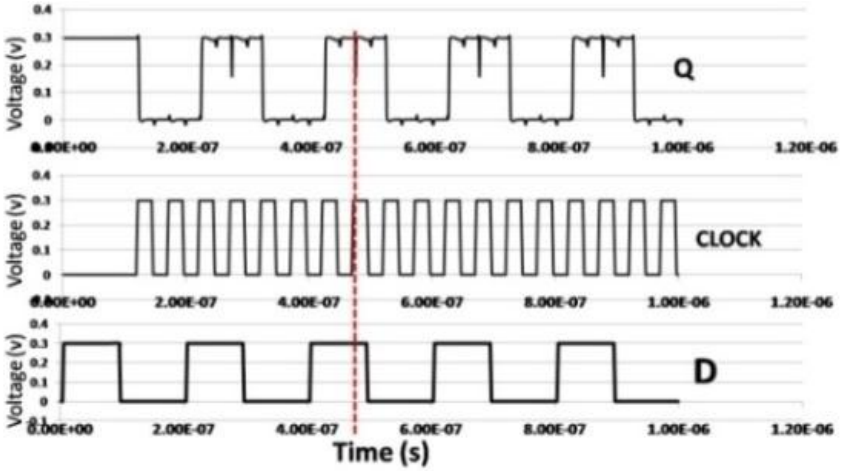


Fig.5.12. HLFF waveforms in V_{DD}=0.3V,T=27°C (TT)

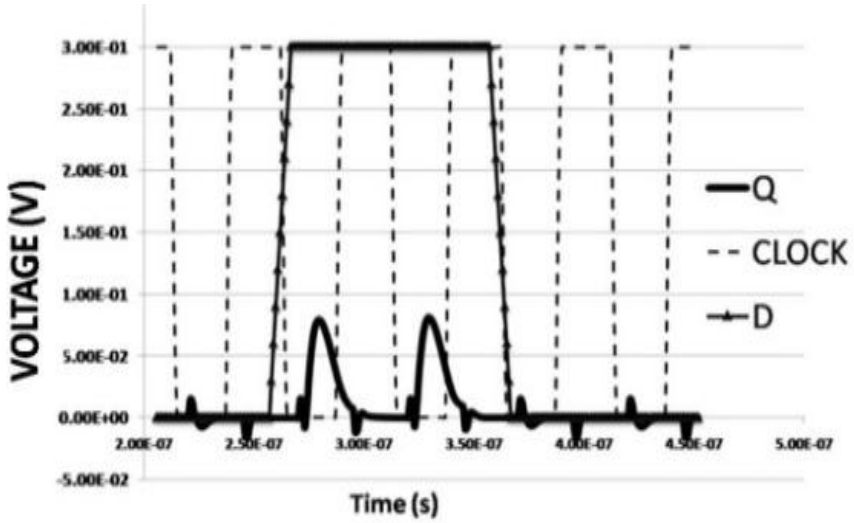


Fig.5.13. Failure in HLFF circuit

Hybrid design appears to be very suitable for high performance systems with little or no penalty in power when compared to classical static structures. As it can be seen in the HLFF circuit, there are stacked NMOS transistors that must evaluate the state of the circuit during the delay for three series inverters. As explained in the previous section, three stacked NMOS configuration has a lower speed than three stacked PMOS transistors in subthreshold region.

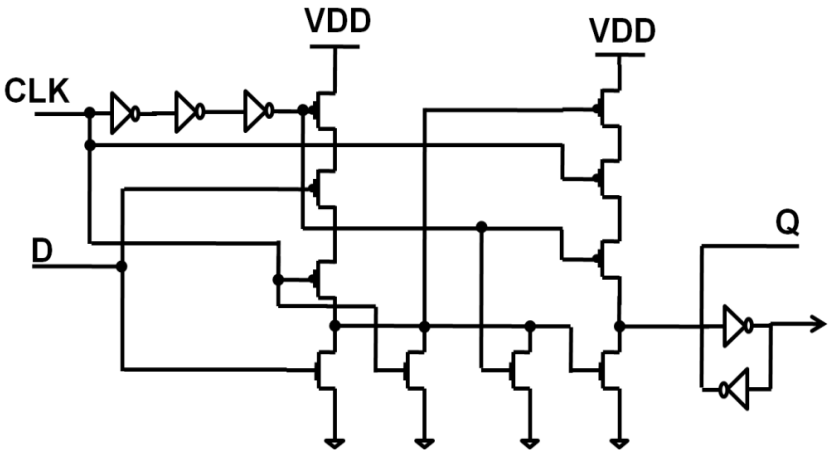


Fig.5.14. Proposed Complementary Hybrid Latch Flip Flop (CHLFF)

Simulations show that this circuit does not work for supply voltages lower than 0.4V. Fig.5.12 shows the results of simulations for HLFF at a supply voltage of 0.3V with higher size stacked NMOS transistors. When we reduce the supply voltage to 0.3V it causes some failures in corners. Fig.5.13 shows failure in output with two different inputs, due to many leakage paths and low active current. Due to the lower current in lower supply voltages, stacked NMOS transistors cannot discharge the related node.

Fig.5.14 shows the schematic of CHLFF (complementary hybrid latch flip-flop) for ultra low power applications. Because of employing PMOS stacked network, the speed of this circuit is higher. Also the technique of FBB is used to increase the speed of the PMOS network. After applying this technique the supply voltage may be reduced to 0.23V.

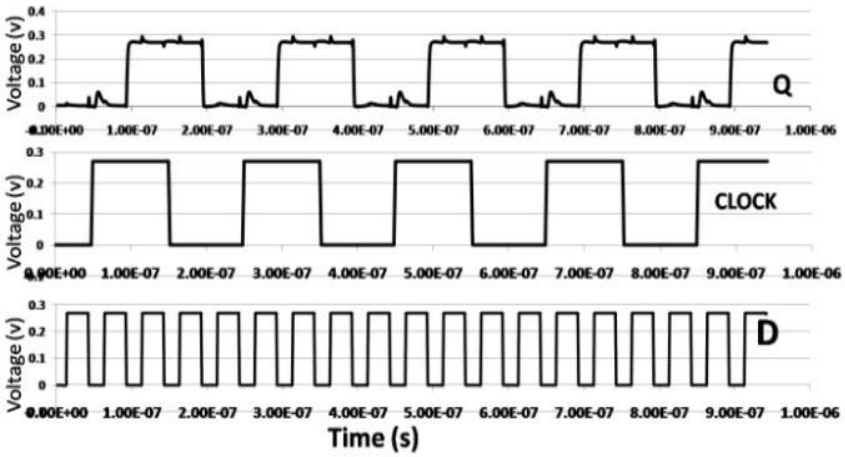


Fig.5.15. Output of CHLFF (TT model, T=27,V_{DD}=270mV)

Table5.2. Simulation results for HLFF (V_{DD}=0.4V)

Temp	FF	SS	TT	SF	FS
110	TCQ= -210ps	610ps	280ps	310ps	160ps
	TDQ=24.19ns	26.88ns	26.55ns	26.58ns	26.43ns
27	300ps	1.39ns	710ps	950ps	640ps
	26.57ns	27.66ns	26.98ns	27.22ns	26.94ns
-40	440ps	4.73ns	1.22ns	1.44ns	1.7ns
	26.71ns	31ns	27.59ns	27.71ns	27.97ns

Table.5.3. Simulation Results for CHLFF (V_{DD}=270mV)

Temp	FF	SS	TT	SF	FS
110	TCQ=59ps	442ps	320ps	1.1ns	500ps
	TDQ=38.81ns	39.19ns	39.07ns	40.21ns	39.75ns
27	300ps	1.28ns	712ps	400ps	1.124ns
	39.20ns	40.03ns	39.46ns	39.15ns	39.87ns
-40	730ps	7.28ns	1.794ps	1.1ns	6.159ns
	39.48ns	46.03ns	40.54ns	39.85ns	44.909ns

Table.5.4. Setup and Hold time for HLFF and CHLFF (V_{DD}=0.3V, TT Model)

FLIP FLOP	T_{SETUP}	T_{HOLD}	$T_{C=>Q\ (Rising)}$
HLFF	1.3ns	750ps	1.81ns
CHLFF	450ps	950ps	1.3ns

Simulation results are shown in Fig.5.15. In this case the proposed circuit is working properly with an area which is two times less than its counterpart for supply voltages close to 0.27V (with two times lower total width compared with conventional HLFF).

Table 5.2 and 5.3 describe the results for HLFF and CHLFF at different supply voltages for five different corners. These results show that the HLFF is fast enough for higher supply voltages, but it fails for lower supply voltages. The HLFF circuit operates properly for V_{DD}=0.4V but with high sized devices that increases the area overhead of this circuit significantly, so it makes it impractical for ultra low supply voltages. However the CHLFF is working as expected for supply voltage even lower than 0.3V and this ability makes this circuit employable in ultra low power applications. Simulations show that the CHLFF circuit has a much higher speed at higher supply voltages like 0.4V. For instance, at the SS corner at Temp=-40, $T_{c\text{-}q}$ is as low as 930ps while HLFF has a $T_{c\text{-}q}$ larger than 4.7ps with higher sized transistors.

FBB is the technique which may be used here to enforce the CHLFF to operate for even lower supply voltages as low as 0.2V. As mentioned in the previous section, for the PMOS transistors of conventional HLFF, increasing the bulk voltage from zero to V_{DD} decreases the current through the device for supply voltages lower than 0.3V. Therefore, if this technique is applied in the proposed circuit combined with higher size devices, the proposed flip flop may operate at supply voltage near 0.2V. As it can be seen in Table.5.4, the speed is increased but for the falling edge a lower speed is observed. It must be taken into account that the simulations are in the worst case. The HLFF cannot operate properly in some corners (like SS, T=-40), the lower supply voltage limit for the HLFF qualifying in all corners is 370mV given that very high sized devices are used.

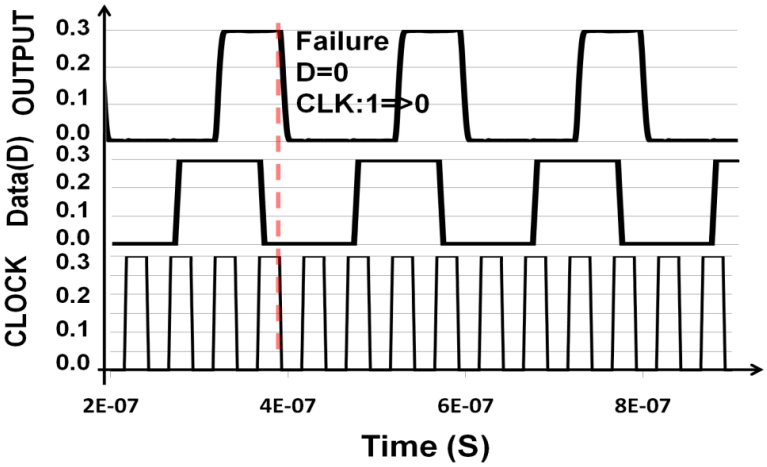


Fig.5.16. Failure in SAFF

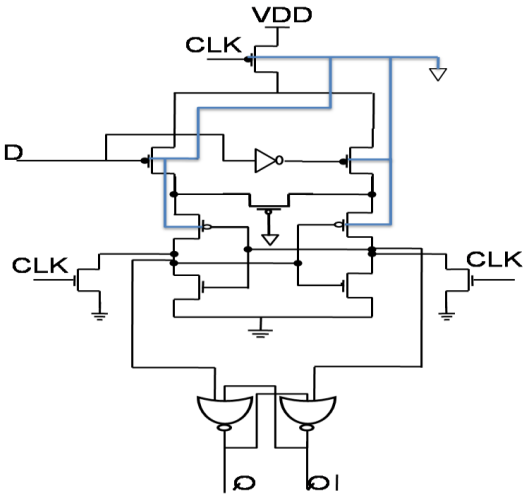


Fig.5.17. Schematic of CSAFF circuit

B. Sense Amplifier Based Flip Flop

The sense-amplifier based flip-flop (SAFF), initially proposed in [134]-[135], is one of the most effective flip-flop available. It consists of a fast differential sense-amplifier stage, followed by a slave latch[136]. The sense-amplifier stage can be seen as a latch whose sampling window closes as soon as the stage switches. This guarantees that the circuit is able to switch independent of circuit sizing. Furthermore the SAFF gives near-zero setup time and reduced hold-time. The main drawback of the SAFF proposed in [133]-[134] is the slave element, composed by a SR NAND latch. While this circuit requires a minimum number of transistors, it results in asymmetrical delays with slow high-to-low clock-to-output propagation. A high speed slave latch for the SAFF has been proposed by Nikolic *et*

al. in [135]. In their design, there is a performance gain. However, it is achieved at the cost of having an increased number of transistors, with 16 MOS devices required by the output stage. For these designs, the main problem is operating in very low supply voltage and as a result the very low active current that causes failure in SAFF in some cases (medium size devices for stacked NMOS network).

As mentioned before, at ultra low supply voltages, stacked PMOS devices show a higher speed than stacked NMOS devices. Thus, a new complementary circuit for application in the SAFF circuit which is optimized for ultra low power applications is proposed.

Fig.5.16 shows the simulation results for the SAFF at low supply voltages. SAFF to operate well must have high size NMOS transistors to evaluate the discharging nodes (S_bar and R_bar) that it in turn decreases the I_{on}/I_{off} ratio. Due to this lowered ratio, data retention in SAFF decreases significantly. For instance, when input D (CLK=low) is low, due to the leakage through upsized NMOS evaluation network, the R_{bar} is discharged causing the Q signal switching to in zero.

Table. 5.5. Simulation results for CSAFF

Temp	FF (T_{C-Q})	SS	TT	SF	FS
110	- 230ps	1.18ns	340ps	10ps	570ps
27	620ps	4.32ns	1.87ns	1.46ns	2.3ns
-40	1.99ns	17.24ns	6ns	5ns	8.05ns

Table.5.6. Results for CSAFF and SAFF in V_{DD}=0.3v (T=27^{Oc}C, TT Models)

FLIP FLOP	T_{SETUP}	T_{HOLD}	$T_{C=>Q}$
SAFF	250ps	1.25ns	2.13
CSAFF	420ps	-210ps	1.6ns

Table.5.7. Results for FFs

D-Flip Flop	Power consumption (x 10^{-7})	W x L
SAFF	2.22	8.28(um^2)
CSAFF	1.44	3.76(um^2)
HLFF	2.07	8.52 (um^2)
CHLFF	1.59	6.9 (um^2)

Fig.5.17 shows the schematic of CSAFF for ultra low power applications. Simulations using 65nm standard CMOS models show that CSAFF has a much higher speed than SAFF and also it has the capability of working in ultra low supply voltage near 0.15V. Table 3 shows the delays (T_{c-q}) for CSAFF. To compare with the SAFF circuit, the CSAFF must be simulated for a higher supply voltage such as 0.4V. Simulations show that CSAFF has a lower delay (4-5 times) compared to the SAFF topology even at lower supply voltages.

Table 5.5, 5.6 and 5.7 show the simulation results for CSAFF compared with SAFF at V_{DD}=0.3V, including setup and hold time, are and power comparisons, for standard 65nm CMOS models at room temperature. Although the CSAFF has a lower effective area than

the SAFF circuit, the CSAFF shows better results for setup and hold times. Another drawback for SAFF at V_{DD}=0.3 is its failure in slow corners (SS models, T=-40°C) while the CSAFF is working properly for supply voltages even less than 0.3V with a satisfactory performance.

The FBB technique was also applied to the CSAFF by connecting the bulks of the

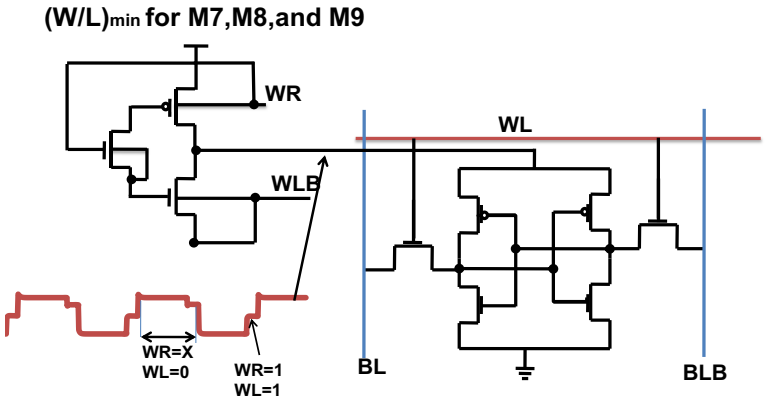


Fig.5.18. Proposed SRAM design

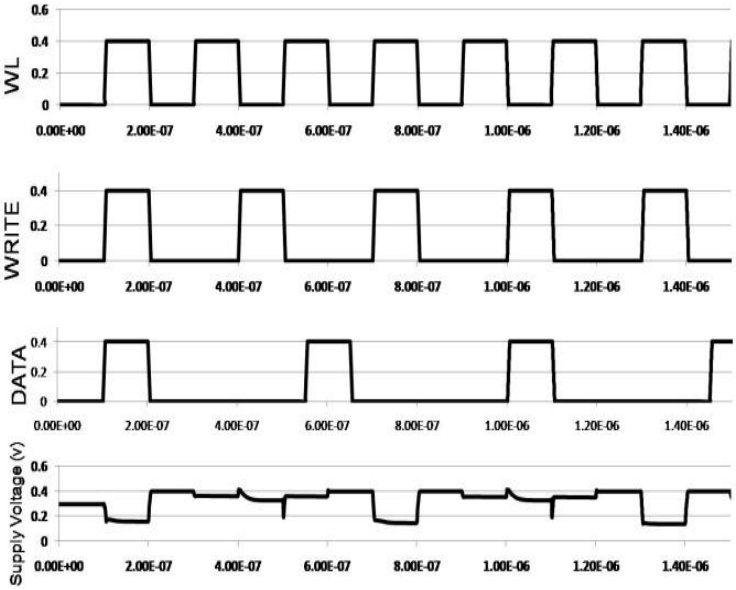


Fig.5.19. Supply voltage waveform

stacked PMOS network (shown with dashed lines in Fig.5.17) to ground which helps to increase the speed of the circuit by 1.2 times. This technique also works for supply voltages near 0.25V by applying the bulk voltage. In this section we showed that due to the higher drive current for PMOS, it is more useful to use a PMOS as evaluation network for

some kind of topologies such as FFs, domino logic circuits, and even SRAM design. As an example, due to lower drive current of NMOS than PMOS, using a PMOS network to evaluate write cycle would be more useful. As we showed through simulations, stacks of NMOS have a lower speed than PMOS stacks.

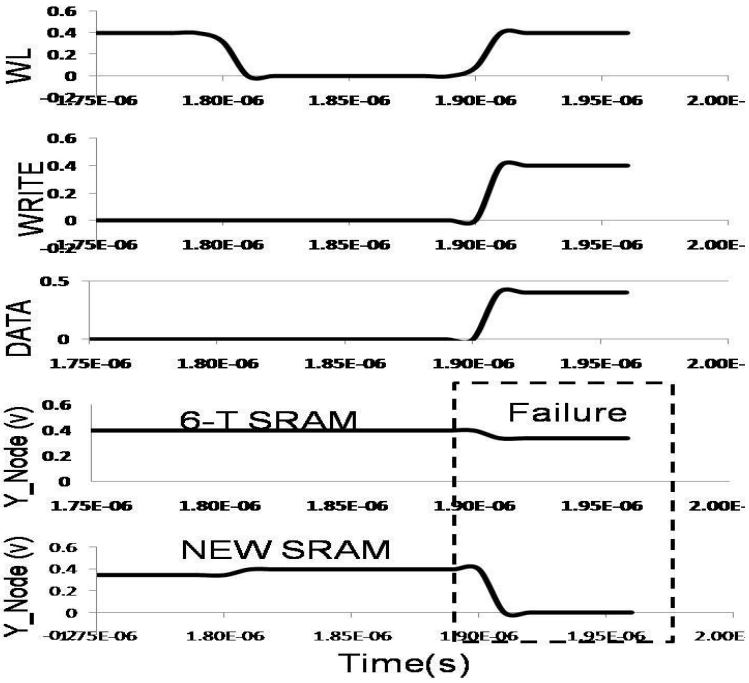


Fig.5.20. Comparison of the proposed technique with conventional 6T-SRAM cell during write cycle

5.4.2. Write Cycle Improved SRAM Design using body-biasing

In this part, we propose an adaptive circuit using body biasing effect to increase the write cycle speed of the SRAM circuit. Fig.5.18 shows the schematic of the proposed topology used in a SRAM design. Proper write operation depends on sizing the access NMOS to win the scaled fight with the PMOS inside the bitcell to write a "0". For a successful write, the bitcell becomes monostable, forcing the internal voltages to the correct values. If the cell retains bistability then the write does not occur, and the SNM is positive on the cell's butterfly plot. Thus, a negative SNM indicates a successful write (mono-stability in the cell) [140]. For ultra low voltage operation, the PMOS transistors are faster than NMOS (the same size) by an order of magnitude, and the write cycle speed decreases due to faster PMOS devices than access NMOS transistors. The best way to increase the write speed is upsizing the NMOS access transistors, but this in turn has an area overhead giving a deteriorative effect on READ and HOLD cycle.

However if we use an adaptive supply voltage during the WRITE cycle that decreases the supply voltage to weaken the PMOS transistors, write evaluation is improved though it has no effect on READ and HOLD, because during the READ cycle and HOLD time the supply voltage is connected to V_{DD}. As it can be seen in Fig.5.19, the supply voltage is

changed due to DATA, WRITE, and WL signals. When WL="1" and WRITE is high, then supply voltage decreases to $V_{DD}/2$ that enable us to decrease the supply voltage for SRAM design even more. This is because the most important constraint that limits designers to work in ultra low supply voltages is write failure due to contention between PMOS transistor and NMOS access transistors. To clarify this technique during the write cycle, dashed square in Fig.5.20, describes the results for the proposed technique and a 6T-SRAM cell with constant supply voltage.

5.5. Conclusions

In this chapter, body-biasing technique was investigated in details. The main result of these analyses is that FBB and RBB techniques have a different behavior in short channel devices compared to long channel devices. Moreover, we found that for any design that uses small channel length devices, it is necessary to find an optimal point to get the best performance or lowest power consumption. For higher supply voltages with short channel length devices, it is better to use body biasing techniques in reverse direction. FBB can be used to increase the threshold voltage and RBB will be used to decreased the threshold voltage and for high speed applications.

Furthermore, the main reason we find that sub-V_{th} current is higher for PMOS than NMOS is that often PMOS are designed to have lower $|V_{th}|$ (which is set by designing the threshold adjust implant of the devices). Above-V_{th}, this helps to compensate their lower mobility, resulting in nearly equal NMOS/PMOS drive; but, in subthreshold, it results in over compensation due to the stronger dependence of drive current on Vth. An additional factor might be that, typically, PMOS devices have higher channel doping concentration. As a result, their channel depletion region in subthreshold is thinner, and therefore results in larger depletion capacitance. As a result, the gate voltage, which affects the channel voltage through a C_{OX}-$C_{depletion}$ capacitive divider, has more effect on the channel voltage, possibly resulting in greater transconductance. Based on these results, in this chapter, a few design challenges for ultra low supply applications in 65nm CMOS technology were presented. Also we employed body biasing and stack effect techniques for flip flop designs for ultra low power applications. Simulation results showed that the setup time for CHLFF is improved by 65% but the hold time is degraded by %12 compared with HLFF design. However the speed of CHLFF is improved by 3X times more than HLFF topology. In CSAFF the speed of circuit is improved by 2X times for high to low of output. Also the effective area of this circuit is 3 times lower than SAFF.

Moreover, new technique based on the FBB technique is proposed. Using this technique, working at ultra low supply voltages is possible for high speed circuits like domino logic circuits. Also, due to the importance of SRAM in processor design, this technique was employed in SRAM design, significantly increasing the speed of write cycle with lower power dissipation due to lower supply voltages. In domino logic circuits, this technique extremely reduces the contention between PMOS (precharge transistor and keeper transistor) transistors and the evaluation network. It is thus not compulsory to upsize NMOS evaluation network to have a successful evaluation. Furthermore, the optimized bulk voltages were found giving a maximum speed for logic circuit designs.

Chapter 6

FinFET Circuit Design

Chapter 6

FinFET Circuit Design

Introduction

Until now we explained different challenges in nano-scale CMOS design especially for low voltage design. However due to significant amount of process variations in nano-scale CMOS technology, considering other devices for low power, high-performance and robust design is important. Aggressive device scaling has led to statistical variability and increased short channel effects (SCE) [141], [142]. Thinner gate oxide helps to improve the short channel effect. However, thinner gate oxide to improve SCE is not a viable option in nanometer scale nodes as it increases the gate leakage exponentially.

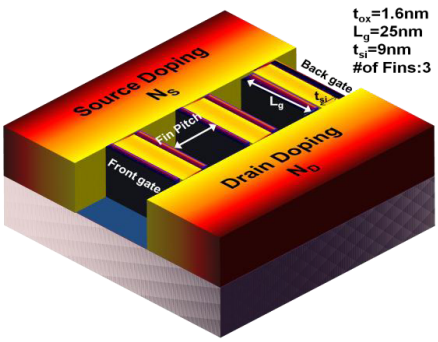


Fig.6.1. Asymmetric-Doping FinFET device

Thus, to overcome SCE, different candidate transistor structures have been investigated to replace the bulk MOSFETs [143-150]. Among them, FinFETs is considered to be a promising candidate for scaled CMOS devices in sub-22-nm technology nodes. This device shows increased immunity to SCE due to improved channel control by the gate voltage [151]. Furthermore, threshold voltage (V_{th}) can be easily controlled by engineering the gate

Table.6.1. FinFET device parameters

Parameter	Nominal Value
Supply Voltage	0.9 V
Physical gate length (L_{gate})	25 nm
Body thickness (t_{si})	9 nm
Physical oxide thickness (t_{oxf}=t_{oxb})	1.6 nm
Body Doping (intrinsic)	1e15 cm^{-3}
Source Doping	1e20 cm^{-3}
Drain Doping	1e19 cm^{-3}
NMOS gate Workfunction	4.63 eV
PMOS gate workfunction	4.84 Ev
Spacer length (Lsp)	20 nm
Spacer dielectric constant (k)	3.9

contact work function. Moreover, V_{th} variations due to random dopants in the channel region (RDF) are reduced due to almost intrinsic channel doping [152], [153].

There are different approaches to improve the electrostatics of the device. Such techniques are sizing, S/D underlap [154], high-K dielectrics [152], high-K spacers [155], and etc [156-158]. Due to the width quantization, transistor sizing is not an efficient approach to improve the cell robustness [159]. Hence, a well-defined device should have a good compromise between speed and leakage power. Using high-k dielectric oxide, the total gate capacitance of device is reduced due to the spatial redistribution of electric field. However, using high-k dielectric at the same oxide thickness value degrades both I_{off} and I_{on} compared to using SiO2 [154]. Upsizing the FinFET by increasing the fin width, improves the ON current, while off current is degraded. For FinFET devices, to increase the current more, increasing the number of fins is an approach. However, optimizing the fin pitch to achieve better electrostatics is important. For instance, at iso-fin width, by increasing the fin pitch by 4 times, total gate capacitance is increased by 50% [160]. Another technique to increase I_{on} current with no penalty on off current is using high-k spacer. This is achieved by some amount of inversion takes place in the underlap region and also because of the encroachment of the fringe fields from the gate through the high - κ dielectric. The coupling of the gate fringe field with the underlap region becomes stronger with increasing dielectric constant. Also, the off current is found to decrease at higher values of spacer κ. Hence, using high-k spacer seems to a promising technique to increase the Ion/Ioff ratio [164]. However, gate leakage is degraded by using this technique. Other techniques are source/drain extension and metal gate work function engineering. Hence, using a well-defined device structure for low leakage and robust design is extremely important.

The effect of leakage current is more pronounced in SRAMs. Furthermore, V_{th} variations due to Random dopant fluctuations (RDF) in sub-45nm standard CMOS devices can be one of the major bottlenecks in high density SRAM design. A high threshold voltage variability compromises SRAM cell stability. Due to the large number of transistors in an array, process variations become more significant, thereby different failures are introduced (such as read, write, and access. Furthermore, due to the increased leakage power in SRAM arrays, the introduction of new devices with lower leakage current with improved read and write margins is desired.

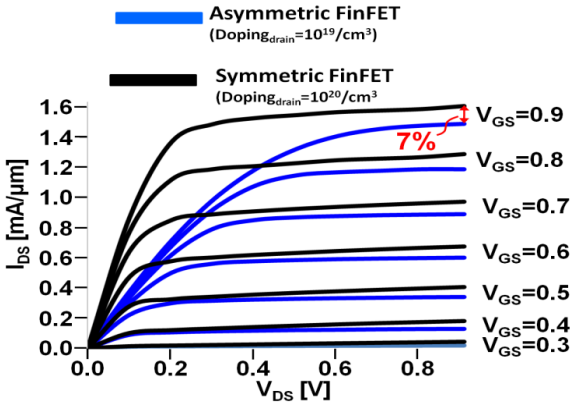
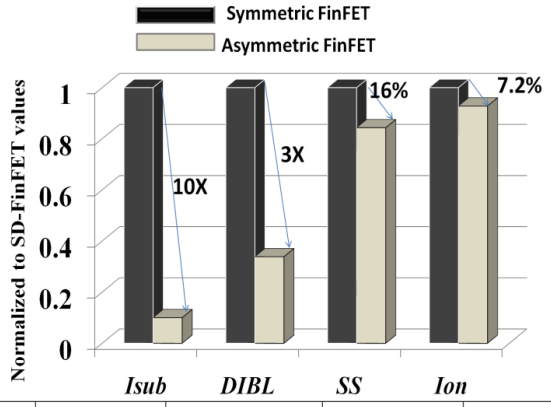


Fig.6.2. I_D-V_{DS} for AD-FinFET and Symmetric FinFET



Device	I_{sub}(A)	DIBL(mV/V)	SS(mV/dec)	Ion(A)
SD-FinFET	1.10E-08	70	83	1.61E-03
AD-FinFET	1.10E-09	22	70	1.49E-03

Fig.6.3. AD-FinFET versus Symmetric FinFET Device

Several techniques have been proposed to improve the SRAM operation by introducing optimized devices [161-164]. Although, these techniques improve I_d-V_g characteristic of the device, the trade-offs are not clearly addressed and improvement in conflict between read and write SNM is marginal. One such technique is the use of asymmetric drain underlap in FinFET [162, 163]. By applying this technique, the Drain-Induced Barrier Lowering (DIBL), sub-threshold swing (SS) and sub-threshold leakage current are improved, but I_{on} degradation is more than 20%,. Furthermore, utilizing Asymmetric spacer FinFET [164] in 6T-SRAM cell shows degradation in read SNM along with degradation in write margin compared to using symmetric spacer FinFET.
Consequently, introducing a new device with improved electrostatics with minimum degradation in I_{on} is desired. Furthermore, designing an SRAM with improved read SNM and write margin *is crucial* [165-168].

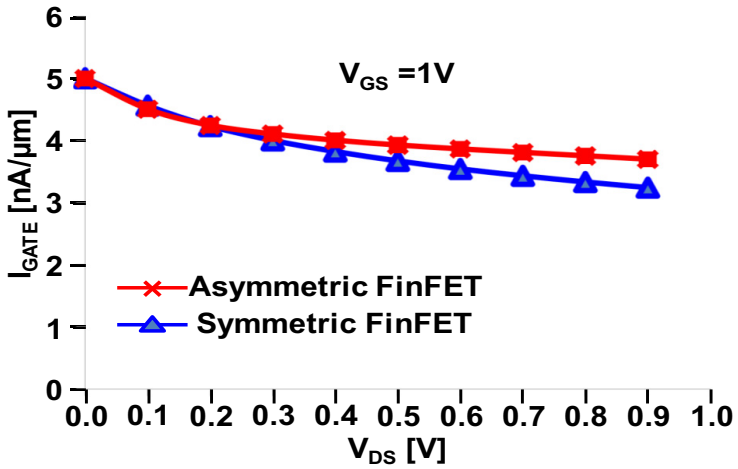


Fig.6.4. Igate vs. drain voltage

In this chapter, we represent a new technique doping the drain and source terminals, asymmetrically. Depending on the device biasing, it gives different currents. Based on that, we design a new FinFET SRAM cell to improve read and write margin for scaled technologies. Our proposed SRAM overcomes the trade-off between read and write margin, and allows simultaneous improvement of both by proper transistor/circuit configuration. Asymmetrically-doped (AD) FinFET SRAM can achieve (a) three times improved DIBL, (b) 10X less subthreshold-leakage, (c) 16% improved SS, and (d) read SNM

improvement by at least 20% with 9% degradation in I_{on} current. Furthermore, the use of asymmetric Drain/Source doping, leads to 9% improvement in write margin.

The remainder of this chapter is organized as follows: In section 6.1 we present the new asymmetric FinFET device explaining the benefits compared to the conventional symmetric FinFET. Conventional FinFET SRAM is briefly explained in section 6.2 in order to compare with the proposed AD-FinFET SRAM cell in terms of read and write SNM. The AD-FinFET based SRAM cell is proposed in section 6.3., and finally, the conclusions are drawn in section 6.4.

6.1. Asymmetric FinFET

The 3-D structure of the conventional FinFET device is shown in Fig.6.1. The physical size of the FinFET device is determined by the height (H_{fin}), length (L_g) and silicon thickness (t_{si}) of the fin. The device shown in Fig.6.1 consists of three vertical fins and each fin consists of asymmetric source-drain diffusion surrounded by the gate, and separated from it by the gate oxide. It is obvious, that the width of a FinFET is an integer multiple of H_{fin}. In addition, the drain and source contacts of the conventional FinFET device have the same doping concentration while the contacts of the Asymmetric-Doping (AD-FinFET) are doped at different concentration levels. This modification in contacts makes the device asymmetric to V_{DS} polarity. The purpose of using asymmetric doping will be explained in details in next section.

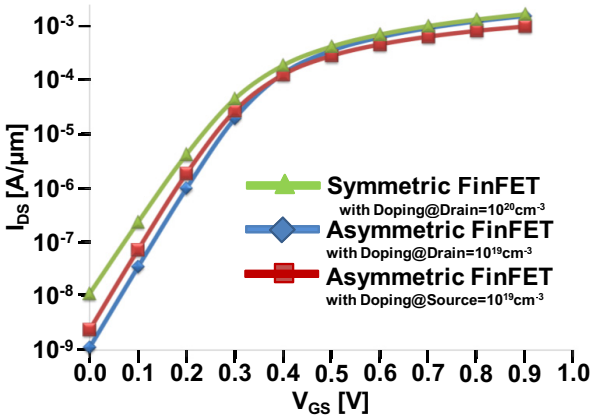


Fig.6.5. ID-VGS for AD-FinFET and symmetric FinFET

The symmetric and asymmetric FinFETs are simulated using the TCAD numerical simulator Taurus [169]. The parameters that we use to simulate both devices are shown in Table 6.1. From these simulations we extract current-voltage and capacitance-voltage characteristics for both AD- and SD-FinFET in order to compare their pros and cons.

The I_d-V_d characteristic curves for both asymmetric and symmetric devices are plotted in Fig.6.2 while in Fig.6.3 key figures of merit are calculated and compared with the conventional FinFET. We observe that AD-FinFET has lower ON current (I_{on}) while the leakage current is reduced significantly.

Specifically, I_{on} decreases by 7% and I_{off} one order of magnitude compared to SD-FinFET. We also observe that for DIBL of the proposed FinFET is improved 3 times and the sub-threshold swing is reduced by 16%.

By introducing the AD-FinFET device, gate leakage is degraded, marginally. Fig.6.4 shows the gate leakage as a function of V_{DS} for both devices. We observe that the AD-FinFET has slightly higher gate leakage current compared to the conventional FinFET. The reason of degradation in gate leakage is due to lower electric field from drain side. This lower electric field is attributed to lower doping at drain side and higher series resistance in extension length of AD-FinFET structure at drain side. Therefore, electric field from gate to drain is increased causing higher gate leakage. In another word, the gate leakage current is dominated by band-to-band tunneling which is proportional to $e^{-W_{dep}}$ (where W_{dep} is the depletion width). In this case, due to higher series resistance in extension length at drain side, voltage drop at the drain side is increased. As a result, depletion width is increased causing lower gate leakage current for SD-FinFET device.

To show the characteristics of the AD-FinFET, we simulate our device in two different structures, lower doping concentration at the drain side (referred to as LDD) and low doping concentration at source side (referred to as LSD). The simulation results are shown in Fig.6.5, where both LDD and LSD devices show much lower off current due to higher barrier height at zero gate bias compared to SD-FinFET device.

To see the effect of doping on I_{on} and I_{off} we simulated the AD-FinFET transistor for different doping components. The simulations results are shown in Fig.6.6 and Fig.6.7. By looking at the results, we can increase the number of dopants at the drain side to improve the I_{on} with a penalty of higher leakage current. The improvement in I_{on} can be ameliorated by 5% if we increase the number of dopants by five times while the capacitance at the drain side is increased. On the other side, we can increase the number of dopants at the source side by ten times to get around 10% increase in I_{on} current.

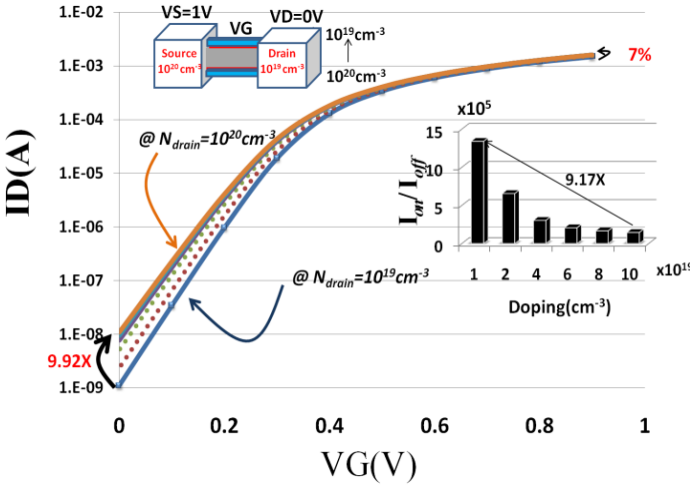


Fig.6.6. Effect of asymmetric FinFET on I_D-V_G characteristics by tuning the doping concentration in the drain contact

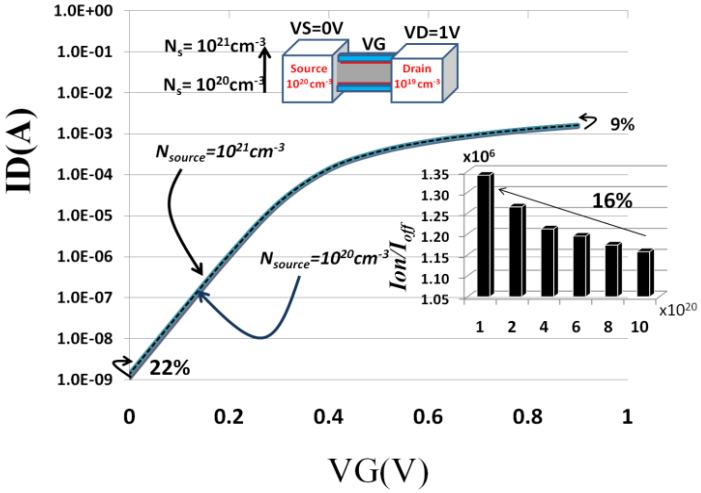


Fig.6.7. Effect of asymmetric FinFET on I_D-V_G characteristics by tuning the doping concentration in the source contact

Therefore, by increasing the doping concentration at the drain terminal of the AD-FinFET transistor to $10^{21}/cm^3$ we can achieve 10% improvement in I_{on} with a penalty of 22% degradation in leakage current. However, the technique still has seven times less leakage current compared to a SD-FinFET device. To get the maximum improvement in I_{on} and less leakage degradations, finding the optimal point is necessary.

The aforementioned improvements achieved by our device will be more comprehensive if we study the band diagrams plotted in Fig.6.8. Hence, the SD-FinFET shows higher leakage current because the top of the conduction band is lower compared to AD-FinFET (note that in Fig.6.8a, V_{GS}=0). This effect is attributed to high electric fields that occur for

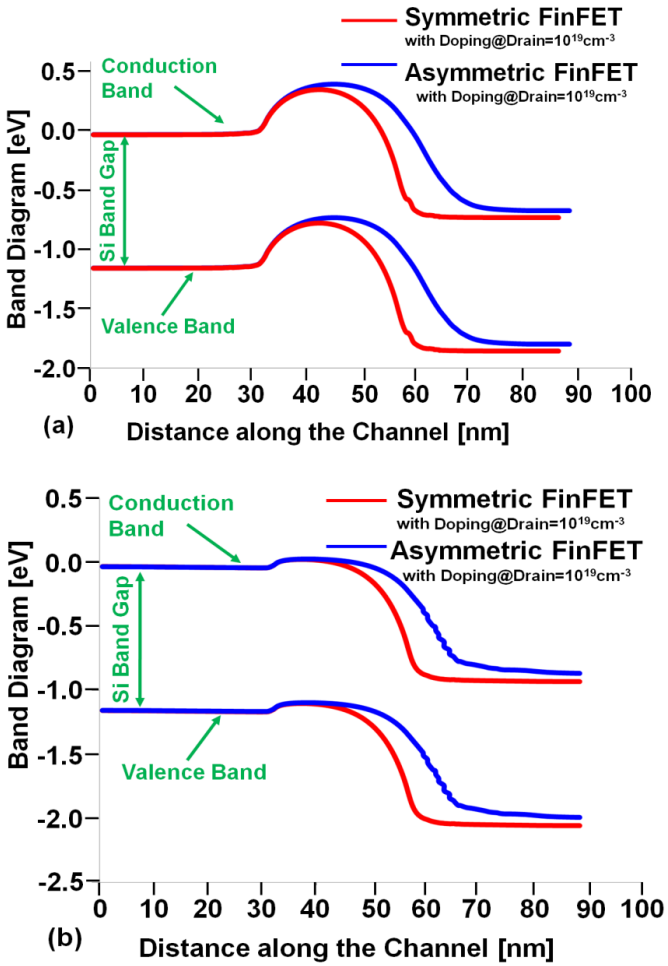


Fig.6.8. Conduction and valence bands for AD- and AS-FinFET: (a) V_{GS}=0, V_{DS}=0.9V and (b)V_{GS}=0.5V, V_{DS}=0.9V

higher doping concentration at the drain contact of the AD-FinFET. Consequently, Due to lower electric field from drain side, barrier height is less affected.

The dynamic behavior of a circuit is determined also by the parasitic capacitances of the corresponding devices. Running AC analysis at different bias points we extract the capacitances of AD-FinFET and SD-FinFET and we show the effect of doping concentration at the drain and source contacts. In Fig.6.9 to Fig.6.12, we demonstrate the capacitance components of the proposed device after performing AC simulations at different bias points. C_{GD} and C_{GS} are the capacitance components from gate to drain and source, respectively. C_{gg} is the total gate capacitance that is, $C_{GD} + C_{GS}$.

In Fig.6.9 the gate and source contacts are grounded and the drain voltage is swept. The change in the electric field due to voltage change in the drain contact does not affect the depletion width at the source side and hence the C_{GS} capacitance should be constant. This

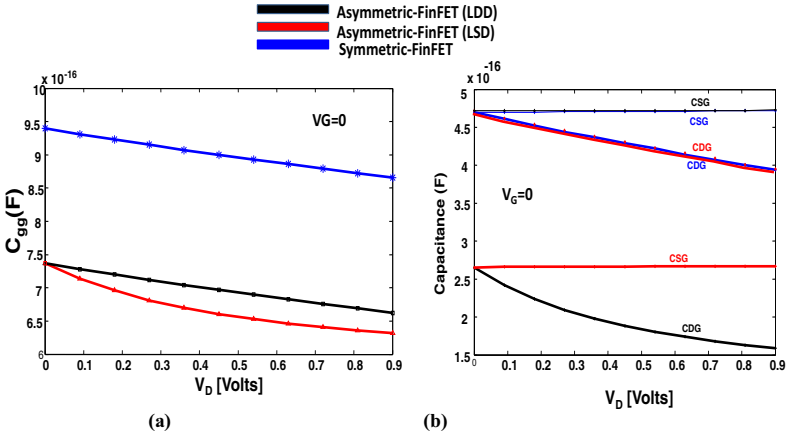


Fig.6.9. (a) Capacitance components and (b) total capacitance for the different device configurations when gate contact is set to zero V_G=0V and the drain contact is swept

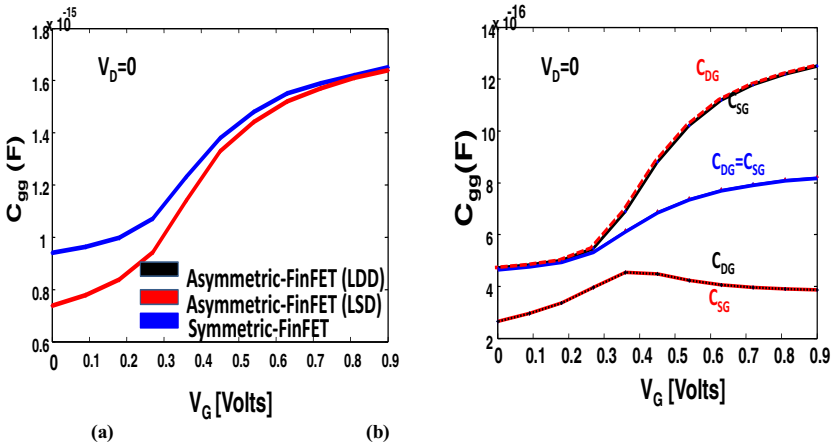


Fig.6.10. (a) Capacitance components and (b) total capacitance for different device configurations when gate contact is set to a constant voltage V_D=0 V and the gate contact is swept.

condition holds true for both AD- and SD-FinFET. Note that the doping concentration determines the depletion width inside the contact and the corresponding capacitance. For instance, the symmetric and LDD cases have same doping and hence, same capacitance. The LSD device shows smaller capacitance because lower doping results in larger depletion width. On the other hand, C_{GD} is affected by the V_D. Increasing V_D the depletion width increases and the capacitance becomes larger. Note that when V_D is equal to zero the C_{GD} and C_{GS} symmetric and LSD configurations are equal. Fig.6.9b shows the total gate capacitance. In Fig.6.10 the source and drain contacts are tied to ground. For the symmetric case the capacitances C_{GD} and C_{GS} will be equal sweeping the gate voltage. Increase of the gate voltage enhances the electric field from gate to S/D contacts and hence the depletion region in the contact decreases. This results to larger capacitances for increased V_G. These capacitances depend also on doping concentration because the depletion width in the contact depends on doping. This relation appears when we simulate the asymmetric device.

In Fig.6.11 we set the gate voltage to 0.9V and sweep the drain voltage. Note that for high Vg the transistor is ON and increasing the drain voltage electric field from drain to source penetrate the channel region and affect the depletion region. For the symmetric

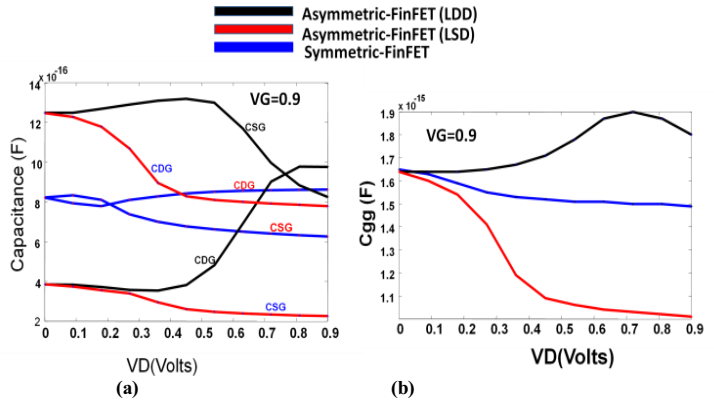


Fig.6.11. (a) Capacitance components and (b) total capacitance for different device configurations when gate contact is set to a constant voltage V_G=0.9V and the drain contact is swept

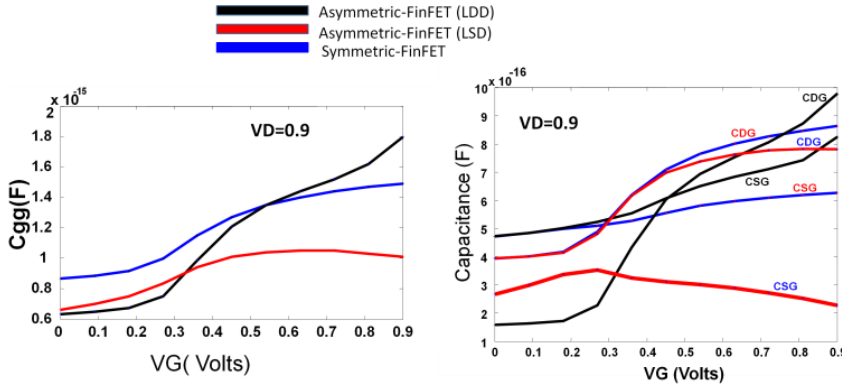


Fig.6.12. (a) Total capacitance (b) Capacitance components for different device configurations when drain contact is set to a constant voltage V_D=0.9V and the gate contact is swept

FinFET the increase of VD decreases the electric field from gate to drain and increases the electric field from drain to source. This electric field pattern results in smaller depletion width at drain and larger at source contact. Hence, C_{DG} increases and C_{SG} decreases.

6.2. FinFET SRAM Design

The layout and schematic diagram of a FinFET SRAM cell is shown in Fig.6.13. To investigate the SRAM challenges, let us consider the failure mechanisms in typical FinFET SRAM cell namely as read, write, and access time failures.

Read failure occurs during the access of the SRAM's internal storage nodes (L or R in Fig.6.13). When voltage at that stores '0' (V_{read}) exceeds a threshold value which is equal to the trip point voltage of the other inverter then the stored data are flipped.

Moreover, read failure probability is increased by either stronger AXR or weaker NR. Hence, to improve the read noise margin, lower ratio of AXR transistor to NR is desired.

On the other hand, write failures occur when the access transistors are not strong enough to change the stored data that is written on storage nodes. To improve SRAM write-ability, weaker PL and stronger AXL are desired. However, stronger access transistor increases the read failure probability. Access time failure is the inability of the cell to produce ΔV_{BL} high enough to be sensed by the sense amplifier.

In next section we propose a new SRAM design in which read, and write failure probabilities are reduced with a small penalty in ON current while the leakage current through access transistors are reduced significantly.

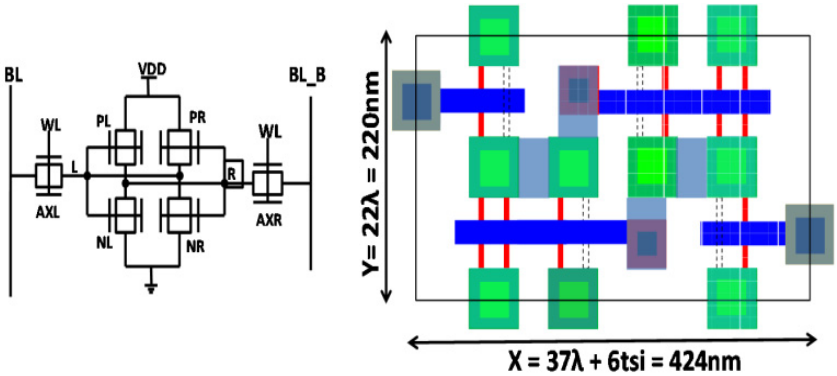


Fig.6.13. 6T-FinFET SRAM cell, layout and schematic (L=22nm)

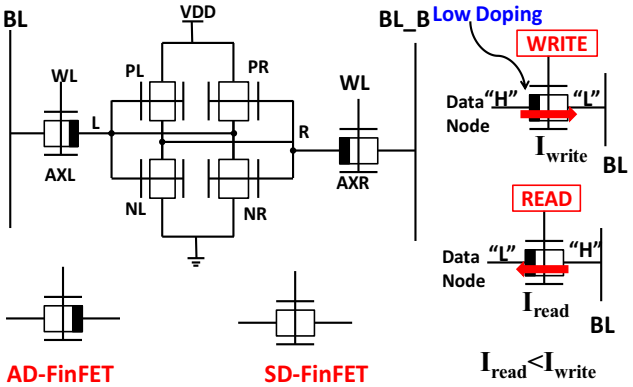


Fig.6.14. AD-SRAM FinFET Design

6.3. Asymmetric-Doping FinFET SRAM

Fig.6.14 represents the schematic diagram of a 6T-SRAM cell using asymmetric S/D doping for the access transistors. As it was shown, the AD-FinFET shows different current depending on forward and reverse current flow. To achieve a better SNM during read, higher

I_{pd}/I_{pg} is desired. By using AD-FinFET as access-transistor (access transistor), a small I_{read} or a large β ratio is achieved for the read operation. However access time is degraded due to the smaller I_{read}. During write operation, the PG transistor becomes stronger which improves the writeability by lowering the I_{pu}/I_{pg} ratio. As mentioned, the difference between write and read current through AD-FinFET is more than 50% for V_G=0.9V and

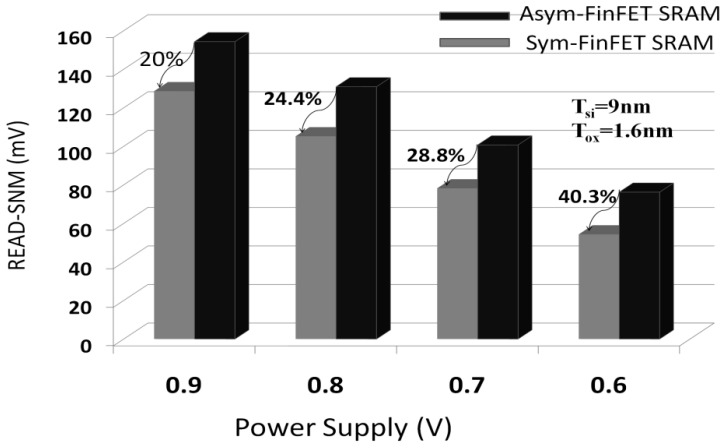


Fig.6.15. Read SNM results for AD-FinFET and SD-FinFET SRAM cells

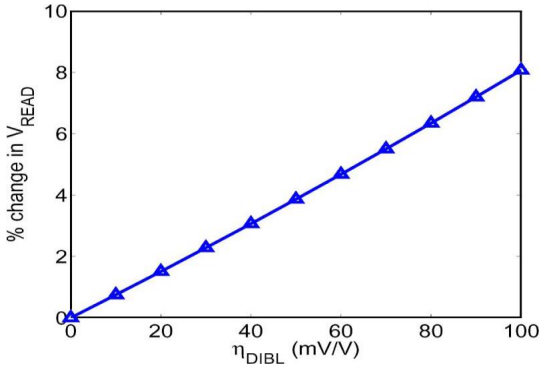


Fig.6.16. Variation in *V*READ with DIBL

VD=0.9V at Temp=27°C. Fig.6.15 shows the results for read static noise margin (SNM) for AD-FinFET SRAM compared to the conventional FinFET SRAM cell. As it can be seen due to the lower current through access-transistors, read SNM is improved. Due to the voltage divider action between right access transistor and pull down transistor, the voltage at node R increases to a positive value V_{read}, during read. By scaling the supply voltage, AD-FinFET shows better SNM compared to SD-FinFET SRAM cell. Improvement in SNM for low supply voltages (e.g. V_{DD}=0.5V) is attributed to decreased current through access transistors. However, degradation in read access time for lower supply voltages is inevitable

Furthermore, due to lower DIBL in access transistors, SNM is improved even more [163]. Fig.6.16 shows the effect of DIBL on V_{read} at different DIBL coefficients η_{DIBL}. Therefore, one important parameter to consider while designing devices for a robust SRAM is DIBL. AD-FinFET SRAM benefits from utilizing very low DIBL access transistors causing improved read SNM.

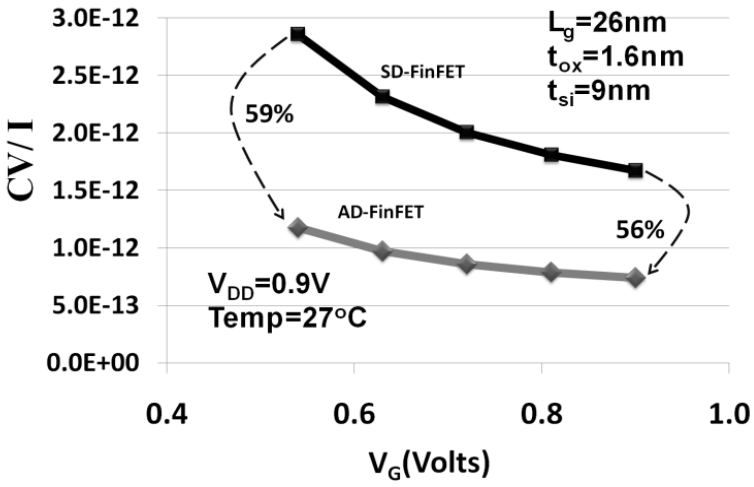


Fig.6.17. CV/I delay versus VG

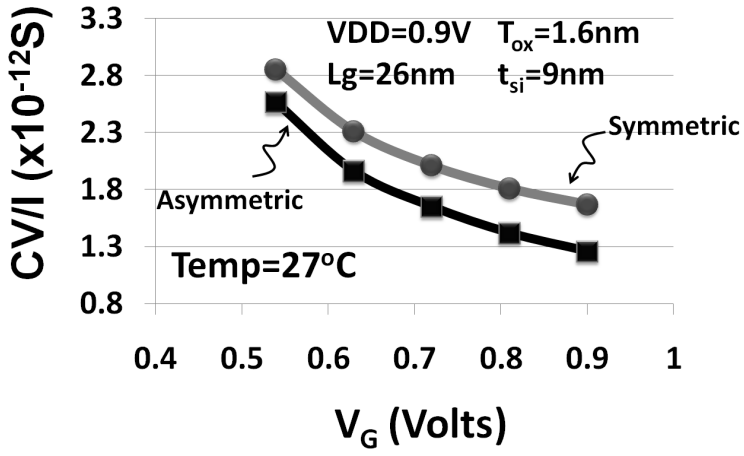


Fig.6.18. Comparison of T_{WL} delay for AD- and SD- FinFETs

In spite of improved read SNM, ON current is degraded in AD-FinFET SRAM. Therefore, let us look at CV/I delay for AD-FinFET SRAM compared to SD-FinFET SRAM design. To calculate the CV/I delay change due to using of AD-FinFET, we calculate the C_{sg} and C_{dg} for read and write cycles, respectively. As we have shown, C_{sg} is reduced significantly that helps to have a faster read depending on the ON current. In this case, Ion is degraded by 9% while Cgs is reduced significantly. As a result CV/I delay is improved significantly. Compared to using SD-FinFET device, simulations show 56% improvement in SRAM read access time by utilizing AD-FinFET device. The results for CV/I delay are shown in Fig.6.17. As it is illustrated due to lower I_{on} for AD-FinFET device, delay increases but at the same time capacitance is decreased significantly. As a result, the total delay is improved by 59% for low supply voltages to 56% for high supply

voltages. To improve the access time (lower delay), we can change the doping in drain/source terminals. However by considering the total gate capacitance of access transistor (C_{WL}) and calculating the T_{WL} (where T_{WL} is the delay from the wordline driver to the SRAM cell) that is given by:

$$T_{WL} = \frac{C_{WL}.V_{DD}}{I_{on.AX}} \quad (5.1)$$

where C_{WL} is:

$$C_{WL} = 2(N_C - 1)C_{g,AX} + C_{wl,\text{int}} \quad (5.2)$$

where Nc is the length of row (or number of columns), $C_{G,AX}$ is the gate input capacitance of the access transistors, and $C_{wl,int}$ is the wordline interconnect capacitance, we can see how we can reduce the access time. Another component of access time is the T_{BL} which is given by:

$$T_{BL} = \frac{(C_{BL} + N_r C_d) \times \delta V_{\min}}{I_{ON,PT} - \sum_{i=1,\ldots,N} I_{sub,PT}(i)} \quad (5.3)$$

where CBL is bitline capacitance, Cd is the drain capacitance of access transistor, and Nr is the number of rows, i.e., length of column. Since C_{BL} is constant, the optimization criterion can be approximated as $C_d/I_{ON,AX}$. Lower drain capacitance (C_d) reduces the access time. However lower access transistor on-current degrades access time. Simulation results show 56% reduction in T_{BL} for our design due to 10X lower leakage current and C_d. As shown in Fig.6.18, due to lower total gate capacitance (C_{gg}), T_{WL} of AD-FinFET SRAM cell improved by 25% compared to the standard FinFET SRAM. This is attributed to 30% reduced total gate capacitance of AD-FinFET during write (LDD) increasing the drivability of wordline driver. Note that, for typical wordline driver, as T_{BL} is dominant, total access time improvement is determined by T_{BL}. For large number of columns (e.g. 256) the total delay is dominated by T_{WL} and leads to improved total access time.

In this chapter we designed a SRAM cell using our proposed AD-FinFET device. Simulations results shows at least 56% improvement in access time due to significantly reduced capacitance at the drain side while Read and Write SNM are improved by at least 20% and 10%, respectively. Our SRAM design shows 59% improvement in low supply voltages (e.g. V_{DD}=0.5V), while SNM for read is improved by 43%.

6.4. Conclusions

In this chapter, we have proposed a design methodology for FinFET devices to achieve improved SRAM cells. Specifically, controlling asymmetrically the doping concentration of source and drain contacts we achieve appropriate transistor properties that are used to improve performance of the SRAM cells. The AD-FinFET replaces the access transistors of a conventional SRAM cell and hence it improves concurrently both read and write SNM. Finally with this technique we have achieved at least 20% read SNM and 9% write SNM. In addition our simulation results show 56% reduced access time compared to conventional FinFET SRAM cell.

Chapter 7

Future Research Trajectory

Chapter 7

Future Research Trajectory

Introduction

For the past decade, digital electronic industry has shown phenomenal growth due to CMOS devices scaling. CMOS technology scaling of transistor dimensions, has led to increased performance and integration. Increased density, leads to higher power density, including dynamic and leakage power, due to higher amount of devices and more switching in circuits.

Today's applications such as wireless sensor nodes, PDAs, cell phones, and medical devices demands lower power consumption. [170][171]. In mobile battery-powered devices, high performance digital circuits with support of low-latency and low-power techniques have been adopted to meet the requirements for running a variety of multimedia applications. On the other hand, the process variation sensitivity increases. Process variation effects are more stringent in small size devices. Variations in channel length, channel width, oxide thickness, threshold voltage, line-edge roughness, and random dopant fluctuations are the sources of the inter-die and the intra-die variations. Random variations in the number and location of dopant atoms in the channel region of the device result in the random variations in transistor threshold voltage. Therefore, robust, low power, and high performance digital design is the main issue in future digital design in nanoscale technologies.

7.1. Contribution of This Work

The contribution of this work is as follows:

A. Memory Design:

Memory design presents an extreme example of corner-based design. To satisfy the functionality of several tens of millions of SRAM cells, the designer has to capture even 5_ or 6_ standard deviations of parameter variations. This is becoming increasingly challenging to satisfy, and may present a problem for continued scaling of memory density. Concurrently, high-end microprocessors use large amount of on-die cache memory arrays to improve the overall performance. Hence, to achieve the highest possible packing density with high parametric yield in bulk-CMOS and SOI technologies, designers use a combination of multi-layered ad-hoc and heuristic techniques that include device sizing, supply and threshold voltage selection, SRAM column height and sense-amplifier optimization, and redundant columns and error correction

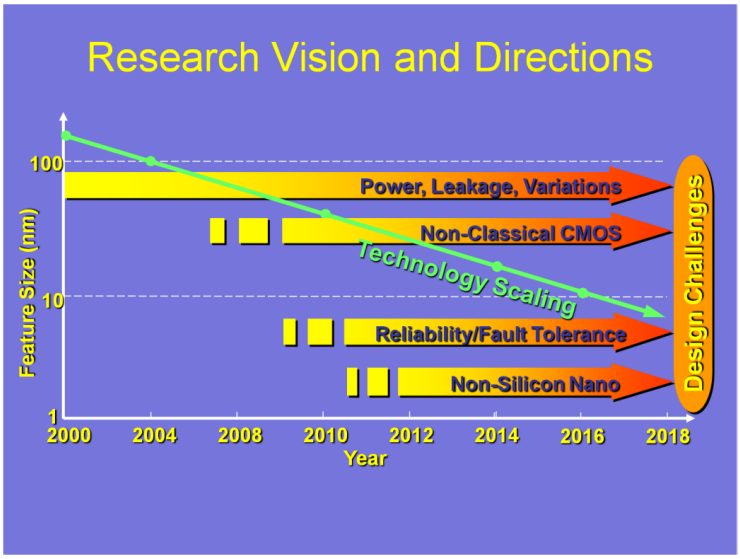


Fig.7.1. Research vision and directions

techniques. Power reduction has also become one of the most challenging design issues in every application domain. We have proposed several designs in SRAM cell modifications, using techniques to improve SRAM parameters such as Static Noise Margin (SNM), leakage power, and access time. Furthermore, we have introduced a multi-level wordline driver scheme to improve 6T SRAM read and write stability [172-174].

B. Low power Digital Circuit/System design:

In this category, our focus has been on designing basic digital blocks such as full adders and Flip-Flops and other logic circuits such as domino logic circuit, domino-based comparator design, and multiplexers. Lowering the power consumption of single flip-flop helps to improve the total power dissipation of some applications like FIR filter and DCT application. In this area, we evaluate different power-reduction techniques in sub- to super-threshold design for portable devices [175]. Due to the dependency of power consumption to switching activity of data, we tried to employ this feature to design new designs to lower the total power of some DSP applications [176].

C. Device Level design:

Aggressive device scaling has led to statistical variability and increased short channel effects (SCE). Thus to overcome SCE , different candidate technologies have been investigated for the nano-scale regime .FinFETs are considered to be a promising candidate for scaled CMOS devices in the sub-22nm technology node. The device shows increased immunity to SCE. The transistor threshold voltage, the Vth can be easily controlled by gate contact work function engineering. Moreover, Vth variations due to random dopant in the channel region (RDF) are reduced due to almost intrinsic channel doping. We introduced an asymmetric doping FinFET device that simultaneously increases read and write margins. Specifically, our FinFET device, exhibits 9X increase of the I_{on}/I_{off} ratio and 3X improved short channel effect 177, 178].

D. Reliability in nanoscale devices

Reliability problems are coming into effect. One of these emerging reliability issues is aging effects that result in device performance degradation over time. NBTI (Negative biased temperature instability) is a well-known aging phenomenon, which is a limiting factor for future scaling of devices. NBTI results the generation of trapped charges, which cause threshold voltage degradation of PMOS. In our research work, we analyzed the combined effect of NBTI, process variation and temperature on the reliability, in terms of SNM, of 6T SRAM cell in 65nm CMOS process. For process variations, we considered the die to die threshold voltage variation. SNM is affected by

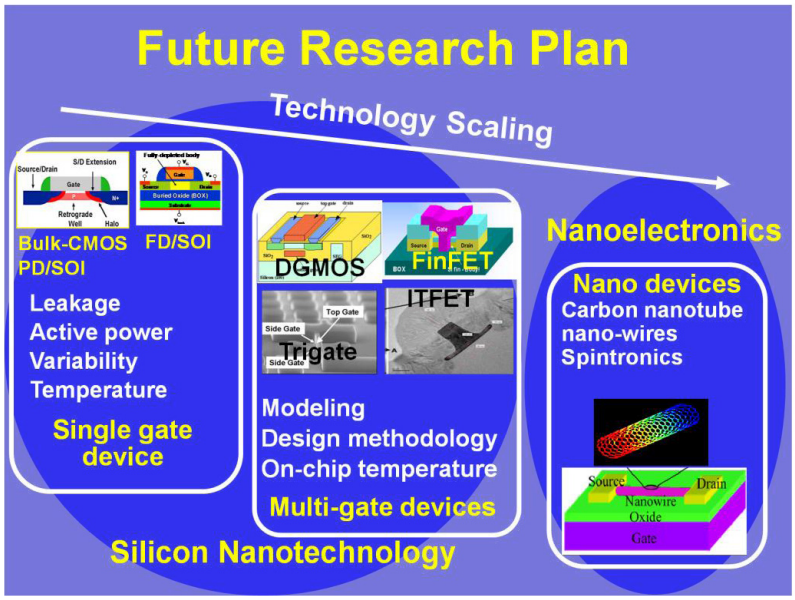


Fig.7.2. Research plan in device level

both the static process variations and the environmental variations such as temperature as well as aging effects such as NBTI. We use a voltage source in series with one of the PMOS transistor in SRAM cell to see the changes of NBTI effect. Also we considered the effect of hot carrier injection (HCI) on ultra thin body devices such as FinFET (Double-gate devices) [179, 180].

7.2. Future Work

Fig.7.1. illustrates a research plan for future research. The research vision of this project is to design new circuit architectures that can achieve greater performance in future CMOS technologies, leveraging the benefits of sealing's advantages. Working on new techniques to enable circuits operating at ultra low supply voltages less than 200mV for biomedical applications, is one of the most important goals for future research. Furthermore, due to the aggressive transistor scaling, process variation effects are non-interestingly increased. As a result to enable a digital design to operate in such a low voltage (e.g. 200mV) in presence of huge process variations (Random dopant fluctuation) is an interesting challenge. In biomedical applications, with very low frequency, lower energy is desired. To achieve lower power consumption, we need to introduce specified circuit design for such applications. Introducing new specified digital circuit design for ultra low voltage applications to be utilized in wireless sensor network nodes for bio-inspired applications is one of the main research topic for future plan. Introducing new challenges in nano-scale technologies and corresponding reliability issues is another hot topic of research. In device level, we have proposed some techniques to improve HCI effect in ultra thin body devices. However, considering new techniques to overcome reliability issues have been limited solely to single device level or circuit level. However, as CMOS scaling continues, we believe there is a need for a new design paradigm device/circuit co-design methodology leading to properly optimized circuits and systems considering new device innovations. As a part of this effort, we will be seeking collaboration with researches in device physics, and will propose an integrated design methodology to achieve adaptive fault/failure tolerant circuit/ system. Achieving new methods to optimize memory arrays is one of the most

important challenges in electronic design. Recently presented memory designs such as Spin Torque Transfer RAM (STT-RAM) shows promising potential to improve the area of memory arrays, speed, and power consumption. Working on Memory arrays is a part of the research plan for future. Fig.7.2 shows corresponding research plan on device level. As it can be seen it is a cumbersome to find the best replica for CMOS in terms of power, performance, and reliability.

REFERENCES

[1] S. Mukhopadhyay, K. Kang, H. Mahmoodi, and K. Roy, "Design of reliable and self-repairing SRAM in nano-scale technologies using leakage and delay monitoring," in International Test Conference, p. 10, 2005.

[2] K. Roy, "Leakage current mechanisms and leakage reduction techniques in deep-submicron CMOS circuits," in Proceedings of IEEE, pp. 305-327, 2003.

[3] S. Gunther, F. Binns, D. Carmean, , and J. Hall, "Managing the impact of in-creasing microprocessor power consumption," in Intel Technology Journal, vol. 5,p. 1, 2001.

[4] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, , and D. Tarjan, "Temperature aware micro-architecture," in International Symposium on Computer Architecture, pp. 2{13, 2003.

[5] C. Poirier, R. McGowen, C. Bostak, and S. Naffziger, "Power and temperature control on a 90nm itanium-family processor," in International Solid State Circuits Conference, vol. 1, pp. 304-305, 2005.

[6] Seongmoo Heo; Asanovic, K.; , "Load-sensitive flip-flop characterizations," VLSI, 2001. Proceedings. IEEE Computer Society Workshop on , vol., no., pp.87-92, May 2001

[7] N. Weste and K. Eshraghian, Principles of CMOS digital design.Reading, MA: Addison-Wesley, pp. 304–307.

[8] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low- power CMOS digital design," IEEE J. Solid-State Circuits, vol. 27, pp. 473–484,Apr. 1992.

[9] A. P. Chandrakasan and R.W. Brodersen, "Minimizing power consumption in digital CMOS circuits," Proc. IEEE, vol. 83, No. 4, pp. 498–523, Apr. 1995.

[10] A. Morgenshtein, A. Fish, I. A. Wanger," Gate-Diffusion Input (GDI):A power efficient method for digital combinational circuits", IEEE Tran.On Very Large Scale Integration (VLSI), Vol. 10, No. 5, 2002, pp. 566-581.

[11] W. C. Athas and N. Tzartzanis, "Energy Recovery for Low-Power CMOS", Proc. Of 16th Conference on Advanced Research in VLSI, 1995, pp. 415-429,

[12] Ricardo Gonzalez, Benjamin M. Gordon, and Mark A. Horowitz, "Supply and Threshold Voltage Scaling for Low Power CMOS", IEEE J. Solid-State Circuits, Vol. 32, No. 8, Aug. 1997, pp. 1210-1216.

[13] Q. Wu, P. Massoud, X. Yu, "Clock-Gating and Its Application to Low Power Design of Sequential Circuits," Proc. of the IEEE Custom Integrated Circuits Conference, 1997, pp. 425-435.

[14] V. Gutnik, A. Chandrakasan, "Embedded Power Supply for Low-Power DSP," IEEE Transactions on VLSI Systems, Vol. 12, pp. 425-435, Dec. 1997

[15] A. Wang, A. P. Chandrakasan, S. V. Kosonocky,"Optimal Supply and Threshold Scaling for Subthreshold CMOS Circuits ", Proc. Of IEEE Computer Society Annual Symposium on VLSI, 2002, pp. 5-9.

[16] R. Shalem, E. John and L. K . John, "A novel low power energy recovery full adder cell", Proc. Of the IEEE Great Lakes Symposium of VLSI, Feb. 1999, pp. 380-383.

[17] Yingtao Jiang, Abdulkarim Al-Sheraidah, Yuke Wang, Edwin Sha, and Jin-Gyun Chung, "A Novel Multiplexer-Based Low-Power Full Adder", IEEE Transactions on Circuits and Systems II: Express Briefs, Vol. 51, No. 7, JULY 2004, PP. 345-348.

[18] H. T. Bui, A. K. Al-Sheraidah, and Y. Wang, "Design and analysis of 10-transistor full adders using novel XOR-XNOR gates," Proc. Of Int. Conf. on Signal Processing (Wold Computer Congress), Beijing, China, Aug. 2000.

[19] Y. Jiang, Y.Wang, and J.Wu, "Comprehensive power Evaluation of Full Adders," Florida Atlantic Univ., BocaRaton, Tech. Rep., 2000.

[20] P. M. Lee, C.H. Hsu, and Y. H. Hung, " Novel 10-T full adders realized by GDI structure", Proc. on Intl. Symposium on Integrated Circuits (ISIC 2007), pp. 115-118.

[21] Mandeep Singh, Christophe Giacomotto, Bart Zeydel, and Vojin Oklobdzija, " Logic Style Comparison for Ultra Low Power Operation in 65nm Technology" Integrated Circuit and System Design, Power and Timing Modeling, Optimization and Simulation book, Vol. 4644, pp. 181-190, 2007.
[22] A. Wang, A. Chandrakasan, "A 180-mV Subthreshold FFT Processor Using a Minimum Energy Design Methodology", IEEE Journal of Solid-State Circuits, Vol. 40, No. 1, January 2005.
[23] T. Vigneswaran, B. Mukundhan, and P. Subbarami Reddy, "A novel lowpower, high speed 14 transistor CMOS full adder cell with 50% improvement in threshold loss problem," Enformatika Trans. Eng. Comp. Tech., vol. 13, pp. 82–85, May 2006.
[24] V. Moalemi, A. Afzali Koosha, "Subthreshold 1-Bit Full Adder Cells in sub-100 nm Technologies ", Proc. of the IEEE Computer Society Annual Symposium on VLSI, pp. 514-515, 2007.
[25] K.Roy, et al., "Leakage tolerant mechanisms and Leakage reduction techniques in deep-submicron CMOS circuits," Proceeding of the IEEE, vol. 91, pp.305-327, Feb. 2003
[26] S.H. Choi, and K. Roy, "Domino noise model: a new crosstalk noise model for dynamic logic circuits" Proc. IEEE Int. Conf. SOC (Systems-on-Chip).2003, pp. 207 - 210
[27] M.W. Allam, M.H. Anis, and M.I. Elmasry, "High-speed dynamic logic style for scaled-down CMOS and MTCMOS technologies," ISLPED, 2000, Page(s): 155-160
[28] A.Alvandpour, et. al., "A sub-130-nm conditional-keeper technique," IEEE JSSC , vol. 37, pp. 633-638, May 2002.
[29] Berkeley Predictive Technology Model. Univ. Berkeley, Berkeley, Available: http://wwwdevices.eecs.berkeley.edu/~ptm
[30] Moradi F., Peiravi A., Mahmoodi H.,"A High Speed and Leakage-Tolerant Domino Logic for High Fan-in Gates" in Proc. GLSVLSI 2005, Ilinois, Chicago. USA
[31] R. K. Krishnamurthy, A. Alvandpour, G. Balamurugan, N. R. Shanbhag, K. Soumyanath, and S. Y. Borkar, "A 130-nm 6-GHz 256 32 bit leakage-tolerant register file," IEEE J. Solid-State Circuits, vol. 37, pp. 624–632, May 2002.
[32] H.Mahmoodi-Meimand, Kauchic Roy, "A Leakage-Tolerant High Fan-in Dynamic circuit Design Style," IEEE Trans 2004
[33] Mahmoodi H., Roy K., "Diode-Footed Domino: A Leakage- Tolerant High Fan-in Dynamic Circuit Design Style" in IEEE Trans. On Circuits and Systems
[34] L.Wang, R.K Krishnamurthy, K. Soumyanath, N.R Shanbhag," An energy efficient leakage-tolerant dynamic circuit technique", ASIC 2000.
[35] J.Kao, "Dual Threshold Voltage Domino Logic," in proc. Eur: Solid State Circuit Conf., 1999, pp.118-121.
[36] V.kursun, E. Friedman, "Node voltage dependent subthreshold leakage current characteristics of dynamic circuits," IEEE Computer and Society Systems, 2004.
[37] B. Chatterjee, et. al, "Leakage control techniques for designing robust, low power wide-OR domino logic for sub-130nm CMOS technologies," IEEE Trans 2004 Computer and Society.
[38] S.O.Jung, K. W. Kim, and S. Kang, "Noise constrained power optimization for dual VT domino logic,"ISCAS, pp.158-161, 2001.
[39] F. Moradi, A. Peiravi, H. Mahmoodi-Meimand,"A new leakage tolerant for high fan-in domino gates," proceeding of International Conference on Microelectronics 2004.
[40] A. Chandrakasan, et al, "Design of high performance microprocessor circuit," IEEE Press, Piscataway, N.J., 2000

[41] G.Yang, Z. Wang, and Sung-Mo Kang, “Low power and high performance circuit technique for high fan-in dynamic gates,” 2004 IEEE Proc. Computer and Society.
[42] V. De, et al, “Technology and design challenges for low power and high performance,” ISLPED, pp. 163-168, 1999.
[43] V. Kursuan, E.G. Friedman, “Low swing dual threshold voltage domino logic”, Proc. on ACM Great Lakes symposium on VLSI (GLSVLSI’02), pp. 47-52, 2002.
[44] S. Borkar, “Low Power Design Challenges for the Decade,”Proceedings of the IEEE/ACM Design Automation Conference, pp. 293-296, June 2001.
[45] P. Srivastava, A. Pua, and L. Welch, “Issues in the Design of Domino Logic Circuits,” Proceedings of the IEEE Great Lakes Symposium on VLSI, pp. 108-112, February 1998.
[46]H. Mostafa, M. Anis, and M. Elmasry, , "Novel Timing Yield Improvement Circuits for High-Performance Low-Power Wide Fan-In Dynamic OR Gates," Circuits and Systems I: Regular Papers, IEEE Transactions on , vol.PP, no.99, pp.1, 0
[47] Appadwedula, T.; et al.;, “Total System Minimization for wireless image transmission”, VLSI Signal Processing, vol 27, 2001, pp. 99-117
[48] Lan, T.; et al.;, “Adaptive low power multimedia wireless communications”, CISS 1997, pp. 377-382
[49] Lueder, E.; , "Generation of equivalent block parallel digital filters and algorithms by a linear transformation," Circuits and Systems, 1993., ISCAS '93, 1993 IEEE International Symposium on , vol., no., pp.495-498, 3-6 May 1993
[50] Jongsun P.; Woopyo J.; Mahmoodi, H.; Yongtao Wang; Choo, H.; Roy, K.; , "Computation sharing programmable FIR filter for low-power and high-performance applications," Solid-State Circuits, IEEE Journal of , vol.39, no.2, pp. 348- 357, Feb. 2004
[51] Bai-Sun K.; Sam-Soo K.; Young-Hyun J.; "Conditional-capture flip-flop for statistical power reduction," Solid-State Circuits, IEEE Journal of, vol.36, no.8, pp.1263-1271, Aug 2001
[52] Kong B. S.; Kim S. S.; Jun Y. H.; , "Conditional-capture flip-flop technique for statistical power reduction," Solid-State Circuits Conference, 2000. Digest of Technical Papers. ISSCC. 2000 IEEE International, vol., no., pp.290-291, 465, 2000
[53] Kumar, R.; Bollapalli, K.C.; Garg, R.; Soni, T.; Khatri, S.P.; , "A robust pulsed flip-flop and its use in enhanced scan design," Computer Design, 2009. ICCD 2009. IEEE International Conference on , vol., no., pp.97-102, 4-7 Oct. 2009
[54] Gerosa, G.; Gary, S.; Dietz, C.; Dac Pham; Hoover, K.; Alvarez, J.; Sanchez, H.; Ippolito, P.; Tai Ngo; Litch, S.; Eno, J.; Golab, J.; Vanderschaaf, N.; Kahle, J.; , "A 2.2 W, 80 MHz superscalar RISC microprocessor," Solid-State Circuits, IEEE Journal of , vol.29, no.12, pp.1440-1454, Dec 1994.
[55] Suzuki Y., Odagawa K., and Abe T., “Clocked CMOS calculator circuitry,” IEEE J. Solid-State Circuits, vol. SC-8,pp. 462-469, Dec. 1973.
[56] Klass, F.; , "Semi-dynamic and dynamic flip-flops with embedded logic," VLSI Circuits, 1998. Digest of Technical Papers. 1998 Symposium on , vol., no., pp.108-109, 11-13 Jun 1998
[57] Nikolić B., et al., “Sense amplifier-based flip-flop,” in ISSCC Dig. Tech. Papers, Feb. 1999, pp. 282-283.
[58] H. Partovi, R. Burd, U. Salim, F. Weber, L. DiGregorio, and D. Draper, “Flow-through latch and edge-triggered flip-flop hybrid elements,” IEEE International Solid State Circuits Conference, ISSCC 1996, Feb. 1996, pp.138–139.
[59] Markovic, D.; Nikolic, B.; Brodersen, R.W.; , "Analysis and design of low-energy flip-flops," Low Power Electronics and Design, International Symposium on, 2001. , vol., no., pp.52-55, 2001

[60] Matsui, M.; Hara, H.; Uetani, Y.; Lee-Sup Kim; Nagamatsu, T.; Watanabe, Y.; Chiba, A.; Matsuda, K.; Sakurai, T.; , "A 200 MHz 13 mm^2 2-D DCT macrocell using sense-amplifying pipeline flip-flop scheme," Solid-State Circuits, IEEE Journal of , vol.29, no.12, pp.1482-1490, Dec 1994
[61] Kawaguchi, H.; Sakurai, T.; , "A Reduced Clock-swing Flip-flop (RCSFF) for 63% Clock Power Reduction," VLSI Circuits, 1997. Digest of Technical Papers., 1997 Symposium on , vol., no., pp.97-98, 12-14 Jun 1997
[62] Rabaey, J. M.; Chandrakasan, A.; and Nikolic, B.; "Digital Integrated Circuits A design Prespective , " second edition, Prentice Hall, 2002.
[63] Tschanz, J.; Narendra, S.; Zhanping Chen; Borkar, S.; Sachdev, M.; Vivek De; , "Comparative delay and energy of single edge-triggered and dual edge-triggered pulsed flip-flops for high-performance microprocessors ," Low Power Electronics and Design, International Symposium on, 2001. , vol., no., pp.147-152, 2001
[64] Zyuban, V.; , "Optimization of scannable latches for low energy," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.11, no.5, pp. 778- 788, Oct. 2003
[65] Shi, K.; , "Area and power-delay efficient state retention pulse-triggered flip-flops with scan and reset capabilities," Computer Design, 2008. ICCD 2008. IEEE International Conference on , vol., no., pp.170-175, 12-15 Oct. 2008.
[66] Teh, C. K.; Hamada, M.; Fujita, T.; Hara, H.; Ikumi, N.; Oowaki, Y.; , "Conditional Data Mapping Flip-Flops for Low-Power and High-Performance Systems," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.14, no.12, pp.1379-1383, Dec. 2006.
[67] Xanthopoulos, T.; Chandrakasan, A.P.; , "A low-power IDCT macrocell for MPEG-2 MP@ML exploiting data distribution properties for minimal activity," Solid-State Circuits, IEEE Journal of , vol.34, no.5, pp.693-703, May 1999
[68] Park, J.; Choo, H.; Muhammad, K.; Roy, K.; , " Nonadaptive and Adaptive filter implementation based on sharing multiplication," ICASSP, June 2000.
[69] Jongsun, P.; Woopyo J.; Hunsoo C.; Mahmoodi-Meimand, H.; Yongtao W.; Roy, K.; , "High performance and low power FIR filter design based on sharing multiplication," Low Power Electronics and Design, 2002. ISLPED '02. Proceedings of the 2002 International Symposium on , vol., no., pp. 295- 300, 2002
[70] G. Mohapatra, D.; Karakonstantis, G.; and Roy, K.; "Significance Driven Computation: A Voltage-Scalable, Variation-Aware, Quality-Tuning Motion Estimator," IEEE International Symposium on Low Power Electronics and Design, 2009.
[71] Synopsys Design Compiler, www.synopsys.com
[72] ST Microelectronic library, www.st.com
[73] Bhaskaran, V.; et al.; , "Image and Video Compression Standard Algorithms and Architecture", Kluwer Academic Publishers, 1996.
[74] Banerjee, N.; Karakonstantis, G.; Roy, K.; , "Process Variation Tolerant Low Power DCT Architecture," Design, Automation & Test in Europe Conference & Exhibition, 2007. DATE '07 , vol., no., pp.1-6, 16-20 April 2007
[75] Karakonstantis, G.; Banerjee, N.; Roy, K,; , "Process-Variation Resilient & Voltage Scalable DCT Architecture for Robust Low-Power Computing," IEEE TVLSI, 2010.

References

[76] A. Chandrakasan, R. Brodersen, "Low Power Digital CMOS Design", Kluwer Academic Publishers, 1995
[77] V. Gutnik and A. Chandrakasan, "Embedded power supply for lowpower DSP," IEEE Trans. VLSI Syst., vol. 5, no. 4, pp. 425–435, Dec.1997.

[78] S. Romanovsky, A. Achyuthan, S. Natarajan, and W. Leung, “Leakage reduction techniques in a 0.13um SRAM cell”, Proceedingsof the 17th International Conference on VLSI Design (VLSID), 2004.
[79] H. Qin, D. Markovic, A. Vladimirescu, and J. Rabaey, “SRAM leakage suppression by minimizing standby supply voltage”, Proceedings of the 5th International Symposium on Quality Electronic Design (ISQED), pp. 55-60, 2004.
[80] Benton H. Calhoun, Alice Wang, Anantha Chandrakasan, “Modeling and Sizing for Minimum Energy Operation in Subthreshold Circuits”, IEEE Jour. of SOLID-STATE CIRCUITS (JSSC), VOL. 40, NO. 9, Sep. 2005, pp. 1178-1786
[81] O. Thomas, M. Reyboz, M. Belleville, “Sub-1V, Robust and Compact 6T SRAM cell in Double Gate MOS technology”, Proc. on IEEE International Symposium on Circuits and Systems(ISCAS), pp.2778 – 2781, May 2007.
[82] S. Yang, W. Wolf, N. Vijaykrishnan, Y. Xie and W. Wang, “Accurate Stacking Effect Macro-modeling of Leakage Power in Sub-100nm Circuits” Proc. of the 18th International Conference on VLSI Design, pp. 165 – 170, 2005.
[83] E. Seevinck, F. J. List, J. Lohstroh, “Static-Noise Margin Analysis of MOS SRAM Cells” IEEE Jour. of Solid-State Circuits, Vol. SC-22, No. 5, Oct. 1987,pp. 748-754.
[84] B. H. Calhoun, A. P. Chandrakasan,” Static Noise Margin Variation for Sub-threshold SRAM in 65-nm CMOS,” IEEE Jour. of Solid-State Circuits VOL. 41, NO. 7, July 2006,pp. 1673-1679.
[85] J. Kwong, A. P. Chandrakasan “VariationDriven Device Sizing for Minimum Energy Subthreshold Circuits”, IEEE Journal of Solid-State Circuits, vol. 41, no. 7, pp. 1673-1679, July 2006
[86] B. H. Calhoun, A. P. Chandrakasan, "A 256-kb 65-nm Sub-threshold SRAM Design for Ultra-low-Voltage Operation," IEEE Journal of Solid-State ircuits, vol. 42, no. 3, pp. 680-688, March 2007.
[87] F. Moradi, Dag T. Wisland, H. Mahmoodi, A. Peiravi, S. Aunet, and T. V. Cao, “New subthreshold concepts in 65nm CMOS technology”, ISQED, pp. 162-166,2009.
[88] E. Seevinck, F. List, and J. Lohstroh, “Static noise margin analysis of MOS SRAM cells,” IEEE J. Solid-State Circuits, vol. SC-22, no. 5, pp. 748–754, Oct. 1987.
[89] J. Wang ,S. Nalam , and B. H. Calhoun , “Analysing Static and Dynamic Write Margin for Nanometer SRAMs”, IEEE ISLPED, pp.129-134, 2008.
[90] A. Bhavnagarwala et al. Fluctuation limits & scaling opportunities for CMOS SRAM cells. IEDM, 659-662, 2005.
[91] K. Zhang et al. A 3-GHz 70-Mb SRAM in 65-nm CMOS technology with integrated column-based dynamic power supply. IEEE J. Solid-State Circuits, 41(1):146-151, 2006.
[92] K. Takeda et al. Redefinition of write-margin for nextgeneration SRAM and write-margin monitoring circuit. ISSCC, 2602-2603, 2006.
[93] C. Wann et al. SRAM cell design for stability methodology. IEEE VLSI-TSA, 21-22, Aug 2004.
[94] E. Grossar et al. Read stability and write-ability analysis of SRAM cells for nanometer technologies. IEEE J. Solid-State Circuits, 41(11):2577-2588, Nov 2006.
[95] Sinangil, M.E.; Verma, N.; Chandrakasan, A.P.; , "A Reconfigurable 8T Ultra-Dynamic Voltage Scalable (U-DVS) SRAM in 65 nm CMOS," Solid-State Circuits, IEEE Journal of , vol.44, no.11, pp.3163-3173, Nov. 2009
[96] Hamzaoglu, F.; Ye, Y.; Keshavarzi, A.; Zhang, K.; Narendra, S.; Borkar, S.; Stan, M.; De, V.; , "Analysis of dual-V_T SRAM cells with full-swing single-ended bit line sensing for on-chip cache," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.10, no.2, pp.91-95, Apr 2002

[97] Hamzaoglu, F.; Ye, Y.; Keshavarzi, A.; Zhang, K.; Narendra, S.; Borkar, S.; Stan, M.; De, V.; , "Dual-V_T SRAM cells with full-swing single-ended bit line sensing for high-performance on-chip cache in 0.13 μm technology generation," Low Power Electronics and Design, 2000. ISLPED '00. Proceedings of the 2000 International Symposium on , vol., no., pp. 15- 19, 2000
[98] J.B. Kuang; H.C. Ngo; K.J. Nowka; S. Ehrenreich; A.J. Drake; J. Pille; S. Kosonocky; R. Joshi; T. Nguyen; I. Vo; , "The Design and Implementation of a Low-Overhead Supply-Gated SRAM," Solid-State Circuits Conference, 2006. ESSCIRC 2006. Proceedings of the 32nd European , vol., no., pp.287-290, Sept. 2006
[99] Morifuji, E.; Yoshida, T.; Kanda, M.; Matsuda, S.; Yamada, S.; Matsuoka, F.; , "Supply and threshold-Voltage trends for scaled logic and SRAM MOSFETs," Electron Devices, IEEE Transactions on , vol.53, no.6, pp. 1427- 1432, June 2006
[100] Kim, C.H.; Jae-Joon Kim; Ik-Joon Chang; Roy, K.; , "PVT-aware leakage reduction for on-die caches with improved read stability," Solid-State Circuits, IEEE Journal of , vol.41, no.1, pp. 170- 178, Jan. 2006
[101] Naveen Verma; Chandrakasan, A.P.; , "A 65nm 8T Sub-Vt SRAM Employing Sense-Amplifier Redundancy," Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International , vol., no., pp.328-606, 11-15 Feb. 2007
[102] Verma, N.; Chandrakasan, A.P.; , "A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy," Solid-State Circuits, IEEE Journal of , vol.43, no.1, pp.141-149, Jan. 2008
[103] Morimura, H.; Shibata, N.; , "A step-down boosted-wordline scheme for 1-V battery-operated fast SRAM's," Solid-State Circuits, IEEE Journal of , vol.33, no.8, pp.1220-1227, Aug 1998
[29] Raychowdhury, A.; Geuskens, B.; Kulkarni, J.; Tschanz, J.; Bowman, K.; Karnik, T.; Shih-Lien Lu; De, V.; Khellah, M.M.; , "PVT-and-aging adaptive wordline boosting for 8T SRAM power reduction," Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International , vol., no., pp.352-353, 7-11 Feb. 2010
[105] Wang, Y.; Bhattacharya, U.; Hamzaoglu, F.; Kolar, P.; Ng, Y.; Wei, L.; Zhang, Y.; Zhang, K.; Bohr, M.; , "A 4.0 GHz 291Mb voltage-scalable SRAM design in 32nm high-κ metal-gate CMOS with integrated power management," Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International , vol., no., pp.456-457,457a, 8-12 Feb. 2009
[106]Nii, K.; Yabuuchi, M.; Tsukamoto, Y.; Ohbayashi, S.; Oda, Y.; Usui, K.; Kawamura, T.; Tsuboi, N.; Iwasaki, T.; Hashimoto, K.; Makino, H.; Shinohara, H.; , "A 45-nm single-port and dual-port SRAM family with robust read/write stabilizing circuitry under DVFS environment," VLSI Circuits, 2008 IEEE Symposium on , vol., no., pp.212-213, 18-20 June 2008
[107] Yabuuchi, M.; Nii, K.; Tsukamoto, Y.; Ohbayashi, S.; Imaoka, S.; Makino, H.; Yamagami, Y.; lshikura, S.; Terano, T.; Oashi, T.; Hashimoto, K.; Sebe, A.; Okazaki, G.; Satomi, K.; Akamatsu, H.; Shinohara, H.; , "A 45nm Low-Standby-Power Embedded SRAM with Improved Immunity Against Process and Temperature Variations," Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International , vol., no., pp.326-606, 11-15 Feb. 2007
[108] Mukhopadhyay, S.; Rao, R. M.; Kim, J.-J.; Chuang, C.-T.; , "SRAM Write-Ability Improvement With Transient Negative Bit-Line Voltage," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.PP, no.99, pp.1-1, 0
[109] Yabuuchi, M.; Nii, K.; Tsukamoto, Y.; Ohbayashi, S.; Nakase, Y.; Shinohara, H.; , "A 45nm 0.6V cross-point 8T SRAM with negative biased read/write assist,"VLSI Circuits, 2009 Symposium on , vol., no., pp.158-159, 16-18 June 2009

[110] Hirabayashi, O.; Kawasumi, A.; Suzuki, A.; Takeyama, Y.; Kushida, K.; Sasaki, T.; Katayama, A.; Fukano, G.; Fujimura, Y.; Nakazato, T.; Shizuki, Y.; Kushiyama, N.; Yabe, T.; , "A process-variation-tolerant dual-power-supply SRAM with 0.179μm^2 Cell in 40nm CMOS using level-programmable wordline driver," Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International , vol., no., pp.458-459,459a, 8-12 Feb. 2009
[111] Nii, K.; Yabuuchi, M.; Tsukamoto, Y.; Hirano, Y.; Iwamatsu, T.; Kihara, Y.; , "A 0.5V 100MHz PD-SOI SRAM with enhanced read stability and write margin by asymmetric MOSFET and forward body bias," Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International , vol., no., pp.356-357, 7-11 Feb. 2010
[112] Fujimura, Y.; Hirabayashi, O.; Sasaki, T.; Suzuki, A.; Kawasumi, A.; Takeyama, Y.; Kushida, K.; Fukano, G.; Katayama, A.; Niki, Y.; Yabe, T.; , "A configurable SRAM with constant-negative-level write buffer for low-voltage operation with 0.149μm^2 cell in 32nm high-k metal-gate CMOS," Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International , vol., no., pp.348-349, 7-11 Feb. 2010
[113] Alorda, B.; Torrens, G.; Bota, S.; Segura, J.; , "Static and dynamic stability improvement strategies for 6T CMOS low-power SRAMs," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2010 , vol., no., pp.429-434, 8-12 March 2010
[114] Hyunwoo Nho; Kolar, P.; Hamzaoglu, F.; Yih Wang; Karl, E.; Yong-Gee Ng; Bhattacharya, U.; Zhang, K.; , "A 32nm High-k metal gate SRAM with adaptive dynamic stability enhancement for low-voltage operation," Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International , vol., no., pp.346-347, 7-11 Feb. 2010
[115] L. Chang, D. Fried, J. Hergenrother, J. Sleight, R. Dennard, R. R. Montoye, L. Sekaric, S. McNab, W. Topol, C. Adams, K. Guarini, and W.Haensch, "Stable SRAM cell design for the 32 nm node and beyond,"in Symp. VLSI Technology Dig., 2005, pp. 128–129.
[116] B. H. Calhoun and A. P. Chandrakasan, "A 256 kb subthreshold SRAM in 65 nm CMOS," in IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig.Tech. Papers, 2006, pp. 628–629.
[117] Kulkarni, J.P.; Kim, K.; Roy, K.; , "A 160 mV Robust Schmitt Trigger Based Subthreshold SRAM," Solid-State Circuits, IEEE Journal of , vol.42, no.10, pp.2303-2313, Oct. 2007
[118] McNairy, C.; Soltis, D.;, "Itanium 2 processor microarchitecture," in Micro, IEEE , vol.23, no.2, pp. 44- 55, March-April 2003
[119] Sharifkhani, M.; Sachdev, M.; , "SRAM Cell Stability: A Dynamic Perspective," Solid-State Circuits, IEEE Journal of , vol.44, no.2, pp.609-619, Feb. 2009
[120] Khalil, D.E.; Khellah, M.; Nam-Sung Kim; Ismail, Y.; Karnik, T.; De, V.K.; , "Accurate Estimation of SRAM Dynamic Stability," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.16, no.12, pp.1639-1647, Dec. 2008
[121] Y. Oowaki et al., "A sub-0.1μm circuit design with substrate-over-biasing," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, 1998, pp. 88–89.
[122] C. Wann et al., "CMOS with active well bias for low-power and RF/analog applications," in Symp. VLSI Technology Dig. Tech. Papers, 2000, pp. 158–159.
[123] M. Miyazaki et al., "A 1000-MIPS/W microprocessor using speed adaptive threshold-voltage CMOS with forward bias," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, 2000, pp. 420–421.
[124] R. K. Krishnamurthy, Alvandpour, A. De, V. Borkar, S., "High-performance and low-power challenges for sub-70 nm microprocessor circuits," in Proc. IEEE 2002 Custom Integrated Circuits Conf., Orlando, Florida, 12-15 May, 2002, pp. 125–128.

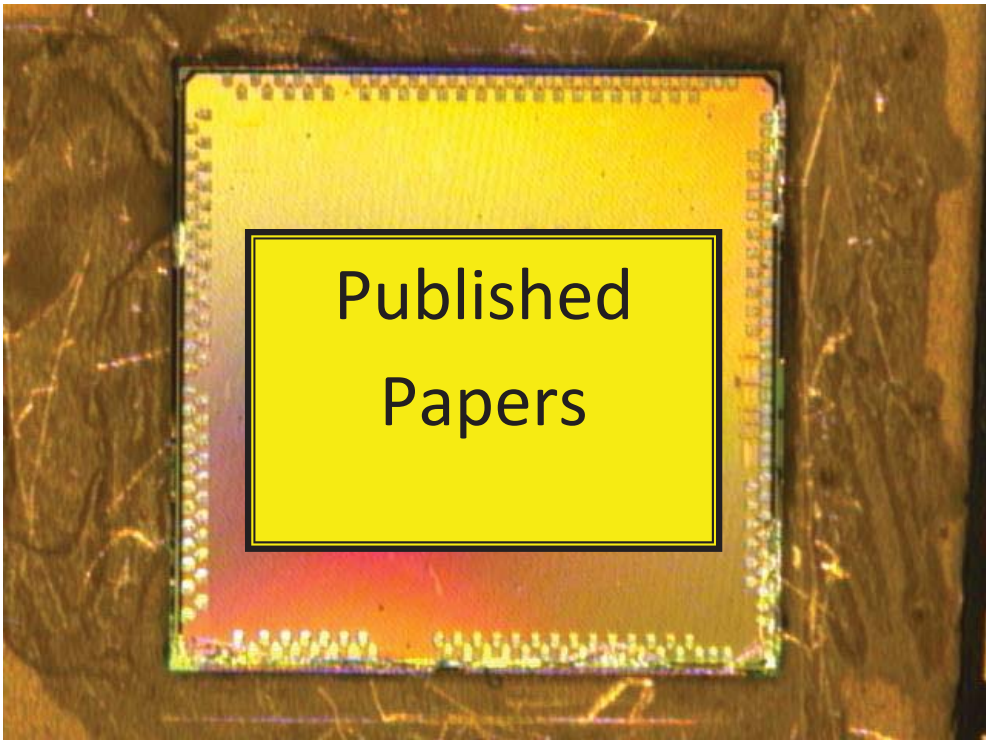
[125] Joseph F. Ryan, Jiajing Wang, and Benton H. Calhoun, “Analyzing and modeling process balance for sub-threshold circuit design”, Proc. of the 17th ACM Great Lakes symposium on VLSI,2007, Stresa-Lago Maggiore, Italy, pp.275 –280
[126] Z. Chen, M. Johnson, L. Wei, K. Roy, "Estimation of Standby Leakage Power in CMOS Circuits Considering Accurate Modeling of Transistor Stacks," Proceedings of the International Symposium on Low PowerElectronics and Design, Monterey, California, United States ,10-12 Aug. 1998, pp. 239-244
[127] Y. Ye, S. Borkar and V. De, "A New Technique for Standby Leakage Reduction in High- Performance Circuits," Symposium on VLSI Circuits, 11-13 June 1998, pp. 40-41.
[128] Siva Narendran, Shekhar Borkar, Vivek De, Dimitri Antoniadis,and Anantha Chandrakasann, “Scaling of Stack Effect and its Application for Leakage Reduction,” In Proceedings of the 2001 International Symposium on Low Power Electronics and Design, Huntington Beach, California, United States , 2001. pp.195-200
[129] J. Kao, S. Narendra and A. Chandrakasan, "Subthreshold leakage modeling and reduction techniques", Proc. of the International conference on Computer-aided design, p.141-148, November 10-14, 2002, San Jose, California
[130] M. Marin, M.J. Deen, M. de Murcia, P. Llinares and J.C. Vildeuil, “Effects of body biasing on the low frequency noise of MOSFETs from a 130nm CMOS technology” IEE Proceedings Circuits, Devices and Systems, pages: 95- 101, 2004
[131] M. Miyazaki, G. Ono, T. Hattori, K. Shiozawa, K. Uchiyama, K. Ishibashi, “A 1000-MIPS/W microprocessor using speed adaptive threshold-voltage CMOS with forward bias,” ISSCC, Digest of Technical Papers, pp. 420-475, February 2000.
[132] T. Kim, H. Eom, J. Keane, and C. Kim, “Utilizing reverse short channel effect for optimal subthreshold circuit design,” in Proc. Int. Symp. Low Power Electronics (ISLPED’06), Oct. 2006, pp. 127–130.
[132] H. Partovi, R. Burd, U. Salim, F. Weber, L. DiGregorio, and D. Draper,“Flow-through latch and edge-triggered flip-flop hybrid elements,” IEEE International Solid State Circuits Conference, ISSCC 1996, Feb. 1996, pp.138–139.
[133] V. Stojanovic and V.G. Oklobdzija ,”Comparative Analysis of Master–Slave Latches and Flip-Flops for High-Performance and Low-Power Systems”, IEEE Jounal of Solid State Circuits, Vol. 34, No.. 4, April 1999, pp.536-548.
[134] M. Matsui et al., “A 200 Mhz 13mm^2 2-D DCT macrocell using Sense-Amplifying Pipeline Flip-Flop Scheme”, IEEE J. Solid State Circuits, Vol.29, No. 12, Dec. 1994, pp.1482-1490.
[135] J. Montanaro et al., “A 160 Mhz 32-b 0.5-W CMOS RISC microprocessor”, IEEE J. Solid-state Circuits, Vol.31, No.11, pp.1703-1714, Nov. 1996.
[136] B. Nikolic, V. Stojanovic, V.G. Oklobdzija, W. Jia, J. Chiu, M. Leung, "Sense Amplifier-Based Flip-Flop," 1999 IEEE International Solid-State Circuits Conference, ISSCC'99, Digest of Technical Papers, San Francisco, CA, February 15-17, 1999, pp. 282-283.
[137] V. Kursuan, E.G. Friedman, “Low swing dual threshold voltage domino logic”, Proc. on ACM Great Lakes symposium on VLSI (GLSVLSI’02), pp. 47-52, 2002.
[138] S. Borkar, “Low Power Design Challenges for the Decade,”Proceedings of the IEEE/ACM Design Automation Conference, pp. 293-296, June 2001.
[139] P. Srivastava, A. Pua, and L. Welch, “Issues in the Design of Domino Logic Circuits,” Proceedings of the IEEE Great Lakes Symposium on VLSI, pp. 108-112, February 1998.
[140] B. Calhon, A. Chandrakasan,” A 256-kb 65-nm Sub-threshold SRAM Design for Ultra-Low-Voltage Operation”, Journal of Solid-State circuit, VOL. 42, NO. 3, pp. 680-688, March 2007.

[141] “International Technology Roadmap for Semiconductors,” 2006.
[142] K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, D. J. Pearson, and N. J. Rohrer, “High-performance CMOS variability in the 65-nm regime and beyond,” IBM J.
Res. Dev., vol. 50, no. 4/5, pp. 433–449, Jul./Sep. 2006.
[143] Collaert, N.; De Keersgieter, A.; Dixit, A.; Ferain, I.; Lai, L.-S.; Lenoble, D.; Mercha, A.; Nackaerts, A.; Pawlak, B.J.; Rooyackers, R.; Schulz, T.; Sar, K.T.; Son, N.J.; Van Dal, M.J.H.; Verheyen, P.; von Arnim, K.; Witters, L.; De Meyer; Biesemans, S.; Jurczak, M.; , "Multi-gate devices for the 32nm technology node and beyond," Solid State Device Research Conference, 2007. ESSDERC 2007. 37th European , vol., no., pp.143-146, 11-13 Sept. 2007
[144] K. Noda, T. Uchida, T. Tatsumi, T. Aoyama, K. Nakajima, H. Miyamoto, T. Hashimoto, and I. Sasake, “0.1 pm delta-doped MOSFET using post-energy implanting selective epitaxy,” in 1994- VLSI Symp. VLSI Technology Dig. Tech. Papers, pp. 19-20.
[145] R. Yan, A. Ourmazd, and K. Lee, “Scaling the Si MOSFET: From bulk to SO1 to bulk,” IEEE Trans. Electron Devices, vol. 39, pp. 1704-1710, July 1992.
[146] T. Hori, “A 0.1 pm CMOS technology with tilt-implanted punchthrough stopper (TIPS),” Int. Electron Devices Meet. Tech. Dig., 1994, pp. 75-78.
[147] F. Assaderaghi, D. Sinitsky, S. A. Parke, J. Bokor, P. KO, and C. Hu, “A dynamic-threshold MOSFET for ultra-low voltage operation,” Znt. Electron Devices Meet. Tech. Dig., 1994, pp. 809-812.
[148] T. Tanaka, K. Suzuki, H. Horie, and T. Sugii, “Ultrafast operation of Vi/th adjusted p+ n+ double-gate SO1 MOSFET’s,” IEEE Electron Device Lett., vol. 15, no. 10, pp. 386-388, Oct. 1994.
[149] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, T.-J. King, J. Bokor, and C. Hu, “A folded-channel MOSFET for deep-sub-tenth micron era,” in IEDM Tech. Dig., 1998, pp. 1032–1034.
[150] K. Sunouchi et al., “A surrounding-gate transistor (SGT) cell for 64/256Mhit DRAM’S,” Int. Electron Devices Meet. Tech. Dig., 1990, pp. 23-26.
[151] Taur Book
[152] B.-Y. Tsui and L.-F. Chin, “A comprehensive study of the FIBL of nanoscale MOSFETs,” IEEE Trans. Electron Devices, vol. 51, no. 10, pp. 1733–1735, Oct. 2004.
[153] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, “Sub-50 nm FinFET: PMOS,” in IEDM Tech. Dig., 1999, pp. 67–70.
[154] Trivedi, V.; Fossum, J.G.; Chowdhury, M.M.; , "Nanoscale FinFETs with gate-source/drain underlap," Electron Devices, IEEE Transactions on , vol.52, no.1, pp. 56- 62, Jan. 2005
[155] Sachid, A.B.; Manoj, C.R.; Sharma, D.K.; Rao, V.R.; , "Gate Fringe-Induced Barrier Lowering in Underlap FinFET Structures and Its Optimization," Electron Device Letters, IEEE , vol.29, no.1, pp.128-130, Jan. 2008
[156] Yang-Kyu Choi; Tsu-Jae King; Chenming Hu; , "Nanoscale CMOS spacer FinFET for the terabit era," Electron Device Letters, IEEE , vol.23, no.1, pp.25-27, Jan 2002
[157] Hui Zhao; Yee-Chia Yeo; Rustagi, S.C.; Samudra, G.S.; , "Analysis of the Effects of Fringing Electric Field on FinFET Device Performance and Structural Optimization Using 3-D Simulation," Electron Devices, IEEE Transactions on , vol.55, no.5, pp.1177-1184, May 2008
[158] Dong-Soo Woo; Jong-Ho Lee; Woo Young Choi; Byung-Yong Choi; Young-Jin Choi; Jong Duk Lee; Byung-Gook Park; ,"Electrical characteristics of FinFET with

vertically nonuniform source/drain doping profile," Nanotechnology, IEEE Transactions on , vol.1, no.4, pp. 233- 237, Dec 2002
[159] J. Gu, J. Keane, S. Sapatnekar, and C. Kim, "Width quantization aware FinFET circuit design," in Proc. IEEE CICC, 2006, pp. 337–340.
[160] Yang-Kyu Choi; Tsu-Jae King; Chenming Hu; , "Nanoscale CMOS spacer FinFET for the terabit era," Electron Device Letters, IEEE , vol.23, no.1, pp.25-27, Jan 2002
[161] Carlson, A.; Zheng Guo; Balasubramanian, S.; Zlatanovici, R.; Tsu-Jae King Liu; Nikolic, B.; , "SRAM Read/Write Margin Enhancements Using FinFETs," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.18, no.6, pp.887-900, June 2010
[162] Bansal, A.; Mukhopadhyay, S.; Roy, K.; , "Modeling and optimization approach to robust and low-power FinFET SRAM design in nanoscale era," Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005 , vol., no., pp.835-838, 21-21 Sept. 2005
[163] Bansal, A.; Mukhopadhyay, S.; Roy, K.; , "Device-Optimization Technique for Robust and Low-Power FinFET SRAM Design in NanoScale Era," Electron Devices, IEEE Transactions on , vol.54, no.6, pp.1409-1419, June 2007
[164] Goel, A.; Gupta, S.; Bansal, A.; Meng-Hsueh Chiang; Roy, K.; , "Double-gate MOSFETs with aymmetric drain underlap: A device-circuit co-design and optimization perspective for SRAM," Device Research Conference, 2009. DRC 2009 , vol., no., pp.57-58, 22-24 June 2009

[165] Lu, D.D.; Chung-Hsun Lin; Niknejad, A.M.; Chenming Hu; , "Compact Modeling of Variation in FinFET SRAM Cells," Design & Test of Computers, IEEE , vol.27, no.2, pp.44-50, March-April 2010
[166] Lu, D; Lin, C; Niknejad, A; Hu, C; , "Compact Modeling of Variation in FinFET SRAM Cells," Design & Test of Computers, IEEE , vol.PP, no.99, pp.1-1, 0
[167] Ming-Long Fan; Yu-Sheng Wu; Hu, V.P.-H.; Pin Su; Ching-Te Chuang; , "Investigation of Cell Stability and Write Ability of FinFET Subthreshold SRAM Using Analytical SNM Model," Electron Devices, IEEE Transactions on , vol.57, no.6, pp.1375-1381, June 2010
[168] O'uchi, S.; Nakagawa, T.; Matsukawa, T.; Liu, Y.X.; Endo, K.; Sekigawa, T.; Sakamoto, K.; Koike, H.; Masahara, M.; , "Impact of FinFET technology on 6T-SRAM performance," SOI Conference, 2009 IEEE International , vol., no., pp.1-2, 5-8 Oct. 2009
[169] Taurus 2005 User Manual, Synopsys Inc. 2005
[170] H. Soeleman and K. Roy, "Ultra-low power digital subthreshold logic circuits", IEEE ISLPED, pp.94-96,Aug. 1999.
[171] A. Wang and A. Chandrakasan, "A 180mV FFT processor using subthreshold circuit techniques", IEEE ISSCC, pp.292-529, 2004.
[172] Moradi, F.; Wisland, D.T.; Mahmoodi, H.; Tuan Vu Cao; , "Improved write margin 6T-SRAM for low supply voltage applications," SOC Conference, 2009. SOCC 2009. IEEE International , vol., no., pp.223-226, 9-11 Sept. 2009
[173] Moradi, F.; Wisland, D.T.; Aunet, S.; Mahmoodi, H.; Tuan Vu Cao; , "65NM sub-threshold 11T-SRAM for ultra low voltage applications," SOC Conference, 2008 IEEE International , vol., no., pp.113-118, 17-20 Sept. 2008
[174] Moradi, F.; Wisland, D.T.; Mahmoodi, H.; Tuan Vu Cao; , "Improved write margin 6T-SRAM for low supply voltage applications," SOC Conference, 2009. SOCC 2009. IEEE International , vol., no., pp.223-226, 9-11 Sept. 2009
[175] Moradi, F.; Wisland, D.T.; Mahmoodi, H.; Peiravi, A.; Aunet, S.; Cao, T.V.; , "New subthreshold concepts in 65nm CMOS technology," Quality of Electronic Design, 2009. ISQED 2009. Quality Electronic Design , vol., no., pp.162-166, 16-18 March 2009

[176] F. Moradi, A. Goel, G. Karakonstantis, H. Mahmoodi, Dag T. Wisland, Kaushik Roy, " DD-SAFF: Data- Dependant Sense Amplifier Flip-Flop", Accepted to CICC 2010, USA
[177] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, "Sub 50-nm FinFET: PMOS," in IEDM Tech. Dig., 1999, pp. 67–70.
178] N. Collaert, A. De Keersgieter, A. Dixit, I. Ferain, L.-S. Lai, D. Lenoble, A. Mercha, A. Nackaerts, B. Pawlak, R. Rooyackers, T. Schulz, K. Sar, N. Son, M. Van Dal, P. Verheyen, K. von Arnim, L.Witters, K. De Meyer, S. Biesemans, and M. Jurczak, "Multi-gate devices for the 32 nm technology node and beyond," in Proc. ESSDERC, 2007, pp. 143–146.
[179] F. Moradi, Dag T. Wisland, Hamid Mahmoodi, " Process Variations in Sub-Threshold SRAM Cells " IEEE ICM 2010
[180] F. Moradi, D. Wisland, H. Mahmoodi, K. Roy, "Asymmetric FinFET SRAM with improved Write and Read margin", IEEE TED, 2010


Published
Papers

Publications:

Published/Accepted Papers:

1. **Moradi, F.;** Wisland, D.T.; Tuan Vu Cao; Peiravi, A.; Mahmoodi, H.; , "1-bit sub threshold full adders in 65nm CMOS technology," *Microelectronics, 2008. ICM 2008. International Conference on* , vol., no., pp.268-271, 14-17 Dec. 2008
2. **Moradi, F**.; Wisland, D.T.; Mahmoodi, H.; Berg, Y.; Tuan Vu Cao; , "New SRAM design using body bias technique for ultra low power applications," *Quality Electronic Design (ISQED), 2010 11th International Symposium on* , vol., no., pp.468-471, 22-24 March 2010
3. **Moradi, F.**; Wisland, D.T.; Mahmoodi, H.; Tuan Vu Cao; Zarre Dooghabadi, M.; , "Adaptive supply voltage circuit using body bias technique," *Mixed Design of Integrated Circuits & Systems, 2009. MIXDES '09. MIXDES-16th International Conference* , vol., no., pp.215-219, 25-27 June 2009
4. **Moradi, F**.; Wisland, D.T.; Mahmoodi, H.; Tuan Vu Cao; , "Improved write margin 6T-SRAM for low supply voltage applications," *SOC Conference, 2009. SOCC 2009. IEEE International* , vol., no., pp.223-226, 9-11 Sept. 2009
5. **Moradi, F**.; Wisland, D.; Berg, Y.; Aunet, S.; Tuan Vu Cao; , "Process variations in sub-threshold SRAM cells in 65nm CMOS," *Microelectronics (ICM), 2010 International Conference on* , vol., no., pp.371-374, 19-22 Dec. 2010
6. **Moradi, F**.; Augustine, C.; Goel, A.; Karakonstantis, G.; Tuan Vu Cao; Wisland, D.; Mahmoodi, H.; Roy, K.; , "Data-dependant sense-amplifier flip-flop for low power applications," *Custom Integrated Circuits Conference (CICC), 2010 IEEE* , vol., no., pp.1-4, 19-22 Sept. 2010
7. **Moradi, F**.; Wisland, D.T.; Aunet, S.; Mahmoodi, H.; Tuan Vu Cao; , "65NM sub-threshold 11T-SRAM for ultra low voltage applications," *SOC Conference, 2008 IEEE International* , vol., no., pp.113-118, 17-20 Sept. 2008
8. **Moradi, F.**; Wisland, D.T.; Mahmoodi, H.; Tuan-Vu Cao; , "High speed and leakage-tolerant domino circuits for high fan-in applications in 70nm CMOS technology," *Devices, Circuits and Systems, 2008. ICCDCS 2008. 7th International Caribbean Conference on* , vol., no., pp.1-5, 28-30 April 2008
9. **Moradi, F.**; Wisland, D.T.; Mahmoodi, H.; Peiravi, A.; Aunet, S.; Cao, T.V.; , "New subthreshold concepts in 65nm CMOS technology," *Quality of Electronic Design, 2009. ISQED 2009. Quality Electronic Design* , vol., no., pp.162-166, 16-18 March 2009
10. **Moradi, F**.; Wisland, D.T.; Mahmoodi, H.; Aunet, S.; Cao, T.V.; Peiravi, A.; , "Ultra low power full adder topologies," *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on* , vol., no., pp.3158-3161, 24-27 May 2009
11. Ali Peiravi, **F. Moradi**, Dag T. Wisland "Leakage Tolerant, Noise Immune domino logic for circuit design in the ultra deep sub-micron CMOS technology for high fan-in gates" Journal of Applied Science, Volume 24, 2010

12. **F. Moradi**, G. Panagopoulos, G. Karakonstantis, H. Mahmoodi, D. Wisland, J. K. Madsen, K. Roy, " Multi-Level Wordline Driver for Low Power SRAMs in Nano-Scale CMOS Technology", Accepted, ICCD 2011
13. **F. Moradi**, D. Wisland, H. Mahmoodi, K. Roy, "Asymmetrically-doped FinFET (AD) for low-power robust SRAMs ", IEEE transactions on Electron Devices (TED), 2011, to appear
14. **Moradi F**., Dag T. Wisland, Hamid Mahmoodi, "Process Variations in Sub-Threshold SRAM Cells " ICM 2010, Egypt
15. **Moradi F.,** G. Panagopoulos, H. Mahmoodi, D. Wisland, and Kaushik Roy, " Multi-Level wordline driver for Robust SRAM design in Scaled CMOS Technology", IEEE Transaction on VLSI Journal (TVLSI) 2011, Accepted
16. **Moradi F**., Tuan Vu Cao, Dag T. Wisland, Hamid Mahmoodi "Optimal Body Biasing for Maximizing Circuit Performance in 65nm CMOS" accepted, MWSCAS 2011

Under Review

17. **F. Moradi**, G. Panagopoulos, D. Wisland, H. Mahmoodi, Kaushik Roy " Flip-Flops for Ultra Low-Power DSP Applications", IEEE TVLSI, 2011, Final Revision
18. **F.Moradi**, D. Wisland, K. Roy " Asymmetric FinFET 6T-SRAM cell for low power applications", Submitted to ISQED 2012
19. **F. Moradi**, G. Panagopolous, Dag T. Wisland, H. Mahmoodi, Kaushik Roy," Pulsed-Flip-Flop design for high performance applications", submitted to ISQED 2012
20. **F. Moradi**, I. Vatajelu, T. V. Cao, H. Mahmoodi, D. Wisland, " Domino Logic Designs for High-Performance and Leakage-Tolerant Applications", Microelectronics Journal, Under review

Papers co-author:

21. Tuan Vu Cao; Wisland, D.T.; Lande, T.S.; **Moradi, F.;** , "Low phase-noise VCO utilizing NMOS symmetric load for Frequency-based Delta-Sigma Modulators," *Electronics, Circuits, and Systems, 2009. ICECS 2009. 16th IEEE International Conference on* , vol., no., pp.767-770, 13-16 Dec. 2009.
22. Tuan Vu Cao; Wisland, D.T.; Lande, T.S.; **Moradi, F**.; , "A bulk-controlled ring-VCO with 1/f-noise reduction for frequency ΔΣ modulator," *Mixed Design of Integrated Circuits & Systems, 2009. MIXDES '09. MIXDES-16th International Conference* , vol., no., pp.187-192, 25-27 June 2009.
23. Tuan Vu Cao; Wisland, D.T.; Lande, T.S.; **Moradi, F.**; , "Low Phase-Noise and Wide Tuning-Range CMOS Differential VCO for Frequency ?S Modulator," *VLSI, 2009. ISVLSI '09. IEEE Computer Society Annual Symposium on* , vol., no., pp.13-18, 13-15 May 2009
24. Tuan Vu Cao; Wisland, D.T.; Lande, T.S.; **Moradi, F**.; , "Low-power, enhanced-gain adaptive biasing based Operational Transconductance Amplifiers," *NORCHIP, 2009* , vol., no., pp.1-4, 16-17 Nov. 2009.

25. Tuan Vu Cao; Wisland, D.T.; Lande, T.S.; **Moradi, F**.; , "Rail-to-rail low-power fully differential OTA utilizing adaptive biasing and partial feedback," *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on* , vol., no., pp.2820-2823, May 30 2010-June 2 2010.
26. Tuan Vu Cao; Wisland, D.T.; **Moradi, F**.; Lande, T.S.; , "Novel low voltage current-mirror sense amplifier based Flip-Flop with reduced delay time," *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on* , vol., no., pp.3166-3169, 24-27 May 2009
27. Tuan Vu Cao; Wisland, D.T.; Lande, T.S.; **Moradi, F**.; Young Hee Kim; , "Novel start-up circuit with enhanced power-up characteristic for bandgap references," *SOC Conference, 2008 IEEE International* , vol., no., pp.123-126, 17-20 Sept. 2008
28. Tuan-Vu Cao; Wisland, D.T.; Lande, T.S.; **Moradi, F**., "Low-Voltage, Low-Power, and Wide-Tuning-Range Ring-VCO for Frequency ΔΣ Modulator," *NORCHIP, 2008.* , vol., no., pp.79-84, 16-17 Nov. 2008