Redundant Skewed Clocking of Pulse-Clocked Latches for Low Power Soft-Error Mitigation

by

Aditya Gujja

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved November 2015 by the
Graduate Supervisory Committee:

Lawrence T. Clark, Chair
Keith E. Holbert
David Allee

ARIZONA STATE UNIVERSITY

December 2015

ABSTRACT

An integrated methodology combining redundant clock tree synthesis and pulse clocked latches mitigates both single event upsets (SEU) and single event transients (SET) with reduced power consumption. This methodology helps to change the hardness of the design on the fly. This approach, with minimal additional overhead circuitry, has the ability to work in three different modes of operation depending on the speed, hardness and power consumption required by design. This was designed on 90nm low-standby power (LSP) process and utilized commercial CAD tools for testing. Spatial separation of critical nodes in the physical design of this approach mitigates multi-node charge collection (MNCC) upsets. An advanced encryption system implemented with the proposed design, compared to a previous design with non-redundant clock trees and local delay generation. The proposed approach reduces energy per operation up to 18% over an improved version of the prior approach, with negligible area impact. It can save up to 2/3rd of the power consumption and reach maximum possible frequency, when used in non-redundant mode of operation.

DEDICATION

To my parents
Ramadevi and Madhava rao Gujja

ACKNOWLEDGMENTS

I would like to express my gratitude to many people who have played an integral role in helping me during this endeavor over all these years.

I would like to thank my parents, Ramadevi and Madhava rao Gujja for their love, encouragement and support throughout my graduate studies and research.

I would also like to thank Dr. L. T. Clark for this opportunity and my committee members Drs. Holbert and Allee for their time and support. I am also indebted to my colleagues Srivatsan Chellappa, Chandarasekaran Ramamurthy, Vinay Vashishtha, Sandeep Shambhulingiah, Yitao Chen, Anudeep Reddy Gogulamudi, Sai Chaitanya Reddy Jakkireddy and Punit Shah for their invaluable contributions, discussions, and support to all our projects. I must also thank the helpful administrative staff at the EE department Toni and Lynn for helping me with all the administrative procedures at ASU.

I would like to thank all my friends who supported and encouraged me in life.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER 1. INTRODUCTION

1.1. Introduction

In modern technologies, electronic circuits used in aerospace, safety-critical and commercial designs are becoming more vulnerable to radiation effects, due to decreasing transistor sizes and supply voltages. When radiation particles such as protons, neutrons, alpha particles or heavy ions strike the sensitive nodes in very large scale integrated (VLSI) circuits, single event effects (SEEs) may occur and cause devices to malfunction. Single event upsets (SEUs) are an important type of SEE that affects the electronic systems. In recent times, single-event transients (SETs) are becoming a primary cause for the malfunctioning of several space applications [Koga93, Ecof94]. Not only space applications but also other critical applications like biomedical, industrial and banking also demand highly reliable systems [Nara06]. The study and analysis of radiation effects on circuits has been a major area of research. The technique of designing and fabricating electronic systems to withstand radiation is called radiation hardening. This chapter provides an overview of the radiation environment, radiation effects on devices and circuits, techniques to achieve radiation hardness and basic sequential element designs.

1.2. Radiation Environment in space

The space environment contains phenomena that are potentially hazardous to human and electronic systems. This environment consists of different kinds of particle that cause SEEs in the modern devices. The space environment and its components are shown in Fig 1.1. The spectrum of radiation environment typically consists of charged particles originating from various sources such as

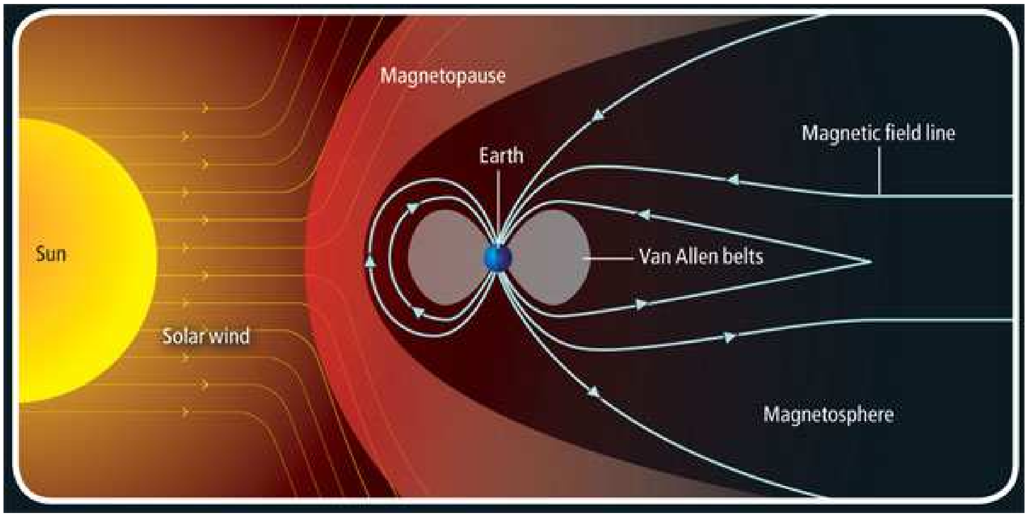


Fig 1.1. Picture illustrating the space radiation environment [miles05]

- Protons and other heavy nuclei associated with solar events
- Trapped radiation particles by the Earth's Van Allen belts
- Galactic cosmic rays that consist of interplanetary protons, electrons and ionized heavy nuclei
- Neutrons (primarily cosmic ray albedo-neutrons or CRAN particles)
- Photons (γ-rays, X-rays, UV/EUV, optical, infra-red and radio waves)

Solar energetic particles (SEP) are the high energy particles that are expelled from the sun due to the solar flares. Primarily SEP consist of the electrons and protons with energies that can range from KeV to GeV. They can also obtain speeds that can be up to 80% of the speed of light. Trapped particles, which are 93% protons, 6% alpha particles, and about 1% heavy nuclei, contribute the most to radiation effects in low and medium Earth orbits that pass through the Van Allen belts [Stass88].

Galactic cosmic rays (GCR) are comprised mostly of protons and heavy ions, i.e., 98% nuclei and 2% electrons. The nuclei component consists of 87% hydrogen and 12% helium and 1% heavier nuclei from the heavy metals [Simp83].

CRAN particles are primarily secondary cosmic ray neutrons produced by the interaction of GCR with the earth's atmosphere about 55km above the Earth surface. These have a half-life of 11.7 minutes beyond which they decay in to an electron, proton and an anti-neutrino. Secondary neutrons are the most important contributor to single event effects at altitudes below 60,000 feet. The rest of the electromagnetic spectrum consists of X-rays (wavelengths 10Å – 100Å), extreme ultraviolet or EUV (100Å – 1000Å), ultraviolet (1000Å – 3500Å), the visible spectrum (3500Å – 7000Å) and the infra-red spectrum (0.7μ – 7mm). Each type of radiation has a characteristic spectrum and preferred interaction mode with matter that give rise to various effects such as photo-ionization, photoelectron emission, Compton effect etc. Photon interactions are not a primary concern for satellites in the natural space environment [Fred96].

1.3. Effect of radiation particles on circuits

Radiation effects may cause malfunction, degradations, processor restarts or even permanent damage to the electronic devices and circuits. The type of damage to circuits depends on the type of particle, mass, energy, charge and state of the particle. The density of the target material is also important. Here we focus on Si and SiO2. These particles loose kinetic energy as they travel through the target material. As they interact with the electrons of the target material, they can also rarely interact with nuclei. The distance required to stop an ion (its range) is both a function of its energy and the properties of the material (primarily its density) in which it is traveling. The stopping power or linear energy transfer (LET) is a function of the material through which a charged particle is traveling and refers to the energy loss of the particle per unit length in the material. The LET (MeV-cm^2/mg) is a function of both the ion's mass and energy and density of the target material, given as

$$LET = \frac{1}{\rho}\frac{dE}{dx} \quad (MeV - cm^2/mg), \tag{1}$$

where $\frac{dE}{dx}$ is the energy loss per unit length and ρ is the material density in mg/cm^3. The maximum LET value near the end of the particle's range is called the Bragg peak [Hsieh81].

These interacting radiation particles create two types of ionizations in the target material. They are direct ionization and indirect ionization. Interacting with the electrons and releasing them from their bonds is direct ionization. This creates a large number of charged particles around their tracks. Interacting with nucleus of the target material and setting them free, where it becomes the ionizing particle, is indirect ionization. The nuclei acts as charged particle and causes more tracks in the target material.

The two major radiation effects on MOS circuits and devices are single event effects (SEEs) [Mavis02] and total ionizing dose (TID) effects [Barn06]. TID hardening is beyond the scope of this work and will not be discussed further. All materials presented in this thesis focuses on mitigating SEEs.

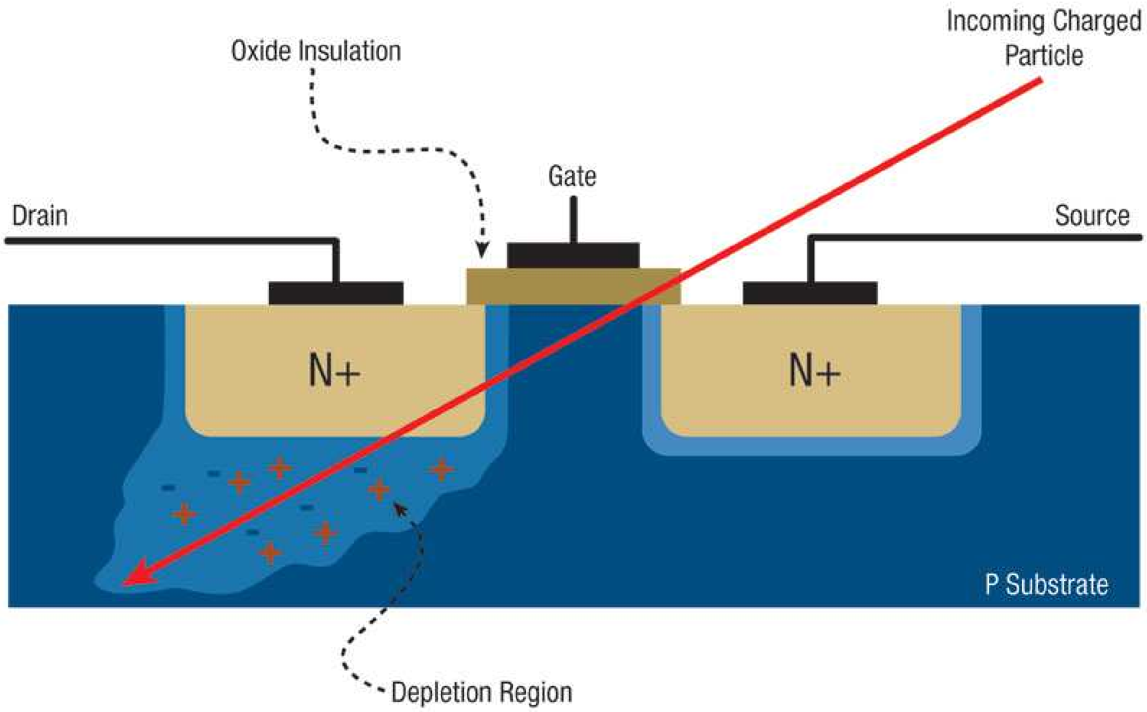


Fig 1.2 Charged particle striking a node of a transistor. Funnel formation and charge collections mechanism in shown following an ion strike.

1.3.1. Single Event effects in CMOS

Single event effects are by definition caused by a single, energetic particle and can take many forms [NASA]. Designers have to be concerned with three main causes of SEEs, cosmic rays, high energy protons and neutrons. For cosmic rays, SEEs are typically caused by its heavy ion component. These heavy ions cause a direct ionization SEE, i.e., if an ion particle traversing a device, deposits sufficient charge, an event such as a memory bit flip or logic voltage transient may occur. Cosmic rays may be galactic or solar in origin. Protons, usually trapped in the earth's radiation belts or from solar flares, may cause direct ionization SEEs in very sensitive devices. However, a proton may more typically collide with nuclei near a sensitive device area, and thus, cause an upset via an indirect ionization effect [Sagg05].

Charge is generated from a single event phenomenon generally within a few microns of the junction. In silicon, one electron-hole pair is produced for every 3.6 eV of energy lost by the impinging radiation. As silicon has a density of 2328 mg/cm^3, it is easy to calculate from equation (1) that an LET of 97 MeV-cm^2/mg corresponds to a charge deposition of 1 pC/µm. Hence, the amount of collected charge (Q) in silicon can be given by the formula

$$Q = 0.01036 * LET \quad pC\ /\mu m\ . \tag{2}$$

Thus, the collected charge for these events is from 1-100 fC depending on the type of ion, its trajectory, and its energy over the path through or near the junction. The most sensitive semiconductor device structure is the reverse-biased junction. In worst-case the junction is floating (as in dynamic logic circuits and some analog designs) and it is extremely sensitive to any charge collected from a radiation event.

The device characteristic that determines the upset sensitivity of a device is its critical charge (Q_{crit}). This is the amount of charge that must be collected at the terminal of a latch to cause an upset.

1.3.2. Types of single event effects

There are many types of Single-event effects but they can be broadly classified into two categories, non-destructive (soft error) and destructive SEEs (hard errors). Soft-errors are due to a non-permanent charge on voltage state change or error in the circuits caused by radiation. This type of errors will be recovered when a new data is written or next cycle of instructions flow through the pipeline. A hard-error is due to physical damage to the device in the circuit which cannot be recovered.

Destructive effects of SEE include Single-event burnout (SEB) and single-event gate rupture (SEGR). These are the permanent effects and cannot be protected by normal circuit design techniques. So, we are not going to discuss these effects further. The non-destructive SEEs consist of single-event transient (SET) and single-event upsets (SEU) which may manifest as multi bit upsets (MBU). These are the main types of SEEs that we are going to discuss in this thesis, since these can be avoided by using circuit design techniques.

SETs are the temporary voltage glitches that occur in the integrated circuits due to LET of the charged particles that are hit on a node [Heil89, Bene04, Gadl04]. Duration depends on the amount of charge carried by the particle and transistor size of the recovering (driving) circuit. These glitches can propagate through the combinational logic. This may travel to the output of the design and may appear at the circuit output. This may also travel through the combinational logic and get captured by a sequential element e.g., latch, FF or memory cell.

When a charged particle strikes one of the sensitive nodes of a memory cell, such as a drain in an off state transistor, it generates a transient that can turn on the gate of the other complementary transistor. This effect can produce an inversion in the stored value, in other words, a bit flip in the memory cell, i.e., a single event upset (SEU) [Sagg05]. So, whenever an SET is captured by a sequential element or memory cell, it may appear to be a SEU. This effects is called SEU as we cannot usually distinguish between a SEU

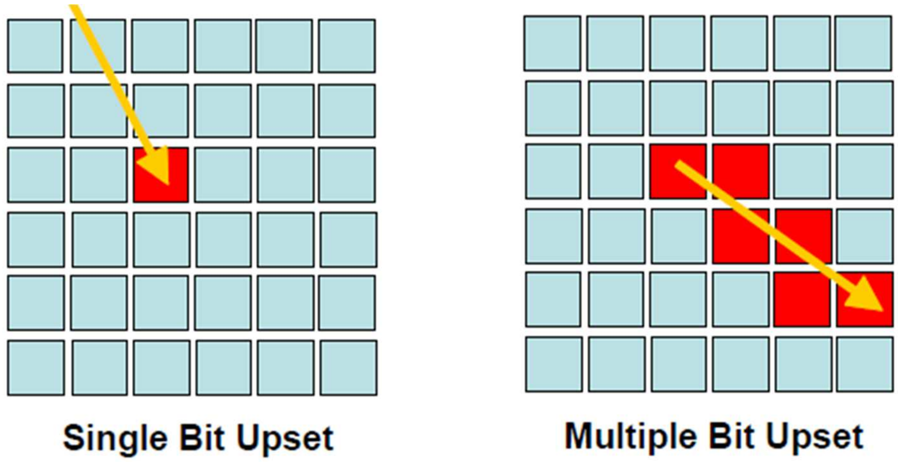


Fig.1.3 A schematic representation of single bit upsets and multi bit upsets in a memory array [Baze97].

on storage node and SET that is propagated form a combinational logic and captured by a storage node.

The rate at which soft errors occur is called soft error rate (SER). The unit of measure commonly used with SER and other hard reliability mechanisms is failure in time (FIT). One FIT is equivalent to one failure in 10^9 device hours.

1.3.3. Multi Bit Upsets

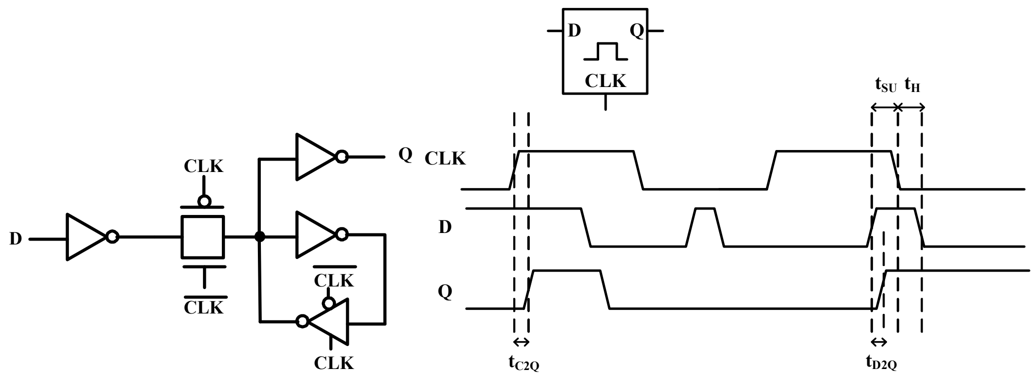
More than a single storage bit might be affected, creating a multi-bit upset (MBU) as opposed to a single bit upset (SBU). MBUs are defined as the occurrence of two or more bit upsets, appearing within the same clock cycle from a single particle hit, to distinguish from random multiple hits within a single cycle [Muss96, David09]. While MBUs are usually a small fraction of the total observed SEU rate, their occurrence has implications for memory architecture in systems utilizing error correction, as well as redundant circuits to mitigate soft-errors.

MBU depends on the node separation of the storage nodes, size of transistors and supply voltage. It also depends on the LET of the particle, angle of incidence, track radius of the particle. Depending on these factors, the charge from a particle can be collected by more than one storage node that are nearby. The previous stored value in those nodes determines which of those nodes will change their values since collected charge can reinforce the correct state. This problem mainly occurs in memories, as they are packed tightly where the storage nodes are side by side. This is also becoming a problem in multi-bit sequential elements where a group of latches or FFs are together. Fig. 1.3 represents the single bit upset and multi-bit upsets in a memory array.

1.4. Sequential Element Design

Sequential elements are widely used in digital VLSI designs for data storage and data synchronization. Sequential elements are the circuits in which the output depends on the previous state and also present state. Finite state machines and pipelining are the two examples where the sequential elements are used [weste04]. They are controlled by the clock signal and store the current data depending on the clock. Clock also controls when the input is stored into the storage node and also when it should be sent to the output.

We discuss only static sequential elements since our work only focuses on the static sequential elements which are mainly used for low power designs. Sequential circuits are also classified into two major categories depending on the way in which data is captured. They are level sensitive and edge sensitive circuits. Level sensitive circuits (such as latches) capture the data at a particular logic level of clock while edge-triggered circuits (e.g. Flip-flops) capture the data at a given change in the clock state such as a rising or a falling edge.

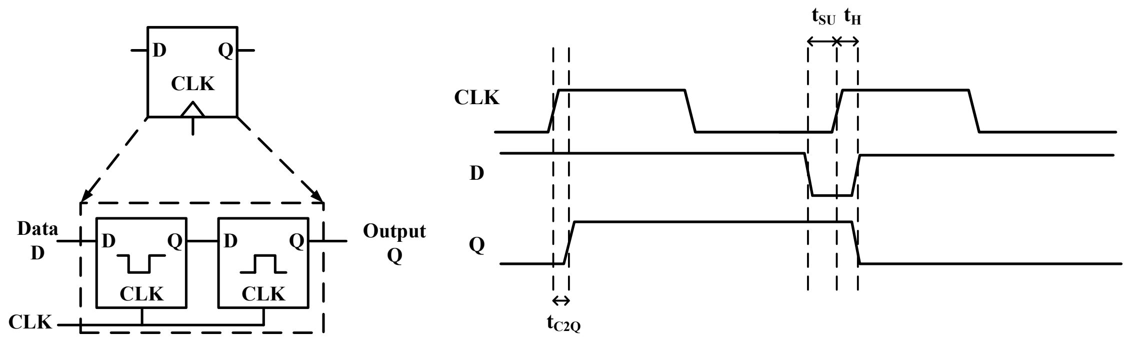


(a) *(b)*

Fig.1.4 (a) Latch Schematic. (b) Latch operation and delays. [Chandra01]

1.4.1. Latch

The latch is the most basic level sensitive sequential storage element in use. A latch is transparent during one of the clock phase, i.e., the input D is open to the output Q. During the other clock phase, the data is stored and continues to send the previous data to the output until a new data over rides it in the coming phase of the clock. Thus the latch operates in two modes, transparent (D propagates to Q) and opaque (Q is retained) depending on the state of the clock. Thus, there will be two kinds of latches, transparent high latch and transparent low latch.



(a) *(b)*

Fig.1.5 (a) D flip-flop constructed from two latches. (b) D flip-flop operation.

The core function of the latch is to store a data bit. It has two inverters connected back-to-back so that it can store logic 0 and 1. The data is stored in the latch by changing the bi-stable circuit to the required state. Fig 1.6(a) shows the most commonly used static latch design. The connection of the clock to the transmission gates determines the type of latch, so it is a positive latch or negative latch. The clock edge at which the latch transitions from the transparent to opaque state is the closing capture edge of the latch. To ensure that correct data has been captured, the data D should set-up to the clock edge so that it can change the state of the storage node before the latch goes to the opaque state. It is called the set-up time (t_{SU}) as shown in Fig. 1.6(b). The data D must also be held stable for a minimum hold-time (t_H) after the closing edge of the clock. The time taken for the data to propagate from D to Q when the latch is transparent is called the data latency (t_{D2Q}) and the time taken for the data to propagate from D to Q at the rising edge (for a positive high latch) of the clock CLK, is called the latch latency (t_{C2Q}).

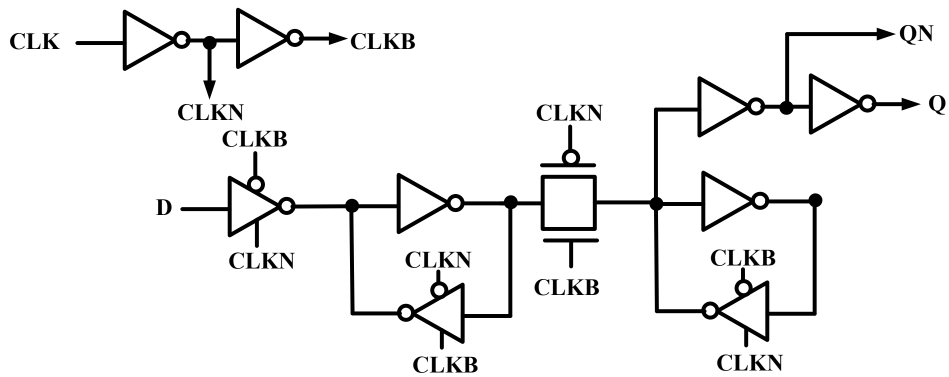
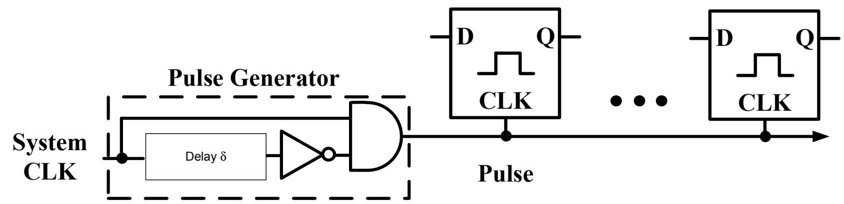


Fig.1.6 Master-slave flip-flop schematic[Chandra01].

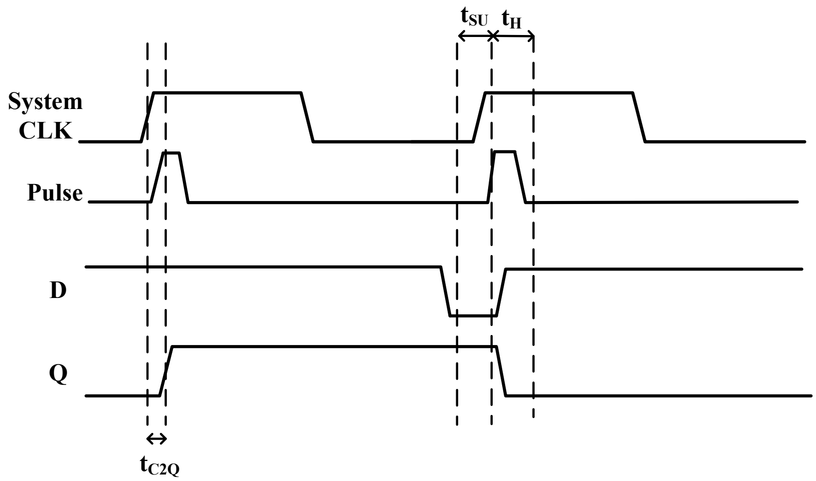
1.4.2. Flip-Flop

D flip-flop is one of the most commonly used sequential element in digital designs. Flip-flops are edge-triggered designs. Simply a flip-flop can be defined as the combination of two different latches connected side-by-side. Generally, this configuration is master-slave configuration. At any given time, master and slave latches will be in opposite modes, transparent and opaque. So, depending on the configuration of master and slave latch, the flip-flops can be differentiated as positive-edge triggered or negative-edge triggered.

Fig 1.7(a) shows the positive edge triggered master slave flip-flop (MSFF) made with a negative latch followed by a positive latch. When the clock is negative (low), the master latch captures the data and the slave latch retains the previous value and propagates it to output. At the positive clock edge, the master latch becomes opaque and slave latch becomes transparent and transmits the data to output. Similar to the latch, the minimum time for which the data has to be held before the rising edge of the clock for it



(a)



(b)

Fig.1.7 (a)Pulse latch schematic and (b) working. Delay δ can be generated by buffers depending on the delay required. Note the pulse generator can be shared across multiple latches.

to be reliably stored is called the flip-flop setup time (t_{SU}) and the minimum time for which the data has to be held constant after the rising edge of the clock, is the hold time (t_H). The output Q is available t_{C2Q} after the rising edge of the clock as shown in Fig 1.7(b). Fig 1.8 shows the standard implementation of the MSFF included in most standard cell libraries.

1.4.3. Pulse-clocked Latch

As the conventional FF consists of two latches operating as a master-slave pair, the overall area and power of the circuit is considerably larger than latch based designs.

Thus a pulse latch or pulse clocked latch with only a single latch is used [Shiba 06] to simulate edge-triggered operation.

A pulse latch is a latch clocked by a pulse clock. A pulsed clock is generated using a pulse generator which generates a small duration high pulse at every raising global clock edge as shown in Fig. 1.9. the brief period of the pulse makes the latch transparent for that small period of time and opaque for the rest of the clock period. This is similar to the edge triggered master-slave flip-flop functionality but with much larger hold time. This saves nearly half the power dissipated by a D flip-flop. The disadvantage is that the pulses need to be generated and propagated to the latches, which dissipates power. This power overhead can be reduced by sharing the pulse between a group of latches. The pulse latch has similar timing parameters as a latch (t_{SU}, t_H and t_{C2Q}), i.e., measured to the closing pulse clock edge.

1.4.4. Timing constraints for sequential designs

In synchronous systems, sequential elements introduce a period of time where no useful logic can be evaluated, known as dead time (t_{DEAD}).

$$t_{DEAD} = t_{C2Q} + t_{SU} \tag{3}$$

The maximum clock frequency (f_{CLK}) or minimum clock period t_{CLK} for the system is then a function of the dead time,

$$\frac{1}{f_{CLK}} = t_{CLK} \geq t_{DEAD} + t_{CLmax} \tag{4}$$

where, t_{CLmax} is the largest worst case combinational logic delay in the chip.

. A setup time violation occurs when data from the previous flip-flop doesn't propagate through the combinational logic to the next FF in time to meet the its setup time. As this violation is due to the propagation delay through the logic elements between the flip-flops, it is a frequency dependent problem and can be addressed by lowering the clock rate. A hold time violation occurs when the hold time constraints imposed by the sequential elements are violated. This means the data from the sending flip-flop races through the shortest combinational logic path (t_{CLmin}) (called the contamination delay) and violates the hold time of the subsequent receiving flip-flop. Thus, the following minimum delay constraint is imposed on the system [Chandra01].

$$t_H \leq t_{C2Q} + t_{CLmin} \tag{5}$$

Hold time errors are frequency independent and thus fixing these violations is of utmost importance when designing the chip. The hold time violation can be solved by adding delays between the stages of the flip flop to increase t_{CLmin}.

Ideally the clock signal at both the sending and receiving flip-flop transition at the same time. However, in practical designs, the clocks to both the sending and receiving flops may be temporally offset with respect to each other changing the design parameters considerably. The difference in clock arrival times between two sequentially adjacent registers is called clock skew t_{SKEW}. The periodicity of the clock signal may also be affected by the deviation of its edges from their expected transition time causing jitter, t_{JIT}.

In the presence of clock skew and jitter, equation (4) becomes

$$\frac{1}{f_{CLK}} = t_{CLK} \geq t_{DEAD} + t_{CLmax} + t_{SKEW} + t_{JIT} \tag{6}$$

and equation (5) becomes

$$t_H \leq t_{C2Q} + t_{CLmin} - t_{SKEW} \tag{7}$$

1.5. Radiation hardening techniques

Radiation hardening is to protect the electronic circuits and systems from being effected by different kinds of radiation by appropriate engineering, as discussed in section 1.1. There are several methods by which we can achieve radiation hardening, they can be classified into two categories. Radiation hardening by process (RHBP) and radiation hardening by design (RHBD). Using these techniques, we can mitigate many radiation effects caused in ICs. While it is not possible to discuss all the published soft error mitigation techniques, the most common techniques are outlined in the subsequent sections.

1.5.1. Radiation hardening by process (RHBP)

Radiation hardening by process (RHBP) are hardware based solutions that make the designs hard to SEE [shin94]. The term RHBP refers to any process deviation from the standard fabrication sequence that are done with the sole purpose of achieving an increase in the radiation tolerance for that particular technology platform. High-resistivity and silicon-on-insulator (SOI) substrates are two popular examples of RHBP solutions.

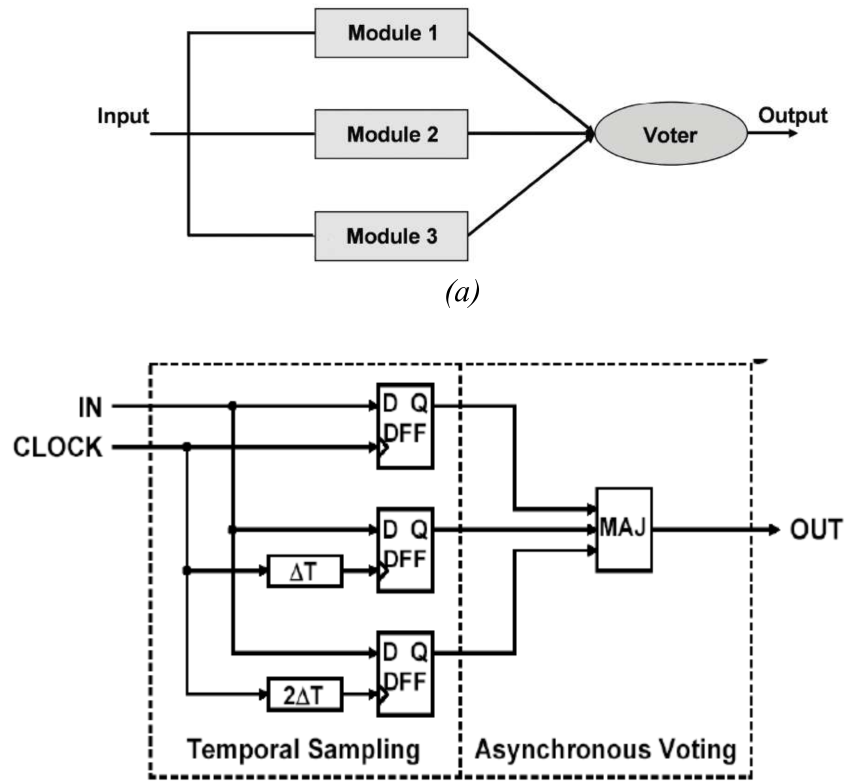
The drawbacks of these techniques are that they are becoming increasingly difficult in modern sub-micron processes. Modifications to the processes is also costly as it requires a dedicated fabrication facility and extreme R and D. Since, this thesis is not based on this techniques, we leave this discussion cursory.

1.5.2. Radiation Hardening by Design (RHBD)

Radiation hardening by design (RHBD) uses circuit design and layout techniques implemented on a standard commercial foundry process to achieve SEE mitigation. This technique reduces the cost per chip compared to RHBP. There are different layout and architectural techniques used to achieve the hardness of the designs. These designs may not be hard to all kinds of radiation effects but they can be varied depending on the type of hardness required by the particular design. This thesis is based on improvements and variations of these RHBD techniques, so we are going to discuss some of the key aspects.

1.5.2.1. Design techniques for mitigating SEE effects

There are different types of techniques that are used to mitigate SEEs. Error correction codes (ECC) can also be included in the designs to detect errors, primarily in memories. Error detection and correction (EDAC) schemes can detect and also correct the errors in the designs to make them hard. For logic, widely used techniques such as triple modular redundancy and temporal redundancy are the techniques on which the proposed designs are based.



(a)

(b)

Fig 1.8 (a)Implementation of a Triple modular redundancy (TMR) based hardware redundancy scheme and (b) Temporal redundancy based on delayed sampling in flip-flops after [Mavis02]. Note ΔT represents the delay introduced.

1.5.2.2. Triple modular redundancy

Triple modular redundancy (TMR) is one of the widely used techniques for hardening the circuits from soft-errors. The basic principle of this technique is that the combinational logic and the sequential elements are triplicated and the output is voted using a majority gate. This technique works on the assumption that no two copies of the logic will be hit by the charged particle on the same node at the same time. With appropriate spatial separation of the key nodes, the probability is very low that it is almost impossible for such a case to occur [Hind09] [Hind11]. This technique can be

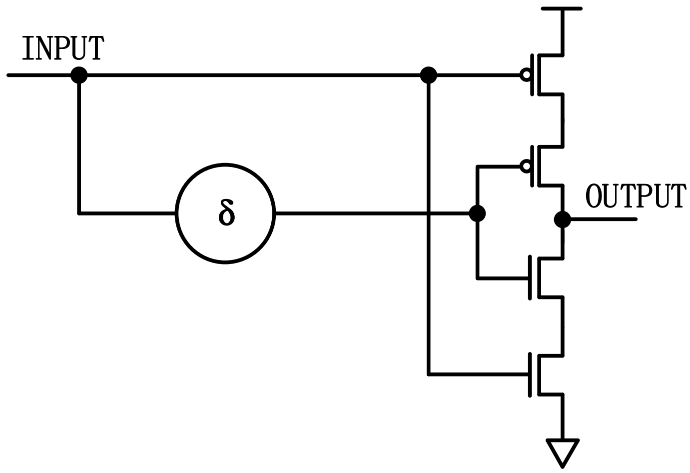


Fig 1.9 Delay filter - delay element in combination with Muller C-element.

implemented in different levels, e.g., triplicating the entire design or pipeline stages etc. This technique is shown in the fig. 1.4(a).

1.5.2.3. Temporal hardening technique

Temporal hardening is another widely used technique. In this approach, data is captured at different times, compared with other instants to ensure that incorrect data does not propagate. We can also use a majority gate to compare the data and get the correct output. One implementation of this technique can be seen in Fig 1.4(b) [mavis02]. This works on the assumption that the upset (SET) width is less than the delay (ΔT) of the delay element. If the upset is more than the delay element delay, then the upset may be captured by two different instances and may propagate error to the output. We can also use delay filters [sandeep15] [Naseer06] to compare two instance of the same input to mitigate the SETs. The delay filter is a combination of delay element and Muller C-element as shown in fig. 1.5.

1.6. Outline

This chapter has provided a brief overview of the radiation environment, the different effects of radiation on circuits and common mitigation techniques, as well as an introduction to various sequential elements. Chapter 2 discusses the previous radiation hardened designs and the prior work for the proposed design. The proposed design and its various modes of operations are explained in Chapter 3. Chapter 4 discusses the AES implemented design with the proposed sequential elements and also its comparisons with some previous hardened designs. Chapter 5 concludes the dissertation specifying the unique feature that are proposed in the new design.

CHAPTER 2. Previous Radiation Hardened Designs

2.1. RHBD Latches

This chapter provides brief overview of some previous RHBD FF designs that are most commonly used. We discuss about different techniques that are adopted by these designs. One of the most widely used circuit is dual interlocked storage cell (DICE) latch.

2.1.1. DICE Latch

Calin introduced the DICE latch in 1996 in [calin96]. The basic principle for the DICE latch is shown in Fig 2.1. Its main principle is to store the same data twice using four storage nodes such that if one of the storage nodes is upset, then the other three nodes will restore those nodes to the correct values. X0-X3 are the four storage nodes, driven by four back-to-back inverter pairs (P0-N3) (P1-N0) (P2-N1) (P3-N2) as shown in Fig 2.1. These four storage nodes will store two pairs of complementary values (1010 or 0101). All these storage nodes can be accessed by separate pass gate transistors. This structure relies on "dual node feedback control" to achieve immunity. This states that each node is protected by two adjacent nodes.

The concept is simplified into only two cross-coupled inverters as shown in Fig. 2.2. The simplified version removes the additional transistors but retains the function. Each storage node is driven by only one PMOS and NMOS transistor, which are each driven by the two adjacent storage nodes, each node driving one transistor. This makes the design hard to SEUs and the only way to upset the state is to flip two nodes at a time. A SEU hitting node X2 of this latch is shown in Fig. 2.3. Since node X2 is changed its value form '1' to '0', transistor P3 turns on and N1 cuts off. This makes node X3 rise

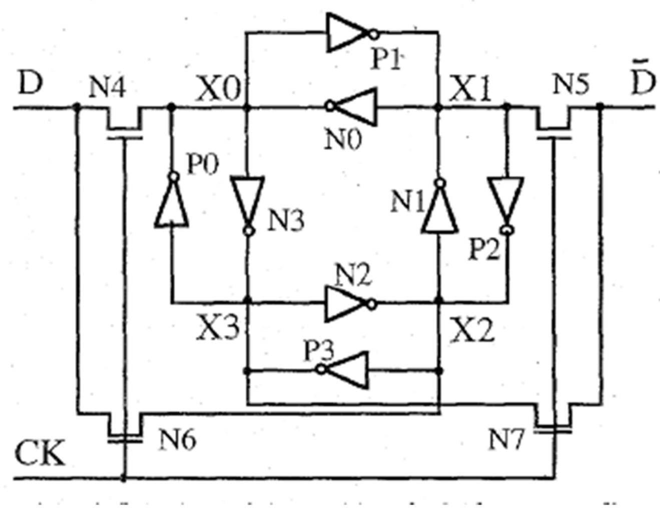


Fig 2.1. Principle of DICE [Calin 96].

above zero and node X1 falls below zero. Node X4 is unaffected. As the deposited charge is completely collected, node X2 is restored to its original value by the adjacent nodes X1 and X3.

Note that an SET on a DICE input will still upset the design. Thus the DICE latch only mitigates SEU. Moreover, it is difficult to provide spatial separation in this design [sandeep15], so MBU type upset is a risk.

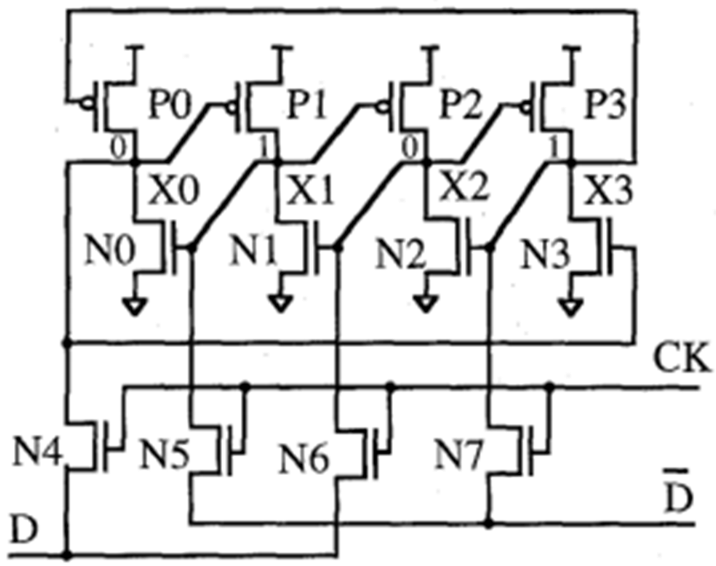


Fig 2.2. DICE Memory cell.

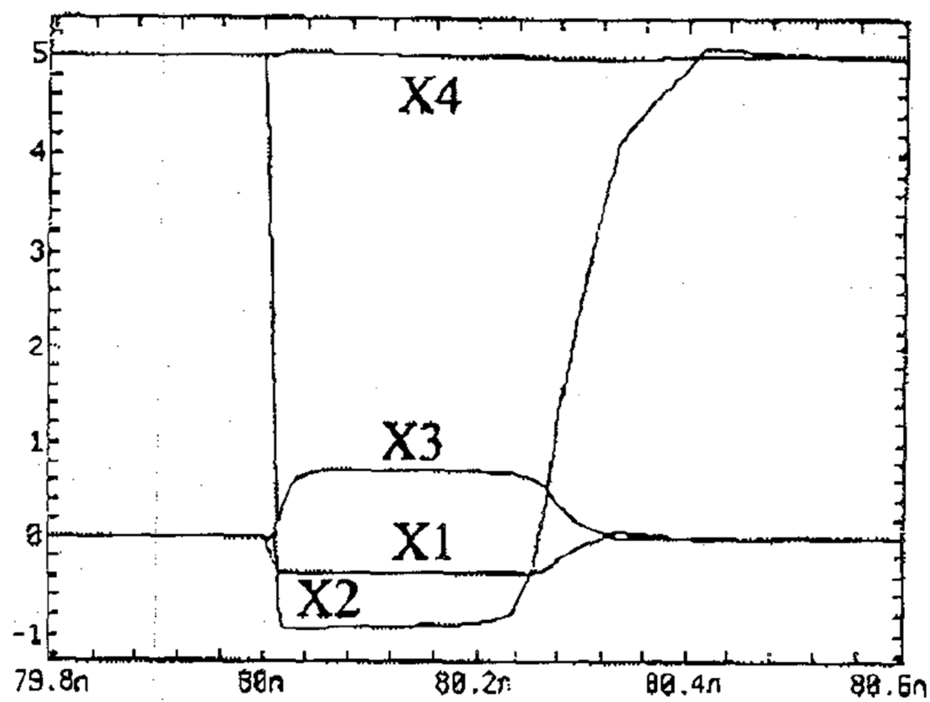


Fig 2.3. Spice simulations showing SET strike at node X2 of the DICE latch.

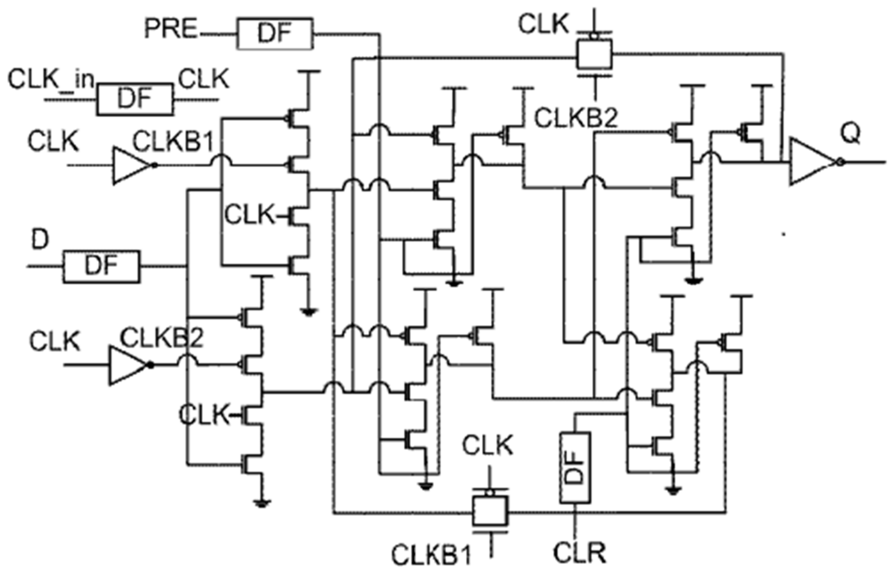


Fig 2.4. Schematic for DF-DICE latch [Naseer 06]).

2.1.2. Delay Filter DICE (DF-DICE)

The delay filter dual interlocked storage cell (DF-DICE) [Naseer 06] is an improved version of the DICE latch. It is shown in Fig 2.4. The DICE latch is only hard to SEUs on the storage nodes. Any SET on input D or the clock CLK may store a wrong state in all four storage nodes. The delay filter (DF) is the combination of a delay element and a Muller C-element as shown in Fig 1.5. This DF circuit will protect propagation of a signal from an SET that is less than the delay used in it. Since this design has DF's on every input including D and CLK, this design is hard to SETs on every input. This design can be scaled to tolerate SETs of higher width by just adjusting the delay of the DF. The cost of protecting the design depends linearly on the width of the SET. As the LET increases the area of the design also increases. However, a SET on the filter output still causes an upset. Thus, the cross-section due to SET is not vanishing. The area

Transient Threshold	DF-DICE latch Area (um^2)	Increase per latch	DF-DICE Flip-flop Area (um^2)	Increase per flip-flop
250ps	82160	54%	117515	31.7%
450ps	94413	77%	129768	45.5%
650ps	131936	100%	142424	59.7%
850ps	119925	124%	154878	73.5%
1200ps	145638	172%	179787	101.5%

Table I. Area comparison between DICE and DF-DICE.

comparison between DICE and DF-DICE and shown in Table I. The layouts were implemented in MOSIS CMOS rules for 6-metal single poly TSMC 0.18 micron technology.

2.1.3. BISER FF

BISER FF [Zhang 06] is a design that only mitigates SEUs in the storage node but not SETs on the data or the clock inputs. Fig. 2.5 shows the block level implementation for the BISER FF.

The BISER FF has two FFs in parallel whose outputs are connected through a Muller C-element and a jam latch at the output of C-element. The concept of this design is that, if one of the FF is affected and stores wrong value, the C-element doesn’t propagate it to the output instead and gets tri-stated. The jam latch, which has the correct value, propagates it to the output. This makes this design hard to SEUs on the storage nodes. Any upset on the input D will make the two copies wrong and propagates it to the output. It is also not hard to any upset on CLK nodes as it may create false edges and capture wrong data. In the original BISER paper as shown in Fig. 2.5, the output inverter

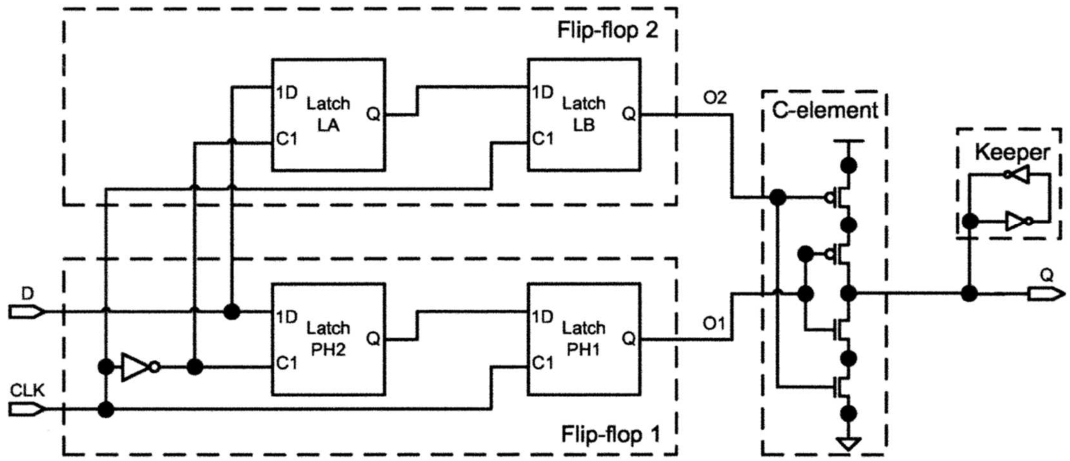


Fig 2.5 Block level implementation of BISER FF [Zhang06].

is missing and jam latch storage node is exposed outside. This should be avoided as this may back couple and change the data in the storage node. An output inverter should be used to decouple the jam latch storage node from the output node.

This design is equivalent to five latches including the C-element and the jam latch. This design is not hard to Multi bit upset (MBU) which causes two or more nodes to get upset at the same time. Since this design has only two storage nodes, upsetting two nodes will propagate wrong values to the output. This can be avoided by separating the two FFs spatially.

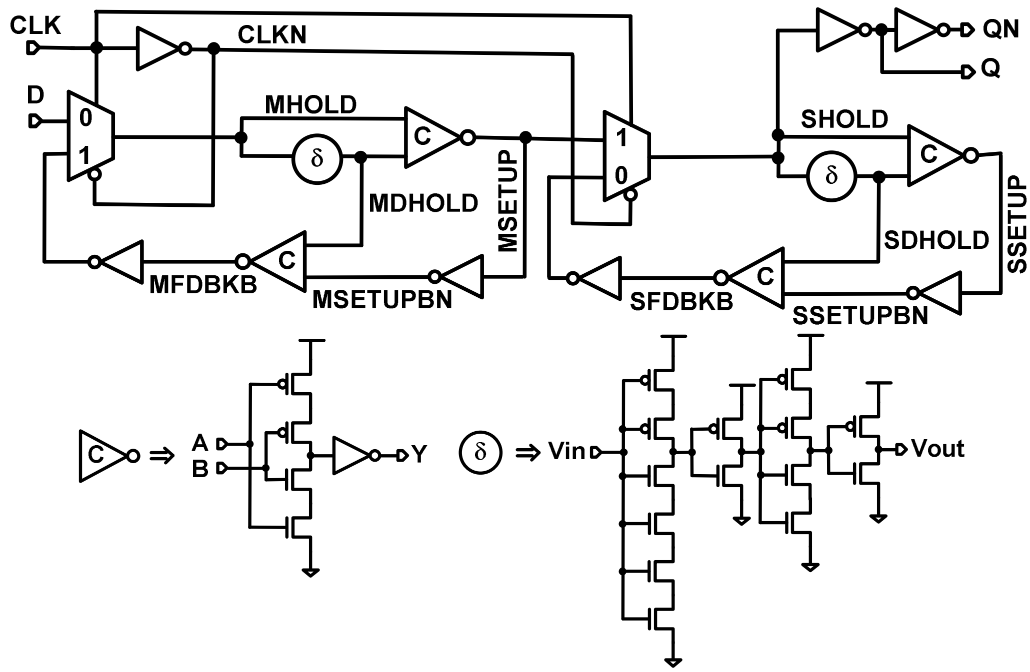


Fig 2.6. Temporal FF using delay elements inside the design.

2.1.4. Temporal FF using internal delay elements

An different temporal FF described in [Sandeep15] is shown in Fig 2.6. This design uses temporal sampling of the data inside the FF by using the delay filters shown in Fig 1.5. The single delay filter used in each latch of the FF is effective in both the transparent and hold modes. Each delay element is shared between the setup and hold nodes to reduce the use of two delay elements for each FF. These delay filters on each of the setup and hold nodes of master slave latches protects the design from the SETs on the clock and Data input. It also protects the storage nodes form SEUs with delay less than or equal to the delay element delay. Unlike the DF-DICE, these designs can be immune to a strike at any node.

As we had previously discussed, designing delay element takes multiple gate stages to match the delay of SET width. Moreover, it consumes more power also. The

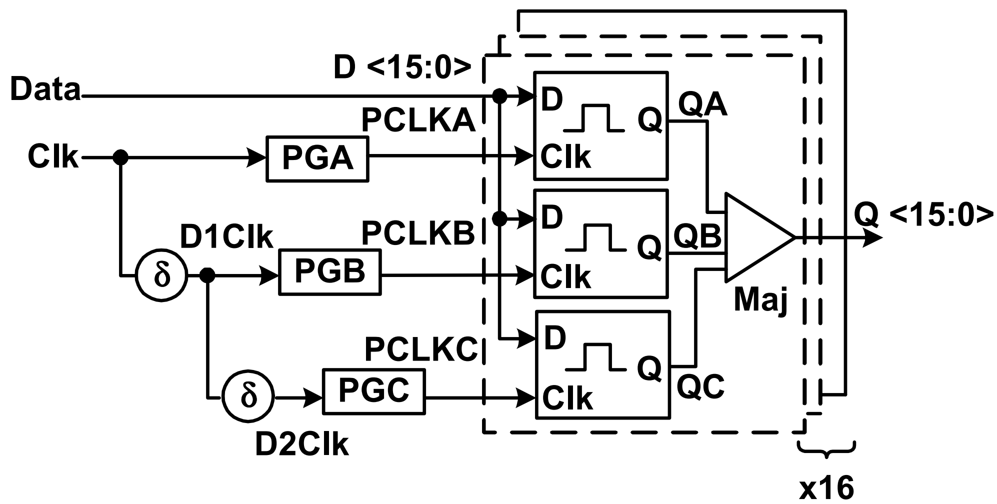


Fig 2.7 Temporally sampled TMR Pulse latch design [sushil15].

temporal FF, shown in Fig 1.6, uses two delay elements per flip-flop. This design was implemented in TSMC 90nm process and was observed that the delay elements occupy about 30% of the total area of the design. Inserting the delay elements in the design is a challenge requiring careful design.

2.2. Temporal Pulse Latch Design

This Temporal Pulse latch design [Sushil15] is the base FF design on which this thesis is based. It has all the properties of this design, so it is discussed in detail here.

The sampling of data by temporally separated clocks to mitigate SEU and SET, initially proposed in [Mavis02], is shown in Fig. 1.4(b). By providing more than one t_δ between each clock edge, any SET at the D inputs is sampled by at most one FF, assuming that the delay t_δ is greater than the SET width. The majority gate output is the majority of the three inputs, which will be the correct the output even if one of the inputs captures wrong data. This design is only hard to SETs on D input and SEUs on the storage nodes. Any SET on Clock input will generate false edges and may capture wrong data and send it to output. In such cases all three FF's will have the wrong data and it propagates to the output.

The initial design approach replaces TMR FFs in Fig. 1.4(b) with TMR pulse-clocked latches as shown in fig. 2.7. As mentioned pulse-clocked latches simulate FFs and can reduce power consumption by over 40%. Three separate pulse generators are used to generate pulses for each TMR copy of the latch, so that the pulse width generated doesn't get degraded by the delay elements. Simulation results showed that power savings is maximized when multiple (16 here) latches share one pulse clock generator and delay elements, since pulse generation is power expensive. 16 latches were chosen as it provides good area utilization while still ensuring pulse fidelity. The pulsed clocks PCLKA, PCLKB, and PCLKC are local to the pulse FF macro, so clock pulse waveform quality is well controlled.

With the addition of pulse latches and sharing the pulse generator between 16 TMR latches in each macro, power consumption is reduced almost 30% but hardness is retained. However, the design is still not hard to clock SETs. To overcome this problem the delay elements used in the design are changed to delay filters (DFs). FFs protect the clocks from SETs as we already discussed. The new design is shown in Fig 2.8.

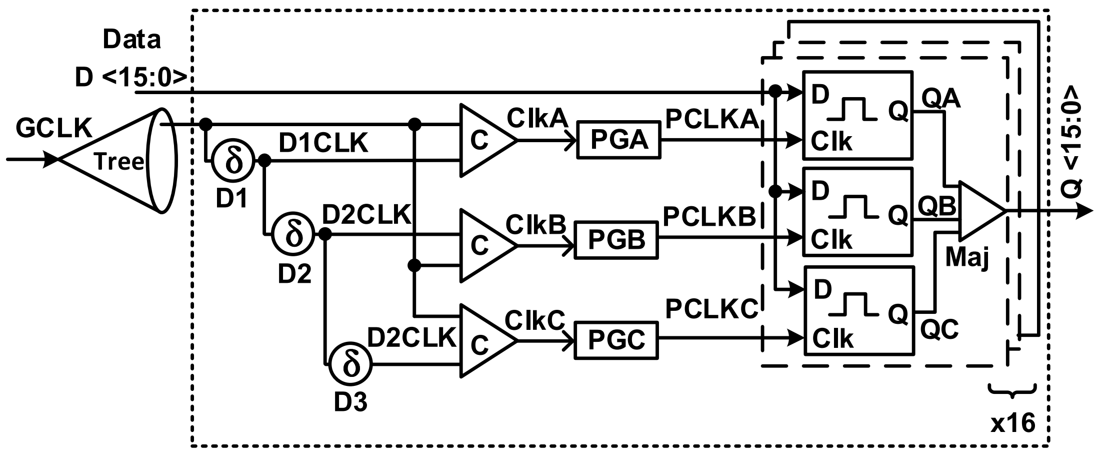


Fig 2.8 Temporal Pulse Latch design.

A global clock tree distributes the clock to the FF macros. The global clock GCLK is delayed using three delay filters (delay element and C-element combination) at each FF macro. These three delay filters will generate three clocks D1CLK, D2CLK and D3CLK that are separated by time t_δ. These temporally separated clocks ensure that SETs of width t_δ and below are not captured by more than one pulse latch. These three clocks are the inputs of the three pulse clock generators to produce temporally separated pulse clocks. The three delay filters along with the PGs and TMR pulse latches macro combined consists of a single layout chain.

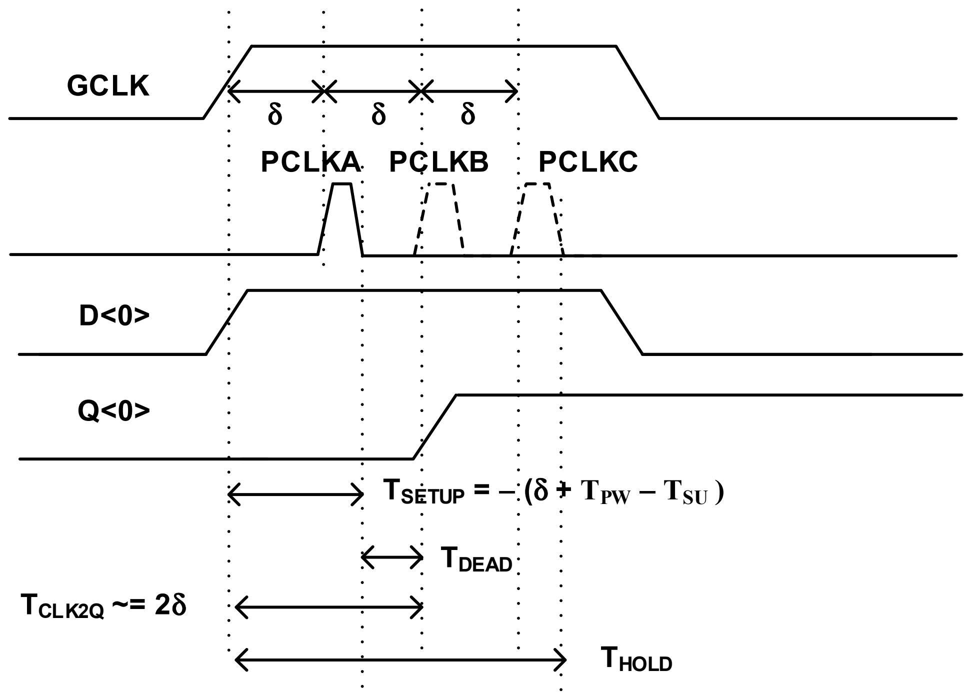


Fig 2.9 Timing parameters of temporal pulse latch design.

2.2.1. Hardness and Timing analysis

The timing parameter of this design are shown in fig 2.9. The D input of the TMR latches should setup to the falling edge of the pulse clock PCLKA. When calculating the setup time with respect to pulse clock PCLKA, it may be just the setup time to falling edge of pulse. Calculating properly with respect to global clock, the setup time is T_{setup} = ~(δ+ Tpw-Tsu), where Tpw is pulse width and Tsu is setup time of latch. It is clear from the waveforms, when there are no abnormalities, the output Q is generated at the falling edge of PCLB, as two copies of majority gate are correct. This may not be the same case with all the cases, like the case were A copy is wrong, then the data should be held till it is captured by the C copy of the latch. So, the hold time of the design considering from GCLK should be T_{hold} = 3δ+Tpw. The dead time of the FF is large (in range of 1.5 ns),

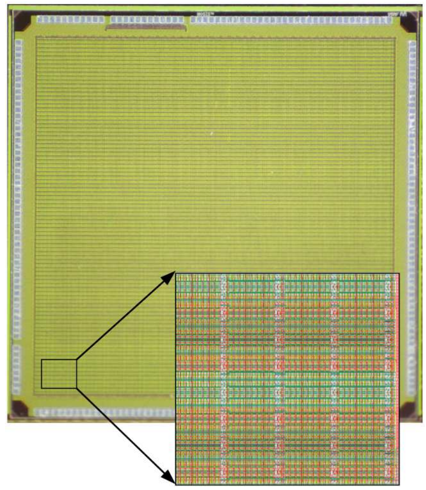


Fig 2.10 Die photo with test structure inset [sushil15].

which makes the design slow. A minimum logic of t_δ is required between the FFs to avoid any hold violation in a pipeline design While designing a shift register using this FF, a t_δ hold buffer was placed in between the FFs.

This design is hard to SEUs and both clock and data SETs with widths upto $t\delta$. TMR copies of the latches make sure that the design is hard to single SEUs on the storage nodes as even one SEU on a storage node can be mitigated by the majority gate at their output. Any SET on data D in this temporal design is only captured by at most one latch and the majority gate makes sure the output is correct. Any SET on the global clock tree before the delay filters will be mitigated near the C-elements. SETs after the delay filters will effect only one local clock by generating false pulses or by diminishing the actual pulse. Neither of these will cause any error as the other two copies are still correct and corrects the output. The hardness of this design depends on the delay that is provided between the clocks. As we increase the delay between the clocks, the speed of the design decrease.

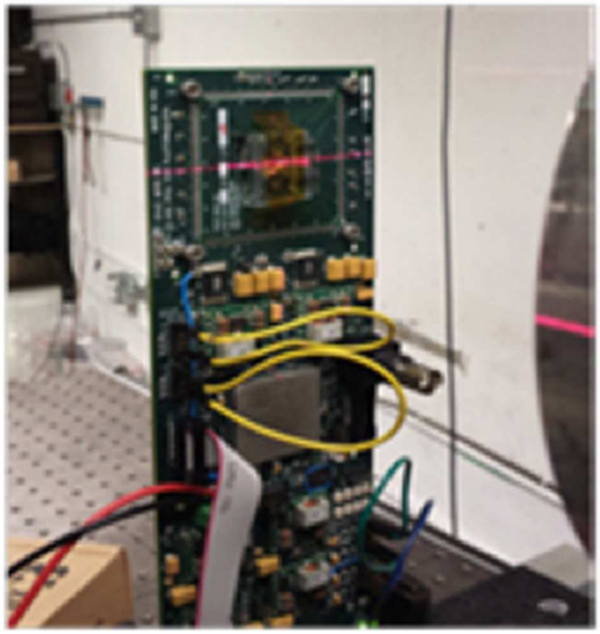


Fig 2.11 Beam testing setup at UC Davis with the DUT in the beam line. The controlling FPGA is at the bottom, away from the beam line.

2.2.2. Pulse Width Determination

A key problem with pulse-clocking is the high quality of the pulse clocks required over systematic (process, voltage and temperature) as well as random process variations. Monte Carlo (MC) simulation were used to determine minimum latch pulse width required, as well as pulse width generated by the PG. The pulse-width (t_{PW}) required by the latch has μ = 97.53 ps and σ = 7.65 ps. A 4σ design requires a 129 ps pulse width. The PG was designed to get pulse width grater that the worst case pulse width required. Monte Carlo simulation determined that the generated clock pulse width has μ = 153.5 ps with σ = 5.7 ps. The resulting target PG provides a margin over 6σ between calculate and generated pulse widths.

2.2.3. Test chip and experimental results

The TPL design was fabricated on a 90-nm low standby power foundry process. Test structures comprised of parallel shift registers with these proposed TPL, as well as unhardened standard foundry library FF designs were included. The Test chip Die photo is shown in Fig. 2.10 along with the FF test structure. The designs were tested by broad beam proton irradiation (Fig. 2.11) at the UC Davis Crocker Laboratory with 63 MeV protons. The die was not lidded (it is a COB as evident in Fig. 2.11) but the plastic protective covering was left in place during irradiation.

The FFs were tested in SEU only and SEU and SET sensitive conditions, i.e., clock held low during irradiation and clocked, respectively at V_{DD} = 1 V. The primary goal of the testing was another design, so the results are limited. But in static testing this design had no failures with a flux of 70.14×10^{6} protons/cm^{2}-s and a total fluence of 41.8×10^{9} protons/cm^{2}, while the unhardened designs, using the foundry supplied flip-flops had 2 errors. In dynamic operation with a flux of 70.14×10^{6} and total fluence of 41.8×10^{9} protons/cm^{2}, the unhardened designs exhibited 14 errors while the TPL hardened designs again had none. Taking into account possible statistical error (i.e., proportional to the square root of the error count) the cross-section of the proposed design in dynamic operation is at least 90.25% less than that of the baseline foundry FF. Static operation has insufficient baseline failure data to make such an estimation. However, given the spatial separation of redundant latches, we believe that the improvement should be even greater, since two latches would have to collect charge.

While protons have been shown capable of upsetting SRAM cells via direct ionization [Heid08], for the energy used here, the proton LET in silicon is under 9×10^{-3}

MeV-cm^2/mg and thus cannot upset these 90-nm latches, which have higher capacitive loading than SRAM cells. Thus the upsets are via indirect mechanisms, either elastic or inelastic scattering. The cross-section of these interactions are about five orders of magnitude lower than that of direct ionization [Paul99], leading to the low failure count in the baseline FF.

2.3. Conclusion

In this chapter we discussed about some of the previous RHBD designs DICE, DF-DICE, BISER etc. A brief literature survey on those designs has been done. A Temporal pulse latches based FF design is described in detail with all its properties and the implemented Test chip and its beam test results. This design is the base design for the novel design that is going to be proposed in the coming chapters.

CHAPTER 3. Proposed TPL design

3.1. Introduction

This chapter describes the proposed temporal pulse latch (TPL) design using redundant skewed clocks. Also presented in this chapter are the various operating modes of the proposed design, which make it more flexible to use.

3.2. Proposed Design Implementation

In the previous chapter, we discussed the temporal pulse latch design [sushil15] that is the starting point for this proposed design. The TPL design, when compared to similar implementation with FFs saves about (30 – 40)% energy. It also reduces the design area, as latches are used in place of FFs. It uses DFs to protect the clock inputs from SET similar to DF-DICE. Compared to DF-DICE, this design uses only three DFs for a 16 FF macro group. Although optimized from a power and area perspective, when considering a processor level design consisting of several thousands of sequential elements, many DFs are still used with the TPL design. Since the DFs are in the clock path, they have a high activity factor. Their power dissipation is also large since they introduce intentional delay and capacitance. Therefore, the focus of this thesis is to improve the power consumption of the design by reducing the number of DFs used in the entire design. Several features are added while reducing power consumption that improves the flexibility of using the design in different operational modes.

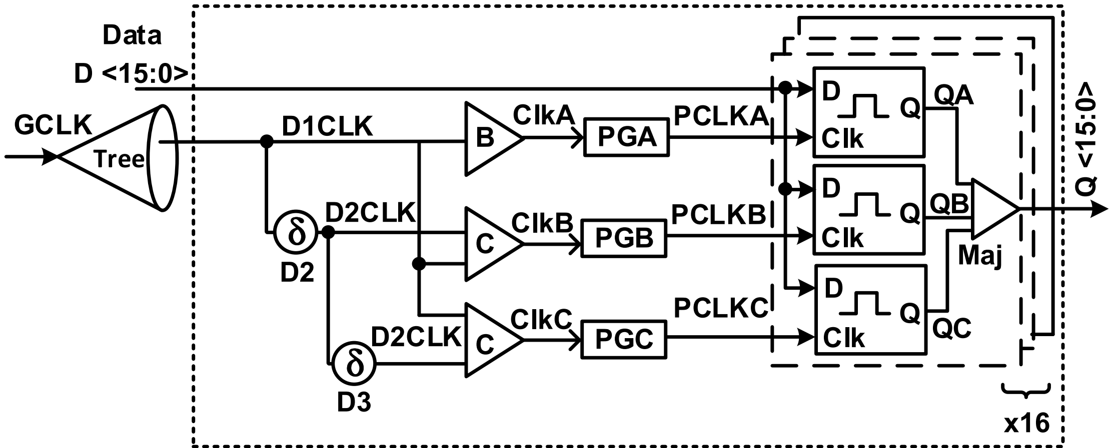


Fig 3.1 Improved version of TPL design.

The fastest and easiest way to reduce the DFs in the design is to remove the DF in the CLKA path (Fig 2.8) by eliminating the first delay element D1 and generating CLKA directly from GCLK. This makes the CLKA vulnerable to SETs, but will not affect the functionality of the design. If GCLK is affected by a SET, it only propagates to CLKA and other two paths remain unaffected because of the DFs. The majority gate at the output corrects this error through majority voting. Fig 3.1 shows the improved TPL design schematic. This reduces the DF per FF group macro by 1/3. This improvement reduces the flip-flop energy per operation by 6% at 100% clock and 25% data activity factor. This is not a great improvement in power, but a step in the right direction.

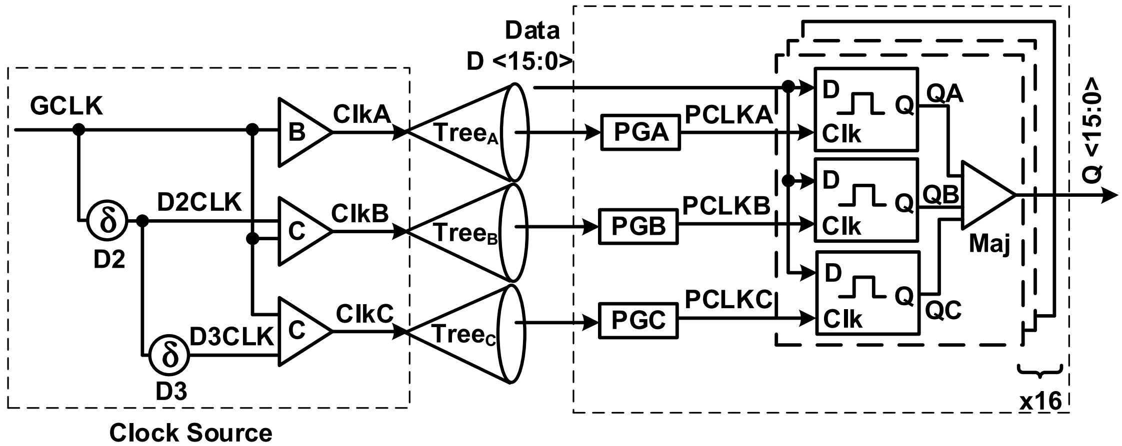


Fig 3.2 Proposed Redundant Skewed clocks based TPL design.

3.2.1. Multiple Clock Implementation

Further improvement at the macro level is a difficult task. The important point that we need to observe in Fig 3.1 is that the pulse generators (PG) are shared across 16 pulse FF's in the macro. The PGs are shared across 16 FF group because they consume a lot of power and are present in every multi-bit macro of the design. In order to save power, our next step is to share them across the entire design. This will save lot of power, however, the problem is ensuring the pulse fidelity when shared across large number of FFs. The variation of the pulse increases and controlling the pulse width will be a cumbersome design task. Since pulse width plays a critical role in the performance of the design, we cannot share the PG's across an entire design or block. Though the PGs cannot be shared, the delay filters that are driving the PGs, can be shared across the entire design.

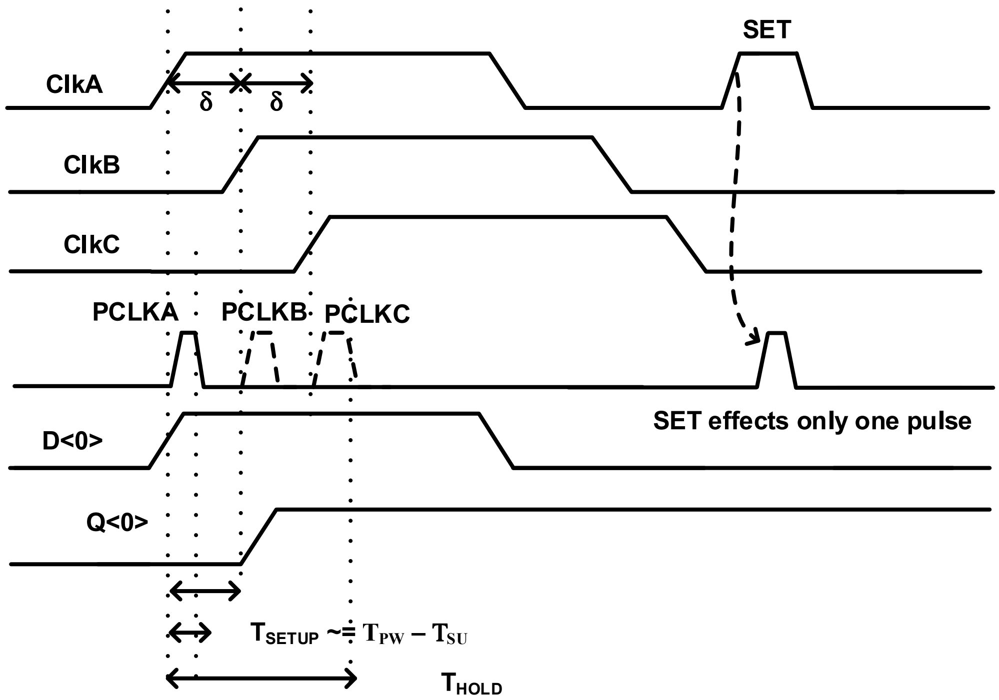


Fig 3.3 Modified Timing Window for the proposed design

We propose to share the two delay elements in the FF macro across an entire design, which may contain any number of such FF macros. This looks like a small improvement from the locally delayed TPL design by just moving the position of the DFs, but this small improvement will make a large impact on the design. Additionally, in this chapter many features are added to the design that helps in design flexibility. All the additional features of the design will be discussed in further sections.

The proposed redundant skewed TPL design is shown in Fig. 3.2. The delay filters are moved to the clock source leaving the pulse generators local to 16-bit FF macro. Now instead of one clock tree, we need to create three separate clock trees CLKA, CLKB and CLKC for the design. These three clocks are supplied to the three copies of the pulse generators PGA, PGB and PGC respectively, which generate the pulsed clocks

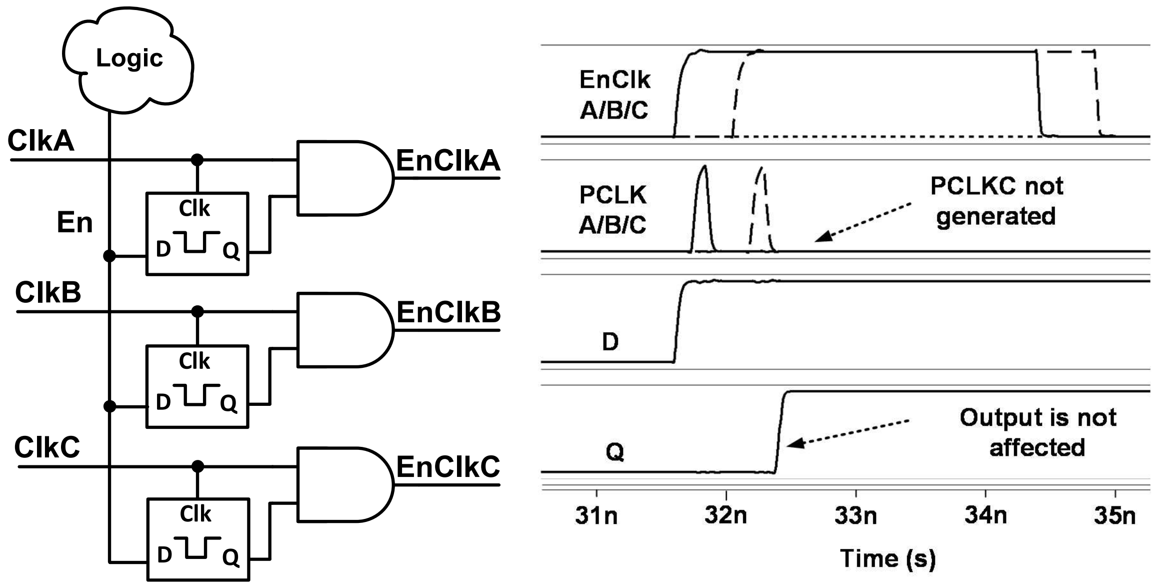


Fig 3.4(a) Clock is gated on each clock. (b) Simulation waveforms showing the design functioning correctly even when one of the clock gater is upset by an SEU or SET.

locally. Only two delay elements are used in the entire design compared to two delay elements for every 16 FFs.

3.2.2. Modified Timing Window

Fig. 3.3 shows the modified timing window for the proposed design. Since one delay filter is removed from CLKA path, the pulse clock PCLKA is generated as soon as the CLKA appears. This improves the setup time of the design to $T_{setup} = \sim (T_{PW} - T_{SU})$ considering from GCLK/CLKA, as it is improved by a delay element delay. The actual setup time to the latch does not change. The rest of the timing parameters remain the same compared to the original design.

3.2.3. Clock Gating

In the previous design where the temporal clocks are generated locally, clock gating is very difficult. Clock gating done on the single tree is prone to soft-errors at the gater control input or in the latch in the integrated clock gating (ICG) cell, propagating the incorrect clock to all redundant copies. To protect them from soft errors we need to place them inside each FF macro. This is also problematic, since it increases vulnerabilities and the macro power dissipation and area, with the previous design, we never found a fully acceptable solution.

The redundant clocks simplify clock gating dramatically. We can gate the clocks separately at any stage of the design as shown in Fig 3.4(a). Since all the three clocks CLKA, CLKB and CLKC are delayed with respect to each other, any SET on the enable signal will not be captured by more than one clock. As shown in Fig 3.4(a), redundant latches in the standard configuration ensure that the enable hold times are met. While the latches are still vulnerable to both SEU and SET at their controlling inputs, the overall system is robust to errors on a single clock copy as shown in Fig 3.4(b). If the clock gater controlling signals are generated by rising edge clocks, there is a significant hold time that is difficult to meet at the third (ClkC) latch closing edge. However, it is less than the hold time required by receiving TPL flip-flop C copies since there is no pulse-generator. These three separate global clocks help in gating each clock separately, which helps in reducing the power and increasing the speed of the design. These are discussed in detail in ensuing sections.

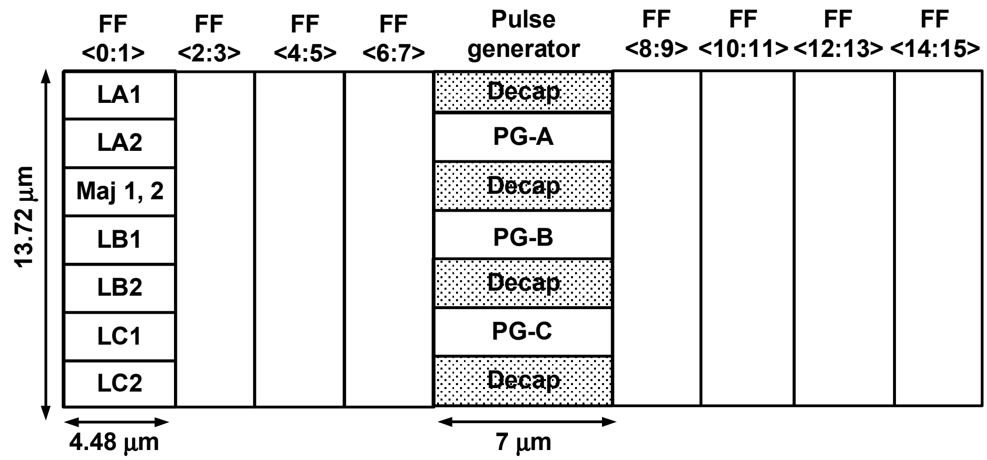


Fig 3.5 Layout of the proposed FF design. The delay elements from the previous TPL design were replaced by de-coupling capacitances to provide spatial separation between pulse generators.

3.2.4. Physical Design for MNCC Robustness

The physical layout of the improved 16-bit FF macro is illustrated in Fig 3.5. It is implemented in a 90 nm low standby power (LSP) process. Each FF in our design consists of three pulse latches. Our design is hard if only one of the three latches hit by an SEU, but it can generate wrong outputs when two are simultaneously upset. So, three storage nodes of the latches are spatially separated to protect the design from multi-node charge collection (MNCC). We combine two FFs (6 latches and 2 majority gates) into each single column. These latches are interleaved such that latches of same FF are separated by at least one standard cell height. LA1, LB1 and LC1 of FF1 are separated with LA2, LB2, and LC2 as shown in Fig 3.5. Vertical interleaving provides an intervening N well between potential upsetting nodes. The N-well is biased at V_{DD}, thereby providing a good charge sink. Together, these layout techniques protect the FFs from MNCC. The two majority gates are combined into one cell to save area. This means that two FFs are combined to form a column seven standard cells high. Each standard cell height is 1.96 μm, so the total height of the column is 13.72 μm.

Our design is a 16-bit FF macro. These 16 FFs are divided into two groups of 8 FFs each and the pulse-generator unit is placed in the middle as shown in the Fig. 3.5. This ensures that the pulsed clock is distributed among FFs with minimal variations. The clocks also need to be protected against MNCC. Thus, the pulse generators are also separated from each other with a de-coupling capacitor standard cell. Removal of the three delay elements from the macro allowed us to reduce the height from eight standard cells to seven standard cells tall. We can remove the two additional de-coupling capacitors (top and bottom) used in the design to save area if needed, resulting in a non-rectangular macro.

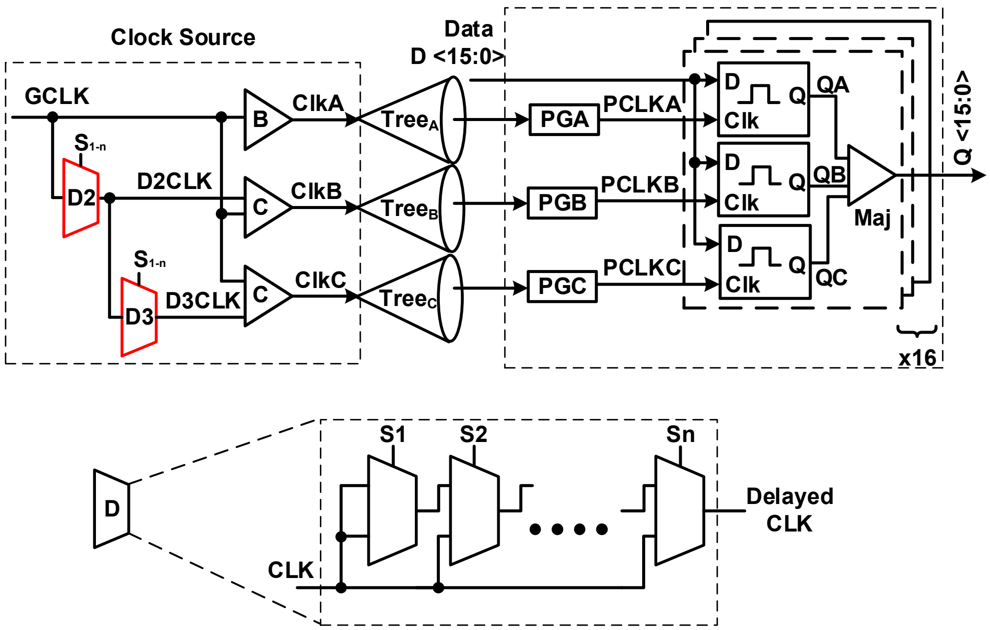


Fig 3.6 Proposed Redundant Skewed clocks based TPL design with Programmable delay elements.

3.3. Programmable Hardness Implementation

One of the important features of a radiation hardened design is the degree of hardness that the design achieves. For temporal designs like our design, hardness depends on the time interval at which the data is sampled, consequently it's a function of the delay between the clocks in this design. In original design, in order to change the hardness, we need to redesign the sequential elements with a different delay element, which cannot be modified once the IC is designed and the chip fabricated. If we wanted to add a programmable or variable hardness to the design, all the delay elements must be programmable. We need additional signals, that are routed to each macro, to change the delay one the fly, which increases the complexity of the design.

The proposed design uses only two delay elements in an entire design, at the root of the clock tree. This makes it easy to change the delay at any stage of the design. We can replace the delay elements with programmable delays, as shown in Fig 3.6, the delay between the clocks can be programmed at any time. We can select the delay between the clocks by selecting the desired multiplexer in the delay line to change the delay. Since there are only two of such programmable delay elements, they can be designed in various configurations without worrying about the size and power that they may consume, since the impact on overall power is minimal. They can also be designed such that there is minimal variation in the delay across different process corners, i.e., large variation resistant elements will not have an overall adverse impact.

3.4. Different Modes of Operation

Most radiation hardened designs, do not require their hardness properties 100% of the operating time. Hardness is often required for some critical operations. Designs in outer space require low power and unhardened operations for most of the time since they are power dissipation limited. For such applications, we would like these hardened designs to be used in low power and unhardened modes. Most current hardened designs do not support low power or non-redundant modes [Zhang06] [Naseer06] [Sandeep15] embedded into them. To add such features to hardened designs is potentially complex and requires additional circuitry and design effort. In some cases, such low power modes simply cannot be added.

In the proposed design, we have the opportunity to add different modes of operations depending on the power, speed and the hardness required by the design.

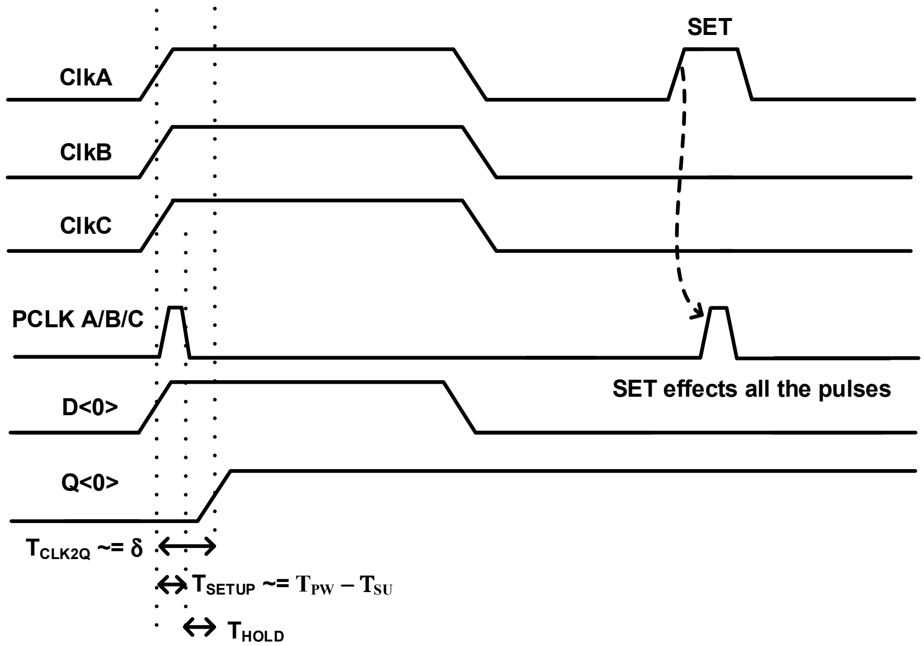


Fig 3.7 Timing waveforms of Fast and SEU hardened mode is shown.

Additionally, we can switch from one mode of operation to another on the fly. This can be achieved with minimal added circuitry as shown next.

3.4.1. Full Hardened Mode

This is the proposed operation of the design without any modifications as shown in Fig. 3.6. In this mode, the design is hard to all the SETs on data and clock, SEUs on the storage nodes. This design is slow since the clocks are delayed and the dead time of the pulse latch is increased.

3.4.2. SEU hardened only mode

This mode can be achieved by reducing the delay between the three clocks to zero such that all three clocks will be identical. The programmable delay element delay can be reduced to zero in Fig. 3.6. This makes the design fast, since the dead time of the design decreased to that of the latch. Now this design can run at maximum speed that is implemented with standard library sequential elements. The improved timing waveforms are shown in Fig. 3.7. We can observe that all the pulses PCLKA, PCLKB and PCLKC are identical and the FF macro generates the output immediately after the three pulses have fallen.

In this mode, the design is only hard to SEUs on the storage nodes. Any SETs on the clock and the data may be captured by all the three copies of the pulse latches (Fig 3.7) and will propagate to the output. The majority gate will not be helpful in this case. This design uses the same amount of power as the original design as it uses all the circuits. This modification of the proposed design also does not need any additional circuitry except the programmable delay elements, which we intend to use in the actual design, regardless.

3.4.3. Non-redundant Low power mode

This section describes one of the most important contributions of this thesis. We can make the design non-redundant with a small modification to the FF design with additional control signals. Since all the three clocks are global, it is very easy to gate each of the clocks separately as we discussed previously. By gating two clocks CLKB and CLKC, we can propagate only CLKA to all the sequential elements in the design. We need to make minimal changes to the proposed design, so that the default values stored in B and C copies of the pulse latches do not affect the majority gate output in this proposed non-redundant mode.

This mode saves almost 2/3rd of the power that is consumed by the sequential elements and achieves almost the same power as the conventional FF design. It can also achieve the maximum possible frequency since there is only one clock and no delay

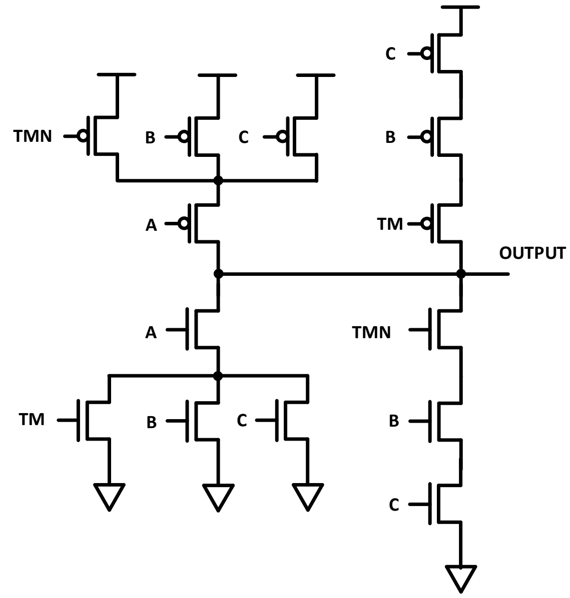


Fig 3.8 Modified Majority gate to include the non-redundant mode.

elements. The FF design can be modified in different ways to achieve the non-redundant mode. Two modifications that can be made to the FF design are discussed here.

3.4.3.1. Modified Majority Gate

The modified majority gate for the design, to operate in non-redundant mode, is illustrated in Fig 3.8. To the actual majority gate design, we added two additional signals NRM and NRMN, which helps in switching the design into non-redundant mode. NRMN is the inversion of NRM. When NRM is 0, then the B and C stack becomes active, making the design work normally as a majority gate. Then the entire design will be in hardened mode and the output of the sequential elements changes depending on all the three copies of the latches. When NRM is 1, then the non-redundant mode is activated. The B and C stack is cut-off from the output. Since A is closer to output node in other CMOS stack, pull-up and pull down are always connected to power rails due to NRM signal, the output depends only on the A input, and is independent of the B and C inputs.

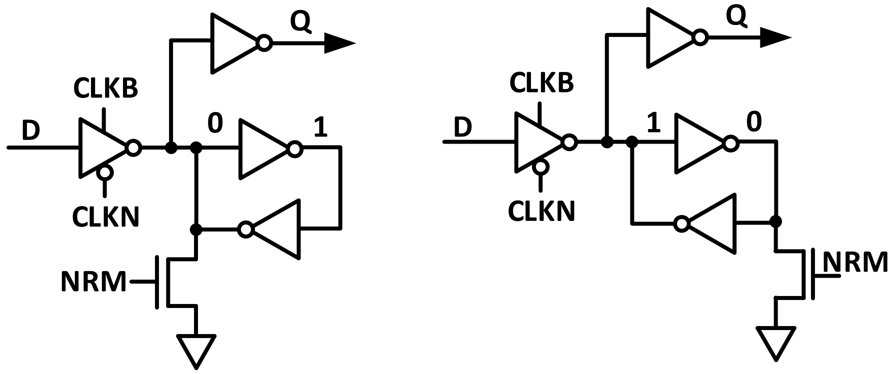


Fig 3.9 Modified latch design to include the non-redundant mode.

We can gate the clocks CLKB and CLKC so that the B and C copies of the latch are inactive. The values stored in the B and C copies of the pulse latches are irrelevant.

3.4.3.2. Modifying the Latches

Modifying the majority gate has some additional overhead. We need to generate NRM and its inverse NRMN for each FF. It also increases the stack height of B and C gate stack from 2 gates to 3 gates. This will increase the delay slightly.

These complications can be eliminated by using the following approach. Instead of modifying the majority gate, we can modify the B and C copies of the pulse latches such that we can store opposite values (0 and 1) in each of the copy. We can store a 0 or 1 in a latch by pulling down one of the storage node by a pull down NMOS transistor. This technique is illustrated in the schematic of Fig 3.9. Using this technique, we can save opposite values in B and C latches such that the majority gate responds to the value of latch A output and generates the majority output based on the value that is stored in latch A. Pull down transistor is controlled by NRM signal.

This approach reduces the number of transistor added for each FF compared to the previous approach. We only need to use NRM signal and the load on this pin is also reduced as we are using only two small transistors for each FF to pull down the storage nodes. Therefore, the complexity of the design reduced with this approach.

3.5. Conclusion

This chapter discussed the idea of globally delaying the redundant clocks. We also showed how the proposed implementation followed the previous locally generated redundant clocks based TPL design. We discussed the advantages of power saving compared to the previous design and the design flexibility of varying hardness. The different modes of operation that can allow for low-power operation that is implemented in the proposed design are also discussed. We are going to compare the power consumed by the proposed design with other FF designs by implementing an advanced encryption standard (AES) with these sequential elements in the next chapter for a detailed power, performance and area metric comparison.

CHAPTER 4. Power and Hardness analysis

4.1. Introduction

In the following chapter, we implemented on an advance encryption standard (AES) design with the proposed global clocked TPL as well as the local clocked TPL design to compare the design metrics of the two designs. The power consumed by the proposed design in different modes is also compared with several other hardened and unhardened designs. Finally, timing and resulting hardness variation are explained using Monte-Carlo simulations.

4.2. AES Implementation

To study the clock quality and compare the power dissipation of the local delay vs. global delay with TMR clocks approaches, we synthesized, placed and routed an advanced encryption system (AES) engine using the two TPL schemes. The AES engine is a 256-bit key, 128-bit data and fully pipelined as described in [chella15]. We used the same foundry low standby power 90-nm process as the TPL test chip. Standard commercial CAD tools (Cadence Encounter, Synopsys PrimeTime and Nanotime) are used for automated place and route (APR) and timing analysis, respectively. Foundry standard cells are used for combinational logic. As mentioned in previous chapters, the two TPL multi-bit macros have similar footprints, only the clock generation is significantly different. The pulse widths required for the robust pulsed-clocking of the latches uses Monte-Carlo analysis with the foundry variation parameters [Sushil15].

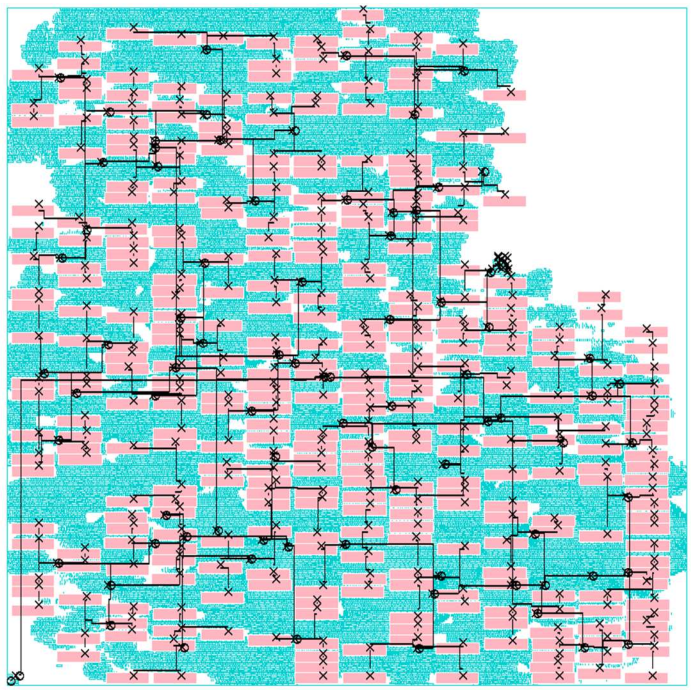


Fig 4.1 AES implemented a single clock tree using TPL with local delay generation [Aditya15].

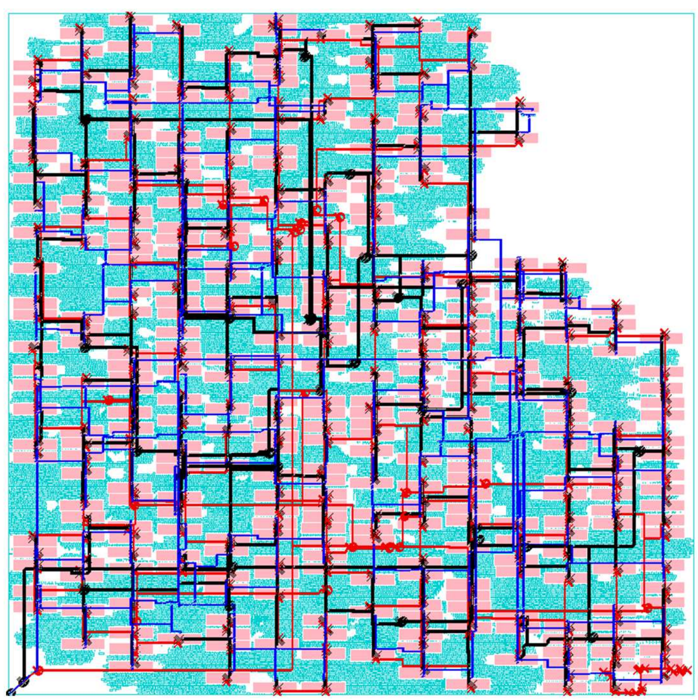


Fig 4.2 TPL FF protected AES implementation with TMR skewed click trees. The redundant clocks are shown in black, red and blue.

The three clock trees are spatially separated during physical design to ensure

	Local delay generation	TMR clocks
Density	74.60%	72.01%
Clock Buffers	83	103
Clock skew (ps)	11	52
Clock tree area (μm^2)	717.28	905.52
Total FF area (μm^2)	259630	247268

Table II: clock tree parameters of the AES implementations: single clock tree with local delay generation in the multi-bit FFs and TMR clocks

MNCC hardness using cell halos, which keep other clock cells from being within a specified distance. We remove the halos after freezing the clock tree post-CTS optimization, freeing the space for logic optimization. Both the designs were implemented using the same floorplan (844 μm by 842 μm) and the same timing constraints, with a clock period of 200 MHz. Figs. 4.1 and 4.2 show the clock trees synthesized for the two implementations of an AES engine. The two designs variants, single clock tree with FF macro integrated delays and triplicated skewed clocks, are compared in Table II. Density is slightly impacted by the use of TMR clock trees, although the total number of buffers was not large. The clock skew, as we originally feared, is substantially greater with TMR clocks. However, this is mitigated by removing random skew variation in the delay filters, which is greater, as discussed in subsequent sections. Note also that the total FF area reduction is greater than the increase in clock buffer area.

Data input Activity Factor (α)	SR tree (pj)	TR tree (pj)	Local delay design (pj)	Proposed TMR clocks design (pj)	BISER (pj)
0	10.32	16.82	353.68	289.46	177.99
0.25	10.32	16.82	415.02	349.10	257.06
0.5	10.32	16.82	479.35	405.33	331.35
0.75	10.32	16.82	540.69	461.14	406.33
1	10.32	16.82	611.41	521.63	481.31

Table III: clock Energy per operation for TPL with local delay generation and global delays with TMR clocks, as well as BISER FFs at different activity factors.

4.3. Power Analysis

Power dissipated in the clock tree is linearly related to the number of clock sinks in a design [Vittal97]. Consequently, the design with single clock tree or triple clocks trees drives the same load and thus theoretically requires almost the same amount of energy to drive the load. However, clock tree synthesis (CTS) minimizes skew by over and under driving clock nodes. The placement is also constrained, resulting in deviations from the ideal power consumption theoritically. In our experiments here, using three redundant trees consistently increased the clock-tree power dissipation by about 60% over using one tree. Fig. 4.2 shows more clock routing than Fig. 4.1, since each macro must now receive CLKA, CLKB, and CLKC.

Data input Activity Factor (α)	Proposed TMR clocks design (hardened mode) (pj)	Proposed TMR clocks design (non-redundant mode) (pj)	Flip-flop design (pj)
0	303.09	109.89	98.66
0.25	356.26	134.43	131.37
0.5	407.38	161.01	164.09
0.75	452.36	187.59	193.53
1	507.57	212.13	226.25

Table IV: clock Energy per operation for TPL in hardened and non-redundant mode compared to the design implemented with standard FFs.

Nonetheless, the TMR clock tree approach saves considerable power dissipation overall when analyzed on an energy per bit basis, due to the elimination of the 1278 delay circuits. The integrated clock and TPL approach reduces the overall energy by 15% at 50% data activity factor and by 18% at vanishing (near zero) data activity factors, over the already improved version of the TPL design (see Table III). Table III also compares these schemes with the BISER flip-flop [Zhang06], which uses two redundant flip-flops to provide SEU, but not SET mitigation. The BISER uses five latches (one jam latch at the output is required to save state when one of the FFs mismatches) per bit of storage. Since our design uses three latches per stored bit, the BISER provides an interesting comparison point. Nonetheless, the BISER flip-flop dissipates around 38% less active power compared to our design at vanishing data activity factors and about 7% less at high data activity factors. This is due to larger clock loading in our design, i.e., the pulse generators and clock buffers, that dominate the energy consumption at low data activity factors. Our proposed TMR clocked TPL scheme cannot meet the energy per bit of this design, but does provide complete SEU and SET mitigation.

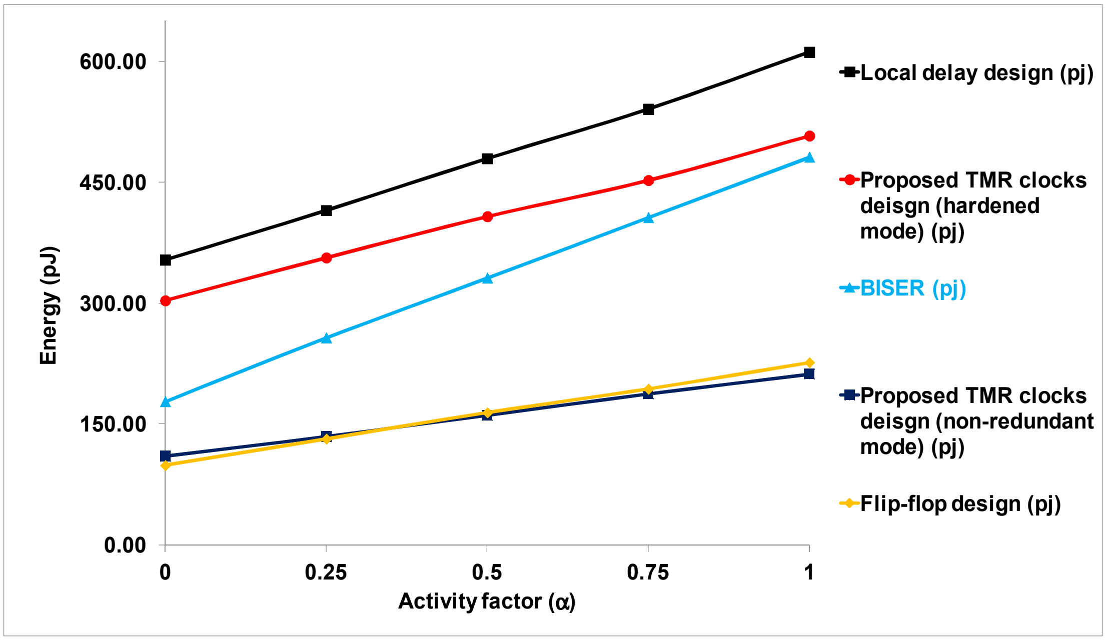
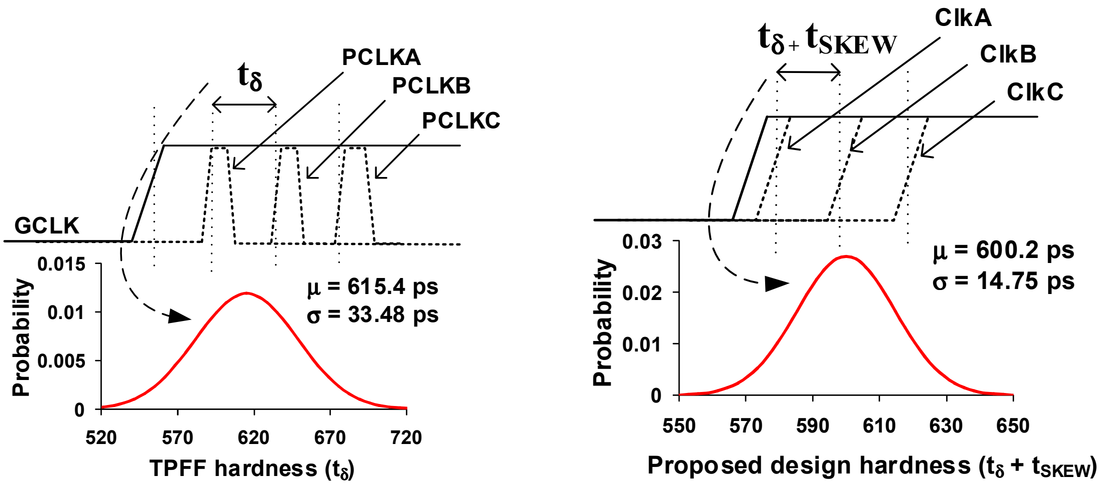


Fig. 4.3 Graph showing energy comparisons between different design implementations

The proposed design is modified to include the different operational of modes that we discussed in the previous chapter. We also compare the power of the design between the hardened mode of operation and the non-redundant mode, where the clocks CLKB and CLKC are gated off, shown in Table IV. In the non-redundant mode, this design consumes about 64% less energy per operation compared to hardened mode at vanishing data activity factors. Even in the worst cases, it consumes 58% less power. This is expected as only one clock tree and one latch is active during non-redundant mode. When compared this design with a design using standard cell library FFs, our design is only consumes 11% more energy per operation at vanishing data activity factors. It consumes 6% less power at high data activity factors. At vanishing data activity factors, our design consumes additional power due to high clock load due to pulse generators. Fig. 4.3 shows



(a) (b)

Fig 4.4 Analysis of the spacing between sampling windows afforded by (a) local delay generation and (b) global delay generation with TMR skewed clocks. The latter exhibits improved variability.

the graph comparing the energy per operation of the various designs [Aditya15] [Sushil15] [Zhang06] that are implemented on the same process for a valid comparison.

4.4. Area Analysis

The AES implementation shows that the single clock tree uses 25% fewer cells than the TMR clock trees combined. However, this is a negligible impact on the overall block area. The resulting original local delay TPL and proposed TRM clock implementations have similar area utilization of 74.6% and 72.1%, respectively. The increase in area due to the added cells in the redundant clock tree can be compensated by 5% smaller macros. For these experiments, to focus on the clock impact solely, we kept the TPL macros the same in these design trials, as stated in previous chapters.

4.5. SET Hardness and Delay Variation analysis

SET hardness can be characterized by the width of the SET glitch that can be mitigated by the design, here the time delay between the redundant clock pulses. This SET hardness varies from latch to latch due to process variability, exacerbated by systematic clock skew between the TMR clock tree endpoints (Fig. 4.4). We investigated this using Monte-Carlo simulated variability of the delay elements, comparing it with the systematic skew between clocks driving the same multi-bit TPL macros.

The delay circuits have a mean delay of 615.4 ps, and a variance (σ) of 33.4 ps. The delays due to the skew in the clock trees however, have a mean of 600.2 ps and a variance (σ) of 14.15 ps. Thus, the clock tree skew, at least for this modest sized design, is less than the variability introduced by many separate delay circuits. The worst-case (i.e., smallest) separation between two pulses is 523 ps and 561 ps for the local delay and global delay generation (TMR clock) designs, respectively. This is due to the relatively small size of the delay circuits required to minimize their energy contribution.

One key advantage of the proposed design is that, the impact of variability on hardness depends on the relative clock skew, which can be controlled to higher degree than delay variations across process corners and due to random variations. Since only two delay elements are required in the top level tree, these can be arbitrarily large to essentially eliminate random variations, without significantly affecting the overall design power dissipation. Another key advantage of the TMR clocks is that the SET delay, the design can mitigate, can be easily calibrated or adjusted. There is only one delay generation circuit so providing programmability is trivial as we discussed in the previous chapter.

4.6. Conclusion

In this chapter, we discussed about various metrics of implementation of the proposed design with local delay elements. We also compared the energy consumed by proposed design with previously published work. The SET hardness variation of the designs due to process variations is also elaborated. The advantages of the proposed globally skewed clocking TPL design over locally delayed clocks TPL design.

CHAPTER 5. Summary

The core idea of this thesis work is to reduce, increasing power consumption of the soft-error mitigating design. As a started point, we took an already robust design, hard to both SEU and SET on all inputs, that itself is a low power design. It is a temporal pulse latch (TPL) design, using pulse latches in place of FFs and sharing the pulse generators and delay filters across a group of 16 pulse latches, which saves around 30-40% energy over FF design approach.

This thesis work proposes an integrated approach to SEU and SET soft-error robustness by using skewed TMR clocks driving TMR pulse-clocked latches. The delay elements, local to FF macro in TPL, are shared across an entire design. This reduces the usage of number of delay elements to only two, compared to three delay elements for every 16 FFs in an entire design.

This approach divides a single global clock into three clock trees, which are skewed to each other by the delay element delay. This approach adds several additional features to the design. Since there are only two delay elements in an entire design, they can be replaced by programmable delays, changing the hardness of the design on the fly. By reducing the delay to zero, we can run the design at maximum speed, but only hard to SEUs. Gating the clocks is also made easy, due to three separate clock trees. This enables us to run the design in non-redundant mode by gating two clocks and making minimal changes to TPL design. In this mode, we can run the design at full speed without any hardness.

In order to verify the designs functionality and energy savings, we implemented both designs, proposed globally skewed clocks and locally delayed clocks TPL designs, on an advanced encryption standard (AES) engine. This is a relatively small design with 6816 FFs in the entire design. Proposed approach saves about 1278 delay elements, saving 5% of total area. TMR clocks rather than local delay generation provides a minimum overall power savings of 15% and up to 18%. Comparing it with an SEU only hard design, BISER, which dissipates around 7% to 38% less energy compared to proposed design, which is SEU and SET hard. When the design is used in non-redundant mode, it saves a minimum of 58% and up to 63%, of FF and clock energy, over redundant mode operation. We also compared the non-redundant mode energy dissipation with that of the standard cell FF approach, which consumes almost same energy with +/- 7% difference with data activity factor. We investigated the hardness variation between the two designs, clocks skew for proposed design and delay elements variation in previous approach, determining the variations to be 14.15 ps and 33.4 ps respectively.

This summarizes the thesis by implementing a novel design approach, which reduces the power consumption, adding flexibility to the design. The proposed scheme is the lowest power published approach to SEU and SET soft-error mitigation.

REFERENCES

[Koga93] R. Koga, S. D. Pinkerton, S. C. Moss, D. C. Mayer, S. Lalumondiere, S. J. Hansel, K. B. Crawford, and W. R. Crain, “Observation of single event upsets in analog microcircuits,” IEEE Trans. Nucl. Sci., vol. 40, no. 6, pp. 1838–1844, Dec. 1993.

[Ecof94] R. Ecoffet, S. Duzellier, P. Tastet, C. Aicardi, and M. Labrunee, “Observation of heavy ion induced transients in linear circuits,” Proc. IEEE NSREC Radiation Effects Data Workshop Record, pp. 72–77, 1994.

[Nara06] Narayanan. V, Xie. Y, "Reliability concerns in embedded system designs," Computer, vol.39, no.1, pp. 118- 120, Jan. 2006.

[Simp83] Simpson, J. A., “Elemental and Isotopic Composition of the Galactic Cosmic Rays” Annual review of nuclear and particle science, Vol. 33, 323-382, Dec 1983.

[Miles05] Miles. M., “The Magnetopause calculated by the unified field” *Journal of Geophysical Research*, vol. 76, no. 34, 2005.

[Stas88] Stassinopoulos, E.G., Raymond, J.P., "The space radiation environment for electronics," Proceedings of the IEEE, vol.76, no.11, pp.1423-1442, Nov 1988.

[Fred96] A. R. Frederickson, “red,” IEEE Trans. Nucl. Sci., vol. 43, no. 2, pp. 426-441, 1996.

[Hsieh81] Hsieh, C. M.; Murley, P. C.; O'Brien, R. R.; , "Dynamics of Charge Collection from Alpha-Particle Tracks in Integrated Circuits," Reliability Physics Symposium, 1981. 19th Annual, vol., no., pp.38-42, April 1981.

[Mavis02] Mavis, D.G.; Eaton, P.H.; , "Soft error rate mitigation techniques for modern microcircuits," Reliability Physics Symposium Proceedings, 2002. 40th Annual, vol., no., pp. 216- 225, 2002.

[Barn06] Barnaby, H. J., "Total-Ionizing-Dose Effects in Modern CMOS Technologies," Nuclear Science, IEEE Transactions on , vol.53, no.6, pp.3103-3121, Dec. 2006.

[NASA] http://radhome.gsfc.nasa.gov/radhome/see.htm

[Sagg05] Saggese, G.P.; Wang, N.J.; Kalbarczyk, Z.T.; Patel, S.J.; Iyer, R.K.; , "An experimental study of soft errors in microprocessors," Micro, IEEE , vol.25, no.6, pp. 30-39, Nov.-Dec. 2005.

[Heil89] S. J. Heileman, W. R. Eisenstadt, R. M. Fox, R. S. Wagner, N. Bordes, and J. M. Bradley, “CMOS VLSI single event transient characterization,” IEEE Trans. Nucl. Sci., vol. 36, no. 6, pp. 2287-2291, 1989.

[David09] David F. H., “Single-Event Upsets and Multiple-Bit Upsets on a 45 nm SOI SRAM” IEEE Trans. Nucl. Sci., vol. 56, no. 6, pp. 3499-3504, 2009.

[shin94] Shinichi Y., “**A** Radiation-Hardened 32-bit Microprocessor Based on the Commercial CMOS Process”, IEEE Trans. Nucl. Sci., vol. 41, no. 6, pp. 2481-2486, 1894.

[Muss96] Musseau, O.; Gardic, F.; Roche, P.; Corbiere, T.; Reed, R.A.; Buchner, S.; McDonald, P.; Melinger, J.; Tran, L.; Campbell, A.B.; , "Analysis of multiple bit upsets (MBU) in CMOS SRAM," Nuclear Science, IEEE Transactions on , vol.43, no.6, pp.2879-2888, Dec 1996.

[Baze97] M.P. Baze and S.P. Buchner, " Attenuation of Single Induces Pulses in CMOS combinatorial Logic," IEEE Trans. on Nuc. Sci., Vol 44 No. 6 , pp. 2217-2223, Dec., 1997.

[Hind11] Hindman, N.D.; Clark, L.T.; Patterson, D.W.; Holbert, K.E.; , "Fully Automated, Testable Design of Fine-Grained Triple Mode Redundant Logic," Nuclear Science, IEEE Transactions on , vol.58, no.6, pp.3046-3052, Dec. 2011

[Hind09] Hindman, N.D.; Pettit, D.E.; Patterson, D.W.; Nielsen, K.E.; Xiaoyin Yao; Holbert, K.E.; Clark, L.T.; , "High speed redundant self-correcting circuits for radiation hardened by design logic," Radiation and Its Effects on Components and Systems (RADECS), 2009 European Conference on , vol., no., pp.465-472, 14-18 Sept. 2009.

[Naseer06] R. Naseer and J. Draper, “DF-DICE: a Scalable Solution for Soft Error Tolerant Circuit Design,” *Proc. ISCAS*, pp. 3890-3893, May 2006.

[Weste04] N. H. E. Weste, D. Harris, CMOS VLSI Design: A Circuits and Systems Perspective, Boston, MA; Pearson Education, Inc, 2005.

[Chandra01] A. Chandrakasan, W. J. Bowhill, and F. Fox, Design of High-Performance Microprocessor Circuits, 2001, IEEE Press.

[Calin96] T. Calin, M. Nicolaidis, and R. Velazco, “Upset Hardened Memory Design for Submicron CMOS Technology,” IEEE Trans. Nuc. Sci., pp. 2874-2878, 43, 6, Dec. 1996.

[Zhang06] M. Zhang et al, “Sequential Element Design with Built-In Soft Error Resilience,” IEEE Trans. VLSI Sys., 14, 12, pp. 1368-1378, Dec. 2006.

[Naseer06] R. Naseer and J. Draper, “DF-DICE: a Scalable Solution for Soft Error Tolerant Circuit Design,” Proc. ISCAS, pp. 3890-3893, May 2006.

[Paul99] Paul E. Dodd, “Basic Mechanisms for Single-Event Effects,” Notes from IEEE Nuclear and Space Radiation Effects Conference Short Course, 1999.

[Heid08] D. F. Heidel, et al., “Low energy proton single-event-upset test results on 65 nm SOI SRAM,” IEEE Trans. Nucl. Sci., vol. 55, no. 6, pp. 3394- 3400, Dec. 2008.

[Sushil15] S. Kumar, S.Chellappa, and L. Clark, “Temporal Pulse-clocked Multibit Flip-flop Mitigating SET and SEU,” Proc. ISCAS, 2015.

[Sandeep15] S. Shambhulingaiah, C. Lieb, and L. Clark, “Circuit Simulation Based Validation of Flip-Flop Robustness to Multiple Node Charge Collection,” IEEE Trans. Nucl. Sci., vol. 62, no. 4, pp. 1577–1588, Aug. 2015.

[Aditya15] *A. Gujja*, S. Chellappa, C. Ramamurthy and L. T. Clark, “ Redundant Skewed Clocking of Pulse-clocked Latches for Low Power Soft Error Mitigation ” Radiation and Its Effects on Components and Systems (RADECS), 2015.

[Vittal97] A. Vittal and M. Marek-Sadowska, “Low-power Buffered Clock Tree Design,” *IEEE Trans. CAD*., vol.16, no.9, pp. 965-975, Sept. 1997.