Design and Analysis of Metastable-Hardened, High-Performance, Low-Power Flip-Flops

by

David Li

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

With rapid technology scaling, flip-flops are becoming more susceptible to metastability due to tighter timing budgets and the more prominent effects of process, temperature, and voltage variation that can result in frequent setup and hold time violations. This thesis presents a detailed methodology and analysis on the design of metastable-hardened, high-performance, and low-power flip-flops.

The design of metastable-hardened flip-flops is focused on optimizing the value of τ mainly due to its exponential relationship with the metastability window δ and the mean-time-between-failure (MTBF). Through small-signal modeling, τ is determined to be a function of the load capacitance and the transconductance in the cross-coupled inverter pair for a given flip-flop architecture. In most cases, the reduction of τ comes at the expense of increased delay and power. Hence, two new design metrics, the metastability-delay-product (MDP) and the metastability-power-delay-product (MPDP), are proposed to analyze the tradeoffs between delay, power and τ. Post-layout simulation results have shown that the proposed optimum MPDP design can reduce the metastability window δ by at least an order of magnitude depending on the value of the settling time and the flip-flop architecture.

In this work, we have proposed two new flip-flop designs: the pre-discharge flip-flop (PDFF) and the sense-amplifier-transmission-gate (SATG) based flip-flop. Both flip-flop architectures facilitate the usage in both single and dual-supply systems as reduced clock-swing flip-flop and level-converting flip-flop. With a cross-coupled inverter in the master-stage that increases the overall transconductance and a small load transistor associated with the critical node, the architecture of both the PDFF and the SATG is very attractive for the design of metastable-hardened, high-performance, and low-power flip-flops. The

amount of overhead in delay, power, and area is all less than 10% under the optimum MPDP design scheme when compared to the traditional optimum PDP design.

In designing for metastable-hardened and soft-error tolerant flip-flops, the main methodology is to improve the metastability performance in the master-stage while applying the soft-error tolerant cell in the slave-stage for protection against soft-error. The proposed flip-flops, PDFF-SE and SATG-SE, both utilize a cross-coupled inverter on the critical path in the master-stage and generate the required differential signals to facilitate the usage of the Quatro soft-error tolerant cell in the slave-stage.

Acknowledgements

First of, I would like to thank Dr.Manoj Sachdev for his great support, guidance, and mentoring as my research supervisor. His advice and support are greatly appreciated. I am also grateful to Professor David Nairn for his insights and suggestions on the metastability research project while proofreading a number of my papers. I would also like to thank Professor Hasan, Professor Anis, and Professor Martin for serving on my Ph.D committee. A special thanks goes out to Professor Gordon Roberts from the McGill University as a member of the external examiner. Thank you for all your positive and valuable comments and suggestions.

I would also like to specially thank Pierce Chuang for being a research collaborator and good friend in this research project, Phil Regier for solving all the computer problems, David Rennie for constantly having valuable discussions and helping me solving various Cadence issues, and everyone else in the CMOS Design and Reliability Group at the University of Waterloo for their support.

In addition, I would like to acknowledge the financial support from the National Sciences and Engineering Research Council of Canada (NSERC).

Also I would like to say a special thanks to Chen Hu, Phillip Woo, Jannie Mak, and Shawn Zhang for being my best friends at the University of Waterloo. Without their help and support, this thesis would not have been possible.

Finally, and most important of all, I would like to thank my family for their support and encouragement throughout my academic careers.

Dedication

To my Mom, Dad, grandparents, and my beloved wife.

Table of Contents

List of Tables

List of Figures

List of Symbols

α	Data Activity Factor
δ	Metastability Window
μ_n	Electron Mobility
μ_p	Hole Mobility
τ	Time Resolving Constant
C_{diff}	Diffusion Capacitance
C_g	Gate Capacitance
C_M	Miller Capacitance
C_Q	Capacitance Associated with the Critical Node
f_{CLK}	Clock Frequency
f_D	Data Frequency
g_m	Transconductance

L_{eff} Effective Transistor Channel Length

T_0 Asymptotic Width of the Metastability Window with No Settling Time

$t_{aperture}$ Flip-Flop Aperture Window

t_{C-Q} Flip-Flop Clock-to-Output Delay

t_{D-Q} Flip-Flop Data-to-Output Delay

t_{hold} Flip-Flop Hold Time

t_{setup} Flip-Flop Setup Time

T_{stage} Minimum Clock Period Requirement

t_s Settling Time

V_{DDH} Nominal Supply Voltage

V_{DDL} Reduced Supply Voltage

V_{DD} Supply Voltage

V_{th} Threshold Voltage

V_{tn} NMOS Threshold Voltage

V_{tp} PMOS Threshold Voltage

V_T Thermal Voltage

List of Abbreviations

ALU	Arithmetic Logic Unit
CCFF	Conditional-Capturing Flip-Flop
CMOS	Complementary Metal-Oxide-Semiconductor
CPN	Clocked-Pseudo-NMOS Flip-Flop
CREST	Circuit for Radiation Effects Self Test
CRFF	Contention-Reduced Flip-Flop
CSSA	Clock-Level Shifted Sense-Amplifier Flip-Flop
CVS	Cluster Voltage Scaling
DCDE	Digital-Controlled Delay Element
DFF	D Flip-Flop
DICE	Dual-Interlocked Cell
DSM	Deep Sub-micron
DVS	Dynamic Voltage Scaling
ECC	Error-Correction Code
EDP	Energy-Delay-Product

FUT	Flip-Flops Under Test
HLFF	Hybrid-Latch Flip-Flop
ITRS	International Technology Roadmap for Semiconductors
LANSCE	Los Alamos Neutron Science Center
LCFF	Level-Converting Flip-Flop
LCPDFF	Level-Converting Pre-Discharge Flip-Flop
LCSATG	Level-Converting Sense-Amplifier Transmission-Gate Flip-Flop
LV	Load Variation
mC^{2}MOS	modified-C^{2}MOS Flip-Flop
MDP	Metastability-Delay-Product
MGHK	Metal-Gate-High-K Technology
MPDP	Metastability-Power-Delay-Product
MTBF	Mean Time Between Failure
MVT	Mixed-V_{th} Design
NDKFF	NAND Keeper Flip-Flop

PDFF	Pre-Discharge Flip-Flop
PDFF-SE	Pre-Discharge Soft-Error Tolerant Flip-Flop
PDP	Power-Delay-Product
PowerPC	PowerPC Flip-Flop
PTM	Predictive Technology Model
PVT	Process, Voltage, Temperature
RBB	Reverse Body-Bias
RCSFF	Reduced Clock-Swing Flip-Flop
RCSPDFF	Reduced Clock-Swing Pre-Discharge Flip-Flop
RCSSATG	Reduced Clock-Swing Sense-Amplifier Transmission-Gate Flip-Flop
RFF	Razor Flip-Flop
RFID	Radio-Frequency Identification
SAFF	Sense-Amplifier Flip-Flop
SATG	Sense-Amplifier Transmission-Gate Flip-Flop
SATG-SE	Sense-Amplifier Transmission-Gate Soft-Error Tolerant Flip-Flop
SD	Standard Deviation
SDFF	Semi-Dynamic Flip-Flop
SER	Soft-Error Rate
SET	Single-Event Transient

SPFF	Self-Precharging Flip-Flop
SRAM	Static Random Access Memory
SSTC	Static Single Transistor Clocked Flip-Flop
Strained-Si	Strained-Silicon Technology
SVT	Single-V_{th} Design
TRIUMF	Tri-University Meson Facility
TSMC	Taiwan Semiconductor Manufacturing Company
TSPC	True Single-Phase Clocked Flip-Flop
TV	Transconductance Variation
VLSI	Very Large Scale Integration
VTC	Voltage Transfer Curve

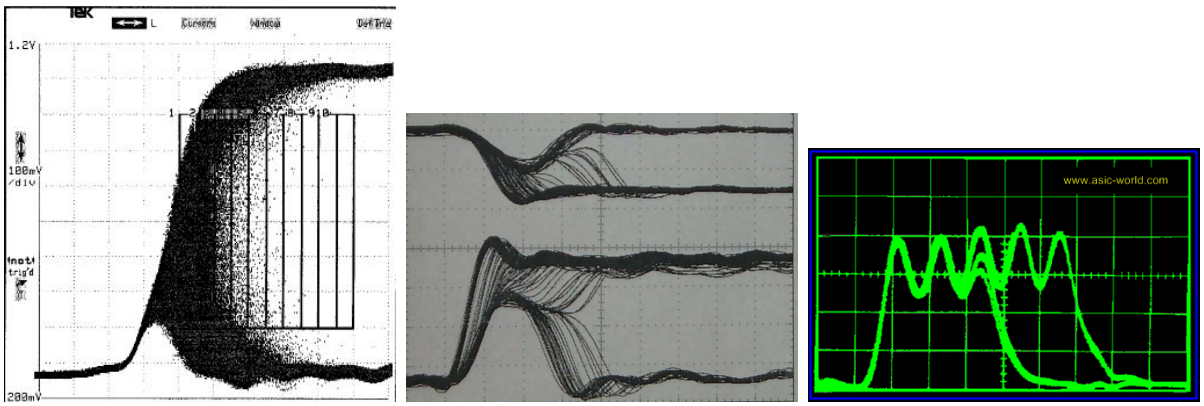
Chapter 1

Introduction

1.1 Design for Reliable, High-Performance, and Low-Power, Flip-Flops

Traditional flip-flop designs have mostly focused on balanced design tradeoff between delay and power, as indicated by the optimum power-delay-product (PDP) value. As the CMOS technology continues to scale, flip-flops are more susceptible to reliability issues such as metastability and soft-errors. While numerous studies have been performed on soft-error tolerant flip-flop designs, the design for metastable-hardened flip-flops has largely been missing in the literature. Metastability is a phenomenon where a bi-stable element enters an undesirable third state in which the output is stuck at an intermediate level between logic "0" and "1". In both synchronous and asynchronous systems, flip-flops are prone to metastability because its two inputs, the input data D and the CLK signal, potentially can make simultaneous transitions and violate the flip-flop setup and hold time constraints such that the resulting state would depend on the order of the input events. In either case,

metastability causes the flip-flop output to behave unpredictably (**Figure 1.1(a)**), taking an unbounded amount of time to settle to a stable state (**Figure 1.1(b)**), or even oscillating several times before settling to a stable state (**Figure 1.1(c)**). Flip-flop metastability can



(a) Random Data [4] (b) Unbounded Settling Time [5] (c) Oscillation [6]

Figure 1.1: Illustration of Metastability

cause corruption of data if the state is not stable before another circuit uses its value. As such, the ability of the flip-flops to resolve from the metastable region is extremely important to maintain a reliable operation by avoiding metastable output that may *(i)* prevent the correct functionality of the handshaking protocol in asynchronous domains, or *(ii)* propagating from stage to stage in the pipeline systems and ultimately results in system failures. As described by the famous Moore's Law, the downscaling of minimum dimensions enables the integration of an increasing number of transistors on a single chip. In fact, Moore predicted that the microprocessor unit (MPU) performance will double every 1.5 to 2 years [7]. The continuous push for higher clock rates and higher performance has led microprocessor designers in recent years to build super-pipelined machines with multiple functional units that can execute operations concurrently. The tighter timing budgets

along with the impact of process, voltage, and temperature (PVT) variations all make the flip-flops more susceptible to metastable output states. Therefore, metastability is becoming an important design consideration for flip-flop designs.

In order for the pipeline system to function correctly, **Equation (1.1)** must be satisfied where T_{stage} represents the minimum clock period, t_{C-Q} and t_{setup} are the delay and setup time of the flip-flop respectively, and t_{logic} is the delay of the logic inserted between the flip-flops.

$$T_{stage} = t_{C-Q} + t_{setup} + t_{logic} \tag{1.1}$$

The aforementioned high clock rates in high-performance microprocessors are often achieved with fine granularity pipelining, for which there are relatively few levels of logic per pipeline stage. One direct consequence of this design trend is that the pipeline overhead, such as the latency of the flip-flop (i.e. t_{C-Q} and t_{setup}) is becoming more significant. Therefore, high-performance flip-flop designs are essential to sustain high latency in deep pipelined systems.

While the performance constraint is an important design consideration in pipelined systems, power consumption has also become an equally critical constraint in high-performance designs. Recent reported power consumption breakups have shown that the clock system consumes anywhere between 20%-50% of the total chip power. This ratio is expected to grow further due to the constant frequency increase trends and the reduction of number of logic gates per pipeline stage. Because the clock systems drive millions of flip-flops in microprocessors, considerable power savings can be achieved on the clock system with low-power flip-flop designs. Among all the techniques in minimizing power consumption, reduction in supply voltage (V_{DD}) is the most effective method due to the quadratic relationship shown in **Equation (1.2)**

$$P = \alpha C V_{swing} V_{DD} f \tag{1.2}$$

where α is the data activity factor, C represents the load capacitance, V_{DD} is the supply voltage, V_{swing} is the value of the signal, and f is the switching frequency. Although direct voltage scaling results in significant performance degradation, a more common approach is to use a dual-supply technique to minimize the performance degradation while achieving reduction in power dissipation. Due to the 100% transition probability, significant power consumption savings can be achieved on the clock system by simply reducing the swing on the clock signal to a lower voltage (V_{DDL}). As such, reduced-clock swing flip-flops (RCSFF) [8][9] have been used to implement such system. Other dual-supply systems including the clustered voltage scaling (CVS) scheme [10][11][12] where lower supply voltage (V_{DDL}) is used in non-critical paths while placing the nominal supply voltage (V_{DDH}) on the critical paths. In such design, level-converting flip-flops (LCFF) are placed at the boundary between the V_{DDL} and the V_{DDH} domains to provide full swing input to the V_{DDH} domain.

1.2 Impact of Technology Scaling

The first CMOS scaling theory [13] is based on a model formulated by Robert Dennard. This theory states that the characteristic of an MOSFET device can be maintained and the basic operational characteristics can be preserved if the critical parameters of a device are scaled by a dimensionless factor S. In general, there are two types of scaling: *constant field scaling* and *constant voltage scaling*. In constant field scaling, all device dimensions, including channel length L, width W, and oxide thickness t_{ox} are reduced by a factor of $1/S$ while the supply voltage V_{DD} is also reduced by the same factor. Since both dimension and voltage are scaled equally, the electric field remains constant. In constant voltage scaling, the electric field is increased in devices because the dimensions are shrunk by 1/S

but the voltage remains unaffected. As the CMOS technology continues to scale into the deep-submicron (DSM) regime, the effect of velocity saturation was significant enough that decreasing feature size no longer improved the device current. This couples with the risks of device breakdown at high field has made constant field scaling a popular choice for modern CMOS technologies. **Table 1.1** summarizes the effect of both constant field and constant voltage scaling.

Table 1.1: Effects of Constant Field and Constant Voltage Scaling

	Relation	Constant Field Scaling	Constant Voltage Scaling
L		1/S	1/S
W		1/S	1/S
t_{ox}		1/S	1/S
Electric Field		1	S
V_{DD}, V_{th}		1/S	1
Current		1/S	1
Gate Capacitance	$\frac{\epsilon_{ox}}{t_{ox}}WL$	1/S	1/S
Intrinsic Delay	$\frac{C_L V_{DD}}{I_{avg}}$	1/S	1/S
Intrinsic Power	$\frac{C_L V_{DD}^2}{t_{delay}}$	$1/S^2$	S
Intrinsic Energy	$C_L V_{DD}^2$	$1/S^3$	S

The benefits of CMOS scaling is reflected in the reductions of transistor parasitic capacitance, lower gate level average power, switching energy, and most importantly, improved propagation delay. If a scaling factor of 0.7 is considered to shrink the feature size from one CMOS generation to the next, based on the expressions shown in **Table 1.1**, the capacitance, average power, energy, and propagation delay should all be decreased by

approximately 30%, 50%, 65%, and 30% respectively.

The energy and delay improvements resulted from CMOS scaling has led to a rapid increasing in frequencies and levels of integration for microprocessors, as indicated by the data forecasted by the International Technology Roadmap for Semiconductors (ITRS) shown in **Table 1.2**. As seen from the table, it is expected that by the year 2021, the

Table 1.2: 2010 ITRS Forecasts [1]

Year	2011	2013	2015	2017	2019	2021
Feature size (nm)	28	23	18	14.2	11.3	8.9
Millions of Transistors/Chip	3092	3092	6184	12368	12368	24736
On-Chip Clock Rate (GHz)	6.329	7.344	8.522	9.889	11.475	14.343
Supply Voltage (V)	0.93	0.87	0.81	0.76	0.71	0.66

CMOS technology will reach the 8.9nm node with an on-die transistor count of 24736 millions and an on-chip clock frequency of 14.343GHz.

1.3 Motivation

While metastability has been present in digital systems for many years, the amount of research is less prevalent when compared to other areas. This is evident in the number of publications relating to metastability in the last 50 years or so (**Figure 1.2**). Past works on metastability have mostly concentrated on theoretical modeling, experimental measurements and the effects of various circuit parameters for a given latch or flip-flop. Works from two decades ago, [14][15][16][17][18], have formed the foundation for metastability analysis

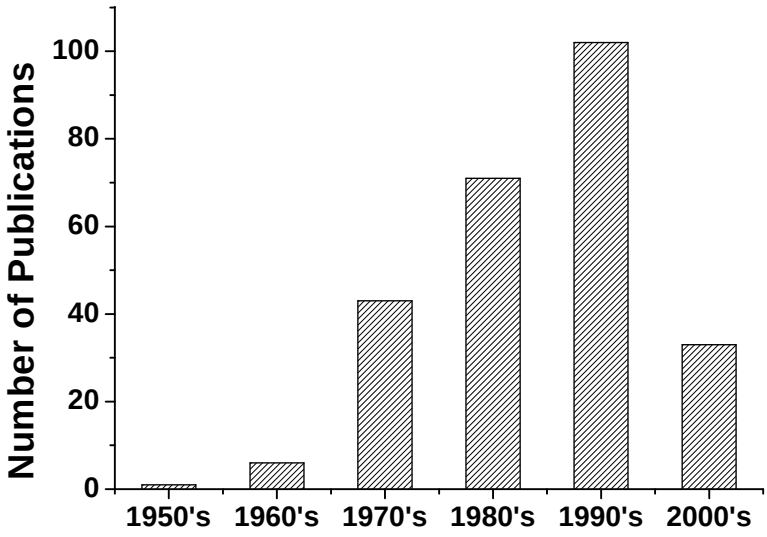


Figure 1.2: Number of Publications on Metastability [2]

by solving small-signal equations for the time-resolving constant τ in the cross-coupled inverter pair. The work presented in [3][19][20] describes the challenges and methodologies involved in on-chip metastability measurement of a particular synchronizer, jamb-latch flip-flop. Different techniques have been proposed in [21][22] to improve metastability in the jamb-latch flip-flop under process variation and in sub-threshold operations. In [23], metastability parameters are extracted from simulation results along with delay and power analysis for various transmission-gate based flip-flops. In the past, metastability typically exists when flip-flops are synchronizing two unrelated signals in asynchronous systems. As CMOS technology continues to scale, tighter timing budgets due to higher clock rates and smaller intrinsic gate delays along with PVT variations have all contributed to the increasing susceptibility of the flip-flops to enter metastability in the synchronous systems. As a result, the number of research work relating to error-resilient design and metastability-correction circuits has shown a steady increase in the past few years [24][25][26][27][28][29].

Overall, the potential to explore metastability-related research topics is rapidly growing.

With various flip-flop architectures proposed in today's VLSI systems to achieve different design objectives, a detailed analysis and design optimization on the flip-flop metastability has largely been missing. While the gate delay may be reduced by a factor of 0.7 for every technology generation, the flip-flop metastability performance may not necessarily follow the same scaling trend, as will be shown later in this thesis. The scaling of supply voltage and threshold voltage V_{th} along with other device parameters such as hole/electron mobility and parasitic capacitances all have a direct impact on the ability of the flip-flops to resolve quickly from the metastable region. Hence, appropriate transistor sizing and novel architectures have become important considerations for metastable-hardened flip-flops designs. In this thesis work, we will provide a detailed methodology and analysis on designing metastable-hardened, high-performance, and low-power flip-flops. We will demonstrate how metastability performance can be improved on previously proposed flip-flop architectures while maintaining an appropriate tradeoff in delay and power. We will also propose two new flip-flop architectures that are suitable for metastable-hardened, high-performance, and low-power design in both the single and the dual-supply systems. In addition, the proposed flip-flops are also able to include the soft-error tolerant feature in the design. Overall, this thesis has made contributions in the following areas.

- Propose two novel flip-flop designs with architectures suitable for metastable-hardened, high-performance, and low-power in both the single and the dual-supply systems.
- Develop a detailed methodology in designing metastable-hardened, high-performance, and low-power flip-flops.
  - Provide qualitative analysis on the metastable behavior for a given flip-flop architecture.

  - Develop transistor sizing methodology to vary the value of the time-resolving constant τ.
  - Apply small-signal modeling on different flip-flop architectures to estimate τ.
  - Propose two new design metrics in analyzing the design tradeoffs between metastability, performance, and power.

- Propose a mixed-V_{th} technique that can dramatically improve flip-flop metastability in the sub-threshold region.
- Studies the flip-flop metastability behavior for CMOS technologies below the 65nm regime using Predictive Technology Modeling.
- Analyze detailed methodology in designing metastable-hardened and soft-error tolerant flip-flops.

1.4 Thesis Overview

This thesis is organized in the following manner. Chapter 2 provides the basic background information on flip-flop metastability including characterization, modeling, past mitigation techniques, simulation techniques, as well as the impact of process, voltage, and temperature (PVT) variation. Chapter 3 proposes two new flip-flop designs as well as reviewing various flip-flop architectures including high-performance and low-power designs along with reduced-clock swing flip-flops (RCSFF) and level-converting flip-flops (LCFF). Chapter 4 offers detailed analysis and design methodologies on metastable-hardened, high-performance, and low-power flip-flops. Chapter 5 analyzes the design methodologies behind metastable-hardened, soft-error tolerant flip-flops. Finally, concluding remarks and future work will be given in Chapter 6.

Chapter 2

Background on Metastability

In this chapter, we present a thorough and detailed background information on flip-flop metastability. The basic timing parameters of the flip-flops will be described in detail. An introduction on both the synchronous and the asynchronous systems is provided to illustrate the respective usage of the flip-flops. Metastability is discussed in terms of its origin, qualitative and quantitative characteristics, and small-signal modeling. Past metastability mitigation techniques in both circuit and system levels are also presented. Finally, the impact of process, voltage, and temperature variation on the value of τ is also described in this chapter.

2.1 Basic Flip-Flop Characteristics

The general timing parameters of a flip-flop (**Figure 2.1**) are provided by [30] and described below.

- C-Q Delay (t_{C-Q}): Propagation delay from the CLK to the output Q, assuming that

the input data D has been set early enough relative to the leading edge of the CLK.

- D-Q Delay (t_{D-Q}): Propagation delay from the input data to the output Q, assuming the CLK has been turned on early enough relative to the transition in D.

- Setup Time (t_{setup}): The minimum time between a transition in D and the sampling edge of the CLK such that, even under worst case conditions, the Q will be guaranteed to change so as to become equal to the new D value.

- Hold Time (t_{hold}): The minimum time that the D must be held constant after the sampling edge of the CLK so that, even under worst case conditions, and assuming that the most recent transition in D occurred no later than t_{setup} prior to the sampling edge of CLK, the Q output will remain stable after the end of the CLK pulse.

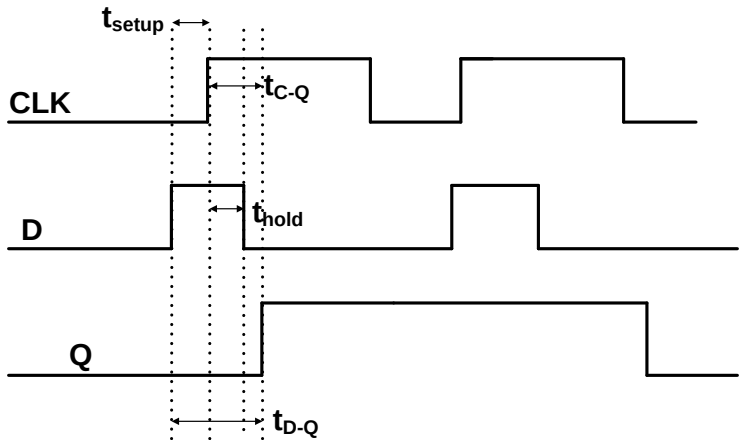


Figure 2.1: Timing Parameters of a Typical Flip-Flop

Figure 2.2 illustrates the timing characteristic curve of a flip-flop. In general, the curve can be divided into three regions: stable, quasi-metastable region, and metastable [31]. In the stable region, the C-Q delay of the flip-flop is constant regardless of the data

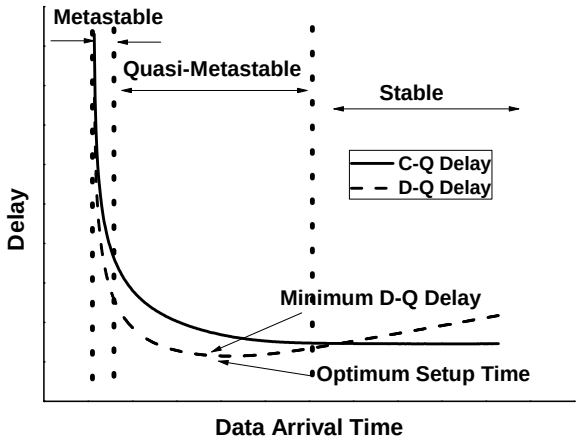


Figure 2.2: Flip-Flop Delay Characteristic Curve

arrival time (t_{D-C}). As t_{D-C} decreases, the C-Q delay starts to rise monotonously in the quasi-metastable region but the D-Q delay reaches its minimum value. We refer the D-C delay at that point as the optimum setup time, which presents the limit beyond which the performance of the flip-flop is degraded and the reliability is endangered. The third region is the region of metastability where the C-Q delay is much larger than the normal delay and increases exponentially. More details on the metastable region of the flip-flop curve will be provided in the next few sections.

In high-performance systems, the amount of cycle time taken out by the flip-flop consists of the sum of setup time (t_{setup}) and clock-output (t_{C-Q}) delay. As a result, the true flip-flop delay, given by **Equation (2.1)** should be measured as the time between the latest point of data arrival and the corresponding output transition such as t_{D-Q} [32].

$$t_{D-Q}|_{min} = t_{C-Q}|_{min_{D-Q}} + t_{setup}|_{min_{D-Q}} \tag{2.1}$$

From the high-performance and reliability point of view, it is also desirable to maintain a smaller aperture window ($t_{aperture}$) value, which is simply the sum of the minimum setup

and hold time requirement, as shown in **Equation (2.2)**. Intuitively, $t_{aperture}$ is the period of time around the clock edge during which the data must not transition if the flip-flop is to produce the correct and stable output.

$$\begin{aligned} t_{aperture0-1} &= t_{setup0-1} + t_{hold1-0} \\ t_{aperture1-0} &= t_{setup1-0} + t_{hold0-1} \end{aligned} \tag{2.2}$$

2.2 Introduction to Synchronous System

In digital logic design, the flow of data in synchronous systems is synchronized with the clock signal such that the data can be sampled directly without any uncertainty. The concept of a positive edge-triggered synchronous system is shown in **Figure 2.3**.

For the system shown in the figure, all the data is sampled at the rising edge of the clock signal for the register. Here, the data signal D_1 is sampled by flip-flop $FF1$ to yield the output signal Out_1. In turn, Out_1 passes through the combinational logic block and produces D_2 after a certain propagation delay. Finally in synchronization with the clock, Out_2 becomes valid after D_2 is sampled by flip-flop $FF2$. The worst propagation delay in the combinational logic block, or the longest time it would take for D_2 to become valid, places an upper bound on the performance of the synchronous system. The requirement for the minimum clock period is discussed in more detail in the following paragraph.

There are three important flip-flop-related timing parameters in any synchronous system: (i) propagation delay of the flip-flop (t_{C-Q}), (ii) setup and (iii) hold time associated with a flip-flop [30]. The other timing parameter that must be considered in a synchronous system includes the maximum delay of the combinational logic (t_{logic}). Under the ideal conditions, the phase of the clock signal at various locations of the system should be exactly identical where the clocks at $FF1$ and $FF2$ shown in **Figure 2.3** should have the same

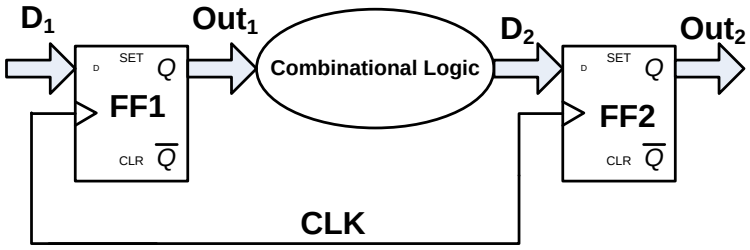


Figure 2.3: Block Diagram of a Synchronous System

period and transition at the exact same time. Under such ideal assumption, the minimum clock period must be long enough for the data to propagate through the flip-flops and logic and be setup for the destination flip-flop before the next rising edge of the clock. This requirement is shown in **Equation (2.3)**.

$$T > t_{C-Q} + t_{logic} + t_{setup} \tag{2.3}$$

Similarly, **Equation (2.4)** shows that the hold time of the destination register must be shorter than the minimum propagation delay through the logic network and the flip-flop in order to avoid the race condition.

$$t_{hold} < t_{logic} + t_{C-Q} \tag{2.4}$$

2.3 Introduction to Asynchronous System

While the synchronous system described in the previous section has some clear advantages such as a structured and deterministic approach as well as robust and easy design, it still presents several disadvantages as stated below.

- Presence of clock skew and jitter, which complicates and restricts certain physical and logical constraints.

- Significant power consumption in the clock network.
- System performance is limited by the slowest stage in the pipeline.

One way to avoid these problems is to eliminate the global CLK signal and adopt an asynchronous design where the logical ordering of the events is dictated by the structure of the transistor network and the relative delays or the signals. In asynchronous designs, careful timing analysis of the network must be performed to ensure a correct circuit operation that avoids all potential race conditions under any operation condition and input sequence.

An example of the asynchronous system is illustrated in **Figure 2.4** where System A is controlled by $CLKA$ and needs to transmit data to System B controlled by $CLKB$. In this system, System A must guarantee that the data is stable when the flip-flops in System B sample the data. It indicates when new data is valid by using a request signal (Req) so System B receives the data exactly once. System B replies with an acknowledge signal (Ack) when it has sampled the data so System A can put new data on the bus. The request and acknowledge signals are called handshaking lines, which can be a two-phase or four-phase protocols. The four-phase handshake is level-sensitive and the two-phase handshake is edge-triggered. In the two-phase handshake, System A places data on the bus, it then changes Req to indicate that the data is valid. System B samples the data when it detects change in the level of Req and toggles Ack to indicate the data has been captured. In the two-phase handshaking system shown in **Figure 2.4**, $CLKA$ and $CLKB$ operate independently at unrelated frequencies. Each system contains a synchronizer, a level-to-pulse converter, and a pulse-to-level converter. System A asserts $ReqA$ for one cycle when $DataA$ is ready, and this will be referred to as a pulse. The XOR gate and the flip-flop form a pulse-to-level converter that toggles the level of Req. This level is synchronized to

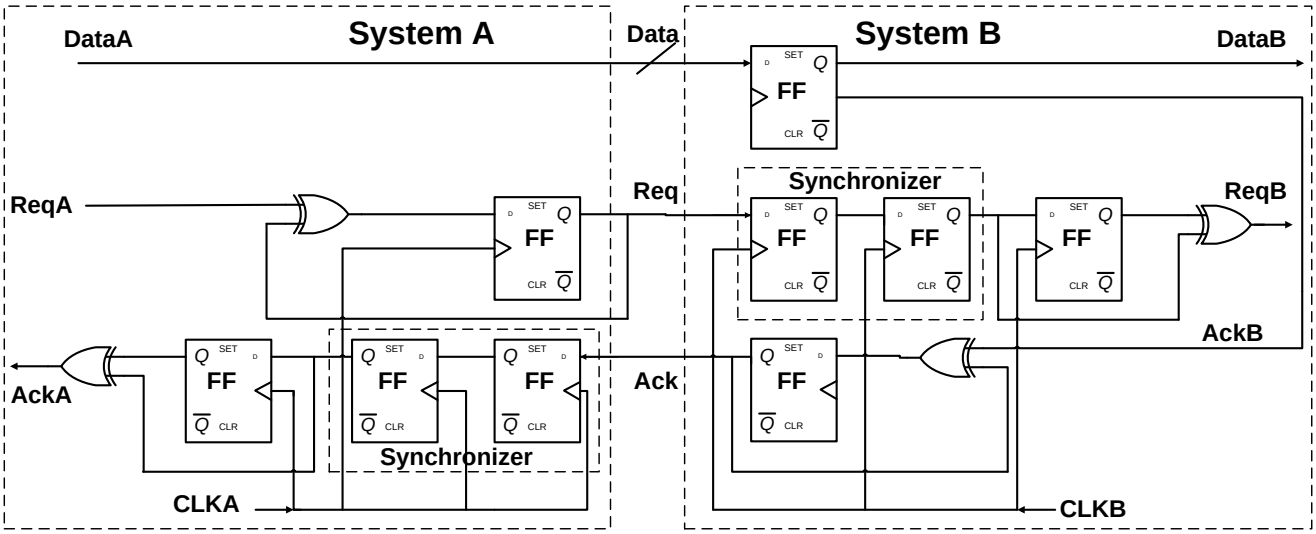


Figure 2.4: Block Diagram of an Asynchronous System

$CLKB$. When an edge is detected, the level-to-pulse converter produces a pulse on $ReqB$. This pulse in turn toggles Ack. The acknowledge level is synchronized to $CLKA$ and converted back to pulse on $AckA$. The usage of the synchronizers add significant latency such that the throughput of asynchronous communication is much lower than that of the synchronous communication.

2.4 What is Metastability

Metastability is a phenomenon where a bi-stable element enters an undesirable third state in which the output is at an intermediate level between logic "0" and "1". Flip-flops, in particular, enter the metastable region when they violate the setup or hold time constraints when the input data D makes a transition within $t_{aperture}$ (**Figure 2.5**).

Because a typical master-slave flip-flop composes of two identical latches, a simplified

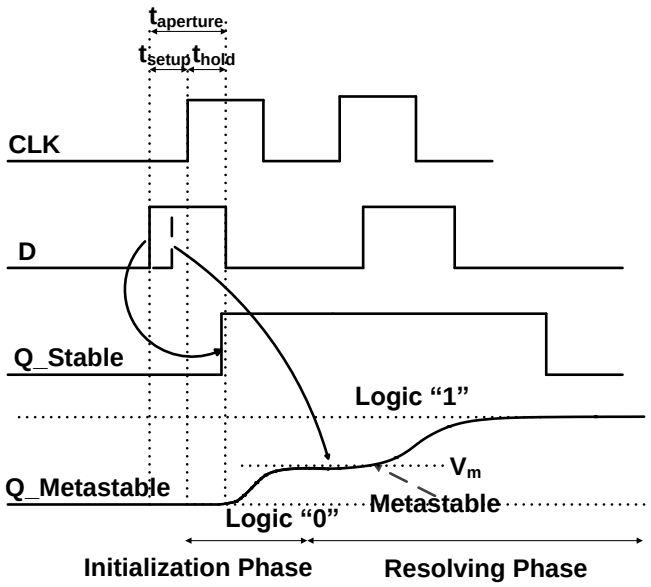
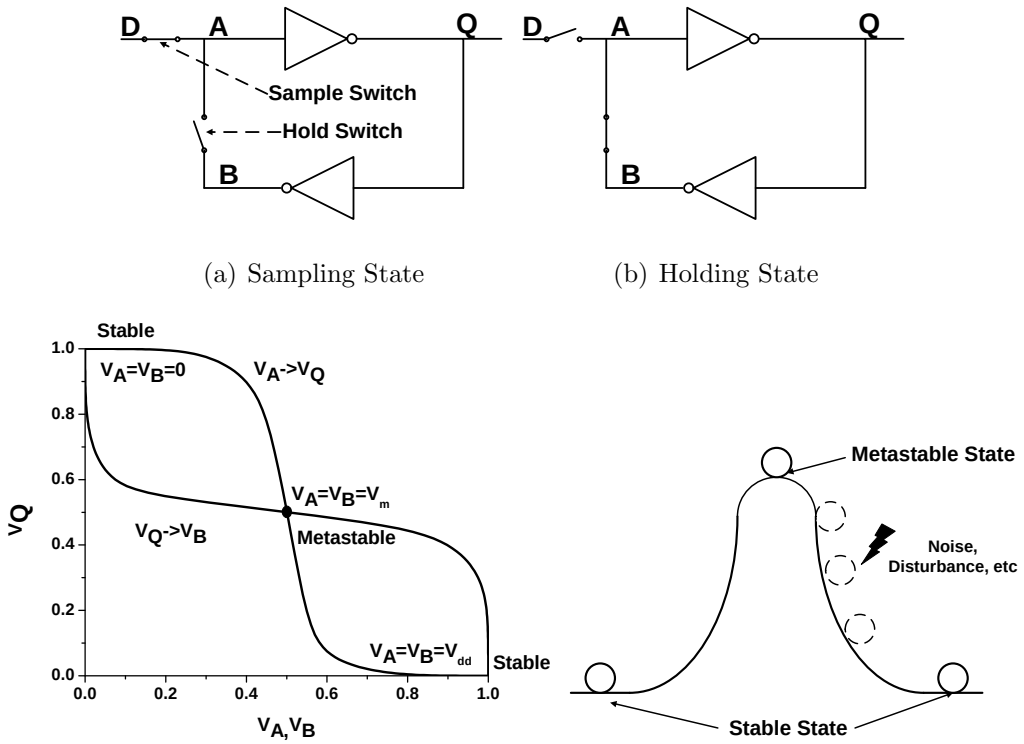


Figure 2.5: Illustration of Metastability using Timing Waveforms

model for a static latch is often used to illustrate the theories behind metastability (**Figure 2.6**). The switches shown in the figure are typically implemented using CLK-controlled transmission gates in practice. When the latch is transparent, the sample switch is closed and the hold switch opens **Figure 2.6(a)**. When the latch becomes opaque, the sample switch opens and the hold switch closes **(Figure 2.6(b))**. The resulting DC transfer characteristic curve of the two inverters is plotted in **Figure 2.6(c)**. When the latch is opaque, V_A=V_B and maintains a stable state of either logic "0" or logic "1". During the voltage transfer, both V_A and V_B can reach the metastable state of V_m, which is an illegal state some where between logic "0" and logic "1". This point is called metastable because the voltages are self-consistent and can remain there indefinitely. However, any noise or other disturbance will cause V_A or V_B to switch to one of the stable states. The idea of metastability can also be illustrated by the location of a ball on a rolling hill (**Figure**



(a) Sampling State (b) Holding State

(c) Voltage Transfer Characteristic Curve (d) Analogy of Metastability

Figure 2.6: Metastability in a Static Latch

2.6(d)). The stable states of the latch is equivalent to the ball being at the bottom of the hill where any disturbance cannot easily alter the stability of the current state. At the top of the hill, however, the ball is at a very fragile state where it can theoretically stay there for an indefinite amount of time. This is the metastable state where the slightest air current would eventually cause the ball to roll down to either side of the hill and reach a stable state. Similarly in a latch, any thermal and induced noise will cause it to move from the metastable state into either the logic "0" or logic "1" state.

In order to achieve high-performance datapaths, flip-flops in the synchronous pipelined

systems may require to operate close at the minimum D-Q delay in order to satisfy the timing constraints. In such case, the ability of the flip-flops to resolve from the metastable region is extremely important to maintain a reliable operation by avoiding metastable output that may ultimately results in system failures. As the integration complexity and clock frequency are rapidly increasing under a tight timing budget, the presence of process, voltage and temperature (PVT) variations cause the flip-flops to become more susceptible to produce metastable outputs when setup and/or hold time violations occur during the intra-domain data transfer [26]. The emergence of various power management techniques such as multiple voltage domains, reduced clock swing, and dynamic voltage scaling (DVS) further aggravates the metastability problem during the data transfer between different domains. A few examples of such scenario are listed below and illustrated in **Figure 2.7**.

- A voltage droop on the combinational logic may prolong the t_{logic} value and cause setup time violation in subsequent flip-flop.
- The impact of clock skew results in race condition and subsequently violation of the flip-flop hold time.
- Presence of glitches during data transfer from the V_{DDL} domain to the V_{DDH} domain.

In the asynchronous system shown in **Figure 2.4**, the signals interfacing the two domains are sampled by synchronizers controlled by the CLK signal. If System A and System B are operating at different frequencies or at the same frequency but with different phases, synchronizers can also produce metastable outputs if the asynchronous and unrelated handshaking signals of Ack and Req make transitions that violate the setup and hold time constraints.

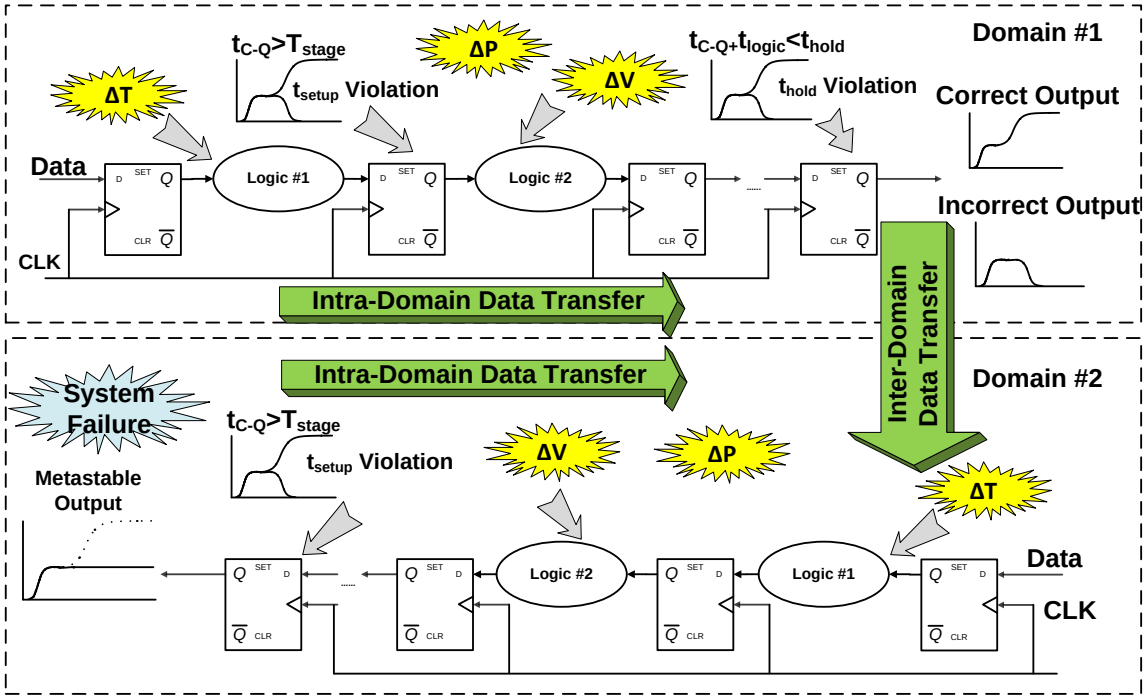


Figure 2.7: Metastability in Synchronous Pipelined Systems

2.5 Characterization of Metastability

Past studies have shown that the flip-flop delay in the metastable region is exponential in nature where two parameters (τ and T_0) can be extracted from simulation to model and analyze the delay behavior in the metastable region (**Figure 2.8**) [17][33]. A common metric used to quantify metastability is the metastability window δ, given in **Equation (2.5)**,

$$\delta = T_0 e^{-t_s/\tau} \tag{2.5}$$

where T_0 is the asymptotic width of the metastability window with no settling time, and τ is the resolution time constant that represents the inverse of the gain-bandwidth product

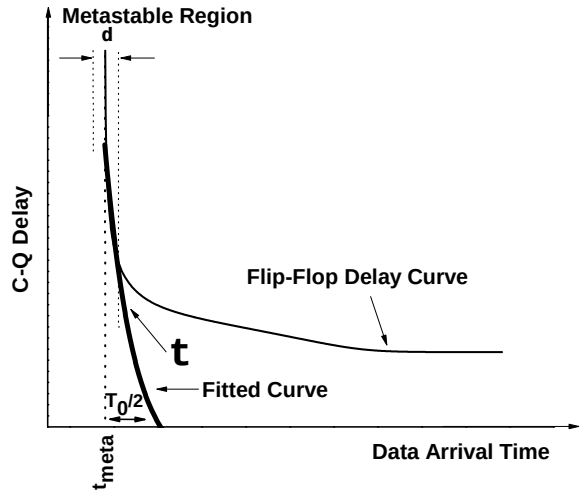


Figure 2.8: Extraction of Flip-Flop Metastability Parameters

of the feedback element in the flip-flop. Intuitively, one can think T_0 is the normalized time aperture when metastability can occur. Hence T_0 is closely related to the aperture window $t_{aperture}$. τ, on the other hand, determines how long the metastable state will last if the device enters such state. In general, metastability window δ can be defined as the time period where data transitions cannot be resolved within a given settling time t_s, and as such it should be kept as small as possible. Since τ is exponentially proportional to δ, a slight improvement in τ can cause a significant reduction in δ. For this reason, a majority of the design effort is focused on minimizing the value of τ.

If the data transitions at a frequency of f_D with respect to the clock which has a frequency of f_{CLK}, the mean-time-between-failures (MTBF) is then given in **Equation (2.6)**.

$$MTBF = \frac{1}{f_D f_{CLK} T_0 e^{-t_s/\tau}} \tag{2.6}$$

In general, the MTBF indicates the average time interval between two successive failures in a system. Hence, a higher MTBF value increases the overall reliability of the system.

Figure 2.9 illustrates the MTBF of three different flip-flop designs assuming the following parameters: f_D=1GHz, f_{CLK}=2GHz, t_s=400ps. Among them, FF#1 has the lowest τ

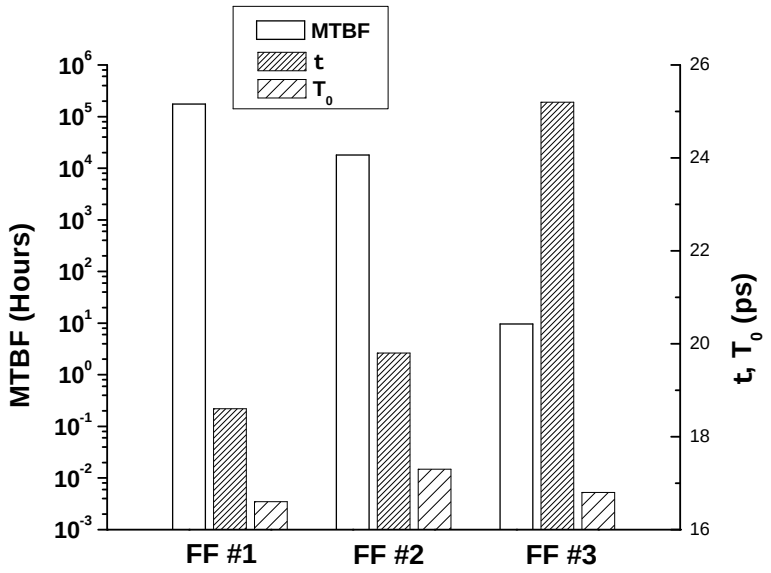


Figure 2.9: Comparison of τ, T_0, and MTBF for Different Flip-Flop Designs

value, and its MTBF is approximately 20 years. On the other hand, the τ of FF#3 is the highest and thus only results in an MTBF of 9.6 hours. Thus, the impact of the time-resolving constant τ on MTBF can be significant due to the exponential relationship. Because MTBF depends in the data and clock frequency of a system, the metastability window δ, given in **Equation (2.5)**, is often used as the main parameter in discussion of flip-flop metastability. **Figure 2.10** illustrates the values of δ as a function of the settling time t_s for each of the flip-flop shown in **Figure 2.9**.

From Equation (2.6), it is evident that the value of MTBF also has an exponential relationship with the settling time t_s. In synchronous pipelined systems, t_s, given by **Equation (2.7)**, is simply the amount of slack time available in a given pipeline stage for

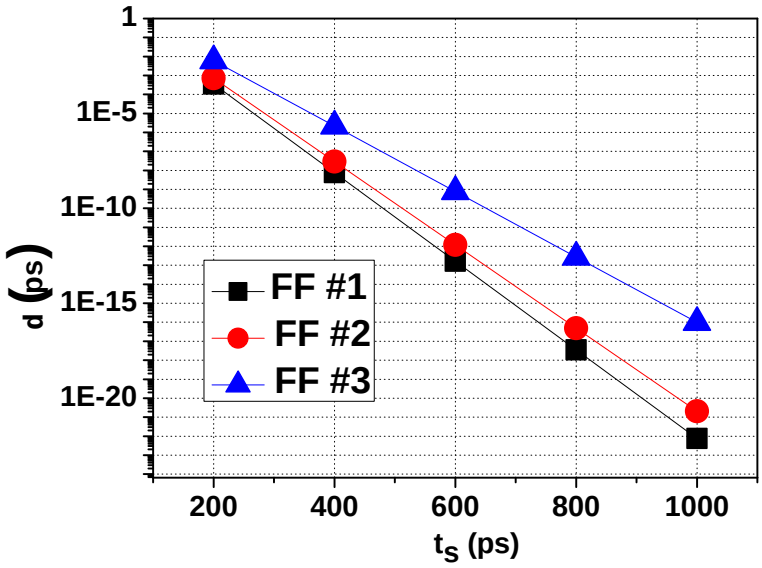


Figure 2.10: Comparison of Metastability Window δ as a Function of the Settling Time t_s

the output to settle to a stable state.

$$t_s = T_{CLK} - t_{C-Q} - t_{setup} - t_{logic} \tag{2.7}$$

t_s may vary from stage to stage depending on the value of the propagation delay in the combination logic (t_{logic}) for a particular stage. For a given flip-flop with T_0=16.6ps and τ=18.6ps, **Figure 2.11** plots the MTBF for various t_s values as a function of three different clock frequencies and assuming $f_D = 0.5 f_{CLK}$. From the data shown, it is evident that the exponential relationship of t_s also has an significant impact on MTBF. For a given clock frequency, the MTBF increases exponentially as t_s increases. As the clock frequency increases, the MTBF decreases as a result of a smaller settling time t_s.

From the analysis provided above, to increase the reliability of a system with higher MTBF values, the designers have the choice of either adjusting the parameters of the overall system or designing for metastable-hardened flip-flops. In the first approach, system

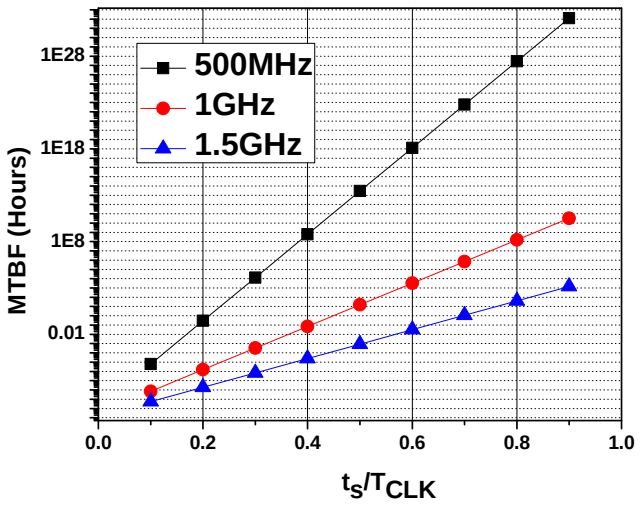


Figure 2.11: Comparison of MTBF as a Function of the Clock Frequency and Settling Time t_s

performance can be decreased by running at slower data (f_D) and clock frequencies (f_{CLK}) along with a higher settling time t_s. However, the overall system performance is often not compromisable, and hence a better approach is to design metastable-hardened flip-flops with smaller T_0 and τ values.

2.6 Metastability Modeling

From **Equation (2.6)**, it is clear that τ has the greatest effect on the MTBF due to the exponential term. A small τ value results in fast flip-flop resolution time from the metastable region and thus increases the MTBF [34]. To model and examine the time resolving constant τ, a simplified CMOS latch composed of cross-coupled inverter pair (**Figure 2.12**) is used. The voltage-transfer curve (VTC) of the back-to-back inverter is

also shown in **Figure 2.12**. During the normal operation of the latch, the outputs of the loop (V_Q and V_Q') will reach a stable state of either logic "0" or logic "1". In the metastable condition, however, the outputs are at a voltage level of V_m, which is an intermediate level somewhere between logic "0" and "1". At V_m, the inverters act as amplifiers with positive feedback causing the loop to eventually settle to one of the two stable logic values.

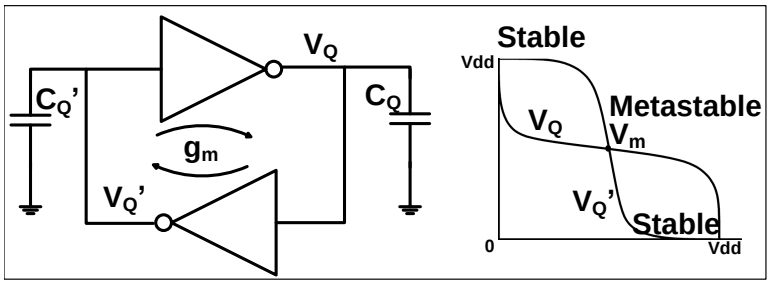


Figure 2.12: Metastability Modeling using Cross-Coupled Inverter

A small signal model (**Figure 2.13**) can be used to perform transient analysis of this situation given the fact that *(i)* the DC bias point can be calculated as the voltage at which VTC intersects $V_Q = V_Q' = V_m$, and *(ii)* VTC behaves approximately linearly around the bias point V_m. g_m represents the total transconductance contribution from both the PMOS and NMOS transistors in the inverter pair. Similarly, R and C_Q are the respective lumped resistance and capacitance values from various sources. The set of equations describing this system can be written in the form of **Equation (2.8)**.

$$\begin{aligned} g_m V_Q + 2C_M \frac{d(V_Q' - V_Q)}{dt} + \frac{V_Q'}{R} + C_Q \frac{dV_Q'}{dt} = 0 \\ g_m V_Q' + 2C_M \frac{d(V_Q - V_Q')}{dt} + \frac{V_Q}{R} + C_Q \frac{dV_Q}{dt} = 0 \end{aligned} \tag{2.8}$$

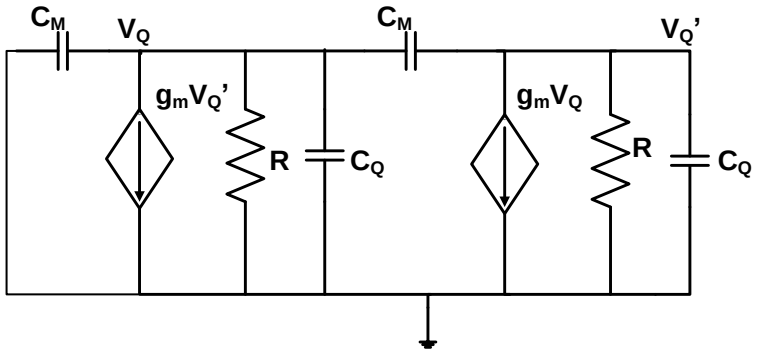


Figure 2.13: Small Signal Modeling for τ

If the solution of V_Q and V_Q' is assumed to be in the exponential form, then

$$V_Q = V_Q(0)e^{\frac{t}{\tau}} \tag{2.9}$$

where

$$\tau = \frac{C_Q + 4C_M}{g_m - \frac{1}{R}} \approx \frac{C_Q + 4C_M}{g_m} \tag{2.10}$$

if we assume $g_m R >> 1$. Typically, C_Q includes the gate and the diffusion capacitances of the transistors while C_M is the Miller capacitance, which is simply the coupling capacitance between the gate and the source/drain terminal of a MOSFET device. **Equation (2.10)** [18][35][36] provides a quick first order calculation of τ based on the value of capacitance and the transconductance.

2.7 Techniques for Metastability Mitigation

Like other reliability issues, metastability is a phenomenon that cannot be completely eliminated but can be mitigated using various design techniques both on the system and the circuit level.

2.7.1 Synchronization Techniques

During asynchronous data transfers, the most common way to tolerate metastability is to cascade one or more successive synchronizing flip-flops in series to the synchronizer. **Figure 2.14** illustrates an example of a one-stage, two-stage, and three-stage synchronizer. This

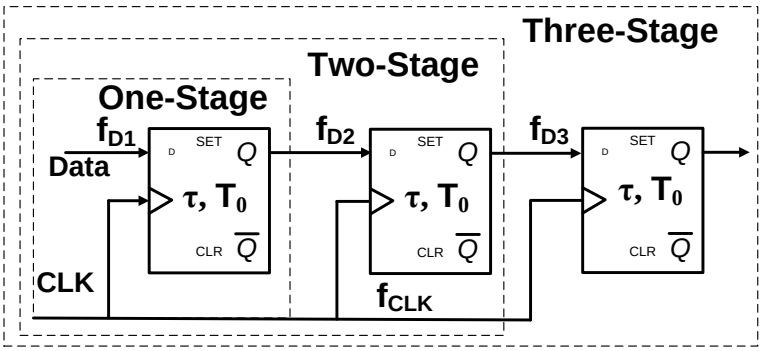


Figure 2.14: Single and Multi-Stage Synchronizer

approach allows the first synchronizing flip-flop to resolve from metastable events for an entire clock period (excluding the setup time of the successive flip-flop), and thus reducing the probability of metastable inputs into the successive flip-flops. Even if the first flip-flop is unable to resolve from the metastable state, the second flip-flop also has an entire clock period to resolve the output to a stable state, and so on. Assume all the flip-flops in the synchronizer have identical parameters (τ, T_0), the MTBF of a single-stage synchronizer can be calculated using **Equation (2.6)**. For the two-stage synchronizer, **Equation (2.11)** calculates f_{D2} by assuming it is the probability that the first flip-flop has not settled to a stable state within one clock cycle. For simplicity, we ignore the setup time of the second flip-flop in our calculation.

$$f_{D2} = \frac{1}{MTBF_1} = f_{D1} f_{CLK} T_0 e^{\frac{1}{f_{CLK}}/\tau} \tag{2.11}$$

Equation (2.12) provides the calculation for a two-stage synchronizer.

$$MTBF_{two-stage} = \frac{1}{f_{D2} f_{CLK} T_0 e^{\frac{1}{f_{CLK}}/\tau}} = \frac{1}{f_D f_{CLK}^2 T_0^2 e^{\frac{2}{f_{CLK}}/\tau}} \tag{2.12}$$

A similar analysis can be extended to calculate the MTBF of a three-stage synchronizer.

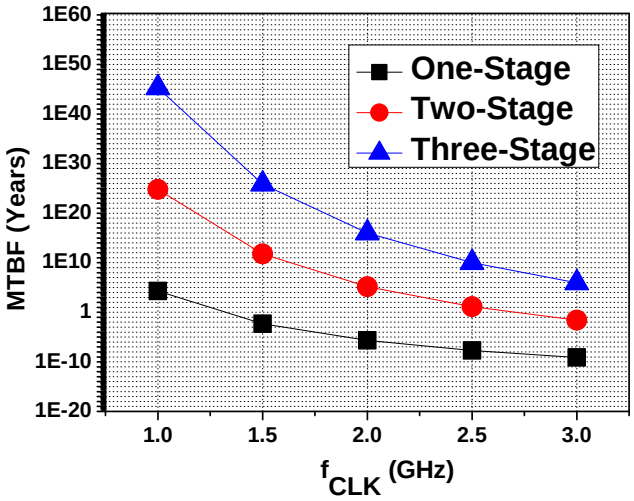


Figure 2.15: MTBF Comparison of Single and Multi-Stage Synchronizer

Figure 2.15 shows the MTBF for three different types of synchronizers as a function of clock frequency ranging from 1GHz to 3GHz. It is evident that using an extra stage synchronizer can improve the MTBF by a least ten orders of magnitude. For example, the MTBF of a single-stage synchronizer is only 0.00506 years (equivalent to 44.4 hours) when the clock frequency is 1.5GHz. When two-stage and three-stage synchronizer are used, the MTBF increases to 6.06×10^{11} and 7.26×10^{25} years, respectively. While the usage of multi-stage synchronizer increases the MTBF of the system significantly, it also increases the overall latency of the system.

2.7.2 Circuit Techniques

Another method to improve metastability is to design metastable-hardened flip-flops with smaller T_0 and τ values. In particular, these flip-flops must have a feedback path loop with a high-gain-bandwidth product to achieve a lower value of τ due to its exponential relationship with the metastability window δ and the MTBF. A common synchronizer used in asynchronous designs is the jamb-latch flip-flop [3], which consists of master and slave jamb-latches (**Figure 2.16**). Each latch is reset to logic "0" while the input data D is low. When D rises before the CLK, the master output X is driven high. This in turn drives the slave output Q high when the CLK rises. The pull-down transistors should be sized large enough to over-power the cross-coupled inverters. The jamb-latch flip-flop exhibits good metastability due to the cross-couple inverter pair and a relatively small loading on the feedback loop. A modified version of the jamb-latch was proposed in [21]

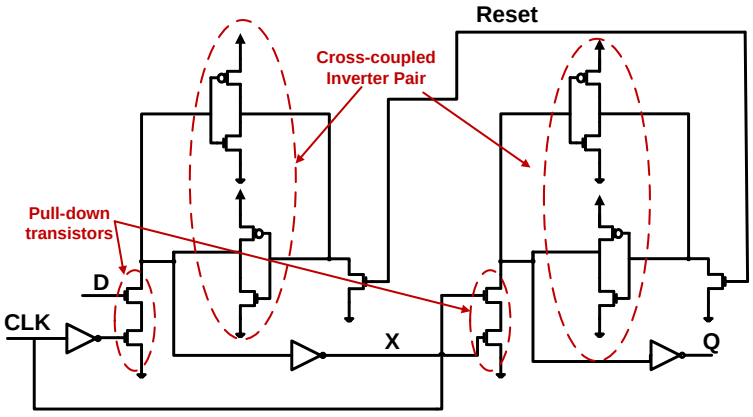


Figure 2.16: Schematic Diagram of the Jamb-Latch Flip-Flop

to exhibit more robustness against voltage and temperature variation. While jamb-latch flip-flop exhibits good metastability, it is not a conventional design that can be used in synchronous pipelined systems because of its inability to sample a logic "0" without a

"Reset" signal.

In synchronous pipelined systems, Razor flip-flop (RFF) shown in **Figure 2.17**, proposed in [37], can be used to provide an in-situ error detection and correction mechanism to recover from timing errors. The RFF composes of a standard D-flip-flop (DFF) , a shadow latch, a metastability detection circuit, and a comparator circuit. While the positive-edge triggered flip-flop samples the data, the input data D is given the duration of the positive CLK phase to settle down to its correct state before the shadow latch samples it at the negative edge of the CLK. An XOR comparator flags a timing error when it detects a discrepancy between the input data sampled at the DFF and the shadow latch. As part of the RFF, an additional detector is required to correctly flag the occurrence of metastability at the output of the DFF. Overall, the outputs of the metastability detector and the error comparator are ORed to generate the error signal of the RFF. Once metastability is detected, a restore signal overwrites the shadow latch data into the main flip-flop, and therefore restoring the correct state in the pipeline following the errant cycle. Although

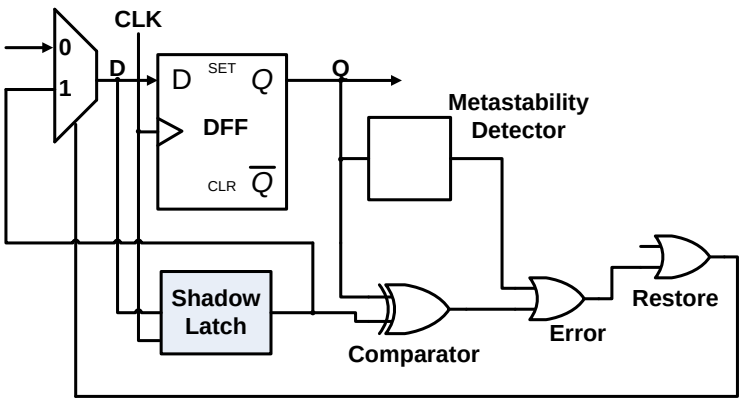


Figure 2.17: Schematic Diagram of the Razor Flip-Flop

the RFF provides error protection and correction mechanism, the amount of power and

area overhead associated with such design can be substantial when compared to a standard D-flip-flop.

2.8 Extraction Method of Flip-Flop Metastability

To extract τ and T_0 from simulation for metastability analysis, the C-Q delay vs.displacement between the input data and the clock signal is plotted for a given flip-flop architecture. In order to obtain accurate results in the metastable region, the data arrival time is varied at an interval of 1fs to generate the corresponding C-Q delay. From the plot, the metastable point, (t_{meta}), at which the flip-flop fails to capture the correct data can be easily obtained, and the last 500 data points before t_{meta} is collected for analysis. The next step is to obtain a theoretical linear curve from the C-Q delay vs. the time displacement between the input data and t_{meta} plot on a semi-log scale (linear scale on Y-axis and log-scale on X-axis). The slope of this line is the time resolving constant τ and the X-intercept is $log(T_0/2)$

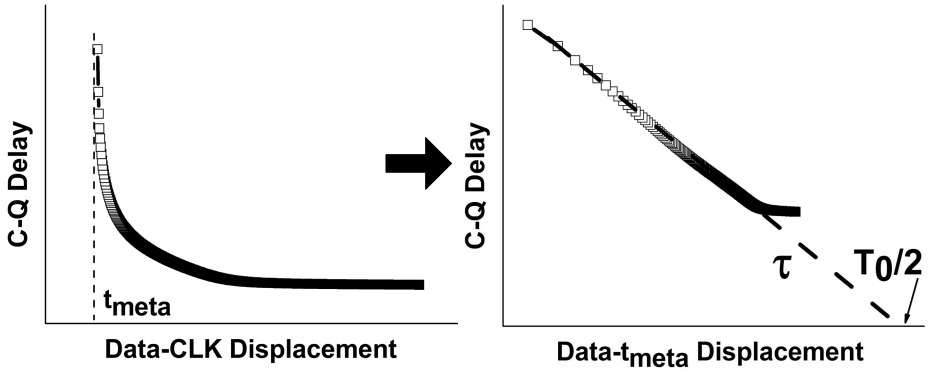


Figure 2.18: Illustration for Extracting Metastability Parameters

(**Figure 2.18**) [23][38][33]. It is possible that the curve obtained is not perfectly linear because the slope in the quasi-metastable and the metastable region could be different. To

be conservative in the analysis, we use the largest slope value and the corresponding X-intercept in the linear curve in extracting τ and T_0. The 500 data points collected translate into a near-metastable region of 0.5ps, which is adequate enough to obtain a meaningful extraction on the metastability parameters [3].

All the flip-flop metastability parameters in this work are extracted using the method described above. **Figure 2.19** shows a sample extraction of different sets of τ and T_0 for a given flip-flop architecture obtained via transistor sizing.

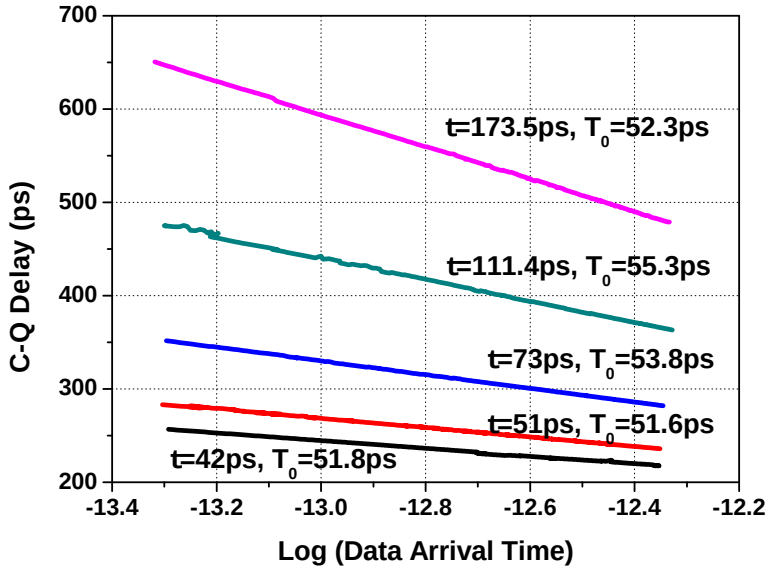
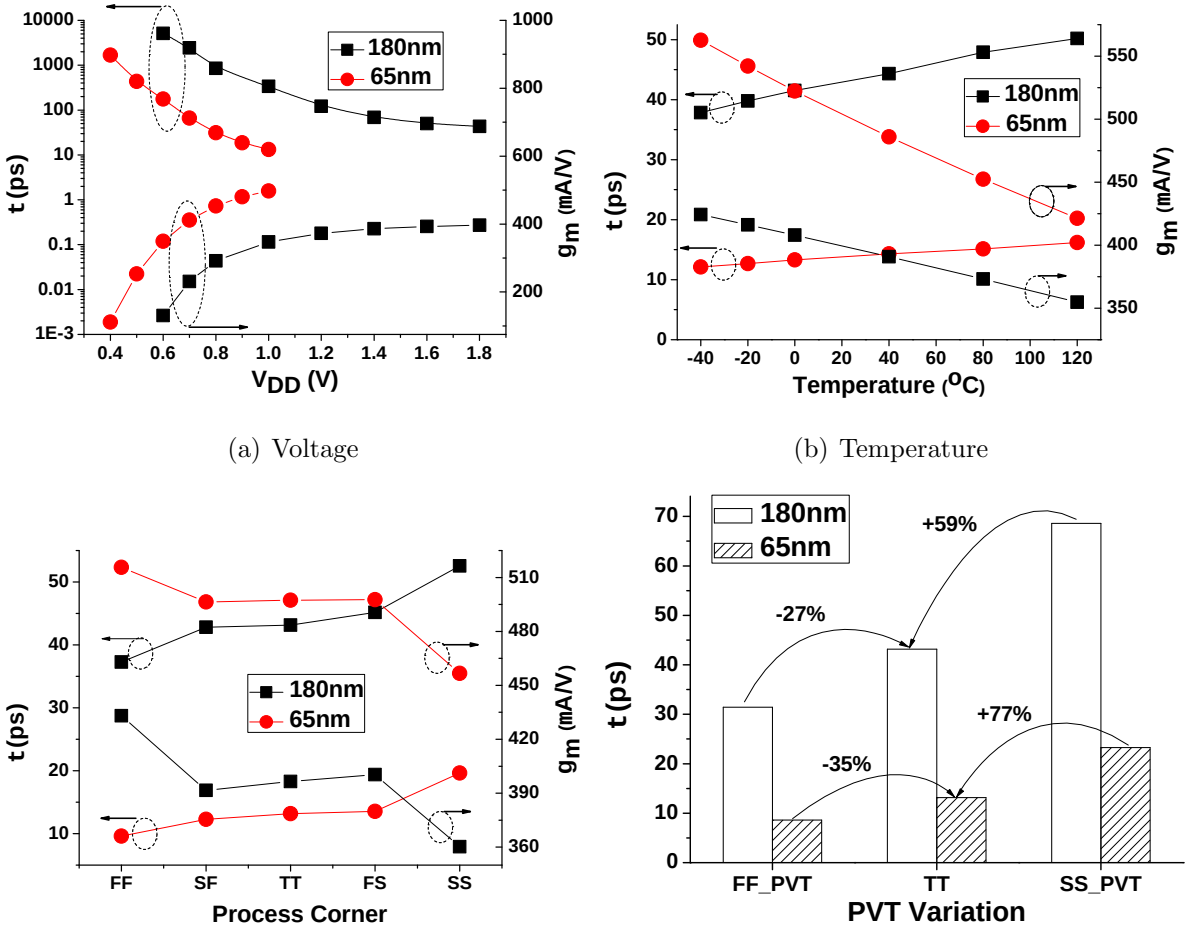


Figure 2.19: Sample Extraction of the Metastability Parameters

2.9 Impact of Process, Voltage, and Temperature Variation On Metastability

In this section, the effects of process, voltage, and temperature variation on the τ of the jamb-latch flip-flop will be illustrated using results obtained from Spice simulation in both $0.18\mu m$ and 65nm technology. We focus on the analysis of τ exclusively because it has the greatest impact on the metastability window δ and the MTBF. As evident from **Equation (2.10)**, τ has an inverse relationship with the transconductance g_m, and **Figure 2.20** plots both the value of τ as well as the sum of g_m for a NMOS and PMOS transistor.

Figure 2.20(a) shows that a reduction in the supply voltage V_{DD} results in an exponential decrease of the g_m, which results in an exponential increase in τ. **Figure 2.20(b)** illustrates a linear relationship between g_m and the temperature, which coincides with the previous studies that demonstrate the dependence of transistor characteristics on temperature [39][40][41][42]. A linear relationship between g_m and temperature also translates to a linear change of τ with respect to the temperature. While the FF and SS corner have resulted in smaller and larger τ values than the TT corner, as shown in **Figure 2.20(c)**, it is interesting to notice that both the SF and the FS corner have resulted in similar τ values as the TT corner. This is because the PMOS and the NMOS transistors under different process variations (i.e. sNfP) in the inverter pair compensate each other to resolve data. For example, the maximum deviation of τ in the SF and FS corner is only around $\pm 7\%$ of the TT corner in both $0.18\mu m$ and 65nm technology. **Figure 2.20(d)** illustrates the simultaneous effect of process, voltage and temperature variation on τ with **Table 2.1** showing the simulation conditions in both $0.18\mu m$ and 65nm technology. While Monte Carlo simulations provide the distribution of τ due to random variations, both the "FF_PVT" and the "SS_PVT" simulation conditions shown in **Table 2.1** provide the lower



(a) Voltage

(b) Temperature

(c) Process

(d) PVT

Figure 2.20: Effects of Process, Voltage, and Temperature Variation on τ

and upper bound limits on τ in order to achieve the appropriate MTBF values under the best and the worst conditions.

Table 2.1: Simulation Conditions for Different Process Corners

Corner	Temperature	V_{DD} ($0.18\mu m$)	V_{DD} (65nm)
FF_PVT	$-40°$C	1.98V	1.1V
SS_PVT	110°C	1.62V	0.9V

2.10 Summary

In this chapter, various background information on flip-flop metastability is examined and analyzed. Flip-Flop metastability can exists in both synchronous and asynchronous systems. It is a phenomenon where the flip-flop violates the setup and hold time constraints and subsequently enters an undesirable third state in which the output is stuck at an intermediate level between logic "0" and "1". Metastability is quantitatively characterized by the metastability window δ and the mean-time-between-failure (MTBF), where both metrics are a function of T_0, τ, and t_s. Among them, T_0 and τ are related to the flip-flop architecture while the settling time t_s depends on the design of the overall system. However, τ and t_s have the greatest impact on metastability due to an exponential relationship. The small signal modeling of a cross-coupled inverter pair provides the foundation for the analysis of τ, which is a function of the transconductance g_m and the parasitic capacitances C_Q. Various metastability mitigation techniques have been proposed both at the circuit and the system level. On the system level, multi-stage synchronizer can significantly increase the MTBF of the system at the expense of increased latency. From the circuit perspective, the Jamb latch-based flip-flop exhibits low value of τ but is exclusively used in synchronizer circuits while the Razor flip-flop enhances the reliability of the pipeline system but encounters significant overhead in both area and power consumption. Finally, the impact of process, voltage, and temperature variation on the value of τ have been shown to have

an inverse relationship with the transconductance g_m. In general, τ has an exponential and linear relationship with the supply voltage V_{DD} and temperature, respectively. As for the process corner, the FF and the SS corner result in lower and higher values of τ than the TT corner while the impact of the FS and the SF corner is negligible when compared to the TT corner. The simultaneous effect of process, voltage, and temperature provides an upper and lower bound on τ for a given flip-flop architecture.

Chapter 3

High-Performance and Low-Power Flip-Flop Architectures

In this chapter, we discuss in details the design of high-performance, low-power flip-flop architectures that can be used in either single and dual-supply systems. Flip-flop is a vital component for high-performance and reliable deep-pipelined systems in digital microprocessors. Various flip-flop architectures have been proposed in the past to facilitate different design objectives such as performance, power, and area constraints. The most notable design techniques include transmission-gate based, tri-state inverter based, pulse-triggered, conditional capturing, and single-clocked phase. While most of the flip-flops are designed for single-supply systems, the recent trend for low-power systems have engaged more flip-flop designs for dual-supply systems. In dual-supply systems, there are mainly two types of flip-flop designs: reduced-clock-swing flip-flops (RCSFF) and level-converting flip-flops (LCFF). In RCSFF, the voltage swing of the CLK is reduced to V_{DDL} while the remaining circuit is still operating on the nominal supply voltage V_{DDH}. In LCFF, the voltage swing

of both the input data D and the CLK signals are reduced to V_{DDL} while the final output Q maintains a voltage swing of V_{DDH}. In any case, flip-flop designs are more challenging in dual-supply systems because special architectures are required such that the reduced swing signals cannot be applied directly to the gate of PMOS transistors to avoid static power dissipation.

Two new high-performance and low-power flip-flop designs are proposed in this work. The main objective behind the proposed designs is to use the same architecture and achieve high-performance and low-power in both the single and dual-supply systems. The first design is called the pre-discharge flip-flop (PDFF) where high-performance is achieved by reducing the number of transistors in the critical path. With fewer transistors in the critical path, the amount of power consumption and total transistor widths have also been reduced accordingly. The second design is called the sense-amplifier transmission-gate flip-flop (SATG) . The master-stage of the SATG utilizes a sense-amplifier like structure with NMOS-pass transistors along with "helper" discharge paths for performance enhancement. While the performance of the SATG is not as good as the PDFF, it exhibits a very good performance and low-power consumption in the dual-supply system. The detailed operation and architecture of the proposed flip-flop designs are provided in this chapter.

Extensive post-layout results will be provided to compare the proposed designs with the previous flip-flop architectures in terms of performance and power consumption. Performance comparison includes propagation delays such as CLK-Q and Data-Q as well as the setup and hold time constraints along with the flip-flop aperture window $t_{aperture}$. Power consumption will be analyzed for various data activities ranging from 0%-100%. The overall comparison merit will be determined by the power-delay-product (PDP), which determines the amount of tradeoff between delay and power consumption.

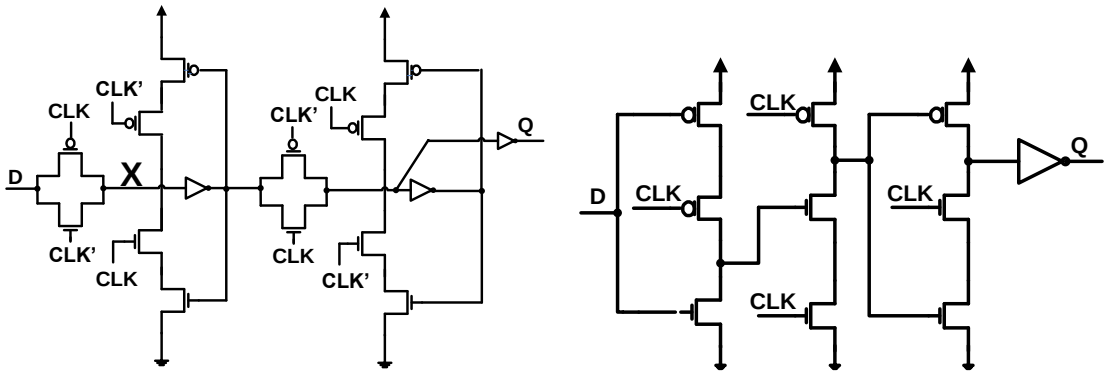
3.1 Single-Supply Flip-Flops

3.1.1 Single-Ended Flip-Flops

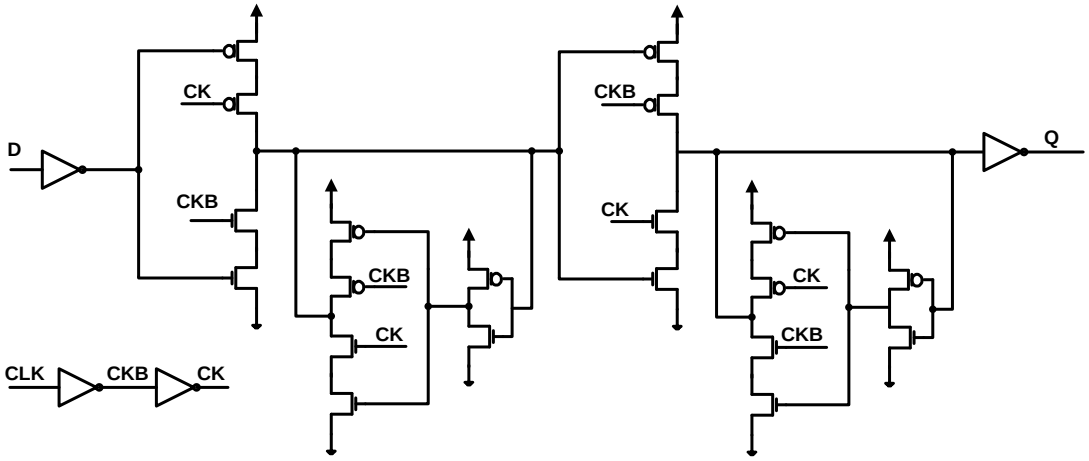
Transmission-gate flip-flops [23] exhibit high-performance and low-power characteristics due to its low-impedance paths. Among them, the PowerPC [43] shown in **Figure 3.1(a)** is a classical single-ended master-slave structure with short direct path and low-power consumption. The good performance of the PowerPC when compared with other transmission-gate based flip-flops comes from the use of complementary pass-gates and low-power feedback. However, the usage of both the CLK and CLK' signal increases the sensitivity to race through in the period of one gate delay in which the two phases overlap. Moreover, its positive setup time makes the overall performance less superior than the pulsed-triggered flip-flops.

The modified C^2MOS (mC^2MOS) [32] is composed of two identical cascaded latches that is insensitive to clock overlap, as long as the rise/fall times of the CLK signal are sufficiently small. Its schematic diagram is shown in **Figure 3.1(c)**. The performance of this flip-flop is slower than the PowerPC because of a large capacitive load associated at the critical nodes. While it exhibits low-power properties featuring small clock load, the local clock buffering still makes its overall power consumption relatively high.

The True Single-Phase Clocked (TSPC) flip-flop, proposed in [44], uses only a single clock phase. While the usage of one clock phase is attractive for many reasons such as smaller clock load and the elimination of clock overlap, its architecture produces a momentary glitch at the dynamic nodes after the rising CLK edge when the D is low for multiple cycles, which increases the overall power consumption. The schematic diagram of the TSPC is shown in **Figure 3.1(b)**.



(a) PowerPC (b) TSPC



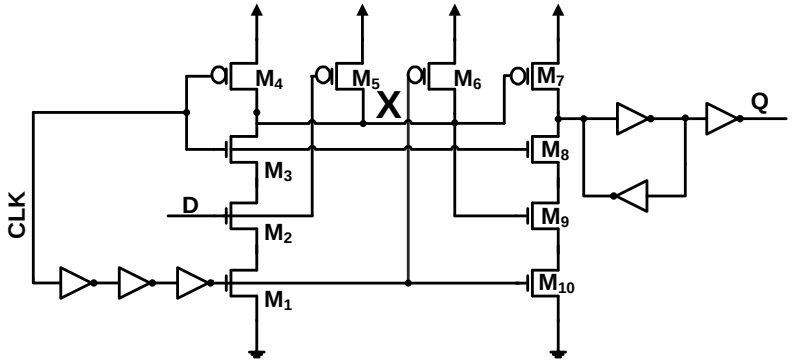
(c) mC^{2}MOS

Figure 3.1: Single-Ended Flip-Flops

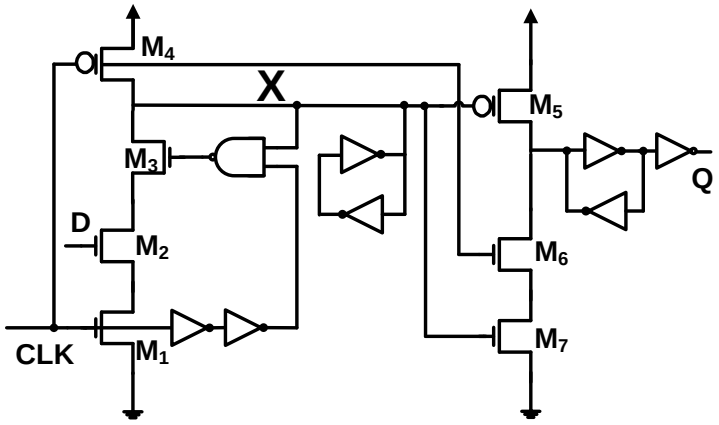
3.1.2 Pulse-Triggered Flip-Flops

The hybrid-latch flip-flop (HLFF) [45] and the semi-dynamic flip-flop (SDFF) [46] best represent the pulsed-triggered flip-flops where performance is greatly enhanced due to negative setup because data is captured during a brief transparent period created by the pulse

generator. Other than very good performance, these flip-flops exhibit soft-edge property where the robustness against clock skew is greatly enhanced. Due to their respective architecture, the overall D-Q delay of the HLFF and the SDFF is a strong function of the negative setup time, and hence resulting in a large hold time as well. The schematic diagram of the HLFF and the SDFF is shown in **Figure 3.2(a)** and **3.2(b)**, respectively.



(a) HLFF



(b) SDFF

Figure 3.2: Pulsed-Triggered Flip-Flops

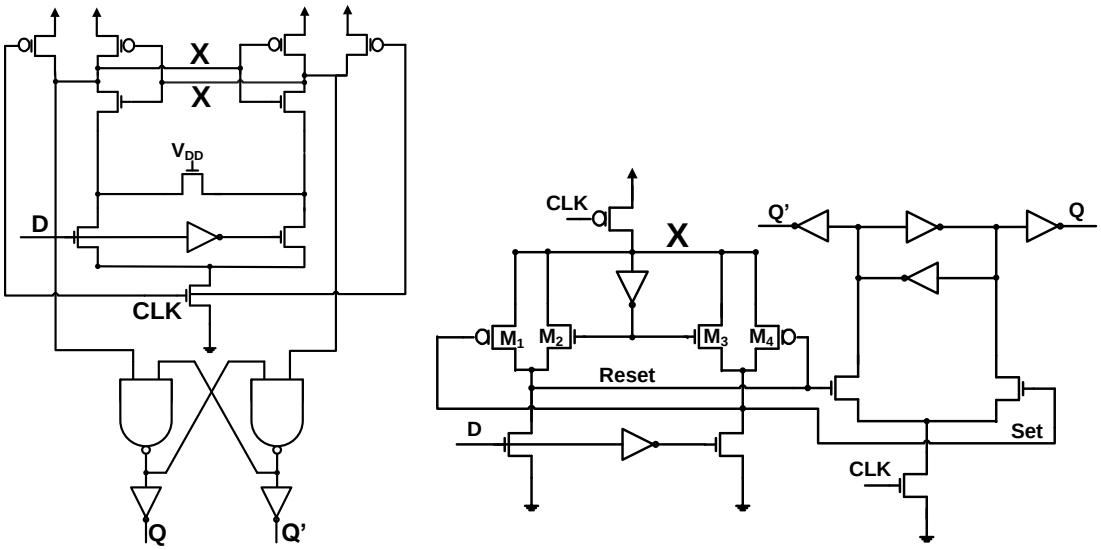
In the HLFF, a local pulse generator was built into the flip-flop itself. When the CLK is

low, transistor M_3 and M_8 are off while M_4 is turned on. Hence, node "X" is pre-charged to logic "1", and the output node Q is decoupled from "X" and holds the previous state. On the rising edge of the CLK, M_3 and M_8 are turned on while M_1 and M_{10} also stay on for a short period of time, which is determined by the delay in the pulse generator. During this interval, the entire flip-flop is transparent as the input data D is sampled. Once the pulsed CLK goes low, node "X" is decoupled from the input and is either remains unchanged or begins to pre-charge to V_{DD} through transistor M_4.

The SDFF is another pulse-triggered flip-flop that exhibits extremely high-performance. It is called semi-dynamic because it combines the dynamic input stage with static operation. When the CLK is low, node "X" pre-charges to logic "1" and the output Q holds the previous state. When the CLK rises, the dynamic NAND gate evaluates. If the D is logic "0", "X" remains at logic "1" and NMOS transistor M_2 is turned off. If the D is logic "1" and "X" starts to discharge to cause an output transition. The SDFF is slightly faster than the HLFF but loses the skew tolerance and time-borrowing capability. Its main disadvantages include bigger clock load and large effective pre-charge capacitance, which results in increased power consumption especially when there is more logic "1" in the input data.

3.1.3 Differential Flip-Flops

The sense-amplifier flip-flop (SAFF) [47] is a pure differential flip-flop that receives differential inputs and produce different outputs. When the CLK is low, the internal nodes "X" are pre-charged to V_{DD}. On the rising edge of the CLK, one of the two nodes is pulled down, and the cross-coupled PMOS transistors act as a keeper for the other node. The slave-stage of the SAFF is composed of a cross-coupled SR NAND gates that captures the



(a) SAFF (b) SSTC

Figure 3.3: Differential Flip-Flops

output and holding through the pre-charge period. This flip-flop is able to amplify and respond to small differential input voltages and has a small clock load and avoids the need for an inverted clock. A modification of the original SAFF design was made in [48] where a weak NMOS transistor is added to fully staticize the flip-flop by avoiding float internal nodes. Another modification to the design was made by [49] where HI-skew inverters replaced the cross-coupled NAND gate in the slave-stage to result in a more even propagation delay for both the 0-1 and 1-0 output transitions. Although the sense amplifier stage is fast, the propagation delay through the cross-coupled slave-stage and the pre-charge activity during every clock cycle hurts its overall performance and power consumption. The schematic diagram of the SAFF is shown in **Figure 3.3(a)**.

The Static Single Transistor Clocked (SSTC) flip-flop [50] is an example of a differential

flip-flop that utilizes just one clock phase. The master-stage of the SSTC asserts the "Set" or the "Reset" signal when the CLK is low. The slave-stage then uses these signals to change the outputs during the evaluation period when the CLK is high. The extra inverter and NMOS transistors in the master-stage discharge the "Set" and "Reset" signal to logic "0" if the inputs change when the CLK is high. SSTC suffers from substantial voltage drop at the outputs due to the capacitive coupling effect between the common node of the slave-stage and the floating high output node of the master-stage. This voltage drop decreases the driving capabilities of the master-stage and this causes an increase in both delay and power consumption. The schematic diagram of the SSTC is shown in **Figure 3.3(b)**.

3.1.4 Conditional Capture Flip-Flops

A new family of low-power flip-flops, namely the conditional-capture flip-flop (CCFF) , was presented in [51]. The motivation behind the conditional capture technique is that considerable portion of power is consumed for driving internal nodes even when the input data activity is low such that the value of the output does not change very often. To accomplish this, the flip-flop conditionally enables the discharge path and turns it off after a brief sampling period. The schematic diagram of the CCFF is shown in **Figure 3.4**.

The CCFF consists of two stages: a differential master-stage with a pair of NOR gates and clocked inverters and a cross-coupled SR latch in the slave-stage. The NOR gates are driven by the outputs to make the discharge of the pre-charge nodes, SB and RB, conditional depending on the input and output data. They are also controlled by the delayed CLK signal to determine the transparency period. The outputs of the master-stage, S/SB and R/RB, are fed into the slave-stage to the SR larch, which captures each

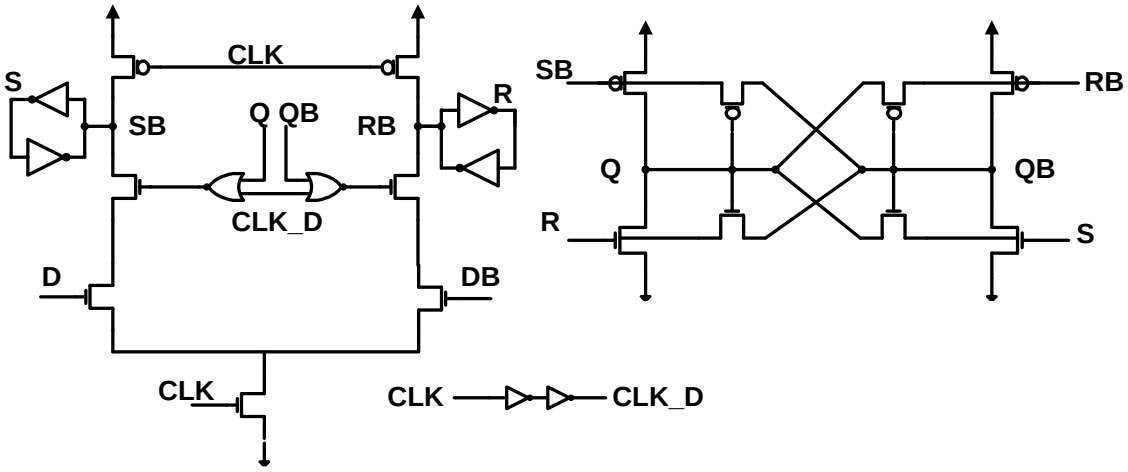


Figure 3.4: Conditional-Capture Flip-Flop

transition and holds the outputs until the next pull-down transition occurs on one of the pre-charged nodes.

While the CCFF achieves statistical power reduction by eliminating redundant internal transitions, the amount of area overhead is substantial. Furthermore, the amount of power overhead due to extra transistors and a large clock load can actually offset the amount of power reduction achieved, even at low data activities. Nonetheless, the conditional-capturing technique is still being utilized in many low-power flip-flop designs.

3.2 Reduced Clock-Swing Flip-Flops

In VLSI systems, a large portion of the power consumption comes from the clock subsystems, including clock generation, distribution, and the final sequential elements load. Due to high frequencies, low skew requirements, and deep pipelining, the clocking power has been increasing with each processor generation [52]. In fact recent studies have shown that

the clock system consumes anywhere between 20-45% of the total chip power with approximately 90% of the clocking power used to drive storage elements such as flip-flops [53][54]. More specifically, a typical arithmetic logic unit (ALU) design in $0.18\mu m$ has shown that the entire clock network contributes to 59.4% of the ALU total energy. This is illustrated in **Figure 3.5**. The significant power consumption of the clock system comes from the fact that the transition probability of the clock signal is 100%. Therefore, reduced-swing clocking, where the clock is distributed at a lower voltage (V_{DDL}) than the rest of the system that is operating at the nominal supply voltage of V_{DDH}, is a viable technique for the overall power reduction. **Equation (3.1)** shows that the amount of power reduction on the clock system can be significant due to the quadratic relationship.

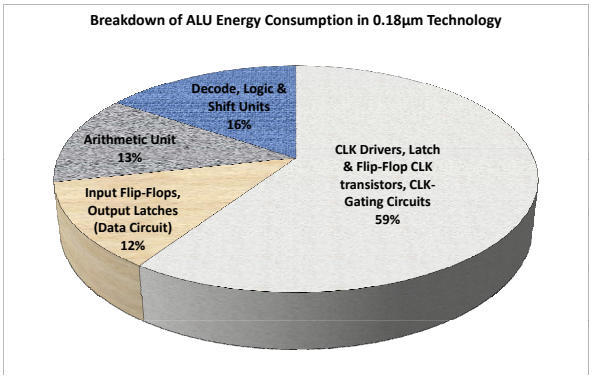


Figure 3.5: Energy Breakdown of an ALU in $0.18\mu m$ Technology

$$P_{clk} = \alpha C_{clk} V_{DD} V_{clk-swing} f_{clk} \tag{3.1}$$

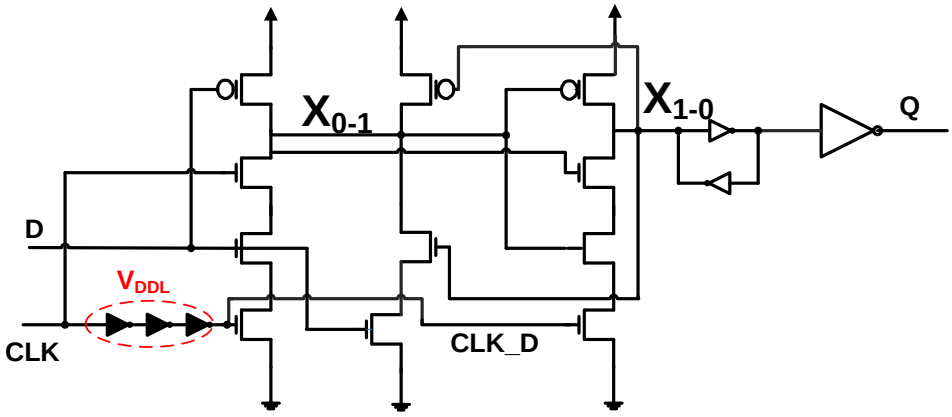
While reduced clock-swing system results in power consumption, it also suffers in circuit performance degradation due to a smaller overdrive voltage driving the gate of the tran-

sistors. Hence, past studies have shown the region of minimum energy operation occurs when $V_{DDL} = 0.7 - 0.75V_{DDH}$ while the region of minimum energy-delay product (EDP) operation is $V_{DDL} = 0.85 - 0.9V_{DDH}$ [55].

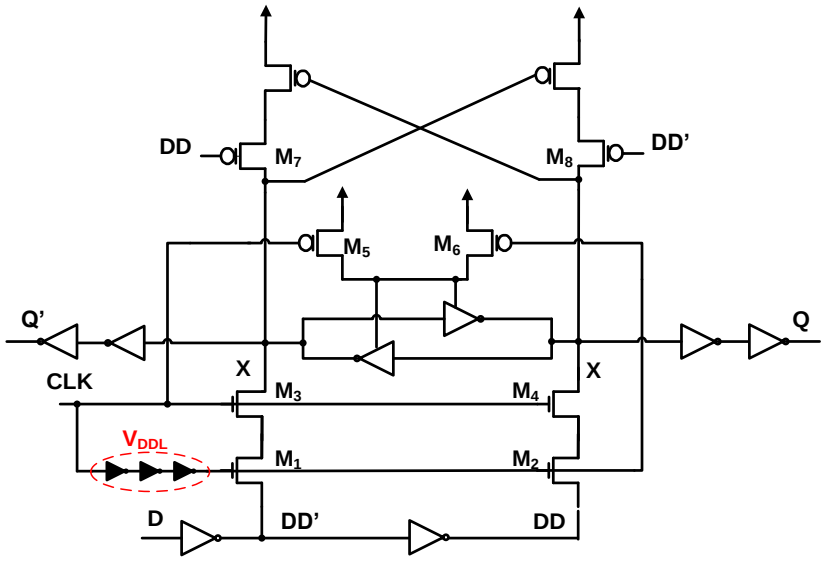
Reduced-swing clocking cannot be implemented simply by scaling down the supply voltage of the clock system. Standard flip-flops for traditional full-swing clocking cannot be used with reduced clock-swing system because any clocked-PMOS transistors will not fully turn off, and thus causing static current consumption and reduced noise margin. The first idea to alleviate this problem is to insert a level converter in front of standard flip-flops to regenerate full clock swings. However, this does not result in much power savings since voltage swings are reduced only on the clock distribution network while the large number of level converters result in significant power overhead. Furthermore, the insertion of level converters result in significant delay penalty in the critical path. Hence, a more efficient approach would be to design flip-flops that can directly receive a reduced clock swing signal.

One of the first proposed reduced clock-swing flip-flop (RCSFF) [54] uses the SAFF architecture. It has only one clocked transistor in the critical path and results in the smallest performance degradation as the clock swing is lowered to V_{DDL}. However, the clocked pre-charge PMOS transistors results in a direct current path and significant power consumption. Although the reverse body-bias (RBB) technique is used to mitigate this problem, it becomes less attractive in smaller technologies due to the extra area requirement for the separate n-well and the reduced effectiveness of the RBB technique in increasing V_{th}.

The NAND-type keeper flip-flop (NDKFF) , **Figure 3.6(a)**, was proposed in [8] where only NMOS transistors are clocked and thus eliminate the leakage power problem for reduced clock swing. This flip-flop modifies the HLFF architecture and results in high-



(a) NDKFF



(b) CRFF

Figure 3.6: Reduced Clock-Swing Flip-Flops

performance due to the pulse-triggered operation. However, two internal nodes ("X_{0-1}" and "X_{1-0}") are subject to contention. This requires larger transistor sizing in the critical path to overcome the feedback transistors and cause the internal nodes to make the correct

transitions. In addition to the issue of contention, the high-stacked transistor in series is undesirable for scaled technologies with smaller nominal supply voltage V_{DD}. For example, simulation results show that the NDKFF fails to function in 65nm technology when V_{DDL} is lowered to $\frac{1}{2}V_{DD} = 0.5V$ because the reduced current drive due to lowered clock swing is unable to switch the huge capacitance associated at nodes "X_{0-1}" and "X_{1-0}".

In [9], a new reduced clock-swing and contention-reduced flip-flop (CRFF) , **Figure 3.6(b)**, is proposed to reduce the effect of stacked transistor in series and contention at highly capacitive node. Contention currents are reduced in two ways. First of all, the pull-up circuit is controlled by the input data D through transistor M_7 and M_8 to reduce the contention with NMOS-pass transistors. Secondly, clock-driven transistors M_5 and M_6 disconnect the cross-coupled latch from V_{DD} during the transparency window. This type of flip-flop is pulsed-triggered and uses NMOS-only transmission-gates in the critical path where the propagation delay of writing a logic "0" and logic "1" can be significantly different. Furthermore, the decreased driving capability of the transmission-gate at reduced clock-swing results in significant performance degradation.

3.3 Level-Converting Flip-Flops

Another method of reducing power consumption in digital systems is to adopt a clustered voltage scaling (CVS) scheme where lower supply voltage (V_{DDL}) is used in non-critical paths while placing the nominal supply voltage (V_{DDH}) on the critical paths [10][11][12]. Such scheme does not degrade system performance while resulting in power reduction. An example of the CVS scheme is shown in **Figure 3.7**. The shaded logics and flip-flops indicate they are operating at V_{DDL}.

To implement the CVS scheme on a chip, a level converter must be used when interfacing

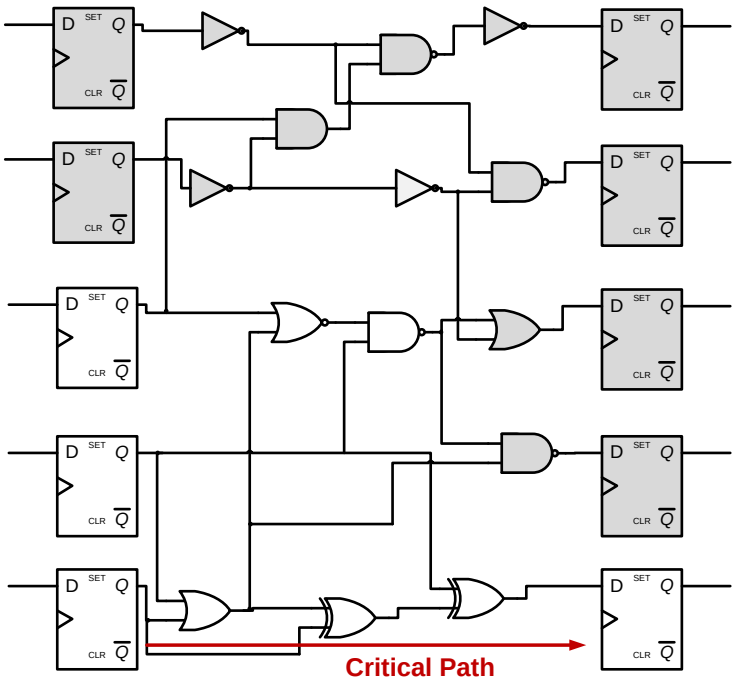


Figure 3.7: Illustration of Cluster Voltage Scheme

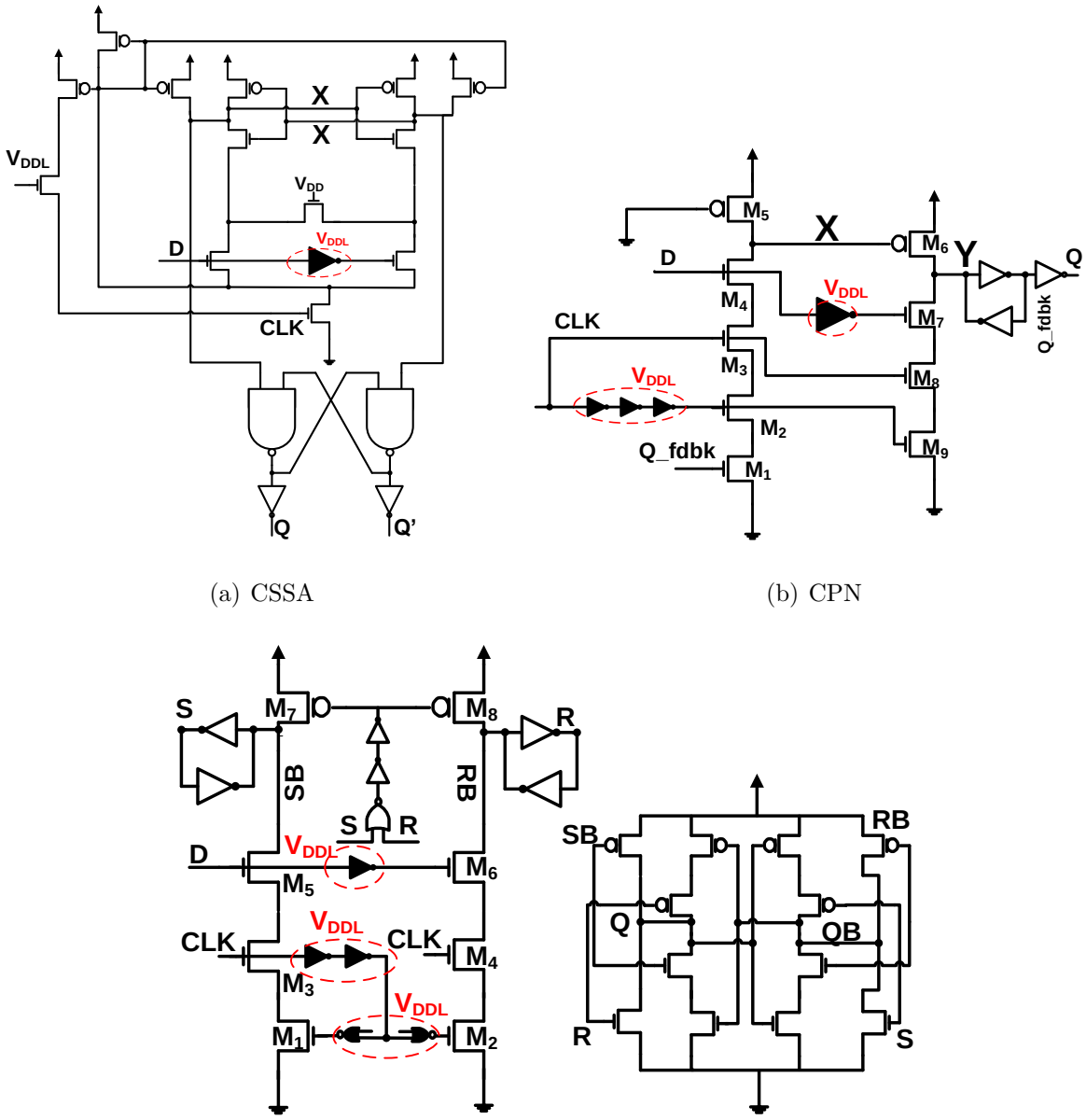
two gates on different supply voltages to avoid static power dissipation. The usage of level converters, however, encounter huge amount of performance and power overhead for the same reasons stated in the previous section. Thus, integrating the level conversion in the flip-flops have become a more popular design choice. In LCFFs, the voltage swing of both the D and the CLK signal is at V_{DDL} while the final output Q is at V_{DDH} [56].

One of the level-converting flip-flop proposed is the Clock-level Shifted Sense Amplifier (CSSA) flip-flop [57]. The CSSA is very much similar to the SAFF described previously except the clock-signal is level shifted to V_{DDH} to avoid static power dissipation during the pre-charge cycle. In addition to the problems associated with the SAFF, the level-converting circuit in this design also consumes a substantial amount of power. While the

potential of static power dissipation is eliminated when lower swing signals are connected to PMOS transistors, the level-shifting circuit itself is consuming a significant amount of power. Hence, the benefit of level-shifting may not actually outweigh the drawback of static power dissipation as the CMOS technology scales deep into the sub-micron regime. The schematic diagram for the CSSA is shown in **Figure 3.8(a)**.

A new improved level-converting flip-flop called Self-Pre-charging Flip-Flop (SPFF) was proposed in [58]. The SPFF, shown in **Figure 3.8(c)**, employs a self-pre-charging technique to pre-charge the dynamic nodes, which eliminates the need for the CLK to drive the PMOS transistors. This flip-flop also employs the conditional capturing technique to remove redundant internal transitions while exhibiting high-performance with negative setup time. The operation of the SPFF is very similar to that of the CCFF described previously. The amount of power saving achieved by internal gating is larger than the incurred power overhead for relative low data switching activities. For high data activities, however, the conditional capturing technique may not be of benefit since there is less chance to prevent redundant internal switching. The order of the transistor stack in the sampling path of the master-stage is based on the arrival time of the signals and increases the flip-flop performance and allows for negative setup time. A clock pulse is generated to control the NMOS transistors M_1 and M_2 to allow enough time for the output to make the correct transition before shutting the discharge paths. The slave-stage of the SPFF is a modified set-reset latch proposed in [49] which allows a balanced delay for 0-1 and 1-0 output transitions. Similar to the CCFF, the main drawback of the SPFF is the substantial area and power consumption overhead encountered.

A clocked-pseudo-NMOS (CPN) level-converting flip-flop was proposed in [59]. The CPN uses a pseudo-NMOS scheme with the conditional discharge technique [60] where a feedback signal Q_fdbk controls the transistor M_1 in the discharge path. A weak pull-up



(a) CSSA

(b) CPN

(c) SPFF

Figure 3.8: Level-Converting Flip-Flops

PMOS device M_5 is used to pre-charge the internal node "X" instead of using clocked pre-charging devices. While M_1 is always on, static current only occurs when the input D makes a 0-1 transition, and the discharge path is disconnected by Q_fdbk. Transistors in the discharge path (M_1, M_2, M_3, M_4) should all be sized appropriately to ensure adequate noise margin. The clock pulse generated must have adequate timing margin to allow a complete discharge of node "X" or "Y" depending on the data transition. Because of the clock pulse, the CPN also has the property of negative setup time to enhance the overall performance. While using fewer transistors than the SPFF, stacking four transistors in the critical discharge path require larger sizing in order to obtain optimum performance. Because of the pseudo-NMOS scheme and high transistor stack, the sizing scheme in the CPN is very critical in maintaining the correct circuit functionality. Furthermore,the design is very sensitive to process variation. The schematic diagram for the CPN is shown in Figure **3.8(b)**.

3.4 Proposed Flip-Flop Designs

3.4.1 Pre-Discharge Flip-Flop (PDFF)

In this work, we propose a pre-discharge flip-flop (PDFF) that exhibits both the characteristic of high-performance and low-power. The master-stage of the PDFF consists of a differential cross-coupled inverter with positive feedback in the critical path. A novel design is proposed to connect the CLK to the drain of the PMOS transistors. An equalizer transistor M_4 is used to discharge the internal nodes "Set" and "Reset" when the CLK is low. When the CLK becomes high, the critical path in the master-stage has been reduced to just a PMOS-pass transistor (M_5 or M_6) to charge one of the internal nodes to logic "1"

while the discharge paths are simply present to prevent false evaluation. A transparency window is created using a pulse generator to allow negative setup time for performance improvement as well as soft-edge robustness against clock skew. Due to the discharging in the master-stage when the CLK is low, the footer clocked-NMOS transistor in the slave-stage can be eliminated to further enhance the flip-flop performance. The output data is retained in the slave-stage by the SRAM-based cross-coupled inverter pair [61]. The schematic diagram of the PDFF is shown in **Figure 3.9**.

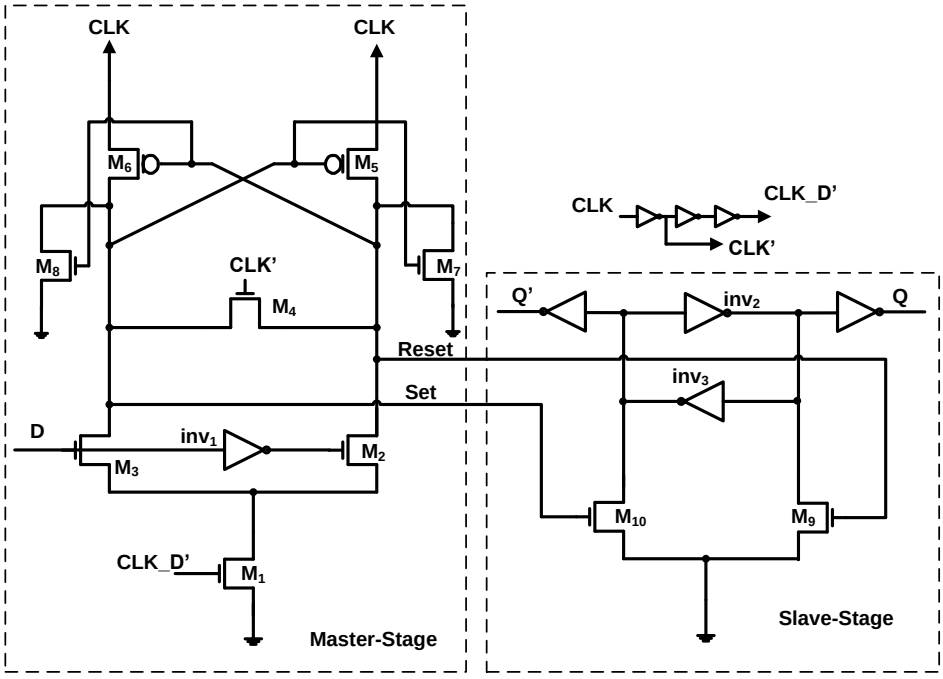


Figure 3.9: Schematic Diagram of the Pre-Discharge Flip-Flop Design

The detailed operation of the PDFF is as follows. During the period in which the transparency window is closed, both pull-down paths in the master-stage is off, and while the CLK is low, CLK' activates M_4 and pre-discharges the nodes "Set" and "Reset" to

logic "0". When both "Set" and "Reset" remain at logic "0", the SRAM-latch in the slave-stage does not turn on the NMOS transistors and holds the data to its current state. When the transparency window is open, M_4 is off, and depending on the input D, one of the pull down path is on while the other is off such that either "Set" or "Reset" will remain at logic "0" and the other node will be pulled up to logic "1" from the cross-coupled PMOS transistors M_5 and M_6. If "Set" remains at logic "0", it will turn on M_5 with the CLK being high and charges "Reset" to logic "1", which then turns on M_9 to bring the output Q to a logic "1". Due to the pre-discharging, the pull-down path in the master-stage is no longer on the critical path because it simply prevents any wrong evaluation outside the transparency period. As soon as M_4 is off, evaluation begins, and the critical path in the master-stage becomes just a single PMOS transistor of either M_5 or M_6 raising the signal "Set" or "Reset" to a logic "1" while the CLK is high. Together, transistor M_5, M_6, M_7, and M_8 form a clocked cross-coupled inverter pair in the master-stage of the PDFF. The timing diagram of the PDFF is given in **Figure 3.10**.

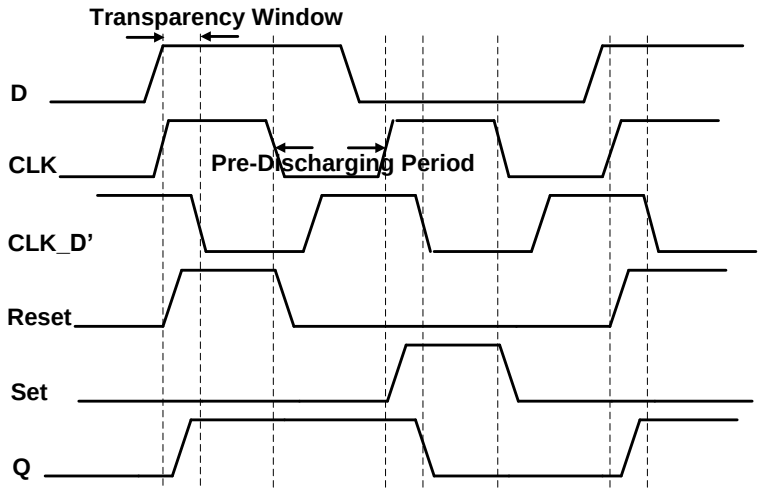
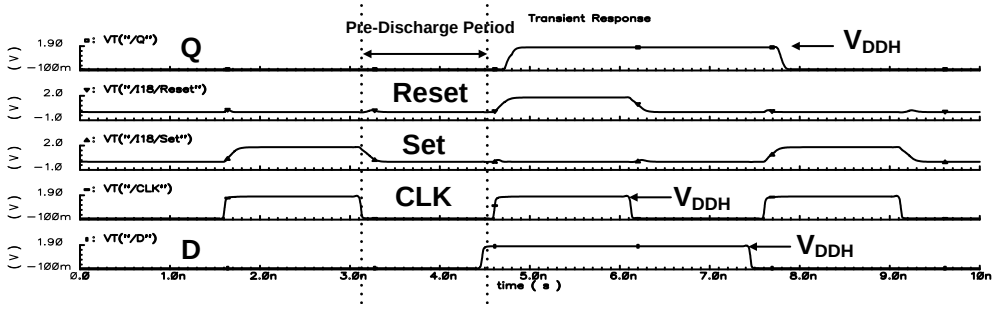
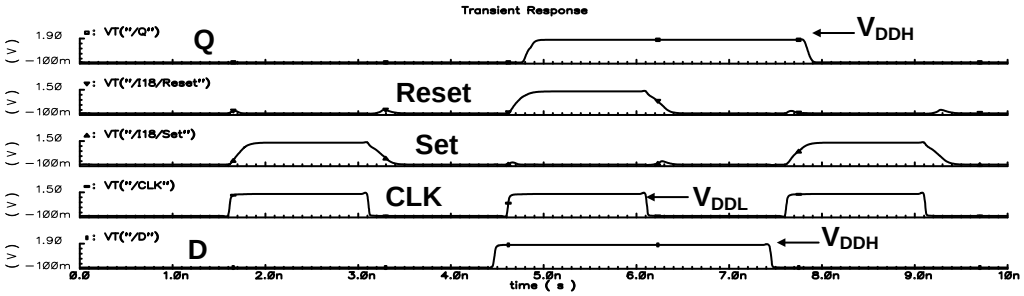


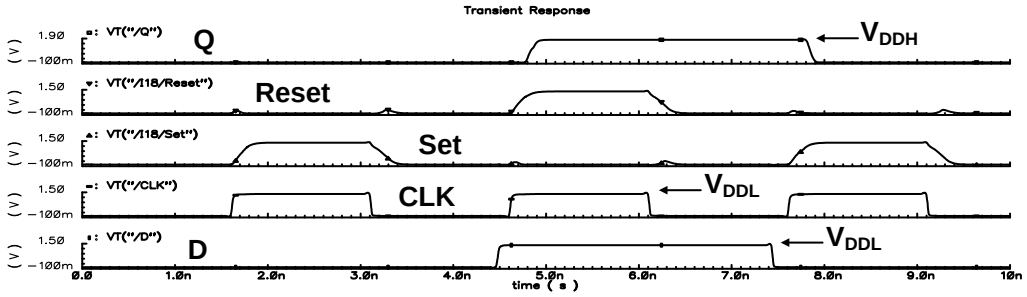
Figure 3.10: Timing Waveform of the Proposed Pre-Discharge Flip-Flop Design



(a) PDFF



(b) RCSPDFF



(c) LCPDFF

Figure 3.11: Simulation Waveforms for the PDFF in Single and Dual-Supply Systems

The high-performance of the PDFF mainly comes from the fact that it has very few transistors in the critical path. Not including the output buffer, the number of critical

transistors in the worst case is only 2P+N. This is less than the worst case delay of 3N+P in the HLFF and the SDFF, which are widely regarded as the fastest flip-flop architectures [32]. Due to fewer transistors in the critical path, the PDFF is also more area-efficient than the other flip-flop architectures when designing for optimal performance such that fewer critical transistors need to be sized up while the rest can be kept at or close to minimum size. A smaller total transistor widths of the PDFF also means its overall power consumption will be lower despite the high-performance characteristics.

The architecture of the PDFF also allows it to function as a high-performance reduced clock-swing flip-flop or level-converting flip-flop because neither the CLK or the data D signal is applied to the gate of PMOS transistors to cause significant leakage power. When the voltage swing on the CLK is reduced to V_{DDL}, the voltage swing on both the "Set" and the "Reset" are also reduced to V_{DDL}. Thus, an important reason that the SRAM-latch is chosen to be the slave-stage is because such architecture allows the internal nodes "Set" and "Reset" from the master-stage to only drive NMOS transistors. When used as a RCSFF and LCFF, the PDFF will be referred to as the RCSPDFF and the LCPDFF in this thesis respectively. **Figure 3.11** illustrates the simulated waveform for the PDFF, the RCSPDFF, and the LCPDFF. Appropriate voltage swing on the input and output signals is indicated for the respective flip-flop designs.

3.4.2 Sense-Amplifier-Transmission-Gate Flip-Flop (SATG)

While considering the design drawback of the SAFF, a new sense-amplifier-transmission-gate (SATG) flip-flop is proposed in this thesis work. As described earlier, the pre-charging of internal nodes that SAFF employs during every clock cycle increases the overall power consumption of the flip-flop. Furthermore, the stacking of three NMOS transistors in the

critical discharge path in the master-stage along with the cross-coupled NAND gate in the slave-stage have significantly impact its performance. In the master-stage of the SATG, transistor $M_1 - M_5$ form a sense-amplifier like architecture with a cross-coupled inverter pair along with the clocked NMOS transistor in the discharge path. The pre-charging transistors in SAFF are replaced with NMOS-pass transistors (M_{10} and M_{11}) that write differential data into the flip-flop. Additional discharge paths are added to enhance the performance by reducing the required setup time. The differential signals produced by the master-stage (Q1 and Q1B) facilitate the usage of the SRAM-latch in the slave-stage. Unlike the PDFF, however, an extra clocked footer NMOS transistor M_{12} must be present in the slave-stage to ensure the correct operation. The schematic diagram of the SATG is shown in **Figure 3.12**.

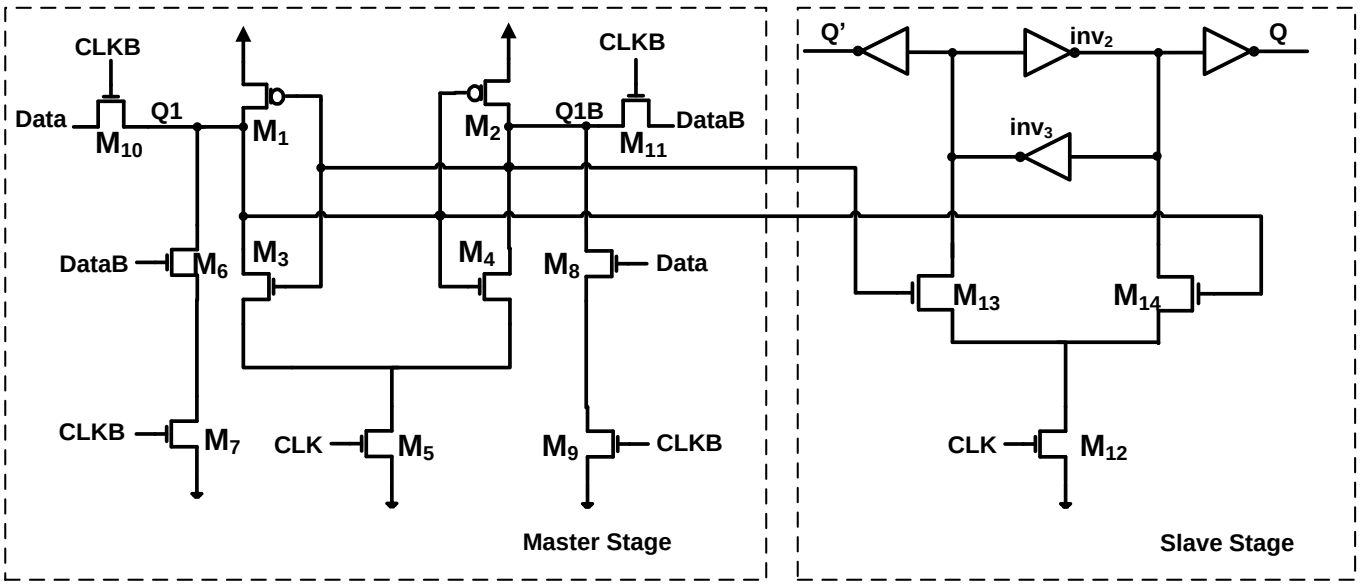


Figure 3.12: Schematic Diagram of the Sense-Amplifier Transmission-Gate Flip-Flop Design

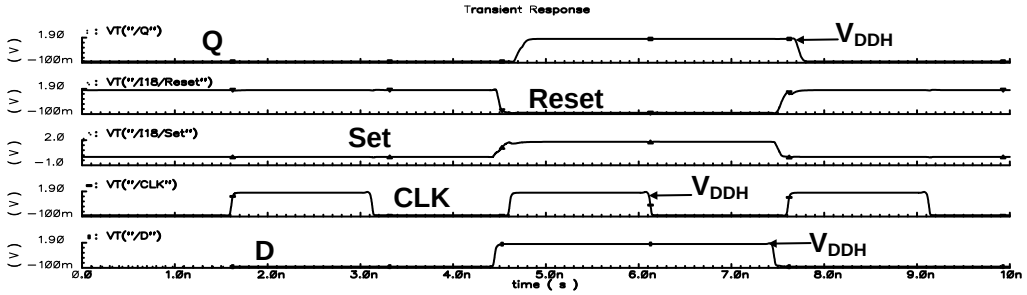
The detail operation of the SATG is given as follows. When the CLK is low, the master-stage becomes transparent as $CLKB$ turns on the NMOS-only pass transistors. If the input data D is logic "1", the differential data allows the cross-coupled inverter pair to restore the voltage swing on node Q1 to full V_{DD} instead of $V_{DD} - V_{thn}$ when only the NMOS pass transistor is present. If the input data D is logic "0", the discharge path further enhances the flip-flop performance by assisting in pulling the node Q1 to a logic "0". When the CLK becomes high, the differential signal of Q1 and Q1B will turn on either M_{13} or M_{14} while the other one is off.

Because both the CLK and D signals are only driving NMOS transistors, the architecture of the SATG also allows it to function as RCSFF and LCFF. When used as a RCSFF and LCFF, the SATG will be referred to as the reduced clock-swing SATG (RCSSATG) and the level-converting SATG (LCSATG) , respectively. While the overall PDP performance of the SATG is not as superior as the PDFF in the single-supply system, its PDP values are much more comparable in the dual-supply systems. As will be discussed in more details in the next chapter, the architecture of the SATG is very suitable for metastable-hardened flip-flop designs, especially in the dual-supply systems. **Figure 3.13** illustrates the simulated waveform for the SATG, the RCSSATG, and the LCSATG.

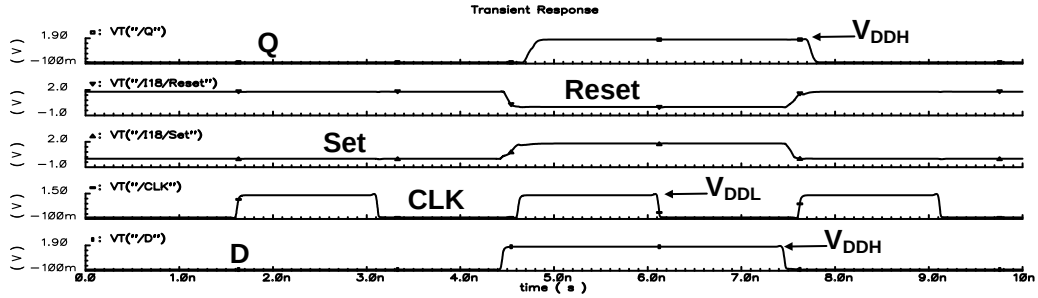
3.5 Design Methodology and Test Bench Setup

3.5.1 Design Methodology

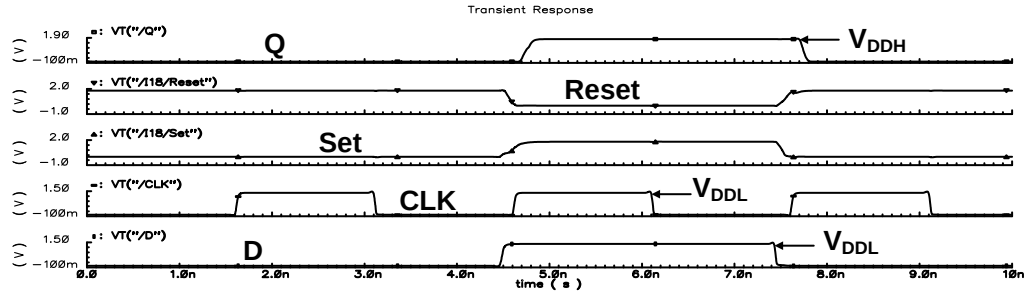
In any digital circuit designs, tradeoff always exists between delay and power. In low-power and high-performance designs, it is important to optimize both criteria. A common metric such as the power-delay-product (PDP) is often used to analyze the tradeoff between



(a) SATG



(b) RCSSATG



(c) LCSATG

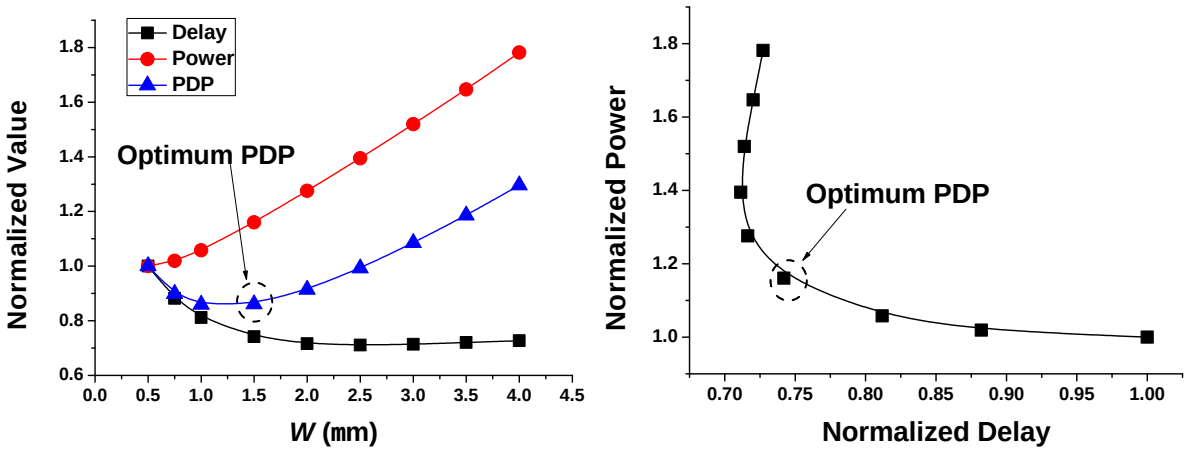
Figure 3.13: Simulation Waveforms for the SATG in Single and Dual-Supply Systems

the propagation delay and power consumption. PDP, given in **Equation (3.2)**, simply

represents the average energy consumed per switching event [62].

$$PDP = Delay \times Power \tag{3.2}$$

Because a typical flip-flop design consists of 20-30 transistors, the role of transistor sizing can result in substantial power-delay tradeoff. We use the mC^{2}MOS flip-flop as an example to illustrate such tradeoff. The architecture of the mC^{2}MOS is rather simple to analyze because it consists of clocked inverters. The feedback transistors are kept to minimum sizes while the feedforward transistors in the critical path are sized as a function of W with an aspect ratio of 1.5 between PMOS and NMOS transistors. The normalized delay, power, and PDP values is shown in **Figure 3.14(a)**.



(a) Delay, Power, and PDP as a Function of W (b) Normalized Delay vs. Normalized Power

Figure 3.14: Tradeoff between Delay and Power in Flip-Flop Design

As the value of W increases, the propagation delay initially decreases and gradually settles to a constant value when the self-loading effect becomes more dominant. The power consumption increases in an almost linear relationship with transistor width W. Because the rate of change for delay and power as a function of W is different, this results in the

minimum PDP point, which also represents the optimal energy design. **Figure 3.14(b)** illustrates the optimum PDP point typically occurs at the knee region of the delay vs. power curve. Using iterative analysis, all the flip-flops analyzed in this chapter are designed to be positive-edge triggered and sized at the optimum PDP point.

3.5.2 Test Bench Setup

The simulation test bench setup [32] is shown in **Figure 3.15**. All simulation runs are done in Cadence environment using $0.18\mu m$ TSMC CMOS bulk technology with 1.8V as the nominal supply voltage V_{DDH} at 27°C. A second supply voltage V_{DDL} is used for RCSFFs and LCFFs. The clock frequency used in the simulation is 100MHz. Input buffers are used to ensure realistic waveforms are being fed into the flip-flop. For performance measurement purposes, the inputs data D and the CLK of the flip-flops are measured at the 50% point of V_{DDH} or V_{DDL} while the rising and the falling edge of output (Q) is measured at the 50% of V_{DDH}. For fair comparison, the output buffer of each flip-flop architecture is sized identically to drive the 20fF output load that simulate the fan-out signal degradation caused by the succeeding stages.

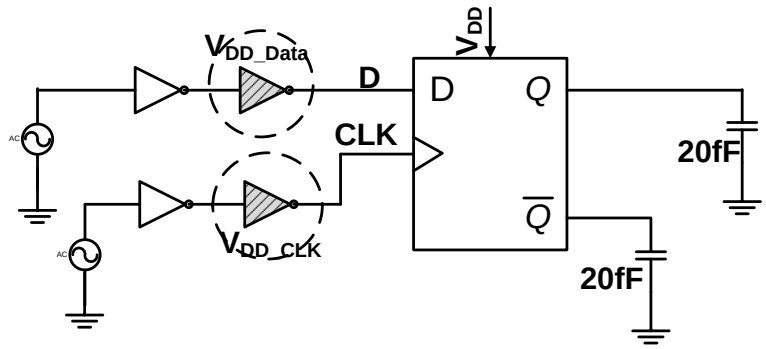


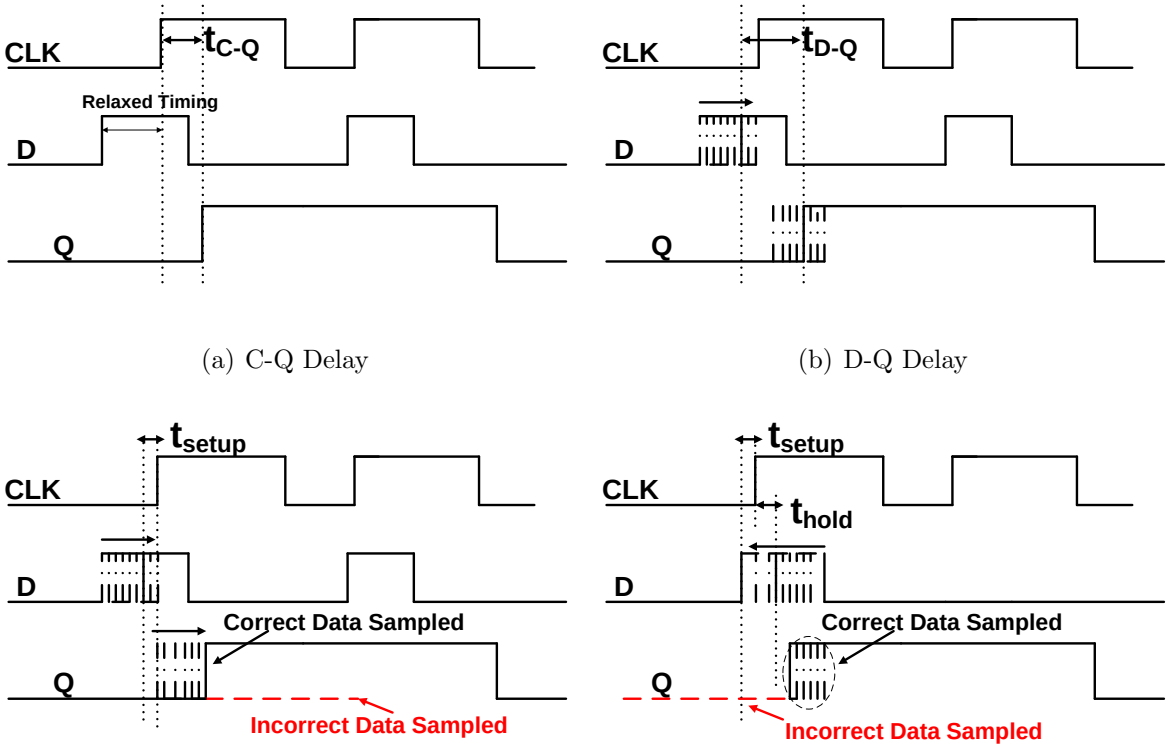
Figure 3.15: Simulation Test Bench

All the performance-related values given in this work are the worst case value of the 0-1 and 1-0 transition measured at 50% delay points. Since some of the flip-flops analyzed have negative setup time, the timing parameter that best characterizes the delay performance of a flip-flop is the minimum D-Q delay [32]. **Figure 3.16** illustrates the methodologies involved in measuring the various flip-flop timing parameters. The C-Q delay is obtained under the relaxed timing condition between the input data D and the CLK signal (**Figure 3.16(a)**). To obtain the minimum D-Q delay, the arrival time of the D with respect to the CLK is varied at an interval of 1ps (**Figure 3.16(b)**). The setup time refers to the last data arrival time when the input data D is correctly captured at the output (**Figure 3.16(c)**). The hold time is obtained by setting the data arrival time at the setup time and varying the width of the pulse to see when the output fails to sample the correct data (**Figure 3.16(d)**). The aperture window ($t_{aperture}$) of the flip-flop is calculated as the sum of the setup and hold time.

Because flip-flop architectures may exhibit different behavior under different input data pattern, four different data activity factors are considered for the analysis of power consumption: 0%, 25%, 50%, and 100%. The data activity is defined by **Equation (3.3)**.

$$DataActivity(\alpha) = \frac{\#ofSignalTransitions}{\#ofSignals \times \#ofClockCycles} \tag{3.3}$$

The power measurement of the flip-flops includes the total power dissipated in the flip-flop as well as the local data and clock power [32]. It is measured over 100 clock cycles. PDP is calculated as the product of the worst D-Q delay and the power dissipation for a given data activity.



(a) C-Q Delay

(b) D-Q Delay

(c) Setup Time

(d) Hold Time

Figure 3.16: Flip-Flop Timing Simulation Waveform

3.6 Post-Layout Simulation Results

In this section, all the analyzed flip-flops are implemented in layout using the $0.18\mu m$ technology. In general, the post-layout results do not deviate very much from the schematic simulation results with an approximately 10% degradation in terms of delay and power.

3.6.1 Flip-Flops in Single-Supply Systems

The performance characteristics of the flip-flops in single-supply supply system are listed in **Table 3.1**. We have limited the analysis to the proposed PDFF along with three other flip-flops (PowerPC, SDFF, and SAFF) because those are some of the most referenced architectures in the literature.

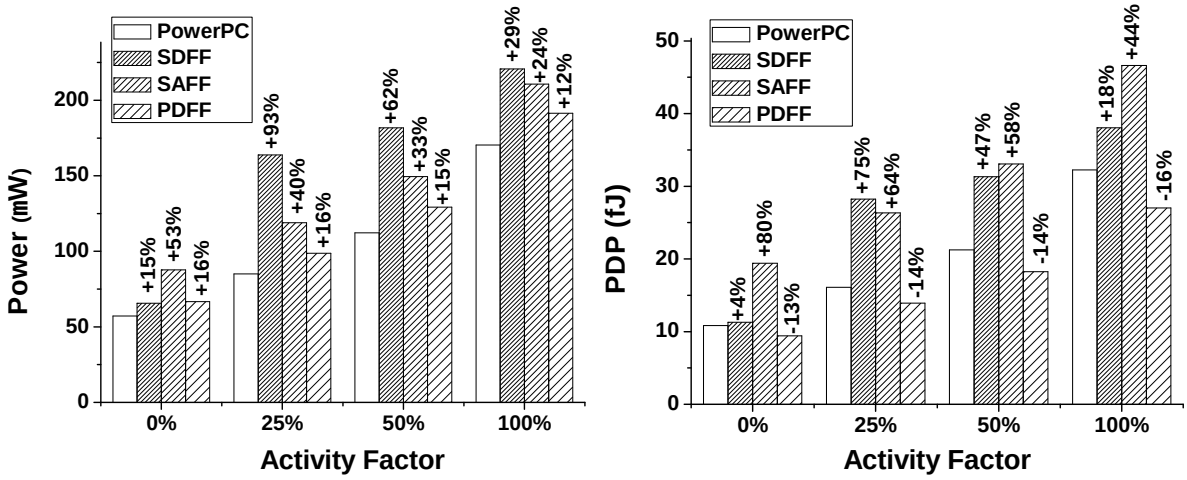
Due to the pre-charging and negative setup time characteristic, the 1-0 D-Q delay in the SDFF is considerably faster than the 0-1 D-Q delay. The cross-coupled NAND in the slave-stage of the SAFF has significantly degraded its overall D-Q delay. Overall, the reduced critical path in the PDFF has resulted in 26%, 36%, and 18% D-Q delay reduction when compared to the PowerPC, the SAFF, and the SDFF respectively. Due to its positive setup time, the aperture window ($t_{aperture}$) of the PowerPC is the smallest among all the flip-flops. When compared to the SDFF, $t_{aperture}$ of the PDFF is considerably smaller despite the negative setup time characteristic.

Table 3.1: Performance Comparison of the Single-Supply Flip-Flops

	C-Q Delay (ps)	D-Q Delay (ps)	Setup Time (ps)	Hold Time (ps)	$t_{aperture}$ (ps)
PowerPC	146.7	189.3	30.23	-38.77	52.15
SDFF	172.5	172.3	-50.28	-195.9	145.62
SAFF	209.9	221.3	-20.36	-77.66	57.3
PDFF	129.9	141.2	-18.26	-89.16	67.12

Figure 3.17(a) and **3.17(b)** illustrate the power consumption and PDP comparison of the single-supply flip-flops at different activity factors. The percentage numbers shown

in the figures indicate the relative power consumption and PDP values when compared to the PowerPC. For example, at activity factor of 50%, the SDFF, the SAFF, and the PDFF consume 62%, 33%, and 15% more power than the PowerPC, respectively.



(a) Power

(b) PDP

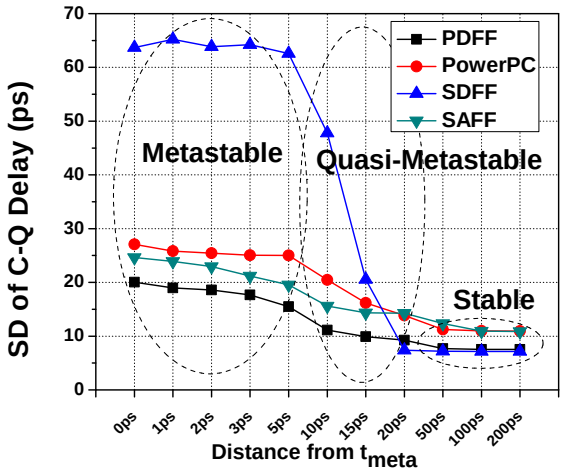
Figure 3.17: Power and PDP Comparison of Flip-Flops in Single-Supply Systems

With its low-impedance paths, the PowerPC exhibits the lowest power consumption for all data activity factors. At low data activity factor (0% and 25%), due to constant pre-charing activities during every clock cycle, the power consumption of the SDFF and the SAFF is anywhere from 40%-93% higher than the PowerPC. Those percentages have decreased significantly with an increase in the data activity factor. Despite the pre-discharging activity, a reduced critical path keeps the overall power consumption of the PDFF lower than the SDFF and the SAFF as well as approximately 16% higher than the PowerPC at all data activity factors. In terms of PDP comparison, the PDP of the PowerPC is much lower than the SDFF and the SAFF. The low-power and high-performance characteristics of the PDFF have resulted in significant PDP reduction for all data activity

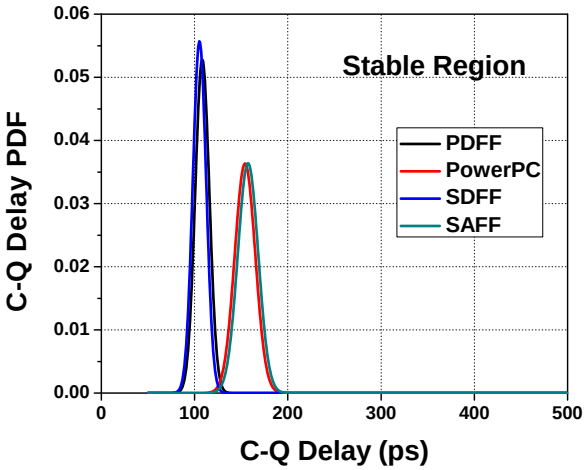
factors. For all data activity factors, the PDP of the PDFF is somewhere between 13%-16% lower than the PowerPC.

In this work, we have also analyzed the behavior of the flip-flop architectures against process variations and mismatches for the three different regions described in Section 2.1. For each flip-flop, the data arrival time in which the flip-flop fails to capture the correct data will be referred to as t_{meta}, the point where the flip-flop is very close to or at the metastable region. For each data arrival time normalized to t_{meta}, a Monte Carlo simulation of 5000 iterations with both process variations and mismatches was performed to analyze the flip-flop C-Q delay distribution. Figure 3.18(a) plots the standard deviation (SD) of the C-Q delay as a function of the normalized data-arrival time with respect to CLK for the four flip-flops analyzed. In the stable region, the SD values are very similar for all the flip-flops where the effects of the random variations and mismatches are less prevalent since the C-Q delay is independent of the data arrival time. As the data-arrival time enters the quasi-metastable and the metastable region, the C-Q delay becomes a strong function of the data arrival time. As a result, the SD value of all the flip-flops is becoming higher because a small variation in the data arrival time due to the effects of variations and mismatches can significantly change the C-Q delay. This effect is especially prominent in the SDFF where the C-Q delay is a strong function of the data arrival time due to the circuit topology and the negative setup time characteristics. Among the flip-flops analyzed, the SD of the PDFF is the lowest across all three regions of operation.

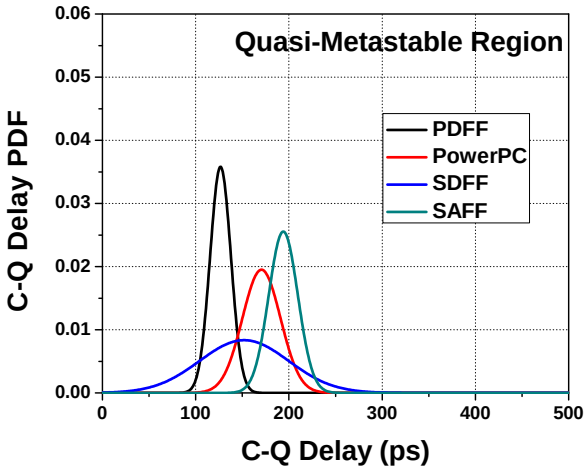
Figure 3.18(b)-3.18(d) illustrate the C-Q delay distribution of the analyzed flip-flops in the three regions of operation. Due to the large SD values, the C-Q delay distribution of the SDFF in the quasi-metastable and the metastable region is the widest among the flip-flops analyzed. Hence, extra timing margins must be provided when using the SDFF in order to meet the timing constraints in the pipeline systems by taking into account the possible



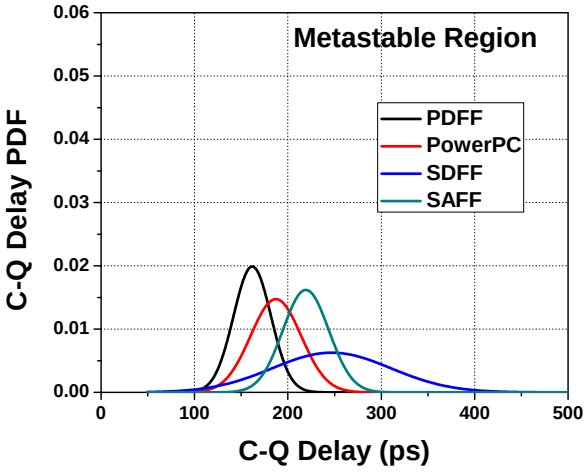
(a) Standard Deviation of C-Q Delay



(b) Delay Distribution in the Stable Region



(c) Delay Distribution in the Quasi-Metastable Region



(d) Delay Distribution in the Metastable Region

Figure 3.18: Comparison of Flip-Flop Robustness against Process Variations and Mismatches

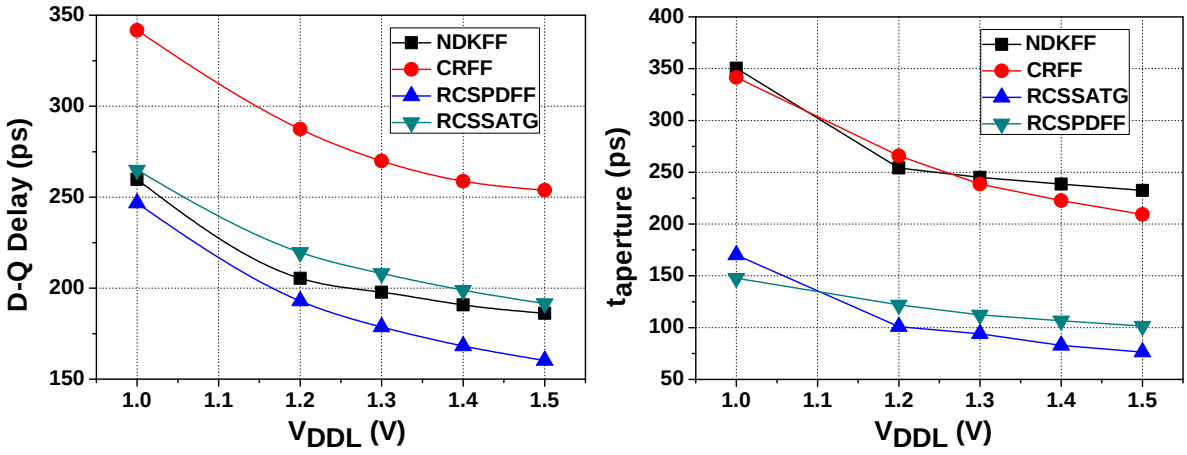
delay variations caused by process variations and transistor mismatches. On the other hand, the PDFF has the smallest SD values and the narrowest C-Q distribution across all three regions. Overall, it demonstrates the best robustness against random process variations and mismatches with less susceptibility in violating the setup and hold time requirements that may result in metastable conditions.

3.6.2 Reduced Clock-Swing Flip-Flops

The performance characteristics of all the reduced clock-swing flip-flops are listed in **Table 3.2**. All the values listed are obtained for $V_{DDL} = 1.3V$, which is approximately equal to $0.7V_{DDH}$ in the $0.18\mu m$ technology. In the RCSPDFF, it is evident that the high-performance characteristic of the PDFF architecture is also extended to the reduced clock-swing flip-flops. The D-Q delay of the RCSPDFF is 13%, 14%, and 34% lower than the NDKFF, the RCSSATG, and the CRFF, respectively. While the D-Q delay of the RCSSATG and the NDKFF is approximately the same, the good performance of the NDKFF comes at the expense of high $t_{apeture}$ due to the negative setup time. In fact, the proposed flip-flops, the RCSPDFF and the RCSSATG, have much lower $t_{aperture}$ values than the NDKFF and the CRFF. Without the usage of a clock pulse generator, the SATG requires a much lower hold time than the other flip-flops, and thus resulting in a smaller $t_{aperture}$ value. While the RCSPDFF uses a clock pulse, it is only present to allow negative setup time and soft-edge property, and has no significant impact on the overall flip-flop performance. Because the clocked transistor is not in the critical path, the hold time required in the RCSPDFF is not as large as the other pulsed-triggered flip-flops. By contrast, both the NDKFF and the CRFF have transistors in the critical path that are controlled by the clock pulse signals, and thus results in a much higher $t_{aperture}$ value.

Table 3.2: Performance Comparison of the Reduced Clock-Swing Flip-Flops at $V_{DDL} = 1.3V$

	C-Q Delay (ps)	D-Q Delay (ps)	Setup Time (ps)	Hold Time (ps)	$t_{aperture}$ (ps)
NDKFF	197.8	205.4	-57.38	-326.3	240.13
CRFF	230.9	269.9	8.574	-258.3	238.47
RCSPDFF	177.4	178.6	-42.06	-156.58	112.5
RCSSATG	120.3	208.1	39.78	-54.22	94



(a) D-Q Delay

(b) $t_{aperture}$

Figure 3.19: D-Q Delay and $t_{aperture}$ Comparison of the Reduced Clock-Swing Flip-Flops

In dual-supply systems, it is also important to analyze the flip-flop characteristics across various V_{DDL} values. **Figure 3.19(a)** and **3.19(b)** illustrate the D-Q delay and $t_{aperture}$ of the different RCSFFs for V_{DDL} values ranging from 1V-1.5V. Due to its unique architecture,

the D-Q delay of the RCSPDFF is the lowest among all the flip-flops across all supply voltages. In the critical path of the RCSPDFF, the reduced clock-swing signal CLK passes through the PMOS transistor (M_5 or M_6) in the master-stage and propagate to the slave-stage to turn on either NMOS transistor M_9 or M_{10}. Thus only one transistor in the critical path is affected by the reduced clock-swing. In contrast, the NDKFF, the CRFF, and the RCSSATG all have two clocked transistors in the critical path. The CRFF has the highest D-Q delay because a reduced clock-swing further degrades the performance of the NMOS-pass transistor in the critical path. Despite its negative setup time, the D-Q delay of the NDKFF is only slightly better than the RCSSATG largely due to the stacking of three NMOS transistors in the critical path.

As evident from the figure, $t_{aperture}$ of all the flip-flops increases with the reduction in the clock-swing. At lower clock-swings, the clock pulse generated in the NDKFF, the CRFF, and the RCSPDFF is becoming larger due to the slower propagation in the inverter chain. While this allows for more negative setup time, the hold time required for the flip-flops is also becoming greater. In general, an overall increase in $t_{aperture}$ suggests the increase in the hold time is greater than the setup time. As in the case when $V_{DDL} = 1.3V$, the $t_{aperture}$ of the RCSPDFF and the RCSSATG is much lower than those of the CRFF and the NDKFF. A smaller $t_{aperture}$ value reduces the likelihood of the flip-flops entering metastability, and thus improves the flip-flop reliability.

Table 3.3 and **3.4** show the power consumption and PDP of the different RCSFFs at four different data activity factors at $V_{DDL} = 1.3V$. With fewer transistors in the critical path, the power consumption of the RCSPDFF is the lowest among the flip-flops analyzed except when there is no data activity. Despite the poor performance, the additional circuitry that the CRFF employs to reduce the amount of contention at reduced clock-swings result in much lower power consumption than the NDKFF and the RCSSATG at most of

the data activity factors. For data activity factor of 100%, the reduced-swing signals from the master-stage in the RCSSATG weakens the discharge paths in the slave-stage. This in turn has resulted in more power dissipation due to the contention in the cross-coupled inverter. At lower data activity factors ($\leq 25\%$), however, the power consumption of the RCSSATG is very much comparable to those of the CRFF and the RCSPDFF.

Table 3.3: Power Comparison of the Reduced Clock-Swing Flip-Flops at $V_{DDL} = 1.3V$

	$\alpha = 0\%$ (μW)	$\alpha = 25\%$ (μW)	$\alpha = 50\%$ (μW)	$\alpha = 100\%$ (μW)
NDKFF	37.74	74.13	109.43	180.43
CRFF	30.163	62.343	92.65	154.89
RCSPDFF	34.68	62.24	88.66	142.26
RCSSATG	21.886	65.027	106.99	191.578

The PDP of the RCSPDFF achieves a minimum of 18%, 27%, and 29% reduction from the other flip-flops for data activity factor of 25%, 50%, and 100%, respectively. Because of the smaller power consumption at lower data activity factors, the PDP of the RCSSATG is 8% and 20% lower than the NDKFF and the CRFF for 25% data activity factor. For data activity factor of zero, the PDP of the RCSSATG is 39% and 44% lower than the NDKFF and the CRFF.

Figure 3.20(a) and **3.20(b)** illustrate the power consumption and PDP of the different RCSFFs for V_{DDL} values ranging from 1V-1.5V at a data activity factor of 25%. We have chosen a low data activity factor for analysis because static logic typically has an activity factor close to 10% [61]. Generally, the power consumption of the CRFF and the RCSPDFF

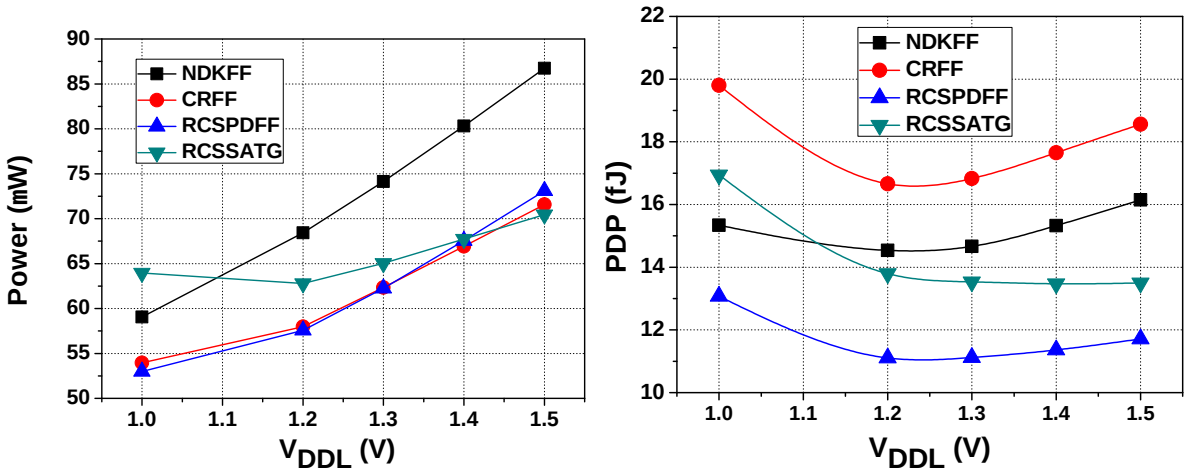
Table 3.4: PDP Comparison of the Reduced Clock-Swing Flip-Flops at $V_{DDL} = 1.3V$

	$\alpha = 0\%$ (fJ)	$\alpha = 25\%$ (fJ)	$\alpha = 50\%$ (fJ)	$\alpha = 100\%$ (fJ)
NDKFF	7.465	14.662	21.645	35.689
CRFF	8.141	16.826	25.006	41.805
RCSPDFF	6.194	11.116	15.835	25.408
RCSSATG	4.554	13.532	22.265	39.867

is approximately 15% lower than the NDKFF across all voltages. At higher V_{DDL} values, the power consumption of the RCSSATG is very similar to those of the CRFF and the RCSPDFF. As V_{DDL} is reduced, however, the power dissipation due to node contention in the RCSSATG offsets the the power reduction resulting from a reduced clock signal. In fact, the minimum power consumption for the RCSSATG occurs when $V_{DDL} = 1.2V$. For a data activity factor of 25%, the PDP of the RCSPDFF is at least 13% lower than the other flip-flops for all V_{DDL} values. The PDP of the RCSSATG is lower than those of the CRFF and the NDKFF for $V_{DDL} \geq 1.2V$. The high delay values of the CRFF have resulted in a much higher PDP values than the other flip-flops. The minimum PDP point of all the flip-flops occurs when $V_{DDL} = 1.2V$, which coincides with the previous studies stating that V_{DDL} should be around $0.7V_{DDH}$ for optimum PDP operation.

3.6.3 Level-Converting Flip-Flops

The performance characteristics of all the level-converting flip-flops are listed in **Table 3.5** for $V_{DDL} = 1.3V$. Like its PDFF counterparts, the LCPDFF also demonstrates the best



(a) Power

(b) PDP

Figure 3.20: Power and PDP Comparison of the Reduced Clock-Swing Flip-Flops for 25% Data Activity Factor

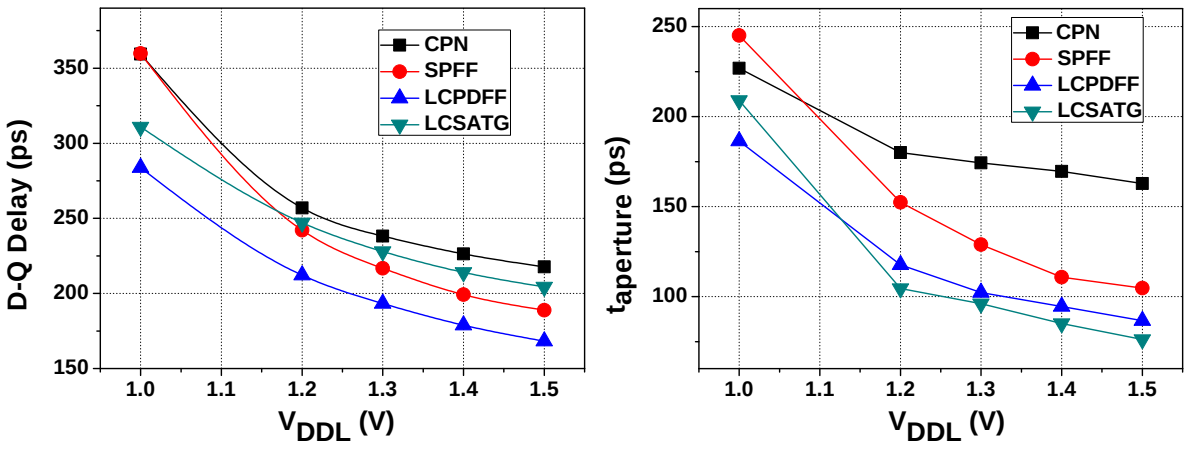
performance among all the LCFFs analyzed in this work. In fact, the D-Q delay of the LCPDFF is 15%, 11%, and 19% lower than the LCSATG, the SPFF, and the CPN, respectively. Since both the CPN and the SPFF employ the technique of conditional capturing and negative setup time, extra hold time on the input data is required to ensure the output makes the correct transition and consequently turn off the corresponding discharge paths. Therefore, both the CPN and the SPFF have higher $t_{aperture}$ values than the LCPDFF and the LCSATG.

Figure 3.21(a) and **3.21(b)** illustrate the D-Q delay and $t_{aperture}$ of the different LCFFs for V_{DDL} values ranging from 1V-1.5V. Overall, the LCPDFF is at least 11% faster in D-Q delay than the other LCFFs. As for $t_{aperture}$, both the CPN and the SPFF have higher values than the LCPDFF and the LCSATG across all V_{DDL} values. Although the architecture of the LCPDFF and the LCSATG is identical to those for the RCSPDFF and

Table 3.5: Performance Comparison of the Level-Converting Flip-Flops at $V_{DDL} = 1.3V$

	C-Q Delay (ps)	D-Q Delay (ps)	Setup Time (ps)	Hold Time (ps)	$t_{aperture}$ (ps)
SPFF	209.8	216.8	-48.55	-193.2	128.85
CPN	242.6	238.2	-66.75	-277.5	174.3
LCPDFF	185.8	193.3	-28.27	-130.5	102.23
LCSATG	133.3	228.0	43.01	-57.03	100.04

the RCSSATG, the additional reduced swing on the input data D has resulted in a slight increase in both the D-Q delay and $t_{aperture}$.



(a) D-Q Delay

(b) $t_{aperture}$

Figure 3.21: D-Q Delay and $t_{aperture}$ Comparison of the Level-Converting Flip-Flops

Table 3.6 and **3.7** shows the power consumption and PDP of the different LCFFs at four different data activity factors at $V_{DDL} = 1.3V$. By employing the conditional capturing

technique, the power consumption of the CPN and the SPFF at low data activity factor is similar to those of the LCPDFF and the LCSATG. For data activity factor greater than 50%, the power consumption of the LCPDFF is at least 13% lower than the rest of the flip-flops. The performance advantage of the LCPDFF has resulted in the lowest PDP values for data activity factor greater than 25%. The PDP values of the LCSATG are very similar to the CPN and the SPFF for all data activity factors.

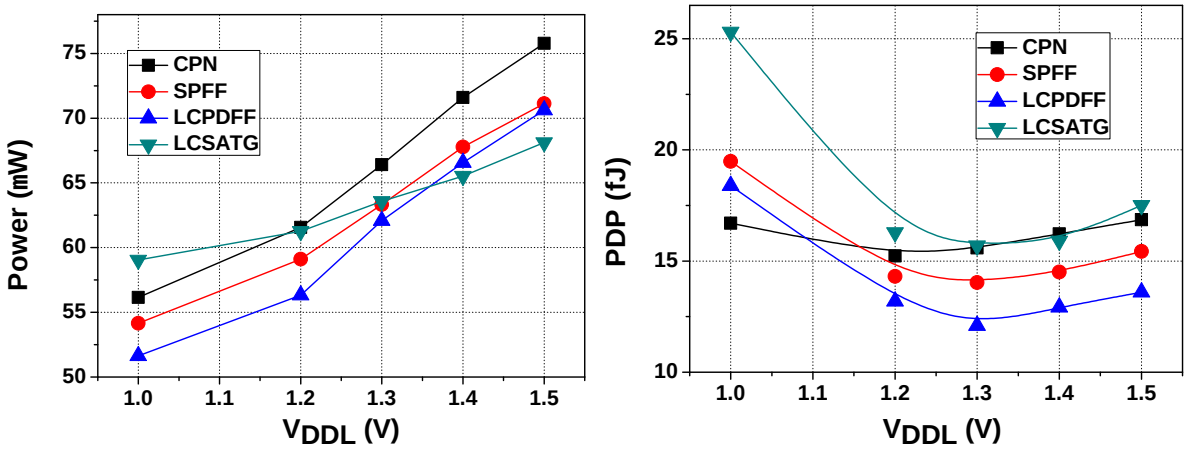
Table 3.6: Power Comparison of the Level-Converting Flip-Flops at $V_{DDL} = 1.3V$

	$\alpha = 0\%$ (μW)	$\alpha = 25\%$ (μW)	$\alpha = 50\%$ (μW)	$\alpha = 100\%$ (μW)
SPFF	24.684	63.331	107.64	169.62
CPN	32.275	66.405	100.94	167.62
LCPDFF	36.037	62.078	87.028	116.73
LCSATG	23.605	63.573	102.415	159.75

Table 3.7: PDP Comparison of the Level Converting Flip-Flops at $V_{DDL} = 1.3V$

	$\alpha = 0\%$ (fJ)	$\alpha = 25\%$ (fJ)	$\alpha = 50\%$ (fJ)	$\alpha = 100\%$ (fJ)
SPFF	5.351	13.73	23.337	36.774
CPN	7.575	15.585	23.691	39.34
LCPDFF	7.254	12.496	17.519	23.498
LCSATG	5.382	14.495	23.351	36.423

Figure 3.22(a) and **3.22(b)** illustrate the power consumption and PDP of the different LCFFs for V_{DDL} values ranging from 1V-1.5V at a data activity factor of 25%. For $V_{DDL} \geq 1.3V$, the power consumption of the CPN is approximately 5% higher than the SPFF, the LCSATG, and the LCPDFF. As V_{DDL} is reduced below 1.2V, the power consumption of the LCPDFF becomes the lowest among all the flip-flops analyzed. Once again, the PDP of the LCPDFF is the lowest for all the V_{DDL} values except when it reaches 1V.



(a) Power

(b) PDP

Figure 3.22: Power and PDP Comparison of the Level-Converting Flip-Flops for 25% Data Activity Factor

3.7 Summary

In this chapter, we examined the architectures and characteristic of various flip-flops in both single and dual-supply systems. We also proposed two new flip-flops, namely the pre-discharge flip-flop (PDFF) and the sense-amplifier-transmission-gate (SATG) flip-flop. The

PDFF achieves very high-performance by adopting a pre-discharge scheme. The SATG uses a sense-amplifier structure along with NMOS pass transistors in the master-stage instead of the traditional pre-charging scheme used in the SAFF, and thus achieves low-power consumption at low data activity factors. The architecture of these flip-flops facilitate the usage in both single and dual-supply systems. A detailed comparison between various flip-flop architectures is performed in terms C-Q delay, D-Q delay, setup and hold time, $t_{aperture}$, power consumption, and power-delay-product (PDP). The high-performance and low-power characteristics of the PDFF have been demonstrated in both single and dual-supply systems. The overall D-Q delay, power consumption, and PDP of the PDFF, the RCSPDFF, and the LCPDFF are much lower than most of the previously proposed flip-flops analyzed in this work. The overall D-Q delay, power consumption, and PDP of the RCSSATG and the LCSATG have also been very much comparable to the analyzed flip-flops in the dual-supply systems. Both proposed flip-flops have shown to have smaller $t_{aperture}$ values, which increases the flip-flop reliability by reducing the likelihood of entering metastability.

Chapter 4

Design and Analysis for Metastable-Hardened, High-Performance, Low-Power Flip-Flops

In this chapter, a detailed analysis and methodologies on designing flip-flops with improved metastability performance while maintaining high-performance and low-power are presented. We will use the term "metastable-hardened" when referring to flip-flops that are less susceptible to metastability by having improved design parameters such as reduced τ. Past flip-flop designs have mainly focused on optimizing the tradeoff between performance and power consumption by designing for optimum PDP through transistor sizing. Of the various flip-flop architectures proposed in today's VLSI systems, a more detailed and in-depth analysis on the flip-flop metastable behavior is largely absent. Us-

ing the fundamental metastability modeling theories, both qualitative and quantitative analysis are provided to demonstrate that flip-flop metastability can be varied accordingly based on transistor sizing. Theoretical calculations will demonstrate the proposed sizing methodology will have a dramatic impact on the value of the time-resolving constant τ. New design metrics called the metastability-delay-product (MDP) and the metastability-power-delay-product (MPDP) are introduced to illustrate the various tradeoffs in flip-flop designs between delay, power, and metastability. The analysis is performed for selected flip-flops in both single and dual-supply systems. In keeping with recent trends of green energy and low-power VLSI designs, flip-flop metastability in the sub-threshold region will be discussed and analyzed. We also examine the impact of technology scaling on τ for technologies below the 65nm regime. Finally, the implementation of an all-digital on-chip metastability measurement circuit will also be given in this chapter as well.

4.1 General Design Methodology

In edge-triggered flip-flops, input data is captured by an intermediate critical node in the master-stage before it is propagated to the output through the slave-stage. The critical nodes that potentially cause metastability due to synchronization of the CLK and the input D signals are stabilized by some form of cross-coupled inverter pair shown in **Figure 2.12**. While T_0 is important to determine the metastability window δ and the MTBF of a flip-flop, the impact of τ is far more greater due to the exponential term in **Equation (2.5)** and **(2.6)**. Hence, the metastability analysis in this work mainly focuses on the optimization of τ where the small signal modeling described in Chapter 2 forms the foundation for the analysis of τ in various flip-flops architectures because each parameter in **Equation (2.10)** can be represented as a function of the transistor width W. As a simple first-

order approximation, the transconductance, diffusion (C_{diff}) and gate capacitance (C_g) of a MOSFET device are given in **Equation (4.1)-(4.3)** [62].

$$g_m = k' \frac{W}{L}(V_{gs} - V_t) \tag{4.1}$$

$$C_g = \frac{2}{3} C_{ox} W L + 2 C_{go} W \tag{4.2}$$

$$C_{diff} = C_j L_s W + C_{js}(2L_s + W) \tag{4.3}$$

The transconductance g_m and the capacitance C_{Crit} associated with the critical nodes are functions of the flip-flop circuit topology, and thus result in different time-resolving constant τ. As such, τ can be varied through transistor sizing for a given flip-flop architecture. Based on **Equation (4.1)**, it is desirable to have large transistor widths to increase g_m in the inverter pair while the width of the load transistors should be kept small to minimize the value of C_{Crit}. The contribution of C_{Crit} at the critical node mainly comes from two different sources: *(i)* the Miller capacitances (C_M) associated with the cross-coupled inverter, and *(ii)* the lumped capacitance (C_Q), which includes all the gate and diffusion capacitances associated with the critical node from both the master and the slave-stage. Because of this, continuous width increase in the inverter pair does not further reduce τ as any increase in g_m is offset by the increase in capacitance C_M. While there may be many transistors associated with the critical node, only those non-minimum transistor sizes will be considered for the analysis of C_Q. In most cases, these transistors will also have a significant impact on the performance of the flip-flops. In our analysis, the value of C_M is considered part of the transconductance g_m variation while C_Q specifically refers to the variation of the load transistors associated with the critical node. **Figure 4.1** illustrates the general design methodology for the metastable-hardened flip-flops. Because each flip-flop may have its own unique architecture, it is sometimes difficult to clearly identify a master and slave stage. In cases where a master and a slave stage can be clearly identified,

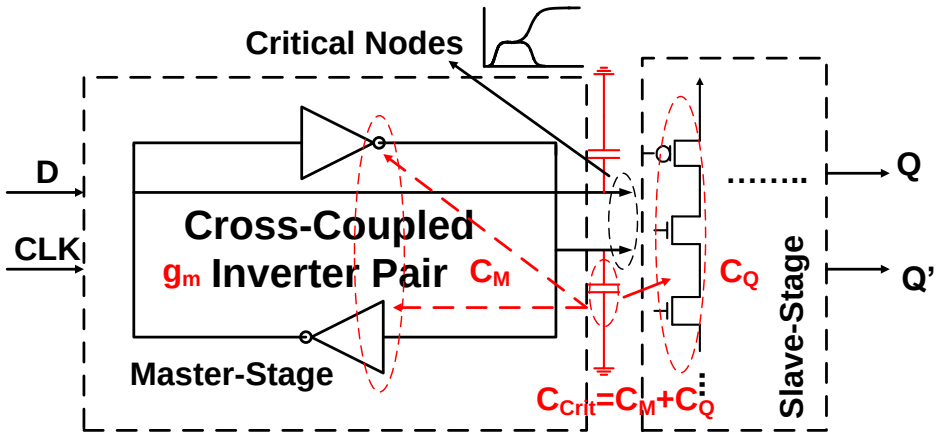


Figure 4.1: Conceptual Diagram of Metastable-Hardened Flip-Flop Design

the sizing variation of the inverter pair is performed in the master-stage with the variation of the load transistors coming both from the master and the slave-stage. The reason we chose to vary g_m in the master-stage is because that is where the initial synchronization occurs. If τ is improved in the master-stage, the probability of metastability happening in the slave-stage can be reduced significantly due to more settling time despite another synchronization with the CLK signal occurs in some flip-flop topologies [22]. When no clear-cut master and slave-stage is present in the flip-flop architecture, τ is varied by simply changing the size of the cross-coupled inverter that stabilizes the critical node which causes contention or changing the size of the load transistors associated with the critical node. Either way, the design and analysis methodology is identical in both cases.

While the proposed designs of the SATG and the PDFF described in Chapter 3 demonstrate the characteristics of low-power and high-performance, their circuit topologies are even more attractive for metastable-hardened flip-flops designs based on the following reasonings. First of all, the $t_{aperture}$ of these flip-flops is significantly smaller than the other analyzed flip-flops in both the single and the dual-supply systems. As previously men-

tioned, a smaller $t_{aperture}$ means the flip-flops are less susceptible to violating the setup and hold time constraints that may result in metastability. With respect to circuit topology, both flip-flops adopt a cross-coupled inverter structure in the master-stage and a small load transistor in the slave-stage. By having the cross-coupled inverter pair on the critical path in the master-stage, transistors can be sized up to increase the transconductance g_m in the loop pair while maintaining high-performance and correct functionality. Furthermore, both proposed flip-flops have similar slave-stage topology to minimize the load capacitance such that the critical nodes from the master-stage only drives a single NMOS transistor. According to Equation (2.10), both of these features are able to reduce the time-resolving constant τ dramatically by minimizing the capacitance terms in the numerator of the equation and increasing the transconductance term in the denominator of the equation. While the τ of the PDFF can be significantly reduced in single-supply systems, it will actually increase in an exponential manner when working as the RCSPDFF or the LCPDFF in the dual-supply systems because the reduced clock-swing is connected to the drain of the PMOS transistors. Therefore, the RCSSATG and the LCSATG are more suitable for metastable-hardened flip-flop designs in the dual-supply systems as V_{DDL} is reduced.

In the rest of this chapter, the metastability of selected flip-flop architectures described in Chapter 3 will be analyzed, and are listed below.

- Single-Supply Flip-Flops
  - PowerPC, SDFF, SAFF, PDFF
- Dual-Supply Flip-Flops
  - Reduced Clock-Swing Flip-Flops
    - NDKFF, CRFF, RCSPDFF, RCSSATG

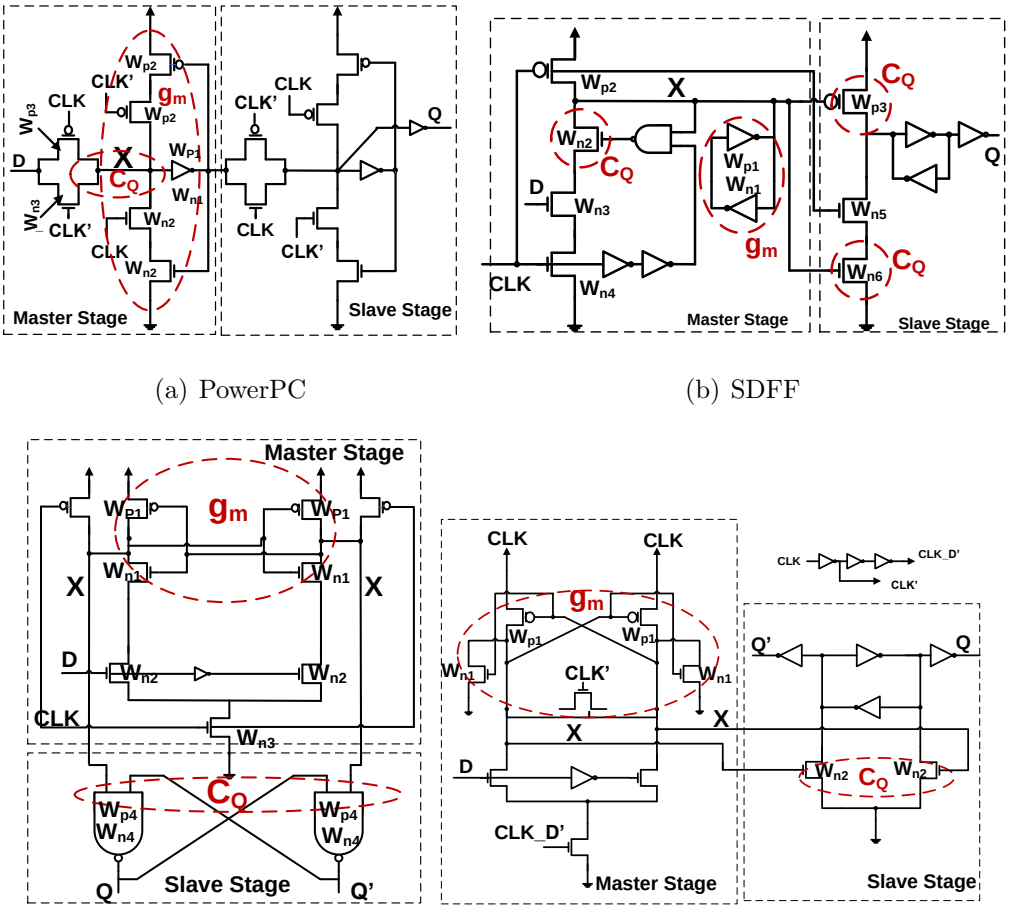
- Level-Converting Flip-Flops
  - CPN, SPFF, LCPDFF, LCSATG

The schematic diagrams of all the flip-flops analyzed for metastability are shown in **Figure 4.2**, **4.4**, and **4.5**. The critical node of each flip-flop architecture is marked by "X" in the schematics. The corresponding transistors that are relevant to flip-flop metastability either through transconductance or load variation have been highlighted by g_m and C_Q on the figures respectively. Due to the identical topology with the PDFF, the schematic diagrams of the RCSPDFF and the LCPDFF are not shown but will be analyzed quantitatively. Similarly, the schematic diagram of the LCSATG is not shown because its topology is identical to the RCSSATG.

4.2 Qualitative Analysis of Flip-Flop Metastability

4.2.1 Flip-Flops in Single-Supply System

In the PowerPC (**Figure 4.2(a)**), the critical node "X" is chosen right after the transmission-gate in the master-stage due to the initial synchronization of the CLK and the input data D. It is stabilized by a CLK-controlled feedback inverter as well as a forward inverter on the critical path. Due to the topology, the transistors in the forward inverter (W_{p1} and W_{n1}) is sized up to maintain high-performance on the critical path. To improve metastability, however, the size of the feedback transistors (W_{p2}, W_{n2}) must be sized up to increase g_m in the cross-coupled inverter pair. Alternatively, the size of the transmission-gate transistors (W_{p3}, W_{n3}) can also be manipulated to obtain different τ values by changing the load capacitance of the critical node. The waveform that demonstrates the contention at node "X" during the metastable period for the PowerPC is shown in **Figure 4.3(a)**.



(a) PowerPC (b) SDFF

(c) SAFF (d) PDFF

Figure 4.2: Schematic Diagram of Single-Supply Flip-Flops for Metastability Analysis

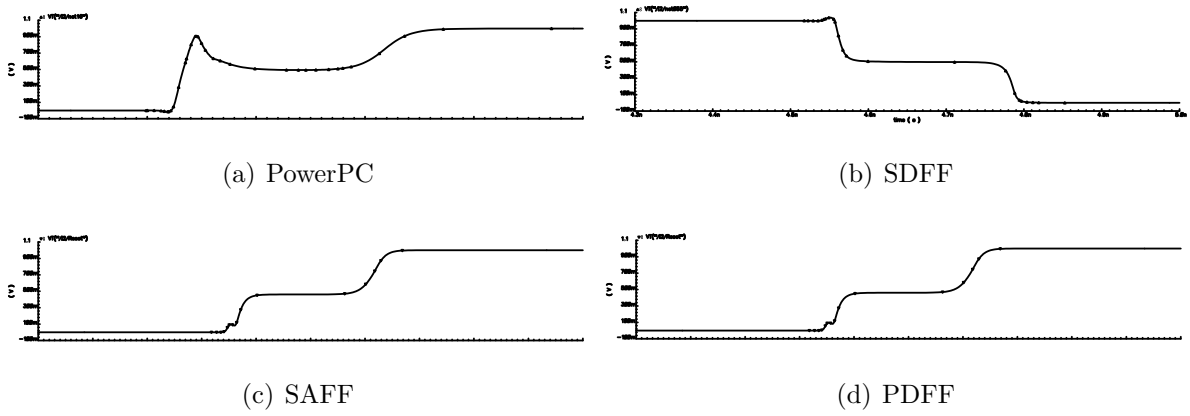
In the SDFF (**Figure 4.2(b)**), metastability occurs because input data is allowed to transition after the rising edge of the CLK, which in turn causes contention at the critical node "X". Such contention is more prominent when input data makes a 1-0 transition. Initially after the rising edge of the CLK, the input data D does not make the 1-0 transition because of the negative setup time, and hence node "X" is falsely discharged until either

the data makes the 1-0 transition or the NAND gate produces a logic "0" which would turn W_{n2} off. Due to the semi-dynamic nature of the SDFF, once node "X" is falsely discharged, the only mechanism that can restore it back to logic "1" is through the weakly sized cross-coupled inverter pair. Thus, the extreme fast performance in the 1-0 output delay of the SDFF comes at the expense of very poor metastability. This is a classic example of demonstrating flip-flop metastability behavior for a given output transition largely depends on the circuit architecture instead of their propagation delay. The metastability of the SDFF can be improved by (i) increasing the size of the cross-coupled inverter pair (both size to W_{p1} and W_{n1}) that stabilizes the critical node "X" or (ii) reduce the transistor size surrounding the critical node to minimize the associated capacitance. The waveform that demonstrates the contention at node "X" during the metastable period for the SDFF is shown in **Figure 4.3(b)**.

The critical node (s) "X" in the sense-amplifier flip-flop (SAFF, **Figure 4.2(c)**) are being pre-charged to logic "1" when the CLK phase is low and stabilized during the evaluation period by a cross-coupled inverter pair formed by transistor W_{p1} and W_{n1}. PMOS transistor W_{p1} is typically designed to have smaller sizes while W_{n1} is on the critical path of the flip-flop and thus sized up to achieve high-performance. To improve τ, however, the size of W_{p1} must be increased to enhance the overall g_m value. Because the transconductance in the master-stage is limited by the fact that path to V_{ss} is formed by three NMOS transistors in series, a more effective method to improve τ is to reduce the transistor size of the NAND gates (W_{p4} and W_{n4}) in order to minimize the amount of capacitance the critical signals drive in the slave-stage. The waveform that demonstrates the contention at node "X" during the metastable period for the SAFF is shown in **Figure 4.3(c)**.

When the CLK is low, the critical nodes "X" in the PDFF have been pre-discharged to logic "0". During the evaluation phase, the cross-coupled inverter pair formed by transistor

W_{p1} and W_{n1} reduces the contention during the synchronization of the CLK and the input data D signal. The load transistor of the PDFF consists of only a single NMOS transistor in the slave-stage. The PDFF demonstrates good metastability because *(i)* a cross-coupled inverter formed by W_{p1} and W_{n1} in the master-stage can be sized up to increase the transconductance without sacrificing much performance, and *(ii)* the load "X" drives in the slave-stage is only a single NMOS transistor. The waveform that demonstrates the contention at node "X" during the metastable period for the PDFF is shown in **Figure 4.3(d)**.



(a) PowerPC (b) SDFF

(c) SAFF (d) PDFF

Figure 4.3: Metastable Contention Nodes for Single-Supply Flip-Flops

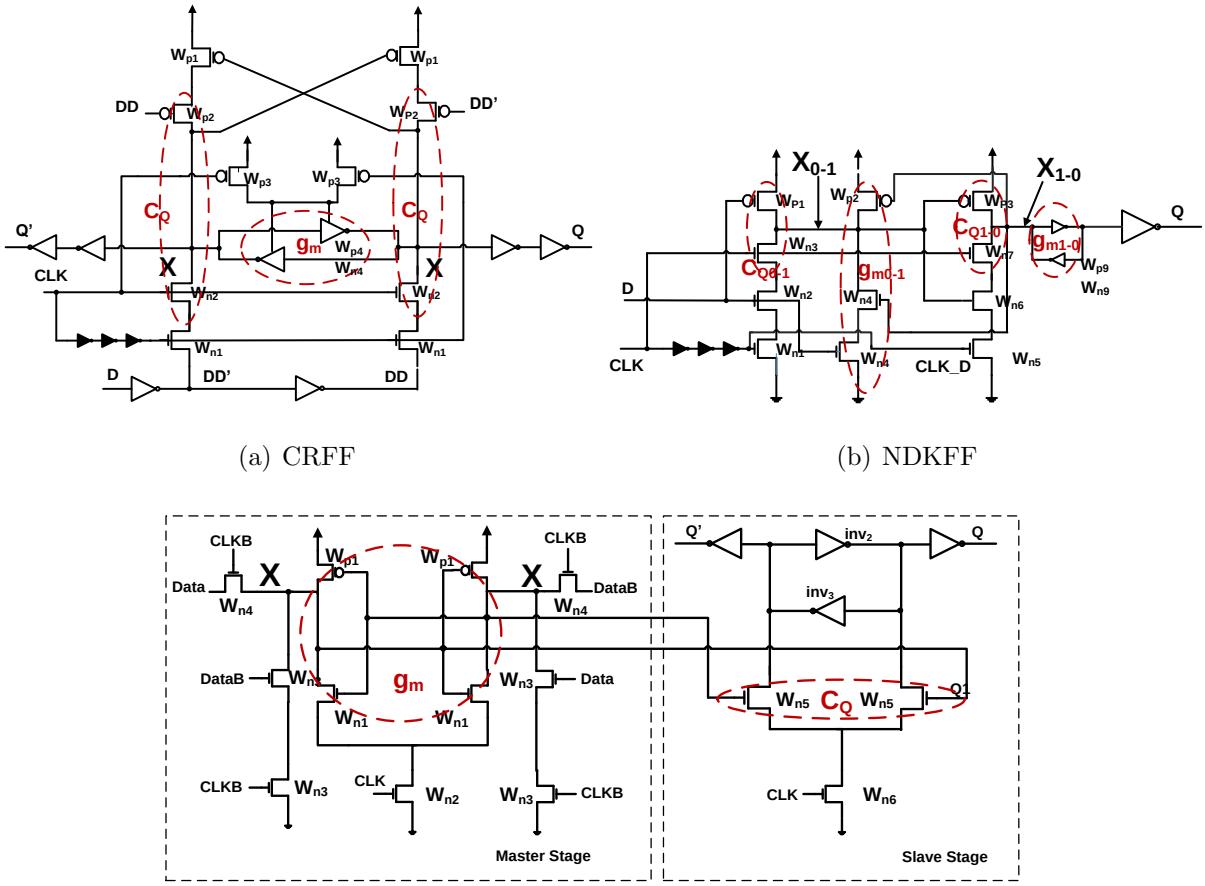
4.2.2 Flip-Flops in Dual-Supply System

In the CRFF (**Figure 4.4(a)**), because data is written into the flip-flop via the NMOS-only pass transistors, its critical nodes "X" are chosen right after the second pass transistor W_{n2} in the critical path. A cross-coupled inverter pair formed by transistor W_{p4} and W_{n4} can be sized up to increase the transconductance while reducing τ. The load transistors

of the cross-coupled inverter pair consists of transistor W_{n2} and W_{p2}. The waveform that demonstrates the contention at node "X" during the metastable period for the CRFF is shown in **Figure 4.6(a)**.

The critical node for metastability in the NDKFF (**Figure 4.4(b)**) is different for 0-1 and 1-0 data transition. In the 0-1 input data transition, the node "X_{0-1}" is under contention because the feedback PMOS transistor W_{p2} is turned on initially by node "X_{1-0}" and fights with the discharge path of the stacked NMOS transistors W_{n1}, W_{n2}, and W_{n3}. Around the metastable region, this contention can last for a long time because W_{p2} is turned off only when "X_{1-0}" completes the 0-1 transition. To improve metastability for 0-1 data transition, *(i)* larger feedback transistors W_{p2} and W_{n4} can be used to increase the transconductance in the inverter pair formed by the feedback transistors along with W_{p3} and W_{n7}, and *(ii)* decrease the size of the load transistor W_{p1} and W_{n3}. The same contention in 0-1 data transition does not exist for the 1-0 input data transition because W_{n4} cuts off the contention path. The critical node in this case is "X_{1-0}" where it takes a certain amount of time to settle to stable value mainly due to the closing of the transparency window from the falling edge of "CLK_D". In this case, the contention at the critical node can be reduced by either sizing up transistor W_{p9} and W_{n9} in the inverter pair or reduce size of the load transistors such as W_{p2}, W_{n4}, W_{p3}, and W_{n7}. The waveform that demonstrates the contention at node "X" during the metastable period for the NDKFF is shown in **Figure 4.6(b)** and **4.6(c)** for 0-1 and 1-0 data transition.

In the RCSSATG, the critical nodes "X" are chosen right after the NMOS-only pass transistor due to the initial synchronization between the CLK and the input data D. Similar to the PowerPC, the metastable period is prolonged because the input data is allowed to pass through via the low-impedance path. The two additional discharge paths improve the setup time of the flip-flops but do not enhance metastability. The g_m in the



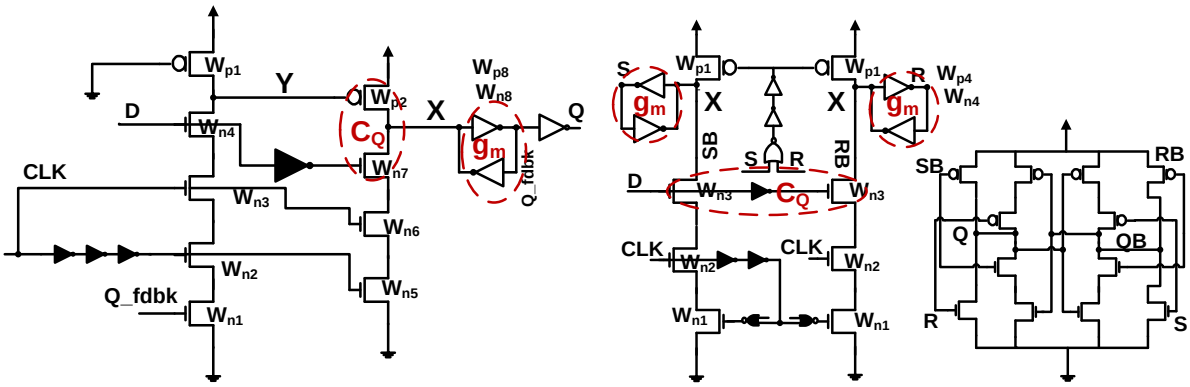
(a) CRFF (b) NDKFF

(c) RCSSATG

Figure 4.4: Schematic Diagram of Reduced Clock-Swing Flip-Flops for Metastability Analysis

cross-coupled inverter pair can be manipulated by varying the size of transistor W_{p1} and W_{n1}. The load transistor that the critical signals in the SATG drive in the slave-stage is only a single NMOS transistor (W_{n5}), which is very desirable for enhanced metastability due to smaller parasitic capacitance values. The waveform that demonstrates the contention at node "X" during the metastable period for the SATG is shown in **Figure 4.6(d)**.

The qualitative metastability analysis for the RCSSATG and the LCSATG is identical. Similarly, the qualitative analysis for the RCSPDFF and the LCPDFF is identical to the PDFF and will not be repeated here.

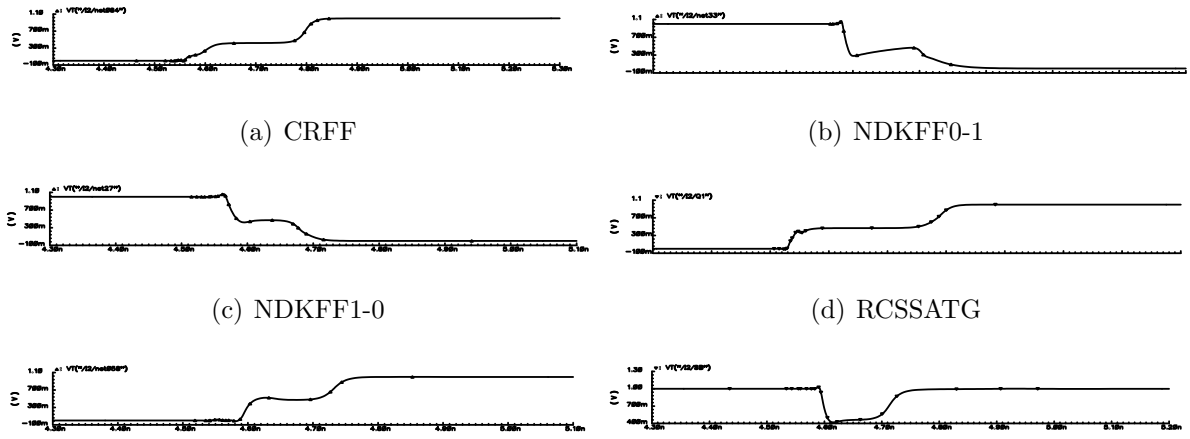


(a) CPN (b) SPFF

Figure 4.5: Schematic Diagram of Level-Converting Flip-Flops for Metastability Analysis

Because both the CPN and the SPFF employ the conditional-capturing technique, a temporary pulse is generated at the critical node during the evaluation phase until the output makes the corresponding transition and consequently cut off the discharge paths. In the CPN, node "Y" is always pre-charged to logic "1" because the weakly sized transistor W_{p1} is always turned on. During the 0-1 data transition, a temporary negative pulse is generated at node "Y", and in turn causes contention at node "X" between transistor W_{p2} and the discharge path formed by transistor W_{n5}, W_{n6}, and W_{n7}. While node "Y" does not contribute to the 1-0 data transition, the same contention exists at node "X" during the metastability period due to the synchronization between the CLK and the input data D signals. The cross-coupled inverter pair formed by transistor W_{p8} and W_{n8} is used to stabilize the critical node. Alternative, metastability can also be enhanced by using smaller transistor W_{p2} and W_{n7} to reduce the amount of capacitance associated at node "X". The

waveform that demonstrates the contention at node "X" during the metastable period for the CPN is shown in Figure **4.6(e)**. In the SPFF, the temporary pulse is generated at the critical node "X", which is stabilized by the cross-coupled inverter pair formed by transistor W_{p4} and W_{n4}. The load transistor for metastability analysis in the SPFF include W_{n3} as well as the SR-latch transistors in the slave-stage. The waveform that demonstrates the contention at node "X" during the metastable period for the SPFF is shown in **Figure 4.6(f)**.



(a) CRFF

(b) NDKFF0-1

(c) NDKFF1-0

(d) RCSSATG

(e) CPN

(f) SPFF

Figure 4.6: Metastable Contention Nodes for Dual-Supply Flip-Flops

4.3 Quantitative Design Methodology for Metastable-Hardened Flip-Flops

4.3.1 Transistor Sizing

In this work, the value of τ is manipulated by varying the transconductance g_m of the inverter pair and the relevant capacitances associated with the critical nodes through transistor sizing. Because flip-flops have two types of data transitions (0-1 and 1-0), transition with the worst τ is chosen for analysis in this section.

Two types of analysis are performed on the variation of τ in flip-flops based on transistor sizing: *(i)* Transconductance (g_m) Variation (TV) and *(ii)* Load Variation (LV) . In the TV analysis, g_m of the inverter pair is varied accordingly while keeping the load capacitance C_Q constant. The LV method changes the value C_Q with a fixed g_m value. While a typical flip-flop design features 20-30 transistors, this work will only focus those that have an impact on τ either through transconductance or load variation. The transistor sizing approach used for analysis of τ is outlined below.

- Size the flip-flop for optimum PDP.
- Vary the size of the transconductance transistors for TV analysis.
- Fix the transconductance transistors sizing based on optimum τ value obtained.
- Vary the size of the load transistors for LV analysis.

We found the sizing for optimum PDP is a very good starting point for analyzing and optimizing τ because it also takes into account the design tradeoff between delay and

power dissipation which will be discussed later in this chapter. All the sizing schemes used in analysis ensure the correct functionality of the flip-flop. For a given analysis, the corresponding transistors used to vary the g_m and C_Q values are listed in **Table 4.1**. It is important to point out the optimum τ value obtained from the TV analysis may not be the absolute minimum value but rather the value around the knee of the curve.

For single-supply flip-flops, the size of W_{p2} and W_{n2} in the PowerPC is varied for TV analysis to change g_m in the inverter pair but always maintain an aspect ratio of 1 in order to yield optimum τ [14][16]. For LV analysis, W_{p2} and W_{n2} are fixed at identical sizes while W_{p3} and W_{n3} are varied to generate different load values at the critical node. For the SAFF, the size of W_{n1} does not change in any scenario in order to maintain high-performance, and the sizing of W_{p1} is responsible for the manipulation of g_m in TV analysis. In the LV scenario, the size of W_{p1} and W_{n1} does not change while the size of W_{p4} and W_{n4} in the NAND gate is varied for different load values. The sizing scenario of the PDFF is very similar to that of the SAFF except W_{n1} replaces W_{p1} as the transistor responsible for g_m variation in the master-stage and NMOS W_{n2} becomes the load transistor. The TV analysis in the SDFF involves changing the size of W_{p1} and W_{n1} with an aspect ratio of 1 while the LV analysis varies the three transistors (W_{p3}, W_{n2}, W_{n6}) that are on the critical path and connecting to the critical node.

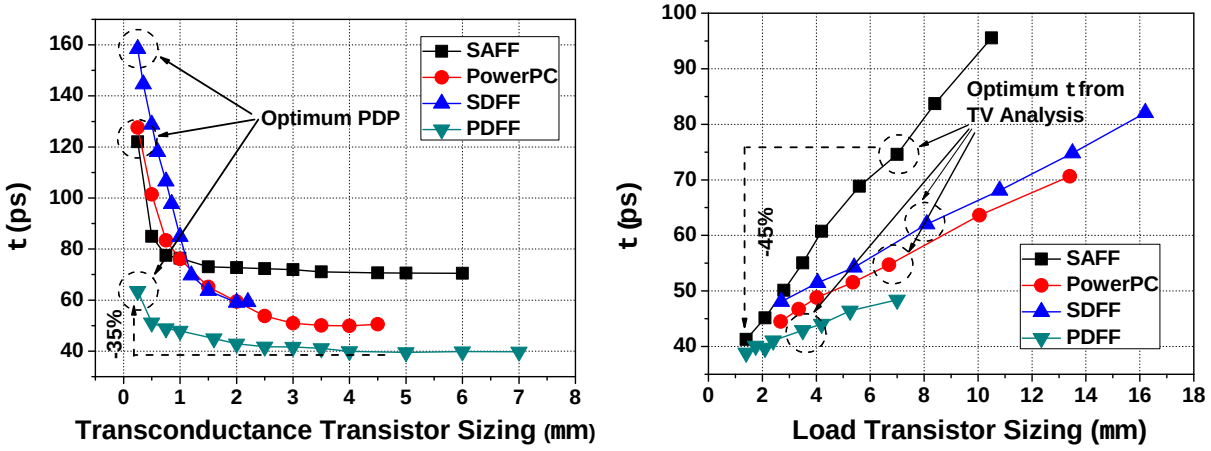
For dual-supply flip-flops, the TV and the LV analysis of the RCSPDFF and the LCPDFF is identical to those of the PDFF. For both the RCSSATG and the LCSATG, the TV analysis involves changing the sizes of transistor W_{p1} and W_{n1} with an aspect ratio of 1 while the LV analysis varies the size of the NMOS transistor W_{n5} with constant g_m transistors. For analysis purposes, the 0-1 output transition in the NDKFF will be used since it has a higher value of τ than the 1-0 transition. The TV analysis of the NDKFF involves changing the size of the feedback transistors W_{p2} and W_{n4} with an aspect ratio of

1 while the LV analysis varies the sizing of transistor W_{p1} and W_{n3} that are connecting to the critical node. In the CRFF, transconductance g_m is manipulated by varying transistor W_{p4} and W_{n4} with an aspect ratio of 1, and the size of transistor W_{p2} and W_{n2} are changed accordingly for LV analysis. The TV and the LV analysis in the CPN and the SPFF is rather similar. The TV analysis involves changing the size of the cross-coupled inverter pair to stabilize the contention at the critical node while the LV analysis deals with the transistors associated with the critical node in the discharge paths (W_{n3} for the SPFF and W_{n7} for the CPN). Although W_{p2} in the CPN and W_{p1} in the SPFF are also connected to the critical node in the respective flip-flop, their size must be kept constant in order to ensure the correct functionality of the flip-flop.

Table 4.1: Flip-Flop Transistor Sizing Schemes for Transconductance g_m and Load C_Q Variation

	Transconductance (g_m) Variation	Load (C_Q) Variation
SAFF	W_{p1}	W_{p4}, W_{n4}
PowerPC	W_{p2}=W_{n2}	W_{p3},W_{n3}
SDFF	W_{p1}=W_{n1}	W_{p3}, W_{n2}, W_{n6}
PDFF, RCSPDFF, LCPDFF	W_{n1}	W_{n2}
RCSSATG, LCSATG	W_{p1}=W_{n1}	W_{n5}
0-1 NDKFF	$W_{p2} = W_{n4}$	W_{p1}, W_{n3}
CRFF	W_{p4}=W_{n4}	W_{p2}, W_{n2}
SPFF	W_{p4}=W_{n4}	W_{n3}
CPN	W_{p8}=W_{n8}	W_{n7}

In the TV analysis of single-supply flip-flops (**Figure 4.7(a)**), increase the relevant



(a) Transconductance Transistor Width Variation

(b) Load Transistor Width Variation

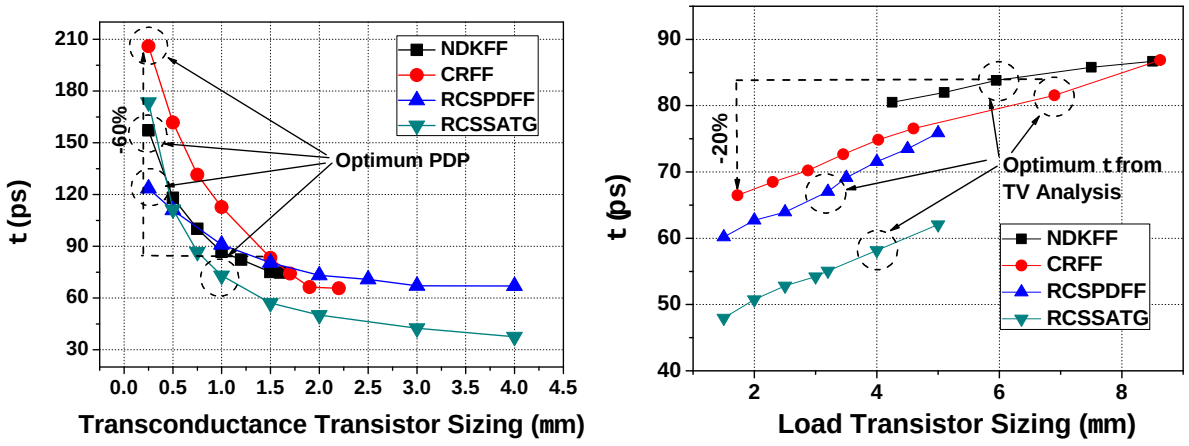
Figure 4.7: Impact of Transistor Sizing on τ using Transconductance and Load Variation in Single-Supply Flip-Flops

transistor widths in the master-stage can significantly reduce τ when compared to the optimum PDP sizing scheme. For example, increase g_m in the master-stage of the PDFF can reduce τ by 35% from the optimum PDP design point. However, further increase in width beyond the values shown in the figure will increase the value of τ because the capacitance terms in the numerator of **Equation (2.10)**, especially the Miller capacitances, begins to dominate over g_m in the denominator. Hence, the knee of the curve is important in determining the optimum sizing scenario for τ in order to prevent over-sizing that can further impact power and performance. Due to their respective architecture, it is clear from **Figure 4.7(a)** that the minimum τ value achieved by the TV analysis in the SAFF and the SDFF is higher than that of the PowerPC and the PDFF. In the SAFF, g_m in the master-stage is limited by the fact that path to V_{SS} is formed by three NMOS transistors in series instead of just of a single transistor. As for the SDFF, the inverter pair that stabilizes

the critical node is not on the critical path, and hence cannot be sized up significantly in order to maintain flip-flop's correct functionality. In general, the minimum τ achieved in TV analysis is limited by the g_m transistor sizes in the master-stage before the saturation occurs. According to **Equation (4.2)** and **(4.3)**, decrease in load transistor width results in a linear decrease in the capacitance, which translates to a linear reduction in τ given a constant g_m. This is illustrated in **Figure 4.7(b)** when the LV analysis is performed by varying the size of the load transistors. In this case, τ reduction is more significant than the TV analysis because the size of load transistors can be reduced continuously as long as the flip-flops retains the correct functionality. For example, using smaller transistors for the cross-coupled NAND gates in the slave-stage of the SAFF can further reduce τ by 45% from the optimum value obtained from the TV analysis.

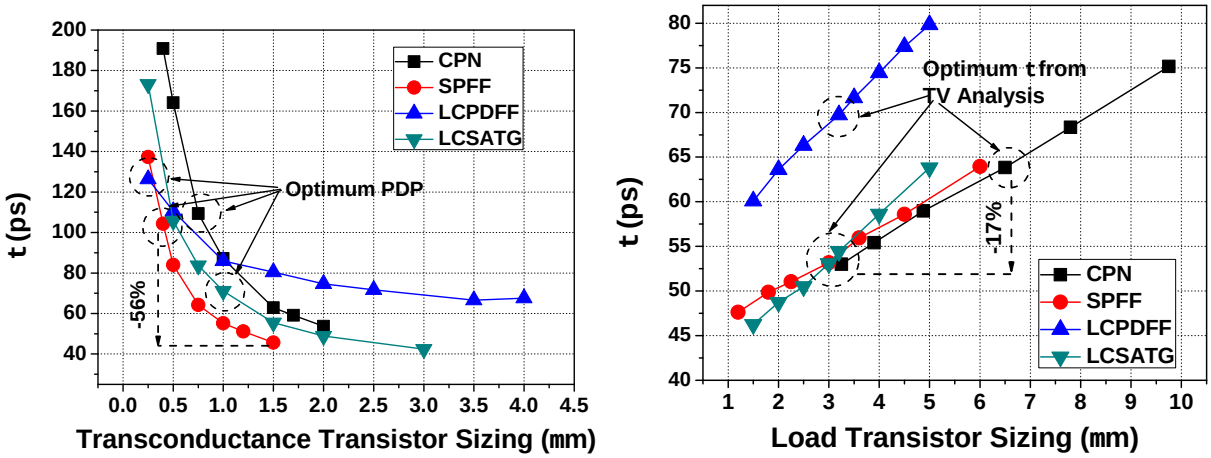
For the dual-supply flip-flops, both the TV and the LV analysis is performed for $V_{DDL} = 1.4V$. However, the analysis can easily be extended to other V_{DDL} values using the same methodology. **Figure 4.8** and **4.9** illustrate the TV and LV analysis of the reduced clock-swing and level-converting flip-flops, respectively.

Similar to the flip-flops in the single-supply system, the reduction of τ in the TV and the LV analysis is also evident in both reduced clock-swing and level-converting flip-flops. For example, increase g_m of the CRFF reduces τ by 60% when compared to the optimum PDP design, and a further 20% reduction can be achieved using the LV analysis. However, it is also clear that the inverter pair (W_{p4}, W_{n4}) in the CRFF cannot be sized above $2.2\mu m$ in order to keep the parasitic capacitances surrounding the critical node small enough to allow the input data to be correctly written via the NMOS-pass transistors. A similar argument can be made for the NDKFF where the g_m transistors cannot be increased beyond $1.5\mu m$ in the TV analysis in order to maintain the correct functionality because the feedback path would then be too strong to prevent the flip-flop from sampling new



(a) Transconductance Transistor Width Variation (b) Load Transistor Width Variation

Figure 4.8: Impact of Transistor Sizing on τ using Transconductance and Load Variation in Reduced Clock-Swing Flip-Flops



(a) Transconductance Transistor Width Variation (b) Load Transistor Width Variation

Figure 4.9: Impact of Transistor Sizing on τ using Transconductance and Load Variation in Level-Converting Flip-Flops

data. The LV analysis of the NDKFF is limited by the effect of transistor stacking where the size of the load transistors (W_{p1}, W_{n3}) cannot be reduced significantly to maintain the flip-flop functionality, and thus only an additional 5% reduction of τ is achieved from the TV analysis. For the CPN and the SPFF, because both the forward and feedback inverter in the cross-coupled inverter are sized identically to ensure optimum τ, the additional parasitic capacitances added to the critical node due to the feedback inverter limits the transconductance value in the inverter pair and subsequently the value of τ. In terms of the LV analysis, an additional 14%-17% of reduction in τ can be achieved in the CPN and the SPFF, respectively, before the flip-flop fails to function. The general trend displayed in the TV and the LV analysis for the RCSPDFF and the LCPDFF is very much identical to those of the PDFF except for the higher τ values because the reduced clock-swing signal connected to the drain terminal of the PMOS devices result in an exponential increase of τ. It is also evident that the architecture of the SATG is very desirable for metastable-hardened flip-flops designs, especially for dual-supply systems because the cross-coupled inverter pair can be sized up to simultaneously achieve good performance and reduce τ without any restriction on maintaining the functionality of the flip-flop. For example, the TV analysis reveals that the value of τ for the RCSSATG and the LCSATG is much lower than the other flip-flops at the optimum PDP design point. Continuous increase in the size of the cross-coupled inverter beyond the optimum PDP point will result in further reduction in τ. Because of the small load transistor in the slave-stage, the TV analysis has been shown to be a more effective method in reducing τ than the LV analysis.

The transistor sizing scheme using transconductance variation (TV) and load variation (LV) have shown consistent results across various flip-flop architectures in both single and dual-supply systems. An initial increase in the size of the cross-coupled inverter that stabilizes the critical nodes will result in dramatic reduction in τ before it saturates to a

constant value. Further linear reduction in τ can be achieved by varying the size of the load transistors surrounding the critical node. Due to the different architectures, the amount of reduction in τ via transistor sizing will vary between flip-flops to ensure the correct circuit functionality is still maintained.

4.3.2 Flip-Flop Metastability Modeling

In this section, a simple "back of the envelope" method is demonstrated for quick estimation and evaluation of τ for different flip-flop topologies as a function of transistor widths. We will also show the proposed transconductance variation (TV) and load variation (LV) analysis can be modeled on the selected flip-flop architectures using the approach described in this section. Although we have included only three flip-flops for analysis in this section (PowerPC, SAFF, and PDFF), the idea can easily be extended to other flip-flop architectures.

The general modeling methodology involves the calculation of the transconductance g_m as well as the modeling of the parasitic capacitances of C_Q and C_M surrounding the critical node of each flip-flop. The calculation of g_m, given in **Equation (4.4)** is identical to the one described in [18].

$$\begin{aligned} g_m &= g_{mn} + g_{mp} \\ &= (\mu_n C_{ox} \frac{W_n}{L} \frac{1}{1+\sqrt{a}} + \mu_p C_{ox} \frac{W_p}{L} \frac{\sqrt{a}}{1+\sqrt{a}})(V_{DD} - |V_{tp}| - V_{tn}) \end{aligned} \tag{4.4}$$

where

$$a = \frac{\mu_n}{\mu_p} \frac{W_n}{W_p}$$

Two types of capacitance are generally considered when modeling a MOSFET device, namely the gate and the diffusion capacitance. The gate capacitance consists of C_{gs},

C_{gd}, and C_{gb}, while diffusion capacitance is composed of C_{sb} and C_{db}. A detailed model of MOSFET capacitance is shown in **Figure 4.10**. For simplicity reasons, we assume

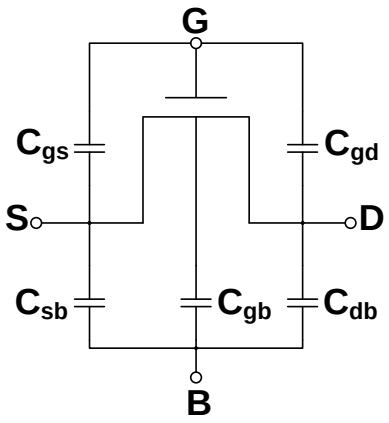


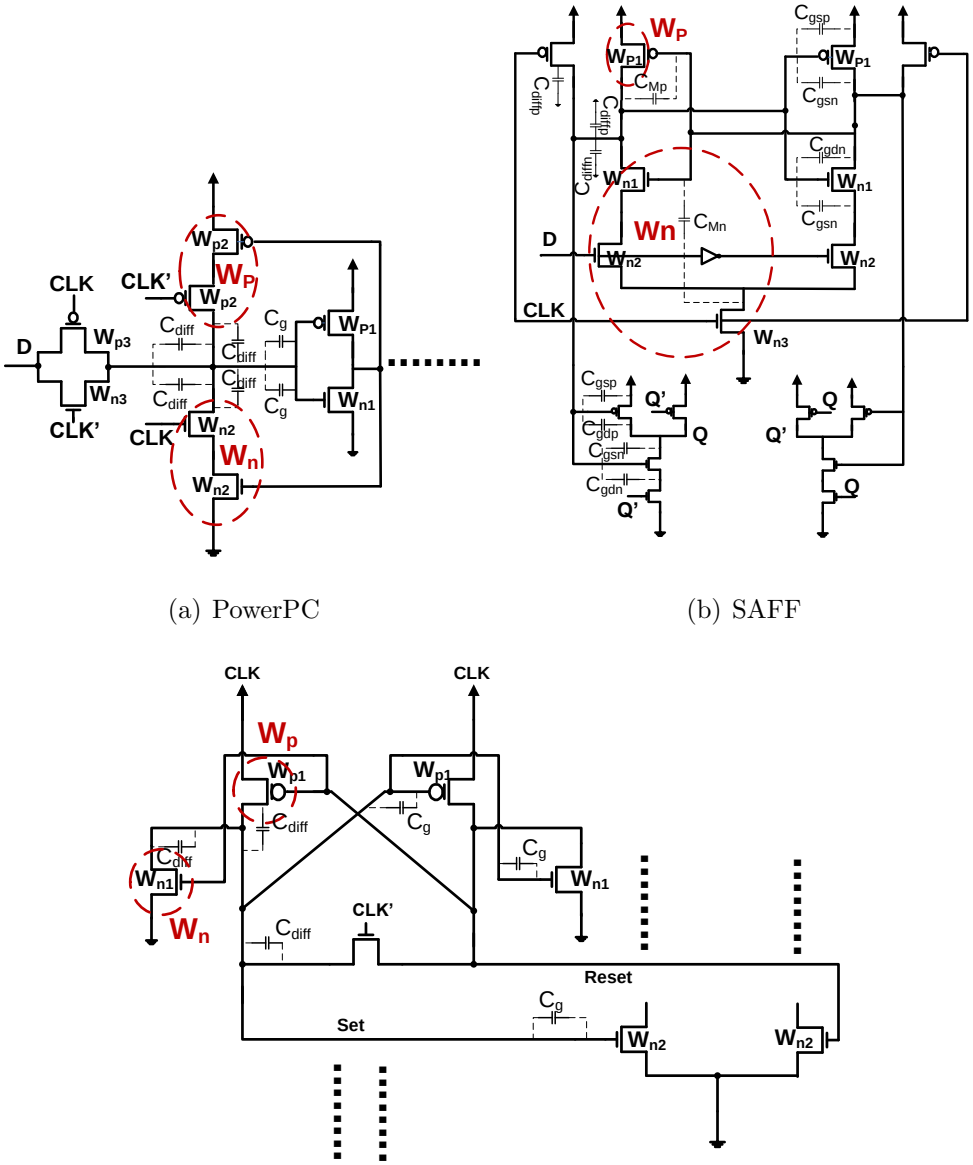
Figure 4.10: Capacitance Modeling of a MOSFET Device

$C_g = C_{gs} = C_{gd} = C_{gb}$ and $C_{diff} = C_{sb} = C_{db}$. The equations for the calculation of C_g and C_{diff} are given in **Equation (4.2)** and (**4.3**). Since the Miller capacitance (C_M) is the coupling capacitor between the gate and the drain terminal of the MOSFET, its value is identical to that of C_{gd}. We also ignored the effect of C_{gb} in our analysis.

Based on the critical nodes identified in **Figure 4.2**, the corresponding g_m and the total capacitance surrounding the critical node can be calculated. In order to apply **Equation (4.2)-(4.4)**, the technology parameters and the parasitic capacitances listed in **Table 4.2** must be available for both the PMOS and NMOS transistors.

Table 4.2: Technology Parameters Required for the Calculation of τ

Technology Parameters	μ_0, C_{ox}, V_{DD}, V_{t0}, L_s
Parasitic Capacitances	C_{go}, C_j, C_{js}



(a) PowerPC (b) SAFF

(c) PDFF

Figure 4.11: Modeling of the Critical Node for Single-Supply Flip-Flops

Figure 4.11 illustrates a detailed modeling of the capacitance at the critical node for each respective flip-flop. The subscript "n" and "p" denote NMOS and PMOS devices respectively. For flip-flops with differential critical signals in the master-stage, the modeling is only illustrated on one of the signals due to symmetry. The relevant transistors used in calculating g_m and the Miller capacitance (C_M) in the inverter pair are labeled as W_p and W_n. In cases where multiple devices are in series, an effective width is used in the calculation. Once the effective transistor width is determined, the calculation of g_m is straightforward using **Equation (4.4)**. The term C_M shown in **Equation (2.10)** is simply the sum of the C_{gdn} and C_{gdp} in the inverter. The parasitic capacitance (C_Q) at the critical node is the lumped value that includes contribution from various diffusion capacitances and the gate capacitances from the inverter pair in the master-stage as well as the load transistors in the slave-stage.

By inputting Equation (4.2)-(4.4) and technology parameters using tools such as Microsoft Excel, the time-resolving constant τ for a given flip-flop topology can be calculated for various transistor sizing scenarios. A sample worksheet for calculating τ in the PDFF is shown in **Table 4.3** where W_S is the size of the load transistor in the slave-stage while W_p and W_n are the transistor size of the inverter pair in the master-stage.

The proposed modeling and estimation tool allows the designers to generate different τ values for various combination of sizing scenarios by simply changing the relevant values in the spreadsheet. The sample data shown in **Figure 4.12(a)** and **4.12(b)** are calculated using the proposed estimation methodology and generated using the spreadsheets. In **Figure 4.12(a)**, the τ of the SAFF is plotted for different series of W_{n1} values as a function of W_{p1}. Similarly, **Figure 4.12(b)** illustrates the τ values of the SAFF for series of different W_{p1} and W_{n1} values with an aspect ratio of 1 as a function of the load transistor sizing (W_{p4}, W_{n4}) in the slave-stage. Using the data generated by the estimation tool, the

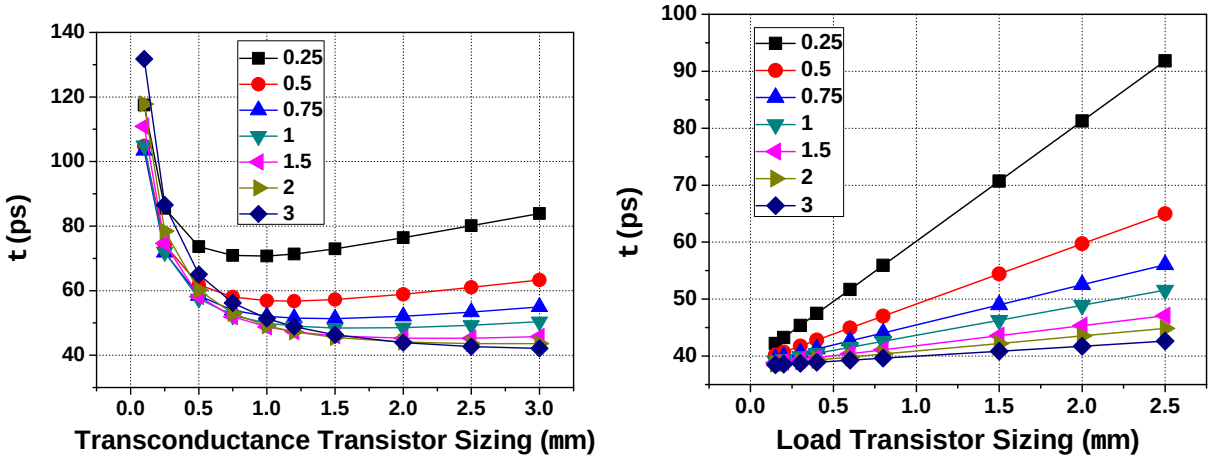
Table 4.3: Sample Microsoft Excel Spreadsheet

W_p (μm)	W_n (μm)	W_s (μm)	g_m ($\mu A/V$)	C_Q (fF)	C_M (fF)	τ (ps)
1.2	0.2	1.5	549.76	5.111	3.03	31.35
1.2	0.25	1.5	614.65	5.208	3.14	28.93
1.2	0.5	1.5	869.24	5.687	3.71	23.6
1.2	0.75	1.5	1064.6	6.166	4.27	21.84
1.2	1	1.5	1229.29	6.645	4.84	21.14
1.2	1.2	1.5	1346.62	7.028	5.29	20.92
1.2	1.5	1.5	1505.57	7.603	5.96	20.89
1.2	2	1.5	1738.48	8.561	7.09	21.24

designers are able to quickly estimate the value of τ and analyze the tradeoffs between τ and other design constraint factors such as area, power, and performance.

To verify our proposed model, we compared the calculated τ values with those obtained in simulation across three different technology nodes: $0.18\mu m$, 90nm, and 65nm. The technology parameters for $0.18\mu m$ technology are obtained from models provided by MOSIS [63]. The parameters for 90nm and 65nm are taken from the BSIM4 model files available in Predictive Technology Model (PTM) [64] . **Table 4.4** summarizes the main technology parameters for the three technology nodes.

Figure 4.13(a)-4.13(c) illustrate the comparison for the calculated and the simulated τ values for each flip-flop architecture across three different technology nodes. The data shown in these figures correspond to the TV analysis where the value of g_m in the inverter pair is changed while the size of the load transistor remains constant. The methodology



(a) Transconductance Variation

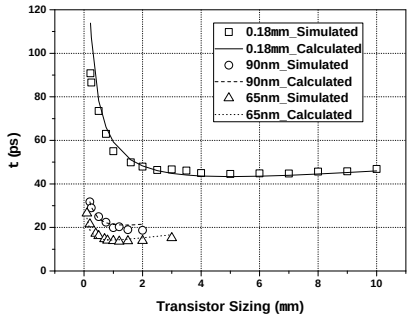
(b) Load Variation

Figure 4.12: Series of SAFF τ Values Generated by the Proposed Modeling Due to Transconductance and Load Variation

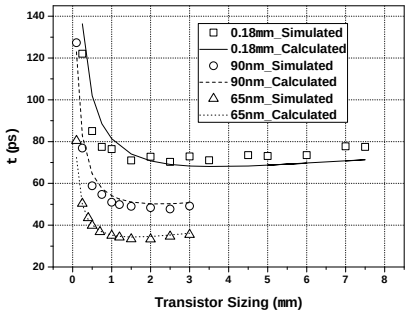
Table 4.4: Selected Process Parameters for Different Technologies

	V_{DD} (V)	V_{thn} (V)	V_{thp} (V)	K'_n ($\mu A/V^2$)	K'_p ($\mu A/V^2$)
$0.18\mu m$	1.8	0.53	0.51	170	37
90nm	1.2	0.397	0.339	687	85
65nm	1	0.368	0.297	1145	127

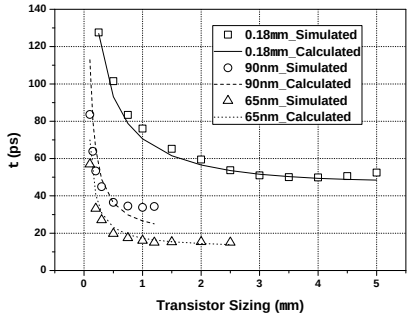
for transistor sizing in each flip-flop is identical to those previously described. From these figures, it is clear that the calculated values match very well with the simulated values across all three technology nodes for the flip-flops analyzed where the maximum deviation is 17%. More importantly, the calculated values accurately estimate the knee of the curve as



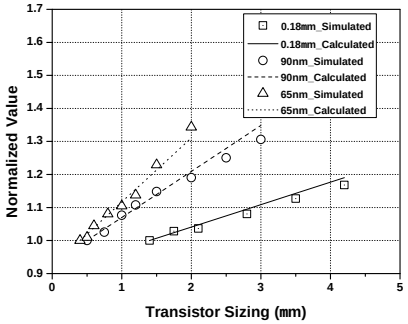
(a) TV Analysis for PDFF



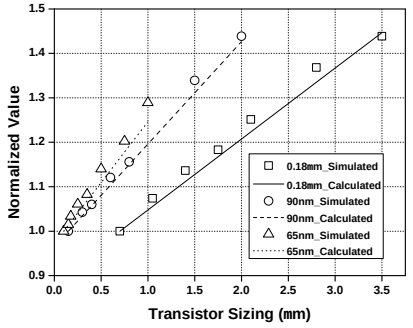
(b) TV Analysis for SAFF



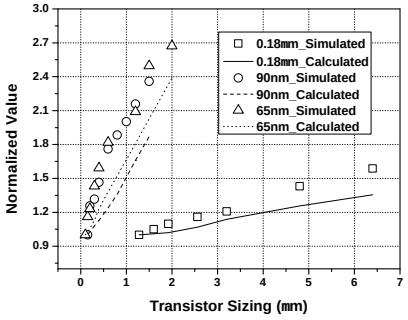
(c) TV Analysis for PowerPC



(d) LV Analysis for PDFF



(e) LV Analysis for SAFF



(f) LV Analysis for PowerPC

Figure 4.13: Comparison between Simulated and Calculated τ values

a function of transistor sizing. This is important in designing reliable systems because the knee point indicates the optimum value on τ for a specific sizing scheme given a particular flip-flop topology.

As evident from **Figure 4.12(b)**, τ changes in a linear manner as the size of the load transistors in the slave-stage varies using the LV analysis. This observation is also evident in **Figure 4.13(d)-4.13(e)**. In these figures, the values obtained from simulation are the normalized τ values while the calculated values are the normalized total capacitance shown in the numerator of **Equation (2.10)** (C_Q+4C_M). Without changing the size in the inverter pair in the master-stage, the value of g_m in the denominator of **Equation (2.10)** remains unchanged. Hence, the value of τ should have a direct linear relationship with the total capacitance value as the size of the load transistors varies. In fact, this is evident in Figure **4.13(d)-4.13(e)** where the amount of change in the simulated τ values closely resembles the percentage change in the total capacitance calculated for various load transistor sizes.

The data shown **Figure 4.13(d)** and **4.13(e)** is obtained from the SAFF and the PDFF where the contribution of load capacitance is dominated by gate capacitances. A similar load transistor variation analysis was also performed on the PowerPC (**Figure 4.13(f)**), but the discrepancy between the simulated and the calculated values was quite large with a maximum deviation of approximately 50%. We believe a couple of reasons may have contributed to this deviation. First of all, majority of the load capacitance contribution in the PowerPC comes from the diffusion capacitance (C_{diff}), and the equation we use to model diffusion capacitance is only a first-order approximation. In reality, the calculation of C_{diff} has a dynamic behavior and can be more accurately modeled by **Equation (4.5)**.

$$C_{diff} = \frac{C_{diff_0}}{(1 + \frac{V_j}{V_0})^m} \tag{4.5}$$

where V_j is the magnitude of the junction reverse-bias voltage, C_{diff_0} is the diffusion capacitance at zero reverse-bias voltage, V_0 is the junction built-in potential, and m is the grading coefficient. Therefore, in order to accurately calculate the diffusion capacitance associated at the critical node of a flip-flop, an important parameter that must be considered is the node voltage, which can only be obtained accurately from simulation. While a value of $V_{DD}/2$ can be assumed as the node voltage during metastability, the exponential relationship shown in **Equation (4.5)** means that a small deviation from that value can potentially result in large deviation from the actual capacitance value. The modeling of the diffusion capacitance does not impact the analysis in the SAFF and the PDFF because its value is generally much smaller than the gate capacitances [62]. A second possible reason that resulted in the deviation is the neglecting of the lumped resistances in our calculation. Because a transmission-gate topology is associated with the critical node in the PowerPC, the effects of the source-drain resistance may have played a more prominent role in determining τ than other flip-flop topologies such as the PDFF and the SAFF.

4.3.3 Proposed Design Metrics

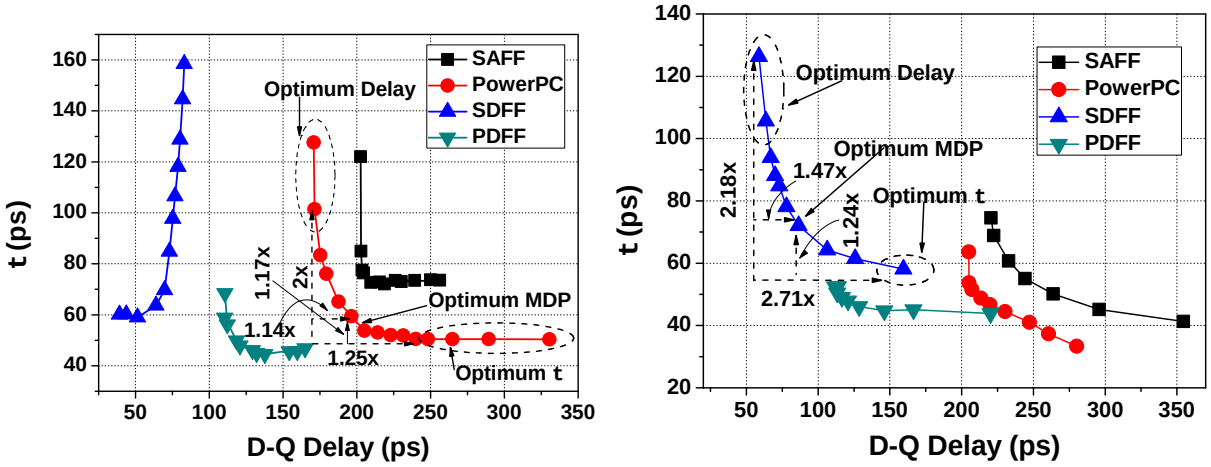
In this section, two new design metrics, namely the metastability-delay-product (MDP) and the metastability-power-delay-product (MPDP), are proposed to provide an illustration in analyzing the design tradeoff between delay, power, and metastability. In traditional flip-flop designs, power vs. delay curve (**Figure 3.14(b)**) is an useful illustration in analyzing the tradeoff between the delay and power consumption. The best design tradeoff usually occurs around the knee of the curve, and thus indicating a minimum PDP value. In this work, both the τ vs. delay and the τ vs. PDP curve illustrate a similar tradeoff analysis and provide a useful illustration in exploring the design space between enhancing

the metastability performance of the flip-flops while still satisfying the timing and power design constraints. The τ vs. delay curve illustrates the design tradeoff between τ and delay, and the knee of the curve usually indicates the optimum MDP design. Likewise, the knee of the τ vs. PDP curve indicates the optimum MPDP design point, which is typically the best design tradeoff point between τ and PDP. These curves can be generated using the aforementioned transistor sizing schemes such as the transconductance and load variation method. For dual-supply flip-flops, both the TV and the LV analysis are again performed for $V_{DDL} = 1.4V$.

MDP:

While appropriate transistor sizing can reduce the value of τ significantly, it often comes at the expense of performance degradation. In the PowerPC, for example, sizing up transistor W_{p2} and W_{n2} increases the g_m to reduce τ but also adds more capacitances to the critical node, and thus increases the delay in the critical path. Although small load transistors in the slave-stage of the SAFF results in smaller τ, it also increases the delay considerably. The PDFF is an example where delay and τ can be simultaneously reduced because transistor W_{p1} is on the critical path and also responsible for g_m in the master-stage. However, increasing the size of W_{n1} in the PDFF to improve g_m will result in higher delay by adding capacitances to the critical node. The SDFF is an interesting case to analyze for the 1-0 data transition. While sizing up the inverter pair (W_{p1}, W_{n1}) improves metastability, it also maintains the pre-charged logic "1" at the critical node from temporary false-discharging and thus enhances the overall flip-flop performance.

Figure 4.14 plots the τ vs. delay curve of the analyzed flip-flops in single-supply system for both the TV and the LV analysis. With the exception to the SDFF, the D-Q delay of the PDFF clearly is lower than the PowerPC and the SAFF while maintaining



(a) Transconductance Transistor Width Variation

(b) Load Transistor Width Variation

Figure 4.14: Illustration of MDP in Single-Supply Flip-Flops using τ vs. Delay Curve via Transistor Sizing

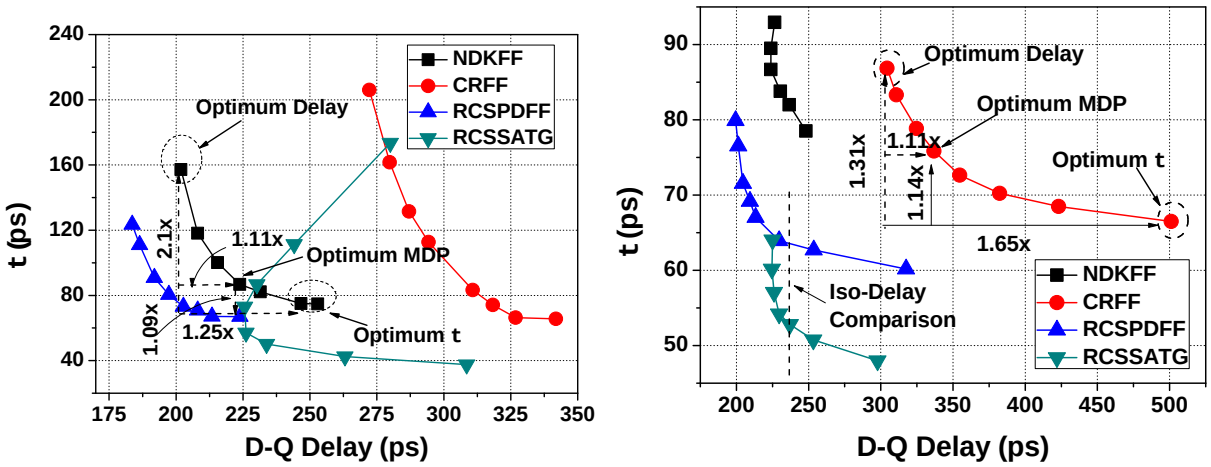
a lower τ value. Although the D-Q delay of the SDFF is lower than the PDFF, it comes largely at the expense of a much higher τ value. In most cases, transistor sizing has an opposite effect on delay and τ, and thus there must exists a point which offers the best design tradeoff between these two parameters. Here, we introduce a new design metric called the metastability-delay-product (MDP) to analyze and balance this tradeoff. As shown in **Equation (4.6)**, MDP is simply the product between τ and the flip-flop delay for a given transistor sizing scheme.

$$MDP = \tau \times Delay \tag{4.6}$$

The design tradeoff between delay and metastability in flip-flops can be illustrated using both the TV and the LV analysis. Using the PowerPC and the SDFF as examples, the region for optimum delay, optimum τ, and optimum MDP are labeled in **Figure 4.14**. In

the case of the PowerPC (**Figure 4.14(a)**), the value of τ at the optimum delay point is 2× greater than the optimum τ value. Similarly, the delay at the optimum τ value is 1.25× higher than the optimum delay value. At the optimum MDP point, however, these values have been reduced to 1.17× and 1.14× respectively, and thus indicating a better design tradeoff between performance and metastability. A similar analysis can be done for the SDFF example shown in **Figure 4.14(b)** for the LV analysis. For either TV or LV analysis, the optimum MDP point in each respective analysis occurs around the knee region of the τ vs. delay curves due to the inherent design tradeoff. However, the absolute lowest MDP value for a given flip-flop can come either from the TV or LV analysis depending on the architecture. For example, the lowest MDP value for the SAFF comes from the LV analysis while the lowest MDP value for the PDFF is from the TV analysis. Overall, MDP is an important metric to consider when designing digital datapaths where reliability and high-performance are the primary objectives.

Figure 4.15 plots the τ vs. delay curve of the reduced clock-swing flip-flops for both the TV and the LV analysis. Using the TV analysis of the NDKFF as an example, the value of τ at the optimum delay point is 2.1× greater than the optimum τ value. On the other hand, the delay at the optimum τ value is 1.25× higher than the optimum delay value. At the optimum MDP point, however, these values have been reduced to 1.09× and 1.11× respectively, and thus indicating a better design tradeoff between performance and metastability. In the LV analysis of the CRFF, the optimum MDP point yields an 1.11× and 1.14× increase from the optimum delay and the optimum τ value, respectively. However, both values still demonstrate better design tradeoff than the 1.65× increase in delay at the optimum τ value and a 1.31× increase in τ at the optimum delay point. While the delay of the RCSPDFF is the lowest among the analyzed reduced clock-swing flip-flops, the RCSSATG exhibits the best design tradeoff between performance and metastability. This



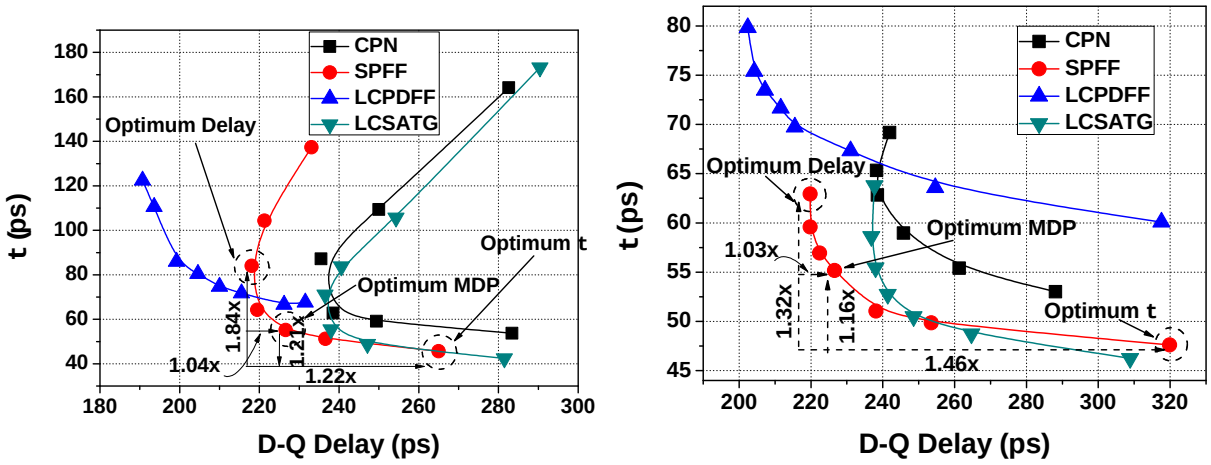
(a) Transconductance Transistor Width Variation

(b) Load Transistor Width Variation

Figure 4.15: Illustration of MDP in Reduced Clock-Swing Flip-Flops using τ vs. Delay Curve via Transistor Sizing

is especially evident in an iso-delay comparison of the LV analysis where the RCSSATG is able to achieve the same delay as the RCSPDFF but at a much lower value of τ.

Figure 4.16 plots the τ vs. delay curve of the level-converting flip-flops for both the TV and the LV analysis. A similar design tradeoff between performance and metastability exists for all the flip-flops analyzed. In the TV analysis, the shape of the τ vs. delay curve is similar for the CPN, the SPFF, and the RCSSATG. Because the forward inverter in the cross-coupled inverter pair of the CPN and the SPFF is also on the critical path, sizing up the inverter pair initially decreases both the delay and τ. As the inverter size continues to increase, the value of τ still decreases but the additional parasitic capacitances added to the critical node due to the feedback inverter subsequently contributes to an increase in the overall delay. A similar analysis can be made for the LCSATG. Regardless of the curve shape, the optimum MDP point still results in the best tradeoff between performance and



(a) Transconductance Transistor Width Variation

(b) Load Transistor Width Variation

Figure 4.16: Illustration of MDP in Level-Converting Flip-Flops using τ vs. Delay Curve via Transistor Sizing

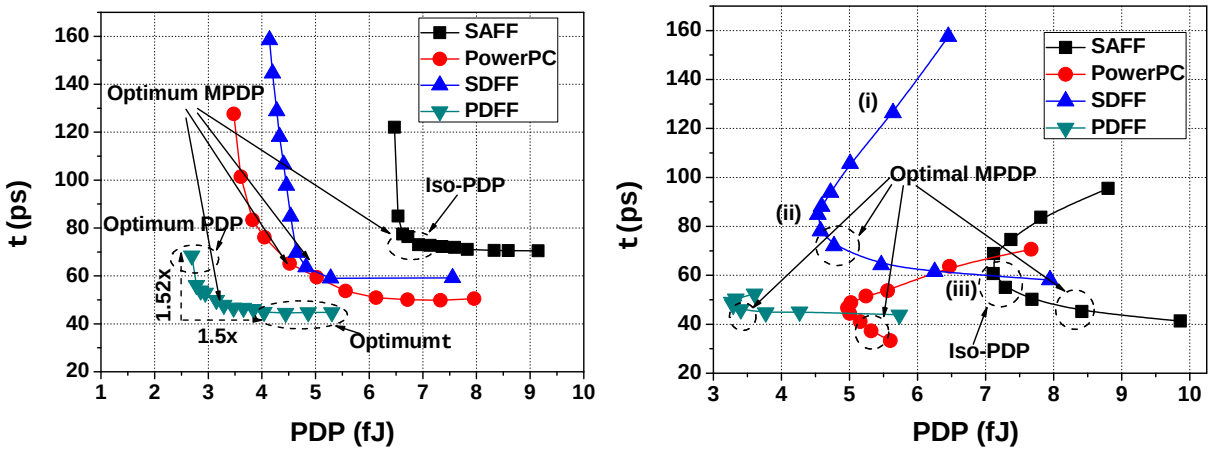
metastability. The TV analysis for the SPFF shows the optimum MDP design results in an 1.04× and 1.21× increase from the optimum delay and optimum τ value as opposed to the 1.84× increase in τ at the optimum delay point or an 1.22× increase in delay at the optimum τ point. A similar design tradeoff is also illustrated for the SPFF using the LV analysis.

MPDP:

Power consumption is another important factor that must be considered in flip-flop designs. Hence, another design metric called the metastability-power-delay-product (MPDP), given by **Equation (4.7)**, is also introduced to design metastable-hardened, high-performance, and low-power flip-flops.

$$MPDP = \tau \times Power \times Delay \tag{4.7}$$

Figure 4.17 illustrates the τ vs. PDP curve for the single-supply flip-flops using both the TV and the LV analysis. In the TV analysis of single-supply flip-flops (**Figure 4.17(a)**),



(a) Transconductance Transistor Width Variation

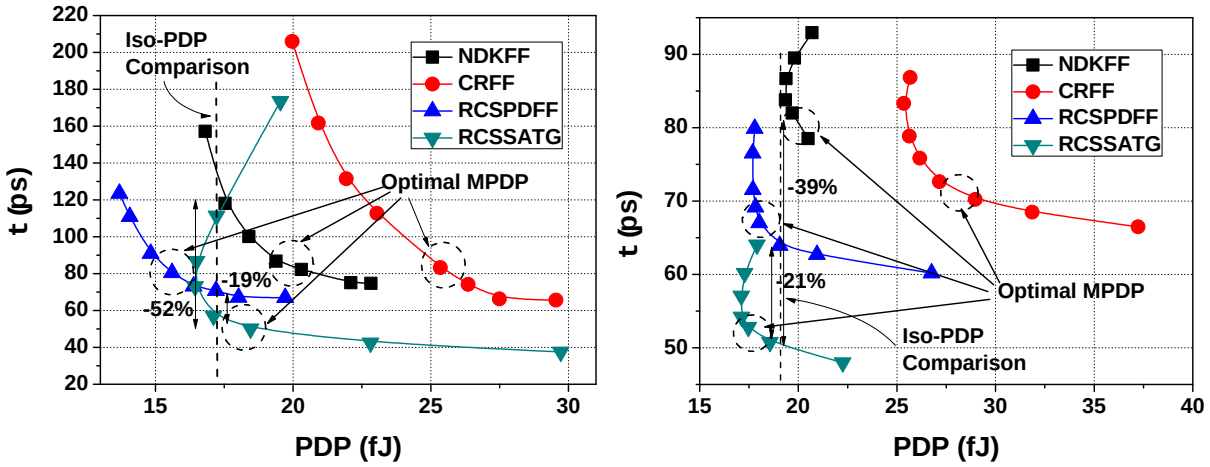
(b) Load Transistor Width Variation

Figure 4.17: Illustration of MPDP in Single-Supply Flip-Flops using τ vs. PDP Curve via Transistor Sizing

the small PDP values indicate the size of the inverter pair in the master-stage is small, which typically results in lower power and delay but higher τ values. As the inverter pair size increases, the reduction of τ comes at the expense of overall PDP increase. Clearly, a tradeoff exists between τ and the PDP in all flip-flop architectures such that the optimum MPDP point is around the knee region of the curve. For the PDFF, the τ and PDP value at the optimum MPDP point is 1.06× and 1.2× higher than the optimum τ and PDP value respectively. This is a better tradeoff than designing at either the optimum τ or the optimum PDP value where the amount of increase in PDP and τ is 1.5× and 1.52×, respectively. In the LV analysis (**Figure 4.17(b)**), a different shape of the τ vs. PDP curve is observed. Generally, the curve can be divided into three regions: *(i)* high-performance,

(ii) optimum PDP, and *(iii)* optimum τ. In the first region, the load transistors have been sized up to achieve high-performance at the expense of higher τ and power consumption. The optimum PDP regions indicates the best tradeoff between performance and power consumption. In the last region, the small load transistors result in a significant increase in delay, which translates into an overall PDP increase. However, simultaneous reduction in power dissipation and τ due to smaller load transistors means the optimum MPDP design point is often found between the optimum PDP and the optimum τ region.

Figure 4.18 plots the τ vs. PDP curve of the reduced clock-swing flip-flops for both the TV and the LV analysis. For the same reasonings stated previously, the shape of the



(a) Transconductance Transistor Width Variation

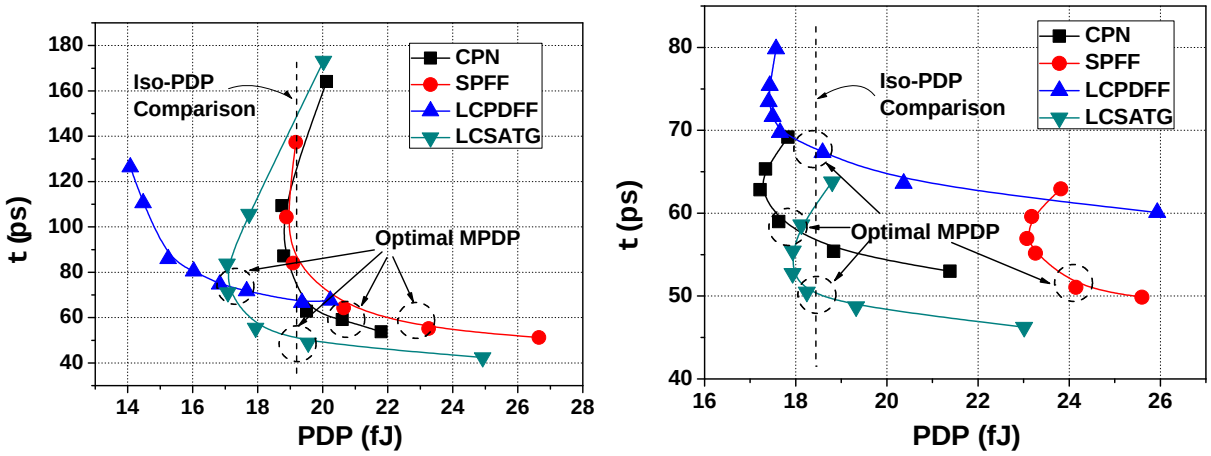
(b) Load Transistor Width Variation

Figure 4.18: Illustration of MPDP in Reduced Clock-Swing Flip-Flops using τ vs. PDP Curve via Transistor Sizing

curve for the NDKFF, the CRFF, and the RCSPDFF all indicate the optimum MPDP point is around the knee of the curve in the TV analysis. The curves in the TV analysis of the RCSSATG and the LV analysis of all the RCSFFs closely resemble the shape of

three regions identified previously: high-performance, optimum PDP, and optimum τ. As such, the optimum MPDP point in these cases occur somewhere between the optimum PDP and the optimum τ region. The less curvature in the LV analysis of the NDKFF and the corresponding high values of τ indicate the sizing scheme is limited due to the flip-flop architecture. While the TV analysis for the CRFF shows a comparable values of τ to those of the RCSPDFF and the NDKFF, the curve is more to the right than the other curves, which indicates its overall PDP and MPDP values are much higher than the other flip-flops. Although the PDP values of the RCSPDFF are the lowest among all the flip-flops, its values of τ are higher than those of the RCSSATG. Overall, it is clear that the RCSSATG is a more metastable-hardened flip-flop than the other flip-flops while maintaining a small PDP value. In the iso-PDP comparison of the TV analysis, the value of τ in the RCSSATG is 19% and 52% lower than the RCSPDFF and the NDKFF, respectively. In the LV analysis, the iso-PDP comparison shows the value of τ in the RCSSATG is 21% and 39% lower than the RCSPDFF and the NDKFF.

Figure 4.19 plots the τ vs. PDP curve of the level-converting flip-flops for both the TV and the LV analysis. The region of high-performance, optimum PDP, and optimum τ is clearly demonstrated in both the TV and the LV analysis of all the LCFFs except the TV analysis in the LCPDFF. Coincides with the previous observations, the optimum MPDP design point for most of the flip-flops occurs somewhere between the optimum PDP and the optimum τ region. Among the flip-flops analyzed, the LCSATG exhibits the best metastability along with the lowest PDP value. In the TV analysis, the iso-PDP comparison shows the value of τ in the LCSATG is 22%, 27%, and 42% lower than the CPN, the LCPDFF and the SPFF, respectively. At the iso-PDP point in the LV analysis, the τ of the LCSATG is 14% and 25% lower than the CPN and the LCPDFF.



(a) Transconductance Transistor Width Variation

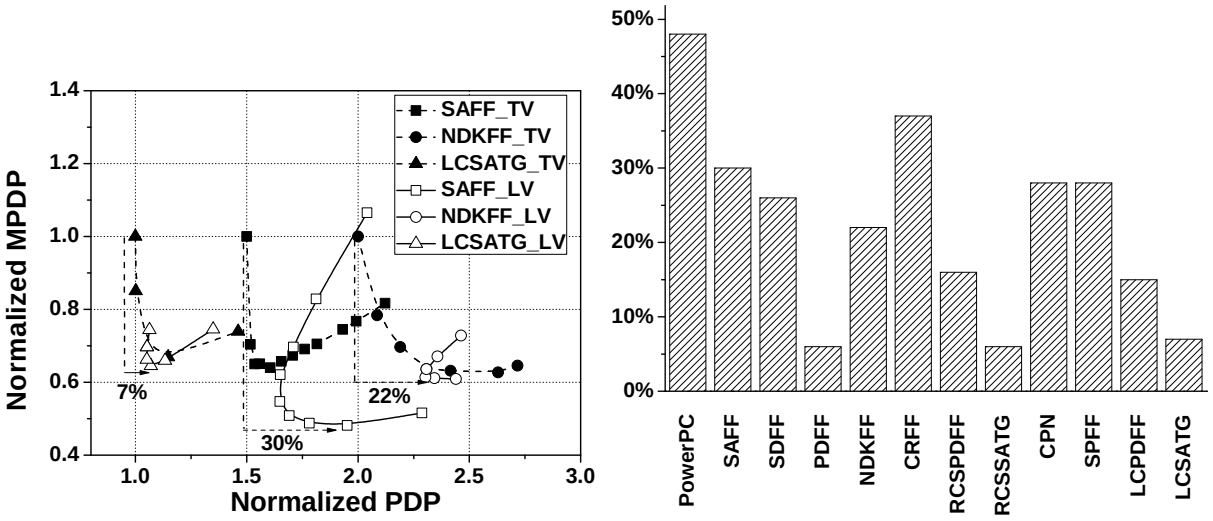
(b) Load Transistor Width Variation

Figure 4.19: Illustration of MPDP in Level-Converting Flip-Flops using τ vs. PDP Curve via Transistor Sizing

Key Remarks:

Because the concept of MDP and MPDP is first introduced in this work, we want to highlight a few key observations in using these new design metrics. First of all, the τ vs. delay and the τ vs. PDP curve allow the circuit designers to explore the design space for the tradeoff between τ, delay, and PDP based on the design requirements. If the delay is the most critical design consideration, one may sacrifice metastability and size the flip-flops to achieve the lowest delay value. Conversely, if reliability is the most important factor such as for systems in spacecrafts and medical equipments, then flip-flops maybe designed to sacrifice significant delay to achieve the optimum τ value for a specific MTBF value. The knee region of the τ vs. delay curve typically yields the best design tradeoff between τ and the flip-flop delay. A similar analysis can be performed for the tradeoff between τ and PDP. The second observation is that the location of the optimum MDP and MPDP point can

be different than the optimum PDP point in flip-flop designs. This means different sizing schemes must be adopted from the traditional PDP design in order to make the flip-flops more metastable-hardened. Thirdly, either the transconductance or the load variation may prove to be the more effective approach in obtaining the optimum MDP or MPDP design, depending on the flip-flop architecture. For instance, the TV method is more attractive for the PDFF mainly because the load in the slave-stage is small and therefore it is more effective to increase the g_m in the master-stage. On the other hand, minimize the load in the slave-stage reduces the τ significantly for the SAFF, and thus yielding the optimum MDP and MPDP design. At the iso-PDP region for the SAFF shown in **Figure 4.17**, the value of τ obtained in the LV analysis is 17% lower than the TV analysis.



(a) MPDP vs. PDP Plot

(b) % of PDP Increase Using the Optimum MPDP Design

Figure 4.20: Comparison between Optimum PDP and Optimum MPDP Designs

Generally, the optimum MPDP design involves sizing up the cross-coupled inverter pair

and/or reducing the size of the load transistors to optimize the value of τ, and hence the resultant PDP value will most likely be higher than the optimum PDP value due to the combination of delay and/or power increase. A small increase in PDP value indicates the flip-flop architecture is suitable for metastable-hardened designs because metastability performance can be improved dramatically without significant sacrifice in delay and power. This observation is illustrated in **Figure 4.20(a)** where the PDP vs. MPDP curve of three sample flip-flops is plotted for both the TV and LV analysis. While all the values are normalized to create enough separation between the curves in order to provide a clear illustration for each flip-flop, the relative difference is not affected by such normalization. For the NDKFF and the SAFF, the optimum MPDP value is obtained through the LV analysis and comes at the expense of an 22% and 30% increase, respectively, in PDP from the optimum PDP value. On the other hand, the LCSATG only encounters an 7% overhead in PDP when designed for optimum MPDP. **Figure 4.20(b)** shows the percentage increase in PDP from the optimum PDP value when the analyzed single and dual-supply flip-flops are designed for optimum MPDP. It is clear that the proposed flip-flop architectures of the PDFF and the SATG are more suitable for metastable-hardened designs, as indicated by the smallest PDP overhead among all the flip-flops analyzed. Overall, the PDP vs. MPDP curve is an useful illustration to analyze the amount of tradeoff in PDP when designed for optimum MPDP. A similar analysis can also be performed for the amount of tradeoff between delay and MDP using the delay vs. MDP curve.

4.4 Post-Layout Simulation Results

4.4.1 Test Bench and Measurement Setup

All the flip-flops analyzed in this chapter are implemented in layout in the $0.18\mu m$ TSMC technology using the optimum MPDP design. The flip-flop layouts can be found in **Appendix A**. The values of all the delay, τ, PDP, MDP, and MPDP given in this work are the worst case value of either the 0-1 or 1-0 data transition. Hence, the PDP, MDP, and MPDP values shown may not necessary be the product of the delay, τ, and power given in the tables. We have chosen 25% data activity factor for power consumption measurement. The simulation test bench setup is identical to the one shown in **Figure 3.15**. The method for extraction of the metastability parameters τ and T_0 is identical to the one described in Chapter 2. The flip-flop area refers to the total transistor widths. All the dual-supply flip-flops are designed under the optimum MPDP scheme specifically for $V_{DDL} = 1.3V$, which is approximately $0.7V_{DDH}$. Once again, the post-layout simulation values of τ, on average, are approximately 10% higher than the schematic simulation results.

4.4.2 Flip-Flops in Single-Supply Systems

Table 4.5 summarizes the simulation results for the analyzed single-supply flip-flops. As evident from the data, the value of τ for the SAFF, the PowerPC, and the PDFF is very similar to each other while the SDFF is approximately 1.37× higher than the SAFF. The high-performance characteristic of the PDFF is demonstrated by the fact that its delay is at least 22% lower than the other flip-flop architectures. With fewer transistors in the critical path, this performance advantage does not come at the expense of significant increase in area or power consumption. The power consumption of the PDFF is only 13% higher

than the PowerPC but 3% and 53% lower than the SAFF and the SDFF, respectively. The total transistor widths of the PDFF is also the lowest among all the flip-flops when designed for optimum MPDP. Overall, PDFF the offers the best design tradeoff between delay, power and metastability, as evident by a 32%, 42%, and 34% reduction in PDP, MDP, and MPDP, respectively, from the next lowest values.

Table 4.5: Simulation Results for Optimum MPDP Designed Single-Supply Flip-Flops

	Delay (ps)	Power (μW)	τ (ps)	PDP (fJ)	MDP (ps^2)	MPDP (fJ·ps)	Area (μm)
SAFF	307.8	109.6	43.9	33.7	13520	1482.1	50.75
PowerPC	246.1	93.76	44.4	23.1	10595	993.3	29.65
SDFF	190.3	224.0	61.0	42.6	10951	2453.4	41.75
PDFF	148	106.4	41.5	15.7	6150	654.5	27.9

Figure 4.21 illustrates the design tradeoff comparison between the optimum PDP and MPDP design for the analyzed single-supply flip-flops. The percentage indicated in the figure refers to the amount of increase (+) or decrease (-) that results from the optimum MPDP design when compared to the optimum PDP design in terms of delay, power, τ and area. While it is true that similar τ values can be achieved in the SAFF, the PowerPC, and the PDFF under the optimum MPDP design, the amount of tradeoff in terms of other design criteria can vary significantly between them. The optimum MPDP design for the SAFF is able to achieve 54%, 12%, and 32% reduction in τ, power, and area respectively but an 39% increase in delay. The usage of smaller transistors in the slave-stage not only reduces τ but also results in lower power dissipation and smaller area. In the PowerPC, the 37% reduction in τ comes at the expense of 30%, 3%, and 23% increase in delay, power,

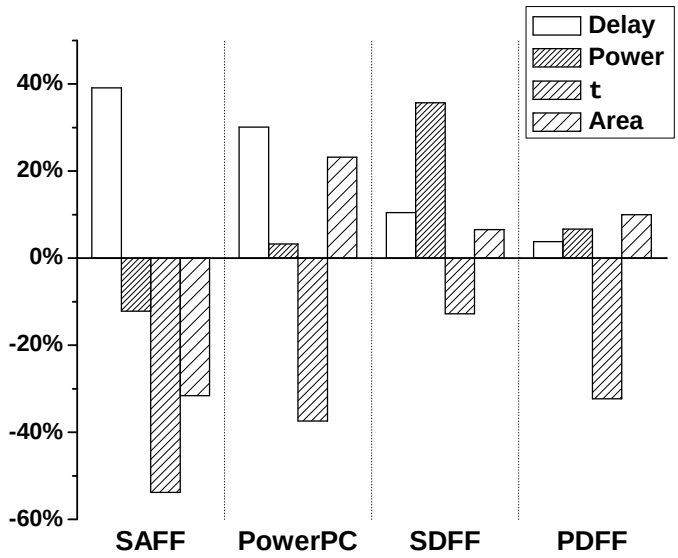


Figure 4.21: Comparison and Analysis between the Optimum PDP and the Optimum MPDP Design for Single-Supply Flip-Flops

and area. This is largely due to sizing up the transistors in the feedback path, which not only increases power and area but also adds capacitances to the critical path to degrade performance. The 13% reduction of τ in the SDFF is achieved by sizing up the inverter pair to stabilize the critical node, but again this translates into 10%, 35%, and 7% increase in delay, power, and area. While the reduction of τ in the PDFF is 32% from the optimum PDP to the optimum MPDP design, the amount of increase in delay, power, and area is all less than 10%, which is significantly less than all the other flip-flops. This suggests the architecture of the PDFF with a cross-coupled inverter pair in the critical path of the master-stage and small load in the slave-stage is very suitable to achieve good metastability without much compromise in delay, power, and area.

Although the majority of analysis in this work is focused on τ, metastability window δ is often used as the main parameter in measuring metastability instead of the mean-time-

between-failure (MTBF) since it is independent of the data and clock frequency, which are determined by the system. **Figure 4.22** plots the metastability window as a function of the settling time (t_s). Two sets of δ values are plotted for each flip-flop: the optimum PDP and the optimum MPDP design. In the optimum PDP designs, the δ of the PDFF is a minimum 1-3 orders of magnitude lower than the other flip-flops. Thus, the architecture of the PDFF is able to achieve good metastability without much optimization when compared to other flip-flops. With the optimum MPDP design, it is very clear that the metastability window is at least few orders of magnitude lower than the optimum PDP design for all the flip-flops because the effect of reducing τ is magnified by its exponential relationship with δ. δ of the PowerPC reduces by three orders of magnitude when t_s is 600ps and six orders of magnitude lower when the settling time increases to 1000ps. While the metastability window of the SDFF is higher than the other flip-flops analyzed in this work, using the optimum MPDP design still achieves a few orders of magnitude lower δ than the optimum PDP design. All in all, the significant reduction in δ using the optimum MPDP design flip-flops greatly reduces the likelihood of the flip-flops with unresolved data in the metastable region.

4.4.3 Reduced Clock-Swing Flip-Flops

Table 4.6 summarizes the simulation results for the analyzed reduced clock-swing flip-flops at $V_{DDL} = 1.3V$. Despite the lowest value of D-Q delay, the τ of the RCSPDFF in dual-supply systems is higher than its counterpart in the single-supply system due to the exponential increase when the clock-swing is reduced. Similar to the SDFF, this is an example of suggesting high-performance flip-flops do not necessarily result in fast time resolving constant τ. It is the circuit architecture that largely determines the flip-flop metastability

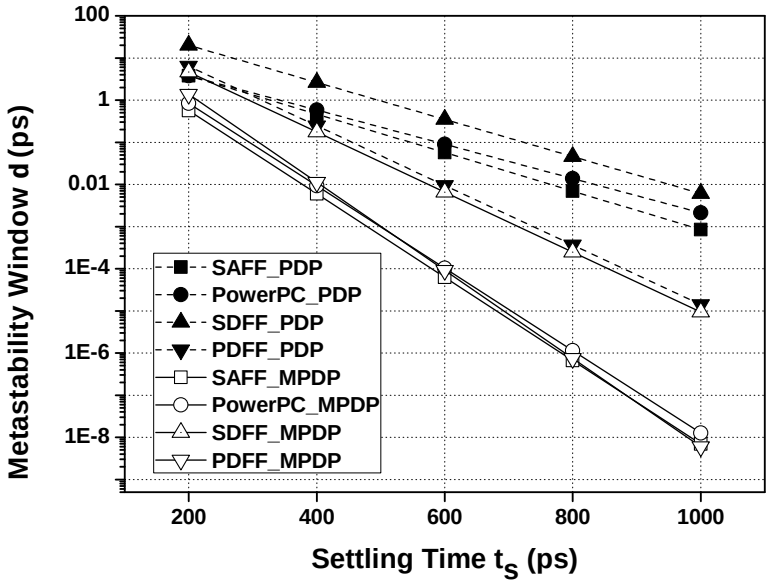


Figure 4.22: Metastability Window Analysis for Single-Supply Flip-Flops

behavior. Despite a higher value of τ, the MPDP value of the RCSPDFF is still 11% and 28% lower than the NDKFF and the CRFF largely due to its high-performance and low characteristics that yield a minimum 8% lower PDP than the other analyzed flip-flops. The post-layout results have clearly shown that the RCSSATG demonstrates the best design tradeoff between delay, power, and metastability in the dual-supply system. While its overall delay and power consumption are almost identical to those of the RCSPDFF such that the overall PDP value is only 8% higher, the value of τ is 40% lower than the RCSPDFF. Subsequently, the RCSSATG achieves a minimum 32% and 35% reduction in MDP and MPDP, respectively, when compared to the other flip-flops. Because the architecture of the RCSPDFF and the RCSSATG are suitable for metastable-hardened flip-flop designs, therefore the total transistor widths of these flip-flops under the optimum MPDP design are much lower than the NDKFF and the CRFF.

Table 4.6: Simulation Results for Optimum MPDP Designed Reduced Clock-Swing Flip-Flops at $V_{DDL} = 1.3V$

	Delay (ps)	Power (μW)	τ (ps)	PDP (fJ)	MDP (ps^2)	MPDP (fJ·ps)	Area (μm)
NDKFF	247.9	73.487	75.893	18.217	17559.5	1382.6	44
CRFF	342	70.351	71.492	24.06	24450.3	1720.1	42.45
RCSPDFF	201.6	68.11	83.612	14.8	18168.9	1237.5	32.3
RCSSATG	215.8	66.886	50.4	15.966	12031.9	804.77	34.25

Figure 4.23 illustrates the design tradeoff comparison between the optimum PDP and the optimum MPDP design for the reduced clock-swing flip-flops. Under the optimum MPDP design, the 13% reduction of τ in the NDKFF is achieved by sizing up the feedback transistors and reducing the size of the load transistors. While the effect on area and power consumption is almost negligible, the optimum MPDP design results in a 40% delay penalty when compared to the optimum PDP design. In the CRFF, the 45% reduction in τ from sizing up the cross-coupled inverter pair comes at the expense of 27%, 14%, and 14% increase in delay, power, and area. While its overall value of τ is higher, the architecture of the RCSPDFF is still suitable for metastable-hardened flip-flop designs, as evident by 14%, 8%, and 18% increase in delay, power, and area while achieving a 35% reduction in τ when comparing the optimum MPDP and the optimum PDP designs. Finally, the RCSSATG encounters the smallest overhead under the optimum MPDP design where the amount of increase in delay, power, and area from the optimum PDP design is only approximately 3% along with a 27% reduction in τ.

Figure 4.24 plots the metastability window δ as a function of the settling time (t_s) for

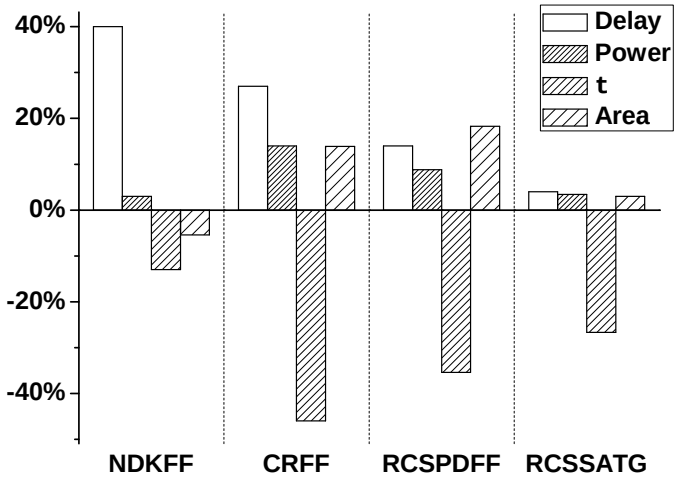


Figure 4.23: Comparison between Optimum PDP and Optimum MPDP Design for Reduced Clock-Swing Flip-Flops at $V_{DDL} = 1.3V$

the reduced clock-swing flip-flops. Two sets of δ values are plotted for each flip-flop: the optimum PDP and the optimum MPDP design. In general, the optimum MPDP design of all the flip-flops has reduced the metastability window δ by at least an order of magnitude from the optimum PDP design. The δ of the RCSSATG under the optimum PDP design is at least one magnitude lower than the other flip-flops under both the optimum PDP and MPDP designs. The optimum MPDP design of the RCSSATG has resulted in a further two orders of magnitude reduction in δ from the optimum PDP design. The significant reduction of the metastability window clearly indicates the architecture of the RCSSATG is very desirable for designing metastable-hardened flip-flops.

4.4.4 Level-Converting Flip-Flops

Table 4.7 summarizes the simulation results for the analyzed level-converting flip-flops at $V_{DDL} = 1.3V$ designed for optimum MPDP. With similar architectures, the values of the

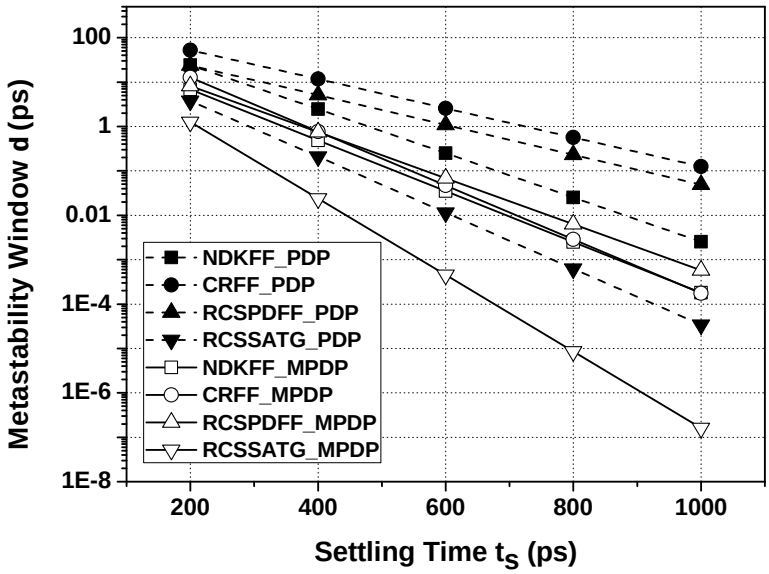


Figure 4.24: Metastability Window Analysis for Reduced Clock-Swing Flip-Flops at $V_{DDL} = 1.3V$

LCPDFF and the LCSATG are very similar to those of the RCSPDFF and the RCSSATG. The high-performance and low-power characteristics of the PDFF architecture resulted in a minimum 13% lower PDP value than the other LCFFs. Its value of τ, however, is 60% higher than the LCSATG. Despite a 15% higher delay value, the overall MDP and MPDP value of the LCSATG is both 28% lower than the LCPDFF. Once again, the total transistor widths of both the LCPDFF and the LCSATG are the lowest among the analyzed flip-flops.

Figure 4.25 illustrates the design tradeoff comparison between the optimum PDP and the optimum MPDP design for the level-converting flip-flops. Due to transistor stacking in the discharge paths, the reduction of τ in the CPN and the SPFF under the optimum MPDP design is only limited to 10% and 15%, respectively, while encountering delay overhead of 27% and 19%. All of these considerations indicate the architecture of the CPN

Table 4.7: Simulation Results for Optimum MPDP Designed Level-Converting Flip-Flops at $V_{DDL} = 1.3V$

	Delay (ps)	Power (μW)	τ (ps)	PDP (fJ)	MDP (ps^2)	MPDP (fJ·ps)	Area (μm)
CPN	303.7	70.174	67.355	21.312	20455.71	1435.46	39.15
SPFF	258.4	72.627	90.726	18.767	20912.34	1702.64	56.2
LCPDFF	223	60.048	81.517	13.39	18178.3	1091.57	31.3
LCSATG	256.7	59.808	51.04	15.353	13102.5	783.63	34.65

and the SPFF is not very attractive for metastable-hardened flip-flop design because the benefit of reducing τ using the optimum MPDP design is outweighed by the overhead in other design considerations. Conversely, the optimum MPDP design in the LCPDFF and the LCSATG achieves 25% and 34% reduction in τ when compared to the optimum PDP design while keeping the overhead in other design considerations less than 10%.

Figure 4.26 plots the metastability window δ as a function of the settling time (t_s) for the level-converting flip-flops. Once again, two sets of δ values are plotted for each flip-flop: the optimum PDP and the optimum MPDP design. With less than 15% reduction in τ, the amount of reduction in δ for both the CPN and the SPFF is less than an order of magnitude under the optimum MPDP design scheme. Depending on the value of the settling time, the optimum MPDP design of the LCSATG can achieve up to three orders of reduction in τ when compared to the optimum PDP design.

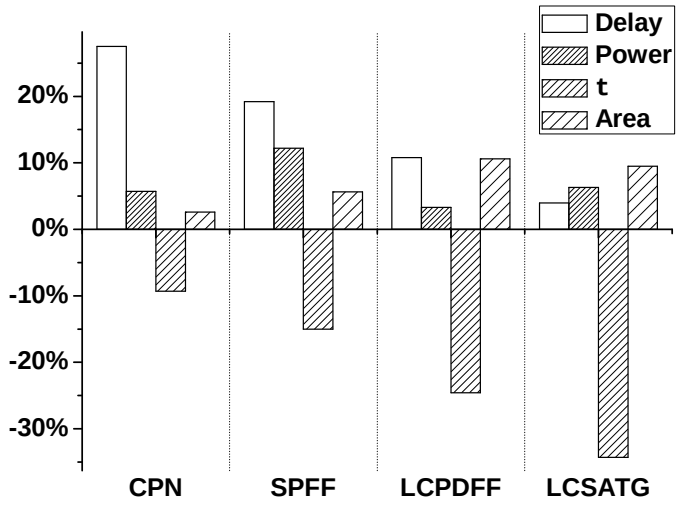


Figure 4.25: Comparison between optimum PDP and optimum MPDP Design for Level-Converting Flip-Flops at $V_{DDL} = 1.3V$

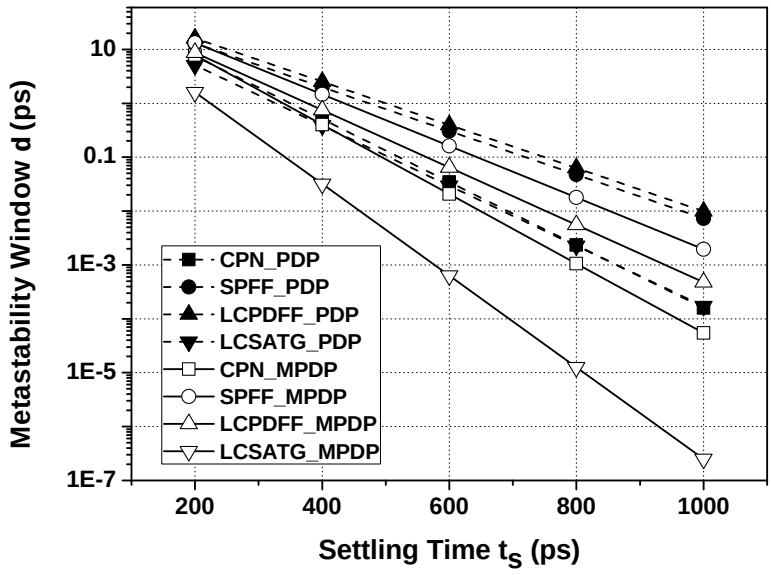


Figure 4.26: Metastability Window Analysis for Level-Converting Flip-Flops

4.5 Metastability in the Sub-Threshold Region

Recently, significant research effort has been made on sub-threshold circuit designs in order to facilitate ultra-low-power applications such as sensor networks, bio-implantables, and RFID tags. Previous work has shown the optimum energy operation occurs in the sub-threshold region where the supply voltage V_{DD} is less than the threshold voltage V_{th} [65][66][67]. While the concept of energy harvesting is attractive in sub-threshold designs, reliability issues should not be overlooked in order to maintain reliable system operations. Past works have analyzed flip-flops in the sub-threshold region in terms of delay and power [68] as well as variability under process variation [69][70]. In this section, flip-flop metastability is analyzed in the sub-threshold region using the optimum MPDP design as well as a proposed mixed-V_{th} technique. We will refer the region where $V_{DD} < V_{th}$ as the sub-threshold region and $V_{DD} > V_{th}$ as the super-threshold region.

Similar to the super-threshold region, the key in designing metastable-hardened flip-flops in the sub-threshold region is to optimize the time resolving constant τ. However, the transconductance equation [41] in the sub-threshold region is dramatically different than the super-threshold region. The current equation in the sub-threshold region is given by

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} (n-1) V_T^2 e^{\frac{V_{GS}-V_{th}}{nV_T}} \tag{4.8}$$

where V_T is the thermal voltage, n is the sub-threshold slope, and V_{GS} is typically close to $V_{DD}/2$ during metastability. Furthermore, the transconductance g_m in the sub-threshold region can be calculated by **Equation (4.9)**.

$$g_m = \frac{I_{DS}}{nV_T} \tag{4.9}$$

Figure 4.27 plots the transconductance g_m and τ as a function of the supply voltage V_{DD} in log-scale on the Y-axis. Simple derivations from **Equation (4.8)** and **(4.9)** will reveal that

g_m has an exponential relationship with V_{GS} in the sub-threshold region. Consequently, this also translates into an exponential relationship between the time-resolving constant τ and V_{DD}. With the exponential relationship, it is evident that slight variation in V_{DD} and/or g_m can result in significant changes in the value of τ.

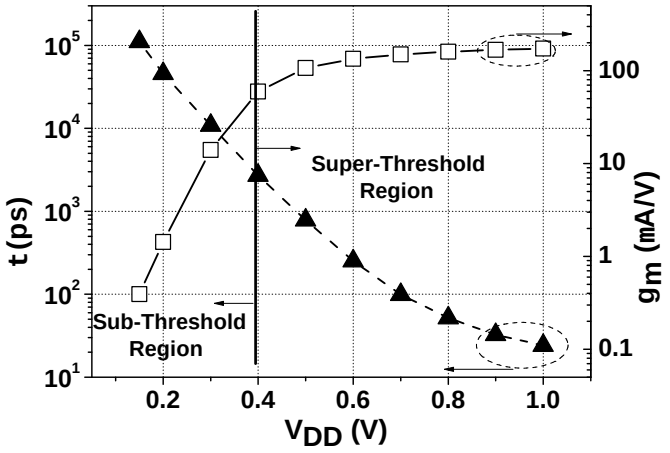


Figure 4.27: Plot of τ and g_m as a Function of V_{DD}

Three flip-flops are chosen for metastability analysis in the sub-threshold region (PowerPC, SAFF, and PDFF) using the TSMC 65nm CMOS technology. The test bench used in the sub-threshold region is similar to the one previously described but with some differences. Four different supply voltage values are used for sub-threshold region analysis: 0.15V, 0.2V, 0.3V, and 0.4V. The clock frequency used for extraction of τ and delay measurement is 300KHz to ensure the output is given enough time to settle to a stable value when it is in the metastable region. In the sub-threshold region, clock frequency (f_{CLK}) has a significant impact on the power measurement because the distribution of dynamic and leakage power can be significantly different. Unless specifically mentioned, the average power is measured over 100 clock cycles by assuming $f_{CLK} = 10t_d$ where t_d represents the worst flip-flop delay for a given supply voltage. When extracting τ, a step size of 1ps is

used in manipulating the data arrival time with respect to the CLK.

While the design for optimum MPDP via transistor sizing is identical to the super-threshold region described previously, a mixed-V_{th} technique will be demonstrated to significantly reduce τ and be more energy efficient than the single standard-V_{th} design if the appropriate supply voltage is selected. As seen from **Equation (4.8)-(4.9)**, lowering the threshold voltage V_{th} in the sub-threshold region results in an exponential increase in g_m. The proposed design methodology in the sub-threshold region is to apply low-V_{th} transistors only on the inverter pair that stabilizes the critical node in order to increase g_m while the remaining circuit uses standard-V_{th} transistors. The low-V_{th} transistors are identical to those listed in **Table 4.1** for the selected flip-flops except W_{p1} and W_{n1} in the PowerPC are also low-V_{th} transistors since they are part of the inverter pair.

Figure 4.28 shows the g_m comparison for a standard-V_{th} and low-V_{th} NMOS/PMOS transistor respectively at V_{DD} ranging from 0.1V to 1V. In the sub-threshold region, low-V_{th}

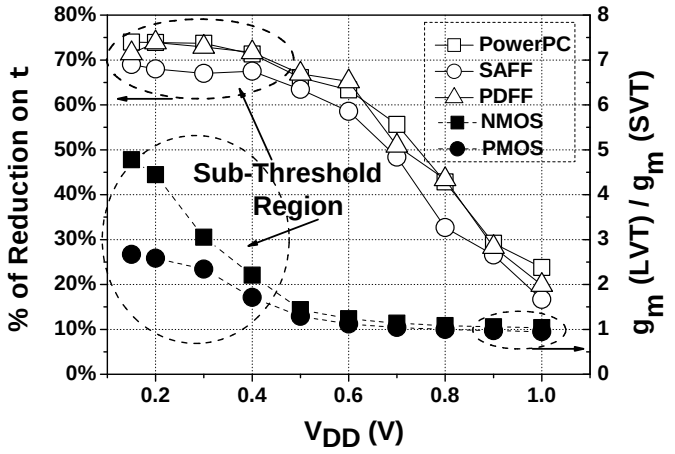


Figure 4.28: Impact of Mixed-V_{th} Design on g_m and τ

NMOS and PMOS result in a minimum 2.2× and 1.7× increase in g_m than the standard-V_{th}

transistors, which is much more significant when compared to the super-threshold region. This suggests using the mixed-V_{th} technique in reducing τ is much more effective in the sub-threshold region than the super-threshold region. For all three flip-flops analyzed, using low-V_{th} transistors results in a minimum of 67% reduction in τ (**Figure 4.28**) for a given V_{DD} in the sub-threshold region.

While the mixed-V_{th} design can significantly reduce τ in the sub-threshold region, its power-delay-product (PDP) must be carefully analyzed to determine if such design is still energy efficient. Using the PowerPC as an example, **Figure 4.29** illustrates the τ vs. PDP plot for both the single-V_{th} (SVT) and mixed-V_{th} (MVT) design at two different clock frequencies under different sub-threshold supply voltages. For the most part, the PDP of the MVT design is very comparable to the SVT design at a given V_{DD} value because the increased power consumption due to low-V_{th} transistors is compensated by an improved performance. At $V_{DD} = 0.15V$, the PDP of the MVT design is only ≈10% higher than the SVT design for both clock frequencies. For V_{DD} above 0.2V, the PDP of the MVT design is about 10% and 13% lower than the SVT design at the slower and faster clock frequency respectively. For iso-PDP comparison, the MVT design is able to achieve significant reduction in τ than the SVT design. **Figure 4.29** also shows that at extremely low supply voltage (i.e. $V_{DD} \leq 0.2V$), the SVT design can be more energy efficient and equally metastable-hardened than the mixed-V_{th} design by selecting an appropriate V_{DD} value. At the iso-τ region shown in the figure, the SVT design at 0.3V has a lower PDP value than the MVT design at 0.2V. For $V_{DD} > 0.2V$, however, MVT design becomes more energy efficient for iso-τ comparison. The cross-over point between the two curves determines the region where the MVT design becomes more energy efficient than the SVT design for iso-τ comparison. Overall, the MVT design is an attractive method to design metastable-hardened flip-flops in the sub-threshold region without much energy

consumption penalty, especially in designs where the supply voltage is fixed at a given value.

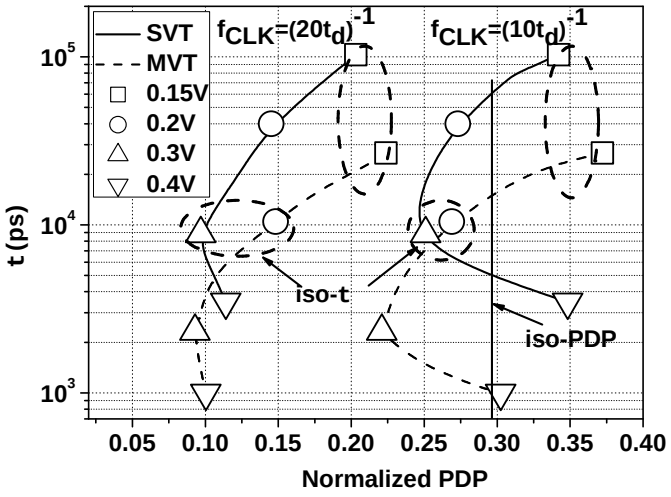


Figure 4.29: Comparison between Single-V_{th} and Mixed-V_{th} Flip-Flop Design

The flip-flops analyzed in this work have been implemented in layout using two different designs: *(i)* optimum PDP sizing using standard-V_{th} transistors and *(ii)* optimum MPDP sizing using mixed-V_{th} transistors. **Figure 4.30** shows the post-layout simulation results of the τ vs. PDP curves. In the PDFF, optimum MPDP design achieves a 4× to 5× reduction in τ while maintaining the same PDP as the optimum PDP design for a given V_{DD}. While the reduction in τ ranges from 6× to 9× in the PowerPC for the optimum MPDP design, it comes with a 10-15% increase in PDP for $V_{DD} \leq 0.2V$. In the SAFF, the delay improvement gained from the optimum MPDP design results in lower PDP values along with a 2× to 7× lower τ than the optimum PDP design for a given V_{DD}. Overall, the optimum MPDP design in iso-PDP comparisons achieves significant reduction in τ for each flip-flop while becoming more energy efficient than the optimum PDP design for $V_{DD} > 0.2V$ in the iso-τ comparison. Under the optimum MPDP design, the value of τ

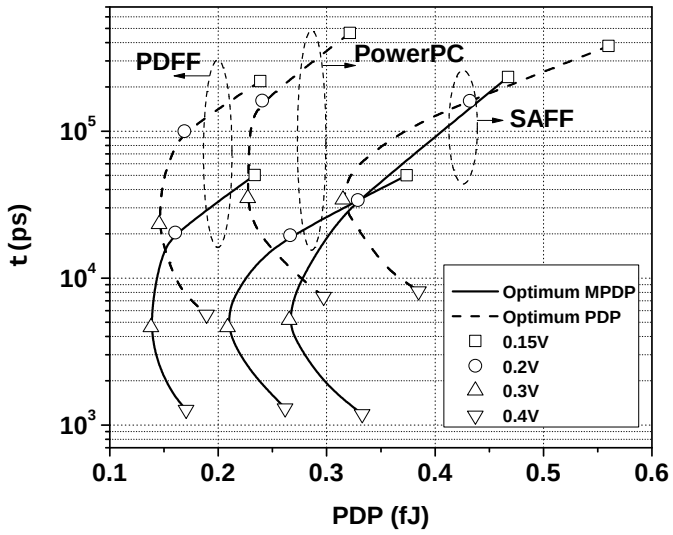


Figure 4.30: τ vs. PDP Curve for Post-Layout Simulation

for all three flip-flops is very similar for $V_{DD} > 0.2V$. At 0.15V and 0.2V, respectively, the τ of the SAFF is approximately 1.7× and 4.6× higher than the PowerPC and the PDFF. This indicates the impact of stacking three NMOS transistors in the inverter pair on τ is much more significant at extremely low voltages. Despite the similarities in τ, the PDP of the PDFF is much lower than the the PowerPC and the SAFF when performing iso-τ comparisons across all supply voltage values. This result coincides with an earlier observation where the architecture of PDFF is able to achieve more balanced design tradeoff between τ, delay, and power, as evident by the lowest MPDP values listed in **Table 4.8** for all supply voltages.

Table 4.9 shows the impact of process variations on τ in the sub-threshold region for the flip-flops analyzed across five different process corners. When compared to the TT corner, the FF and the SS corner results in approximately 2.8× to 3.3× improvement and 3.4× to 3.8× degradation, respectively, in τ for all flip-flops at voltages ranging from 0.2V

Table 4.8: Post-Layout Simulation Results of MPDP ($fJ \cdot ns$) in the Sub-Threshold Region

	0.15V	0.2V	0.3V	0.4V
PowerPC	18.67	5.20	0.96	0.34
SAFF	109.00	11.147	1.373	0.395
PDFF	11.74	3.27	0.64	0.22

to 0.4V. At 0.15V, the effect of these corners becomes more prominent in the PowerPC and the SAFF where the master-stage consists of stacked transistors in the inverter pair. Because the inverter pair in the PDFF consists of a single transistor to V_{DD} and V_{SS}, the impact of the SS and the FF corners is mostly consistent across all supply voltages relative to the TT corner. Due to the three NMOS transistors stacked in series in the master-stage, the impact of the SF corner results in higher τ value than the FS corner in the SAFF, especially at 0.2V and below. For the PDFF, the τ value for the FS and the SF corners does not deviate too much from the TT corner because the PMOS and NMOS transistors under different process variations in the inverter pair compensate each other. In the PowerPC, the effect of stacking two transistors in series becomes more prominent at lower supply voltages, and results in greater deviation of τ between the FS and the SF corner.

4.6 Impact of Technology Scaling on Metastability

Using the predictive technology modeling (PTM) provided by [64], this section examines the impact of technology scaling on the metastability time-resolving constant τ for advanced technologies below the 65nm regime. In this work, we consider two types of

Table 4.9: Post-Layout Simulation Results of τ (ns) in the Sub-Threshold Region under Different Process Corners at 27°C

	PowerPC				SAFF				PDFF			
	0.15V	0.2V	0.3V	0.4V	0.15V	0.2V	0.3V	0.4V	0.15V	0.2V	0.3V	0.4V
FF	15.69	5.90	1.50	0.43	29.67	7.44	1.48	0.43	16.20	6.35	1.48	0.46
FS	92.88	33.89	5.93	1.58	137.1	22.22	3.94	1.12	75.32	22.51	4.91	1.37
TT	49.94	19.54	4.63	1.30	233.19	33.91	5.17	1.19	50.21	20.38	4.62	1.27
SF	44.8	17.45	5.09	1.52	466.4	61.39	7.40	1.43	51.30	20.28	4.87	1.30
SS	267.7	73.78	17.92	4.67	1283.9	128.9	18.72	4.00	189.5	75.13	17.12	4.83

CMOS technology: *(i)* CMOS bulk technology with high-K/metal gate (MGHK) , and (ii) CMOS bulk technology with high-K/metal gate and strained-silicon (Strained-Si) . In sub-100nm regime, MOSFETs with strained-Si structures are promising for high-performance, low-power CMOS applications because of the high electron and hole mobility caused by strained-induced band splitting [71]. The MGHK technology model files are available from 65nm to 22nm while the Strained-Si model files are available from 45nm to 16nm. For the Strained-Si model, we have chosen the high-performance (HP) kit over the low-power (LP) kit to engage a fair comparison with the MGHK model.

While the gate delay is expected to be reduced by 30% for each generation of technology scaling [72], the value of τ may not necessarily scale by the same amount. Based on **Equation (2.10)**, a first order approximation reveals a general equation of τ if we simplify the capacitance value equal to the gate capacitance given by **Equation (4.2)** and a loop transconductance (g_m) equal to the summation of the transconductance of the NMOS

(g_{mn}) and the PMOS (g_{mp}) transistor, as given by **Equation (4.1)**.

$$\begin{aligned}\tau \approx \frac{C}{g_m} &\approx \frac{\frac{2}{3}WL_{eff}C_{ox}}{g_{mn}+g_{mp}} \\ &= \frac{\frac{2}{3}WL_{eff}C_{ox}}{\mu_n(V_{gs}-V_{tn})+\mu_p(V_{gs}-V_{tp})} \\ &= \frac{\frac{2}{3}{L_{eff}}^2}{\mu_n(V_{gs}-V_{tn})+\mu_p(V_{gs}-V_{tp})}\end{aligned} \tag{4.10}$$

$V_{DD}/2$ is substituted into the equation for V_{gs} to emulate the critical node voltage during metastability [73]. To include the effects of the short-channel and the drained-induced barrier lowering, **Equation (4.11)** is used to calculate the threshold voltage V_{tn} and V_{tp} across different technologies from the model file.

$$V_{th} = V_{th0} - V_{ds} \cdot exp(-\frac{L_{eff}}{l'}) \tag{4.11}$$

l' is the characteristic length that can be derived by **Equation (4.12)** [74] where each parameter can be found in the model file.

$$l' = \sqrt{\frac{\epsilon_{si} \cdot TOXE \cdot X_{dep}}{EPSROX \cdot \eta}} \tag{4.12}$$

Table 4.10 displays the values of the calculated threshold voltage as well as the supply voltage and the electron and hole mobility provided by the model files for the two types of transistor models. As expected, the electron and hole mobility of the Strained-Si devices are much higher than those of the MGHK devices. By inputting these values into **Equation (4.10)**, **Figure 4.31** plots the calculated values of τ for both the MGHK and the Strained-Si devices. All MGHK values are normalized with respect to the 65nm node and all the Strained-Si values are normalized to the 45nm node. Because the values shown are approximated using the first-order equation, the absolute values are not as important as the relative values. For the MGHK model, an infliction point is observed at the 32nm

Table 4.10: Device Parameters for Different Technology Nodes

		MGHK				Strained-Si			
	V_{DD} (V)	μ_n $(m^2/V \cdot s)$	μ_p $(m^2/V \cdot s)$	V_{tn} (V)	V_{tp} (V)	μ_n $(m^2/V \cdot s)$	μ_p $(m^2/V \cdot s)$	V_{tn} (V)	V_{tp} (V)
65nm	1	0.049	0.006	0.368	0.297	N/A	N/A	N/A	N/A
45nm	1	0.044	0.004	0.376	0.307	0.054	0.02	0.404	0.413
32nm	0.95	0.039	0.003	0.386	0.31	0.05	0.014	0.4	0.383
22nm	0.9	0.018	0.002	0.408	0.23	0.04	0.01	0.38	0.326
16nm	0.8	N/A	N/A	N/A	N/A	0.03	0.006	0.341	0.28

node where the τ of the 22nm node is higher than the 32nm. In the case of Strained-Si, however, no infliction point is observed as τ continues to decrease from the 45nm node to the 16nm node, although the value calculated at the 22nm and the 16nm node is very similar. The theoretical model derived from **Equation (4.10)** can provide some insights into this phenomena. From previous analysis, it is clear that τ has an inverse relationship with the transconductance g_m, which is a function of the overdrive voltage $V_{gs} - V_{th}$. For analysis purposes, we can assume V_{th} is the sum of V_{tn} and V_{tp}. During metastability, the overdrive voltage is around $V_{DD} - V_{th}$ in a cross-coupled inverter pair. With rapid technology scaling, the value of V_{DD} is decreasing faster than V_{th} because the latter cannot be scaled as aggressive for reasons such as suppressing the leakage power. Therefore, the value of V_{DD} is quickly approaching the value of V_{th}, and at the same time reducing the effective value of g_m. The other parameters in **Equation (4.10)**, L_{eff} and μ, are also contributing to the calculation of τ. As L_{eff} is decreasing by a factor of 0.7 for each

technology generation, the numerator of the equation is decreasing in a quadratic manner. However, the hole and electron mobility in the denominator of the equation is also scaling with the technology. At the 22nm node of the MGHK model, the amount of reduction in the denominator exceeds the amount of reduction in the numerator when compared to the 32nm node, and thus results in an infliction point at the 32nm node. For the Strained-Si model, the higher hole and electron mobility values contribute to the continuous decrease in τ to the 16nm node, and thus no infliction point is observed.

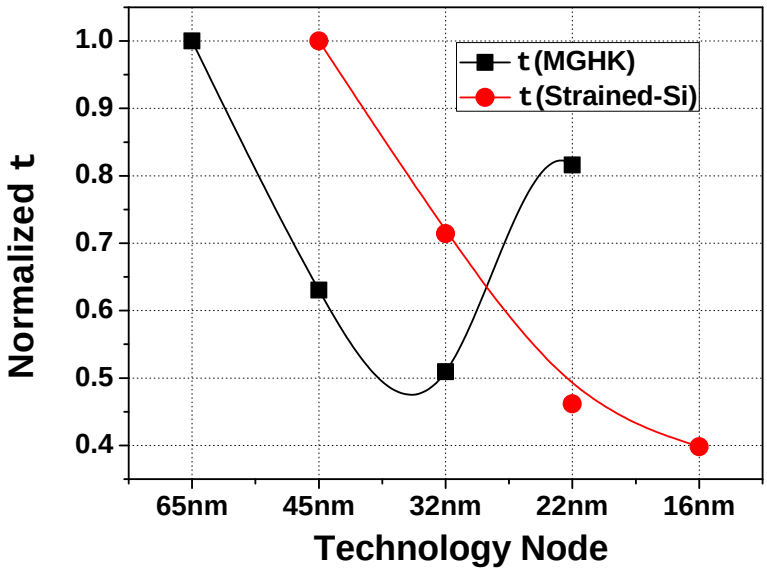
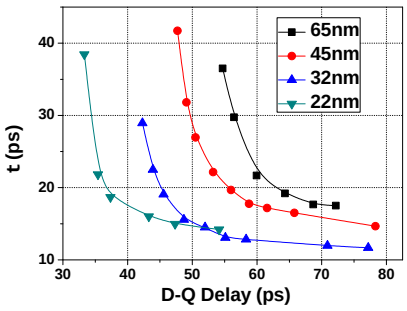
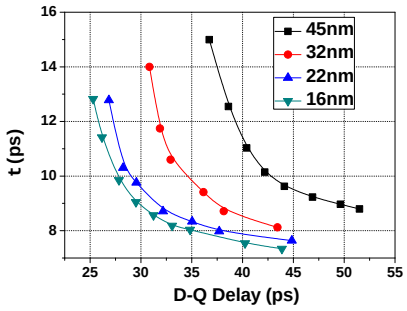


Figure 4.31: Impact of Technology Scaling on τ

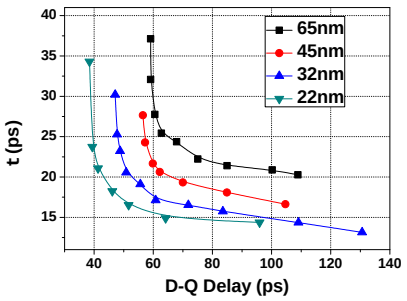
To verify the observations from **Figure 4.31**, Spice simulations were performed on three flip-flop architectures (PowerPC, SAFF, and PDFF) using the model files for both the MGHK and the Strained-Si technology. For each technology node, both the transconductance variation (TV) and the load variation (LV) analysis are performed on each flip-flop via transistor sizing. Subsequently, the τ vs. delay curves (**Figure 4.32**) are plotted to determine the optimum MDP point using either the TV or the LV analysis. As evident



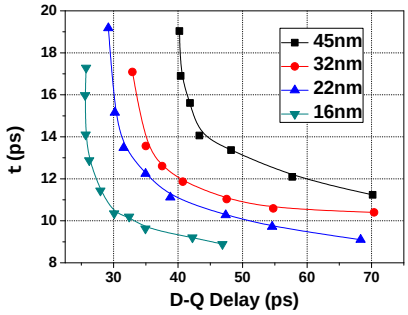
(a) MGHK Results for PowerPC



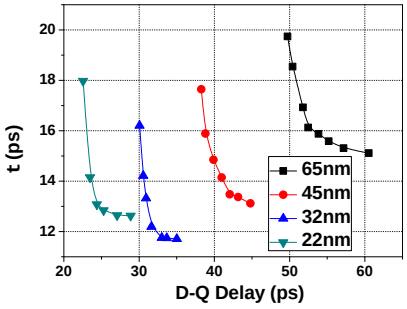
(b) Strained-Si Results for PowerPC



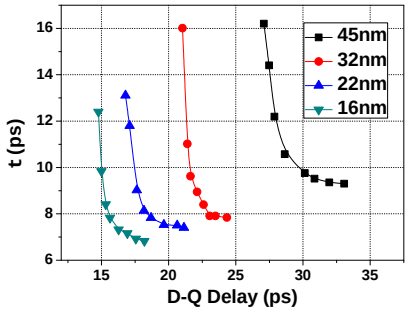
(c) MGHK Results for SAFF



(d) Strained-Si Results for SAFF



(e) MGHK Results for PDFF



(f) Strained-Si Results for PDFF

Figure 4.32: Simulation Results of τ for Flip-Flops in MGHK and Strained-Si Technology

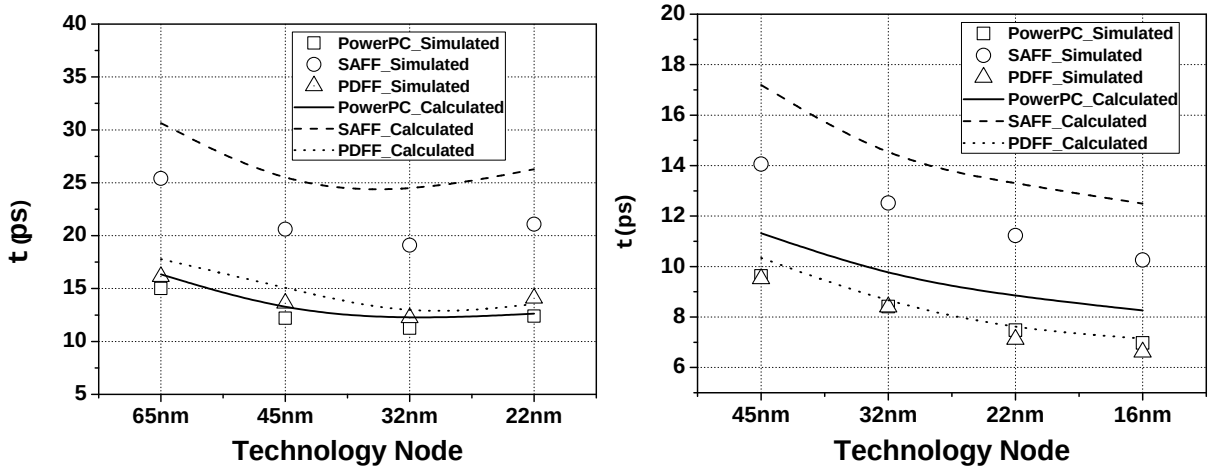
from **Figure 4.32**, the D-Q delay of the flip-flops is decreasing with each technology generation for both models. However, the behavior of τ varies. For example, the τ of all the flip-flop in the 22nm MGHK model is higher than the 32nm, and thus coincides with the earlier observation of the infliction point. Similarly, the τ of all the flip-flops in the 16nm Strained-Si model is lower than the 22nm.

To obtain a fair comparison, the value of τ for each respective flip-flop obtained at the optimum MDP design point for each technology node in both models is used for comparison and analysis. In addition, the transistor sizes at the optimum MDP point are used to calculate a set of theoretical values of τ based on the methodology described in Section 4.3.2. **Figure 4.33** shows both the simulated and the calculated values of τ at the optimum MDP design point for the three flip-flops analyzed. Overall, the calculated values are slightly higher than the simulated values but both sets display a consistent trend in the variation of τ in all flip-flops across different technology nodes for both models. In **Figure 4.33(a)**, the infliction point at the 32nm node of the MGHK model is evident for all three flip-flops. **Figure 4.33(b)**, on the other hand, shows a continuous reduction in τ with respect to the scaling of the technology node.

4.7 An All-Digital On-Chip Flip-Flop Metastability Measurement Test Chip

4.7.1 Test Chip Design

An all digital on-chip flip-flop metastability measurement test chip was designed and fabricated in TSMC 0.18μm CMOS technology. The main block diagram of the test chip is



(a) MGHK

(b) Strained-Si

Figure 4.33: Simulated and Calculated Values of τ at Different Technology Nodes for MGHK and Strained-Si Models

shown in **Figure 4.34**. The test chip can be divided into the following components:

- Input Circuitry
- Flip-Flops under Test
- Timing Block
- Metastability Detector Circuitry
- Counter Circuitry

Input Circuitry

The input circuitry consists of a digitally-controlled delay line as well as a distribution network to provide appropriate signal buffering for both the CLK and the input data D

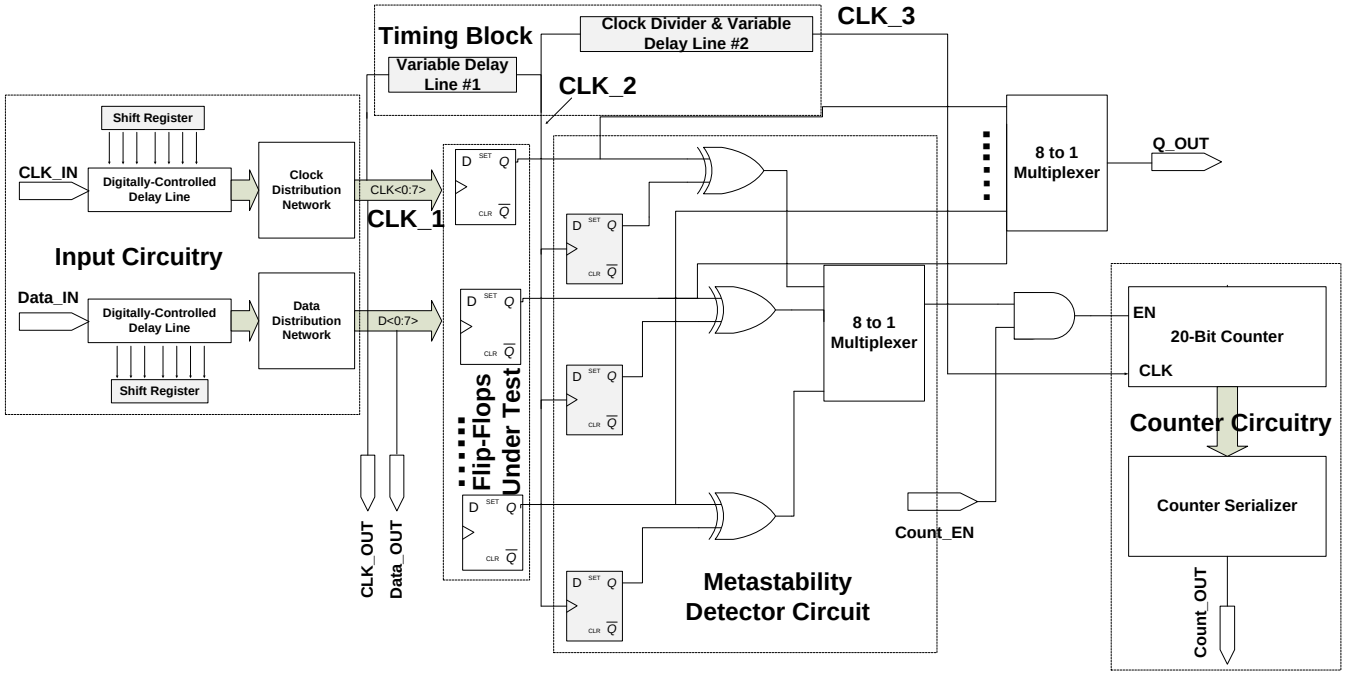


Figure 4.34: Schematic Diagram of an All-Digital On-Chip Flip-Flop Metastability Measurement Circuit

signal. The main purpose of the delay lines is to control the relative timing difference between the CLK and the D signal in order to generate metastable events on the flip-flops. The delay line is controlled by a 21-bit digital code that provides both coarse and fine delay adjustments for a total delay range of 500ps. The first two bits of the digital code are binary code that provides the coarse delay of 400ps with a step size of 100ps. A 19-bit thermometer code is used to provide fine delay step of 1ps and 10ps. The thermometer code is used to provide monolithic linearity on the delay line [75]. The delay line is composed of seven digitally-controlled delay element (DCDE) [76] that is responsible for the fine delay and a chain of inverters that provide the coarse delay range. The schematic diagram of the DCDE and the digital coding scheme are illustrated in **Figure 4.35**.

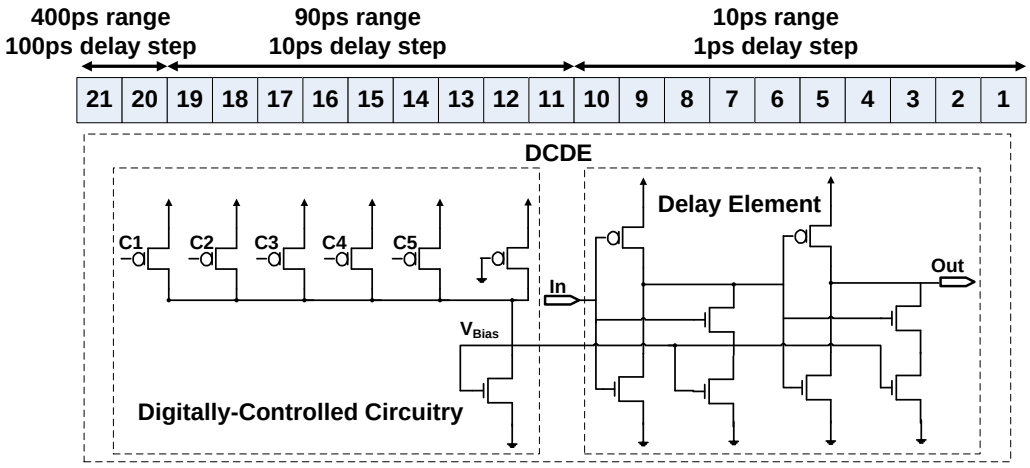


Figure 4.35: Schematic of the Delay Element and the Digital Coding Scheme

A reset mechanism must be implemented on the flip-flop output to ensure a possible metastable event is always generated and can be detected. For instance, if the flip-flop correctly samples a logic "1" on the current CLK edge, then the output must be reset to logic "0" prior to the next CLK edge in order to detect if another logic "1" is correctly sampled on the next CLK edge. In this design, the period of the input data D is set to twice as much as the period of the CLK signal with a non-50% duty cycle to ensure the output is reset appropriately. While a possible metastable event is created only once every two clock cycles, it eliminates additional circuitry required to implement an asynchronous reset mechanism. The timing waveform is illustrated in **Figure 4.36**.

Flip-Flops Under Test

Table 4.11 lists the 16 flip-flops under test (FUT) that have been implemented on the test chip for metastability testing. The word after the "_" indicates weather the flip-flop is designed for optimum PDP or optimum MPDP. In addition to the flip-flops analyzed in

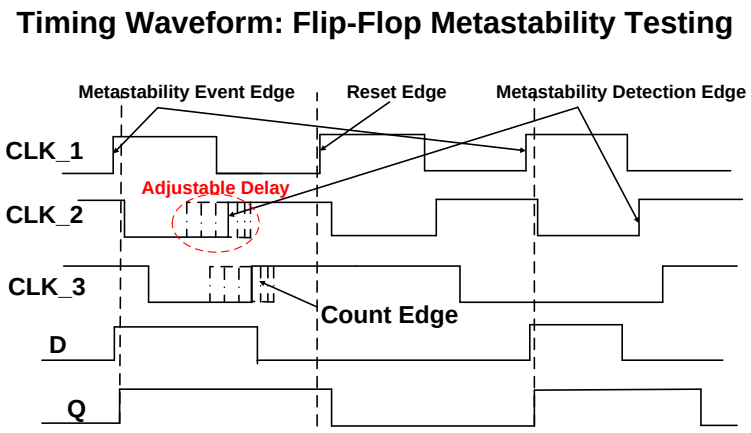


Figure 4.36: Metastability Testing Waveform for the Input Circuitry

this section, a few other flip-flop architectures, indicated by "SE", designed for metastable-hardened and soft-error tolerant are also included on the test chip. For the SAFF and the PowerPC, a third sizing scheme design was also implemented.

Table 4.11: Flip-Flops Under Test

1	SDFF_MPDP	9	SAFF_SE
2	SDFF_PDP	10	PDFF_MPDP
3	PowerPC_MPDP	11	PDFF_PDP
4	PowerPC_PDP	12	PDFF_SE
5	PowerPC_Size3	13	SATG_MPDP
6	SAFF_MPDP	14	SATG_PDP
7	SAFF_PDP	15	SATG_SE
8	SAFF_Size3	16	Hazucha_SE

Timing Block

The timing block consists of two variable delay line circuits that provide the CLK signal for the metastability detector circuit and the counter circuit. The input of the first delay line comes from the CLK_1 signal that is also sent to the FUT circuit. It generates the CLK_2 for the metastability detector circuit. Because it controls the shadow flip-flop, the phase of CLK_2 is inverted from that of CLK_1 with an adjustable delay offset between the two signals. The second delay line generates the CLK_3 signal for the counter circuit with the input coming from the CLK_2. Because possible metastable event only occurs once during two clock periods, a clock divider is used to divide the frequency of the CLK_3 to be half as much as CLK_1 with a certain delay offset with respect to CLK_2 to ensure the enable signal is properly generated. The timing waveform of the CLK signals is illustrated in **Figure 4.36**.

Metastability Detector

The metastability detector circuit is very similar to the one implemented for the Razor flip-flop where a shadow flip-flop is adopted to double sample the output data of the FUT. In the possible event of metastability, the output of the FUT is given a certain amount of time, depending on the adjustable delay values, to settle to a stable value since the shadow flip-flop is triggered by CLK_2. The output of the shadow flip-flop is compared with the output of the flip-flop under test using an XOR gate to generate a signal to indicate if a metastability event has been detected. In the testing circuitry, a total of eight flip-flops are under test, which requires a total of eight metastability detectors. Therefore, an 8-to-1 multiplexer is used to select the appropriate enable signal to be sent to the counter circuit.

Counter Circuitry

A 20-bit synchronous counter is designed to count the number of flip-flop metastable events detected. An AND gate combines a global enable signal with the enable generated from the metastability detector circuit to produce an output that activate the counter. The global signal is essentially a positive pulse signal that can last from a few seconds to a few minutes to determine the time period of the metastability testing. While the counter generates a 20-bit output, it is impossible to simultaneously output every single bit due to the limitations on the number of available I/O pins. Hence, a counter serializer, or simply an output shift register, is used to output the count in a bit-by-bit fashion. The detailed schematic diagram of the counter serializer and the corresponding waveform is illustrated in **Figure 4.37**.

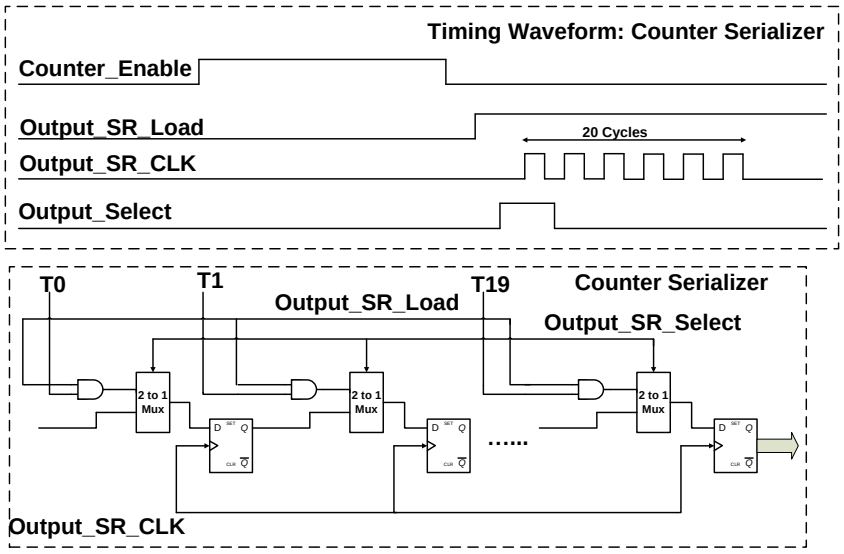


Figure 4.37: Metastability Testing Waveform for the Output Circuitry

Chip Layout

The overall layout of the test chip is shown in **Figure 4.38**. Two identical testing circuits are implemented on the chip to facilitate the testing of 16 flip-flops. The appropriate components of the testing circuit are labeled on the layout: (1) delay line, (2) flip-flops under test, (3) metastability detector, and (4) counter.

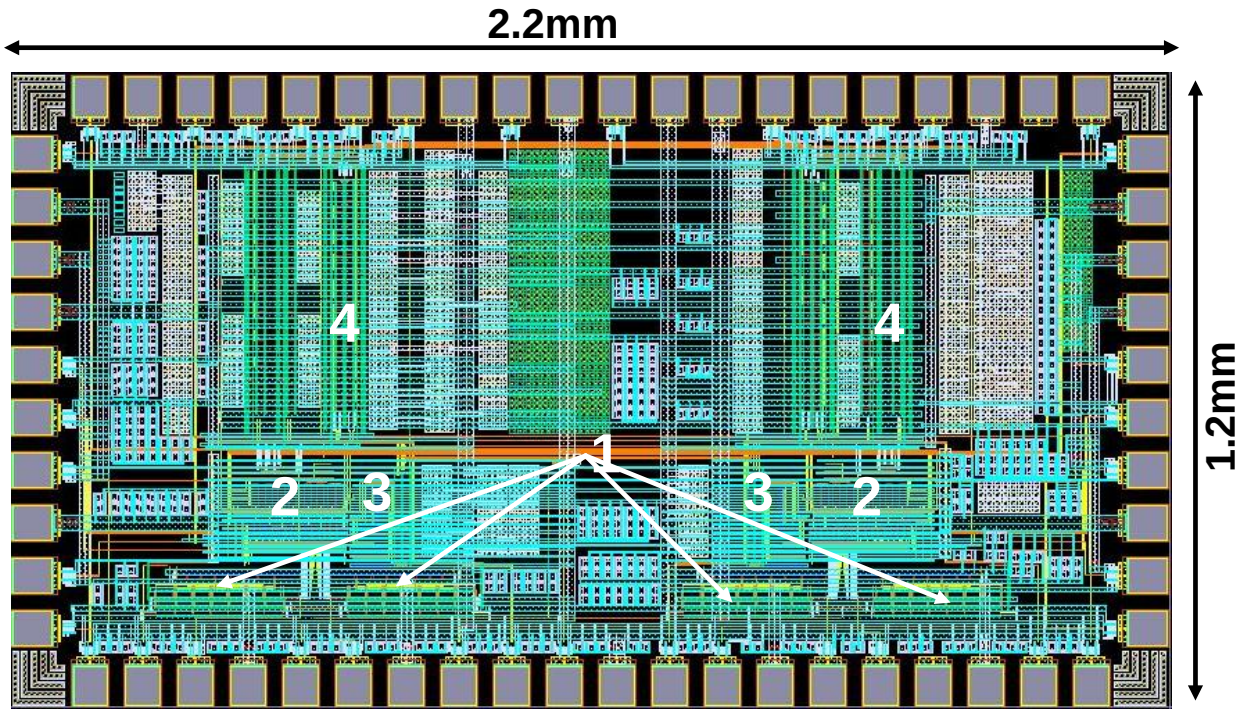


Figure 4.38: Layout of the Flip-Flop Metastability Testing Chip

4.7.2 Testing Methodology

Three different methodologies can be used to measure the time-resolving constant τ of the flip-flops. By outputting the CLK, D, and the output Q signal, the first measurement

methodology is identical to the one presented in [77] and [3]. It involves the usage of a clock pulse generator and a data pulse generator that runs at slightly different frequencies. This method ensures any input data transition that caused an output transition of the FUT occurred within a certain period of the clock edge, and thus resulting in a uniform distribution of the output data around the clock edge. For example, if the clock is running at 101MHz and the data is running at 100MHz, the input data transition will always occur within approximately 1ns of the clock edge. This method requires the usage of an oscilloscope and digital timing system to collect millions of data points such that a histogram (**Figure 4.39**) can be generated to extract the value of τ.

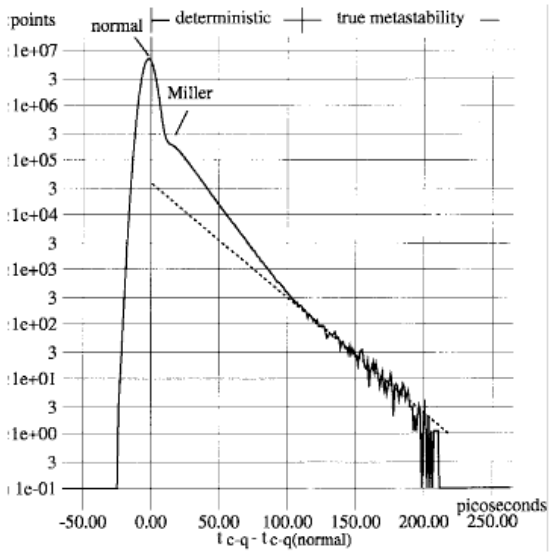


Figure 4.39: Sample Histogram for Metastability Testing [3]

The second measurement method is a digitally-controlled technique that combines various testing features from the previous works [20][78][79][80]. The delay lines should be characterized to determine the relative delays between the CLK and the input data D signal for different combinations of digital code. By fixing the digital code to the CLK

delay line, the code for the input data D signal can be manipulated to vary the data arrival time with respect to the CLK. For a given data arrival time, the counter will count the number of metastable events detected. τ can be extracted as a function of the data arrival time and the number of metastable events detected.

In the third method, the digital code of the CLK and the D delay line can be adjusted accordingly and fixed such that the data time arrival time is always around the metastable region. After that, the variable delay line #1 should be adjusted to change the settling time given for the output of the FUT to reach a stable state while the counter counts the number of metastable events detected. τ can be extracted as a function of the settling time and the number of metastable events detected.

4.8 Summary

In this chapter, a detailed analysis and methodologies for the design of metastable-hardened, high-performance, and low-power flip-flops in both the single and the dual-supply systems is presented. Because the metastability window δ and the MTBF of a flip-flop is largely determined by its time-resolving constant τ, the design of metastable-hardened flip-flops is focused on optimizing the value of τ. Through small-signal modeling, τ is determined to be a function of the load capacitance and the transconductance in the cross-coupled inverter pair for a given flip-flop architecture. In this work, we have shown two ways that can result in significant variation of τ in a flip-flop: *(i)* vary the transconductance by changing the size of the cross-coupled inverter, and *(ii)* vary the size of the load transistors associated with the critical node.

In most cases, the reduction of τ through transistor sizing comes at the expense of increased delay and power. Hence, metastability-delay-product (MDP) and metastability-

power-delay-product (MPDP) are introduced to analyze the design tradeoffs between delay, power and τ. Depending on the flip-flop architecture, either the transconductance or the load variation method will yield the optimum MDP and MPDP design, which, in most cases, is different than the traditional optimum PDP design. With a cross-coupled inverter pair in the critical path of the master-stage and a small load in the slave-stage, the architecture of the PDFF and the SATG is very attractive to achieve good metastability while maintaining high-performance and low-power. For the PDFF in the single-supply system and the SATG in the dual-supply systems, the amount of compromise in delay, power, and area to achieve the optimum MPDP design when compared to the traditional optimum PDP design are all less than 10% , which is significantly less than the other flip-flops analyzed in this work. For all the analyzed flip-flops, simulation results have shown that the optimum MPDP design can reduce the metastability window δ by at least an order of magnitude depending on the value of the settling time and the flip-flop architecture.

In the sub-threshold region, the proposed mixed-V_{th} technique can reduce the τ of the flip-flops by more than 2$\times$ depending on the flip-flop architecture and be more energy efficient than the single standard-V_{th} design if the appropriate supply voltage is selected. The metastable-hardened characteristic of the PDFF is also demonstrated in the sub-threshold region with the lowest MPDP value among the flip-flops analyzed.

The study on the impact of technology scaling has shown that the value τ does not necessarily scale in the same fashion as the gate delay with each generation of the technology node. While τ continues to decrease from the 45nm node down to the 16nm node when the Strained-Si model is used, an infliction point in τ is observed at the 32nm node for the MGHK model. This trend in τ is shown in both the simulated and the theoretical calculated values for the three flip-flops analyzed.

A detailed description on an all-digital on-chip flip-flop metastability testing circuit is

also given in this chapter. The chip is implemented in TSMC $0.18\mu m$ technology with various flip-flop architectures implemented using both the optimum MPDP and the optimum PDP design schemes. The main components of the chip design include digitally-controlled delay line, flip-flops under test, metastability detector, and a 20-bit digital counter.

Chapter 5

Design for Metastable-Hardened, Soft-Error Tolerant Flip-Flops

As size and complexity of chip design are rapidly growing, reliability is becoming an important factor to consider when designing nanometer circuits and systems. In addition to metastability, another reliability concern associated with flip-flop design is soft-errors. In this chapter, we will analyze the techniques involved in designing high-performance and low-power flip-flops while addressing the reliability issues of metastability and soft-error. By extending the methodology for metastable-hardened flip-flop designs, soft-error tolerant cells will also be incorporated into the flip-flop designs. We will apply the idea of using cross-coupled inverter and soft-error tolerant cells on various past flip-flop architectures as well as the two proposed designs, namely the PDFF-SE and SATG-SE . Following our main design approach, both PDFF-SE and SATG-SE use a cross-coupled inverter on the critical path in the master-stage to achieve good metastability while generating differential signals to facilitate the usage of the Quatro cell in the slave-stage to protect against soft-

errors. PDFF-SE is designed to achieve very high performance with good metastability while SATG-SE is a low-power design also with good metastability. Detailed analysis and simulation results will be given on the techniques and issues involved in designing reliable and robust flip-flops.

5.1 Background on Soft-Errors

Cosmic radiation-induced single-event transient (SET) , also known as soft-error, has become a major reliability concern in today's integrated circuits (**Figure 5.1**). Consequently, factors such as increasing clock frequencies and decreasing node capacitances and supply voltage all contribute to a drastic increase in the soft-error susceptibility of both combinational and sequential circuits to alpha particle and cosmic neutron strikes. In combinational

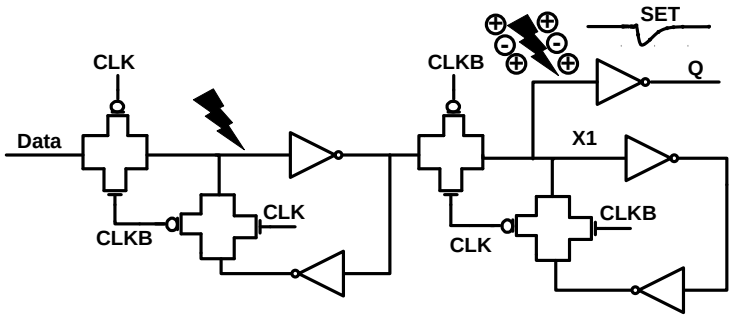


Figure 5.1: Illustration of Soft-Error in Flip-Flop

circuits, phenomenon such as logical masking, electrical masking, and latch-window masking can all mask the glitches caused by soft-errors [81]. Such masking, however, does not exist in sequential elements such as latches and flip-flops, which contribute to approximately 50% of the soft-errors observed in various processors [82]. Recently, the usage of tolerant cells [83][84][85] has emerged as a more popular technique for soft-error protection

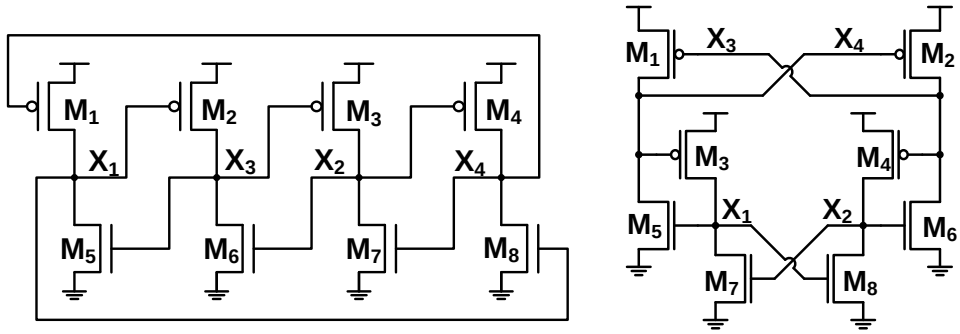
in flip-flops over other techniques such as error-correction code (ECC) and redundancy due to more design robustness along with less delay, power, and area overhead. For example, more than 99% of the latches in the system interface are soft-error protected in the state-of-the-art microprocessor design [86].

5.2 Analysis of Soft-Error Tolerant Cells

5.2.1 Operation

A number of soft-error tolerant cells have been proposed in the past. In this work, we will focus on two particular cells: DICE [87] and Quatro [88]. The Dual-Interlocked Cell (DICE) (**Figure 5.2(a)**) stores a logic "0" or "1" as a combination of four node voltages: two nodes holding the original data and two nodes retain the complement of the data. When the value stored at any node (i.e. X_1) is modified due to SET, other unaffected nodes (X_2, X_3, and X_4) will help to restore the correct value of the affected node because one transistor of each inverter driving one of the affected nodes is driven by one unaffected node. The Quatro cell (**Figure 5.2(b)**) also has four storage nodes. Each of these nodes is driven by an NMOS and a PMOS transistor with their gates connected to two different nodes. If an SET upsets a node voltage, the affected node is restored by the corresponding "ON" PMOS (NMOS) transistor connected to the node and driven by an unaffected node. A detailed operation and simulation waveforms on the usage of the Quatro cell in SRAM and flip-flop design is given in [88] and [89].

In soft-error tolerant flip-flops, the critical internal nodes are protected by being written into the tolerant cells. When writing into the DICE cell, the two nodes must have the same phase and written into cell location of either X_1 and X_2 or X_3 and X_4. Hence, the usage



(a) DICE Cell (b) Quatro Cell

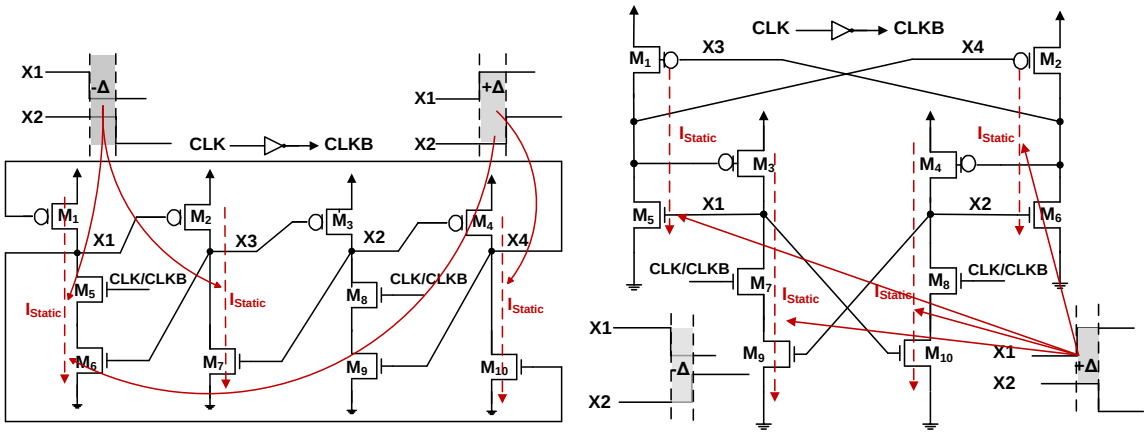
Figure 5.2: Soft-Error Tolerant Cells

of the DICE cell requires the flip-flop architecture to produce identical signals, which is typically accomplished by using duplicated datapath [83]. The Quatro cell, on the other hand, facilitates many differential flip-flop architectures because it requires differential signals to be written into the cell location of either X_1 and X_2 or X_3 and X_4.

5.2.2 Performance

While the addition of the tolerant cells increases the immunity of the flip-flops against soft-errors, it also impacts its performance by adding more resistivity in terms of changing the values stored at the critical nodes during the normal operation of the flip-flops. Hence, a modified version of the DICE and the Quatro cell is shown in **Figure 5.3** where two additional CLK-controlled transistors are added to the DICE (M_5 and M_8) and the Quatro (M_9 and M_{10}) cell respectively in order to maintain high-performance.

Depending whether the cells are used in the master or the slave-stage, these transistors are controlled either by the CLK in the master-stage or the $CLKB$ in the slave-stage.



(a) Modified DICE Cell

(b) Modified Quatro Cell

Figure 5.3: Modified Soft-Error Tolerant Cells

Assuming the flip-flop is positive-edge triggered, for example, during the evaluation period in the slave-stage, the $CLKB$ cuts off the NMOS path that holds a logic "0" in the hardened cell, which allows the node to be flipped to logic "1". If these two transistors are not present, contention exists between the flip-flop and the hardened cell in changing the node value from 0-1, which results in significant performance degradation. Alternatively, two more clocked-transistors can be added in the PMOS paths that holds a logic "1", however the amount of performance degradation without these two transistors is not as significant when changing the node value from 1-0 due to the relative weaker strength of the PMOS transistors when compared to the NMOS transistors. Simulation results have shown that the presence of these transistors in the tolerant cells improve the performance by at least 10% depending on the flip-flop architecture. Such performance enhancement, however, may come at the expense of reduced soft-error immunity of the tolerant cell. Simulation results have shown that the implementation of the clocked-transistors in the

Quatro cell reduces the critical charge by approximately 1.4×-1.5× for the 0-1 and 1-0 data transition.

5.2.3 Power Consumption

The power consumption of the DICE and the Quatro cell is also analyzed in this work. Ideally, there is no phase offset between the signals being written into the DICE cell. Due to PVT variations and transistor mismatches, it is possible that the two signals can have a small static offset of Δ (i.e. X_1 arrives earlier than X_2 or vice versa) and consequently results in a few static power dissipation paths in the cell for a given data transition. In the Quatro cell, a static offset of Δ exists even without the presence of PVT variations and mismatches due to the inverter delay required to generate the differential signal such that the signal transition of X_2 will always arrive later than that of X_1. If X_1 makes a 0-1 transition, X_2 will make a 1-0 transition after an inverter delay. During this period, however, four potential paths in the Quatro cell could result in static power dissipation by simultaneously turning on both the PMOS and NMOS transistors. The same scenario does not occur when X_1 is making a 1-0 transition and X_2 is making a 0-1 transition because all the NMOS transistors are turned off. The potential static power dissipation paths have been marked in red in Figure 5.3(a) and Figure 5.3(b), respectively, for the DICE and the Quatro cell.

A simple test bench was setup to measure the power consumption (**Figure 5.4**) of the DICE and the Quatro cell using a data activity of 25% with equal number of 0-1 and 1-0 data transitions for input signal (X_1 and X_2) having two sets of rise/fall time: 50ps and 100ps. $+\Delta$ indicates signal X_2 arrives later than X_1 in both the DICE and the Quatro cell respectively, and vice versa for $-\Delta$. From the figure, it is evident that the power

consumption in the DICE cell is symmetrical about the point where the static phase offset Δ is 0, which means power consumption only depends on the absolute value of the phase offset and indifferent to the arrival order of the input signals. In the Quatro cell, the power consumption for a rise/fall time of 50ps is symmetrical about the Δ=10ps point, which is roughly equivalent to an inverter delay for the corresponding signal rise/fall time. The symmetry point moves to 40ps when the rise/fall time is 100ps, which suggests the inverter delay degrades with input signals having a higher rise/fall time. Once again, the power consumption is irrelevant to the arrival of the input signals in the Quatro cell as long as the number of 0-1 and 1-0 data transitions is equal. If the input data vector has more 0-1 transitions, then the power dissipated will be significantly higher than when there is more 1-0 transitions. Under such scenario, the power consumption will no longer be symmetrical for $+\Delta$ and $-\Delta$ offset. Finally, a faster rise/fall time will result in significant power saving in both the DICE and the Quatro cell, as evident by the data comparison between 50ps and 100ps rise/fall time shown in **Figure 5.4**. The effect of higher rise/fall time is more prominent in the Quatro cell where both the short-circuit and static power dissipation contribute to the overall power consumption. Based on the above analysis, it is clear that the power consumption of the Quatro cell is generally higher than that of the DICE cell.

5.2.4 Radiation Testing

As part of the research collaboration, a test chip was designed and fabricated in the TSMC 40nm CMOS technology to provide some insights in comparing the flip-flop soft-error rates of using the DICE and the Quatro tolerant cells. Three types of flip-flops are implemented on the test chip: *(i)* a master-slave C^2MOS configuration without any soft-error protection, *(ii)* a master-slave C^2MOS configuration using the DICE cell on the slave-stage, and *(iii)*

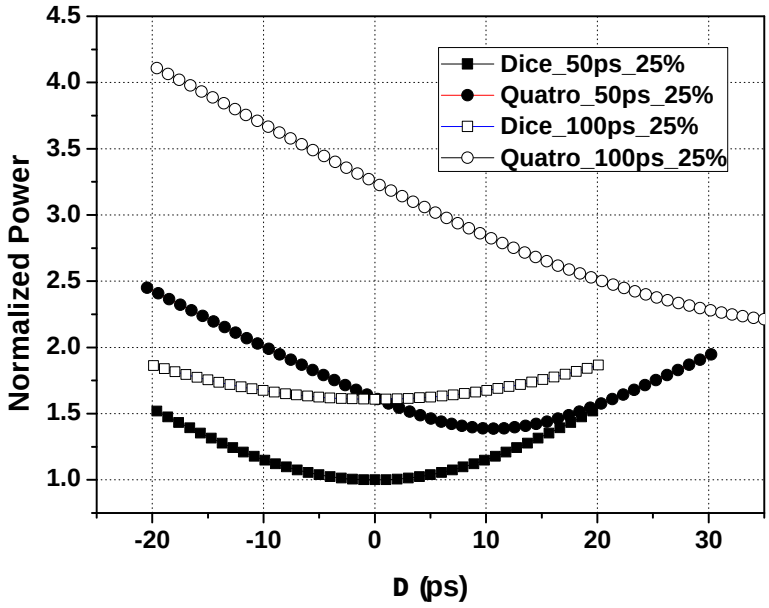
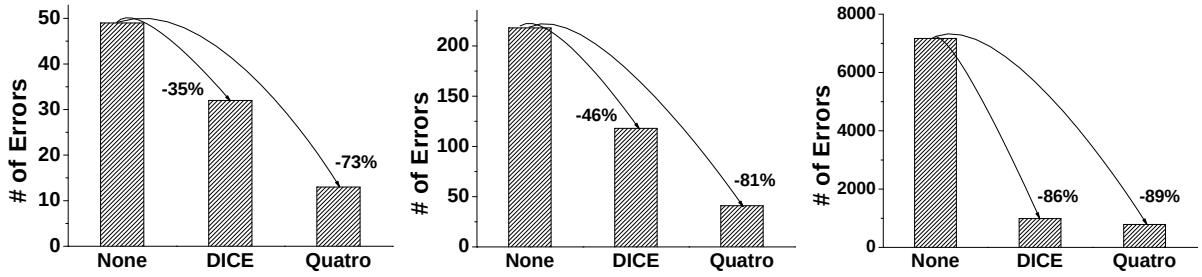


Figure 5.4: Power Consumption of the Soft-Error Tolerant Cells

a master-slave C^2MOS configuration using the Quatro cell on the slave-stage. A shift-register test structure was utilized because it is the densest array of flip-flops, and as such is commonly used to validate the SET robustness of flip-flops in an area-efficient manner [90]. Shift registers were created utilizing the previously described flip-flops, with each shift register contained 8000 flip-flops. The test chip implemented the Circuit for Radiation Effects Self Test (CREST) methodology [91], to enable at-speed soft-error rate testing.

A total of three different sets of testing were conducted in this study. The accelerated radiation testing was performed by Vanderbilt University. Accelerated neutron radiation testing was conducted at the Tri-University Meson Facility (TRIUMF) at the University of British Columbia, Vancouver as well as the Los Alamos Neutron Science Center (LANSCE) in Los Alamos, New Mexico, USA. The results of the neutron radiation experiments and

the alpha radiation experiments are illustrated in **Figure 5.5**. From the results of the



(a) TRIUMF (b) LANSCE (c) Vanderbilt

Figure 5.5: Results of Radiation Testing

neutron radiation testing and the alpha radiation testing, it is clear that the usage of the DICE and the Quatro cell have reduced the soft-error rate significantly when compared to the flip-flop without any protection. Furthermore, the Quatro cell yields a lower soft-error rate (SER) than the DICE cell, as evident by the percentage of SER reduction shown in **Figure 5.5**.

5.3 Analysis and Design Methodology

In this work, we have analyzed the usage of the DICE and the Quatro cells along with the cross-coupled inverter structure on various flip-flops architectures in order to simultaneously achieve good metastability and soft-error protection while maintaining the characteristic of high-performance and low-power. The main approach is to resolve metastability in the master-stage with a cross-coupled inverter pair while adding the soft-error tolerant cell in the slave-stage to protect the output nodes against possible SET (**Figure 5.6**). Because duplicated or redundant signals must be generated to use the DICE cell while differential

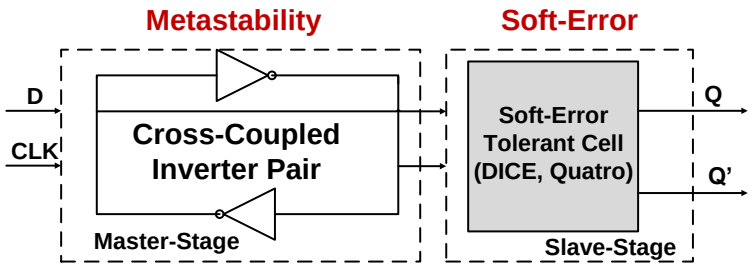


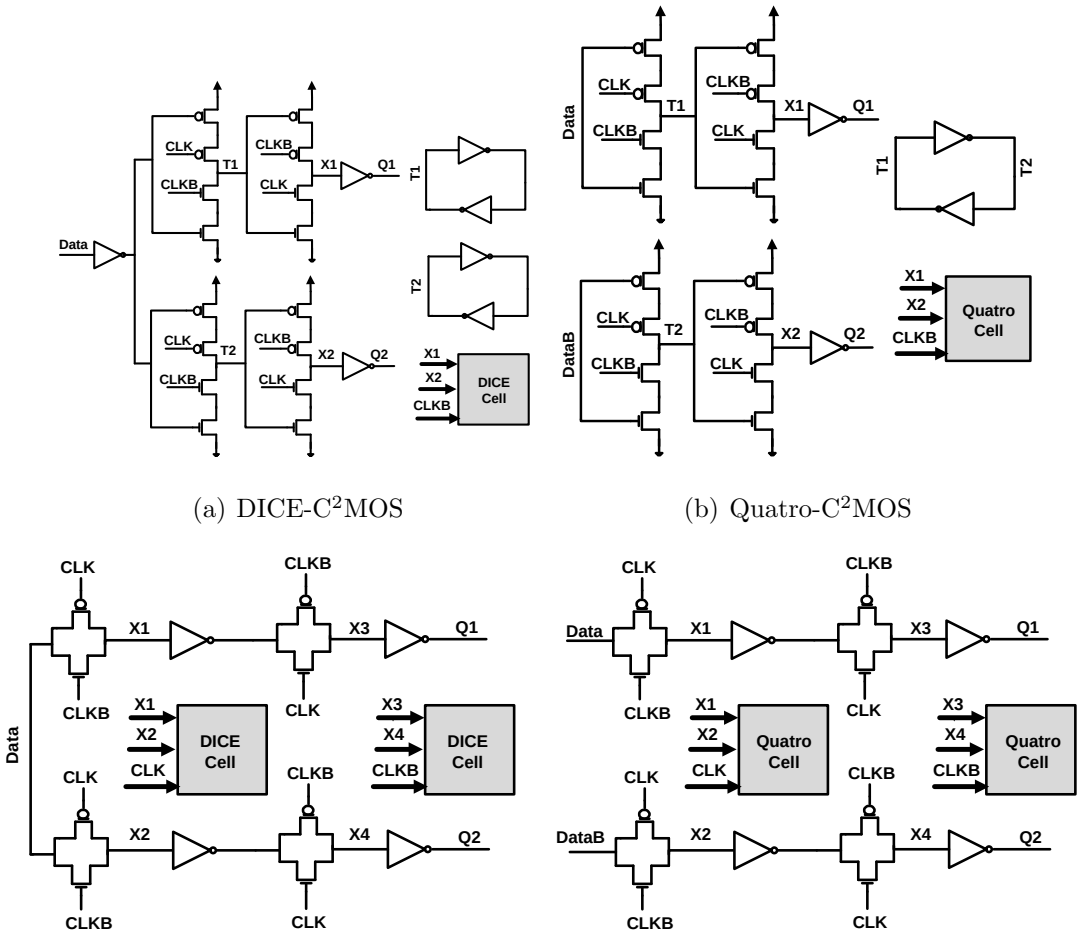
Figure 5.6: Design Methodology of Metastable-Hardened, Soft-Error Tolerant Flip-Flops

signals are required for the Quatro cell, special flip-flop architectures are required to facilitate the usage of these cells. As such, flip-flops analyzed in the previous chapter may not be suitable for metastable-hardened and soft-error tolerant designs because the amount of area and power overhead associated can be substantial.

Two C^{2}MOS-based architectures are analyzed in this work (**Figure 5.7(a)** and **5.7(b)**). In the Quatro-C^{2}MOS configuration, a cross-coupled inverter pair is used to stabilize the dynamic nodes of T1 and T2 while improving metastability. The DICE-C^{2}MOS configuration does not produce differential signals, and hence separate inverter pairs are used on each datapath to improve the metastability in the master stage. The value of τ is limited by the size of the feedback inverter, which must be kept close to minimum size to reduce the amount of parasitic capacitance at critical nodes in order to maintain good performance and functionality.

A special soft-error robust latch based on transmission-gate and DICE cell was proposed in [85]. In this work, we modify the design slightly to create a Hazucha flip-flop (**Figure 5.7(c)** and **5.7(d)**) using both the DICE and the Quatro tolerant cell by cascading two identical latches. Instead of using the traditional cross-coupled inverter in the master-stage to improve metastability, the DICE and the Quatro cell are used in each respective design

because the cross-coupled inverter structure with feedback paths still exists in these cells to improve the immunity against soft-errors.



(a) DICE-C^{2}MOS (b) Quatro-C^{2}MOS

(c) DICE-Hazucha (d) Quatro-Hazucha

Figure 5.7: Metastable-Hardened, Soft-Error Tolerant Flip-Flop Designs

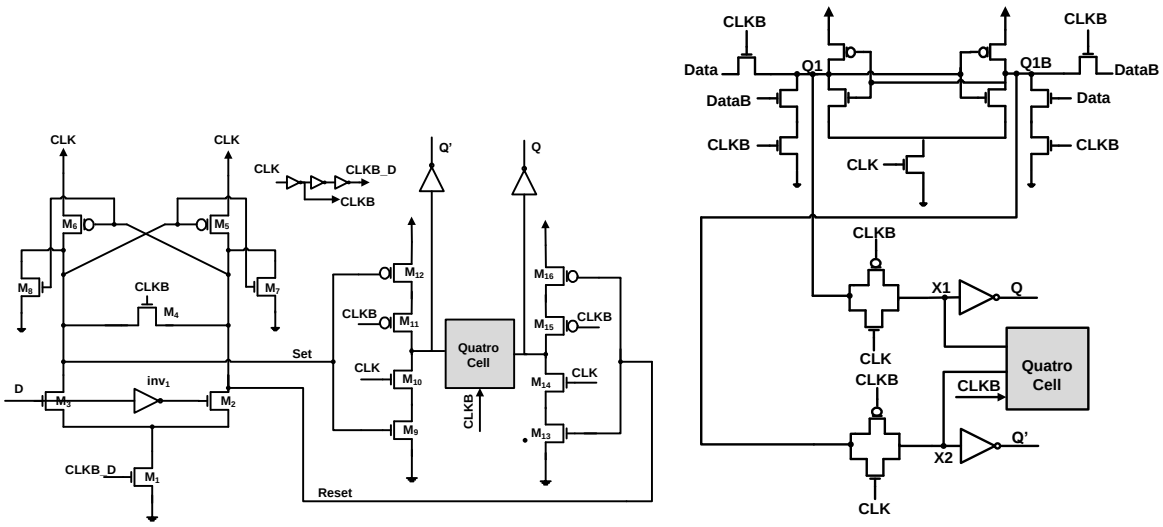
From the data illustrated in **Figure 5.5**, it is evident that the usage of the Quatro cell has resulted in a lower SER than the DICE cell in the nanoscale CMOS technologies. Furthermore, the usage of a cross-coupled inverter pair in the critical path of the master-stage

can significantly improve metastability while generating the differential signals required for the Quatro cell. The combination of these attractive features suggest the master-stage of the PDFF and the SATG are very attractive in designing metastable-hardened and soft-error tolerant flip-flops. Hence, two new differential flip-flops are proposed in this work: (i) pre-discharge soft-error tolerant flip-flop (PDFF-SE, **Figure 5.8(a)**) (ii) sense-amplifier transmission-gate soft-error tolerant flip-flop (SATG-SE, **Figure 5.8(b)**). Both designs can achieve good metastability with a cross-coupled inverter in the master-stage and soft-error protection by using the Quatro cell in the slave-stage. The cross-coupled inverter structure in the master-stage can be sized up to simultaneously achieve good performance and metastability while the differential nature facilitates the usage of the Quatro cell in the slave-stage. The design of the PDFF-SE is targeted towards very high-performance with good metastability while the SATG-SE is designed to have low-power consumption also with good metastability.

While the master-stage of the PDFF-SE and the SATG-SE is identical to that of the PDFF and the SATG described previously, the design of the slave-stage is modified in order to minimize the power consumption by balancing the arrival time of the input signals written into the Quatro cell. PDFF-SE utilizes a tri-state inverter architecture and SATG-SE uses the CLK-controlled transmission-gates architecture. With careful design considerations, the power consumption of and PDFF-SE and SATG-SE can be reduced significantly with reasonable performance despite the usage of the Quatro cell.

5.4 Results and Discussion

Table 5.1 summarizes the schematic simulation results of delay, power, and τ for all the metastable-hardened and soft-error tolerant flip-flops analyzed in this work. All the results



(a) PDFF-SE (b) SATG-SE

Figure 5.8: Proposed Metastable-Hardened, Soft-Error Tolerant Flip-Flop Designs

are obtained using the 65nm STM CMOS bulk technology. The simulation test bench setup used is identical to the one shown in **Figure 3.15**. Two sets of data activity factors are used for analysis: 10% and 50%. Iterative process was used in transistor sizing in order to achieve the optimum MPDP flip-flop design for the best possible combination between delay, power, and τ. Minimum-sized transistors are used in all the DICE and Quatro cells along with the implementation of the CLK-controlled transistors.

The addition of the minimum-sized cross-coupled inverter on the dynamic nodes of the C^2MOS flip-flops to enhance metastability significantly degrade its performance. Without these inverters, however, the value of τ can be as much as 40× higher. The performance of the Quatro-Hazucha is worse than the DICE-Hazucha because the Quatro cell is more resistant in writing data for 1-0 transition which consequently results in higher setup time when it is used in the master-stage. Based on the reasonings from earlier analysis,

the power consumption of the Quatro-C^{2}MOS and the Quatro-Hazucha are shown to be higher, especially at higher data activity, than the same flip-flop architectures when DICE cell is used. Minimum transistor sizing is used on the cross-coupled inverter structure in the C^{2}MOS and Hazucha architectures, and therefore their respective τ is very similar with the difference coming from the parasitic capacitance surrounding the critical node. The proposed PDFF-SE results in at least 17% performance improvement over the other flip-flop architectures, but the pre-discharging of the internal nodes during every clock cycle makes its power consumption higher than the other flip-flops, especially at low data activity. The proposed SATG-SE maintains a very comparable performance to the other analyzed flip-flops while achieving a minimum 18% and 6% power reduction for 10% and 50% data activity respectively. The size of the cross-coupled inverter in both the PDFF-SE and the SATG-SE can be sized up significantly higher than minimum size because they are on the critical path, and thus results in a lower value of τ. The τ of the PDFF-SE and the SATG-SE is at least 21% and 14% lower than the other flip-flops respectively.

Table 5.2 summarizes the design metrics of PDP, MDP, and MPDP for all the metastable-hardened and soft-error tolerant flip-flops analyzed in this work. Once again, data activity of 10% and 50% are used for the analysis of PDP and MPDP. The PDP of the SATG-SE and the PDFF-SE is the lowest among all flip-flops for data activity of 10% and 50% respectively. The PDP value of the DICE-Hazucha is also small for both data activity factors. Since the PDFF-SE exhibits both the best performance and metastability, its MDP value is significantly lower than the other flip-flops such as a 43% reduction than the DICE-Hazucha, and thus indicating a well-balanced design tradeoff between performance and metastability. While higher than the PDFF-SE, the MDP of the SATG-SE is still at least 12% lower than the other flip-flops. For 10% data activity, the MPDP of the PDFF-SE and the SATG-SE is 16% and 32% lower than the DICE-Hazucha flip-flop. At 50%

Table 5.1: Simulation Results of Metastable-Hardened, Soft-Error Tolerant Flip-Flops: Delay, Power, τ

	Delay (ps)	10% Power (μW)	50% Power (μW)	τ (ps)
DICE-C^{2}MOS	79.34	3.61	6.95	24.26
Quatro-C^{2}MOS	73.55	3.88	7.95	27.49
DICE-Hazucha	52.57	3.89	6.88	25.2
Quatro-Hazucha	89.58	4.19	8.32	28.12
PDFF-SE	39.68	5.66	8.05	19.2
SATG-SE	56.29	2.97	6.48	20.86

data activity, the minimum MPDP reduction of the PDFF-SE and the SATG-SE from other flip-flops is 33% and 17% respectively.

In this work, we also analyze the robustness of each flip-flop architecture against process variations and mismatches when it is operating near or at the metastable region. For each flip-flop, the data arrival time in which the flip-flop first fails to capture the correct data was determined and will be referred to as t_{meta}, the point where the flip-flop is very close or at the metastable region. A Monte Carlo simulation of 2000 iterations with both process variations and mismatches was performed to determine the number of clock cycles where the correct data was sampled. Then the data arrival time of the flip-flop is relaxed from t_{meta} by a certain value, and another set of Monte Carlo simulation is performed. This procedure (**Figure 5.9**) is repeated for a number of data arrival time values until the sampled data is 100% correct. Based on previous studies and simulation results, a 20ps

Table 5.2: Simulation Results of Metastable-Hardened, Soft-Error Tolerant Flip-Flops: PDP, MDP, MPDP

	MDP (ps^2)	10% PDP (fJ)	50% PDP (fJ)	10% MPDP (fJ·ps)	50% MPDP (fJ·ps)
DICE-C^2MOS	1614.97	0.286	0.551	5.829	13.367
Quatro-C^2MOS	1865.13	0.285	0.584	7.239	16.054
DICE-Hazucha	1324.76	0.204	0.362	5.153	9.114
Quatro-Hazucha	1605.09	0.375	0.745	6.720	20.949
PDFF-SE	761.86	0.225	0.319	4.312	6.133
SATG-SE	1174.10	0.167	0.365	3.488	7.609

flip-flop metastable region from t_{meta} is assumed.

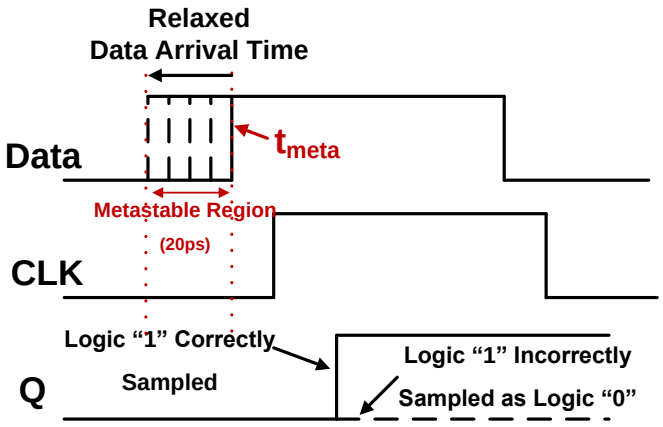
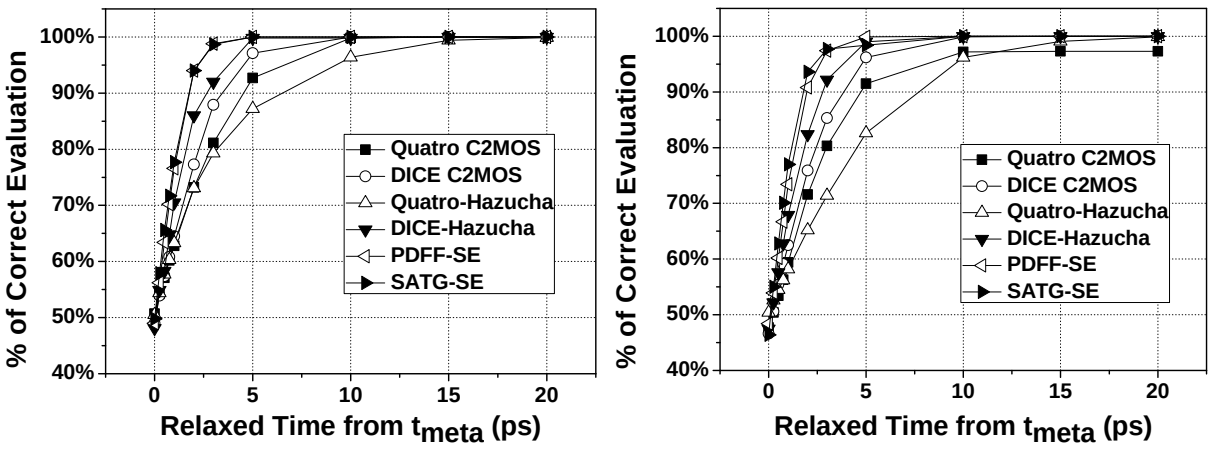


Figure 5.9: Waveform for Monte Carlo Simulation

Figure 5.10 shows the Monte Carlo simulation results of each flip-flop architecture

for both the 0-1 and 1-0 data transition at various data arrival times. At t_{meta} for each respective flip-flop, the percentage of correctness is approximately 50%, which suggests total randomness when the flip-flop is going under metastability [92]. As the data arrival time is relaxed, the percentage gradually increases at various rates for different flip-flops depending on their resolving time constant. It is interesting to note that the flip-flops with a lower τ value have an overall higher percentage of correctness, and thus are more robust against process variations and mismatches. For example, the PDFF-SE and the SATG-SE have an overall 83% and 81% correctness respectively in the metastable region for 0-1 data transition and 81% and 78% for 1-0 data transition. the Quatro-C^{2}MOS and the Quatro-Hazucha, on the other hand, have the highest τ values and consequently yield the lowest overall percentage of 75% and 74% for 0-1 data transition and 75% and 71% for 1-0 data transition respectively.



(a) 0-1 Data Transition

(b) 1-0 Data Transition

Figure 5.10: Flip-Flop Robustness against Process Variations and Mismatches

5.5 Summary

In this chapter, we have analyzed the design of metastable-hardened and soft-error tolerant master-slave flip-flops as well as proposing two new flip-flop designs. The main approach is to resolve metastability in the master-stage with a cross-coupled inverter pair while adding the soft-error tolerant cell in the slave-stage to protect the output nodes against possible soft-errors. To achieve good metastability, it is desirable to have the cross-coupled inverter on the critical path of the flip-flops in order to increase the overall loop gain and lower the value of τ. The DICE and the Quatro cell are the two soft-error tolerant cells used in the flip-flop design to provide protection against soft-errors. The former requires the flip-flop to generate duplicated signals to be written into the cell while a differential signal is needed in the latter cell. Additional clocked-transistors are added to both cells in this work when compared to the traditional design in order to maintain high-performance. The power dissipation of the Quatro cell is higher than the DICE cell due to an inverter delay that generates the differential path as well as more leakage paths.

The design of the proposed flip-flop PDFF-SE and SATG-SE uses a cross-coupled inverter on the critical path in the master-stage to achieve good metastability while generating differential signals to facilitate the usage of the Quatro cell in the slave-stage to protect against soft-error. The PDFF-SE is designed to achieve very high-performance with good metastability while the SATG-SE is a low-power design also with good metastability. Simulation results have shown that both designs achieve significant reduction in MDP and MPDP when compared to other flip-flop architectures analyzed in this work. Monte Carlo simulation results demonstrate that the two proposed flip-flops are very robust against process variations and mismatches.

Chapter 6

Conclusions and Future Work

In this thesis, we present a detailed analysis and designed methodology on metastable-hardened, high-performance, and low-power flip-flops. While the design of high-performance and low-power flip-flops has been a popular research topic, the issue of flip-flop metastability has rarely being dealt with. The following points summarize the key contributions of this research work.

- Proposed flip-flop architectures that achieve high-performance, low-power, and good reliability that can function in both the single and the dual-supply systems.
- Proposed and developed methodologies in analyzing flip-flop metastability in both qualitative and quantitative manner.
  - Developed methodology of transconductance and load variation to vary flip-flop metastability performance using transistor sizing.
  - Developed calculation methodologies to model the value of τ for a given flip-flop architecture.

  - Proposed design metrics (MDP and MPDP) that analyze the tradeoff between performance, power, and metastability.
  - Demonstrated the proposed flip-flop architectures of the PDFF and the SATG are suitable for metastable-hardened designs with small penalties in delay and power consumption.
  - Proposed a novel mixed-V_{th} technique that can improve flip-flop metastability in the sub-threshold region.
  - Analyzed the impact of scaling in sub-65nm technologies on flip-flop metastability.

- Analyzed methodologies of designing metastable-hardened and soft-error tolerant flip-flops and proposed two new flip-flop designs.

6.1 High-Performance, Low-Power Flip-Flop Designs

The proposed pre-discharge flip-flop (PDFF) has demonstrated low-power and high-performance characteristics in both the single and the dual-supply systems. The worst critical path has been reduced to a maximum of three transistors, and thus results in a smaller D-Q delay. With fewer transistors on the critical path, the total transistor widths of the PDFF is reduced and also results in smaller power consumption. When comparing to the single-supply flip-flops, post-layout simulation results have shown the PDFF yields a minimum of 18% and 13% reduction in D-Q delay and PDP, respectively, than the other flip-flops. The power consumption of the PDFF is only 15% higher than the PowerPC but more than 15% lower than the other analyzed flip-flop architectures. When functioning as a reduced clock-swing flip-flop, along with comparable power consumption across all data activity

factors, the RCSPDFF also results in a minimum 40% and 18% reduction in D-Q delay and PDP, respectively, when compared to other flip-flops for $V_{DDL} = 1.3V$. In the case of level-converting flip-flops, the LCPDFF outperforms its counterparts by at least 11% in D-Q delay, 18% reduction in PDP, and 15% reduction in power consumption for data activity factor higher than 50%.

The sense-amplifier-transmission-gate (SATG) flip-flop was proposed specifically for the dual-supply systems to function both as reduced clock-swing and level-converting flip-flops. While its overall performance and power characteristics are not as superior as those of the RCSPDFF and the LCPDFF, both the RCSSATG and the LCSATG still exhibit high-performance as well low-power characteristics at low data activity factors. At $V_{DDL} = 1.3V$, the D-Q delay of RCSSATG is only 1.3% higher than the previously proposed reduced clock-swing flip-flops with a 28% lower power consumption at zero data activity. At 0% and 25% data activity factor, the PDP of the RCSSATG is 39% and 8% lower than the previous designs. Detailed comparisons with level-converting flip-flops reveal that the delay of the LCSATG is 5% higher than the previous design with very similar power consumption values at 0% and 25%. With similar delay and power values, the PDP of the LCSATG across all data activities is almost identical to the previous level-converting flip-flop designs.

An important flip-flop design criteria that is often overlooked in past designs is the flip-flop aperture window. A smaller aperture window reduces the likelihood of the flip-flop entering metastability, and thus increases the reliability of the flip-flop. In this work, we have shown that the PDFF and the SATG have demonstrated a very small aperture window value in both the single and the dual-supply systems.

6.2 Metastable-Hardened Flip-Flop Designs

Unlike past works where performance and power are the main design criteria, this research work also incorporates the element of metastability into the flip-flop designs. Various design and analysis methodologies are proposed in order to design metastable-hardened, high-performance, and low-power flip-flops. Because the time-resolving constant τ has the greatest impact on the mean-time-between-failure (MTBF) of the flip-flop due to its exponential relationship, the design of metastable-hardened flip-flops is focused exclusively on the optimization of τ. τ can be varied via transistor sizing in two ways: (i) vary the transconductance changing the size of the cross-coupled inverter that stabilizes the critical node, and (ii) vary the size of the load transistors associated with the critical node. Depending on the flip-flop architecture, appropriate transistor sizing can reduce the value of τ by a minimum of 30% from the traditional optimum PDP design point. By applying small-signal modeling, the manipulation of τ due to transconductance and load variation analysis of a given flip-flop architecture can be theoretically modeled by calculating the transconductance in the cross-coupled inverter pair and the amount of parasitic capacitances surrounding the critical code.

While appropriate transistor sizing can improve the flip-flop metastability with the reduction of τ, it often comes at the expense of an increase in delay and power consumption. Therefore, both the τ vs. delay and the τ vs. PDP curve can be used to illustrate the tradeoff between delay, power, PDP, and τ. Subsequently, two new design metrics, the metastability-delay-product (MDP) and the metastability-power-delay-product (MPDP), are proposed in this work to analyze the optimum tradeoff between τ and delay as well as τ and PDP. Depending on the flip-flop architecture, either the transconductance or the load variation analysis may result in the optimum MDP and MPDP design point, which

is usually different from the optimum PDP point under the traditional design scheme.

With a cross-coupled inverter in the master-stage that increases the overall transconductance and a small load transistor associated with the critical node, the architecture of both the PDFF and the SATG is very suitable for the design of metastable-hardened, high-performance, and low-power flip-flops. The amount of overhead in delay, power, and area is all less than 10% under the optimum MPDP design scheme when compared to the traditional optimum PDP design. For single supply flip-flops, the optimum MPDP design of the PDFF has produced a minimum reduction of 42% and 34% in MDP and MPDP, respectively. While the MDP and MPDP of the RCSPDFF and the LCPDFF are still lower than the previous reduced clock-swing and level-converting flip-flops, the amount of reduction in the RCSSATG and the LCSATG for the optimum MPDP design is even greater. For example, the MDP and MPDP for the RCSSATG is 34% and 35% lower than the RCSPDFF while the LCSATG is 28% and 25% lower than the LCPDFF for the level-converting flip-flops.

For all the flip-flop architectures analyzed, the reliability of all the analyzed flip-flops under the optimum MPDP design scheme is greatly improved when compared to the traditional optimum PDP design, as evident by a minimum of one order of magnitude reduction in the metastability window δ.

In the sub-threshold region, the proposed mixed-V_{th} technique can reduce the τ of the flip-flops by more than $2\times$ depending on the flip-flop architecture and be more energy efficient than the single standard-V_{th} design if the appropriate supply voltage is selected. The metastable-hardened characteristic of the PDFF is also demonstrated in the sub-threshold region with the lowest MPDP value among the flip-flops analyzed.

The study on the impact of technology scaling has shown that the value τ does not

necessarily scale in the same fashion as the gate delay with each generation of the technology node. While τ continues to decrease from the 45nm node down to the 16nm node when the Strained-Si model is used, an infliction point in τ is observed at the 32nm node for the MGHK model. This trend in τ is shown in both the simulated and the theoretical calculated values for the flip-flops analyzed.

6.3 Metastable-Hardened and Soft-Error Tolerant Flip-Flop Designs

The work presented in this thesis also attempts to increase the reliable operation of the flip-flops by incorporating soft-error mitigation techniques into the design of metastable-hardened flip-flops. The main design approach is to resolve metastability in the master-stage with a cross-coupled inverter pair in the critical path while adding the soft-error tolerant cell in the slave-stage to protect the output nodes against possible soft-errors. In particular, two soft-error tolerant cells, DICE and Quatro, are analyzed in detail from the perspective of performance, power consumption, and immunity against soft-errors. For both cells, the addition of clock-transistors to cut the feedback paths during a particular clock cycle can improve the flip-flop performance by at least 10% but suffers a decrease in soft-error immunity with an approximately 1.5× reduction in critical charge. The overall power consumption of the Quatro cell is higher than the DICE mainly due to the different arrival time of the two signals being written into the cell. With careful design considerations, however, the power consumption of the Quatro cell can be minimized. Radiation testings have shown the soft-error rate (SER) is much lower when the Quatro cell is applied on the slave-stage of the flip-flops.

Based on the above analysis, two new flip-flop designs are proposed: PDFF-SE and SATG-SE. Both flip-flops utilize a cross-coupled inverter on the critical path in the master-stage and generate the required differential signals to facilitate the usage of the Quatro cell in the slave-stage. While being soft-error protected, the optimized τ of both flip-flops is a minimum of 14% lower from the other analyzed flip-flops. and subsequently results lower MDP and MPDP. The MDP of the PDFF-SE is a minimum 43% lower than the other flip-flops. At 50% data activity factor, the minimum MPDP reduction of the PDFF-SE and the SATG-SE from other flip-flops is 33% and 17% respectively. Finally, both flip-flops have shown better robustness against process variations near the metastable region.

6.4 Future Work

In this work, we have focused on the analysis and optimization of flip-flop metastability from the perspective of circuit design on the transistor level. Further research work can be explored on developing techniques on multi-stage synchronizer designs to further improve the mean-time-between-failure of the system without compromising in the overall latency of the system. The principle of metastability can be extended to other applications such as phase-detectors. When the zero-crossing points of the recovered clock fall in the vicinity of data transitions, the flip-flops comprising the phase detector (PD) may experience metastability, and therefore generating an output lower than the full logic level for an extended period of time [93]. A detailed study can be performed to establish a relationship between the metastability parameters of T_0 and τ and the phase errors generated by the PD using various flip-flop architectures with different metastability behaviors.

Appendix A

Flip-Flop Layouts

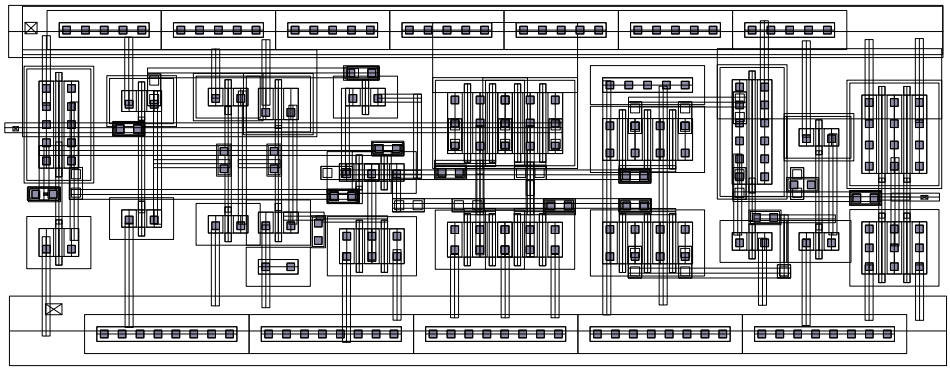


Figure A.1: Layout Diagram of the PDFF

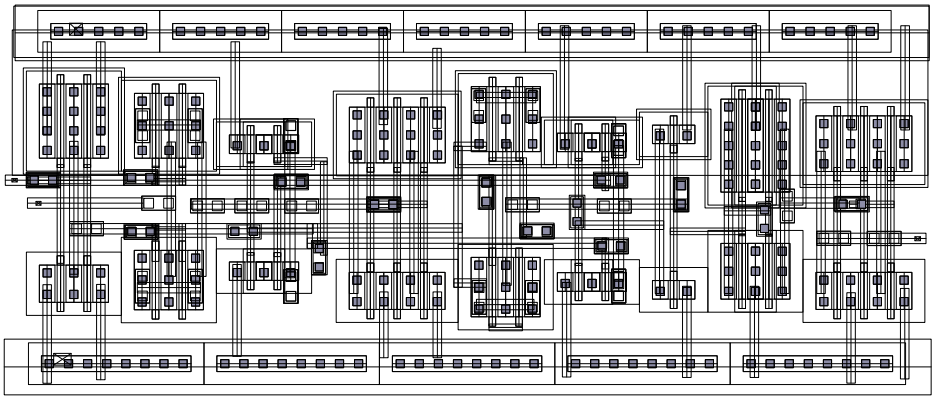


Figure A.2: Layout Diagram of the PowerPC

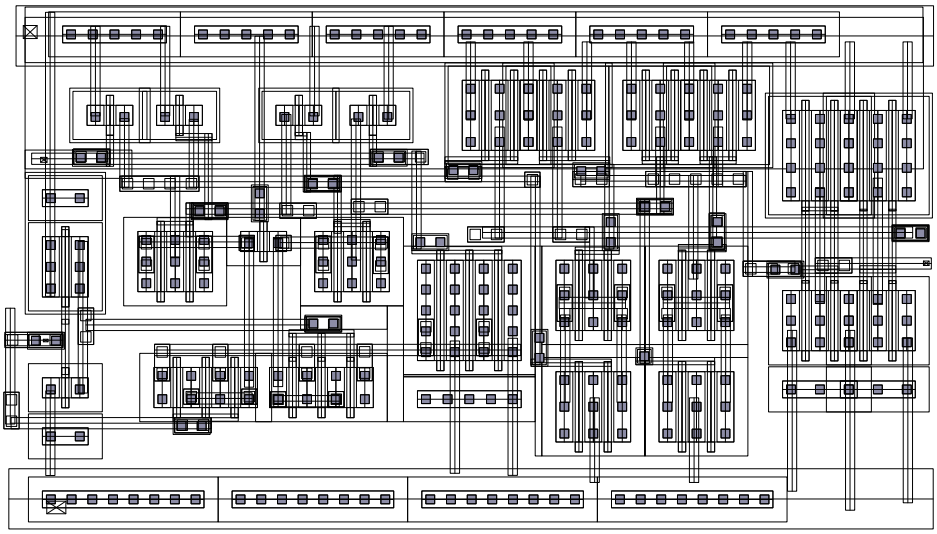


Figure A.3: Layout Diagram of the SAFF

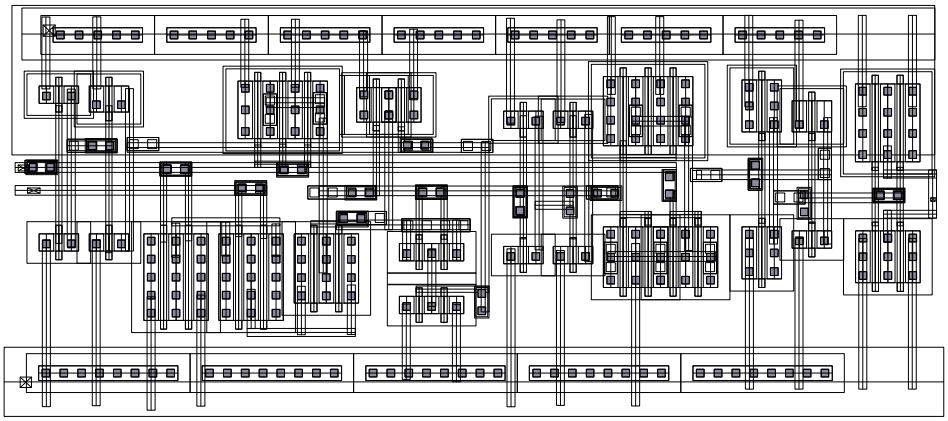


Figure A.4: Layout Diagram of the SDFF

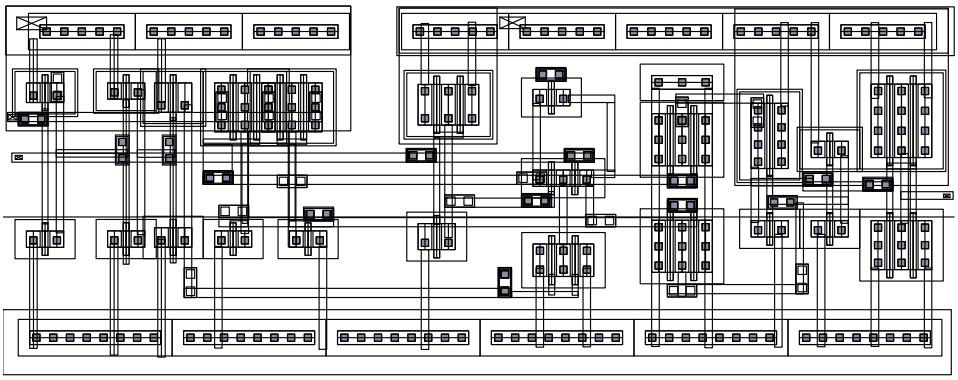


Figure A.5: Layout Diagram of the RCSPDFF

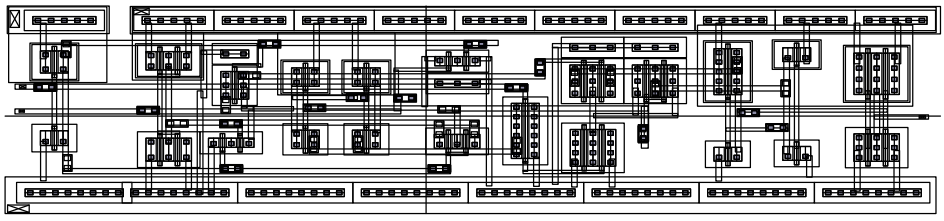


Figure A.6: Layout Diagram of the RCSSATG

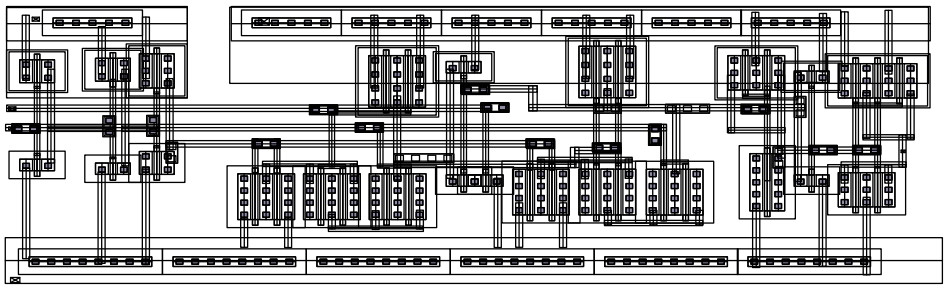


Figure A.7: Layout Diagram of the NDKFF

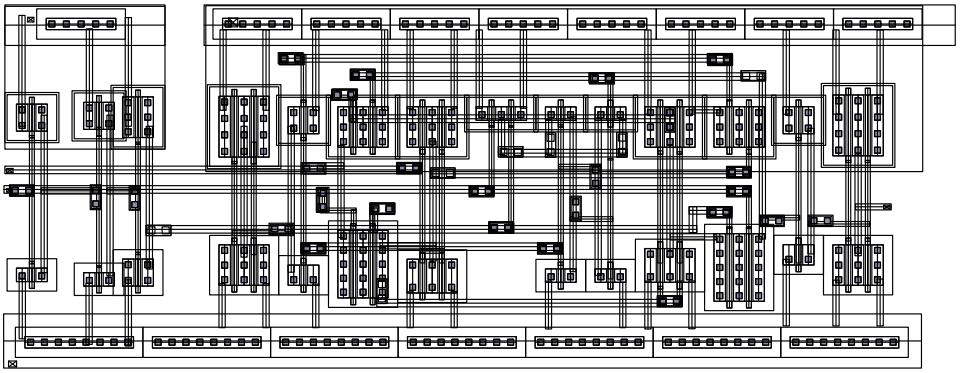


Figure A.8: Layout Diagram of the CRFF

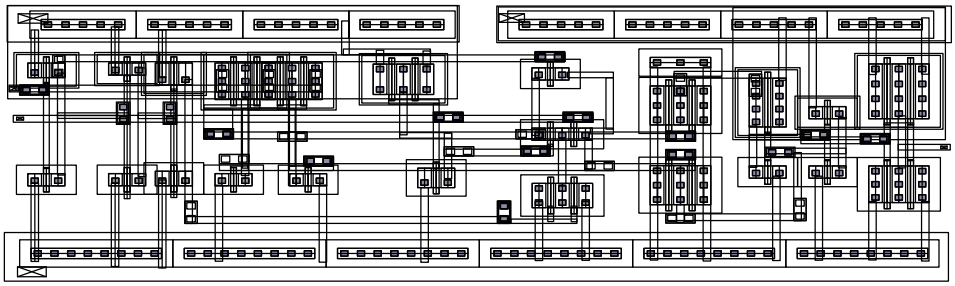


Figure A.9: Layout Diagram of the LCPDFF

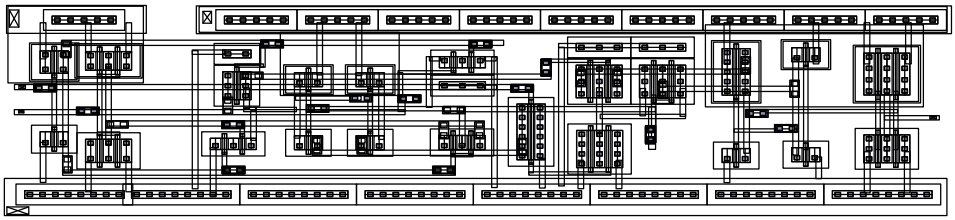


Figure A.10: Layout Diagram of the LCSATG

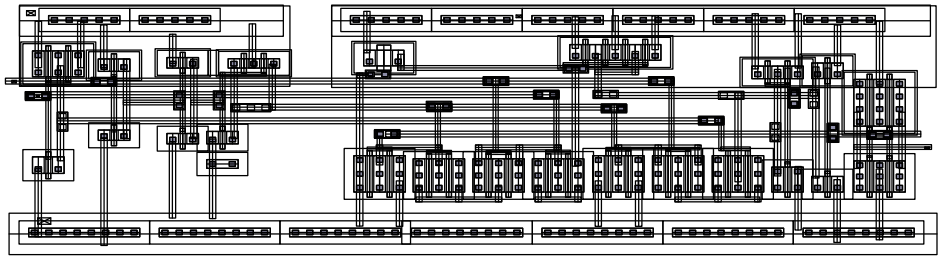


Figure A.11: Layout Diagram of the CPN

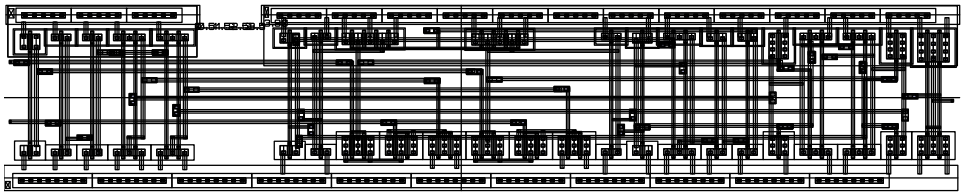


Figure A.12: Layout Diagram of the SPFF

References

[1] S.I.A, "2010 Executive Summary," *http://public.itrs.org*, 2010. xii, 6

[2] I. Clark, "Metastability Bibliography." `http://iangclark.net/metastability.html`, 2008. [Online; accessed 8-April-2011]. xiv, 7

[3] C. Dike and E. Burton, "Miller and Noise Effects in a Synchronizing Flip-Flop," *Journal of Solid-State Circuits*, vol. 34, pp. 849–855, June 1999. xix, 7, 29, 32, 149

[4] C. Foley, "Characterizing Metastability ," International Symposium on Advanced Research in Asynchronous Circuits and Systems, pp. 175–184, March 1996. 2

[5] P. Freidin, "FPGA-FAQ 0017: Tell Me About Metastability." `http://www.fpga-faq.com/FAQ_Pages/0017_Tell_me_about_metastables.htm`, 2000. [Online; accessed 8-April-2011]. 2

[6] D. Tala, "What is Metastability." `http://resalpes.grenoble.cnrs.fr/tutorat/vhdl_altera/divers/metastablity.html`, 2005. [Online; accessed 8-April-2011]. 2

[7] G. Moore, "Cramming more Components into Integrated Circuits," *Electronics*, vol. 38, pp. 82–85, April 1965. 2

[8] M. Tokumasu, H. Fujii, M. Ohta, T. Fuse, and A. Kameyama, "A New Reduced Clock-Swing Flip-Flop: NAND-Type Keeper Flip-Flop (NDKFF)," IEEE Custom Integrated Circuits Conference, pp. 129–132, May 2002. 4, 47

[9] D. Levacq *et al.*, "Half V_{DD} Clock-Swing Flip-Flop with Reduced Contention for up to 60% Power Saving in Clock Distribution," European Solid State Circuits Conference, pp. 190–193, September 2007. 4, 49

[10] M. Igarashi *et al.*, "A Low-Power Design Method Using Multiple Supply Voltages," International Symposium on Low Power Electronics and Design, pp. 36–41, March 1997. 4, 49

[11] T. Kuroda *et al.*, "Variable Supply-Voltage Scheme for Low-Power High-Speed CMOS Digital Design," *Journal of Solid-State Circuits*, vol. 33, pp. 454–462, March 1998. 4, 49

[12] T.Kuroda and M. Hamada, "Low-Power CMOS Digital Design with Dual-Embedded Adaptive Power Suppies," *Journal of Solid-State Circuits*, vol. 35, pp. 652–655, April 2000. 4, 49

[13] R. H. Dennard *et al.*, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions,," *Journal of Solid-State Circuits*, vol. 9, pp. 256–268, October 1974. 4

[14] T. Sakurai, "Optimiazation of CMOS Arbiter and Synchronize Circuits with Submicrometer MOSFETs," *Journal of Solid-State Circuits*, vol. 23, pp. 901–906, August 1988. 6, 93

[15] J. U. Horstmann, H. W. Eichel, and R. L. Coates, "Metastability Behavior of CMOS ASIC Flip-Flops in Theory and Test," *Journal of Solid-State Circuits*, vol. 24, pp. 145–157, February 1989. 6

[16] T. Kacprzak and A. Albicki, "Analysis of Metastable Operaiton in RS CMOS Flip-Flops," *Journal of Solid-State Circuits*, vol. sc-22, pp. 57–64, February 1987. 6, 93

[17] S. Flannagan, "Synchronization Reliability in CMOS Technology," *Journal of Solid-State Circuits*, vol. sc-20, pp. 880–882, August 1985. 6, 20

[18] L. Kim and R. Dutton, "Metastability of CMOS Latch/Flip-Flop," *Journal of Solid-State Circuits*, vol. 25, pp. 942–951, August 1990. 6, 26, 99

[19] D. Kinniment *et al.*, "Measuring Deep Metastability and Its Effect on Synchronizer Performance," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 15, pp. 1028–1039, September 2007. 7

[20] J. Zhou *et al.*, "On-Chip Measurement of Deep Metastability in Synchronizers," *Journal of Solid-State Circuits*, vol. 43, pp. 550–557, February 2008. 7, 149

[21] J. Zhou, D. Kinniment, G. Russell, and A. Yakovlev, "A Robust Synchronizer," IEEE Symposium on Emerging VLSI Technologies and Architectures, pp. 442–443, March 2006. 7, 29

[22] J. Zhou, M. Ashouei, D. Kinniment, J. Huisken, and G. Russell, "Extending Synchronization from Super-Threshold to Sub-threshold Region," IEEE Symposium on Asynchronous Circuits and Systems, pp. 85–93, May 2010. 7, 82

[23] U. Ko and P. Balsara, "High-Performance Energy-Efficient D-Flip-Flop Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 8, pp. 94–98, February 2000. 7, 31, 39

[24] O. S. Unsal *et al.*, "Impact of Parameter Variations on Circuits and Microarchitecture," *Micro*, vol. 26, no. 6, pp. 30–39, 2006. 7

[25] S. Dighe *et al.*, "Within-Die Variation-Aware Dynamic-Voltage-Frequency-Scaling With Optimal Core Allocation and Thread Hopping for the 80-Core TeraFLOPS Processor," *Journal of Solid-State Circuits*, vol. 46, pp. 184–193, January 2011. 7

[26] K. Bowman *et al.*, "Energy-Efficient and Metastability-Immune Resilient Circuits for Dynamic Variation Tolerance," *Journal of Solid-State Circuits*, vol. 44, pp. 49–63, January 2009. 7, 19

[27] J. Tschanz *et al.*, "Resilient Design in Scaled CMOS for Energy Efficiency," 15th Asia and South Pacific Design Automation Conference, p. 625, January 2010. 7

[28] K. Bowman *et al.*, "Circuit Techniques for Dynamic Variation Tolerance," ACM/IEEE Design Automation Conference, pp. 4–7, July 2009. 7

[29] J. Tschanz *et al.*, "A 45nm Resilient and Adaptive Microprocessor Core for Dynamic Variation Tolerance," IEEE International Solid-State Circuits Conference Digest of Technical Papers, pp. 282–283, February 2010. 7

[30] S. Unger and C. Tan, "Clocking Schemes for High-Speed Digital Systems," *IEEE Transction on Computer*, vol. C-35, pp. 880–895, October 1986. 10, 13

[31] M. Baghini and M. Desai, "Impact of Technology Scaling on Metastability Performance of CMOS Synchronizing Latches," 7th Asia and South Pacific Design Automation Conference and 15th International Conference on VLSI Design, pp. 317–322, January 2002. 11

[32] V. Stojanovics and V. G. Oklobdzija, "Comparative Analysis of Master-Slave Latches and Flip-Flops for High-Performance and Low-Power Systems," *Journal of Solid-State Circuits*, vol. 34, pp. 536–548, April 1999. 12, 39, 57, 62, 63

[33] C. Portmann and T. Meng, "Metastability in CMOS Library Elements in Reduced Supply and Technology Scaled Applications," *Journal of Solid-State Circuits*, vol. 30, pp. 39–46, January 1995. 20, 31

[34] F. Rosenberger and T. Chaney, "Flip-Flop Resolving Time Test Circuit," *Journal of Solid-State Circuits*, vol. sc-17, pp. 731–738, August 1982. 24

[35] R. Dutton, "Reply to "Comments on 'Metastability of CMOS Latch/Flip-Flop'"," *Journal of Solid-State Circuits*, vol. 27, pp. 131–132, January 1992. 26

[36] F. Rosenberger and C. Molnar, "Comments on "Metastability of CMOS Latch/Flip-Flop"," *Journal of Solid-State Circuits*, vol. 27, pp. 128–130, January 1992. 26

[37] D. Ernst *et al.*, "Razor: A Low Power Pipeline based on Circuit Level Timing Speculation," International Symposium on Microarchitecture, pp. 7–18, December 2003. 30

[38] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective, 3rd Edition*. Addison-Wesley, 2005. 31

[39] R. Cobbold, *Theory and Application of Field Transistors*. New York, NY: Wiley Interscience, 1970. 33

[40] L. Vadasz and A. Grove, "Temperature Dependence of MOS Transistor Characteristics Below Saturation," *IEEE Transction on Electron Devices*, vol. ED-13, pp. 863–866, December 1966. 33

[41] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor, 3rd Edition*. New York, NY: Oxford University Press, 2011. 33, 129

[42] E. Gutierrez, J. Deen, and C. Claeys, *Low Temperature Electronics: Physics, Devices, Circuits, and Applications*. New York, NY: Academic Press, 2001. 33

[43] G. Gerosa *et al.*, "A 2.2W, 80MHz Superscalar RISC Mircoprocessor," *Journal of Solid-State Circuits*, vol. 29, pp. 1440–1452, December 1994. 39

[44] J. Yuan and C.Svensson, "High-Speed CMOS Circuit Technique," *Journal of Solid-State Circuits*, vol. 24, pp. 62–70, February 1989. 39

[45] H. Partovi, "Flow-Through Latch and Edge-Triggered Flip-Flop Hybrid Elements," IEEE International Solid-State Circuits Conference Digest of Technical Papers, pp. 138–139, February 1996. 40

[46] F. Klass, "Semi-Dynamic and Dynamic Flip-Flops with Embedded Logic," Symposium on VLSI Circuits, Digest of Technical Papers, pp. 108–109, June 1998. 40

[47] M. Matsui *et al.*, "A 200MHz 13mm^2 2-D DCT Macrocell using Sense Amplifying Pipeline Flip-Flop Scheme," *Journal of Solid-State Circuits*, vol. 29, pp. 1482–1490, December 1994. 42

[48] J. Montanaro *et al.*, "A 160-MHz, 32-B, 0.5-W CMOS RISC Microprocessor," *Journal of Solid-State Circuits*, vol. 31, pp. 1703–1714, November 1996. 43

[49] B. Nikolic, V. G. Oklobdzija, V. Stojanovic, W. Jia, J.-S. Chiu, and M.-T. Leung, "Improved Sense-Amplifier-Based Flip-Flop: Design and Measurements ," *Journal of Solid-State Circuits*, vol. 35, pp. 876–884, June 2000. 43, 51

[50] J. Yuan and C.Svensson, "New single-clock CMOS Latches and Flip-Flops with Improved Speed and Power Savings," *Journal of Solid-State Circuits*, vol. 32, pp. 62–69, January 1997. 43

[51] B.-S. Kong, S.-S. Kim, and Y.-H. Jun, "Conditional-Capture Flip-Flop for Statistical Power Reduction," *Journal of Solid-State Circuits*, vol. 36, pp. 1263–1271, August 2001. 44

[52] D. Markovic, J. Tschanz, and V. De, "Feasibility Study of Low-Swing Clocking," International Conference on Microelectronics, pp. 547–550, May 2004. 45

[53] D. Duarte, V. Narayanan, and M. Irwin, "Impact of Technology Scaling in the Clock System Power," International Symposium on Low Power Electronics and Design, pp. 52–57, April 2002. 46

[54] H. Kawaguchi and T. Sakurai, "A Reduced Clock-Swing Flip-Flop (RCSFF) for 63% Power Reduction," *Journal of Solid-State Circuits*, vol. 33, pp. 807–811, May 1998. 46, 47

[55] B. Chatterjee, M. Sachdev, and R. Krishnamurthy, "A CPL Based Dual Supply 32-bit ALU Design for Sub-180nm CMOS Technologies," International Symposium on Low Power Electronics and Design, pp. 248–251, August 2004. 47

[56] F. Ishihara, F. Sheikh, and B. Nikolic, "Level Conversion for Dual-Supply Systems," *IEEE Transaction on Very Large Scale Integration (VLSI) Systems*, vol. 12, pp. 185–195, February 2004. 50

[57] M. Hamada *et al.*, "A Top-Down Low Power Design Technique Using Clustered Voltage Scaling with Variable Supply-Voltage Scheme," IEEE Custom Integrated Cirucits Conference, pp. 495–498, May 1998. 50

[58] H. Mahmoodi-Meimand and K. Roy, "Self-Precharging Flip-Flop (SPFF): A New Level Converting Flip-Flop," 28th European Solid-State Circuits Conference, pp. 407–410, September 2002. 51

[59] P. Zhao *et al.*, "Low-Power Clocked-Pseudo-NMOS Flip-Flop for Level Conversion in Dual Supply Systems," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, pp. 1196–1202, September 2009. 51

[60] P. Zhao, T. Darwish, and M. Bayoumi, "High-Performance and Low-Power Conditional Discharge Flip-Flop," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 12, pp. 477–484, May 2004. 51

[61] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design, 2nd Edition*. Reading, MA: Addson-Wesley, 1993. 54, 72

[62] J. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective, 2nd Edition*. Upper Saddle River, NJ: Prentice-Hall, Inc, 2003. 61, 81, 107

[63] C. Pina, "Evolution of the MOSIS VLSI Educational Program," Electronic Design, Test, and Application Workshop, pp. 187–191, 2002. 103

[64] PTM, "Predictive Technology Model." `http://ptm.asu.edu`, 2008. [Online; accessed 3-Oct-2010]. 103, 135

[65] B. Calhoun and A. Chandrakasan, "Characterizing and Modeling Minimum Energy Operation for Subthreshold Circuits," IEEE International Symposium on Low-Power Electronic Designs, pp. 90–95, August 2004. 129

[66] A. Wang and A.Chandrakasan, "A 180-mV Subthreshold FFT Processor Using a Minimum Energy Design Methodology," *Journal of Solid-State Circuits*, vol. 40, pp. 310–319, January 2005. 129

[67] B. Calhoun, A. Wang, and A.Chandrakasan, "Modeling and Sizing for Minimum Energy Operation in Subthreshold Circuits," *Journal of Solid-State Circuits*, vol. 40, pp. 1778–1786, September 2005. 129

[68] B. Fu and P. Ampadu, "Comparative Analysis of Ultra-Low Voltage Flip-Flops for Energy Efficiency," IEEE International Symposium on Circuits and Systems, pp. 1173–1176, May 2005. 129

[69] N. Lotze, M. Ortmanns, and Y. Manoli, "Variability of Flip-Flop Timing at Sub-Threshold Voltages," IEEE International Symposium on Low-Power Electronic Designs, pp. 221–224, August 2008. 129

[70] H. Mostafa, M. Anis, and M. Elmasry, "Comparative Analysis of Power Yield improvement Under Process Variation of Sub-Threshold Flip-Flops," IEEE International Symposium on Circuits and Systems, pp. 1739–1742, May 2010. 129

[71] T. Maeda *et al.*, "Device Characterizations and Physical Models of Strained-Si Channel CMOS," The International Conference on Microelectronic Test Structures, pp. 133–138, March 2004. 136

[72] S. Borkar, "Design Challenges of Technology Scaling," *Micro*, vol. 19, pp. 23–29, July-August 1999. 136

[73] S. Yang and M. Greenstreet., "Computing Synchronizer Failure Probabilities," Design, Automation and Test in Europe Conference, pp. 1–6, April 2007. 137

[74] Y. Ye, F. Liu, S. Nassif, and Y. Cao, "Statistical Modeling and Simulation of Threshold Variation under Dopant Fluctuations and Line-Edge Roughness ," Design, Automation and Test in Europe Conference, pp. 900–905, June 2008. 137

[75] M. Maymandi-Nejad and M. Sachdev, "A Monotonic Digitally Controlled Delay Element," *Journal of Solid-State Circuits*, vol. 40, pp. 2212–2219, November 2005. 143

[76] M. Maymandi-Nejad and M. Sachdev, "A Digitally Programmable Delay Element: Design and Analysis," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 11, pp. 871–878, October 2003. 143

[77] J. Jex and C. Dike, "A Fast Resolving BiNMOS Synchronizer for Parallel Processor Interconnect," *Journal of Solid-State Circuits*, vol. 30, pp. 133–139, February 1995. 149

[78] D. Kinniment, K. Heron, and G. Russell, "Meausring Deep Metastability," IEEE International Symposium on Asynchronous Circuits and Systems, pp. 10–11, March 2006. 149

[79] Y. Semiat and R. Ginosar, "Timing Measurements of Synchronization Circuits," IEEE International Symposium on Asynchronous Circuits and Systems, pp. 68–77, May 2003. 149

[80] J. Kalisz and Z. Jachna, "Metastability Tests of Flip-Flops in Programmable Digital Circuits," *Microelectronics Journal*, vol. 37, pp. 174–180, February 2006. 149

[81] Y. Dhillon, A. Diril, A. Chatterjee, and A. Singh, "Analysis and Optimization of Nanometer CMOS Circuits for Soft-Error Tolerance," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 14, pp. 514–524, May 2006. 154

[82] S. Mitra, N. Seifert, M. Zhang, Q. Shi, and K. Kim, "Robust System Deisgn with Built-In Soft-Error Resilience," *Computer*, vol. 38, pp. 43–52, February 2005. 154

[83] R. Naseer and J. Draper, "DF-DICE: a Scalable Solution for Soft Error Tolerant Circuit Design," International Symposium on Circuits and Systems, pp. 3890–3893, May 2006. 154, 156

[84] W. Wang and H. Gong, "Edge Triggered Pulse Latch Design with Delayed Latching Edge for Radiation Hardened Application," *IEEE Transactions on Nuclear Science*, vol. 51, pp. 3626–3630, December 2004. 154

[85] P. Hazucha *et al.*, "Measurements and Analysis of SER-Tolerant Latch in a 90nm Dual-V_T CMOS Process," *Journal of Solid-State Circuits*, vol. 39, pp. 1536–1543, September 2004. 154, 162

[86] D. Krueger, E. Francom, and J. Langsdorf, "Circuit Design for Voltage Scaling and SER Immunity on a Quad-Core Itanium Processor," IEEE International Solid-State Circuits Conference Digest of Technical Papers, pp. 94–95, February 2008. 155

[87] T. Calin, M. Nicolaidis, and R. Velazco, "Upset Hardened Memory Design for Submicron CMOS Technology," *IEEE Transactions on Nuclear Science*, vol. 43, pp. 2874–2878, December 1996. 155

[88] S. Jahinuzzaman, D. Rennie, and M. Sachdev, "A Soft Error Tolerant 10T SRAM Bit-Cell with Differential Read Capability," *IEEE Transactions on Nuclear Science*, vol. 56, pp. 3768–3773, December 2009. 155

[89] S. Jahinuzzaman, D. Rennie, and M. Sachdev, "Soft Error Robust Impulse and TSPC Flip-Flops in 90nm CMOS," 2nd Microsystems and Nanoelectronics Research Conference, pp. 45–48, October 2009. 155

[90] T. Karnik and P. Hazucha, "Characterization of Soft Errors Caused by Single Event Upsets in CMOS Processes," *IEEE Transaction on Dependable and Secure Computing*, vol. 1, pp. 128–143, April-June 2004. 160

[91] P. Marshall *et al.*, "Autonomous Bit Error Rate Testing at Multi-Gbit/s Rates Implemented in a 5AM SiGe Circuit for Radiation Effects Self Test (CREST)," *IEEE Transaction on Nuclear Science*, vol. 52, pp. 2446–2454, December 2005. 160

[92] C. Tokunaga, D. Blaauw, and T. Mudge, "True Random Number Generator With a Metastability-Based Quality Control ," *Journal of Solid-State Circuits*, vol. 43, pp. 78–85, January 2008. 169

[93] J. Lee, K. Kundert, and B. Razavi, "Analysis and Modeling of Bang-Bang Clock and Data Recovery Circuits," *Journal of Solid-State Circuits*, vol. 39, pp. 1571–1580, September 2004. 177

[94] K. Bowman *et al.*, "Dynamic Variation Monitor for Measuring the Impact of Voltage Droops on Microprocessor Clock Frequency," IEEE Custom Integrated Circuits Conference, pp. 1–4, September 2010.